INTERNATIONAL SYMPOSIUM ON UNCERTAINTIES IN HYDROLOGIC and WATER RESOURCE SYSTEMS

UNCERTAINTY IN THE RETURN PERIOD OF MAXIMUM EVENTS: A BAYESIAN APPROACH

by

Donald Davis
Lucien Duckstein
Chester Kisiel
Department of Hydrology & Water Resources
University of Arizona

and

Martin Fogel
Department of Watershed Management
University of Arizona
Tucson, Arizona 85721

Abstract

In previous papers by the co-authors, it was found that the return period of maximum point rainfall could be derived from the following event-based process valid for the summer precipitation season:

(a) the number of events per season is Poisson distributed,
(b) the amount of point rainfall per event may be exponentially distributed,
(c) the above two random variables are independent.

The purpose of this paper is (1) to obtain a correct distribution function of the return period when uncertainty arises in the parameters of the Poisson and exponential distributions and (2) to obtain the probability density function of the event corresponding to a given return period and that of the return period corresponding to a given event stemming from the same uncertainty. The methodology employs conjugate distributions for the Poisson and exponential distributions to permit direct computation of posterior distributions of rainfall amounts given point rainfall data. Computer simulation shows that errors in estimated return periods lead to underdesign even with 20 years of data (for the example used in our analysis). Extensions of the method are possible to encompass effects of uncertainties in rainfall data samples on runoff estimates and associated economic factors.

1.0 Introduction

Return periods of extreme rainfalls and flows are typically derived from empirical frequency analyses of available data. Use of plotting positions on probability paper (for example, Gumbel or log-normal) give frequency curves. These are in turn used either to interpolate for intermediate return period T_R or to extrapolate for return periods beyond the range of the data. There appears to be a general awareness of the limitations of this approach but, in general, there exists no generally

accepted prescription for coping with those uncertainties resulting from small data samples at a single sampling site. The above procedures and specific design return periods become enshrined in handbooks, codes, published charts (U.S. Weather Bureau, 1961), and administrative regulations. It may be that users of the techniques may make allowances for the uncertainties but, if done, it may be quite subjective and inconsistent from designer to designer.

It is well known that maximal events with small return period T_R are more accurately estimated than those with much larger T_R (Metler, et al., 1972). The confidence band flares out at larger T_R's so as to have no meaning. It is uncommon for confidence limits to be constructed because of this reason and, more importantly, because sampling properties for non-Gaussian distributions of interest to hydrologists are not generally available.

An awareness of the above situation has prompted the pursuit of at least three pathways:

(a) Use of the theory of stochastic processes and event-based data analysis (Fogel & Duckstein, 1969; Duckstein, et al., 1972).

(b) Use of small sample theory on regional regression methods to augment the information content of short records (for the case of floods, see Metler, et al., 1972 and Metler, 1972).

(c) Use of decision theory; for example by consideration of Type I and II errors (Duckstein and Kisiel, 1971), by use of Bayesian decision theory (Davis, et al., 1972), and by use of methods to "design" a factor of safety as that given in the Induced Safety Algorithm (Bogardi & Szidaroszky, 1972). In these approaches, economic loss functions are used in conjunction with the encoded uncertainty, that is, the posterior distribution of the parameters in question. In this way, the approaches customize the evaluation of uncertainty for each design situation.

Earth scientists (geologists, geomorphologists, paleo-climatologists, and palynologists) express concern that the blind reliance of surface water hydrologists on gaging records is tantamount to assuming zero knowledge or data at the initiation of a new data gathering station. Of course, this is not true as is well emphasized by Dalinsky (1970) but as yet there is no operational method for integrating such prior information into existing continuous records. Assignment of T_R to past extreme events identified by indirect data like (old river banks, uprooted trees, tree rings and mud varves) is fraught with much inconsistent guesswork. Bayes rule has the potential for removing some of this inconsistency (Bernier, 1967).

Two important facts are invoked to justify further development of the combined use of the first two of the three pathways. First, increased human intervention in the rainfall-runoff process makes historical records less representative of future conditions under which design structures must function, unless a case can be made for insensitivity of

extreme large events (their magnitudes and return periods) to human activity. Second, the designer, planner and other decision makers are increasingly compelled to make decisions irrespective of how much data are at hand or how deficient our methods (models) may be. The first fact motivates the more intensive study of the theory of event-based stochastic processes; the second fact motivates the study of decision theory.

2.0 Objectives of the Paper

In previous papers by the co-authors (Fogel and Duckstein, 1969; Duckstein, et al., 1972), it was found that the return period $T_R(k)$ of maximum point rainfall $\underset{\sim}{R}$ (with k the value of the random variable $\underset{\sim}{R}$) could be derived from the following event-based process judged to be valid for the summer precipitation season (characterized by sequences of thunderstorms):

(a) the number $\underset{\sim}{N}$ of events per season is Poisson distributed with mean m (of number of events per season): $f_N(n|m) = \frac{e^{-m}m^n}{n!}$

(b) the amount $\underset{\sim}{R}$ of point rainfall per event is exponentially distributed with parameter u (equal to reciprocal of mean amount of rainfall per event):

$$f_R\ (k|u) = u\ e^{-uk}$$

(c) $\underset{\sim}{N}$ and $\underset{\sim}{R}$ are independent.

Then, it may be shown (Duckstein, et al, 1972; Fogel, et al., 1972) that the return period of x units of rain given the event-based parameters m and u is

$$T_R\ (k|m,\ u) = [\ 1 - \exp(-m\ e^{-uk})\ ]^{-1} \qquad (1)$$

Given this relationship, the purpose of the paper is fourfold (knowing that the estimates of m and u are uncertain due to small sample size).

(a) to obtain a density function of the return period by introducing uncertainty of estimate into the parameters m and u.

(b) fr. the above density to obtain the expected value of the ret n period considering the uncertainty in the parameter estimation.

(c) to show how this approach can be generalized to the return period of flows $T_Q(x)$, (see Davis, et al., 1973)

(d) to demonstrate the necessity of following this approach in terms of a Bayesian decision-theoretic analysis of a water resources design problem.

3.0 Methodology

The return period given by eq.(1) is correct for the true values of m and u. Since we do not know the true values of m and u, we must estimate them from the data available. m is estimated by $\hat{m}$; the average

number of events per season, and $1/u$ is estimated by $1/\hat{u}$, the mean point rainfall over the r seasons considered. The sample data could have been provided by parent distributions with values of m and u not equal to the estimates, (with probability one this is the case). $\hat{u}$ and $\hat{m}$ are merely those values of m and u that have the maximum likelihood of having produced the data. Other values of m and u have a large likelihood of having produced the sample. The likelihood of the values of m and u which produced the data is given by the conjugate distributions (De Groot, 1970, Chap. 9) for the exponential and Poisson distributions.

The distribution conjugate to both the exponential and to the Poisson is the gamma:

$$F(x|a, b) = \frac{b^a \, x^{a-1} \, e^{-bx}}{\Gamma(a)}$$

The values of a and b are different for the exponential and the Poisson. For the Poisson:

$x = m$	the parameter of the Poisson
$b = r$	the number of seasons
$a = \hat{m}r$	the total number of rainfall events in r seasons

For the exponential:

$x = u$	the parameter of the exponential
$a = \hat{m}$	the total number of rainfall events in r seasons
$b = \hat{m}/\hat{u}$	the total amount of rainfall for the $\hat{m}r$ events

These distributions are called posterior distributions; they represent the likelihood that various values of m and u are the values describing the rainfall process which we are observing, after getting the data.

What we would really like is the posterior distribution of the return periods associated with various rainfalls produced by this process. These return periods may be produced by simulation. Equally likely values of m and u are generated by use of the posterior distributions. This enables equally likely return periods to be calculated corresponding to k inches of rainfall: $T_R(k)$. These equally likely return periods may be used to form an empirical distribution for $T_R(k)$. The mean of this distribution, easily calculated, is the expected return period, $E[T_R(k)]$, for a k-inch rainfall.

Similar methods can be used to generate a posterior distribution of rainfall amounts corresponding to a given return period.

4.0 Example

Thunderstorm rainfall in excess of 0.3 inches may be assumed to be runoff producing rainfall (Fogel and Duckstein, 1970). Rainfall in excess of 0.3 inches is distributed approximately as an exponential

distribution. For this example, we assume data are available for 10 and 20 summer seasons, that the average rainfall for these runoff producing storms is 0.69 inches and that the average number of storms per season is 14. Simulations to produce the posterior distribution of the return periods of 2 and 2.5 inch rainfall were done based both on 10 and 20 years of record. The results are shown in Figures 1 and 2. The return periods calculated by equation (1) for the two cases were 6.07 and 20.8 years respectively. The mean return periods of the simulated distributions were 6.64 and 23.2 years, respectively, for 10 years data and 6.76 and 21.9 years for 20 years data. The variances of the simulated pdfs were 6.21, 190.0, 3.00 and 63.8, respectively, for the above four values. With increasing data sample, the variance of the posterior distribution continues to drop; for the 2.5-inch storm the variance of the posterior distribution drops to 25.7 when based on 40 years of data. This is less than 15% of the variance based on 10 years of data.

In actual usage the sample average rainfall per storm event and the sample average number of storms per season will change as more data are obtained. As a general rule the posterior distribution will have smaller variance with an increasing data base as illustrated by this example.

5.0 Application

The economic value of a structure may be related to the return period of a specific k-inch rainfall. To evaluate an existing structure the estimation of the return period for k inches of rain, $T_R(k)$, by equation (1) is a first approximation. At first glance it would seem adequate since the mean of the simulated return periods is a little greater than that calculated by equation (1). However, it is the reciprocal of the return period that is used in many situations to calculate expected yearly loss. Simulation indicates that the posterior distribution of the reciprocals of the return periods has a mean slightly higher than that estimated by the reciprocal to equation (1); see Table 1.

TABLE 1. Reciprocal Return Periods by Equation (1) and Simulation (Mean Value).

Data (years)	Rainfall			
	2 inches		2.5 inches	
	Eq. (1)	Simulation	Eq. (1)	Simulation
10	.1647 (6.07)	.1764 (5.67)	.0480 (20.8)	.0556 (17.99)
20	.1647 (6.07)	.1654 (6.03)	.0480 (20.8)	.0531 (18.8)

Number in parentheses is reciprocal of number above.

For this type of calculation the approximation offered by equation (1) is not always on the safe side. For the cases where shorter return periods have a disproportionate economic impact for the evaluation of the structure than do longer return periods, the estimation afforded by equation (1) will clearly be inadequate. A posterior distribution of the factor(s) affected by the return period would definitely be needed.

A common calculation with regard to existing structures which will fail if there is a rainfall in excess of k inches is the probability of failure (PF) in the next N years:

$$PF(k,N) = 1 - \left(1 - \frac{1}{T_R(k)}\right)^N .$$

As N gets large, small uncertainties in $T_R(k)$ will grow exponentially, and may produce larger uncertainties in the probability of failure. When such probabilities are a design consideration, a posterior distribution of such probabilities should be taken into account before the design is specified.

For many structures the design is given in terms of return period. For a constant return period, simulation can be used to produce a posterior distribution of likely rainfalls. Examination might call for a change in the design flow if the posterior distribution differed appreciably from a spike (delta function), that is, from a distribution with very small variance or spread.

Of course, the fullest use of this type of simulation would be in a complete Bayesian Decision Theoretic analysis (Davis, et al, 1972a). Such an analysis would lead to an optimal decision based on the economics and the uncertainty of the situation; it would also evaluate the expected economic cost of the uncertainty and calculate the value of additional data to reduce the uncertainty.

6.0 Discussion

For small watersheds, areal precipitation is closely related to point rainfall. For larger watersheds, simulation must be used to obtain a posterior distribution of areal precipitation given specific point rainfall data (Davis, et al, 1972b). The return period corresponding to a specific runoff then is uncertain due to at least three uncertainties: the small sample error, the distribution of rainfall over the watershed and the rainfall-runoff relationship. These points will be examined by Davis, et al. (1973).

We believe that such uncertainties are most easily examined by simulation procedures. This type of simulation differs from the usual case in that it is conditioned on the sample statistic rather than on a parameter assumed to be true.

7.0 Conclusions

(a) Calculation of return periods by the use of estimated parameters can be in error. In the example given, the error for the reciprocal of the return period was such as to lead to underdesign.

(b) The return period of point rainfall is subject to considerable uncertainty even with 20 years of data.

(c) Simulation techniques are most useful for studying uncertainties due to small samples. The simulation technique can produce posterior distributions not only of the parameters in question but also of associated economic factors.

8.0 Acknowledgments

This work was supported in part by the following grants: National Science Foundation grant (GK-35791) on Bayesian Decision Theory and two grants from the Office of Water Resources Research (U.S. Interior Dept.) on (a) Decision Analysis of Watershed Management Alternatives and (b) Use of Models to Validate Hydrologic and Other Models. Acknowledged is the enthusiastic, prompt and efficient help of Joel Friedman in the programming, numerical analysis and running of the computer simulation.

9.0 References

Bogardi, I. and F. Szidarovsky. 1972. Safety factor for compensating losses due to uncertainties in hydrologic statistics. Proceedings of this Symposium (Paper No. 8.2).

Dalinsky, J. S. 1970. An unconventional approach to flood frequency analysis. Bull. International Assoc. Scientific Hydrology, Vol. 15(3), pp. 55-59.

Davis, D. R., L. Duckstein, C. Kisiel and M. Fogel. 1973. A decision-theoretic approach to uncertainty in the return period of maximum flow volumes using rainfall data. Proceedings (to appear), International Symp. on Design of Water Resource Projects with Inadequate Data, UNESCO, Madrid, Spain.

Davis, D. R., C. C. Kisiel, L. Duckstein. 1972a. Bayesian decision theory applied to design in hydrology. Water Resources Res., Vol. 8, No. 1, pp. 33-41.

Davis, D. R., C. C. Kisiel, L. Duckstein. 1972b. Optimum design of mountainous raingage networks using Bayesian decision theory. Proceedings, International Symp. on Distribution of Precipitation in Mountainous Areas, World Meteorological Organization, Geilo, Norway.

DeGroot, Morris H., 1970. Optimal Statistical Decisions. McGraw Hill, New York.

Duckstein, L., M. M. Fogel and C. C. Kisiel. 1972. A stochastic model of runoff-producing rainfall for summer type storms. Water Resources Res., Vol. 8, No. 3, June 1972.

Duckstein, L. and C. C. Kisiel. 1971. Efficiency of hydrologic data collection systems: Role of Type I and II Errors. Water Resources Bulletin, Vol. 7, No. 3, pp. 592-604.

Fogel, M. and L. Duckstein. 1969. Point rainfall frequencies in convective storms. Water Resources Res., Vol. 5, No. 6, pp. 1229-1237.

Fogel, M. and L. Duckstein. 1970. Prediction of convective storm runoff in semi-arid regions. Proceedings, International Symp. on Representative and Experimental Watersheds, International Association of Scientific Hydrology, New Zealand.

Fogel, M., L. Duckstein and C. C. Kisiel. 1972. Choosing hydrologic models for management of changing watersheds. Proceedings, Symposium on Watersheds in Transition, Amer. Water Res. Assoc., Fort Collins, Colo.

Hershfield, D. M. 1961. Rainfall frequency atlas of the United States for durations from 30-minutes to 24-hours and return periods from 1 to 100 years. U.S. Weather Bureau, Technical Paper No. 40, 115 pp.

Huff, F. A. 1968. Spatial distribution of heavy storm rainfalls in Illinois. Water Resources Res., Vol. 4, No. 1, pp. 47-54.

Metler, W. 1972b. Bayes risk analysis of regional regression estimates of floods. Master of Science thesis, Dept. of Systems & Industrial Engineering, Univ. of Arizona, Tucson, Arizona.

Metler, W., L. Duckstein and C. C. Kisiel. 1972a. Worth of flood data in culvert design: an expected value approach. Proceedings, Second Annual Meeting of Arizona Section - Amer. Water Res. Assoc. and Hydrology Section - Arizona Academy of Science, Prescott, Arizona.

Bernier, J., "Les Methods Bayesiennes En Hydrologie Statistique," Proceedings of the Int. Hydr. Symp., Sept. 1967, Colo. State University, Fort Collins, Colorado, pp. 454-470.

FIGURE 1

DISTRIBUTION OF POSSIBLE RETURN PERIODS FOR 2.0 INCH SUMMER RAINFALL BASED ON 10 AND 20 YEARS OF DATA

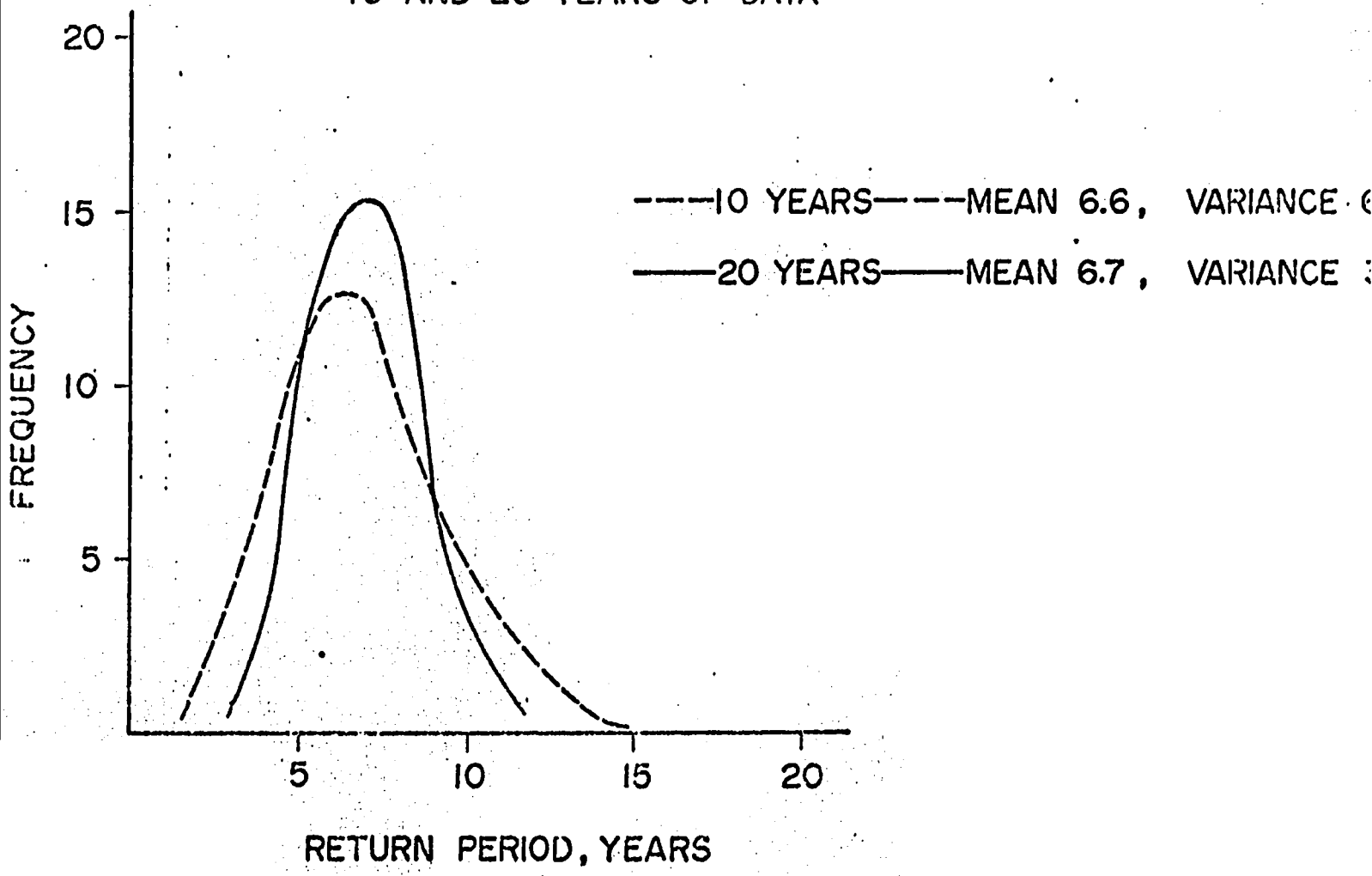
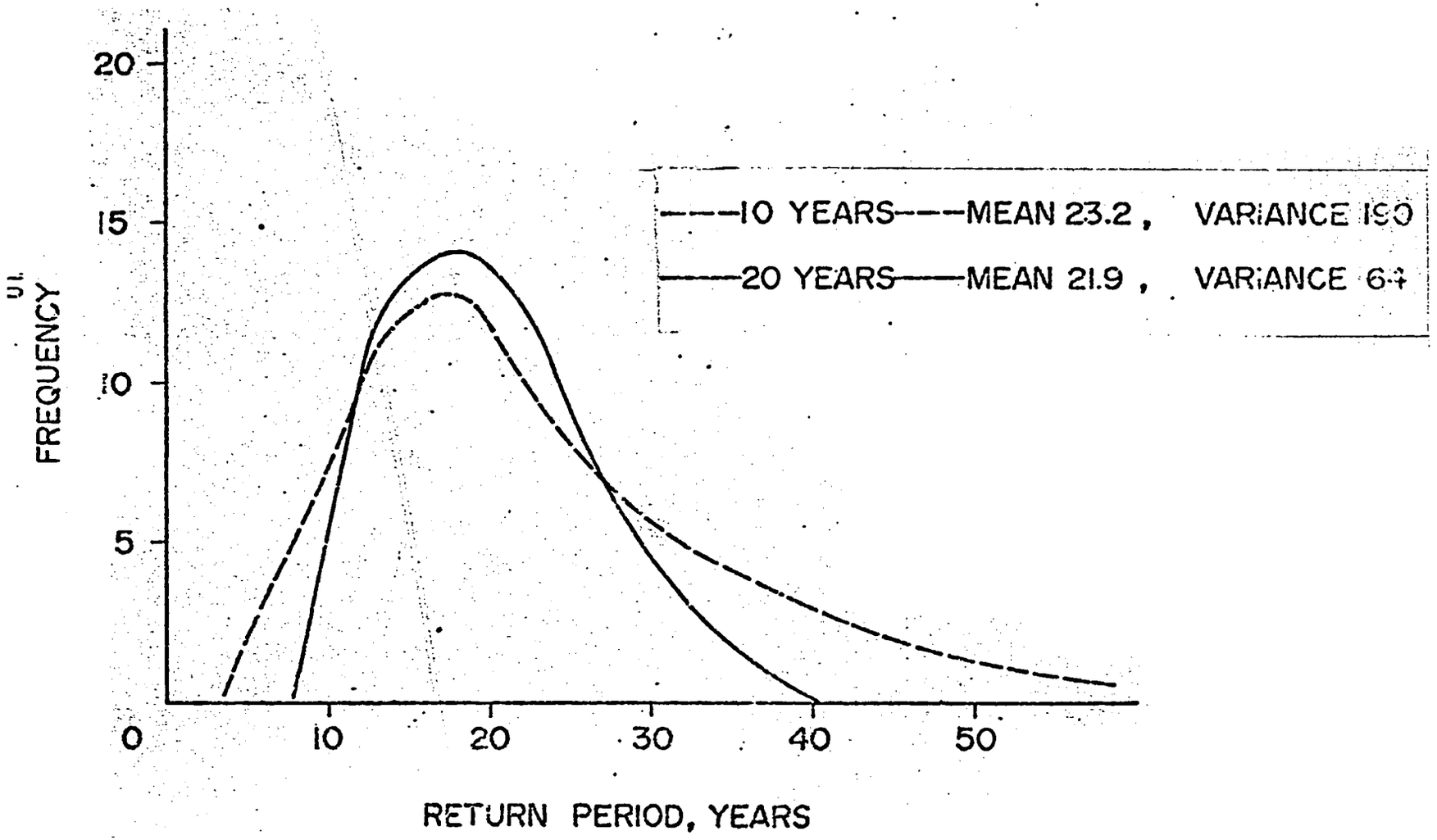


FIGURE 2

DISTRIBUTION OF POSSIBLE RETURN PERIODS FOR 2.5 INCH SUMMER RAINFALL BASED ON 10 AND 20 YEARS OF DATA