1-1998

Effect of Parameter Distributions on Uncertainty Analysis of Hydrologic Models

C. Thomas Haan
Oklahoma State University

Daniel E. Storm
Oklahoma State University

T. Al-Issa
Oklahoma State University

Sandeep Prabhu
Oklahoma State University

George J. Sabbagh
Texas A & M University

See next page for additional authors

Authors
C. Thomas Haan, Daniel E. Storm, T. Al-Issa, Sandeep Prabhu, George J. Sabbagh, and Dwayne R. Edwards

Effect of Parameter Distributions on Uncertainty Analysis of Hydrologic Models

Notes/Citation Information
Published in *Transactions of the ASAE*, v. 41, issue 1, p. 65-70.

Digital Object Identifier (DOI)
https://doi.org/10.13031/2013.17158

EFFECT OF PARAMETER DISTRIBUTIONS ON UNCERTAINTY ANALYSIS OF HYDROLOGIC MODELS

C. T. Haan, D. E. Storm, T. Al-Issa, S. Prabhu, G. J. Sabbagh, D. R. Edwards

ABSTRACT. *Increasing concern about the accuracy of hydrologic and water quality models has prompted interest in procedures for evaluating the uncertainty associated with these models. If a Monte Carlo simulation is used in an uncertainty analysis, assumptions must be made relative to the probability distributions to assign to the model input parameters. Some have indicated that since these parameters can not be readily determined, uncertainty analysis is of limited value. In this article we have evaluated the impact of parameter distribution assumptions on estimates of model output uncertainty. We conclude that good estimates of the means and variances of the input parameters are of greater importance than the actual form of the distribution. This conclusion is based on an analysis using the AGNPS model.*
Keywords. *Hydrologic model, Uncertainty, Parameters, Water quality.*

Developers and users of hydrologic and water quality (H/WQ) models are becoming increasingly concerned about the accuracy of predictions made with these models. Experience has shown that predictions may contain substantial errors. Rather than providing a point estimate of a particular quantity, it may be preferable to provide an interval estimate with an associated probability that the actual value of the quantity will be contained by the interval. Haan et al. (1995) set forth a statistical procedure for evaluating H/WQ models. Two techniques were presented: First Order Analysis (FOA) and Monte Carlo Simulation (MCS). Haan and Zhang (1996) have applied the procedures to watersheds in the Lake Okeechobee Basin in Florida. The importance of incorporating uncertainty analysis into H/WQ models has been emphasized by many authors (Beck, 1987; Reckhow, 1994; Haan et al., 1995; Kumar and Heatwole, 1995; Hession et al., 1996). Morgan and Henrion (1990) provide an excellent general reference on uncertainty analysis of this type.

Article was submitted for publication in March 1997; reviewed and approved for publication by the Soil & Water Div. of ASAE in November 1997.

The work reported here was supported in part through the Oklahoma State University Water Resources Research Institute by the U.S. Department of Interior as authorized by the Water Resources Development Act of 1979 (p.L. 95-467); The U.S. Department of Agriculture, CSRS, Grants Program; and the Oklahoma Agricultural Experiment Station as a contribution to Regional Project S-249. The contents of the article do not necessarily represent the views of the supporting agencies.

The authors are **C. Thomas Haan,** ***ASAE Fellow Engineer***, Professor, **Daniel E. Storm,** ***ASAE Member Engineer***, Associate Professor, **T. Al-Issa,** ***Graduate Research Assistant***, and **Sandeep Prabhu**, Graduate Research Assistant, Biological & Agricultural Engineering Dept., Oklahoma State University; **George J. Sabbagh,** ***ASAE Member Engineer***, Assistant Professor, Agricultural Engineering Dept., Texas A&M University; and **Dwayne R. Edwards,** ***ASAE Member Engineer***, Associate Professor, Biosystems and Agricultural Engineering Dept., University of Kentucky. **Corresponding author:** Prof. C. Thomas Haan, Biological & Agricultural Engineering Dept., 124 Ag Hall, Oklahoma State University, Stillwater, OK 74078; tel: (405) 744-8398; fax: (405) 744-6059; e-mail: haan@agen.okstate.edu.

Parameter values used as input to models are only estimates, since the actual values are not known with certainty. Several researchers have compared the accuracy, applicability and computational demands of various sensitivity and uncertainty analysis techniques. Thomas (1982) discussed the use of Latin Hypercube sampling as a means of obtaining an output probability distribution function (pdf) and cumulative density function (cdf). Doctor (1989) summarized various sensitivity and uncertainty analysis procedures. Rejeski (1993) referred to “modeling honesty” as the truthful representation of model limitations and uncertainties. Reckhow (1994) suggested that all scientific uncertainties must be estimated and included in modeling activities. Beven (1989) and Binley and Beven (1991) have outlined a general strategy for model calibration and uncertainty estimation in complex models. Beven (1993) and Haan et al. (1995) suggested that the inclusion of uncertainty analysis in modeling activities interjects “intellectual honesty” into the effort. Although the concept of random parameter values may be initially rather abstract to a model user, the influence of parameter uncertainty on model outputs can be conveyed quite vividly (Edwards and Haan, 1989).

Determining the uncertainty to assign to input parameters is one of the major hurdles that must be addressed in overall evaluation of uncertainty associated with hydrologic and water quality modeling. We might get some guidance from a user manual, from our own experience, or from the literature. Since many parameters are not directly measurable, it is generally not possible to collect a large, random sample of parameter values and test various pdfs for their ability to describe uncertainty in the parameters. The same can be said about the correlation structure among the parameters of a particular model.

It is often difficult to determine the correct pdf to use to describe particular parameters. Thus, the objective of this study was to investigate the impact of the form of model input parameter pdfs on model output probability distributions predicted by MCS.

PROCEDURE

This section briefly describes the procedures that were used in the study. Added detail can be found in Prabhu (1995) and Al-Issa (1995). The field that was used in this analysis is field WA, a 1.46 ha watershed in northwestern Arkansas with an average slope of 4%. Edwards et al. (1994) presents more details on the watershed. The crop cover for this field is predominantly tall fescue and it has predominantly a Linker Loam soil. The Linker series consist of well-drained, moderately permeable soils. The runoff is medium and the erosion hazard is severe with these soils.

An event based model, AGNPS (Young et al., 1989), was selected for this study. Even though the model is a distributed parameter model, only one cell was used so it functioned as a lumped parameter model rather than a distributed parameter model. AGNPS has many parameters that must be estimated. Prabhu (1995) conducted a sensitivity analysis on the model considering 28 parameters and found that the following eight parameters were the most sensitive in terms of prediction of the model outputs described below: curve number (CN); slope; the Universal Soil Loss parameters P, K and C; soil nitrogen; nitrogen extraction coefficient for runoff; and nitrogen extraction coefficient for leaching. These are the parameters that were considered uncertain in the current study. The CN was included through the use of the retention parameter, S, which is a transform of the CN (S = 25400/CN – 254). Based on Prabhu (1995), the means, standard deviations and most likely pdfs for the eight most sensitive parameters are shown in table 1. The pdfs for K, C, and P were selected as triangular because little information on the actual distribution of these parameters is available and they are bounded on the right and left. Likewise a lognormal distribution was used for the remaining parameters because they are bounded on the left by zero and positively skewed. Since data were not available to define the correlation structure among the parameters, they were assumed independent of each other. A precipitation event of 95 mm on 30 July 1992, was used.

The outputs of concern in this analysis are runoff volume (mm), sediment yield (kg), soluble N in runoff (kg/ha), sediment-bound N in runoff (kg/ha), and sediment-bound P in runoff (kg/ha). Uncertainty in these outputs was investigated using MCS. A batch file procedure was used to generate random observations from the pdfs of the input parameters assuming the parameters were mutually independent. AGNPS was run for each parameter set. The number of runs used for each simulation was 1,500 based on results of Prabhu (1995). The means and standard deviations of the resulting 1500 values of the five outputs were determined and probability plots developed.

Table 1. Initial model parameters and their distributions

Parameter	Mean	Standard Deviation	pdf
S (mm)	67.6	33.8	Lognormal
Slope (%)	4.00	1.20	Lognormal
K	0.24	0.033	Triangular
C	0.012	0.0024	Triangular
P	0.90	0.045	Triangular
Soil N (kg/ha)	0.00112	0.00056	Lognormal
N RO coef	0.05	0.025	Lognormal
N Leach coef	0.25	0.125	Lognormal

Table 2. Probability distributions used*

Parameter	Sim 2	Sim 3	Sim 5	Sim 7	Sim 1	Sim 4	Sim 6
S (mm)	ln	u	nor	tri	ln	ln	ln
Slope (%)	u	u	nor	tri	ln	ln	ln
K	u	u	u	tri	tri	tri	tri
C	u	u	u	tri	tri	tri	tri
P	u	u	u	tri	tri	tri	tri
Soil N (kg/ha)	u	u	nor	tri	ln	ln	ln
N RO coef	u	u	nor	tri	ln	ln	ln
N Leach coef	u	u	nor	tri	ln	ln	ln

* pdf designations:
ln lognormal
tri triangular
u uniform nor normal

The objective was to evaluate the effect that changing pdf assumptions for the input parameters has on the uncertainty of estimated model outputs as quantified in the form of means, standard deviations and pdfs. The base pdfs that were used are shown in table 1. Variations of lognormal, triangular, uniform, and normal distributions were used for a total of seven simulations. The parameter means shown in table 1 were used for all seven of the simulations with the parameter pdfs shown in table 2. Input parameter standard deviations for the various simulations were obtained by multiplying the base standard deviations of table 1 by the ratios shown in table 3.

Simulations 1, 2, 3, 5, and 7 used input parameters having the same means and standard deviations but different pdfs. Simulations 1, 4, and 6 used input parameters having the same pdfs and means but different standard deviations. The standard deviations for simulations 4 and 6 were ½ and ¼, respectively, of the standard deviations for simulation 1 (table 3).

Table 3. Standard deviation ratios with respect to simulation 1

	Sim 2	Sim 3	Sim 5	Sim 7	Sim 1	Sim 4	Sim 6
Std dev ratio	1	1	1	1	1	0.50	0.25

Using a uniform pdf reflects little knowledge of parameter uncertainty except the range of the parameter. With a uniform pdf, any value in the defined range is as likely as any other value. Thus, values near the ends of the range are as likely as values near the center of the range. The normal distribution is problematic for many parameters in that the range of the normal pdf is the whole real line, both positive and negative. The probability of a negative value from a normal pdf can be determined based on the mean and standard deviation of the variable. For example, if the standard deviation is ½ of the mean (the coefficient of variation is ½), the probability of randomly selecting a negative value from the normal distribution is 0.0228. This is the case for the retention parameter, S. Since 1,500 runs were made for each simulation, 34 negative values of S could be expected (0.0228×1500). This actually occurred in 38 of the simulations in which case S was set to zero since a negative S is not possible. If S is zero, runoff equals the rainfall of 95 mm. Simulation 5 actually resulted in 38 values of runoff equal to 95 mm.

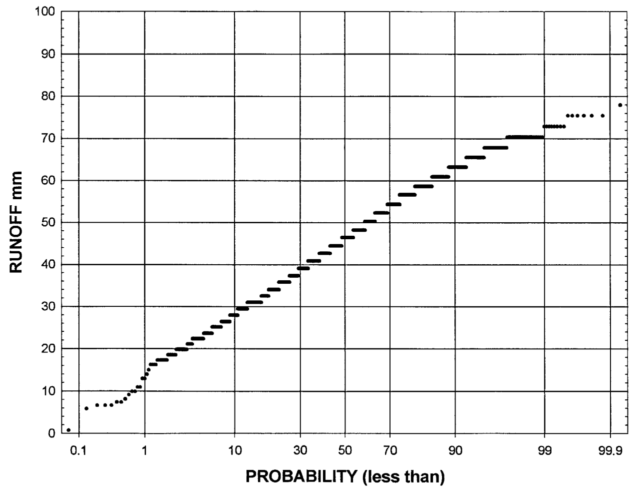

Figure 1–RO distributions from simulation 1.

Table 4. Means and standard deviations of the simulated outputs

		Sim 2	Sim 3	Sim 5	Sim 7	Sim 1	Sim 4	Sim 6
RO	Mean	46	47	47	46	46	44	44
(mm)	Std dev	13	17	17	14	13	7	4
Sed	Mean	264	264	259	261	280	234	231
(kg)	Std dev	157	157	156	156	183	82	46
N RO	Mean	2.8	3.3	3.6	3.1	2.5	1.6	1.3
(kg/ha)	Std dev	3.3	4.2	6.2	4.0	2.5	1.1	0.5
N Sed	Mean	0.54	0.54	0.52	0.53	0.56	0.49	0.49
(kg/ha)	Std dev	0.40	0.40	0.39	0.39	0.44	0.19	0.10
P Sed	Mean	0.26	0.26	0.26	0.26	0.28	0.24	0.24
(kg/ha)	Std dev	0.13	0.13	0.12	0.12	0.14	0.07	0.04

RESULTS

Figure 1 is an example of the output that is generated from each simulation showing 1,500 estimates for RO based on Simulation 1. The 1,500 estimates were ranked and probability plotting positions were determined. Figure 1 is the resulting probability plot. Similar plots were made for each simulation and for each model output. Table 4 shows the means and standard deviations for the 1,500 simulations for each of the outputs and simulations.

Confidence intervals (CIS) can be determined from figure 1. The 90% confidence intervals are found by reading the values at probabilities of 5% and 95%. The width of the CI is the difference in the values at 95% and 5%. For example for RO for Simulation 1, the 95% value is 67 mm and the 5% value is 22 mm resulting in a 90% CI width of 45 mm.

Tables 2 and 3 show that Simulations 1, 2, 3, 5, and 7 had different pdfs for the various input parameters but the

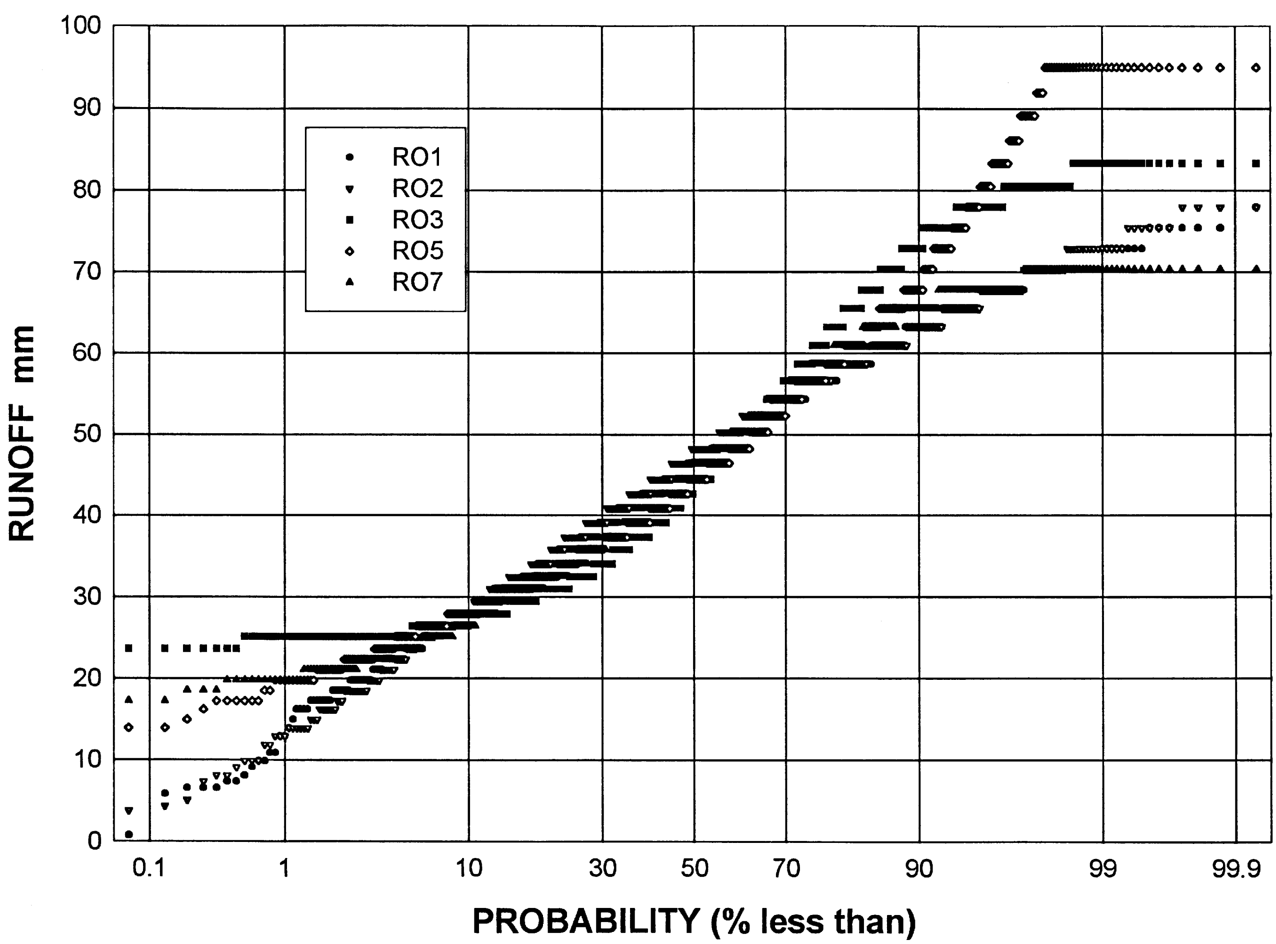

Figure 2–Comparison of RO distributions using various pdfs.

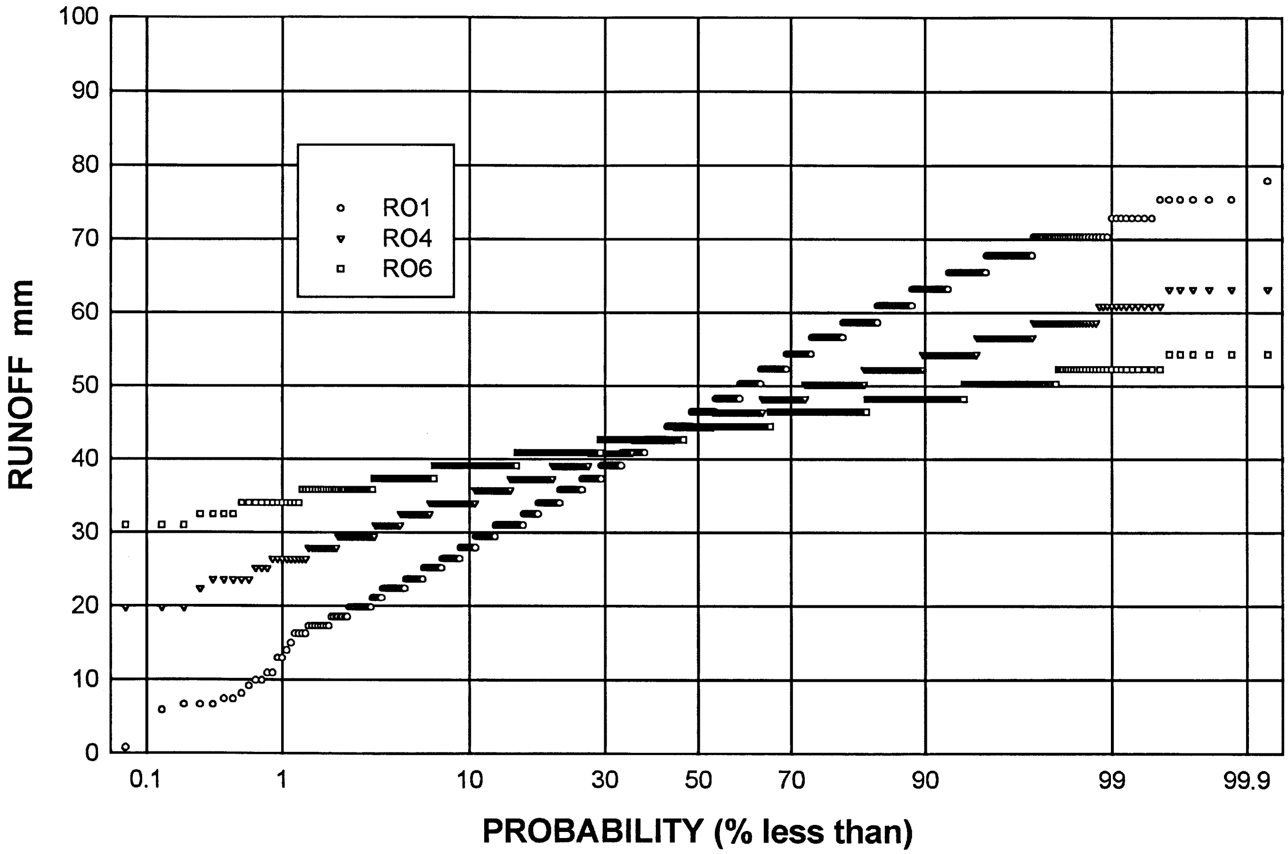

Figure 3–Comparison of runoff distributions using different standard deviations.

same means and standard deviations. It can also be seen that simulations 1, 4, and 6 had the same pdfs and means but different standard deviations. Again using RO as an example, figures 2 and 3 show the pdfs for Simulations 1, 2, 3, 5, and 7 and Simulations 1, 4, and 6, respectively.

The question arises as to how to determine if the various pdfs differ from each other by more than what might be expected by chance. Analysis of Variance techniques and homogeneity of variance tests can be used but these tests may not be valid for this situation. For example, it certainly appears that a normality assumption would not be in order because the pdfs do not plot as straight lines.

We elected to use the nonparametric Kruskal-Wallis (KW) one-way analysis of variance test based on ranks (Conover, 1980). This tests the hypothesis that all of the simulations result in identical distribution functions (pdfs) versus the alternative that at least one of the simulations tends to yield larger values than at least one of the other simulations.

The results of the KW tests were the same for each of the outputs. When all seven simulations were included, the hypothesis of identical pdfs was rejected. When only simulations using parameters with the same means and standard deviations were included (1, 2, 3, 5, 7) regardless of the pdfs, we failed to reject the hypothesis. When simulations with the same pdfs but different standard deviations (1, 4, 6) were used, the hypothesis was rejected. table 5 indicates that the widths of the CIs tend to follow the same trend as the hypothesis tests as one would expect. Except for random variability, the width of the CIs are similar for Simulations 1, 2, 3, 5, and 7 and are different for Simulations 1, 4, and 6.

Table 5. Width of 90% confidence intervals

Output	Sim 2	Sim 3	Sim 5	Sim 7	Sim 1	Sim 4	Sim 6
RO (mm)	46	53	55	44	45	24	13
Sed (kg)	509	499	490	490	554	245	127
N RO (kg/ha)	9.7	12.2	16.3	11.6	7.4	3.3	1.6
N sed (kg/ha)	1.23	1.23	1.23	1.21	1.22	0.58	0.31
P sed (kg/ha)	0.41	0.40	0.39	0.40	0.44	0.21	0.11

These results lead us to conclude that in this particular situation, knowledge of the variance of the input parameters is far more important than knowledge of the exact pdfs. The pdfs used in Simulations 3 were very different than those used in Simulation 7 (uniform versus triangular), yet the output means, standard deviations, width of the 90% CIs, and pdfs tended to be quite similar. The same can not be said when Simulations 1, 4, and 6, having the same pdfs but different standard deviations, are compared.

To determine if the observed differences in pdfs could be ascribed solely to the random nature of MCS, all simulations were repeated three times. Figure 4 shows a typical set of pdfs resulting from the three replications based on runoff. Obviously the variation from simulation to simulation is very small. This indicates that variations that might be attributed to the MCS sample size are quite small in comparison to variations due to the pdf assumption.

Even though we have only used the AGNPS model, we feel the results obtained are typical of what might be found with similar models of this type.

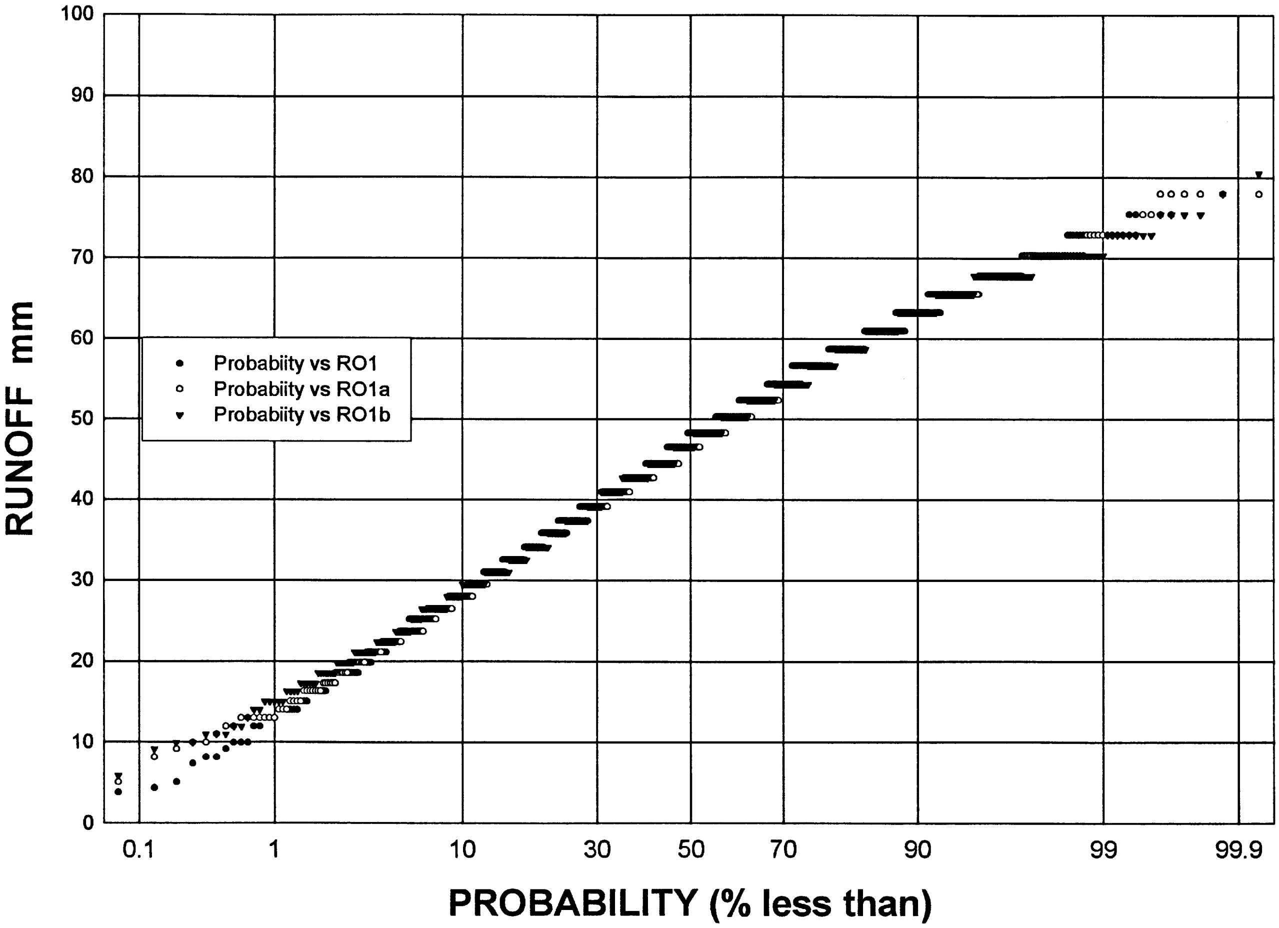

Figure 4–RO distributions showing three replications using same parameters and parameter pdfs.

CONCLUSIONS

MCS using reasonable pdfs to describe model input uncertainty is a valid and powerful method for evaluating the impact of parameter uncertainty on model output uncertainty. A good estimate of the mean and variance of the parameters is more important than the actual pdfs chosen to represent parameter uncertainty. For our study, a sample size of 1,500 simulations produced stable and reproducible results.

Uncertainty analysis using MCS enables one to place confidence limits on model outputs. These confidence limits can then be used as a guide to determine whether a model has the capability to simulate hydrologic and water quality outputs with sufficient accuracy for a given application.

REFERENCES

Al-Issa, T. A. 1995. Impact of parameter probability distribution on model output uncertainty using the AGNPS model. M.S. thesis. Stillwater, Okla.: Biosystems and Agricultural Engineering Dept., Oklahoma State University.

Beck, M. B. 1987. Water quality modeling: A review of the analysis of uncertainty. *Water Resour. Res.* 23(8):1393-1442.

Beven, K. J. 1989. Interflow. In *Unsaturated Flow in Hydrological Modeling*, ed. H. J. Morel-Seytoux. Dordrecht, the Netherlands: D Reidel.

Beven, K. J. 1993. Prophecy, reality and uncertainty in distributed hydrologic modeling. *Adv. Water Resour.* 16:41-51.

Binley, A. M., and K. J. Beven. 1991. Physically-based modeling of catchment hydrology: A likelihood approach to reducing predictive uncertainty. In *Computer Modeling in the Environmental Sciences*, 75-88, eds. D. G. Farmer, and M. J. Rycroft. Oxford, England: Clarendon Press.

Conover, W. J. 1980. *Practical Nonparametric Statistics*. New York, N.Y.: John Wiley & Sons.

Doctor, P. G. 1989. Sensitivity and uncertainty analysis for performance assessment modeling. *Engng. Geol.* 26:411-429.

Edwards, D. R., and C. T. Haan. 1989. Incorporating uncertainty into peak flow estimates. *Transactions of the ASAE* 32(1):113-119.

Edwards, D. R., V. W. Benson, J. R. Williams, T. C. Daniel, J. Lemunyon, and R. G. Gilbert. 1994. Use of the Epic model to predict runoff transport of surface-applied inorganic fertilizer and poultry manure constituents. *Transactions of the ASAE* 37(2):403-409.

Haan, C. T. 1977. *Statistical Methods in Hydrology*. Ames, Iowa: Iowa State University Press.

Haan, C. T., and J. Zhang. 1996. Impact of uncertain knowledge of model parameters on estimated runoff and phosphorus loads in the Lake Okeechobee basin. *Transactions of the ASAE* 39.

Haan, C. T., B. Allred, D. E. Storm, G. J. Sabbagh, and S. Prabhu. 1995. Statistical procedure for evaluating hydrologic/water quality models. *Transactions of the ASAE* 38(3):725-733.

Hession, W. C., D. E. Storm, and C. T. Haan. 1996. Two-phase uncertainty analysis: An example using the Universal Soil Loss Equation. *Transactions of the ASAE* 39(4):1309-1319.

Kumar, D., and C. D. Heatwole. 1995. Probabilistic procedures for evaluating nonpoint source pollution models. In *Proc. Intnl. Symp. on Water Quality Modeling* , ed. C. D. Heatwole, 164-176. St. Joseph, Mich.: ASAE.

Morgan, M. G. and M. Henrion. 1990. *Uncertainty*. Cambridge, Mass.: Cambridge University Press.

Prabhu, S. 1995. Application of stochastic model evaluation protocol on EPIC and AGNPS. M.S. thesis. Stillwater, Okla.: Biosystems and Agricultural Engineering Dept., Oklahoma State University.

Reckhow, K. H. 1994. Water quality simulation modeling and uncertainty analysis for risk assessment and decision making. *Ecol. Modeling* 72:1-20.

Rejeski, D. 1993. GIS and Risk: A three-culture problem. In *Environmental Modeling with GIS*, eds. M. F. Goodchild, B. O. Parks, and L. T. Steyaert, 318-331. New York, N.Y.: Oxford University Press.

Thomas, R. E. 1982. Uncertainty analysis. Battelle Columbus Laboratories, Columbus, Ohio. U.S. Department of Energy, Office of Nuclear Waste Isolation Contract DE-AC606-76-PAO1830-ONWI.

Young, R. A., C. A. Onstad, D. D. Bosch, and W. P. Anderson. 1989. AGNPS: A nonpoint-source pollution model for evaluating agricultural watersheds. *J. Soil and Water Conserv.* 44(2):168-173.