

Worth of extreme flood events in site specific flood frequency analysis

B. P. PARIDA

Department of Environmental Science, University of Botswana, Private Bag 0022, Gaborone, Botswana. (email: paridab@mopipi.ub.bw)

Abstract

To obtain reliable at-site flood quantile estimates from a small sample of observations is a trivial exercise. To overcome this problem, attempts to (i) use Probability Weighted Moment (PWM) method of parameter estimation and (ii) to increase the length of database through the incorporation of discontinuous extraordinary events by making the new sample a censored one have been in vogue. Though the use of PWM method has generally been accepted as a means of reduction of bias in the estimates, the usefulness of the later is yet to be established. This study, through a computer simulation, tries to bring out the worth of incorporation of the extraordinary events, particularly while such events as well as the observed sample are likely to have been corrupted with varying degrees of measurement error. It is found that incorporation of extraordinary event(s) yielded reliable estimates of quantiles, particularly with small sample sizes, even if the extraordinary event and the sample had measurement error up to 30 and 10% respectively.

INTRODUCTION

Flood, among others, is probably the most feared disaster with the highest public profile. In spite of extensive research in the last twenty five to thirty years, in the area of flood prediction, this century has witnessed extensive loss of human and animal lives including loss property worth billions of dollars caused by extraordinarily large magnitude of floods.

To model these floods and prepare management strategies therefrom needs an appropriate choice of distribution and method of parameter estimation. Though it is generally agreed that method of Probability Weighted Moments (PWM) can provide unbiased estimates of design floods, the choice of an underlying parent distribution describing the extreme flood behaviour at a site is made a priori and has always been a problem. Because of the limited experience a small sample of floods has gone through at a gauged site, it can provide only limited information about the flood frequency distribution at that site. To obtain more precise estimates of design flood would require incorporation of additional information on extraordinary events, which might have taken place in the near future, or in the distant past, such that extension of record could be possible (Condie and Lee, 1982). Such additional information however, can be obtained from paleo-hydrologic studies or records and chronicles or even from visible flood marks at the sites of interest.

Further, as the uncertainty in peak flow is related to the uncertainty in various hydraulic parameters involved in computation of discharge, it can involve measurement errors in

the observations, which in turn can influence the choice of distribution. This can be much different from the choice of distribution made apriori. It is reported by Herschy (1985) that uncertainty in measurements taken using current meters are of the order of 5 % and can even go up to 20% while computing discharges through indirect methods such as slope-area or fall-discharge methods. Since, the magnitudes of extraordinary events are usually estimated using the indirect methods, it is natural that they are likely to be corrupted with large measurement errors. It is in this context, worth of incorporation of such additional information need to be assessed so that the estimates from flood frequency analysis remain reliable.

In one of the recent researches, Hosking & Wallis (1986) undertook a simulation exercise and showed that inclusion of extraordinary information was useful for at-site flood frequency analysis, particularly when the site contained small sample of observations. But, they used the method of maximum likelihood (MML) procedure of parameter estimation as suggested by Prescott & Walden (1983) for undertaking flood frequency analysis with the discontinuous extraordinary observations (i.e. censored samples), as no procedure to undertake such an analysis using PWM was available.

To overcome this difficulty, Ding & Yang (1988) not only suggested a PWM procedure for parameter estimation incorporating extraordinary information but also showed that these estimators performed better both in terms of bias and efficiency even when compared to the results from MML, besides others. But their simulation study used different magnitude of extraordinary values in different samples and also has not considered the effect of measurement error in extraordinary values, which are not the normal case.

In this study, an attempt has been made to study the worth of incorporation of extraordinary event(s) to a small sample of observations, both of which are likely to be infested with varying degrees of measurement error in the estimated quantilesevaluated through the use of a PWM procedure. To assess their usefulness particularly with some commonly used two and three-parameter distributions, analyses have been carried out using the Extreme Value Type 1 (EV-1) and the Pearson Type III (PT-III) distributions.

THEORY

If $x_i(N)$, $i = 1, 2, 3, \dots, n, \dots, N$ is the sample of annual maximum floods observed over n years ranked in ascending order of magnitude and followed by say h numbers of discontinuous extraordinary events observed in a total period N years, the of sample PWMs, (b_r , $r = 0, 1, 2$) can be estimated (Ding & Yang, 1988) as:

$$b_r = (1/N) \left[\{(N-h)/n\} \sum_{i=1}^n \frac{(i-1)(i-2)\dots(i-r)}{(n-1)(n-2)\dots(n-r)} \cdot \frac{(N-h-1)(N-h-2)\dots(N-h-r)}{(N-1)(N-2)\dots(N-r)} \cdot x_i \right. \\ \left. + \sum_{i=n+1}^{n+h} \frac{(N-n-h+i-1)\dots(N-n-h+i-r)}{(N-1)\dots(N-r)} \cdot x_i \right] \quad (1)$$

In particular, the first three moments can then be defined by substituting $r = 0, 1$ and 2 as:

$$b_0 = (1/N) \left[\{(N-h)/n\} \sum_{i=1}^n x_i + \sum_{i=n+1}^{n+h} x_i \right] \quad (2)$$

$$b_1 = (1/N) \left[\{(N-h)/n\} \sum_{i=1}^n \frac{(i-1)}{(n-1)} \cdot \frac{(N-h-1)}{(N-1)} \cdot x_i + \sum_{i=n+1}^{n+h} \frac{(N-n-h+i-1)}{(N-1)} \cdot x_i \right] \quad (3)$$

$$b_2 = (1/N) \left[\{(N-h)/n\} \sum_{i=1}^n \frac{(i-1)(i-2)}{(n-1)(n-2)} \cdot \frac{(N-h-1)(N-h-2)}{(N-1)(N-2)} \cdot x_i + \sum_{i=n+1}^{n+h} \frac{(N-n-h+i-1)(N-n-h+i-2)}{(N-1)(N-2)} \cdot x_i \right] \quad (4)$$

In a closer perspective, it can be seen that the first part in each of the above three equations is the contributed by the sample flood observations while the second part corresponds to the contributions of extraordinary flood values. Now these sample PWMs can be used to estimate the parameters for the two chosen distributions as given below :

PT III Distribution

Using the first three sample sample PWMs viz: b_0 , b_1 and b_2 , the unbiased estimate of the scale(β), location(x_0) and shape(θ) parameters for the PT III can be computed (Parida, 1993) from:

$$\theta = 4 / C_s^2 \quad (5)$$

$$\beta = (1/3) [\{ (3b_2 - b_0) \cdot B_{(1/2, \theta)} \} / \{ I_{\downarrow}(\theta, 2\theta) \}] \quad (6)$$

$$\text{and } x_0 = b_0 - \beta\theta \quad (7)$$

where, C_s = Co-efficient of skewness of the observed sample,

$I_{\downarrow}(\theta, 2\theta)$ = Incomplete Beta Function evaluated at $\downarrow(\theta, 2\theta) = (\downarrow)[(3b_2 - b_0)/(2b_1 - b_0)]$

and $B_{(1/2, \theta)}$ = Complete Beta Function evaluated at $(1/2, \theta)$

Then the flood quantile for any return period T can be computed from

$$X_T = x_0 + \beta \theta + \beta \theta^{1/2} \cdot K_T(\theta) \quad (8)$$

where, $K_T(\theta)$ is the Chow's frequency factor for the given θ (or C_s), and can be obtained using Wilson-Hilferty transformation.

EV-1 Distribution

For this case however the first two sample PWMs viz: b_0 and b_1 can be used to compute the location (u) and scale(α) parameters of the EV 1 distribution (Greenwood et al., 1979) from Eqns (9) and (10) as given below:

$$\alpha = (2b_1 - b_0) / \ln 2 \quad (9)$$

$$u = b_0 - 0.5772\alpha \quad (10)$$

Then the flood quantile for any return period T can be computed from

$$X_T = u + \alpha [-\ln \{-\ln (1-1/T)\}] \quad (11)$$

STATISTICAL EXPERIMENT AND ANALYSIS

Assuming the flood sample to have emerged from a Pearson Type III parent, 10000 flood like values with a mean of $100 \text{ m}^3/\text{s}$, Co-efficient variation = 0.3 and Coefficient of skewness = 2.0, were generated. From the so generated values, different samples of size 10(10) 60(20) 100 were obtained. Using the population statistics, magnitude of an extraordinary event having an assumed return period of 150 years ($N=150$) was computed. To observe the usefulness of this value on the flood estimates two return periods were chosen, one within the period of within the period of observation of the extraordinary event, N , (say $T=100$) and the other beyond N (say $T=500$). Using the population parameters true values of 100 year (x_{100}^{true}) and 500 year (x_{500}^{true}) flood values were computed. Each of the sample so drawn for a given sample size, were then subjected to six different cases of measurement error as given in Table 1 and for each case estimates at $T=100$ and $T=500$ were computed separately using PTIII and EV 1 as the underlying distributions.

To arrive at the amount of error in sample and the error in the extraordinary event(s), though a first order analysis as suggested by Tung and Mays (1981) can be undertaken, values based on the experience from the field have been used as a simple exercise. Accordingly most error values of 10 and 30 % have been considered as the most likely values in current and extraordinary observations respectively. Extensive simulation studies have then been carried out with a chosen parent distribution having measurement errors of various magnitudes with or without the presence of extraordinary historic information having various degrees of measurement error as illustrated in Fig. 1.

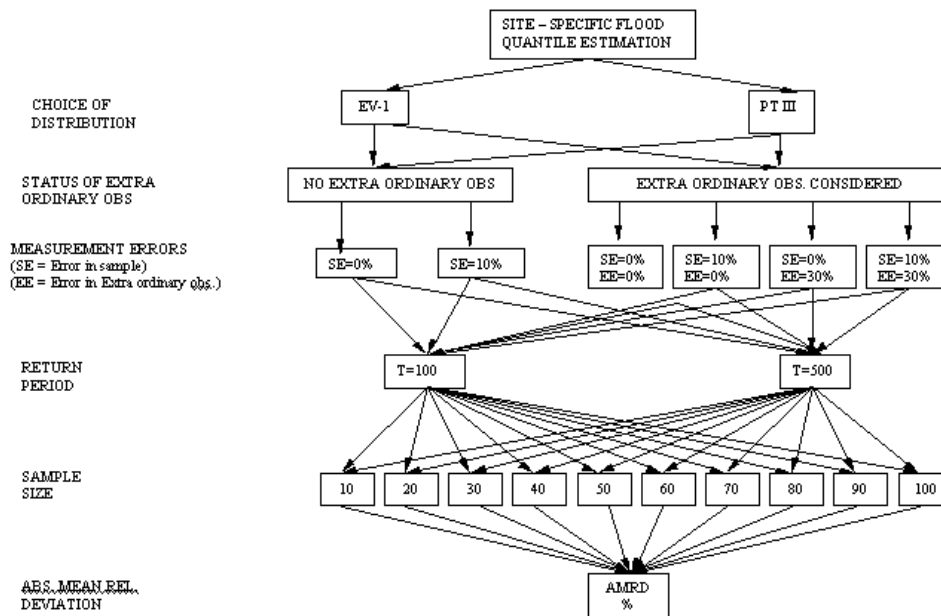


Figure 1. Simulation scheme with different combinations.

Table 1. Scheme for statistical experiment.

Case	Sample		Extraordinary Event(s)	
	Without Error	With !0% Error	Without Error	With 30% Error
1	x			
2		X		
3	x		X	
4		X	X	
5	x			x
6		X		x

Table 2. Effect of extraordinary observation(s) on Mean Absolute Relative Deviation in 100 year flood estimates computed using PT III and EV-1 distributions with varying sample sizes.

Sample Size (n) & No. of Samples (ns)	Distri- bution	Mean Absolute Relative Deviation in Estimated 100-Yr Flood Quantile (%)					
		No ExtraOrd Ob- servation		ExtraOrd Observation (True 150-Yr Flood) Considered			
		S.E.= 0%	S.E.=1 0%	S.E.= 0% +E.E.= 0%	S.E.=10% +E.E.= 0%	S.E.= 0% +E.E.=30%	S.E.=10% +E.E.=30%
10 (1000)	PT III	25.72	27.62	17.63	19.45	21.35	24.04
	EV - 1	18.30	18.66	14.31	14.92	17.43	17.46
20 (500)	PT III	16.53	17.37	11.97	12.22	15.30	15.90
	EV - 1	15.04	15.62	11.81	13.06	14.40	14.80
30 (333)	PT III	14.83	15.30	10.91	10.54	13.80	14.28
	EV - 1	13.34	14.05	9.97	11.82	12.43	12.89
40 (250)	PT III	12.12	12.53	8.27	8.35	10.55	11.33
	EV - 1	12.55	13.28	9.62	11.28	11.71	12.02
50 (200)	PT III	10.60	11.01	7.20	7.46	9.05	9.52
	EV - 1	11.65	12.31	8.88	10.58	11.31	11.36
60 (166)	PT III	8.80	9.46	6.70	7.02	7.93	8.38
	EV - 1	11.29	11.73	8.51	10.11	10.86	11.01
80 (125)	PT III	8.09	8.62	5.77	6.02	6.98	7.41
	EV - 1	11.02	11.38	8.17	9.18	9.80	10.86
100 (100)	PT III	7.16	7.31	5.34	5.40	6.57	7.01
	EV - 1	10.76	11.06	7.81	8.58	9.15	10.63

S.E. = Observation Error in Sample(s) ; E.E. = Observation Error in Extraordinary (ExtraOrd) Event(s)

In natural sequences, the errors (ϵ) usually have the properties of being random and multiplicative. And to achieve this, these values were assumed to be log-normally distributed with a zero mean and a standard deviation of $\log(1+\epsilon)$.

Then the PWM's for each sample of the assumed sample size were estimated with the help of equation (2), (3) and (4). Using the equations (5), (6) and (7), the parameters x_0 , β and θ for the PTIII distributions were obtained and in turn used to estimate the flood quantile values for T = 100 and 500 years from equation (8).

Table 3. Effect of extraordinary observation(s) on Mean Absolute Relative Deviation in 500 year flood estimates computed using PT III and EV-1 distributions with varying sample sizes.

Sample Size (n) & No. of Samples (ns)	Distribu-tion	Mean Absolute Relative Deviation in Estimated 500-Yr Flood Quantile (%)					
		No ExtraOrd Obser-vation		ExtraOrd Observation (True 150-Yr Flood) Considered			
		S.E.= 0%	S.E.=10 %	S.E.= 0% +E.E.= 0%	S.E.=10% +E.E.= 0%	S.E.= 0% +E.E.=30%	S.E.=10% +E.E.=30%
10 (1000)	PT III	35.15	35.25	23.04	24.70	27.18	30.66
	EV - 1	21.52	21.57	17.64	18.67	19.29	20.92
20 (500)	PT III	20.86	22.16	14.96	15.17	18.63	19.50
	EV - 1	19.50	20.01	16.77	17.49	18.58	18.98
30 (333)	PT III	18.87	19.61	13.86	13.30	17.43	17.84
	EV - 1	16.99	18.51	15.62	16.03	16.54	16.83
40 (250)	PT III	15.64	16.19	10.72	11.90	13.38	14.83
	EV - 1	16.37	17.65	15.40	15.85	16.15	16.28
50 (200)	PT III	13.42	14.22	10.43	9.63	11.71	12.68
	EV - 1	15.48	17.05	15.05	15.20	15.28	15.33
60 (166)	PT III	11.51	12.45	9.04	9.91	10.39	11.01
	EV - 1	15.38	16.88	14.97	15.13	15.17	15.24
80 (125)	PT III	10.29	11.18	7.26	7.99	9.00	9.50
	EV - 1	15.02	15.66	14.18	14.54	14.73	14.82
100 (100)	PT III	8.72	9.02	6.60	6.77	7.81	8.20
	EV - 1	14.35	15.35	13.45	13.54	13.72	14.17

S.E. = Observation Error in Sample(s) ; E.E. = Observation Error in Extraordinary (ExtraOrd) Event(s)

The mean of the quantile values thus obtained were used to compute the Absolute Mean Relative Deviation as

$$AMRD (\%) = \frac{1}{ns} \sum_{i=1}^{ns} \left| \frac{(x_T)_i - x_T^{true}}{x_T^{true}} \right| \times 100 \quad (12)$$

where $(x_T)_i$ = Computed flood quantile for return period T for the sample i.
 x_T^{true} = True value of flood quantile for return period T
 ns = Number of samples

For arriving at the values of x_T^{true} , the parent distribution was assumed to be unknown. The two parameter EV-1 distribution was used in the given data set to historic true 100-year and true 500-year flood values in the manner narrated below.

First the sample PWM's viz b_0 and b_1 were computed using equations (2) and (3), which were then incorporated into equations (9) and (10) to obtain parameters u and α . The

quantile x_T for return period T was obtained for the EV-1 distribution using equation (11).

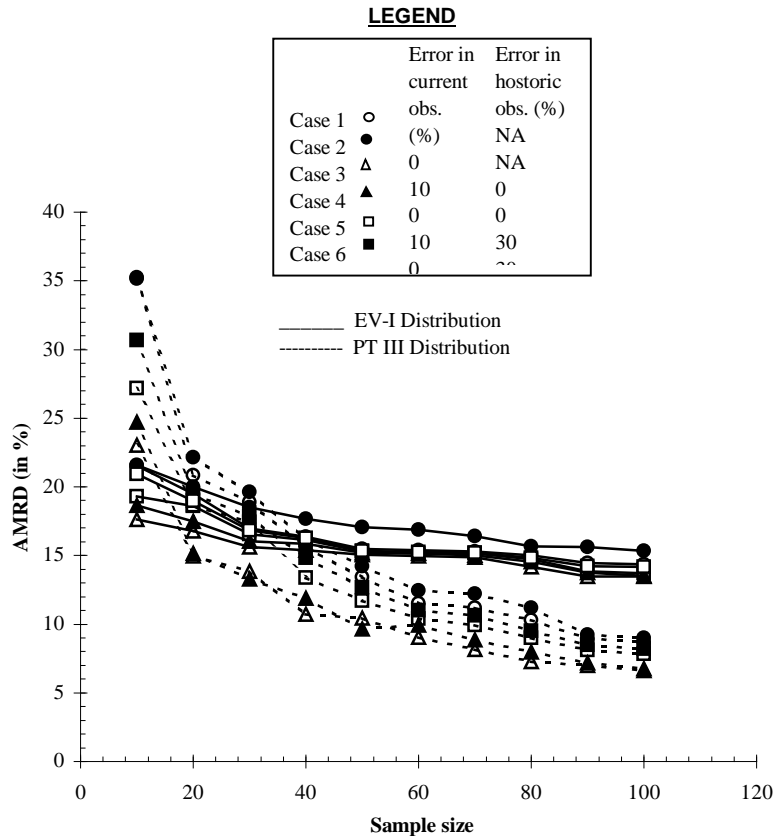


Figure 2. Changes in AMRD values of quantiles estimated at $t = 100$ years with sample size using EV-I and PT III distribution for six different cases.

Effect of Sample Size or AMRD Values

The overall examination of the Tables 2 and 3 as well as the Figs. 2 and 3 reveal that AMRD values decreased with increase in sample size. This decrease was significant when the samples varied from 10 to 50 in size. Thereafter, the decrease was marginal. Thus, it can be concluded that for flood quantile estimation for large return periods, it is preferable to use observations with large sample size.

Effect of Extraordinary Observations and Errors on AMRD Values

As shown in column (4) and (5) of Table 2, for a sample size of 10, where no extraordinary observation(s) was available, there was an increase of 2% in the AMRD values just by the introduction of a measurement error of 10%. On the inclusion of extraordinary observation of 150 years with the rest of the sample and assuming that both the extraordinary and sample observations were free from any error, it was observed that the AMRD

values decreased by 2.7%. However, with the consideration of error to the extent of 30% in the extraordinary observations and to the extent of 10% in the observed sample the AMRD values were comparable with the results obtained from consideration of sample observations alone without the consideration of extraordinary observation(s).

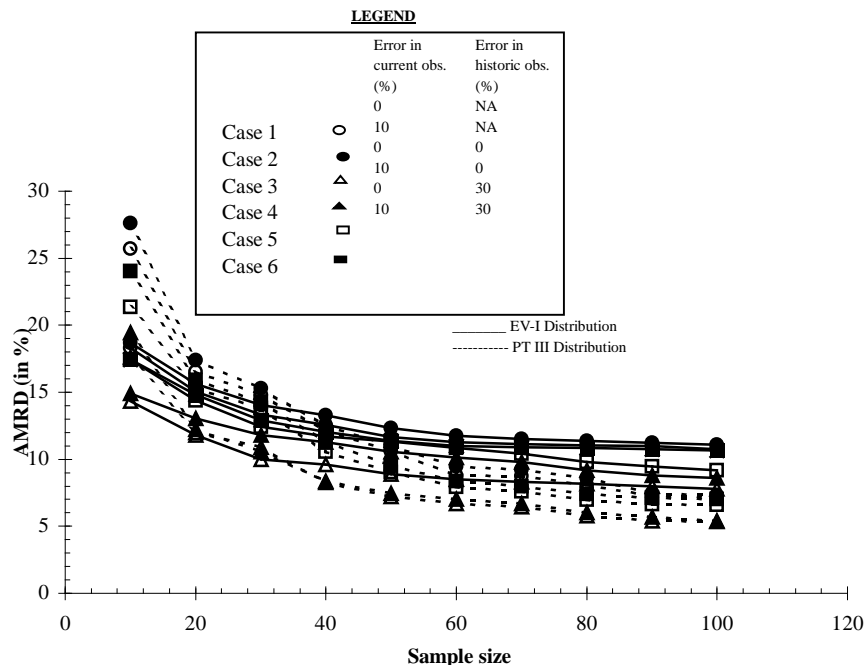


Figure 3. Changes in AMRD values of quantiles estimated at $t = 500$ years with sample size using EV-I and PT III distribution for six different cases.

The negligible differences between the two sets of results shown in column (3) and column (8) of Table 2 and 3, give a clear indication that incorporation of historic observations into sample with measurement error upto 10% was better than non-consideration of historic events, even though the samples were free from measurement error. It was further observed that introduction of more error in the sample and/or the extraordinary observation was counter-productive.

Effect of Choice of Distribution on AMRD Values

The above generated flood sample were subjected to analysis using EV-1 / PWM method. The AMRD values computed in this case were smaller than the values obtained from using PT-III / PWM distribution, for all cases at small sample sizes up to 30, even when the parent distribution was PT-III. In cases 3 to 6 i.e. the ones, which take into account the incorporation of extraordinary observations, EV-1 / PWM yielded higher AMRD values than the PT-III / PWM distribution for higher sample sizes. In summary it can be said that a two-parameter EV-1 / PWM method yielded less biased estimate of flood quantiles when the sample size was less than or equal to 30, and no historic obser-

vations available. However, for larger sample size, PT-III / PWM procedure was preferred to get better results.

Table 4. Effect of Historic Events on Mean Absolute Relative Deviation of Quantiles Estimated from Different Sample Sizes and Coefficients of Skewness.

Sample size , n & No. of Samples (ns)	Coefficient of Skewness C_s	Absolute Mean Relative Deviation (in %) of Estimated Flood Quantiles *					
		No Extraordinary Observation		One Extraordinary Observation** with 10% Error		One Extraordinary Observation** with 30% Error	
		T=100	T=500	T=100	T=500	T=100	T=500
10 (1000)	3.0	31.05	40.01	20.76	27.06	24.59	31.09
	2.0	25.72	35.15	17.63	23.04	21.35	27.18
	1.0	19.97	26.01	14.60	19.44	17.86	22.65
20 (500)	3.0	20.57	26.12	14.59	18.35	17.42	21.45
	2.0	16.53	20.86	11.97	14.96	15.30	18.63
	1.0	13.67	16.08	9.32	11.97	11.81	15.76
30 (333)	3.0	18.26	23.21	13.32	16.80	15.80	20.04
	2.0	14.83	18.87	10.91	13.86	13.80	17.43
	1.0	11.17	14.37	8.47	11.07	11.07	15.22
40 (250)	3.0	15.45	19.85	10.34	13.34	12.15	15.44
	2.0	12.12	15.64	8.27	10.72	10.55	13.38
	1.0	9.03	11.97	6.58	8.92	9.02	12.75

* Based on 10000 PT III flood like values generated with parameters average=100 m³/sec, $C_v=0.3$, $C_s=1.0(1.0)3.0$; Period of Extraordinary observation N=150 years and random multiplicative error.

** Extraordinary value is considered as true 150 year value i.e. 220.02 m³/s for $C_s=3.0$, 208.40 m³/s for $C_s=2.0$, 191.0m³/s for $C_s=1.0$

Effect of Skewness on AMRD Values

To study the effect of incorporation of extraordinary observations on samples which have been drawn from populations with varying skew properties, flood like values were generated separately with different coefficients of skewness, keeping the location and scale parameters same as before. For a meaningful comparison, two sets of 10,000 Pearson Type III distributed flood like values with $C_s = 1.0$ (less than the originally assumed value) and $C_s = 3.0$ (more than the originally assumed value) were generated. To each of these series, appropriate values of extraordinary events with recurrence interval of 150 years had been incorporated. Each series containing the flood like values were then divided into 1000, 500, 333 and 250 number of samples of size 10, 20, 30, 40, respectively and were subjected to analysis under three categories viz:

- i. No extraordinary observation considered
- ii. One extraordinary observation and no measurement error
- iii. One extraordinary observation and 30% measurement error

and for each case, quantiles were estimated at T=100 and T=500 years, as earlier.

AMRD values were then computed for the above 3 cases and for varying sample skewness values and have been listed in Table 4. For the case with higher skewness i.e. $C_s = 3.0$, AMRD showed a marked decrease with increase in sample size, in all the three cate-

gories as listed above. But for skewness as low as 1.0, though reduction in the AMRD values were observed, but the extent of reduction was not that pronounced as in the case of large C_s and when the sample size was small. In conclusion it is observed that results from $T = 500$, $T = 100$, though were similar but the reduction in AMRD value were more pronounced in case of quantiles estimated at $T=500$ years than at $T=100$ years.

CONCLUSIONS

From the present study aimed at identifying some of the benefits of incorporation of extraordinary information in flood frequency analysis, the following conclusions can be drawn.

1. While undertaking at-site flood frequency analysis with some of the commonly used distributions such as EV-1 or PT-III distribution even with PWM method of parameter estimation, incorporation of extraordinary or paleologic events yielded results that are close to reality.
It is also evident from the result of the statistical experiment that it is worthwhile to incorporate the extraordinary events even if these events contained upto 30% measurement error while the sample contained up to 10% measurement error.
2. It is also inferred that, when the sample size was small and the skewness was large (typically more than 2.0), the incorporation of extraordinary information significantly reduced the likely error in the quantile estimates.
3. When no extraordinary observations were available and the sample sizes were small, typically less than or equal to 30, the use a two parameter EV-1 distribution are preferable to the three parameter PT III distribution.

Based on the above findings, it can be concluded that at the sites which have experienced unprecedented extreme floods in the near past, the design procedure be revisited to incorporate these events and obtain a reliable estimate which can be used for evolving flood management strategies in these areas as well as to bring back confidence in the minds of people.

References

- Condie, R. and Lee, K.A., 1982. Flood Frequency Analysis with Historic Information, *J. of Hydrology*, 58, 47 – 61.
- Greenwood, J.A., Landwehr, J.M., Matalas, N.C. and Wallis, J.R., 1979. Probability Weighted Moments: Definition and Relation to Parameters of Several Distributions Expressible in Inverse Form, *Water Resources Research*, 15(5), 1049-1054.
- Herschy, R.W., 1985. *Streamflow Measurement*, Elsevier Applied Science Publications, London, 553.
- Hosking, J.R.M and Wallis, J.R., 1986. The Value of Historical Data in Flood Frequency Analysis; *Water Resources Research*, 22(11), 1606 – 1612.
- Jing, D and Yang, R., 1988. The determination of Probability Weighted Moments with the Incorporation of Extra-ordinary values into Sample Data and their Application to Estimating Parameters for the Pearson Type Three Distribution, *Jr. of Hydrology*, 101, 63-81.
- Parida, B.P., 1993. *Statistical and Physically Based Flood Frequency Analysis*, Ph.D. Thesis, IIT, Delhi, 96 – 112,.

- Prescott, P. and Walden, A.T., 1983. Maximum-Likelihood Estimation of the Parameters of the Three-Parameter Generalized Extreme Value Distribution from Censored Samples, *Jr. of Statistical Computing and Simulation*, 16, 241-250.
- Tung, Y.K. and Mays, L.W., 1981. Risk Models for Flood Levee Design, *Water Resources Research*, 17(4), 833-841.