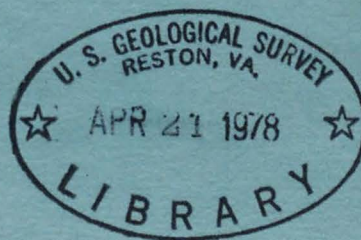
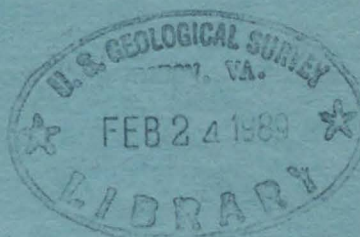


— FLOOD FREQUENCY — EXPECTED AND UNEXPECTED PROBABILITIES

U.S. GEOLOGICAL SURVEY
Open-File Report 76—775



(200)
R 290
no. 76-775



— FLOOD FREQUENCY —

EXPECTED AND UNEXPECTED PROBABILITIES

By ^{onald} D.M. Thomas ^{cat 63} 1927-

Thomas, Donald M

NOTH

✓
✓ T. Wana

U.S. GEOLOGICAL SURVEY,
Open-File Report 76-775

[Reports-Open file series]

286519



U.S. DEPARTMENT OF THE INTERIOR

Thomas S. Kleppe, Secretary

GEOLOGICAL SURVEY

V.E. McKelvey, Director

For additional information write to:

Chief Hydrologist
U.S. Geological Survey, WRD
415 National Center
Reston, Virginia 22092

FLOOD FREQUENCY — EXPECTED AND UNEXPECTED PROBABILITIES

By DAN THOMAS

CONTENTS

	Page
Abstract	1
Introduction	1
Frequency concepts	1
Discharge variations	1
Probability variations	3
Expected probability	4
Discussion	5
Summary and conclusions	6
References	7

ILLUSTRATIONS

	Page
Figure 1. Time-sampling variations in 10-year records	2
2. Discharge histograms	3
3. Frequency curves on rectangular coordinate graph	4
4. Probability histograms	5

TABLES

	Page
Table 1. Theoretical number of years in a 100-year period that floods exceed a 3 percent chance design discharge estimated from a 10-year record at 20 homogeneous but independent sites	6

— FLOOD FREQUENCY —

EXPECTED AND UNEXPECTED PROBABILITIES

By D.M. Thomas

ABSTRACT

Flood-frequency curves may be defined either with or without an "expected probability" adjustment; and the two curves differ in the way that they attempt to average the time-sampling uncertainties. A curve with no adjustment is shown to estimate a median value of both discharge and frequency of occurrence, while an expected probability curve is shown to estimate a mean frequency of flood years. The attributes and constraints of the two types of curves for various uses are discussed.

INTRODUCTION

Flood-frequency curves may be defined either with or without an "expected probability" adjustment. This optional use of expected probability adjustments has caused problems when different analysts have attempted to compare and coordinate flood-frequency estimates. From observations at several meetings intended to achieve such coordination, it is clear that few hydrologists and fewer users of flood frequency estimates have a thorough understanding of the assumptions and limitations associated with frequency curves and expected probability adjustments.

This report explains and discusses various aspects of frequency curves with special emphasis on expected probability. This nonmathematical explanation attempts to illuminate concepts through use of graphical displays in the hope that the explanation will be clear to those with limited knowledge of statistical processes.

A simple time-sampling analysis by Benson (1960) provided the data for examples in this report. Although the type of frequency analysis used by Benson is no longer in vogue, the information was available in a form readily usable for preparing illustrations, and is appropriate for demonstrating concepts.

FREQUENCY CONCEPTS

Flood frequency curves are used to assess the characteristics of future floods. Through analysis of records of past floods we define a relation between flood magnitudes and the frequency of years in which they occur. When using a defined frequency curve for planning or

designing we assume, often implicitly, that the curve is based upon a representative sample of long-term flood experience and therefore is an adequate predictor of the future.

Such an assumption often is tenuous. We should recognize that there is quite likely some variation between frequency curves based upon short records and curves based upon a much longer record. Benson (1960) studied those variations. He assumed that if he had a 1,000-year record it might define a curve identified as the 'base curve' in figure 1. He recorded on a separate piece of paper each of the 1,000 flood magnitudes that would define the base curve perfectly, placed all in a hat, and mixed them thoroughly. He then drew the slips out of the hat, one at a time, and recorded chronologically each magnitude to produce a simulated 1,000-year flood sequence. Then he defined 100 frequency curves based upon the first '10-year' period, the second "10-year" period, and so fourth. From figure 1 it is obvious that any given 10-year period may define a frequency curve that differs considerably from the long-term curve. This variation commonly is called the time-sampling error.

An equally viable interpretation of figure 1 is that the base curve averages the time-sampling errors in a regional sample of homogeneous and independent flood records. Consider that we might have a flood record of 10-year length at each of the 100 sites, all having identical flood-producing characteristics, but sufficiently remote from each other to assure independence of the annual floods. Under these conditions we could expect the 100 individual frequency curves to display the scatter shown in figure 1.

DISCHARGE VARIATIONS

The usual assessment of time-sampling variations expected in frequency curves defined from short-term records considers the range in discharge estimates for a selected probability of exceedance. That is, the exceedance probability is specified and the time-sampling error reflects differences in discharge estimates. Variations of discharge estimates by the 100 curves in figure 1

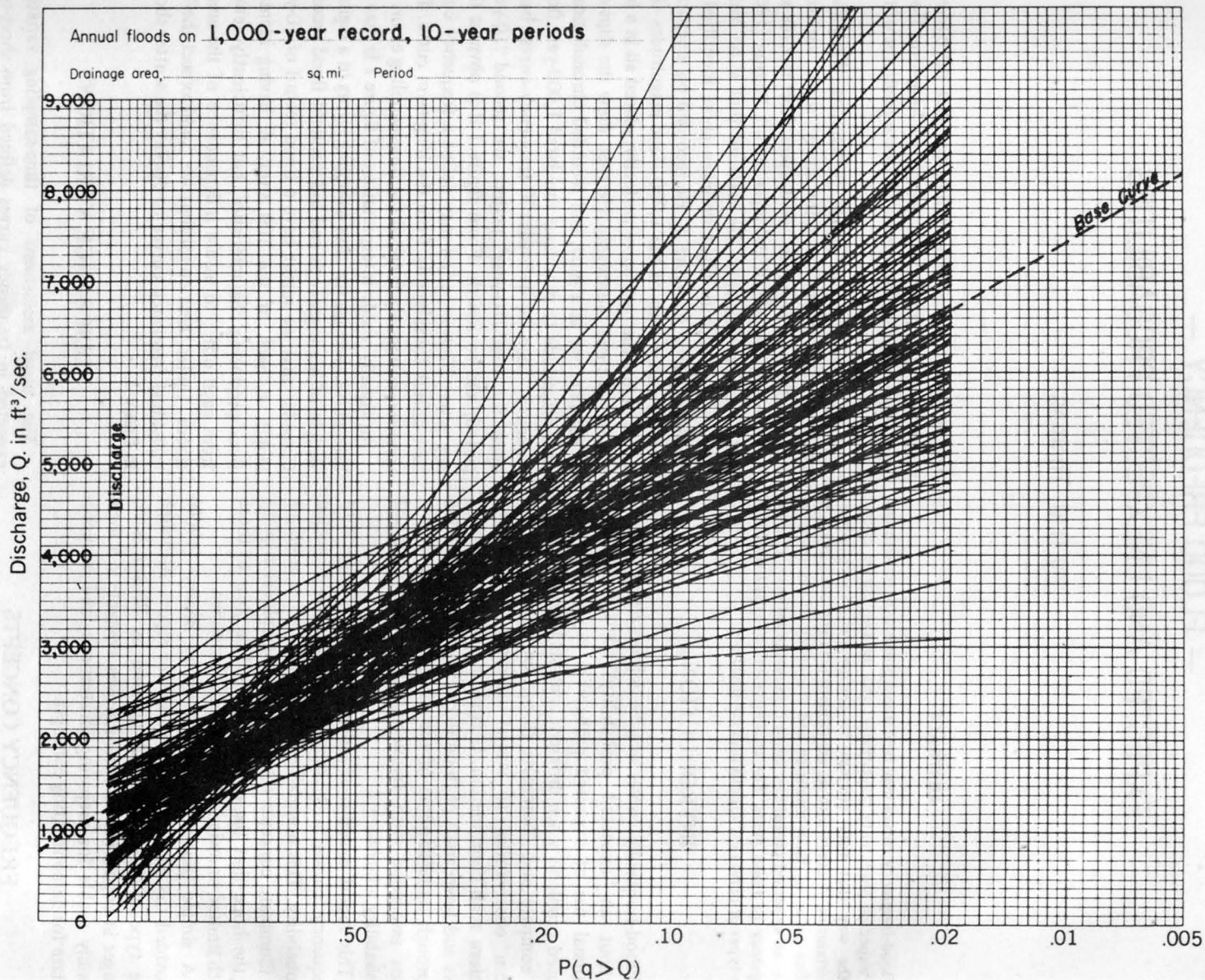


Figure 1.—Time sampling variations in 10-year records.

are displayed in figure 2. These histograms show, for example, that 14 of the 100 curves defined from 10-year samples estimated the 4-percent chance (25-year) flood magnitude between 5,000 and 5,499 ft³/s (cubic feet per second) while 18 curves estimated the value between 5,500 and 5,999 ft³/s. From the base curve the 4-percent chance flood is 5,890 ft³/s. Some statistical measures may be computed to define the time-sampling variability and to aid in the comparison of variability between different probability levels. Again for the 4-percent chance flood, the median estimate (\tilde{X}_{04}) of 5,850 ft³/s is exceeded by 50 percent of the estimates; another measure of the central value of the distribution histogram is the mean (\bar{X}_{04}) of 6,160 ft³/s; an index of the variation in estimates is the standard deviation (S_{04}) of 1,482; and a measure of the nonsymmetry of the distribution histogram is the coefficient of skew (g_{04}) which is 1.275.

Several attributes of time-sampling errors observable from the limited data of Benson's study are true of time-sampling errors in general. Note, for example, that as the probability of exceedance decreases, increases occur both in the scatter and in the nonsymmetry of estimates, as indicated by the standard deviation and the coefficient of skew. Note also that the mean of the estimates exceeds the median and that the median estimate is always a better estimate of the long-term curve than the mean. Not shown here but easily observable in Benson's data is the significant reduction that occurs in the variation of estimates as record length increases.

Time-sampling error distributions vary with the characteristics of the parent population of floods, the length of the sample flood record, and the probability of flood exceedance. Statistical techniques are available for describing these time-sampling errors for X-percent chance floods known to be from a Gaussian (normal) population. The Water Resources Council (1969,1976) and Hardison (1969,1976) have proposed expressions for evaluating the errors and confidence bands for Pearson Type III populations. In many cases it is possible, therefore, to provide probability statements about the population value of an X-percent flood lying within a specified range of a value computed from a sample.

PROBABILITY VARIATIONS

In contrast to the time-sampling errors in discharge estimates for a given probability of exceedance, consider the variations in estimates of exceedance probabilities for a specified discharge. Time-sampling errors are then viewed in a horizontal direction on the usual flood-frequency curve.

The horizontal scale of frequency curves invariably is distorted in an attempt to linearize the magnitude-frequency relation. This scale distortion tends to obscure the nature of the time-sampling distribution of

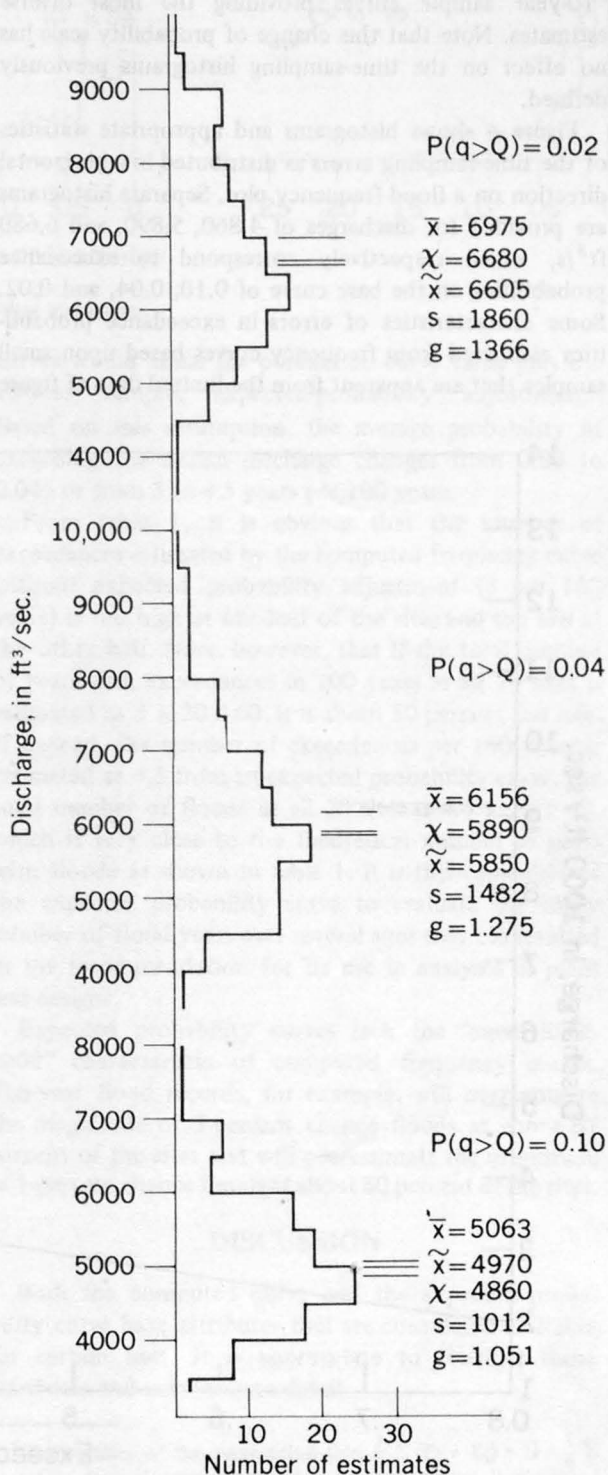


Figure 2.—Discharge histograms.

exceedance probabilities; and it is informative, therefore, to view the data of figure 1 plotted on a rectangular coordinate graph. Figure 3 shows Benson's complete base curve, but for simplicity and clarity shows only the "10-year" sample curves providing the most diverse estimates. Note that this change of probability scale has no effect on the time-sampling histograms previously defined.

Figure 4 shows histograms and appropriate statistics of the time-sampling errors as distributed in a horizontal direction on a flood-frequency plot. Separate histograms are provided for discharges of 4,860, 5,890, and 6,680 ft^3/s , which respectively correspond to exceedance probabilities on the base curve of 0.10, 0.04, and 0.02. Some characteristics of errors in exceedance probabilities estimated from frequency curves based upon small samples that are apparent from the limited data of figure

4 and true in general are:

- The nonsymmetry of the error distribution, as indicated by the skew coefficient, increases as the probability of exceedance decreases.
- The mean exceedance probability, \bar{p} , is always greater than both the median, \tilde{p} , and the long-term or population probability Π .
- The median probability, \tilde{p} , is always near the long-term or population probability, Π , indicating that the commonly computed frequency curve has about an equal chance of estimating the frequency of flood exceedance too high or too low.

EXPECTED PROBABILITY

A frequency curve defined as the mean probability value (\bar{p} in figure 4) for all possible discharge values is

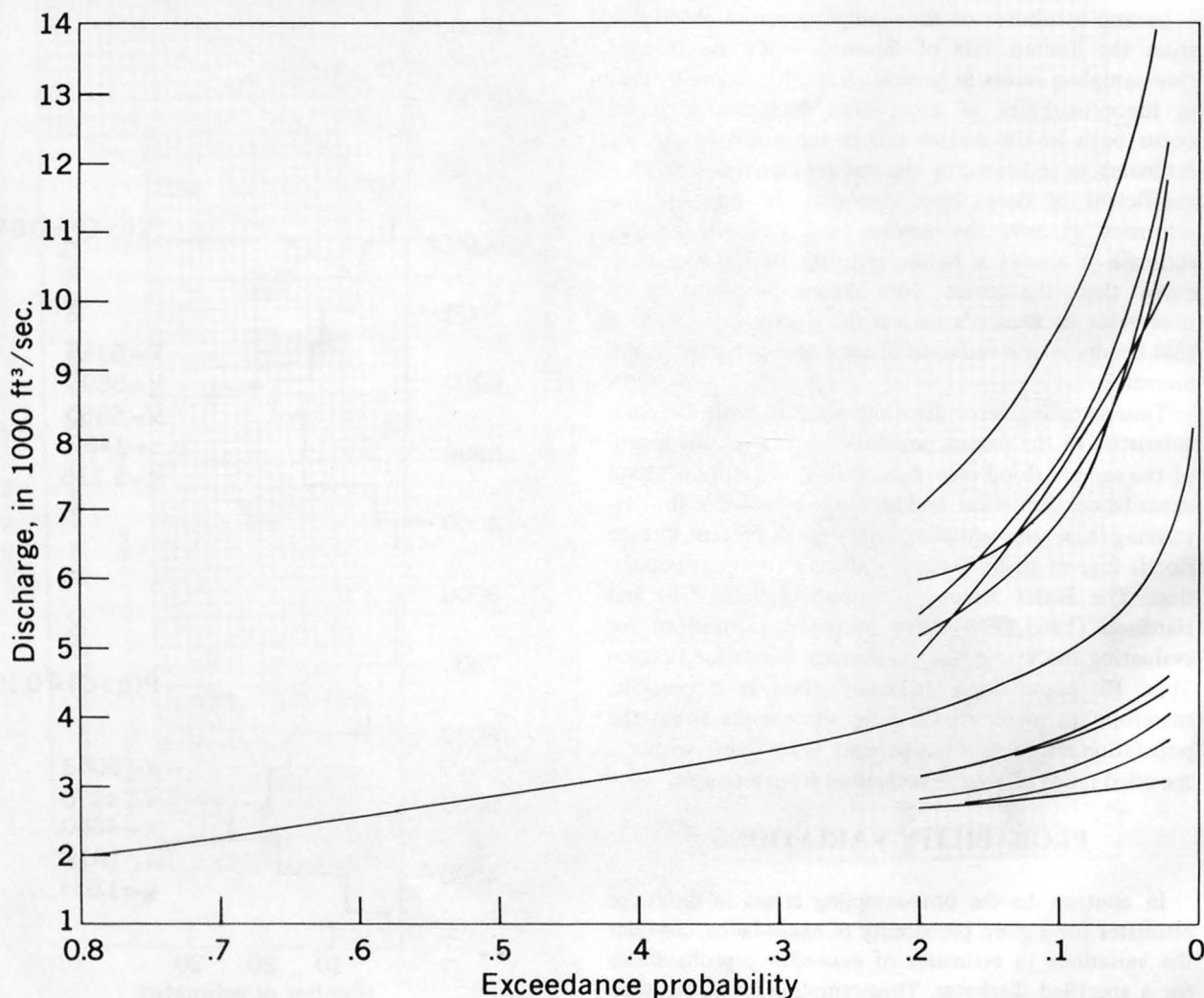


Figure 3.—Frequency curves on rectangular coordinate graph.

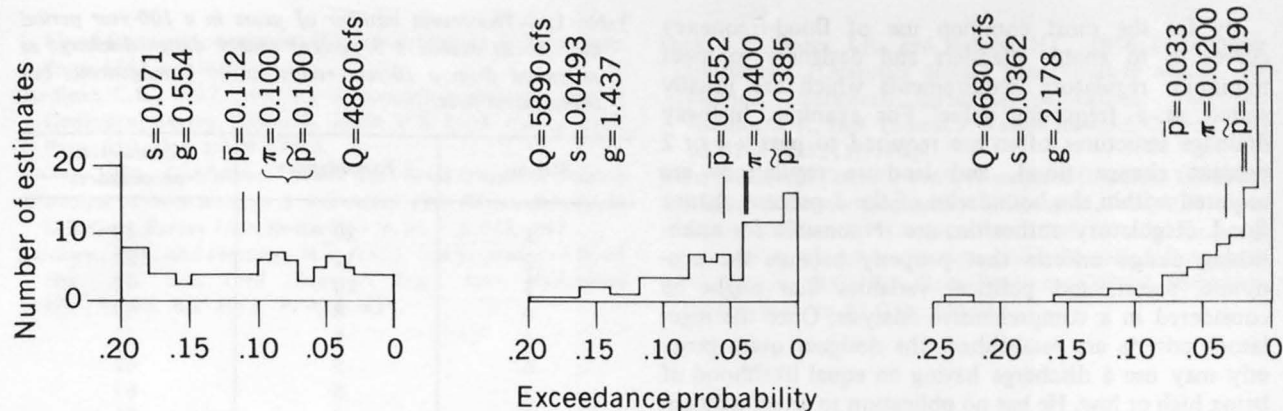


Figure 4.—Probability histograms.

the “expected probability” curve. In practice, the expected probability curve is not computed directly but is evaluated as an adjustment to the usual discharge-frequency relation. The magnitude of the adjustment can be mathematically described only for floods that belong to a Gaussian (normal) probability distribution. Hardison and Jennings (1972) and Slack, Wallis, and Matalas (1975) employed Monte Carlo techniques to suggest adjustments for use with skewed distributions, but the Water Resources Council (1976) recommends that “published adjustments applicable to the normal distribution be used,” when defining expected probability curves from log Pearson Type III computed curves.

The expected probability curve has been proposed as the proper curve for use in analysis of project plans and designs because it has the “best capability” of estimating the number of years having floods exceeding selected discharges (Beard, 1959, 1960, 1974). To illustrate this point, consider that projects might be designed on the basis of 10-year flood records at 20 independent sites having identical flood-producing characteristics. On the assumption that the floods sample a normal population, Beard (1974) computed for each of the 20 sites the theoretical number of years per 100 years that floods exceed a design discharge having a 0.03 exceedance probability (33-year flood) as defined by the population curve. Table 1 lists the number of years with design-flood exceedances arranged in descending order from the site where the 10-year record most underestimated the frequency.

The number of exceedance years shown in table 1 might be expected when using the commonly computed frequency curve. If the expected probability adjustment has been applied to each of the 20 curves, the estimated number of design-flood exceedances would be altered. To appraise the effect of the adjustment, we might assume that the average frequency of the design discharge as estimated by the 20 expected probability

curves would equal the population curve value plus the 10-year sample, expected-probability adjustment.¹ Based on this assumption, the average probability of exceeding the design discharge changes from 0.03 to 0.045 or from 3 to 4.5 years per 100 years.

From table 1, it is obvious that the number of exceedances estimated by the computed frequency curve without expected probability adjustment (3 per 100 years) is too high at one-half of the sites and too low at the other half. Note, however, that if the total number of years with exceedances in 100 years at all 20 sites is estimated as $3 \times 20 = 60$, it is about 50 percent too low. If instead, the number of exceedances per 100 years is estimated as 4.5 from an expected probability curve, the total number of floods at all 20 sites is $4.5 \times 20 = 90$, which is very close to the theoretical number of years with floods as shown in table 1. It is this capability of the expected probability curve to evaluate the likely number of flood years over several sites that has resulted in the recommendation for its use in analyses of plans and designs.

Expected probability curves lack the “equal likelihood” characteristic of computed frequency curves. Ten-year flood records, for example, will overestimate the magnitude of 3-percent chance floods at about 67 percent of the sites and will overestimate the magnitude of 1-percent chance floods at about 80 percent of the sites.

DISCUSSION

Both the computed curve and the expected probability curve have attributes that are considered desirable for certain uses. It is appropriate to consider those attributes and uses in some detail.

¹The validity of the assumption that $\frac{1}{n} \sum (P_i + E_i) = \Pi + \frac{1}{n} \sum E_i$, where E_i is the expected probability adjustment, is questionable; but effects of errors in the assumption have not been investigated.

By far the most common use of flood-frequency curves is to enable planners and designers to meet minimum regulatory requirements which are usually stated as a frequency value. For example, highway drainage structures often are required to pass a 4 or 2 percent chance flood, and land-use regulations are required within the boundaries of the 1 percent chance flood. Regulatory authorities are responsible for establishing design criteria that properly balance the economic, social, and political variables that might be considered in a comprehensive analysis. Once the regulatory criteria are established, the designer quite properly may use a discharge having an equal likelihood of being high or low. He has no obligation to select a design discharge having, say, an 80- to 20-percent chance of being too high.

Owners of individual sites are unlikely to be intrigued by the possibility of averaging the number of their expected floods with the floods of distant neighbors. They quite likely are interested in larger than minimum regulatory requirements only on the basis either of optimizing economic returns or of risk aversion; both of which may be assessed from a computed frequency curve.

The greatest demand for expected probability curves comes from governmental construction agencies that build and own numerous structures. By evaluating the mean number of years likely to experience various flood discharges, the expected probability curve provides a logical tool for evaluating future flood damages associated with various designs and, thereby, can lead to economically optimum plans and designs.

Users of expected probability curves should recognize, however, that the highly unbalanced distribution of flood exceedances between sites may limit the reality of an economic analysis. From table 1 it is apparent that nearly one-half of all the design flood exceedances occur at but 3 of 20 sites. It is likely that when damaging floods, which are expected to occur in only 3 (from a computed curve) to 5 (expected probability curve) years, per 100 years, actually occur at time intervals averaging 5 to 10 years, the under-designed structure or plan will be revised to accommodate a larger discharge and thereby invalidate the earlier economic analysis.

A case of special interest is the utility of flood-frequency curves as a tool in the Federal Insurance Administration (HUD) flood insurance program. Obviously an expected probability curve is an ideal basis for establishing equitable flood insurance rates. Conversely, land owners forced to comply with the arbitrarily established criteria for land-use regulation have a right to expect that the flood-plain boundaries will be established with an equal likelihood of being too high or too low—an obvious application for a computed frequency curve without the expected probability adjustment. Two different frequency curves can thus be justified for one program—a result at odds with the

Table 1.—*Theoretical number of years in a 100-year period that floods exceed a 3 percent chance design discharge as estimated from a 10-year record at 20 homogeneous but independent sites.*

Site no.	Exceedances	Cumulative exceedances
1	20	20
2	12	32
3	10	42
4	8	50
5	7	57
6	5	62
7	5	67
8	4	71
9	4	75
10	3	78
11	3	81
12	2	83
13	2	85
14	2	87
15	1	88
16	.9	89
17	.8	90
18	.5	90
19	.3	90
20	.1	91

goals and desires of the Hydrology Committee of the Water Resources Council (1969) for a uniform or unique frequency curve.

SUMMARY AND CONCLUSIONS

Two types of flood-frequency curves are in use. The curves differ because they attempt to average in different ways the uncertainty in defining the magnitude-frequency relations on the basis of short flood records. The commonly computed frequency curve approaches the median value of the time-sampling error on both the discharge and frequency scales. The expected probability curve estimates the mean value of the time-sampling variation on the frequency scale. Users of frequency and(or) probability curves should recognize the attributes and limitations of each curve, and then apply the appropriate one for their problem.

REFERENCES

- Beard, L.R., 1959, Probability estimates based on small normal-distribution samples: U.S. Army Corps Engineers Civil Works Inv. Proj. CW-151, West Coast Flood Volume Studies, Research Note No. 2, Sacramento, Calif., 14 p.
- , 1960, Probability estimates based on small normal-distribution samples: Jour. Geophys. Research, v. 65, no. 7, p. 2143-2148.
- , 1974, Flood-flow frequency techniques: Austin, Texas Univ. Center For Water Resources Research Tech. Rept. CRWR-119, 27 p, 22 tables, 5 app.
- Benson, M.A., 1960, Characteristics of frequency curves based on a theoretical 1000-year record, in Dalrymple, Tate,

- Flood-frequency analyses: U.S. Geol. Survey Water-Supply Paper 1543, p. 51-74.
- Hardison, C.H., 1969, Accuracy of streamflow characteristics, *in* Geological Survey Research, 1969: U.S. Geol. Survey Prof. Paper 650-D, p. D210-D214.
- 1976, Confidence limits for flood-frequency curves computed from samples from Pearson Type III populations, *in* U.S. Geol. Survey Jour. Research, v. 4, no. 5, p. 545-547.
- Hardison, C.H., and Jennings, M.E., 1972, Bias in computed flood risk: Am. Soc. Civil Engineers Proc., Jour. Hydraulics Div., v. 98, no. HY3, p. 415-427.
- Slack, J.R., Wallis, J.R., and Matalas, N.C., 1976, Distribution functions for statistics derived from bivariate normal and bivariate 2-parameter log-normal populations: Yorktown Heights, N.Y., IBM Thomas J. Watson Research Center Rept. RC 5794 (#25111), 178 p.
- Water Resources Council, 1967, A uniform technique for determining floodflow frequencies: Washington, D.C., U.S. Water Resources Council, Hydrology Comm. Bull. No. 15, 15 p.
- 1976, Guidelines for determining flood-flow frequency: Washington, D.C., U.S. Water Resources Council, Hydrology Comm. Bull. No. 17, 25p., 14 app.

USGS LIBRARY RESTON



3 1818 00011003 9