PJET, Vol. 3, No. 2, July 2025: 13-29

Poly Journal of Engineering and Technology

journal homepage: https://journals.bdu.edu.et/index.php/PJET/index

ORGINAL ARTICLE

Utilization of Catchment Morphometric Features to Estimate Design Peak Flow at Ungauged Catchments: Insights from Lake Tana Sub-Basin, Ethiopia

Walelign K. Endalew[1*], Mulugeta A. Belete[2]

[1] Department of Natural Resource Management, Debre Tabor University, Debre Tabor, Ethiopia.

[2] Faculty of Civil and Water Resource Engineering, Bahir Dar Institute of Technology, Bahir Dar University, Bahir Dar, Ethiopia.

ABSTRACT

Accurately estimating design peak flow is a fundamental and frequently encountered task in engineering hydrology. This estimation is crucial for planning civil infrastructure related to water storage, diversion, conveyance, and crossing of waterways. The aim of this study was to propose a new model for estimating design peak flow in ungauged catchments, using morphometric characteristics of the catchment as input variables. To achieve this, daily streamflow records from gauged catchments were analyzed, and the annual maximum daily flows were extracted. Flood frequency analysis at each site was performed to determine peak flow values corresponding to return periods of 10, 25, 50, 100, and 200 years. Drawing on prior research, key morphometric parameters namely catchment area, longest stream length, channel slope, and circularity ratio were identified as significant predictors of peak flow. These variables were then linked to the flow quantiles using regression analysis. The findings revealed that the Pearson Type III, General Extreme Value, Log-Normal, and General Pareto distributions provided the best statistical fit for the sub-basin data. The regression results demonstrated a strong correlation between design peak flows and the selected morphometric features, with coefficients of determination (R^2) ranging from 0.962 to 0.993. The resulting models offer a practical and simplified approach for estimating design peak flows in ungauged catchments within the Lake Tana sub-basin and other regions with similar hydrological characteristics.

Keywords: *Design Peak flow, Morphometric Features, Flood Frequency Analysis, Regression Analysis, Lake Tana Sub-basin.*

DOI: https://doi.org/10.20372/pjet.v3i2.1850

Corresponding Author: Walelign K. Endalew

Department of Natural Resource Management, Debre Tabor University, Debre Tabor, Ethiopia.

Email: walelignkassie@gmail.com

1. Introduction

Efficient and effective water resources management and development requires reliable estimates of the hydrologic variables like low, mean and peak flows that require the use of appropriate engineering hydrology techniques. Some of these hydrologic variables, for example, floods, have recurrent phenomenon that causes many socio-economic losses. Even in April 2024 we observed many devastating floods in east Africa and Asia that brought the loss of many lives. This phenomenon commonly occurs across Ethiopia, varying in timing and intensity (Tegegne et al., 2020). This is due to heavy rainfall from highland areas causing extensive large-scale flooding, also known as riverine floods, in the nation's low-lying regions. When rivers breach their banks and submerge downstream floodplain land, resulting in flood disasters (Desalegn et al., 2016; Ukumo et al., 2022). In 2006, severe flooding affected parts of the country, with Lake Tana remaining one of the regions most prone to recurring inundation (Tegegne et al., 2020). Accurate peak flood estimation is vital for safeguarding human lives and infrastructure, optimizing water resource utilization, and preserving environmental integrity (Rabba, 2023).

A variety of techniques are available for estimating peak floods, including the rational method, empirical formulas, the unit hydrograph approach, and statistical or probabilistic flood frequency analyses (Othman et al., 2013; Rabba, 2023; Stuckey & Reed, 2000; Vassova, 2013). Flood frequency statistics have been applied in various water resource projects, including floodplain management, bridge engineering, and the design of flood control infrastructure (Vassova, 2013). Obtaining techniques that generate dependable flood flow data are essential for engineers and planners to incorporate into the planning and execution of these projects (Stuckey & Reed, 2000). The selection of an appropriate method for peak flood estimation is influenced by the availability and reliability of data, as well as the type and expected lifespan of the project (Ahn et al., 2014; Topaloğlu, 2002).

Traditional flood frequency analysis is straightforward to apply when estimating design peak flow in gauged catchments with adequate historical flow records. In contrast, for ungauged catchments, empirical approaches that rely on morphometric characteristics such as catchment area, slope, and shape factor are commonly employed to estimate peak discharge (Stuckey & Reed, 2000; Vassova, 2013). Jha, R., and Smakhtin, V. (2008) identified a range of techniques for estimating flood characteristics and long-term mean annual flow, primarily through regression models that incorporate catchment parameters.

Numerous studies have explored the relationship between catchment morphometric attributes and various flow characteristics, including low, mean, and peak flows. For instance, Dubreuil (1986) highlighted research conducted in tropical regions demonstrating that drainage area and slope significantly influence most hydrological parameters, particularly peak flow. Similarly, Garzon et al. (2023) identified catchment area and average slope as critical factors governing peak flow magnitudes. In another study, Sandrock et al. (1992) employed multiple regression analysis to compare flood peaks across different return periods with various morphometric features. Further contributions by Sandrock et al. (1992), Stuckey and Reed (2000), and Topaloğlu (2002) also utilized regression techniques to establish predictive relationships between flow statistics and watershed characteristics.

Hydro-climatic measurements in Ethiopia are very scanty. In Ethiopia, the majority of catchments requiring design peak flow estimates lack gauging data (Mulugeta, 2004). Tassew et al. (2021) highlighted the scarcity and unreliability of hydro-meteorological data within the Tana sub-basin. Notably, nearly all streamflow gauging stations are concentrated in the upper reaches, leaving the lower portion accounting for approximately 40% of the sub-basin completely ungauged (Engida, 2010). Besides, there are no sufficient, validated, and standardized alternative peak flow estimation methods in the sub-basin and elsewhere in Ethiopia.

The Lake Tana sub-basin is rich in natural and cultural resources including abundant water, fertile land, hydropower potential, livestock, forests, fisheries, and heritage sites that contribute significantly to both local and national livelihoods. Its vast potential for irrigation, hydroelectric generation, agriculture, and ecotourism makes it a vital asset for regional and national development (Setegn et al., 2008). It is one of the identified growth corridor areas by the federal democratic Republic of Ethiopia (Belete, 2013). However, flooding is one of the most serious problems in the sub-basin (Tassew et al., 2021). The objective of this paper was to develop an alternative design peak flow estimation model for the ungauged catchments using catchment morphometric features as input variables. Thus, findings will support the water resources management and development efforts through the provision of applicable engineering hydrology techniques.

2. Materials and Methods

2.1. Study Area

The Lake Tana sub-basin serves as the headwaters of the Abbay (Blue Nile) River, contributing over 60% of the Nile's total water flow (Tassew et al., 2021). Geographically, it lies on Ethiopia's northwestern plateau, spanning latitudes 10.9° N to 12.8° N and longitudes 36.7° E to 38.2° E (Fig. 1). At the mouth of the sub-basin sits Lake Tana, a natural treasure of Ethiopia (Belete, 2013). The sub-basin covers a drainage area of approximately 15,321 km², including the lake itself (Ashenafi, 2013). Elevation within the basin ranges from 1,765 to 4,097 meters above sea level, with the terrain around the lake being relatively flat, typically between 1,750 and 1,850 meters. The region experiences an average annual rainfall of about 1,350 mm, with values ranging from 964 mm to 2,000 mm (Awlachew et al., 2007). Characterized by a tropical highland monsoon climate, the wet season extends from June to September. The sub-basin has an average annual temperature of 20°C and a mean yearly evapotranspiration of 773 mm.

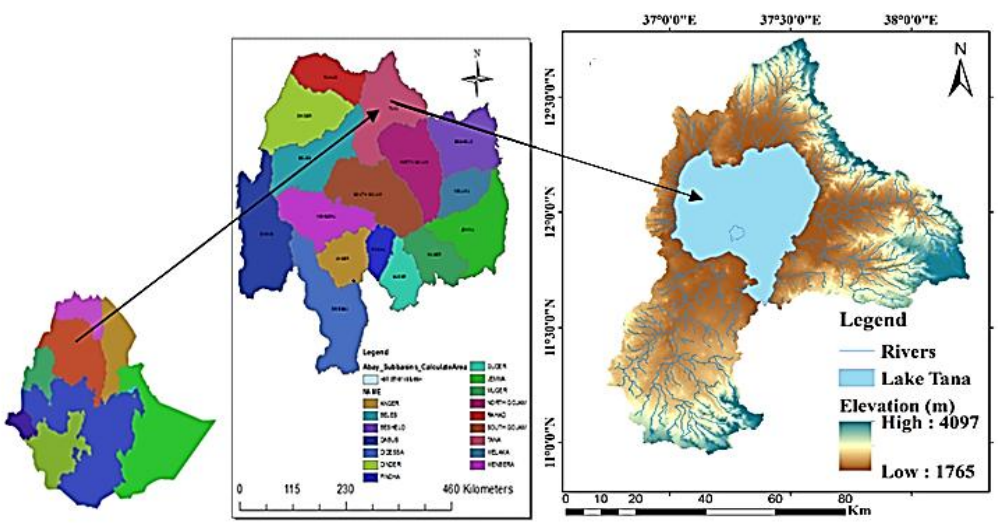

Fig. 1. Location map of the study area

2.2. Data types and collections

Acquiring reliable streamflow measurements is a fundamental and preliminary step in hydrological studies. In the Lake Tana sub-basin, thirteen streamflow gauging stations were initially identified, and daily flow data from these sites were obtained from the Ethiopian Ministry of Water and Energy. Based on criteria such as data completeness, record length, and consistency, eight of the thirteen stations were selected for detailed analysis. The geographic locations of these selected stations, along with maps of their respective catchments, are presented in Fig. 2. From the daily datasets, annual maximum daily flow series were extracted and utilized for flood frequency analysis.

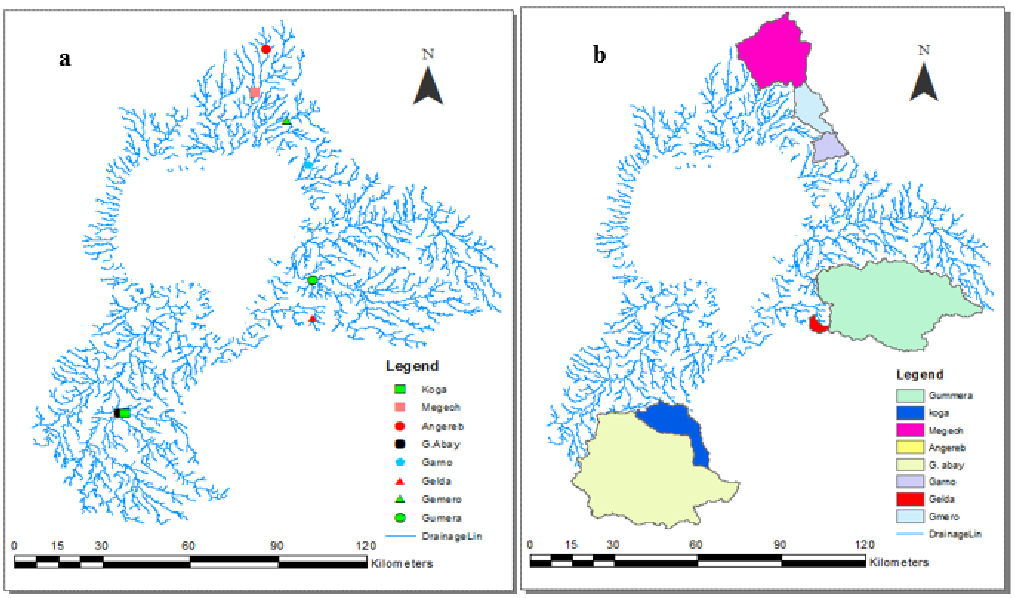

Fig. 2. Location of selected gauging stations (a) and map of selected catchments in the sub-basin (b)

Stations with less than 20 years of continous records and stations that have homogeneiuty problems were screened out intitially. Years of recording periods for gauging stations are shown in Table 1. As illustrated in Fig. 3, the time series plots for the Ribb and Kility Rivers reveal inconsistencies and potential issues in the measurement of annual maximum flow data. Following the preliminary data processing and screening homogeneity and trend tests were done before using the data sets for quantile estimation.

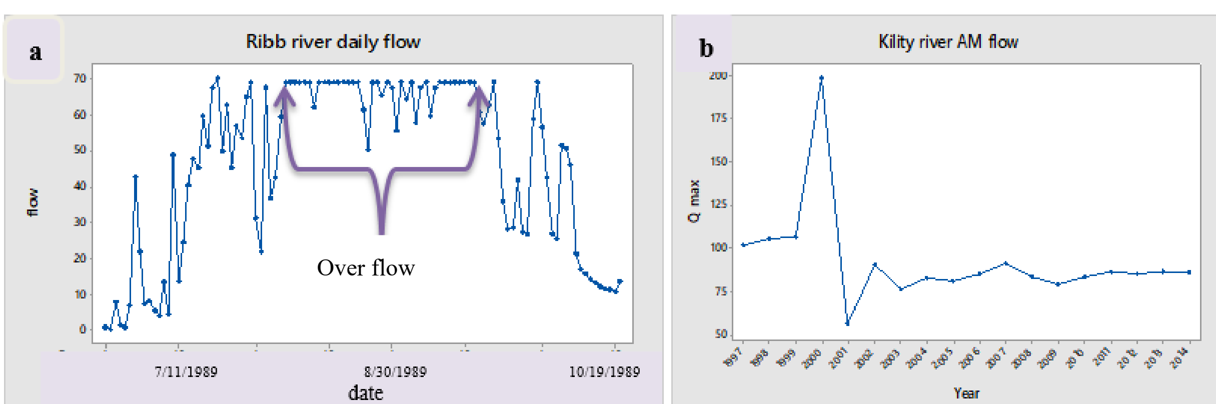

Fig. 3. Ribb River daily discharge (a) and Kility River annual peak flow hydrograph (b)

Table 1. Years of recording periods of gauging stations

Station Name	G.Abay	Garno	Gemero	Gumara	Megech	Qushini	Gelda	koga
Recording length (years)	49	20	23	30	21	23	28	27

A Digital Elevation Model (DEM) with a spatial resolution of 30 by 30 meters for the Lake Tana sub-basin was obtained from the GIS department of the Ministry of Water Resources and Energy. The DEM was processed and analyzed using ArcGIS 10.3 and Arc Hydro 10.3 to extract key morphometric characteristics of the catchment.

2.3. Exploratory Data Analysis (EDA)

Exploratory data analysis (EDA) is a method that utilizes graphical tools to investigate, interpret, and effectively communicate the underlying patterns and characteristics of a dataset. Zbigniew W. K. and Alice J. R. (2004) have noted that a study that does not include a detailed exploratory data analysis is not complete. EDA can give strong insight into patterns and relationships. Time series plots, box plots, and probability plots were used to get insights into the patterns and distributions of the data sets.

2.4. Trend and Homogeneity Tests

Prior to quantile estimation, tests for trend and homogeneity were conducted on the annual maximum daily streamflow data series. Dahmen and Hall (1990) recommended the use of Spearman's rank-correlation method for detecting linear trends due to its simplicity and non-parametric nature. This method is widely applicable and maintains consistent statistical power across both linear and non-linear trend scenarios. In this study, Spearman's rank-correlation was employed to assess the presence of linear trends in the annual maximum daily streamflow series. The approach is based on the Spearman rank-correlation coefficient, Rsp, which is defined as:

$$R_{sp} = 1 - \frac{6 * \sum_{i=1}^{n} Di^2}{n\,(n * n - 1)} \quad (1)$$

Where n represents the total number of observations, D denotes the difference between the ranks of paired values, and i corresponds to the chronological order of the data points.The difference between rankings is computed with:

$$D_i = K_{xi} - K_{yi} \quad (2)$$

Where:Kxi is the rank of the variable in the chronological order (x), Kyi is the rank of the corresponding streamflow value (y). The streamflow series y is transformed into its rank equivalent Kyi by assigning the chronological order number from the original series to the corresponding position in the ranked series. In cases of tied values—where multiple observations share the same rank the average rank is assigned. To assess the presence of a trend, the null hypothesis $H0$: Rsp = 0 (no trend) is tested against the alternative hypothesis $H1$: Rsp≠0 (presence of a trend). The test statistic is:

$$t_t = R_{sp}\left[\frac{n-2}{1-R_{sp}^2}\right]^{0.5} \quad (3)$$

Where t_t, has Student's t-distribution with v = n-2 degrees of freedom.

The distribution is symmetric around $t = 0$. At a 5% significance level (two-tailed), the critical region U is defined as: $U = (-\infty, tv, 2.5\%) \cup (tv, 97.5\%, +\infty)$. If the calculated tt falls outside this range, the null hypothesis is rejected, indicating a significant trend in the time series. In such cases, the data series is considered unsuitable for frequency analysis (Dahmen and Hall, 1990).

To evaluate the homogeneity of the streamflow data, the Mann–Whitney test was employed, as recommended by Salas (1993). This non-parametric test compares two segments of the dataset—of sizes $n1$and $n2$, where $n1<n2$. The combined dataset of size $N=n1+n2$ is ranked in ascending order. The test assesses whether the means of the two segments are statistically equivalent using the Mann–Whitney test statistic Uc.

$$U_c = \frac{\sum_{t=1}^{n_1} R(x_1) - \frac{n_1(n_1+n_2+1)}{2}}{\sqrt{\left(\frac{n_1 n_2 (n_1+n_2)}{12}\right)}} \quad (4)$$

In the Mann–Whitney test, $R(xt)$ represents the sum of the ranks assigned to the elements of the first sample within the combined, ordered dataset. The test statistic Uc is then used to evaluate the hypothesis of homogeneity between the two samples. To determine whether the data series is homogeneous, Uc is compared against the critical value from the standard normal distribution at a chosen significance level α. The null hypothesis that the two segments have equal means is rejected if: $|Uc| > u1-\alpha/2$. Where $u1-\alpha/2$ is the critical value corresponding to the $1-\alpha/2$ quantile of the standard normal distribution. If this condition is met, it indicates a statistically significant difference between the two segments, suggesting that the data series is not homogeneous.

2.5. Descriptive Statistics

For a dataset containing N observations $\{x1, x2, \ldots, xN\}$, key statistical measures such as central tendency, variability, and distribution shape are calculated using standard formulas. These include the arithmetic mean, variance, standard deviation, coefficient of skewness, and coefficient of kurtosis.The sample mean, $\bar{x}$, is estimated by:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_N}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad (5)$$

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 \quad (6)$$

The sample standard deviation, denoted as s, is calculated as the positive square root of the unbiased sample variance $s2$, and is expressed by the following formula:

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2} \quad (7)$$

The coefficient of skewness g is a dimensionless number given by:

$$g = \frac{N}{(N-1)(N-2)} \frac{\sum_{i=1}^{N}(x_i - \bar{x})^3}{s^3} \quad (8)$$

The coefficient of kurtosis k is calculated by:

$$k = \frac{N(n+1)}{(N-1)(N-2)(N-3)} \frac{\sum_{i=1}^{N}(x_i - \bar{x})^4}{s^4} - 3\left[\frac{(N-1)^2}{(N-2)(N-3)}\right] - 1 \quad (9)$$

2.6. Probability Distribution Selection and Quantile Estimation

The most suitable probability distribution for each gauging station was identified using the L-Moment Ratio Diagram (L-MRD) and probability plots, as recommended by Fiorillo & Rolla (1989) and Roa & Hamed (2000). Distribution parameters were estimated using the Probability-Weighted Moments (PWM) method. To estimate the flood quantile QT, it is assumed that the probability of this value being equaled or exceeded in any given year is $1/T$, where T is the return period. Consequently, the probability of non-exceedance of QT is expressed as:

$$F = 1 - 1/T \tag{10}$$

Where: F is the probability of having a flood of magnitude Q_T or smaller, and T is the return period.

2.7. Catchment morphometric analysis

The magnitude of peak flow within a watershed is influenced by its morphometric attributes, such as area, slope, shape, and drainage patterns (Stuckey & Reed, 2000) mainly area, the length of the longest flow path, slope, elongated ratio, circularity ratio, and perimeter, which are the most dominant features (Alemu, 2011; Stuckey & Reed, 2000; Topaloğlu, 2002; Tassew et al., 2021). As pointed out in 2.3 above, the catchments' morphometric characteristics are estimated by processing the 30m * 30m DEM using Arc GIS and Arc Hydro 10.3 version Tools.

Best subset analysis is used to select the most influential morphometric features among the six and lessen the independent varibles in the regression equations. Reducing the independent variables is crucial since a regression equation with a smaller independent variables is preferable and practicable (Sandrock et al., 1992; Stuckey & Reed, 2000). Best subset analysis is a simple techniques that offers separate regression anaysis for each of the six catchment morphometric features.

According to Alemu (2011), the shape of a catchment can be characterized using circularity and elongation ratios. The circularity ratio (Cr) compares the area of the drainage basin to that of a circle with an equivalent perimeter, providing insight into how closely the basin resembles a circular form. The elongation ratio (Er) is defined as the ratio between the diameter of a circle with the same area as the basin and the basin's maximum length, offering a measure of the basin's stretch or compactness (Tassew et al., 2021; Sandrock et al., 1992). The watershed slope reflects the rate of change of elevation along the principal flow path (Eldho, 2007).

2.8. Regression Analysis

Regression helps to investigate the existence of any linear or nonlinear functional relationships among a set of variables. Hence, hydrological studies have made extensive use of multivariate regression methods, including factor analysis, best subset analysis, stepwise regression, and principal component regression (Sandrock et al., 1992; Topaloğlu, 2002). From previous similar work (Topaloğlu, 2002), the general form of the expected relationship is to be non-linear of the following type.

$$Q_T = b_0 * A_1^{\,b1} * A_2^{\,b2} * \ldots * A_i^{\,bi} \tag{11}$$

Where QT represents the estimated flood quantile associated with a return period of T years. The variables $A1$, $A2$, …, Ai denote catchment characteristics, while $b1$, $b2$, …, bi are the corresponding regression coefficients, and $b0$ is a constant term. Since the original relationship is nonlinear, a logarithmic transformation is applied to convert it into a linear form. This transformation is a widely used and straightforward technique that enables linear regression to approximate curves typically handled by nonlinear models. As a result, the nonlinear equation can be reformulated as a linear expression:

$$\log Q_T = b_0 + b_1 \log A + b_2 \log Cr + b_3 \log LL + b_4 \log S \quad (12)$$

2.9. Evaluation of the Regression Models

Once a regression equation is established, various criteria can be used to assess its effectiveness (Philippe, 2012). One of the primary metrics for evaluating model performance is the coefficient of determination (R^2). This statistic is calculated as the square of the ratio between the covariance of observed and predicted values and the product of their standard deviations. As noted by Gebiaw et al. (2017), R^2 ranges from 0 to 1, where values closer to 1 indicate stronger predictive accuracy. It provides a concise measure of the model's explanatory power and is derived from the sums-of squares components in the regression analysis.

$$R^2 = 1 - \frac{SSE}{SST} \quad (13)$$

Where: SSE is the sum of squares error, SST is the sum of squares total

3. Results and discussion

3.1. Result of data quality test

Statistical flood frequency analysis relies on two fundamental assumptions: the independence of data points and the homogeneity or stationarity of the data series (Rao & Hamed, 2000). Test for independency is not necessary as annual hydrological data series are usually expected to be independent. Dependencey is expected in sub-annual data series like monthly and daily data sets. However, annual data series might have trend and the series might not be homogeneous hence test for trend and homogeneity is necessary. The trend and homogeneity test results are shown below in Table 2.

Table 2. Homogeneity test and trend test results of gauged rivers

Rivers	**Homogeneity test (Üc)**	***T_t* *(Trend test statistic***	**Critical value at 5% level (normal distribution)** ($u_{1-\alpha/2}$)	***Critical value at 5% significant level(t-distribution)(t_{cr})***	**Decision**
G.Abay	0.98	0.789	±1.96	2.021	Homogeneous/no trend
Garno	0.302	0.65	±1.96	2.101	Homogeneous/no trend
Gemero	-1.415	1.94	±1.96	2.08	Homogeneous/no trend
Gumera	-0.02	0.86	±1.96	2.048	Homogeneous/no trend
Koga	-3.44	3.03	±1.96	2.06	**Heterogeneous/has a trend**
Megech	-0.14	-0.21	±1.96	2.093	Homogeneous/no trend
Gelda	-1.65	1.31	±1.96	2.056	Homogeneous/no trend
Qushini	1.693	-1.04	±1.96	2.086	Homogeneous/no trend

As indicated in Table 2, the streamflow data for the Koga River exhibit both a significant trend and heterogeneity. To assess homogeneity, the test statistic Uc is applied at a 5% significance level (α = 0.05) and compared against the corresponding critical value from the standard normal distribution. As shown from the table, the value of |Uc| is 3.44; this value is greater than the critical value at the 5% level (standard normal distribution), which is 1.96. The trend test result also shows that the Koga river flow data has a trend, and no trend for other rivers. The t-test value is 3.03, which is greater than the critical value at 5% significance level and n-2 degrees of freedom(t-distribution) ($t_{0.025,25}$), which is 2.06. Consequently, to develop a representative model, we removed those stations before passing further flood frequency analysis, and then we used seven stations only.

3.2. Descriptive statistics

Descriptive statistical measures essential for selecting appropriate probability distributions and estimating their parameters were computed. These include the central moments: mean ($m1$), variance ($m2$), skewness ($m3$), and kurtosis ($m4$).

Table 3. Descriptive statistics of the available streamflow data

Rivers	Descriptive statistic			
	Mean	**Variance**	**Sekewness**	**Kortosis**
G.Abay	336.72	6023.32	816964.36	288934459
Garno	18.81	224.70	7033.67	382326.6431
Gmero	67.96	3308.65	251233.33	56332347.62
Gumera	252.84	2797.29	103595.66	31928421.89
Megech	138.87	5666.37	360090.83	114904416.1
Gelda	45.13	2051.29	198350.84	31905389.53
Qushini	12.16	22.72	94.99	2213.77

Additional key metrics for determining the appropriate at-site probability distribution include the L-moments and L-moment ratios. The calculated values of these sample statistics are presented in Table 4.

Table 4. Sample L-moments of the stream flow data

Rivers	B_0	β_1	β_2	β_3	λ_1	λ_2	λ_3	λ_4	τ	τ_3	τ_4
G.Abay	336.72	181.29	127.52	98.99	336.72	25.86	14.07	-6.87	0.08	0.54	-0.27
Garno	18.82	12.86	10.15	8.53	18.82	6.90	2.57	1.45	0.37	0.37	0.21
Gmero	67.97	48.78	38.70	32.25	67.97	29.59	7.48	1.41	0.44	0.25	0.05
Gumera	252.84	140.66	98.71	76.48	252.84	28.47	1.17	3.39	0.11	0.04	0.12
Megech	138.87	89.64	67.50	54.55	138.87	40.40	6.07	2.80	0.29	0.15	0.07
Gelda	45.13	32.94	26.81	22.96	45.13	20.75	8.35	5.14	0.46	0.40	0.25
Qushini	12.16	7.35	5.37	4.27	12.16	2.53	0.27	0.35	0.21	0.11	0.14

3.3. Selected Probability distributions

The L-Moment Ratio Diagram (L-MRD) is a widely recognized and effective tool for selecting the most suitable at-site probability distribution. In this study, a range of candidate distributions including Generalized Extreme Value, Generalized Logistic, Lognormal, Generalized Pareto, Pearson Type III, Uniform, Exponential Gumbel (EV1), Logistic, Normal, and Wakeby were initially considered and evaluated using the L-MRD approach. Fig. 4 illustrates the L-moment ratio diagram highlighting the five most appropriate distributions identified for the G. Abay River.

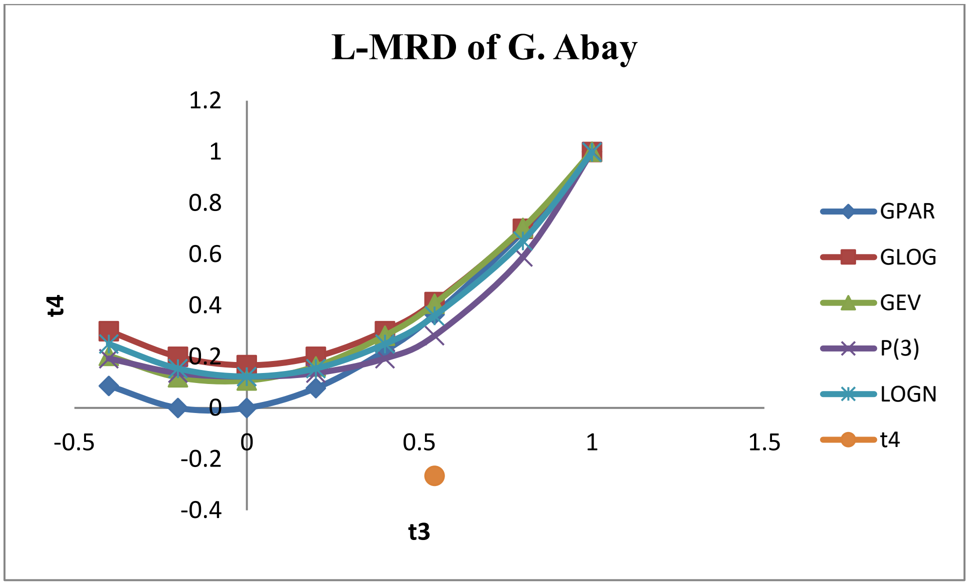

Fig. 4. L-moment ratio diagram for G.Abay River

Note: - GEV: Generalized Extreme Value, GLOG: Generalized Logistic, LOGN: Lognormal, GPAR: Generalized Pareto, P-3: Person type III, uni: Uniform.

When the plotted point of L-skewness ($t3$) versus L-kurtosis ($t4$) for a station aligns with the theoretical curve of a specific distribution on the L-Moment Ratio Diagram, that distribution is considered the best fit for the station. In the case of the G. Abay River, the L-MRD shows that the point corresponding to $t3$ and $t4$ lies on or near the curve of the Pearson Type III (P3) distribution, indicating it as the most suitable model. As summarized in Table 5, the P3, Generalized Pareto (GPAR), Lognormal (LOGN), and Generalized Extreme Value (GEV) distributions emerged as the best-fit models for the gauged rivers within the Lake Tana sub-basin.Table 5. Best fit distribution of each station

Rivers	**G.Abay**	**Garno**	**Gmero**	**Gumera**	**Megech**	**Qushini**	**Gelda**
fitted distribution	P3	GPAR	GPAR	GEV	GPAR	LOGN	LOGN

On the other hand, observed data are plotted against the values estimated from the fitted distributions selected from L-MRD. The relationship appears as a straight line through the origin with 45^0 slopes for all rivers, which means all fitted distributions are exact parent distributions.

3.3.1. Probability distribution parameter estimation

Following the selection of an at-site probability distribution, its parameters were estimated using PWM. The estimated parameters shown in Table 6 are used to estimate the quantile for different return periods.

Table 6. Estimated Parameter value of the selected distribution in each station

Station	Selected distribution	Parameters estimated by PWM		Station	Selected distribution	Parameters estimated by PWM	
		Parameter	Value			Parameter	Value
G.Abay	P3	K	0.34	Megech	GPAR	K	0.48
		α	110.71			α	147.93
		ε	298.75			ε	38.77
Garno	GPAR	K	-0.09	Gelda	LOGN	μ	3.40
		α	12.07			σ	0.94
		ε	5.61				
Gmero	GPAR	K	0.19	Qushini	LOGN	μ	2.43
		α	77.40			σ	0.40
		ε	3.08				
Gumera	GEV	c	0.03				
		K	0.21				
		α	6.11				
		u	262.54				

3.4. Quantile estimation

Estimation of peak flow requires the specification of the frequency of occurrence of the event by specifying its return period. In this study peak discharge (Q) was obtained for the following return periods. The 1 in 10 years = Q_{10}, the 1 in 25 years = Q_{25}, the 1 in 50 years = Q_{50}, the 1 in 100 years = Q_{100} and the 1 in 200 years = Q_{200}.

Table 7. Shows the derived Peak flow frequency for all gauged stations

Station	Selected distribution	Return period				
		10 year	25 year	50 year	100 year	200 year
G.Abay	P3	403.97	476.26	559.09	603.53	770.00
Garno	GPAR	36.35	50.39	61.76	73.84	86.66
Gmero	GPAR	146.99	188.69	215.68	239.30	259.96
Gumera	GEV	276.38	279.11	280.81	282.26	283.51
Megech	GPAR	245.34	281.87	300.63	314.09	323.76
Gelda	LOGN	41.70	57.95	71.49	86.22	102.23
Qushini	LOGN	12.96	14.83	16.17	17.46	18.72

Note: Column 2 of Table 7 indicates the probability distribution that yielded the best fit for each dataset. Columns 3 through 7 present the estimated peak flow values corresponding to the specified return periods.

AS shown in Table 7, the peak flow is increased with increasing return flow in all stations. This is general truth in flood frequency analysis. The peak flow is highly dependent on climate and physiographic catchment characteristics.

3.5. Catchment Morphometric Characteristics

DEM-Hydro process through GIS with Arc Hydro Tool 10.3 software using outlet location coordinates of the catchments, the area, perimeter, and other selected catchment characteristics have been extracted.

Table 8. Summary of catchment physiographic characteristics result

Rivers	Area (km2)	Perimeter (km)	Circularity ratio	Elongated ratio	Slope (m/km)	Longest flow path (km)
G.Abay	1664	309.68	0.22	0.60	9.55	77.18
Garno	94	67.84	0.26	0.55	48.43	19.79
Gmero	174	97.36	0.23	0.61	36.20	24.52
Gumera	1394	297.48	0.20	0.49	23.62	86.06
Megech	462	151.64	0.25	0.55	23.91	43.93
Gelda	32	36.48	0.30	0.55	35.44	11.66
Qushini	42	25.32	0.82	0.82	25.50	8.91

As illustrated in Table 8, the Gumara and G. Abay watersheds exhibit relatively higher circularity compared to the other catchments. This suggests that rainfall within the Garno and Megech watersheds has a shorter flow path to reach the stream network and exit the basin at its outlet. Qushini and Gelda watersheds are more elongated than other watersheds. Alemu, (2011) reported that the circularity ratio of the Gumara River is 0.35 there is some difference with this study, this is because of the software used for estimating catchment characteristics. Tassew et al., (2021) also estimated the physiographic catchment characteristics of major watersheds of the Lake Tana sub-basin. He reported that the circularity ratio of G. Abay, Megech, and Gumara watersheds are 0.28, 0.25, and 0.41 respectively and the elongated ratios are 0.49, 0.52, and 0.51 respectively. The results are relatively similar to the result computed by this study but due to the selection of the outlet points of the watersheds there are some differences.

3.6. Developed Regression Equations

3.6.1. Selection of independent variables

Best subset analysis was employed to identify the most suitable explanatory variables for the regression model, as it allows for evaluating multiple regression solutions across various physiographic factors. The selection process was constrained to models containing up to four explanatory variables, using a 5% significance threshold (α = 0.05). Models were chosen based on the highest coefficient of determination (R^2) and by inspecting residual versus fitted value scatter plots. To determine the optimal subsets, three key criteria were considered: maximum R^2, maximum adjusted R^2, and the Mallows Cp

statistic.Table 9. Best subset analysis results for 10 year return period

Variables	R-Sq	R-Sq(adj)	Mallows Cp	A	Cr	Er	LL	S

1	93.0	91.5	10.6	X				
1	81.8	78.2	32.1				X	
2	95.0	92.5	8.7	X				X
2	94.6	91.9	9.5	X	X			
3	98.6	97.3	3.6		X		X	X
3	97.9	95.8	5.0	X	X			X
4	**99.5**	**98.4**	**4.0**	**X**	**X**		**X**	**X**
4	99.0	97.1	4.9		X	X	X	X
5	99.5	96.9	6.0	X	X	X	X	X

Not: Key variables and statistical indicators used in the regression analysis include: R-Sq (coefficient of determination), R-Sq (adj) (adjusted coefficient of determination), Cp (Mallows' statistic), A (basin area), LL (stream length), S (basin slope), Cr (circularity ratio), and Er (elongation ratio). Based on the best subset regression analysis presented in Table 9, it was found that models incorporating four variables are adequate for estimating flood peaks across all return periods. The most influential variables identified were basin area (A), longest stream length (LL), channel slope (S), and circularity ratio (Cr).

Table 10. Result of best subset analysis for all return periods

Peak flow with return period	Variables	R-Sq	R-Sq(adj)	Mallows Cp
Q10	4	99.5	98.4	4
Q25	4	99.3	97.8	4
Q50	4	99.1	97.3	4
Q100	4	98.9	96.6	4.2
Q200	4	98.6	95.7	4.4

3.6.2. Development of regression equations

Multiple general regression analyses were conducted to formulate regression equations for each return period, guided by the outcomes of the best subset analysis. Prior to modeling, a logarithmic transformation was applied to the data to linearize the relationships. The regression modeling was performed using Minitab software. The resulting equations are presented in a compact form as follows:

$$\mathbf{Q_n = k_n * A^{an} * (Cr)^{bn} * (LL)^{cn} * S^{dn}} \quad (14)$$

Where: Q_n Peak flow for n year return period, k_n is constant, a_n is the regression coefficients of area, b_n is the regression coefficient of circularity ratio, c_n is the regression coefficient of main stream length, d_n is the regression coefficient of average channel slope, n is the return period.

For example, the final regression equations for 50 recurrence intervals with the coefficient of determination (R^2) are shown below in nonlinear form.

$$Q50 = 204.17 * A^{1.2} * Cr^{-2.60} * LL^{-2.31} * S^{-0.78} \qquad (R^2) = 97.95\%$$

Table 11. Coefficients and respective R^2 of the developed regression equations for all recurrence intervals

Qn	kn	a_n	b_n	c_n	d_n	Coefficient of Determination (R^2)
Q10	36.31	1.07	-1.97	-1.63	- 0.58	96.23%
Q25	91.20	1.17	-2.35	-2.08	- 0.67	97.11%
Q50	204.17	1.2	-2.60	-2.31	- 0.78	97.95%
Q100	346.74	1.17	-2.75	-2.37	- 0.84	98.43%
Q200	1000	1.18	-2.99	-2.57	- 1.04	99.25%

Topaloğlu (2002) has developed similar regional regression equations for estimating instantaneous peak flow for the Seyhan River Basin. He used three independent variables, such as annual peak daily mean areal precipitation (P_T), area of the basin (A_B), and stream frequency (S_F). However, the coefficients are relatively similar to the coefficients of this study, for example, 10-year return period, a_n = 2.03, b_n = 1.35, and cn = 2.13. The variation of those coefficients is because of the spatial and temporal variation of the physiographic catchment characteristics.

3.7. Model Performance Evaluation and Validation

According to the result of statistical measures, nonlinear regression analysis with four variables gives the best result. In addition, the nonlinear regression analysis relationship is more effective in the Lake Tana sub-basin. The coefficient of determination (R^2) of the model evaluation criteria shows that the developed model has very good model efficiency.

Table 12. Summary of model evaluation

Model	R^2 (%)	RMSE	MAE	NSE
Q_{10}	99.29	7.166	5.239	0.996
Q_{25}	96.46	7.637	4.393	0.997
Q_{50}	97.76	9.726	9.726	0.995
Q_{100}	98.22	12.474	4.592	0.993
Q_{200}	99.35	13.899	8.524	0.992

As presented in Table 12, the regression equations developed in this study demonstrate strong reliability, supported by consistent performance across all evaluation metrics. Following model development, validation was conducted using data from one measuring station, while six gauging stations were used to construct the model. The validation process, performed with flow data from the Megech River, revealed a close match between observed and predicted peak flows for return periods of 10, 25, 50, and 100 years. The high coefficient of determination (R^2=98.1%) confirms the model's accuracy and robustness for estimating flood peaks.

4. Conclusions

Accurately estimating design peak flows remains a major challenge for hydrologists and engineers, particularly when planning hydraulic structures in ungauged catchments. This study addresses that challenge by developing regression equations that link key catchment morphometric characteristics—namely basin area (A), channel slope (S), circularity ratio (Cr), and longest stream length (LL)—to design peak flows for return periods of 10, 25, 50, 100, and 200 years. The general form of the developed equations is expressed as: $Q_n = k_n{\cdot}A^{an}{\cdot}Cr^{bn}{\cdot}LL^{cn}{\cdot}S^{dn}$. These four morphometric parameters are sufficient to estimate design peak flows at ungauged sites. On-site flood frequency analysis revealed that the Pearson Type III (P3), General Logistic (GLOG), and General Pareto (GPAR) distributions provided the best fit for rivers within the sub-basin. The regression models demonstrated high predictive performance, with coefficient of determination (R^2) values approaching 1 across all return periods. As such, these models offer a practical and simplified alternative for estimating design peak flows in ungauged catchments within the Lake Tana sub-basin and other hydrologically similar regions.

Authors' contributions

Walelign K. Endalew performed material preparation, data collection, analysis, and wrote the first draft of the manuscript and it has been rewritten and thoroughly edited by **Mulugeta A. Belete.** Both authors read and approved the final manuscript.

Funding: No funding source for this study.

Competing Interests

The authors affirm that they have no known financial or non-financial conflicts of interest, nor any personal relationships that could have influenced the research presented in this paper.

Conflicts of interests

This is to inform you that we the authors of the article entitled *"Utilization of catchment morphometric features to estimate design peak flow in ungauged catchments: Insights from Lake Tana sub-basin, Ethiopia"* have decided to submit the manuscript and are responsible for its content. Thus, the authors have not declared any conflict of interest.

Availability of data: The datasets used and/or analyzed during the current study were available from the corresponding author upon reasonable request.

Reference

Ahn, J., Cho, W., Kim, T., Shin, H., & Heo, J. H. (2014). Flood frequency analysis for the annual peak flows simulated by an event-based rainfall-runoff model in an urban drainage basin. Water (Switzerland), 6(12), 3841–3863. https://doi.org/10.3390/w6123841

Alemu, E. (2011). School of Graduate Studies Addis Ababa Institute of Technology Department of Civil Engineering msc Thesis on Effects of Watershed characteristics on River Flow for the Case of Ribb And Gummara Catchments Addis Ababa Institute of Technology Thesis Submitte.

Ashenafi, A. B. (2013). Watershed modeling and uncertainty analysis on Gumara watershed. 117.

Awulachew, Seleshi B.; Yilma, A. D.; Loulseged, M.; Loiskandl, W., Ayana, M.; Alamirew, T. (2007): Water resources and irrigation development in Ethiopia. Colombo, Sri Lanka: International Water Management Institute. (Working Paper 123)

Belete, Mulugeta A. (2013): Modeling and Analysis of Lake Tana Sub-Basin Water Resources Systems, Ethiopia. Dissertation. Faculty of Agriculture and Environmental Science, Rostock

University, Rostock.

Bitew G. Tassew, Mulugeta A. Belete & K. Miegel (2021): Assessment and analysis of morphometric characteristics of Lake Tana sub-basin, Upper Blue Nile Basin, Ethiopia, International Journal of River Basin Management, DOI: 10.1080/15715124.2021.1938091

Dahmen, E. R., & Hall, M. J. (1990). *Screening of hydrological data: tests for stationarity and relative consistency* (No. 49). ILRI.

Desalegn, A., Demissie, S., & Admassu, S. (2016). Extreme Weather and Flood Forecasting and Modelling for Eastern Tana Sub Basin, Upper Blue Nile Basin, Ethiopia. Journal of Waste Water Treatment & Analysis, 7(3). https://doi.org/10.4172/2157-7587.1000257

Dubreuil, P.L., 1986. Review of relationships between geophysical factors and hydrological characteristics in the tropics. J; Hydrol., 87: 201-222.

Eldho, T. I. (2007). Watershed management lecture note. Indian Institute of Technology Bombay, India.

Engida, A. N. (2010). Hydrological and suspended sediment modeling in the Lake Tana sub basin, Ethiopia. PHD Thesis. Universities Joseph-Fourier-Grenoble

Fiorillo, G., & Rolla, P. A. (1989). Regional flood frequency analysis for the Tre Venezie rivers. Energ. Elettr., 66(5, May 1989), 201–216.

Gebiaw, T., Ayele, Engidasew, Z., Teshale, Bofu, Y., Ian, D., . . . Jaehak, J. (2017). Streamflow and Sediment Yield Prediction for Watershed Prioritization in the Upper Blue Nile River Basin, Ethiopia. Journal Vol.9, no.10.

Garzon, L. F. L., Johnson, M. F., Mount, N., Gomez, H., 2023. Exploring the effects of catchment morphometry on overland flow response to extreme rainfall using a 2D hydraulic-hydrological model (IBER).J;Hydrol.,627.

Hamed, K. H., & Roa, A. R. (2000). Flood Frequency Analysis. CRC press LLC, Florida.

Jha, R.; Smakhtin, V. 2008. A review of methods of hydrological estimation at ungauged sites in India. Colombo, Sri Lanka: International Water Management Institute. 24p. (IWMI Working Paper 130)

Mulugeta, A. (2004). Development of synthetic unit hydrographs for watersheds in the Upper Awash and Upper Tekeze Basin, Ethiopia. M.Sc Theisis.

Othman, F., Sadeghian, M. S., Ebrahimi, F., & Heydari, M. (2013). A Study on Sedimentation in Sefidroud Dam by Using Depth Evaluation and Comparing the Results with USBR and FAO

Methods. International Proceedings of Chemical, Biological and Environmental Engineering, 51(9), 6. https://doi.org/10.7763/IPCBEE

Philippe, C. (2012). Flood frequency estimation for ungauged catchments in Iceland by combined hydrological modiling and regional frequency analysis. Icelandic Met Office, Vol. 2014-001, ISSN 1670-8261.

Rabba, Z. A. (2023). Flood frequency analysis with PyTOPKAPI model-simulated stream flows from Aweitu river in Jimma town, Ethiopia. Sustainable Water Resources Management, 9(2), 1–14. https://doi.org/10.1007/s40899-023-00825-2

Sandrock, G., Viraraghavan, T., & Fuller, G. A. (1992). Estimation of peak flows for natural ungauged watersheds in southern saskatchewan. Canadian Water Resources Journal, 17(1), 21–31. https://doi.org/10.4296/cwrj1701021

Setegn, S. G., Srinivasan, R., & Dargahi, B. (2008). Hydrological Modelling in the Lake Tana Basin, Ethiopia Using SWAT Model. The Open Hydrology Journal, 2(1), 49–62. https://doi.org/10.2174/1874378100802010049

Stuckey, M. H., & Reed, L. A. (2000). Techniques for estimating magnitude and frequency of peak flows for Pennsylvania streams. 47.

Tegegne, G., Melesse, A. M., Asfaw, D. H., & Worqlul, A. W. (2020). Flood frequency analyses over different basin scales in the Blue Nile River Basin, Ethiopia. Hydrology, 7(3). https://doi.org/10.3390/HYDROLOGY7030044

Topaloğlu, F. (2002). Estimation of instantaneous peak flows in Seyhan river basin using regional regression procedures. Turkish Journal of Agriculture and Forestry, 26(1), 47–55. https://doi.org/10.3906/tar-0012-2

Ukumo, T. Y., Lohani, T. K., Edamo, M. L., Alaro, M. A., Ayele, M. A., & Borko, H. B. (2022). Application of Regional Climatic Models to Assess the Performance Evaluation of Changes on Flood Frequency in Woybo Catchment, Ethiopia. Advances in Civil Engineering, 2022. https://doi.org/10.1155/2022/3351375

Vassova, D. (2013). Comparison of rainfall-runoff models for design discharge assessment in a small ungauged catchment. Soil and Water Research, 8(1), 26–33.

Zbigniew W. Kundzewicz & Alice J. Robson (2004). Change detection in hydrological records:a review of the methodology. Hydrological Sciences Journal.49:1, 7-19.