Civil Engineering Beyond Limits 1 (2025) 1991

Civil Engineering Beyond Limits

journal home: www.acapublishing.com/journals/1/cebel

Research Article

Estimation of River Peak Flows Using Different Hydrological Modeling Methods: A Case Study of the Demirci River Basin

Naci Büyükkaracığan*,[1]

SBMYO, Selcuk University, Turkey

Keywords

Annual peak flows,
Flood,
Goodness of fit tests,
Probability distribution,
Synthetic methods.

Abstract

In recent years, sudden floods caused by global warming and climate change have increasingly impacted social life in Türkiye, leading to significant loss of life and property. At the same time, protecting and managing water resources has become essential to sustainably meet the growing demand for drinking water, particularly in metropolitan areas. The Demirci River Basin, located within the Gediz Basin and spanning the provinces of Manisa, Kütahya, and Uşak, serves as a critical source for both irrigation and, to a lesser extent, drinking water. Accurate estimation of annual peak flows is crucial, as these values are fundamental inputs in the planning and design of water resource projects and flood risk management. In this study, both statistical and synthetic methods were applied to estimate annual peak discharges in the Demirci River Basin. In the statistical approach, nine different probability distributions—including Log-normal-2, Gumbel, Pearson-3, Log-Pearson-3, Log-Boughton, log-logistic, Wakeby, Pareto and Weilbull were analyzed using six different parameter estimation techniques, resulting in a total of 28 model combinations. The suitability of these models was evaluated using the Kolmogorov-Smirnov, Chi-square, and Cramér-von-Mises tests. Additionally, synthetic methods such as Snyder, Kirpich, Mockus, and SCS were employed. Annual maximum flow data were used for the statistical methods, while 50- and 100-year rainfall records were used as input for the synthetic methods. The findings indicate that the Log-Pearson Type III distribution yielded the most reliable results among the statistical methods, while the Kirpich method was the most effective among the synthetic approaches.

1. Introduction

Climate change, rapid urbanization, and significant alterations in land use have led to a marked increase in both the frequency and intensity of flood events. Floods rank among the most devastating natural disasters, causing substantial economic damage and loss of life. This growing threat necessitates the early identification of flood risks and the implementation of both structural and non-structural mitigation measures. In this context, hydrological modeling methods serve as essential tools for the reliable estimation of flood discharges [1].

Flood events represent the leading cause of economic losses among natural disasters on a global scale. In particular, the accurate and reliable estimation of flood discharge is of critical importance for the development of flood hazard maps, the design of early warning systems, infrastructure planning (such as bridges, culverts, and dams), and urban development strategies. Flood discharges serve as essential input data in hydraulic modeling studies, playing a key role in the identification of flood-prone areas. Moreover, these data are also required for developing flood risk management strategies, formulating climate adaptation policies, and supporting disaster insurance practices [2]. Today, increasing population density and the pressures of urbanization have made the precise estimation of flood discharges even more essential. Consequently, flood discharge is not only vital for engineering applications but is also regarded as a strategic parameter in terms of environmental sustainability, disaster risk reduction, and public safety.

Annual peak flow values represent the maximum discharge observed in a river or watercourse within a year and are considered a key parameter in hydrological analyses. These values are critically important for flood management, water resources planning, and infrastructure design. Annual peak flows play a central role in flood risk assessments, as they are used to estimate the frequency and magnitude of flood events. Determining flood discharges is essential in the design of river channels, bridges, dams, and other hydraulic structures. Moreover, annual peak flow data are required for the planning of waterways and drainage systems. These values help ensure that channel dimensions and pumping systems are designed with adequate capacity. In infrastructure design, the durability and safety of structures are assessed based on their ability to withstand expected flow capacities [3].

In climate change analyses, annual peak flow values are used to monitor the effects of climate change on the water cycle. Rising temperatures and shifting precipitation patterns can lead to changes in flow patterns. These data play a crucial role in climate change predictions and the development of adaptation strategies. Consequently, annual peak flow values are an indispensable parameter in hydrological modeling and planning processes, facilitating decision-making in critical areas such as water management and flood control [4].

Hydrological modeling aims to mathematically represent the water cycle by utilizing physical and meteorological data such as rainfall, flow, evaporation, soil characteristics, and basin morphology. The methods used to estimate flood discharges are generally classified into two main categories: statistical methods and synthetic (empirical or deterministic) methods. Statistical methods use probability distributions based on

Received 07 Apr 2025; Revised 09 Apr 2025; Accepted 09 Apr 2025 https://doi.org/10.36937/cebel.2025.1991

historical flow data to estimate flood magnitude, while synthetic methods focus on modeling flood characteristics by considering the physical and hydrological properties of the watershed [5].

In statistical flood frequency analysis, the most commonly used probability distributions include the two- and three-parameter log-normal distribution, Gumbel Extreme Value Distribution, Pearson- 3, Log-Pearson -3, Log-logistic, Wakeby, and Pareto distributions. These distributions are particularly useful for conducting flood analyses based on annual maximum flow data [6]. The selection of the appropriate distribution is a critical factor that directly influences the accuracy of flood predictions. Therefore, testing different distributions on the same dataset is essential for assessing model performance.

On the other hand, synthetic methods are alternative modeling approaches, particularly preferred when long-term flow data is unavailable or insufficient. Synthetic flood hydrographs are of critical importance in the design of hydraulic structures and flood risk assessment. These hydrographs represent flood peaks and volumes for a given return period and are typically derived using quantiles from flood frequency analyses. However, it is important to note that the same flood peak may occur with different volumes and various hydrograph shapes. To address this, synthetic design hydrographs based on different flood types have been developed. This approach considers the relationship between flood peak and volume and determines probabilities for each flood type. For example, a study conducted on 162 basins in Germany created various hydrograph shapes using different probability density functions and derived best- and worst-case scenario models [7].

Accurate estimation of flood peaks, volumes, and hydrographs is essential for the safe and cost-effective design of hydraulic structures. When generating synthetic flood hydrographs, flood-type-specific hydrograph shapes and peak-volume relationships should be considered. These methods are important tools in flood risk management and the design of hydraulic structures [8]. These models typically combine precipitation data with the physical characteristics of the watershed to produce flood hydrographs [9]. Specifically, the Kirpich method is a simple yet effective approach recommended for small basins, while the Snyder method can be used for larger and more complex basins with appropriate parameter adjustments.

The Demirci River Basin is one of the significant sub-basins of the Gediz River, frequently facing flood risks. The topographic features, land use, and climatic characteristics of the basin are conducive to the formation of floods. In this context, flood predictions for the Demirci River Basin are of critical importance for regional risk management.

This study applies various hydrological modeling methods to predict flood discharges in the Demirci River Basin. The methods used are grouped into two categories: (1) Statistical methods, including Log-normal- 2, Gumbel, Pearson-3, log-Pearson-3, Log-Boughton, log-logistic, Wakeby, Pareto and Weilbull distributions; and 4 Synthetic methods, which include the Snyder, Kirpich, Mockus, and S.C.S methods. For statistical methods, annual maximum flow data were used, while for synthetic methods, rainfall data corresponding to 50 and 100 year return periods were used as model inputs. Thus, a total of 32 probability distribution models and 4 synthetic models were applied to flow and rainfall data through the implementation of different parameter methods with 10 distribution models.

The aim of this study is to demonstrate how different hydrological modeling methods can produce varying results for the same basin and to compare the accuracy and applicability of these methods in light of these results. Through this, the study aims to provide a scientific contribution to the selection of the most appropriate methods for flood risk analysis.

2. Material

The Demirci River Basin is located in the western part of Turkey, within the borders of Demirci district in Manisa province. The basin consists of Demirci River, which flows from north to south, and its tributaries. The main tributaries of Demirci River include the Gümele River, originating from the south and east of Asi Tepesi, the Değirmen River, the Iklıkçı River, and the Alaağaç River. These rivers converge to feed Demirci River and form the basin's hydrological dynamics. The basin's topography is situated on the southern slopes of the Demirci Mountains, with significant elevations such as Akçakertik Ridge (1475 m), Türkmen Mountain (1487 m), and Ziyaret Hill (1795 m). These elevations define the watershed boundary and affect the region's hydrological regime [10].

Demirci River is one of the significant tributaries of the Gediz River and is located in the Upper Gediz Sub-Basin. Due to this position, the water potential and flood risk of the basin have a direct impact on the general hydrological dynamics of the Gediz River. The geological structure of the basin is also noteworthy. Studies conducted around İçikler have revealed that the region's geological characteristics, particularly lateritic and saprolitic formations, influence hydrogeomorphological processes [11]. Due to its hydrological and geological features, the Demirci River Basin is an important region for water resource management, flood risk reduction, and environmental planning.

For the probability methods in the study, maximum flow values obtained from station 05A022, operated by the General Directorate of State Hydraulic Works (DSİ), were selected as the data. The rainfall area of the station is 818.80 km², and the maximum instantaneous flow recorded during the observation period is 890,000 m³/s (on 15.12.1981), while the minimum instantaneous flow is 0.000 m³/s (on 19.07.1972). For synthetic methods, rainfall data obtained from the 17746 observation station operated by the General Directorate of Meteorology (DMI) were used. According to the data from the Demirci Station, an automatic meteorological observation station, the average monthly rainfall is 53.2 mm/month, with a total of 264.8 mm/month. The distance between the two stations, as shown in Figure 1, is 21 km in a straight line. For this study, river flow and rainfall data from the period 1960-2015 for both stations were used. The study area and the locations of the observation stations used in the study are shown in Figure 1.

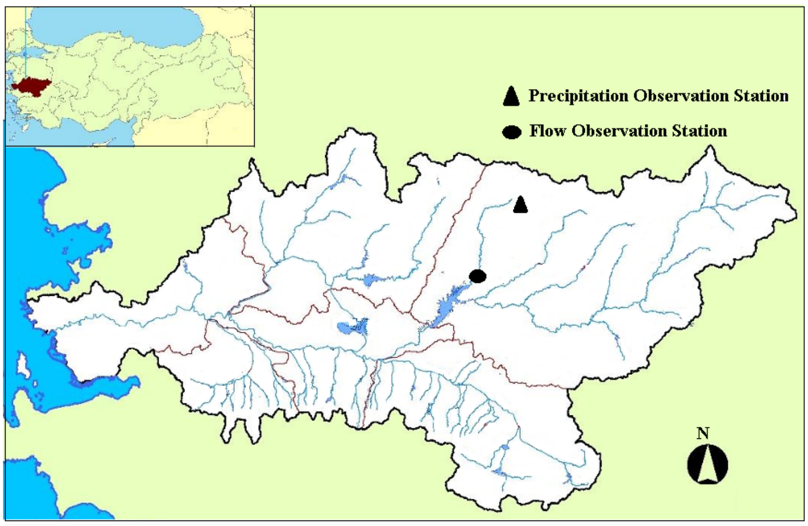

Figure 1. Demirci River Basin and the locations of the observation stations

3. Methods

3.1. Hydrological modeling methods

3.1.1. Statistical methods

A probability distribution defines the likelihood associated with each measurable subset of a sample space in a random process, such as those encountered in experiments, surveys, or statistical inference tasks. For example, categorical distributions are used for sample spaces that are non-numeric; discrete random variables rely on probability mass functions; while continuous random variables are represented by probability density functions. In more complex scenarios, such as continuous-time stochastic processes, more extensive probability measures are applied.

In the field of hydrology, a random variable's observed value (x) corresponds to a stochastic flow event. Although the exact value may fluctuate, it is possible to estimate the event's recurrence probability by constructing a theoretical probability model for the variable. The peak flow series associated with such hydrological events can be expressed using the following equation:

$$P\ (Q \leq Qt\ ) = F(Q = Q+\ ) = SQ + F\ (Q)\ d\ Q \quad (1)$$

In the equation mentioned earlier, $SQ + F\ (Q)\ d\ Q$ represents the annual maximum peak discharge. The return period T (measured in years) is a widely used metric in hydrology, often favored over exceedance probability. It indicates the average time interval between occurrences of a specific flow magnitude or higher, and can be mathematically expressed as:

T=1/Prob(Q<Q_+) ve Prob (Q≤Q_+)=1-1/T (2)

In defining probability distributions, it is essential to differentiate between discrete and continuous random variables, particularly in elementary cases. For discrete variables, assigning probabilities to individual outcomes is straightforward. When the variable is real-valued—or when its possible values have a defined total order—the cumulative distribution function (CDF) represents the probability that the variable takes on a value less than or equal to a specified threshold. In the continuous case, provided a probability density function (PDF) exists, the CDF is obtained by integrating the PDF over the relevant domain.

Frequency analysis in hydrology relies on the application of probability distribution models. This analytical approach uses observed annual flow data to derive key statistical parameters, such as the mean, standard deviation, skewness, and return period. These parameters are then utilized to generate frequency distributions, which present the probability of various flow magnitudes as functions of recurrence intervals or exceedance probabilities. The form of the frequency distribution varies depending on the underlying statistical model employed in the analysis [12].

3.1.1.1. Log-Normal- 2 distribution

In a log-normal distribution X, the parameters μ and σ correspond to the mean and standard deviation of the natural logarithm of the variable. This means that the variable's logarithmic transformation follows a normal distribution. Mathematically, this is expressed as ln(X)~N(μ,σ2) ndicating that standard statistical techniques applicable to normal distributions can be used after log-transformation [14]. This property makes the log-normal model particularly useful in hydrological studies where flow values are strictly positive and skewed.

$$X = e^{\mu + \sigma Z} \quad (3)$$

In contrast, the arithmetic mean, standard deviation, and variance of the original (non-log-transformed) data are denoted in this study as m, s.d., and v, respectively. These empirical parameters differ from the parameters μ and σ of the log-transformed data. However, a mathematical relationship exists between the two sets of parameters, allowing for conversion between the log-scale and the original scale (see also: arithmetic moments).

$$\mu = \ln \frac{m^2}{\sqrt{v+m^2}} \ , \ \sigma = \sqrt{ln\ (1+\frac{v}{m^2})} \quad (4)$$

The probability density function (PDF) of a log-normal distribution is given by:

$$f_x = (x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}}\ e^{-\frac{(lnx-\mu)^2}{2\sigma^2}} \quad , \quad x > 0 \quad (5)$$

Where μ and σ are the mean and standard deviation of the natural logarithm of the variable, respectively. This function describes the likelihood of a random variable taking a particular value x, assuming the variable follows a log-normal distribution.

This result is derived by applying the change-of-variables rule to the probability density function (PDF) of the normal distribution. The cumulative distribution function (CDF) of a log-normal distribution, which gives the probability that the random variable X is less than or equal to a certain value x, is expressed as:

$$f_x = (x; \mu, \sigma) = \frac{1}{2}\left[1 + erf\ (\frac{\ln x - \mu}{\sigma\sqrt{2}})\right] = \varphi\left(\frac{\ln x - \mu}{\sigma}\right)\ x > 0 \quad (6)$$

Where Φ(·)represents the cumulative distribution function of the standard normal distribution. This CDF indicates the probability that the log-transformed variable ln(X) falls below a certain threshold [14].

3.1.1.2. Gumbel distribution

The Gumbel distribution is used to model the distribution of the maximum (or minimum) values from a set of samples with various distributions. It can represent, for example, the maximum river level in a given year, based on past maximum values. This distribution is particularly valuable in predicting the likelihood of extreme events, such as earthquakes, floods, or other natural disasters. The relevance of the Gumbel distribution in extreme value theory lies in its suitability for data with normal or exponential distributions [15].

The standard Gumbel distribution corresponds to the case where μ=0and β=1, and its cumulative distribution function (CDF) is:

$$F(x) = \exp(-exp\ (\frac{-x}{\beta})) \quad (7)$$

In this case, the parameters μ and β simplify, making the Gumbel distribution useful for modeling extreme values in various fields such as hydrology and engineering.

The probability density function (PDF) of the standard Gumbel distribution, where μ=0 = andβ=1, is given by:

$$f(x) = \frac{1}{\beta}\exp(-\left(\frac{x}{\beta}\right) - \exp\left(\frac{-x}{\beta}\right) \quad (8)$$

This function describes the likelihood of extreme values, making it suitable for modeling the distribution of maxima in various phenomena.

In the Gumbel distribution, the mode is 0, the median is ln(ln(2))≈0.3665\ln(\ln(2, the mean is γ , and the standard deviation is π√6≈1.2825. The quantile function, which is the inverse of the cumulative distribution function, is given by:

$$Q(p) = \mu - \beta \ln(-\ln(p)) \quad (9)$$

Here, Q(u) follows a Gumbel distribution with parameters μ\ and β when the random variable U is drawn from a uniform distribution on the interval (0,1) (0,1) (0,1).

3.1.1.3. Pearson-3 distribution

The Pearson system was developed to model skewed data. While it was known how to adjust a theoretical model to match the first two cumulants (mean and variance), the challenge was to adjust skewness (third cumulant) and kurtosis (fourth cumulant) freely. This became evident when fitting theoretical models to skewed observed data, such as survival data, which are often asymmetric [16].

The functions of the distribution are as follows:

$$\lambda = \mu_1 + \frac{b_0}{b_1} - (m+1)b_1 \quad (10)$$

$$\Gamma(m+1, b_1^2) \quad (11)$$

$$b_0 + b_1(x - \lambda) \quad (12)$$

In the given formulas, λ represents a parameter related to the location or central tendency of the distribution, while μ1 denotes the first moment (mean) of the distribution, which is often used to indicate the expected value. The parameter mmm refers to the shape or order, typically used in the context of distributions like the Gamma distribution. The term 〖Γ(m+1,b〗_1^2) represents the Gamma function, where m+1 is the shape parameter and b_1^2 is the square of the scale parameter, commonly used in probability and statistics to describe continuous, positive-valued distributions. Finally, b0 and b1 are parameters that affect the distribution's shape and scale, with b0typically representing a location shift and b1 determining the spread or scale of the distribution.

3.1.1.4. Log Pearson-3 distribution

The Pearson- 3 (Gamma) distribution is used to estimate the frequency of extreme events when all events, large and small, follow a log-normal distribution. This occurs when the events result from the product of many independent random variables. In hydrology, the log-normal distribution has been shown to effectively model variables such as individual storm precipitation depth and annual peak discharges [17].

The probability density function (PDF) using the shape-scale parameterization is:

$$f(x;k;\theta) = \frac{x^{k-1} e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)} \text{ for } x > 0 \ \ and \ k, \theta > 0 \tag{13}$$

In the given formula, x is the random variable, k is the shape parameter, θ is the scale parameter, and Γ(k)) is the Gamma function, which normalizes the distribution such that the total probability is 1, with x>0x and k, θ>0.

The cumulative distribution function is the regularized gamma function:

$$(x;k;\theta) = \int_0^x f(u;k;\theta)du = \frac{\gamma(k, \frac{x}{\theta})}{\Gamma(k)} \tag{14}$$

In the given formula, the cumulative distribution function (CDF) of the Gamma distribution is expressed using the regularized Gamma function, f (k; x; θ) is the lower incomplete Gamma function, and Γ(k) is the Gamma function, normalizing the result so that the CDF ranges from 0 to 1 as x increases.

3.1.1.5. Log-Boughton distribution

The Log-Boughton distribution is a probability distribution that is commonly used in various fields such as hydrology, meteorology, and economics. It is derived by transforming a standard distribution (often a normal distribution) using the logarithmic function. This distribution is useful for modeling skewed data, where the variable of interest is the result of multiplying several independent random variables, leading to a log-normal behavior.

The Log-Boughton (LB) distribution is based on the relationship:

$$(\mathrm{Kb} - \mathrm{A})(\mathrm{B} - \mathrm{A}) = \mathrm{C}\ (40) \tag{15}$$

In this formula, Kb represents the frequency factor, while A and C are parameters of the distribution. These parameters are derived empirically to minimize the mean square error of C. The LB distribution does not have a clearly defined analytical probability density function. It has been suggested to perform as well as popular flood frequency analysis models. However, methods such as Maximum Likelihood (ML) and Probability Weighted Moments (PWM) are not yet applicable to this distribution. Jain & Singh (1987) found the LB distribution to be one of the more effective models in their study [18].

3.1.1.6. Log-Logistic distribution

The log-logistic distribution describes a random variable whose logarithm follows a logistic distribution. It shares a similar shape with the log-normal distribution but has heavier tails. Unlike the log-normal distribution, its cumulative distribution function can be expressed in a closed form [19]. The cumulative distribution function F(x;α;β) of the log-logistic distribution is given by:

$$F(x;\alpha;\beta) = \frac{1}{1+\left(\frac{x}{\alpha}\right)^{-\beta}} \tag{16}$$

where α is the scale parameter, and β is the shape parameter of the distribution. Here is the probability density function (PDF) for the log-logistic distribution:

$$F(x;\alpha;\beta) = \frac{\left(\frac{\beta}{\alpha}\right)\left(\frac{x}{\alpha}\right)^{\beta-1}}{1+\left(\frac{x}{\alpha}\right)^{-\beta}} \tag{17}$$

where α is the scale parameter and β is the shape parameter. This function describes the probability of observing a given value x in the log-logistic distribution.

3.1.1.7. General Extreme Value Distribution (GEV)

This distribution is a family of continuous probability distributions developed within extreme value theory to combine the Gumbel, Fréchet and Weibull families also known as type I, II and III extreme value distributions. By the extreme value theorem the GEV distribution is the only possible limit distribution of properly normalized maxima of a sequence of independent and identically distributed random variables. Note that a limit distribution need not exist: this requires regularity conditions on the tail of the distribution. Despite this, the GEV distribution is often used as an approximation to model the maxima of long (finite) sequences of random variables [20].

The General Extreme Value (GEV) distribution is a family of continuous probability distributions that integrates the Gumbel, Fréchet, and Weibull distributions, also referred to as type I, II, and III extreme value distributions. According to extreme value theory, the GEV distribution is the only possible limiting distribution for the properly normalized maxima of a sequence of independent and identically distributed random variables. It's important to note that a limit distribution may not always exist, as it requires certain regularity conditions on the tail of the distribution. Nevertheless, the GEV distribution is frequently used as an approximation to model the maxima of long (finite) sequences of random variables [20].

The General Extreme Value (GEV) distribution has a cumulative distribution function (CDF) given by:

$$F(x;\mu,\sigma,\xi) = \exp\left(-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right) \tag{18}$$

where μ, σ, and ξ are the location, scale, and shape parameters, respectively. The CDF is valid when $1 + \xi\left(\frac{x-\mu}{\sigma}\right) > 0$.

The probability density function (PDF) of the GEV distribution is expressed as:

$$f(x;\mu,\sigma,\xi) = \frac{1}{\sigma}\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi-1}} \exp\left(-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right) \quad (19)$$

This PDF provides the likelihood of observing a specific value x in the GEV distribution, and both functions are essential for analyzing extreme values in various fields.

3.1.1.8. Wakeby distribution

The Wakeby distribution is a flexible probability distribution often used to model extreme values in various fields. It is a generalized distribution derived from transformations of other known distributions. The Wakeby distribution is particularly useful for modeling data that has both light and heavy tails [21].

The distribution is defined by the transformation of the generalized Pareto distribution and the beta distribution. It allows for more flexibility by combining features from both of these distributions. The cumulative distribution function (CDF) of the Wakeby distribution is given by:

$$F(x;\alpha,\beta,\gamma,\delta) = 1 - (1 + (\tfrac{x-\gamma}{\delta})^{\alpha})^{-\beta} \quad (20)$$

In the Wakeby distribution, α is the shape parameter, β is the shape parameter that controls the tail behavior, γ is the location parameter, and is the scale parameter.

The probability density function (PDF) is the derivative of the CDF and is given by:

$$f(x;\alpha,\beta,\gamma,\delta) = \frac{\alpha\beta}{\delta}(\tfrac{x-\gamma}{\delta})^{\alpha-1}\,(1 + (\tfrac{x-\gamma}{\delta})^{\alpha})^{-\beta-1} \quad (21)$$

3.1.1.9. Pareto distribution

The Pareto distribution is a skewed, heavy-tailed distribution often used to model phenomena like income distribution. The cumulative distribution function (CDF) of a Pareto random variable with parameters α\alphaα (shape parameter) and xmx_mxm (scale parameter) is defined as:

$$F_x(x) = \begin{cases} 1 - (x_m)/x^{\alpha}, & x \geq x_m \\ 0 & , \quad x < x_m \end{cases} \quad (22)$$

In the Pareto distribution, α\alphaα is the shape parameter, and xmx_mxm is the minimum value of x.

When the Pareto distribution is plotted on linear axes, it exhibits a characteristic J-shaped curve that asymptotically approaches both axes. The segments of this curve are self-similar, provided appropriate scaling factors are applied. However, when plotted on a log-log scale, the distribution is represented by a straight line [22].

By differentiating the cumulative distribution function, the probability density function (PDF) is obtained, as follows:

$$F_x(x) = \begin{cases} \frac{\alpha\, x_m^{\alpha}}{x^{\alpha+1}} & x \geq x_m \\ 0 & x < x_m \end{cases} \quad (23)$$

3.1.1.10. Weilbull distributions

The Weibull distribution is a continuous probability distribution introduced by Waloddi Weibull in 1951. It is commonly used in reliability analysis and life data modeling. The probability density function (PDF) of a Weibull random variable is given by:

$$f(x;\lambda;\mathrm{k}) = \begin{cases} \frac{k}{\lambda} - (\frac{x}{\lambda})^{k-1}\, e^{-(\frac{x}{\lambda})^{\mathrm{k}}}, & x \geq 0 \\ 0 & , x < x_m \end{cases} \quad (24)$$

In the Weibull distribution, k>0 is the shape parameter, and λ>0is the scale parameter. The complementary cumulative distribution function follows a stretched exponential form. The Weibull distribution is including moments, maximum likelihood estimation (MLE), probability-weighted moments (PWM), maximum entropy, mixed moments, and individual probability-weighted moments, in that order [24].

Specifically, when k=1, it reduces to the exponential distribution, and when k=2k, it represents the Rayleigh distribution.

The cumulative distribution function (CDF) for the Weibull distribution is given by:

f(x;λ;k)=1-e^(-(x/λ)k)

3.1.2. Parameter estimation methods

In parameter estimation, neutrality is a key characteristic. However, it is advisable to use the sample with the least variance for more accurate estimates. This study applies several methods for parameter estimation, Moments Method involves using the sample moments (mean, variance, skewness, etc.) to estimate the parameters of a probability distribution. The idea is to equate the sample moments with the theoretical moments of the distribution and solve for the parameters.

Maximum Likelihood Estimation (MLE) is a method that estimates the parameters by maximizing the likelihood function. This function represents the probability of observing the given data, and the parameters that maximize this likelihood are considered the best estimates.

Probability Weighted Moments (PWM) is a more generalized version of the method of moments, where moments are weighted by the cumulative distribution function (CDF). It is particularly useful in modeling extreme events, as it can better capture the behavior of the tails of the distribution.

The maximum entropy method estimates parameters by maximizing the entropy subject to known constraints (e.g., the mean, variance). This method is based on the principle of making the least biased estimate, assuming as little as possible about the underlying distribution.

Mixed moments combine the concepts of both raw moments and weighted moments. These are used when both theoretical and empirical data need to be considered in estimating parameters. Individual Probability Weighted Moments is a variant of PWM where individual weighted moments are calculated for specific probability levels, which can be useful for distributions that exhibit significant skewness or heavy tails.

3.1.3. Goodness of fit tests

Several goodness of fit tests are available in the literature, though many may not be suitable for hydrological series. A straightforward and reliable method to assess whether an observed data set aligns with a theoretical probability distribution is by comparing the cumulative distribution of the observed data with the cumulative intensity function of the proposed distribution. If the two functions align closely, the theoretical distribution is considered a good fit for the data [25].

The Kolmogorov-Smirnov (K-S) test is a commonly used appropriateness test that can only be applied to continuous random variables. It determines if the distribution is a reasonable fit by comparing the observed and expected distributions, accepting or rejecting the hypothesis based on the test statistic.

The Chi-square (χ^2) test is another statistical test that can be applied to both continuous and discrete random variables. Unlike the K-S test, it compares the probability density functions instead of the cumulative intensity functions.

The Cramér-von-Mises (CvM) test compares the total probability distribution of the observed data with the theoretical frequency curve. It is considered reliable, especially with sample sizes around 20 cases. However, for smaller sample sizes, both the K-S and Chi-square tests may not be as robust in accurately accepting or rejecting hypotheses when the data is incorrect.

3.2. Synthetic flood estimation methods

When sufficient flow data is available for a river basin, statistical methods are generally effective in estimating flood discharges. However, many catchment areas lack available rainfall-runoff data necessary to derive unit hydrographs. In cases where rainfall-runoff data is inadequate or absent, various synthetic unit hydrograph methods have been developed.

3.2.1. Snyder method

When sufficient flow data is available for a river basin, statistical methods are generally effective in estimating flood discharges. However, many catchment areas lack available rainfall-runoff data necessary to derive unit hydrographs. In cases where rainfall-runoff data is inadequate or absent, various synthetic unit hydrograph methods have been developed [26].

In basins without rainfall and runoff records, the unit hydrographs are derived using various physical characteristics of the basin. The time difference between the centroid of the river basin and the peak of the flood hydrograph (tp) is calculated by the equation:

tp = 0.75 x Ct x (L x Lc)x0,3 (25)

Where: Ct is the coefficient related to the basin's storage capacity and slope, L is the length of the basin (km), Lc is the longest distance from the centroid of the basin to its entry or exit point (km).

The value of Ct is taken as 1.2 for mountainous areas, 0.72 for flat areas, and 0.35 for valleys. The value of Cp, which is empirically derived, is given by:

Cp=0.89×Ct (26)

The rainfall duration (tr) for the unit hydrograph is calculated using:

tr=tp5.5tr (27)

The peak discharge qp is determined using:

qp =2760× Cp / tp (lt/sn/km2/cm) (28)

Finally, the peak discharge for the flood (Qp) is calculated as:

Qp = qp * A * 10-3 (m3/sn/cm) (29)

A is the basin area (km²). The calculated value is then multiplied by the 100-year rainfall height to obtain the 100-year flood discharge.

3.2.2. Kirpich method

The Kirpich method, also known as the triangular unit hydrograph method, is favored due to its simplicity and its close resemblance to the Snyder method [27]. The main formulas for the Kirpich method are as follows:

The effective rainfall duration (tf) for the flood hydrograph is calculated by:

tf= te / 2 + tp (30)

Where: te is the effective rainfall duration and is accepted as equal to trt_rtr, the rainfall duration, and tp is the time difference between the centroid of the catchment and the peak of the flood hydrograph.

For a uniform rainfall distribution, the peak discharge Q_p is calculated using:

Qp = k.A.ha /tf (31)

In this formula: A is the catchment area (km²),h_a is the 100-year rainfall height (cm),k is an empirically derived coefficient.

3.2.3. Mockus method

The Mockus method is often preferred due to its practicality and the ease of drawing triangular hydrographs. Triangular hydrographs yield results comparable to more complex curve hydrographs, especially when the tail of the hydrograph does not significantly influence the design. The Mockus method can be applied to drainage areas with a concentration time (tct) of up to 30 hours, and for larger areas, the basin is subdivided and hydrographs are superimposed based on the delay times [28].

The unit rainfall duration (ΔD) is a critical parameter in flood estimation, calculated as:

$\Delta D=(t_c/5)$ (32)

The transition time t_c is calculated using:

$tc=0,00032*(L_h\ 0,77/S0,385)$ (33)

The flood attenuation time is calculated using the equation (34):

$tr=H_c \times tp$ (34)

Here, H_c is an empirical coefficient that varies between 1 and 2 based on the basin characteristics, and it is generally accepted as 1.60.

The flow caused by a 1 mm rainfall is determined using equation (35):

$qp=(K \times A_t)/ 2$ (35)

Here, K represents the basin coefficient, and it ranges between 0.21 and 1.60. For this basin, this value is taken as 1. [4]

The obtained qpq_pqp value is then multiplied by the 100-year maximum rainfall height (hah_aha) as shown in equation (36) to find the 100-year flood discharge (Qp):

$Q_p=q_p \times h_a$ (36)

3.2.34. S.C.S. method

The S.C.S. method is frequently used for planning, designing, and managing water resources in basins smaller than 30 km². It is particularly useful for estimating runoff and peak discharge values, which are generally needed for simulation models. The method is preferred for its simplicity, relationship to physical basin characteristics, and accurate results[29]. The formulas used in the SCS method are as follows:

The transition time tct_ctc is calculated as:

$t_c = 0.066 * (L_{h2} / S)\ 0.385$ (hour) (37)

Lh is the hydraulic length of the drainage area (km), S is the slope of the drainage area (%).

The total rainfall duration D is calculated as:

$D = 0.133 \times t_c$ (38)

The basin delay time L is derived using:

$L = 0.6 \times t_c$ (hour) (39)

The time to peak discharge t_p is:

$t_p = (D / 2) + L$ (hour) (40)

The runoff curve number (CN) is used to calculate the maximum retention capacity S as:

S = (1.000 / CN) – 10 (mm) (41)

The peak discharge Q_p for the 100-year flood is then calculated as:

$Qp = (0.2083 \times A / t_p) * h_e$ [20] (m3/sn) (42)

Where A is the basin area (km²), and hehehe is the maximum peak flow height.

4. Results and Discussion

To estimate the probability of recurrence over a specified period, probability distribution models require particular statistical parameters. In this research, distributions for periods of 50 and 100 years were computed using a software tool developed by Haktanır [19].

The measured peakflow values at a station for each year (represented as C_{sx}) ranged from 0.10 to 1.21, with all values consistently greater than zero. These values followed a right-skewed distribution. The skewness and variation coefficients varied across different observations, which can be explained by differences in the physical properties of the watersheds, streambed characteristics, and climatic conditions.

The analysis revealed that, for various probability distribution models, the calculated current values for the same recurrence periods showed differences. Larger discrepancies were observed for longer return periods. Generally, for periods under 50 years, the differences between the calculated flow values were minimal, but these differences became more pronounced as the return period increased. For periods shorter than 100 years, the calculated current values were mainly influenced by the distributions employed. However, for return periods exceeding 100 years, Pearson Type-3 and General Extreme Value (GEV) distributions yielded the highest values, whereas the log-logistic distribution provided the lowest.

The Moments method produced reliable results for all series. The Maximum Likelihood method offered the best fit, particularly for Log Normal, Pearson- 3, and Gumbel distributions, surpassing Log-Logistic and Log-Pearson- 3. The Probability Weighted Moments method performed well with Pearson- 3 and GEV. On the other hand, the Maximum Entropy, Mixed Moments, and Individual Probability Weighted Moments methods were less effective than the others.

As a result, the goodness-of-fit tests were performed with 95% confidence limits. The accepted (suitable) and rejected (unsuitable) distributions based on the probability distributions are shown in Table 1. As seen in Table 1, according to all three goodness-of-fit tests, the distributions that best fit the station data are Pearson-3 and GEV.

Table 1. Evaluatıon of distributions as to goodness of fits

Distrubiton	χ^2 test	K-S test	CvM
Log-Normal 2	accepted	rejected	rejected
Gumbel	rejected	accepted	accepted
Pareto	rejected	rejected	rejected
Log-Logistic	accepted	rejected	rejected
Pearson 3	accepted	accepted	accepted
Log-Pearson 3	rejected	accepted	accepted
Log-Boughton	rejected	accepted	rejected
Wakeby	rejected	rejected	rejected
GEV	accepted	accepted	accepted
Weilbull	rejected	rejected	accepted

In conclusion, based on the Pearson Type III distribution, the flood discharges calculated for the 50-year and 100-year return periods were found to be 898.000 m3/sn m³/s and 1021 m³/s, respectively. For the GEV distribution, the flood discharges for the same return periods were calculated as 1171 m³/s and 1422 m³/s.

Using the Snyder method, the maximum flood discharge for a 50-year return period was found to be 1423 m³/s, and for a 100-year return period, it was 1928 m³/s. With the Kirpich method, the 100-year flood discharge was calculated to be 1034 m³/s, and the 50-year flood discharge was 1284 m³/s. The peak discharge estimates for 50 and 100 years using the Marcos method were calculated to be 1122. m³/s and 1681 m³/s, respectively. Finally, using the SCS method, the 50-year discharge was found to be 1432 m³/s, and the 100-year discharge was 1884. m³/s.

The 50-year and 100-year peak flow prediction values obtained in the study are presented in Table 2. The results of Pearson Type III and GEV distributions, which provided the best outcomes among the probability distributions, were considered, while the results of other distributions that did not pass the goodness-of-fit tests were excluded from the table.

Table 2. Peak Flow Values (m³/s) for 50 and 100-Year Return Periods (years)Obtained by Different Methods.

Return Period	Pearson- 3	GEV	Snyder	Kirpich	Marcos	S.C.S:
50	898.000	1171	1523	1034	1122	1391
100	1021	1422	1928	1284	1681	1884

As seen in the table, the flow predictions obtained through statistical methods are closest to the values provided by the Kirpich and Marcos methods. The flow-year graph derived from the station records also highlights the suitability of the Pearson Type III and Kirpich method results.

5. Conclusion

Hydrological modeling offers valuable opportunities for better understanding the natural processes occurring in watersheds, playing a crucial role in water resource management and planning. Despite the inherent uncertainties within these models, technological advancements in recent years have improved their accuracy. Particularly, the widespread use of satellite technologies and the expansion of observation networks have enhanced the models' reliability, making them more compatible with observed data. Additionally, the lack of river gauging stations (RGS) in certain basins makes hydrological models even more significant in understanding the spatial distribution of events like droughts and floods. In this context, hydrological modeling holds the capacity to effectively simulate these natural events.

Flow predictions are of great importance in rivers without gauging stations. This situation can pose challenges in various fields, including water resource management, flood risk assessment, and infrastructure planning. In areas where flow observation data is scarce, accurately predicting flood discharges and flow behaviors becomes critical. In such cases, flow predictions using rainfall data and various hydrological models are essential for efficient water resource management and mitigating flood risks. This approach provides valuable insights for water resource planners and engineers, contributing to the development of effective water management strategies.

Determining annual peak flow values is vital for flood risk management and infrastructure planning, as these values inform the design of water structures and the efficient use of water resources. Flow forecasting enables the calculation of annual peak flows even in areas without gauging stations, allowing for more effective flood risk management and water resource direction.

In this study, the success of synthetic methods was compared to statistical methods, and it was found that two synthetic methods, in particular, provided successful results for the specific river basin in question. In conclusion, the modeling studies carried out for the Demirci Stream Basin offer significant insights for climate-related and watershed infrastructure planning. Although the models produce varying results, they will prove essential for future planning, especially in regions lacking observational data.

Declaration of Conflict of Interests

The author declares that there is no conflict of interest. They have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1.] Singh, V. P., Hydrologic Modeling: Progress and Future Directions. Water Resources Management, 31 (2017) 2951–2973.

[2.] Yazdi, J., Sheikh Khozani, Z., & Rezaei, H. , A novel hybrid framework for flood hazard and risk mapping using machine learning algorithms and GIS. Journal of Environmental Management, 287 (2021) 112281.

[3.] Freire Diogo, A., & Antunes do Carmo, J., Peak flows and stormwater networks design—current and future management of urban surface watersheds. Water, 11 (2019) 759-772.

[4.] Johnston, R., & Smakhtin, V., Hydrological modeling of large river basins: how much is enough. Water resources management, 28 (2014) 2695-2730.

[5.] Büyükkaracığan, N., & Ödük, M. N. Application of artificial neural network modeling to monthly flows of gediz river. In International Conference on Electrical, Computer, Electronics and Communication Engineering, (2013) Amsterdam.

[6.] Awadallah,A.G. & Ayman G., Evolution of the Nile River drought risk based on the streamflow record at Aswan station, Egypt. Civil Engineering and Environmental Systems, 31 (2014) 260-269.

[7.] Fischer, S., & Schumann, A. H., Generation of type-specific synthetic design flood hydrographs. Hydrological Sciences Journal, 68 (2023) 982-997.

[8.] Beneyto, C., Aranda, J. Á., Benito, G., & Francés, F. New approach to estimate extreme flooding using continuous synthetic simulation supported by regional precipitation and non-systematic flood data. Water, 12 (2020). 3174.

[9.] Al-Humoud, J. M., & Esen, I. I. , Approximate methods for the estimation of Muskingum flood routing parameters. Water resources management, 20 (2006), 979-990.

[10.] Ünlü, M., Hydrogeomorphological Study of Demirci River Basin and Its Surrounding Area, PhD Thesis, Marmara University, (1998) 2-32.

[11.] ÇYGM, Gediz Basin Pollution Prevention Action Plan, Ministry of Environment and Urbanization, 2015, 8-10.

[12.] SU, Techical Report, Analysis Techniques: Flood Frequency Analysis, (2024) available: http://streamflow.engr.oregonstate.edu/analysis/floodfreq

[13.] Limpert, E., Stahel, W. A., & Abbt, M.., Log-normal distributions across the sciences: Keys and clues. BioScience, 51 (2001).) 341–352.

[14.] Satyanaga, A., Rahardjo, H., Zhai, Q., Moon, S. W., & Kim, J., Modelling particle-size distribution and estimation of soil–water characteristic curve utilizing modified lognormal distribution function. Geotechnical and Geological Engineering, 42 (2024). 1639-1657.

[15.] Gumbel, E.J., Les valeurs extrêmes des distributions statistiques. Ann. Inst. Henri Poincaré, 5 1935 115–158.

[16.] Okoth, J. M., Otim, D., & Kamalha, E, Distribution of floods frequency of Manafwa River, Uganda. East African Journal of Engineering, 7 (2024): 1-20.

[17.] Zamani, N. D., Badyalina, B., Abd Jalal, M. Z. H., Mohamad Khalid, R., Ya'acob, F. F., & Chang, K. L., 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 (2024) 105-115.

[18.] Karım, S., Das, S., & De, S. K., Statistical analysis of annual maxımum flows of a controlled river: a study on the Kangsabati Rıver, Indıa. Annals of the National Association of Geographers, India, 41 (2021) 234-255.

[19.] Shoukri,, M.M., Mian,and, I.U.M. & Tracy, D.S., Sampling properties of estimators of the Log-Logistic Distribution with application to Canadian Precipitation Data", The Canadian Journal of Statistics, ,. 16 (1988) 223–236.

[20.] Basso, S., Botter, G., Merz, R., & Miniussi, A., The PHYsically-based Extreme Value distribution of river flows. Environmental Research Letters, 16 (2021) 1-11.

[21.] Jamal, M., Ebrahimi, H., & Jahromi, H. M. , Effect of selecting the superior probability distribution in modifying streamflow drought index (SDI). Arabian Journal of Geosciences, 15 (2022) 785- 802.

[22.] Filatov, S. V., Poplevin, A. V., Levchenko, A. A., Lebedeva, E. V., & Mezhov-Deglin, L. P., The formation of Pareto distribution in tracer systems on the water surface. Results in Physics, 27 (2021) 104446.

[23.] Nguyen, V. T., & Papavassiliou, D. V., Velocity magnitude distribution for flow in porous media. Industrial & Engineering Chemistry Research, 60 (2021) 13979-13990.

[24.] Herrera, P. A., Marazuela, M. A., & Hofmann, T. , Parameter estimation and uncertainty analysis in hydrological modeling. Wiley Interdisciplinary Reviews: Water, 9 (2022) e1569.

[25.] Büyükkaracığan, N., & Ödük, M. N. ,Application of artificial neural network modeling to monthly flows of gediz stream. In International Conference on Electrical, Computer, Electronics and Communication Engineering, (2013), Amsterdam.

[26.] Barid, B., & Setyaningrum, D. N., Increasing surface impermeability influences peak discharge with the Snyder method in the Ngadipiro sub-watershed of the Bengawan Solo watershed. In E3S Web of Conferences 517 (2024) 04003.

[27.] Biantoro, A. W., Wahyudi, S. I., Ni'am, M. F., & Mahardika, A. G. Time of Concentration Estimated Using Some Methods and Application in The Ciliwung Watershed, Jakarta. International Journal of Advanced Research in Engineering Innovation, 4 (2022). 17-25.

[28.] Bantchina, B. B., & Gündoğdu, K. S., Watershed Characteristics and Synthetic Unit Hydrographs Determination using Geographical Information Systems. J. BIOL. ENVIRON. SCI, 15 (2021) 27-34.

[29.] Niyazi, B., Khan, A. A., Masoud, M., Elfeki, A., Basahi, J., & Zaidi, S.,. Optimum parametrization of the soil conservation service (SCS) method for simulating the hydrological response in arid basins. Geomatics, Natural Hazards and Risk, 13 (2022) 1482-1509.

How to Cite This Article

Büyükkaracığan, N., Estimation of River Peak Flows Using Different Hydrological Modeling Methods: A Case Study of the Demirci River Basin, Civil Engineering Beyond Limits, 1(2025),1991.
https://doi.org/10.36937/cebel.2025.1991