

**INCORPORATION OF UNCERTAINTY  
INTO HYDROLOGIC MODELS**

**C. T. Haan, Regents Professor and  
Sarkeys Distinguished Professor**

**and**

**B. N. Wilson, Associate Professor**

**Agricultural Engineering Department  
Oklahoma State University  
Stillwater, Oklahoma**

**University Center for Water Research  
Oklahoma State University  
Stillwater, Oklahoma**

**December, 1989**

**Acknowledgement:**

The activities on which this report is based are financed in part by the Department of the Interior, U.S. Geological Survey, through the Oklahoma Water Resources Research Institute. Additional Support was provided by the Oklahoma Agricultural Experiment Station as a contribution to Regional Research Project S-211. The contents of this publication do not necessarily reflect the views and policies of the Department of the Interior, nor does mention of trade names or commercial products constitute their endorsement by the United States Government.

## **INCORPORATION OF UNCERTAINTY INTO HYDROLOGIC MODELS**

### **INTRODUCTION**

The design of water resources facilities to alleviate extreme hydrologic events (floods or droughts) is generally based on a frequency analysis of rainfall data. This allows the designer to select the magnitude of the design rainfall event on a probabilistic basis and then estimate the resulting runoff. This practice is well established in hydrologic design. The probabilistic occurrence of rainfall amounts and intensities is but one of several factors that contributes to uncertainty in the expected performance of a water resources system and thus to the risk associated with a failure of the system to meet its intended purpose. For example, catastrophic failure of flood retarding structures may result from the joint occurrence of a rainfall event of severity no greater than the event used in the design of the structure and the occurrence of other factors such as very high antecedent soil water conditions. The net result of this joint occurrence may be a runoff magnitude far exceeding the expected magnitude from the rainfall event.

Improved procedures for quantifying the hazards associated with water resources projects due to rainfall in combination with other factors are needed so that risk analyses can be incorporated into project planning and design. This is especially critical in evaluating the risk of flooding using hydrologic frequency analyses in combination with other factors which contribute to uncertainty in the performance of a hydraulic structure. In simple terms we need answers for the controversy that has existed for a long time concerning the relationship between the return period of rainfall events and the return period of peak discharges or flood flows.

During the three-year life of this project, several technical papers were presented and published. This report summarizes the project accomplishments and refers to other project publications where additional detail on the research can be found.

### **OBJECTIVES**

The design process for flood control projects involves many steps which can basically be reduced to determining the magnitude of the flow against which protection is to be provided and sizing the hydraulic facilities to provide the desired degree of protection. This report addresses the first of these steps. Uncertainty and probabilistic variation in parameters of existing hydrologic procedures were investigated to determine their impact on estimates of return period flows.

The likelihood and consequences of the joint occurrence of such things as intense rainfall and saturated antecedent conditions were studied and quantified. The end product of the research should enable one to construct a relationship between the selected design capacity of a water resources system and the risk of failure of the system taking into account not only precipitation probabilities but uncertainty in the parameters of the hydrologic procedure employed.

The specific objectives of the research were: (1) develop procedures for estimating flood flow frequencies that consider precipitation probabilities and the uncertainty in hydrologic model parameters and (2) investigate the impact of parameter uncertainty on estimates made using hydrologic models.

## METHODOLOGY AND FINDINGS

### Flood Flows

Currently the majority of hydrologic frequency analyses are conducted by either fitting a probability distribution to observed peak flow data or by assuming flood flows of a given return period are produced by rainfalls having the same return period and using a hydrologic model to convert rainfall to streamflow. Rainfall magnitudes for given return periods are generally estimated from published data.

In most situations flow data is either not available or is of such short duration that standard flow frequency analysis is not recommended. Thus the rainfall-hydrologic model approach is commonly employed especially in urban areas. Hydrologic model is being used here in a general sense to include procedures ranging from single equations to complex simulation models.

Rather than assume that the return period of flow is equal to the return period of rainfall, the joint probabilities of occurrence of rainfall and values of parameters in hydrologic models such as antecedent soil water conditions can be incorporated into the hydrologic frequency analysis.

As an example of the approach that may be used, the relationship used by the Soil Conservation Service (1972) to relate runoff volume to rainfall volume is

$$Q = \frac{(R-0.2S)^2}{(R+0.8S)} \quad (1)$$

where Q is runoff volume, R is rainfall volume and S is a measure of available soil water storage. Equation (1) can be solved for S resulting in

$$S = 5R + 10Q - 10 \sqrt{Q^2 + 1.25RQ} \quad R > Q \quad (2)$$

Denoting the joint probability density function for R and S as  $p_{R,S}(r,s)$ , the probability that an observed Q will exceed  $Q_*$  is given by

$$\text{prob } (Q > Q_*) = 1 - \int_0^{Q_*} \int_Q^{\infty} p_{R,S}(r,s) |J| dy dQ \quad (3)$$

where y is a defined variable and |J| is the absolute value of the Jacobian of the transformation between R and S and Q and y.

If R and S are independent, equation (3) simplifies to

$$\text{prob } (Q > Q_*) = 1 - \int_0^{Q_*} \int_Q^{\infty} p_R(r) p_S(s) |J| dy dQ \quad (4)$$

To get the magnitude of Q associated with any return period T, either equation (3) or (4) is solved for  $Q_*$  such that

$$\text{prob } (Q > Q_*) = \frac{1}{T} \quad (5)$$

The distribution of rainfall,  $p_R(r)$ , can be easily evaluated from the Type I Extreme Value distribution and the data published in U.S. Weather Bureau TP 40 (Hershfield, 1961). The distribution of S,  $p_S(s)$ , was estimated from 30 years of data collected on experimental watersheds near Stillwater, Oklahoma, and from data published by the Agricultural Research Service in a series of publications on hydrologic data for small watersheds. Table 1. contains some basic information on the catchments from which data were used.

Table 1. Catchments used for initial studies

Ref #	ARS #	Location	Area(km <sup>2</sup> )	Years
C2630	26030	Coshocton, OH	1.23	29
C2635	26035	Coshocton, OH	10.45	29
H4401	44001	Hastings, NB	1.96	29
H4403	44003	Hastings, NB	8.48	29
S4501	45001	Safford, AZ	2.11	31
S4503	45003	Safford, AZ	3.11	31
Stillwater	37001	Stillwater, OK	0.07	21

The results of this part of the study have been published in Haan and Schulze (1987) and Haan and Edwards (1988). The first of these papers presents a technique for estimating confidence intervals on return period flows if uncertainty in the parameters of the runoff model is considered. It was found that if the SCS curve number approach is used, the 80% confidence intervals on storm water runoff volume can be determined using the antecedent I and antecedent III curve numbers and treating the antecedent II predictions as the expected value of the runoff volume for a given return period.

The paper by Haan and Edwards (1988) extends this work to include uncertainty in extreme rainfall as well as uncertainty in antecedent conditions. Using this approach the joint probability of extreme rainfall and various catchment antecedent conditions is considered through the use of equations 3 - 5. Flow frequency curves based on the standard approach of using a return period rainfall and an average value for the parameter S had the same shape as the curves based on the joint frequency approach. The magnitude of the return period flows estimated by the joint probability approach always exceeded the magnitude of the same return period flow estimated by the standard approach. Based on the seven catchments studied and recognizing the difficulty of comparing frequency curves, it appears that the estimates based on the joint frequency approach are in closer agreement with the observed frequency curves than are the estimates derived through the more conventional approach. Figure 1 is an example of the resulting curves.

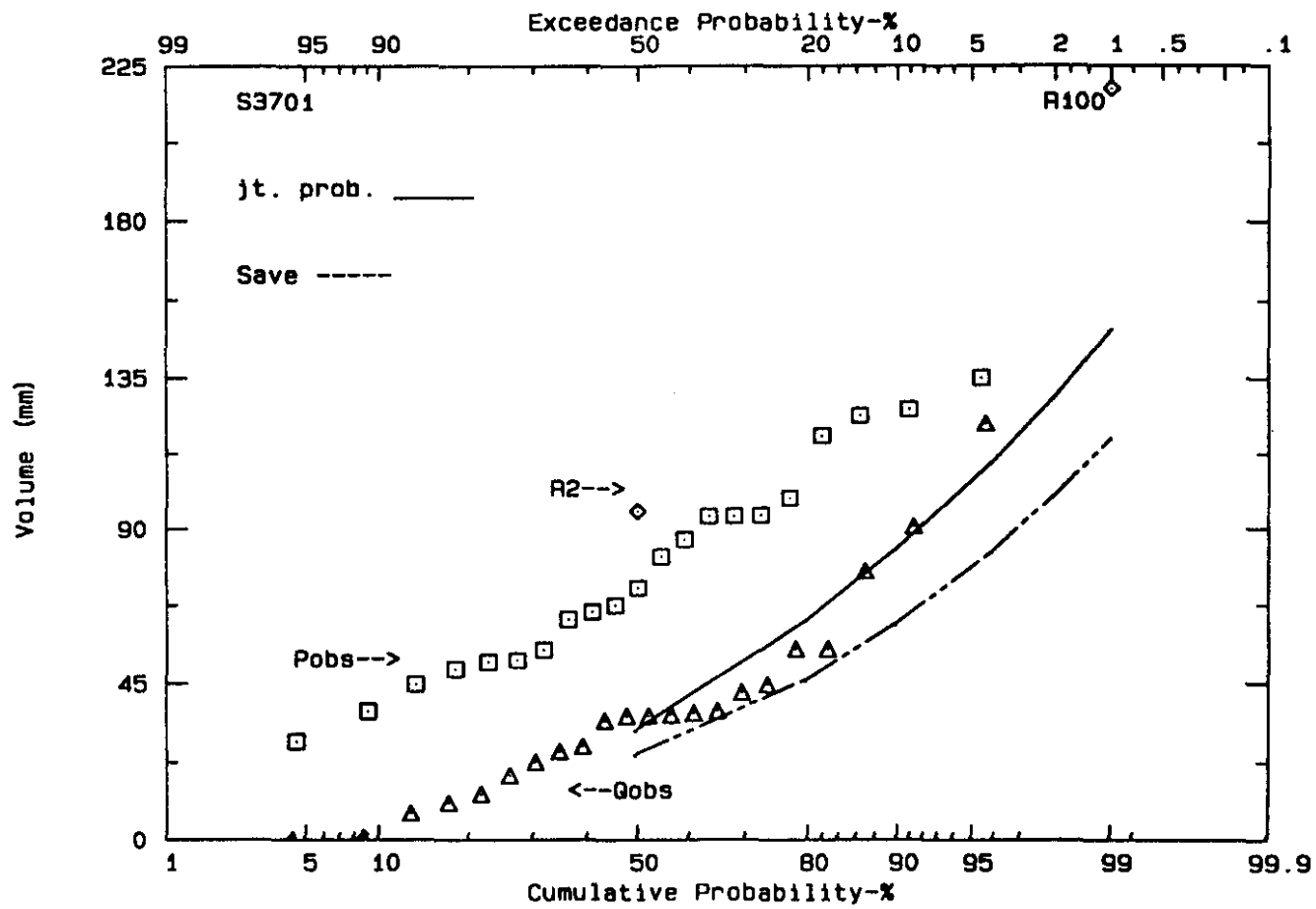


Figure 1. Rainfall-Runoff Frequency Plot for Watershed 53701

The next phase of the research was an attempt to incorporate uncertainty into peak flow estimates from small catchments. The SCS unit hydrograph model (SCS, 1972) was selected for use in this study. The model convolutes incremental runoff volume with a unit hydrograph in order to produce a storm hydrograph. For the purposes of this study, the model is considered as producing two distinct outputs: peak flow (Q) and runoff volume (V). The model parameters which are taken as uncertain are the parameters S (maximum potential soil moisture storage, mm) and  $T_p$  (time to peak of the unit hydrograph, h). Additional model inputs are watershed area and the specification of the rainfall event in terms of its depth, duration, and temporal distribution.

This study used data from 15 watersheds in the Washita River basin of south-central Oklahoma. These watersheds were USDA-ARS experimental watersheds and ranged from eight to 15,747 ha in area. Periods of record for the watersheds ranged from 5 to 15 years. Table 2 summarizes some of the relevant characteristics of these watersheds. Observed values of peak flow and runoff volume were obtained for 50 rainfall-runoff events per watershed. With very few exceptions, these events occurred during the months of April through September, inclusive, and were randomly selected from the sets of all rainfall-runoff events occurring during these months. The exceptions to this rule corresponded to situations in which fewer than 50 rainfall-runoff events occurred in these months over the period of record of a watershed. In these instances, the rainfall-runoff events occurring between April and September were augmented with those occurring nearest these months.

A Bayesian statistical approach was used to analyze uncertainty in the parameters S and  $T_p$  of the SCS unit hydrograph model. Probability density functions of S and  $T_p$ , as well as optimal estimates of S and  $T_p$ , were derived for the 15 watersheds using data on both peak flow and runoff volume. A Monte Carlo procedure was developed to estimate Bayesian flood frequency curves and confidence intervals on these curves for watersheds with short records while explicitly accounting for uncertainty in S and  $T_p$ . The Bayesian flood frequency curves were found to be statistically indistinguishable from the observed curves and compared well to flood frequency curves derived using USGS and unmodified SCS procedures. The theoretical framework and details of this aspect of the study can be found in Edwards and Haan (1989a) and Edwards (1988).

The Bayesian procedure was used to determine estimates of S and  $T_p$  and the associated residuals of peak flow and runoff volume for each of the 15 watersheds. The assumptions of the Bayesian estimation procedure were then checked to determine whether the estimates of S and  $T_p$  could be taken as statistically optimal. In general, residual plots for the 15 watersheds indicated non-constant variance of the residuals violating one of the assumptions of the Bayesian procedure.

Table 2. Summarized Watershed Characteristics

Watershed	Area, ha	Pre-Dominant soil	Cultivation %	Pasture %	Wooded %	Misc. %
111	6734.2	Sandy loam	10	83	4	3
131	10384.6	Sandy loam	21	49	28	2
311	6153.9	Silt loam	36	64	0	0
411	13832.6	Silt loam	75	23	0	0
511	15746.9	Silt loam	58	38	1	3
513	4983.5	Loam	7	85	4	4
5142	145.8	Loam	3	100	0	0
5143	196.6	Silt loam	0	100	0	0
5145	102.3	Loam	0	100	0	0
515	655.6	Loam	31	51	9	10
611	1960.8	Loam/ Silt loam	22	72	5	1
R5	9.6	Silt loam	0	100	0	0
R6	11.0	Silt loam	0	100	0	0
R7	7.8	Silt loam	0	100	0	0
R8	11.2	Silt loam	0	100	0	0

A square root transformation of model predictions and observations was used to attempt to correct the problem of non-constant variance. This corrective action is implemented by defining transformed errors as

$$\eta_Q = Q_o^{1/2} - \hat{Q}^{1/2}$$

$$\eta_V = V_o^{1/2} - \hat{V}^{1/2}$$

where the subscript o denotes the observation and the symbol ^ denotes the estimate. The variance of the transformed residuals was found to be relatively constant.

Figure 2 shows the estimated flood frequency curves (henceforth referred to as the Bayesian flood frequency curves) and their 90% confidence intervals for one of the 15 watersheds. Also shown in this figure is the observed flood frequency curve and the curve resulting from using the USGS procedures (Tortorelli and Bergman, 1985) and the unmodified SCS procedure (SCS, 1972). The observed curve, which is considered the true curve, was determined by fitting the partial duration series of peak flows to a Log-Pearson Type III distribution and adjusting the distribution as described by Chow (1964). The USGS curve is derived from a set of equations which predict the T-year peak flow as a function of mean annual precipitation and watershed area. The SCS flood frequency curve was derived for the eight watersheds having areas of less than 1200 ha using procedures described in NEH-4 (SCS, 1972). The values of S and Tp used to produce the SCS curves were developed following SCS guidelines and are shown in Table 3. A major difference in how the estimated and SCS curves were derived is that the SCS curves are based upon rainfall events having duration of 24 h. The 2- and 100-year, 24-h rainfall used to determine the SCS curves were 94 and 222 mm, respectively.



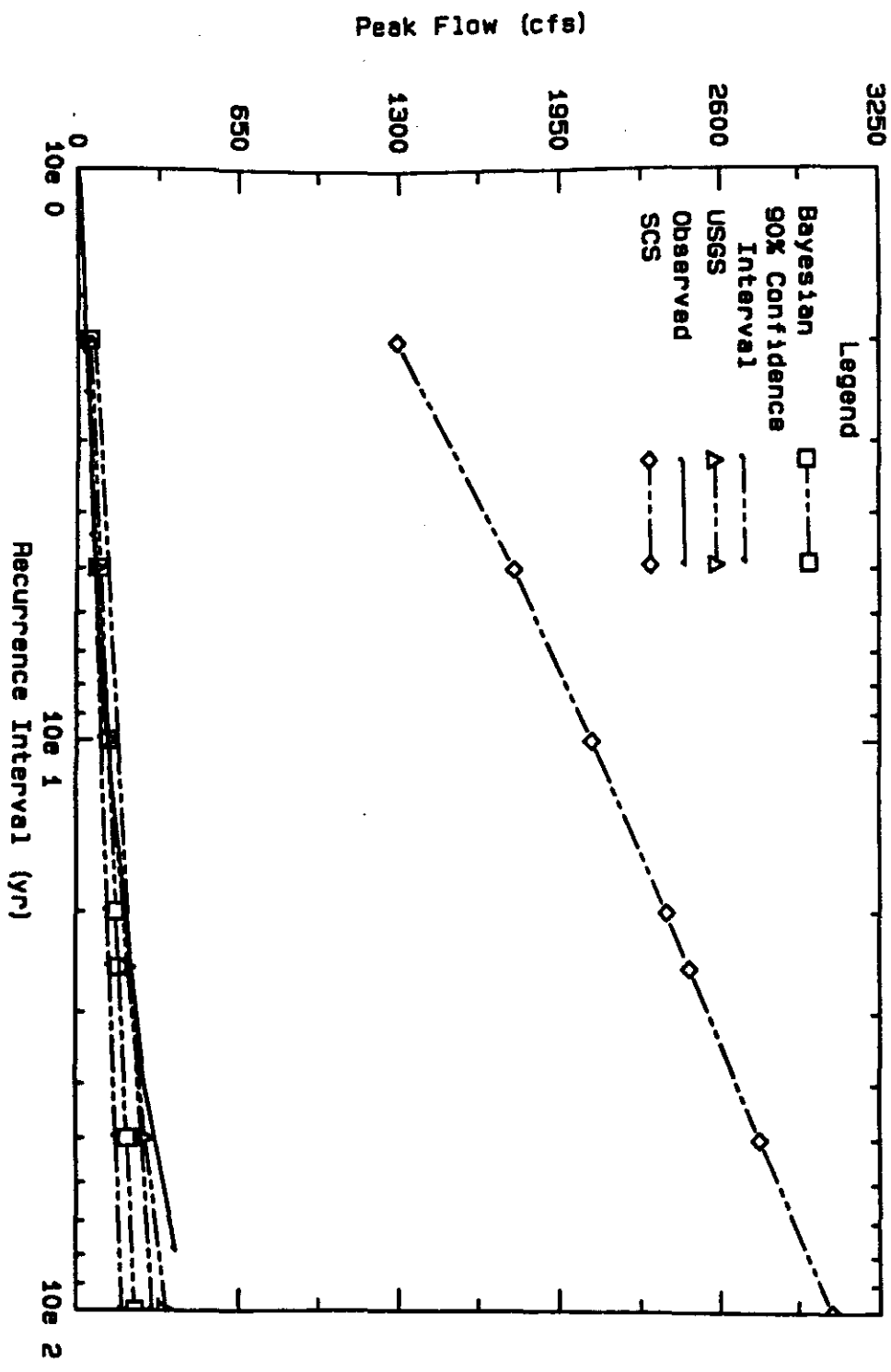


Figure 2. Flood Frequency Curves for Watershed R8

Table 3. Values of S and Tp used to Derive SCS Flood Frequency Curves

Watershed	S, mm	Tp, h
5142	73.7	0.61
5143	78.7	0.81
5145	50.8	0.55
515	63.5	1.41
R5	63.5	0.27
R6	55.9	0.25
R7	33.0	0.13
R8	43.2	0.15

The accuracy of the Bayesian flood frequency curves relative to the observed flood frequency curves varied between watersheds, but no trends of overprediction or underprediction of the T-year flood were observed. Figure 2 presents an example of these curves. Kolmogorov-Smirnov goodness of fit tests conducted at the 0.05 significance level indicate that the Bayesian and observed flood frequency curves may be taken as equal in all 15 cases. However, due to the weakness of the test, each of the USGS and SCS flood frequency curves may also be taken as equal to their respective observed curves. A direct comparison of Kolmogorov-Smirnov test statistics yielded the result that in 8 of 15 cases, the test statistic of the Bayesian flood frequency curve was lower than that of the USGS curve. This should not be considered conclusive evidence that the Bayesian curves perform better than the USGS curves. It only suggests that the Bayesian and USGS curves perform equally well for recurrence intervals in the range of approximately 2 to 10 years. The Kolmogorov-Smirnov test statistic was lower for the Bayesian curves than for the SCS curves for each of the eight watersheds for which SCS curves were computed. Again, this does not prove the Bayesian curves perform better than the SCS curves; it only suggests it. The means of S and Tp were similar to the SCS estimates of these parameters. The apparent relatively poor performance of the SCS flood frequency curves is therefore due in large part to the differences in the rainfall events used to produce the curves.

The uncertainty in S and Tp is transferred to the Bayesian flood frequency curves in the form of the confidence intervals about the curves. High parameter uncertainty translates to wide  $(1-\alpha)\%$  confidence intervals; conversely, low parameter uncertainty will lead to relatively narrow confidence intervals. The capability of specifying these confidence intervals permits one to address questions of risks associated with hydrologic design alternatives from a different perspective than would be possible without the confidence intervals. To illustrate, suppose that one wishes to design a structure capable of handling the 100-year peak flow. Define structural failure as the inability of the structure to accommodate the 100-year peak flow. Simply designing the structure for the point estimate of the 100-year flow does not mean that there is no risk of failure associated with the design. Actually, there is approximately a 50% probability that the true 100-year peak flow is greater than the point estimate. Thus there is a 50% risk of failure should the 100-year event occur. However, if the structure is designed for the upper bound of the 90% confidence interval about the 100-year peak flow, then there is only approximately a 5% risk of failure. The degree of parametric uncertainty has a direct impact in this type of scenario in that it determines the design that must be adopted in order to avoid unacceptable risk, whatever the

magnitude of that unacceptable risk.

Edwards and Haan (1989b) extended this work to ungaged catchments. Uncertainty in model parameters was quantified by considering the parameter to be a random variable and by deriving its probability density function. In the common situation of multiple model parameters, parameter uncertainty is embodied within the joint probability density function of the model parameters, which may be determined by applying principles of Bayesian statistical theory.

The impact of uncertainty in  $S$  and  $T_p$  on peak flow estimates may be assessed by deriving the probability density function of peak flow for a given rainfall event. Because peak flow as computed by the SCS unit hydrograph model is determined from convolution rather than from a single equation, a direct application of the theory of derived distributions is not possible. As an alternative, the probability density function of peak flow is determined by a Monte Carlo simulation which generates multiple values of  $S$  and  $T_p$  from their respective distributions and uses them in conjunction with one rainfall event to obtain an empirical distribution of peak flow. In this study, 2000 pairs of values of  $S$  and  $T_p$  were generated for each rainfall event of interest. The point estimate of peak flow for a given rainfall event is taken as the mean of all peak flows computed for that rainfall event. Additionally, the multiple computations of peak flow allow the specification of  $(1-\alpha)\%$  confidence intervals on the peak flow resulting from a given rainfall event.

If certain assumptions are made, one may make use of the uncertainty in  $S$  and  $T_p$  to estimate flood frequency curves, with confidence intervals, for ungaged watersheds. One such assumption made in this study was that the recurrence interval of a peak flow is equal to the recurrence interval of the rainfall event which produced that peak flow; i.e., a  $T$ -year rainfall event will produce the  $T$ -year peak flow. A  $T$ -year rainfall event must be specified in terms of its duration, depth, and temporal distribution. In this study, the appropriate rainfall duration to use for an ungaged watershed was taken as the time of concentration of the watershed. Upon rearranging relationships between the time of concentration and  $T_p$  given by the SCS (1972), one finds that the time of concentration may be estimated as  $1.5 T_p$ . Given the rainfall duration, one may select a recurrence interval  $T$  and find the depth of the  $T$ -year rainfall event from U.S. Weather Bureau TP-40 (Hershfield, 1961). The temporal distribution of rainfall events was taken as the SCS Type II distribution (SCS, 1986) and rescaled for the appropriate duration. By selecting various recurrence intervals and determining the point estimate and  $(1-\alpha)\%$  confidence intervals for the peak flow of that recurrence interval, one may derive a flood frequency curve, with confidence intervals, for an ungaged watershed.

One may make use of the confidence limits about the Bayesian flood frequency curves to assess risks associated with alternative hydrologic structure designs from other than traditional perspectives. Take, for example, the situation of designing a structure for the 100-year peak flow from Watershed 5142. Define a structural failure as the inability of that structure to accommodate the 100-year peak flow from Watershed 5142. The traditional approach to this problem would be to identify the point estimate of the 100-year peak flow which is 47 cms and to design the structure for a flow of that magnitude. If the 100-year peak flow were known with certainty, this procedure would be appropriate for avoiding structural failure as it has been defined. However, the 100-year peak flow is not known with certainty; it is a random variable whose behavior is described in part by the 90% confidence limits about the flood frequency curve. The true value of the 100-year flood is greater than 47 cms with approximately 50% probability; conversely, it is less than 47 cms with approximately 50% probability. Hence there is about a 50% risk of structural failure if the structure is designed for a flow of 47 cms. The risk of failure can be decreased by

designing the structure for higher capacity. For example, if the structure is designed for a flow of 143 cms (the upper bound of the 90% confidence limit about the 100-year peak flow), then there is only about a 5% risk of failure. Parameter uncertainty affects the capacity that must be designed for in order to meet a given level of risk. As parameter uncertainty increases, the width of the 90% confidence bounds will also increase. This means that a structure must be designed for a relatively greater capacity to have the same risk of failure.

### **Infiltration**

The infiltration rate of water into a soil profile depends on several soil parameters. Soil parameters have been found to be spatially variable and uncertain. Often a lognormal distribution is used to describe the variability in the parameters.

The relationships that govern the movement of water in soil are nonlinear equations. When several sets of parameter values are available for computing infiltration, the average value of infiltration determined by using all of the parameters individually and then computing the average infiltration is not the same as the result obtained by using the average values of the parameters and a single application of the infiltration equations.

This result was demonstrated by Haan (1987). In this paper analytic and Monte Carlo techniques for transforming the probability density functions of soil parameters into probability density functions of infiltration were demonstrated. The importance of preserving the correlations among parameters was also shown.

These results were further confirmed and reported in Haan, Ben Jemaa and Nofziger (1989). In this work it was shown that using the mode of the distribution of model parameters was superior to using the numerical average as far as computing average infiltration is concerned. The work was based on a numerical solution of the Richard's equation governing unsaturated flow through soil.

### **Hydraulic Structures**

Ben Salem and Haan (1989) investigated the effect of hydrologic model parameter uncertainty on the required design flood water storage capacity of a small flood water retarding structure. They found that using average model parameters in a design procedure underestimated the required storage compared to considering the probabilistic and uncertain nature of the model parameters. The uncertainty in model parameters is passed through the model and results in uncertainty in flood storage requirements.

### **Parametric Uncertainty**

Many of the results of this study have been summarized in Haan (1989). The paper addresses uncertainty, parameter estimation, parameters as random variables, and incorporating uncertainty into model results.

## **SUMMARY**

Hydrologic modeling has become commonplace. Currently most hydrologic models provide point estimates. The current trend in modeling is to address problems of parameter estimation and uncertainty in a manner that enables the prediction of the probability density functions of model outputs. Specifying these probability density functions enables one to evaluate system performance in a probabilistic manner.

This approach provides decision makers with considerably more information and makes it possible to estimate the risk associated with a particular decision. Probability density functions of model results can also be used as a model evaluation and selection criteria since models with smaller error variances would generally be preferred all other things constant.

Several techniques are currently available that enable a modeler to estimate the probability density functions of model parameters and model predictions. The techniques will see increasing application over the next several years.

## BIBLIOGRAPHY

- Ben Salem, E. and C. T. Haan. 1989. "Impact of Parameter Uncertainty on Small Flood Water Retarding Structures." Paper presented at 1989 ASAE/CSAE Meeting, Quebec, Canada.
- Chow, V. T., ed. 1964. Handbook of Applied Hydrology. New York: McGraw-Hill.
- Edwards, D. R. 1988. "Incorporating parametric uncertainty into flood estimation methodologies for ungaged watersheds and for watersheds with short records." Unpublished Ph.D. thesis. Agricultural Engineering Department, Oklahoma State University, Stillwater, OK.
- Edwards, D. R. and C. T. Haan. 1989a. "Incorporating Uncertainty into Peak Flow Estimates." Transactions American Society of Agricultural Engineers 32(1):113-119.
- Edwards, D. R. and C. T. Haan. 1989b. "Confidence Limits on Peak Flow Estimates for Ungaged Watersheds." Transactions American Society of Agricultural Engineers 32(1):120-126.
- Haan, C. T. 1987. "The Effect of Parameter Uncertainty on Runoff Volume Estimates." Proceedings International Conference on Infiltration Development and Application. Honolulu, Hawaii, January 6-8, 1987.
- Haan, C. T. 1989. "Parametric Uncertainty in Hydrologic Modeling." Transactions American Society of Agricultural Engineers 32(1):137-146.
- Haan, C. T. and D. R. Edwards. 1988. "Joint Probability Estimates of Return Period Flows." Transactions American Society of Agricultural Engineers 31(4):1115-1119.
- Haan, C. T., F. Ben Jemaa and D. L. Nofziger. 1989. "Impact of Parameter Uncertainty on Infiltration." Paper presented at 1989 ASAE/CSAE Meeting, Quebec, Canada.
- Haan, C. T. and R. E. Schulze. 1987. "Return Period Flow Prediction With Uncertain Parameters." Transactions American Society of Agricultural Engineers 30(3):665-559.
- Hershfield, D. M. 1961. "Rainfall Frequency Atlas of the United States." Technical Paper 40, U. S. Department of Commerce, Weather Bureau, Washington, DC.
- Soil Conservation Service. 1972. "Hydrology" Section 4, Soil Conservation Service National Engineering Handbook, U.S. Department of Agriculture, Washington, D.C.
- Soil Conservation Service. 1972. "Urban Hydrology for Small Watersheds," Technical Release 55. U. S. Department of Agriculture, Washington, D.C.
- Tortorelli, R. L. and D. L. Bergman. 1985. "Techniques for Estimating Flood Discharges of Oklahoma Streams." U. S. Geological Survey Water Resources Investigations Report 84-4358.