International Seminar on Hydrology of Extremes 1-3 December, 1988
Roorkee (India)

RELATIVE PERFORMANCE OF REGIONAL FLOOD PREDICTION MODELS

S. H. Moharram B. P. Parida P. N. Kapoor

Civil Engineering Department
Indian Institute of Technology, New Delhi-110016

SYNOPSIS

Prediction of extreme hydrologic events, such as, peak flood magnitude for design decision in water resources constitutes a serious problem throughout the world. The accuracy attained is limited by the available data at a given site and lack of complete understanding of the underlying mechanism originating the extreme flood events.

Regional flood frequency techniques are used mainly to estimate a specific return period flood flow at a gauging site or to provide site specific estimates based on limited single site data. Regionalisation technique associates annual flood characteristics with physiographic and climatic causative factors. Based on the distributional assumption, degree of spatial heterogeneity and intersite correlation, many regional approaches have been suggested in the recent past. These include estimation of N-dimensional location parameter by James-Stein estimator subject to Lindley modification, use of Wakeby distribution, coupling the Index-Flood method with estimation by probability weighted moments ,(PWM), and regionalising the parameter of Box-Cox transformation.

For a data sparse country like India, regional frequency analysis can be of great value. Hence identification of a suitable technique from amongst the available techniques reported in literature is desirable. In this paper an attempt has been made to evaluate the relative performance of techniques that use (i) Index-Flood method with PWM estimators, (ii) Wakeby distribution with James-Stein estimators for corrected means, (iii) Log-Boughton distribution, and (iv) Method of power transformation, in estimating flood quantiles.

1.0 INTRODUCTION

Regional flood frequency analysis as a tool for flood estimation at any site within a hydrometeorological homogenous region, with either fully or partially available record or with even no record, is gaining recognition in contrast to the traditional flood frequency analysis techniques for site specific quantile estimations. Several approaches to regional flood frequency analysis have been explored, (Gries and Wood, 1981; Kuczera 1982; Hosking et al. 1985; Seth and Singh 1987).

Cunnane (1988) provided a comprehensive overview of several regional flood estimation methods. One of the simplest approach in regoinal flood frequency analysis is to use a multiple regression model to develop a regional relationship between flood quantiles and the catchment characteristics and extend the relationship to estimate the desired quantile at an unguaged/guaged site (Benson, 1962; Tasker, 1980; Stedinger and Tasker, 1985). Another widely used approach is that of the "index flood", (NERC, 1975), in which a dimensionless regional frequency curve is rescaled at the site of interest by a median scaling factor. Dalrymple (1960) introduced the index-flood method along with a method for examining homogeneity of the gauging sites and used them for deriving a regional flood frequency curve also known as the growth curve.

More recently, attention has been directed towards improving the flood quantiles by using index-flood approach through use of (PWM) as the method of parameter estimation instead of the traditional methods, (Wallis, 1980; Gries and Wood, 1981; Wallis and Wood, 1985).

In the present study, four candidate regional flood estimation methods have been used for making specific return period quantile estimates. There are: (i) Index-Flood method with probability weighted moment estimators, (INDF/PWM), (ii) Wakeby distribution with James Stein estimator for corrected means, (WA/PWM), (iii) Power transformation, (PT), based on assimilation of dimensionless data from various sites, and assuming them as a single homogenous sample, and (iv) Log-Boughton distribution, (LB), based oncombination of standardized frequency factors from various sites, and assuming the same as a single sample.

Since no comparative performance of the above stated methods for regional quantile estimation is available, an attempt has been made in this study to evaluate their reliability and efficiency. Further to bring out a judicious comparison, this study has specifically been carried on data from 10 sites of Northern India region satisfying the Dalrymple homogeneity test.

2.0 METHODS OF ANALYSIS

T-year flood quantile-magnitudes, Q_T, at any site are estimated using the above listed four methods. All these methods, which use the annual flood peaks in their analysis, are described as under.

2.1 Index-Flood Method with PWM Estimators:

The index flood method (Dalrymple, 1960) uses a graphical procedure to estiamte the parameters and hence the quantiles of the assumed Gumbel/EV1 distribution and computes an averaging ratio of the estimated quantiles and the index flood. In order to overcome the possible errors in the estimated quantiles due to inaccurate estimation of parameters, an improved parameter estimation technique (PWM) as suggested by Greenwood et al. (1979) has been adopted by Gries and Wood (1981). This forms the basis for consideration and adoption in this study.

The annual peak discharges in a region are first tested for homogeneity at 10-year period level using EV1 with PWM estimator. For each site, a probability plot is then prepared and the following steps followed: (1) Flood frequency curve is established for future use in estimation of quantiles, (2) The mean annual flood, Q_m, corresponding return period 2.33 yr calculated from each of the frequency curves, (3) Using EV1/PWM, the recurrence interval of floods at various return period, Q_T, (say 2,5,10,20,50 and 100 yrs.) are estimated, (4) Standardised quantile estimates, Q_T/Q_m, are then computed from which the median, the arithmatic mean, and the weighted mean are evaluated and plotted on Gumbel or EV1 probability paper for each return period, T. This plot is termed as the regional frequency curve, (5) Relationship between the mean annual annual flood, Q_m, and catchment area, CA, is established through a linear regression analysis using Eq(1) given below,

$$Q_m = a_o (CA)^{b_o} \qquad .....(1)$$

where a_o and b_o are constants, and (6) Q_T for any gauged or ungauged site in the region is then estimated using information from the regional growth curve and Eq. (1).

2.2 Wakeby Distribution with James-Stein Estimators for Corrected Means:

The Wakeby distribution (Hougthon, 1978), defined in its inverse form, is represented as under

$$X(F) = m + a\,(1 - (1-F)^{b}) - c(1 - (1-F)^{-d}) \qquad .....(2)$$

where, F is the probability of non-exceedence ($X \leq x$) and a,b,c,d and m are the parameters for Wakeby distribution.

Parameter estimation procedure using this distribution is based on the concept of PWM suggested by Greenwood et al. (1979), and is represented through Eq. (3) below:

$$M_{j,k} = 1/N(j) \sum_{j=1}^{N(j)} x_i \binom{n-i}{k} \Big/ \binom{n-1}{k} \qquad .....(3)$$

where:

$j = 1,2,......, Ns$
$k = 0,1,2,3,4$

N_s = Nos. of guaging stations
$M_{j,k}$ = kth order PWM for jth gauging station
$N(j)$ = No. of annual peak flows at jth gauging station
$x_{i,j}$ = ith item in the sample of ranked discharges in ascending order at jth gauging station.

To obtain parameter estimates, moment ratios of the form M_0/M_0, $M_1/M_0, \ldots\ldots,$ M_4/M_0 are computed for each station instead of just the traditional moments $M_0, M_1, \ldots\ldots, M_4$. Average of these moment ratios obtained for various catchments, (Wallis, 1980), give the regional parameter which is used to estimate the regional quantile for specific return period.

For ungauged site, a relationship between the James-Stein corrected means, instead of annual means, and catchment area is developed as given in Eq. (4).

$$Q_{JMS} = a_1 (CA)^{b_1} \qquad \ldots\ldots(4)$$

Eq. (4) is linear in Log-domain where a_1 and b_1 are coefficients obtained from regression analysis.

2.3 Method of Power Transformation:

This method is also considered as a potential technique for regional frequency analysis (Kuczera, 1982; Perumal and Seth, 1985). By this method transformation of observed annual peak discharges to near normality is done such that normal distribution, can be used for estimation of quantiles. For this, the annual peak discharges are grouped together in a standardised form dividing them by the mean annual flood at each site.

The transformation is carried out with the scheme given by Box-Cox (1964) and Chander et al. (1978),

$$\begin{aligned} Z_i &= (Q_i^{\lambda} - 1)/\lambda \qquad \text{for } \lambda \neq 0 \\ &= \text{Log } Q_i \qquad \text{for } \lambda = 0 \end{aligned} \qquad \ldots\ldots(5)$$

where:

Q_i = the variate of a given series
Z_i = the transformed variate, and
λ = a constant of transformation

From a standardised sample, the regional value of constnt, λ , is determined such that its coefficient of skewness is nearly zero. Flood quantiles are then estimated by

$$Q_T = (\lambda Z_T + 1)^{1/\lambda} \qquad \ldots\ldots(6)$$

in which $Z_T = \overline{Z} + K_T\,\sigma_z$, K_T is the frequency factor corresponding to return period, T, (Chow, 1964), and $\overline{Z}$ and σ_z are the mean and the standard deviation of the transformed series, respectively.

2.4 Log-Boughton Distribution:

This is a three parameter distibution (Boughton, 1980) based on the nonlinear relaionship between the frequency factor, K, and the function of recurrence interval, T.

The relation is written as

$$(K - A)\ (G - A) = C \qquad \ldots\ldots(7)$$

where:

K = the frequency factor
$G = \mathrm{Ln}\ (\mathrm{Ln}\ (T/\ (T-1)))$
$\quad = \mathrm{Ln}\ (-\ \mathrm{Ln}(F))$
A = the shape parameter, and
C = constant.

For each site, the frequency factors are determined by using the notation of the Water Resources Council (1976) guidelines. The base 10 logarithm of the discharge, Q, for selected recurrence interval, T, is given by:

$$\mathrm{Log}\ Q = \overline{X} + K\ S \qquad \ldots\ldots(8)$$

in which $\overline{X}$ = the mean of the log-transformed annual discharge series, S = the standard deviation of the above.

However, the standardized series are obtained by combining the frequency factors of all the sites together. Using the fitting procedure of Boughton (1983), the parameters A and C can be estimated as regional parameters. The T-year flood discharge is then calculated by:

$$\mathrm{Log}\ Q_T = \overline{X}^* + K_T\ S^*$$

$$\ldots\ldots(9)$$

$$Q_T = 10^{(\mathrm{Log}\ Q_T)}$$

where:

K_T = the frequency factor at various return periods,
$\overline{X}^*$ = the new mean at site, and
S^* = the new standard deviation at site.

For unguaged site, the relations between catchment area and both new mean and new standard deviation are calculated using linear regression and can be used to obtain the $\bar{X}^*$ and S^*. Hence the estimation flood quantiles can be made using Eq. (9).

3.0 COMPARISON CIRTERIA

For comparison of results obtained from four regional flood estimation methods, two criteria viz the mean absolute relative deviation and root mean square error as adopted by Wallis and Wood (1985), and Jain and Singh (1987), have been used. Main reason of choosing these criteria is to obtain indices on the goodness of fit from the four methods.

The mean absolute relative deviation (MARD), is defined as

$$\text{MARD} = 1/n \sum \left| \frac{Q_o - O_c}{Q_o} \right| * 100 \qquad \text{.....(10)}$$

and the root mean square error (RMSE), is given by

$$\text{RMSE} \; (1/n \sum (\frac{Q_o - Q_c}{Q_o})^2)^{0.5} \qquad \text{.....(11)}$$

where:

Q_o = the T-year flood using single at site estimate,
Q_c = the T-year flood at site using the regional approach.

4.0 STUDY REGION

A region located in the border district of Chamoli, U.P., of northern Indian has been selected for this study. The region has 10 guaging stations with catchment areas varying from 1,600 to 56,885 sq. km., and the annual flood record ranging from 8 to 79 years. Details of these guaging stations are presented in Table 1.

5.0 RESULTS AND DISCUSSION

Using the available data at the ten sites, a homogeneity test suggested by Dalrymple (1960) has been carried out. 10-year flood quantile at each site have been computed using EV1/PWM estimators for computation of the test-statistics instead of the traditional graphical procedure. Based on this test, eight gauging stations have been short-listed as given in Fig. 1, out of which seven have been used for development of the regional growth curves and one used for test purpose (Ravi at Madhopur). The sites (Rudra; Rudra Prayag) not falling within the test limits have been indicated by asterix against them in Table 1.

Based on the mean annual flood data computed at each site using the candidate distributions with or without James-Stein correction for means and their respective catchment areas, the relations developed are as under:

For INDF/PWM

$$Q_{2.33} = 257.75\ (CA)^{0.306} \qquad .....(12)$$

For WA/PWM

$$Q_{JMS} = 270.97\ (CA)^{0.3024} \qquad .....(13)$$

For PT

$$Q_{m} = 162.065\ (CA)^{0.354} \qquad .....(14)$$

For LB

$$\bar{X}^{*} = 2.199\ (CA)^{0.052} \qquad .....(15)$$

$$S^{*} = 1.0065\ (CA)^{-0.177} \qquad .....(16)$$

As per the procedure discussed earlier, regional growth curves have been developed using, (i) median ratios, (ii) arithmatic mean ratios, and (iii) weighted mean ratios. These curves, have been presented in Fig. 2. Also, flood quantiles based on regional parameters have been estimated based on the relations developed for each candidate distributions as given in Table 2.

For various return periods flood quantiles have been estimated using the regional parameters, growth curves based on median, arithmatic mean and weighted mean, and are tabulated in Table 3. Values of two comparison indices, i.e. MARD; RMSE, are given in Table 4.

Based on the MARD and RMSE values, performance of candidate distributions INDF/PWM and WA/PWM can be grouped into one category, and PT and LB into another.

It is thus seen that the MARD and RMSE values on particular group are of the same order irrespective of the estimation procedure used. Growth curves developed on the basis of weighted mean ratios using INDF/PWM estimates, yielded the least value of both MARD and RMSE, with almost similar results from the WA/PWM along with James-Stein correction for means.

6.0 CONCLUSION

It is concluded from the study that the Index-Flood method with PWM estimators can be used to develop regional growth curve based on weighted mean growth factors for various return period. Before it's potentiality as the best of four methods established, further study is suggested using data from other regions.

REFERENCES

Benson, M.A., (1962), 'Evolution of method for evaluating the occurrence of floods', U.S.G.S. Water Supply Pap. 1580-A, 30 pp.

Boughton, W.C., (1980), 'A frequency distribution for annual floods', Water Resources Res., 16 (2), pp 347-354.

Boughton, W.C., and Shirley, E.D., (1983), 'Fitting procedures for the Log-Boughton distribution', J. of Hydraulic Engg., ASCE, 109(HY4), pp 579-589.

Box, G.E.P., and Cox, D.R., (1964), 'Analysis of transformation', J. of the Royal Statistical Society, Vol. B26, pp 211-252.

Chander, S., Spolia, S.K., and Kumar, A., (1978), 'Flood frequency analysis by power transformation', J. of Hydraulic Engg., ASCE, 104(HY11), pp 1495-1504.

Chow, V.T., (1984), Part I: 'Frequency analysis': in :V.T. Chow, Handbook of Applied Hydrology.

Cunnane, C., (1988), 'Methods and merits regional flood fequency analysis', J. of Hydrology, Vol. 100, pp 269-290.

Dalrymple, T., (1960), 'Flood frequency analysis', U.S. Geol. Surv. Water Supply Pap., 1543-A, 80 pp.

Greenwood, J.A., Landwehr, J.M., Matalas, N.C., and Wallis, J.R., (1979), 'Probability weighted moments: Definition and relation to parameters of several distribution expressable in inverse form', Water Resources Res., 15(5), pp 1049-1054.

Greis, N.P., and Wood, E.F., (1981), 'Regional flood estimation and network design', Water Resources Res., 17(4), pp 1167-1177.

Hosking, J.R.M., Wallis, J.R., and Wood, E.F., (1985), 'An appraisal of the regional flood frequency procedure in the U.K. Flood Studies Report, Hydrol. Sci. J., 30(1), pp 85-109.

Houghton, J.C., (1978), 'Birth of a parent: The Wakeby distribution for modelling flood flows', Water Resources Res., 14(6) pp 1105-1110.

Jain, D., and Singh, V.P., (1987), 'Estimation parameters of EV1 distribution for flood frequency analysis', Water Resources Bull. 23(1), pp 59-71.

Kuczera, G., (1982), 'Robust flood frequency models', Water Resources Res., 18(2), pp 315-324.

Natural Environment Research Council, (1975), 'Flood Studies Report', Her Majesty's Stationary Office, London.

Perumal, M., and Seth, S.H., (1985), 'Regional flood frequency analysis using power transformation -a case study', Seminar on Flood Frequency Analysis, N.I.H., Roorkee.

Seth, S.M., and Singh, R.D., (1987), 'Flood estimation in India using regional frequency analysis', National Symposium on Hydrology, Roorkee.

Stedinger, J.R., and Tasker, G.D., (1985), 'Regional hydrologic analysis 1. Ordinary, Weighted, and Generalized least squares compared', Water Resources Res., 21(9), pp 1421-1432.

Tasker, G.D., (1980), 'Hydrologic regression and weighted least squares', Water Resources Res., 16(6), pp 1107-1113.

Wallis, J.R., (1980), 'Risk and uncertainties in the evaluation of flood events for the design of hydrologic structures', Keynote address at "Seminar and Extreme Hydrylogical Events-Flood and Droughts", Erice, Italy, 33 pp.

Wallis, J.R., and Wood, E.F., (1985), 'Relative accuracy of Log-person type III procedures', J. of Hydraulic Engg., ASCE 111(7), pp 1043-1056.

Water Resources Council, (1976), 'Guidelines for determining flood flow frequency', Bull. 17, Hydrology Committee, Washington, D.C.

Table 1 : Physical Details Including Statistics and Homogeneity Test Computations for Ten Gauging Stations under Chamoli Region, U.P. (INDIA)

				SAMPLE STATISTICS								HOMOGENEITY TEST COMPUTATIONS			
Sl. No.	River, Station	Record size	Catchment Area (sq. km)	Mean Q_m (m^3/sec)	Standard Deviation (m^3/sec)	Coefficient			$Q_{2.33}$ (m^3/sec)	Q_{10} (m^3/sec)	Ratio, $\varepsilon = Q_{10}/Q_{2.33}$	$\hat{Q}_{10} = Q_{2.33} \times$ ave. ε (m^3/sec)	$\hat{T}$ for $\hat{Q}_{10}$ yrs.	Adjusted period of record yrs.
						Variation Cv	Skew Cs	Kurtosis Ks						
1.	Mandakini, Rudra	1971-78	1600	578	187	0.324	1.378	3.0	579	805	1.392	1212	193.55	8*
2.	Alakhanda, Rudra Prayag	1967-79	9933	2064	496	0.240	-0.142	2.0	2065	2755	1.334	4322	424.31	13*
3.	Bhagirathi, Tehri	1963-78	7511	3024	1387	0.458	1.487	5.0	3026	4758	1.572	6333	43.86	16
4.	Sarada, Banbasa	1960-78	14975	7790	2792	0.358	0.676	3.0	7802	18403	2.359	16329	7.35	49
5.	Sutlej, Nichari	1909-51	56885	4523	1647	0.364	0.561	3.0	4529	9529	2.104	9479	9.85	43
6.	Ganga, Raiwala	1901-79	22396	5832	2665	0.457	1.936	9.0	5845	17168	2.937	12234	5.10	79
7.	Beas, Pong dam	1935-76	12540	8560	3529	0.412	0.557	2.0	8571	18100	2.112	17939	9.73	42
8.	Yamuna, Tajewale	1924-78	11120	5607	3916	0.698	1.567	5.0	5617	14461	2.275	11756	6.20	55
9.	Tons, Kishau	1943-78	4885	1843	1904	1.033	2.168	6.0	1846	4035	2.186	3864	8.83	36
10.	Ravi, Madhopur	1918-62	6075	4413	3738	0.847	2.031	6.0	4421	10448	2.363	9253	7.33	44

* REJECTED STATIONS

Table 2 : Regional Parameters for the Candidate Distributions

Candidate Distribution	Values of Regional Parameters									
	a	b	c	d	m	λ	$\bar{Z}$	σ_z	A	C
INDF/PWM	.7573	.5629								
WA/PWM	0.091	24.0	2.748	0.249	0.0					
PT						-.107	.138	.495		
LB									4.163	19.746

Table 3 : Estimates Various Return Period Floods, m^3/sec, for Test Site; Ravi at Madhopur

Quantile Estimation	Candidate Distribution	Return Period, T, year					
		2	5	10	20	50	100
Using At Site Parameters	INDF/PWM	3654	7741	10448	13044	16404	18922
	WA/PWM	2540	6648	10381	14734	21601	27783
	PT	3469	6413	8697	11098	14487	17226
	LB	3718	6255	7836	9240	10883	11993
Using Regional Parameters	INDF/PWM	3115	6296	8401	10422	13036	14996
	WA/PWM	2300	5467	8392	11869	17493	22688
	PT	3087	4697	5894	7141	8904	10346
	LB	2612	4023	4923	5741	6726	7408
Using Growth Factor, Median Ratio	INDF/PWM	3146	6126	8101	9995	12446	14283
	WS/PWM	2232	6261	9470	12661	16898	20119
	PT	3303	4598	5597	6616	8036	9179
	LB	3356	4588	5239	5791	6460	6910
Using Growth Factor, Arithmatic Mean	INDF/PWM	3109	6285	8390	10406	13015	14972
	WS/PWM	2401	5739	8549	11505	15621	18910
	PT	3073	4556	5738	7115	9569	12500
	LB	3250	4588	5540	6121	6988	7590
Using Growth Factor, Weighted Mean	INDF/PWM	3039	6589	8942	11199	14116	16306
	WS/PWM	2096	6060	9398	12884	17894	21504
	PT	2644	4263	5735	7080	9416	12080
	LB	3236	4604	5452	6241	7190	7860

Table 4 : Values of MARD and RMSE for the Candidate Distributions

Candidate Distribution	Using Regional Parameters		Using Growth Factor					
			Median Ratios		A. Mean Ratios		W. Mean Ratios	
	MARD (%)	RMSE	MARD (%)	RMSE	MARD (%)	RMSE	MARD (%)	RMSE
INDF/PWM	19.06	0.192	21.54	0.218	19.20	0.193	14.67	0.147
WA/PWM	17.20	0.175	15.02	0.168	19.72	0.216	14.89	0.156
PT	30.69	0.322	33.39	0.363	28.61	0.298	32.07	0.323
LB	36.15	0.363	31.65	0.335	29.13	0.303	28.43	0.294

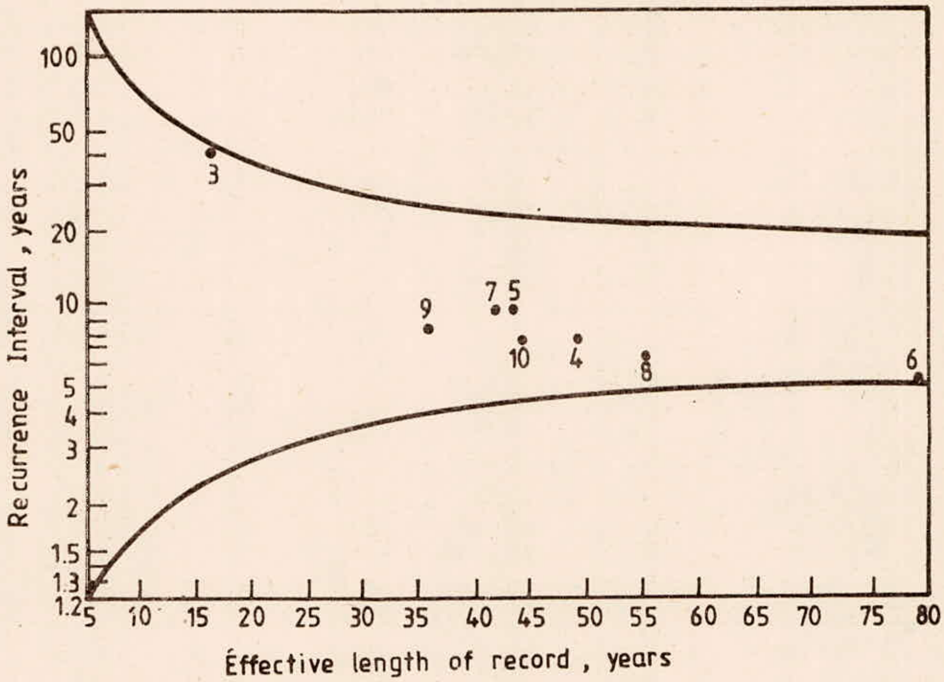

FIG. 1. HOMOGENEITY TEST PLOT FOR THE GAUGING STATIONS OF CHAMOLI REGION

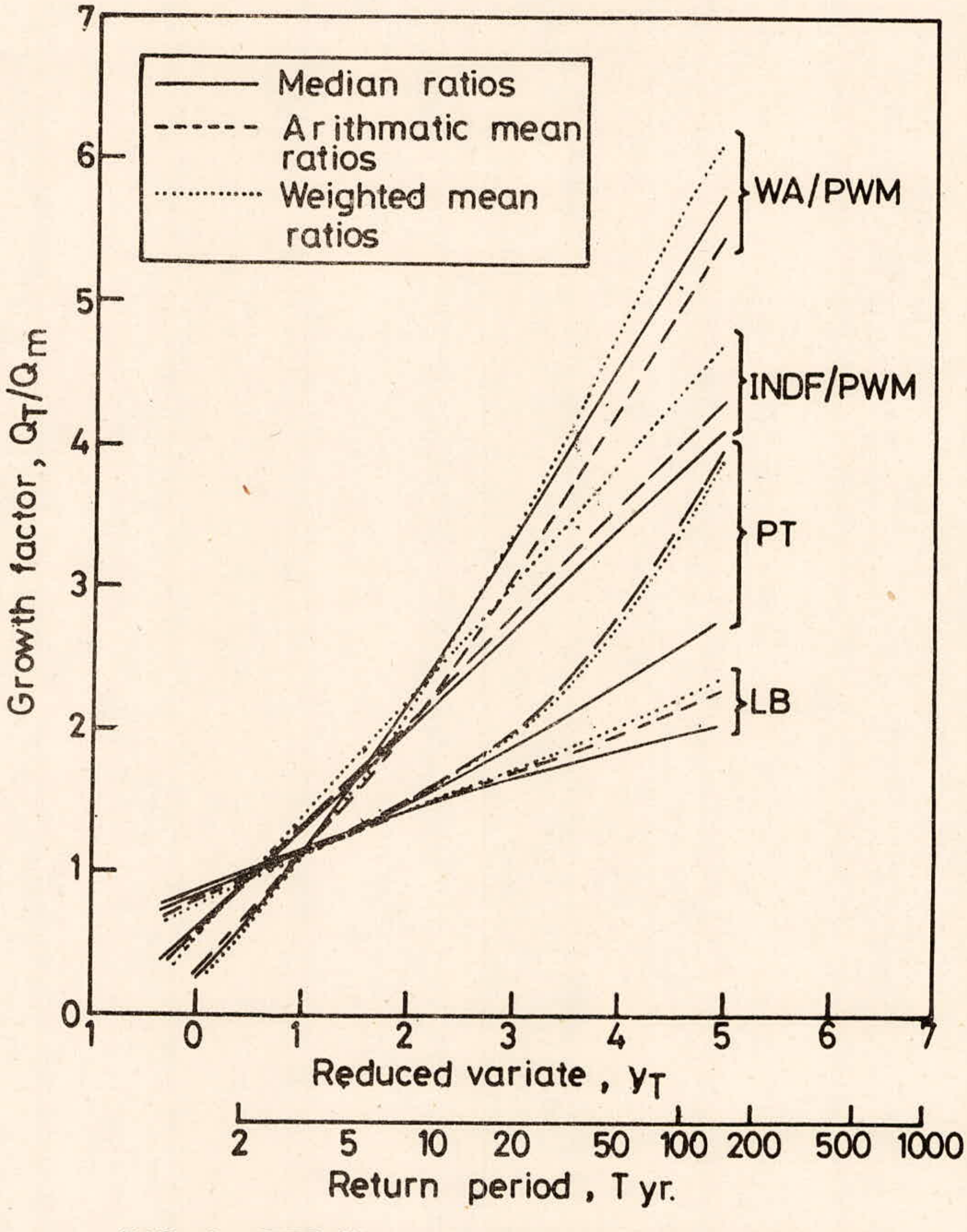

FIG. 2. REGIONAL FREQUENCY CURVES