



A COMPARISON OF PLOTTING FORMULAS FOR THE PEARSON TYPE III DISTRIBUTION

ANI SHABRI*

Abstract. Unbiased plotting position formulas are discussed to fit the Pearson Type 3 distribution (PIII). The best quantile estimate made from the plotting position should be unbiased and should have the smallest root means square error among all such estimates. Probability plot correlation coefficient (PPCC) is used to evaluate goodness of fit to test the PIII distribution hypothesis. Results obtained using the annual maximum flow data from Peninsular of Malaysia based on PPCC show the plotting position formulas consistently produced linear probability plots with correlation coefficient near to one. Based on root mean square error (RMSE) and root mean absolute error, the Weibull formula performs better than the other formulas.

Keywords: Plotting Position, quantile, unbiased, root means square error

Abstrak. Formula kedudukan memplot tanpa bias dibincangkan untuk di padankan dengan taburan Pearson 3 (PIII). Penganggar kuantil kedudukan memplot terbaik seharusnya tanpa bias dan mempunyai min punca ralat terkecil antara penganggar-penganggar yang lain. Pekali korelasi kedudukan memplot digunakan sebagai ujian pemadanan cocokan untuk menguji hipotesis taburan PIII. Hasil keputusan menggunakan data aliran maksimum daripada Semenanjung Malaysia berdasarkan ujian PPCC menunjukkan rumus kedudukan memplot menghasilkan plot kebarangkalian yang linear dengan pekali korelasi menghampiri satu. Berdasarkan punca min ralat kuasa dua dan punca min ralat mutlak, formula Weibull adalah terbaik antara formula-formula yang lain.

Kata Kunci Kedudukan Memplot, kuantil, tanpa bias, punca min ralat kuasa dua

1.0 INTRODUCTION

Probability plotting positions are used for the graphical display of annual maximum flood series and serve as estimates of the probability of exceedance of those series. Probability plots allow a visual examination of the adequacy of the fit provided by alternative parametric flood frequency models. They also provide a non-parametric means of forming an estimate of the data's probability distribution by drawing a line by hand and or automated means through the plotted points. Because of these attractive characteristics, the graphical approach has been favoured by many hydrologists and engineers. It has been widely used both in hydraulic engineering and water resources research [1, 3, 4 and 5].

* Jabatan Matematik, Fakulti Sains UTM Skudai, Johor, Malaysia, e-mail: anis@ittho.my, shabri99@hotmail.com.



Probability plotting positions have been discussed by hydrologists and statisticians for many years. To date, more than ten plotting position formule have appeared in the literature. Cunnane [2] and Stedinger *et. al* [7] published a very comprehensive review of the existing plotting formula. They postulated that a plotting formula should be unbiased and should have the smallest mean square error among all estimates.

Many distributions and various ways of fitting them are suitable. The selection distribution for any given flood records from among the alternative distributions is still a subject of continuing investigations. In hydrology many distributions for flood frequency analysis most often used, namely Extreme Value Type I (EV1), General extreme value (GEV), Pearson Type III (PIII), Log-Pearson Type III (LPIII), Log Normal (LNIII), General Pareto (GP), Wakeby and Weibull. Similarly, there are many plotting formula available, several of which are summarized in Table 1.

The choice of plotting position formula for fit to the distributions has been discussed many times in hydrology and statistical literature. Different plotting positions attempt to use to achieve almost quantile-unbiasedness for different distributions. In this paper, the focus is to find the best plotting position formula to fit the PIII distribution. In order to determine which plotting position formula is the most suitable for PIII distribution, the probability plot correlation coefficient test and RMSE and RMAE were used. The parameters for each distribution was estimated using moment method.

2.0 PEARSON TYPE III DISTRIBUTION

The Pearson Type III (PIII) distribution is used widely by hydrologists for modeling flood flow frequencies [5] and [8]. The Pearson type III probability density function may be expressed as

$$f(x) = \frac{|\beta|}{\Gamma(\alpha)} (\beta(x - \xi))^{\alpha-1} e^{-\beta(x-\xi)} \quad (1)$$

where α , β and ξ are parameters. The parameters α , β and ξ are related to the first three moments of the random variable X as follows:

$$\mu = \xi + \frac{\alpha}{\beta} \quad (2)$$

$$\sigma^2 = \frac{\alpha}{\beta^2} \quad (3)$$

$$\gamma = \frac{2\beta}{|\beta|\alpha^{1/2}} \quad (4)$$

3.0 THE INVERSE OF A PEARSON TYPE III DISTRIBUTION

The cumulative distribution function of PIII random variable is defined as



$$\begin{aligned} F(x) &= \int_{\xi}^x f(x) dx \quad \gamma > 0 \\ F(x) &= \int_{-\infty}^x f(x) dx \quad \gamma > 0 \end{aligned} \quad (5)$$

which given the complex form of $f(x)$ in (1), is not easily inverted. Many investigations have developed approximation inversion formula. Stedinger [7] found the good approximation for inverse of standardized PIII random variable is

$$x_{p_i} = \mu + \sigma K_{p_i} \quad (6)$$

where K_{p_i} is referred to as frequency factor for the PIII distribution and can be written as

$$K_{p_i} = \frac{2}{\gamma} \left[1 + \frac{\gamma z_{p_i}}{6} - \frac{\gamma^2}{36} \right]^3 - \frac{2}{\gamma} \quad (7)$$

where μ , σ and γ are mean, standard deviation and skew coefficient respectively, while z_{p_i} is the p th quantile of the zero-mean and unit-variance standard normal distributions.

4.0 PLOTTING POSITION

Many investigators have advocated the use of quantile unbiased plotting positions when constructing probability plots. A quantile-unbiased plotting position is defined as [6]

$$p_i = F[E(X_i)]$$

where

$$E[X_i] = F^{-1}(p_i) \quad \text{for } i = 1, 2, \dots, n \quad (8)$$

In situations where no historical floods are considered, most of them may be expressed as a special case of general form

$$p_i = \frac{i - a}{n + 1 - 2a} \quad (9)$$

where p_i is the plotting probability of the i th order statistic, n is the sample size and a is the plotting position parameter yielding approximately unbiased plotting positions for different distributions[1, 8]. For example, $a = 0$ for all distributions (Weibull formula), 0.44 for extreme value and exponential distribution (Gingorten formula), 0.5



for extreme value distribution (Hazen formula) and 3/8 for normal distribution (Blom formula)[7]. The approximation unbiased plotting position for PIII developed by Nguyen *et. al* takes the form [8]

$$p_i = \frac{i - 0.42}{n + 0.3\gamma - 0.05} \quad (10)$$

and is suitable for skews in the range $-3 \leq \gamma \leq 3$ and samples in the range $5 \leq n \leq 100$. All of the plotting position formulas in this study are summarized in Table 1.

Table 1 Plotting Position Formulas (Cunnane, [2], Stedinger et al. [7])

Proponent	Formula	a	Parent Distribution
Weibull (1939)	$\frac{i}{n+1}$	0	All distributions
Beard (1943)	$\frac{i - 0.3175}{n + 0.365}$	0.3175	All distributions
APL	$\frac{i - 0.35}{n}$	~0.35	Used with Probability Weighted Moments Method (PWM)
Blom (1958)	$\frac{i - 3/8}{n + 1/4}$	0.375	Normal distributions
Cunnane (1977)	$\frac{i - 0.40}{n + 0.2}$	0.40	GEV and PIII distributions
Griengorten (1963)	$\frac{i - 0.44}{n + 0.12}$	0.44	Exponential, EV1 and GEV distributions
Hazen (1914)	$\frac{i - 0.5}{n}$	0.50	Extreme Value distributions
Nguyen <i>et.al</i> (1989)	$\frac{i - 0.42}{n + 0.3\gamma + 0.05}$		PIII distribution

5.0 PROBABILITY PLOT CORRELATION COEFFICIENT TEST

A probability plot is defined as a graphical representation of the i th order statistic of the sample, x_i as a function of a plotting position. The i th order statistic is obtained by ranking the observed sample from the smallest ($i = 1$) to the largest ($i = n$) value, then x_i equals the i th largest value.

A simple but powerful goodness-of-fit test is the probability plot correlation coefficient (PPCC) test developed by Filliben in 1975, [7, 9]. The test uses the correlation r between the ordered observations and the corresponding fitted quantiles

$x_{p_i} = F^{-1}(x)$, determined by plotting position p_i for each x_i . The PPCC test is a



measure of linearity of a probability plot. If the sample to be tested is actually drawn from the hypothesized distribution, it is expected to be nearly linear and the correlation coefficient will be near to one. If \bar{x} denotes the average value of the observations and \bar{w} denotes the average value of the fitted quantiles, the correlation coefficient sample can then be defined as

$$r = \frac{\sum (x_i - \bar{x})(x_{p_i} - \bar{w})}{\sqrt{\sum (x_i - \bar{x})^2 (x_{p_i} - \bar{w})^2}} \quad (11)$$

The 5% critical values of PPCC test statistic of the PIII distribution can be approximated using

$$r_{0.05} = \exp[3.77 - 0.0290\gamma^2 - 0.000670n]^{(0.105\gamma - 0.748)} \text{ for } |\gamma| \leq 5 \quad (12)$$

as given by Vogel *et. al* [8]. One rejects the hypothesized PIII distribution if the observed value, r , is smaller than the critical value.

6.0 ROOT MEAN SQUARE ERROR AND ROOT MEAN ABSOLUTE ERROR

Root mean square errors (RMSE) and root mean absolute error (RMAE) are used to compare the efficiency of the different plotting positions formulas. The RMSE is calculated by the equation

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - x_{p_i}}{x_i} \right)^2} \quad (13)$$

while RMAE is calculated by the equation

$$RMAE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - x_{p_i}}{x_i} \right|} \quad (14)$$

where x_i and x_{p_i} are observed and quantile values, respectively for a given value of i .

7.0 APPLICATION TO ANNUAL FLOOD DATA

The selected case study involved annual maximum flow in Peninsular of Malaysia. The data was obtained from the Department of Irrigation and Drainage Malaysia. Data from 31 stations were collected for the present study. A list of these stations number, years of record and PPCC test of plotting position formula is presented in Table 2. These data were selected on the basis of length completeness and independence of record. The lengths of record are between 14 to 34 years. The parameters α ,

Table 2 Correlation Coefficient Value (r) of the Plotting Position Formulas and 5% Critical Value for the PIII Distribution

Station Number	Site	n	Plotting Position Formula						5% Critical Value
			Weibull	Beard	APL	Blom	Cunnane	Gringorten	
1732401	1	16	0.981	0.979	0.975	0.978	0.978	0.977	0.976
1737451	2	32	0.983	0.984	0.984	0.984	0.984	0.984	0.983
1836402	3	18	0.985	0.988	0.992	0.989	0.989	0.989	0.986
2130422	4	21	0.982	0.985	0.987	0.986	0.986	0.986	0.986
2235401	5	18	0.970	0.973	0.978	0.974	0.974	0.975	0.970
2237471	6	34	0.872*	0.890*	0.931	0.895*	0.898*	0.902*	0.777*
2527411	7	22	0.965	0.971	0.978	0.972	0.973	0.975	0.967
2928401	8	15	0.934	0.947	0.965	0.949	0.951	0.953	0.956
3030401	9	16	0.987	0.988	0.987	0.988	0.988	0.988	0.988
3224433	10	21	0.983	0.987	0.992	0.988	0.988	0.989	0.990
3329401	11	14	0.980	0.979	0.979	0.978	0.978	0.978	0.977
3519426	12	29	0.992	0.993	0.994	0.993	0.993	0.993	0.993
3629403	13	23	0.955	0.951*	0.944*	0.944*	0.949*	0.948*	0.946*
4019462	14	34	0.963	0.971	0.982	0.973	0.974	0.975	0.977
4023412	15	26	0.982	0.981	0.981	0.981	0.981	0.980	0.981
4121413	16	21	0.986	0.983	0.979	0.982	0.981	0.981	0.980
4131453	17	14	0.920	0.939	0.965	0.943	0.945	0.948	0.953
4218416	18	15	0.975	0.978	0.981	0.979	0.980	0.980	0.981
4219415	19	16	0.984	0.988	0.992	0.988	0.989	0.989	0.990
4223450	20	19	0.970	0.976	0.983	0.977	0.977	0.978	0.979
4232452	21	19	0.987	0.987	0.988	0.987	0.987	0.987	0.987
4732461	22	16	0.982	0.982	0.983	0.982	0.982	0.982	0.982
4832441	23	25	0.916*	0.934	0.956	0.937	0.939	0.942	0.956
5129437	24	18	0.984	0.987	0.990	0.987	0.987	0.988	0.985
5130432	25	32	0.962	0.967	0.970	0.968	0.968	0.969	0.961
5229436	26	15	0.988	0.987	0.985	0.986	0.986	0.985	0.986
5330443	27	23	0.970	0.977	0.985	0.979	0.979	0.980	0.982
5428401	28	18	0.987	0.988	0.989	0.988	0.988	0.988	0.987
5721442	29	37	0.985	0.989	0.992	0.990	0.990	0.991	0.985
5724411	30	20	0.805*	0.834*	0.880*	0.841*	0.844*	0.854*	0.761*
6019411	31	31	0.986	0.987	0.987	0.987	0.987	0.987	0.965

* The hypotheses of plotting position formula is rejected at 5% significant level.

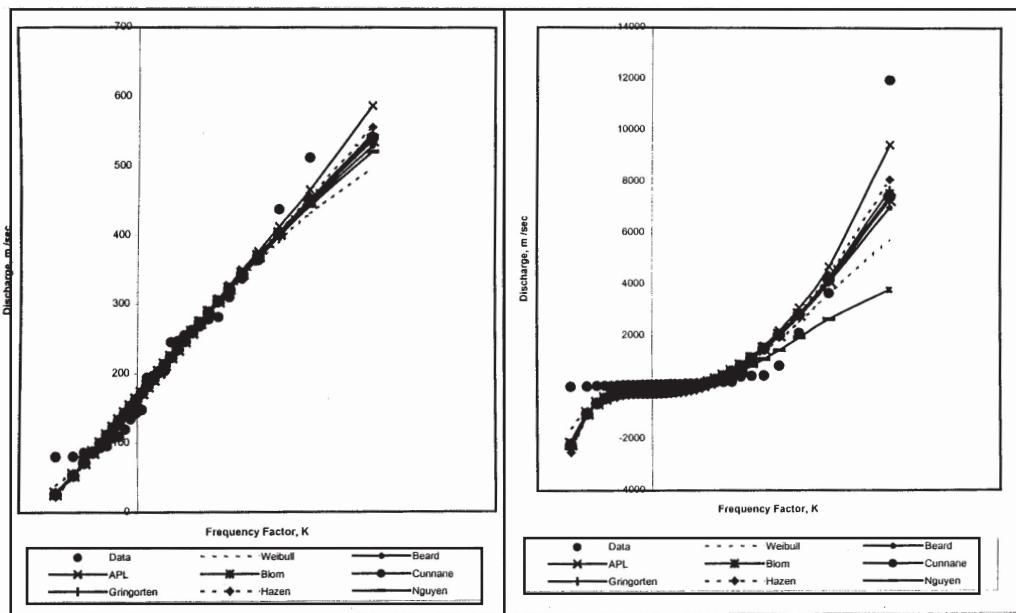
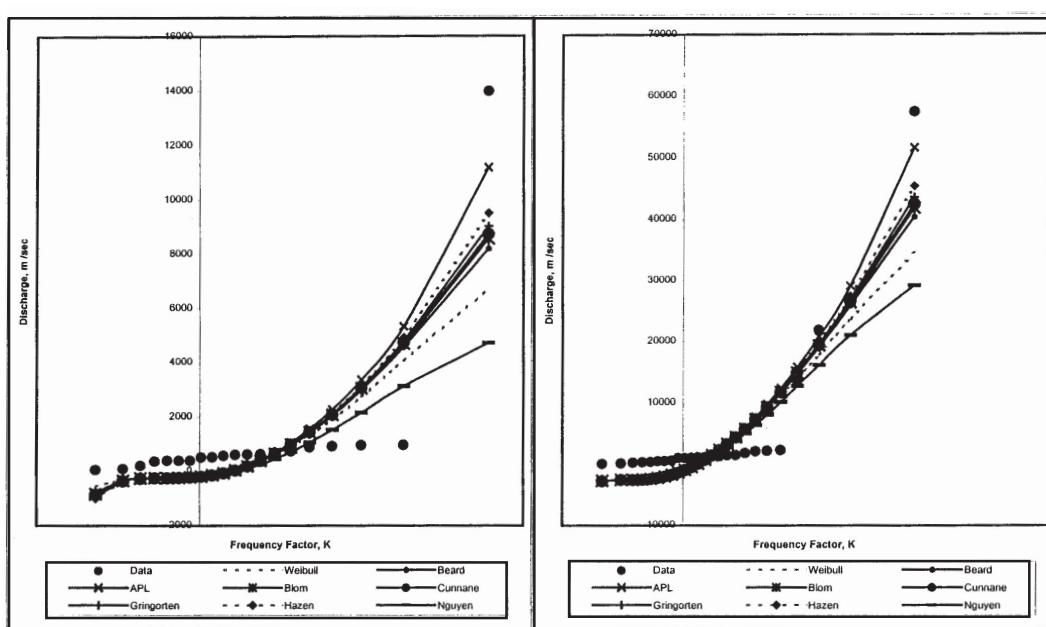
(a) Station number 12, For $r > 0.991$ (b) Station number 6, With $0.776 < r < 0.932$ (c) Station number 23 With $0.898 < r < 0.957$ (d) Station number 30 With $0.760 < r < 0.888$

Figure 1 Comparison of Observed and Computed Frequency Curves For The 4 Stations With Difference r



β and ξ of the Pearson Type 3 distribution were estimated by using the method of moment.

Two criteria were used for comparing the eight plotting positions. The first criterion is defined as the probability plot goodness of fit. Table 2 shows the probability plot correlation coefficient, r , for 8 plotting position formulas and the 5% critical value of the PPCC test statistic using equation (11).

The correlation coefficient values of the plotting position formulas for each stations corresponding with 5% critical values are shown in Figure 1. Table 2 and Figure 1 show that the all plotting position formulas fall in accepted region at 5% critical values at all stations except the APL is rejected at two stations, Nguyen is rejected at four stations and the other formulas are rejected at three stations.

Two sets of observed data were selected for numerical demonstration. Figure 3 and Figure 4 show a demonstration comparison of plotting position formulas for r are accepted for station 12 and rejected for station 30 at 5% critical values. From Figure 3, it can seen that plots based on all of plotting position formulas are closed to data. However Figure 4 shows that the PIII using these plotting position formulas do not show good fit to the data especially at the largest data.

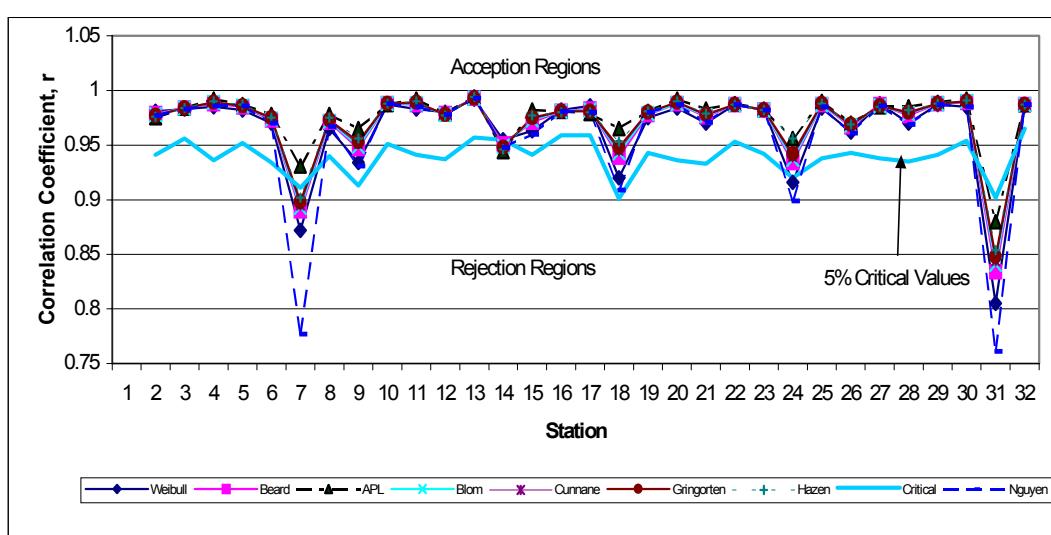


Figure 3 The Probability Plot Correlation Coefficient for the 8 Plotting Position Formulas and 5% Critical Values

The second criterion is the defined as the RMSE and RMAE. Table 3 and 4 list the values of RMSE and RMAE for PIII by using the plotting position formulas.



Table 3 Values of Root Means Square Error

Plotting Position Formula						
Station	Weibull	Beard	APL	Blom	Cunnane	Gringorten
1	0.153	0.175	0.182	0.183	0.187	0.194
2	0.190	0.211	0.214	0.215	0.218	0.221
3	0.100	0.115	0.117	0.120	0.122	0.126
4	0.171	0.126	0.123	0.117	0.113	0.107
5	0.271	0.317	0.324	0.327	0.332	0.340
6	7.011	8.796	8.993	9.247	9.460	9.828
7	7.011	8.796	8.993	9.247	9.460	9.828
8	0.092	0.088	0.089	0.088	0.088	0.087
9	0.169	0.110	0.110	0.103	0.100	0.098
10	0.075	0.079	0.080	0.084*	0.086	0.090
11	0.126	0.131	0.136	0.135	0.137	0.141
12	0.109	0.115	0.117	0.118	0.119	0.122
13	0.358	0.436	0.448	0.455	0.463	0.478
14	0.115	0.113	0.112	0.113	0.113	0.113
15	0.199	0.238	0.244	0.248	0.253	0.261
16	0.005	0.005	0.006	0.005	0.005	0.005
17	0.754	0.778	0.785	0.782	0.784	0.787
18	0.097	0.070	0.070	0.066	0.064	0.061
19	0.005	0.004	0.003	0.004*	0.004	0.003
20	1.106	1.250	1.263	1.280	1.293	1.315
21	0.134	0.125	0.131	0.132	0.136	0.144
22	0.169	0.174	0.180	0.181	0.185	0.192
23	5.071	5.184	5.216	5.218	5.235	5.266
24	0.253	0.321	0.331	0.343	0.353	0.372
25	0.336	0.356	0.359	0.359	0.361	0.363
26	0.133	0.183	0.193	0.205	0.216	0.234
27	0.120	0.119	0.120	0.119	0.120	0.120
28	0.189	0.591	0.644	0.740	0.809	0.923
29	0.252	0.274	0.276	0.278	0.280	0.284
30	2.466	3.061	3.182	3.226	3.306	3.447
31	0.141	0.121	0.121	0.120	0.119	0.119



Table 4 Values of Root Means Absolute Error

Station	Plotting Position Formula						Nguyen
	Weibull	Beard	APL	Blom	Cunnane	Gringorten	
1	0.334	0.336	0.342	0.338	0.339	0.341	0.348
2	0.359	0.363	0.368	0.364	0.364	0.365	0.366
3	0.294	0.299	0.297	0.3	0.301	0.301	0.303
4	0.28	0.262	0.254	0.258	0.256	0.254	0.249
5	0.45	0.463	0.462	0.466	0.467	0.469	0.472
6	1.849	1.969	1.986	1.996	2.009	2.03	2.065
7	0.239	0.235	0.234	0.235	0.235	0.234	0.233
8	0.274	0.274	0.271	0.274	0.274	0.274	0.276
9	0.321	0.273	0.271	0.262	0.256	0.258	0.265
10	0.24	0.256	0.253	0.261	0.263	0.266	0.272
11	0.315	0.309	0.319	0.309	0.31	0.312	0.316
12	0.263	0.266	0.261	0.267	0.267	0.268	0.269
13	0.514	0.532	0.54	0.536	0.538	0.542	0.548
14	0.238	0.24	0.238	0.24	0.241	0.241	0.248
15	0.373	0.386	0.388	0.389	0.391	0.394	0.398
16	0.064	0.063	0.065	0.063	0.063	0.063	0.064
17	0.715	0.736	0.751	0.739	0.741	0.743	0.747
18	0.255	0.235	0.233	0.23	0.228	0.224	0.217
19	0.06	0.054	0.051	0.053	0.053	0.053	0.052
20	0.792	0.821	0.819	0.826	0.829	0.833	0.839
21	0.311	0.307	0.307	0.31	0.311	0.313	0.317
22	0.337	0.345	0.351	0.347	0.348	0.35	0.353
23	1.801	1.818	1.829	1.823	1.825	1.828	1.834
24	0.409	0.417	0.416	0.419	0.42	0.422	0.426
25	0.47	0.478	0.488	0.48	0.48	0.481	0.484
26	0.326	0.366	0.373	0.375	0.38	0.388	0.402
27	0.316	0.32	0.319	0.321	0.321	0.321	0.32
28	0.37	0.49	0.504	0.525	0.541	0.565	0.604
29	0.423	0.43	0.429	0.432	0.433	0.434	0.435
30	1.297	1.375	1.401	1.395	1.404	1.419	1.445
31	0.294	0.282	0.281	0.282	0.283	0.283	0.284



The eight plotting position formulas were ranked for all stations according to the values of RMSE and RMAE on scale 1 to 8, with one being the best method.

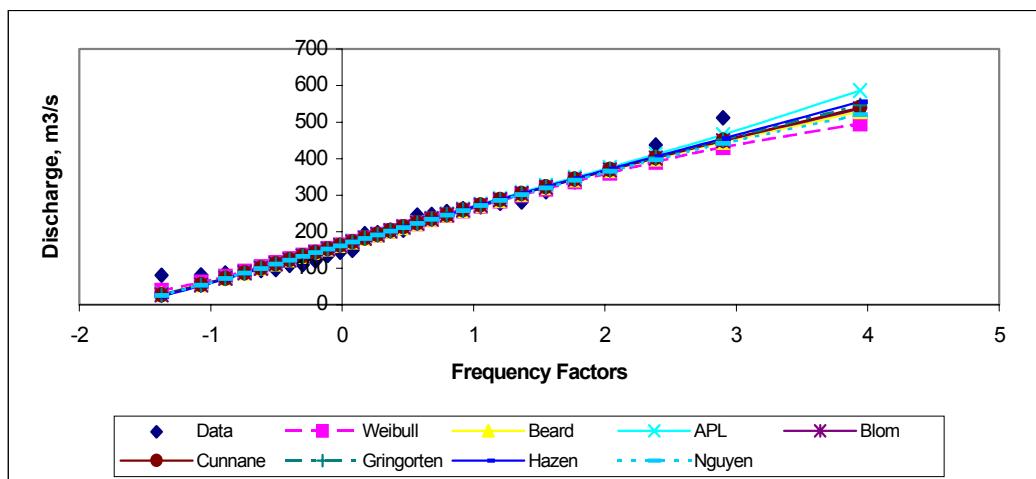


Figure 4 Comparison of Observed and Quantile Using The Plotting Position Formulas (r Are Accepted At 5% Critical Values, Station 12)

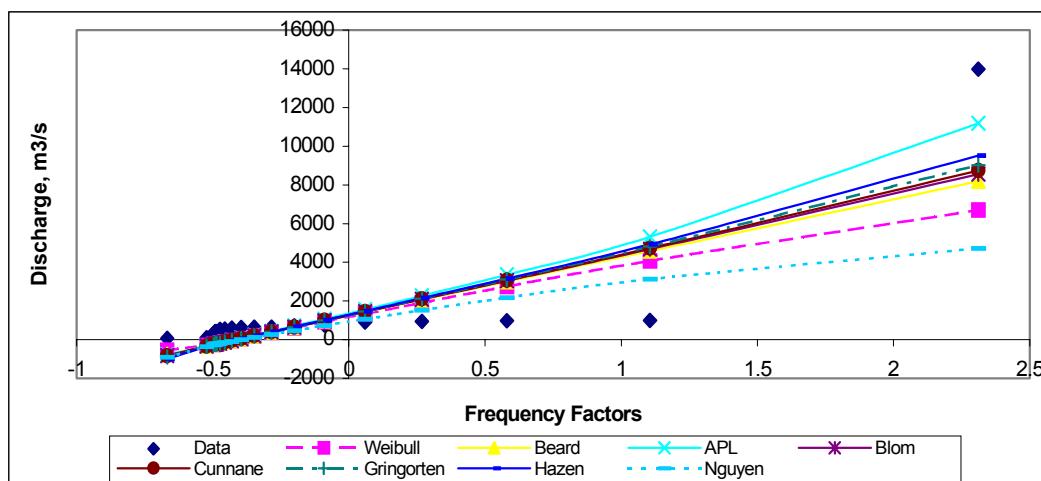


Figure 5 Comparison of Observed and Quantile Using The Plotting Position Formulas (r Are Rejected At 5% Critical Values, Station 30)

Table 5 ranks the eight plotting position formulas according to RMSE. It can be seen that the Weibull formula was the best, followed by APL, Beard, Blom, Cunnane, Gringorten, Nguyen and Hazen formulas in descending order of their performance.

The ranking of the eight plotting position formulas according to RMAE is given in Table 6. Clearly Weibull formula was the best of all, followed by APL, Beard, Blom,



Table 5 Ranking of the Plotting Position Formulas for 31 Stations by Root Means Square Error (RMSE) on a scale of 1 to 8 with 1 being the best method

Plotting Position	Number of Stations Receiving Ranking							
	1	2	3	4	5	6	7	8
Weibull	13	0	0	0	0	3	6	8
Beard	3	13	3	1	0	8	3	0
APL	10	3	7	3	2	3	2	1
Blom	2	2	8	9	10	0	0	0
Cunnane	2	0	4	15	10	0	0	0
Gringorten	0	6	6	0	5	8	4	1
Hazen	4	5	2	0	3	4	5	9
Nguyen	2	2	3	2	2	5	7	7

Table 6 Ranking of the Plotting Position Formulas for 31 Stations by Root Means Absolute Error (RMAE) on a scale of 1 to 8 with 1 being the best method

Plotting Position	Number of Stations Receiving Ranking							
	1	2	3	4	5	6	7	8
Weibull	20	2	0	1	1	1	0	6
Beard	1	13	12	0	0	0	5	0
APL	5	9	4	1	2	2	3	5
Blom	1	1	8	16	3	2	0	0
Cunnane	0	1	2	8	18	2	0	0
Gringorten	0	1	3	1	3	15	8	0
Hazen	3	1	0	0	1	1	11	14
Nguyen	1	3	2	4	3	8	4	6

Cunnane, Gringorten, Hazen and Nguyen formulas in descending order of their performance. Again, the previous conclusions hold. However those differences between plotting positions were not too great and therefore these plotting positions could be considered comparable for practical purpose.



8.0 CONCLUSIONS

Probability plots and the probability-plot correlation coefficient test statistic are used for testing the PIII using plotting position formula to fit annual maximum flow data. The PPCC test statistic was found to be a useful tool for discriminating among competing probability and plotting position formula. Eight plotting position formulas were compared for their ability to fit flood flow data. Overall these plotting position formulas consistently produced linear probability plots with r nearly one as measured by the PPCC test statistics. If an unbiased plotting position formula is required for the PIII distribution, then the Weibull formula would be the best selection.

REFERENCES

- [1] Adamowski, K. 1981. Plotting Formula For Flood Frequency, *Water Resour. Bulletin*, 17(2): 197-202.
- [2] Cunnane, C. 1978. Unbiased Plotting Positions- A Review, *Journal of Hydrology*. 37: 205-222.
- [3] Guo, S. L. 1990. A Discussion On Unbiased Plotting Positions For The General Extreme Value Distribution, *Journal of Hydrology*. 121: 33-44.
- [4] Guo, S. L. 1990. Unbiased Plotting Position Formula For Historical Floods, *Journal of Hydrology*. 121: 45-61.
- [5] Ji Xuewu, Ding Jing, H. W. Shen, and J. D.Salas. 1984. Plotting Positions For Pearson, *Journal of Hydrology*. 74: 1-29.
- [6] Kotegoda N. T., and R. Rosso. 1997. *Probability, Statistics, and Reliability for Civil and Environmental Engineers*. Mc-Graw Hill Book Co., New York.
- [7] Stedinger, J. R., R. M. Vogel, and G. E. Foufoula. 1993. Frequency Analysis of Extreme Events. *Handbook of Applied Hydrology*. Mc-Graw Hill Book Co., New York, Chapter 18.
- [8] Vogel, R. M., and D. M. McMartin. 1991. Probability Plot Goodness-of-Fit and Skewness Estimation Procedures for the Pearson Type 3 Distribution, *Water Resour. Res.*, 27(12): 3149-3158.
- [9] Vogel, R. M. 1986. The Probability Plot Correlation Coefficient Test for the Normal, Lognormal, and Gumbel Distribution Hypotheses, *Water Resour. Res.*, 22(4): 587-590.