Environmental Health
Engineering and
Management Journal

Original Article

Environmental Health
Engineering and Management

Comparison of hybrid regression and multivariate regression in the regional flood frequency analysis: A case study in Khorasan Razavi province

Seyed Hashem Hosseini[1], Mohammad Reza Khaleghi[1*], Hasan Jami[2], Sajjad Baygi[2]

[1]Department of Range and Watershed Management, Torbat-e-Jam Branch, Islamic Azad University, Torbat-e-Jam, Iran
[2]Department of Civil Engineering, Torbat-e-Jam Branch, Islamic Azad University, Torbat-e-Jam, Iran

Abstract

Background: Magnitude, rate and frequency of the stochastic and unexpected events are of great significance and importance in hydrology. Nowadays, for economic planning of the projects, the use of analytical methods of unexpected events in hydrology is unavoidable. The aim of this study was to compare hybrid regression and multivariate regression to estimate flood peak discharge in the province of Khorasan Razavi and in the selected water measured stations.

Methods: For this purpose, 19 hydrometric stations were selected and analyzed. In the first step, the rate of peak discharge was estimated with different return periods and by selecting the best regional distribution (lognormal distribution type III). In the next step, independent and important variables including area, mean annual rainfall, the average height of the watershed and its slope were determined using functional analysis and using SPSS software version 22. Then, two hydrologically homogeneous regions were determined by homogeneity test using cluster analysis, and accordingly, two models were presented for the whole area and also for homogeneous areas. To compare and evaluate the accuracy and efficiency of the estimated models, the rates of discharges were estimated and compared with observational rates using three control watersheds. To compare models, it was used from the average absolute values of the relative error index.

Results: It was revealed that the hybrid method was more accurate than the multivariate regression method in the return period of 50 years and provides better results of flood discharges for the area. Homogenous areas had a higher coefficient of determination (R^2) and lower relative standard error (RSE) compared to the whole area. It was also revealed that with increase of return period, the rates of R^2 decreased but the rates of relative standard error increased.

Conclusion: The accuracy of multivariate regression and hybrid methods was the same in the 25-year return period. In the present study, the importance and necessity of homogenous areas compared with the model of the whole area are completely evident.

Keywords: Floods, Regression analysis, Khorasan Razavi province, Statistical distributions, Hybrid regression

Citation: Hosseini SH, Khaleghi MR, Jami H, Baygi S. Comparison of hybrid regression and multivariate regression in the regional flood frequency analysis: A case study in Khorasan Razavi province. Environmental Health Engineering and Management Journal 2018; 5(2): 93–100. doi: 10.15171/EHEM.2018.14.

Introduction

Flood is one of the most important natural disasters in the world that makes main problems for the development and reclamation of many countries. This natural phenomenon has been converted to a continuous phenomenon that can destroy capital and financial facilities of countries. Flood management in a basin will not be successful unless the hydrological behaviors of catchments are predicted (1-4). Hence, the determination of flood peak discharge with different return periods is essential for conducting

flood control plans, designing hydraulic structures, preventing erosion and designing storages in an area (5-8). For this reason, different methods such as regional flood analysis have been presented. All these methods are made based on the recognition of factors affecting flood. To estimate flood frequency at ungauged drainage basins, flood frequency analysis using regression analysis is performed (9). In regional methods for flood prediction, the estimation of the index flood is a crucial point in order to attain reliable predictions in ungauged basins

(10). In recent years, the regression methods have been successfully employed and demonstrated satisfactory results (11-13). Kjeldsen and Rosbjerg compared different methods for estimating T-year events. Their simulation study revealed that in regional index-flood estimation, the method of probability-weighted moments is preferable to the method of moment estimation with regard to bias and RMSE (14). Kuldkepp in a study on watersheds located in arid areas of Nevada State, presented a new method called hybrid method to estimate peak flood discharge in areas with deficient hydrologic statistics (15). Kjeldsen et al, in a flood frequency analysis in South Africa, suggested that the effective factors in the peak flood discharge like the physiographic and climatic characteristics including longitude, latitude, watershed area, mean annual rainfall, rainfall concentration and soil compaction coefficient (16). Eng et al estimated the peak flow of 50 years using hybrid regression in the Southeast of the United States and concluded that this method has less error compared with geographic methods or methods based on the predictive variables (17). Song and Singh performed flood frequency analysis in Illinois by comparing curves of flood frequency obtained from flood simulation series with those obtained from equations of regional regression. They found that the curves obtained from flood simulation series have a higher accuracy compared to those obtained from the equations of regional regression (18). Eng et al used a hybrid region-of-influence (HRoI) regression method (as a hybrid geographic region-of-influence approach) to estimate the 50-year peak flow (Q50) in the Southeastern United States (17). They concluded that by increasing the accuracy of the geographic proximity of the stations, the accuracy of the model also increases. Marofi showed that Gumbel method is the best-fitted distribution in 10 regions of Hamadan province (19). Shukla et al dealt with the probabilistic estimates of extreme maximum rainfall using the regional generalized extreme value (GEV) distribution and concluded that the GEV model satisfied the selection criteria (20). Haberlandt and Radtke evaluated a hybrid rainfall model for the flood frequency analysis and reported a good performance in capturing average and extreme rainfall characteristics (21). Liu et al in downscaling some data-driven methods to forecast the daily precipitation and daily maximum and minimum temperature series concluded that time-lagged feed forward neural network (TLFFNN) and evolutionary polynomial regression (EPR) have more efficiency than statistical downscaling model (SDSM) (22). Smith and Karr investigated the flood frequency analysis using the Cox regression model and considered the assessment of the relative importance of physical processes such as snowmelt or soil moisture storage on flood frequency at an area and derivation of time-varying flood frequency estimates (23). Speyrer and Ragas examined the impact of flood risk and mandatory flood insurance on property values (24). Several studies performed the popular SPARROW model using a hybrid approach combining conventional regression methods with spatial data based on landscape characteristics and stream properties to predict continuous water quality from point observations (25-27). Firat compared different models for daily river flow forecasting and concluded that adaptive neuro-fuzzy inference system (ANFIS) has higher accuracy and reliability than other models (28). Soong and Soong compared flood-frequency curves derived from two different methods and concluded that flood-frequency curves derived from regional regression equations was slightly sloppy than those derived from the simulated flood-frequency curves (29). The aim of this study was to determine the physical and climatic factors affecting regional flood analysis, to find the best statistical distribution for determining the maximum flood discharges and also to compare two models of multivariate regression and hybrid method in the regional flood frequency analysis.

Material and Methods
Study area
The study area was located in Northeast of the country and from geological view, the sedimentary basin of Hezar Masjed- Kopedagh and Binalood as two separate tectonic, was the main part of the study area. Based on Domartan, the climate of these areas is arid and semi-arid. Figure 1 shows the location of the study stations in Khorasan Razavi.

Selection of the most suitable regional frequency distribution
After testing and data reconstruction, the flood frequency analysis included Pearson type Π distribution, Log Pearson type $\Pi\Pi$, two-parameter gamma, three-parameter Lognormal, two-parameter normal and Lognormal, were fitted with peak discharge of each station by HYFA software. To select the best statistical distribution, Chi-square test was used, so that its minimum value was considered $\chi^2=1$ and its maximum value was considered $\chi^2=6$.
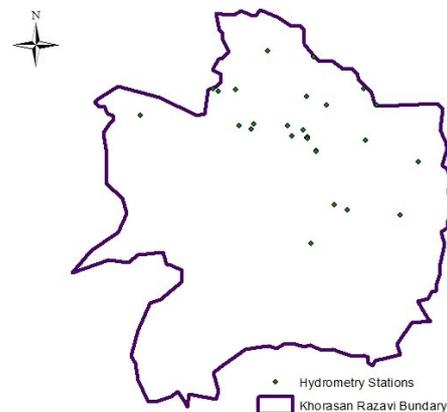

Figure 1. Location of the study stations in Khorasan Razavi.

Selection of the independent variables and limiting variables

In this study, a method of factor analysis was performed for 17 variables in the selected watersheds using SPSS software. Variables include different characteristics of watershed such as area, perimeter, average slope, the slope of the main river, length of the watershed, length of the main river, average height, average rainfall, maximum 24-hour precipitation, drainage density, maximum and minimum height and finally Gravelius, Horton, Miller and Schumm coefficients. As a unit of measurement, each variable differed from another one, therefore, for proper comparison of the variables with each other, all of them were standardized to be the same unit. As the results obtained from factor analysis was complex at first and did not produce an optimal solution, so to maximize the variance of each factors and to simplify the expression of variables of functional structure, functional axis has been rotated using a varimax rotation, one of the conventional methods, to be as an independent factor. Meanwhile, the naming of factors is done on the basis of the rotated functional values. Then, the matrices of functional scores of the stations were extracted using regression estimation. To confine the number of factors, Kaiser-Meyer-Olkin (KMO) measure was used to determine the rate of proportionality of a number of selective factors. Nonessential variables were extrapolated using the anti-image correlation matrix by measurement system analysis (MSA). These variables were diameter elements of the correlation matrix. In the elimination of data, the rate of KMO and percentage of the variance should be considered, because by eliminating one variable, KMO value and percentage of variance increase or decrease. The coefficient values of about 0.9 were considered very suitable, 0.8, suitable, 0.7, moderate, 0.6, medium, and less than 0.5 were not suitable. After determination of essential variables (12 factors), functional analysis was performed based on these variables.

Identification of homogeneous regional classes

The first step in the analysis of flood-prone area is to determine the homogeneous areas that were done using multivariate regression and index flood methods by cluster analysis. For this purpose, in the first step independent variables were determined by applying the process of factor analysis among input variables. In the next step, the homogenous areas were identified by cluster analysis using cumulative classes in SPSS software version 22. In the hybrid method, homogeneous areas were partitioned based on the most important parameters of the watershed (obtained from factor analysis), that affect the flood characteristics.

Hybrid method

Hybrid method is on the basis of the station-year method. In this method, all statistics available to the stations were used to overcome the lack of statistics. In the hybrid method, standardization was done by dividing each statistical data by a power of hydrological parameter of the watershed and then statistics of each class was combined with each other. The hybrid method included two sections: In the first section, the area was partitioned into at least three homogenous areas on the basis of the most important parameter of the watershed, and the flood peak statistics in each class were combined with each other. In the second section, statistics of annual flood peak discharge were standardized using an approximate factor, and then at the end of the hybrid process, an approximate factor of the standard was corrected using a technique combined from regression and flood frequency analysis. The process of standardization was done on using one of the parameters of the watershed in each replication. The model used in the hybrid method (equation 1), is like those that are used in different methods of regional flood analysis.

$$Q_T = aAbBcCd... \tag{1}$$

Where Q_T is peak discharge in return period of T years, A, B, C are independent parameters of the watershed, B, c, d, symbols in the regression equation and A is the fixed rate of the equation.

Partitioning of homogenous areas was done on the basis of the most important parameter of the watershed that affects flood characteristics. So, the area of the watershed, as the most important parameter of the watershed, was used in the first process. Partitioning must be done in a way that in each class, total data would be at least 100 classes and sum of a number of data in different could be near to each other and otherwise, by weighing to class data could be near as far as possible. A maximum number of classes with respect to the hypothesis of at least 100 data in each class were determined using equation 2:

$$J \leq \frac{Nf}{100} \tag{2}$$

Where J is the maximum number of classes while its minimum value is three classes and Nf is the number of total statistical data in each class. Then, the average weight of the catchment area (first parameter) was determined according to equation 3:

$$\overline{A}_j = antilig \frac{\sum_{j=1}^{g} \sum_{k=1}^{h} \log A_{ijk}}{gh} \tag{3}$$

Where $\overline{A}_J$ is average weight of the area of the watershed in class I, A_{ijk}, area of the watershed in class i and at the station j and i=1, 2, 3…f, number of classes, J=1, 2, 3… g, number of stations in ith class, and K, number of years in the jth station. It should be noted that the number of classes should be determined on the basis of the second and nth physiographic parameters in order and the average weight of the area in classes should be also calculated.

Standardization of peak discharge was replicated from equation 4 and it was performed through dividing peak discharge by the average weight of the area in each class; in the first replication, b was considered equal to 1.

$$S_{ijk} = \frac{Q_{ijk}}{A_{ijk}^b}$$

(4)

Then, S_{ijk} of each class was fitted by the curve of intermittent flood or by an empirical equation. For each class, a suitable probability distribution was determined, then, S_{ijk} with different return periods for each class, was obtained using a suitable distribution (S_{ti}). To obtain peak discharge with a return period of T years, the rates of S_{ti} was non-standardized as equation 5:

$Q_{ti} = S_{ti} (\bar{A}_i) b$ (5)

The average weight of the area was fixed along with the replication but the values of b vary to a constant value. The new exponent in each return period was obtained as equation 6:

$$b_t = \frac{\sum_{i=1}^{F} \bar{A}_i Q_{ti} - \sum_{i=1}^{F} A_i \sum_{i=1}^{F} Q_{ti}}{\sum_{i=1}^{F} \bar{A}_i^2 - \frac{\left[\sum_{i=1}^{F} A_i\right]^2}{F}}$$

(6)

In fact, if a regression equation is plotted between Q_{ti} and A_i, b_t is the slope of the equation, in other words, it is the coefficient variation of A_i. By the new exponent of calculation, the next stage was replicated so that the rates of Q_{ijk} was standardized through dividing A_{ijk} by exponent of b_t, and stages of determination of a suitable distribution and calculation of S_{ti} were replicated to get a newer b_t. This replication was continued until the variation of b_t reached less than 1% . After application of a parameter, if b was not fixed, it might be indicated that there was no linear relationship between parameter and peak discharge, so it was not included in the model. When the most important parameter of the watershed (area) was included in the model and fixed exponent and final b_t were calculated, then, it was time to determine the second parameter of the watershed, for example, the annual rainfall of the watershed. After determination of the second parameter, all stages performed for the first parameter, were also replicated for the selected parameter with a difference that for standardization of discharges of each rainfall class, Q_{ijk} of each station was divided by ($A^b P^c$) to obtain b. Exponent of c was also equal to 1 in the first replicate. After standardizing the discharges, the steps for determining the S_{ti} and C_t are repeated. The repetition process continues as long as the power of each new parameter is fixed. The rate of coefficient A was determined based on the correlation.

Method of multivariate regression

In this method, some ratios are presented between discharges in different return periods and physiographic characteristics of the watershed (12). The general relation of multiple regression is as equation 7:

$Q_T = F(A^a, B^b, C^c, \ldots, Z^n)$ (7)

Estimation of the regional distribution parameters

In this method, regression equations were developed between characteristics of watershed and parameters of the probability distribution. Firstly, after frequency analysis, suitable parameters were obtained for each station, then; regression equations were developed if necessary, to estimate the parameters of the probability distribution. In fact, using characteristics of the watershed, distribution parameters were obtained for different areas with deficient statistics, then, by application of obtained distribution parameters, different Q_{Ts} were calculated (1,4).

Results

Functional analysis

In this stage, essential variables were determined based on the value of KMO = 0.721. After determination of the essential variables (14 factors), the functional analysis was performed based on these variables. Table 1 shows the rates of invisible matrix root and variance percentage of the factors.

By application of functional analysis, from input variables, four factors including area, mean annual rainfall, the average height of watershed and slope of the main river, were identified as the most important factors. The total variance of these factors was considered 91.18%.

Determination of the homogenous areas

In the study area, the cluster analysis method was carried out through cumulative classes. Factors included area, mean annual rainfall, the average height of watershed and pure slope of the river was done using hierarchical clustering method proposed by Wardez. By considering the maximum coefficient of similarity, two groups were obtained as follows:

The homogenous group 1: Watersheds of Golmakan, Dolat Abad, Mooshang, Sarasiab, Chakaneh, kooshk bala, Golestan, Hesar, anderkh, Zoshk, Zarandeh, Taghoon, Dorood, Nashib.

The homogenous group 2: Band Sarooj, Sang Divar, Ghara Tikan, Kalateh Rahman, Kartian, Agh Darband, Bagh Abbasi, Jaafar Mashhadi, Chahchaheh, Kardeh, Kashafrood, and Hatiteh.

The abovementioned processes are shown by a tree diagram (Figure 2).

Hybrid method

In the study area, 19 watersheds that had at least 10 statistical years of peak flood discharge, were selected. These 19 watersheds had totally 328 years of discharge statistics. After functional analysis, the factor of area was identified as the most important factor and also the factor of rainfall was identified as the second effective factor on the peak discharge. So, in the first step, the

Table 1. Values of invisible matrix root and variance percentage of the factors

Component (Factor)	Special primary values			Total	Variance percentage (%)	Cumulative variance
	Total	Variance percentage (%)	Cumulative variance			
1	6.581	54.839	54.839	4.401	36.357	36.67
2	1.966	16.385	71.223	2.766	23.05	59.72
3	1.313	10.941	82.164	2.536	21.137	80.86
4	1.083	9.023	91.187	1.239	10.32	91.187
5	0.371	3.09	94.276			
6	0.268	2.235	96.512			
7	0.213	1.77	98.29			
8	0.088	0.234	99.025			
9	0.044	0.354	99.379			
10	0.036	0.302	99.681			
11	0.029	0.242	99.992			
12	0.009	0.077	100			

```
* * * * * * H I E R A R C H I C A L   C L U S T E R   A N A L Y S I S * *
* * * *

Dendrogram using Ward Method

                    Rescaled Distance Cluster Combine

  C A S E    0         5        10        15        20        25
Label  Num  +---------+---------+---------+---------+---------+

     6    ─┐
     5    ─┤
     7    ─┤
     1    ─┤
    10    ─┤
     2    ─┤
     8    ─┤
     9    ─┤
     1    ─┤
     3    ─┤
     4    ─┤
    18    ─┤
    13    ─┤
    14    ─┤
    15    ─┤
    17    ─┤
    24    ─┤
    12    ─┤
    19    ─┘
    21    ─┐
    29    ─┤
    31    ─┤
    20    ─┤
    27    ─┤
    30    ─┤
    23    ─┤
    26    ─┤
    25    ─┤
    22    ─┤
    16    ─┤
    28    ─┘
```
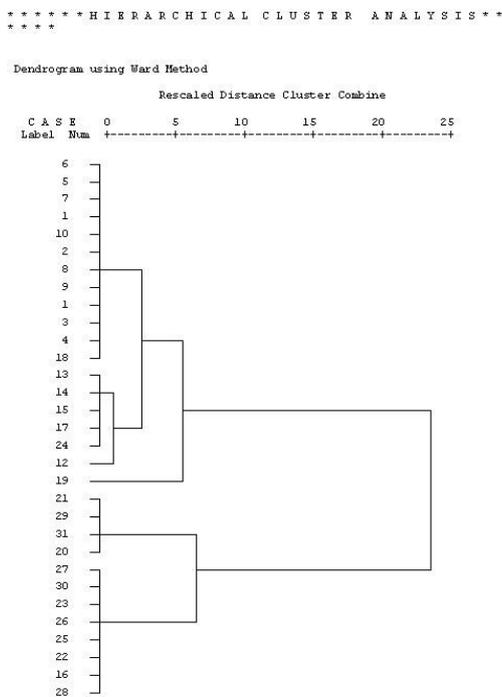
Figure 2. Tree diagram for area, mean annual rainfall, and mean height of watershed and pure slope of the river.

area was homogenized based on the basin area, then, the average weight of area for each class was calculated using equation 3. In the process of the hybrid method, In the process of hybrid, the instantaneous flood observations of the studied classes were standardized according to equation 4. Also the amount of b in the first replication was equal to 1. In other words, peak flood discharge of each station was divided by the exponent of its area to obtain the standardized discharges. As in the subsequent replication, the area exponent (b) would change, thus, for regional flood frequency analysis in each class; a long-term standardized compound statistic was obtained.

Then, based on the distribution of the three-parameter Lognormal (the best regional distribution), standard floods with return periods of 2, 5, 10, 25, 50 and 100 years were extracted for each class. After this stage, peak discharge of t years in each class (Q_{ti}) was determined as equation 5.

The second replication of the hybrid method was begun from equation 4. In this stage, instead of b that in the first stage was equal to 1, the rates of Table 2 that have been determined for each return period was used and replication process continued until exponent of b was fixed and if the power (b) is not fixed, the corresponding variable is abandoned and it indicating that there was no linear relationship between the related factor and discharge. In this research, the parameter b in the third replication means that the fourth step was fixed. These rates have been shown in Table 3.

After the factor of watershed area, mean annual rainfall was used. In the first step, the number of classes was determined and the stages done for the area, were replicated. As shown in equation 4, the method used for obtaining standard discharges for the first factor (area) and the second one (mean annual rainfall) is different. For subsequent agents, this process is repeated.

After determination of regression coefficients of the area in different return periods, the stages for subsequent important factors were replicated to complete the abovementioned relations. For discharges with different return periods, the abovementioned stages for the factor

Table 2. The rates of regression of watershed area factor (b) for the first replication in the study area

Return Period (y)	2	5	10	25	50	100
B-value	0.1612	0.2747	0.3389	0.4089	0.4534	0.4889

Table 3. The rates of regression coefficient of watershed area factor (b) for various replications in the study area.

Return Period (y)	2	5	10	25	50	100
First step	1	1	1	1	1	1
Second step	0.1612	0.2747	0.3389	0.4089	0.4534	0.4889
Third step	0.1373	0.2498	0.3391	0.4587	0.5339	0.6158
Fourth step	0.1401	0.2518	0.3391	0.4461	0.5184	0.5788

of mean annual rainfall were replicated and when the regression coefficient of these factors was fixed, the repetition was stopped and coefficients were replaced in the equation. In this study, the rates of c in different return periods were fixed after four steps (Table 4).

After the factor rainfall factors, average height and slope of watershed were determined and as the regression coefficients of these factors did not reach a fixed and logical value, they were not included in the model and put aside. After determination of c and b and other factors (average slope of the watershed, the average height of watershed) that could not include in the model, the fixed rate of the factor was determined. To determine regression fixed value (a) in each station, first, the rates of Q_t were obtained using the dominant distribution of the area, and simultaneously, with respect to physiographic characteristics included in the model, the rates of $(A^b P^c L^d)_t$ in different return periods for each station were calculated separately. Then, the ratio of Q_t to $(A^b P^c L^d)$ was determined for each station, finally, the median of these rates in different return periods, as regression fixed value (a), was included in the model. These rates are presented in the Table 5.

Multivariate regression

All models used for the estimation of peak flood discharge in a return period of 2, 5, 10, 25, 50 and 100 years for the whole area and homogenous areas of watersheds, are presented in Tables 6 and 7, respectively.

Discussion

The first step in the analysis of regional flood is to determine homogenous areas that were done using multivariate regression method by cluster analysis. For homogenizing the areas, firstly, functional analysis was performed on the input variables, and four factors including area, mean annual rainfall, average height of the watershed and slope of the main river, were identified as the most important factors. The total variance of these factors was considered 91.18%.

In the next stage, homogeneity of the watershed was determined by cluster analysis and two homogenous areas were determined. Also, with respect to four main variables, two series of models were estimated: one in the whole area and another one in the homogenous areas. Homogenous areas had a higher coefficient of determination (R^2) and less relative standard errors (RSE) compared to the models for the whole area. It was also revealed that with increase of return period, the rates of R^2 decreased but the rates of RSE increased. Figure 3 shows the mean relative error in the whole area and in the homogenous areas in the multivariate regression method.

Kjeldsen and Rosbjerg performed a flood frequency analysis in South Africa and suggested that the physiographic and climatic characteristics of the area including the area of the watershed, mean annual rainfall, rainfall concentration, the average slope of the watershed, altitude, edaphic characters and finally soil compaction coefficient, are effective in maximum flood discharges (14). Ouarda et al introduced some effective characteristics of the watersheds in peak discharges like area, length of the canal, the slope of the main canal and the mean annual rainfall (30). Honarbakhsh, in the analysis of regional flood in watersheds of Salt Lake,

Table 4. The rates of regression coefficient of mean annual rainfall factor (b) for various replications in the study area

Return Period (y)	2	5	10	25	50	100
First step	1	1	1	1	1	1
Second step	0.0716	0.0436	0.0716	-0.0089	-0.0146	-0.0124
Third step	0.0908	0.0806	0.057	0.01	0.0036	0.0012
Fourth step	0.09	0.0709	0.0578	0.0095	0.0032	0.001

Table 5. Regression fixed rates of hybrid model in various return periods

Return Period (years)	2	5	10	25	50	100
Regression fixed	54.17	57.96	44.49	35.9	111.7	100.9

Table 6. Multivariate regression models of peak discharge for the whole area

Return Period	Model	SE	R²
2	$Q = 480.9 + 0.154\,A - 0.151\,H$	0.17	0.88
5	$Q = 975.6 + 0.284\,A - 0.321\,H$	0.18	0.88
10	$Q = 1303 + 0.380\,A - 0.439\,H$	0.2	0.87
25	$Q = 1725.9 + 0.51\,A - 0.594\,H$	0.19	0.86
50	$Q = 2048 + 0.154\,A - 0.716\,H$	0.2	0.8
100	$Q = 2237 + 0.154\,A - 0.841\,H$	0.21	0.73

Table 7. Multivariate regression models of peak discharge for the homogenous areas

Return Period	Model	SE	R²
2	$Q = 1085 + 0.126\,A - 0.372\,H$	0.17	0.87
5	$Q = 1856 + 0.216\,A - 0.63\,H$	0.17	0.89
10	$Q = 2335 + 0.277\,A - 0.79\,H$	0.18	0.88
25	$Q = 2915 + 0.355\,A - 0.981\,H$	0.19	0.86
50	$Q = 3335 + 0.413\,A - 1.12\,H$	0.19	0.82
100	$Q = 3752 + 0.47\,A - 1.25\,H$	0.24	0.75

introduced physiographic characteristics like the area of the watershed, average height and the average slope of the watershed, annual rainfall and length of the watershed as the most important parameters affecting the flood (31). In this study, according to the functional analysis, area, mean annual rainfall, the average height of watershed and slope of the main river, were identified as the most important factors. Figure 3 shows that SE of the model of the whole area was higher than that of the homogenous areas and R² was also lower than it. Therefore, it is concluded that models of homogenous areas have higher efficiency compared to those of the whole area.

Conclusion

The aim of this study was to determine the physiographic and climatic characteristics affecting analysis of regional flood, to find the best statistical distribution for determining the maximum discharges and also to compare two models of multivariate and hybrid in the analysis of regional flood. Many studies have shown the importance of hydrologically homogenous areas in higher accuracy and efficiency of the regional analysis models. In his study, the importance and necessity of homogenous areas compared with the whole area were completely evident. Generally, homogenizing in most cases makes effective factors in the flow, to show higher accuracy in the model. The comparison showed that in the hybrid method, by increasing the return period, the rate of error decreased at first and then increased. Hybrid method in the return period of 50 years had a higher accuracy compared to the multivariate regression method and presented better results of flood discharge for the area.
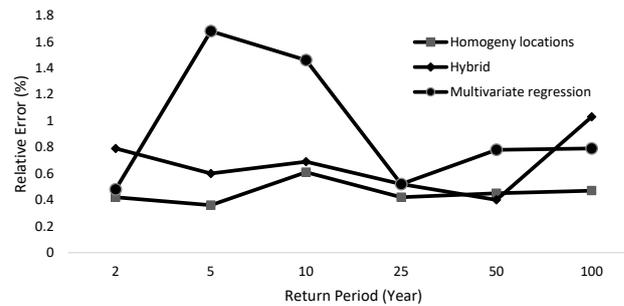
Figure 3. Comparison of the mean relative error in the multivariate regression (whole area) and the multivariate regression (homogenous areas) and hybrid regression.

Methods of multivariate regression and hybrid in a 25-year return period have similar accuracy.

Acknowledgements

Ethical issues

The author hereby certifies that all data collected during the study is as stated in the manuscript, and no data from the study has been or will be published separately elsewhere.

Competing interests

The author declares that he has no competing interests.

Authors' contributions

The author was involved in the suggestion of the problem, design of experiments, data collection, and article approval.

References:

1. Bhadra A, Panigrahy N, Singh R, Raghuwanshi NS, Mal BC, Tripathi MP. Development of a geomorphological instantaneous unit hydrograph model for scantily gauged watersheds. Environ Model Softw 2008; 23(8): 1013-25. doi: 10.1016/j.envsoft.2007.08.008.

2. Khaleghi MR, Gholami V, Ghodusi J, Hosseini H. Efficiency of the geomorphologic instantaneous unit hydrograph method in flood hydrograph simulation. Catena 2011; 87(2): 163-71. doi: 10.1016/j.catena.2011.04.005.

3. Khaleghi MR, Ghodusi J, Ahmadi H. Regional analysis using the geomorphologic instantaneous unit hydrograph (GIUH) method. Soil Water Res 2014; 9(1): 25-30.

4. Khaleghi MR. The influence of deforestation and anthropogenic activities on runoff generation. J For Sci 2017; 63(6): 245-53. doi: 10.17221/130/2016-JFS

5. Jain V, Sinha R. Derivation of unit hydrograph from GIUH analysis for a Himalayan River. Water Resources Management 2003; 17(5): 355-76. doi: 10.1023/a:1025884903120.

6. Gholami V, Khaleghi MR. The impact of vegetation on the bank erosion (case study: the Haraz River). Soil Water Res

2013; 8(4): 158-64.

7. Khaleghi MR, Varvani J. Simulation of relationship between river discharge and sediment yield in the semi-arid river watersheds. Acta Geophys 2018; 66(1): 109-19. doi: 10.1007/s11600-018-0110-9.

8. Khaleghi MR, Varvani J. Sediment rating curve parameters relationship with watershed characteristics in the semiarid river watersheds. Arab J Sci Eng 2018; 43(7): 3725-37. doi: 10.1007/s13369-018-3092-7.

9. Senaratne S, Cunnane C. Assessment of model error in regression estimates of floods at an ungauged site. Integrated Water Resources Management; 2000 Apr; LAHS, California; 2000.

10. Gioia A, Fiorentino M, Iacobellis V, Margiotta MR. Theoretical derivation of the index flood. Adv Geosci 2005; 2: 249-53. doi: 10.5194/adgeo-2-249-2005.

11. Yu PS, Chen CJ, Chen SJ, Lin SC. Application of Grey Model toward Runoff forecasting. J Am Water Resour Assoc 2001; 37(1): 151-66. doi: 10.1111/j.1752-1688.2001.tb05482.x.

12. Trivedi HV, Singh JK. Application of grey system theory in the development of a Runoff prediction model. Biosyst Eng 2005; 92(4): 521-6. doi: 10.1016/j.biosystemseng.2005.09.005.

13. Chou CM. Applying multi-resolution analysis to differential hydrological grey models with dual series. J Hydrol 2007; 332(1-2): 174-86. doi: 10.1016/j.jhydrol.2006.06.031.

14. Kjeldsen TR, Rosbjerg D. Comparison of regional index flood estimation procedures based on the extreme value type I distribution. Stoch Environ Res Risk Assess 2002; 16(5): 358-73. doi: 10.1007/s00477-002-0104-6.

15. Kuldkepp M. Rudolf Kjellén: geopolitiken och konservatismen edited by Ragnar Björk, Bert Edström and Thomas Lundén Stockholm, Hjalmarson & Högberg, 2014.337 pp., 198 kr, ISBN 978-91-7224-196-1. J Balt Stud 2017; 48(4): 549-51. doi: 10.1080/01629778.2017.1394699.

16. Kjeldsen TR, Smithers JC, Schulze RE. Regional flood frequency analysis in the KwaZulu-Natal province, South Africa, using the index-flood method. J Hydrol 2002; 255(1-4): 194-211. doi: 10.1016/S0022-1694(01)00520-0.

17. Eng K, Milly PC, Tasker GD. Flood regionalization: a hybrid geographic and predictor-variable region-of-influence regression method. J Hydrol Eng 2007; 12(6): 585-91. doi: 10.1061/(ASCE)1084-0699(2007)12:6(585).

18. Song S, Singh VP. Meta-elliptical copulas for drought frequency analysis of periodic hydrologic data. Stoch Environ Res Risk Assess 2010; 24(3): 425-44. doi: 10.1007/s00477-009-0331-1.

19. Marofi M. Investigation of the probability distributions fitting to extreme values of the flood. Geophysical Research Abstracts, EGS - AGU - EUG Joint Assembly; 6-11 April 2003; Nice, France.

20. Shukla RK, Trivedi M, Kumar M. On the proficient use of GEV distribution: a case study of subtropical monsoon region in India. Anale. Seria Informatică 2010; 83(1):81-92.

21. Haberlandt U, Radtke I. Hydrological model calibration for derived flood frequency analysis using stochastic rainfall and probability distributions of peak flows. Hydrol Earth Syst Sci 2014; 18(1): 353-65. doi: 10.5194/hess-18-353-2014.

22. Liu X, Coulibaly P, Evora N. Comparison of data-driven methods for downscaling ensemble weather forecasts. Hydrol Earth Syst Sci 2008; 12(2): 615-24. doi: 10.5194/hess-12-615-2008.

23. Smith JA, Karr AF. Flood frequency analysis using the cox regression model. Water Resour Res 1986; 22(6): 890-6. doi: 10.1029/WR022i006p00890.

24. Speyrer JF, Ragas WR. Housing prices and flood risk: An examination using spline regression. The Journal of Real Estate Finance and Economics 1991; 4(4): 395-407. doi: 10.1007/bf00219506.

25. Smith RA, Schwarz GE, Alexander RB. Regional interpretation of water-quality monitoring data. Water Resour Res 1997; 33(12): 2781-98. doi: 10.1029/97WR02171.

26. Lyon SW, Seibert J, Lembo AJ, Steenhuis TS, Walter MT. Incorporating landscape characteristics in a distance metric for interpolating between observations of stream water chemistry. Hydrol Earth Syst Sci 2008; 12(5): 1229-39. doi: 10.5194/hess-12-1229-2008.

27. Samadi MT, Sadeghi S, Rahmani A, Saghi MH. Survey of water quality in Moradbeik river basis on WQI index by GIS. Environ Health Eng Manag 2015; 2(1): 7-11.

28. Firat M. Comparison of artificial intelligence techniques for river flow forecasting. Hydrol Earth Syst Sci 2008; 12: 123-39.

29. Soong TW, Straub TD. Flood-Frequency Curves Estimated by Regional Regression Equations and Simulated Flood Series in the Ungaged Areas of Blackberry Creek Watershed, Illinois. World Water and Environmental Resources Congress; 2003 Jun 23-26; ASCE, Philadelphia; 2003. doi: 10.1061/40685(2003)357.

30. Ouarda T, Girard C, Cavadias GS, Bobee B. Regional flood frequency estimation with canonical correlation analysis. J Hydrol 2001; 254(1-4): 157-73. doi: 10.1016/S0022-1694(01)00488-7.

31. Honarbakhsh A. Regional analysis of flood in Namak lack watershed [dissertation]. Tehran: University of Tehran; 1996. [In Persian].