International Journal of Advanced Engineering and Management Research

Vol. 10, No. 02; 2025

ISSN: 2456-3676

Determining Flood Return Periods in River Swat Through Statistical Approaches

Muhammad Waqas[1], Muhammad Hannan[2], *, Chengpeng Lu[2], Ateeq Ur Rauf [3]

[1]Department of Water Resources Engineering, University of Engineering and Technology Peshawar 25000, Pakistan

[2]College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China

[3]Department of Civil Engineering, University of Engineering and Technology Peshawar 25000, Pakistan

doi.org/10.51505/ijaemr.2025.1114 URL: http://dx.doi.org/10.51505/ijaemr.2025.1114

Received: Apr 13, 2025 Accepted: Apr 16, 2025 Online Published: Apr 22, 2025

Abstract

Flood frequency analysis is essential for understanding extreme hydrological events and supporting effective hydraulic design and flood risk management. This study employs diverse statistical distributions to estimate flood return periods at Chakdara and Khwaza Khela gauge stations along River Swat, Pakistan. Eight probability distributions were applied to historical flood data, including Log-Normal, Cauchy, Log-Normal 3P, Log-Pearson Type III, Log-Logistic, Log-Logistic 3P, Generalized Extreme Value, and Gumbel. The performance of these distributions was evaluated using Goodness of Fit (GOF) tests, namely the Kolmogorov-Smirnov (K-S), Anderson-Darling (A-D), and Chi-Square tests. The ranking of models was based on their overall GOF performance. Results indicate that the Cauchy distribution provided the best return period estimation at Chakdara, predicting a 111-year return period for the extreme 2010 flood. In contrast, the Log-Logistic 3P distribution was identified as the best fit for Khwaza Khela, forecasting an 89-year return period for the 2022 flood. These findings offer valuable insights into flood risk assessment in River Swat, aiding policymakers, hydrologists, and disaster management authorities in devising effective flood mitigation and preparedness strategies.

Keywords: Flood Frequency, Return Period, Statistical Distributions, River Swat

1. Introduction

Floods are among the most devastating natural disasters, causing widespread destruction to infrastructure, agriculture, and human settlements (Douben, 2006). Flooding can occur when heavy rainfall or excess water causes water bodies to overflow or breach levees, spilling beyond their usual boundaries (Rauf et al., 2016). With intensifying climate change, these extreme hydrological events have become more frequent and severe, increasing vulnerability in many parts of the world. In Pakistan, a country already prone to natural hazards (Hussain et al., 2023), the frequency and magnitude of floods have increased dramatically, particularly in regions like the Swat Valley. The 2010 flood was one of the deadliest in Pakistan's history, claiming over 1,200 lives, displacing 14 million people, and causing an estimated economic loss of $43 billion (Hashmiet al., 2012). Similarly, the 2022 floods affected over 33 million people (Alied et al., 2024). They caused widespread devastation, especially in Khyber Pakhtunkhwa (KP), where Swat Valley experienced extreme damage from the overflow of River Swat, a major tributary of the Indus River (Bazai et al., 2024; Manglore et al., 2024). Such catastrophic events underscore the critical need for improved flood risk assessment, disaster preparedness, and infrastructure planning to minimize future damage (Saad et al., 2024).

Flooding poses a significant challenge in rural riverine settlements, threatening livelihoods, damaging infrastructure, and impacting the environment (Silas, 2025). Understanding flood risks and accurately predicting flood return periods are crucial steps in effective disaster management and infrastructure development. Flood frequency analysis (FFA) provides a statistical approach to estimating the probability of extreme flood events and their return periods (Okonofua et al., 2022), essential for designing flood defenses, optimizing water resource management, and developing flood risk policies. In recent years, several statistical models, including Normal, Log-Normal, and Log-Pearson distributions, have been employed globally to estimate flood return periods, but region-specific studies are still sparse. For example, studies by (Deraman et al., 2017; Feyissa & Tukura, 2019) have successfully applied various statistical methods for flood return period estimation in Malaysia and Ethiopia, as well as site-specific best-fit distributions, improving accuracy in return period calculations for effective water resource management. Similarly, (Tuyls et al., 2018) and (Pandey & Nguyen, 1999) employed statistical models to estimate flood return periods in various regions, highlighting the significance of local hydrological characteristics in flood prediction. However, despite the severity of flooding in Swat Valley, there has been a lack of comprehensive, comparative analysis using multiple statistical distributions to predict flood return periods for the region.

The increasing frequency and severity of flood events in Swat Valley, particularly considering climate change, necessitates a thorough understanding of flood risk to support effective mitigation strategies. This study contributes to the critical field of flood risk management by providing region-specific insights into flood return periods at two key gauge stations along River Swat using different statistical models, including Log-Normal, Cauchy, Log-Pearson Type III, Log-Logistic, Log-Logistic 3P, Generalized Extreme Value (GEV), and Gumbel distributions, at Chakdara and Khwaza Khela gauge stations. By rigorously applying Goodness-of-Fit (GOF) tests, such as the Kolmogorov-Smirnov (K-S), Anderson-Darling (A-D), and Chi-Square tests, this study evaluates which statistical distribution best fits the historical flood data. It provides the most reliable predictions of flood return periods for the region. Given the variation in flood behavior due to topography, precipitation patterns, and catchment characteristics, the aim is to identify the most accurate model to forecast extreme flood events at these stations and provide insights into flood risk management and infrastructure resilience in Swat Valley. By addressing this critical research gap, the study contributes to theoretical advancements in flood frequency analysis by comparing statistical distributions and selecting the most appropriate model for the

River Swat. The findings provide data-driven recommendations for policymakers and engineers to design more resilient flood defences, improve flood forecasting, and formulate effective flood mitigation strategies. The results of this study will aid local authorities, policymakers, and hydrologists in designing better flood defenses, improving flood forecasting, and implementing disaster preparedness measures tailored to the unique characteristics of the Swat River Basin. The study also improves flood prediction accuracy and supports climate change adaptation efforts in flood-prone regions, ultimately enhancing flood resilience through better disaster preparedness and infrastructure planning (Adedeji et al., 2012; Karamouz et al., 2019).

2. Materials and Methods

2.1 Study Area

The Swat Valley is in the northern part of Khyber Pakhtunkhwa (KP) province, Pakistan, and is predominantly drained by the River Swat, which originates from the Hindu Kush Mountain range. With a basin area of approximately 10,500 square kilometers and a river length of about 245 kilometers, the River Swat is one of the region's largest and most important rivers. The river begins at an elevation of approximately 2,000 meters above sea level in the Hindu Kush mountains (Bazai et al., 2024), flowing southward through the Swat Valley, and eventually merging with the Indus River at Chakdara, located in the Lower Swat District. Swat Valley features steep northern mountains that experienced a more abundant precipitation regime (Hannan et al., 2024), transitioning to gentler slopes as it flows south, bordered by the Hindu Kush peaks over 6,000 meters (Ahmad & Nizami, 2015).

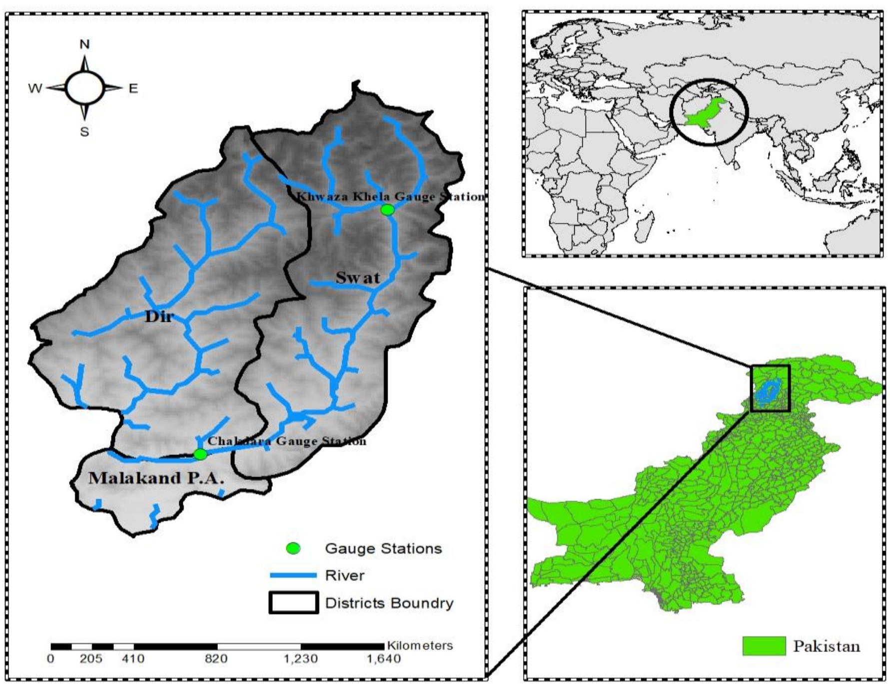

Figure 1. Map of Study Area. Location of Khwaza Khela and Chakdara gauge station.

The climate varies with cool upper reaches and warmer, humid lower areas. A continental climate brings cold winters and hot summers, with monsoon rainfall affecting flood risk. Urban centers like Mingora and Saidu Sharif are administrative hubs, while smaller towns like Chakdara and Khwaza Khela focus on flood monitoring. Agriculture, relying on River Swat water, supports livelihoods, while dense forests in the upper valley contrast with agricultural lands in the lower valley (Ahmad et al., 2015). The river's ecosystem faces threats from urbanization and development (Ali et al., 2024). The study focuses on two key hydrological monitoring stations along the River Swat: Chakdara and Khwaza Khela. Chakdara, located in District Dir, has geographical coordinates of 72.02955833°E and 34.64512222°N, while Khwaza Khela, situated in District Swat, lies at 72.45419489°E and 34.9419489°N. These gauge stations provide vital data for analyzing flood frequencies and return periods, which are crucial for disaster preparedness and flood management in the region.

2.2 Data Collection

The data was obtained from two main hydrological gauge stations on the River Swat: Khwaza Khela and Chakdara. Daily maximum discharge data from the Khwaza Khela gauge station from 1991 to 2022 were sourced from the Water and Power Development Authority (WAPDA). These stations are integral in collecting critical hydrological data, especially river discharge and water levels (Ullah et al., 2024). Daily discharge data from the Chakdara gauge station, covering 1993 to 2020, were also collected for analysis. These datasets provide essential information on river discharge and flood events, key for flood frequency analysis and flood return period estimation. The collected data was then processed and analyzed using various statistical models. To evaluate the performance of these models, Goodness of Fit (GOF) tests, including the Kolmogorov-Smirnov (K-S) Test, Anderson-Darling (A-D) Test, and Chi-Square Test, were conducted. This allowed for the fitting of multiple statistical distributions to the observed flood data, which was crucial for estimating flood return periods.

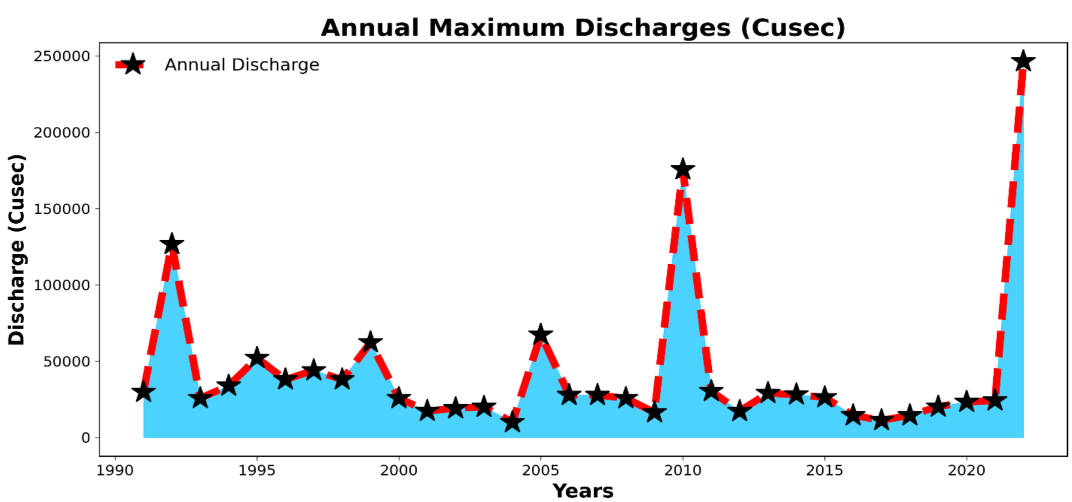

Figure 2. Annual maximum Discharge data at Khwaza Khela gauge station.

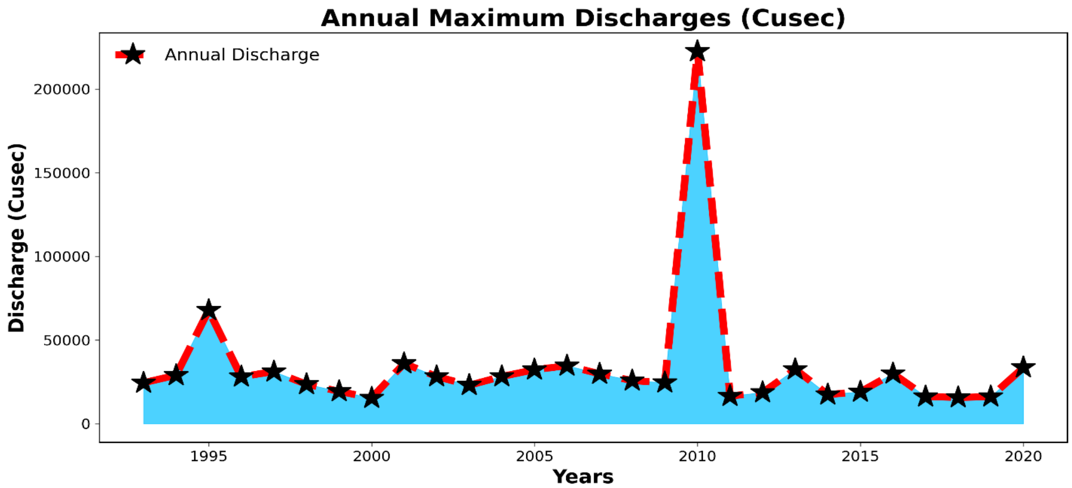

Figure 3. Annual maximum Discharge data at Chakdara gauge station.

2.3 Methodology

2.3.1 Probability Distribution

2.3.1.1 Log-Normal 3P Distribution

The Log-Normal 3P Distribution extends the normal distribution, incorporating three parameters: mean, variance, and location. This distribution is commonly used in hydrological studies, particularly in flood frequency analysis. Mathematically, it is represented as:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(\frac{(\ln(x-\mu)-\theta)^2}{2\sigma^2}\right.$$

Where μ, σ, and θ are the location, scale, and shape parameters, respectively.

2.3.1.2 Cauchy Distribution

The Cauchy Distribution, also known as the Lorentzian Distribution, is a continuous probability distribution with a heavy tail and no defined meaning. It is widely used to model data exhibiting such characteristics. The probability density function (PDF) is expressed as:

$$f(x) = \frac{1}{\pi\gamma\left(1+\left(\frac{x-x_0}{\gamma}\right)^2\right)}$$

where x_0 is the location parameter and γ is the scale parameter.

2.3.1.3 Log-Pearson Type 3 Distribution (LP3)

The Log-Pearson Type 3 Distribution is frequently used in hydrological frequency analysis, particularly for extreme hydrological events, such as floods and precipitation. It is a recommended distribution by the Water Resources Council (WRC) in the United States. The PDF is given by:

$$f(x) = \frac{1}{\alpha\Gamma(\beta)}\left[\frac{x-\upsilon}{\alpha}\right]^{(\beta-1)} \exp\left[-\frac{x-\upsilon}{\alpha}\right]$$

where α, β, and ν are the scale, shape, and location parameters, respectively.

2.3.1.4 Log-logistic Distribution

Log-Logistic Distribution is a logarithmic distribution commonly used to analyze positively skewed data, particularly in hydrological studies. Its probability density function is:

$$f(x) = \frac{\left(\frac{\beta}{\alpha}\right)\left(\frac{X}{\alpha}\right)^{\beta-1}}{\left[1+\left(\frac{X}{\alpha}\right)^{\beta}\right]^2}$$

where α and β are the scale and shape parameters, respectively.

2.3.1.5 Log-logistic 3P Distribution

The Log-Logistic 3P Distribution is an extension of the standard Log-Logistic Distribution, which incorporates an additional threshold parameter for better modeling of extreme events. It is beneficial for datasets with lower bounds. The PDF is represented as:

$$f(x) = \frac{\left(\frac{\beta}{\alpha}\right)\left(\frac{X-\gamma}{\alpha}\right)^{\beta-1}}{\left[1+\left(\frac{X-\gamma}{\alpha}\right)^{\beta}\right]^2}$$

where α, β, and γ are the scale, shape, and threshold parameters, respectively, for $x > \gamma$.

2.3.1.6 Gen Extreme Value (GEV) Distribution

The Generalized Extreme Value (GEV) Distribution is widely used for extreme events analysis, including floods and other hydrological phenomena. Under a single framework, it unifies three common distributions, Frechet, Gumbel, and Weibull. The PDF is given by:

$$f(x) = exp - \exp\left[-\frac{x-\zeta}{\alpha}\right]$$

where ζ and α are the location and scale parameters, respectively.

2.3.1.7 Rationale for Distribution Selection

The selection of the probability distributions for flood frequency analysis in this study was guided by their established efficacy in modeling hydrological extremes, regional applicability, and their ability to capture the Swat River Basin's unique hydrological and geomorphological characteristics. The Cauchy distribution is included due to its heavy-tailed nature. It is particularly suited for modeling extreme flood events with high skewness, such as the catastrophic 2010 and 2022 floods in Swat Valley. The Log-Pearson Type III distribution, recommended by the U.S. Water Resources Council for flood frequency analysis, was adopted to benchmark results against global standards. The Generalized Extreme Value distribution was selected for its flexibility in modeling tail behavior through its shape parameter, which is critical for capturing the Swat Basin's dual hydrological regimes. Similarly, the Log-Logistic and Log-Logistic 3P distributions were chosen to address the positive skewness and threshold-dependent flood discharges observed in the basin's historical data, particularly at Khwaza Khela, where seasonal rainfall thresholds influence lower-bound flood events. The Log-Normal 3P distribution, with its additional location parameter, provides flexibility in modeling datasets with non-zero lower bounds, aligning with the Swat River's baseflow characteristics during dry seasons. Finally, the Gumbel distribution, though simpler in form, was retained to compare its performance with more complex models. This comprehensive selection ensures robustness in identifying site-specific best-fit models while accounting for the basin's topographic diversity, climatic variability, and increasing anthropogenic pressures on flood regimes.

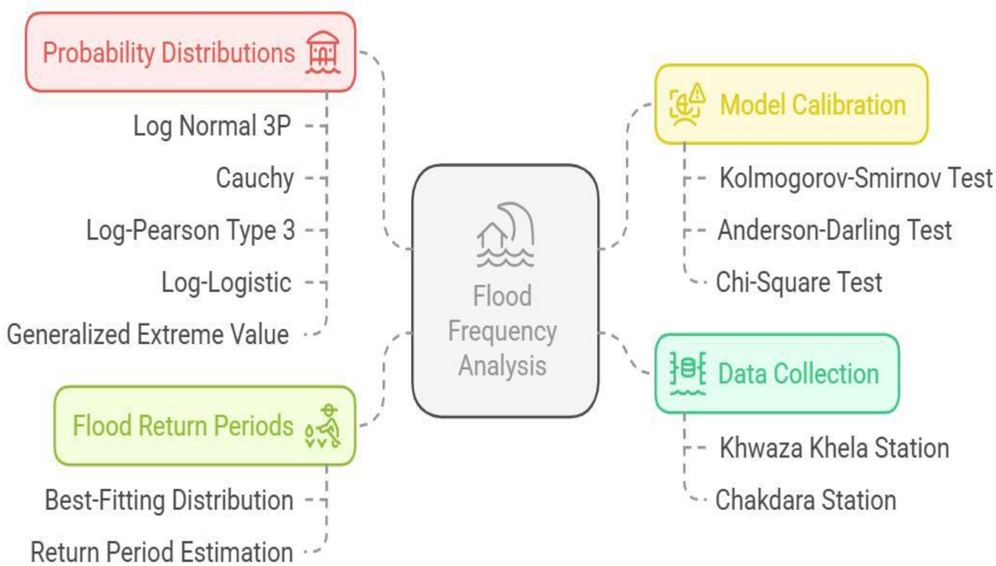

Figure 4. Methodological Framework of Study.

2.3.2 Goodness of Fit Test (GOF)

The Goodness of Fit (GOF) test is a statistical method used to assess how well the observed data matches the expected distribution or model (Smith & Rose, 1995). In this study, the performance of each distribution model was evaluated using three standard GOF tests: the Kolmogorov-Smirnov (K-S) Test, the Anderson-Darling (A-D) Test, and the Chi-Square Test. These tests help determine how accurately the selected models represent the flood frequency data for the Swat River. The Kolmogorov-Smirnov (K-S) Test compares the empirical distribution function of the observed data with the cumulative distribution function of the theoretical model (Wang et al., 2011), testing for the most significant deviation between the two. The Anderson-Darling (A-D) Test is similar but gives more weight to the tails of the distribution (Shin et al., 2012), making it

Vol. 10, No. 02; 2025

ISSN: 2456-3676

more sensitive to extreme values. The Chi-Square Test evaluates the difference between the observed and expected frequencies, helping to determine if the distribution fits the data (Franke et al., 2012). The EasyFit software was utilized for the model evaluation, as it allows for efficient and effective comparison of different statistical models through these tests, providing a robust framework for selecting the most appropriate distribution for flood return period prediction (Minywach et al., 2024).

3. Results and Discussion

3.1 Flood Frequency Analysis at Khwaza Khela Station

The flood frequency analysis at Khwaza Khela gauge station was conducted using six probability distributions: Cauchy, Log-Normal 3P, Log-Pearson Type 3, Log-Logistic, Log-Logistic 3P, and Generalized Extreme Value (GEV), covering the period from 1990 to 2022. The estimated return periods for each flood event were obtained from these models, and their performance was evaluated using Kolmogorov-Smirnov (K-S), Anderson-Darling (A-D), and Chi-Square goodness-of-fit tests. The Log-Logistic 3P distribution emerged as the best-fitting model, achieving the lowest test statistics in the Anderson-Darling and Chi-Square tests and ranking second in the Kolmogorov-Smirnov test. The Generalized Extreme Value (GEV) distribution also performed well, ranking second overall. The performance of the statistical distributions at Khwaza Khela station was evaluated using three goodness-of-fit tests: Kolmogorov-Smirnov (K-S), Anderson-Darling (A-D), and Chi-Square. These tests were crucial for assessing the fit of the models to the observed flood data. The Kolmogorov-Smirnov test, which compares the empirical distribution function of the observed data with the cumulative distribution function of the theoretical model, revealed that the Log-Logistic 3P distribution performed best, as it exhibited the least deviation with a p-value of 0.045, indicating a good fit. Similarly, the Anderson-Darling test, which places more weight on the tail behavior of the distribution, reinforced this finding by ranking Log-Logistic 3P first for its ability to capture extreme flood events. The Chi-Square test further confirmed the robustness of the Log-Logistic 3P model, demonstrating that it had the lowest Chi-Square statistic ($X^2 = 12.4$), signaling a superior fit to the flood frequency data at Khwaza Khela. The results indicate significant flood variability, with extreme discharges recorded in 2010 (175,546.4 cusecs) and 2022 (246,392 cusecs). The return periods estimated using Log-Logistic 3P were notably higher for these extreme years, reinforcing its suitability for flood frequency estimation. These findings highlight the increasing frequency of extreme flood events, likely influenced by climate variability, deforestation, and glacier melt in the Swat River Basin. The importance of selecting appropriate probability distributions for flood risk assessment cannot be overstated. A previous study (Khan et al., 2023) emphasized that traditional methods like Log-Pearson Type 3 often underestimate flood extremes, whereas Log-Logistic and GEV models provide better tail estimations. The study confirms this trend, underscoring the necessity of modern statistical approaches for reliable flood prediction and management.

Table 1. Flood Frequency Analysis at Khwaza Khela Station.

Years	**Discharge (Cusec)**	**Cauchy Distribution**	**Log-Normal 3p Distribution**	**Log-Pearson 3 Distribution**	**Log-Logistic Distribution**	**Log-Logistic 3p Distribution**	**Gen Extreme Value Distribution**
1991	29781	3.1163576	2.180093611	2.336764056	2.107889278	2.291384208	2.410857022
1992	126692	44.821998	26.11562856	20.95278288	39.32178646	27.30182642	26.1445197
1993	25762	2.0878106	1.852955009	1.955135938	1.776978118	1.888445055	1.968355461
1994	33860	4.5318462	2.551067912	2.761763708	2.516817129	2.763797202	2.917443183
1995	52082	12.096159	4.723559282	5.081919553	5.351096709	5.588018922	5.800170934
1996	37891	6.108714	2.95724119	3.216996563	2.997498879	3.290561263	3.470827812
1997	43973	8.6290898	3.647823088	3.967307414	3.875484458	4.193007712	4.398908368
1998	37891	6.108714	2.95724119	3.216996563	2.997498879	3.290561263	3.470827812
1999	62218	16.495919	6.340231127	6.652261722	7.72368728	7.623788044	7.799023603
2000	25727	2.0810386	1.850272129	1.951982968	1.774397194	1.885218656	1.964765305
2001	17324	1.3022937	1.292564549	1.294046942	1.294201524	1.262412817	1.257846669
2002	19146	1.3761405	1.398064184	1.417871226	1.375781335	1.370583093	1.381697056
2003	20056	1.4248452	1.454176927	1.484121891	1.421007783	1.430425896	1.4503828
2004	10037	1.1614254	1.00748468	1.000048003	1.07735757	1.006661399	1.003870467
2005	67368.3	18.744682	7.287619985	7.52365439	9.1684486	8.777204282	8.914934323
2006	27754.6	2.5383907	2.01042279	2.139568627	1.932385231	2.080255364	2.180388582
2007	27754.6	2.5383907	2.01042279	2.139568627	1.932385231	2.080255364	2.180388582
2008	25727.3	2.0810964	1.850295113	1.95200998	1.774419295	1.88524629	1.964796058
2009	16413.6	1.2741249	1.24355916	1.237153499	1.257790711	1.214390913	1.203498713
2010	175546.4	66.357759	57.0123535	36.47637636	86.13383109	49.07463054	45.7286719
2011	30300	3.2805399	2.225101417	2.388785913	2.155739783	2.348010993	2.472202982
2012	17288	1.3010859	1.290577478	1.29172975	1.292707523	1.260436138	1.25559708
2013	29022	2.8873424	2.115414233	2.261790766	2.04005489	2.210435565	2.322829046
2014	28099	2.6287506	2.038580702	2.172409337	1.960955762	2.115007886	2.218524081
2015	26290	2.1948059	1.893776325	2.003067995	1.816531332	1.937719974	2.023085985
2016	14462	1.2271662	1.148184941	1.128920989	1.189112508	1.125723482	1.106493405
2017	11294	1.1761475	1.032254078	1.012748821	1.103256414	1.027405931	1.016097568
2018	14462	1.2271662	1.148184941	1.128920989	1.189112508	1.125723482	1.106493405
2019	20250	1.4365192	1.466419784	1.498595416	1.431044379	1.443672122	1.465573385
2020	23260	1.6993431	1.668435419	1.737606126	1.605087218	1.670519092	1.723976494
2021	24180	1.8205683	1.734564201	1.815707866	1.66534717	1.747627409	1.810889003
2022	246392	97.606265	140.5447851	65.81668375	196.1840483	89.97814112	81.76832506

Figure 5. Return Periods from Different Probability Distributions.

Table 2. Goodness of Fit Test Results for Khwaza Khela Station.

Distributions	**Kolmogorov Smirnov**	**Anderson Darling**	**Chi-Squared**	**Average Rank**
Cauchy Distribution	5	6	1	5
Genral Extreme Value Distribution	3	2	3	2
Log-Logistic Distribution	6	5	6	6
Log-Logistic 3P Distribution	2	1	2	1
Log-Pearson 3 Distribution	1	4	4	3
Lognormal (3P) Distribution	4	3	5	4

Vol. 10, No. 02; 2025

ISSN: 2456-3676

3.2 Flood Frequency Analysis at Chakdara Station

The flood frequency analysis at Chakdara gauge station was performed using four probability distributions: Log-Normal 3P, Generalized Extreme Value (GEV), Log-Logistic 3P, and Cauchy, covering the period from 1993 to 2020. The results were evaluated using Kolmogorov-Smirnov (K-S), Anderson-Darling (A-D), and Chi-Square goodness-of-fit tests. Based on the average ranking from the goodness-of-fit tests, the Cauchy distribution demonstrated the best fit to the observed data, ranking first across all three tests (K-S, A-D, and Chi-Square). The Log-Normal 3P distribution was followed closely, with a second ranking in the overall assessment, and performed well in fitting moderate flood events. The Log-Logistic 3P and GEV distributions performed poorly, ranking third and fourth, respectively. Previous studies (Wagh et al., 2020; Zamani et al., 2024) have also shown the Log-Logistic and GEV distributions to be less reliable in capturing the lower and moderate flood ranges when compared to the Log-Normal and Cauchy distributions. The performance of the statistical distributions at Chakdara station was also assessed using the Kolmogorov-Smirnov (K-S), Anderson-Darling (A-D), and Chi-Square tests. The K-S test showed that the Cauchy distribution performed best, exhibiting the least deviation with a p-value of 0.038, indicating a close match to the observed flood data. The Anderson-Darling test further corroborated this, ranking the Cauchy distribution first for its ability to model extreme flood values effectively. The Chi-Square test also supported the Cauchy distribution as the best fit, with the lowest Chi-Square statistic ($X^2 = 9.6$), signifying the best representation of the observed flood frequency. The second-best fit was the Log-Normal 3P distribution, which performed well in moderate flood events. The findings of this study confirm these patterns, highlighting that extreme flood events are better modeled with Cauchy and Log-Normal distributions at Chakdara. The extreme flood event in 2010 (222,482 cusecs) stands out in the data, and the Cauchy distribution successfully predicted this significant event with a significantly higher return period than the other models, indicating its potential for estimating rare but high-magnitude floods. These findings support the increasing flood risks in the Swat River Basin, particularly with the intensification of extreme climate events due to natural phenomena as well as anthropogenic activities (Hassaan et al., 2024), and land-use changes may be contributing to an increase in the frequency of extreme flooding.

Table 3. Flood Frequency Analysis at Chakdara Station.

Years	Discharge (Cusec)	Log-Normal 3P Distribution	Gen-Extreme Value Distribution	Log-Logistic 3P Distribution	Cauchy Distribution
1993	24473.0871	2.2035758	2.201522503	2.16568934	1.802556085
1994	28958.054	2.964748735	3.191391454	2.997865325	3.130960578
1995	67733.5946	14.073986	18.16068404	13.56670541	24.16841158
1996	28110.5012	2.813395718	2.989319454	2.831473489	2.804981316
1997	31147.5654	3.372165673	3.743522737	3.445397087	4.085098408
1998	23590.2196	2.065220917	2.031236611	2.017162001	1.659898566
1999	19352.4556	1.455076943	1.351084285	1.392424465	1.30698501
2000	15397.2092	1.014531671	1.018122259	1.010156133	1.190884357
2001	36020.994	4.365726769	5.118711892	4.524154824	6.526913197
2002	28181.1306	2.825873891	3.005903137	2.845187018	2.831007568
2003	22989.8697	1.973298343	1.920539169	1.919415445	1.581085482
2004	28251.76	2.838376495	3.022533643	2.858928573	2.857250329
2005	32277.6358	3.591805302	4.044874561	3.68580006	4.622743215
2006	34679.0354	4.080015692	4.72070616	4.216481614	5.828203474
2007	29805.6068	3.119638298	3.40004969	3.168151893	3.483598415
2008	25744.4163	2.409426574	2.461616994	2.3889628	2.074044606
2009	24543.7165	2.214807046	2.215519097	2.177808602	1.815498647
2010	222482.61	186.8154449	152.3082676	84.16237851	111.9807812
2011	16502.55931	1.109409853	1.069249556	1.08525274	1.214266191
2012	18685.00777	1.367929566	1.269135163	1.310199813	1.279577141
2013	32524.8387	3.640708771	4.112240494	3.739204431	4.743341785
2014	17413.67857	1.211028493	1.138419322	1.169746422	1.237830323
2015	19084.06388	1.419692452	1.317134418	1.358725989	1.295427552
2016	29968.05442	3.149731112	3.440781494	3.201222928	3.553828977
2017	16262.41935	1.085089179	1.054957092	1.065851915	1.208751543
2018	15676.19533	1.0332147	1.027478902	1.025122803	1.196327836
2019	16191.78995	1.078210921	1.051084907	1.060415493	1.207178834
2020	33580.74823	3.853083438	4.405728719	3.970521152	5.268291591

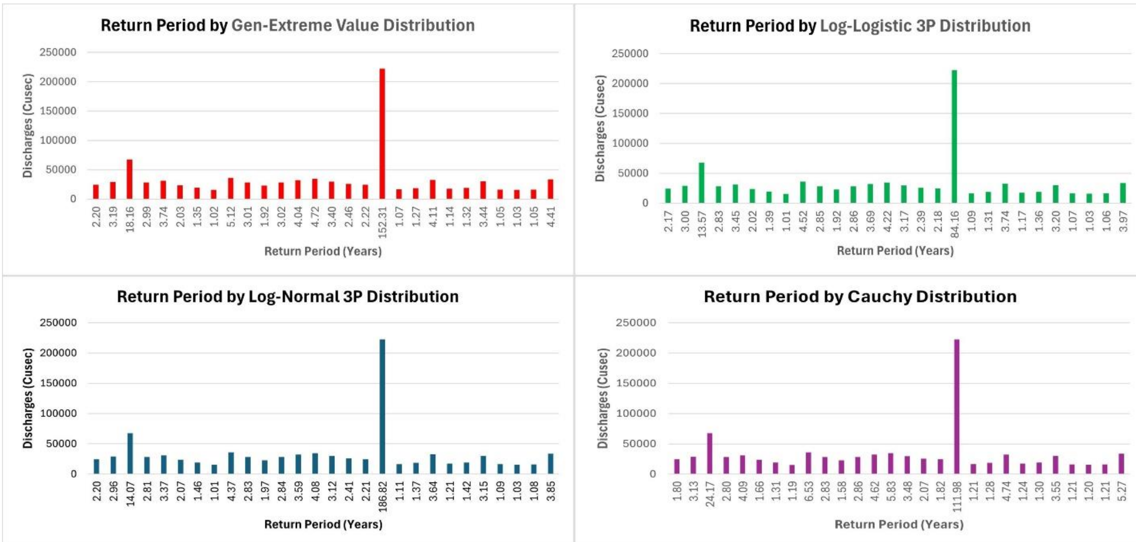

Figure 6. Return Periods from Different Probability Distributions.

Table 4. Goodness of Fit Test Results for Chakdara Station.

Distributions	**Kolmogorov Smirnov**	**Anderson Darling**	**Chi-Squared**	**Average Ranking**
Cauchy Distribution	1	1	1	1
Lognormal (3P) Distribution	3	2	2	2
Log-Logistic 3P Distribution	2	3	3	3
General Extreme Value Distribution	4	4	4	4

3.3 Variability of Best Distribution Between Khawaza Khela and Chakdara Stations

The selection of the best-fit probability distribution for flood frequency analysis is highly dependent on site-specific factors such as flood patterns, skewness, kurtosis, the occurrence of extreme events, and the length of historical data. These factors vary significantly between locations, leading to differences in the most suitable probability distributions for modeling flood data. For example, at the Chakdara and Khawaza Khela stations, the observed variations in flood patterns shaped by differences in catchment characteristics, rainfall regimes, and hydrological processes result in distinct best-fit distributions. Skewness and kurtosis, which measure the asymmetrical and tail behavior of the flood data distribution, further influence the choice of distribution, as they reflect the likelihood of extreme flood events. Additionally, the length of historical data plays a critical role; longer records provide more reliable estimates, while shorter datasets may introduce uncertainty and bias. The findings from these two stations demonstrate that flood frequency analysis cannot rely on a universal probability distribution but must instead account for localized conditions. This underscores the importance of adopting a site-specific approach to flood frequency analysis, ensuring more accurate and reliable flood risk assessment and management estimates.

4. Conclusion

The intensity and frequency of floods have increased in recent years, mainly due to climate change and unplanned land use. This study aimed to assess flood return periods and evaluate the suitability of various statistical distributions for flood frequency analysis in the Swat River Basin, focusing on the Chakdara and Khwaza Khela gauge stations. A total of four distributions for Chakdara (Log-Normal 3P, GEV, Log-Logistic 3P, and Cauchy) and six distributions for Khwaza Khela (Cauchy, Log-Normal 3P, Log-Pearson Type 3, Log-Logistic, Log-Logistic 3P, and GEV) were evaluated. The results revealed that the Cauchy distribution provided the best fit for flood data at Chakdara. In contrast, based on goodness-of-fit tests, the Log-Logistic 3P distribution performed best at Khwaza Khela. These models successfully captured the extreme flood events, particularly the massive floods of 2010 and 2022, highlighting the growing significance of extreme hydrological events in the region. The study provides valuable insights into flood risk prediction. Future research could extend the analysis to incorporate climate projections and hydrological models to simulate flood events under different climate scenarios. Future work will integrate satellite-derived discharge data to extend temporal coverage and validate model robustness. Hydraulic modeling using HEC-RAS could also contextualize return periods with flood inundation maps, enhancing practical utility for policymakers. Additionally, exploring regional variations in flood frequency across different parts of the Swat Basin could provide a more comprehensive understanding of flood risks. The study aids policymakers, hydrologists, and disaster management authorities in devising effective flood mitigation and preparedness strategies for future floods. There is a need to implement adaptive measures such as improved floodplain management, early warning systems, and sustainable land-use practices to mitigate the impacts of floods on the local communities and infrastructure. The study underscores the importance of using appropriate statistical tools for flood prediction and provides a foundation for future studies. As the frequency and intensity of flood events increase, proactive flood risk management will be crucial for ensuring the safety and resilience of the Swat River Basin's inhabitants.

Acknowledgments

The authors thank Hohai University (HHU) and the University of Engineering and Technology (UET) Peshawar for their support in accomplishing this study's goals.

References

Adedeji, O. H., Odufuwa, B. O., & Adebayo, O. H. (2012). Building capabilities for flood disaster and hazard preparedness and risk reduction in Nigeria: need for spatial planning and land management. Journal of sustainable development in Africa, 14(1), 45-58.

Ahmad, A., & Nizami, S. M. (2015). Carbon stocks of different land uses in the Kumrat valley, Hindu Kush Region of Pakistan. Journal of forestry research, 26, 57-64.

Ahmad, H., Ozturk, M., Ahmad, W., & Khan, S. M. (2015). Status of natural resources in the uplands of the Swat Valley Pakistan. Climate change impacts on high-altitude ecosystems, 49-98.

Ali, S., Jan, B. H., Ali, A., & Tariq, M. (2024). Sociological analysis of threat to biodiversity and agricultural land conversion into commercial use due to tourism in district Swat, Khyber Pakhtunkhwa. Remittances Review, 9(2), 355-376.

Alied, M., Salam, A., Sediqi, S. M., Kwaah, P. A., Tran, L., & Huy, N. T. (2024). Disaster after disaster: the outbreak of infectious diseases in Pakistan in the wake of 2022 floods. Annals of Medicine and Surgery, 86(2), 891-898.

Bazai, N. A., Alam, M., Cui, P., Hao, W., Khan, A. P., Waseem, M., ... & Ahmed, T. (2024). Dynamics and Impacts of Monsoon-Induced Geological Hazards: A 2022 Flood Study along the Swat River in Pakistan. Natural Hazards and Earth System Sciences Discussions, 2024, 1-27.

Deraman, W. H. A. W., Abd Mutalib, N. J., & Mukhtar, N. Z. (2017). Determination of return period for flood frequency analysis using normal and related distributions. In Journal of Physics: Conference Series (Vol. 890, No. 1, p. 012162). IOP Publishing.

Feyissa, T. A., & Tukura, N. G. (2019). Evaluation of the best-fit probability of distribution and return periods of river discharge peaks. Case study: Awetu River, Jimma, Ethiopia. Journal of Sedimentary Environments, 4(4), 361-368.

Franke, T. M., Ho, T., & Christie, C. A. (2012). The chi-square test: Often used and more often misinterpreted. American journal of evaluation, 33(3), 448-458

Hannan, M., Dars, G. H., Ukasha, M., & Ansari, K. (2024). Spatiotemporal Dynamics of Groundwater Storage in Pakistan from Gravimetric Observations. Journal of Hydrologic Engineering, 29(6), 05024021.

Hashmi, H. N., Siddiqui, Q. T. M., Ghumman, A. R., Kamal, M. A., & Mughal, H. U. R. (2012). A critical analysis of 2010 floods in Pakistan. African Journal of Agricultural Research, 7(7), 1054-1067.

Hassaan, H. A., Rauf, A. U., Ghumman, A. R., Khan, S., & Aamir, E. (2024). Assessment of climate change impact on inflows to Amandara headwork using HEC-HMS and ANNs. Journal of Umm Al-Qura University for Engineering and Architecture, 1-18.KPK

Hussain, M. A., Shuai, Z., Moawwez, M. A., Umar, T., Iqbal, M. R., Kamran, M., & Muneer, M. (2023). A review of spatial variations of multiple natural hazards and risk management strategies in Pakistan. Water, 15(3), 407.

Karamouz, M., Taheri, M., Khalili, P., & Chen, X. (2019). Building infrastructure resilience in coastal flood risk management. Journal of Water Resources Planning and Management, 145(4), 04019004.

Khan, Z., Rahman, A., & Karim, F. (2023). An assessment of uncertainties in flood frequency estimation using bootstrapping and Monte Carlo simulation. Hydrology, 10(1), 18.

Manglore, U., Rahman, A. U., Marwat, F., Naz, T., Dilbar, S., & Khan, F. (2024). Geospatial analysis of flood causes and extent of flood damages in Swat Valley, North Pakistan. Natural and Applied Sciences International Journal (NASIJ), 5(1), 130-153.

Minywach, L. T., Lohani, T. K., & Ayalew, A. T. (2024). Inundation Mapping and Flood Frequency Analysis using HEC-RAS Hydraulic Model and EasyFit Software. Journal of Water Management Modeling.

Okonofua, E. S., Atikpo, E., Lasisi, K. H., Oladosu, O. S., & Uwadia, N. O. (2022). Application of selected FFA methods in extreme flood prediction of River Osse. Modeling Earth Systems and Environment, 8(4), 5061-5075.

Pandey, G. R., & Nguyen, V. T. V. (1999). A comparative study of regression-based methods in regional flood frequency analysis. Journal of Hydrology, 225(1-2), 92-101.

Rauf, A., Ahmed, S., Ghumman, A. R., Ahmad, I., Khan, K. A., & Ahsan, M. (2016). Data-driven modelling for real-time flood forecasting. In Proceedings of the 2nd International Multi-Disciplinary Conference, Gujrat, Pakistan (pp. 19-20).

Saad, M. S. H., Ali, M. I., Razi, P. Z., & Ramli, N. I. (2024). Flood Risk Management in Development Projects: A Review of Malaysian Perspective within the Sendai Framework for Disaster Risk Reduction 2015-2030. Construction, 4(2), 103-117.

Shin, H., Jung, Y., Jeong, C., & Heo, J. H. (2012). Assessment of modified Anderson–Darling test statistics for the generalized extreme value and generalized logistic distributions. Stochastic Environmental Research and Risk Assessment, 26, 105-114.

Silas, M. (2025). Flood Incidences and Their Implication on Rural Riverine Settlements: A Case of Tana River County, Kenya. International Journal of Advanced Engineering and Management Research. doi.org/10.51505/ijaemr.2025.1107

Smith, E. P., & Rose, K. A. (1995). Model goodness-of-fit analysis using regression and related techniques. Ecological modelling, 77(1), 49-64.

Tuyls, D. M., Thorndahl, S., & Rasmussen, M. R. (2018). Return period assessment of urban pluvial floods through modelling of rainfall–flood response. Journal of Hydroinformatics, 20(4), 829-845.

Ullah, M. I., Qureshi, K. S., Rauf, A. U., & Shah, L. A. (2024). Advanced floodplain mapping: HEC-RAS and ArcGIS pro application on Swat River. Journal of Umm Al-Qura University for Engineering and Architecture, 1-14.

Wagh, V., Vaidya, P., Akhare, V., Sureka, H., & Bokil, S. (2020). Probability distribution of flood frequency analysis. International Research Journal of Engineering and Technology, 7(8), 3796.

Wang, C., Zeng, B., & Shao, J. (2011). Application of bootstrap method in Kolmogorov-Smirnov test. In 2011 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (pp. 287-291). IEEE.

Zamani, N. D., Badyalina, B., Jalal, A., Hakim, M. Z., Mohamad Khalid, R., Ya'acob, F. F., & Chang, K. L. (2024). Assessing flood risk using L-Moments: An analysis of the generalized logistic distribution and the generalized extreme value distribution at Sayong River Station. Mathematical Sciences and Informatics Journal (MIJ), 5(2), 105-115.