FLOOD FREQUENCY ANALYSIS FOR SARAWAK USING WEIBULL, GRINGORTEN AND L-MOMENTS FORMULA

Onni S. Selaman[1], Salim Said[1] and F.J. Putuhena[1]
[1]*Faculty of Engineering, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak*
E-mail: onni70@yahoo.com, E-mail: ssalim@feng.unimas.my

ABSTRACT
The objective of this study is to determine the magnitude and frequency of floods for Sarawak using Gumbel distribution. Nineteen stations were selected for the study based on the criteria stated in Hydrological Procedure No. 4 (HP4). The probability plot and flood-frequency curves by Gumbel distribution of each individual station are prepared using three different plotting position formulas (i.e. Weibull, Gringorten and L-Moments). From the results and analysis of each individual station, Gumbel distribution based on L-Moments always give the least ratio of peak discharge of T year's recurrence interval over mean annual flood (QT/MAF) but at some stations, it gives unreasonable return period (T) and reduced variate (y) range. The appropriateness of L-moments with Gumbel distribution had some limitation. It is only good for small samples data. If compared between Gumbel distribution by Weibull formula and Gumbel distribution by Gringorten formula, the latter is better because it gives the least ratio (which is in agreement with the literature). Therefore, it could be concluded that for some stations, L-Moments method is the best, but since L-Moments method had some limitations, Gringorten formula is still the best plotting position method to be applied with Gumbel distribution.

Keywords: *Gringorten, Gumbel Distribution, L-Moments, Weibull*

1. INTRODUCTION

In the planning and design of water resources projects, engineers and planners are often interested to determine the magnitude and frequency of floods that will occur at the project areas. Besides the rational method, unit hydrograph method and rainfall-runoff models method, frequency analysis is one of the main techniques used to define the relationship between the magnitude of an event and the frequency with which that event is exceeded.

As a guidelines to determine the magnitude and frequency of Floods in Peninsular Malaysia, the Department of Drainage and Irrigation (DID) of Malaysia has published a hydrological procedure called Hydrological Procedure No 4 (HP4) [1]. The procedure is based on the regional frequency analysis method used by the Natural Environmental Research Council (NERC) [2]. In NERC method, the flood frequency analysis of individual station flood data is determined using Gumbel distribution and the theoretical fits are determined by the method of moments. The plotting position of each sample is calculated using the Weibull formula.

Cunnane [3] had studied various plotting position methods using the criteria of unbiasedness and maximum variance. He found that the Weibull plotting position formula was biased, and it plotted the largest values of a sample at too small a return period. He said, for data distributed according to the Extreme Value Type I distribution (or Gumbel distribution), the Gringorten formula (b = 0.44) was the best.

No such procedure has been developed for Sabah and Sarawak but there was a prior research on regional flood estimation for ungauged basins in Sarawak by Lim and Lye [4]. They had examined the flood records in Sarawak using an index-flood estimation procedure based on L-moments technique. They adopted four-parameter Kappa distribution to simulate the flood data. From the simulation, they obtained two homogeneous flood frequency regions. The two regions were described by the Generalised Extreme Value and the Generalised Logistic distributions.

This paper focuses on the application of Gumbel distribution with Weibull Formula, Gringorten Formula and L-Moments Method. It is hoped that the findings from this study could contribute to the knowledge of the application of Gumbel distribution in flood-frequency analysis study in Sarawak.

2. LITERATURE REVIEW

2.1. Probability Distributions for Hydrologic Variables

According to Arnell [5], in principle, it is possible to estimate the frequency of a given magnitude event by using an empirical distribution function (because in the empirical distribution function the magnitude of event are plotted against the proportion of events greater than or equal to that event), but in practice where too few data are available, the empirical distribution produced could not be used to estimate the frequency of occurrence of events larger than the maximum records. He suggested, as an alternative, the samples of data are fitted using a theoretical frequency distribution.

There are several types of theoretical probability distributions (or frequency distribution functions) that have been successfully applied to hydrologic data [6]. Some of the probability distributions commonly used for hydrologic variables were Normal Distribution, Lognormal Distribution, Exponential Distribution, Gamma Distribution, Pearson Type III Distribution, Log-Pearson Type III Distribution and Extreme Value Distribution. Extreme Value Distribution which is further subdivided into three form – EVI (Gumbel Distribution), EVII (Frechet Distribution) and EVIII (Weibull Distribution) [7].

The most popular theoretical probability distributions (or frequency distribution functions) have been the lognormal, log-Pearson Type III and Gumbel distributions [6, 8]. In the United

States and Australia the log Pearson Type III (LPIII) distribution has been selected as a standard by federal agencies [9]. The general extreme value (GEV) distribution is the standard method for flood-frequency analysis in the U.K. [7].

The benefit of using probability distribution in accordance to Gordon et al. [9] is that the estimated parameter values compactly summarise the characteristics of the distribution. Arnell [5] says, the distributions used in hydrology tend to have two or three parameters (and rarely four or five). In general, the more parameters a distribution has, the better it will fit a set of data and the more flexibility it has for fitting many different sets of data [9]. Referring to Haan, [10], often, these parameters are related to factors such as catchments area, rainfall, topography and other physiographical and meteorological measures. Parameters can be estimated from sample data using a range of procedures, including the methods of moments, maximum-likelihood and L-moments [5].

Commenting on the credibility of an estimated frequency distribution, Arnell [5] says, it depends on the degree to which the assumed distribution fits the data, the robustness of the procedure used to estimate parameters from the data, the extent to which the data conform to the assumption that they can be described by a single, smooth theoretical frequency distribution and, perhaps most importantly, the assumption that the nature of the relationship between magnitude and frequency does not change over time. He then concluded, as a general rule, most procedures give similar results within the range of the data, but can give very different magnitude/frequency relationships when extrapolating beyond the largest observed events.

Gordon et al. [9] say, although no one distribution will fit all flood data, specifying the distribution used and the method of fitting it will allow other researchers to obtain some results from the same set of data. The procedure is thus much more objective than geographical methods using eye-fitted curves.

2.2. Generalise Extreme Value (GEV) Distribution

DID and NAHRIM [11] had made a summary of the recommended frequency types based on various location in the world. The summary is as given in Table 1. From the summary, it could be seen that Generalised Extreme Value (GEV) had been accepted throughout the world.

Table 1: Summary of the Recommended Frequency Types

Location	Recommended Frequency Distribution	Site Nos.
UK (1999)	GLO	98
India (1999)	GPA	93
Indiana, USA (1997)	LN3, GEV, LP3	1490
Continental, USA (1996)	LN3, GEV, LP3	19
World (1995)	GEV	
Ontario-Quebec, Canada (1997)	GEV	183
Saskatchewan, Canada (1994)	LN3, PE3	180
Bangladesh (1993)	GEV	31
Australia (1993a)	GEV, GPA, LP3, LN3	61
Southwestern, USA (1993b)	LN3, LN2, GEV, LP3	383
N. Brunswick, Canada (1992)	GEV	53
Nova Scotia, Canada (1992)	GEV	25
New Zealand (1991) – RAINFALL	EV1, EV2, GEV	275
Central Victoria, Australia (1991)	GEV	53
Eastern, USA (1998)	GEV	55

As had been mentioned in Section 2.1, Extreme Value Distribution had been further subdivided into three forms called EVI (or Gumbel Distribution), EVII (or Frechet Distribution) and EVIII (or Weibull Distribution), following the name of the person further developing them. The three limiting forms were shown by Jenkinson [11] to be special cases of a single distribution called the General Extreme Value (GEV) distribution [7].

The probability distribution function for the GEV is :

$$F(x) = \exp\left[-\left(1 - k\,\frac{x-u}{a}\right)^{1/k}\right] \tag{1}$$

where k, u, and α are parameters to be determined. The three limiting cases are : k = 0 for EVI; k < 0 for EVII with $(u + \alpha/k) \leq x \leq \infty$; and k > 0 for EVIII with $-\infty \leq x \leq (u + \alpha/k)$. In all three cases a is assumed to be positive. EVI distribution has no upper or lower limits; EVII distribution is bounded on the below/lower end (by $u + \alpha/k$); whereas EVIII distribution is bounded on the above/upper end (by $u + \alpha/k$). There has been little interest in the EVII distribution in hydrology. The EVI distribution is often used in flood frequency analysis, and a form of the EVIII distribution is commonly used in the analysis of low flows [7, 9]. Detailed information on applied extreme value statistics could be found in Kinnison [13].

2.3. Extreme Value Type I (also known as EVI or Gumbel Distribution)

EVI distribution (or Gumbel distribution) is a double-exponential distribution. According to Ponce [14], the cumulative density function, F(x) of the Gumbel method is:

$$F(x) = e^{-e^{-y}} \tag{2}$$

in which, F(x) is the probability of non exceedence. He added, in flood frequency analysis, the probability of interest is the probability of exceedence (i.e. the complementary probability to F(x)):

$$G(x) = 1 - F(x) \tag{3}$$

Subramanya [8] expresses Equation (3) in the form of Equation (4). He says, according to Gumbel's theory of extreme events, the probability of occurrence of an event equal to or larger than a value x_o is :

$$P\,(X \geq x_o) = 1 - e^{-e^{-y}} \tag{4}$$

in which y is a dimensionless variable given by:

$$y = \alpha\,(x - a)$$
$$a = \bar{x} - 0.45005\sigma_x$$
$$\alpha = 1.2825 / \sigma_x$$

Thus,

$$y = \frac{1.2825\,(x-\bar{x})}{\sigma_x} + 0.577 \tag{5}$$

where $\bar{x}$ = mean and σ_x = standard deviation of the variate X. The return period T is the reciprocal of the probability of exceedence (i.e. T=1/P). Therefore, Equations (3) and (4) could also be written as below:

$$\frac{1}{T} = 1 - e^{-e^{-y}} \quad (6)$$

Subramanya [8] also highlighted that in practice it is the value of X for a given P that is required and as such Equation (4) could be transposed as :

$$y_P = 1n\,[\,-\,1n\,(1 - P)] \quad (7)$$

Noting that T=1/P, he then noticed that by designating y_T = the value of y, commonly called the reduced variate, for a given T, the following equations could be produced :

$$y_T = -1n\left(1n\,\frac{T}{T\text{-}1}\right) \quad (8)$$

or

$$y_T = -\left[0.834 + 2.303 \log\log \frac{T}{T\text{-}1}\right] \quad (8a)$$

Gordon et. al. [9] stated that EVI is described by two parameters, a scale parameter and a location parameter, where the latter is the mode of the distribution. The Extreme Value Type I (EVI) probability distribution function could also be written in the form below [7]:

$$F(x) = \exp\left[-\exp\left(-\frac{x - u}{\alpha}\right) - x \leq x \geq x\right] \quad (9)$$

The parameters are estimated as:

$$\alpha = \frac{\sqrt{6s}}{\pi} \quad (10)$$

$$u = \bar{x} - 0.5772\alpha \quad (11)$$

A reduced variate y can be defined as:

$$y = \frac{x - u}{\alpha} \quad (12)$$

According to Hosking and Wallis [15], u is the location parameter and a the scale parameters. Ponce [14] stated, x is the value of flood discharge and s is the standard deviation. Chow et. al. [7] revealed, substituting the reduced variate into Equation (9) yields :

$$F(x) = \exp\,[-\exp\,(-y)] \quad (13)$$

Note that Equation (13) is the same as Equation (2). Solving for y :

$$y = 1n - \left[1n\left(\frac{1}{F(x)}\right)\right] \quad (14)$$

Further, according to Chow *et.al*. [7], values of return period T as an alternate axis to y:

$$\frac{1}{T} = P\,(x \geq x_T)$$
$$= 1 - P(x < x_T)$$
$$= 1 - F(x_T)$$

So,

$$F(x_T) = \frac{T\text{-}1}{T}$$

and, substituting F(*xT*) into Equation (14),

$$y = 1n - \left[1n\left(\frac{T}{T\text{-}1}\right)\right] \quad (15)$$

Note that Equation (15) is the same as Equation (8). Chow *et. al*. [7] then further elaborated that for the EVI distribution, x_T is related to y_T by Equation (12), or

$$x_T = u + \alpha y_T \quad (16)$$

According to Ponce [14], in the Gumbel Method, values of flood discharge are obtained from the frequency formula:

$$x = \bar{x} + Ks \quad (17)$$

The frequency factor K is evaluated with the frequency formula:

$$y = \bar{y}_n + K\sigma_n \quad (18)$$

in which y = Gumbel (reduced) variate, a function of return period; $\bar{y}_n$ = the mean of the Gumbel variate; σ_n = the mean standard deviation of the Gumbel variate; $\bar{y}_n$ and σ_n and values are a function of record length n. Gumbel [16] define $\bar{y}_n$ and σ_n and values as a function of record length n.

Ponce [14] added, in Equation (17), for K = 0, x is equal to the mean annual flood $\bar{x}$. Likewise, in Equation (18), for K = 0, the Gumbel variate y is equal to its mean $\bar{y}_n$. The limiting value of $\bar{y}_n$ for n approaching ∞ is the Euler constant, 0.5772. In Equation (6), for y = 0.5772, the return period, T = 2.33 years. Therefore, the return period of 2.33 years is taken as the return period of the mean annual flood. From Equations (17) and Equation (18),

$$x = \bar{x} + \frac{y - y_n}{\sigma_n}\,\sigma \quad (19)$$

and, with Equation (8) or Equation (15),

$$x = \bar{x} - \frac{1n1n\,\frac{T}{T\text{-}1} + y_n}{\sigma_n}\,\sigma \quad (20)$$

According to Ponce [14], the following steps are necessary to apply the Gumbel Method : (i) Assemble the flood series; (ii) Calculate the mean $\bar{x}$ and standard deviation s of the flood series; (iii) Use Table 2 to determine the mean $\bar{y}_n$ and standard deviation σ_n of the Gumbel variate as a function of record length n; (iv) Select several return periods T_j and associated exceedence probabilities P_j; and (v) Calculate the Gumbel variate yj corresponding to the return periods T_j by using Equations (8) or

(15), and calculate the flood discharge $Q_j = x_j$ for each Gumbel variate (and associated return period) using Equation (19). Alternatively, the flood discharges can be calculated directly for each return period by using Equation (20).

Table 2: Mean y_n and Standard Deviation σ_n of Gumbel Variate (y) Vs Record Length (n)

n	$\bar{y}_n$	σ_n	n	$\bar{y}_n$	σ_n	n	$\bar{y}_n$	σ_n
8	0.4843	0.9043	35	0.5403	1.1285	64	0.5533	1.1793
9	0.4902	0.9288	36	0.5410	1.1313	66	0.5538	1.1814
10	0.4952	0.9497	37	0.5418	1.1339	68	0.5543	1.1834
11	0.4996	0.9676	38	0.5424	1.1363	70	0.5548	1.1854
12	0.5035	0.9833	39	0.5430	1.1388	72	0.5552	1.1873
13	0.5070	0.9972	40	0.5436	1.1413	74	0.5557	1.1890
14	0.5100	1.0095	41	0.5442	1.1436	76	0.5561	1.1906
15	0.5128	1.0206	42	0.5448	1.1458	78	0.5565	1.1923
16	0.5157	1.0316	43	0.5453	1.1480	80	0.5569	1.1938
17	0.5181	1.0411	44	0.5458	1.1499	82	0.5572	1.1953
18	0.5202	1.0493	45	0.5463	1.1519	84	0.5576	1.1967
19	0.5220	1.0566	46	0.5468	1.1538	86	0.5580	1.1980
20	0.5236	1.0628	47	0.5473	1.1557	88	0.5583	1.1994
21	0.5252	1.0696	48	0.5477	1.1574	90	0.5586	1.2007
22	0.5268	1.0754	49	0.5481	1.1590	92	0.5589	1.2020
23	0.5283	1.0811	50	0.5485	1.1607	94	0.5592	1.2032
24	0.5296	1.0864	51	0.5489	1.1623	96	0.5595	1.2044
25	0.5309	1.0915	52	0.5493	1.1638	98	0.5598	1.2055
26	0.5320	1.0961	53	0.5497	1.1653	100	0.5600	1.2065
27	0.5332	1.1004	54	0.5501	1.1667	150	0.5646	1.2253
28	0.5343	1.1047	55	0.5504	1.1681	200	0.5672	1.2360
29	0.5353	1.1086	56	0.5508	1.1696	250	0.5688	1.2429
30	0.5362	1.1124	57	0.5511	1.1708	300	0.5699	1.2479
31	0.5371	1.1159	58	0.5515	1.1721	400	0.5714	1.2545
32	0.5380	1.1193	59	0.5518	1.1734	500	0.5724	1.2588
33	0.5388	1.1226	60	0.5521	1.1747	750	0.5738	1.2651
34	0.5396	1.1255	62	0.5527	1.1770	1000	0.5745	1.2685

Ponce [14] proclaimed, the values of Q could be plotted against y or T (or P) on Gumbel probability paper, and a straight line could be drawn through the points. Gumbel probability paper has an arithmetic scale of Gumbel variate y in the abscissas and an arithmetic scale of flood discharge Q in the ordinates. To facilitate the reading of frequencies and probabilities, Equation (6) can be used to superimpose a scale of return period T (or probability P) on the arithmetic scale of Gumbel variate y.

2.4. Several Modifications to Gumbel Distribution

According to Ponce [14], since its inception in the 1940s, several modifications to the Gumbel method have been suggested. Cunnane [3] for example, had studied various plotting position methods using the criteria of unbiasedness and maximum variance. An unbiased plotting method is one that, if used for plotting a large number of equally sized samples, will result in the average of the plotted points for each value of m (i.e. rank) falling on the theoretical distribution line; whereas a minimum variance plotting method is one that minimises the variance of the plotted points about the theoretical line. He had concluded that the Weibull plotting position formula is biased and plots the largest values of a sample at too small a return period. He said, for data distributed according to the Extreme Value Type I distribution (or Gumbel distribution), the Gringorten [17] formula (b = 0.44) is the best.

Chow et. al [7] notified that most plotting position formulas are represented in the following form:

$$P\ (X \geq x_m) = \frac{m - b}{n+1-2b} \qquad (21)$$

where m is the rank of annual extreme series arranged in descending order of magnitude, b is a parameter and n is the number of years of record.

Plotting positions should lie between 0 and 1 (or 0% and 100%). The different plotting position formulas tend to give similar values near the middle of the data, but can vary considerably at the tail ends [9]. Table 3 shows summary of some available plotting position formula [8, 9].

Table 3: Plotting position formula

Method	Probability of Exceedence, P	Average Recurrence Interval, T
California	m/n	n/m
Hazen	(m-0.5)/n	n/(m-0.5)
Weibull	m/(n+1)	(n+1)/m
Chegodayev	(m-0.3)/(n+0.4)	(n+0.4)/(m-0.3)
Blom	(m-3/8)/(n+1/4)	(n+1/4)/(m-3/8)
Gringorten	(m-0.44)/(n+0.12)	(n+0.12)/(m-0.44)

Another modification of Gumbel Method according to Ponce [14] is by Lettenmaier and Burges [18]. They have suggested that better flood estimates are obtained by using the limiting values of mean and standard deviation of the Gumbel variate (i.e. those corresponding to n = ∞) in Equation (2.18), instead of basing these values on the record length. In this case, $\bar{y}_n = 0.5772$, and $\sigma_n = \pi / \sqrt{6} = 1.2825$. Therefore, Equation (19) reduces to:

$$x = \bar{x} + (0.78y - 0.45)\ \sigma \qquad (21)$$

and Equation (20) reduces to:

$$x = x - (0.781n\ 1n\ \frac{T}{T\text{-}1} + 0.45)\ \sigma \qquad (22)$$

Lettenmaier and Burges [18] have also suggested that a biased variance estimate, using n as the divisor in the second moment about the mean (i.e. the variance, σ^2), yields better estimates of extreme events that the usual unbiased estimate, that is, the divisor n - 1. The common variance formula is as shown below:

$$\sigma^2 = \frac{1}{n - 1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \qquad (23)$$

2.5. Various Approaches to Fit Probability Distributions

Probability distributions are defined by their parameters. Therefore it is necessary to understand the concepts underlying parameter estimation for known theoretical frequency distributions to further understand the theoretical probability distribution method. The graphical method, regression analysis (or least-squares method), maximum likelihood method and method of moments (including probability-weighted moments) are the approaches most often used for fitting probability distribution (or to estimate parameters of frequency distributions) [19, 9].

According to Wanielista and Yousef [19], the advantage of the graphical method is its simplicity with visual appeal whereas its disadvantage is that the method is highly subjective and is usually not reproducible. Gordon et al. [9] and Chow et al. [7] agree that the cumulative probability of a theoretical distribution may be represented graphically on probability paper designed for the distribution (e.g. Gumbel probability paper). Haan [10] gives information on constructing probability paper.

The least-squares method uses mathematical formulas to determine the parameters of an empirical distribution, such as the slope and intercept of the distribution. The results are reproducible among users. A best fit is achieved when the sum of squares of all deviations between the observed point and some theoretical function is minimised. The function is calculated for each point, and then the difference between the observed and calculated is squared such that the sum is minimised. This method has gain popularity and is especially useful if the theoretical function can be made linear. Such is the case for the Weibull distribution. [19]

For large sample sizes, method of maximum likelihood is superior to others since the resulting estimators of population parameters are considered to be more efficient and accurate [23]. Devore [23] says, in this method a likelihood function is derived which indicates how likely the observed sample is, assuming that it is from a certain distribution with a range of possible parameter values. He then concluded that by maximising this likelihood function yields parameter values which agree most closely with the observed data. Gordon et. al. [9] commented, solving the maximum likelihood equations to obtain parameter value normally requires an iterative procedure and can need a substantial amount of computer time. They said, since an efficient estimate will not necessarily exist, a solution may or may not be found.

Wanielista and Yousef [19] say, the method of moment is similar to the concept of moments and used in basic physics. They explain, the mean value of a distribution is the first moment about the origin, the variance is the second moment about the mean, the third moment measures the skewness of a distribution and the fourth, is called kurtosis of a distribution, which is a measure of the peakedness. In most frequency analyses using the method of moments, the relation between flood magnitude and probability can be reduced to a simple equation given by Chow [22] as:

$$x_T = \bar{x} + K_T^{\sigma} \quad (24)$$

where x_T is the magnitude of an event with an average recurrence interval T years, x and σ are the mean and standard deviation of sample values, respectively, and K_T is a frequency factor. Method of moments provides an exact theoretical fitting, but the accuracy is substantially affected by errors in the tail of the distribution (i.e. events of long return period) [14]. The disadvantage of the method is the uncertainty regarding the adequacy of the chosen probability distribution [14].

In accordance with Hosking and Wallis [15], it is hard to perform estimation by conventional methods such as maximum likelihood or the method of moments. Hosking [24] had defines a type of probability weighted moments (PWMs) called L-moments, in which the moments are linear functions of the data values, hence the "L". In comparison to conventional moments in which the data values are squared, cubed, etc., with L-moments it is the probabilities which are manipulated. Gordon et al. [9] say, this gives less weight to the very high or very low data values. Hosking and Wallis [15] said, Cunnane [25] had reviewed twelve different methods of regional frequency analysis and rated the regional PWM algorithm as the best. L-moments could be interpreted as measures of the location, scale, and shape of probability distributions and formed the basis for a comprehensive theory of the description, identification, and estimation of distributions [15]. With reference to Hosking [24] and Gordon et. al. [9], the

Table 4: The 19 Selected River Stations

Index	Station No.	Station Name	Latitude (D,M,S)	Longitude (D,M,S)	Elevation (m)
1	1301426	Boring	001 23 21	110 06 39	0
2	1301427	Buan Bidi	001 23 54	110 06 46	67
3	1302428	Kpg Git	001 21 20	110 15 50	1
4	1204441	Kpg Ma'ang	001 15 54	110 24 33	-
5	1304439	Batu Gong	001 20 46	110 26 23	4
6	1104438	Krusen	001 04 11	110 29 52	3
7	1005447	Meringgu	001 03 00	110 33 10	-
8	1114422	Entulang	001 09 00	111 25 32	-
9	1210401	Tuba	001 17 50	110 04 50	-
10	1018401	Lubuk Antu	001 02 35	111 49 35	21
11	1415401	Nanga Lubau	001 29 50	111 35 20	-
12	1813401	Sebatan	001 48 15	111 20 00	1
13	1932408	Telok Buing	001 59 50	113 13 20	-
14	2130405	Nanga Benin	002 09 55	113 04 10	0
15	2421401	Stapang	002 04 00	112 08 05	0
16	3152408	Lio Matu	003 10 10	115 13 20	-
17	4448420	Nanga Insungai	004 24 00	114 53 30	-
18	1108401	Sabal Kruin	001 08 35	110 53 35	-
19	3946411	Long Terawan	003 59 35	114 37 50	-

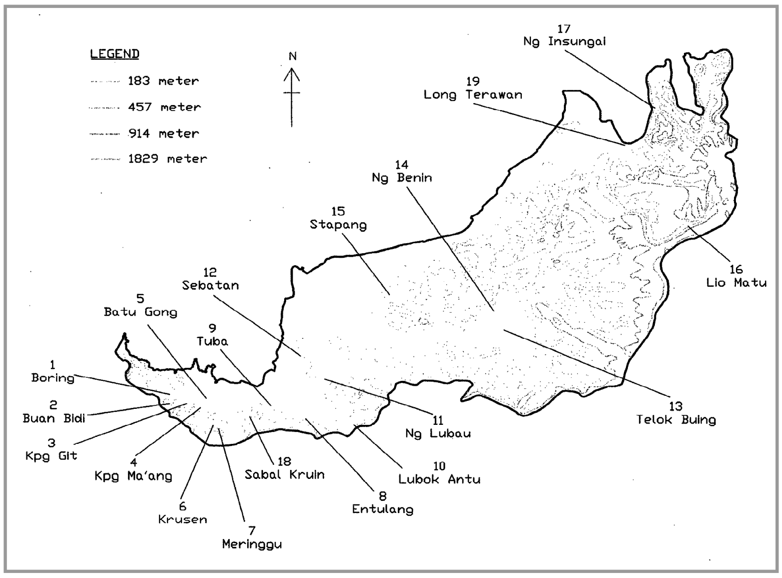

Figure 1: Location of the River Gauges Selected in the Study

Table 5: Example of Frequency Analysis of Individual Station

Station No : 1301426
Station Name : **Boring (Sg.Pedi)**
River : Sg. Pedi
Basin : Sungai Sarawak
Zero of Gauge : 9.98 m M.S.L.
Type of Gauge : Stick gauge
B.M. Value : 19.92 m M.S.L.
Rating Curve Formula : Q = 7.44 (H - 0.92) ^ 1.81
Effective Range of Rating Curve Formula : 1.29 - 3.69 m
Catchment's Area : 124.5 sq.km.

Year	Date	Max. WL (Reading above zero of gauge in metre), H	Records from incomplete years is indicated with #	Q	i	Sorted Qi	MAF =	291.86	T (Weibull)	QT	QT/MAF	y
1970	5-Nov	7.31		213.56	1	487.34			29.00	487.34	1.670	3.350
1971	9-Jan	9.75		383.50	2	432.79			14.50	432.79	1.483	2.639
1972	23-Jan	9.45		360.24	3	408.21			9.67	408.21	1.399	2.215
1973	28-Dec	9.20		341.36	4	383.50			7.25	383.50	1.314	1.908
1974	1-Mar	8.93		321.48	5	373.34			5.80	373.34	1.279	1.665
1975	24-Dec	10.36		432.79	6	371.79			4.83	371.79	1.274	1.462
1976	12-Jan	10.06		408.21	7	369.46			4.14	369.46	1.266	1.286
1977	6-Feb	9.60		371.79	8	360.24			3.63	360.24	1.234	1.131
1978	24-Jan	8.96		323.66	9	353.39			3.22	353.39	1.211	0.990
1979	27-Dec	8.05		260.41	10	341.36			2.90	341.36	1.170	0.861
1980	22-Feb	7.92		251.88	11	327.31			2.64	327.31	1.121	0.740
1981	7-Feb	9.57		369.46	12	326.58			2.42	326.58	1.119	0.627
1982	2-Mar	9.01		327.31	13	323.66			2.23	323.66	1.109	0.520
1983	25-Jan	11.00		487.34	14	321.48			2.07	321.48	1.101	0.417
1984	6-Mar	9.00		326.58	15	271.76			1.93	271.76	0.931	0.317
1985	4-Mar	8.02		258.43	16	260.41			1.81	260.41	0.892	0.220
1986	7-Jan	9.62		373.34	17	258.43			1.71	258.43	0.885	0.125
1987	24-Dec	7.67		235.84	18	255.81			1.61	255.81	0.876	0.031
1988	6-Dec	8.22		271.76	19	251.88			1.53	251.88	0.863	-0.063
1989	14-Dec	9.36		353.39	20	235.84			1.45	235.84	0.808	-0.157
1990	11-Feb	6.67		176.43	21	225.82			1.38	225.82	0.774	-0.253
1991	29-Jan	6.32		157.47	22	213.56			1.32	213.56	0.732	-0.352
1992	19-Jan	7.98		255.81	23	203.39			1.26	203.39	0.697	-0.455
1993	16-Mar	6.47		165.48	24	183.72			1.21	183.72	0.629	-0.564
1994	24-Jan	7.14		203.39	25	176.43			1.16	176.43	0.605	-0.684
1995	25-Dec	5.81		131.59	26	165.48			1.12	165.48	0.567	-0.819
1996	7-Feb	7.51		225.82	27	157.47			1.07	157.47	0.540	-0.984
1997	20-Feb	6.80	#	183.72	28	131.59			1.04	131.59	0.451	-1.214
	Station closed on May 1997					8172.04						

advantage of L-moments is that they are less sensitive to sampling variability and less subject to bias. They are robust in the presence of outliers, meaning that they give consistent results even if the extreme values contain measurement errors. For small samples they produce parameter estimates which are sometimes more accurate than even maximum likelihood estimates.

The four fitting methods can be rated in ascending order of effectiveness: graphical, least square, moments, and maximum likelihood. The latter, however, is somewhat more difficult to apply [20, 21]. Ponce [14] said, in practice, the method of moments is the most commonly used curve-fitting method.

3. THE RESEARCH METHODOLOGY

Data for analysis are extracted from Drainage and Irrigation Department (DID) of Sarawak. A total of nineteen (19) sample stations had been selected for the

Table 6: Results of Gumbel Distribution for Station Boring by Weibull Formula, Gringorten Formula and L-Moments Method

Weibull (1939)			Gringorten (1963)			L-Moments (Hosking and Wallis 1997)				
								Euler	0.577	
								Ln 2	0.693	
								Lamda 1	291.859	
								Lamda 2	-33.345	
								Alpha	48.106	
								Epsilon	264.091	
Q_T/MAF	T_W	y_W	Q_T/MAF	T_G	y_G	Q/AMAF	Discharge	Y_{LM}	F(x)	T_{LM}
1.670	29.000	3.350	1.670	50.21	3.906	1.670	487.340	4.641	0.990	104.124
1.483	14.500	2.639	1.483	18.03	2.863	1.483	432.790	3.507	0.970	33.844
1.399	9.667	2.215	1.399	10.98	2.349	1.399	408.210	2.996	0.951	20.507
1.314	7.250	1.908	1.314	7.90	2.000	1.314	383.500	2.482	0.920	12.475
1.279	5.800	1.665	1.279	6.17	1.732	1.279	373.340	2.271	0.902	10.198
1.274	4.833	1.462	1.274	5.06	1.513	1.274	371.790	2.239	0.899	9.891
1.266	4.143	1.286	1.266	4.29	1.326	1.266	369.460	2.190	0.894	9.448
1.234	3.625	1.131	1.234	3.72	1.161	1.234	360.240	1.999	0.873	7.891
1.211	3.222	0.990	1.211	3.29	1.013	1.211	353.390	1.856	0.855	6.913
1.170	2.900	0.861	1.170	2.94	0.878	1.170	341.360	1.606	0.818	5.501
1.121	2.636	0.740	1.121	2.66	0.753	1.121	327.310	1.314	0.764	4.244
1.119	2.417	0.627	1.119	2.43	0.636	1.119	326.580	1.299	0.761	4.188
1.109	2.231	0.520	1.109	2.24	0.525	1.109	323.660	1.238	0.748	3.974
1.101	2.071	0.417	1.101	2.07	0.418	1.101	321.480	1.193	0.738	3.822
0.931	1.933	0.317	0.931	1.93	0.316	0.931	271.760	0.159	0.426	1.743
0.892	1.813	0.220	0.892	1.81	0.216	0.892	260.410	-0.077	0.340	1.515
0.885	1.706	0.125	0.885	1.70	0.118	0.885	258.430	-0.118	0.325	1.481
0.876	1.611	0.031	0.876	1.60	0.021	0.876	255.810	-0.172	0.305	1.439
0.863	1.526	-0.063	0.863	1.52	-0.076	0.863	251.880	-0.254	0.276	1.380
0.808	1.450	-0.157	0.808	1.44	-0.173	0.808	235.840	-0.587	0.165	1.198
0.774	1.381	-0.253	0.774	1.37	-0.273	0.774	225.820	-0.796	0.109	1.122
0.732	1.318	-0.352	0.732	1.30	-0.375	0.732	213.560	-1.050	0.057	1.061
0.697	1.261	-0.455	0.697	1.25	-0.483	0.697	203.390	-1.262	0.029	1.030
0.629	1.208	-0.564	0.629	1.19	-0.598	0.629	183.720	-1.671	0.005	1.005
0.605	1.160	-0.684	0.605	1.14	-0.726	0.605	176.430	-1.822	0.002	1.002
0.567	1.115	-0.819	0.567	1.10	-0.874	0.567	165.480	-2.050	0.000	1.000
0.540	1.074	-0.984	0.540	1.06	-1.062	0.540	157.470	-2.216	0.000	1.000
0.451	1.036	-1.214	0.451	1.02	-1.365	0.451	131.590	-2.754	0.000	1.000

analysis. The selection is based on the criteria stated in HP4. Details of the selected data and the approximate location of the 19 selected stations are as shown in Table 1 and Figure 1 [extracted from Sarawak Hydrological Year Book Series (SHYB) [26]].

Raw data from DID come in water level form. These values are then converted into discharge, Q form by using the discharge rating curve establish by DID. After the conversion, the annual extreme series are arranged in descending order of magnitude. Then the arithmetic mean of the annual flood series is calculated. After that, the plotting position of each sample is determined. In this study, three plotting position formula are applied onto the samples. The three plotting position formula are Weibull formula [see Table 3 and Equations (8) or (15)], Gringorten formula [again see Table 3 and Equation (8) or (15)] and L-Moments method [see Equation (12), (13) and (14)]. As to construct the Gumbel distribution by L-Moment Method with QT/MAF as the y-axis and Gumbel reduce varite (y) as the x-axis, a calculation of L-moments parameters in a Fortran Programming form is needed. (Refer to Hosking and Wallis [15] for the details). The

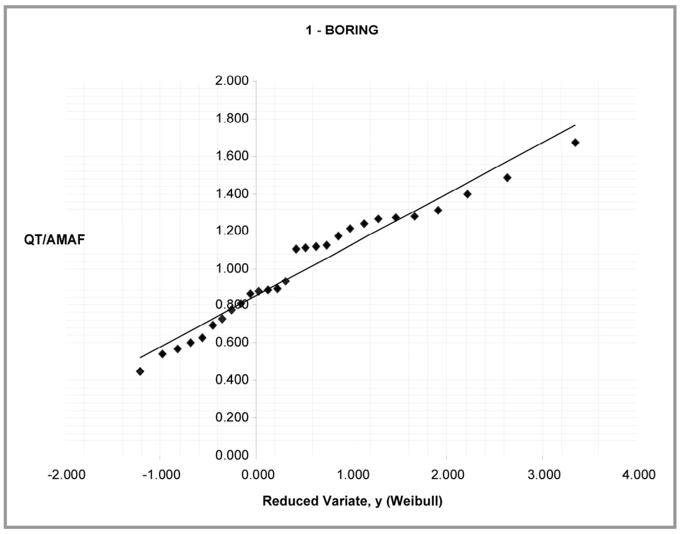

Figure 2: Gumbel Distribution Using Weibull Formula for Station Boring

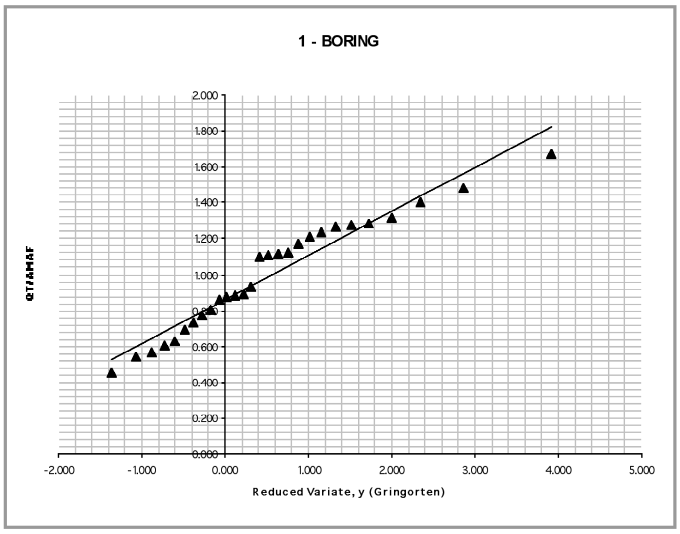

Figure 3: Gumbel Distribution Using Gringorten Formula for Station Boring

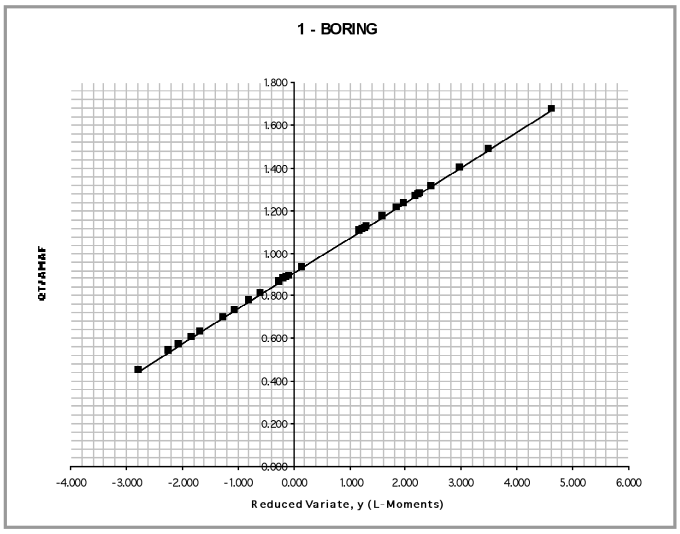

Figure 4: Gumbel Distribution Using L-Moments Method for Station Boring

parameters and results from the programming are then used as the inputs for the calculations of Gumbel reduced variate, y.

The values of annual peak discharge over the arithmetic mean of the annual flood series, Q/MAF or QT/AMAF are

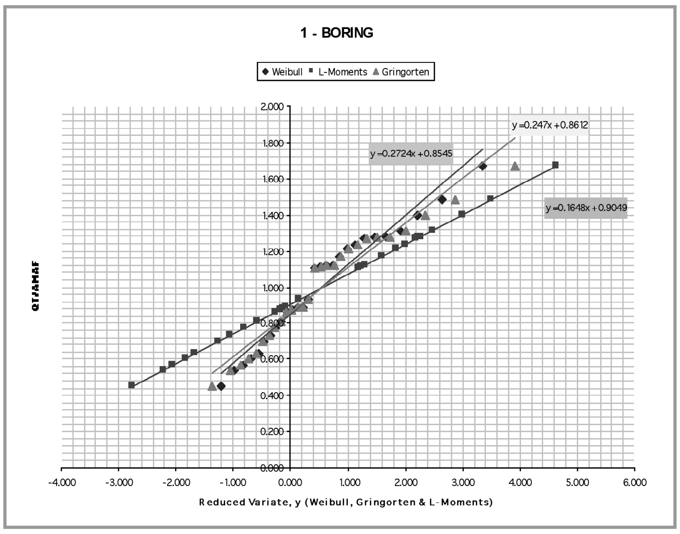

Figure 5: Gumbel Distribution Using Weibull, Gringorten and L-Moments Method

then plotted against the reduce variate, y. Finally, dimensionless flood-frequency curve of each individual station was constructed. Then, comparison of Gumbel distribution by the three plotting positions will be made. Comparison of the overall 19 station using each method will not be discussed in this paper.

4. RESULTS AND ANALYSIS

This section presented the results and analysis of Gumbel distribution for one of the individual station (i.e. Station Boring) using Weibull formula, Gringorten Formula and L-Moments Method. The calculation of flood frequency curve for Stations Boring using Gumbel distribution (Weibull Formula) is as tabulated in Table 5. Summary of Gumbel distribution from the three methods for Station Boring is as shown in Table 6. The results are utilised to produce the probability plot and flood-frequency curves for Stations Boring. Figure 2 illustrates the probability plot and flood frequency curve of Gumbel distribution using Weibull Formula for Station Boring. Illustration of the probability plot and flood-frequency curve of Gumbel distribution of the station using Gringorten formula is as shown in Figure 3. The flood-frequency curve of the station by L-moments method is shown in Figure 4. The discharge and reduce variate (y) data shown in Table 6 when plotted together in one graph could contribute on the comparison of the three plotting position method. (See Figure 5)

In this paper, as an example, only the results and analysis of Station Boring are presented. The results and analysis of the overall 19 sample stations superimposed together in one graph for each of the method (i.e. Gumbel Distribution by Weibull Formula, Gumbel Distribution by Gringorten Formula and Gumbel Distribution by L-Moments Method), will be discussed in another paper.

5. DISCUSSIONS

From the results and analysis of Gumbel distribution using the three plotting position formula/method in Section 4, these few trends had been identified: (i) Gumbel distribution with Weibull Formula is always the steepest followed by Gumbel distribution with Gringorten Formula and then Gumbel distribution by L-moment method, and (ii) Flood Frequency Curve of Gumbel distribution by L-moment always fit nicely to probability plot compared to the other two cases.

5.1. Steepness of the Flood Frequency Curves

According to Arnell [5] the steeper the slope of the flow duration curve the greater the variability in flow. Referring to Figure 5, the following equations had been produced in the plot of Gumbel distribution with Weibull, Gringorten and L-moments Formula:

Weibull	:	$y = 0.2724x + 0.8545$
Gringorten	:	$y = 0.2470x + 0.8612$
L-moments	:	$y = 0.1648x + 0.9049$

It shows that Gumbel distribution with Weibull Formula is always the steepest followed by Gumbel distribution with Gringorten Formula and then Gumbel distribution by L-moments method. If we relate these results with findings from Arnell [5] finding, we could presume that the flow variability for Station Boring using Gumbel distribution by Weibull and Gringorten formula is greater than the flow variability of the Boring Station using Gumbel distribution by L-Moments formula. According to Hosking [24], L-Moments are less sensitive to variability.

Cunnane [3] discovered that Weibull plotting formula was biased and plots the largest values of a sample at too small a return period. For data distributed according to the Extreme Value Type I distribution (or Gumbel distribution), he then recommended the Gringorten formula ($b = 0.44$) as the most appropriate. The example shown in Figure 5 was in agreement with Cunnane's findings where he had concluded that for Gumbel distribution, Gringorten plotting position formula is better than Weibull formula.

5.2. Probability Plots by L-Moments Method Fit Nicely to The Flood Frequency Curve

Hosking [24] had highlighted that the advantage of L-moments is that they are less sensitive to sampling variability, less subjective to bias and they are robust in the presence of outliers (i.e. they give consistent results even if the extreme values contain measurement errors). And, according to Gordon et. al. [9], with L-moments, it is the probabilities which were manipulated. As a result, it gives less weight to the very high or very low data values. Most probably, this is the reason why the probability plots by L-moments method in Figure 4 and Figure 5 fit nicely to the flood frequency curves.

6. CONCLUSIONS

In this study, the magnitude and frequency of floods for Sarawak is analysed using Gumbel distribution with three plotting position formula, namely Weibull, Gringorten and L-Moments. Amongst the three methods, L-moments always give the least ratio of peak discharge of T year's recurrence interval / mean annual flood (QT/MAF). Even though L-Moments always give the least ratio, at some stations, it gives unreasonable return period and reduced variate range. The literature [24,9] did say that the method is good for small samples. Therefore, the appropriateness of L-moments with Gumbel distribution had some limitations. If compared between Weibull and Gringorten formula, Gumbel distribution by Gringorten formula is better than Gumbel distribution by Weibull formula because the former always gives the least ratio. The literature [3] says, Gringorten formula is more suitable to be used with Gumbel distribution. Therefore, it could be concluded that for some stations, L-Moments method is the best, but since L-Moments method had some limitations, Gringorten formula is still the best plotting position method to be used with Gumbel distribution. ■

REFERENCES

[1] Ong C.Y. (1987). *HP 4 - Magnitude and Frequency of Floods in Peninsular Malaysia*, Drainage and Irrigation Department, Ministry of Agriculture, Kuala Lumpur.

[2] Natural Environmental Research Council (NERC). (1975). *Flood Studies Report, Vol. 1, Hydrological Studies*, Natural Environment Research Council, London (available from Institute of Hydrology, Wallingford, Oxon, England) in Ong, C.Y. (1987). *HP4 - Magnitude and Frequency of Floods in Peninsular Malaysia*, Drainage and Irrigation Department, Ministry of Agriculture, Kuala Lumpur.

[3] Cunnane C. (1978). "Unbiased Plotting Positions – A Review." Journal of Hydrology.,Vol. 37, pp. 205-222 in Chow V.T., Maidment D.R. and Mays L.W. (1988). *Applied Hydrology*, McGraw-Hill, New York.

[4] Lim Y. H. and Lye L. M. (2003). "Regional Flood Estimation for Ungauged Basins in Sarawak, Malaysia." Hydrological Sciences Journal, 48(1) February 2003, 79-94.

[5] Arnell N.W. (2002). *Hydrology and Global Environmental Change*, Prentice-Hall, Harlow.

[6] Mays L.W. (2004). *Urban Storm water Management Tools*, McGraw-Hill, New York.

[7] Chow V.T., Maidment D.R. and Mays L.W. (1988). *Applied Hydrology*, McGraw-Hill, New York.

[8] Subramanya K. (2002). *Engineering Hydrology*, Tata McGraw-Hill, New Delhi.

[9] Gordon N.D., McMahon T.A. and Finlayson B.L. (1993). *Stream Hydrology : An Introduction For Ecologists,* John Wiley & Sons, New York.

[10] Haan C. (1977). *Statistical Methods in Hydrology*, The Iowa State University Press, Ames, Iowa, U.S.A. (1977). in Gordon N.D., McMahon T.A. and Finlayson B.L. (1993). *Stream Hydrology : An Introduction For Ecologists,* John Wiley & Sons, New York.

[11] DID and NAHRIM. (2003). "Frequency Analysis of Rainstorm and Flood (Rainstorm Analysis)". In a Seminar - "Frequency Analysis of Flood and Low Flows and Introduction to MASMA Concept" at Kuching, Sarawak, 15-17 July 2003.

[12] Jenkinson A.F. (1955). "The Frequency Distribution of the Annual Maximum (or Minimum) Values of Meteorological Elements." Quart. Jour. Roy. Met. Soc., Vol. 81, pp. 158-171 in Chow V.T., Maidment D.R. and Mays L.W. (1988). *Applied Hydrology*, McGraw-Hill, New York.

[13] Kinnison R. R. (1985). *Applied Extreme Value Statistics,* MacMillan Publishing Company, New York.

[14] Ponce V. M. (1989). *Engineering Hydrology : Principles and Practices*, Prentice-Hall Inc., New Jersey.

[15] Hosking J.R.M and Wallis J. R. (1997). *Regional Frequency Analysis,* Cambridge University Press, United Kingdom.

[16] Gumbel, E.J. (1958). *Statistics of Extremes*, Irvington, New York: Columbia University Press in Ponce V. M. (1989). *Engineering Hydrology : Principles and Practices*, Prentice-Hall Inc., New Jersey.

[17] Gringorten I.I. (1963). "A Plotting Rule for Extreme Probability Paper." Journal of Geophysics Resources, Vol. 68, No. 3, pp. 813-814 in Chow V.T., Maidment D.R. and Mays L.W. (1988). *Applied Hydrology,* McGraw-Hill, New York.

[18] Lettenmaier D.P., and Burges S.J. (1982). "Gumbel's Extreme Value Distribution: A New Look," Journal of the Hydraulics Division, ASCE, Vol 108, No HY4, April, p.p 503-514 in Ponce V. M. (1989). *Engineering Hydrology : Principles and Practices*, Prentice-Hall Inc., New Jersey.

[19] Wanielista M.P. and Yousef A.Y. (1993). *Storm water Management*, John Wiley & Sons, New York.

[20] Chow V.T., (1964). *Handbook of Applied Hydrology,* McGraw-Hill, New York in Ponce V. M. (1989). *Engineering Hydrology : Principles and Practices*, Prentice-Hall Inc., New Jersey.

[21] Kite G.W. (1977). *Frequency and Risk Analyses in Hydrology*, Fort Collins, Colorado: Water Resources Publications in Ponce, V. M. (1989). *Engineering Hydrology : Principles and Practices,* Prentice-Hall Inc., New Jersey.

[22] Chow V.T., (1964b). 'Statistical and probability analysis of hydrologic data. Part I. Frequency analysis', in Handbook of applied Hydrology (Ed. V.T. Chow), pp.8-1 to 8-42, McGraw-Hill, New York in Gordon N.D., McMahon T.A. and Finlayson B.L. (1993). *Stream Hydrology : An Introduction For Ecologists,* John Wiley & Sons, New York.

[23] Devore J.L. (1982). *Probability and Statistics for Engineering and the Sciences,* Brooks/Cole, Monterey, California in Gordon N.D., McMahon T.A. and Finlayson B.L. (1993). *Stream Hydrology : An Introduction For Ecologists,* John Wiley & Sons, New York.

[24] Hosking J.R.M. (1989). "The Theory of Probability Weighted Moments", Res. Rep. RC 12210 (#54860), IBM Research Division, T.J. Watson Research Center, Yorktown Heights, NY 10598 in Gordon N.D., McMahon T.A. and Finlayson B.L. (1993). *Stream Hydrology: An Introduction For Ecologists,* John Wiley & Sons, New York.

[25] Cunnane, C. (1978). "Unbiased plotting position – a review." Journal of Hydrology, 37, 205-22 in Hosking J.R.M and Wallis J. R. (1997). R*egional Frequency Analysis*, Cambridge University Press, United Kingdom.

[26] Department of Irrigation & Drainage, Sarawak. *Sarawak Hydrological Year Book,*Vol. 1-25, Kuching, Sarawak.