



Establishing regional intensity-duration-frequency (IDF) relationships by using the L-moment approach and genetically based techniques for the Euphrates-Tigris basin

Mehmet Ali Hınıs¹ · Kadri Yurekli² · Muberra Erdogan²

Received: 16 February 2023 / Accepted: 27 September 2023 / Published online: 18 October 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2023

Abstract

Regional intensity-duration-frequency (IDF) relationships for the Euphrates-Tigris basin were established using genetic programming (GP) and multi-gene genetic programming (MGGP). The regional homogeneity of the study area was provided with two sub-regions (SRI and SRII) using the L-moment method. Estimated intensity values for various recurrence periods from selected regional distributions, new IDF relationships were established through GP and MGGP approaches, and the successful results were compared with the results obtained from the distributions. In addition, the parameters of 11 empirical equations commonly used in the literature for rainfall intensities were determined according to particle swarm optimization (PSO), artificial bee colony (ABC), genetic algorithm (GA), and flow direction algorithm (FDA) optimization methods. The rainfall intensity results of both the new IDF equations established with GP and MGGP techniques and the highest-performing empirical equations showed that the closest findings to the data set from regional distributions were obtained with MGGP for SRI and GP for SRII.

1 Introduction

It is important to protect the natural resources that provide water, which is absolutely necessary for living things to survive, and to develop them in accordance with the purpose in order to benefit from these resources in the most effective way. In particular, the threat of human-induced pressures (such as global warming, uncontrolled water consumption, and pollutants) on natural water resources is felt much more today than in the past. Global warming which caused the disruption of the normal functioning of the hydrological cycle has led to a variety of hydro-meteorological events in the historical process. Due to this change, different parts

of the globe have been exposed to heavy rains (followed by floods) and drought (IPCC 2007; IPCC 2012; Zeder and Fischer 2020; Lestari et al. 2019). Moreover, increases in temperatures have also given rise to an upward change in global and continental flow amplitude (Labat et al. 2004; Zhang et al. 2009). Considering that hydro-meteorological events occur under the influence of many factors, it is necessary to predict what the future amounts of a hydro-meteorological variable would be in order to prevent possible loss of property and life. This is vital for a water-related structure in order to implement the most appropriate preparation plans (Anlı 2009). The estimation of the design value required for a hydraulic structure should be realized by frequency analysis of the reliable variables necessary for the purpose in question. This effort prevents water-related natural disasters as well as also enables optimal utilization of water resources (Yurekli et al. 2009). The main goal of frequency analysis is to reveal the theoretical probability distribution that best fits the available data. Thus, it would be ensured that the hydraulic structure to be planned would reliably fulfill the expected functions by the design value from the selected probability distributions (Yurekli 2022).

Extreme events are of great interest in the literature due to their key importance in atmospheric, climatic, and hydrological events (Coles 2001). Maximum rainfall data (MRD)

✉ Mehmet Ali Hınıs
mhınıs@aksaray.edu.tr

Kadri Yurekli
kadiyurekli@yahoo.com

Muberra Erdogan
muberra.erdogan@gop.edu.tr

¹ Department of Civil Engineering, Aksaray University, Aksaray 68100, Turkey

² Department of Biosystem Engineering, Gaziosmanpaşa University, Tokat, Turkey

in standard times has the most common use in estimating design value in the development of water resources and prevention of water-related natural disasters (Asikoglu and Benzedem 2007; Karahan and Ozkan 2012). In the construction of hydraulic structures, the successful estimation of the design criteria based on the MRD is important in eliminating the negativities that would put the relevant structure at risk. For this reason, it is of critical importance to correctly determine the theoretical probability distribution that would best follow the data, and the statistical approaches considered in the estimation of its parameters (Hosking and Wallis 1993).

Frequency analysis of hydrological data is carried out on a point and regional scale. Although the frequency analysis of the data of any site is easier than the regional one, the main handicap that is frequently encountered in this context is not being able to reach the desired data length and quality on a point basis. Because of such concerns, regional frequency analysis is preferred in terms of reliability. Regional frequency analysis is based on the basic assumption that the data of sites in a region that is considered homogeneous show statistically similar behavior (i.e., having the same frequency distribution). With this assumption, it is emphasized that the data of sites in the homogeneous region could be combined as a single data for regional frequency analysis. This allows for a more reliable estimation of the region (Cunnane 1989; Hosking and Wallis 1997; Anlı 2009; Yurekli et al. 2009).

The commonly used approach in the context of regional frequency analysis is the L-moment method (Hosking and Wallis 1997). The L-moment approach is still one of the most suitable and currently used methods in frequency analysis of hydro-meteorological data (e.g., Gocic et al. 2021; Haddad 2021; Hassan et al. 2021; Nain and Hooda 2021; Khan et al. 2020; Gado et al. 2021; Yurekli et al. 2021).

Achieving the intensity-duration-frequency (IDF) curve, which is a reference tool in the design of hydraulic structures, is of paramount importance. The IDF curve represents the relationship between the intensity, duration, and recurrence interval of rainfall. The recurrence interval of the IDF relationship is determined based on the theoretical probability distribution best fit to the relevant rainfall data (Dupont and Allen 2000). It is possible to find some efforts in the literature in obtaining IDF curves (Bell 1969; Chen 1983; Koutsoyiannis et al. 1998; Nhat et al. 2006; Raiford et al. 2007; Ouali and Cannon 2018; Okonkwo and Mba-jorgu 2010; Elsebaie 2012; Chang et al. 2013; Paola et al. 2014; Al-Wagdany 2020). With the effect of global climate change, the intensity-duration-frequency relations of maximum rainfalls have also changed (Fadhel et al. 2017; Gebru 2020). Therefore, the successful establishment of the IDF relationship would lead to the reliable formation of the long-term management strategies of the water structures in question. Mathematically, the IDF is defined as the relationship

between the intensity, duration, and probability of exceedance (or recurrence period) of the maximum rainfall event considered. The IDF relationship is defined by curves due to their ease of use, and both statistical and empirical approaches are used for this purpose. The weighting parameters in the IDF relationship often require mathematical transformations and/or statistical analysis, and generally, establishing which distribution fits the observation data better requires a large number of trials or the use of software developed for this purpose. Due to the time-consuming and difficulty of estimating the weight parameters with traditional approaches, researchers have been led to search for alternative approaches to obtaining the parameters in question easily and simply.

Apart from traditionally obtaining IDF equation parameters, different approaches have recently come to the fore in the literature. According to Karahan et al. (2007) and Basakın et al. (2021), while analyzing the IDF relationship with the genetic algorithm approach, Gorkemli et al. (2022) used artificial bee colony programming. On the other hand, Rasel and Islam (2015) and Elsebaie (2012) estimated the IDF parameters by using multiple nonlinear regressions. Agbazo et al. (2016) used the scaling methodology to derive the IDF relationships and compared the results with the empirical methods. Zakwan (2016) estimated the parameters of the IDF equations with an optimization approach and reported that the IDF curves obtained were more successful than the traditional multiple regression method.

Karahan (2012) applied a Particle Swarm Optimization (PSO) approach to model the IDF relationship to the data sets with a length of 50 years and 68 years. He implied that the length of the data sequences and the selection of the empirical IDF equation were influential on model performance. Depending on the number of parameters in the empirical IDF equations, it has been found that the PSO-based estimations are more successful than those of the formulas from the Genetic algorithm (GA) technique. Similarly, Citakoglu and Demir (2023) also used the PSO and GA optimization approaches for the calibration of the IDF equations. In addition, new IDF relations were established with the multi-gene genetic programming (MGGP). The results of their study pointed out that the MGGP approach had more successful performance than those of the PSO and GA methods. Farzin et al. (2022) used the Harris Hawk optimization algorithm to establish a hybrid with a bi-directional long-short-term model in predicting the groundwater table and drought analysis. The results of the study indicated that the hybrid model algorithm had more accuracy than the other considered simulating algorithms. Farzin and Anaraki (2021) considered the combination of a new optimization technique, the flower pollination algorithm based on the pollination behavior of flowers, and a hybrid least-squares support vector machine to evaluate the impact

of global climate change on flow and suspended sediment load. It was stated that this combination algorithm produced more successful results than other model algorithms used in the study. Karami et al. (2021) proposed the flow direction algorithm (FDA), a physics-based algorithm, in their study. Comparing the suggested FDA technique with other optimization algorithms, including the GA, PSO, artificial bee colony (ABC), gray wolf optimization (GWO), and whale optimization (WOA), the FDA, in solving challenging optimization problems, indicated superior performance. The first aim of this study is to investigate the application possibilities of L-moment methods for regional frequency analysis of annual maximum rainfall for different durations (0.5 h, 1 h, 2 h, 3 h, 4 h, 5 h, 6 h, 8 h, 12 h, 18 h, and 24 h) in the Euphrates-Tigris basin. Another purpose is to compare the results of the IDF relationships forming with genetic programming (GP) and multi-gene genetic programming (MGGP) with those of conventionally derived IDF equations whose parameters would be estimated by genetic algorithm (GA), artificial bee colony (ABC), particle swarm optimization (PSO), and Flow direction algorithm (FDA) approaches. In the context of the setup of the study, estimating the weighting parameters of the IDF relationships, which are widely accepted in the literature, with four different optimization methods and revealing their performance, as well as the formation of new IDF relationships with the GP and MGGP techniques constitute the originality of the study. On the other hand, from the point of view of the studied region, there is no study on this subject for the Euphrates-Tigris basin in the literature. Therefore, in this study, research that will be beneficial in the design of any hydrological structure planned to be built in the basin or measures to be taken against natural events such as floods and droughts that may occur in the future for the Euphrates-Tigris basin or similar basins has been presented and research that can shed light on such studies.

2 Material and method

2.1 Study area and data

The upper basin within the borders of Turkey of the Euphrates-Tigris rivers, which is the main water source of Mesopotamia, was chosen as the study area. The Euphrates-Tigris rivers, which form the two main tributaries of the basin, merge in Şatt-ül-Arab and pour into the sea in the Persian Gulf (Ozis and Ozdemir 2008). With the melting of the snow at the beginning of March, the discharge of the Euphrates River starts to increase and reaches its peak level in April. The river, whose discharge decreases gradually since May, has the lowest level in September (Degirmenci 2007). The Euphrates river, with a length of 2780 km and a catchment

area of 720,000 km², is formed by the merging of Karasu, Muratsuyu, and many small streams. The average annual water potential of the river, which has an average discharge of 909 m³/s, is approximately 34 billion m³ and 33 billion m³ of this amount is gained within the borders of Turkey (Müftüoğlu 1997; Akbaş, 2015). The Tigris river, with a total length of 1900 km, arises around the Hazar Lake, and its annual average flow is 360 m³/s. Its discharge rises to 2263 m³/s in February and drops to 55 m³/s in September. While Turkey's contribution to the annual water volume of the Tigris River is 51%, Iraq and Iran contribute 39% and 10%, respectively (Demir and Pamukçu 1996; Akbaş 2015).

In the Euphrates-Tigris basin, the Southeastern Anatolia Project (GAP), the most important regional development project of the Republic of Turkey, was realized. The project area covers 9 provinces, named Adıyaman, Batman, Diyarbakır, Gaziantep, Kilis, Mardin, Siirt, Şanlıurfa, and Şırnak located in the Euphrates-Tigris basin and the upper Mesopotamian plains. GAP is a multi-sectoral regional development project including all sectors related to development such as agriculture, industry, transportation, urban and rural infrastructure, health, and education, by utilizing the resources of the Southeastern Anatolia Region (Altınbilek 2004). With this project, 22 dams and 19 hydroelectric power plants were targeted in the Euphrates-Tigris basin (Kaygusuz 1999).

In the study area (the upper Euphrates-Tigris basin), the annual maximum rainfall series with the duration of 0.5 h, 1 h, 2 h, 3 h, 4 h, 5 h, 6 h, 8 h, 12 h, 18 h, and 24 h obtained from 18 rain gauge stations under the control of the General Directorate of State Meteorology was used as material to achieve the intended purpose. Some features of the related stations were given in Table 1, and their geographic locations were also shown in Fig. 1.

2.2 Regional frequency analysis

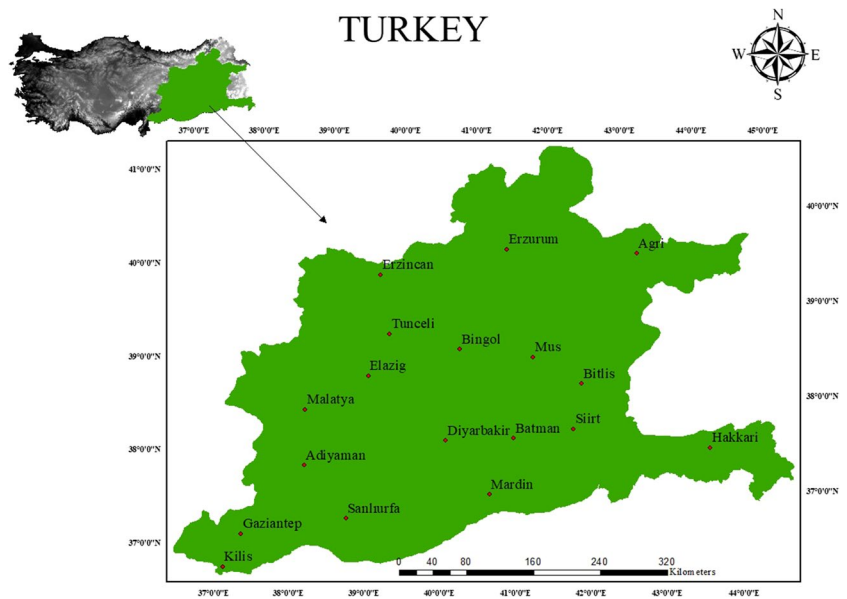
To form the regional and site-based intensity-duration-frequency curves of the maximum rainfall data with different durations obtained from 18 rainfall gauging stations, the regionalization algorithm based on L-moments, whose details were described by Hosking and Wallis (1997), was considered in the study.

The L-moment method has some advantages in the calculation of their parameters characterizing the theoretical probability distributions. These advantages are less sensitivity to outliers, and the ability to make reliable inferences about a theoretical probability distribution even under small sample size conditions (Hosking 1990; Park et al. 2001; Gubareva and Gartsman 2010). To describe L-moments, which are linear combinations of probability-weighted moments, let us take a random variable x and define its quantile function

Table 1 Some characteristics of the stations used in the study area

Sites	Longitude (E)	Latitude (N)	Elevation (m)	Mean (mm)	Data period
Adiyaman	38°17′	37°46′	669	713	1965–2020
Agri	43°03′	39°44′	1640	526	1967–2020
Batman	41°07′	37°52′	550	492	1969–2020
Bingol	40°29′	38°53′	1177	953	1966–2020
Bitlis	42°06′	38°22′	1545	1234	1966–2020
Diyarbakir	40°14′	37°55′	677	487	1942–2019
Elazig	39°14′	38°41′	1015	410	1957–2020
Erzincan	39°29′	39°44′	1214	380	1957–2020
Erzurum	41°17′	39°55′	1893	434	1956–2020
Gaziantep	37°22′	37°05′	840	554	1957–2020
Hakkari	43°45′	37°34′	1720	783	1965–2020
Kilis	37°05′	36°44′	680	498	1966–2020
Malatya	38°19′	38°21′	977	377	1958–2020
Mardin	40°44′	37°18′	1150	678	1966–2020
Mus	41°30′	38°44′	1300	765	1966–2020
Siirt	41°57′	37°56′	895	718	1959–2020
Sanliurfa	38°46′	37°08′	547	453	1959–2020
Tunceli	39°32′	39°07′	914	884	1968–2020

Fig. 1 The geographic position of the stations in the study area



as $x(u)$. The mathematical description of the L-moments related to the variable x in question is as follows.

$$\lambda_r = \int_0^1 x(u)P_{r-1}^*(u)du \tag{1}$$

In Eq. (1), $P_{r-1}^*(u)$ designates the r th shifted Legendre polynomial. Dimensionless L-moment ratios (τ_r) that describe the shape of a probability distribution represent the proportional relationship between the higher-order L-moment (λ_r) and the scale measure (λ_2). The L-moment ratios are as

$$\tau_r = \lambda_r/\lambda_2 \quad r \geq 3 \tag{2}$$

The first L-moment (λ_1) termed as a measure of location. The L-moment ratios denoted τ_3 and τ_4 correspond to the coefficient of L-skewness and coefficient of L-kurtosis. The ratio of the second L-moment (λ_2) to the first L-moment (λ_1) also specifies the coefficient of L-variation (τ). The sample L-moment ratios, abbreviated and symbolized as L-CV (τ), L-CS (τ_3), and L-CK (τ_4), are formulated mathematically as.

$$\tau = l_2/l_1, \tau_3 = l_3/l_2, \tau_4 = l_4/l_2 \tag{3}$$

In Eq. (3), “ l ” is symbolic of the sample L-moments.

Performing L-moment-based regional frequency analysis involves three stages. The first of these stages is to scrutinize the data in detail in terms of its suitability for analysis. The changes (homogeneity condition) during the process of obtaining the data that would affect the frequency analysis at any rainfall station should be closely examined. Homogeneity analysis is performed in order to detect the variability of statistical properties associated with the considered data over time due to natural or man-made reasons. The causes such as deterioration of the technical conditions required when installing the observation station, the relocation of the station, changes in the land use in the basin, natural disasters such as fire, and changes caused by human factors can be counted among these (Rougé et al. 2013; Belay et al. 2019). However, the fundamental assumption that the data is homogeneous is made when performing frequency analysis of hydrological data (Adeloye and Montaseri 2002; Fernando and Jayawardena 1994). In the context of performing a reliable analysis, with the suspicion that there may be variation in the measured records of a rainfall gauge station during the recording period, the inconsistency in rainfall data should be checked. The presence of inconsistency in a hydro-meteorological record could be detected using the widely preferred Mann–Whitney U non-parametric approach. Details on this approach are available in Nachar (2008). The basis of the approach is based on determining whether the difference between the medians of the two data sets is statistically significant. Its test statistic (Z_U) in Eq. (4) is compared with the critical table value (Z_{table}) for a 5% significance level of the standard normal distribution. The test statistic (Z_U) is mathematically formulated as follows:

$$Z_U = (U - \mu_U) / \sigma_U \tag{4}$$

where μ_U and σ_U are the mean and standard deviation associated with U distribution. These parameters are calculated according to the number of observations in the groups. In the calculation of the U parameter, the sum of the rank of each group is taken into account, as well as the number of observations of the groups. The null hypothesis (H_0) based on the assumption that the two data sets belong to the same population is either accepted or rejected according to the critical table value. If the relevant data has inhomogeneous, Belay et al. (2019) and Amjadi (2015) recommended the double-mass curve approach to eliminate the inconsistency in the data. In this study, the double-mass curve methodology was applied to the rainfall data sequences in which inhomogeneity was detected.

The other stage is the effort to assign sites (rainfall stations) to the region, where the frequency distributions of the sites are presumed to be approximately the same. The general recommendation to initially identify a tentative region

is to place sites into the relevant region(s) based on the characteristics belonging to the sites. For this purpose, cluster analysis, which is a very practical method to classify sites with similar characteristics in the study area into groups, is widely preferred (Modarres and Sarhadi 2010). Before performing the homogeneity analysis of the proposed tentative region according to the L-moment approach, bringing out the sites showing discordancy with the whole sites in the region is mandatory for the reliability of the frequency analysis. The detection of discordant sites is determined by the relationship between the L-moment ratios dealing with a site and the average L-moment ratios belonging to a group of similar sites (Rao and Hamed 2000; Šimková 2017). The discordancy measure (D_i) for a region with the N -sites has in the form of the equation given below:

$$D_i = \frac{1}{3} N (u_i - \bar{u})^T A^{-1} (u_i - \bar{u}) \tag{5}$$

$$A = \sum_{i=1}^N (u_i - \bar{u})(u_i - \bar{u})^T \tag{6}$$

In the equations, u_i is a vector consisting of the τ, τ_3 , and τ_4 associated with the site i , \bar{u} is the unweighted group average, and “ A ” denotes the covariance matrix of the example. To detect the discordant site(s) within the suggested region, the calculated D_i value for any site i is compared with the value of the D_{critic} . When $D_i > D_{critic}$, the site i is deemed to be discordant. The values of D_{critic} according to the number of sites exist in Hosking and Wallis (1997).

The homogeneity check of the tentatively selected region where there are no sites with discordancy is compulsory in terms of accepting that the sites in that region have similar frequency distribution. The realization of this check-in question is mostly by assigning the sites in the studied area to sub-regions. Acceptance of regional homogeneity for any region is achieved by applying the heterogeneity measure (H), which is given in Eq. (7) (Hosking and Wallis 1997).

$$H = \frac{(V - \mu_V)}{\sigma_V} \tag{7}$$

$$V = \left\{ \frac{\sum_{i=1}^N n_i (\tau^i - \tau^R)^2}{\sum_{i=1}^N n_i} \right\}^{\frac{1}{2}} \tag{8}$$

where n_i and τ^i are observation length and the coefficient of L-variation for site i , τ^R is the coefficient of L-variation for the relevant region. The μ_V and σ_V are the mean and standard deviation of the “ V ” values calculated based on the data estimated with the Monte Carlo simulation technique. The homogeneity of the formed region in terms of the calculated H value is decided according to three conditions that are classified as “acceptably homogeneous if $H < 1$, possibly

heterogeneous if $1 \leq H < 2$, and definitely heterogeneous if $H \geq 2$.”

The final stage is to bring out the regional frequency distribution that best represents the homogeneous region data. The selection of the regional distribution is based on the goodness-of-fit test, designated as Z^{DIST} . The analysis associated with Z^{DIST} is carried out in accordance with the difference between the L-kurtosis (τ_4^{DIST}) of relevant theoretical distribution and the regional average L-kurtosis (τ_4^R) of the sites in the region. Mathematically, the formulation of the goodness-of-fit test is as

$$Z^{DIST} = (\tau_4^{DIST} - \tau_4^R + \beta_4) / \sigma_4 \tag{9}$$

In the equation, DIST denominates the considered probability distribution. β_4 and σ_4 correspond to the bias and standard deviation that are calculated by the four-parameter Kappa distribution. The candidate distributions considered in the study are Generalized Logistic (GLOG), Generalized Extreme Values (GEV), Generalized Normal (GNO), Pearson Type III (PIII), and Generalized Pareto (GPA) distributions. Among these theoretical distributions, those that fulfill the condition of the $|Z^{DIST}| \leq 1.64$ are selected as a regional distribution. If this condition is detected in more than one theoretical distribution, the distribution having the smallest $|Z^{DIST}|$ value for the region should be picked up as the best fit.

The quantile estimation at the T return period for regional and site is realized according to the index-flood or index-storm approach, named with regard to flood and rainfall data, of Dalrymple (1960). The index-storm approach is based on the assumption that the data of sites in the statistically approved homogeneous region are similarly distributed apart from a site-specific scaling factor (Hosking and Wallis 1997). Its mathematically formulated form for the site i is given in Eq. (10).

$$Q_i(F) = \mu_i q(F) \tag{10}$$

In the equation, μ_i is the site-specific scaling factor of the site i , F is non-exceedance probability, and q is the value dealing with the growth factor.

2.3 Genetic algorithm

In order to find high-success solutions to optimization and search problems, genetic algorithm (GA) has been intensively dwelled on recently. The GA is a search heuristic approach inspired by the process of natural selection in which the ideal individuals are selected for reproduction to produce the next generation. The process of natural selection in the GA starts with the selection of the fittest individuals from a population corresponding to all of the individuals (chromosomes consisting of genes) who include possible

solution information. These selected individuals allow the production of new individuals bearing their characteristics and being added to the next generations. The process in question is kept on iterating until a generation is formed with the fittest individuals. Solving a problem with the GA is finalized in five steps, which are initial population, fitness function, selection, crossover, and mutation. When being focused on solving a problem in a genetic algorithm, the set of genes belonging to an individual (chromosome) is formed as a string. Usually, binary values (bits) consisting of ones and zeros are used. In other words, this activity is the process of encoding genes on chromosomes. Then, the competitiveness of an individual with other individuals is found out by the fitness function. According to the fitness score assigned to each individual, the probability of that individual being selected for reproduction is determined. The two individuals with the highest fitness score are selected to pass on their genes to the next generation. In the crossover being the most important step in a genetic algorithm, the crossover point is randomly determined among the genes for the selected pairs of individuals. Some of the genes of the new individuals formed in the crossover stage are subjected to a random mutation with low probability. This allows some bits in the bit string to be flipped. Details of the optimal solution process to a problem with the GA within these stages are available in Goldberg (1989).

2.4 The particle swarm optimization

Empirical IDF relationships (Yukseket al. 2022; Karahan et al. 2008; Basakin et al. 2021), whose mathematical formulations were given in Table 2, were considered in this study. The estimations of the parameters in these equations

Table 2 Rainfall duration and frequency-based empirical intensity equations

Equation cod	Equations	Parameters
EQ1	$I = \frac{a \cdot T^b}{(D+c)^d}$	a, b, c, d
EQ2	$I = \frac{a \cdot T^b}{D^c}$	a, b, c
EQ3	$I = \frac{a \cdot T^b}{(D^c+d)}$	a, b, c, d
EQ4	$I = \frac{(a+b \cdot \ln T)}{D^c}$	a, b, c
EQ5	$I = \frac{(a+b \cdot \ln T)}{(D^c+d)}$	a, b, c, d
EQ6	$I = \frac{(a+b \cdot \ln T)}{(D+c)^d}$	a, b, c, d
EQ7	$I = \frac{(a+b \cdot \ln(\ln T))}{(D+c)^d}$	a, b, c, d
EQ8	$I = \frac{(a+b \cdot \ln(\ln T))}{(D)^c}$	a, b, c
EQ9	$I = \frac{(a+b \cdot \ln(\ln T))}{(D^c+d)}$	a, b, c, d
EQ10	$I = \frac{(a+b \cdot \ln(T))+c[\ln(T)^2]}{d+e \cdot \ln(D)+f[\ln(D)^2]}$	a, b, c, d, e, f
EQ11	$I = \frac{a \cdot T^b}{(c+D^d)^f}$	a, b, c, d, e

were determined by optimization techniques such as particle swarm optimization (PSO), genetic algorithm (GA), artificial bee colony (ABC) algorithm, and flow direction algorithm (FDA).

I , rainfall intensity (mm/min); T , recurrence period (year); D , rainfall duration (min)

The PSO approach introduced by Kennedy and Eberhart (1995) is based on the principle that flocks of fish, birds, or insects share the experience of all their members to achieve their goals. In other words, in this optimization technique, swarm intelligence is brought to the fore. Social scientists have stated that the random movements of animals that move in herds, in situations such as food and safety, enable them to reach their goals more easily. Social knowledge sharing between individuals is fundamental in the PSO. Each individual is defined as a particle and a population of particles as a swarm. Each particle adjusts its position toward the best position in the swarm, taking advantage of its previous experience. The main target of the PSO is to approximate the position of the individuals in the herd to the best-positioned individual of the herd. A particle j in the D -dimensional search space is defined with three vectors, which are its position $[\vec{x}_j = (x_{j1}, x_{j2}, \dots, x_{jD})]$, velocity $[\vec{v}_j = (v_{j1}, v_{j2}, \dots, v_{jD})]$, and the best position that is experienced individually $[\vec{p}_j = (p_{j1}, p_{j2}, \dots, p_{jD})]$. In iterations of the algorithm associated with PSO, the current positions of these particles forming the swarm are analyzed as a solution to the problem. Thus, the herd at each time step is updated depending on the position and velocity of each individual. From the iterations, the best position (x_{pbest}), which corresponds to the local best for each individual, is estimated. Then, the global best (x_{gbest}) is brought out among the local bests. To carry out the PSO process, first, a starting swarm is created with randomly generated starting positions and velocities for each particle (Blackwell et al. 2007; Bratton and Kennedy 2007; Jain et al. 2018). The algorithm rules for the updated swarm, which is formed by updating the velocity and position of each individual or particle, are as follows:

$$v_{ij}^{k+1} = w * v_{ij}^k + C_p r_p (x_{(pbest)ij}^k - x_{ij}^k) + C_g r_g (x_{(gbest)ij}^k - x_{ij}^k) \tag{11}$$

$$x_{ij}^{k+1} = x_{ij}^k + v_{ij}^{k+1} \quad i = 1, 2, \dots, N, j = 1, 2, \dots, d \tag{12}$$

In Eqs. (11) and (12), k , C_p , and C_g are numbers of iteration and cognitive and social acceleration coefficients, respectively. N is the total number of variables, and r_p and r_g are numbers randomly estimated between 0 and 1. v_{ij}^k and x_{ij}^k correspond to the velocity and position of the j th particle of the i th variable, respectively, where $x_{(pbest)ij}^k$ is the best position of the j th particle of the i th

variable at the k th iteration. The $x_{(gbest)ij}^k$ denotes the global best value belonging to the i th variable, where w is the inertia weight value.

2.5 Optimization with artificial bee colony algorithm

The artificial bee colony algorithm (ABC) is a heuristic optimization approach such as PSO, inspired by the methods honey bees use when searching for food. This approach has recently gained widespread use in optimization problems. The sources that the bees go to in search of food represent the possible solutions of the problem to be solved in the algorithm, and the amount of nectar in the sources expresses the quality of the solution. In the ABC algorithm, there are three types of bees in a colony: employed bees, onlooker bees, and scout bees (Karaboga et al. 2014). The ABC optimization approach tries to reach the most optimal one among the possible solutions for the problem by finding the source with the most nectar. More detailed information about the ABC optimization process can be obtained from Karaboga (2005) and Akay (2009). Depending on the three bee groups in the colony, the process is briefly summarized below.

- At the beginning of the foraging process, scout bees try to find food by searching randomly in the environment.
- After the food sources are found, the scout bees become the employed bees and carry nectar from the source they find to the hive. These bees return to the source after emptying the nectar they carry or transfer the information they have about the source to the onlooker bees with the dance they perform in the dance area. In the event that a food source (a solution) is not improved with a predetermined number of trials, called “limit” which is a necessary control parameter in the ABC, the solution (food source) is depleted by its employed bee, the employed bee becomes a scout bee and seeks a new resource.
- Onlooker bees in the hive watch dances that indicate rich sources and prefer a source depending on the dance frequency proportional to the quality of the food.

Considering the search space directed to solve a problem as a hive environment containing food sources, the ABC algorithm starts by generating random food source locations corresponding to the solutions in the search space. This is illustrated below in a mathematical function.

$$x_{ij} = x_j^{\min} + \text{rand}(0, 1)(x_j^{\max} - x_j^{\min}), i = 1 \dots SN, j = 1 \dots D \tag{13}$$

“ SN ” is the number of food sources, while “ D ” is the number of parameters to be optimized. x_j^{\max} and x_j^{\min} respectively correspond to the upper and lower limits of the

parameter j . The “rand(0,1)” represents the generated number between 0 and 1.

Employed bees identify a food source in the neighborhood where the food source is located in their memory and evaluate its quality. When the information that the quality of the food source is more satisfactory is reached, the new source is stored in the memory. In the ABC algorithm, this situation is formulated as follows.

$$v_{ij} = x_{ij} + \varphi_{ij}(x_{ij} - x_{kj}) \tag{14}$$

In the equation, v_i and x_i are the new source and the source in the memory of the employed bee, respectively. “ j ” is a random number generated in the range of $[1, D]$, and φ_{ij} is a random number generated in the range of -1 and 1 . x_k is a source randomly formed in the neighborhood of the current resource. The fitness value of the new source is determined by the following equation.

$$\text{fitness}_i = \begin{cases} 1/(1 + f_i) & f_i \geq 0 \\ 1 + \text{abs}(f_i) & f_i < 0 \end{cases} \tag{15}$$

where f_i is the quality of the neighboring source solution, that is, the objective function value. When choosing between the current source and the neighboring source, the greedy selection method is applied according to the fitness value. When all the employed bees return to the hive after completing their research, they convey information to the onlooker bees about the nectar amount of the sources. In light of this information, the onlooker bee chooses a source with a probability proportional to the amount of nectar in the source. This probabilistic selection process is performed according to the fitness value corresponding to the amount of nectar in the algorithm. In other words, the ratio of the fitness value of a resource to the cumulative fitness value of all resources defines the relative probability of the source in question being selected compared to other sources (Eq. (16)).

$$P_i = \frac{\text{fitness}_i}{\sum_{i=1}^{SN} \text{fitness}_i} \tag{16}$$

where “fitness _{i} ” denotes the quality of the source i , and SN indicates the number of employed bees. In the algorithm, random numbers are generated in the range of 0 and 1, taking into account the calculated probability values of each source. If this value is greater than the probability value, onlooker bees are included in Eq. (14) and the following process. These stages in the algorithm continue until a pre-defined criterion or the maximum number of iterations is reached and the algorithm is terminated.

2.6 Flow direction algorithm

The flow direction algorithm (FDA) introduced by Karami et al. (2021) is a physics-based optimization algorithm for

solving global optimization problems. It simulates the direction of flow (runoff) to the drainage basin outlet point with the lowest height. Based on the aspect slope, the runoff flows to the outlet basin. By creating different cells within the basin, this movement is simulated. Based on the height and slope of its surrounding cells, each cell transfers the amount of runoff to the others. The difference between each cell’s height and distance from neighboring cells determines the direction of flow. The flow then moves to the cell with the highest slope after each cell’s slope has been calculated. After determining the direction of flow for the entire basin, each cell is given a value equal to the number of cells entering that cell.

The catchment exit point is, therefore, given the highest number. Additionally, a cell is said to have a hole and needs to be filled if its height is lower than that of its adjacent cells. Drainage basin is considered as problem search space. Flowing flow to a lower altitude outlet point is aiming to achieve the optimum answer. There are a number of predetermined positions which have height or objective functions around each flow. The slope affects the flow movement velocity, which is directed toward the lowest height. The neighbor cell with the least objective function is where the flow is moving at the fastest rate. Additionally, no neighbor’s objective function is allowed to be less than a flow, which is comparable to how a sink fills up to determine the flow direction. If the objective function of the flow is less than that of the present flow, it will move in the same direction to escape local optima in FDA; if not, it will move in the direction of the dominating slope.

Based on the slope to each individual’s surrounding neighbors, the velocity of each individual is updated. The slope to the neighbor is directly related to the separation (difference in positions) from the neighbors and the difference in the objective functions. Consequently, the objective function has an effect on the updating velocity of the individual in addition to position. The position of the initial flow is determined based on the following equation:

$$\text{FlowX}(i) = lb + rnd * (ub - lb) \tag{17}$$

where $\text{FlowX}(i)$ is the position of the i -th flow; lb and ub are the lower and upper limits of the decision variables; and rnd is the random number with uniform distribution in the range of $[0,1]$. The neighbor flow position is built around each flow with the following relationship:

$$\text{NeighborX}(j) = \text{FlowX}(i) + rndn * \Delta \tag{18}$$

where $\text{NeighborX}(j)$ is the j -th neighbor position; $rndn$ is a random value with a normal distribution in the range of $[0,1]$.

$$\Delta = [\text{rand} * Xrand - \text{rand} * \text{FlowX}(i)] * \|\text{Best}_X - \text{FlowX}(i)\| * W \tag{19}$$

where rand is in the range of $[0,1]$, Best_X is the global optimal solution, $Xrand$ is a random position, and W is the

nonlinear weight. Equation (19) shows that $\text{Flow}X(i)$ moves to a random position (X_{rand}). The second term shows by increasing iteration; $\text{Flow}X(i)$ is close to Best_X and the Euclidian distance between Best_X and $\text{Flow}X(i)$.

$$W = \left[\left(1 - \frac{\text{iter}}{\text{Max_iter}} \right)^{2 * \text{randn}} \right] * \left(\overline{\text{rand}} * \frac{\text{iter}}{\text{Max_iter}} \right) * \overline{\text{rand}} \quad (20)$$

where iter is the current iteration, Max_iter is the global iteration, and $\overline{\text{rand}}$ is a random vector with uniform distribution. W has a large variation that guaranties the escaping from local optimum in FDA. The new position of the flow is calculated by the following relationship:

$$\text{Flow_new}X(i) = \text{Flow}X(i) + V \frac{\text{Flow}X(i) - \text{Neighbor}X(j)}{\|\text{Flow}X(i) - \text{Neighbor}X(j)\|} \quad (21)$$

where $\text{Flow_new}X(i)$ shows the i -th new flow position. V is the velocity of the flow that moves to the neighbor with the least objective function and is related to its slope. V is given by Eq. (22) as follows:

$$V = \text{rand} * S_0 \quad (22)$$

$$S_0(i, j, d) = \frac{\text{Flow_fitness}(i) - \text{Neighbor_fitness}(i)}{\|\text{Flow}X(i, d) - \text{Neighbor_fitnes}(j, d)\|} \quad (23)$$

where $\text{Flow_fitness}(i)$ and $\text{Neighbor_fitness}(j)$ represent different values of the i -th flow and the j -th neighbor, respectively; d indicates the dimensions of the problem.

The flow will move to the r -th flow if the fitness of the r -th flow is less than the fitness of the current flow. Karami et al. (2021) provide more comprehensive details on the FDA.

2.7 Genetic programming

Genetic programming (GP) is an evolutionary programming technique that is accepted as an extension of genetic algorithms, formed from the building blocks of the problem at hand and aimed to optimize by evolving according to a certain adaptation criterion of the possible primitive solution styles (Koza 1992). The basis of the approach was laid based on “tree-based genetic programming” in 1985 by Michael L. Cramer. Then, it was developed by Koza (1992) for use in optimization and search problems. The aim of GP is to produce mathematical functions that will serve a certain purpose by using evolutionary processes. Genetic programming is a sub-branch of genetic algorithms, which encode a potential solution to a particular problem on a simple chromosome-like data structure and applies recombination operators to these structures in a way that preserves critical information. The representation of chromosomes in GP

is in the form of a tree. There are operators in the nodes of the tree and terminals in their leaves. For programming languages, operators can be commands and terminals can be parameters or variables. The GP forms a mathematical model in three basic stages. These are the beginning population creation, crossover, and mutation stages. The steps to be followed to develop a program code with the GP is as follows (Riccardo et al. 2008).

- A random society is produced. Individuals in society should be in a tree structure. Tree nodes should consist of functions, and leaves should consist of terminals.
- All programs in the society are compared, and their performance values are calculated.
- A new society (reproduction) is formed using crossover and mutation operations.
- The best of the existing programs in any generation is determined as the result of genetic programming.

In the present study, the multi-gene genetic programming (MGPP) technique, being a new variant of the traditional GP outlined above, was taken into consideration as an alternative approach to comparing the success of the GP. The basis of MGPP is based on linearly combining low-depth GP trees (gene) in order to improve the performance of the traditional GP approach (that is, to produce the most approximate results to the available data). That is, the solution in the MGPP is a weighted linear combination of the outputs from a number of GP trees, where each tree corresponds to a “gene,” whereas the GP evolves a population of trees to solve the problem. The MGPP enables linear combinations of nonlinear transformations of the input variables. By limiting the GP tree depth, transformations are forced to be low-order. This activation, in contrast to that of GP, allows the development of accurate, relatively compact mathematical models of predictor-response (input-output) datasets. This possibility is even provided in the case of a large number of input variables. A detailed description of MGPP can be found in Searson et al. (2010) and Gandomi and Alavi (2012). Hinchliffe et al. (1996) stated that the MGPP could be more accurate and computationally efficient than the GP approach.

In the study, the deviation between the maximum rainfall amounts (quantiles) to be estimated from the regional distribution and the maximum rainfall amounts to be obtained from the 11 empirical equations whose parameters would be calculated by optimization approaches PSO, ABC, and GA will be analyzed according to some error metrics. These are the correlation coefficient (R), Nash–Sutcliffe efficiency (NSE), root mean square error (RMSE), mean percentage error (MPE), mean absolute percentage error (MAPE), mean absolute error (MAE), and mean squared error (MSE). On the other hand, these error metrics were also used to compare

Table 3 Discordancy results (D_j) of the sites in choice of the two sub-regions

Region	Site	Rainfall durations in hours											
		0.5 h	1 h	2 h	3 h	4 h	5 h	6 h	8 h	12 h	18 h	24 h	
SR ₁	Adiyaman	0.69	1.18	0.36	0.25	0.60	0.93	0.66	0.26	0.35	1.71	1.78	
	Bingol	1.70	1.84	1.06	1.51	1.09	0.60	1.07	1.32	0.27	0.91	0.85	
	Bitlis	0.50	2.35	0.48	0.81	1.30	1.60	1.63	1.41	1.21	1.91	1.75	
	Elazig	1.22	0.26	0.43	1.19	0.66	0.19	0.65	0.66	1.33	0.60	0.32	
	Erzincan	1.48	1.38	1.62	1.86	1.95	0.88	1.00	0.26	1.22	1.29	0.36	
	Erzurum	1.12	0.17	0.28	0.38	0.54	0.64	1.09	1.28	1.24	0.29	0.34	
	Malatya	0.56	0.28	0.46	0.71	1.11	1.86	1.59	1.67	0.52	0.30	1.68	
	Siirt	1.25	0.65	1.64	1.63	0.60	0.35	0.92	0.82	1.14	–	–	
	Tunceli	0.48	1.32	1.48	0.68	1.15	1.94	0.39	1.32	1.71	–	–	
Hakkari	–	0.55	2.19	–	–	–	–	–	–	–	0.92		
H-Measure		0.06	0.97	0.70	0.54	0.00	0.44	0.24	0.72	0.03	0.09	0.51	
SR ₂	Agri	1.51	1.23	0.90	1.03	1.38	1.45	1.67	1.83	1.87	1.72	0.78	
	Batman	0.55	1.50	0.99	0.79	1.56	0.88	1.35	1.07	1.06	1.58	0.82	
	Diyarbakir	0.71	1.03	0.57	0.42	0.75	0.88	0.41	0.28	0.10	0.10	0.97	
	Gaziantep	1.14	0.93	0.49	0.82	0.62	0.60	0.34	0.64	0.09	0.24	0.76	
	Kilis	0.35	0.52	1.09	1.12	0.78	1.14	1.07	0.60	1.30	0.99	1.24	
	Mus	1.15	1.45	1.61	1.62	1.03	0.72	0.61	0.94	1.31	1.31	1.10	
	Sanliurfa	1.58	0.34	1.36	1.19	0.88	1.34	1.54	1.65	1.27	1.40	1.01	
	Siirt	–	–	–	–	–	–	–	–	–	0.66	1.34	
H-Measure		0.36	0.95	0.95	0.28	0.37	0.07	0.29	0.25	0.56	0.10	0.81	

Values in bold indicate homogeneity is acceptable since H-Measure<1

the maximum rainfall amounts estimated from the new IDF relationships formed with GP and MGGP with the maximum rainfall amounts calculated from the regional distribution and 11 empirical relationships based on the optimization methods mentioned above.

3 Results and discussion

The main aim of this study was to establish regional and site-based intensity–duration–frequency relationships for the annual maximum rainfall series with the duration of 0.5 h, 1 h, 2 h, 3 h, 4 h, 5 h, 6 h, 8 h, 12 h, 18 h, and 24 h belonging to 18 precipitation stations in the Euphrates-Tigris basin. In line with this goal, before starting the analysis, the missing rain amounts in the maximum rainfall series of 11 different durations of each rainfall station were completed with the normal ratio method. Then, the homogeneity of all rainfall series was statistically checked with the Mann–Whitney U (MWU) approximation, which was detailed by Yurekli

(2015). The maximum rainfall series at some of the considered durations in the study for the 6 rainfall stations could not satisfy the homogeneity condition. The homogeneity of these series in question has been made homogeneous with the double-mass curve approach.

After the existing data has been brought into a form that could be used in the analysis in order to realize the regionalization process based on L-moments, the first attempt for this purpose was to calculate L-moments and rates belonging to maximum rainfall sequences of all durations. The prevalent choice that comes to mind in the regionalization process of hydro-meteorological data sequences has been to investigate the possibility of whether the entire region selected as the study area could statistically satisfy the homogeneous region conditions. It was focused on the case of forming a single region of the rainfall series for each duration in the study. Discordancy was not detected in all maximum rainfall data series with the remaining durations except those of the 0.5-h duration for the Mardin site and the 18-h duration for the Tunceli site. However, none of the rainfall

Table 4 Goodness-of-fit results for the sub-regions based on the Z^{DIST} criterion

Region	Rainfall durations in hours											
	0.5 h	1 h	2 h	3 h	4 h	5 h	6 h	8 h	12 h	18 h	24 h	
SR1	GEV	GEV	GEV	GLOG	GLOG	GLOG	GEV	GEV	GNO	GNO	GLOG	
SR2	GNO	PIII	GEV	GEV	GNO	GNO	GNO	PIII	PIII	GNO	GEV	

series fulfilled the regional homogeneity condition according to the heterogeneity criterion. According to this finding, the studied area should be divided into the sub-regions. As a first alternative, the option of dividing the entire region into two sub-regions called SR₁ and SR₂, based on cluster analysis, was considered. In this preference, it was decided to exclude the Mardin site from the analysis because it

brought about either discordancy or heterogeneity in the sub-region to which it was assigned. Similar findings were also detected in the maximum rainfall series (excluding data of the 1-, 2-, and 24-h durations) belonging to the Hakkari site (Table 3). On the other hand, the maximum rainfall series with the 18-h duration for Siirt and Tunceli sites failed to fulfill the necessary conditions in terms of discordancy and

Table 5 Comparisons of the results of PSO, GA, ABC, and FDA algorithms for SRI and SR II regions

		SR I (sub-region I)						
		Correlation	NS	RMSE	MPE	MAPE	MAE	MSE
EQ1	PSO	0.9900	0.9753	4.3309	0.3734	0.4488	3.5716	18.7570
EQ2		0.9900	0.9753	4.3309	0.3734	0.4488	3.5717	18.7570
EQ3		0.9900	0.9753	4.3309	0.3734	0.4488	3.5717	18.7570
EQ4		0.9920	0.9787	4.0188	0.3923	0.4481	3.2792	16.1511
EQ5		0.9994	0.9989	0.9244	0.0279	0.0815	0.6923	0.8546
EQ6		0.9920	0.9787	4.0188	0.3923	0.4481	3.2792	16.1511
EQ7		0.9744	0.9441	6.5144	0.3912	0.4779	4.5700	42.4377
EQ8		0.9744	0.9441	6.5144	0.3912	0.4779	4.5700	42.4377
EQ9		0.9818	0.9639	5.2363	0.0571	0.1821	2.7273	27.4185
EQ10		0.9996	0.9992	0.8008	-0.0019	0.0885	0.6424	0.6414
EQ11		0.8956	0.5225	19.0435	0.6163	0.6412	9.3203	362.6532
EQ1	GA	0.9839	0.9592	5.5672	0.4788	0.5718	4.7366	30.9933
EQ2		0.9900	0.9753	4.3309	0.3734	0.4488	3.5717	18.7570
EQ3		0.9908	0.9729	4.5386	0.4541	0.5150	4.0303	20.5989
EQ4		0.9920	0.9787	4.0188	0.3923	0.4481	3.2792	16.1511
EQ5		0.9880	0.9748	4.3771	0.2488	0.3465	2.8926	19.1591
EQ6		0.9920	0.9787	4.0188	0.3923	0.4481	3.2793	16.1511
EQ7		0.9774	0.9457	6.4196	0.4587	0.5165	4.7121	41.2106
EQ8		0.9744	0.9441	6.5144	0.3912	0.4779	4.5700	42.4377
EQ9		0.9686	0.9329	7.1400	0.4101	0.5165	4.9300	50.9801
EQ10		0.9996	0.9992	0.8008	-0.0019	0.0885	0.6425	0.6414
EQ11		0.8740	0.5778	17.9057	0.7128	0.7676	9.6648	320.6147
EQ1	ABC	0.9900	0.9753	4.3309	0.3734	0.4488	3.5717	18.7570
EQ2		0.9900	0.9753	4.3309	0.3734	0.4488	3.5717	18.7570
EQ3		0.9900	0.9753	4.3309	0.3734	0.4488	3.5717	18.7570
EQ4		0.9920	0.9787	4.0188	0.3923	0.4481	3.2792	16.1511
EQ5		0.9920	0.9787	4.0188	0.3923	0.4481	3.2792	16.1511
EQ6		0.9920	0.9787	4.0188	0.3923	0.4481	3.2792	16.1511
EQ7		0.9744	0.9441	6.5144	0.3912	0.4779	4.5700	42.4377
EQ8		0.9744	0.9441	6.5144	0.3912	0.4779	4.5700	42.4377
EQ9		0.9794	0.9467	6.3595	-0.0099	0.2093	2.9080	40.4432
EQ10		0.9993	0.9985	1.0843	-0.0136	0.1044	0.7936	1.1756
EQ11		0.8956	0.5225	19.0422	0.6163	0.6412	9.3196	362.6043
EQ1	FDA	0.9900	0.9753	4.3309	0.3734	0.4488	3.5709	18.7570
EQ2		0.9900	0.9753	4.3309	0.3734	0.4488	3.5717	18.7570
EQ3		0.9900	0.9753	4.3309	0.3734	0.4488	3.5717	18.7570
EQ4		0.9920	0.9787	4.0188	0.3923	0.4481	3.2792	16.1511
EQ5		0.9920	0.9787	4.0188	0.3923	0.4481	3.2792	16.1511
EQ6		0.9920	0.9787	4.0188	0.3923	0.4481	3.2792	16.1511
EQ7		0.9744	0.9441	6.5144	0.3912	0.4779	4.5700	42.4377
EQ8		0.9744	0.9441	6.5144	0.3912	0.4779	4.5700	42.4377
EQ9		0.9744	0.9441	6.5144	0.3912	0.4779	4.5700	42.4377
EQ10		0.9996	0.9992	0.8008	-0.0018	0.0885	0.6426	0.6414
EQ11		0.8956	0.5225	19.0435	0.6163	0.6412	9.3203	362.6533

Table 5 (continued)

		SR II (sub-region 2)						
		Correlation	NS	RMSE	MPE	MAPE	MAE	MSE
EQ1	PSO	0.9881	0.9723	4.9148	0.3461	0.4216	3.7885	24.1553
EQ2		0.9881	0.9723	4.9148	0.3461	0.4216	3.7885	24.1553
EQ3		0.9881	0.9723	4.9148	0.3461	0.4216	3.7885	24.1553
EQ4		0.9913	0.9781	4.3662	0.3707	0.4212	3.3863	19.0638
EQ5		0.9991	0.9982	1.2603	0.0064	0.0911	0.9049	1.5883
EQ6		0.9913	0.9781	4.3662	0.3707	0.4212	3.3863	19.0638
EQ7		0.9744	0.9451	6.9208	0.3699	0.4493	4.7472	47.8970
EQ8		0.9744	0.9451	6.9208	0.3699	0.4493	4.7472	47.8970
EQ9		0.9820	0.9643	5.5790	0.0364	0.1906	2.8234	31.1249
EQ10		0.9989	0.9978	1.3780	-0.0300	0.1191	1.0130	1.8988
EQ11		0.8883	0.5005	20.8746	0.6029	0.6305	10.1988	435.7507
EQ1	GA	0.9881	0.9723	4.9148	0.3461	0.4216	3.7885	24.1553
EQ2		0.9881	0.9723	4.9148	0.3461	0.4216	3.7885	24.1553
EQ3		0.9855	0.9625	5.7205	0.4584	0.5329	4.7895	32.7239
EQ4		0.9913	0.9781	4.3662	0.3707	0.4212	3.3863	19.0638
EQ5		0.9881	0.9749	4.6749	0.2722	0.3532	3.2277	21.8547
EQ6		0.9913	0.9781	4.3662	0.3707	0.4212	3.3863	19.0638
EQ7		0.9744	0.9451	6.9208	0.3699	0.4493	4.7472	47.8970
EQ8		0.9744	0.9451	6.9208	0.3699	0.4493	4.7472	47.8970
EQ9		0.9643	0.9274	7.9580	0.3114	0.4421	4.9701	63.3300
EQ10		0.9989	0.9978	1.3756	-0.0327	0.1193	1.0114	1.8923
EQ11		0.8873	0.4637	21.6311	0.7472	0.7682	11.3446	467.9057
EQ1	ABC	0.9881	0.9723	4.9148	0.3461	0.4216	3.7892	24.1553
EQ2		0.9881	0.9723	4.9148	0.3461	0.4216	3.7892	24.1553
EQ3		0.9881	0.9723	4.9148	0.3461	0.4216	3.7892	24.1553
EQ4		0.9913	0.9781	4.3662	0.3707	0.4212	3.3863	19.0638
EQ5		0.9990	0.9980	1.3240	0.0258	0.0829	0.8694	1.7529
EQ6		0.9913	0.9781	4.3662	0.3707	0.4212	3.3863	19.0638
EQ7		0.9744	0.9451	6.9208	0.3699	0.4493	4.7472	47.8970
EQ8		0.9744	0.9451	6.9208	0.3699	0.4493	4.7472	47.8970
EQ9		0.9744	0.9451	6.9208	0.3699	0.4493	4.7472	47.8970
EQ10		0.9988	0.9975	1.4727	-0.0121	0.1327	1.1394	2.1688
EQ11		0.8882	0.5009	20.8672	0.6017	0.6296	10.1899	435.4384
EQ1	FDA	0.9894	0.9671	4.9962	0.2987	0.4086	3.4795	24.9619
EQ2		0.9894	0.9671	4.9962	0.2987	0.4086	3.4795	24.9619
EQ3		0.9894	0.9671	4.9962	0.2987	0.4086	3.4795	24.9619
EQ4		0.9913	0.9713	4.6700	0.3233	0.4128	3.3849	21.8089
EQ5		0.9913	0.9713	4.6700	0.3233	0.4128	3.3849	21.8089
EQ6		0.9913	0.9713	4.6700	0.3233	0.4128	3.3849	21.8089
EQ7		0.9736	0.9366	6.9396	0.3179	0.4501	4.7361	48.1582
EQ8		0.9736	0.9366	6.9396	0.3179	0.4501	4.7361	48.1582
EQ9		0.9736	0.9366	6.9396	0.3179	0.4501	4.7361	48.1582
EQ10		0.9994	0.9916	2.5238	-0.0823	0.1256	1.4451	6.3696
EQ11		0.8943	0.5472	18.5435	0.5800	0.6134	8.9666	343.8597

The most suitable model is indicated in bold

Table 6 The parameter values of the empirical equations selected for the sub-regions

Optimization method	Parameters of EQ10 for SR1						Parameters of EQ5 for SR2			
	a	b	c	d	e	f	a	b	c	d
PSO	174.5	96.6	3.00	20.1	24.1	5.94	2.71	2.60	0.44	-0.61
ABC	196.3	114.5	4.67	24.7	29.1	6.08	3.12	3.05	0.50	-0.56
GA	-184.0	-101.7	-3.20	-21.2	-25.4	-6.26	-	-	-	-
GA***	-132.3	-98.1	-1.61	-16.6	-19.7	-4.80				
FDA	181.964	100.67	3.12	20.95	25.14	6.196				
FDA***	164.64	122.02	1.999	20.64	24.52	5.97				

***Parameter values of EQ10 based on the GA and FDA approach for the SR2 region

heterogeneity in the first sub-region. The way of assigning these sites to the other sub-regions was performed. With this preference, only the Siirt site was covered in the sub-region, and regionalization conditions were satisfied. Tunceli station was out of operation. The 24-h duration maximum rainfall data sequence of the Tunceli site induced both sub-regions to assume a statistically heterogeneous character. Although the Siirt site did not pose any problem in terms of the regionalization process when it remained in the first sub-region, it was assigned to this region in order to ensure the homogeneity of the second sub-region. The final results of homogeneous regions for both 18-h and 24-h durations are given in Table 3. It is detectable in Table 3 that none of the sites in the two sub-regions formed for each rainfall duration in question shows discordancy with each other, and the results of the heterogeneity measure calculated for these regions emphasize acceptable homogeneity ($H < 1$). The discordancy measure (D_i) estimated for each site was smaller than the critical value ($D_{critical}$) determined according to the number of sites in each region (7, 8, 9, and 10 sites for this study). The $D_{critical}$ values are 1.92, 2.14, 2.33, and 2.49 for 7, 8, 9, and 10 sites, respectively.

The selection of the regional distribution that could best represent the data of each sub-region whose homogeneity was confirmed statistically was determined according to the Z^{DIST} value, which was the goodness-of-fit test. Although more than one theoretical distribution with a smaller value

of 1.64, which was the critical value of the goodness-of-fit test, was found for the sub-regions formed for each duration considered in the study, the distribution with the smallest Z^{DIST} value was chosen among them. Table 4 shows the results of goodness-of-fit tests. The quantile estimations corresponding to the recurrence periods of 2 years, 10 years, 20 years, 100 years, and 1000 years for the two sub-regions formed from annual maximum rainfall sequences with different durations were calculated based on the index-storm approach.

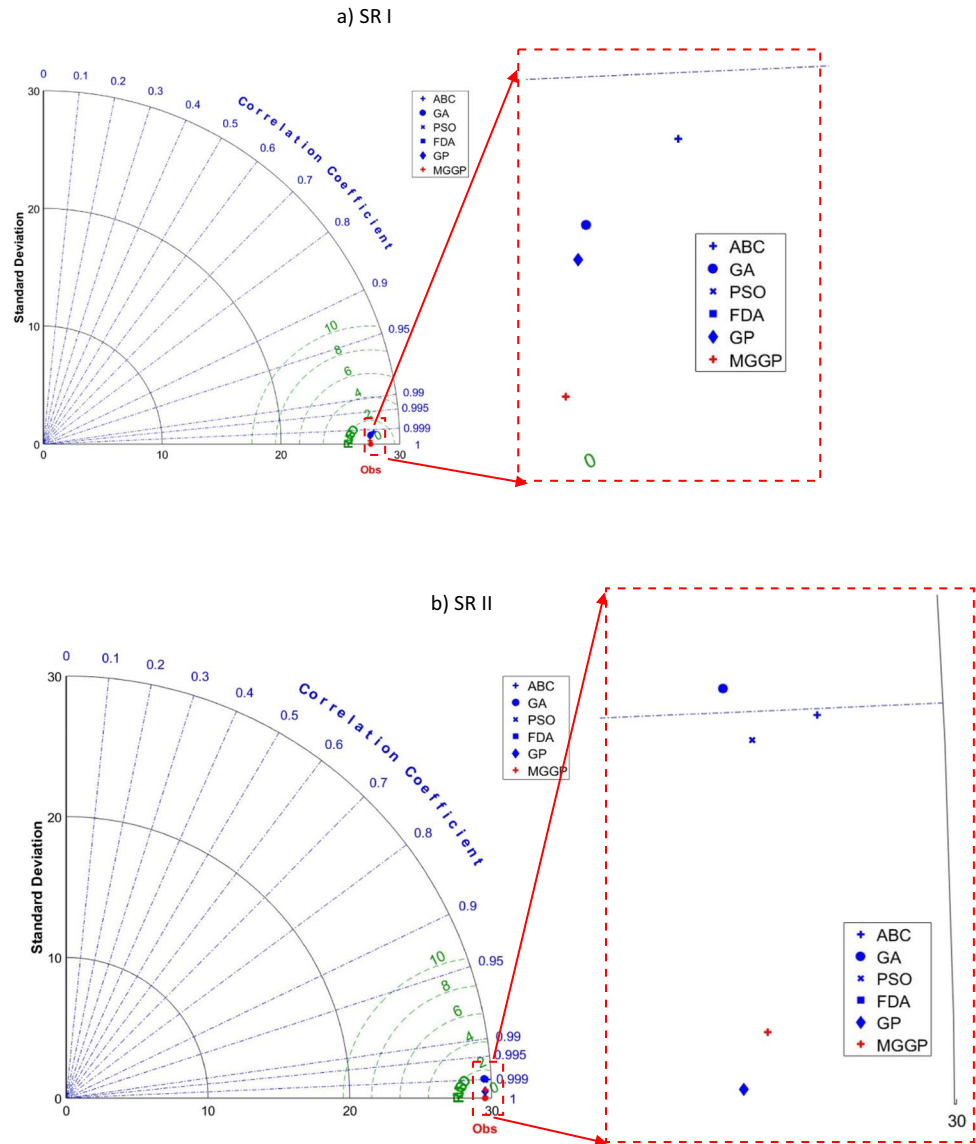
GLOG, generalized logistic; *GEV*, generalized extreme values; *GNO*, generalized normal; *PIII*, Pearson type III.

Considering the calculated quantile values for recurrence periods mentioned above, the parameters of the empirical equations in Table 2 were estimated by using the PSO, ABC, GA, and FDA methodologies. Based on seven error metrics, the most successful result for the SR1 sub-region was satisfied with the EQ10-coded relationship among the 11 empirical equations using the PSO, ABC, GA, and FDA optimization techniques. In this context, the most successful result for the SR2 sub-region was achieved in the EQ5-coded equation with PSO and ABC (only at EQ10 according to the MPE metric) optimization approaches. The results were given in Table 5. In the SR2, the GA approach provided the most successful result in the equation with code EQ10. FDA approach has the minimum MPE metric in the SR1 region. The parameter values of the empirical relationships with

Table 7 The new IDF relationships formed for the sub-regions through the GP and MGGP

Region	Genetic programming (GP)
SRI	$I = 3.62 + 0.98 * \ln(T) + \frac{(246.15+0.10*T+180.76*\ln(T))}{(D-19.28)} - 0.31 * \ln(D) - 0.00057 * D * \ln(T)$
SRII	$I = 1.39 + 0.000512 * T + \frac{(467.20+456.54*\ln(T))}{D} + \frac{-1.89E5*\ln(2.16+T)}{(D^3-1.83E3*D)}$
Region	Multi-gene genetic programming (MGGP)
SRI	$I = 0.00905 * D + 0.00452 \exp\left(D^{\left(\frac{1}{2}\right)}\right) - 0.696 * \log(D^3) + 0.159 * \log\left(\frac{T^2}{\log(D)}\right) - \frac{(0.36*\log(T))}{D^4} + \frac{(5.2*\log\left(\frac{T}{D}\right))}{D} + \frac{(0.101*\log\left(\frac{T}{D}\right))}{D^7} + \frac{0.00452*T}{D} + 7.94$
SRII	$I = \frac{(9.94*\log(T)+1.9*D*D^2+6.83*D*\log(T)-8.34*T*D^2*\exp(-1*D^2))*\log(T)}{(T*t^2)} + \frac{(6.67*\log(T))}{(D*\tanh(\tanh(D)))}$

Fig. 2 Taylor diagrams of best-fit equations for PSO, GA, ABC, FDA, GP, and MGGP approaches for (a) sub-region 1 (SR I) and (b) sub-region II (SR II)



the most successful results, the parameters of which were estimated according to the three optimization approaches, are available in Table 6.

Another focus of this study became to establish new IDF relationships as an alternative to the 11 empirical equations for homogeneous regions with GP and MGGP techniques. In Table 7, the new IDF relationships with the highest performance based on the above-mentioned error metrics among the models formed with GP and MGGP approaches were presented. For the SRI and SRII sub-homogeneous regions, the maximum gene numbers of the models (in Table 7) with the highest performance brought into existence by the MGGP technique were 6 and 4, respectively. The maximum gene depths for these models were also determined as 6 for sub-region I and 4 for sub-region II. In order to compare the best-fit equations of PSO, GA, ABC, FDA, GP, and MGGP approaches, the Taylor diagram was

used for evaluation. Results are presented in Fig. 2. As seen in Fig. 2, the results of the PSO, AG, ABC, FDA, GP, and MGGP models gave results very close to the observation values; among them, the GA, PSO, and FDA models for SR I overlap, and the model with the best results for SR I sub-region was MGGP; on the other hand, for the SR II region, GP and MGGP gave the best results by far different from the others, and the GP formula gave the best result. The violin plot was used to analyze how well the data estimated by PSO, GA, ABC, FDA, GP, and MGGP methods matched up with observed data. Using the violin plot, more statistical comparisons between the models were available. A violin plot for the best result of the PSO, GA, ABC, FDA, GP, and MGGP approaches is shown in Fig. 3. Violin plots for both SR I and SR II sub-regions revealed that results of all methods have similar distributions except for ABC algorithm.

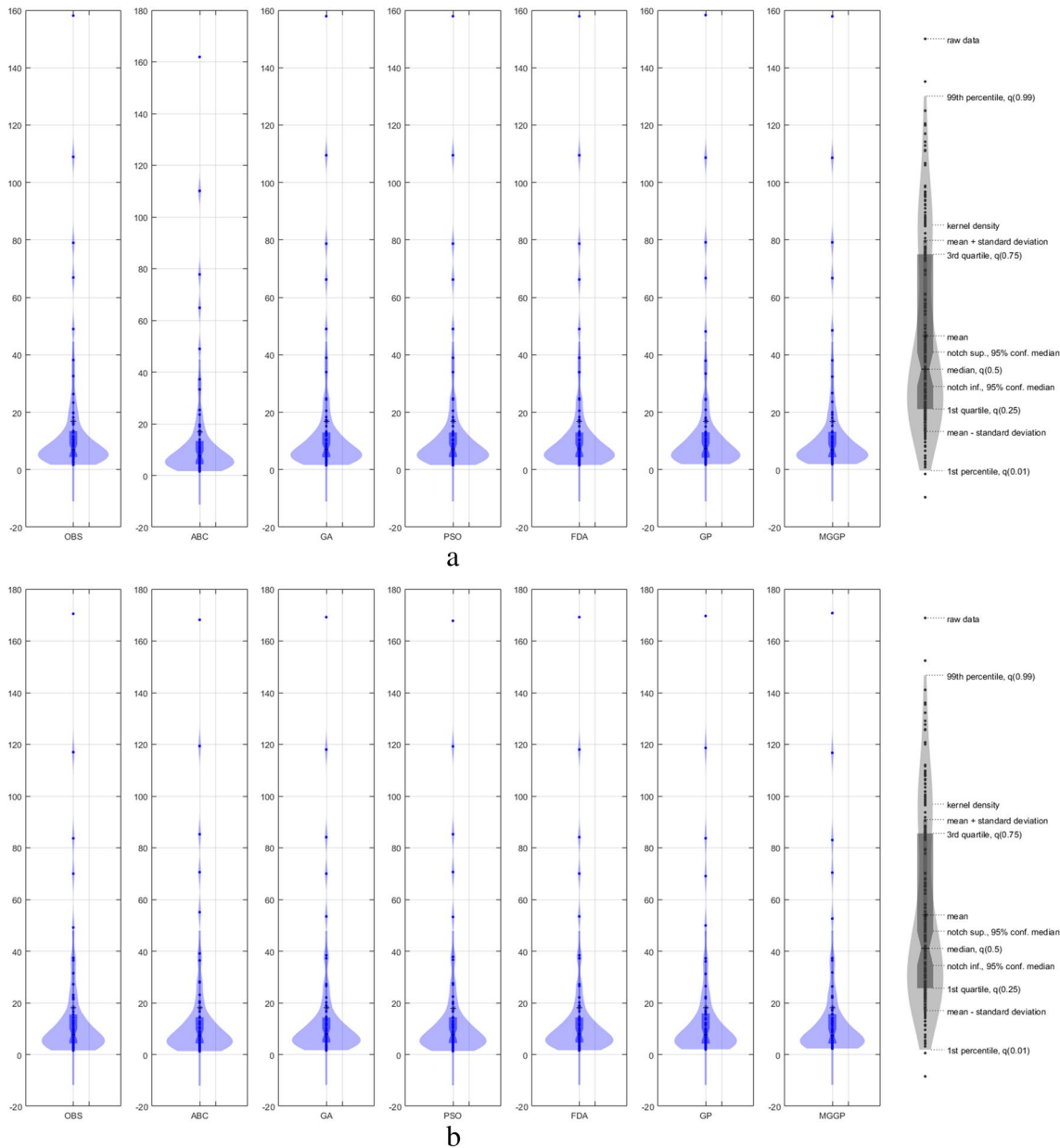


Fig. 3 Volin plot for (a) SR I sub-region and (b) SR II sub-regions

I , rainfall intensity (mm/min); T , recurrence period (year); D , rainfall duration (min)

The rainfall intensity data sets obtained from the GP and MGGP and empirical equations selected based on the PSO, ABC, GA, and FDA were first subjected to normality analysis to check whether these data groups came from the same population as those estimated from the regional distributions. It was concluded that all data sets are not normally distributed according to Kolmogorov-Smirnov and Shapiro-Wilk tests. The non-parametric Mann-Whitney U (MWU) test was used to check whether the rainfall intensity amounts estimated from the GP, MGGP, and

empirical equations for different durations and recurrence intervals were picked out from the same population with the amounts obtained based on regional distributions. The values of the MWU test statistic (Z_U) regarding the GP, MGGP, and selected empirical equations became between -0.033 and -0.009 for the SRI and between -0.182 and -0.25 for the SRII. The absolute Z_U values were less than the critical value of 1.960 taken from the standardized normal distribution table at the 5% significance level. The MWU test results revealed that the estimated rainfall intensity amounts were statistically from the same population as those from the regional distribution.

4 Conclusion

The IDF curves and the equations describing the relationship between the intensity, duration, and frequency of rainfall are widely used tools in the planning, design, and operation of hydraulic structures. In addition, human factor interventions such as global climate change have led to differentiation in IDF relationships. In this context, IDF relationships establishing successfully would cause reliable forming of the long-term management strategies of the structures in question. Apart from the traditional derivation of IDF relationships, genetically inspired approaches have been widely used recently in solving the current problem in hydrology, as in many other fields. In this study, genetic inspiration-based techniques, called particle swarm optimization (PSO), artificial bee colony (ABC), genetic algorithm (GA), flow direction algorithm (FDA), genetic programming (GP), and multi-gene genetic programming (MGGP) approaches, are considered in establishing IDF relationships. In this context, this study, designed for the Euphrates-Tigris river basin, was founded on three main topics, namely performing the regional frequency analysis based on the L-moment algorithm, estimating the parameters of empirical equations in the literature with PSO, ABC, GA, and FDA techniques, and developing new IDF relationships with GP and MGGP approaches.

The homogeneity of the study area for the maximum rainfall series of 11 different durations was ensured by forming two sub-regions with the L-moment algorithm. The GEV, GNO, GLOG, and the PIII theoretical distributions in the eight, seven, four, and three of 22 maximum rainfall series with different durations for the two sub-regions provided the most approximate regional fit, respectively. Rainfall intensity magnitude was estimated for the desired recurrence period from the regional distribution of each duration, and these estimates were taken into account in the PSO, ABC, GA, GP, FDA, and MGGP approaches as observed values. Rainfall intensity values from 11 empirical IDF equations, whose parameters were calculated by using PSO, ABC, GA, and FDA optimization techniques, were compared with the intensities estimated from regional distributions based on error metrics, and the empirical relationships having the highest performance were revealed for both sub-regions. With the GP and MGGP approaches, new IDF relationships were formed, and the ones that gave the most approximate results to intensity amounts obtained from the regional distribution were determined with error metrics. When the results obtained with both the highest-performing empirical equations and the GP and MGGP approaches were compared with each other with the Taylor diagram, the most successful result for SRI was achieved with the equation formed

with MGGP and for SRII with the equation obtained from GP. The results obtained from this study show that GP and MGGP methods can provide in-depth information on the internal properties of nonlinear IDF equations and generate equations that reflect the nonlinear dynamic process. Therefore, both methods used have proven to be effective tools for hydrological forecasting and can be used to solve similar water resource problems in other basins.

Acknowledgements The authors thank the General Directorate of Meteorology for helping to provide the data used in this study.

Author contribution All authors contributed to the study's conception and design. Data collection and data curation were performed by Kadri Yurekli. The methodology was performed by Mehmet Ali Hinis, Kadri Yurekli, and Muberra Erdogan. The investigation and writing – review and editing were performed by Mehmet Ali Hinis and Kadri Yurekli. Supervision was performed by Mehmet Ali Hinis and Kadri Yurekli. Analysis was performed by Mehmet Ali Hinis and Muberra Erdogan. All authors have read and agreed to the final version of the manuscript.

Data availability Not applicable.

Code availability Not applicable.

Declarations

Competing interests The authors declare no competing interests.

References

- Adeloye AJ, Montaseri M (2002) Preliminary streamflow data analyses prior to water resources planning study. *Hydrol Sci J* 47:679–692. <https://doi.org/10.1080/02626660209492973>
- Agbazo M, Koton'Gobi G, Kounouhewa B, Alamou E, Afouda A (2016) Estimation of IDF curves of extreme rainfall by simple scaling in Northern Oueme Valley, Benin Republic (West Africa). *Earth Sci Res J* 20:1–7. <https://doi.org/10.15446/esrj.v20n1.49405>
- Akay B (2009) Performance analysis of artificial bee colony algorithm on numerical optimization problem. Ph.D. Thesis, Erciyes University, Graduate School of Natural and Applied Science
- Akbaş Z (2015) Türkiye'nin Fırat ve Dicle sınıraşan sularından kaynaklanan güvenlik sorunu ve çatışma riski (Turkish). *J Soc Sci Turkic World* 72:93–116
- Altınbilek D (2004) Development and management of the Euphrates-Tigris basin. *Int J Water Resour Dev* 20:15–33. <https://doi.org/10.1080/07900620310001635584>
- Al-Wagdany AS (2020) Intensity-duration-frequency curve derivation from different rain gauge records. *J King Saud Univ Sci* 32:3421–3431. <https://doi.org/10.1016/j.jksus.2020.09.028>
- Amjadi M (2015) Statistic and probabilistic variations and rainfall predictions of TRNC. Master Thesis, Eastern Mediterranean University, Gazimagusa North Cyprus
- Anlı AS (2009) Regional frequency analysis of rainfall data in Ankara Province via L-moment methods. Ph.D. Thesis, Ankara University, Graduate School of Natural and Applied Sciences
- Asıkoğlu ÖL, Benzeden E (2007) Stable frequency distribution models for the annual maximum rainfalls of standard durations. *Sci Eng J Fırat Univ* 19:543–551
- Basakın EE, Ekmekcioglu O, Ozger M, Citakoglu H (2021) Determination of intensity-duration-frequency relation by particle swarm

- optimization and genetic programming. In: II. International Applied Statistics Conference (UYIK-2021). Tokat, Turkey, pp 1–8
- Belay GW, Azeze M, Melesse AM (2019) Reservoir operation analysis for Ribb reservoir in the Blue Nile basin (Chapter 16), 191–211. In: Melesse AM, Abteu W, Senay G (eds) Extreme hydrology and climate variability: monitoring, modelling, adaptation and mitigation
- Bell FC (1969) Generalized rainfall-duration-frequency relationships. *J Hydraul Div ASCE* 95:311–327. <https://doi.org/10.1061/JYCEAJ.0001942>
- Blackwell T, Kennedy J, Poli R (2007) Particle swarm optimization. *Swarm Intell* 1:33–57. <https://doi.org/10.1007/s11721-007-0002-0>
- Bratton D, Kennedy J, (2007) Defining a standard for particle swarm optimization. *IEEE Swarm Intelligence Symposium*, Honolulu, HI, USA, pp 120–127. <https://doi.org/10.1109/SIS.2007.368035>
- Chang KB, Lai SH, Faridah O (2013) RainIDF: automated derivation of rainfall intensity–duration–frequency relationship from annual maxima and partial duration series. *J Hydroinformatics* 15:1224–1233. <https://doi.org/10.2166/hydro.2013.192>
- Chen CL (1983) Rainfall intensity-duration-frequency formulas. *ASCE J Hydraulic Eng* 109:1603–1621. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1983\)109:12\(1603\)](https://doi.org/10.1061/(ASCE)0733-9429(1983)109:12(1603))
- Citakoglu H, Demir V (2023) Developing numerical equality to regional intensity–duration–frequency curves using evolutionary algorithms and multi-gene genetic programming. *Acta Geophys* 71:469–488. <https://doi.org/10.1007/s11600-022-00883-8>
- Coles S (2001) An introduction to statistical modeling of extreme values. *Springer Series in Statistics*. London, Springer-Verlag, p 208. <https://doi.org/10.1007/978-1-4471-3675-0>
- Cunnane C (1989) Statistical distributions for flood frequency analysis. *WMO Report No. 718*. World Meteorological Organization, Geneva
- Dalrymple T (1960) Flood frequency methods. U. S. Geological Survey. *Water Supply Paper* 1543A:11–51
- Degirmenci S (2007) Turkey's transboundary waters and water problem in Middle East within the context of Firat, Dicle, Asi Rivers. M.Sc. Thesis, Pamukkale University Graduate School of Social Sciences
- Demir AF, Pamukcu ÖK (1996) Firat Dicle Havzasında Türkiye'nin Su Politikası. *Değişen Dünya ve Türkiye Der. Faruk Sönmezoglu. Bağlam Yayınları, İstanbul*, pp 267–286
- Dupont BS, Allen DL (2000) Revision of the rainfall intensity duration curves for the commonwealth of Kentucky. Kentucky Transportation Center, College of Engineering, University of Kentucky, Research Report: KTC-00–18
- Elsebaie IH (2012) Developing rainfall intensity–duration–frequency relationship for two regions in Saudi Arabia. *J King Saud Univ Eng Sci* 24:131–140. <https://doi.org/10.1016/j.jksues.2011.06.001>
- Fadhel S, Rico-Ramirez MA, Han D (2017) Uncertainty of intensity–duration–frequency (IDF) curves due to varied climate baseline periods. *J Hydrol* 547:600–612. <https://doi.org/10.1016/j.jhydrol.2017.02.013>
- Farzin S, Anaraki MV (2021) Modeling and predicting suspended sediment load under climate change conditions: a new hybridization strategy. *J Water Clim Chang* 12:2422–2443. <https://doi.org/10.2166/wcc.2021.317>
- Farzin S, Anaraki MV, Naeimi M, Zandifar S (2022) Prediction of groundwater table and drought analysis; a new hybridization strategy based on bi-directional long short-term model and the Harris hawk optimization algorithm. *J Water Clim Chang* 13:2233–2254. <https://doi.org/10.2166/wcc.2022.066>
- Fernando DAK, Jayawardena AW (1994) Generation and forecasting of monsoon rainfall data. 20th WEDC Conference on Affordable Water Supply and Sanitation, Colombo, pp. 310–313
- Gado TA, Salama AM, Zeidan BA (2021) Selection of the best probability models for daily annual maximum rainfalls in Egypt. *Theoret Appl Climatol*. <https://doi.org/10.1007/s00704-021-03594-0>
- Gandomi AH, Alavi AH (2012) A new multi-gene genetic programming approach to nonlinear system modeling—part I: materials and structural engineering problems. *Neural Comput Appl* 21:171–187
- Gebru TA (2020) Rainfall intensity-duration-frequency relations under changing climate for selected stations in the Tigray region Ethiopia. *J Hydrol Eng* 25:05020041. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001999](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001999)
- Gocic M, Velimirovic L, Stankovic M, Trajkovic S (2021) Regional precipitation–frequency analysis in Serbia based on methods of L-moment. *Pure Appl Geophys*. <https://doi.org/10.1007/s00024-021-02688-0>
- Goldberg DE (1989) Genetic algorithms in search, optimization and machine learning, 1st edn. Addison-Wesley Longman Publishing Co., Inc, USA
- Gorkemli B, Citakoglu H, Haktanir T, Karaboga D (2022) A new method based on artificial bee colony programming for the regional standardized intensity–duration–frequency relationship. *Arab J Geosci*. <https://doi.org/10.1007/s12517-021-09377-1>
- Gubareva TS, Gartsman BI (2010) Estimating distribution parameters of extreme hydrometeorological characteristics by L-moment method. *Water Resour* 37:437–445. <https://doi.org/10.1134/S0097807810040020>
- Haddad K (2021) Selection of the best fit probability distributions for temperature data and the use of L-moment ratio diagram method: a case study for NSW in Australia. *Theoret Appl Climatol* 143:1261–1284. <https://doi.org/10.1007/s00704-020-03455-2>
- Hassan MU, Noreen Z, Ahmed R (2021) Regional frequency analysis of annual daily rainfall maxima in Skåne, Sweden. *Int J Climatol* 41:4307–4320. <https://doi.org/10.1002/joc.7074>
- Hinchliffe MP, Willis MJ, Hiden H, Tham MT, McKay B, Barton GW (1996) Modelling chemical process systems using a multi-gene genetic programming algorithm. In *Genetic Programming: Proceedings of the First Annual Conference*, pp 56–65
- Hosking JRM (1990) L-moments: analyzing and estimation of distributions using linear combinations of order statistics. *J R Stat Soc B* 52:105–124. <https://doi.org/10.1111/j.2517-6161.1990.tb01775.x>
- Hosking JRM, Wallis JR (1993) Some statistics useful in regional frequency analysis. *Water Resour Res* 29:271–281. <https://doi.org/10.1029/92WR01980>
- Hosking JRM, Wallis JR (1997) Regional frequency analysis: an approach based on L-moments. Cambridge University Press, Cambridge
- IPCC (2012) Managing the risks of extreme events and disasters to advance climate change adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change. Cambridge and New York: Cambridge University Press p 582
- IPCC (2007) IPCC Climate Change 2007: The Physical Science Basis (eds Solomon, S. et al.) Cambridge Univ. Press
- Jain NK, Nangia U, Jain J (2018) A review of particle swarm optimization. *J Instit Eng (India) Ser B* 99:407–411
- Karaboga D, Gorkemli B, Ozturk C, Karaboga N (2014) A comprehensive survey: artificial bee colony (ABC) algorithm and applications. *Artif Intel Rev* 42:21–57. <https://doi.org/10.1007/s10462-012-9328-0>
- Karaboga D (2005) An idea based on honey bee swarm for numerical optimizasyon technical report-TR06, Erciyes Üniversitesi, Engineering Faculty, Computer Engineering Department
- Karahan H (2012) Determining rainfall-intensity-duration-frequency relationship using particle swarm optimization. *KSCE J Civ Eng* 16:667–675. <https://doi.org/10.1007/s12205-012-1076-9>
- Karahan H, Ozkan E (2012) Best fitting distributions for the standard duration annual maximum precipitations in the Aegean Region. *Pamukkale Univ J Eng Sci* 19:152–157. <https://doi.org/10.5505/pajes.2013.29392>

- Karahan H, Ceylan H, Tamer A, Ayvaz M (2007) Predicting rainfall intensity using a genetic algorithm approach. *Hydrol Process* 21:470–475. <https://doi.org/10.1002/hyp.6245>
- Karahan H, Ayvaz MT, Gürarlan G (2008) Determination of intensity-duration-frequency relationship by genetic algorithm: case study of GAP. *Techn J* 19:4393–4407 (In Turkish)
- Karami H, Anaraki MV, Farzin S, Mirjalili S (2021) Flow direction algorithm (FDA): a novel optimization approach for solving optimization problems. *Comput Ind Eng* 156:107224. <https://doi.org/10.1016/j.cie.2021.107224>
- Kaygusuz K (1999) Energy and water potential of the southeastern Anatolia project (GAP). *Energy Sources* 21:913–922. <https://doi.org/10.1080/00908319950014281>
- Kennedy J, Eberhart RC (1995) Particle swarm optimization. *Proc IEEE Int Conf Neural Netw* 4:1942–1948. <https://doi.org/10.1109/ICNN.1995.488968>
- Khan MSR, Hussain Z, Ahmad I (2020) Regional flood frequency analysis, using L-moments, artificial neural networks and Ols regression, of various sites of Khyber-Pakhtunkhwa, Pakistan. *Appl Ecol Environ Res* 19:471–489. https://doi.org/10.15666/aeer/1901_471489
- Koutsoyiannis D, Kozonis D, Manetas A (1998) A mathematical framework for studying rainfall intensity-duration-frequency relationships. *J Hydrol* 206:118–135. [https://doi.org/10.1016/S0022-1694\(98\)00097-3](https://doi.org/10.1016/S0022-1694(98)00097-3)
- Koza JR (1992) Genetic programming: on the programming of computers by means of natural selection. MIT Press, Cambridge
- Labat D, Godderis Y, Probst JL, Guyot JL (2004) Evidence for global runoff increase related to climate warming. *Adv Water Resour* 27(631):642. <https://doi.org/10.1016/j.advwatres.2004.02.020>
- Lestari S, King A, Vincent C (2019) Seasonal dependence of rainfall extremes in and around Jakarta. *Indones Weather Clim Extrem* 24:100202. <https://doi.org/10.1016/j.wace.2019.100202>
- Modarres R, Sarhadi A (2010) Statistically-based regionalization of rainfall climates of Iran. *Global Planet Change*. <https://doi.org/10.1016/j.gloplacha.2010.10.009>
- Müftüoğlu F (1997) Ortadoğu Su Meseleleri ve Türkiye. Marifet Yayınları, İstanbul (In Turkish)
- Nachar N (2008) The Mann Whitney U: a test for assessing whether two independent samples come from the same distribution. *Quant Meth Psych* 4:13–20
- Nain M, Hooda BK (2021) Regional frequency analysis of maximum monthly rainfall in Haryana State of India using L-moments. *J Reliab Stat Stud* 14:33–56. <https://doi.org/10.13052/jrss0974-8024.1413>
- Nhat L, Tachikawa J, Takara K (2006) Establishment of intensity-duration-frequency curves for precipitation in the monsoon area of Vietnam. *Annuals of Disas Prev Res Inst, Kyoto Univ.*, No. 49 B
- Okonkwo GI, Mbajioru CC (2010) Rainfall intensity-duration-frequency analyses for South Eastern Nigeria. *Agric Eng Int CIGR Ej. Manuscript* 1304. Vol. XII
- Ouali D, Cannon AJ (2018) Estimation of rainfall intensity–duration–frequency curves at ungauged locations using quantile regression methods. *Stoch Environ Res Risk Assess* 32:2821–2836. <https://doi.org/10.1007/s00477-018-1564-7>
- Ozis U, Ozdemir Y (2008) Euphrates-Tigris rivers basin and Turkey. (In Turkish) TMMOB 2. Water Policy Congress, IMO Congress and Cultures Center, Ankara, Turkey, pp 443–445
- Paola FD, Giugni M, Topa ME, Bucchignani E (2014) Intensity-duration-frequency (IDF) rainfall curves, for data series and climate projection in African cities. *Springerplus* 3:1–18. <https://doi.org/10.1186/2193-1801-3-133>
- Park JS, Jung HS, Kim RS, Oh JH (2001) Modelling summer extreme rainfall over the Korean Peninsula using wakeby distribution. *Int J Climatol* 21:1371–1384. <https://doi.org/10.1002/joc.701>
- Raiford JP, Aziz NM, Khan AA, Powell DN (2007) Rainfall depth-duration-frequency relationships for South Carolina, North Carolina, and Georgia. *Am J Environ Sci* 3:78–84
- Rao AR, Hamed KH (2000) Flood frequency analysis. CRC Press, Boca Raton
- Rasel M, Islam M (2015) Generation of rainfall intensity-duration frequency relationship for north-western region in Bangladesh. *J Environ Sci, Toxicol Food Technol (IOSR-JESTFT)* 9:41–47
- Riccardo P, William BL, Nicholas FM, Koza JR (2008) A field guide to genetic programming, Lulu Enterprises. UK Ltd. Reprint edition. <http://www.gp-field-guide.org.uk>
- Rougé C, Ge Y, Cai X (2013) Detecting gradual and abrupt changes in hydrological records. *Adv Water Resour* 53:33–44. <https://doi.org/10.1016/j.advwatres.2012.09.008>
- Searson DP, Leahy DE, Willis MJ (2010) GPTIPS: an open source genetic programming toolbox for multigene symbolic regression. In: Proceedings of the International MultiConference of Engineers and Computer Scientists (IMECS '10), March 17–19, Hong Kong, pp 77–80
- Šimková T (2017) Homogeneity testing for spatially correlated data in multivariate regional frequency analysis. *Water Resour Res* 53:7012–7028. <https://doi.org/10.1002/2016WR020295>
- Yukse O, Anılan T, Saka F, Örgün E (2022) Rainfall intensity-duration-frequency analysis in Turkey, with the emphasis of Eastern Black Sea basin. *Techn J* 33:12087–12103. <https://doi.org/10.18400/tekderg.727085>
- Yurekli K (2015) Impact of climate variability on precipitation in the upper Euphrates-Tigris rivers basin of Southeast Turkey. *Atmos Res* 154:25–38. <https://doi.org/10.1016/j.atmosres.2014.11.002>
- Yurekli K (2022) Identification of possible risks to hydrological design under non-stationary climate conditions. *Nat Hazards*. <https://doi.org/10.1007/s11069-022-05686-0>
- Yurekli K, Modarres R, Ozturk F (2009) Regional daily maximum rainfall estimation for Cekerek Watershed by L-moments. *Meteorol Appl* 16:435–444. <https://doi.org/10.1002/met.139>
- Yurekli K, Erdogan M, Ismail AH, Shareef MA (2021) Hydrochemical characteristics of surface water and its suitability for drinking and irrigation: a case study of the Euphrates river basin, Turkey. *Sustain Water Resour Manag* 7:1–11. <https://doi.org/10.1007/s40899-021-00507-x>
- Zakwan M (2016) Application of optimization technique to estimate IDF parameters. *Water Energy Int* 59:69–71
- Zeder J, FisCher EM (2020) Observed extreme precipitation trends and scaling in Central Europe. *Weather Clim Extrem* 29:100266. <https://doi.org/10.1016/j.wace.2020.100266>
- Zhang Q, Xu C, Tao H, Tao J, Chen YD (2009) Climate changes and their impacts on water resources in the arid regions: a case study of the Tarim River basin, China. *Stoch Environ Res Risk Assess* 24:349–358. <https://doi.org/10.1007/s00477-009-0324-0>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.