Invention Journal of Research Technology in Engineering & Management (IJRTEM)
ISSN: 2455-3689
www.ijrtem.com Volume 2 Issue 11 ǁ November 2018 ǁ PP 16-26

Comparison of Various Frequency Distributions (Chindwin Basin), Myanmar

Mi Yin San[1], Zin Zin Zaw[2]
[1](Civil Department, West Yangon Technological University, Myanmar)
[2](Civil Department, Pyay Technological University, Myanmar)

ABSTRACT : *In this paper, computed flood values obtained by frequency analysis for each station in the selected basin are compared with the values by HEC statistical software package whether or not to be reliable. Flood frequency analysis refers to the application of frequency analysis to study the occurrence of floods. Using the annual flood peak data series of three hydrological stations in Chindwin basin, the various probability distributions are analyzed with four methods such as Normal, Log-Normal, Pearson Type-III and Log-Pearson Type III. Chi-square and Kolmogorov-Smirnov (K-S) tests, the most commonly used tests of goodness of fit, are applied to check the fit of probability distribution used in this study. Standard errors are also calculated for each kind of probability distributions. Based on standard errors and reliability of analysis, it can be concluded that Pearson type III distribution function is the most appropriate distributions for the area under study*.

KEYWORDS: *Normal distribution, Log-Normal distribution Pearson Type III distribution, Log-Pearson Type-III distribution, Chi-square and Kolmogorov-Smirnov tests.*

I. INTRODUCTION

Flood is a natural disaster that has caused extensive damages in the world over the past years. In Myanmar, flood occurrences are used to face yearly and causes of life, injury and inconvenience. All extreme flood is usually caused by heavy rainfall at a time when conditions allow the highest possible rates of runoff. Flood frequency analysis imparts the import role in forecasting of flood events up to certain return period. The relation between the magnitude of flood events and their frequency of occurrence is very useful in the construction of flood control structure. Flood frequency analysis is stated by G.W Kite (1997) that is not only of use as an aid in averting disaster but also a means of introducing efficient designs. In this study, four methods such as Normal, Log-Normal, Pearson Type-III and Log-Pearson Type III are used to get the probability distributions using the annual data series from three stations of Chindwin basin. Chi-square and Kolmogorov-Smirnov (K-S) tests are also applied to check the fit of probability distribution.

In summary, the objectives of this study are (i) to find the annual maximum discharge of the representative recurrence interval by applying HEC Statistical Software Package, (ii) to test the goodness of fit and to compare standard error for different probability distributions applied in this study and (iii) to compare the results from various distributions to determine how distributions are fitted for each data series.

II. MATERIALS AND METHODS

Characteristics of Chindwin Basin : The Chindwin basin occupies almost the entire North Western part of Myanmar. The Chindwin with its tributaries all together comprises the upper Chindwin catchment area. The large tributaries of Chindwin river are U Yu and Myittha, where U Yu flows into Chindwin near Homalin and Myittha near Kalewa. Chindwin river is 877.51 km long and its junction with the Ayeyarwady is situated about 16.09 km north-east of Pakokku. The width of the river varies from 91.44 to 3048 m. Chindwin catchment area covers 110350 sq. km. The Chindwin basin has approximately 120,000 acres of cultivated land. About 90% of the basin is thickly forested by valuable species of wood. The location of Chindwin Basin and each of stations are shown in Fig 1.

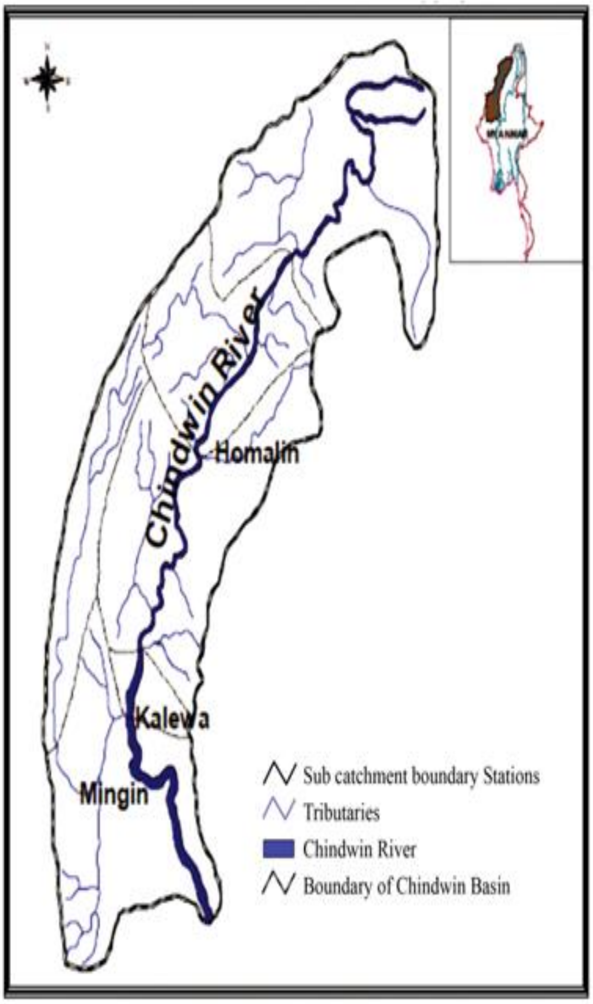


Figure 1. Map Showing the location of each station

Hydrological Data Collection : The respective annual data series for each station in Chindwin basin is collected from the Meteorological and Hydrological Department, Ka Bar Aye in Yangon. Although it is collected the periods from 1970 to 2009 for Homalin and Kalewa stations, only data from 1989 t0 2009 can get for Mingin station. According to Shaw (1994), at least 20 years of data should be obtained in order to achieve reliable results. The series of values from a period of observation (ie., 40 years for Homalin and Kalewa stations and 21 years for Mingin station) is shown in Table 1.

Statistical Properties of Hydrological Data : The statistical characteristics of Annual Maximum Discharges (cumecs) of three gate stations along the Chindwin river are given in Table 2.

Statistical Analysis of Data Series : Most of statistical methods used in hydrologic studies are based on the assumption that the observations are independently distributed in time and space. The occurrence of an event is assumed to be independent of all previous events. Therefore, statistical analysis is made whether the hydrological data behave as is assumed by the hypothesis, especially whether they are independent, random and homogeneous.

Table 1. Annual Flood Peak Data Series (Cumecs) of Chindwin River.

Year	Homalin Station	Kalewa Station	Mingin Station
1970	17470 ON AUG 1	19820 ON AUG 3	
1971	13440 ON AUG 7	17573 ON AUG 10	
1972	14370 ON JUL 31	15323 ON AUG 4	
1973	18760 ON AUG 9	22170 ON AUG 12	
1974	17990 ON AUG 4	24543 ON AUG 15	
1975	15407 ON AUG 2	17550 ON AUG 5	
1976	20703 ON JUL 18	25713 ON JUL 22	
1977	15480 ON SEP 1	23753 ON SEP 3	
1978	16730 ON JUN 30	18193 ON JUL 2	
1979	17690 ON SEP 10	21050 ON SEP 13	
1980	17270 ON OCT 7	25877 ON OCT 8	
1981	15668 ON JUL 7	15601 ON JUL 10	
1982	16073 ON JUL 30	23243 ON AUG 3	
1983	15000 ON AUG 5	20975 ON AUG 6	
1984	18139 ON JUL 17	22451 ON JUL 19	
1985	15865 ON JUL 11	20330 ON JUL 16	
1986	12540 ON OCT 11	14120 ON SEP 13	
1987	15830 ON SEP 29	20090 ON AUG 23	
1988	17417 ON JUL 31	25061 ON AUG 4	
1989	19230 ON AUG 1	23580 ON AUG 6	23170 ON AUG 6
1990	16360 ON JUL 30	21820 ON JUL 24	21540 ON AUG 2
1991	20087 ON JUL 20	25063 ON JUL 22	24930 ON AUG 17
1992	11103 ON OCT 12	13373 ON JUL 12	14980 ON JUL 13
1993	16660 ON SEP 1	22180 ON SEP 5	21960 ON SEP 6
1994	8239 ON AUG 16	10867 ON JUL 31	12360 ON AUG 28
1995	16980 ON JUL 20	22213 ON JUL 22	21260 ON JUL 22
1996	14000 ON JUL 22	16737 ON JUL 26	16940 ON JUL 27
1997	19470 ON JUL 16	25817 ON JUL 19	25140 ON SEP 30
1998	17830 ON SEP 7	23033 ON SEP 9	22640 ON SEP 10
1999	16730 ON AUG 30	26083 ON SEP 1	24740 ON SEP 2
2000	13007 ON OCT 4	19740 ON OCT 5	19690 ON OCT 5
2001	11287 ON AUG 30	12863 ON AUG 6	13777 ON AUG 6
2002	19233 ON AUG 15	26220 ON AUG 18	23250 ON AUG 18
2003	17280 ON JUL 7	18767 ON JUL 19	18780 ON JUL 18
2004	18370 ON JUL 24	20900 ON JUL 25	21893 ON JUL 26
2005	15870 ON AUG 28	15900 ON AUG 31	16620 ON AUG 31
2006	12820 ON SEP 15	14770 ON SEP 21	16012 ON SEP 20
2007	18295 ON JUL 29	21494 ON SEP 10	20000 ON SEP 10
2008	18700 ON JUL 8	23720 ON JUL 10	17870 ON AUG 23
2009	14768 ON AUG 24	17005 ON AUG 27	16704 ON AUG 28

Test for Outliers: Outliers are data points that depart significantly from the trend of the data. The following frequency equations (1) can be used to detect for high outliers and low outliers (Chow et al. 1988).

Table 2. Statistical Characteristics of Annual Maximum Discharges (cumecs)

Sr No	Station	N	$\bar{x}$	$\bar{y}$	Sx	Sy	Csx	Csy	Cv	Ck
1	Homalin	40	16204	4.2030	2653	0.0799	-0.8699	-1.5079	0.1637	4.0596
2	Kalewa	40	20389	4.2997	4134	0.0961	-0.4672	-0.8612	0.2027	2.5018
3	Mingin	21	19820	4.2885	3900	0.0901	-0.2350	-0.5575	0.1968	2.4066

$$Y_H = \bar{y} + K_n S_y\,,\quad Y_L = \bar{y} + K_n S_y \qquad (1)$$

Each data series for outliers and results are shown in Table 3.

Table 3. Results for Outliers Test

Sr No	Station	Kinds of Outlier	Computed Value	Observed Value	No. of Detected Outlier
1	Homalin	High	26134	20703	0
		Low	9745	8239	1
2	Kalewa	High	36072	26220	0
		Low	11018	10867	1
3	Mingin	High	32020	25230	0
		Low	11793	12360	0

Test for Independence: Spearman Test (Mc Cuen 1992) is used for testing independence of the data series. The restriction of independence assures that a hydrologic event such as a single large storm does not enter the data set more than once. If the 't' values are less than the tabulated value with 95% confidence limit, the data are assumed to be a random sequence. The results are shown in Table 4.

Table 4. Results for Independent Test

Sr No.	Station	Computed 't'	Tabulated 't'	Remark
1	Homalin	-1.4473	1.69	Accept, A
2	Kalewa	-1.7627	1.69	A
3	Mingin	-1.5446	1.73	A

Test for Randomness: Many statistical methods assume that the data values of a random variable that are in sequence, but with independence between the measured values. The run test can be used to test a sample of data for randomness (Mc Cuen 1992). The results are shown in Table 5.

Table 5. Results for Randomness Test

Sr No.	Station	No. of runs	Observations Above Median	Observations below Median	Range	Remark
1	Homalin	22	20	20	15 to 27	A
2	Kalewa	23	20	20	15 to 27	A
3	Mingin	12	10	10	7 to 15	A

Test for Homogeneity: The Mann-Whitney test may be used for testing homogeneity of data series. The Mann-Whitney U test is a nonparametric test for independent sample and can be used to test whether two independent samples have been taken from the same population. The restriction of homogeneity assures that all the observations are from the same population. The results are shown in Table 6.

Table 6. Results for Homogeneity Test

Sr No.	Station	Computed 'z'	Tabulated 'z'	Remark
1	Homalin	-0.2032	1.96	A
2	Kalewa	-0.4878	1.96	A
3	Mingin	-0.9155	1.96	A

Application of Flood Frequency Analysis

Frequency Factor Method: The magnitudes of extreme events can be calculated by various methods. In this study frequency analysis is carried out by frequency factor method. A general equation for frequency analysis of hydrologic events is given as (Chow, 1964)

$x_T = \bar{x} + K\,\sigma$……………………………………………………….(2)

where

K = frequency factor

$\bar{x}$ = mean value of variate

σ = standard deviation

The various probability distributions, are analyzed with four methods such as Normal, Log-Normal, Pearson Type-III and Log-Pearson Type III. Normal and Log-normal distributions are adopted from Normal family and each has two parameters. Pearson Type III and Log-Pearson Type III distributions are derived from Gamma family and each has three parameters.

Statistical Analysis of Fitting Probability Distributions: The goodness of fit of a probability distribution can be tested by comparing the theoretical and samples values of the relative frequency or the cumulative frequency distribution. The most commonly used tests of goodness of fit, Chi-square and Kolmogorov (K-S) tests are applied for checking the fit of probability distributions used in this study. For the application of the Chi-square method, the number of class intervals, k, should not be less than 5 and the expected absolute frequency in each class is at least 5. The Kolmogorov statistic, which is denoted by D_n , is the maximum absolute difference between the values of the cumulative distribution of a random sample and a specified probability distribution function. The K-S test may be used for small samples. It is generally more efficient than the Chi-square test when the sample is small (Chow 1988).

The class interval (CI) were computed for various distributions as follows

$CI = \bar{x} + K_T\, S$……………………………………………………(3)

Table 7 through 9 are the lists of the computed class limits for each distribution together with calculated Chi-square value and calculated value of D_n . From these tables, it can be seen that all the computed valued of Chi-square are less than the tabulated values o 95% confidence limit and all distributions are well within the acceptance limit of D_n value.

Table 7. Comparison of Computed and Tabulated Results obtained by Chi-square and K S tests (Homalin)

Class Interval	Probability	Class limits for different distributions			
		Normal	Log-normal	Pearson III	Log-Pearson III
1	0.1667	13686	13404	13813	13706
2	0.3333	15069	14753	15423	15424
3	0.5000	16204	15959	16580	16665
4	0.6667	17345	17273	17606	17725
5	0.8333	18770	19066	18704	18771
6	1.0000	∞	∞	∞	∞
Chi-square		3.2018	5.6026	0.7984	1.3984
D_n		0.0583	0.1083	0.0333	0.0417
x^2 (0.95,k-m-1)		7.810	7.81	5.99	5.99
D (0.05, 40)		0.210	0.210	0.210	0.210

Table 8. Comparison of Computed and Tabulated Results obtained by Chi-square and K S tests (Kalewa)

Class Interval	Probability	Class limits for different distributions			
		Normal	Log-normal	Pearson III	Log-Pearson III
1	0.1667	16466	16162	16538	16330
2	0.3333	18621	18138	18904	18677
3	0.5000	20389	19939	20709	20568
4	0.6667	22168	21933	22408	22408
5	0.8333	24387	24694	24367	24564
6	1.0000	∞	∞	∞	∞
Chi-square		2.6012	4.1016	1.9982	1.3984
D_n		0.0917	0.1000	0.05	0.05
x^2 (0.95,k-m-1)		7.810	7.81	5.99	5.99
D (0.05, 40)		0.210	0.210	0.210	0.210

Table 9. Comparison of Computed and Tabulated Results obtained by Chi-square and K S tests (Mingin)

Class Interval	Probability	Class limits for different distributions			
		Normal	Log-normal	Pearson III	Log-Pearson III
1	0.2	16588	16361	16644	16504
2	0.4	18837	18442	18983	18789
3	0.6	20806	20478	20946	20830
4	0.8	23101	23137	23137	23206
5	1.0	∞	∞	∞	∞
Chi-square		1.6190	0.6667	1.6190	2.0952
D_n		0.0762	0.0762	0.0762	0.0762
x^2 (0.95,k-m-1)		5.99	5.99	3.84	3.84
D (0.05, 21)		0.2859	0.2859	0.2859	0.2859

Results of Flood Frequency Analysis: Annual maximum discharges of Chindwin river at Homalin, Kalewa and Mingin stations are analyzed in this study. Four probability distributions; namely, Normal, Log-normal, Pearson Type III and Log-Pearson Type III are applied to data series to calculate recurrence interval values for different return period of 1.01, 1.05, 1.11, 1.25, 2, 5, 10, 20, 50, 100 and 500 years by HEC. Statistical Software Package. The results are given in Table 10 through 12.

Table 10. Comparison of T year Event Magnitude Using Various Frequency Distribution (Cumecs) for Homalin Station

Return Period (year)	Exceedence Probability	Normal	Log-Normal	Pearson III	Log-Pearson III
1.01	0.990	10031	10405	8409	8643
1.05	0.950	11839	11794	11288	11147
1.11	0.900	12803	12609	12653	12490
1.25	0.800	13971	13672	14155	14060
2	0.500	16204	15960	16584	16684
5	0.200	18437	18632	18472	18574
10	0.100	19604	20202	19263	19240
20	0.050	20569	21598	19823	19639
50	0.020	21654	23285	20360	19950
100	0.010	22377	24482	20667	20091
200	0.005	23039	25632	20915	20183
500	0.002	23841	27097	21177	20260

Table 11. Comparison of T year Event Magnitude Using Various Frequency Distribution (Cumecs) for Kalewa Station

Return Period (Year)	Exceedence Probability	Normal	Log-Normal	Pearson III	Log-Pearson III
1.01	0.990	10773	11918	9382	10424
1.05	0.950	13590	13857	13087	13239
1.11	0.900	15092	15017	14927	14830
1.25	0.800	16910	16551	17037	16803
2	0.500	20389	19938	20710	20573
5	0.200	23869	24017	23928	24088
10	0.100	25687	26472	25437	25737
20	0.050	27189	28687	26597	26975
50	0.020	28880	31402	27813	28222
100	0.010	30006	33354	28571	28963
200	0.005	31038	35246	29230	29578
500	0.002	32288	37684	29985	30241

Table 12. Comparison of T year Event Magnitude Using Various Frequency Distribution (Cumecs) for Mingin Station

Return Period (Year)	Exceedance Probability	Normal	Log-Normal	Pearson III	Log-Pearson III
1.01	0.990	10747	11994	10080	11041
1.05	0.950	13405	13815	13156	13413
1.11	0.900	14822	14897	14735	14760
1.25	0.800	16538	16320	16591	16451
2	0.500	19821	19433	19973	19810
5	0.200	23103	23140	23140	23214
10	0.100	24819	25351	24710	24963
20	0.050	26236	27336	25965	26372
50	0.020	27831	29756	27331	27907
100	0.010	28894	31487	28215	28893
200	0.005	29867	33159	29006	29766
500	0.002	31046	35305	29942	30783

Figure 2 through 4 shows the event magnitudes from each of four distributions together for each station. It will be noted from the figure that the distributions are very closely grouped except for Log-normal in all stations. The Pearson Type III and Log-Pearson Type III are so closely as to be indistinguishable at this scale for all stations.

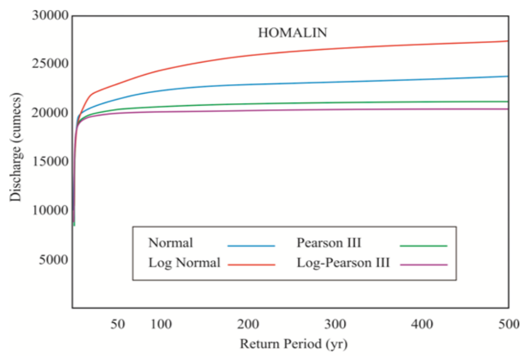


Figure 2. Comparison of Fequency Curves From Various Distibution (Homalin)

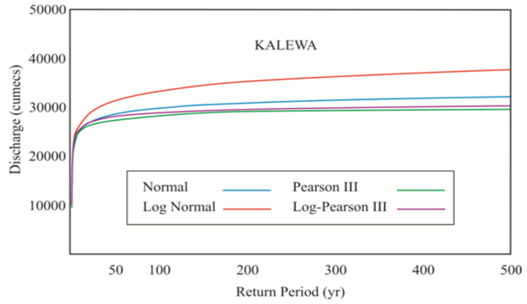


Figure 3. Comparison of Fequency Curves From Various Distibution (Kalewa)

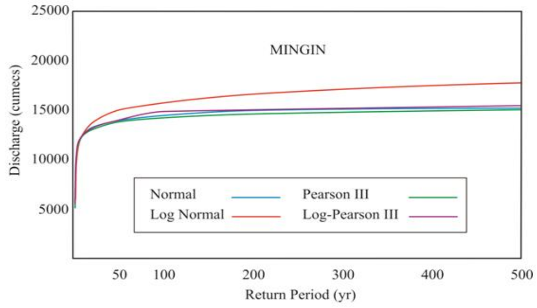


Figure 4. Comparison of Fequency Curves From Various Distibution (Mingin)

III. RESULTS OF STANDARD ERROR

Each distribution gives a different standard error of estimate for different return period of 5, 10, 20, 50, 100, 200 and 500 years by method of moment (MOM) (Karl Pearson, Chow et al 1998). The most efficient distribution is that which gives the smallest standard error of estimate. Therefore, Pearson III distribution is fitted well for all stations and log-Pearson Type III distributions gives the largest standard error of estimate for all stations Table 13. Comparison of Standard Errors of T Year Events Using Various Frequency Distributions (Cumecs) For Homalin Station

ReturnPeriod (Yrs)	Normal	Log-Normal	Pearson III	Log-Pearson III
5	488	561	368	1702
10	566	690	388	2219
20	643	816	512	2873
50	740	975	746	3546
100	808	1090	937	3932
200	872	1202	1128	4229
500	951	1346	1375	4521

Table 14 Comparison of Standard Errors of T Year Events Using Various Frequency Distribution (Cumecs) For Kalewa Station

Return Period(Yrs)	Normal	Log-Normal	Pearson III	Log Pearson III
5	760	903	666	2788
10	882	1129	719	3056
20	1003	1349	863	3669
50	1153	1631	1144	4629
100	1258	1837	1393	5325
200	1358	2038	1657	5967
500	1482	2298	2017	6742

Table 15 Comparison of Standard Errors of T Year Events Using Various Frequency Distribution (Cumecs) For Mingin Station

Return Period (Yrs)	Normal	Log-Normal	Pearson III	Log-Pearson III
5	990	1170	932	2756
10	1149	1459	1048	3057
20	1305	1741	1259	3558
50	1501	2100	1634	4368
100	1638	2362	1962	5013
200	1768	2618	2311	5640
500	1930	2950	2793	6450

Reliability of Flood Frequency Analysis: The reliability of computed flood values obtained by frequency analysis using all four probability distributions are checked for each basin by HEC Statistical Software Package. The graphical representations of flood estimate obtained by all studied distributions together with confidence limits are shown in fig 5 through 7. From these figures, all studied probability distributions seem to be well fit for all studied basins by visual aspect.

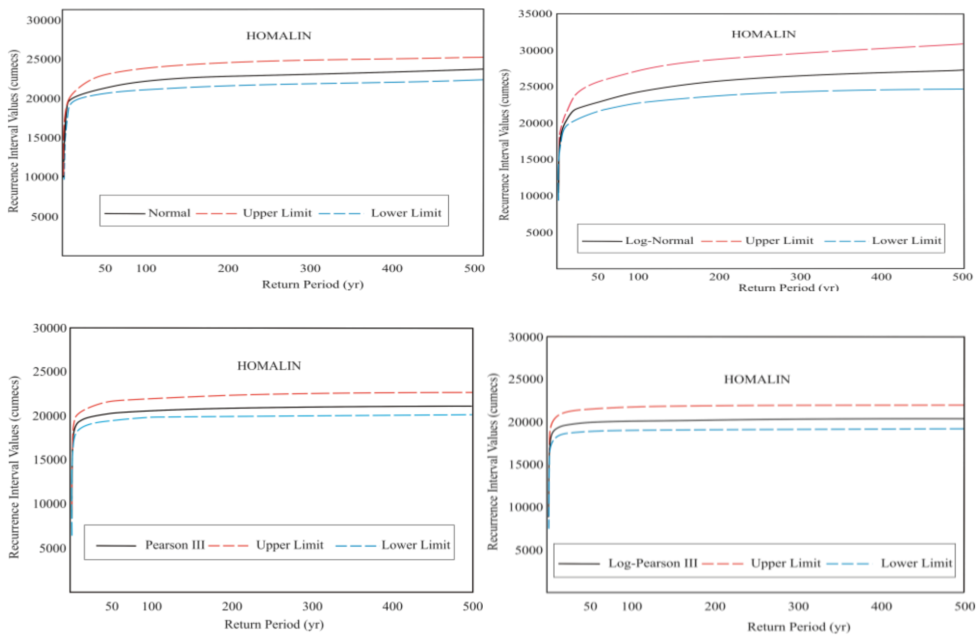
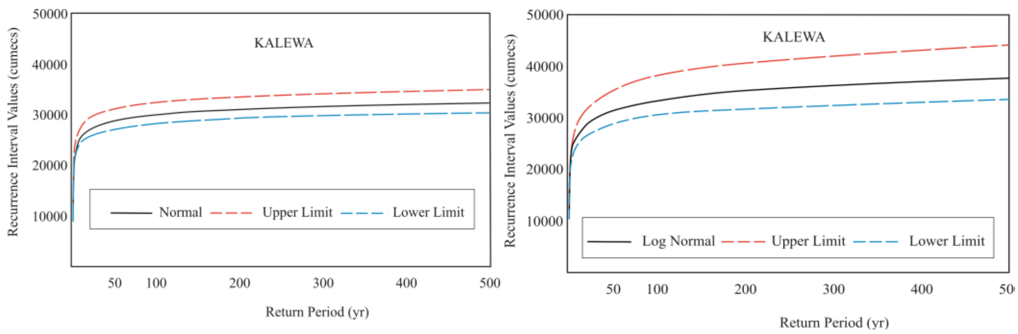


Figure 5. Comparison of Fequency Interval Vs Return Period For Homalin Station



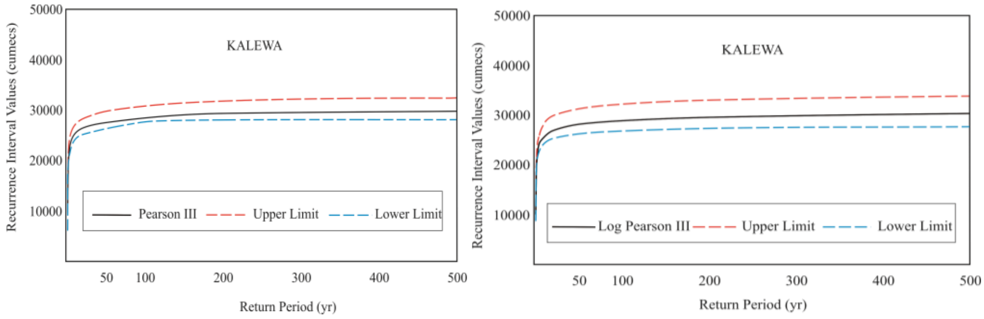


Figure 6. Comparison of Fequency Interval Vs Return Period For Kalewa Station

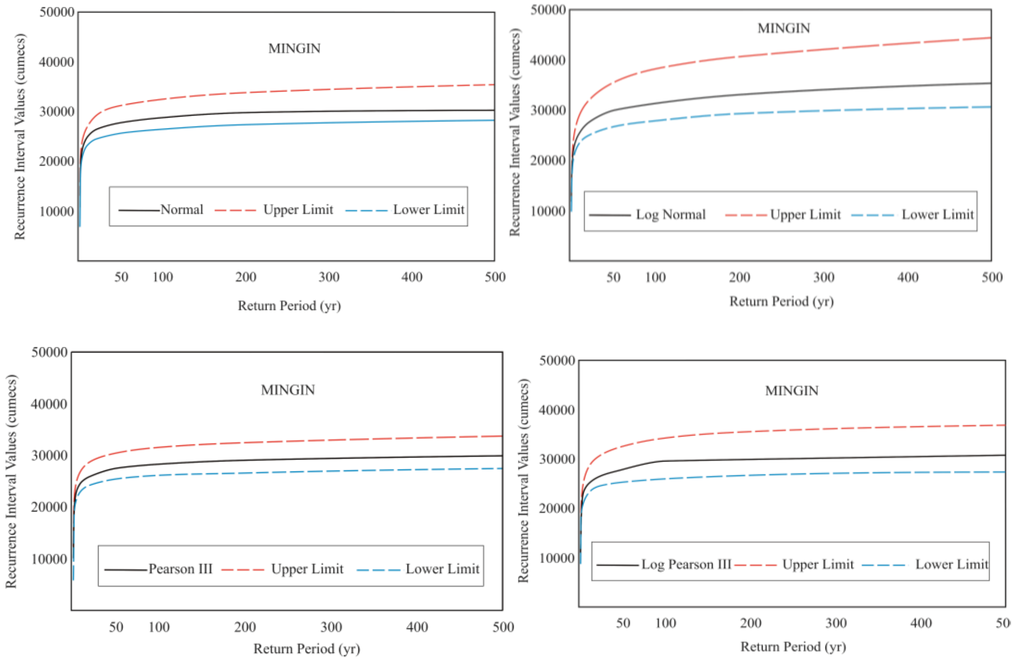


Figure 7. Comparison of Fequency Interval Vs Return Period For Mingin Station

IV. DISCUSSION AND CONCLUSIONS

Annual flood peak data series of Homalin, Kalewa and Mingin stations are used in this study to be compared for various frequency distributions. All studied distributions give statistically more or less the same values for the recurrence interval values up to 10 years. All distributions predict a 500 years event considerably larger than the maximum event recorded in all sample data except for the flood value obtained by Log-Pearson Type III distribution for Homalin station. Four theoretical distribution functions have been tested on annual floods in three stations of Chindwin river. Three parameter distribution function of Pearson Type III have generally been found to give the best fit for all studied stations. If two parameter distribution function has to be applied, Log-normal distribution can be chosen as the best fit for short-recorded sample, as in Mingin. In conclusion, there is no rigorous theoretical foundation for the choice of a distribution function

ACKNOWLEDGEMENT

The author is very thankful to Dr. NiLar Aye, Professor, Department of Civil Engineering, Mandalay Technological University, for indispensable guidance in the preparation of this study. The author is grateful to all teachers from the department of Civil Engineering for their useful advices. The author would like to express special thanks to U Sein Lin, Meteorological and Hydrological Department (Ka Bar Aye) for sharing his experiences.

REFERENCES

1. Rao, A.R. and Hamed, K.H. *Flood Frequency Analysis*, CRC Press, Boca Raton, Florida. 1999
2. Kite, G.W. *Frequency and Risk Analyses in Hydrology*. Fort Collins, Colorado: Water Resources Publications. 1999

3. Nilar Aye, Ma. *Flood Regionalization using rainfall and basin characteristics of Chindwin Catchments*. Ph. D Preliminary Research Department of Civil Engineering, Yangon Technological University. 2001
4. Shaw, E.M. Hydrology in Practice. U.K, Chapman and Hall 1994
5. Chow, V.T., Maidment, D.R. and L.W.1998. Applied Hydrology. New York: Mc Graw Hill Co
6. N.K. Geol, Htay Htay Than, Flood hazard Mapping in the lower part of Chindwin River Basin, Myanmar, Indian Institute of Technology, India. 2005.