Acta Geophysica (2023) 71:469–488
https://doi.org/10.1007/s11600-022-00883-8

RESEARCH ARTICLE - HYDROLOGY

Developing numerical equality to regional intensity–duration–frequency curves using evolutionary algorithms and multi-gene genetic programming

Hatice Citakoglu[1] · **Vahdettin Demir**[2]

Received: 22 April 2022 / Accepted: 16 July 2022 / Published online: 24 August 2022

Abstract
This study aims to carry out regional intensity – duration – frequency (IDF) equality using the relationship with IDF obtained from point frequency analysis. Eleven empirical equations used in the literature for seven climate regions of Turkey were calibrated using particle swarm optimization (PSO) and genetic algorithm (GA) optimization techniques and the obtained results were compared. In addition, in this study, new regional IDF equations were obtained for each region utilizing Multi-Gene Genetic Programming (MGGP) method. Finally, Kruskal–Wallis (KW) test was applied to the IDF values obtained from the methods and the observed values. As a result of the study, it was observed that the coefficients of 11 empirical equations calibrated with PSO, and GA techniques were different from each other. The mean absolute error (MAE), root mean square error (RMSE), mean absolute relative error (MARE), coefficient of determination (R^2), and Taylor diagram were used to evaluate the performances of PSO, GA, and MGGP techniques. According to the performance criteria, it has been determined that the IDF equations obtained by the MGGP method for the Eastern Anatolia, Aegean, Southeastern Anatolia, and Central Anatolia regions are more successful than the empirical equations calibrated with the PSO and GA method. The empirical IDF equations produced with PSO and the IDF equations acquired with MGGP have similar findings in the Mediterranean, Black Sea, and Marmara. In addition, the KW test results showed that the data of all models were from the same population.

Keywords Multi-gene genetic programming · Genetic algorithm · Particle swarm optimization · Turkey · Regional intensity–duration–frequency analysis

Introduction

Climatic extremes and weather have long been thought to be the root causes of environmental dangers including food shortages and landslides (Stephenson et al. 2016; Zahiri et al. 2016; Deb et al. 2018; Zeder and Fischer 2020; Şen and Kahya 2021). Extreme precipitation events could occur more common in the future as a result of global climate change (Dastagir 2015; Shahid et al. 2016; Hay et al. 2016; Lestari et al. 2019; Barbero et al. 2019). Climate change causes extreme precipitation events' intensity, duration, and frequency (IDF) (Mirhosseini et al. 2013; Yilmaz et al. 2014; Fadhel et al. 2017; Buba et al. 2017; Al-Wagdany 2021). The way the current and future urban drainage systems are constructed, operated, and maintained would be impacted by changes in IDF linkages (Sillmann et al. 2017; Moujahid et al. 2018; Cook et al. 2020; Gebru 2020). As a result, a better knowledge of the potential changes in the relationship between the characteristics of extreme rainfall events as a result of climate change is critical for developing long-term urban stormwater management strategies.

Many studies on hydrology and water resources fields, such as rainfall–runoff models, flood control structures, stormwater drainage projects, and watershed modeling, are need to IDF (Tyralis and Langousis 2019; Anılan et al. 2022). IDF must also be discovered or predicted in urban hydrological applications (Borga et al. 2005; Egodawatta

Edited by Dr. Ankit Garg (ASSOCIATE EDITOR) / Dr. Michael Nones (CO-EDITOR-IN-CHIEF).

✉ Vahdettin Demir
vahdettin.demir@karatay.edu.tr

[1] Civil Engineering Department, Engineering Faculty, Erciyes University, Kayseri, Turkey

[2] Civil Engineering Department, Faculty of Engineering and Natural Sciences, KTO Karatay University, Konya, Turkey

et al. 2007; Aly et al. 2009). The capacity of these civil infrastructure systems must be adequately sized to avoid overdesign, which could result in financial losses, greater property damage, and possibly increased danger of death (Anılan et al. 2022). As a result, having accurate IDF estimations is critical.

The IDF is a mathematical formula that connects the rainfall intensity I, duration D, and frequency of exceedance F. This relationship is commonly represented with curves that depict the change in rainfall intensity over time. Empirical and statistical methods are commonly used to create these curves.

The first IDF studies in the literature date back to 1932 (Bernard 1932). Rainfall contour maps were created by Hershfield in 1963 to determine rainfall design depths for various return intervals and durations (Hershfield 1963). Chen (1983) devised a generalized IDF formula for the USA based on three different precipitation depths: ten years one hour, ten years twenty-four hours, and one hundred years one hour. Froehlich (1995) developed duration equations—rainfall intensity for periods of 1–24 h for four US Weather Bureau-defined geographical regions. García-Bartual and Schneider (2001) operating frequency analyses to a series of several extreme rainfall events in Spain, measured between 1925 and 1992, and developed nine different empirical IDF equations. Yu et al. (2004) created a regional IDF connection for ungauged sites. Chang et al. (2013) developed Rain-IDF, a software tool for deriving an IDF relationship with an Excel add-in using Visual Basics for Applications (VBA). Ouali and Cannon (2018) estimated regional IDF curves at ungauged locations based on quantile regression. They compared linear quantile regression and nonlinear quantile regression using canonical correlation analysis and nonlinear canonical correlation analysis at sites across Canada for the IDF curve. They stated that nonlinear QR framework leads to the best results for estimation of IDF curves at ungauged sites. Adarsh and Janga Reddy (2018) developed an alternate methodology employing the scale invariance property of rainfall, the empirical method, and extreme value model formulations to derive the hourly and sub-daily IDF connections for data-scarce regions. They claimed that the proposed approach was effective in extracting short-duration IDF correlations from daily rainfall data.

Many researchers have conducted research around the world for empirical precipitation density prediction equations (Awadallah et al. 2017; Hayder and Al-Mukhtar 2021; Bulti et al. 2021; Galiatsatou and Iliadis 2022; Alramlawi and Fıstıkoğlu 2022). For example, Bell (1969) described IDF curves for specific locations in Russia. For the development of IDF curves by various authors, there are academics from the Kingdom of Saudi Arabia (Al-Shaikh 1985; Al-Khalaf 1997; Elsebaie 2012; Subyani and Al-Amri 2015; Al-Amri and Subyani 2017; Ewea et al. 2017; Elsebaie et al. 2021). Several researchers analyze and estimate the IDF curves for the southern and northern parts of Iraq (Lamia Abdul Jaleel and Maha Atta Farawn 2013; Hamaamin 2016; Mahdi and Mohamedmeki 2020; Hasan and Saeed 2020; Kareem et al. 2022). Eman Ahmed Hassan El-Sayed (2011) created a collection of regional IDF curves for Egypt's Sinai Peninsula using isohyet maps. Liew et al. (2014) used meteorological records from nearby stations to adapt the IDF curve in remote areas in Malaysia.

In the past, various academics have presented a variety of IDF curve creation methods, depending on the available data in Turkey. Karahan et al. (2007) used a genetic algorithm (GA) method which could consistently calculate the IDF connection to predicting rainfall intensity and used mean square error at the cost function. They also reported a strong match between measured and predicted values, and they used the method to analyze IDF relationships in Turkey's GAP region. According to these researchers, the GA approach was an effective tool for determining mathematical model parameters and successfully representing the supplied data (Karahan et al. 2008). Based on the easy generalization approach and the robust estimation procedure, Asikoglu and Benzeden (2014) developed two different IDF equations for the western region of Turkey using the Gumbel distributions and two-parameter lognormal distributions in frequency studies. The RMSE values of both methods were compared, and the former procedure produced superior rainfall intensity readings than the latter. Şen (2019) developed a novel method for creating IDF curves using data from the Annual Daily Maximum Rainfall (ADMR). Four ADMR recordings from meteorology stations in Turkey's Ceylanpinar district, near the Syrian border, were subjected to the proposed approach. Başakın et al. (2021) used genetic programming (GP) to build an equation that used information from IDF associations obtained as the result of statistical frequency investigations to calculate severe rainfall intensities for four meteorological stations in the interior of Turkey. They also compared their model to one of the accessible empirical IDF equations in the literature. The particle swarm optimization (PSO) scheme was used to optimize the parameters of the empirical equation. Finally, they concluded that the equation developed using GP was more accurate than the equation derived using PSO. Görkemli et al. (2022) developed nine different equations for the regional IDF curves using the artificial bee colony programming (ABCP) method for the interior of Turkey. They also calibrated nine empirical IDF equations commonly used in the literature for the interior of Turkey with an Excel solver. According to the performance criteria, they stated that the ABCP equations gave better results than the empirical equations for the entire IDF relationship between 2 and 10,000 years. Anılan et al. (2022) determined the regional IDF curves to the Black Sea region in the north of Turkey using the Log-Pearson Type 3

distribution. They calibrated the coefficients of the empirical formulas in the literature with the SPSS package program and obtained new equations with multiple nonlinear regression analysis. They stated that the equations they obtained as a result of their studies had sufficiently high reliability. They expressed that they concluded that the IDF relationship would open new horizons and motivate future studies.

Scopus' database was used to examine significant concentrations in the literature. In this database, 876 studies were identified related to intensity, duration and frequency keywords. A list of essential keywords for this research area was generated using the VOSviewer software (Fig. 1a) (VOSviewer 2022). Furthermore, within the adopted research timeline (Fig. 1a), it seems that many studies have been published from 2005 to the present and IDF studies are a current issue. The main focuses and most researched topics in these studies are climate change, IDF curves, extreme precipitation, and the use of machine learning techniques in recent studies. The primary areas investigated by IDF studies are shown in Fig. 1b. Figure 1b shows that the USA is the most studied region (183), followed by the Canada (98), France (28), Brazil (65), India (48), Italy (46), etc. However, only (24) studies were conducted in Turkey, and it is in the 12th place.

The motivation of this research is to examine the point frequency analyses of three stations selected from each region to quantitatively describe the regional IDF relations with seven different climate regions in Turkey. In the study, empirical equations whose parameters are adjusted using PSO and GA and MGGP method, which generates equations according to parameters, were used.

Material

Turkey is between 36°–42° north latitudes, and 26°–45° east longitudes and its surface area is 783,562 km^2. Turkey is located between the temperate zone and the subtropical zone and is a region where four seasons are experienced. The region is surrounded by the sea on three sides, and the range and height of the mountains are different, resulting in different climate types. There are different climates in the interior and coastal regions of Turkey, with more temperate climate characteristics in the coastal regions and continental climate in the interior regions.

In the current study, 21 meteorology stations (three from each region) of the General Directorate of Meteorology (MGM) in Turkey are observed; the annual maximum precipitation intensity values of various durations (5 min, 15 min, 30 min, 1 h, 3 h, 6 h, 12 h, 24 h) were used. The series of 14 standard T periods of 2, 5, 10, 25, 50, 100, 200, 500, 1000, 2000, 5000, and 1000 years were used at the 21 sites in this investigation.

Between 1991 and 2020 (last climatic period), the average annual precipitation in Turkey is 573.4 mm. Since this average precipitation amount corresponds to approximately 450 billion m^3 of water, it is very important to know the relationship between precipitation intensity (i), precipitation duration (t), and recurrence period (T) (IDF) in the planning and operation of water resources projects. The locations of 21 meteorology stations used in the study are given in Fig. 2 (MGM 2022).

Figure 2 shows three stations randomly chosen from 7 geographical regions of Turkey. All analyses representing the basins were analyzed using data from three stations. Irrigation is the most important factor affecting the efficiency of agricultural activities in Turkey. Especially in the summer months, the need for water is quite high with the effect of arid and semi-arid climates. In our country, irrigation is very important due to the wide agricultural areas in Central Anatolia, Southeastern Anatolia, Aegean, and Mediterranean regions and the dry climate. In Turkey, 75% of the land cannot be irrigated and therefore the desired yield cannot be obtained (Yavuz 2018). As a result, the control and correct planning of water resources is very important for the study area. The precipitation data used in this study were obtained from the study conducted by Yavuz (2018).

Method

Frequency analyses

Statistical frequency analysis was created so that practitioners may make probabilistic future predictions. It is also used to calculate the relationship between the magnitude and likelihood of hydrologic random variables reaching or not exceeding their limitations. Frequency analyses are commonly used in hydrology for extreme events such as peak flows and precipitations. A suitable probability distribution is used for frequency analysis, and one of the three-parameter probability distributions of generalized normal (GNO), also known as the three-parameter log–normal (LN3), Gumbel, 3-parameter Gamma distribution (P3) general extreme values (GEV), generalized logistic (GLO), and generalized Pareto (GPA) is usually used. The magnitude forecasting of the parameters of a probability distribution using observed series is an important step in frequency analysis, and also the traditional methods of moments and maximum likelihood are widely utilized (Görkemli et al. 2022).

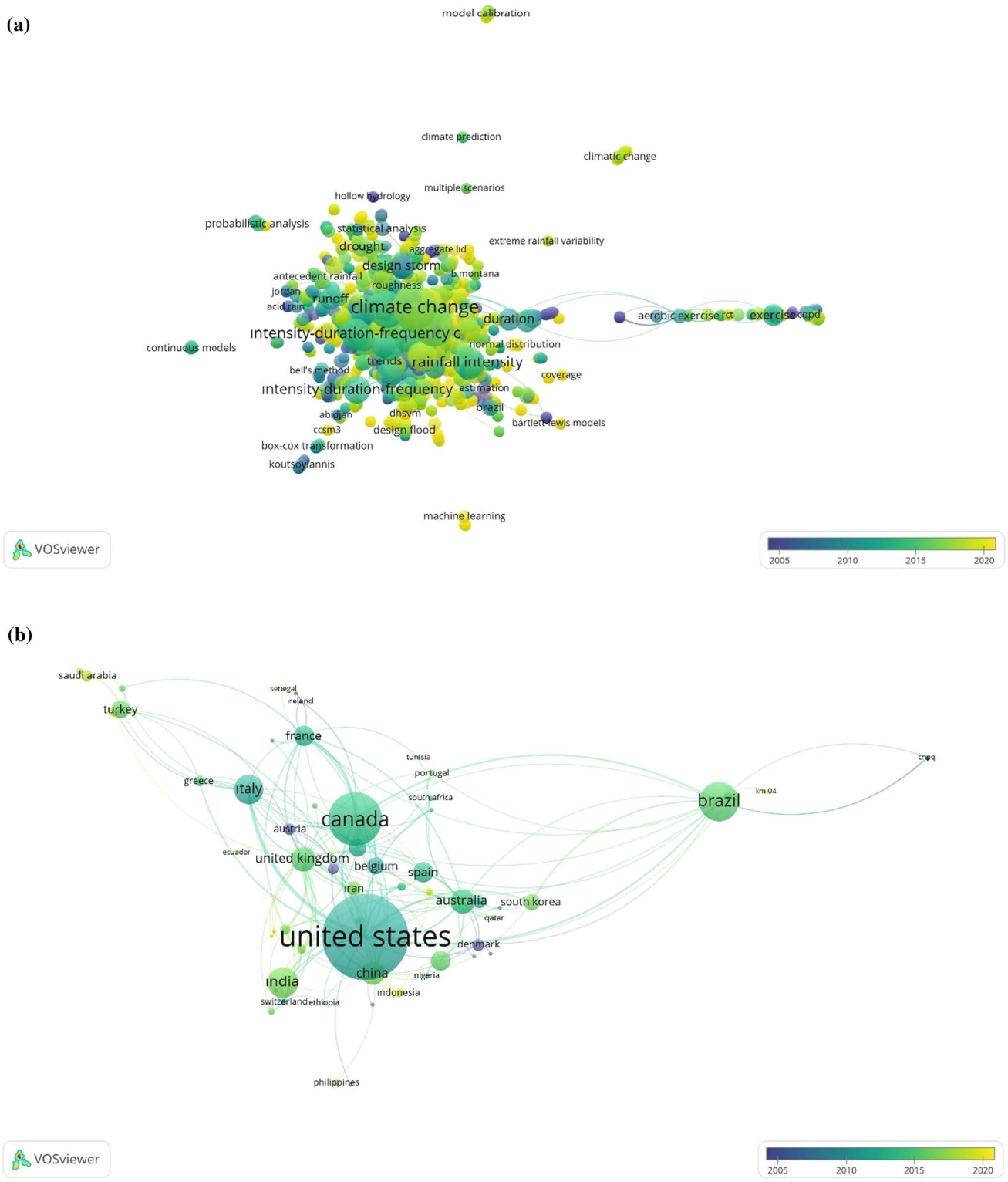

Fig. 1 **a** Keywords and **b** research regions in Scopus' database

Fig. 2 Location of stations and seven geographical regions of Turkey

Table 1 As a result of frequency analysis, appropriate probability distribution functions and parameter estimation methods belong to 7 meteorological regions Yavuz (2018)

Regions	Appropriate distributions	Parameter estimation methods
Mediterranean region	GEV	Probability weighted moments
Eastern Anatolia region	Gumbel	Probability weighted moments
Aegean region	GEV	Moments
Southeastern Anatolia region	P3	Moments
Central Anatolia region	GEV	Probability weighted moments
Black sea region	LN3	Probability weighted moments
Marmara region	P3	Moments

The frequency analyses and precipitation intensity values used in this study were obtained from the study conducted by Yavuz (2018). For appropriate distributions (Table 1) and tests, this study can be reviewed.

Optimization with particle swarm optimization (PSO)

PSO is an algorithm inspired by swarm behavior (Kennedy and Eberhart 2010). Being population-based, PSO is based on information sharing within the herd. Individuals in the swarm are called particles and the position and velocity information of each particle is recorded. Optimization work is carried out by updating this speed and position information in each iteration. This update is adjusted according to the particle's current velocity information, its position, the best position of the swarm, and the best position in that iteration. X represents particle position (direction) and v represents particle flight speed. Each x in the particle swarm is given a point, depending on the solution approach of the problem. Pbest represents the local best for each particle; Gbest represents the global best in the current generation. Particles record their best position (Pbest) and can be capable of finding the optimal position (Gbest) known to all particles in the swarm. In order to determine the position of the particle one step further, the velocity and the new position vector are obtained by utilizing the particle's information up to that point (Eqs. 1 and 2).

$$v_{id}^{t+1} = wv_i^t + c_i r_i (p_{id} - x_{id}^t) + c_2 r_2 (p_{gd} - x_{id}^t), \quad (1)$$

$$x_{id}^{t+1} = x_{id}^t + v_{id}^{t+1}. \quad (2)$$

In these equations, the inertia value of w is the constant to prevent the motion of the particle, the acceleration coefficients c_1 and c_2, and r_1 and r_2 are a randomly generated values in the range of [0–1]. The algorithm starts the search process in the solution space with a random velocity and

position information. The fitness value of each particle is determined with the help of the position information fitness function. The Pbest value is updated by comparing the determined fitness value with the Pbest value, which is the best fitness value obtained by the particle so far. The Gbest value is updated by comparing the best fitness value that the particle has so far with the best fitness value that the swarm has achieved so far (Başakın et al. 2021). Detailed information on the subject and hybrid applications can be found in (Elbaz et al. 2020; Shaban et al. 2021).

Optimization with genetic algorithms (GA)

Genetic algorithms are inspired by Darwin's theory of evolution (Karahan et al. 2008). The solution of any problem with a genetic algorithm is done by virtually evolving the problem. GAs have been widely used in engineering problems after the books of Goldberg (1989), Gen and Cheng (1997), and Gen et al. (2008).

One of the main features of GAs is to change the population represented by chromosomes by using some operators. Chromosomes can be represented by chains of characters of a given length l. Each chromosome represents a suitable solution to the problem. Chromosomes are made up of chains of symbols, and each symbol is called a bit (digit). Chromosomes are formed by arranging each bit in order of which parameter it represents. For example, if chains are created in the binary number system, each chain takes the value 0 and 1. The connection between the GA and the problem is provided by the objective function. The objective function allows the chromosomes to be converted to real numbers. If the objective function is maximized, a large objective function value indicates that the solution represented by this chromosome is better than the other chromosomes.

GAs are a sequential generative method. GAs use three basic parameters: reproduction, crossover, and mutation. Each generation made in the GA process creates a new society from the existing society. For example, let the initial population size be (pr). Each (pr) individual is assigned an integer. This assignment can be arbitrary or deterministic. The reproduction process selects the most suitable individuals from the existing society based on the objective function and using selection operators, for example, roulette wheel or tournament (Bell 1969). The reproduction operator selects the best individuals from the current generation to pass them on to the next generation. After completing this operator operation, crossover and mutation operators come into play and make other manipulations in the society in line with the given crossover (pc) and mutation (pm) probabilities. Detailed information on the subject and hybrid applications can be found in (Goldberg 1989; Goldberg and Deb 1991; Karahan et al. 2008; Elbaz et al. 2019; Shaban et al. 2021; Başakın et al. 2021).

Multi-gene genetic programming (MGGP)

The initial population achievement is judged on training data based on the objective function minimization to obtain a better solution using the MGGP approach. The most common objective function is the root mean squared error.

MGGP is evolved from genetic programming (GP), which is a mathematical model that is used to further empirical mathematical modeling. MGGP is made up of a weighted number of GP trees created by combining set genes and regression genes which employs the least squares approach for coefficient estimation. The MGGP estimates the output value based on the problem's input variables (Eray et al. 2018). The user specifies the maximum number of genes (G) and maximum gen depth (D) to gain control over this model complexity. The size and quantity of models to be evaluated in the global space are influenced by the G and D parameters. As a result, there exist ideal G and D values that result in an easily portable model (Citakoglu et al. 2020; Citakoglu 2021). In MGGP, the starting population is created by randomly selecting different genes from people that already exist in GP trees. Genes are inserted or removed by the crossover operator during MGGP execution, which is defined as a two-point high-level crossover that facilitates gene exchange across individuals. Each parent's gene is chosen at random. The comprising tree is then swapped with parent trees using the normal subtree crossover operator (Gandomi and Alavi 2012; Eray et al. 2018). The user determines the probability of the crossover operator. The steps of the MGGP algorithm are as follows:

Step 1: Identify the issue.
Step 2: Initialize the settings.
Step 3: Using the least square method, construct a model.
Step 4: Examine the models' performance.
Step 5: Create a new population using genetic operators.
Step 6: Examine individual performance in the new population.
Step 7: If termination conditions are provided, complete the actions. If not, proceed to step 5.

Detailed information about MGGP was obtained from (Searson 2009; Searson et al. 2010).

Empirical equations

In this study, Lopcu (2007), Karahan et al. (2008), Yavuz (2018), and Başakın et al. (2021) suggested empirical connections were used. The relations of the intensity value are as shown in Eqs. 3–13:

$$i = \frac{\left(a \cdot T^{b}\right)}{(t+c)^{d}}, \tag{3}$$

$$i = \frac{\left(a \cdot T^{b}\right)}{t^{c}}, \tag{4}$$

$$i = \frac{\left(a \cdot T^{b}\right)}{\left(t^{c} + d\right)}, \tag{5}$$

$$i = \frac{(a + b \cdot \ln T)}{t^{c}}, \tag{6}$$

$$i = \frac{(a + b \cdot \ln T)}{\left(t^{c} + d\right)}, \tag{7}$$

$$i = \frac{(a + b \cdot \ln T)}{(t + c)^{d}}, \tag{8}$$

$$i = \frac{(a + b \cdot [\ln(\ln T)])}{(t + c)^{d}}, \tag{9}$$

$$i = \frac{(a + b \cdot [\ln(\ln T)])}{t^{c}}, \tag{10}$$

$$i = \frac{(a + b \cdot [\ln(\ln T)])}{\left(t^{c} + d\right)}, \tag{11}$$

$$i = \frac{(a + b \cdot [\ln(T)] + c.\left[\ln\left(T\right)^{2}\right])}{(d + e \cdot [\ln(t)] + f.\left[\ln\left(t\right)^{2}\right])}, \tag{12}$$

$$i = \frac{\left(a.T^{b}\right)}{(c + t^{d})^{e}}. \tag{13}$$

Here, i stands for the standard rainfall intensity (mm/minute); T stands for the frequency (years); t stands for the rainfall duration (min); and *a, b, c, d, e,* and *f* stands for the coefficients in these equations.

Performance metrics

The accuracy of the models proposed in this research was evaluated using widely known performance metrics (Voyant et al. 2017). MAE, MARE, RMSE, and R^2 were used in model evaluations (Eqs. 14–17). Low MAE, MARE, and RMSE values, as well as R^2 values near 1 suggest accurate and dependable estimations.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|i_{\mathrm{predicted}} - i_{\mathrm{measured}}\right|, \tag{14}$$

$$\mathrm{MARE} = 100\frac{1}{n}\sum_{i=1}^{n}\frac{\left|i_{\mathrm{predicted}} - i_{\mathrm{measured}}\right|}{SR_{\mathrm{predicted}}}, \tag{15}$$

$$\mathrm{RMSE} = \frac{1}{n}\sum_{i=1}^{n}\sqrt{\left(i_{\mathrm{predicted}} - i_{\mathrm{measured}}\right)^{2}}, \tag{16}$$

$$R^{2} = \frac{\sum_{i=1}^{n}(i_{i\mathrm{measured}} - \underline{i_{i\mathrm{measured}}})^{2}.(i_{i\mathrm{predicted}} - \underline{i_{\mathrm{predicted}}})^{2}}{\sum_{i=1}^{n}\left(i_{i\mathrm{measured}} - \underline{i_{i\mathrm{measured}}}\right)^{2}.\sum_{i=1}^{n}\left(i_{i\mathrm{predicted}} - \underline{i_{\mathrm{predicted}}}\right)^{2}}, \tag{17}$$

where i_{measured} is rainfall intensity variables measured by MGM; $i_{\mathrm{predicted}}$ is rainfall intensity variables predicted by approaches; $\underline{i_{\mathrm{measured}}}$ is average of rainfall intensity variables; and n is amount of data. In this study, scatter plots, Taylor diagram was used to compare approaches. These diagrams graphically summarize how close the models are to the observations (Taylor 2001; Legouhy 2021; Uncuoğlu et al. 2021; Demir 2022).

Applications and results

In the study, firstly, the empirical equation parameters were optimized by the PSO method for the calculation of precipitation intensity values. The optimization process was carried out by rearranging the codes in the Licensed MATLAB package program and library (MATLAB 2022a). There are a total of 6 coefficients (*a, b, c, d, e,* and *f*) that need to be optimized in the empirical equation. To perform intensity–duration–frequency analysis, firstly, frequency analysis of the largest precipitation values should be done for each station one by one. In this study, the largest possible precipitation values, which are suitable for the selected distribution, are used in 2, 5, 10, 25, 50, 100, 200, 500, 1000, 2000, 5000, and 1000 years. These values are named the measured precipitation value and represent the actual precipitation intensity values of the station. Obtained precipitation intensity values are introduced to the methods as output, precipitation duration, and precipitation frequency as inputs. Then, coefficients for 11 empirical equations under the title of empirical equations were introduced to the models and the models were asked to optimize these coefficients. During the optimization process, the RMSE was used as the objective function (Eq. 16). The values that minimize the objective function are calculated by the PSO algorithm. During the calculation process, inspired by Başakın et al. (2021), the algorithm was started by taking the population size value of 100 and the inertia value of 0.9 acceleration coefficients c1 and c2 "2" belonging to the PSO. The coefficients and

Table 2 PSO optimized coefficients and performance criteria

Region	Equation	*a*	*b*	*c*	*d*	*e*	*f*	RMSE	MAE	MARE	R^2
Mediterranean region	**1**	249.110	0.229	1.194	0.636			25.507	13.161	22.246	0.943
	2	220.468	0.230	0.615				26.561	13.306	22.297	0.943
	3	300.197	0.227	0.664	0.912			25.100	13.073	22.230	0.944
	4	74.003	156.890	0.616				17.751	10.315	21.892	0.972
	5	**105.358**	**188.343**	**0.648**	**0.748**			**16.708**	**10.114**	**21.778**	**0.973**
	6	95.768	172.151	1.476	0.635			16.725	10.138	21.795	0.973
	7	356.785	444.523	0.816	0.685			28.123	16.604	30.137	0.935
	8	332.887	400.815	0.671				28.230	16.773	30.194	0.932
	9	376.682	468.864	0.693	0.473			28.154	16.614	30.137	0.935
	10	17.007	4.558	1.423	1.000	− 0.778	− 0.218	28.629	14.782	23.933	0.931
	11	280.788	0.227	1.201	0.815	0.804		25.082	13.064	22.237	0.944
Eastern Anatolia region	1	277.329	0.147	1.602	0.756			10.470	4.276	16.895	0.963
	2	**230.852**	**0.146**	**0.723**				**9.596**	**4.323**	**17.077**	**0.966**
	3	264.546	0.147	0.748	0.521			9.960	4.257	16.919	0.965
	4	147.290	65.159	0.702				9.605	3.496	14.783	0.972
	5	152.524	68.824	0.710	0.166			9.870	3.496	14.756	0.971
	6	150.861	68.145	0.356	0.709			9.963	3.505	14.753	0.971
	7	274.701	194.918	1.084	0.740			14.639	5.998	19.019	0.940
	8	244.448	169.815	0.719				13.950	5.981	19.112	0.941
	9	286.835	203.319	0.747	0.648			14.679	6.014	19.012	0.940
	10	7.684	3.596	− 0.015	0.775	− 0.677	− 0.191	12.042	5.289	22.608	0.951
	11	273.933	0.147	1.405	0.987	0.764		10.297	4.259	16.892	0.963
Aegean region	**1**	698.337	0.141	5.366	0.852			20.555	10.214	22.299	0.936
	2	395.103	0.141	0.752				21.744	11.150	23.193	0.924
	3	760.035	0.139	0.863	3.823			20.510	10.241	22.325	0.936
	4	280.649	103.265	0.725				20.233	10.116	21.534	0.930
	5	**463.262**	**167.131**	**0.807**	**2.425**			**19.730**	**9.407**	**20.918**	0.940
	6	412.254	149.415	3.661	0.790			19.913	9.409	20.900	0.940
	7	552.134	404.026	3.698	0.788			24.820	10.879	21.715	0.913
	8	365.028	264.306	0.711				24.001	11.192	22.331	0.905
	9	609.543	442.302	0.802	2.401			25.016	10.958	21.712	0.913
	10	8.452	4.476	− 0.084	0.825	− 0.675	− 0.169	28.762	12.929	27.409	0.882
	11	706.562	0.141	5.478	1.000	0.854		20.613	10.216	22.300	0.936
Southeastern Anatolia region	**1**	380.130	0.175	1.892	0.738			43.672	15.445	18.583	0.869
	2	311.586	0.173	0.701				41.782	15.315	18.841	0.871
	3	401.243	0.174	0.745	0.999			43.510	15.404	18.562	0.870
	4	181.872	115.195	0.689				44.453	14.798	14.927	0.877
	5	195.884	125.401	0.702	0.286			45.241	14.859	14.893	0.877
	6	**195.328**	**125.286**	**0.653**	**0.702**			**45.051**	**14.788**	**14.887**	**0.877**
	7	406.512	381.413	2.451	0.735			55.256	17.994	19.121	0.831
	8	309.540	294.271	0.689				52.528	17.649	19.332	0.834
	9	412.285	382.309	0.734	1.000			54.359	17.850	19.122	0.833
	10	6.581	7.293	− 0.183	0.787	− 0.665	− 0.176	48.313	16.781	22.690	0.835
	11	211.145	0.173	− 0.040	0.068	9.961		41.626	15.308	18.867	0.871

performance metrics obtained for the regions are given in Table 2. Bold numbers in Table 2 show the equation coefficients of the lowest MARE performance criterion.

In the second phase of the present study, obtaining the optimum equation coefficients with the GA method was carried out using the MATLAB library (MATLAB 2022b). At the stage of creating GA estimation models, the population size was chosen as 100, the number of generations was 1000, and the StallGenLimit was 100. The RMSE was employed as the objective function during the GA

Table 2 (continued)

Region	Equation	*a*	*b*	*c*	*d*	*e*	*f*	RMSE	MAE	MARE	R^2
Central Anatolia region	**1**	247.392	0.294	4.861	0.746			49.104	18.868	13.846	0.869
	2	192.604	0.293	0.705				48.213	17.464	14.368	0.874
	3	**265.618**	**0.294**	**0.755**	**2.333**			**48.087**	**18.652**	**13.857**	**0.872**
	4	45.174	193.763	0.723				50.494	21.525	21.748	0.867
	5	64.714	319.475	0.804	2.519			54.369	21.933	21.219	0.872
	6	60.436	292.069	4.088	0.792			55.017	22.080	21.216	0.871
	7	411.669	480.723	2.977	0.781			81.021	33.541	34.958	0.774
	8	313.828	353.536	0.731				76.846	32.826	35.167	0.767
	9	451.881	527.608	0.795	1.924			80.864	33.507	34.957	0.775
	10	96.517	− 0.527	13.522	− 0.999	0.000	− 0.919	62.279	28.747	34.627	0.786
	11	255.952	0.294	3.375	0.876	0.857		48.603	18.766	13.850	0.870
Black Sea region	**1**	311.864	0.228	0.964	0.680			26.836	11.444	16.741	0.959
	2	281.637	0.229	0.663				30.229	11.956	16.858	0.957
	3	320.633	0.227	0.683	0.418			26.916	11.385	16.744	0.959
	4	115.546	185.859	0.674	–			19.009	9.208	15.741	0.974
	5	138.345	230.983	0.709	0.736			17.198	8.965	15.601	0.977
	6	**132.632**	**218.216**	**1.438**	**0.702**			**17.096**	**9.009**	**15.598**	**0.977**
	7	529.690	606.276	4.620	0.725			37.832	17.755	25.794	0.920
	8	341.903	389.423	0.656				34.724	18.125	26.202	0.920
	9	606.211	697.307	0.743	2.456			36.772	17.509	25.774	0.922
	10	15.835	10.347	0.797	1.272	− 0.986	− 0.260	27.590	13.533	19.361	0.940
	11	318.609	0.227	0.690	0.837	0.816		26.731	11.403	16.742	0.959
Marmara region	**1**	696.688	0.138	5.870	0.797			9.371	4.650	9.453	0.984
	2	396.971	0.139	0.708				17.016	7.128	11.770	0.971
	3	782.600	0.138	0.814	3.760			9.303	4.642	9.426	0.984
	4	276.378	117.808	0.711				16.008	6.269	8.717	0.980
	5	**518.121**	**200.861**	**0.801**	**3.255**			**6.200**	**2.915**	**5.841**	**0.993**
	6	463.234	179.457	5.180	0.785			6.221	2.938	5.859	0.993
	7	609.929	546.833	5.021	0.796			10.414	5.433	10.816	0.982
	8	390.437	322.905	0.714				15.304	7.483	12.643	0.966
	9	706.085	620.882	0.815	3.430			10.479	5.451	10.797	0.982
	10	10.426	6.756	− 0.194	0.802	− 0.674	− 0.179	13.335	7.981	20.994	0.969
	11	756.662	0.137	4.611	0.885	0.914		9.311	4.637	9.436	0.984

optimization procedure. The coefficients and performance metrics obtained for the regions are given in Table 3. Bold numbers in Table 3 show the equation coefficients of the lowest MARE performance criterion.

Comparison of the best results obtained from Eqs. 1–11 with the measured outputs through the use of MARE error criteria is provided in Table 4. As can be inferred from Table 4, with empirical equations, RMSE values varied between 6.2–81.02, 9.463−114.177; MAE values between 0.78–18.86, 0.78−19.85; MARE values between 2.15–21.77, 2.17−22.53; and the R^2 values between 0.86−0.99, 0.86–0.99 for PSO and GA, respectively. The experimental equations for these error criteria are given in Table 5.

Comparing the empirical equations in Table 4, the generally best results were obtained with the equations where 3 or 4 coefficients were optimized. Although the optimized equations differ regionally, the best results were generally observed in Eqs. 2, 3, 5, and 11. Regionally, the lowest RMSE errors are seen in the Eastern Anatolia Region and Marmara Region; the highest RMSE errors are seen in the Central Anatolia Region and Southeastern Anatolia Region.

Table 3 GA optimized coefficients and performance criteria

Region	Equation	*a*	*b*	*c*	*d*	*e*	*f*	RMSE	MAE	MARE	R^2
Mediterranean region	**1**	292.901	0.222	1.000	0.661			27.213	13.595	22.539	0.949
	2	413.898	0.196	0.711				40.665	16.986	25.595	0.957
	3	**458.558**	**0.226**	**0.731**	**2.979**	**–**		**25.777**	**13.513**	**22.531**	**0.937**
	4	91.434	87.521	0.541				36.115	19.180	27.168	0.976
	5	51.035	84.490	0.525	− 0.925			20.676	11.647	23.565	0.963
	6	17.333	85.563	− 1.000	0.601			27.079	16.549	29.181	0.964
	7	332.756	85.561	− 1.000	0.602			63.155	34.648	44.384	0.854
	8	216.964	92.694	0.537				72.645	38.514	43.756	0.899
	9	493.179	89.555	0.675	− 0.551			53.417	31.386	45.352	0.820
	10	32.980	0.681	2.831	0.998	−0.822	0.277	32.349	15.777	25.161	0.933
	11	57.785	0.302	− 0.354	0.618	0.749		48.007	25.307	31.561	0.854
Eastern Anatolia region	**1**	238.793	0.148	0.609	0.731			10.061	4.295	16.971	0.965
	2	264.190	0.139	0.745				9.611	4.605	17.537	0.966
	3	438.717	0.145	0.837	3.137			11.959	4.690	17.428	0.955
	4	**161.970**	**64.028**	**0.704**				**9.478**	**3.581**	**14.948**	**0.971**
	5	242.769	83.000	0.765	1.000			10.513	3.841	15.784	0.967
	6	174.370	72.457	0.998	0.724			10.448	3.588	14.885	0.970
	7	204.438	85.322	0.263	0.653			24.692	10.252	23.887	0.919
	8	200.903	91.427	0.656				23.350	9.682	23.005	0.924
	9	167.573	90.458	0.641	− 0.880			16.007	7.231	20.989	0.927
	10	103.302	0.919	2.567	0.431	− 0.999	0.895	17.496	10.124	33.026	0.906
	11	52.402	0.201	− 0.997	0.739	0.698		26.597	13.210	31.887	0.929
Aegean region	**1**	413.796	0.141	0.480	0.759			20.838	10.886	22.987	0.928
	2	400.958	0.142	0.755				21.985	11.156	23.196	0.924
	3	658.534	0.143	0.843	2.862			20.122	10.321	22.357	0.936
	4	278.085	97.893	0.717				20.128	10.080	21.586	0.931
	5	**302.341**	**98.392**	**0.719**	**0.363**			**20.475**	**9.898**	**21.554**	**0.935**
	6	295.142	91.259	0.599	0.713			21.142	10.117	21.682	0.934
	7	321.635	83.437	− 0.980	0.644			38.434	18.635	29.761	0.856
	8	385.299	90.625	0.672				38.666	18.602	29.824	0.856
	9	482.507	83.947	0.705	− 0.188			34.899	17.462	30.920	0.833
	10	139.352	0.966	3.965	0.383	− 1.000	0.896	35.014	20.412	39.240	0.820
	11	223.348	0.160	− 0.593	0.915	0.722		24.212	12.983	25.789	0.925
Southeastern Anatolia region	**1**	327.843	0.175	0.619	0.711			42.567	15.322	18.706	0.871
	2	355.085	0.168	0.722				39.973	15.242	19.103	0.871
	3	590.057	0.175	0.809	3.243			46.047	16.045	18.807	0.862
	4	188.546	82.658	0.651				54.142	17.907	17.092	0.871
	5	**205.806**	**87.914**	**0.663**	**− 0.115**			**49.942**	**16.122**	**16.305**	**0.870**
	6	212.562	86.368	− 0.235	0.662			50.612	16.308	16.683	0.869
	7	265.724	89.900	− 0.999	0.605			73.583	28.947	31.013	0.779
	8	177.426	93.747	0.557				82.084	32.813	30.033	0.808
	9	399.418	87.297	0.660	− 0.359			67.291	26.818	33.931	0.748
	10	116.272	0.791	4.870	0.379	− 0.901	0.773	50.629	24.059	34.325	0.807
	11	100.357	0.213	− 0.148	0.690	0.789		68.522	28.251	29.409	0.846

Table 3 (continued)

Region	Equation	*a*	*b*	*c*	*d*	*e*	*f*	RMSE	MAE	MARE	R^2
Central Anatolia region	**1**	196.616	0.294	0.355	0.709			47.192	17.460	14.251	0.875
	2	249.446	0.283	0.739				55.314	18.061	15.240	0.875
	3	**332.048**	**0.295**	**0.789**	**4.620**			**53.060**	**19.853**	**13.940**	**0.862**
	4	62.849	89.473	0.615				80.517	32.655	26.722	0.869
	5	71.571	88.165	0.624	− 0.986			60.725	26.663	25.511	0.845
	6	178.253	82.276	− 0.448	0.667			79.516	33.633	32.542	0.836
	7	185.900	97.273	− 1.000	0.604			106.919	46.568	43.103	0.729
	8	141.463	87.072	0.565				114.177	49.960	43.844	0.747
	9	131.106	89.332	0.559	− 1.000			98.602	42.811	40.437	0.736
	10	17.489	0.631	2.652	0.825	− 0.788	0.283	48.832	21.136	27.356	0.865
	11	176.672	0.296	0.328	0.874	0.792		46.005	17.806	14.452	0.874
Black Sea region	**1**	**305.621**	**0.228**	**0.771**	**0.676**			**27.160**	**11.463**	**16.749**	**0.959**
	2	380.825	0.198	0.695				28.154	11.632	18.265	0.966
	3	462.743	0.227	0.738	2.280			24.168	11.774	17.172	0.955
	4	119.232	96.643	0.573				40.360	19.951	22.416	0.976
	5	71.821	85.400	0.545	− 0.997			22.454	11.872	18.539	0.961
	6	211.311	89.090	− 0.986	0.600			33.918	17.410	23.816	0.960
	7	281.764	88.918	− 0.989	0.557			73.329	36.635	40.113	0.850
	8	298.885	89.457	0.565				76.804	38.104	40.808	0.849
	9	241.743	80.553	0.540	− 1.000			59.953	31.254	38.417	0.835
	10	27.993	0.762	2.010	0.902	− 0.757	0.231	29.420	13.533	21.390	0.954
	11	156.623	0.253	− 0.970	0.915	0.630		28.541	14.214	19.141	0.940
Marmara region	**1**	**606.459**	**0.139**	**4.486**	**0.776**			**9.463**	**4.713**	**9.588**	**0.984**
	2	397.397	0.139	0.708				16.988	7.123	11.770	0.971
	3	660.936	0.146	0.793	3.234			10.183	5.262	9.878	0.981
	4	289.153	88.684	0.678				9.871	5.525	10.981	0.983
	5	272.509	87.709	0.680	− 0.234			12.021	6.130	10.565	0.978
	6	309.389	94.434	0.333	0.694			9.546	5.373	10.017	0.984
	7	306.078	85.885	− 0.985	0.610			31.490	16.566	25.085	0.927
	8	413.294	92.790	0.651				29.097	15.412	25.301	0.920
	9	447.823	76.690	0.660	− 0.441			26.046	14.586	26.881	0.894
	10	133.912	0.589	3.798	0.198	− 0.665	0.728	31.156	18.694	34.633	0.853
	11	159.528	0.193	− 0.103	0.741	0.816		23.049	13.127	21.798	0.943

The MGGP technique was utilized to estimate intensity levels in the study's final phase. The generalization ability of the equations developed with the MGGP approach is influenced by parameter selection. The function node set employed in this study is +, *x*, *y*, *e*, abs, sin, and log. The maximum number of generations and population size is 300. In addition, the maximum number of genes and maximum gene depth range from 4 to 5, as inspired by Citakoglu et al. (2020). Maximum number of genes (*Gmaks*) and maximum gene depth (*Dmaks*) determine the size of the search space and the solutions to be sought in this search space, and high values for population size and number of generations increase study duration. Different values for *Gmaks* and *Dmaks* were explored to discover the model with more accurate outcomes, and the best models were achieved when *Dmaks* was set to 4–5 and *Gmaks* was set to 4–5. The fitness values were assessed using RMSE values, and the model with the lowest RMSE value was chosen as the best. As a termination condition, the maximum number of generations was chosen. To evaluate model performance, we used all of the error criteria listed in the "Error Criteria" section (Table 5). The equations generated for the estimation of intensity with the use of *T*, and *t* data and MGGP are given Table 6.

Table 4 Regional empirical equations

Region	Equations	
	PSO	GA
Mediterranean region	$i(5) = \frac{(105.358+188.343\cdot \ln T)}{(t^{0.648}+0.748)}$	$i(5) = \frac{(51.035+84.490\cdot \ln T)}{(t^{0.525}+-0.925)}$
Eastern Anatolia region	$i(2) = \frac{(230.852\cdot T^{0.146})}{t^{0.723}}$	$i(4) = \frac{(161.970+64.028\cdot \ln T)}{t^{0.704}}$
Aegean region	$i(5) = \frac{(463.262+167.131\cdot \ln T)}{(t^{0.807}+2.425)}$	$i(3) = \frac{(658.534\cdot T^{0.143})}{(t^{0.843}+2.862)}$
Southeastern Anatolia region	$i(11) = 211.145.T^{0.173} / \left(-0.040 + t^{0.068}\right)^{9.961}$	$i(2) = \frac{(355.085\cdot T^{0.168})}{t^{0.722}}$
Central Anatolia region	$i(3) = \frac{(265.618\cdot T^{0.294})}{(t^{0.755}+2.333)}$	$i(11) = 176.672.T^{0.296} / \left(0.328 + t^{0.874}\right)^{0.792}$
Black Sea region	$i(6) = \frac{(132.632+218.216\cdot \ln T)}{(t+1.438)^{0.702}}$	$i(5) = \frac{(71.821+85.400\cdot \ln T)}{(t^{0.545}-0.997)}$
Marmara region	$i(5) = \frac{(518.121+200.861\cdot \ln T)}{(t^{0.801}+3.255)}$	$i(1) = \frac{(606.459\cdot T^{0.139})}{(t+4.486)^{0.776}}$

Table 5 Performance criteria for MGGP

Region	RMSE	MAE	MARE	R^2
Mediterranean region	14.927	9.066	23.909	0.9781
Eastern Anatolia region	8.531	3.942	16.555	0.9724
Aegean region	18.156	9.861	25.546	0.9422
Southeastern Anatolia region	36.746	14.770	19.325	0.8813
Central Anatolia region	42.563	16.031	22.210	0.8905
Black Sea region	16.150	8.520	19.891	0.9784
Marmara region	5.561	3.097	8.288	0.9942

Comparison of the results obtained from MGGP method with the measured outputs through the use of MARE error criteria is provided in Table 5. As can be inferred from Table 5, with MGGP method, RMSE values varied between 5.561 and 42.56, MAE values between 3.09 and 16.03, MARE values between 8.28 and 25.54, and the R^2 values between 0.88 and 0.99. The experimental equations for these error criteria are given in Table 6.

The results of the experimental equations and MGGP method are examined; in all regions, the MGGP method gave the lowest RMSE error values compared to other methods. The PSO method, on the other hand, gave very close results with the MGGP. The disadvantage of the MGGP method is that the equations obtained as a result of the model are more complex than the experimental equations.

The scatter plots of the best results by regions are presented in Fig. 3. In the scatterplot given in Fig. 3, it is seen that the values of the output variables obtained with, are distributed above the trend line despite small deviations. The output variables obtained in the time series plot presented in Fig. 3 coincide with the measured values.

In present study, MGGP approach, PSO optimization, and GA optimization were used for estimation of i variable. The method in which experimental equations are optimized with the MGGP approach was selected as the best model for estimation of *i* since it had the least performance metrics values (RMSE, MAE, MARE).

Taylor diagram is used as another comparison criterion. In this graphical method, in which the closeness to the

Table 6 Regional equations obtained by MGGP

Region	Equations
Mediterranean region	$i_{\text{Mediterranean}} = 1.02 \cdot \sqrt{\left\lvert \frac{T}{t} \right\rvert} - 188.53 \cdot e^{(3.275-t)} + \frac{94.46\cdot \log(\lvert T \rvert)}{\sqrt{\lvert t \rvert}} + 184.453$
Eastern Anatolia region	$i_{\text{Eastern Anatolia}} = \frac{\sqrt{\lvert t \rvert}}{0.0119\cdot t-0.0087} + \frac{13.86\cdot \log(\lvert T \rvert)}{\sqrt{\lvert t \rvert}} + \frac{\log(\lvert T \rvert)}{0.0112\cdot t-0.0088} - 1.52$
Aegean region	$i_{\text{Aegean}} = 1.763\frac{\log(\lvert t \rvert)}{\lvert T \rvert} + \frac{860.4}{t+6.133} + 74.50\frac{\log(\lvert T \rvert)}{\sqrt{\lvert t \rvert}} - 10.18$
Southeastern Anatolia region	$i_{\text{Southeastern Anatolia}} = \frac{\log(\lvert T \rvert)}{0.0031\cdot t-0.006} + \frac{36.27\cdot \log(\lvert t \rvert)+23.76\cdot \log(\lvert T \rvert)}{\sqrt{\lvert t \rvert}} - 7.48$
Central Anatolia region	$i_{\text{Central Anatolia}} = 2 \cdot e^{-t} \cdot (2900 \cdot t - 12188) + \frac{\log(\lvert T \rvert)}{0.0017\cdot t+0.0125} + 5.7 \cdot \sqrt{\left\lvert \frac{T}{t} \right\rvert} + 2.75$
Black Sea region	$i_{\text{Black Sea}} = \frac{\log(\lvert T \rvert)}{0.005\cdot t+0.016} + 204.6\frac{\log(\lvert t \rvert)}{T+t} + 87.6\frac{\log(\lvert T \rvert)}{\sqrt{\lvert t \rvert}} - 4.35$
Marmara region	$i_{\text{Marmara}} = \frac{394.28\cdot \log(\lvert T \rvert)}{t+10.23} + \frac{209.99\cdot \sqrt{\lvert t \rvert}}{t-0.4849} - 144.56 \cdot \frac{\sqrt{\lvert t \rvert}}{t\cdot T} - 2.32$

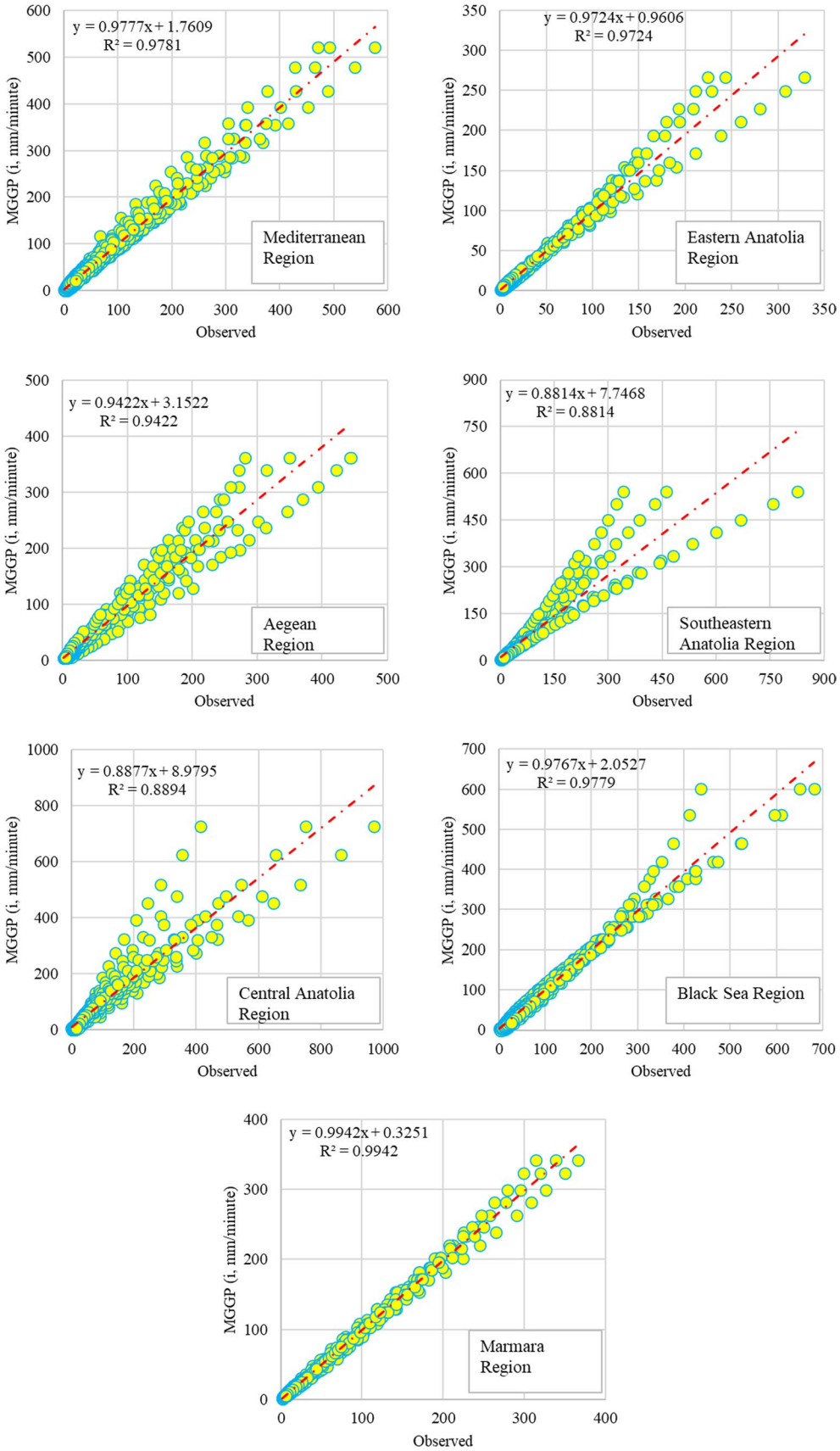

Fig. 3 The scatter plots of the best results for regions

observed values determines the best model, the RMSD error criterion and the correlations coefficient are used. The Taylor diagram prepared according to the regions is given in Fig. 4.

The closeness to the observed data in Taylor diagrams is expressed by the RMSD and the correlation coefficient. Closeness to the observed point indicates that the model is more accurate or better expresses the measured values (Citakoglu 2021; Demir 2022). In the Taylor diagrams of the best results, it is seen in many regions where PSO is superior to GA. But the most striking results were found in MGGP. MGGP points are very close to PSO and gave better results than PSO in all regions according to RMSD criterion. For this reason, Taylor diagrams adds a different interpretation feature to the studies. As a result, the MGGP approach made estimations with less errors than the coefficients obtained with PSO, and GA and the equations obtained using these coefficients. Therefore, MGGP is superior to other methods. However, when the IDF relationship was plotted on a graph in Fig. 5, it was visually observed that the empirical equation in Table 6 obtained for Marmara Region was partially successful in representing the overall IDF relationship. In Fig. 5, it can be seen that it shows close agreement for all ranges of frequencies. It gave unsuccessful results for 2−, 5−, 10-year frequencies and annual frequencies such as 2000, 5000, and 10,000, but remarkably successful results were observed for other frequencies. The overall result was that although the error values of the relevant models for the IDF relationship were low, even the best ones could not accurately represent the regionalized IDF relationship.

In addition to the traditional performance measurements, the Kruskal–Wallis (KW) test for best empirical equations and the MGGP technique were used at the end of the study. The KW test is a nonparametric alternative to the traditional one-way ANOVA. It compares data sets' medians to see if the samples are from the same population (or, equivalently, different populations with the same distribution). The KW test calculates test statistics using rows of data rather than numerical values. The ranks are determined by sorting the data into groups from smallest to largest and then taking the numerical index of that order. The average rank of all associated observations is equal to the rank of a linked observation. In KW test, the p value measures the significance of the Chi-square statistic, which replaces the F-statistic used in traditional one-way ANOVA. The KW test assumes that all samples come from the same continuous distribution, with the exception of possible differences in location due to group effects, and that all observations are mutually independent (Citakoglu 2021; Demir 2022; MATLAB 2022c). Table 7 shows the results of the KW test. The empirical equations and the MGGP technique passed the KW test at the 95% significance level, as shown in Table 7. The empirical equations and the MGGP technique both produce results from the same population.

Discussion

When many of the worldwide articles mentioned in the introduction are examined, three coefficients belonging to the $i = \frac{c1 \times T^{c2}}{t^{c3}}$ empirical equation, which is widely used in the literature, to the IDF relations obtained by point frequency analysis were determined with a single method.

Several articles on IDF equality have been prepared by researchers in Turkey. Alramlawi and Fistikoğlu (2022), Şen (2019), Karahan et al. (2007), and Karahan et al. (2008) performed frequency analysis on the extreme precipitation intensity data of a single meteorology station in Turkey. In these studies, the probability distribution and IDF relationship determined for each meteorology station were primarily determined. Afterward, they developed point IDF equations by using only one optimization technique, the coefficients of only one of the empirical equals commonly used in the literature for each determined IDF relationship.

Anılan et al. (2022) performed frequency analysis for each of the stations in the Black Sea region one by one and showed that each of these stations fit different probability distributions. They optimized the coefficients of 9 empirical equations in the literature with the SPSS program to the IDF relations obtained as a result of the frequency analysis. They did not mention the optimization technique they used in the analysis, and they used a single optimization method like other studies for Turkey. The IDF equations obtained in this study are regional and it has not been checked whether the results represent the original IDF relations. Acar et al. (2008) applied frequency analysis to the extreme precipitation data of İzmir (a station in the Aegean Region of Turkey) and estimated the precipitation intensity with artificial neural networks by using the precipitation duration and frequency as input data. This study is a machine learning based study using a single station data. Ouali and Cannon (2018) performed regional frequency analysis for Canada and obtained an IDF relationship. In addition, they obtained a single IDF relationship for Canada by making separate Canonical correlation analysis and nonlinear canonical correlation analysis, which are used for multiple variables in their studies. Thus, they obtained three different IDF associations for Canada. For each IDF relationship, the rainfall duration and frequency were taken as input data and predicted rainfall intensity with quantile regression and quantile regression neural network. This study is regional and an artificial neural network-based study. Haktanir et al. (2016) performed regional frequency analysis using the L-moments method in their study. L-moments method produces a single IDF relationship using the weighted averages of the L-variation

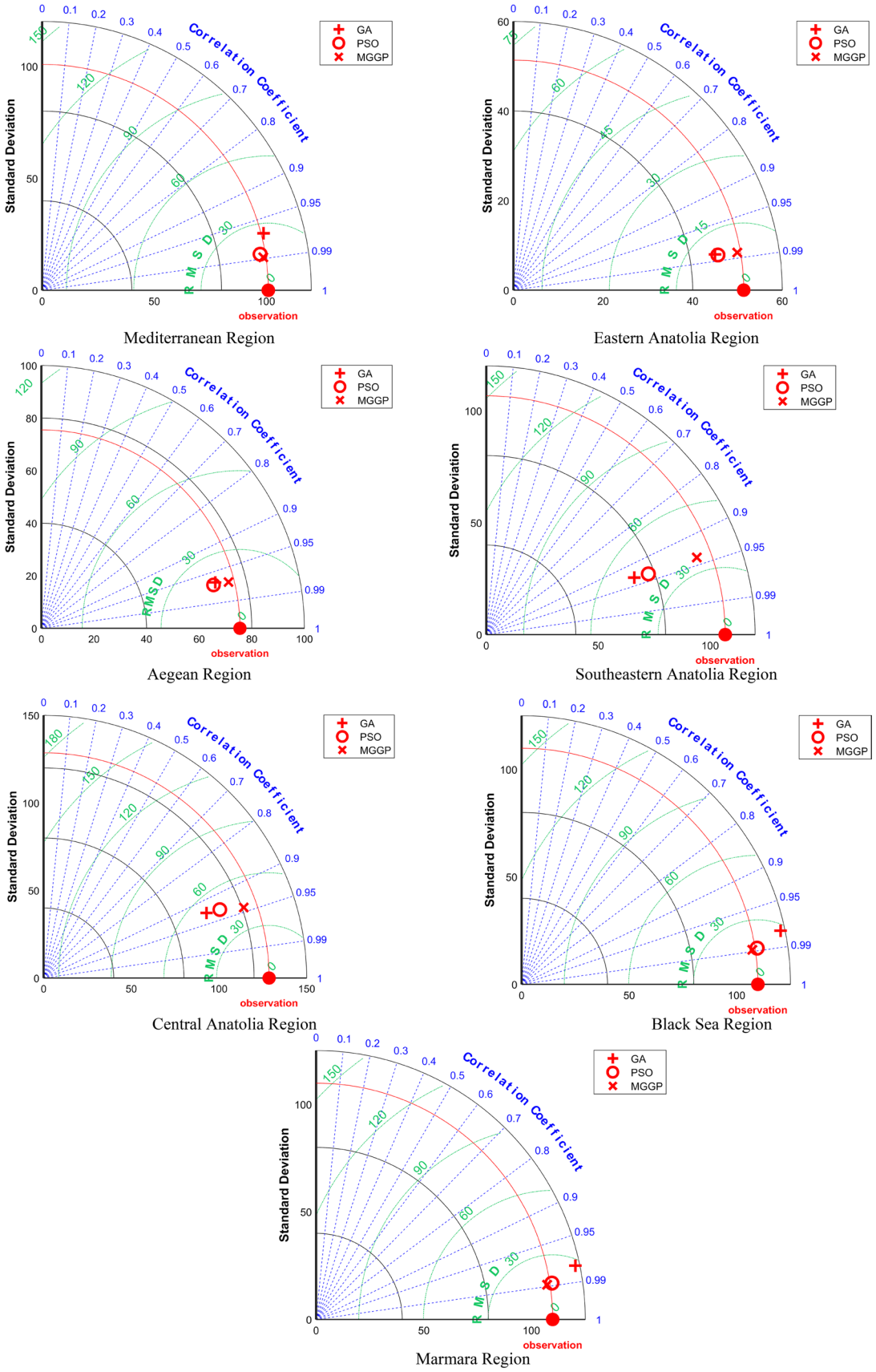

Fig. 4 Taylor diagrams for approaches

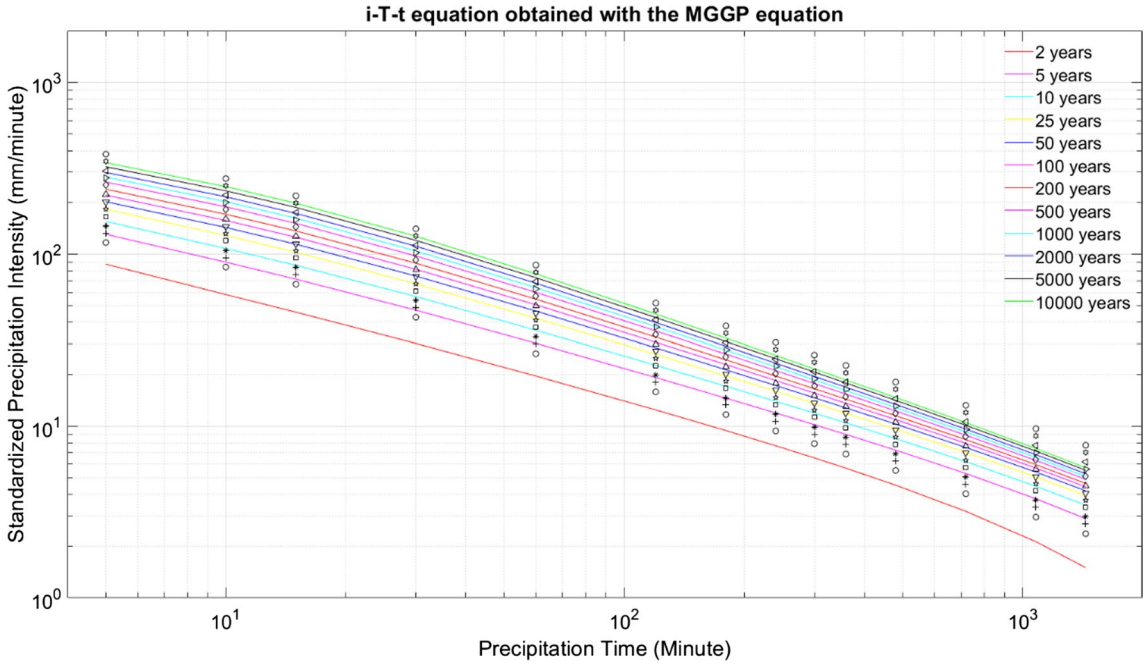

Fig. 5 Graphical presentation of numerical values obtained by MGGP model with IDF relationship for Marmara Region

Table 7 The p values of the null hypothesis (H_0) of the KW test at 95 percent confidence level

Region	Method	P value	Significance level	H_0*
Mediterranean region	PSO	0.8946	0.05	Reject
	GA	0.4499	0.05	Reject
	MGGP	0.9570	0.05	Reject
Eastern Anatolia region	PSO	0.5362	0.05	Reject
	GA	0.5517	0.05	Reject
	MGGP	0.8451	0.05	Reject
Aegean region	PSO	0.2909	0.05	Reject
	GA	0.4760	0.05	Reject
	MGGP	0.7598	0.05	Reject
Southeastern Anatolia region	PSO	0.5148	0.05	Reject
	GA	0.4889	0.05	Reject
	MGGP	0.8365	0.05	Reject
Central Anatolia region	PSO	0.5091	0.05	Reject
	GA	0.4648	0.05	Reject
	MGGP	0.9132	0.05	Reject
Black Sea region	PSO	0.1602	0.05	Reject
	GA	0.3005	0.05	Reject
	MGGP	0.8476	0.05	Reject
Marmara region	PSO	0.9307	0.05	Reject
	GA	0.9581	0.05	Reject
	MGGP	0.7682	0.05	Reject

coefficient, L-skewness coefficient, and L-moment ratios of the stations in the region. Görkemli et al. (2022) developed several new equations using the ABCP method using the regional curves obtained in Haktanir et al. (2016) in their study. The difference of this study from Görkemli et al. (2022) is that it makes frequency analysis for each station separately and determines the probability distribution for each station separately. It has been taken into account that the structure of the extreme precipitation data at each station will be different and this difference will cause different probability distributions. For this reason, it is aimed to obtain regional IDF equations by combining the IDF equations determined by each point frequency analysis. For this purpose, the coefficients of the 9 empirical equations used in the literature for the numerical determination of the IDF relationship were optimized with GA and PSO. In addition, new equations have been developed using the MGGP method, since it gives fast results for the IDF relationship by taking Görkemli et al. (2022) as an example. In addition, while only one region of Turkey is studied in the literature for Turkey, empirical equations for seven climate regions of Turkey have been optimized and new IDF equations have been developed in this study. It was investigated whether the IDF results obtained represent the original IDF relationships. One of the most important features that distinguishes the present study from other studies is the use of two different optimization techniques, which are widely used in the literature. By using two different optimization techniques, the study has been enriched and it has been possible to examine it from different

angles in global minimums. In addition, a comparison of these two techniques was made.

Conclusions

IDF equations have a very important place in water management and water structures design. In this study, the annual maximum precipitation data of 3 each meteorology stations belonging to 7 Regions of Turkey, which have semi-arid climates, were regionalized and examined. Frequency analysis was performed using probability distribution functions and parameter estimation methods in the literature. Precipitation intensity values obtained with 11 different equations whose parameters were estimated by PSO and GA and MGGP method were compared with precipitation intensity values obtained by frequency analysis. According to the performance metrics values, the best $i-t-T$ relationship was determined with the PSO method Eqs. 2, 3, 5, 6, 11 and the GA method Eqs. 2, 3, 4, 5, 11. According to the RMSE criterion, the most successful method is MGGP, followed by PSO and GA. Also, this result was supported Taylor diagrams and RMSD value. In KW suitability test performed for the methods, it was determined that the data came from a similar population in all regions and the methods gave successful results. As a result, MGGP method is more complex and more successful, PSO method is simpler and successful, and GA is less successful than other methods.

The four main limitations of this study can be mentioned as follows:i) the usage of three meteorological stations data representing each region in Turkey, (ii) using data extending up to 35 years, (iii) using two input parameters (precipitation times and frequencies) in IDF estimation, (iv) two different nature-inspired optimization techniques and using a machine learning method.

This study is an effort to estimate IDF, which is of great importance in sewerage and infrastructure works, in a more realistic way. In future studies, the accuracy of the regional study can be increased by providing more station data from each geographic region. In addition, newer equations can be obtained by adding parameters such as latitude, longitude, and altitude with the MGGP method.

Acknowledgements The data used in this study were obtained from the master's thesis titled "Determination of the Appropriate Probability Distribution Function and Formula of the Relationship Between the Period of Intensity-Rainfall Duration-Return Period for Standard Rainfall in Turkey" written by Kemal Yavuz. The authors thank Kemal Yavuz for providing data.

Funding No funding or support was received for research from any funding institutions.

Availability of data and material The data were provided by the General Directorate of State Meteorology Affairs (MGM).

Code availability The codes were written by Hatice Citakoglu and the analyses were done by Vahdettin Demir. Accessible MATLAB codes have been adapted to study. Code availability is not accessible.

Declarations

Conflicts of interest The authors declare no conflicts of interest.

Ethical approval In the study, the scientists paid close attention to ethical guidelines. There was no ethical infraction.

Consent for publication This manuscript can be published in the Acta Geophysica.

References

Acar R, Çelik S, Senocak S (2008) Rainfall intensity-duration-frequency (IDF) model using an artificial neural network approach. J Sci Ind Res (india) 67:198–202

Adarsh S, Janga Reddy M (2018) Developing hourly intensity duration frequency curves for urban areas in India using multivariate empirical mode decomposition and scaling theory. Stoch Environ Res Risk Assess 32:1889–1902. https://doi.org/10.1007/s00477-018-1545-x

Al-Amri NS, Subyani AM (2017) Generation of rainfall intensity duration frequency (IDF) curves for ungauged sites in arid region. Earth Syst Environ 1:8. https://doi.org/10.1007/s41748-017-0008-8

Al-Khalaf HA (1997) Predicting short duration, high-intensity rainfall in Saudi Arabia. Faculty of the college of graduate studies

Al-Shaikh AA (1985) Rainfall Frequency Studies for Saudi Arabia. M.S. Thesis, Dept. C.E., King Saud University, Riyadh, p 156

Al-Wagdany AS (2021) Construction of IDF curves based on NRCS synthetic rainfall hyetographs and daily rainfall records in arid regions. Arab J Geosci 14:527. https://doi.org/10.1007/s12517-021-06922-w

Alramlawi K, Fıstıkoğlu O (2022) Estimation of intensity-duration-frequency (IDF) curves from large scale atmospheric dataset by statistical downscaling. Tek Dergi 33:11591–11615. https://doi.org/10.18400/tekderg.874035

Aly A, Pathak C, Teegavarapu RSV et al (2009) Evaluation of Improvised spatial interpolation methods for infilling missing precipitation records. World environmental and water resources congress 2009. American Society of Civil Engineers, Reston, VA, pp 1–10

Tuğçe A, Ömer Y, Fatih S, Emrah O (2022) Rainfall intensity-duration-frequency analysis in Turkey, with the emphasis of eastern black sea basin. Teknik Dergi. https://doi.org/10.18400/tekderg.727085

Asikoglu OL, Benzeden E (2014) Simple generalization approach for intensity-duration-frequency relationships. Hydrol Process 28:1114–1123. https://doi.org/10.1002/hyp.9634

Awadallah AG, Magdy M, Helmy E, Rashed E (2017) Assessment of rainfall intensity equations enlisted in the Egyptian code for designing potable water and sewage networks. Adv Meteorol 2017:1–10. https://doi.org/10.1155/2017/9496787

Barbero R, Fowler HJ, Blenkinsop S et al (2019) A synthesis of hourly and daily precipitation extremes in different climatic regions. Weather Clim Extrem 26:100219. https://doi.org/10.1016/j.wace.2019.100219

Başakın EE, Ekmekcioğlu Ö, Özger M, Citakoglu H (2021) Determination of intensity-duration-frequency relation by particle swarm optimization and genetic programming. In: In II. International

Applied Statistics Conference (UYIK-2021). Tokat, Turkey, pp 1–8
Bell FC (1969) Generalized rainfall-duration-frequency relationships. J Hydraul Div ASCE 95:311–327
Bernard MM (1932) Formulas for rainfall intensities of long duration. Trans Am Soc Civ Eng 96:592–606. https://doi.org/10.1061/taceat.0004323
Borga M, Vezzani C, Fontana GD (2005) Regional rainfall depth–duration–frequency equations for an alpine region. Nat Hazards 36:221–235. https://doi.org/10.1007/s11069-004-4550-y
Buba LF, Kura NU, Dakagan JB (2017) Spatiotemporal trend analysis of changing rainfall characteristics in Guinea Savanna of Nigeria. Model Earth Syst Environ 3:1081–1090. https://doi.org/10.1007/s40808-017-0356-2
Bulti DT, Abebe BG, Biru Z (2021) Climate change–induced variations in future extreme precipitation intensity–duration–frequency in flood-prone city of Adama, central Ethiopia. Environ Monit Assess 193:784. https://doi.org/10.1007/s10661-021-09574-1
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 C (1983) Rainfall intensity-duration-frequency formulas. J Hydraul Eng 109:1603–1621. https://doi.org/10.1061/(asce)0733-9429(1983)109:12(1603)
Citakoglu H (2021) Comparison of multiple learning artificial intelligence models for estimation of long-term monthly temperatures in Turkey. Arab J Geosci 14:2131. https://doi.org/10.1007/s12517-021-08484-3
Citakoglu H, Babayigit B, Haktanir NA (2020) Solar radiation prediction using multi-gene genetic programming approach. Theor Appl Climatol. https://doi.org/10.1007/s00704-020-03356-4
Cook LM, McGinnis S, Samaras C (2020) The effect of modeling choices on updating intensity-duration-frequency curves and stormwater infrastructure designs for climate change. Clim Change 159:289–308. https://doi.org/10.1007/s10584-019-02649-6
Dastagir MR (2015) Modeling recent climate change induced extreme events in Bangladesh: a review. Weather Clim Extrem 7:49–60. https://doi.org/10.1016/j.wace.2014.10.003
Deb P, Babel MS, Denis AF (2018) Multi-GCMs approach for assessing climate change impact on water resources in Thailand. Model Earth Syst Environ 4:825–839. https://doi.org/10.1007/s40808-018-0428-y
Demir V (2022) Enhancing monthly lake levels forecasting using heuristic regression techniques with periodicity data component: application of Lake Michigan. Theor Appl Climatol 148:915–929. https://doi.org/10.1007/s00704-022-03982-0
Egodawatta P, Thomas E, Goonetilleke A (2007) Mathematical interpretation of pollutant wash-off from urban road surfaces using simulated rainfall. Water Res 41:3025–3031. https://doi.org/10.1016/j.watres.2007.03.037
Elbaz K, Shen SL, Zhou A et al (2019) Optimization of EPB shield performance with adaptive neuro-fuzzy inference system and genetic algorithm. Appl Sci. https://doi.org/10.3390/app9040780
Elbaz K, Shen SL, Sun WJ et al (2020) Prediction model of shield performance during tunneling via incorporating improved particle swarm optimization into ANFIS. IEEE Access 8:39659–39671. https://doi.org/10.1109/ACCESS.2020.2974058
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
Elsebaie IH, El Alfy M, Kawara AQ (2021) Spatiotemporal variability of intensity–duration–frequency (idf) curves in arid areas: wadi al-lith, Saudi Arabia as a case study. Hydrology 9:6. https://doi.org/10.3390/hydrology9010006
Eman Ahmed Hassan El-Sayed (2011) Generation of rainfall intensity duration frequency curves for ungauged sites. Nile Basin Water Sci Eng J 4:112–124
Eray O, Mert C, Kisi O (2018) Comparison of multi-gene genetic programming and dynamic evolving neural-fuzzy inference system in modeling pan evaporation. Hydrol Res 49:1221–1233. https://doi.org/10.2166/nh.2017.076
Ewea HA, Elfeki AM, Al-Amri NS (2017) Development of intensity–duration–frequency curves for the Kingdom of Saudi Arabia. Geomat Nat Hazards Risk 8:570–584. https://doi.org/10.1080/19475705.2016.1250113
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
Froehlich DC (1995) Long-duration–rainfall intensity equations. J Irrig Drain Eng 121:248–252. https://doi.org/10.1061/(asce)0733-9437(1995)121:3(248)
Galiatsatou P, Iliadis C (2022) Intensity-duration-frequency curves at ungauged sites in a changing climate for sustainable stormwater networks. Sustainability 14:1229. https://doi.org/10.3390/su14031229
Gandomi AH, Alavi AH (2012) A new multi-gene genetic programming approach to non-linear system modeling. Part II: geotechnical and earthquake engineering problems. Neural Comput Appl 21:189–201. https://doi.org/10.1007/s00521-011-0735-y
García-Bartual R, Schneider M (2001) Estimating maximum expected short-duration rainfall intensities from extreme convective storms. Phys Chem Earth, Part B Hydrol Ocean Atmos 26:675–681. https://doi.org/10.1016/S1464-1909(01)00068-5
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
Gen M, Cheng R (1997) Genetic algorithms and engineering design. John Wiley, Hoboken
Gen M, Cheng R, Lin L (2008) Network models and optimization: multiobjective genetic algorithm approach, 1st edn. Springer Publishing Company, Incorporated
Goldberg DE (1989) Genetic algorithms in search, optimization and machine learning, 1st edn. Addison-Wesley Longman Publishing Co., Inc, USA
Goldberg DE, Deb K (1991) A Comparative analysis of selection schemes used in genetic algorithms. In Found Genet Algorithms 1:69–93
Görkemli 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
Haktanir T, Citakoglu H, Seckin N (2016) Regional frequency analyses of successive-duration annual maximum rainfalls by L-moments method. Hydrol Sci J 61:647–668. https://doi.org/10.1080/02626667.2014.966722
Hamaamin YAH (2016) Developing of rainfall intensity-duration-frequency model for Sulaimani city. J Zankoy Sulaimani - Part A 19:93–102. https://doi.org/10.17656/jzs.10634
Hasan I, Saeed Y (2020) Analysis of rainfall data for a number of stations in northern Iraq. Al-Rafidain Eng J (AREJ)

25(2):105–117. https://doi.org/10.33899/rengj.2020.127531.1044
Hay JE, Easterling D, Ebi KL et al (2016) Conclusion to the special issue: observed and projected changes in weather and climate extremes. Weather Clim Extrem 11:103–105. https://doi.org/10.1016/j.wace.2015.11.002
Hayder AM, Al-Mukhtar M (2021) Modelling the IDF curves using the temporal stochastic disaggregation BLRP model for precipitation data in Najaf City. Arab J Geosci 14:1957. https://doi.org/10.1007/s12517-021-08314-6
Hershfield DM (1963) Estimating the probable maximum precipitation. Trans Am Soc Civ Eng 128:534–551. https://doi.org/10.1061/taceat.0008684
Karahan H, Ayvaz MT, Gürarslan G (2008) Determination of intensity-duration-frequency relationship by genetic algorithm: case study of GAP. Tek Dergi/technical J Turkish Chamb Civ Eng 19:4393–4407
Karahan H, Ceylan H, Tamer Ayvaz M (2007) Predicting rainfall intensity using a genetic algorithm approach. Hydrol Process 21:470–475. https://doi.org/10.1002/hyp.6245
Kareem DA, Rahman A, Amen M et al (2022) Comparative analysis of developed rainfall intensity-duration-frequency curves for Erbil with other Iraqi Urban Areas. Water 14:1–17. https://doi.org/10.3390/w14030419
Kennedy J, Eberhart R (2010) Particle swarm optimization. In: Proceedings of ICNN'95 - International Conference on Neural Networks. IEEE, pp 1942–1948
Jaleel LA, Farawn MA (2013) Developing rainfall intensity-duration frequency relationship for Basrah city. Kufa J Eng 5:105–112
Legouhy A (2021) Al_goodplot - boxblot & violin plot. In: MATLAB Cent. Mathworks. https://www.mathworks.com/matlabcentral/fileexchange/91790-al_goodplot-boxblot-violin-plot
Lestari S, King A, Vincent C et al (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
Liew S, Raghavan SV, Liong S-Y (2014) Development of intensity-duration-frequency curves at ungauged sites: risk management under changing climate. Geosci Lett 1:8. https://doi.org/10.1186/2196-4092-1-8
Lopcu Y (2007) Modeling the intensity−duration−frequency relationships of annual maximum storms. Dokuz Eylul University
Mahdi ES, Mohamedmeki MZ (2020) Analysis of rainfall intensity-duration-frequency (IDF) curves of Baghdad city. IOP Conf Ser Mater Sci Eng 888:012066. https://doi.org/10.1088/1757-899X/888/1/012066
Matlab (2022a) Particleswarm. In: Introd. R2014b. https://www.mathworks.com/help/gads/particleswarm.html
Matlab (2022b) Genetic Algorithm. https://www.mathworks.com/help/gads/genetic-algorithm.html
Matlab (2022c) Matlab. In: MATLAB Cent. Mathworks. https://www.mathworks.com/help/stats/kruskalwallis.html
MGM (2022) Annual areal precipitation in Turkey. In: Turkish state Meteorol. Serv. https://mgm.gov.tr/veridegerlendirme/yillik-toplam-yagis-verileri.aspx
Mirhosseini G, Srivastava P, Stefanova L (2013) The impact of climate change on rainfall intensity–duration–frequency (IDF) curves in Alabama. Reg Environ Chang 13:25–33. https://doi.org/10.1007/s10113-012-0375-5
Moujahid M, Stour L, Agoumi A, Saidi A (2018) Regional approach for the analysis of annual maximum daily precipitation in northern Morocco. Weather Clim Extrem 21:43–51. https://doi.org/10.1016/j.wace.2018.05.005
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
Shaban WM, Elbaz K, Yang J, Shen SL (2021) A multi-objective optimization algorithm for forecasting the compressive strength of RAC with pozzolanic materials. J Clean Prod 327:129355. https://doi.org/10.1016/j.jclepro.2021.129355
Searson DP, Leahy DE, Willis MJ (2010) GPTIPS: an open source genetic programming toolbox for multigene symbolic regression. In: In Proceedings of the International multiconference of engineers and computer scientists Citeseer. pp 77–80
Searson DP (2009) GPTIPS: Genetic programming and symbolic regression for MATLAB
Şen O, Kahya E (2021) Impacts of climate change on intensity–duration–frequency curves in the rainiest city (Rize) of Turkey. Theor Appl Climatol 144:1017–1030. https://doi.org/10.1007/s00704-021-03592-2
Şen Z (2019) Annual daily maximum rainfall-based IDF Curve derivation methodology. Earth Syst Environ 3:463–469. https://doi.org/10.1007/s41748-019-00124-x
Shahid S, Wang X-J, Bin HS et al (2016) Climate variability and changes in the major cities of Bangladesh: observations, possible impacts and adaptation. Reg Environ Chang 16:459–471. https://doi.org/10.1007/s10113-015-0757-6
Sillmann J, Thorarinsdottir T, Keenlyside N et al (2017) Understanding, modeling and predicting weather and climate extremes: challenges and opportunities. Weather Clim Extrem 18:65–74. https://doi.org/10.1016/j.wace.2017.10.003
Stephenson AG, Lehmann EA, Phatak A (2016) A max-stable process model for rainfall extremes at different accumulation durations. Weather Clim Extrem 13:44–53. https://doi.org/10.1016/j.wace.2016.07.002
Subyani AM, Al-Amri NS (2015) IDF curves and daily rainfall generation for Al-Madinah city, western Saudi Arabia. Arab J Geosci 8:11107–11119. https://doi.org/10.1007/s12517-015-1999-9
Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. J Geophys Res Atmos 106:7183–7192. https://doi.org/10.1029/2000JD900719
Tyralis H, Langousis A (2019) Estimation of intensity–duration–frequency curves using max-stable processes. Stoch Environ Res Risk Assess 33:239–252. https://doi.org/10.1007/s00477-018-1577-2
Uncuoğlu E, Latifoğlu L, Özer AT (2021) Modelling of lateral effective stress using the particle swarm optimization with machine learning models. Arab J Geosci 14:2441. https://doi.org/10.1007/s12517-021-08686-9
VOSviewer welcome to VOSviewer. In: 2022 Cent Sci Technol Stud Leiden Univ Netherlands. https://www.vosviewer.com/
Voyant C, Notton G, Kalogirou S et al (2017) Machine learning methods for solar radiation forecasting: a review. Renew Energy 105:569–582. https://doi.org/10.1016/j.renene.2016.12.095
Yavuz K (2018) Determination of the appropriate probability distribution function and formula of the relationship between the period of intensity-rainfall duration-return period for standard rainfall in Turkey. Erciyes University
Yilmaz AG, Hossain I, Perera BJC (2014) Effect of climate change and variability on extreme rainfall intensity–frequency–duration relationships: a case study of Melbourne. Hydrol Earth Syst Sci 18:4065–4076. https://doi.org/10.5194/hess-18-4065-2014

Yu PS, Yang TC, Lin CS (2004) Regional rainfall intensity formulas based on scaling property of rainfall. J Hydrol 295:108–123. https://doi.org/10.1016/j.jhydrol.2004.03.003

Zahiri E-P, Bamba I, Famien AM et al (2016) Mesoscale extreme rainfall events in West Africa: the cases of Niamey (Niger) and the Upper Ouémé Valley (Benin). Weather Clim Extrem 13:15–34. https://doi.org/10.1016/j.wace.2016.05.001

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