Probability Distributions of Best Fit to South African Flood Data

P.T. ADAMSON
[HYDROLOGY DIVISION, DEPARTMENT OF WATER AFFAIRS,
PRIVATE BAG X313, PRETORIA, 0001]

Abstract

Four statistical tests for goodness of fit are applied to fifty sets of South African annual flood peak maxima with a minimum period of record of twenty five years. Five statistical models are proposed of which two are shown to provide equally acceptable results in terms of the criterion of goodness of fit. However, it is shown that this criterion does not imply similarity of performance since estimates of events for predefined return periods are, on the average, widely divergent and particularly at the upper tails of the two favoured models. It is further shown that the Extreme Value Type 1 (Gumbel) distribution should be discouraged as a vehicle for the statistical analysis of South African flood data. Finally some comparisons are made between South African and British data.

Introduction

Central to the design of hydraulic structures and river works is the concept of the design flood, and it is the task of the analytical hydrologist to provide the engineer with an estimate of the frequency and risk associated with a particular extreme discharge. Thus, if a dam spillway is designed to pass the one in two hundred year event then the hydrologist faces the problem of estimating that flood which will occur once, on the average, every two hundred years. Since such problems involve elements of uncertainty the sciences of probability and statistics are invoked for they are concerned with such phenomena in which extract prediction is practically impossible. A probabilistic model must be fitted to the sample of flood data at a site in order to theoretically specify the unknown population properties.

In addition to graphical methods there are three basic objective methods of fitting probability distributions to data samples, namely moments, least squares and maximum likelihood. Of these, the method of moments is the most mathematically tractable although not the most statistically elegant. Comparative studies of the relative efficiencies of the methods are given by Lowery and Nash (1970), Matalas *et al.* (1973) and Bobee and Robitaille (1977) for example.

Each method of fitting has its own protagonists, although for three parameter models there is general agreement that maximum likelihood provides the most statistically efficient results. However, this method can be computationally daunting and requires weighty excursions into the esoteric fields of abstract mathematics (Matalas & Wallis 1973). Consequently for routine analysis moments estimators are to be preferred since a balance must be struck between the statistically efficient, the tractable, and the comprehensible. Certainly, improved moments estimators are available (Sangal and Biswas 1970) and if they are combined with a quantitative statement about their fallibility then there seems to be little reason not to recommend their application in general practice.

Perhaps the most contentious aspect of the statistical analysis of floods is exactly which distribution is THE BEST. In other words which model best fits the data. It is towards this debate that a contribution is made in this paper.

Selected Models

A basic reason behind many studies of probability models of floods and their relative performances in terms of fit, is to arrive at some form of procedural standardization. However, floods are generated by a considerable variety of meteorological phenomena from tropical cyclones to spring snowmelt and it is this causal non-homogeneity which compounds the problem of finding the most efficient model for general practice. Kite (1975) addresses the problem as follows:

> 'There is no general agreement amongst hydrologists as to which of the various theoretical distributions available should be used. Benson (1968) has also indicated that the present state of the art is such that no general agreement has been reached as to preferable techniques and no standards have been established for design purposes. As examples of this divergence of choice: Spence (1973) compared the fit of the Normal, log-Normal, type 1 Extremal and log type 1 Extremal distributions to annual maximum flows on the Canadian Prairies and found that the log-Normal was the best fitting: Cruff and Rantz (1965) compared six probability distributions in California and found that the Pearson type 3 was the most desirable. In other studies, Santos (1970) has found the log-Normal distribution better than the Pearson type 3, and Gumbel (1966) has explained as follows: "It seems that the rivers know the (extreme value) theory. It remains to convince the engineers .... of the validity of this analysis". Benson (1962) has found in a study of 100 long term flood records that no one type of frequency distribution gives consistently better results. In short, no one distribution is acceptable to all hydrologists.'

To add to the apparent confusion the American Water Resources Council, reported on by Benson (1968) concluded that: 'The log-Pearson type 3 distribution has been selected as the base method, with provisions for departures from the base method where justified.' The British Flood Studies Report (1975) favours the General Extreme Value (GEV) distribution, a universal form of the extreme value distributions types 1, 2 and 3, attributable to Jenkinson (1955).

In South Africa there exists no recommended procedure for the statistical analysis of floods and indeed little light has been shed on the performance of the various models in respect of the considerable amount of data available.

The distributions selected for study are given below in terms of their probability density functions.

Two parameter log-Normal distribution (LN2)

$$z = \ln x \tag{1}$$

$$f(z) = \frac{1}{\sigma_z\sqrt{2\pi}} e^{-\frac{1}{2}[(z-\mu_z)/\sigma_z]^2} \tag{2}$$

where μ_z and σ_z are the mean and standard deviation of ln x.

Three parameter log-Normal distribution (LN3)

$$z = \ln(x - x_o) \tag{3}$$

$$f(x) = \frac{1}{(x - x_o)\sigma_z\sqrt{2\pi}} e^{-\frac{1}{2}[(\ln(x-x_o)-\mu_z)/\sigma_z]^2} \tag{4}$$

where μ_z and σ_z are the mean and standard deviation of ln $(x-x_o)$, x_o being the lower bound.

Log-Pearson type 3 distribution (LP3)

$$z = \ln x \tag{5}$$

$$f(z) = \frac{(z-z_o)^{\gamma-1} e^{-(z-z_o)/\beta}}{\beta^\gamma \Gamma(\gamma)} \tag{6}$$

where β and γ are scale and shape parameters respectively, and z_o a location parameter.

Extreme Value type 1 distribution (EV1)

$$f(x_1) = \frac{1}{\alpha} \exp \left[-(x_1-u)/\alpha - e^{-(x_1-u)/\alpha}\right] \tag{7}$$

where 'u' and 'α' are location and scale parameters respectively.

Extreme Value type 2 distribution (EV2)

$$f(x_2) = \frac{1}{\alpha}\{1 - \frac{x_2-u}{\alpha}K\}^{1/K - 1} e^{-[1 - K(x_2-u)/\alpha]^{1/K}} \tag{8}$$

where 'u' and 'α' are as defined above and 'K' is a shape parameter.

All models were fitted by the method of moments as given in the British Flood Studies Report (1975) and discussed with examples of South African data by Adamson (1978). Fifty South African data sets were so fitted to the above models, they have a minimum length of record of 25 years, a mean record length of 46 years and span 2 317 years. No tests of homogeneity were applied largely because the correction of the effects of dams, urbanization, improved land drainage, channel improvements, vegetation changes and improvements to hydrometric structures and ratings would have provided a massive study in itself. Besides the purpose of the exercise is to gain a general insight into the relative performance of the proposed models as applied to floods in South Africa rather than present the evidence for a methodological edict.

Tests of fit

The usual vehicles for testing hypotheses that a sample has been drawn from a specified population are tests of goodness of fit. However, whatever means are used to recommend a model, be they statistical, empirical, or computational attractiveness, they are generally subject to censure. Goodness of fit is a necessary but not a sufficient condition for acceptance, it can illustrate the non-acceptability of a proposed model and provide considerable insight into its relative adequacy. Kite (1975) puts the problem as follows:

> 'If goodness of fit were the only criterion, then high order polynomials would often provide a much better fit than any of the standard distributions, and yet this method is not used because there is no hydrologic justification. The most important criteria in the selection of a distribution are that there be a sound theory describing the phenomenon and that the distribution abstract the maximum information from the data by using proper estimation techniques.'

A number of prominent methods are available for the testing of goodness of fit, the Chi-square and Kolmogorov-Smirnov tests being the most widely used in practice. The other tests available are less well documented and lack a complete statistical theory. Of these the Anderson-Darling and Cramer-von Mises tests have found some application in hydrology (Cicioni *et al.*, 1973) and are selected for use here.

The Chi-square and Kolmogorov-Smirnov tests are dealt with in considerable detail in almost every standard introductory text on statistics, e.g. Kirkpatrick (1974). Methods for computing class intervals for the Chi-square test and critical values of the Kolmogorov-Smirnov statistic in addition to basic computations are given by Adamson (1978) using samples of South African annual floodpeak maxima.

The Anderson-Darling test uses the actual observations without grouping and, most importantly for hypothesis testing and goodness of fit using flood data, it is sensitive to discrepancies at the tails of the distribution rather than near the median. In addition, the test can be weighted to accentuate the differences between the empirical and cumulative probability densities where it is desired to have particular sensitivity. The theoretical derivation of the test criterion is given by Anderson and Darling (1954), and it is defined by:

$$W_n^2 = n \int_{-\infty}^{\infty} [E(Q) - F(Q)]^2 \cdot \Psi \cdot (F(Q)) \cdot dF(Q) \qquad (9)$$

where E(Q) and F(Q) are the empirical and theoretical cumulative distribution functions respectively and Ψ is a non-negative weight function which in this discussion is set to unity. A more convenient form for computation is given by:

$$W_n^2 = -n - \frac{1}{n}\sum_{i=1}^{n} (2i-1)\left[\ln u_i + \ln(1-u_{n-i+1})\right] \qquad (10)$$

where 'ln' is the natural logarithm and $u_i = F(Q_i)$. If W_n^2 is too large then the particular hypothesis under investigation is rejected. Asymptotic significance points are given by Anderson and Darling (1954).

The Cramer-von Mises test is similar to the Anderson Darling test and has similar characteristics. The test statistic is defined by:

$$nw_n^2 = \int_{-\infty}^{+\infty} [E(Q) - F(Q)]^2 \cdot dF(Q) \qquad (11)$$

which reduces to:

$$nw_n^2 = 1/12n + \sum_{i=1}^{n} [u_i - (2i-1)/(2n)]^2 \qquad (12)$$

with definitions as before.

The theoretical development of the test statistic is given by Anderson and Darling (1952) from which critical values of nw_n^2 are also available.

TABLE 1
NUMBER OF TIMES OUT OF 50 THAT DISTRIBUTIONS WERE REJECTED BY THE χ^2 TEST

Significance level	LN2	LN3	LP3	EV1	EV2
0,10	11	32	14	42	43
0,05	8	30	11	42	39
0,01	5	24	4	35	29

TABLE 2
NUMBER OF TIMES OUT OF 50 THAT DISTRIBUTIONS WERE REJECTED BY THE KOLMOGOROV-SMIRNOV TEST

Significance level	LN2	LN3	LP3	EV1	EV2
0,10	1	11	3	26	22
0,05	0	6	0	23	18
0,01	0	4	0	13	15

TABLE 3
NUMBER OF TIMES OUT OF 50 THAT DISTRIBUTIONS WERE REJECTED BY THE ANDERSON-DARLING TEST

Significance level	LN2	LN3	LP3	EV1	EV2
0,10	0	16	0	27	18
0,05	0	10	0	22	17
0,01	0	4	0	12	10

TABLE 4
NUMBER OF TIMES OUT OF 50 THAT DISTRIBUTIONS WERE REJECTED BY THE CRAMER-VON MISES TEST

Significance level	LN2	LN3	LP3	EV1	EV2
0,10	0	14	0	26	18
0,05	0	7	0	20	16
0,01	0	2	0	12	8

Test Results

Tables 1 to 4 show the number of times out of 50 that the selected distributions were rejected using the four goodness of fit tests referred to above.

The most obvious point to be drawn from the analysis is that the LN2 and LP3 models provide the best overall fit. The LN3 model is less well favoured, no doubt a consequence of a poor estimate of the lower bound which is probably directly attributable to the sample error inherent in the skew coefficient. (Sangal and Biswas, 1970.)

The EV1 distribution, as expected, is found to be a most unsuitable model for the analysis of South African flood data. This model has an implied skew coefficient of 1,14 and the further the departure of sample skew from this value the less efficient the model becomes. Figure 1, shows a histogram of the skew coefficients of forty two of the fifty samples investigated. These have a mean of 2,25 and most data sets have a skew considerably in excess of 1,14, implying that the EV1 (Gumbel) will considerably underestimate a flood event for a particular return period. Despite this the EV1 distribution has reached such a level of popularity in hydrology that, as Van Montfort (1970) puts it, the estimation of extreme probabilities seems more influenced by the choice of distribution than by the data itself. Rarely when the EV1 distribution is used is there any statistical support for the choice. As an example the South African Weather Bureau (1974) published extreme value data that has been fitted to the EV1 distribution with no test criterion or justification at all. (Adamson, 1977.)

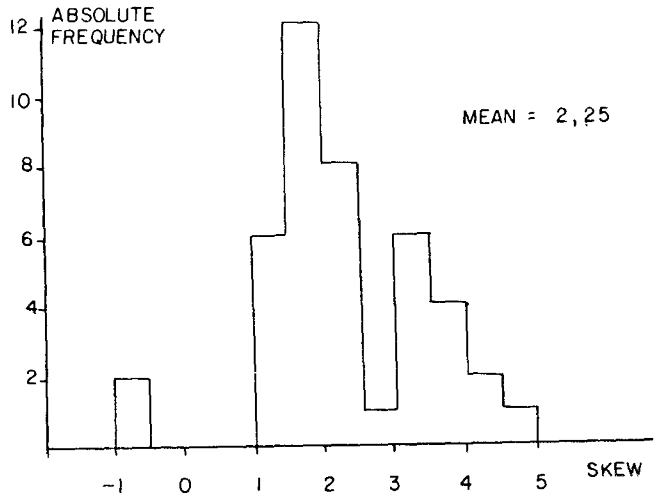

Figure 1
Frequency of skew coefficients for 42 South African data sets of annual flood maxima (N > 30)

Surprisingly perhaps, the EV2 distribution is rejected as a potential model far more times than one might expect for such a flexible distribution. The reason possibly lies with the fact that moment estimators are particularly inefficient and/or the tests of fit are overly influenced by the fact that the lower 37% of a data sample are not distributed as extreme values at all (Jenkinson, 1969).

The performance of the individual tests is interesting. The Kolmogorov-Smirnov, Anderson-Darling and Cramer-von Mises tests are comparable in the number of rejections at each significance level, whilst the χ^2 test rejects a much higher number of cases at all levels. Although the LN2 and LP3 models are conclusively shown to perform better than the others, at least in terms of the test criteria used, the χ^2 test is obviously over discriminatory when compared to alternative tests of fit. The reason is suggested to lie in the loss of information in the grouping of the data which possibly becomes critical for sample sizes of less than $n = 40$.

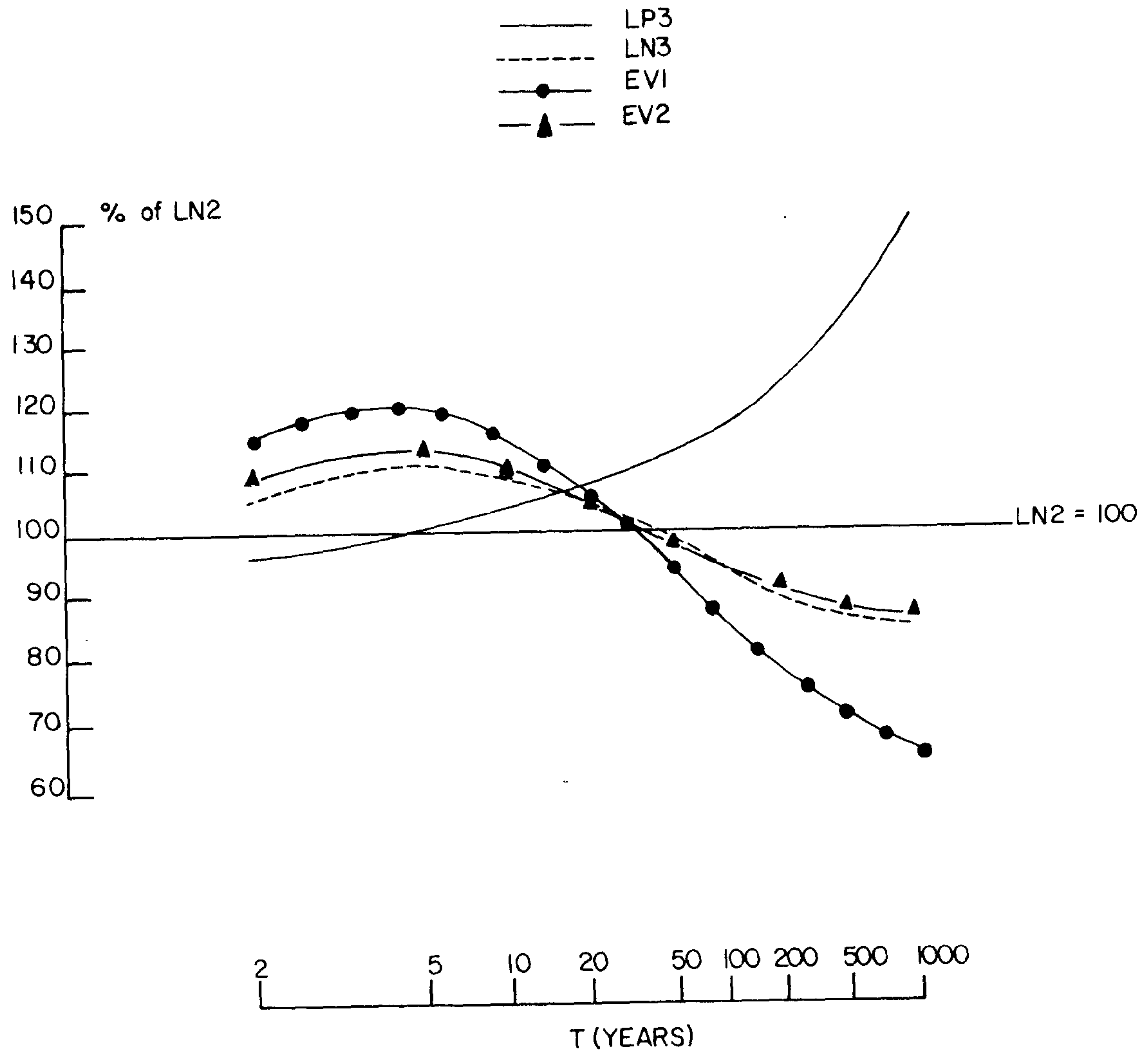

Figure 2
Mean of 50 moment estimates of Q(T) expressed as a percentage of the mean for the LN2 distributions

Comparative performance of models

The real area of interest in the relative merits of models fitted to maxima lies in the fit at the upper tail, and in 1957 Moran wrote:

> 'In the first place the form of the distribution of floods is not known and any distribution used must be guessed. This may have a considerable effect since the part of the distribution we are interested in is well away from the part where the observations provide some information about the shape. It is therefore easy to construct two different distributions both of which fit the observations closely but for which the tails are of quite different shape. This difficulty cannot be surmounted, but we can try to fit several different distributions and see how the choice affects the result. It would also be desirable to use some measure of goodness of fit on the fitted distributions.'

The point made in this quote is quite evident for the South African data so far analysed, for apparently the LN2 and LP3 distributions are equally successful models on the evidence of goodness of fit, but they behave quite differently at the upper tails. This is shown plainly in Figure 2, where the mean estimate for each model and for all fifty samples is expressed as a percentage of the mean estimate using the LN2 distribution. Figure 2 is summarized in Table 5.

From this analysis, with the LN2 as the standard, it is evident that, on the average:

(1) The LP3 model gives significantly higher estimates of floods for return periods in excess of 100 years than the other models. The model is particularly sensitive to high outliers and the sample error in the skew coefficient. More conservative estimates will result if some attempt is made to correct this error.

(2) As expected, the EV1 model is far from satisfactory, grossly underestimating flood peaks with return periods greater than 100 years.

(3) The LN3 and EV2 models are broadly comparable in results, both giving slightly higher estimates than the LN2 model for the more commonly occuring events and lower estimates, on the average, for the rarer maxima.

These results emphasise that the criterion of goodness of fit does not imply similarity of performance, particularly in the upper tails where sampling variability is inevitably at its highest.

South African flood data are in fact characterised by particularly long upper tails in the samples analysed to date; in other words the range of annual floodpeak maxima is much greater than, for example, that which may be expected in a temperate region. To illustrate this point $Q/\bar{Q}$ was calculated for each station and the whole assemblage of 2 317 values for the fifty stations were fitted by moments to the EV2 and LN2 models. The results are compared with a similar curve available for 420 stations in the United Kingdom and presented in the British Flood Studies Report (Volume 1 (1975) p.149). The results are tabulated in Table 6 and plotted in Figure 3.

It is immediately apparent that the ratio between the mean of a sample of flood data and the estimate of the T-year flood is much higher in South African than in Britain. This is to be expected for a largely semi-arid country with distinct wet and dry seasons where the intensity-duration-frequency structure of flood generating storms is quite different and where there is inevitably a larger range in antecedent catchment conditions from 'saturated' to 'parched'. In short the range of sample floods at a site, as defined by their coefficient of variation, is particularly great as is evidenced in Table 7 where the basic statistics of $Q/\bar{Q}$ are compared for South Africa and Great Britain.

These results predicate that high outliers are the rule rather than the exception in South African samples of flood data. Consequently those distributions which are particularly sensitive to high outliers, such as the LP3 model, must be used with singular caution and some objective attempt must be made to correct the sample error in the estimate of the skew coefficient. Furthermore, whatever the model used the most refined moments estimators must be invoked, since a failure to do so is tantamount to throwing away a significant portion of the information contained in the sample.

TABLE 5
MEAN OF 50 MOMENT ESTIMATES OF Q(T) EXPRESSED AS A PERCENTAGE OF THE MEAN OF 50 FITTED LN2 DISTRIBUTIONS

Distribution	Return Period: T (years) 2	5	10	20	50	100	200	500	1 000
LN2	100	100	100	100	100	100	100	100	100
LN3	105	111	108	104	99	95	91	85	85
LP3	96	101	104	108	114	115	126	135	150
EV1	115	121	114	105	94	86	79	70	67
EV2	109	11[illegible]	109	105	99	94	91	86	85

TABLE 6
RELATIONSHIP BETWEEN Q/$\bar{Q}$ AND T (YEARS) AFTER FITTING THE LN2 AND EV2 TO 50 SOUTH AFRICAN STATIONS AND FITTING THE EV2 TO 420 BRITISH STATIONS

Distribution	T (years) 2	5	10	20	25	50	100	200	500	1000
South Africa										
LN2	0,6	1,4	2,1	3,0	–	4,4	5,7	7,3	9,6	11,7
EV2	0,7	1,6	2,3	3,1	–	4,3	5,4	6,6	8,4	10,1
Great Britain										
EV2	0,9	1,2	1,5	–	1,9	2,2	2,6	3,1	3,8	4,4

TABLE 7
COMPARATIVE STATISTICS FOR $Q/\bar{Q}$

	Mean	Standard deviation	Skew	Coefficient of variation
South Africa	–	1,19	3,40	1,20
Great Britain	–	–	3,54	0,44

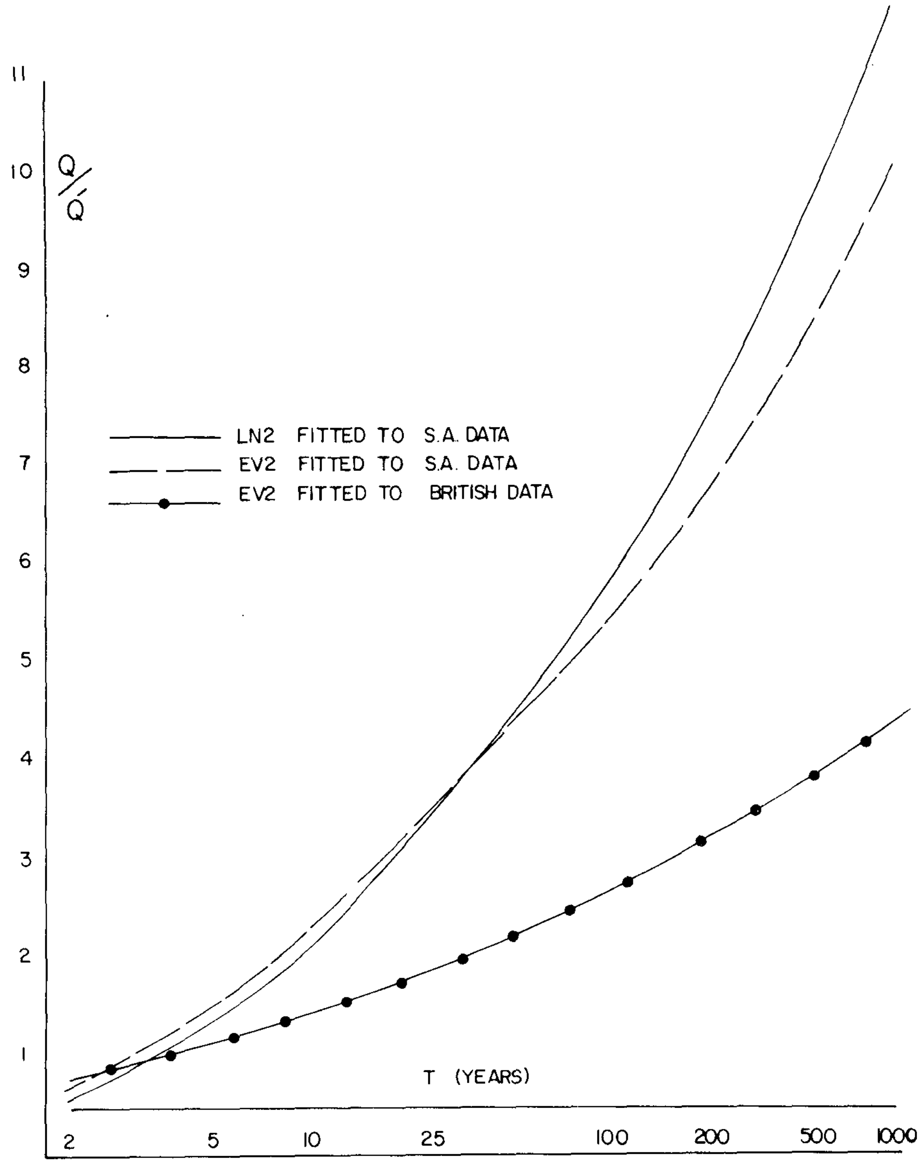

Figure 3
Region curve based on all S.A. records (50) and all British records (420)

Conclusions

(1) The Gumbel (EV1) distribution should be discouraged as a means of statistical annual flood peak analysis in South Africa as the implied skew of 1,14 is a limitation that gives estimates that are generally too low.

(2) The log-Normal 2-parameter and log-Pearson type 3 models, on the evidence of statistical methods of fit alone, provide the 'best' models.

(3) The log-Pearson type 3 model gives much higher estimates, on the average, for return periods greater than 100 years than the log-Normal 2-parameter model.

(4) On the average the log-Normal 3-parameter and Extreme Value type 2 models, when fitted by moments give similar results.

(5) Two parameter models have the limitation of a fixed skew, for example the Extreme Value type 1. However, the log-Normal 2-parameter model when fitted by moments has the versatility to overcome this limitation. (Chow, 1954.)

Finally, on the evidence it is difficult to make any recommendations for standard practice in South Africa. The insight gained in this investigation has been of a general nature. For routine analysis moment estimators, with suitable corrections, are attractive and the log-Normal 2-parameter model is evidenced as worthy of consideration. This said, the process of cognition in hydrology must combine the statistical evidence with sound engineering judgement.

References

ADAMSON, P.T. (1977). Extreme values and return periods for rainfall in South Africa. *Department of Water Affairs. Division of Hydrology. Technical Note No* **78.**

ADAMSON, P.T. (1978). The statistics of extreme values. The analysis of floods in South Africa. *Department of Water Affairs. Research Report No* **4/78.**

ANDERSON, T.W. and DARLING, D.A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *Annals of Mathematical Statistics.* **23** 193–212.

ANDERSON, T.W. and DARLING, D.A. (1954). A test of goodness of fit. *Journal of the American Statistical Association.* **49** 765–769.

BENSON, M.A. (1962). Evolution of methods for evaluating the occurence of floods. *USGS Water Supply Paper.* **1580-A.**

BENSON, M.A. (1968). Uniform flood frequency estimating methods for Federal Agencies. *Water Resources Research.* 4(5) 891–908.

BOBEE, B.B. and ROBITAILLE, R. (1977). The use of the Pearson type 3 and Log-Pearson type 3 distributions revisited. *Water Resources Research* 13(2) 427–443.

BRITISH FLOOD STUDIES REPORT. (1975). Published in four volumes by the *Natural Environment Research Council.* Charing Cross. London.

CHOW, V.T. (1954). The log-probability law and its engineering applications. *Proceedings American Association of Civil Engineers.* **80** 1–25.

CICIONI, G., GIULIANO, G. and SPAZIANI, F.M. (1973). Best fitting probability functions to a set of data for flood studies. *Proceedings Second International Symposium in Hydrology, Floods and Droughts.* Water Resources Publications, Fort Collins. Colorado. 304–314.

CRUFF, R.W. and RANTZ, S.E. (1965). A comparison of methods used in Flood frequency studies for coastal basins in California. *USGS. Water Supply Paper.* **1580-E.**

GUMBEL, E.J. (1966). Extreme Value Analysis of hydrologic data. *Proceedings Hydrology Symposium No. 5.* NRC. Ottawa. 147–169.

JENKINSON, A.F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society.* **87** 158–171.

JENKINSON, A.F. (1969). Estimation of maximum floods. Chapter V. *World Meteorological Office.* Technical Note No 98.

KIRKPATRICK, E.G. (1974). *Introductory statistics and probability for engineering, science and technology.* Published by Prentice Hall. New Jersey.

KITE, G.W. (1975). Confidence limits for design events. *Water Resources Research.* **11** (1) 48–53.

LOWERY, M.D. and NASH, J.E. (1970). A comparison of methods of fitting the double exponential distribution. *Journal of Hydrology.* **10** 259–275.

MATALAS, N.C. and WALLIS, J.R. (1973). Eureka! It fits a Pearson Type III distribution. *Water Resources Research.* **9** (2) 281–289.

MORAN, P.A.P. (1957). Statistical treatment of flood flows. *Transactions American Geophysical Union.* **38** 519–523.

SANGAL, B.P. and BISWAS, A.K. (1970). The 3-parameter log-Normal distribution and its application in hydrology. *Water Resources Research.* **6** (2) 505–515.

SANTOS, A. (1970). The Statistical treatment of flood flows. *Water Power.* **22** (2) 63–67.

SOUTH AFRICAN WEATHER BUREAU. (1974). Climate of South Africa. Part II. Extreme values of rainfall, temperature and wind for selected return periods. *Weather Bureau publication no 36.* Pretoria.

SPENCE, E.S. (1973). Theoretical frequency distributions for the analysis of Plains streamflow. *Canadian Journal of Earth Sciences.* **10** 130–139.

VAN MONTFORT, M.A.J. (1970). On testing that the distribution of extremes is of type 1, when type II is the alternative. *Journal of Hydrology.* **11** 421–427.