THEME 3

REGIONAL FLOOD FREQUENCY ANALYSIS

FLOOD ESTIMATION USING REGIONAL FLOOD FREQUENCY ANALYSIS

R. J. Garde
Professor

U. C. Kothyari
Lecturer

Civil Engineering Department, University of Roorkee, Roorkee (India).

SYNOPSIS

Annual flood peak series at 93 stations located all over India have been analysed. Out of the various frequency distributions available, it is found that GEV type I fits the data more closely. It is then found that flood with return period 2.33 yr ($Q_{2.33}$) is related to catchment area A, catchment slope S, percent area covered by Forest A_F and P, where, P is 2 yr; 45 mt, 3 hr, 6 hr or 24 hr rainfall depending on A/S value. It is also found that the coefficient of variation of flood series is related to coefficient of variation of maximum monthly rainfall series in the catchment.

1.0 INTRODUCTION

The estimation of flood magnitude and its return period is one of the most important information required for flood management in a basin. Design of various hydrologic structures such as spillways, barrages, levees requires a realistic estimation of flood of given return period.

Flood Frequency Analysis approach is commonly used for estimation of design flood. When the data record of sufficient length are available, flood frequency analysis can be based on that record alone. However, historical records of floods are not available for all the sites of interest in the country and quite often the record length is not adequate for the single site frequency analysis. When the data available are limited or when there is no record available at a site, the regional flood frequency approach provides reasonable estimate of expected floods. In order to have a consistent approach for estimation of design floods, it is necessary to develop standard procedure and guidelines for deciding about the frequency distribution that the flood series follow and relation of parameters of frequency distribution with climatic, hydrometeorologic and catchment characteristics.

At present some empirical formulae such as those of Dicken (10), Ryves (12), Inglis (11), Ali Nawab Jang Bahadur (12), the rational formula and envelope curves of Kanwar Sain and Karpov (12) are available for the estimation of flood magnitude when little or no data are available, for any catchment. Because of the limited data on which they are based and because of the fact that they do not include all the variable on which flood discharge of given return period depends, these formulae are not expected to give results with satisfactory accuracy. Use of design rainfall alongwith synthetically generated unit hydrograph for the catchment can be alternatively used to obtain the peak discharge when little or no data exist. It may be mentioned that the results of such procedure are greatly dependent upon the accuracy of unit hydrograph that will be used. However, at present enough information is not available so that a reliable synthetic unit hydrograph can be obtained for any catchment in India.

Keeping the above mentioned points in view, all the available flood data for different rivers in the country have been analysed using the regional flood frequency approach. The results of this study are presented herein.

2.0 DATA

Data for yearly flood peaks of several catchments were obtained from various sources, which are listed in reference (1). The length of data for yearly flood peak varied from a minimum of 10 years to 83 year. In all, flood data were available for 93 catchments. The data were from small, medium and large catchments from all over the country. For 43 catchments the data length was twenty years or more; hence these data were used for flood frequency distribution studies. Since all the flood data used in this study have been already analysed by various investigators and also published else-where, need for reexamination of data for likely error was not felt.

The catchment areas were also available from the literature. The average rainfall values having recurrence interval of 2 years, duration of 45 minutes, 3 hours, 6 hours and 24 hours were obtained for all the catchments from the maps to a scale of 1:6000000 given in the Indian Meterologic Department (IMD) monograph (13) The coefficient of variation of maximum monthly rainfall C_{VR} for different catchments was obtained from the Rainfall Atlas of India to a scale of 1:6000000. The catchment slope was defined as $S = (1/A) \Sigma A_i S_i$, where S_i is the slope of the portion of catchment having area A_i. The slope for a large number of catchments was available from previous studies made by the investigator(1). The slope for the remaining catchments was obtained from the maps given in the National Atlas of India and in the Irrigation Atlas of India, to a scale of 1:6000000. The land usage of four types viz. arable land (A_A), grass and scrub (A_S), waste land (A_W) and forest land (A_F) were found in earlier studies by investigators to have influence on the hydrologic variables. The data on land usage of several catchments were available with the investigators from earlier studies; for remaining catchments the portion lying under each of these types was measured from the maps given in the Agricultural Atlas of India and in the Forest Atlas of India to a scale of 1:1000000. It is realised that land usage must have changed during the period of record of data However, in the absence of any other information, information available in the Agricultural Atlas and Forest Atlas was considered average over the period and used. The range of various quantities for the data used in the present study is given in Table 1.

3.0 ANALYSIS

3.1 Testing of Frequency Distributions

The frequency distributions that are generally fitted to the annual peak flood data are the 2 and 3 parameter Log-normal, GEV type I (Gumbel's distribution) or GEV type II, PT-3 and LPT-3. Recently (3) five parameter Wakeby distribution has also been fitted to the annual peak flood series. In addition to these the use of transformations to convert the series to near normal has also been advocated. However, there is no single distribution which is in universal use. The five parameter Wakeby distribution is the most flexible distribution but the parameter estimation based upon single site data is poor even when the probability weighted moment method is used (3). Moreover, in the regional flood study using Wakeby distribution, five different relationships for the five parameters of the distribution will have to be developed. In the family of GEV distribution, the GEV type I or Gumbel's distribution has been most popular. The Log Normal, The LPT-3, PT-3 distributions have also been used quite frequently. It has been found that the normalisation procedures applied to peak annual flood spell trouble frequently(1).

Therefore, it is worthwhile to make comparative study about the closeness with which distributions GEV type-I, PT-3, LPT-3, 2-parameter and 3-parameter Log-normals can be fitted to the annual peak flood series from Indian Catchments.

The computer programme given by Kite (16) was used to test the distributions. The method of maximum likelihood was used for estimating the parameters of all the distributions.

The closeness of fit of these distributions to the observed data was evaluated by computing standard error of the distributions. The standard error of distribution is defined as

$$S_E = [\sum_{i=1}^{n} (x_i - y_i)^2) / (N - m_j)]^{1/2} \quad (1)$$

where S_E is the standard error of the distribution x_i, i = 1, ...n are the standard events, y_i i= 1......n are the event magnitudes computed from particular frequency distribution at probabilities computed from the stored ranks of x_i, i 1....n and m is the number of parameters estimated for the jth distribution. The data from 43 catchments having minimum length of 20 years were used in this analysis. The catchments considered have been taken from all parts of the country.

A close study of the standard error of the distribution for the data of all the sites was made. It was noticed that for most of the catchments the standard error for all distributions was nearly the same. It can also be mentioned that for 20 catchments the skewness of log values were negative; therefore the application of LPT-3 distribution to these catchments would result in lower estimates of the flood magnitudes.

In order to check which of the distributions fits better to the upper tail of the data, the standard error for each distribution was computed using the highest ten percent values of the series of all catchments. It was noticed that for 33 out of 43 series the standard error was minimum for Gumbel's distribution and in case of remaining series also the standard error of Gumbel's distribution is not much higher than that of the other distributions.

It is known that there is a reasonable mathematical basis for the annual peak flood series to follow the Gumbel's distribution. The Gumbel's distribution has been extensively used in flood frequency analysis else where (2, 4, 6, 7, 9, 5). However, having only two parameters and a constant skewness the Gumbel's distribution may not be flexible enough to represent adequately a number of flood sequence. Nevertheless, it remains a useful model for many flood series, particularly treated in the context of single-site or at site analysis. In addition to this, the distribution is a simple one and the long term data from 43 Indian Catchments reveal that this distribution fits more closely to them than the others; hence for further investigations only the Gumbel's distribution with the method of maximum likelihood for parameters estimation has been used.

3.2 Relationships for the Parameters of Gumbel's Frequency Distribution

Mean Annual Flood

The flood magnitude having a recurrence interval of 2.33 yrs. i.e. $Q_{2.33}$ becomes the mean annual flood when the Gumbel's distribution is fitted to the flood series. The values of $Q_{2.33}$ were computed using the Gumbel's distribution with procedure of maximum likelihood for the data from 93 catchments. Based upon the review of literature the following functional relationship was assumed for the mean annual flood.

$$Q_{2.33} = f(A, S, P, F_V)$$

where A is the catchment area, S is the catchment slope, P is the rainfall term , and F_V is the vegetal cover factor. Based on previous hydrologic studies carried out by investigators, the term F_V is defined as follows :

$$F_V = \frac{a_1 A_A + a_2 A_S + a_3 A_F + a_4 A_W}{A}$$

where a_i, i = 1,4 are weighting factors. These were determined using the grid search technique. Different sets of values were given to the coefficients a_i, i = 1,4 and the correlation coefficient of F_V with $Q_{2.33}$ obtained. The combination of values of these weighting factors that gave maximum correlation coefficient are $a_1 = a_2 = a_4 = 0.0$ and $a_3 = 1.0$. Thus F_V becomes equal to the forest area in the catchment as a fraction to total area. The other morphometric terms such as drainage density, shape factor, etc. have not been included in the functional relationship for mean annual flood because the authors, earlier investigation (1) had shown these to be strongly interrelated with A and/or S in case of Indian Catchments. The lithologic variable and/or soil type variables were also found to be strongly correlated with the vegetal cover and rainfall. The rainfall values having recurrence interval of 2 years and duration 24 hour were considered in the begining of analysis for all catchments. As a first step in the analysis the influence of various factors on mean annual flood was studies by the simple correlation analysis and graphical plotting. Table 2 gives the correlation coefficients between the independent variables and their log-transforms with the dependent variable and its log transform.

Table 1 : Ranges of Data Used

Sl.No.	Variable	Abbreviations	Range
1.	Mean annual flood	$Q_{2.33}$	37.4 m^3/s to 72915 m^3/s
2.	Coefficient of Variation of annual peak flood	C_{VF}	0.15 to 1.3
3.	Coefficient of Variation of maximum monthly rainfall	C_{VR}	0.3 to 0.95
4.	2 years, 24 hour rainfall	P_{24}	65 mm to 200 mm
5.	2 year, 3 hour rainfall	P_3	10 mm to 50 mm
6.	2 year, 6 hour rainfall	P_6	20 mm to 80 mm
7.	2 year, 45 minute rainfall	P_{45}	15 mm to 35 mm
8.	Catchment area	A	20 km^2 to 1×10^6 km^2
9.	Catchment slope	S	0.004 to 0.50
10.	Percent forest area	$(\frac{100 A_F}{A})$	1% to 90%

Table 2 : Correlation Coefficients

Variables	A/LogA	P_{24}/LogP_{24}	S/LogS	F_F/Log F_F
$(Q_{2.33}/A)$	-0.28	0.45	0.59	-0.47
$\text{Log}(Q_{2.33}/A)$	-0.84	0.59	0.84	-0.57

As expected the log-transforms have higher correlation coefficients than the linear ones. Hence for further analysis only log-transforms have been considered. Further, all other independent variables are strongly correlated with $Q_{2.33}/A$. The closeness of association between the dependent variable $\text{Log}(Q_{2.33}/A)$ and each of the independent variable after the effect. of other variables on the dependent variable has first been removed by evaluation of the β -coefficient (14) which is given by

$$\beta_i = b_i \; \sigma_i / \sigma_1$$

where β_i, b_i and σ_i are the β-coefficient, regression coefficient and the standard deviation of the i^{th} independent variable respectively and σ_1 is the standard deviation of dependent variable. The β -coefficients of log-transforms of each independent variable with the dependent variables were obtained through multiple regression analysis (Table 3). A close study of β -coefficients reveal that the catchment area is the most influential factor on the dependent variable followed by the 2 years 24 hours rainfall, forest area and catchment slope. However, all the independent variables strongly influence $Q_{2.33}$; hence none of them can be dropped from analysis without loss of accuracy, for further simplification.

Table 3 : β - Coefficients

Variable	A	P_{24}	S	F_F
β - Coefficient	-0.84	0.71	0.49	-0.45

The equation obtained for $Q_{2.33}$ using all the data is

$$Q_{2.33} = 0.393 \; A^{0.81} \; P_{24}^{0.69} \; S^{0.28} \; A_F^{-0.28} \qquad (2)$$

The multiple correlation coefficient for Eq.(2) is 0.95. The comparison of $Q_{2.33}$ observed and computed using Eq.(2) showed that for 70 percent of the data, error was less than ±50 perçent.

In order to obtain a more accurate relationship, the data were divided according to the catchment area. Establishing different relationships for small catchments (Area $\leq$ 500 km^2), medium catchments (500 km^2 < Area $\leq$ 10000 km^2), the large catchments (10000 km^2 > area), did not result in better accuracy than finally obtained. Then the data were divided according to the geographical locations. Five divisions of the country were made based upon geography viz. Northern, Central, Eastern, Western and Southern India. But this also did not produce relationships giving better accuracy as compared to the final equation discussed below.

It can be mentioned here that in any flood formula the duration of rainfall should be slightly larger than the time of concentration for given catchment. From the consideration of Manning's equation the term A/S should be proportional to the time of concentration. The plot between ($Q_{2.33}$ observed) / ($Q_{2.33}$ computed), using Eq.(2) with A/S supported this notion (See Fig. 1). It suggests that duration of the rainfall term in Eq. 2 should vary with the ratio of catchment area to catchment slope. Hence different values of P were used depending on A/S values. The equation for $Q_{2.33}$ that gave best results is as follows :

$$Q_{2.33} = 0.185\, P^{1.03}\, A^{0.74}\, S^{0.36}\, A_F^{-0.30} \qquad (3)$$

where,

P	=	2 years, 45 minute, rainfall in mm if $A/S \leq 300$
P	=	2 years, 3 hour rainfall in mm if $300 < A/S \leq 3000$
P	=	2 years, 6 hour rainfall in mm if $3000 < A/S \leq 6 \times 10^4$
P	=	2 years, 24 hour rainfall in mm if $A/S > 6 \times 10^4$

A is the catchment area in km^2, S is the catchment slope and A_F is the forest area in the catchment as percentage of catchment area and $Q_{2.33}$ is the mean flood. Comparison of $Q_{2.33}$ values computed using Eq.(3) with $Q_{2.33}$ observed values revealed that Eq (3) gives results within ±30 percent error for all the data, see Fig.(2).

It can be mentioned here that data used in present analysis are from all the parts of the country. Therefore Eq.(3) is having countrywise use. Since the data used in the analysis are from different parts of the country, the accuracy obtained has been considered, adequate.

3.3 Coefficient of Variation

As mentioned earlier, for 43 catchments the flood series had length of minimum 20 years. The data from these catchments were used for establishing the relationship for the coefficient of variation of flood series C_{VF}. Since the length of flood series is more than 20 years the sampling error in the computation of C_{VF} should be small.

The coefficient of variation of flood series is found to vary with the climatic conditions and the catchment area (8,15). Following functional relationship was assumed for the coefficient of variation of flood series.

$$C_{VF} = f(C_{VR}, A)$$

where C_{VR} is the coefficient of variation of maximum monthly rainfall series in the catchment. The term C_{VR} takes into account the yearly deviations in the flood causing rainfall. The relationship obtained for C_{VF} is as follows :

$$C_{VF} = 1.92\, A^{-0.06}\, C_{VR}^{1.16} \tag{4}$$

The coefficient of multiple regression of Eq.(4) is 0.96. The comparision of C_{VF} computed using Eq.(4) with observed values of C_{VF} showed that Eq.(4) gives results within ±25 percent error for all the data (See Fig. 3). As the data used from different parts of the country the accuracy obtained has been assumed to be satisfactory.

CONCLUSION

Testing of the various frequency distributions viz. 2-parameter and 3-parameter Log-Normal, GEV type-1, PT-3 and LPT-3 using data from 43 catchments having minimum record length of 20 years has shown that GEV type-1 distribution can be assumed to be most closely followed by annual peaks of Indian catchments. The statistical parameters of Gumbel distribution viz. mean and the coefficient of variation obtained using the method of maximum liklihood for 93 catchments have been found to be related to the catchment characteristics and the flood producing rainfall characteristics, as shown by Eq.(3) and Eq.(4) which have countrywise applications.

REFERENCES

1. Garde, R.J. and U.C. Kothyari (1987). Flood Estimation Using Regional Flood Frequency Analysis, Hydraulic Engg. Report No. HYD 8702, Civil Engg. Deptt., University of Roorkee, Roorkee.

2. Raynal, J.A. and J.D. Salas (1986). Estimation Procedures for the Type-I Extreme Value Distribution, Journal of Hydrology, Vol. 87, No. 3/4.

3. Wallies, R. and E.F. Wood (1985). Relative Accuracy of Log-Pearson Type-III Procedure, Journal of Hydraulic Div., ASCE, Vol. III, No. 7.

4. Jain, D. and V.P. Singh (1987). Estimating Parameters of EVI Distribution for Flood Frequency Analysis, Water Resources Bulletin, Vol. 23, No.1.

5. Matalas, N.C., Slack, J.R. and J.R. Wallis (1975). Regional Skew in Search of a Parent, Journal Water Resources Research, Vol. II, No.6.

6. Phin, H.N. (1987). A Review of Methods of Parameter Estimation for the Extreme Value Type-I Distribution, Journal of Hydrology, Vol. 90, No. 3/4.

7. Cicioni, G.G. (1973). Best Fitting of Probability Functions to a set of Data for Flood Studies, Second Int. Symp. on Floods and Droughts, Water Resources Publications, Fort Collins, Colorado, U.S.A.

8. Natural Environment Research Council (1975). Flood Studies Report, Vol. I, Hydrological Studies, London, U.K.

9. Schiller, E.J. and J.O. Nagana (1981). Estimation of Flood Peaks with Limited Hydrologic Data, Journal of Hydrology, Vol. 57, No. 11/4.

10. Dicken (1965). Professional Paper on Indian Engineering, Vol.II.

11. Inglis, C.C. (1940). Technical Paper No. 30, Bombay, P.W.D.

12. Varshney, R.S. (1986). Engineering Hydrology, Nem Chand Publishers, Roorkee.

13. Indian Meteorological Department (IMD), Monograph, (1974). Rainfall Frequency Maps of India, No. 6, IMD, New Delhi.

14. Ezkiel, M. (1956). Methods of Correlation Analysis, John Wiely and Sons, New York.

15. Rao, P.R. and R.S. Goel (1986). Flood Frequency Analysis - Assumptions and Limitations Workshop on Flood Estimation in Himalayan Region.

16. Kite, G.W. (1977). Frequency and Risk Analysis in Hydrology, Water Resources Publications Fort Collins, Colorado, U.S.A.

FLOOD ESTIMATION

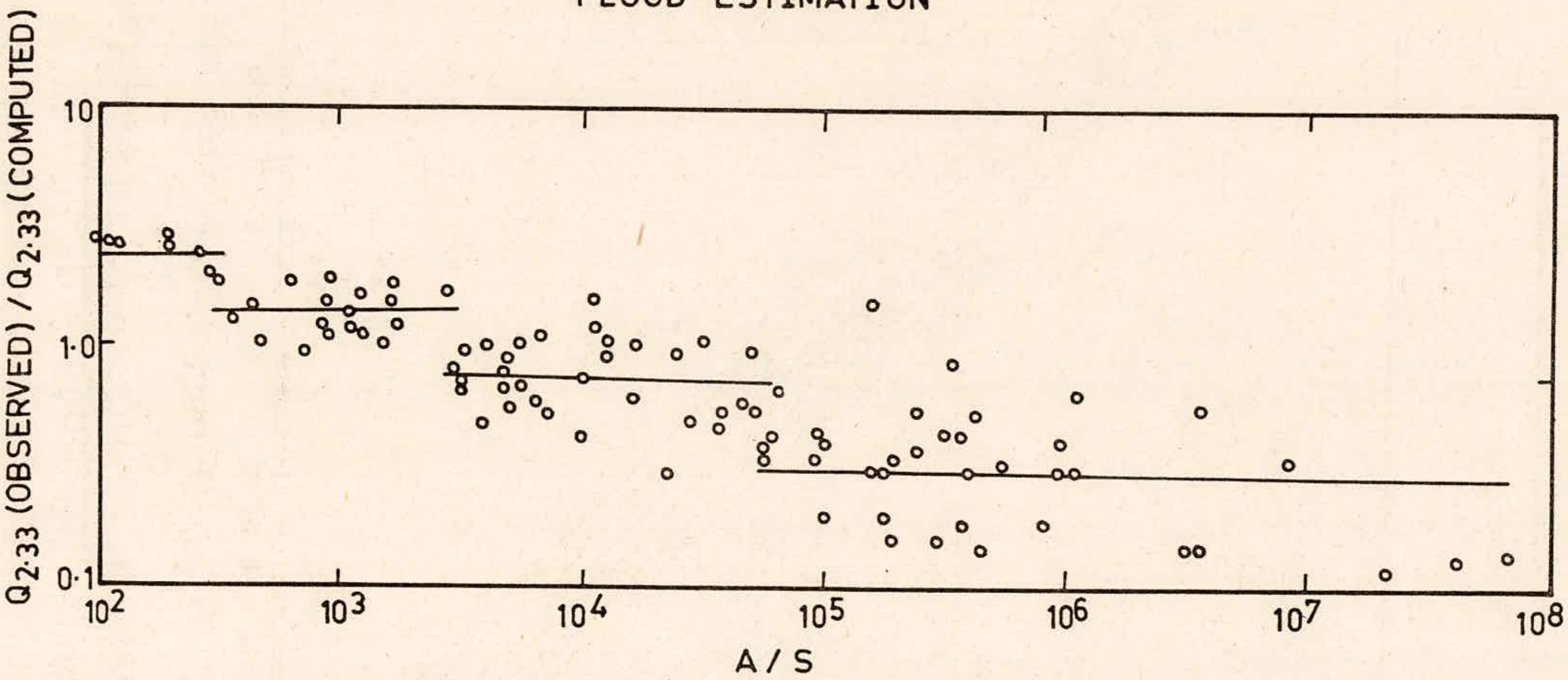

FIG. 1 — EFFECT OF A/S ON $Q_{2\cdot33}$ (OBSERVED)/$Q_{2\cdot33}$ (COMPUTED)

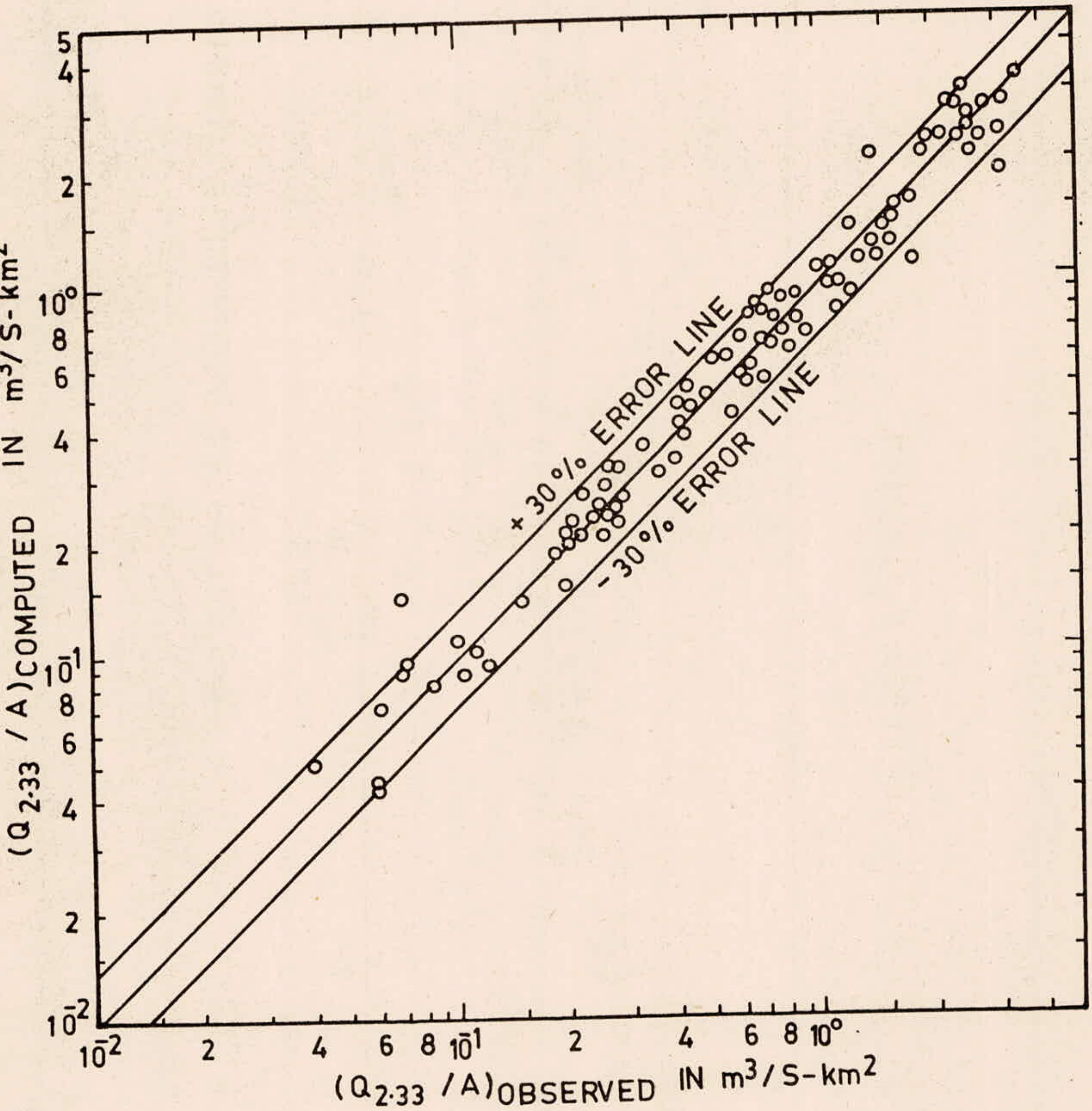

FIG. 2 — COMPARISION OF $(Q_{2\cdot33}/A)$ COMPUTED USING EQ. 3 WITH OBSERVED $(Q_{2\cdot33}/A)$

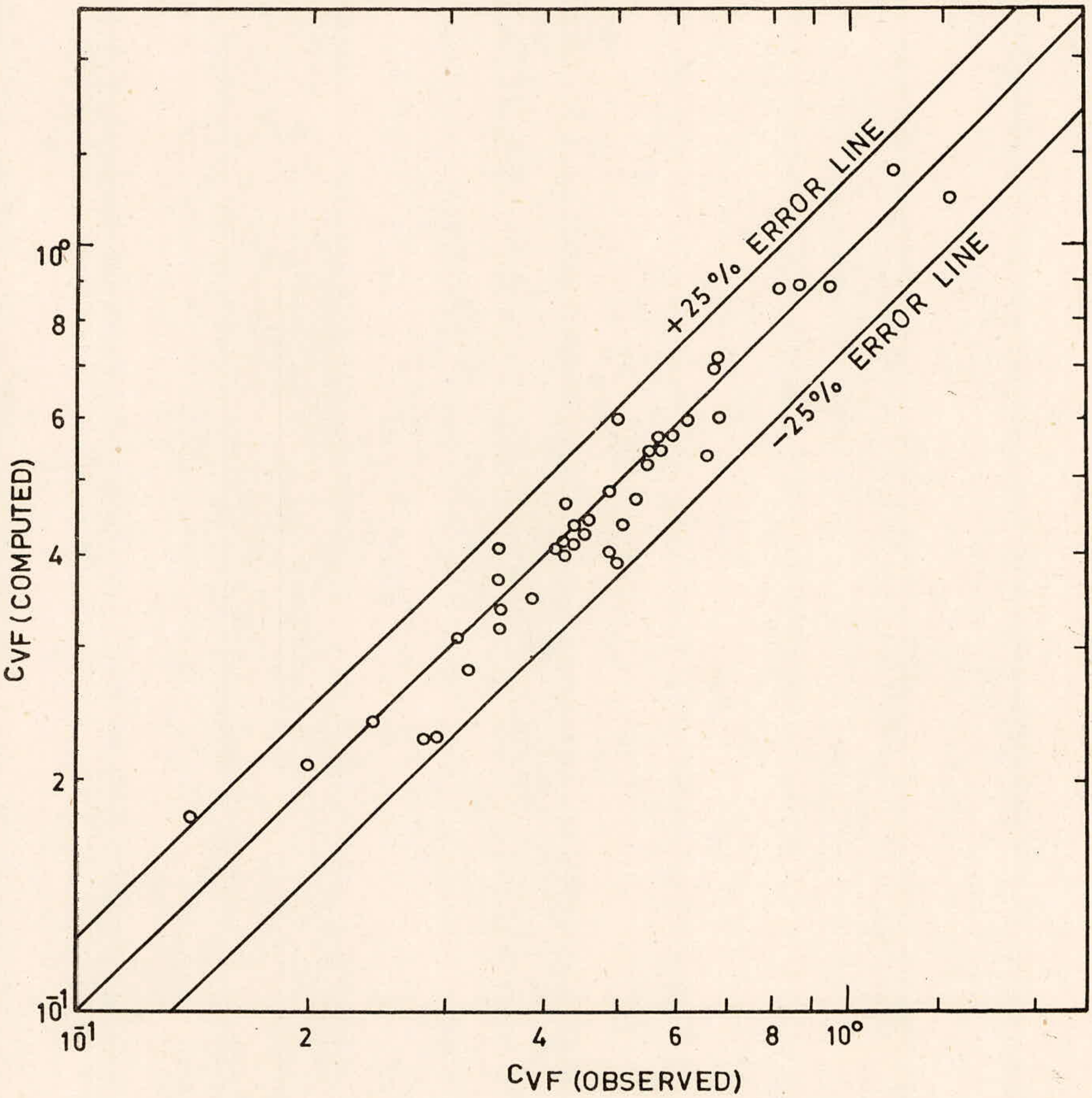

FIG. 3 - COMPARISON OF C_{VF} COMPUTED USING EQ. 4 WITH OBSERVED C_{VF}