Regional flood frequency analysis for flood index estimation in hydrologic regions with limited flood data

Dr. Abdulwahab M. Younis
Dams and Water Resources Engineering, College of Engineering, University of Mosul, Iraq. E-mail: amyggv56@yahoo.com

Abstract:
The objective of this study is to establish a regional relationship between mean annual peak flood and the catchment area based on the annual peak flood frequency analysis. Annual peak flood time series for various gauging sites of six basins in north part of Iraq are used in analysis. The procedure of analysis is approached by scaling a regional flood frequency distribution (index-flood of catchment) which is defined here as the mean annual maximum flood discharge and estimated by linear regression using physiographic and climatic catchment properties. The index-flood method was found to have slightly better prediction accuracies over the direct-regression method. The coefficient of determination $R^2$ between reference and estimated index-floods is relatively high in most cases for all the gauging stations. All regions give very high $R^2$ correlation and the best model is the model (2), which is also associated with minimum value of RMSE.

Keywords:
Design flood; Index-flood method; Homogeneity test; ungauged sites. Direct regression

1. Introduction
Design flood has not received enough attention and has been adherent with emerging a large gap between the theoretical and practical design flood. Therefore, an improvement in estimation of design flood is imperative for flexible practical needs and uses.

The basic idea of flood frequency analysis is to obtain an estimate of flood quantile magnitude $Q$ ($T$), for selected locations on river system. Quantile estimates are widely used in planning, design, construction and operation of water resources projects. This approach in particular has been used for decision process related to hydraulic works, flood alleviation programs and water resources management within river system [9].

Estimation of floods arises in hydrology for forecasting and predicition purposes and aspects. Problems of forecasting approach for flood estimation arise most directly in the operation of hydrological controls in the broadest sense including flood warning. While predication approach issues are associated with the designs rather than the operation of such controls.

Direct-regression based approach has been widely used for developing the relationship between the regional flood and frequency. Index-flood based method using L-moments, however, has introduced more reasonable flood predictions due to providing a reasonable approach for statistical flood estimation of extreme events and presenting the characteristics of basin.
Estimation of design floods in urban hydrology has been approached by different ways. One of these methods is well known as the rational formula that determines the design flood by converting extreme precipitation statistics into extreme flood statistics. Driving streamflow statistics using distributed hydrological modeling is another way to estimate the design flood after calibrating the model for a gauged catchment. Such a procedure was adopted by [5]. The literature of relevant studies indicated that the quality of the estimated T-year flood is associated with the capacity of the hydrological model that uses for a proper extreme flood simulation and the physiographic and hydrologic similarities between calibrated and uncalibrated catchments.

The present work addresses the problem associated with estimation of index-flood, particularly in a region with limited observed streamflow data. An overview on study area and data availability has been presented in the next section. The methodology section introduces the general approaches used in deriving index-flood relationship.

2. Methodology

Index-flood estimation methods depend on whether the site of interest is gauged or ungauged. For gauged site, the estimation methods usually called by direct method and use information of annual maximum discharge series at the site of interest. In this approach, the mean of observed sample is used as an estimator of index-flood.

For the ungauged site in a hydrologic homogeneous region, the index-flood estimation methods may call as an indirect method and use information on regional basins related to the flood values and the corresponding physiographic/climatic properties. Thus, the present study uses the indirect approach to determine the index-flood estimation using different techniques as shown in the following sections [10,6].

The index-flood method basically assumes that the floods from different catchments within a region can be normalized by their mean annual floods and originate from a single distribution. The procedure of this method is characterized with standardizing the flood data from sites with different flood magnitudes. Further, the used data series is divided by the estimate of mean at site to obtain dimensionless data [8].

The basic relationship for estimating design flood \( \hat{Q}_i(T) \) of return period T at site i in index-flood (IF) based on the analysis of regional flood frequency which is expressed as Eq.(1):

\[
\hat{Q}_i(T) = q_r(T)(IF) ........................................(1)
\]

Where:
\( \hat{Q}_i(T) \) denotes the estimated flood peak discharge during T-year for a catchment i. \( q_r(t) \) is the growth factor which is the mean dimensionless regional T-year flood of the representative region.

\[
q_t(j) = \frac{Q_t(j)}{IF} ............................................(2)
\]

Where:
\( Q_t(j) \) is the observed maximum flood discharge during the year j for gauged catchment i.

\[
IF = E[Q_t]............................................(3)
\]
\[ E[Q_i] = \frac{1}{N} \sum_{j=1}^{N} Q_i(j) \]  

Index-flood at any location is mainly governed by its drainage area, \( A \) which is usually available for most catchments. To establish relationship for index-flood estimation in terms of drainage area, enough numbers of gauged stations in the interested hydrologic region should be available. At first, index-flood at each gauged station is computed as mean of the annual maximum discharge series. These mean values are called observed index-flood values \(^{[2,5]}\).

The regional flood-frequency relationship is power function and can be expressed as:

\[ Q_T = aA^b \]  

where \( a \) and \( b \) are regression parameters.

The present study intends to perform simple regression for all the hydrologic regions.

To obtain better index-flood estimation, an expression with more than one independent variable is derived. In addition to drainage area, the further information such as the drainage slope, rainfall, elevation of the basin, perimeter of the basin, etc. can be integrated in estimation of the index-flood. Such integration could be approached using multiple regression technique. The multi regression model expresses the estimate of index flood [IF] as:

\[ IF = C_0A_1^{c_1}A_2^{c_2} \ldots A_n^{c_n} \]  

In Eq.(5), \( A_1, A_2, \ldots, A_n \) represent an appropriate set of physiographic/climatic attributes of the basin. The constant \( C_0 \) and exponent \( c_1, c_2, \ldots, c_n \) are estimated from observed index-flood of each basin and corresponding physiographic/climatic information. In the present study, the researcher intends to perform multiple regressions for different possible combinations of physiographic/climatic attributes to establish a better index-flood relationship.

The regional growth factor \( q_r(T) \) is estimated using Generalized Extreme Value (GEV) distribution and Probability Weighted Moment (PWM) regionalization algorithm proposed by Hosking \ et al.\((1997-7)\). The GEV distribution of the annual maximum flood is determined for each gauged site \( i \), at homogenous region by estimating the PWM. Where, the estimated flood quantile \( \bar{Q}_i(T) \) at a given site \( i \), is calculated using Eq. (1). The index-flood (IF) is calculated either by Eq. (4) or Eq.(6) and \( q_r(T) \) is given by:

\[ \bar{q}_r(T) = m + \frac{\alpha}{\kappa} [1 - (-ln(p))^K], \ldots \ldots \ldots \]  

Where \( k, m \) and \( \alpha \) are the parameters of shape, location and scale, respectively. The case with \( \kappa=0 \) refers to the Gumbel distribution. Eq(7) can also be written as follow:

\[ \bar{q}_r(T) = m - \alpha ln(-ln(1 - 1/T)), \text{if } K = 0 \ldots \ldots \ldots \]  

Parameters of the GEV distribution can be estimated using different approaches such as the Maximum Likelihood (ML) and the Probability Weighted Moments (PWM) or the equivalent L-moments (LMOM)\(^{[4,11,13]}\).

3. Study Area and Data Availability

Annual peak flood series are determined from the peak flood records by selecting only one event a year. The mean of the annual maximum flood discharge data series for six catchments of the Tigris River tributaries were obtained from the gauging stations. The drainage area of selected catchments varies from 1020 km\(^2\) to 17330 km\(^2\). The mean elevation within the study area varies from 86 m to 570 m a.s.l consequently, the precipitation over the catchment under study where the average varies
between 320 mm and 900 mm. The length of historical streamflow records varies from 14 years to 50 year. Fig (1) shows the characteristics (lengths, slopes and drainage areas) and general location map of the study basins. Each gauged station is extracted according to their coordinates that extracted from the digital elevation model (DEM) databases (with 30 meters’ horizontal accuracy). The digital elevation model data are downloaded from the United State Geological Survey USGS (https://earthexplorer.usgs.gov/). Morphological characteristics of the basin (catchment area, stream network, delineated sub-basins within each watershed, basin drainage outlet …etc) are identified and determined by employing the DEM of area of interest and WMS v7.1. The basins with sufficient available annual peak flood series were selected for this study purposes.

Digital Elevation Model (DEM) was also used to identify catchments characteristics (catchment area, mean catchment slope, mean elevation of the catchment, station elevation, stream length and stream slope). These catchment characteristics are used to drive multiple regression equations that show the relationship between index-flood and catchment characteristics as independent variables.

4. Results and Discussions

Index-flood relationship has been established by computing the mean annual maximum discharge values at all the stations that averaged to call as observed index-flood. A simple regression was derived between observed index-flood and corresponding drainage area ($A$). To obtain better index-flood estimation, a multiple regression technique was then applied. The index-flood parameter was determined using the sample mean approach (Eq. (4)) and modeled with Eq. (6) including the following physiographic parameters of catchment: drainage area: $A$, mean annual rainfall: $P$, mean catchment slope: $S$, mean catchment altitude: $H$, catchment perimeter: $L$. The five following models have been tested as shown in Table 1

Figure 2 shows that the growth curves have consistent shapes in two basins (Jundian and Balikian) and indicating the homogeneity of the catchments. The growth curves for the other basins are relatively close to each other. The variability of growth curves among these basins may result from variabilities in period of data time series and climate data.
Figure 1: Catchment areas under study

Table 1. Index –flood with catchment physiographic parameters

| Model 1     | Model 2     | Model 3     | Model 4     | Model 5     |
|-------------|-------------|-------------|-------------|-------------|
| $E[Q^{-}] = a(A)^b$ | $E[Q^{-}] = a(AP)^b$ | $E[Q^{-}] = a(AP/H)^b$ | $E[Q^{-}] = a(A/L)^b$ | $E[Q^{-}] = a(AS)^b$ |

Table 2 shows the index-flood (IF in $m^3/s$) relationships for different hydrologic regions. It was obtained after analyzing the results of simple and multiple regressions. To assess predictive power of relationships, the estimated index-flood values were compared with observed index-flood values. In this manner, Nash efficiency coefficient (Eq.9) has been used to evaluate predictive power of the derived relationship. The Nash coefficient ($E$) for $N$ observation basins with average index-flood ($Q^{-}$) is (Eq. 9):

\[
E = 1 - \frac{\sum_i^N (Q_i - E[Q^{-}]_i)^2}{\sum_i^N (Q_i - Q^{-}_i)^2} \ldots \ldots \ldots \ldots \ldots (9)
\]

Here, $Q_i$ and $E[Q^{-}_i]$ are observed index flood and modeled index flood in basin $i$, respectively.

Figure 2 shows the growth curves for stations under study.

Table 2. Instantaneous flood Quantiles: $R^2$, Nash coefficient ($E$) and Root Mean Squared Error [RMSE] for each catchment under study.

| Gauging Station | $E[Q^{-}]$ | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|-----------------|------------|---------|---------|---------|---------|---------|
| Manquba         | 55.81      | 78.08*  | 82.00   | 141.10  | 95.57   | 80.43   |
| Eske-Kelek      | 105.91     | 67.46   | 66.98   | 124.02  | 83.86   | 109.54* |
| Balikian        | 121.13     | 191.11  | 101.44* | 65.69   | 160.17  | 99.53   |
| Jundian         | 879.93     | 898.34* | 794.96  | 714.77  | 1386.82 | 1245.46 |
| Dokan us        | 591.81     | 392.28  | 556.34* | 766.83  | 314.70  | 383.20  |
| Zerdala         | 675.75     | 729.51* | 586.06  | 421.02  | 409.60  | 592.87  |
| $R^2$           | 0.90       | 0.95    | 0.77    | 0.81    | 0.93    |
The frequency distribution of flood was calculated with Eq. (1) and compared with the reference one that calculated directly from the observed annual maximum flood data set based on the GEV/PWM method. The coefficient of determination $R^2$ between reference and estimated index-floods is relatively high in most cases. Also, the best parameter sets to use in Eq. (6) for modeling is the mean annual maximum flood may vary from basin to another. All regions give very high $R^2$ correlation and the best model is the model (2) which is also associated with minimum value of RMSE.

| E     | 0.79 | 0.97 | 0.67 | 0.28 | 0.68 |
|-------|------|------|------|------|------|
| RMSE  | 27.38| 20.07| 51.29| 78.32| 48.1 |

*a* represents the best regression model.

Figure 3 presents the results of the index-flood parameter were estimated based on the Eq. of models (1- 5), for the six regions as compared with observed flood. The coefficient of determination, $R^2$, is relatively high and accepted in most cases. Also, the best parameter sets to use in Eq. (6) for modeling is the mean annual maximum flood may vary from basin to another. All regions give very high $R^2$ correlation and the best model is the model (2) which is also associated with minimum value of RMSE.

Figures 4-10 show the comparison between the predicted and observed floods for 2, 5, 10, 20, 50, 100 and 200-years return periods, respectively, derived from direct-regression and index-flood methods for each site. These figures show that some stations have larger predictive discrepancies because of their location at the boundary of the hydrologic regions and may be influenced by other regions and also may lead to a significant discrepancy in the estimated values. These plots show similarity for the predicted floods at most of the stations for both used methods.

Absolute error at each site was computed for both the regional methods [2] by using Eqs. (10-11), to identify which regional method is better

$$\Delta Q^{IF}_T = \left| \frac{Q^{IF}_T - Q^{At\ site}_T}{Q^{At\ site}_T} \right| \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (10)$$

$$\Delta Q^{DR}_T = \left| \frac{Q^{DR}_T - Q^{At\ site}_T}{Q^{At\ site}_T} \right| \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (11)$$

where, $\Delta Q^{IF}_T$ is relative absolute error in index-flood

$\Delta Q^{DR}_T$ is relative absolute error in direct regression

$Q^{IF}_T$ is index-flood

$Q^{DR}_T$ is direct regression

$Q^{At\ site}_T$ is at-site flood-frequency analysis

For any of the used stations, the maximum absolute percentage error between the index-flood based regional estimates and the at-site flood-frequency analysis estimates were found to be 28%. While, the maximum absolute percentage error in direct regression regional estimates were found to be 61%. All
variables mentioned before based estimates of T-year return period. All the rivers and its tributaries within the study area do not have any hydraulic structures to control the regulation of runoff or flood control; but they have gauging stations to measure the discharges during varying periods. The results of the study showed that there is high correlation between the model and observed flow index at gauging stations.

5. Conclusion and future research
   a) 1- The growth curves have consistent shapes in two basins (Jundian and Balikian) indicating the homogeneity of the catchments.
   b) 2- The growth curves for the other basins are relatively close to each other. The variability of growth curves among these basins may result from variabilities in period of data time series and climate data.
   c) 3- The coefficient of determination $R^2$ between reference and estimated index-floods is relatively high in most cases for all the gauging stations. Because of errors associated with the regional growth factor estimation and estimation of the index-flood, the best results are not coherent with the best index-flood that is derived by the sample mean flood. The regression model is giving the best index-flood estimate.
   d) 4- The coefficient of determination, $R^2$, is relatively high and accepted in most cases. Also, the best parameter sets is the mean annual maximum flood that may vary from basin to another. All regions give very high $R^2$ correlation and the best model is the model (2) which is also associated with minimum value of RMSE.
   e) 5- Flood estimates in some stations are with larger predictive discrepancies because these stations are located at the boundary of the hydrologic regions and may be influenced by other regions and also may lead to a significant discrepancy in the estimated values, the predicted floods are close of the stations for both used flood methods.
Figure 3 Shows the relation between observed and estimated mean annual maximum instantaneous flood using five different models.
Figure 4 Mean annual maximum instantaneous flood verse catchment characteristics using the five models mentioned in Table 1.

\[
E[Q] = a(A)^b
\]

\[
y = 0.051x + 23.74 \\
R^2 = 0.994
\]

\[
E[Q] = a(AP)^b
\]

\[
y = 1E-07x - 58.83 \\
R^2 = 0.9985
\]

\[
E[Q] = a(AP/H)^b
\]

\[
y = 4E-10x - 88414. \\
R^2 = 0.997
\]

\[
E[Q] = a(A/L)^b
\]

\[
-770 36.89x = y \\
0.95 = R^2
\]

\[
E[Q] = a(AS)^b
\]

\[
y = 0.3503x + 18564. \\
R^2 = 0.9997
\]
Figure 5 Comparison of design floods for T=2 years

Figure 6 Comparison of design floods for T=5 years

Figure 7 Comparison of design floods for T=10 years
Figure 8 Comparison of design floods for T=20 year

Figure 9 Comparison of design floods for T=50 year

Figure 10 Comparison of design floods for T=100 year
Figure 11 Comparison of design floods for T=200 year

References

[1] Akshay R T and Manoj M M June 2011 Design flood estimation for upper Krishna Basin through Rffa, *International Journal of Engineering Science and Technology (IJEST)*, Vol. 3 No. 6

[2] Binaya K M, Kaora T, Yosuke Y and Yasuto T February 2010 An Assessment of predictive accuracy for two regional flood frequency estimation method *Annual Journal of Hydraulic Engineering, JSCE*, Vol. 54

[3] Bocchiola D, De Michele C and Rosso R 2003 Review of recent advances in index flood estimation, *Hydrology and Earth System Sciences*, Vol.7 (3) pp 283-296

[4] Brath A, Castellarin A, Franchini M and Galeati G 2001 Estimating the index flood using indirect methods, *Hydrological Sciences-Journal*, Vol. 46(3), pp 399-418

[5] Chetty K and Smithers J 2005 Continuous simulation modeling for design flood estimation in South Africa: Preliminary investigations in the Thukela catchment, *Physics and Chemistry of the Earth*, Vol. 30, pp 634-638

[6] George A Griffiths and Alistair I McKerchar 2012 Estimation of mean annual flood in New Zealand. *Journal of Hydrology* (NZ) 51 (2): 111-120 2012, © New Zealand Hydrological Society

[7] Hosking J R M and Wallis J R 1997 *Regional Frequency Analysis: An approach based on L-moments*, Cambridge University Press, Cambridge, UK

[8] Rao A R and Hamed K H 2000 *Flood Frequency Analysis*, CRC press LLC, Florida, USA

[9] Leuleseg Tadesse, Mohamed A Sonbol and Willems P. At-site and regional flood frequency analysis of the upper awash sub-basin in the Ethiopian Plateau, Ministry of Water Resources, Addis Ababa 5744, Ethiopia

[10] Mishra B K, Takara K and Tachikawa Y 2008 NRCS curve number employed hydrologic homogeneous regionalization in regional flood frequency analysis, *Proceedings on International Engineering*, Nagoya, Japan

[11] Mishra B K, Takara K, Yamashiki Y and Tachikawa Y 2009 Estimation of index flood in hydrologic regions with limited flood data availability, *Annual Journal of Hydraulic Engineering, JSCE*, Vol. 53, pp 50-60

[12] Perica S and Stayner M 2004 Regional flood frequency analysis for selected basins in Utah, University of Utah, Report No. UT-4.12

[13] Philippe C. Estimating the flood frequency distribution for ungauged catchments using an index flood procedure. Application to ten catchments in Northern Iceland, Icelandic Met Office, Report no.VÍ 2012-005