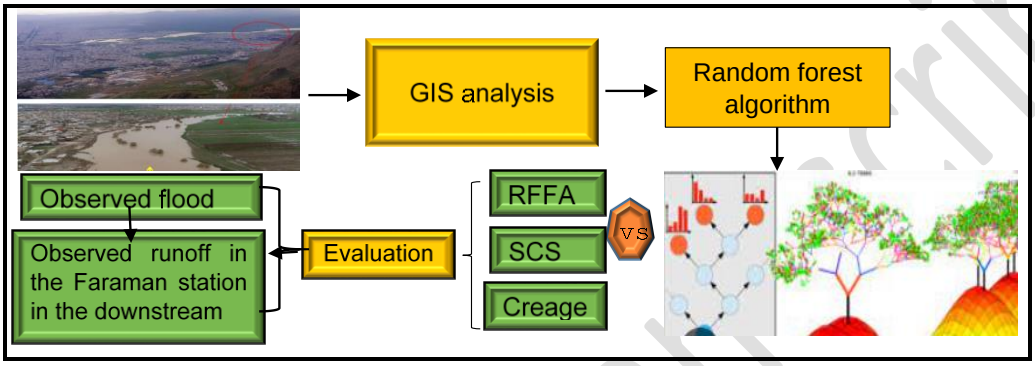
Journal of Applied Research in Water and Wastewater
E-ISSN: 2476-6283
Journal homepage: https://arww.razi.ac.ir

Comparing random forest flood frequency analysis with regional flood frequency, Creager and SCS in Doab-Qazanchi, Kermanshah

Maryam Hafezparast Mavaddat*, Sadaf Gord, Rasool Ghobadian

Water Engineering Department, Faculty of Agriculture, Razi University, Kermanshah, Iran.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article type:
Research Article

Article history:
Received xx Month xxx
Received in revised form xx Month xxx
Accepted xx Month xxx
Available online x Month xx

Keywords:
Creager
Doab-Qazanchi
Flood
Regional flood frequency analysis (RFFA)
Random forest (RF)
Soil conservation service (SCS)

Publisher: Razi University

ABSTRACT

Due to climate change and rising global temperature, the occurrence of extreme floods and drought events has intensified. In this regard, in 2019, heavy rainfall occurred in Kermanshah province. The Gharasoo river runs through the city of Kermanshah in western Iran. The Doab-Qazanchi area is located on the Gharasoo River at the crossroads of the Razavar and mereg Rivers to the Gharasoo River and there is no hydrometric station in this area. In this research, floods with different return periods of 2, 5, 20, 50, 100, 200, 500 and 1000 years with Creager and regional flood frequency analysis (RFFA), and the random forest machine learning method using the physical and hydrological characteristics of the surrounding watersheds are predicted. The SCS method was implemented for the flood on 03/04/2019 and it showed that the occurred flood is equivalent to a 25-year flood in this region. The predicted values estimated a lower discharge than the soil conservation service (SCS) method. The random forest (RF) method, as a machine learning method compared to old statistical methods, has a good performance in predicting the flood discharge using the physical and hydrological indicators of the catchment area, and by determining the priority of different features, it predicts the flood discharge well.

1. Introduction

Floods are among the most devastating natural disasters, causing significant damage to infrastructure, the environment, and human life. Accurate flood frequency analysis (FFA) is crucial for effective flood risk management, infrastructure design, and emergency planning. Traditional methods such as the Creager, soil conservation service (SCS), and regional flood frequency analysis (RFFA) have been widely used for predicting flood events. However, these methods often rely on assumptions that may not hold true in all scenarios, leading to inaccuracies in flood predictions.

In recent years, machine learning techniques have emerged as powerful tools for analyzing complex environmental data. The RF method, in particular, has shown promise in various hydrological applications due to its ability to handle large datasets and model non-linear relationships. This study aims to compare the performance of the RF method with traditional FFA methods in the Doab-Qazanchi region of Kermanshah, Iran, an area prone to significant rainfall and urban development.

By evaluating the strengths and limitations of each approach, this research seeks to enhance flood prediction accuracy and inform better flood management strategies (Kundzewicz, 2012). Thus, for flood risk management purposes, periodic assessment of rivers is essential, especially regarding long-term discharge patterns. Accurate flood prediction enables emergency management agencies to develop effective response plans. By knowing the potential magnitude and timing of flood events, authorities can allocate resources, evacuate vulnerable areas, and coordinate emergency services in a timely manner. This improves the overall preparedness and response capabilities, ultimately saving lives and minimizing the impact of floods on affected communities (Gavrilović, Milanović Pešić, and Urošev, 2012; Hafezparast Mavadat and Marabi, 2021; Hamaamin *et al*., 2022).

*Corresponding author Email: m.hafezparast@razi.ac.ir

Cite this article: M. Hafezparast Mavaddat, S. Gord, R. Ghobadian, Comparing random forest flood frequency analysis with regional flood frequency, Creager, and SCS in Doab-Qazanchi, Kermanshah, *Journal of Applied Research in Water and Wastewater*, x (x), 202x, xx-xx. doi: https://doi.org/10.22126/arww.2025.11095.1349

The FFA is a dimensionless method for determining the relationship between the magnitude of peak flow events and their frequency using probability distributions derived from observed flow data at various gauge stations along the river (Topaloglu, 2005; Shahabi and Hessami Kermani, 2015). Analyzing flood frequency is particularly important for rivers like the Gharasoo River in Kermanshah province, which flows through the city. The return period is a crucial hydrological tool for estimating the time interval between events of similar size or intensity. However, estimating this return period can be challenging due to issues such as missing data, short data series, or unknown probability distribution functions (Oosterbaan, 1994).

The RFFA method is being utilized across Europe, with research conducted in Germany (Bormann, Pinter and Elfert, 2011), Poland (Rutkowska *et al.,* 2017), Norway (Hailegeorgis and Alfredsen, 2017), and the Danube River basin (Morlot, Brilly and Šraj, 2019). This method enhances accuracy and provides opportunities for further development in decision-making. Sharifi Garmdareh, Vafakhah and Eslamian (2018) applied RFFA to data from 55 hydrometric stations in the Namak Lake Basin, Iran, from 1992 to 2012, calculating flood discharges for specific return periods using the log PIII distribution, deemed the best regional option. They extracted physiographic, meteorological, geological, and land use variables to predict peak flood discharges for return periods of 2, 5, 10, 25, 50, and 100 years using SVR, ANFIS, ANN, and NLR. The GT + ANFIS and GT + SVR models outperformed both ANN and NLR in RFFA.

Recently, machine learning methods have been significantly applied in the estimation of hydrological parameters. Many studies have been done in the estimation of peak discharge with these methods, among which we can refer to (Allahbakhshian-Farsani *et al*., 2020; Mosavi Ozturk and Chau, 2018; Gizaw and Gan, 2016; Al-Fawa'Reh *et al*., 2021; Sharifi Garmdareh, Vafakhah and Eslamian, 2018).

The study by Allahbakhshian-Farsani *et al.,* (2020) analyzed data from fifty-four hydrometric stations over 21 years (1993-2013) in the Karun and Karkheh watersheds of southwest Iran to estimate flood discharge using various statistical methods. The RFFA was implemented according to US Federal Agencies Bulletin 17 B, selecting GNO PDF through the L-moment method from multiple PDFs. The researchers identified twenty-five predictive variables related to physiography, climate, geology, and land use as suitable inputs via GP. Results revealed that the SVR, PPR, and MARS models provided more accurate flood discharge estimates for expected return periods compared to NLR and BRT.

In Wadi Al Wala, Jordan, a study utilized 13 rain gauge stations with 38 years of daily data for real-time rainfall forecasting and flood control. ML methods were assessed, with DT and RF achieving the best flood forecasts (Al-Fawa'Reh *et al.,* 2021). Additionally, Lee *et al*., (2020) developed an ML-based model for estimating design floods in ungauged watersheds, enhancing the design rainfall–runoff analysis. The model focused on flood prediction by frequency and demonstrated that the RF algorithm significantly reduced errors, achieving an average design flood estimation accuracy of 99%.

Several studies have explored different flood discharge methods based on observed watersheds. Mustamin, Maricar and Karamma. (2021) compared the Creager and SCS methods to determine the most suitable artificial unit hydrograph and peak discharge in a specified area. Their calculations, based on rainfall data using synthetic unit hydrographs of Nakayasu, ITB I, ITB II, and SCS, indicated that the SCS method closely matched design flood discharge with measured discharge and Q1000 from the Creager method. Salami *et al.* (2017) developed runoff hydrographs for rivers in the Ogun-Osun river catchment, Nigeria, utilizing Snyder and SCS methods to determine discharges for various return periods: 20-yr (112.63 m³/s), 50-yr (e.g., 13364.30 m³/s), 100-yr, 200-yr, and 500-yr. The SCS method provided discharge values ranging from 304.43 m³/s to 6466.84 m³/s across eight watersheds. It is recommended due to its consideration of morphometric parameters, such as basin slope and Curve Number (CN), alongside the characteristics of soil and vegetation in estimating peak flow. Flooding in the Doab_Qazanchi region is mostly due to the flow of moist western air masses or the rapid melting of snow from the Zagros Mountains, which mostly occurs in April. In addition, the erosion and connection of the Razavar and Mereg Rivers cause much flooding in this area. In the April 2019, the amount of rainfall in the Kermanshah synoptic station was 67mm on 1/4/2019 and the peak flow in the Faraman hydro station, downstream of the selected area was 490 m^3/s on 3/4/2019. So, this area was impacted by a flood event unparalleled in the hydrological records of the region. There isn't any hydro station in Doab_Qazanchi, so the goal of this paper is to estimate flood hydrograph and flood frequency based on the most cited empirical methods, SCS, RFFA with index FFA, and Creager, in previous research and compare it to the RF ML method (Al-Fawa'Reh *et al*., 2021; Salami *et al.,* 2017; Morlot *et al*., 2019).

The objective of this research is to estimate flood hydrographs and flood frequency in the Doab-Qazanchi region, which lacks hydrometric stations. The study aims to address the knowledge gap in flood prediction methods by comparing the performance of traditional methods (Creager, SCS, and RFFA) with RF ML method. By utilizing these methods, the research aims to provide accurate flood discharge values for different return periods (DRPs) in the absence of direct measurements in the study area.

The significance of this study lies in the importance of accurate flood prediction for effective hydrological and hydraulic planning, particularly in flood control projects. By estimating flood hydrographs and frequency, the research contributes to improved flood risk management in the Doab-Qazanchi region. The comparison of different methods allows for the evaluation of their performance and identification of the most suitable approach for flood prediction in the absence of hydrometric stations.

Overall, this research fills the knowledge gap by exploring the applicability of ML method in flood frequency analysis and comparing them with traditional methods. The findings of this study will provide valuable insights for researchers, practitioners, and decision-makers involved in flood control and management efforts in the Doab-Qazanchi region and similar areas.

2. Material and methods

2.1. Study area

The watershed is a part of the Gharasoo River watershed, which is a part of the Karkheh watershed and is located in Kermanshah province in Iran. It's located between 34° 22´´ to 34° 55´ 10´´ latitude and 46° 22´ 12´´ to 47° 22´ 12´´ of eastern longitude. This basin is limited to the Gaveroud, Ravand, Zemkan, and Gamasyab watersheds, from the north, south, west, and east, respectively. The most important rivers in the basin are Gharasoo, Razavar, and Merg. Among the important urban areas, we can mention the city of Kermanshah (the center of the province) and the cities of Javanrud, Ravansar, Mahidasht, and Kozaran, and among the important tourist areas are the ancient monuments of Taq-e Bostan. The basin area up to the exit point, i.e. Pol-Kohneh hydrometric station, is equal to 5340 km^2. Among the important heights in the basin are Sefidkoh, Weiss, Kehjar, Shahkoh, and Mila with a height of 2486, 1865, 1644, 2486, and 2440 meters, respectively. The annual rainfall in the Kermanshah Plain is between 400 and 500 mm and in the heights of the basin, it is 600 to 660 mm per year. Among the hydrometric stations on the studied rivers are Doab Merg, Khersabad, Hojatabad, and Pol-kohne stations in Fig.1.

2.2. Data preparation

Historical IMF and maximum daily discharge from 1966 to 2019 for five stations around Doab-Qazanchi area and their basins characteristics are used in this research. Geographic information of basins is presented in Table 1.

Table 1. Hydrometry stations.

Hydrometric station	River	Geographic information			Area, km^2
		Longitude	Latitude	Altitude	
Khersabad	Mereg	46° 44´	34° 30´	1320	1460
Doabmereg	Gharasoo	46° 47´	34° 33´	1290	1243
Hojatabad	Razavar	46° 00´	34° 29´	1290	1338
Pol-kohne	Gharasoo	46° 08´	34° 19´	1260	5340
Faraman	Gharasoo	46° 15´	34° 14´	1260	5370

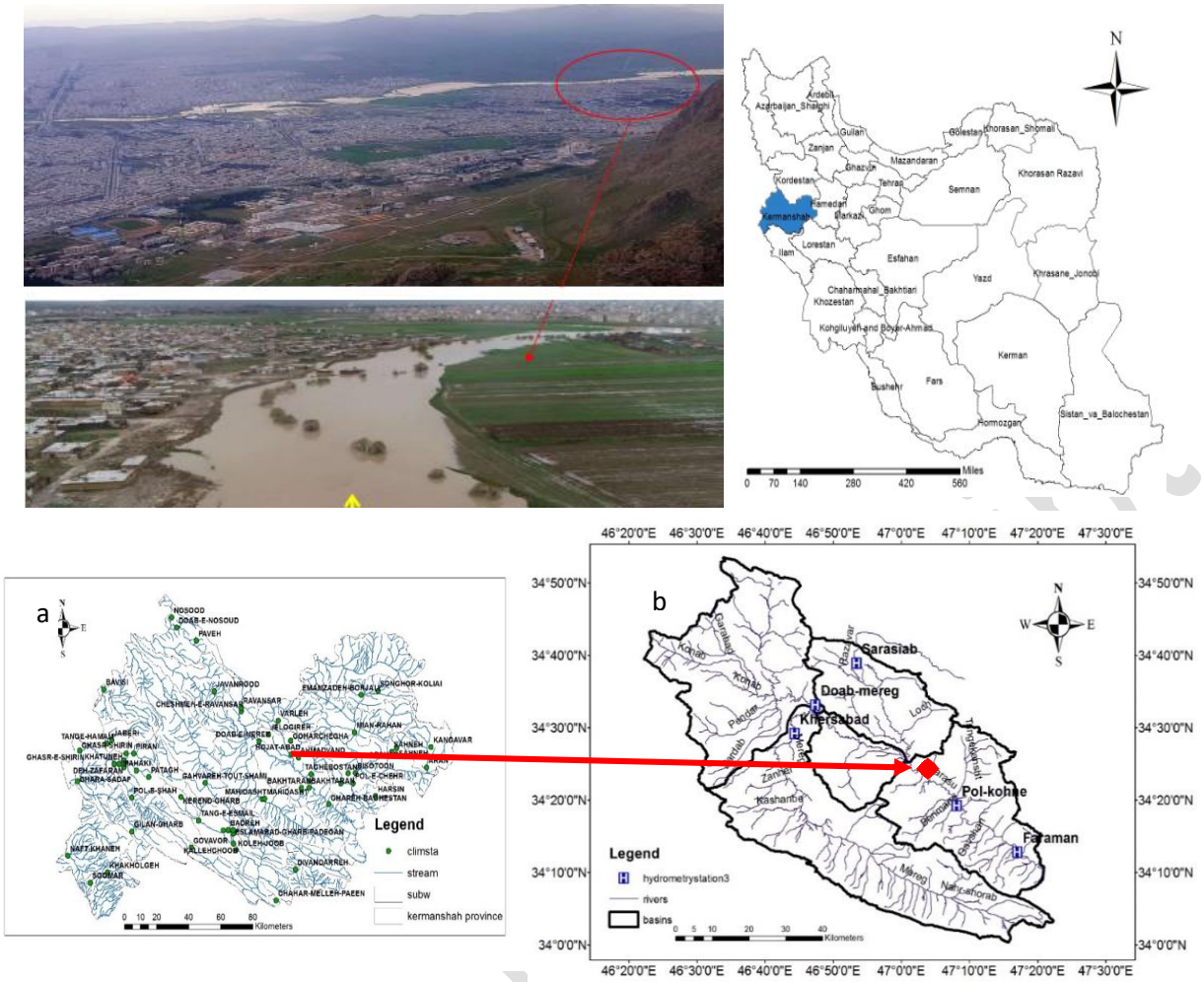


Fig. 1. Kermanshah province and watersheds (a), Gharasoo river and hydro stations including red point as the DoabQazanchi area (b).

2.3. Outlier, randomness, trend and hurst analysis

Due to the historical nature of the flood event, the instantaneous maximum flow of the flood in the hydrometric stations does not need an outlier detection test. The homogeneity of the data is that the data are related to a certain random statistical population. To check the randomness of the data, the run test method was used in the Minitab software.

While the presence of a significant trend in a climatic variable time series does not conclusively prove climate change in a region, it does support the assumption of its occurrence. This is because the climate system is influenced by multiple controlling factors. The presence or absence of trends and analysis of time series and climate change are divided into two categories: parametric and non-parametric methods.

The nonparametric Mann-Kendall test, initially introduced by Mann (1945) and further developed by Kendall (1975), ranks data in a time series to analyze trends in hydrological and meteorological series. This widely used method offers an advantage in its ability to handle extreme values observed in certain time series. The null hypothesis of this test indicates randomness and the absence of a trend in the data series, and accepting the null hypothesis (rejection of the null hypothesis) indicates the existence of a trend in the data series. This method was implemented with the Pymannkendall python package. For removing the trend, the linear regression differencing method was used.

Hurst coefficient is a statistic coefficient to measure the adequacy of information in terms of the length of the statistical period. This coefficient is used to measure the long-term memory of a time series (Hurst, 1951). Different probability distribution functions have been fitted to the constructed discharge data of each station.

2.4. Standardization of the data

Before the training of the RF model, both input and output variables were normalized within the range of 0.1 to 0.9 as Eq.1.

$$N_i = 0.8 \times \frac{(x_i - x_{min})}{(x_{max} - x_{min})} + 0.1 \quad (1)$$

Where, N_i is the normalized value of a certain parameter, x_i is the measured value, and x_{min} and x_{max} are the minimum and maximum values in the dataset, respectively (Dogan *et al.,* 2010).

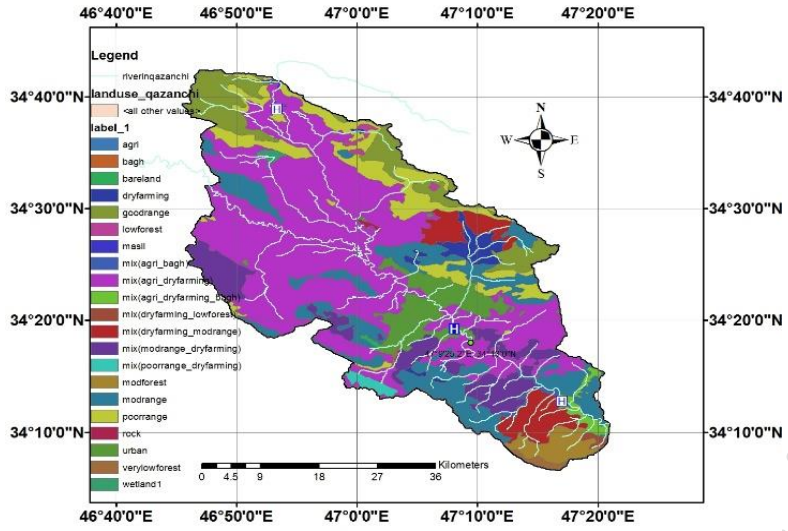
2.5. Random forest

RF consists of multiple decision trees. Each decision tree is trained and predicts outcomes based on training data. The unique aspect of RF is that it generates multiple training datasets and decision trees through diverse training, and by combining the results, the predictive power is enhanced (Lee *et al*., 2020). RF feature selection is used to select the best independent variables for modeling. The importance of a feature is calculated according to its ability to increase the purity of leaves in each tree of the RF. The higher the purity of the leaves, the higher the importance of the trait. This is done for each tree, then averaged across all trees, and finally normalized to one. Therefore, the sum of the importance scores calculated by a RF is one. For this reason, RFECV from feature selection in the Sklearn library in the python environment is considered (Chen *et al.,* 2020).

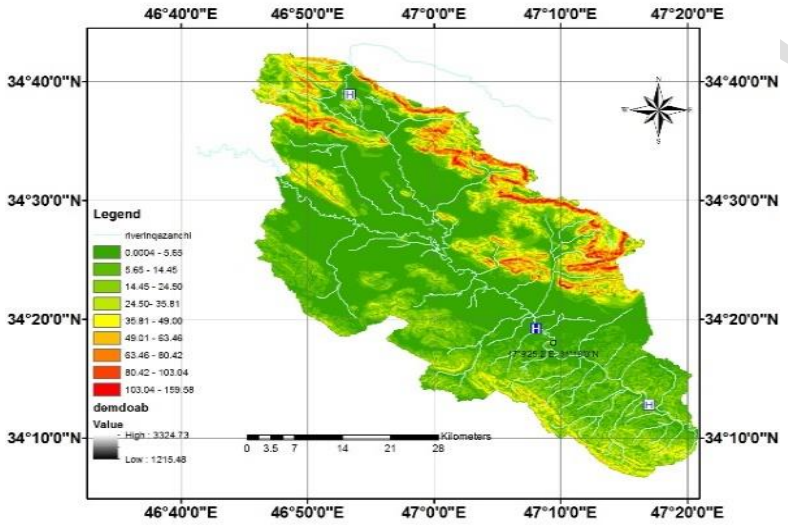
To calculate FFA in Doab Qazanchi using RF, 12 features, including Watershed Characteristics extracted in the GIS environment, cumulative three days rainfall, and maximum and minimum of maximum daily discharge for five stations in the area, are considered (Eq.2, Table 2, and Fig. 2).

$$Q_{T,i} = f(x_{1i}, x_{2i}, ..., x_{13i}) \quad (2)$$

Where, T is DRPs, i is the certain hydrometric station, $x_1$to $x_{13}$are basin features.



(a)



(b)

Fig. 2. Landuse map (a) and slop map (b) for the study area.

Table 2. Watershed characteristics.

Basin features	Sub-basin				
	Khersabad	**Doabmereg**	**Hojatabad**	**Pol-kohne**	**Faraman**
x_1: basin circumference (km)	364.42	285.87	281.67	591.87	680
x_2: Gravelius factor	2.67	2.29	2.01	2.26	1.46
x_3: Length of Watershed (km)	121.85	74.302	96.11	172.98	207.8
x_4: Shape factor	3.83	1.65	2	1.33	5.75
x_5: Circulatory ratio	0.13	0.18	0.24	0.19	0.47
x_6: Elongation ratio	0.35	0.52	0.45	0.47	0.4
x_7: Basin average Slop (%)	7.98	13.53	18.16	13.34	7.8
x_8: time of concentration	22.66	13.99	12.42	24.66	30.96
x_9: Mean elevation (m)	1591.1	1547.7	1651.5	1561.97	1450
x_{10}: Area (km^2)	1456.06	1314.7	1526.72	5339.2	5460

2.6. Flood frequency analysis

Generally, the steps followed in flood frequency analysis are as follows: Selection of the data including, instantaneous peak flow, Area, CN, annual maxima daily rainfall (Ghanbarpour *et al*., 2011). In addition, the three-day rain flood discharge is the most critical duration for designing and evaluating flood mitigation (National Research Council, 1999). Step 2: Fitting the probability distribution. Step 3: Goodness of fit test to identify the best fitting distribution. Software development for statistical extreme value analysis has been rapid (Gilleland, Ribatet, and Stephenson, 2013). Different program packages and pre-defined Excel sheets are used to perform frequency analysis. In between the most used of them are: RAINBOW (Raes, Mallants, and Song, 1970); Peak FQ (Flynn, Kirby, and Hummel, 2006); Hydrognomon (Kozanis *et al*., 2010); HYFRAN (El Adlouni and Bobée, 2015) and, the EASY-FIT (Schittkowski, 1980). Since RFFA estimates strongly depend on the shape of the selected distribution of the manufacturer's data, accurate models are needed to determine the best distribution. Therefore, EASY-FIT model and Index flood frequency method have been chosen to determine the best distributions in each station and calculate discharge in DRPs in Doab-Qazanchi.

2.7. Creager method

Creager is a kind of assessment of specific flood and this method delivered nonlinear equations based on a relationship between the basin area and PMF. Creager's equation in the metric system is given by the following formula Eq.3 (Creager, Justin, and Hinds, 1945).

$$q = 0.503C(0.368A)^{(0.936A^{-0.48}-1)} \quad (3)$$

A is the basin area in Km^2, c is the Creager coefficient, q is the specific discharge in $m^3/s/km^2$, and the specific discharge value is obtained from the following relationship Eq. 4.

$$q = \frac{Q}{A} \quad (4)$$

Q is the annual IMF in m^3/s for each station per return period Tr (year) that can be calculated according to the selected distribution.

2.8. SCS method

Soil Conservation Service uses this method, which is called the SCS method or dimensionless unit hydrograph, Presented in Eqs.5-6. It showed that the model can be used on any type of urban, natural, and mixed watershed. Cumulative rainfall in synoptic stations around Doab-Qazanchi is considered as an input data for the SCS method and is presented in Fig. 3.

$$s = 25.4\left(\frac{1000}{CN} - 10\right) \quad (5)$$

$$Q = \frac{(P - 0.2S)^2}{(P - 0.8S)} \quad (6)$$

S is the potential maximum retention (mm), CN is the curve number and it is taken from a table that is related to soil group, surface cover, and antecedent moisture condition (USDA, 1972) and it's compared to studies conducted by Karkouti *et al.,* (2010), So the best value for CN is 73. P is the cumulative rainfall (mm) and Q is the runoff (mm).

2.9. Performance assessment

The RMSE, MAE and the NSE is a normalized statistic that determines the relative magnitude of the residual variance compared to the measured data variance (Nash and Sutcliffe, 1970). Were used as predictive performance evaluation indices for the design flood estimation model and presented in Eqs. 7-9.

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T}(P_t^{est} - P_t^{obs})^2}{T}} \quad (7)$$

$$MAE = \frac{\sum_{t=1}^{T}|P_t^{est} - P_t^{obs}|}{T} \quad (8)$$

$$\text{NSE}=1-\frac{\sum_{t=1}^{T}(P_t^{est}-P_t^{obs})^2}{\sum_{t=1}^{T}(P_t^{obs}-\overline{P_t^{obs}})^2} \quad (9)$$

In the above relationships, P_{obs} and P_{est} are the observed and estimated discharge, respectively.

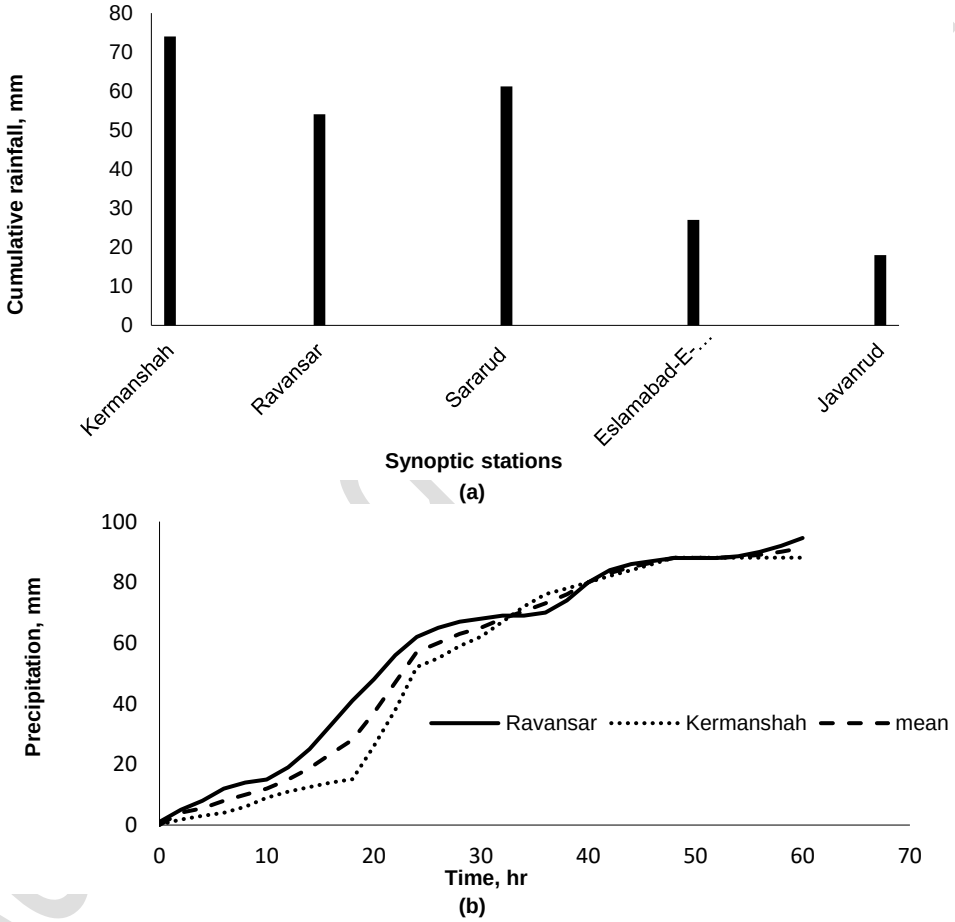


Fig. 3. Cumulative precipitation (Left) and cumulative rainfall in synoptic stations from 01/04/2019 to 03/04/2019 (right).

3. Results and discussion

3.1. Data analysis

Controlling the randomness of the instantaneous maximum discharge of the hydrometric stations of the study area using the Run test in Minitab software shows that all data are random in current stations at the level of 5% and it's presented in Table 3. Pymankendall package in the python environment shows that all stations have decreasing trends and the trend equations are shown in the last column of Table 4. The trend was removed by trendline differencing. At the end of FFA calculations, the trend parts were added to the flood values. For completing data in the pol-kohne station with Hurst coefficient < 0.5 the linear regression between instantaneous maximum flow in Faraman station as a reference station and, Pole-kohne station calculated with R^2 = 0.97 and the required data was constructed Figs.4-5.

Table 3. Instantaneous maximum discharge randomness in all stations.

Station	Number of runs		P-value
	Observed	Expected	
Khersabad	20	25.3	0.11
Doabmereg	20	23.66	0.24
Hojatabad	19	19.35	0.607
Pol-kohne	17	21.3	0.12
Faraman	19	20.37	0.607

Table 4. Trend analysis with pymankendall package for hydrometric stations.

Stations	Trend	Z	P-value	Trend line parameters
Khersabad	Decreasing	-4.069	4.714e-05	Slope=-1.24, intercept=71.74
Doabmereg	Decreasing	-4.15	3.3162e-05	Slope=-1.0969, intercept=61.69
Hojatabad	Decreasing	-3.33	0.0009	Slope=-2.2819, intercept=154.57
Pol-kohne	Decreasing	-3.47	0.0005	Slope=-2.375, intercept=160.375
Faraman	Decreasing	-2.25	0.0244	Slope=-1.026, intercept=98.65

The quadratic equation for maximum daily flood between Doab mereg and Khersabad stations was created with R^2=0.8 and required data was constructed for khersabad station, then the linear regression with R^2=0.85 is used to complete IMF in khersabad station.

3.2. Feature selection and random forest

Once the importance of each feature is determined, feature selection is done using a procedure called recursive feature elimination. In this paper, the k-fold CV version is used. The procedure is to remove the less relevant feature after fitting the model and calculate the average value of some performance measures in the CV. Then the last important feature is removed and the model is run again and the average performance is calculated. This procedure continues until there are no features left. The set of features that maximizes performance in CV is the set of features that are selected. The whole procedure should work with the same values for the hyperparameters. Fig.6 presented the importance of all stations characteristics based on RFECV. As it is clear, the shape factor feature is the most important one, followed by 72hr rainfall, elongation ratio, and the other features. Finally, the discharge in DRPs was predicted based on selected features for Doab-Qazanchi with a trained and tested RF model. The results presented in Tables 5 and 6 showed that the model is well-trained and tested but the flood predictions in all DRPs are mostly less than other methods.

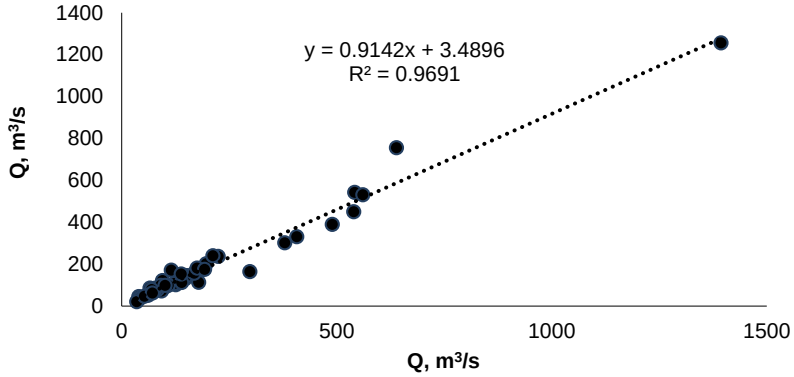


Fig. 4. IMF regression in the Faraman and Pole-kohne stations.

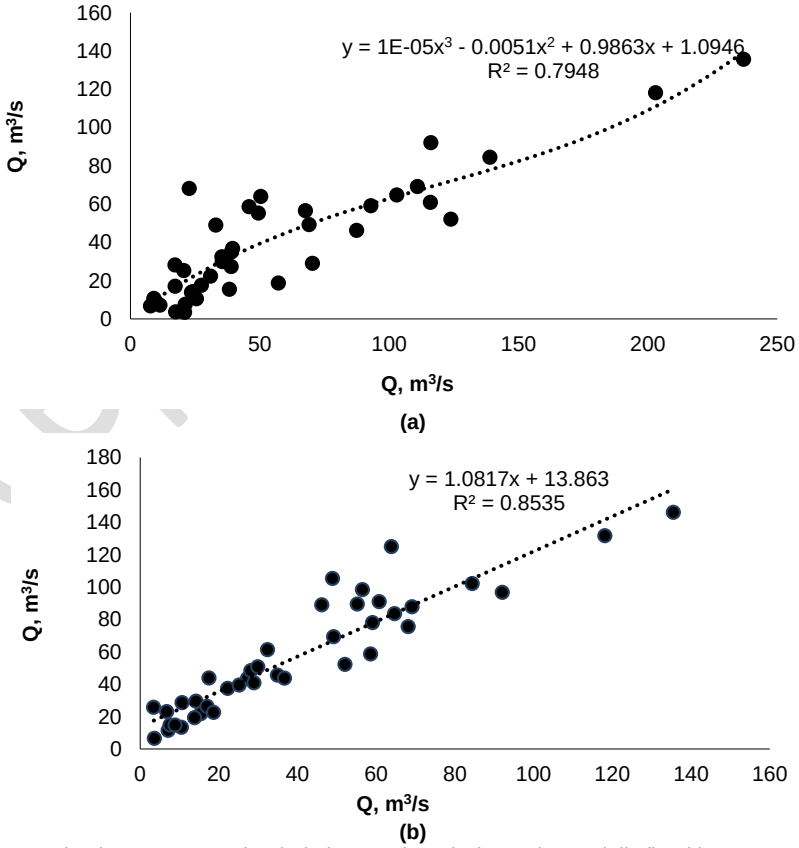


Fig. 5. IMF regression in Faraman and Pole-kohne stations (up), maximum daily flood between Doab mereg and Khersabad stations (Down).

Using RF, despite the old statistical methods for flood calculation in DRPs, the physical and hydrological characteristics of the basin area are used and the priority of each feature is determined in the calculation of peak discharge. ML methods have an effective role in predicting flood events in different basins, especially in basins without hydrometric stations. In this research the value of peak flood in 25y flood is less than other methods but it's still useful for the area without hydro station.

Table 5. Performance criteria.

Error criteria	RMSE	MAE	NSE
Trained RF model	0.92	1.23	0.95
Tested RF model	0.83	3.21	0.89

3.3. Regional flood frequency

To estimate the discharge with DRPs in the studied stations, the IMF discharge data had a downward trend. The trend was removed in all stations, and different distributions were fitted in Easy fit software. The total ranking of three tests of Kolmogorov Smirnov, Chi Square and, Anderson Darling were calculated and the selected distributions were preferred based on the lowest rank number. The flood values were calculated with DRPs in Table 7. To calculate the discharge with DRPs in Doab Qazanchi area which lacks a hydrometric station, RFFA with index FFA method was used.

Table 6. Calculated Discharge with RF Method for DRPs in Doab-Qazanchi

DRPs	T=2 y	T=3 y	T=5 y	T=10y	T=25y	T=50	T=100y	T=200y	T=500y	T=1000y
Doab-Qazanchi	93.8	137.7	220.4	360.6	718.9	1160.0	1820.6	2987.3	4750.5	8973.7

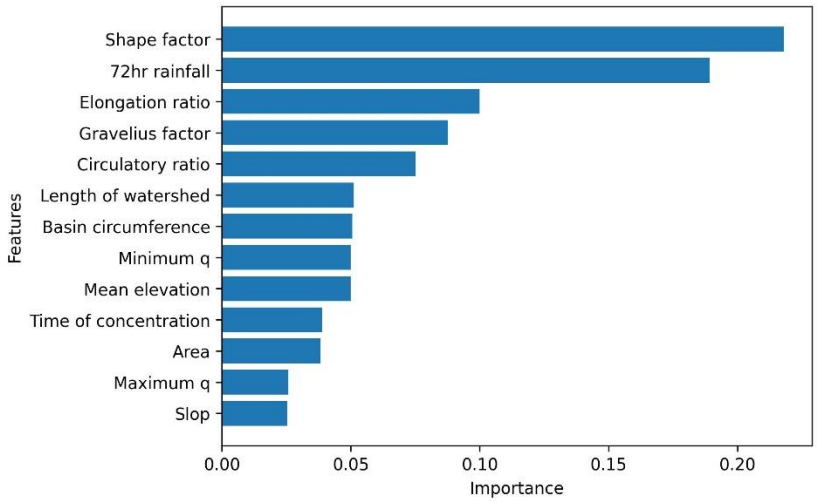


Fig. 6. Feature importance plot.

In this way, dimensionless values are calculated by obtaining the ratio of discharge with DRPs to discharge with a two-year return period at each station in Table 8. On the other hand, the regression relationship between the area and the two-year return period discharge in the nearby stations was calculated, which is shown in Fig.7.

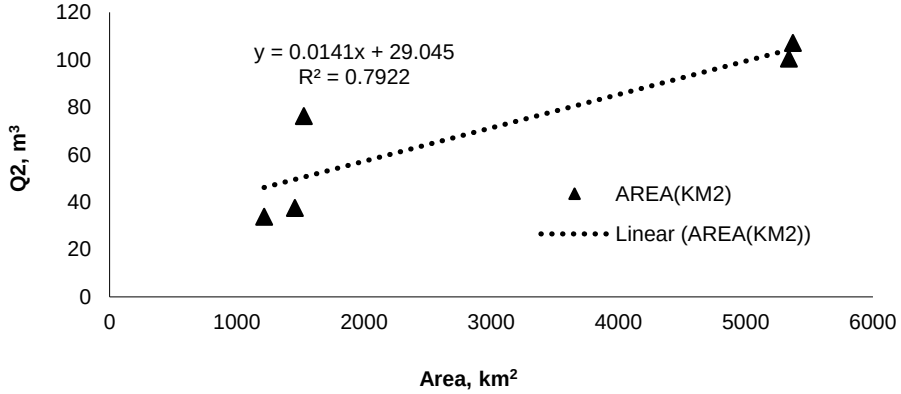


Fig. 7. Linear regression between AREA.

Table 7. Flood discharge with FFA for DRPs in all stations.

Stations	Best distribution	Return period, Year									
		2	3	5	10	25	50	100	200	500	1000
Doab mereg	Lognormal (3P)	33.8	51.7	80.9	128.6	211.4	291.5	389.5	507.9	700.7	878.3
Khersabad	Lognormal (3P)	37.5	52.5	72.9	100.8	141.0	174.4	210.6	250.0	307.2	354.7
Hojatabad	Frechet 3p	76.2	104.6	147.9	217.1	340.7	468.5	637.1	859.8	1268.3	1695.3
Pol-kohne	log-logistic 3p	100.5	149.9	236.9	402.5	775.1	1256	2026.4	3263.6	6121.5	9849.6
Faraman	log_person3	107.1	160.1	247.6	395.2	665.3	943.1	1302	1761.4	2563.9	3356.7

Table 8. The ratio of discharge with DRPs to discharge with a 2-year return period.

stations	Area, Km²	Q3/Q2	Q5/Q2	Q10/Q2	Q25/Q2	Q50/Q2	Q100/Q2	Q200/Q2	Q500/Q2	Q1000/Q2
Doab mereg	1214.72	1.53	2.40	3.81	6.26	8.64	11.54	15.04	20.76	26.02
Khersabad	1456.06	1.40	1.94	2.69	3.76	4.65	5.62	6.67	8.20	9.46
Hojatabad	1526.72	1.37	1.94	2.85	4.47	6.15	8.36	11.28	16.64	22.24
Pol-kohne	5339.2	1.49	2.36	4.01	7.72	12.50	20.17	32.49	60.93	98.04
Faraman	5460	1.49	2.31	3.69	6.21	8.81	12.16	16.44	23.94	31.34

Finally, the area of the Doab-Qazanchi as an input and the discharge with a 2-year return period as output is obtained based on the calculated regression equation in Fig. 7. Discharge with a 2-year return period and Pol-kohne dimonsionless ratio are used to calculate DRPs in Doab- Qazanchi.

Table 9. Calculated discharge with index FFA for DRPs in Doab-Qazanchi.

Desired point	Area (km^2)	T=2 y	T=3 y	T=5 y	T=10y	T=25y	T=50y	T=100y	T=200y	T=500y	T=1000y
Doab-Qazanchi	4590	93.8	139.9	221.1	375.6	723.5	1172.3	1891.3	3046.1	5713.5	9193.1

3.4. Calculating flood discharge with Creager coefficients

The Creager method was used to estimate peak flood in the Doab-Qazanchi area, and Creager coefficient (C) values were determined for this specific region. Using these coefficients in Creager's Eqs. 1-2, peak flood values for Doab-Qazanchi were estimated (Table 10). Given Doab-Qazanchi's proximity to the Pol-kohne station, the Pol-kohne Creager coefficient was used to calculate Doab-Qazanchi flood discharge for DRPs (Table 12). Peak discharge results from the Creager and RFA methods align with findings from previous studies (Karkouti et al. 2010; Jahandideh et al. 2011; Jabbari, Ghobadian, and Ahmadi Melaverdi, 2017) regarding flood events on 19/03/1998 and 29/03/2005. A three-parameter log-normal frequency distribution best fit the studied stations, with Creager coefficients ranging from 0.41 to 31.98 for return periods of 2 to 1000 years.

Table 10. Specific discharge calculated for each station in DRPs.

Stations	Area, Km^2	T=2 y	T=3 y	T=5 y	T=10y	T=25y	T=50y	T=100y	T=200y	T=500y	T=1000y
Doab mereg	1214.72	0.03	0.04	1.58	1.61	1.67	1.39	1.35	1.31	1.39	1.26
Khersabad	1456.06	0.03	0.04	0.05	0.07	0.10	0.11	0.13	0.15	0.17	0.19
Hojatabad	1526.72	0.05	0.07	0.09	0.13	0.21	0.29	0.41	0.57	0.87	1.20
Pol-kohne	5339.2	0.02	0.03	0.04	0.06	0.11	0.16	0.23	0.34	0.57	0.84
Faraman	5460	0.02	0.03	0.04	0.07	0.12	0.18	0.26	0.38	0.61	0.88

Table 11. The Creager coefficient for each station in DRPs

Stations	T=2 y	T=3 y	T=5 y	T=10y	T=25y	T=50y	T=100y	T=200y	T=500y	T=1000y
Doab mereg	0.42	0.65	24.53	24.93	25.81	21.56	20.86	20.35	21.57	19.54
Khersabad	0.44	0.64	0.89	1.21	1.63	1.94	2.25	2.55	2.95	3.25
Hojatabad	0.86	1.17	1.62	2.35	3.70	5.16	7.16	9.91	15.22	21.05
Pol-kohne	0.67	0.97	1.44	2.26	3.87	5.73	8.43	12.38	20.52	30.06
Faraman	0.70	1.03	1.57	2.52	4.41	6.53	9.54	13.81	22.33	31.98

Table 12. Creager coefficient and discharge for DRPs in Doab-Qazanchi

Doab-Qazanchi	T=2 y	T=3 y	T=5 y	T=10y	T=25y	T=50y	T=100y	T=200y	T=500y	T=1000y
C	0.67	0.97	1.44	2.26	3.87	5.73	8.43	12.38	20.52	30.06
Q	94.2	140.7	222.2	377.5	727.1	1178.1	1900.7	3061.2	5741.	9238.7

3.5. Flood hydrograph and peak discharge with SCS method

The flood hydrograph was calculated based on the average of cumulative precipitation of the synoptic stations of Kermanshah and, Ravansar from 01/04/2019 to 03/04/2019 using the SCS method in the Doab-Qazanchi region.

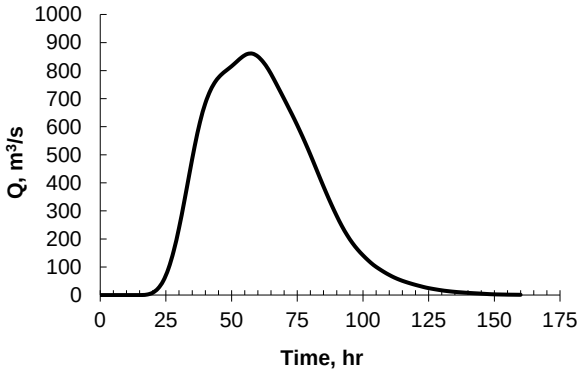


Fig. 8. SCS flood hydrograph (Right).

By considering the CN equal to 73 with GIS analysis and the concentration-time of 30 hours for the upstream basin of Doab-Qazanchi and the calculations were made in the Excel hydrograph spreadsheet environment, the flood hydrograph was calculated .As shown in Fig. 8, the peak of this flood is equivalent to 860 m^3/s. As expected, a comparison of the results of the SCS simulation of the desired flood hydrograph with the methods of RFFA, Creager, and RF, shows that the peak discharge of this flood is close to the return period of 25 years flood and it was a significant flood has occurred on 03/04/2019.

The observed peak flood which is registered in the Faraman station on 03/04/2019 is equal to 490 m^3/s that is shown in Fig.9. The Faraman station is in the downstream of Gharasoo River so it's reasonable to have a peak flood equal to 860 m^3/s in the Doab-Qazanchi area and it's decreased to 490 m^3/s in Faramn station due to flood routing.

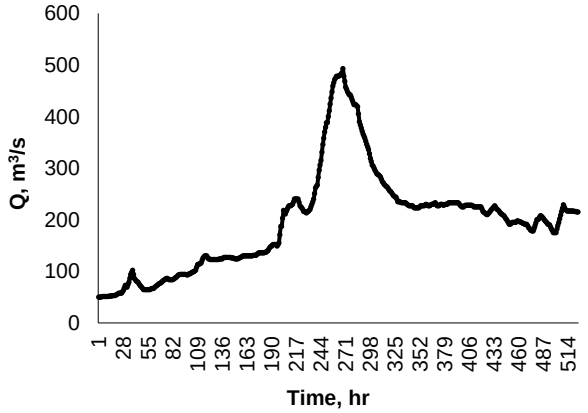


Fig. 9. Observed run off in Faraman station on 03/04/2019.

4. Conclusions

Estimating the amount of runoff and flood hydrograph is the first and the most important step in the design and implementation of hydrological and hydraulic plans, especially in flood control projects. Data preprocessing including outlier, randomness, trend and hurst analysis was performed with related equations and models, in this way, data losses created with high correlation regression equations. In this research, flood flow estimation with DRPs were created through the FFA method, then to calculate these values in the Doab-Qazanchi region, which has no hydrometric station, the RFFA with the flood index method was used. The results showed that Lognormal (3P), Frechet 3p, log-logistic 3p, and log_person3 are the best distributions for flood prediction in the DRPs. On the other hand, to compare the results with Creager method, the Creager coefficient was calculated and the discharge in a certain area was calculated. The Creager coefficient (C)

values were determined for all hydrometric stations. The Pol-kohne creager coefficient was determined for the Doab-Qazanchi point and based on this value the flood was predicted in the DRPs. Considering the importance of machine learning methods and based on the studies, the RF method for training, testing, and forecasting discharge in DRPs based on the basin characteristics around the area was considered by the REFCV method in the Python environment. It showed that the shape factor and 72hr rainfall are the most important features for discharge prediction with RF. The results are similar but it's nearly less than other methods. The flood hydrograph was calculated with the SCS method based on cumulative rainfall from 01/04/2019 to 03/04/2019 in Kermanshah and Ravansar synoptic stations. The results indicated that the peak flow on 03/04/2019 was equal to 860 m^3/s which is close to 25y flood in the other methods but it's a bit more than others. The results showed that the RF method, utilizing machine learning techniques, demonstrated good performance in predicting flood discharge based on physical and hydrological indicators of the catchment area. The Creager method and RFFA also provided accurate predictions. However, the SCS method tended to overestimate peak flow. Despite the absence of a station in the Doab-Qazanchi region, the importance of the region in terms of its proximity to Kermanshah city, the joining of Razavar, and Doab Mereg rivers to this place and increasing the flood discharge, RFFA, Creager, RF, and SCS methods are predicted the flood discharge values in DRPs. The peak flood that is obtained with SCS method which is nearly consistent with the 25y flood is higher by 18.27%, 18.86%, and 19.63% than that of Creager, FFA and RF, respectively. The maximum 25y peak flood is predicted by the SCS and the lowest one is by RF. The research highlighted the importance of accurate flood prediction for effective flood risk management. Accurate estimation of flood hydrographs and frequency is crucial for designing flood control measures, land-use planning, emergency response planning, and assessing flood insurance policies.

Author Contributions

Sadaf Gord: Conceptualization, methodology, software implementation, data collection and analysis, visualization.
Maryam Hafezparast Mavaddat: Supervisor, original draft preparation, Supervision, validation, critical review and editing of the manuscript, project administration, academic guidance.
Rasool Ghobadian: Technical advisement, resource provision, methodological feedback, manuscript revision for technical accuracy.

Acknowledgment

The authors would like to express their sincere gratitude to the Regional Water Company of Kermanshah, Kermanshah, Iran, for providing the necessary data and support for this research. We also extend our thanks to the Meteorological Organization of Kermanshah Province, Kermanshah, Iran, for their valuable climate data and cooperation. Their contributions were essential to the success of this study.

Conflicts of Interest

The authors declare no conflict of interest.

Data Availability Statement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Nomenclature

ANFIS	Adaptive neural inference system
ANN	Artificial neural network
BRT	Boosted regression trees
CN	Curve number
CV	Cross validation
DRPs	Different return periods
DT	Decision tree
FFA	Flood frequency analysis
GEV	Generalized extreme value
GL	Generalized logistic
GNO	The generalized normal
GP	Genetic programming
IMF	Instantaneous maximum flow
MARS	Multivariate adaptive regression spline
MAE	Mean absolute error
ML	Machine learning
NSE	Nash-sutcliff efficiency
NLR	Nonlinear regression
PPR	Projection pursuit regression
P III	Pearson type 3
RFFA	Regional flood frequency analysis
RMSE	Root mean square error
PDF	Probability density function
RF	Random forest
RFECV	Recursive feature elimination cross validation
SVR	Support vector regression

References

Allahbakhshian-Farsani P. *et al*. (2020) 'Regional flood frequency analysis through some machine learning models in semi-arid regions', *Water Resources Management*, 34, pp. 2887–2909. doi: https://doi.org/10.1007/s11269-020-02590-9

Al-Fawa'reh, M., et al. (2021) 'Intelligent methods for flood forecasting in Wadi al Wala, Jordan', *Proceedings of the International Congress of Advanced Technology and Engineering (ICOTEN)*. Taiz, Yemen, 1–3 July. Piscataway, NJ: IEEE, pp. 1–9. doi: https://doi.org/10.1109/ICOTEN52080.2021.9493425

Bormann, H., Pinter, N. and Elfert, S. (2011) 'Hydrological signatures of flood trends on German rivers: Flood frequencies, flood heights and specific stages', *Journal of Hydrology*, 404(1–2), pp. 50–66. doi: https://doi.org/10.1016/j.jhydrol.2011.04.019

Chen R.C., *et al*. (2020) 'Selecting critical features for data classification based on machine learning methods', *Journal of Big Data*, 7, 52. doi: https://doi.org/10.1186/s40537-020-00327-4

Creager, W.P., Justin, J.D. and Hinds, J. (1945) *Engineering for Dams, Vol. 1: General Design*. New York: John Wiley.

Dogan E., *et al.* (2010) 'Modelling of evaporation from the reservoir of Yuvacik dam using adaptive neuro-fuzzy inference systems', *Engineering Applications of Artificial Intelligence*, 23(6), pp. 961–967. doi: https://doi.org/10.1016/j.engappai.2010.03.007

El Adlouni, S. and Bobée, B. (2015) *Hydrological Frequency Analysis Using HYFRAN-PLUS Software*. pp. 1–71. Available at: https://www.scribd.com/document/307191717/ (Accessed date: 5 June 2024.

Flynn, K.M., Kirby, W.H. and Hummel, P.R. (2006) *User Manual for Program PeakFQ, Annual Flood-Frequency Analysis Using Bulletin 17B Guidelines* (No. 4-B4). USGS. Available at: https://pubs.er.usgs.gov/publication/tm4B4 (Accessed date: 2 June 2024).

Gavrilović, L., Milanović Pešić, A. and Urošev, M. (2012) 'A hydrological analysis of the greatest floods in Serbia in the 1960–2010 period', *Carpathian Journal of Earth and Environmental Sciences*, 7(2), pp. 107–116. Available at: https://www.cjees.ro/viewTopic.php?topicId=274 (Accessed date: 6 June 2024).

Ghanbarpour M.R., *et al.* (2011) 'Calibration of river hydraulic model combined with GIS analysis using ground-based observation data', *Research Journal of Applied Sciences, Engineering and Technology*, 3(5), pp. 456–463. Available at: https://portal.research.lu.se/en/publications/calibration-of-river-hydraulic-model-combined-with-gis-analysis-u#:~:text=Abstract,used%20for%20many%20practical%20applications (Accessed date: 2 June 2024).

Gilleland, E., Ribatet, M. and Stephenson, A.G. (2013) 'Software review for extreme value analysis', *Extremes*, 16(1), pp. 103–119. doi: https://doi.org/10.1007/s10687-012-0155-0

Gizaw, M.S. and Gan, T.Y. (2016) 'Regional flood frequency analysis using support vector regression under historical and future climate', *Journal of Hydrology*, 538, pp. 387–398. doi: https://doi.org/10.1016/j.jhydrol.2016.04.041

Hailegeorgis, T.T. and Alfredsen, K. (2017) 'Regional flood frequency analysis and prediction in ungauged basins including estimation of major uncertainties for mid-Norway', *Journal of Hydrology: Regional Studies*, 9, pp. 104–126. doi: https://doi.org/10.1016/j.ejrh.2016.12.084

Hafezparast Mavadat, M., and Marabi, S. (2021). 'Prediction of SAR and TDS parameters using LSTM- RNN model: A case study on Aran station, Iran', *Journal of Applied Research in Water and Wastewater*, 8(2), pp. 88-97. doi: https://doi.org/10.22126/arww.2021.5708.1188

Hamaamin D., *et al*. (2022). 'The simulation of flood hydrograph under uncertain conditions of rainfall extreme values in different return periods: A case study on Gharesoo basin', *Journal of Applied*

Research in Water and Wastewater, 9(1), pp. 91-99. doi: https://doi.org/10.22126/arww.2022.7902.1251

Hurst, H.E. (1951) 'Long-term storage capacity of reservoirs', *Transactions of the American Society of Civil Engineers*, 116(1), pp. 770–799. doi: https://doi.org/10.1061/TACEAT.0006518

Jabbari, I., Ghobadian, R. and Ahmadi Melaverdi, M. (2017) 'The relationship between the LFH index and the flood zones with different return periods (Case study: Gharasoo River)', *Journal of Geographic Space*, 17(58), pp. 191–207. Available at: http://geographical-space.iau-ahar.ac.ir/article-1-447-fa.html (Accessed date: 10 June 2024).

Jahandideh, K., *et al.* (2011) 'Evaluation and calibration model WMS/HEC-HMS in the drainage basin of Gharesoo', *1st National Conference on Coastal Water Resources Management*. Sari Agricultural Sciences and Natural Resources University, Iran, 9-10 December. Civilica, NCCLWRM01_025. Available at: https://civilica.com/doc/105739/(Accessed date: 1 January 2024).

Karkouti A. *et al.* (2010). 'Determination of Maximum Flood Flow by use of Sampling (observe) Creager and SCS Method (Case study: Gharasoo river, Kermanshah, Iran)', *Journal of Environmental Studies*, 36(55), pp. 99-110. dor: https://dor.org/20.1001.1.10258620.1389.36.55.10.7

Kendall, M.G. (1975) *Rank correlation methods*. 4th ed. London: Charles Griffin.

Kozanis S. *et al*. (2010) 'Hydrognomon – opensource software for the analysis of hydrological data', *Proceedings of the European Geosciences Union General Assembly*, Vienna, Austria, 2-7 May, EGU General Assembly, p. 12419. doi: https://doi.org/10.13140/RG.2.2.21350.83527

Kundzewicz, Z.W. (Ed.). (2012). *Changes in flood risk in Europe* . Boca Raton: CRC Press.

Lee, J., *et al.* (2020) 'Estimating design floods at ungauged watersheds in South Korea using machine learning models', *Water*, 12(11), 3022. doi: https://doi.org/10.3390/w12113022

Mann, H.B. (1945) 'Non-parametric test against trend', *Econometrica*, 13, pp. 245-259. doi; http://dx.doi.org/10.2307/1907187

Morlot, M., Brilly, M. and Šraj, M. (2019) 'Characterisation of the floods in the Danube River basin through flood frequency and seasonality analysis', *Acta Hydrotechnica*, 32(56), pp. 73–89. doi: https://doi.org/10.15292/acta.hydro.2019.06

Mosavi, A., Ozturk, P. and Chau, K.W. (2018) 'Flood prediction using machine learning models: Literature review', *Water*, 10(11), 1536. doi: https://doi.org/10.3390/w10111536

Mustamin, M.R., Maricar, F. and Karamma, R. (2021) 'Hydrological analysis in selecting flood discharge method in watershed of Kelara River', *INTEK: Jurnal Penelitian*, 8(2), pp. 141–150. doi: https://doi.org/10.31963/intek.v8i2.2874

Nash, J.E. and Sutcliffe, J.V. (1970) 'River flow forecasting through conceptual models part I — A discussion of principles', *Journal of Hydrology*, 10(3), pp. 282–290. doi: https://doi.org/10.1016/0022-1694(70)90255-6

National Research Council. (1999). *Improving American river flood frequency analyses*. Washington, DC: The National Academies Press.

Oosterbaan, R.J. (1994) 'Frequency and regression analysis of hydrologic data', in Ritzema, H.P. (ed.) *Drainage principles and applications*. 2nd edn. Wageningen, The Netherlands.

Raes, D., Mallants, D. and Song, Z. (1996) 'RAINBOW: a software package for analysing hydrologic data', in *Proceedings of the 6th International Conference on Hydraulic Engineering Software (HYDROSOFT 96)*, George Town, Malaysia, 1 September. Edited by Blain, W.R. Southampton: Computational Mechanics Publications, pp. 525-534. Available at: https://kuleuven.limo.libis.be/discovery/fulldisplay?docid=lirias4049382&context=SearchWebhook&vid=32KUL_KUL:Lirias&lang=en&search_scope=lirias_profile&adaptor=SearchWebhook&tab=LIRIAS&query=any,contains,LIRIAS4049382&offset=0 (Accessed date:12 September 2024)

Rutkowska, A. *et al.* (2017) 'Regional L-moment-based flood frequency analysis in the upper Vistula River basin, Poland', *Pure and Applied Geophysics*, 174(2), pp. 701–721. doi: https://doi.org/10.1007/s00024-016-1428-3

Salami, W.A. *et al*. (2017) 'Runoff hydrographs using Snyder and SCS synthetic unit hydrograph: A case study of selected rivers in south west Nigeria', *Journal of Ecological Engineering*, 18(1), pp. 25–34. doi: https://doi.org/10.12911/22998993/66258

Schittkowski, K. (1980) *Nonlinear programming codes: Information, tests, performance*. lecture notes in economics and mathematical systems. 1st edn. Berlin: Springer. Available at: https://www.amazon.com/Nonlinear-Programming-Codes-Information-Mathematical/dp/3540102477 (Accessed date: 20 September 2024)

Shahabi, S., and Hessami Kermani, M. R. (2015) 'Flood frequency analysis using density function of wavelet (Case study: Polroud River)', *Journal of Applied Research in Water and Wastewater*, 2(1), pp. 122-130. doi: https://doi.org/10.22126/arww.2015.122

Sharifi Garmdareh, E., Vafakhah, M. and Eslamian, S.S. (2018) 'Regional flood frequency analysis using support vector regression', *Journal of Hydrology*, 565, pp. 14-24. https://doi.org/10.1016/j.jhydrol.2018.07.072

Topaloğlu, F. (2005) 'Regional flood frequency analysis of the basins of the East Mediterranean region', *Turkish Journal of Agriculture and Forestry*, 29(4), pp. 281-288. doi: https://doi.org/10.3906/tar-0409-8