Joint Probabilities in the Rainfall-Runoff Relation

LOURENS A. V. HIEMSTRA, The Pennsylvania State University

A vexing problem in surface drainage of highways is assigning of probabilities to peak rates and total volumes of runoff, predicted from rainfall with known probability of occurrence. An attempt to solve this problem was made through considering the rainfall on a catchment and the moisture status of the catchment as stochastic variables. Probability distributions were derived for these stochastic variables. By a random sampling technique using the probability distributions, it was possible to define input and initial conditions, for a probable flood event. A deterministic water budget approach makes it possible to translate the rainfall input into a runoff hydrograph. In this way, series of probable floods can be generated. Analysis of these series of generated floods permits assigning probabilities to the most probable corresponding floods.

It was possible to describe the time pattern of rainfall intensity by means of the incomplete beta-function. The two-parameter log-normal distribution function was a suitable descriptor for all the necessary probability distributions.

The results of this exploratory investigation were tested against annual series of observed flood peaks on two very small rural catchments. Results are positive and promising.

ONE of the vexing problems in surface drainage of highways is the assigning of probabilities to peak rates and total volumes of storm runoff from rural and urban areas. When long historical records of consistent and homogeneous catchment yields are available, this problem simplifies to a problem in the probabilities of a single random variable which is subject to rigorous statistical treatment. However, gaging of runoff from all catchments seems to be an impossible task. For example, in 1963 the estimated 846,000 tributary catchments within the conterminous United States with areas between one and two square miles were represented by less than 60 streamgages (4). Furthermore, the impact of man's cultural practices on catchments might be so severe that it becomes difficult to predict streamflow responses from historical records of streamflow. Hence, trustworthy methods for the estimation of yields from catchments and their probabilities of occurrence based on factors other than direct measurements of runoff are necessary.

Because rainfall records are more freely available and are usually applicable to much larger areas than streamflow records many methods have been proposed to estimate flood peaks from rainfall (2). Most of these methods purposely try to calculate the magnitudes of floods with specified probabilities on the assumption that the resulting flood will have the same probability as the causative rainfall. On most catchments, however, the processes that modify rainfall before it emerges as runoff are of such complexity and variability in both time and space that no presently available method can account for all causative factors with deterministic certainty. In a recent study (8) return periods were attached to 134 observed flood peaks on small catchments in terms of Gumbel's analysis (5) performed on series of annual maxima. Correspondingly observed rainfall return periods for both 30-min and 1-hr durations were assigned

Paper sponsored by Committee on Surface Drainage of Highways and presented at the 48th Annual Meeting.

from the Rainfall Frequency Atlas (6). No relationship was apparent between the return period of an individual rainfall and its associated flood peak.

Consideration of the statistics involved shows that a flood peak will very seldom have the same return period as its causative rainfall (15). For example, let the flood peak Q be a function of the rainfall amount K_1, and other variables K_2, $K_3, \ldots, K_n$, then the probability of occurrence of any peak discharge Q will be a function of these variables, i. e.,

$$P(Q) = f(K_1, K_2, K_3, \ldots, K_n) \tag{1}$$

When the probability distributions of each of these variables are known and the variables are independent of each other, then the probability distribution of Q would be a convolution of the probability distributions of the variables. Let us assume the two-parameter log-normal function is a good descriptor for all these probability distributions and denote the probability distributions by $F_1(K_1), \ldots, F_n(K_n)$ where the probability $P(K_1 \leq K_1^1) = F_1(K_1), \ldots, P(K_n \leq K_n^1) = F_n(K_n)$, respectively. Then the joint probability $P(K_1 \leq K_1^1, \ldots, K_n \leq K_n^1) = F_K(K_K)$ and

$$F_K(K_K) = \int_0^{K_1^1}, \ldots, \int_0^{K_n^1} \frac{1}{K_1, \ldots, K_n \ \hat{\delta}_1^2, \ldots, \hat{\delta}_n^2 \ (\sqrt{2\pi})^n}$$

$$\exp\left[-\sum_{i=1}^{n} \frac{(\ln K_i - \ln \hat{\mu}_i)^2}{2\,\hat{\delta}_i^2}\right] dk_1, \ldots, dk_n \tag{2}$$

where K^1 represents the desired magnitude of the variable K whose probability of occurrence is needed.

An estimate of the mean is represented by $\hat{\mu}$ and an estimate of the standard deviation of the variable under consideration is represented by $\hat{\delta}$.

If the probability distributions of the variables $K_1, \ldots, K_n$ are not identical, the mathematics may become difficult and if the variables are not independent a solution in closed form would border on the intractable. It is also clear from Eq. 1 that the only time when it can be expected that Q and K_1 have the same probability is when K_2, $K_3, \ldots, K_n$ are all constants.

The solution of Eq. 1 is further complicated by the problem of selection of the variables $K_1, \ldots, K_n$, because an exhaustive number of probability distributions for relevant variables is obviously impossible. The runoff-producing system is, furthermore, not static, but dynamic and as the process continues, interdependence between some variables may develop.

Clearly then, a satisfactory solution to the problem of assigning probabilities to flood peaks from known rainfall probabilities cannot be found by purely statistical methods, nor by a purely deterministic approach. A combination of the two approaches, by using statistical methods to find the initial conditions from which to start a deterministic bookkeeping of the water status of the catchment, seems to present a promising avenue of approach.

THEORETICAL CONSIDERATIONS

Runoff from a catchment can be considered as the output from a system with input, system parameters, and state variables. Rainfall, the input, has a stochastic nature. It is fully defined only when the amount of rain, the duration of the storm, the time distribution of rainfall intensities and the areal distribution of the rain at any time are known. All these descriptors can possibly be defined in probabilistic terms, derived from suitable historical records of rainfall. Some of the system parameters can remain constant in engineering time, like the area and slope of the catchment while other

system parameters like the surface roughness, vegetation, etc., may show a seasonal variation as well as the effects of natural catastrophes or man-made changes on the catchment. Selection of system parameters for a particular event is theoretically possible through physical measurement. Finally, the state variables, such as soil moisture status, depth to groundwater and available surface storage at any time, need estimation for their initial values at the onset of the storm. For subsequent values a water budget method should suffice. The initial values of the state variables can again be estimated in terms of the probabilities of occurrence of quantities in historical records.

With the input and state variables defined and the system parameters selected for a particular event, calculation of the output is a purely deterministic process, for example, the reproduction of a recorded historical event. For purposes of prediction, however, the input and the state variables are largely unknown but the system parameters can be estimated. If the probability distributions of the input variables and the state variables are known it is possible to develop a sampling procedure through the probability distributions which may define a probable combination of variables for an event. With this set of variables an output can be generated. This sampling and generation can be repeated until a series of outputs with varying probability are available. Interpretation of the generated outputs will depend on the sampling procedure used, and with a realistic sampling procedure it is possible to assign probabilities to the output.

Sampling Procedure

In the design of drainage structures attention is mostly focused on rare events with return periods in excess of ten years. For such events, it seems reasonable to assume independence between input and state variables; hence a random sampling technique through the probability distributions of the variables is applicable. One such a sampling technique which is conceptually easy to understand, although not necessarily the most economical method, is shown in Figure 1. Say, for example, a flood hydrograph with a 100-yr return period is desired for a problem catchment. Assume further that Figure 1a shows the probability distribution of storm durations for this area; Figure 1b, the probability distribution of the rainfall time intensity patterns; Figure 1c, the probability distribution of the areal pattern; and Figure 1d, the probability distribution of the soil moisture status of the catchment. If these probability distributions fully define the input and state variables of the runoff producing system, the following procedure can be followed:

1. Select randomly a storm duration from Figure 1a. Calculate, or read from known data (such as the Rainfall Frequency Atlas of the United States) the amount of rainfall which can be expected on the average to occur once in 100 years over the selected storm duration.
2. Select randomly from Figure 1b the intensity time pattern to be used.
3. Select randomly from Figure 1c the areal pattern to be used.
4. Select randomly from Figure 1d the soil moisture status of the catchment to be used.
5. With suitable system parameters known from the catchment, generate a hydrograph.
6. Repeat the process for as many selections through Figure 1a-d as necessary, to establish a probability distribution of generated flood peaks.
7. From the generated hydrographs select the hydrograph whose peak rate of discharge forms the mode of the peak distributions. This hydrograph is the most likely hydrograph to have a 100 year return period.

PROBABILITY DISTRIBUTIONS FOR INPUT AND STATE VARIABLES

In the study (7) on which this paper is based, probability distributions were derived for storm durations, rainfall intensity patterns and soil moisture status of catchments. The data used did not lend itself to an analysis of the areal pattern of rainfall over catchments. The derived probability distributions were based on data available in

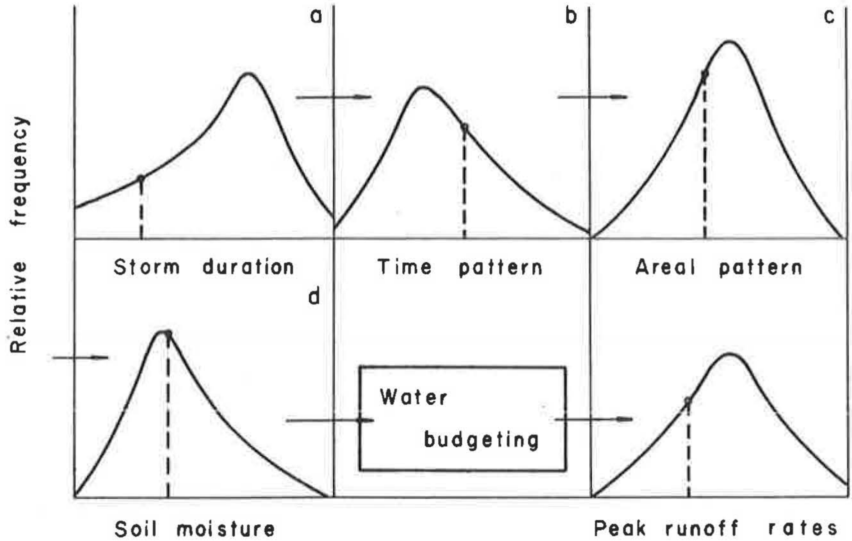

Figure 1. Schematic representation of random sampling procedure.

Colorado State University's collection of flood events on small rural basins (3). Of the 428 suitable events, 388 were selected which were produced by thunderstorm type of rainfall. Most of these storms were from the selected runoff events published by the Agricultural Research Service (17, 18, 19). The selected events were observed on 108 catchments in 25 localities of both the humid and arid regions of the conterminous United States. It was assumed that each thunderstorm event was drawn from a single population of thunderstorms and therefore the statistical parameters of the thunderstorm model were not related to the geographical location.

The catchment areas range from 0.11 through 43.90 sq mi with a mean catchment area of 5.30 sq mi. The rainfall durations were from 12 min through 32.5 hr with a mean duration of 2.65 hr. Storm-rainfall amounts varied from 0.7 through 7.14 in. with a mean of 1.53 in. Soil moisture values were available for only 3 catchments and only for clay and silt-loam soil textures.

Storm Duration

The basis for most design storms in calculations of flood predictions is the optimum storm duration for the problem catchment. This optimum storm duration is often assumed to be a constant for a particular catchment. However, consideration of the flood producing system shows that the optimum storm duration depends on the values of the remaining input variables, the changing system parameters and the state variables, and hence, varies from event to event. Furthermore, the probability that the actual storm duration of a particular event equals the optimum storm duration of the catchment is extremely small. Instead of trying to calculate the optimum storm duration for each event, it was decided to derive a probability distribution for the storm durations of the 388 selected storms.

Criteria for the selection of a probability distribution function for storm durations are (a) the function should be continuous and defined for all positive values of storm durations, (b) the upper end of the function should be unbounded, (c) the density curve must be asymptotic to the axis for large storm durations, and (d) the mathematical form of the curve must be simple enough for easy manipulations. The log-normal density function with two parameters should satisfy all these criteria. If K represents the observed values of the variable under consideration, this function has the following form (1).

$$f(K) = \frac{1}{K\sigma\sqrt{2\pi}} \exp\left[- \frac{(\ln K - \ln \mu)^2}{2\delta^2} \right] \tag{3}$$

with $O \leq K \leq \infty$ and where $\ln \mu$ represents the population mean and σ the standard deviation of the population ln K values.

Estimation of Parameters—According to Eq. 3 and the concept of maximum likelihood the estimator of the population mean is

$$\ln \hat{u} = \frac{1}{N} \sum_{i=1}^{N} \ln K_i \tag{4}$$

and the estimator of the population variance is

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (\ln K_i - \ln \hat{u})^2 \tag{5}$$

where N represents the sample size.

Test of Goodness of Fit on Observed Data—To test the goodness of fit of the theoretical function to the observed data, the concept of equal probabilities of class intervals was used (11). According to this method, with the number of class intervals chosen and using the fact that the total value of the probability integral is unity, the probability of each class interval is determined by

$$P_j = \frac{1}{n} \quad j = 1, 2, \ldots, n \tag{6}$$

where n represents the number of class intervals.

For this value of probability, the required length of any class interval can be obtained.

The Chi-square test was used as a measure of goodness of fit. If the observed class frequency of the n mutually exclusive and exhaustive class intervals is denoted by O_j and the corresponding expected class probability by E_j, then:

$$\frac{(O_j - E_j)^2}{E_j}$$

is Chi-square distributed and with n-1-ν degrees of freedom, where ν represents the number of parameters estimated from the sample data.

By expanding the equation,

$$\chi^2 = \sum_{j=1}^{n} \frac{(O_j - E_j)^2}{E_j} \tag{7}$$

and noting that $\Sigma O_j = \Sigma E_j = N$ (the sample size), and $E_j = P_j N$ where $P_j = \frac{1}{n}$, the following equation is obtained.

$$\chi^2 = \frac{n}{N} \sum_{j=1}^{n} (O_j^2 - N) \tag{8}$$

which is a convenient form for tests of goodness of fit (12).

Results for Observed Storms—Using this procedure on the 388 observed storm durations, it was found that the two-parameter log-normal density function (Eq. 3) with the mean of the log of the modular values

$$\ln \hat{u}_1 = -0.519$$

and the variance

$$\hat{\sigma}_l^2 = 0.909$$

gives a suitable description when tested with the Chi-square test at the 95 percent level. The mean storm duration = 159 minutes.

Time Pattern of Rainfall Intensities

A suitable description for the time pattern of rainfall intensities is necessary before a probability distribution for the patterns can be derived. A suitable mathematical distribution function must be bounded by zero at the one end and by unity at the other end, to describe a dimensionless mass curve of rainfall. The density function must be single peaked with a variety of skewness and kurtosis. Such a function is the beta-function. A suitable form of the beta-function, which is extensively tabulated (14), was derived by representing the complete beta-function by

$$\beta(p,\ q) = \int_0^1 K^{p-1}(1-K)^{q-1}\,dk \tag{9}$$

with $p \geq 0$ and $q \geq 0$ and the incomplete beta-function by

$$\beta_K(p,\ q) = \int_0^K K^{p-1}(1-K)^{q-1}\,dk \tag{10}$$

with $p \geq 0$ and $q \geq 0$; in which case the wanted distribution function is given by

$$F(K) = \frac{\beta_K(p,\ q)}{\beta\ (p,\ q)} \tag{11}$$

where K represents the variable under consideration and p and q the shape and scale parameters, respectively. Figures 2, 4, and 5 show the influence of p and q on the resulting curves.

Estimation of Parameters of the Beta-Function—The theoretical mean of Eq.11 is given by

$$\mu = \frac{p}{p+q} \tag{12}$$

and the variance by

$$\sigma^2 = \frac{pq}{(p+q)^2\ (p+q+1)} \tag{13}$$

from which

$$p = \frac{u\ (u - u^2 - \sigma^2)}{\sigma^2} \tag{14}$$

and

$$q = \frac{p\ (1 - u)}{u} \tag{15}$$

Estimators for p and q are

$$\hat{p} = \frac{\hat{u}\,(\hat{u} - \hat{\mu}^2 - \hat{\sigma}^2)}{\hat{\sigma}^2} \quad \text{and} \quad \hat{q} = \frac{\hat{p}\,(1 - \hat{\mu})}{\hat{u}} \tag{16}$$

where $\hat{u}$ and $\hat{\sigma}^2$ represents estimators for the population mean and variance respectively.

In this study, $\hat{p}$'s and $\hat{q}$'s were calculated for 428 available rainfall events. The method of moments was used to calculate $\hat{\mu}$ and $\hat{\sigma}^2$ on an electronic digital computer.

Goodness of Fit—It can be expected that Eq.11 would describe smooth single-peaked patterns of rainfall intensities very well. Most of the hyetographs used in this study are single peaked but jagged. As no convenient statistical test for the goodness of fit of Eq.11 on the observed hyetographs exists, it was decided to observe visually how well the equation describes 12 randomly selected hyetographs, as shown in Figure 2.

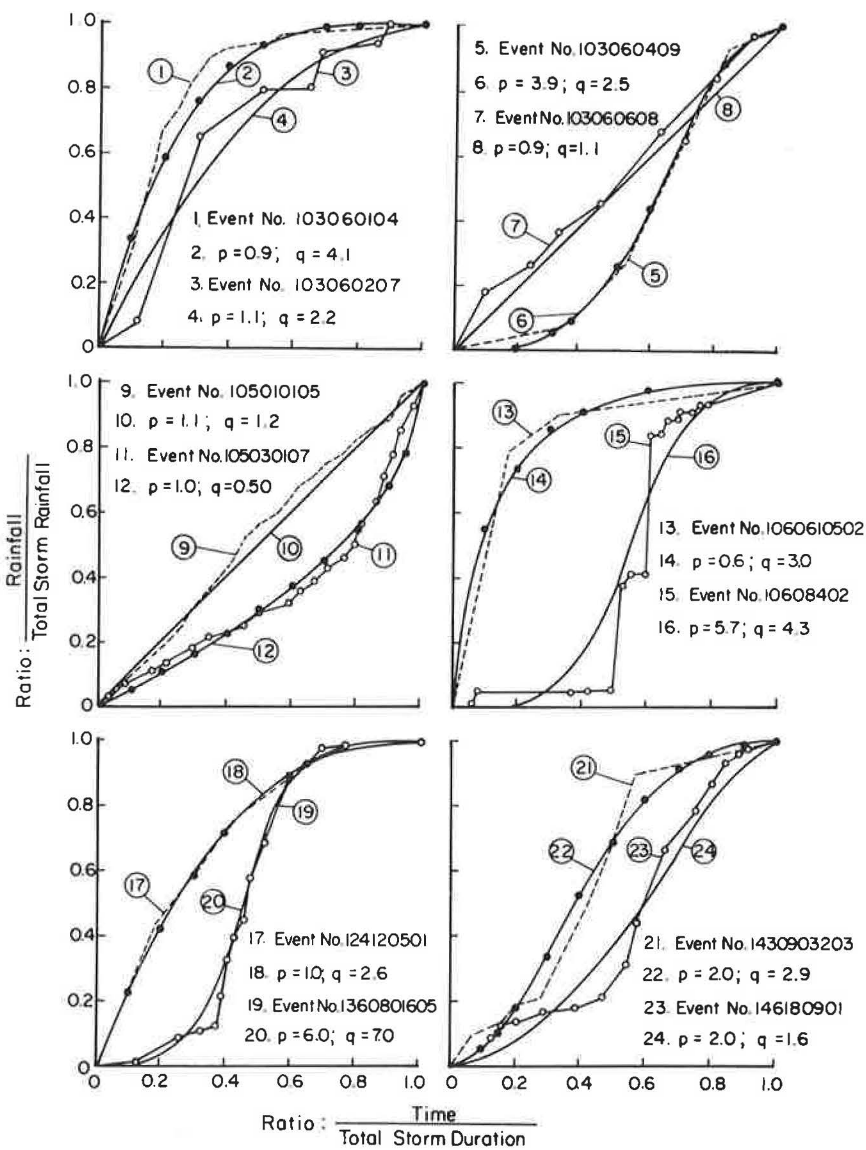

Figure 2. The incomplete beta-function fitted to observed hyetographs.

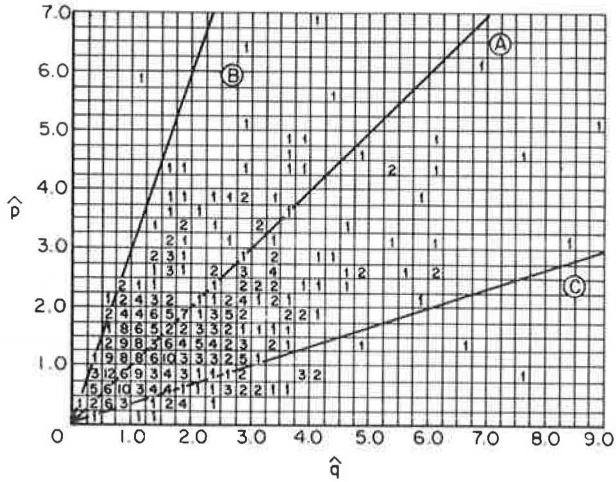

Figure 3. Bivariate frequency distribution of the shape and scale parameters of the beta-function.

The fit on most hyetographs is good, but the maximum intensities on hyetographs Nos. 15 and 23 seem to be significantly reduced by the beta-function. However, the overall fit seems to warrant further pursuance of the method.

Probability Distribution of Shape and Scale Parameters—Since $\hat{p}$ and $\hat{q}$ are paired for each storm the joint distribution of $\hat{p}$ and $\hat{q}$ is of interest in this study. Figure 3 shows a plot of the frequency of occurrences of paired $\hat{p}$'s and $\hat{q}$'s within class intervals with 0.25 spacings. The number in the class interval block represents the frequency of occurrences of each pair within the class intervals which describe the block. The modal paired value is $\hat{p} = 0.9$ and $\hat{q} = 0.6$ which is the description of a late peaking storm.

Since $\hat{p}$ and $\hat{q}$ are not independent but paired, the ratio of $\hat{p}/\hat{q}$ will simplify the bivariate distribution of $\hat{p}$ and $\hat{q}$ into a univariate distribution of this ratio. If the distribution of $\hat{p}$ and $\hat{q}$ within each ratio can be ignored, the simplification obtained will be considerable.

Probability Distributions Within Descriptive Ratios of $\hat{p}/\hat{q}$—In Figure 3, the heavy line A represents the ratio $\hat{p}/\hat{q} = 1$. Line B represents $\hat{p}/\hat{q} = 3$ and line C represents $\hat{p}/\hat{q} = 1/3$. Lines B and C roughly envelope the bivariate distribution of $\hat{p}$ and $\hat{q}$, hence a graphical plot of the hyetographs described by the ratios $\hat{p}/\hat{q} = 3$ and $\hat{p}/\hat{q} = 1/3$ will give an indication of the extremes of late and early peaking storms observed in the sample of 428 storms. A ratio of 2 variables, however, does not define each variable uniquely and variations in the hyetographs with the same ratio can be expected. Hyetographs with $\hat{p}/\hat{q} = 1$ are shown in Figure 4 and hyetographs with $\hat{p}/\hat{q} = 3$ in Figure 5. Hyetographs with $\hat{p}/\hat{q} = 1/3$ will be an early peaking mirror-image of the late-peaking hyetographs of Figure 5 and are not shown. In Figures 4 and 5, the mass curves are plotted as continuous curves and the hyetographs as discrete blocks, in accordance with the common form of published rainfall intensity throughout a storm.

When Figure 4 is studied with Figure 3, it is clear that most of the observed $\hat{p}$'s and $\hat{q}$'s with a ratio of 1 are somewhere between hyetographs 2 and 4. These hyetographs can be expected to have quite different effects on the resulting flood hydrograph, and hence, it is also necessary to find the frequency distribution of the $\hat{p}$'s and $\hat{q}$'s within each ratio. A univariate distribution will result if the product of each pair of $\hat{p}$'s and $\hat{q}$'s with the same ratio is considered. The distribution of $\hat{p}\hat{q}$, when considered together with the ratio of $\hat{p}/\hat{q}$, will then uniquely define each $\hat{p}$ and $\hat{q}$ to be expected.

Probability Distribution of the Ratio $\hat{p}/\hat{q}$ for the Observed Storms—If the time pattern of precipitation intensities in single storms varies according to climatic regions, this variation can easily be detected through a study of the distributions of the $\hat{p}/\hat{q}$ ratios of observed storms in each region.

The conterminous United States can conveniently be divided into two major climatic regions, eg., arid and humid. This was done for the 428 observed storms according to an available published map (9).

The next consideration was the uplifting mechanism responsible for precipitation, which broadly divides orographic from frontal mechanism. Hence, three groups of

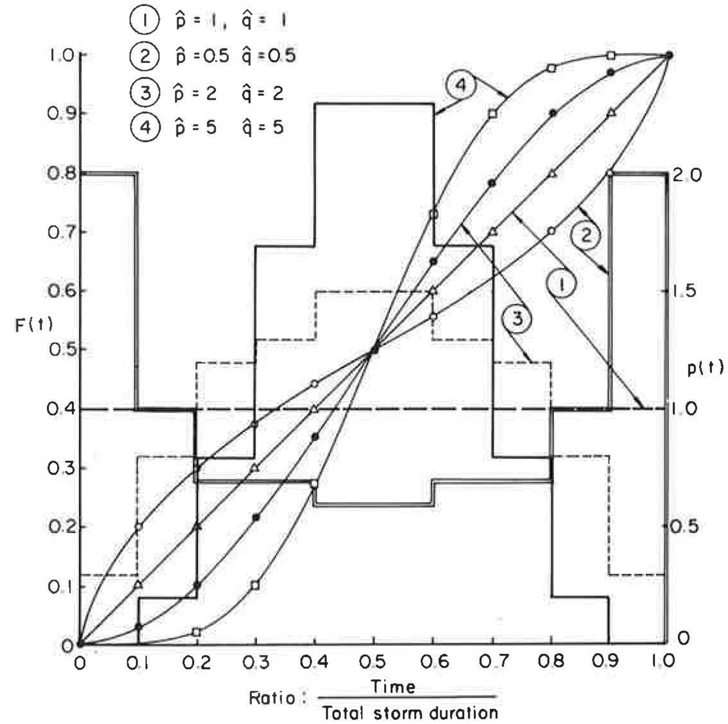

Figure 4. Hyetographs and mass curves with $\hat{p}/\hat{q} = 1$.

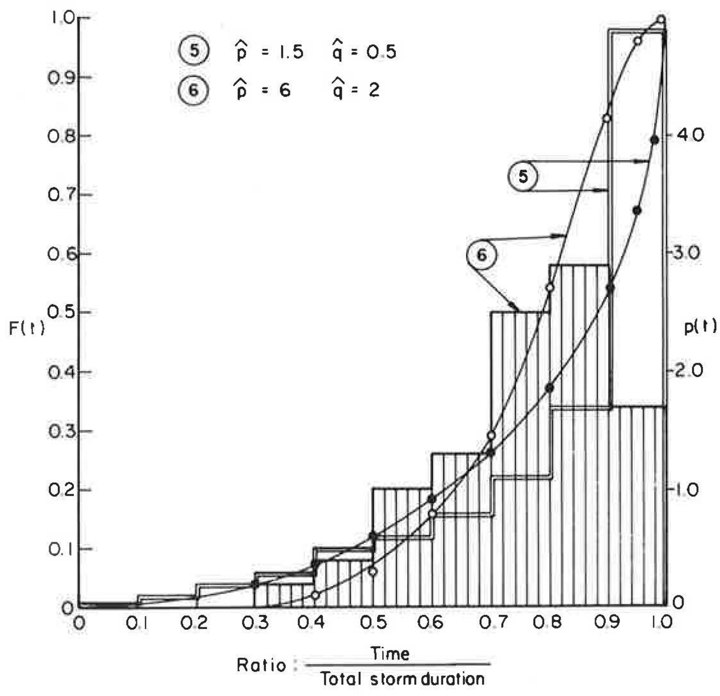

Figure 5. Hyetographs and mass curves with $\hat{p}/\hat{q} = 3$.

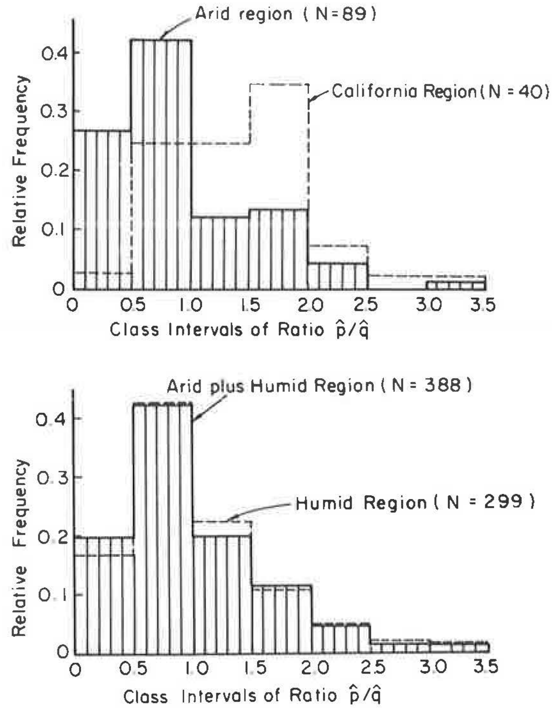

Figure 6. Histograms of $\hat{p}/\hat{q}$.

storm data were formed from the 428 observed storms. First the California watersheds were separated from the rest, as they were mostly in regions under orographic precipitation. The rest of the watersheds were then classified into arid and humid regions. Figure 6 shows the histograms obtained for each classification.

Figure 6a shows a marked difference in the frequency distributions for the arid region and the California watersheds. The storms on the arid and humid watersheds do not show a great difference in frequency distributions, and hence, it seems reasonable to eliminate the California data from the rest and to pool the data from the arid and humid watersheds. In this way, one sample of thunderstorm-type precipitation consisting of 388 observed events was obtained.

Results for Observed Storms—The two parameter log-normal function with $\ln \hat{u}_2 = -0.136$ and $\hat{\sigma}_2^2 = 0.384$ was found to give a satisfactory fit to the observed probability function of the 388 ratios $\hat{p}/\hat{q}$ when tested by the Chi-square test at the 95 percent level.

For the probability distribution of the product $\hat{p}\hat{q}$ within class intervals of the ratio $\hat{p}/\hat{q}$, the 388 selected storms were again used. The two parameter log-normal function was again a suitable description for the observed probabilities.

The three class intervals of the ratio $\hat{p}/\hat{q}$ from, 0 to 0.5, 0.5 to 1.0 and 1.0 to 1.5 were used to get an indication of how well the theoretical function can describe the observed distributions. Results for the three class intervals are given in Table 1.

To test whether the difference between the estimators of the population means or the differences between the estimators of the population variances are significant, the Student -t distribution is convenient.

A useful form of this distribution for this purpose is the following (10):

$$t = \frac{(\hat{u}_a - \hat{u}_b) - (u_a - u_b)}{\sqrt{\frac{\hat{\sigma}_a^2}{N_a} + \frac{\hat{\sigma}_b^2}{N_b}}} \tag{17}$$

and for a test of the variances

$$t = \frac{\hat{\sigma}_a^2 - \hat{\sigma}_b^2}{\sqrt{\frac{\hat{\sigma}_a^2}{N_a} + \frac{\hat{\sigma}_b^2}{N_b}}} \tag{18}$$

If the populations are normal, these statistics follow approximately the Student-t distribution with the degrees of freedom given by

$$\text{d.f.} = \frac{\left(\frac{\hat{\sigma}_a^2}{N_a} + \frac{\hat{\sigma}_b^2}{N_b}\right)^2}{\frac{\left(\frac{\hat{\sigma}_a^2}{N_a}\right)^2}{N_a - 1} + \frac{\left(\frac{\hat{\sigma}_b^2}{N_b}\right)^2}{N_b - 1}} \qquad (19)$$

TABLE 1

ESTIMATORS FOR THE MEAN AND THE VARIANCE OF THE DISTRIBUTION OF THE PRODUCT $\hat{p}\hat{q}$ WITHIN CLASS INTERVALS OF THE RATIO $\hat{p}/\hat{q}$

Class Interval for $\hat{p}/\hat{q}$	N	ln $\hat{\mu}_2$	$\hat{\delta}_2^2$
0 to 0.5	74	0.882	1.143
0.5 to 1.0	163	1.038	1.446
1.0 to 1.5	77	0.746	1.278

Using Eqs. 17 and 18 and assuming the population means $\mu_a = \mu_b$, t = 1.8 with d.f. = 5100. This t is smaller than the t = 1.96 with probability of 0.05, which is a level of significance commonly used.

Hence, the conclusion can be made that there is no significant difference between the means of the distributions for $\hat{p}\hat{q}$ for different ratios of $\hat{p}/\hat{q}$.

Likewise, no significant difference exists between the variances of the $\hat{p}\hat{q}$ distributions. This means a single theoretical distribution can describe the distributions of the products $\hat{p}\hat{q}$ for any ratio of $\hat{p}/\hat{q}$.

With the product $\hat{p}\hat{q}$ and the ratio $\hat{p}/\hat{q}$ known for a rainstorm, the time pattern of rainfall intensities for the storm is uniquely defined in terms of $\hat{p}$ and $\hat{q}$. The probability distribution for the ratios $\hat{p}/\hat{q}$ for the observed storms has ln $\hat{u}_2$ = -0.136 and $\hat{\sigma}_2^2$ = 0.384, and the probability distribution for the products $\hat{p}\hat{q}$ has ln $\hat{u}_3$ = 0.889 and $\hat{\sigma}_3^2$ = 1.289.

Soil Moisture Content

As a third step toward generation of flood peaks, an effort was made to derive a frequency distribution for observed soil moisture contents for the different soil textural classes. Only limited data were available, showing the variations in soil moisture content throughout the year (16). However, the assumption that the soil moisture content on each soil textural class is normally bounded by the saturated soil moisture content at one end and by the soil moisture content at wilting point on the other end, simplified the problem considerably. It is true that the top 12 inches of soil, of importance in this study, may often dry out beyond the wilting point. But this study is concerned with floods which are often more severe on very wet soil than on very dry soil; hence, the assumption of a lower bound at wilting point should be acceptable.

Variations in Soil Moisture Through the Year—The soil moisture variation throughout the year is shown for silt loam in Figure 7 and for clay in Figure 8. Both figures show a characteristic decrease in soil moisture during the active plant growing season, which results in somewhat typical bimodal histograms for soil moisture shown in Figure 9.

Probability Distribution for Soil Moisture Content—If the probability distribution is unimodal, the problem of fitting a theoretical distribution becomes easy, following the procedure already discussed. Hence, it was decided to consider only the higher peaks of the frequency distributions (Fig. 9) and to pool the data for the two years for each of the soil textural classes. This procedure might be acceptable if only wet soil is under consideration for predictions of storm runoff.

If one theoretical probability distribution function can describe the variations in soil moisture for all the soil textural classes, another worthwhile simplification is possible. This will be possible if no significant difference exists between the mean and standard deviations of the modular values for soil moisture used. Following this reasoning, it was found that the soil moisture also follows the log-normal distributions with a χ^2 value = 8.99 which corresponds with a probability $P(\chi^2)$ which is acceptable at the 95 percent level. It was also found that the mean of the log of modular values for clay, ln $\hat{u}$ = -0.0097 and for silt loam ln $\hat{u}$ = -0.0089 which are not significantly different at the 95 percent level, using the Student-t test. Likewise, the variance of the log of the modular values $\hat{\sigma}^2_{clay}$ = 0.0199 and for silt loam $\hat{\sigma}^2_{silt\ loam}$ = 0.0183 which is also not significantly different.

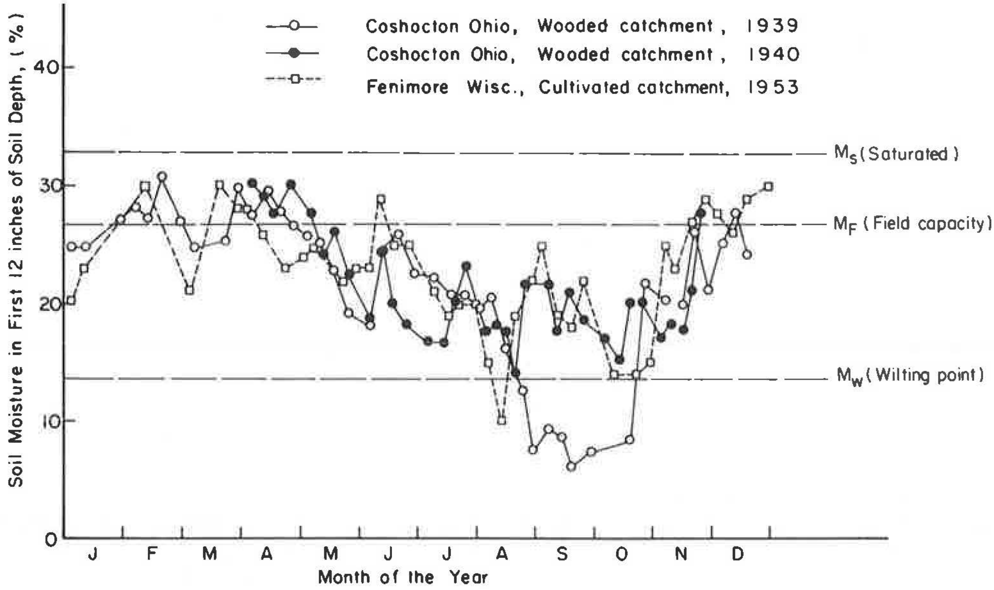

Figure 7. Soil moisture variations throughout the year for silt-loam.

Finally, the mean of the actually observed soil moisture values, for clay is 32.9 percent which is 1.63 times the soil moisture value for clay at the wilting point; and $\hat{u}_{\text{silt loam}}$ = 24.3 which is 1.79 times the soil moisture value for silt loam at wilting point.

Hence, the final assumption necessary to enable extrapolation of these results to other soil textural classes was that the mean soil moisture content is simply 1.71 times the soil moisture content at wilting point.

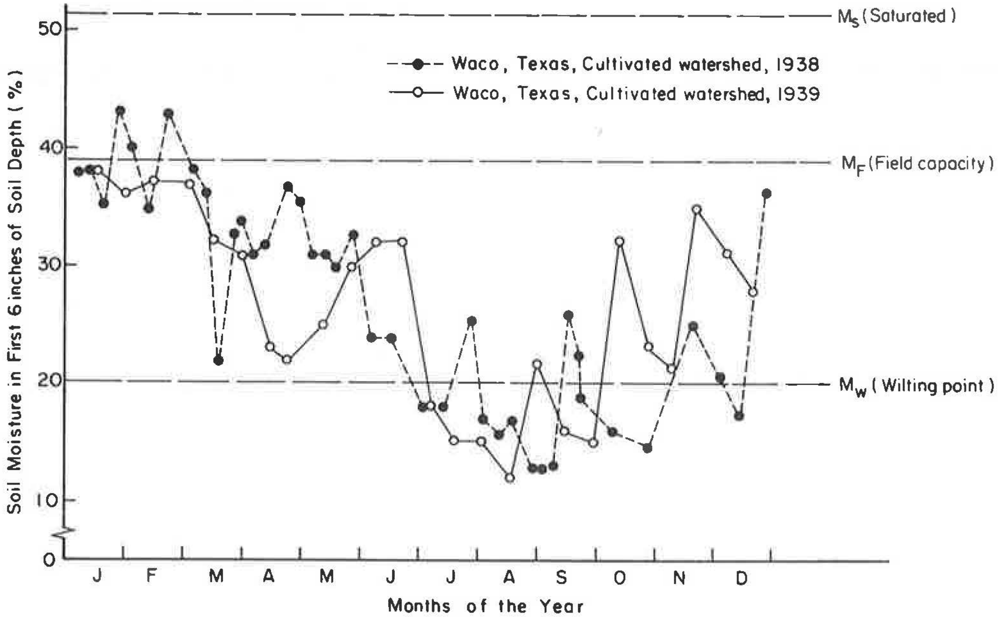

Figure 8. Soil moisture variations throughout the year for clay.

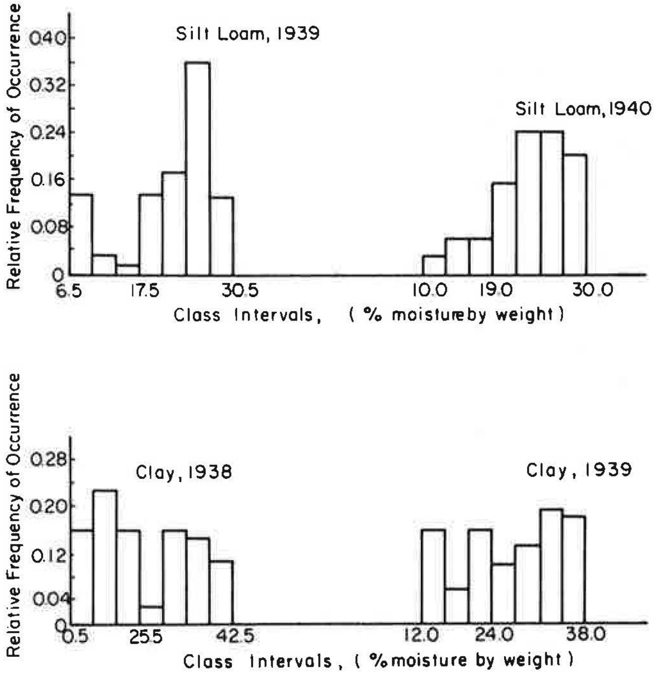

Figure 9. Histograms of variations in soil moisture content.

Conclusion— From the limited data available on variations in soil moisture content for different soil textural classes and by ignoring the moisture content of relatively dry soil, it was found that the observed frequency distribution is log-normal with modular values, $\ln \hat{u}_4 = -0.0093$ and $\hat{\sigma}_4^2 = 0.0191$.

To obtain the actual soil moisture for any soil textural class, the modular value needs multiplication with the mean soil moisture content which was assumed equal to $1.71 \times M_W$, where M_W represents the soil moisture content at wilting point.

Rainfall Amounts

In the generation of flood peaks with desired return periods from rainfall, the rainfall amount to be used is of vital importance. It was decided to use rainfall amounts as calculated for fixed durations by the United States Weather Bureau (6). These calculations were based on Gumbel analysis of partial duration series of extreme rainfall intensities for fixed durations, observed at selected stations in the United States. This analysis is not ideally suited to the purpose of this study, because single-peaked storms with varying intensities during their durations are used. Hence, instead of analyzing only extreme rainfall intensities with fixed durations, this generation of flood peaks needs an analysis of extreme single-peaked rainstorms of fixed durations. As such an investigation was not yet done, it was decided to use the values given in the Weather Bureau publication without any modification. This may result in overprediction of floods resulting from very short duration storms, but as the storm duration increases, the error in predictions should decrease.

In addition, rainfall amounts for fixed return periods must be known for any storm duration. This necessitates interpolation for unpublished values, which was easily done by means of a logarithmic type of regression equation:

$$R_t = g + u \log T_t \tag{20}$$

where R_t represents the total rainfall amount, g and u are constants, and T_t represents the storm duration.

A separate equation linking storm duration with rainfall amount is necessary for each return period at every location where flood peaks are to be generated.

Areal Variation in Rainfall—The storm hyetographs used in this study were Thiessen mean hyetographs. As such the areal variations were taken into considerations. On larger catchments, however, it may be necessary to invent a method which can handle both the areal distribution of rainfall and the storm movement over the catchment. Such a study was beyond the scope of this paper and apart from using Thiessen mean hyetographs the areal distribution of rainfall on the catchments was ignored.

Groundwater Flow—Groundwater flow was also ignored in this study because interest was focussed on the generation of single isolated events. As groundwater flow is a function of the depth of the groundwater table, it should be easy to derive a probability distribution of expected groundwater flows from continuous recordings on bore-holes. If such a probability distribution can be included in this type of analysis it will, without doubt, increase the accuracy of the final results.

FLOOD GENERATION

At this stage, enough is known about the input and state variables of the system so that an attempt can be made to generate floods on catchments with known catchment parameters. With the input, initial conditions and system parameters defined, it remains only to decide on a method which can translate the rainfall into runoff. A deterministic conceptual model (7) was used in this study.

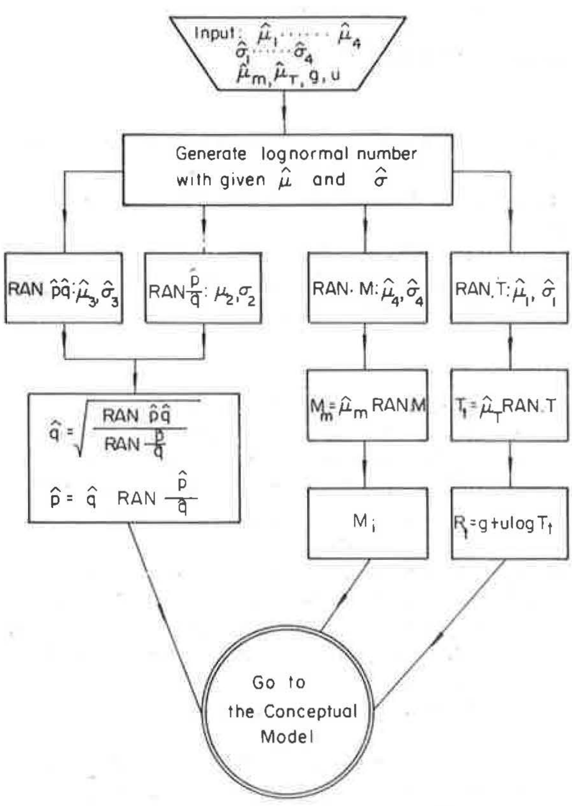

Figure 10. Flow chart for the generation of floods using random numbers (M represents soil moisture content; T, storm duration; R_t, rainfall amount; and g and u, constants for a particular catchment).

Random Sampling Through the Probability Functions—Making use of uniformly distributed random numbers and using suitable transformations, series of log-normally distributed random numbers with previously defined mean and standard deviation can easily be generated on the electronic computer. The computer program for this purpose described by Naylor et al was used (13). The flow chart used in the generation of flood peaks is shown in Figure 10.

Success of Flood Generation—The generated flood peaks for two selected catchments were compared with annual series of observed flood peaks. The results are shown in Figures 11 and 12. In Figure 11, only one exploratory flood peak was obtained by means of random sampling through the probability distributions. The modal value of this generated series of flood peaks for a return period of ten years was 277 cfs. The range of flood peaks obtained was from 80 cfs through 310 cfs. Also shown in Figures 11 and 12 are flood peaks generated by using only the modal values of the probability distributions.

The results obtained, although not very good, are promising. Very good results cannot be expected when the

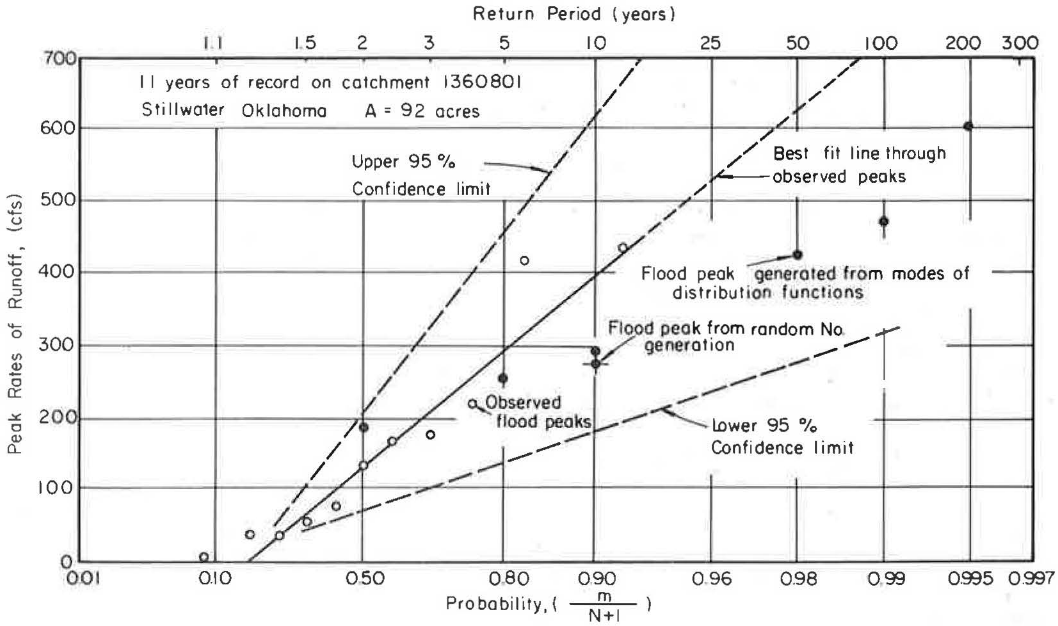

Figure 11. Comparison of observed and generated flood peaks.

probability distributions of the stochastic variables are derived from such a mixed sample of events. Regional analysis of these factors in hydrological homogeneous areas should improve the results considerably.

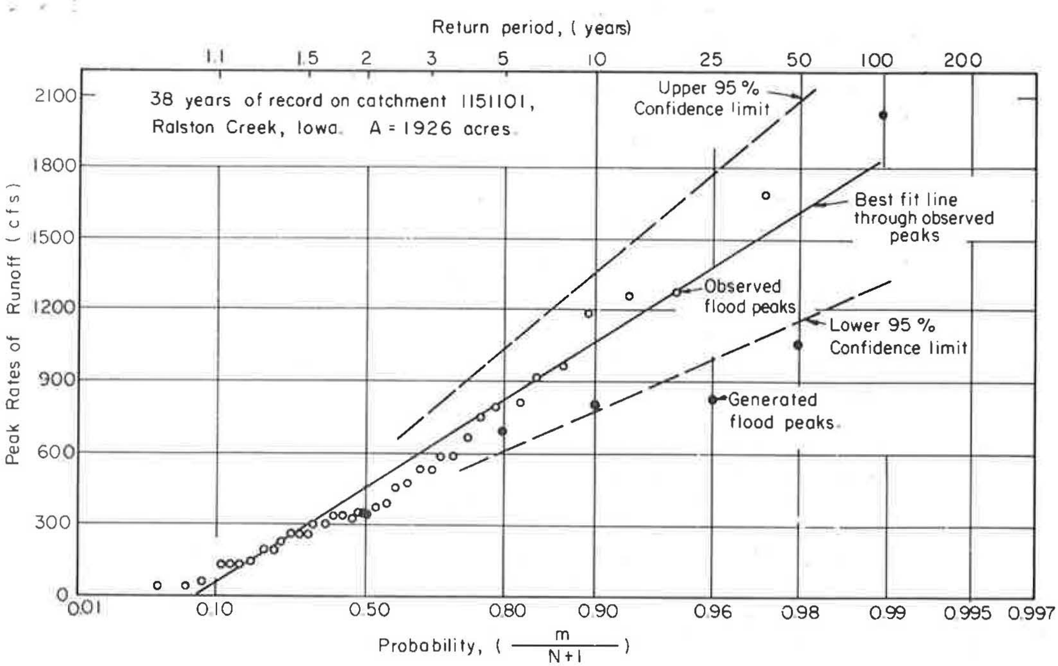

Figure 12. Comparison of observed and generated flood peaks.

CONCLUSION

The results obtained in this exploration of a new technique to assign probabilities to design floods are promising. Refinement in the analysis of historical data on runoff and rainfall is needed before this new technique will reach full fruition.

ACKNOWLEDGMENTS

This paper is based largely on research conducted while the author was working towards his Doctor of Philosophy degree at Colorado State University. He wishes to acknowledge his indebtedness to Walter U. Garstka, Professor of Civil Engineering, and David A. Woolhiser of the United States Department of Agriculture, Agricultural Research Service, for their advice and guidance during the pursuance of this study. Brian M. Reich, Associate Professor of Civil Engineering, The Pennsylvania State University, also earned acknowledgment for his valuable comments on the presentation of this material. The author gratefully acknowledges financial support by the Bureau of Land Management, United States Department of the Interior, and the Colorado Agricultural Experiment Station which made this study possible.

Additional work was done on this project at The Pennsylvania State University by the author. This additional work was supported in part by the Office of Water Resources Research, USDI, under P. L. 88-379 on Project A-015-PA, and the Institute for Research on Land and Water Resources of The Pennsylvania State University. The research is continuing on this general concept and further results will be reported in the final report on Project A-015-PA.

REFERENCES

1. Aitchison, J., and Brown, J. A. C. The Lognormal Distribution. Cambridge University Press, Cambridge, 1963.
2. Chow Ven Te. Hydrologic Determination of Waterway Areas for the Design of Drainage Structures in Small Drainage Basins. Univ. of Illinois, Engineering Experiment Station Bulletin No. 462, 1962.
3. Colorado State University, Department of Civil Engineering. Research Data Assembly for Small Watershed Floods, Part II. Small Watershed Hydrology Program. Publication CER 67-68-13, Sept. 1967.
4. Guisti, E. V. Distribution of River Basin Area in the Conterminous United States. Int. Assoc. of Scientific Hydrology, Gentbrugge, Belgique, Vol. VII, No. 3, p. 20-29, 1963.
5. Gumbel, E. J. Statistics of Extremes. Columbia University Press, New York, 1960.
6. Hershfield, D. M. Rainfall Frequency Atlas of the United States for Duration from 30-Minutes to 24-Hours and Return Periods from 1 to 100 Years. Technical Paper No. 40, U.S. Weather Bureau, Washington, D.C., 1961.
7. Hiemstra, Lourens A. V. Frequencies of Runoff from Small Basins. PhD. Dissertation, Colorado State University, Ft. Collins, Colorado, Publication CED 67-68LAVH21, March 1968.
8. Hiemstra, Lourens A. V., and Reich, Brian M. Engineering Judgment and Small Area Flood Peaks. Colorado State University Hydrology Paper No. 19, April 1967.
9. Hodge, Carl (editor). Aridity and Man. American Assoc. for the Advancement of Science, Washington, D.C., 1963.
10. Li, Jerome C. R. Statistical Inference. Vol. I, p. 142-149, Edwards Brothers, Inc., Ann Arbor, Mich., 1964.
11. Mann, H. B., and Wald, A. On the Choice of the Number of Class Intervals in the Application of the Chi-square Test. Annals of Mathematical Statistics, Vol. 13, p. 306-317, 1942.
12. Markovic, Radmilo E. Probability Functions as Best Fit to Distributions of Annual Precipitation and Runoff. Colorado State University Hydrology Paper No. 8, Ft. Collins, Colorado. Aug. 1965.

13. Naylor, Thomas H., Balintfy, Joseph L., Burdick, Donald S., and Kong Chu. Computer Simulation Techniques. John Wiley and Sons, New York, 1966.
14. Pearsons, Karl. Tables of the Incomplete Beta-function. Cambridge University Press, Cambridge, 1956.
15. Quimpo, Rafael. Discussion of a paper "Purpose and Performance of Peak Predictions" by Brian M. Reich and Lourens A. V. Hiemstra. Proc. The International Hydrology Symposium, Ft. Collins, Colo., p. 365, Sept. 1967.
16. United States Department of Agriculture, Hydrologic Division. Hydrologic Bulletins No. 1, 1941, p. 25; No. 2, 1942, pp. 58-60; No. 4, 1942, pp. 22-23. Office of Research Soil Cons. Service, U. S. Government Printing Office, Washington, D. C.
17. United States Department of Agriculture, Agricultural Research Service. Selected Runoff Events for Small Agricultural Watersheds in the United States. Washington, D. C., 1960.
18. United States Department of Agriculture. Hydrologic Data for Experimental Agricultural Watersheds in the United States, 1956-1959. Compiled by H. W. Hobbs, Miscellaneous Publication No. 945, Washington, D. C., Nov. 1963.
19. United States Department of Agriculture. Hydrological Data for Experimental Agricultural Watersheds in the United States, 1960-1961. Compiled by H. W. Hobbs and Florence B. Crammatte. Miscellaneous Publication No. 994, Washington, D. C., May 1965.