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

UNCERTAINTIES IN PROBABILISTIC MODELLING OF FLOODS

F. T. Mathew
Senior Research Officer

M. M. Hasija
Research Officer

Central Water and Power Research Station, Pune

SYNOPSIS

The paper discusses the various components of hydrologic and hydraulic uncertainties that combinedly constitute the hydraulic risk of a water resources structure; and specifically examines in detail model and parameter uncertainties inherent in generally used statistical extrapolation procedures for estimation of extreme flood magnitudes for different design-frequencies. The study illustrates the use of moment-ratio diagram, wherein coverage areas/points of distributions commonly used for analysis of hydrologic data are shown, for reducing model uncertainty. Further, the paper presents a procedure, which can potentially reduce parameter uncertainty of the identified model by developing generalised values of the parameters through the use of a weighting between sample and regional parameter estimates. A numerical example, reported in the paper, illustrates the general applicability of the procedures detailed.

1. INTRODUCTION

In engineering studies concerned with the design and certification of water resources projects, it is essential to have estimates of extreme flood conditions expected during the lifetime of the structure. The problem is very often solved by a suitable statistical extrapolation procedure involving the recorded data of extreme events. The existence of uncertainties in such procedures has long been recognised.

The uncertainties afore-mentioned can be classified as : i) the uncertainty due to the inherent randomness of the hydrologic process(es) giving rise to the flood-events, ii) the model uncertainty resulting from the choice of the probability-distributional form chosen from the potentially representative ones, and iii) the parameter uncertainty resulting from the method of computation of them, as well as the limited data from which the parameter estimation is done.

In a design-situation, uncertainties of the type aforesaid are termed hydrologic uncertainties. This paper presents certain procedures observed useful for analysing, and consequently reducing, model and parameter uncertainties in probabilistic modelling of flood-data.

The paper presents a graphical method, which can aid in the identification of probability-functions to model flood-data by combining conceptual (known theory-based) and empirical (data-based) approaches to statistical flood-modelling. Further, a procedure is presented which can potentially reduce parameter uncertainty of the identified model by developing generalised values of the parameters through the use of a weighting between sample and regional parameter estimates.

2. HYDRAULIC RISK

Hydrologic uncertainties, together with hydraulic uncertainties, constitute the hydraulic risk [1,2] of a structure for water resources utilisation. The hydraulic uncertainties arise from such factors as : i) operational conditions of flows imposed on the structure, ii) material used, and construction methods, iii) geotechnical factors, iv) model uncertainties that can result from the use of hydraulic model(s) to describe the flow conditions through or over the structure, and v) loading due to flood flows, and the interaction between hydrologic loading and hydraulic resistances of the structure.

The uncertain features of a structure as a whole is very difficult in practice to determine. However, it is easy and economical to analyse the uncertainty features of the different constituents of hydraulic risk of a structure while making decisions in design situations.

The problem of determining one extreme flood-value for specified design-frequency is often solved by extrapolating through statistical distributional modelling procedures,measured data to events that have not been measured. The present study restricts itself to the model and parameter uncertainties inherent in such procedures.

3. DATA BASE

The data-base available for flood analysis is generally limited. For large river basins in the country, a common observation is that the flood-data available for frequency analysis rarely extend over 20 to 30 years. A typical example is that of Narmada river, the data in respect of which have been made use of for exemplification in the present study. There are only three sites, namely Jamtara, Mortakka, and Garudeshwar on the main river for which the recorded flood data are available for more than thirty years. For three more sites, namely Barmanghat, Hoshangabad and Mandleshwar, data for 10-12 years are available.

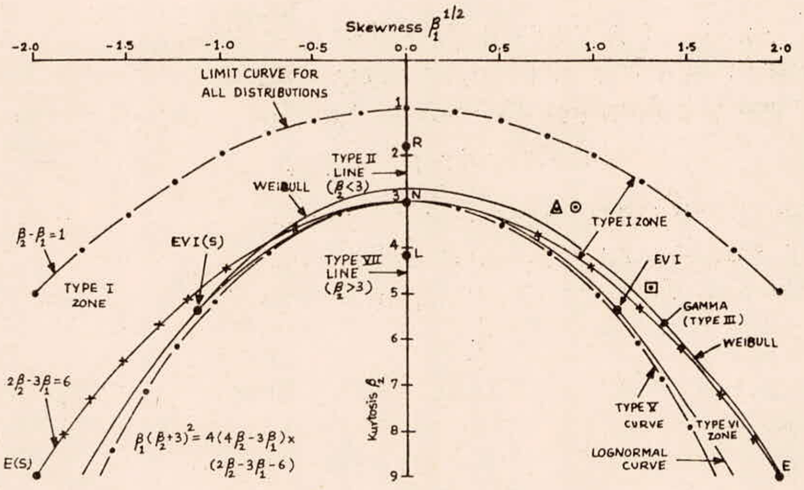
4. MODEL IDENTIFICATION

In empirical approach to flood modelling, model identification is based on the posterior information from the sample data. Various statistical parameters involving the cumulants can aid this process. A drawback of this method is that the variance of estimates of parameters can be large for samples that are generally small. This brings to the fore the importance of model selection in conjunction with the prior information emanating from an understanding of the mechanism giving rise to the data and the known theory thereof.

In conceptual approach to flood modelling, the model identification is done based on a rational-theoretical analysis of the phenomenon taking into account its characteristics of evolution. Some standard distributions have been identified to be good in representing extreme flood events. Some of the common distributions frequently used for flood modelling are : Lognormal, Extreme Value Type I (EV I), Exponential, Pearson Type I-VII, and Log Pearson Type III (LP III).

Graphical methods are considered as the single-most efficient, robust statistical tool. They play an important role in all aspects of statistical investigation - from the beginning exploratory plots, through various stages of analysis. Graphical techniques can come to the aid in the identification of probability functions of flood-flows from field data, and in assessing functional adequacy.

An eclectic model selection procedure in flood analysis is to combine the prior information with the posterior information obtained from observed data. A possible way of doing this is by using the moment-ratio diagram, wherein coverage areas/points of distributions frequently used for flood modelling are shown, to locate that distribution which best satisfies the criteria dictated by the posterior and prior information bases. Figure 1 gives such a diagram.



LEGEND

AFS OF NARMADA AT:

JAMTARA △
MORTAKKA ⊡
GARUDESHWAR ⊙

R - RECTANGULAR
N - NORMAL
EV I - EXTREME VALUE TYPE I (GUMBEL)
L - LOGISTIC
E - EXPONENTIAL
EVI(S) - EVI (SMALLEST)
E(S) - E (SMALLEST)

FIG.1 : MOMENT-RATIO DIAGRAM OF KURTOSIS AGAINST SKEWNESS WHICH GIVES COVERAGE AREAS/POINTS OF SOME COMMON DISTRIBUTIONS. ESTIMATED VALUES OF KURTOSIS AND SKEWNESS FOR HISTORICAL FLOOD DATA OF NARMADA RIVER ARE ALSO SHOWN.

Figure 1 is based on the theory that for many common distributions, coefficients of skewness ($\beta_1^{1/2}$) and kurtosis (β_2) are either constant, or plot as a curve/line in the ($\beta_1^{1/2}$, β_2) plane [4]. Additional distributions like Log Pearson Types I-III can be included in the model identification process by calculating the respective statistical parameters for the log-transformed data series, and conducting checks to see where the point falls.

5. REDUCING PARAMETER UNCERTAINTY

This section gives a methodology that can be used in hydrologic frequency analysis to reduce the parameter uncertainty. The methodology, briefly stated, consists of using a weighting between sample and regional parameter estimates. The weights are functions of the variances of the sample parameters and regionalised parameter values.

The regionalised parameter estimates are computed by relating, through the use of multiple regression equations, the individual station parameters for the chosen statistical model to the physiographic and meteorologic characteristics of the basin such as contributing area, channel slope, channel length, and stream frequency (defined as the number of stream junctions per square kilometre of the catchment). If θs is a sample parameter estimate and θr the regional parameter estimate, the weighted parameter θw is defined as $\theta w = W\theta s + (1-W)\theta r$, where $W = Var(\theta r)/[Var(\theta s) + Var(\theta r)]$. $Var(\theta s)$ and $Var(\theta r)$ represent the variances of θs and θr respectively . These variances are estimated by the procedure described below.

$Var(\theta s)$ is computed by Jackknife method, which is a nonparametric method. The method consists in : estimating the parameter θs using the N sample observations, and also computing $\theta_s^i (i = 1,2...N)$ from N-1 observations with the ith observation deleted from the data set. The Jackknife variance of θs is given by $Var(\theta s) = [(N-1/N)[\sum_{i=1}^{N} (\theta s^i - \theta s)^2]$.

The variance of θr is derived from a multiple regression analysis linking θr to various catchment characteristics. If the regression model is written as $Y = X\beta + \epsilon$, where Y is an nx1 vector of response variables, X is an nxk matrix of k independent variables, β is a kx1 vector of unknown regression parameters, and ϵ is an nx1 vector of random errors. If $e = Y - X\hat{\beta}$, where $\hat{\beta} = (X'X)^{-1} XY$, then the variance of Y given X is $\hat{\sigma}^2 = e'e/(n-k)$. This theoretical procedure can be used in estimating the variance of regionalised parameter θr.

6. EXEMPLIFICATION

In Figure 1, the points representing the estimated values of skewness ($b_1^{1/2}$) and kurtosis (b_2) for the data on observed floods of Narmada river at the sites Jamtara, Mortakka, and Garudeshwar are also shown. The three parameter-points lie in the Pearson type I (P I) zone. As such, P I is an initial choice of the probability distribution to model the flood data at each of the three sites. Further, the three ($b_1^{1/2}$, b_2) points for the data studied are also seen to lie close to the EV I point. A statistical hypothesis testing involving $b_1^{1/2}$ and b_2 showed that the deviations from the respective theoretical values were not significant at 95 per cent confidence level in each of the three cases. Hence EV I was chosen as an alternative model. The test procedure [5] used is this : $b_1^{1/2}$ and b_2 evaluated from sample data drawn from the EV I distribution are distributed around the values 1.14 and 5.40 with respective standard errors of $5.63/\sqrt{n}$ and $41.00/\sqrt{n}$. Choice of the EV I and P I distributions to serve as alternative models to characterise observed flood data of Narmada river is thus based on the mechanism giving rise to the data, and examination of the observations themselves. In the present study, for illustrative purposes, the EV I distribution was fitted to the annual flood series (AFS) of Narmada at the Garudeshwar site.

The sample parameter estimates $\hat{\alpha}_s$ and $\hat{\beta}_s$ for the model were obtained as 0.09244 and 23.51371 respectively. The variances of the parameters were determined by using the Jackknife method. Var ($\hat{\alpha}_s$) was obtained to be 0.0001862, and that of Var ($\hat{\beta}_s$), 3.9917741.

Using the physiographic and hydrologic data of the basin pertaining specifically to six gauging sites on the main river, the generalised parameter estimates $\hat{\alpha}_r$ and $\hat{\beta}_r$ for the EV I model were computed. The factors considered are : contributing area (A), channel length (L), channel slope (S), and stream frequency (F).

$\hat{\alpha}_r$ and $\hat{\beta}_r$ were linked to the various catchment characteristics as described earlier. The multiple regression equations were obtained to be :

$$\hat{\alpha}_r = -0.60694 + 0.14885 \times 10^{-4} A - 0.10650 \times 10^{-2} L + 0.445558 S + 0.18306 \times 10^{3} F \text{ and}$$

$$\hat{\beta}_r = 0.58848 - 0.47296 \times 10^{-4} A + 0.41790 \times 10^{-2} L - 0.18635 \times 10 S - 0.14574 \times 10^{5} F.$$

Variances of $\hat{\alpha}_r$ and $\hat{\beta}_r$ were obtained as 0.0045 and 0.5714. The weights w_α and w_β for generalised parameter estimation of $\hat{\alpha}_\omega$ and $\hat{\beta}_\omega$ were obtained to be 0.96027 and 0.12522. The weighted parameters $\hat{\alpha}_\omega$ and $\hat{\beta}_\omega$ are : 0.09258 and 23.3993.

In flood modelling, an often-adopted practise is to report the values of several quantiles so as to give an idea of the range of values the flood can take for various probabilities of nonexceedance. The pth quantile is defined as a value x such that : $\Pr[x < x_p] \leqslant p \leqslant \Pr[x \leqslant x_p]$, where x represents the random variable. The quantiles are obtained by the solution of $F(x_p) = p$. Frequently reported quantiles in flood analysis are : upper and lower quantiles $x_{.25}$ and $x_{.75}$, mean annual flood $x_{.571}$, 50-year flood, 100-year flood, and 1,000-year flood.

For Narmada river, different quantiles were worked out using sample-parameter estimates and generalised parameter estimates. The following table gives the results.

Quantile (X_p)	Return Period (Yr)	Flood-estimate($10^3 m^3/s$) given by: Sample parameters	Generalised parameters
$X_{.25}$	1.33	19.922	19.813
$X_{.5}$	2	27.479	27.358
$X_{.571}$	2.33	29.773	29.649
$X_{.96}$	25	58.115	57.948
$X_{.98}$	50	65.724	65.546
$X_{.99}$	100	73.277	73.088
$X_{.999}$	1,000	98.234	98.007

The values of the various quantiles given by the above table gives an idea of the range of values the flood can take in the watershed studied. The values given by the sample-parameters and the generalised parameters are seen to be close to each other, with the maximum variation being less than five percent. This result shows that the parameter uncertainty in fitting the EVI model to Narmada flood data is minimal. Though the sample-parameters and generalised parameters in the particular case studied in the paper gives flood-estimates that are close to each other, theoretically the values given bv the regional parameters are preferable on account of the larger data-base that have gone into them.

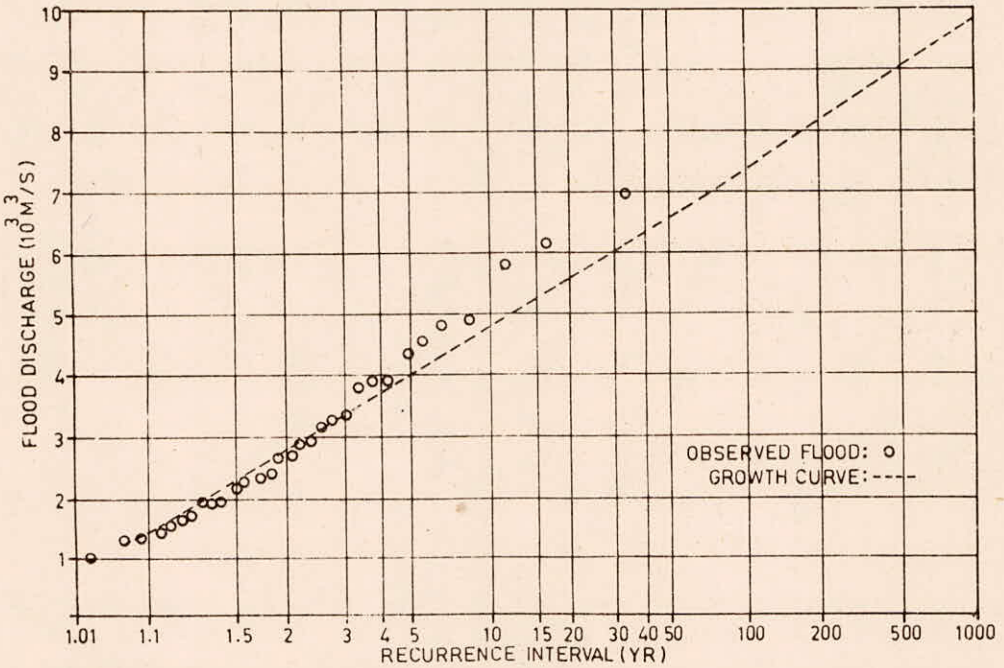


FIG. 2: PROBABILITY PLOT OF OBSERVED FLOODS OF NARMADA AT GARUDESHWAR VERSUS GROWTH CURVE GIVEN BY FLOOD ESTIMATES FOR DIFFERENT RETURN PERIODS COMPUTED FROM GENERALISED PARAMETERS

The growth-curve given by the flood estimates for different return periods was observed to agree well with the trend exhibited by the data-points corresponding to observed floods, when plotted on a probability-plot. Figure 2 gives this probability plot.

7. CONCLUDING REMARKS

The present study illustrated the use of moment-ratio diagram, wherein coverage areas/points of distributions commonly used for flood analysis are shown, in reducing model uncertainty in flood modelling. A method for reducing parameter uncertainty by regionalisation using physiographic and meteorological characteristics of the basin contributing to the flood events. A case study is also presented, together with the results, which demonstrates the applicability of the procedures suggested to any watershed for which the data-availability is good.

REFERENCES

1. Plate, Eric J., 'Extreme Value Models', Proc.of the International Symposium on Risk and Reliability in Water Resources' held at University of Waterloo, Water Resources Publications, Colorado, 1979.

2. Tung, Yeon-Kong,and Mays, Larry W., 'Optimal Risk-Based Hydraulic Design of Bridges', Jnl. of Water Resources Planning and Management Division, ASCE, Vol. 108, No. WR2, June 1982. pp 843-852.

3. Mathew, F.T., and Rao, B.C. , 'An Eclectic Approach to Statistical Modelling of Flood Events', Proc. of the Fifth Annual Convention and Seminar on Hydrology held by the Association of Hydrologists of India at Bhopal, 15-17 July 1986.

4. Miller, Rupert G., 'The Jackknife - a Review', Biometrika, Vol. 61, No.1, April 1974, pp 1-15.

5. Tung, Yeon-Kong,and Mays, Larry W., 'Reducing Hydrologic Parameter Uncertainty', Jnl. of the Water Resources Planning and Management Division, ASCE, Vol. 107, No. WR1, March 1981, pp 245-262.