

# Comparative Study of Flood Frequency Analysis on Selected Rivers in Myanmar

Ni Lar Win and Khin Maung Win

**Abstract** A number of probability distributions are used to estimate flood discharges with various return periods of extreme hydrologic events. Among others, Log-Normal, Gumbel, Pearson Type III and Log-Pearson Type-III distributions are used in this study. The parameters of these distributions are estimated from the given data by the method of moments. The objectives of this study are: to carry out the flood frequency analysis using the different distributions to selected rivers in Myanmar, namely the Chindwin and Yenwe Rivers and to identify the most appropriate probability distribution for the basins under study. Log-Pearson Type III distribution can be recommended for both Chindwin and Yenwe basins among other distributions under study. The estimated flood values can be used in the engineering design of hydraulic structures such as dams, bridges, culverts, levees and other structures located along rivers and streams and for the effective management of flood plains.

**Keywords** Flood frequency analysis · Probability distribution · Chindwin and Yenwe rivers

## 1 Introduction

Estimation of the magnitude and frequency of occurrence of extreme hydrologic events, such as severe storms, floods, and droughts are important in water resources planning and management. Frequency analysis of hydrologic data is one

---

N. L. Win (✉) · K. M. Win

Faculty of Science, Technology, Engineering and Mathematics, INTI International University, Putra Nilai, Malaysia  
e-mail: larwin.ni@newinti.edu.my

K. M. Win

e-mail: maungwin.khin@newinti.edu.my

of the widely used methods of estimating the return period associated with a flood of a given magnitude through the use of probability distributions [1–8].

A number of probability distributions are used to fit the probability of occurrence of flood series. Details of these distributions are available in the literature including [9–11] among others. Selection of the most appropriate distribution for annual maximum series has been paid attention worldwide. Log-Pearson Type III distribution is the recommended technique for flood frequency analysis by the U.S. Water Advisory Committee on Water Data, USGS [3]. It is used to estimate the flood discharges for the streams in Iowa, USA [4]. The Generalized Extreme Value distribution is the standard method for UK as reported in [2] and it is adopted in regional flood frequency analysis in Northern Iceland [5] and Nile equatorial basins [6]. Gumbel distribution is used in Nyanyadzi river in Zimbabwe [7]. A World Meteorological Organization survey of 54 agencies in 28 countries reveals that Log-Normal distribution with 3 parameters is not a standard in any country, Generalized Extreme Value is a standard in one country, and Log-Pearson Type III is a standard in seven countries [8].

Numerous goodness-of-fit procedures exist for comparing the fit of alternative probability distributions to streamflow sequences. Since the introduction of L-moments, numerous investigators have recommended them to assess the goodness-of-fit of various probability distributions to regional samples of streamflow [8, 12, 13] and others. Vogel and Wilson [8] constructed L-moments diagrams for annual maximum streamflow at more than 1,455 rivers in the United States and concluded that the Generalized Extreme Value, three-parameter Log-Normal and the Log-Pearson Type III distributions provide good approximations to the distribution of annual maximum flood series. Since Log-Pearson Type III has been selected as the base method in the U.S., they suggested the agencies and countries to reevaluate their standards with respect to the choice of a suitable model for flood frequency analysis. Stedinger and Griffis [14] also recommended that the regional skew map given in [3] be updated to use the additional 30 years of data now available, to appropriately adjust for low outliers identified in the samples used to estimate the regional skew, and to use new and powerful statistical estimation procedures developed to use such data set.

The objectives of the study are hence twofold: (1) to carry out the flood frequency analysis using Log-Normal, Gumbel (Extreme Value Type I), Pearson Type III and Log-Pearson Type III which are commonly used frequency distributions to selected rivers in Myanmar, namely the Chindwin and Yenwe rivers and (2) to identify the most appropriate probability distribution for the basins under study. The estimated flood values can be used in the engineering design of hydraulic structures such as dams, bridges, culverts, levees and other structures located along rivers and streams and for the effective management of flood plains.

**Table 1** Location of gauging stations under study

No.	Name of the river	Gauging station	Location	Data period	Area (km <sup>2</sup> )
1	Chindwin	Hkamti	N26°00' E94°42'	1972–1995	27,420
2	Chindwin	Htamathi	N25°50' E95°17'	1971–1995	39,410
3	Chindwin	Homalin	N24°52' E94°54'	1968–1995	43,124
4	Yenwe	Myochau-ng	N17°57' E96°44'	1968–1995	790

## 2 Data Used in This Study

In this study, two rivers namely Chindwin and Yenwe rivers in Myanmar are selected. The Chindwin river has floods several times in a year from the onset of rainy season to the end of the monsoon. Annual maximum discharges are collected at Hkamti, Htamathi and Homalin stations for Chindwin river. Average annual rainfall at Hkamti, Htamathi and Homalin stations are 3,847, 3,262 and 2,314 mm respectively. The mean monthly temperature varies from 28 °C in May to 18 °C in January.

Myochaung station for Yenwe river which is a tributary of Sittoung is also selected in this study. It lies in a tropical monsoon area with an average annual rainfall of about 2,000 mm. The mean monthly temperature varies from 31 °C in April to 25 °C in January. The location and area of these stations are shown in Table 1. Series of annual maximum discharge of each basin is plotted in Fig. 1.

It can be seen from Fig. 1 that the variation of the flood values from the mean in Chindwin is small compared to the variation in Yenwe basin. Statistical characteristic of each series is given in Table 2. Annual maximum series of Chindwin basins have negative skews while Yenwe has a positive skew.

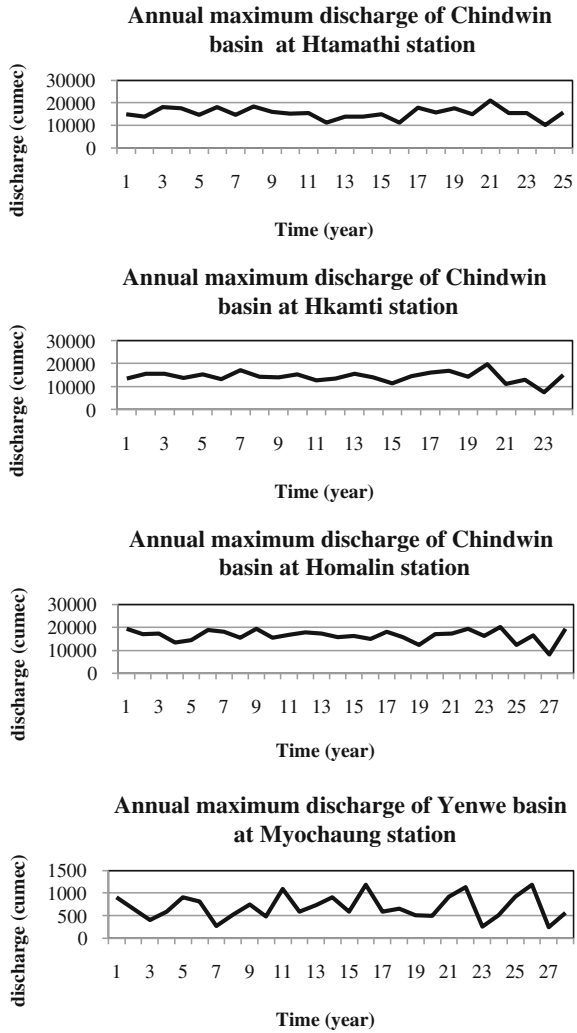
## 3 Methodology

### 3.1 Testing for Outliers

Reference [9] recommends that adjustments be made for the outliers of the data. If the station skew is greater than +0.4, tests for high outliers are considered first and if the station skew is less than -0.4, tests for low outliers are considered first. Where the skew is between ±0.4, tests for both high and low outliers should be applied before eliminating any outliers from the data set. The following equation is used to detect high outliers:

$$y_H = \bar{y} + K_N s \quad (1)$$

**Fig. 1** Annual maximum discharge series of Chindwin and Yenwe basins



**Table 2** Statistical characteristics of each basin

No.	Gauging station	Mean (m <sup>3</sup> /s)	Standard deviation (m <sup>3</sup> /s)	Coefficient of variation	Skewness coefficient
1	Hkamti	14,254	2342.0	0.16	-0.6
2	Htamathi	15,476	2407.9	0.16	-0.17
3	Homalin	16,485	2593.0	0.16	-1.25
4	Myochaung	692	274.0	0.40	0.19

where

- $y_H$  high outlier threshold in log units
- $\bar{y}$  mean logarithm of variate
- $s$  standard deviation of  $y$ 's
- $K_N$  10 % significance level  $K$  values.

The following equation is used to detect low outliers:

$$y_L = \bar{y} - K_N s \tag{2}$$

where  $y_L$  = low outlier threshold in log units and other terms are as defined earlier.

### 3.2 Frequency Analysis Using Frequency Factors

Chow [9] proposed the frequency factor equation shown in (3) which is applicable to many probability distributions used in hydrologic frequency analysis.

$$x_T = \mu + K_T \sigma \tag{3}$$

which may be approximated by

$$x_T = \bar{x} + K_T s \tag{4}$$

where

- $x_T$  value of the variate  $x$  of a random hydrologic series with a return period  $T$ ,
- $\bar{x}$  mean of the variate,
- $s$  standard deviation of the variate,
- $K_T$  frequency factor which depends upon the return period  $T$  and assumed frequency distribution

In the event that the variable analyzed is  $y = \log x$ , then the same method is applied to the statistics for the logarithms of the data using

$$y_T = \bar{y} + K_T s \tag{5}$$

and the required value of  $x_T$  is found by taking the antilog of  $y_T$ .

Table 3 summarizes the probability density function, the range of the variable, and equations for estimating the distribution's parameters from sample moments for each distribution [9]. The theoretical  $K_T$  relationships for the selected probability distributions: Log-Normal, Pearson Type III, Log-Pearson Type III and Gumbel distributions extracted from [9] are given below.

**Table 3** Probability distributions for fitting hydrologic data

Distribution	Probability density function	Equations for parameters
Log-normal	$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-\mu_y)^2}{2\sigma_y^2}\right)$ where $y = \log x, x > 0$	$\mu_y = \bar{y}$ $\sigma_y = S_y$
Extreme value type I (Gumbel)	$f(x) = \frac{1}{x} \exp\left[-\frac{x-u}{\alpha} - \exp\left(-\frac{x-u}{\alpha}\right)\right]$ $-\infty < x < \infty$	$\alpha = \frac{\sqrt{6}S_x}{\pi}$ $u = \bar{x} - 0.5772\alpha$
Pearson type III	$f(x) = \frac{\lambda^\beta (x-\varepsilon)^{\beta-1} e^{-\lambda(x-\varepsilon)}}{\Gamma(\beta)}$ $x \geq \varepsilon$	$\lambda = \frac{S_y}{\sqrt{\beta}}$ $\beta = \left(\frac{2}{C_s}\right)^2$ $\varepsilon = \bar{x} - S_x - \sqrt{\beta}$
Log-Pearson type III	$f(x) = \frac{\lambda^\beta (y-\varepsilon)^{\beta-1} e^{-\lambda(y-\varepsilon)}}{x\Gamma(\beta)}$ where $y = \log x$ $\log x \geq \varepsilon$	$\lambda = \frac{S_y}{\sqrt{\beta}}$ $\beta = \left(\frac{2}{C_s(y)}\right)^2$ $\varepsilon = \bar{y} - S_y - \sqrt{\beta}$

1. Gumbel distribution

The frequency factor  $K_T$  for different return period T is given in (6).

$$K_T = -\frac{\sqrt{6}}{\pi} \left\{ 0.5772 + \ln \left[ \ln \left( \frac{T}{T-1} \right) \right] \right\} \tag{6}$$

2. Log-Pearson Type III distribution

The frequency factor depends on the return period T and the coefficient of skewness  $C_s$ . When  $C_s = 0$ , the frequency factor is equal to the standard normal variable z. The value of z corresponding to an exceedence probability of p ( $p = 1/T$ ) can be calculated by finding the value of an immediate variable w:

$$w = \left[ \ln \left( \frac{1}{p^2} \right) \right]^{1/2} \quad (0 < p < 0.5) \tag{7}$$

then calculating z using the approximation

$$z = w - \left[ \frac{(2.515517 + 0.802853w + 0.010328w^2)}{(1 + 1.432788w + 0.189269w^2 + 0.001308w^3)} \right] \tag{8}$$

$$K_T = z + (z^2 - 1)k + \frac{1}{3}(z^3 - 6z)k^2 - (z^2 - 1)k^3 + zk^4 + \frac{1}{3}k^5 \tag{9}$$

where  $k = C_s/6$ , and  $C_s =$  coefficient of skewness of logarithms of the series.

3. Pearson Type III distribution

For this distribution, the same equation in Log-Pearson Type III applies except that  $C_s$  is skewness of the original series (without taking the logarithmic values).

#### 4. Log-Normal distribution

For this distribution, the same equation in Log-Pearson Type III applies using  $C_s = 0$ .

### 3.3 Statistical Analysis

Goodness of fit between the observed events and the fitted distribution is tested by the two most commonly used tests: Chi square and Kolmogorov–Smirnov (K–S) [15].

#### 1. Chi square test

The statistic is calculated by

$$\chi^2 = \sum_{j=1}^k \frac{(O_j - E_j)^2}{E_j} \quad (10)$$

where  $O_j$  is the observed number of events in the  $j$ th class interval and  $E_j$  is the number of events that would be expected from the theoretical distribution.

#### 2. Kolmogorov–Smirnov (K–S) test

The statistic  $D_n$  is evaluated by observing the deviation of the sample distribution function  $P(x)$  from the completely specified continuous hypothetical distribution function  $P_0(x)$ , such that

$$D_n = \text{Max}|P(x) - P_0(x)| \quad (11)$$

The test requires that the computed value  $D_n$  from the sample distribution be less than the tabulated value of  $D_n$  at the required significance levels of 0.05 and 0.1.

### 3.4 Reliability of Analysis

The reliability of the results of frequency analysis depends on how well the assumed probability model applies to a given set of hydrologic data. This can be done by calculating confidence limits which are upper and lower boundary values of the confidence interval. For estimating the event magnitude for return period  $T$ , the upper limit  $U_{T,\alpha}$  and lower limit  $L_{T,\alpha}$  may be specified by adjustment of the frequency factor equation given in [9]:

$$U_{T,\alpha} = \bar{x} + sK_{T,\alpha}^U \quad (12)$$

$$L_{T,\alpha} = \bar{x} + sK_{T,\alpha}^L \quad (13)$$

where  $\alpha$  is a significance level and is obtained by  $\alpha = \frac{1-\beta}{2}$ ,  $\beta =$  confidence interval.

$K_{T,\alpha}^U$  and  $K_{T,\alpha}^L$  are upper and lower confidence factors, which can be determined for normally distributed data using the noncentral t distribution. Approximate values for these factors are given by the following equations:

$$K_{T,\alpha}^U = \frac{K_T + \sqrt{K_T^2 - ab}}{a} \quad (14)$$

$$K_{T,\alpha}^L = \frac{K_T - \sqrt{K_T^2 - ab}}{a} \quad (15)$$

in which  $a = 1 - \frac{Z_\alpha^2}{2(n-1)}$  and  $b = K_T^2 - \frac{Z_\alpha^2}{n}$ .

The quantity  $Z_\alpha$  is the standard normal variable with exceedence probability  $\alpha$ .

## 4 Results

### 4.1 Flood Discharges

Annual maximum discharge series of Chindwin River at Hkamti, Htamathi and Homalin stations and Yenwe at Myochaung station are checked whether outliers exist in the data series before using them. Based on the skewness coefficient of the data series used, outliers are calculated and given in Table 4.  $K_N$  value based on the number of data in the series is obtained from the table for outlier test in [3].

It is observed that the annual maximum discharges in the year 1994 are smaller than the low outliers for Chindwin at all stations. However, useful historical information is not available to adjust for low outliers and therefore they are retained in the series. There is no high outlier observed in all series.

Flood discharges with recurrence intervals of 2, 5, 10, 20, 50 and 100 years are calculated using four probability distributions: Log-normal, Gumbel, Pearson Type III and Log Pearson type III distributions. The results are given in Tables 5, 6, 7, 8 that show large differences in results obtained by the different methods, particularly at the larger return periods. It can be seen that Gumbel distribution gives the highest estimated flood discharges for Chindwin and Log-Normal distribution gives the highest for Yenwe. The estimated discharges for high return period (50 and 100 years) using Log-Normal and Gumbel distributions yield about 10–20 % higher than the flood values estimated by Pearson Type III and Log-Pearson Type III distributions for all basins under study.



**Table 4** Outliers (cumec) for Chindwin and Yenwe basins

River	Gauging station	No. of data	$K_N$	Low outlier	High outlier
Chindwin	Hkamti	24	2.467	8,922	22,105
Chindwin	Htamathi	25	2.486	10,197	22,920
Chindwin	Homalin	28	2.534	10,220	25,831
Yenwe	Myochaung	28	2.534	211.21	1,920

**Table 5** Flood discharges  $X_T$  (cumec) obtained by different distributions for Chindwin at Hkamti station

Return period (years)	Log-normal	Gumbel	Pearson type III	Log-Pearson type III
2	14,043	13,902	14,485	14,724
5	16,394	16,346	16,258	16,285
10	17,776	17,963	17,065	16,811
20	19,005	19,515	17,671	17,110
50	20,489	21,520	18,294	17,324
100	21,542	23,029	18,674	17,408

**Table 6** Flood discharges  $X_T$  (cumec) obtained by different distributions for Chindwin at Htamathi station

Return period (years)	Log-normal	Gumbel	Pearson type III	Log-pearson type III
2	15,287	15,111	15,544	15,587
5	17,534	17,582	17,520	17,572
10	18,837	19,216	18,516	18,522
20	19,987	20,785	19,320	19,253
50	21,364	22,813	20,204	20,012
100	22,334	24,338	20,782	20,479

**Table 7** Flood discharges  $X_T$  (cumec) obtained by different distributions for Chindwin at Homalin station

Return period (years)	Log-normal	Gumbel	Pearson type III	Log-Pearson type III
2	16,248	16,091	17,002	17,163
5	18,952	18,751	18,654	18,696
10	20,542	20,512	19,270	19,114
20	21,954	22,202	19,670	19,306
50	23,660	24,385	20,015	19,401
100	24,870	26,028	20,192	19,420

### 4.2 Statistical Analysis

The two most commonly used tests of goodness of fit namely Chi square and Kolmogorov–Smirnov tests are applied to check the fit of probability distributions used in this study. Tables 9 and 10 list the values of Chi square statistics ( $\chi^2$ ) and Kolmogorov–Smirnov ( $D_n$ ) respectively.

**Table 8** Flood discharges  $X_T$  (cumec) obtained by different distributions for Yenwe at Myo-chaung station

Return period (years)	Log-normal	Gumbel	Pearson type III	Log-Pearson type III
2	636.8	650.6	681.0	667.6
5	918.7	931.7	918.8	924.3
10	1112.9	1117.8	1049.8	1069.2
20	1303.8	1296.3	1161.3	1191.6
50	1558.0	1527.0	1290.6	1330.0
100	1754.4	1700.6	1379.1	1421.1

**Table 9** Comparison of Chi square statistics for different distributions for the basins under study

Basin	Log-normal	Gumbel	Pearson type III	Log-Pearson type III
Hkamti	4.75	5.86	4.96	5.56
Htamathi	6.03	3.98	4.71	3.51
Homalin	3.01	3.01	3.48	5.91
Yenwe	2.5	2.81	4.17	1.84

**Table 10** Comparison of Kolmogorov–Smirnov statistics for different distributions for the basins under study

Basin	Log-normal	Gumbel	Pearson type III	Log-Pearson type III
Hkamti	0.13	0.13	0.091	0.18
Htamathi	0.11	0.11	0.10	0.14
Homalin	0.12	0.14	0.14	0.21
Yenwe	0.11	0.06	0.10	0.11

It can be observed from the Chi square values of four distributions for all basins given in Table 9 that the values are less than the critical value at 95 % confidence level. From Table 10, it can be seen that all the computed values of  $D_n$  are less than the critical value at 95 % confidence level. From both statistical tests, no distinction among all distributions can be made for the basins under study. All distributions are acceptable to fit to the annual maximum series of Chindwin and Yenwe basins at 95 % confidence interval.

If the goodness of fit is the criterion used to select a specified distribution, then the distribution with three parameters will usually provide a much better fit than two parameters [16]. In this study, Log-Pearson Type III is selected as the recommended distribution since it includes the skew coefficient as a variable and therefore is more flexible than the Log-Normal which has a skew of zero of the logarithms. Log-Normal is considered as the special case of Log-Pearson Type III distribution. Another reason is that Pearson Type III is capable of fitting frequency relations that may, for hydrologic reasons, be highly skewed [1].

### 4.3 Reliability Analysis

Reliability analysis for the basins under study is performed using log-Pearson Type III distribution. The confidence limits with 90 % confidence interval are calculated and shown in Tables 11 and 12.

Graphical presentations of estimated flood discharges using all probability distributions under study together with the confidence limits using Log-Pearson Type III for each basin are shown in Fig. 2.

It can be seen from Fig. 2 that the confidence interval is quite wide especially in Yenwe basin since the sample size of each series is small in this study.

## 5 Conclusions

Based on the analysis of statistical tests, Gumbel, Pearson Type III and Log-Pearson Type III and Log-Normal distributions are suitable to estimate flood discharges for Chindwin and Yenwe basins.

Log-Pearson Type III distribution can be recommended for Chindwin and Yenwe basins since it includes the skew variable and therefore is more flexible than other distributions.

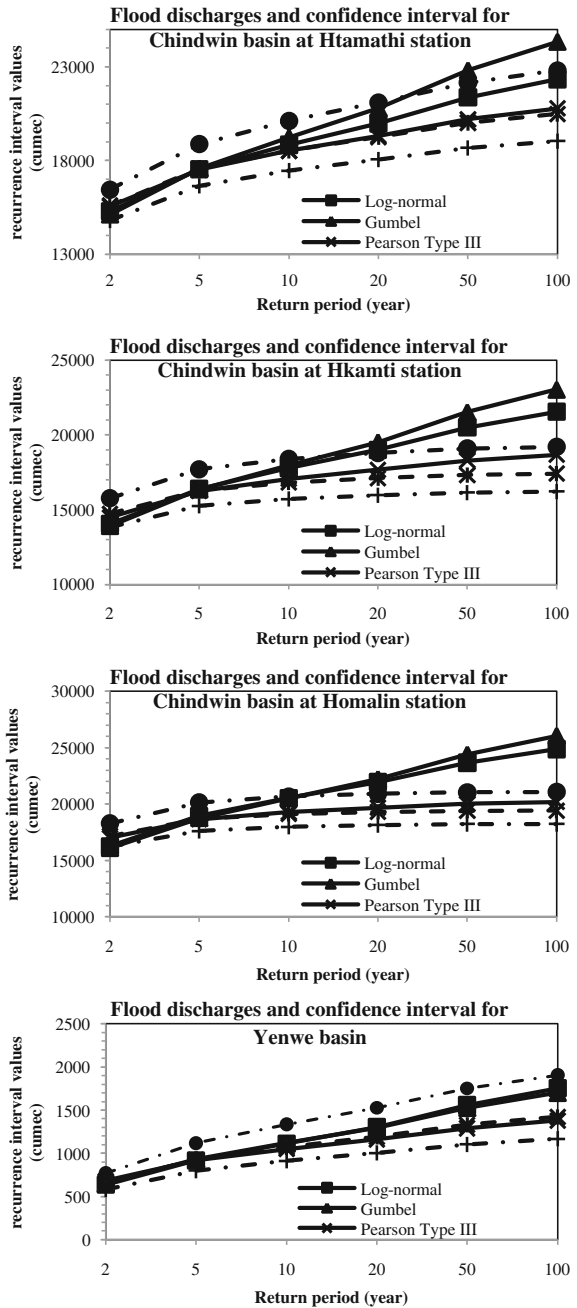
**Table 11** Confidence limits (cumec) for Chindwin at Hkamti and Htamathi stations

Basin	Chindwin at Hkamti		Chindwin at Htamathi	
	Upper limit	Lower limit	Upper limit	Lower limit
2	15755.7	13840.6	16444.4	14809.3
5	17703.1	15261.3	18867.0	16639.6
10	18393.6	15713.8	20115.7	17450.6
20	18793.5	15967.1	21103.9	18055.3
50	19082.1	16146.3	22151.7	18670.2
100	19196.0	16216.0	22805.6	19042.8

**Table 12** Confidence limits (cumec) for Chindwin Basin at Homalin and Yenwe Basin

Basin	Chindwin at Homalin		Yenwe	
	Upper limit	Lower limit	Upper limit	Lower limit
2	18273.3	16214.5	770.0	582.1
5	20141.2	17612.7	1117.9	798.9
10	20672.2	17979	1334.7	911.2
20	20917.8	18144.7	1526.8	1002.3
50	21041.2	18227.2	1752.3	1102.1
100	21065.8	18243.7	1905.6	1166.5

**Fig. 2** Flood discharges and confidence interval vs return period for Chindwin and Yenwe basins



A study on frequency analysis for other gauging stations of Chindwin basin as well as the other river basins needs to be performed to develop the regional frequency analysis. Future research can be carried out by using other parameter estimator such as L moments in frequency analysis for the river basins in Myanmar.

**Acknowledgments** Thanks are due to staff officers from Hydrology section, Irrigation Department and Meteorology and Hydrology Department, Myanmar for their friendly collaboration in data gathering.

## References

1. M.A. Benson, Uniform flood-frequency estimating methods for federal agencies. *Water Resour. Res.* **4**(5), 891–908 (1968)
2. European Cooperation in Science and Technology, *Review of applied-statistical methods for flood-frequency analysis in Europe*, Centre for Ecology and hydrology, Natural Environmental Research Council, (2012)
3. Interagency Advisory Committee on Water Data, Guidelines for determining flood flow frequency: Hydrology Subcommittee Bulletin #17B, available from Office of Water data Coordination, U.S. Department of the Interior Geological Survey, Reston, VA, 1982
4. D.A. Eash, Techniques for estimating flood-frequency discharges for the streams in IOWA, US Geological Survey, Water-Resources Investigation Reports 00-4233, IOWA Department of Transportation and the IOWA Highway Research Board (Project HR-395A), 2001
5. P. Crochet, Evaluation of two delineation methods for regional flood frequency analysis in Northern Iceland, Report, Icelandic Meteorological Office, (2012)
6. A.O. Opere, S. Mkhandi, P. Willems, At site flood frequency analysis for the Nile equatorial basins. From [http://www.unesco.org/fileadmin/MULTIMEDIA/FIELD/Cario/pdf/AT\\_SITE\\_FLOOD-FREQUENCY\\_ANALYSIS.pdf](http://www.unesco.org/fileadmin/MULTIMEDIA/FIELD/Cario/pdf/AT_SITE_FLOOD-FREQUENCY_ANALYSIS.pdf)
7. N. Mujere, Flood frequency analysis using the Gumbel distribution. *Int. J. Comput. Sci. Eng.* **3**(7), 2774–2778 (2011)
8. R.M. Vogel, I. Wilson, Probability distribution of annual maximum, mean, and minimum streamflows in the United States. *J. Hydrologic Eng.* **1**, 69–76 (1996)
9. V.T. Chow, D.R. Maidment, L.W. Mays, *Applied Hydrology* (McGraw-Hill Inc, New York, 1988)
10. K.C. Patra, *Hydrology and Water Resources Engineering*, 2nd edn. (Alpha Science International Ltd, Oxford, 2008)
11. K. Subramanya, *Engineering Hydrology*, 3rd edn. (Tata McGraw Hill Publishing Company Limited, New Delhi, 2009)
12. O.S. Selaman, S. Said, F.J. Putuhena, Flood frequency analysis for Sarawak using Weibull, Gringorten and L-moments formula. *J. Inst. Eng. Malaysia* **68**(1), 43–52 (2007)
13. U.N. Ahmad, A. Shabri, Z.A. Zakaria, Flood frequency analysis of annual maximum stream flows using L-Moments and TL-Moments. *Appl. Math. Sci.* **5**(5), 243–253 (2011)
14. J.R. Stedinger, V.W. Griffis, Flood frequency analysis in the United States: Time to update. *J. Hydrologic Eng. ASCE* **13**, 199–204 (2008)
15. A. Hayter, *Probability and Statics for Engineers and Scientists*, Thomson Higher Education, 3rd edn. (2007)
16. R.M. Vogel, D.E. McMartin, Probability plot goodness-of-fit and skewness estimation, procedure for the Pearson type 3 distribution. *Water Resour. Res.* **27**(12), 3149–3158 (1991)