CHAPTER 7

PROBABILITY AND RISK

DESIGN STORMS

Storms and floods are unpredictable. The magnitude and frequency of floods cannot be calculated in advance, although a statistical assessment is possible. The question thus arises as to what discharge rate to design any drainage structure for. The design flow may seldom, if ever, occur during the life of the structure. Fortunately the flow rate at other times will usually be less than the capacity of the system. For extreme events the drain may overflow. This may cause damage or inconvenience and is to be avoided. The decision as to what discharge to design a structure for is usually based on an economic risk analysis. The cost of a larger structure is balanced against the probable cost of damage due to larger floods than the structure can accommodate.

The probability of different storm magnitudes and durations occuring must be obtained from a statistical analysis. The ideal situation for a hydrological analysis would be one where a continuous flow record over many years was available. A direct analysis of peak flows could then yield probabilities of different flow rates at the site.

Unfortunately adequate flow records are seldom available at the site in question, or even anywhere in the catchment in question. Even if there were historical records, it is likely that the catchment has undergone, or will undergo, changes in surface cover and drainage patterns. The hydrologist will therefore need to resort to rainfall records, and from these synthesize the necessary runoff pattern at the site of the proposed drain, culvert of conduit. Regional or local rainfall records are usually come by and are independent of development in the basin. There may, however, be a change in the method of recording at some stage, and tests for stationarity in the records should be done.

Many standard hydrological techniques, such as the rational method, the Lloyd-Davies method and the tangent method, are based on the assumption that the probability of a storm of a particular duration is the same as that of the runoff computed using the method. Any discrepancy in the correspondence is supposedly built into the runoff coefficient. The assumption is suspect as there are many variables likely to affect runoff besides the storm intensity. These include antecedent moisture conditions in the catchment, including the state of the surface or retention storage and soil saturation. The storm distribution in space and its variation in time will also affect the peak rate of runoff. The length of record will also affect the distribution pattern especially if recurrence intervals are extrapolated. Beard (1978) indicated that risk is traditionally underestimated with short records.

There is a tendency to design drainage structures to discharge safely the flood which will be exceeded on an average of once in a specified number of years. Thus mean storm return periods of 2, 5, 10, 20, 50 or 100 years, depending on the severity of exceedance, are often used as the basis for the design flood. In general areas subject to more extreme events plan for more remote possibilities. Whereas such rules of thumb are useful design guides it is invariably worthwhile selecting a probability of exceedance based on hydro-economic risk analysis.

PROBABILITY DISTRIBUTIONS

Although rainfall and runoff are to some degree random they have limits imposed by the climate and environment and even follow some trend. Rainfall or storm intensity generally follows a distribution pattern with a mean and variation. Total storm precipitation depth for any selected storm duration can be correlated with probability as in Fig. 7.1.

Rainfall follows a more distinctive probability distribution than runoff. Runoff or stream flow can be described with a distribution curve but the parameters are more complex than for storm distribution. No mathematical expression can be fitted to runoff distribution or even the relationship between runoff and rainfall, on account of such factors as antecedent ground water conditions, alternative sources of flow (groundwater, surface runoff, unnatural discharges), changing land use, storage-discharge relationships for the catchment, and complex topography, all of which result in a non-linear rainfall-runoff relationship

Various mathematical approximations have been attempted to fit flood discharges (and, not of interest in this context, drought flows). Instantaneous runoff has some frequency distribution (e.g. Fig. 7.2) which may or may not coincide with a mathematical function. Parameters which describe mathematical or other distributions are indicated below. The storm drain designer is primarily interested in the upper extreme values of flow.

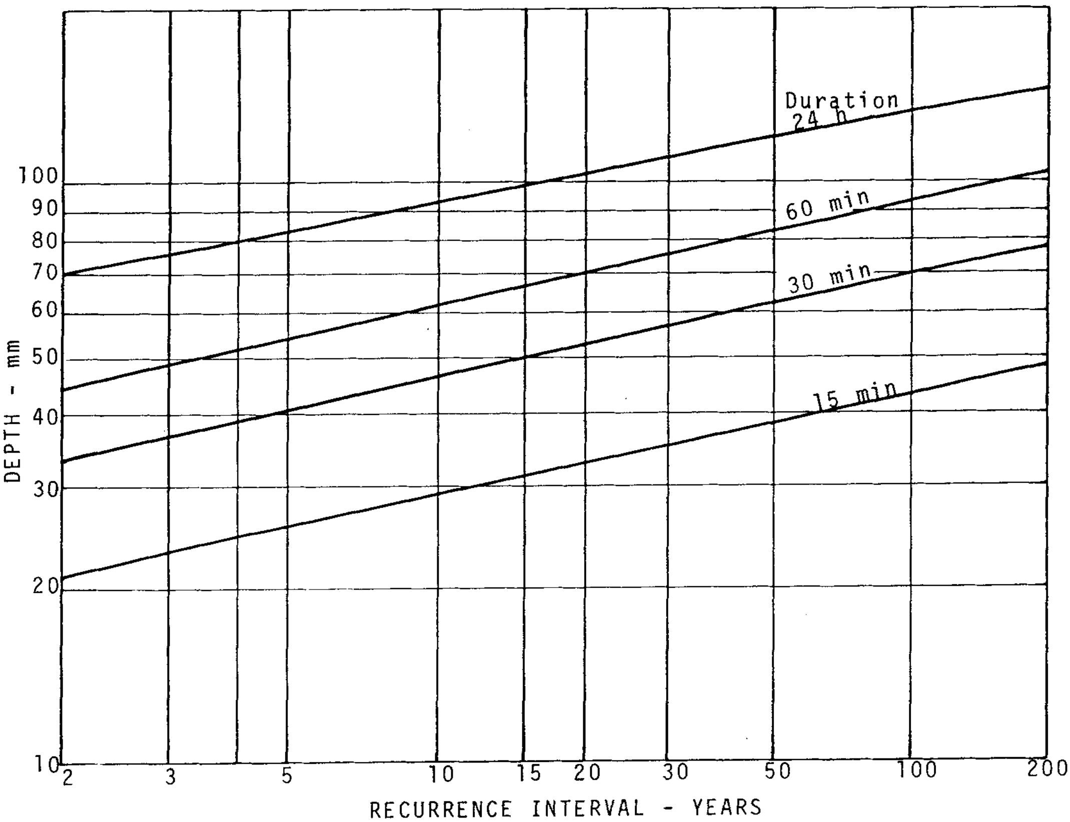


Fig. 7.1 Rainfall depth - Duration - Frequency relationships

In the expressions, the following terms are employed.

Arithmetic mean or mean: The centre of gravity of the distribution. The population mean is designated μ, and the sample mean $\bar{x}$.

Polulation mean $\mu = \int x dp$ (7.1)

Sample mean $\bar{x} = \frac{\Sigma x}{N}$ (7.2)

where x is the variate and N the number of observations and p is the expectance of x.

Median: Middle value of variate, such that it divides the frequency distribution into two equal portions.

Mode: Value of variate which occurs most frequently.

Standard deviation: A measure of variability. Variance is the square of the standard deviation.

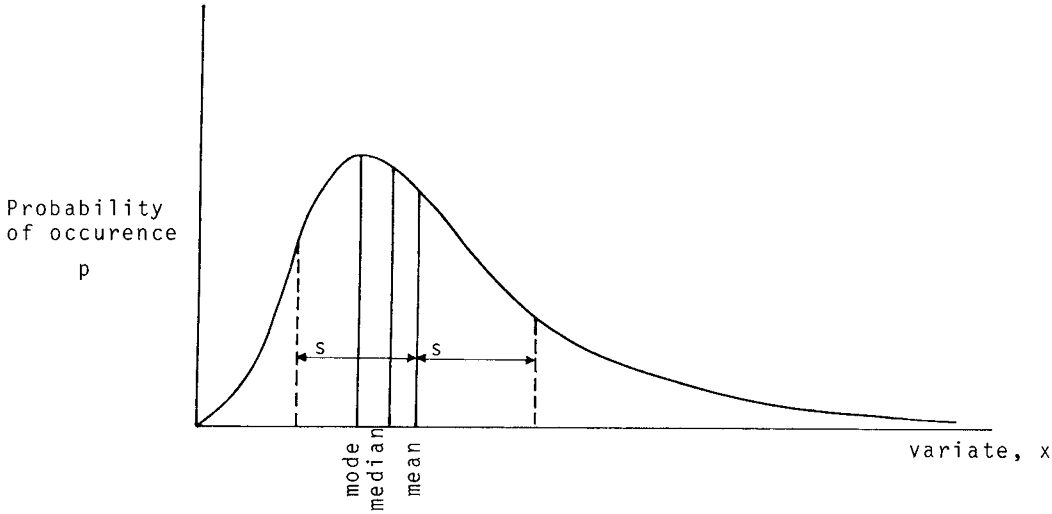


Fig. 7.2 A probability distribution

Polulation standard deviation $\sigma = \sqrt{\frac{\Sigma(x-\mu)^2}{N}}$ (7.3)

Sample estimate $s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{N-1}}$ (7.4)

Skewness:

Polulation skewness $\alpha = \frac{1}{N}\Sigma(x-\mu)^3$ (7.5)

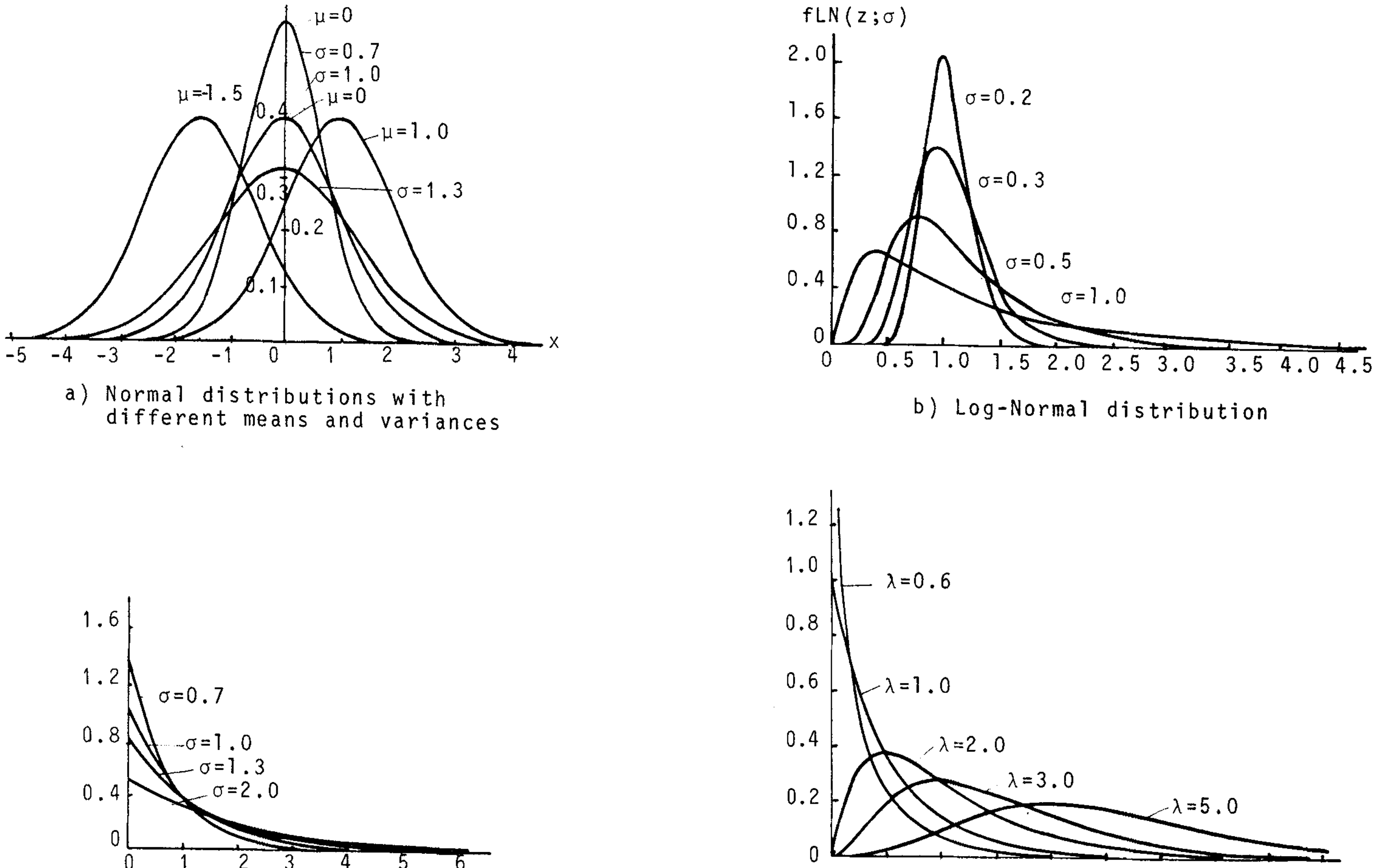
Sample estimate $a = \frac{N}{(N-1)(N-2)}\Sigma(x-\bar{x})^3$ (7.6)

The distributions in Fig. 7.3 are often used in hydrological analysis (Haan, 1977, Yevjevich, 1972a, Bury, 1975).

The *Normal distribution* represents the distribution of a completely random number about a mean. It is a symmetrical bell-shaped distribution with the area under the curve equal to unity.

The *Log Normal distribution* represents a similar distribution of the log of the variate.

The *Gamma distribution* has a skewness and passes through zero. It is used in drought flow analysis in particular for estimating reservoir capacity.



a) Normal distributions with different means and variances

b) Log-Normal distribution

c) Exponential distribution

d) Gamma distribution

Fig. 7.3 Shapes of various mathematical probability distributions

Pearson distributions (types I, II and III) were devised by Karl Pearson (1930) to fit virtually any distribution and are of an exponential type. Log Pearson distributions are also used.

Extreme value distributions (types I, II and III). Fisher and Tippet (1928) found that the extreme values of many distributions approached a limiting exponential form as the number of points in the sample increased. They fitted equations to the upper extremes. Gumbel (1941) first applied the Type I extreme value theory to floods. His work is now standard in the analysis of hydrological extremes (1958).

The distribution is of the form

$$p = 1-e^{-e^{-y}} \tag{7.7}$$

where p is the probability of the flood being equalled or exceeded, e is the base of Naperian logarithms and y is a mathematical function of probability. The equation for the extreme value of the variate as a function of recurrence interval T resulting from Gumbel's theory is

$$x = \bar{x} - \frac{\sqrt{6}}{\pi} s \{0.5772+\ln[-\ln(1-1/T)]\} \tag{7.8}$$

Gumbel went so far as to prepare graph paper which causes the variate (flood peaks), to plot as a straignt line against probability or its inverse, return period (e.g. Fig. 7.4). Alternatively the variate may be plotted to a log scale (Fig. 7.1). The distribution, together with the log Pearson type III is often applied in flood hydrology.

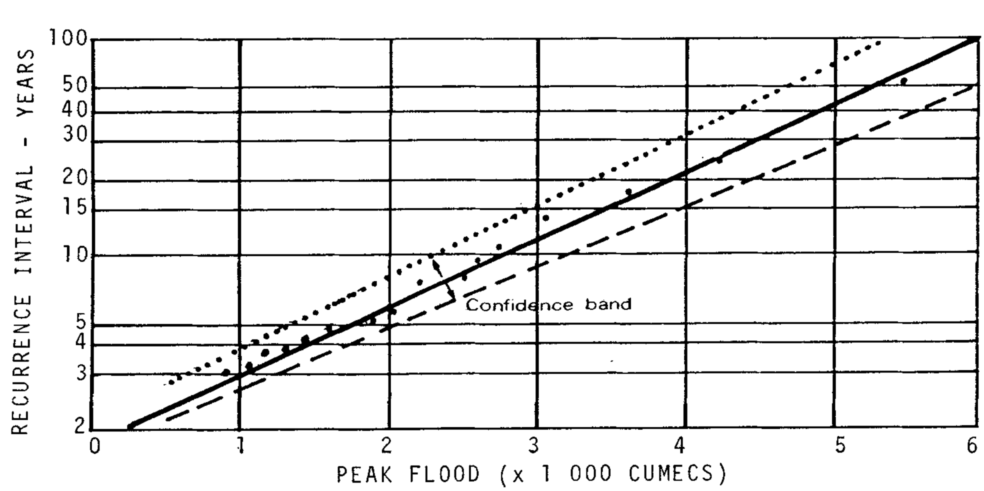


Fig. 7.4 Frequency distribution of flood peaks

ANALYSIS OF RECORDS

Presuming that flow records are available or can be synthesized (e.g. Yevjevich, 1972b; Fiering, 1967), then the extreme flow distributions can be analysed by means of the following procedure. The record is usually produced in chronological order as a complete duration series with peak flows for each day or month or year identified. The record should be divided into years (with the beginning of the hydrological year preferably at the start of the wettest season). The annual maxima should be selected and ranked, taking only the maximum flood in any one year. If the annual flood peaks are arranged in order of magnitude (Fig. 7.5b) we have an extreme value exceedance distribution. One thus has what is termed an annual partial series. If every flood on record was included we would have a complete series. The difference is only of interest for low recurrence intervals giving a lower recurrence interval for partial series than for annual series. For recurrence intervals over 10 years both series yield practically the same results.

The probability of a flow being equalled or exceeded in any hydrological year is the inverse of the frequency of it occuring. Thus a flood which is equalled or exceeded on an average once every 50 years has a two-percent probability. It is common to use the frequency, or recurrence interval or return period in hydrological analysis, in preference to probability. Thus

$$P(X \geq x) = \frac{1}{T} \tag{7.9}$$

So the probability that the annual maximum flood X is less than x is

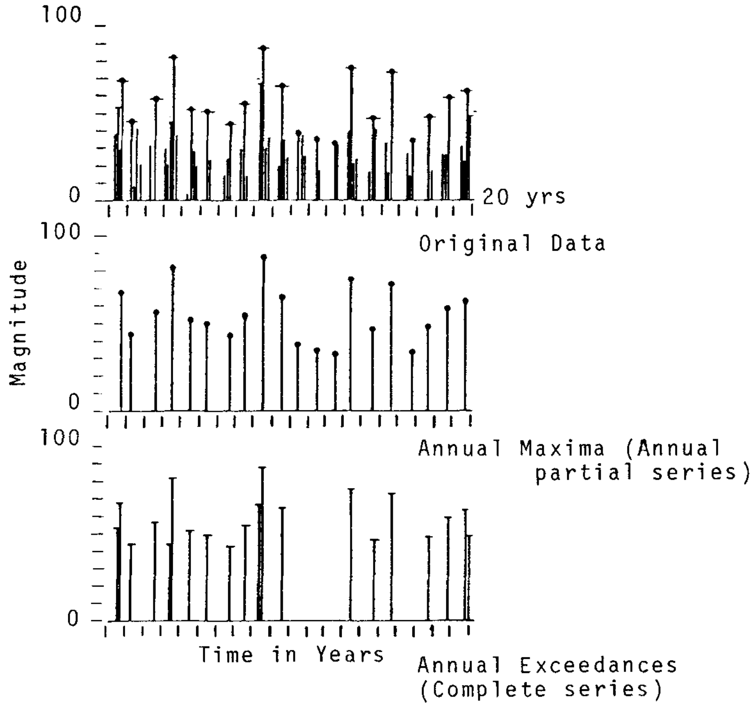
$$p(X < x) = 1 - \frac{1}{T} \tag{7.10}$$

where T is the recurrence interval of the flood of magnitude x.

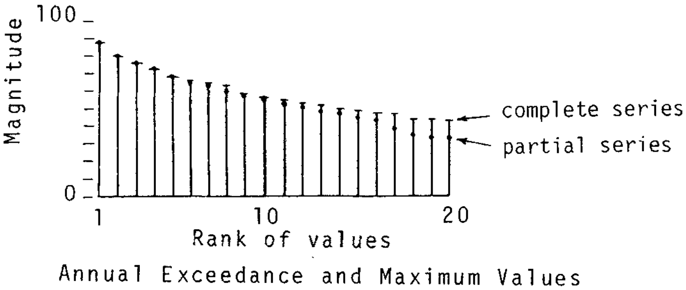
The estimation of T from the ranked sample has been done in different ways. Once a sample has been ranked in descending order of magnitude, the recurrence interval T may be estimated from the Weibull formula

$$T = \frac{N+1}{m} \tag{7.11}$$

where N is the number of events (or years of record) and m is the rank proceeding from 1 for the highest value. Some other formulae for estimating recurrence interval and the corresponding origin are indicated in Table 7.1.



a. Arranged in the order of occurrence



b. Arranged in the order of magnitude

Fig. 7.5 Hydrological data series.

TABLE 7.1 Formulae for recurrence interval T

Formula		T	Distribution	T for N=50, m = 1
California	(1923)	$\frac{N}{m}$		50
Hazen	(1930)	$\frac{2N}{2m-1}$		100
Weibull	(1939)	$\frac{N+1}{m}$	Normal & Pearson III	51
Blom	(1958)	$\frac{N+0.25}{m-0.375}$	Normal	80.4
Beard	(1962)	$\frac{N+0.4}{m\ 0.3}$	Pearson III	72
Gringorten	(1963)	$\frac{N+0.12}{m-0.44}$	Exponential, Extreme Value I	89.5

The constants in Gringorten's equation actually vary slightly depending on length of record.

The resulting recurrence intervals or so-called plotting positions may be plotted on a suitable graph such as in Fig. 7.4. The process of fitting a smooth curve through the resulting data will then eliminate many deviations. It should be borne in mind that values will deviate from the distribution near the mean, as extreme value theory as its name implies, is inapplicable there. In fact different distributions apply on either side of the mean.

CONFIDENCE BANDS

An infinite length of record only, will yield a true distribution with a very low minimum and an extremely high maximum. Any record of finite duration will have a more limited range and can only approximate the true distribution. The shorter the record, the less representative of the true distribution it is likely to be. In fact the variation of a number of sample means about the true mean will be distributed as a normal distribution with a mean μ and variance σ^2/N.

The possible extent of the true distribution each side of the available data can be described in terms of confidence limits. There will be a confidence band above and below the plotted line on an extreme value plot such as Fig. 7.4. The width of the band will depend on the degree of confidence accepted and on the scatter and sparseness of the data.

Confidence limits can be estimated in the case of a normal distribution from the data set (Haan, 1977). Thus there is a 68% probability that a sample mean will be within $\pm\sigma/\sqrt{N}$ of the true population mean. The probability of lying within a certain band about the mean, F(c) for different c is given in Table 7.2.

TABLE 7.2 Confidence band factors

Degree of confidence, c (%)	95	90	80	68
F(c) (about mean)	2.0	1.7	1.3	1.0
Recurrence interval, T(years)	2	10	100	1000
G (T)	1.0	1.5	2.2	2.7

For other plotting positions the band width increases in accordance with a factor G. The factor is dependent on the population distribution, length of record and recurrence interval. It may be deduced that G is dependent primarily on recurrence interval T, and the values in Table 7.2, apply for over 20 years of record. Thus the confidence band width about the line drawn through the plotted positions is very approximately

$$BW = F(c)G(T)s/\sqrt{N} \qquad (7.12)$$

As an approximation the band width on the upper side of the line is 0.6BW and on the lower side 0.4BW. For example for a sample of 20 peak flows with a standard deviation of 200 m^3/s, then the width of the 90% confidence band about the 100 year value plotted is $BW = 1.7 \times 2.2 \times 200/\sqrt{20} = 168\ m^3/s$. For plotting positions below the mean the band width again expands but these positions are normally of little interest to the drainage engineer. More accurate figures for the 90% confidence band are presented by Viessman et al (1977) and Beard (1978).

Yen (1974) presented a chart from which it is possible to read the probability of an event of rank m = 1, 2 or 3 in n years of record corresponding to an event of average return period T.

DESIGN DISCHARGE

The construction of a culvert, bridge waterway or even drain to pass a flood of a selected recurrence interval, involves some risk. The probability of the discharge capacity being exceeded at least once during the operation of the structure is greater than 1/T where T is tne recurrence interval of the design flood.

Although a minor overtopping may result in only inconvenience, a severe overflow of an embankment could cause scour and wash-away. This would cause economic damage as well as risk to traffic and life. A probalistic approach to the selection of design capacity is therefore desirable. Young et al (1974) applied risk analyses to the design of highway culverts and indicated damage costs. Allowance for cost uncertainty and interaction between culvert capacity and flood magnitude was studied in detail by Mays (1979).

SPREAD RISK

The simplest approach to the computation of design flood in the case of a structure which is to function over a long time or indefinitely is on an annual basis. The costs of the structures which can discharge different flood magnitudes are added to the corresponding probable cost of damage. All cost figures are converted to a common time basis, eg. annual costs. The method is explained with an example below.

Example

A highway authority is to construct a low bridge over a river. The flood frequency distribution is indicated in Fig. 7.4. The capital cost of the bridge is a function of the discharge capacity of the waterway beneath. The capital cost corresponding to various design discharge rates is converted to an annual amount representing interest and redemption on the loan to meet the cost, and added to annual maintenance cost. Annual interest plus redemption figures can be obtained from interest tables or from a formula (Institution of Civil Engineers, 1969). The capital cost is multiplied by the factor

$$\frac{r(1+r)^n}{(1+r)^n-1} \qquad (7.13)$$

where r is the interest rate (a fraction) on the loan, assumed here equal to the interest rate on the redemption or sinking fund, and n is the loan period in years. To get a true picture the loan should be renewed for as long as the economic life of the structure.

The resulting annual cost, as a function of waterway capacity, is plotted in Fig. 7.6 (curve A). Now corresponding to each of the trial waterway design capacities is a probability of exceedance in any year, as indicated in Fig. 7.4. There is a different probability corresponding to either the best estimate or the upper confidence limit of the flood frequency curve.

The probabilities of each flow being exceeded are multiplied by the cost of a flooding. This may include repair costs to the bridge, damage to surroundings, and interference with transport. Assuming in this case that the cost of exceeding the design capacity is $100 000, the probably cost is this figure multiplied by the probability of exceedance of the discharge. The resulting annual costs are plotted in Fig. 7.6 as curve B, for both best estimate and upper confidence limit.

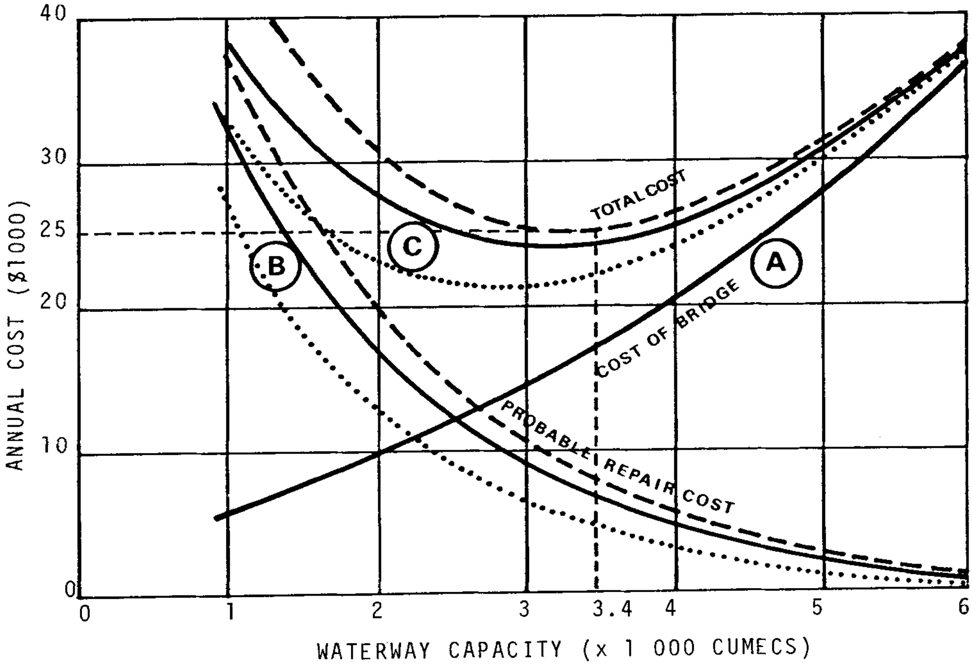


Fig. 7.6 Overall costs for spread risk example

The total average annual cost is the sum of A and B and this is indicated as curve C. The minimum cost corresponds to a waterway capacity of 3400 m^3/s, provided the upper flood curve is chosen. This is the usual procedure as it provides a margin for uncertainty in flood estimates. It will be observed however that the best estimate curve will result in a slightly smaller optimum waterway capacity. Thus the effect of uncertainty in hydrological data is to increase expenditure.

In real situations the computations will be more complicated. There is always the risk to human life which is difficult to evaluate. The damage costs may very well increase with increasing flood magnitude above the design capacity. In this case the probable economic loss must be evaluated. It is the integral of the probability of various extreme floods occuring multiplied by the corresponding cost, i.e.

$$\text{probable annual cost} = \Sigma\ P\ C \qquad (7.14)$$

where P is the probability of the flood being in a certain range and C is the corresponding damage cost.

The aspect of cost escalation also arises. If damage costs are likely to increase over the years, they should be evaluated separately for successive years, and discounted to a present value. Thus each annual cost is multiplied by

$$\frac{1}{(1+R)^n} \qquad (7.15)$$

where R is the inflation rate as a fraction. The total of the present values for each year is obtained and this may be added to the capital cost of the bridge. Again the capacity with least total cost is selected.

ISOLATED RISK

A flood with a return period of T will occur or be exceeded in any year with a probability of 1/T. The probability that it will occur at least once in n years is greater than 1/T since there is more chance for it to occur. The actual probability of the flood being equalled or exceeded at least once in a year may be evaluated as follows:

The probability that it will not occur in any one year is 1-1/T. Hence the probability that it will not occur in n years is the product of n such probabilities or $(1-1/T)^n$.

The probability that it will occur at least once in n years is therefore

$$P(Q \geq Q_T) = 1-(1-1/T)^n \qquad (7.16)$$

This probability represents the risk that a flood with return period T will occur or be exceeded at least once in n years. If the life of a structure is n years, there is a risk P that the capacity of the structure will be exceeded sometime in its life. P is greater than 1/T but less than n/T except for very high recurrence intervals, when P approaches n/T. It can be calculated that for large T there is a 63% chance that a flood of any recurrence interval T will occur in T years.

The application of (7.16) is in short-term projects such as temporary diversion works eg. bypass culverts or cofferdams (see eg. Linsley et al, 1975). An overtopping could result in damages of greater order than the actual cost of the diversion works and may in fact entail a completely new start. Thus the average loss is not so much of concern as the risk of overtopping. Nevertheless in view of the short useful life a more frequent design flood may be selected than for permanent structures.

The constructor of a temporary diversion works or cofferdam will be not so much interested in whether the structure will be overtopped during its useful life at all, as to what the probability is of it being overtopped once, twice or any other number of times. Each flooding will cause damage and he needs to know the total probable damage. The partial series method of ranking the floods excludes the investigation of more than one flood in any year though (unless the n worst events are selected in strict order of magnitude without only selecting one maximum a year i.e. a complete series).

The probability that a flood Q_T of recurrence interval T will be exceeded exactly r times in n years ($r \leq n$) can be calculated from the expression

$$P(Q>Q_T)_r = \frac{n!}{r!(n-r)!} \left(\frac{1}{T}\right)^r \left(1 - \frac{1}{T}\right)^{n-r} \qquad (7.17)$$

This equation is derived using the binomial distribution describing the probability of occurrence of independent events, and was solved by Yen (1970) for various cases.

Use of the equation is demonstrated by way of an example involving the construction of a temporary diversion culvert. Assuming the flood frequency distribution in Fig. 7.4 again, the probable damage costs associated with a 5-year design life are to be investigated. The total cost of exceeding the design capacity or overtopping the embankment is estimated to be \$100 000 per flooding. The cost of the culvert and embankment for various capacities is indicated in Fig. 7.7, line A. The risk of exceeding the diversion works capacity 1, 2 or 3 times is computed from (7.17) and tabulated in Table 7.3.

TABLE 7.3 Risk of flooding for a 5-year diversion works

Recurrence interval, years		5	10	20	30	50
Discharge capacity, m³/s:		1700	2800	3800	4500	5300
No. of failures	Total cost of failures					
0	0	0.33	0.59	0.77	0.85	0.90
1	\$100 000	0.39	0.32	0.20	0.135	0.09
2	\$200 000	0.20	0.075	0.025	0.013	0.009
3	\$300 000	0.07	0.014	0.005	0.001	0.001

The cost of failures multiplied by the corresponding risk is indicated in Fig. 7.7 (lines marked B are filled in after plotting the individual risks).

The total cost of construction and the associated risk of damage is then computed and plotted at the top of Fig. 7.7 (curves C). This figure presents more information than the average cost data derived in the example under the spread risk approach. In fact it gives the

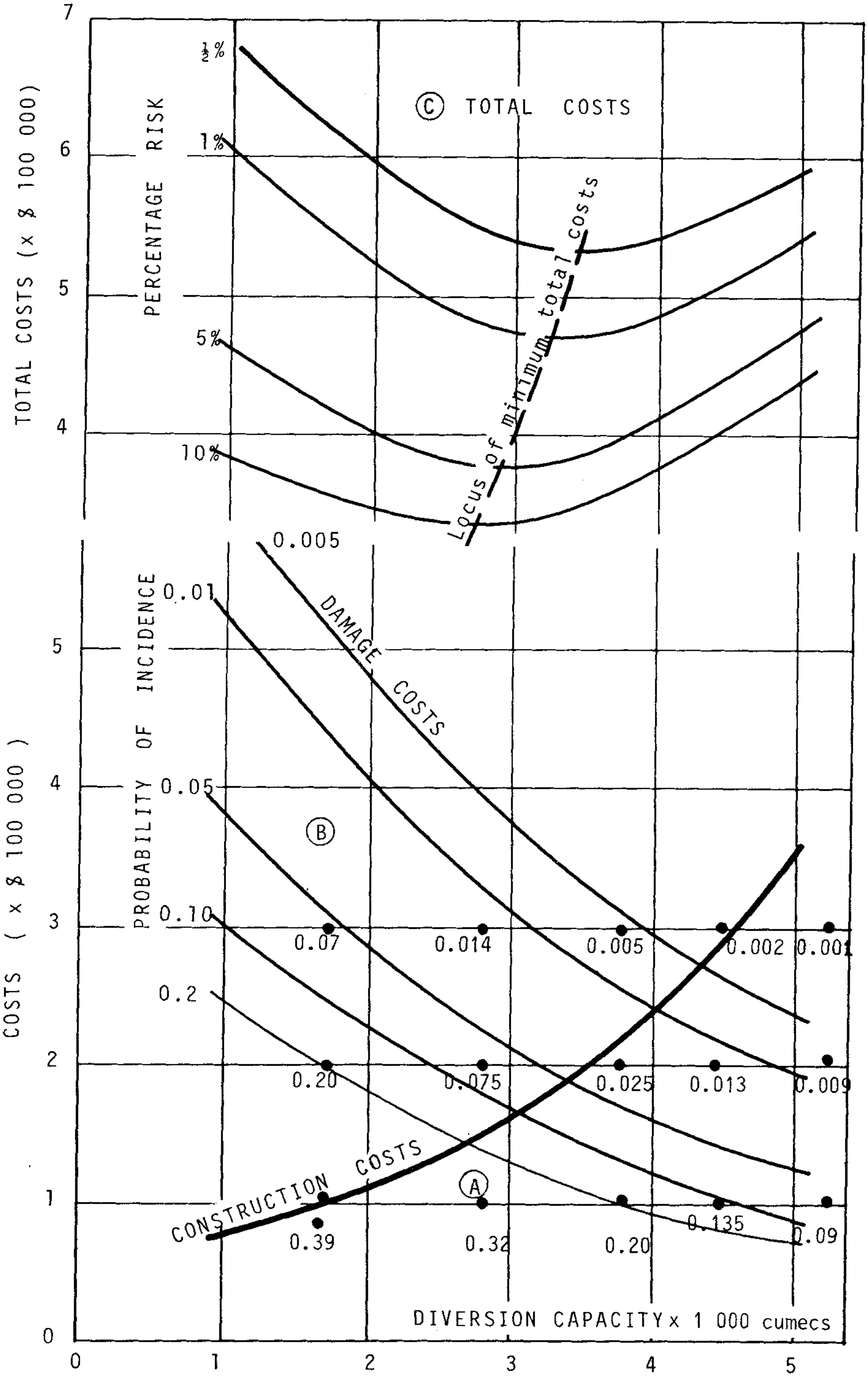


Fig. 7.7 Hydro-Economic analysis of diversion works; isolated risk example

risk of different expenditures corresponding to various design capacities. Thus for a design capacity of 3000 m^3/s there is a 1% risk that the cost will be as high as $480 000.

The additional or marginal information is useful in cases where the contractor or constructor must bear an excess on insurance premiums. In particular the constructor may have $400 000 available to cover cost of construction and damage, and the balance of damage must be covered by insurance. If the works was constructed to a capacity of 3000 m^3/s (corresponding to minimum total cost of $400 000), then there is a 4% chance the insurance company will have to meet a proportion of costs.

REFERENCES

Beard, L.R., Nov., 1978. Impact of hydrological uncertainties on flood insurance. Proc. ASCE, 104(HY11), 14168, p1473-1484.
Bury, K.V., 1975. Statistical Models in Applied Science. John Wiley & Sons. N.Y.
Fiering, M.B., 1967. Streamflow Synthesis. Harvard Univ. Press, Cambridge, Mass., 139pp.
Fisher, R.A. and Tippett, L.H.C., 1928. Limiting forms of the frequency distribution of the smallest and largest members of a sample. Proc. Cambridge Phil. Soc., 24, p180-190.
Gringorten, I.I., Feb. 1963. A plotting rule for extreme probability paper. J. Geophys. Res., 68(3), p813-814.
Gumbel, E.J., June, 1941. The return period of flood flows. Ann. Math. Statist., XII (2), p163-190.
Gumbel, E.J., May, 1958. Statistical theory of floods and drought, J. Inst. Water Engrs. 12 (3), p157-184
Haan, C.T., 1977. Statistical Methods in Hydrology Iowa State Univ. Press, Iowa. 378pp.
Institution of Civil Engineers, 1969. An Introduction to Engineering Economics, London. 182pp.
Linsley, R.K., Kohler M.A., and Paulus J.L.H., 1975. Hydrology for Engineers, 2nd Ed. McGraw Hill, N.Y. 482pp.
Mays, L.W., May, 1979. Optimal design of culverts under uncertainties. Proc. ASCE, 105(HY5), 14572. p443-460.
Pearson, K., 1930. Tables for Statisticians and Biometricians, 3rd Ed. Part I. The biometric Lab., Univ. College, Cambridge Univ. Press, London.
Viessman, W.,Knapp, J.W., Lewis, G.L. and Harbaugh, T.E., 1977. Introduction to Hydrology, 2nd Ed. Harper & Row, N.Y. 704pp.
Yen, B.C., April, 1970. Risk in hydrologic design of engineering projects. Proc. ASCE., 96(HY4), 7229, p959-966.
Yevjevich, V., 1972a. Probability and Statistics in Hydrology. Water Resources Publications, Fort Collins, 302pp.
Yevjevich, V., 1972b. Stochastic Processes in Hydrology. Water Resources Publications, Fort Collins, 276 pp.
Young, G.K., Childrey, M.R., and Trent, R.E., July, 1974. Optimal design for highway drainage culverts. Proc. ASCE, 100(HY7), 10676, p971-993.