Assessment of a Probabilistic Scheme for Flood Prediction

Faisal Hossain, M.ASCE,[1] and Emmanouil N. Anagnostou, M.ASCE[2]

Abstract: This study presents the development of a probabilistic discharge prediction scheme based on an uncertainty framework called generalized likelihood uncertainty estimation (GLUE). By being explicit about a hydrologic model's parameter uncertainty, historical data is used adaptively on a storm-to-storm basis to derive ensembles of representative parameter sets, along with the corresponding likelihood weights of discharge prediction quantiles. The quantile with highest likelihood weight represents the most probable discharge hydrograph, with upper/lower uncertainty limits represented by the various upper/lower likelihood weight quantiles. On the basis of new data, the Bayesian theorem is used to update for the posterior representative parameter sets and likelihood weights of prediction quantiles. The probabilistic scheme is evaluated using 15 flood-inducing storms over a medium-sized watershed in northern Italy. The scheme's discharge predictions on the basis of its highest likelihood quantile are evaluated comparatively to the conventional single optimum parameter set prediction. It is observed that the two methods have comparable accuracy in terms of the overall hydrograph prediction, but the probabilistic scheme is subject to 50% less variability in time to peak error. The probabilistic scheme has an added value important to decision making and risk assessment, which is its ability to provide consistent assessment of uncertainty in such major flood parameters as peak runoff and time-to-peak. The procedure is simple in design, model-independent, and can be easily implemented in real-time for computationally efficient rainfall-runoff models.

DOI: 10.1061/(ASCE)1084-0699(2005)10:2(141)

CE Database subject headings: Floods; Predictions; Uncertainty analysis; Probabilistic methods; Decision making.

Introduction

The need to address hydrologic prediction of flood events in a probabilistic framework has been argued for a long time. Such a framework should have two main aspects: (1) characterization of the sources of uncertainty and, (2) verification of the predictive probability distribution functions on the basis of observed data. Although much research has been conducted in the last 20 years to address these two aspects (Kitanidis and Bras 1980; Day 1985; Georgakakos 1987, among others), most operational systems worldwide produce largely deterministic hydrologic flood predictions. Murphy and Carter (1980) have shown that probabilistic prediction schemes can enable the evaluation of predictive uncertainty and thus support a more rational decision making. In view of this realization, the U.S. National Weather Service (NWS) first implemented an ensemble streamflow prediction system (ESP) that is based on the framework of Day (1985). Most recently, a small-scale prototype end-to-end probabilistic prediction system was implemented by NWS (Schaake et al. 2001) and is based on two distinct approaches. The first is the Bayesian forecasting system (BFS) of Krzysztofowicz (1999), and the second is an enhancement of the earlier ESP model. These two approaches are probably the only probabilistic systems developed and implemented in an operational discharge prediction system (Seo et al. 2000).

Both BFS and ESP approaches to probabilistic prediction of discharge have limitations of their own. The BFS approach amounts to solving quasianalytically (with via the total probability law) for the conditional probability distribution of the future river stage, given initial and boundary conditions. The key to the solution is to decompose the conditional probability distribution into two components—one because of hydrologic model uncertainty and the other because of uncertainty in input. These two components are integrated in the final stage of river-stage prediction. The basic idea behind the BFS scheme is to blend prior and posterior information using Bayes' theorem. Schaake et al. (2001), however, notes that at present, it is not entirely clear what procedure is best for such blending of information on an adaptive basis. Apart from being analytically very complex, the BFS requires a very long climatic record of streamflow data to determine the mean and variance of the prior distribution of river discharges (Krzysztofowicz 1999). This also imposes an additional limitation, since many model applications need to be made for watersheds without surface hydrologic data. Earlier Bayesian analyses of various forecast data also have suggested that certain version of a BFS system can be unsuitable for river-stage prediction of flood events (Kelly and Krzysztofowicz 1994) but may be suitable for seasonal snowmelt runoff volumes (Krzysztofowicz and Reese 1991). The ESP system, on the other hand, amounts to generating a number of ensembles of traces of future precipitation on the basis of an uncertainty framework and running them numerically through a hydrologic model. From the resulting multiple hydrographs, several probabilistic statements can be made about the future river discharge (Schaake and Larson 1998). Although the ESP scheme is simpler in design, it suffers from two main limitations: (1) quantitative estimation of precipitation fore-

[1]Dept. of Civil and Environmental Engineering, Tennessee Technological Univ. Cookeville, TN 38505-0001. E-mail: fhossain@tntech.edu

[2]Associate Professor, Dept. of Civil and Environmental Engineering, Univ. of Connecticut, 261 Glenbrook Rd., U 2037, Storrs, CT 06269 (corresponding author). E-mail: manos@engr.uconn.edu

Note. Discussion open until August 1, 2005. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on September 8, 2003; approved on April 29, 2004. This paper is part of the ***Journal of Hydrologic Engineering***, Vol. 10, No. 2, March 1, 2005. ©ASCE, ISSN 1084-0699/2005/2-141–150/$25.00.

cast uncertainty is often not available for most watersheds, and (2) the method typically underestimates the uncertainty in river-stage prediction because it does not account for all sources of uncertainty (Krzysztofowicz 1999).

This study was prompted by the need to address some of the limitations of present probabilistic schemes. It is approached through developing and assessing an alternative probabilistic scheme with the following desirable features as design objectives:

- It should have moderate computational efficiency.
- It should be independent of a hydrologic model.
- It should be conceptually simple to understand and implement without needing a long climatic record of streamflow and precipitation data.

The scheme builds upon an established framework for uncertainty assessment of hydrologic models called generalized likelihood uncertainty estimation (GLUE; Beven and Binley 1992). GLUE is a Bayesian approach to uncertainty estimation for nonlinear hydrological models that recognizes explicitly the equivalence, or near-equivalence, of different parameter sets in the model's representation of hydrological processes. GLUE was chosen because of its widespread use and ease in implementation, as has been demonstrated in a number of studies dealing with hydrologic prediction uncertainty (Romanowicz et al. 1994; Freer et al. 1996; Fisher and Beven 1996; Franks and Beven 1997; Franks et al. 1998; Beven and Freer 2001; Hossain et al. 2004a, among others).

The paper is organized as follows. In the next section, we present a description of data, the watershed, and the hydrologic model used in the study. This section is followed by a brief description of the GLUE method, a description of the GLUE-based probabilistic discharge prediction scheme, and a description of the assessment framework used for evaluating the probabilistic and deterministic model predictions. Next, we discuss the results, and the last section summarizes our major conclusions and discusses extensions of this work.

Study Area, Hydrologic Model, and Data

The watershed chosen for this study (named Posina) is located in northern Italy, close to Venice (Fig. 1, right panel). Posina has an area of 116 km^2 and altitudes ranging from 2,230 to 390 m at the outlet (Fig. 1, left panel). Within a radius of 10 km from the center of the watershed, a network of seven rain gauges provide representative estimates of the basin-averaged hourly rainfall. The mean annual precipitation accumulation derived from the gauge network is estimated to be in the range of 1,600–1,800 mm. Posina is 68% forested and saturation-excess is the main rainfall-runoff generation mechanism. Further details about the study area, including its terrain characteristics and rain climatology, can be found in the work by Borga et al. (2000).

The rainfall runoff model TOPMODEL (Beven and Kirkby 1979) was chosen to simulate the rainfall-runoff processes of floods in the Posina watershed. It is a semidistributed watershed model that can simulate the variable source area mechanism of storm runoff generation and incorporates the effect of topography on flow paths. This model makes a number of simplifying assumptions about the runoff generation process that are thought to be reasonably valid in this wet and humid environment. The model is premised on the following two assumptions: (1) that the dynamics of the saturated zone can be approximated by successive steady-state representations; and (2) that the hydraulic gradient of the saturated zone can be approximated by the local surface topographic slope. These assumptions are most likely to be appropriate for mountainous regions with dense and humid vegetation where the typical rain rates rarely exceed the potential infiltration rate of the soils. Detailed background information of the model and applications can be found in Beven et al. (1995). The model has been applied in past studies in the study region and found adequate to simulate the rainfall-runoff transformation process of the watershed with a correlation coefficient exceeding 0.90 (Borga et al. 2000; Hossain et al. 2004a, 2004b). As with many other TOPMODEL applications, the topographic index $\ln(a/\tan\beta)$ is used as an index for hydrological similarity, where a is the area draining through a point, and $\tan\beta$ is the local surface slope. In this study, the derivation of the topographic index from a 20 m grid size digital terrain model used the multiple flow direction algorithm by Quinn et al. (1991, 1995). The important model parameters common to any watershed are T_0, $[\ln(m^2/h^{-1})]$ the lateral transmissivity; XK_0, $(m \cdot h^{-1})$, the vertical

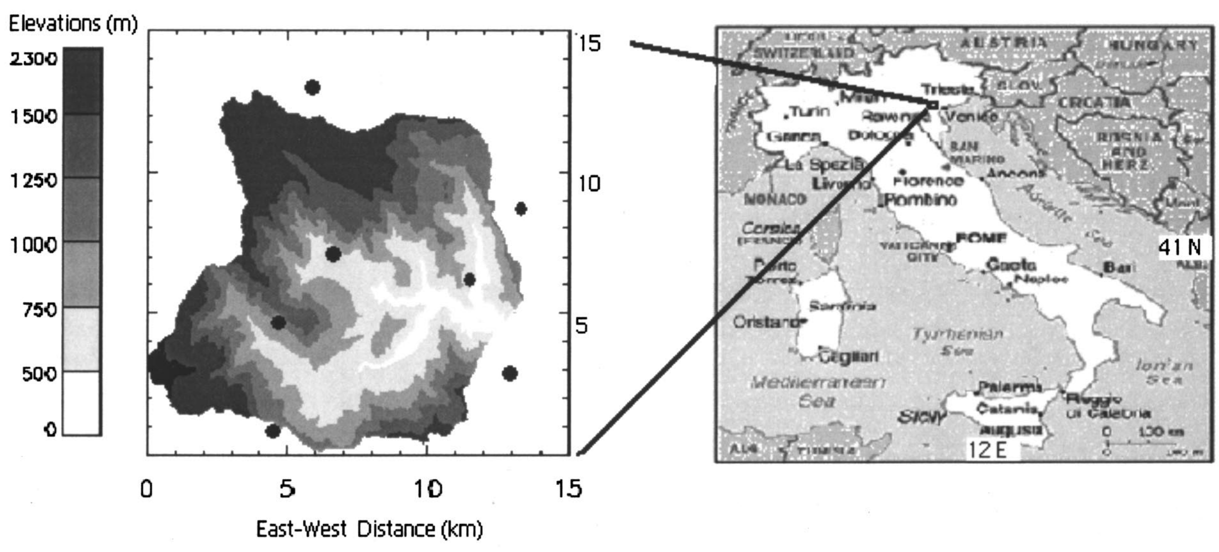

Fig. 1. Geographic location of Posina watershed (right panel) and watershed elevation map (left panel), overlaid by rain-gauge network locations (in solid circles)

conductivity; SZM (m), the exponential decay rate of hydraulic conductivity with depth; SRMAX (m), the maximum storage capacity of the root zone, interpreted here as the soil moisture at field capacity; TD ($h \cdot m^{-1}$), the time-delay parameter used to simulate the vertical unsaturated drainage flux; RV ($m \cdot h^{-1}$), the overland flow velocity parameter; and CHV ($m \cdot h^{-1}$)-the channel-flow velocity parameter. The rainfall input to the model, like the model parameters, was lumped (i.e., basin-averaged) in this study. The generated runoff was routed to the main channel using an overland flow delay function. The main channel routing effects are considered using an approach that is based on an average flood wave velocity for the channel network.

A series of 15 widespread storms (3–7 days long) that took place from 1987 to 1997 and caused flooding in the nearby region was studied. Statistics of the floods are shown in Table 1, where rainfall is basin-averaged and discharge measured at the outlet of the Posina watershed.

Probabilistic Discharge Prediction Scheme

GLUE Methodology

GLUE, first proposed by Beven and Binley (1992), is based on Monte Carlo simulations: a large number of hydrologic model runs are performed, each with parameter values randomly selected from probability distributions representing the uncertainty in model parameters. This uncertainty of model parameters is argued on the grounds of the physical theory that there should be sufficient interactions among the components of a hydrologic system that, unless the detailed characteristics of these components can be specified independently, many representations may be equally acceptable (Beven and Freer 2001). The acceptance of each run is assessed through a likelihood measure evaluated on the basis of predicted to observed differences. Parameter sets that achieve likelihood values below a certain likelihood threshold are rejected as "nonbehavioral" (i.e., nonrepresentative of the hydrologic process). The assignment of likelihood thresholds is subjective and is based on the hydrologist's understanding of what constitutes an acceptable simulation of the physical processes. Following the rejection of nonbehavioral runs, the likelihood values (hereafter called *likelihood weights*) of the retained parameter sets are rescaled so that their cumulative total is 1 (Freer et al. 1996). In this study, the GLUE method was applied to characterize the uncertainty in streamflow prediction by TOPMODEL. At each time step, the predicted discharge values from the retained parameter sets are weighted according to the corresponding likelihood weights and ranked to determine discharge quantiles at selected exceedance probability levels. We used the exponential efficiency measure for evaluating the likelihood weight of a parameter set, which is,

$$E_{\mathrm{exp}} = \exp\left[\frac{-\sigma_e^2}{\sigma_{\mathrm{obs}}^2}\right] \tag{1}$$

where σ_e^2 = variance of the errors in prediction for a given hydrologic parameter set and σ_{obs}^2 = variance of the observed discharge. The variance is computed over the whole simulation period. The choice for the preceding likelihood measure was governed by the following reasons: (1) it is always nonnegative and bounded between 0 and 1; (2) it increases monotonically as the closeness in behavior between the predicted and observed variable increases.

To implement the GLUE methodology, each parameter of TOPMODEL was specified a range of acceptable values. Table 2 lists the four TOPMODEL parameters used in GLUE and the ranges assigned to each. Constant (calibrated) values were used for the other three less-sensitive parameters. These four parameters (SZM, TD, T_0, and RV) were found most sensitive (for Posina) by applying the generalized sensitivity analysis (Spear and Hornberger 1980), which is indicated by the sensitivity of the behavioral parameter values' cumulative distribution function to the varying likelihood threshold. Fig. 2 shows the cumulative distribution functions (CDF) of these four sensitive parameters for three different likelihood thresholds (i.e., $E_{\mathrm{exp}} > 0.2$, 0.4, and 0.6), indicating differences in the shapes of the CDF functions. The CDFs of the three nonsensitive parameters were almost identical (not shown here). Following the GLUE methodology, multiple simulations were conducted by sampling randomly and uniformly from the specified ranges of the sensitive parameters (Table 2). Discharge simulations were carried out for each randomly generated parameter set, and the likelihood measure was calculated on the basis of Eq. (1). A large number of random simulations were conducted to allow the selection of 500 behavioral parameter sets characterized by simulation efficiency greater than an assigned threshold of 0.4. Ensemble runs based on the selected behavioral parameter sets characterize the range of uncertainty in the model prediction. Beven and Binley (1992) have argued that this procedure of GLUE would typically reflect all sources of uncertainty collectively in discharge simulation and would allow the uncertainty to be carried forward into the predictions. For further details about the GLUE method, the reader is

Table 1. Statistics of Flood Events Used in This Study

	Date	Rainfall statistics			Discharge (m^3/s)
Storm number	Month and year	Duration (h)	Rainfall volume (mm)	Maximum rain rate (mm/h)	Peak discharge (m^3/s)
1	08 1987	72	128.8	26.7	54.4
2	10 1987	96	137.3	11.2	75.7
3	07 1989	96	125.2	10.9	31.5
4	11 1990	96	120.8	8.7	64.4
5	12 1990	108	191.4	13.9	76.4
6	03 1991	72	77.4	8.3	32.6
7	10 1991	84	185.2	16.9	117.4
8	04 1992	120	113.0	8.7	56.9
9	10 1992	120	440.3	18.0	192.5
10	12 1992	144	118.6	6.8	41.6
11	09 1993	132	286.2	8.4	49.4
12	11 1994	72	146.0	11.7	106.9
13	10 1996	96	299.8	12.9	156.5
14	11 1996	120	179.9	10.2	70.8
15	12 1997	84	126.1	9.3	70.3

Table 2. Parameter Value Ranges Used in GLUE Sampling

Parameter	Minimum value	Maximum value	Sampling strategy
SZM (m)	0.001	0.25	Uniform
TD ($h \cdot m^{-1}$)	0.001	15.0	Uniform
T_0 [$\ln(m^2 \cdot h^{-1})$]	0.001	10.0	Uniform
RV ($m \cdot h^{-1}$)	200.000	2,000.0	Uniform

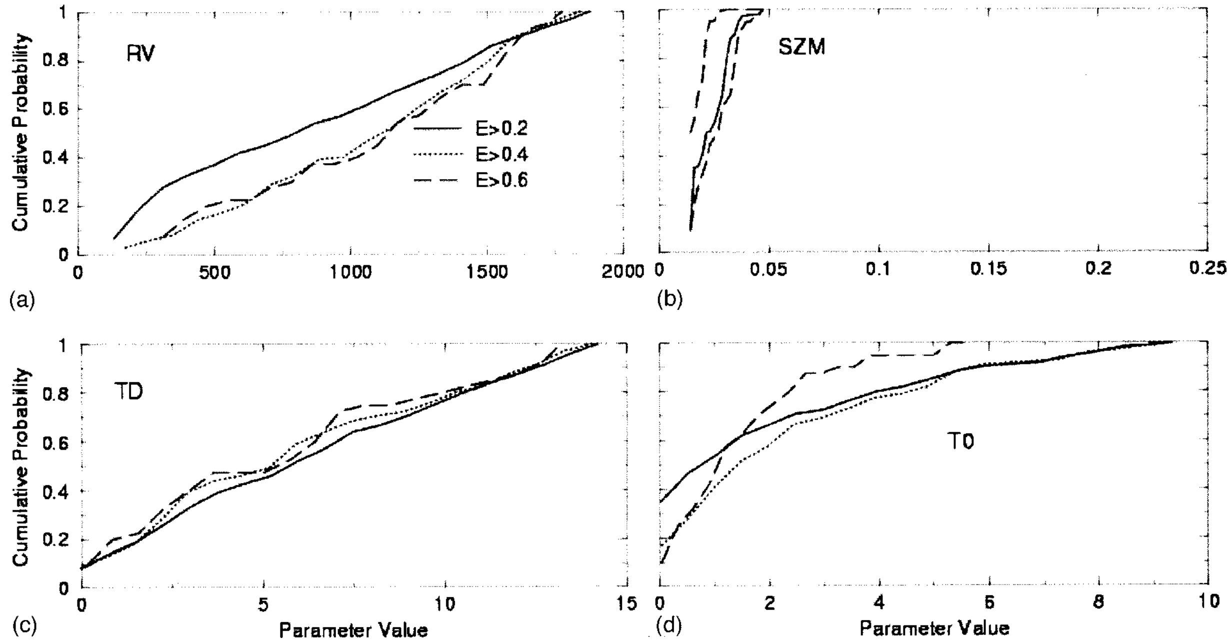

Fig. 2. Generalized sensitivity analyses (GSA; Spear and Hornberger 1980) for the four most sensitive TOPMODEL parameters: (a) overland flow velocity parameter RV (m h^{-1}); (b) exponential decay rate parameter SZM (m); (c) time delay parameter TD ($h\,\mathrm{m}^{-1}$); and (d) lateral transmissivity parameter T_0 ($\ln(\mathrm{m}^2\,h^{-1})$). Three different curves represent the three different cumulative probability distribution functions achieved with behavioral threshold E_{exp} of 0.2, 0.4, and 0.6, each having equal numbers of model simulation runs.

referred to the original work of Beven and Binley (1992) and subsequent follow-ups in Freer et al. (1996) and Beven and Freer (2001).

Prediction Scheme

As previously discussed, the behavioral ensemble parameter sets can be used to generate a range of discharge predictions that can be weighted by the corresponding likelihood weights of the parameter sets (hereafter called *likelihood-weighted discharges*). The likelihood-weighted discharges are then ranked to determine discharge prediction quantiles (ranging from 1st to 99th percentile). Comparison with observed runoff is used to evaluate the likelihood weight [based on Eq. (1)] of each quantile. The hydrograph resulting from the quantile associated with the highest likelihood value (hereafter called *optimum quantile*) is termed the *most probable hydrograph*. The most probable hydrograph does not necessarily correspond to the mean (expected) or median of all hydrograph simulations. Rather, it is an attempt to forecast a likelihood-weighted discharge that is likely to be most accurate since it reflects the most current prior knowledge. Thus, by virtue of the a priori evaluated model parameter sets and the likelihood of the prediction quantiles, we can predict for a subsequent storm event the most probable discharge, along with an estimate of its uncertainty. When new observations become available, each of the behavioral parameter sets and likelihood of the prediction quantiles are updated on the basis of Bayes' equation, as follows:

$$L_P(Y \mid \theta_i) = \frac{L(\theta_i \mid Y)L_0(\theta_i)}{C} \tag{2}$$

where θ_i refers either to the ith parameter set or the ith prediction quantile; $L_0(\theta_i)$=prior likelihood weight; $L(\theta_i|Y)$=likelihood weight calculated with the set of observed (current) data Y; $L_P(Y|\theta_i)$=posterior likelihood weight for the simulation by Y given θ_i; and C=scaling constant calculated such that the posterior weights add up to one.

For successive storm events, the calculated posterior likelihood distribution is used to project the uncertainty associated with the predictions to future events, thus becoming the prior distribution in Eq. (2). The posterior likelihood distribution is also used directly to evaluate the uncertainty limits for future events for which observed streamflow data is not available to validate model predictions. In successive prediction instances, the most probable hydrograph may not necessarily correspond to the same quantile. Nevertheless, the optimum quantile may be expected to eventually converge to a certain percentile depending on the worth of model, additional data, and stationarity of the hydrologic process. The previously described Bayesian updating may also have the effect of reducing the size of the sample of behavioral parameter sets, i.e., sets that were behavioral a priori may become nonbehavioral a posteriori because of the multiplicative updating process of Eq. (2). Thus, it may be necessary to resample the parameter distribution to reflect the regions of high likelihood values in parameter space, which can be expected to vary from storm to storm (Beven and Binley 1992). In resampling, it is ensured that the sample size remains constant (500 sets) to maintain consistency. In this study, we have employed the resampling procedure of nearest neighborhood search proposed by Beven and Binley (1992).

We summarize the probabilistic discharge prediction scheme with a flowchart (Fig. 3). The scheme starts with a historical storm-runoff data set to determine 500 behavioral parameter sets and the derived optimum quantile. In a subsequent storm event, the preceding information of the optimum quantile is used to evaluate its most probable hydrograph. The associated uncertainty is characterized by the ensemble discharge prediction generated by the behavioral parameter sets. When runoff observations for the new event become available, the likelihood weights of the

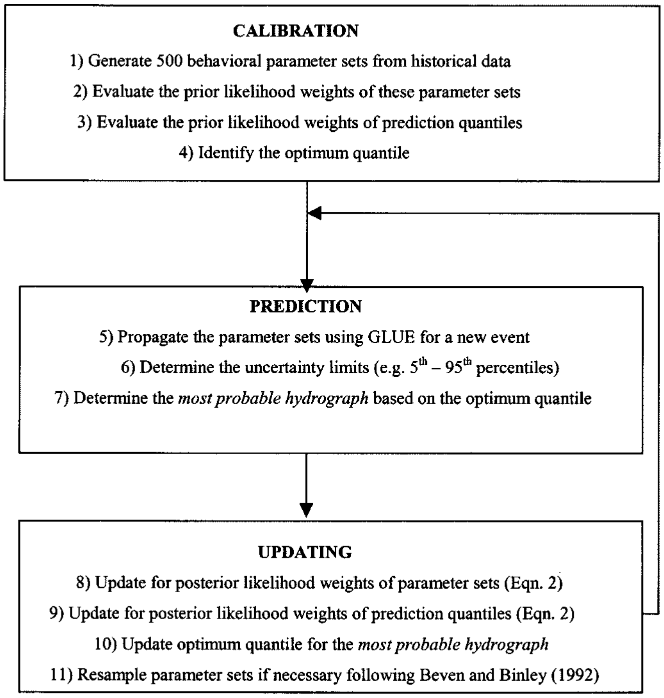

Fig. 3. Flowchart for the probabilistic discharge prediction scheme

parameter sets and prediction quantiles are updated using Bayes' equation. The procedure is repeated for successive storm events.

Assessment Framework

Assessment of the probabilistic discharge prediction scheme is performed on series of 15 flood-inducing storm events shown in Table 1. Comparisons were made against discharge predictions derived from a best parameter set (hereafter called *deterministic scheme*) derived on the basis of the optimization algorithm of Duan et al. (1992). The first storm of the series was used as an a priori data set for generating the ensemble of behavioral parameters and deriving the optimal parameter set for the deterministic scheme. For consistency in comparison, the deterministic scheme employed Eq. (1) for the cost function and the same parameter search space shown in Table 2. The comparison between the two schemes is performed as follows. On the basis of the chronological series of 15 floods, one can evaluate 14 prediction/update instances. To generate statistically sufficient cases for comparison, we shuffled the original chronological storm series to synthesize more hypothetical storm sequences. Since is it impossible to account for all possible combinations of storm sequences (in the order of 10^{12}), limited shuffling was performed to generate 14 more distinct storm sequences which are shown in Table 3. This leads to a total of 210 prediction/update cases for assessment. To minimize potential dependence on the initialization of the hydrologic model, no two sequences in the selected set had a common initial storm. The deterministic scheme was recalibrated, i.e., updated, after prediction for every new storm in a sequence considering all prior storm cases and using the values used in prediction as a first guess of the parameter set. For the probabilistic scheme, the likelihood weights of the ensemble of behavioral parameter sets and prediction quantiles were updated with each new observation, as described in the section describing the probabilistic discharge scheme and summarized in the flowchart of Fig. 3. For

Table 3. The 15 Storm Sequences

	Storm sequences														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	2	1	1	6	1	1	1	1	10	1	13	1	1	1	1
	3	3	8	13	2	2	2	2	11	6	14	2	2	2	2
	4	4	7	3	3	3	3	3	15	3	15	3	3	3	3
	5	5	9	11	12	4	4	9	4	4	4	4	4	4	4
	6	6	4	1	6	5	5	5	5	5	5	5	5	5	5
Storm number	7	7	7	10	7	7	6	6	6	2	6	6	6	11	6
	8	8	2	8	8	14	8	7	13	7	7	10	7	7	7
	9	9	9	9	9	9	9	4	8	8	8	8	8	8	8
	10	10	10	7	10	10	10	10	1	9	9	9	9	9	9
	11	11	11	5	11	11	11	13	2	11	10	7	10	10	10
	12	15	12	15	4	12	12	12	12	12	12	11	11	6	11
	13	14	13	2	13	13	13	11	7	13	1	13	12	12	12
	14	12	14	14	14	8	14	14	14	14	2	14	14	13	13
	15	13	15	12	15	15	15	15	3	15	3	15	15	15	14

Note: The first column is the original chronological series; the numbers refer to the storm numbers in Table 1.

event simulations, we recognize that the initial moisture conditions of the watershed can be a sensitive issue and that the model parameter associated with this initial condition may be uncertain. To ensure consistency in comparison, we constrained the initial moisture condition for each event to physically representative values common to both schemes.

We define the relative error (ε) in three runoff parameters: runoff volume (RV), peak runoff (PR), and time to peak (TP), as

$$\varepsilon_X = \frac{X_{\mathrm{sim}} - X_{\mathrm{obs}}}{X_{\mathrm{obs}}} \tag{3}$$

where X is defined as one of the runoff parameters (i.e., RV, PR, or TP). For the probabilistic scheme, the simulated runoff parameters were derived from the most probable hydrograph; and for the deterministic scheme, they were obtained from the optimum parameter set–based hydrograph. We evaluated the probabilistic scheme's prediction limits (derived on the basis of the different quantiles) in terms of its ability to envelope the observed peak runoff and time-to-peak parameters. For this purpose, we analyzed the sensitivity of exceedance probability in peak runoff and time to peak with respect to different quantile ranges. Consistent with the arguments of Beven and Binley (1992), our aim is to assess whether our probabilistic scheme predicts uncertainty bounds that are consistent with validation statistics, which represent both modeling and input uncertainty.

Results and Discussion

Fig. 4 shows comparisons between deterministic and probabilistic scheme predictions for the chronological storm sequence (first column in Table 3). The most probable hydrograph is shown by the thick dashed line, whereas the deterministic prediction is shown by the thin dashed line. It is observed that both schemes perform with the same accuracy. For certain storm cases (such as storms 2, 3, and 13), the most probable hydrograph appears superior in prediction to the deterministic scheme, whereas in other cases (such as storms 5, 9, and 14), peak runoff is moderately overestimated while the deterministic scheme underestimates it. Fig. 5 presents the prediction limits associated with the 90% confidence bounds (5th and 95th percentile) for prediction of peak

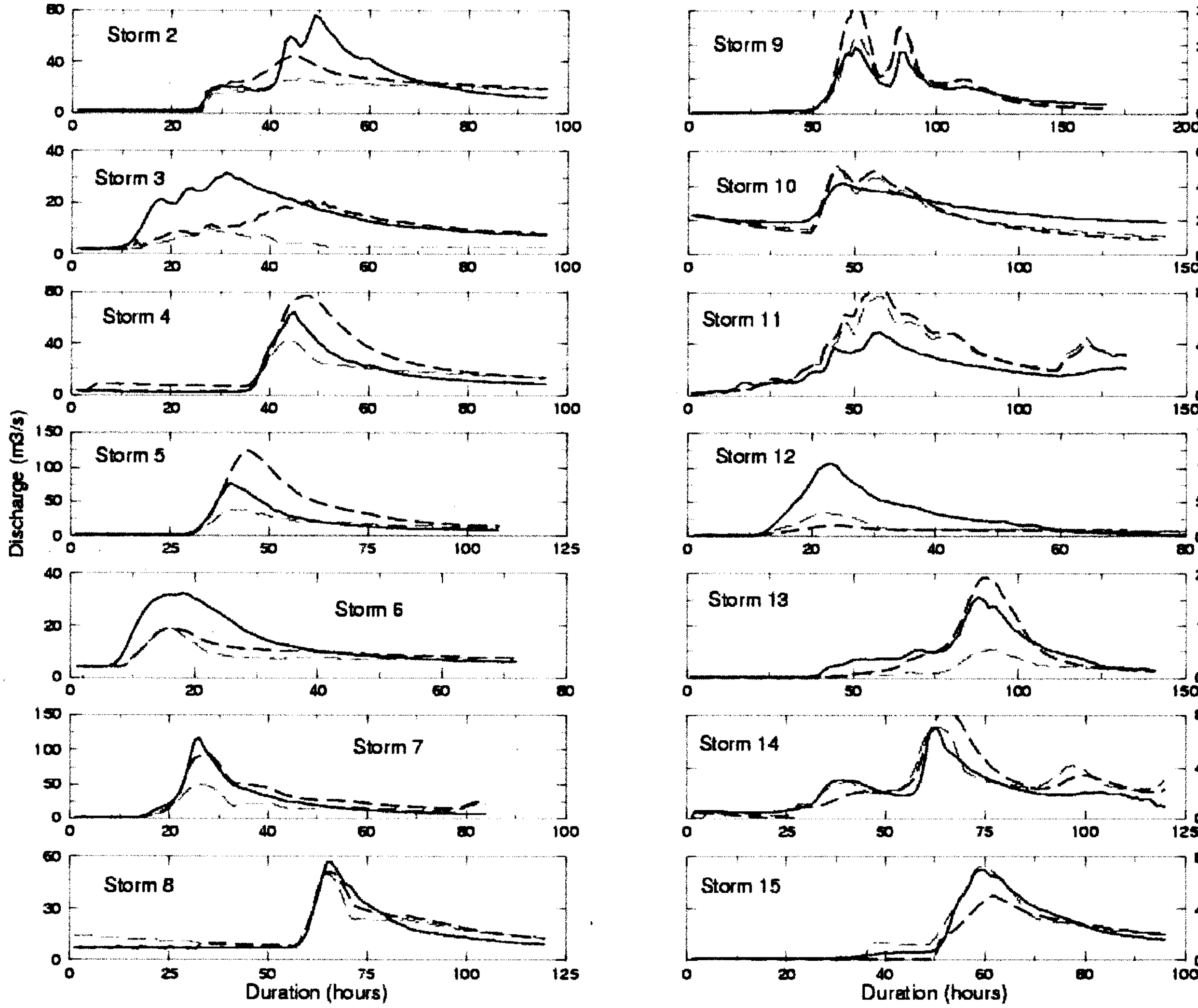

Fig. 4. Observed (solid line) versus predicted runoff hydrographs derived from deterministic (thin dashed line) and probabilistic (thick dashed) schemes for chronological sequence of 15 storm events

runoff and time to peak. The observed peak runoff is bounded within the 90% confidence bounds for 10 out of 14 prediction instances (storms 4, 5, 7, 8, 9, 10, 11, 13, 14, 15). This result indicates that the upper and lower uncertainty limits corresponding to the 90% confidence bounds of the probabilistic scheme have sufficient value to the decision makers for assessing the probable level of river stages during a flood. Similarly, for predicting the arrival time of the peak flood wave (i.e., time to peak), we observe that it is sufficiently bounded by the 90% prediction limits in 13 storms out of 14 (compare the vertical lines with shading with the thick solid line in Fig. 5). The probabilistic scheme can therefore adequately predict the lower and upper limits of the two most important runoff parameters during a flood: time to peak and peak runoff.

Table 4, summarizes the mean and standard deviation of the prediction accuracy of the two schemes on the basis of the 210 prediction cases. The mean prediction efficiency [Eq. (1)] for the probabilistic scheme (corresponding to the most probable hydrograph) is slightly higher (0.57) than the deterministic scheme (0.55). The standard deviation of the probabilistic scheme's efficiency of prediction is about 15% less than that of the deterministic scheme. For mean simulation error of runoff volume, both schemes have very low bias (around 6%), whereas peak runoff is underestimated by about 13% by the probabilistic scheme compared with the 6% overestimation by the deterministic scheme. The most noticeable difference is observed in the prediction of time to peak. The probabilistic scheme has 50% less variability than the deterministic scheme in time-to-peak error, even though the mean errors are comparable.

Table 5 shows the reliability of the probabilistic scheme's confidence bounds in terms of runoff and time to peak predictions. The *runoff exceedance probability* is defined as the number of hours that the observed discharge exceeds the 90% confidence bounds derived from the probabilistic scheme normalized by the total number of hours. *The time-to-peak exceedance probability* is evaluated by the number of prediction cases where the observed time to peak is beyond the 90% confidence bounds of the predicted time to peak normalized by the total number of cases (i.e., 210). A high exceedance probability would indicate the probabilistic scheme's inability to explain the uncertainty in runoff prediction. In Table 5, we observe that the probabilistic scheme has a low time to peak exceedance probability (12.4%). The runoff exceedance probability is moderately low (47%), indicating that the

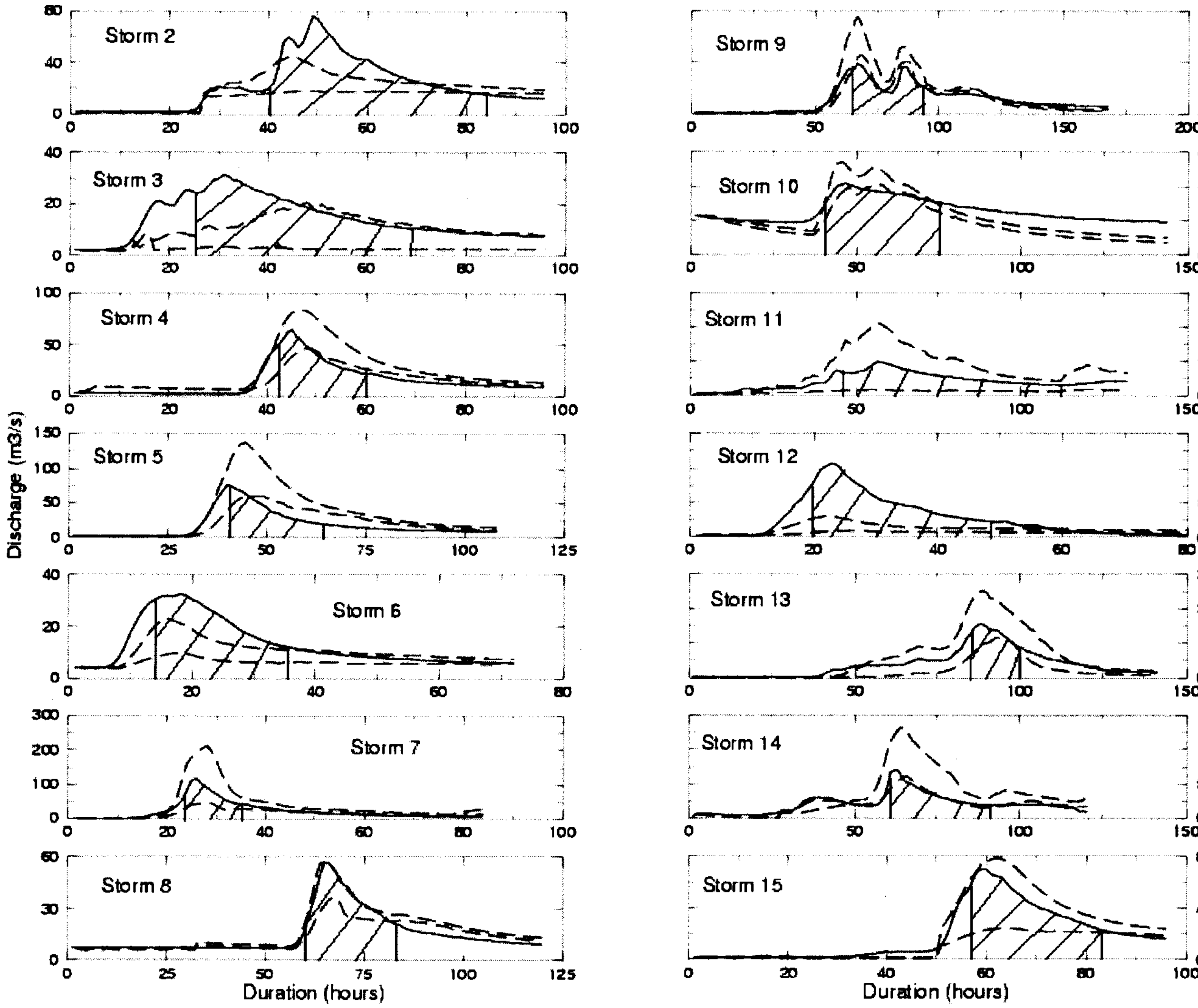

Fig. 5. Probabilistic prediction with uncertainty limits corresponding to 90% confidence bounds. The thick solid line represents observed discharge. Dashed lines represent the 5th and 95th percentiles of runoff prediction. Vertically drawn lines with shading represent the 5th and 95th percentiles of time to peak.

runoff (including the peak runoff) error bounds can capture 53% of the variability in true runoff.

In Fig. 6, we evaluate the sensitivity of runoff and time-to-peak exceedance probabilities and uncertainty range in time to peak (in hours) with respect to error bounds varied from 10% (45th to 55th percentile) to 90% (5th to 95th percentile) quantile ranges. We define this range as the "width of prediction limits." We observe a monotonic decrease of the Runoff exceedance probability for an increasing width of prediction limits (uppermost panel). The lowermost panel shows a similar pattern (rate of increase) in the sensitivity of time-to-peak uncertainty, expressed in relative terms as a fraction of the storm duration to the width. The lowermost panel shows the variation of time to peak exceedance probability with respect to the width of prediction limits. When the middle and uppermost panels are assessed jointly, it is concluded that at nearly 70% confidence bounds, we get the best prediction uncertainty scenario, which gives an optimal combination of low runoff exceedance probability with relatively low uncertainty range in time to peak. Analyses of all three panels combined can potentially assist decision makers in identifying the

Table 4. Comparison of Prediction Accuracy of Probabilistic Scheme with Deterministic Scheme

	Deterministic scheme		Probabilistic scheme	
Prediction/error	Mean	Standard deviation	Mean	Standard deviation
Prediction efficiency	0.552	0.318	0.517	0.280
Error in runoff volume	−0.063	0.502	0.063	0.591
Erro in peak runoff	0.064	0.515	−0.134	0.584
Error in time to peak	−0.043	0.428	0.094	0.205

Note: Prediction efficiency is based on Eq. (1). For the probabilistic scheme, it is computed from the most probable hydrograph.

Table 5. Exceedance Probabilities for Runoff and Time to Peak for the Probabilistic Scheme

Runoff exceedance probability (mean)	Time-to-peak exceedance probability (mean)
0.467	0.124

optimum uncertainty quantile range, which is not too wide, yet reliable.

Fig. 7 shows the impact of additional data on the convergence toward the optimum parameter set for the deterministic scheme (upper panel) and optimum prediction quantile for the probabilistic scheme (lower panel). The purpose of this analysis is to identify whether the schemes (probabilistic and deterministic) are capable of reaching steady-state prediction conditions. In the upper panel, the convergence toward the global parameter set is evaluated by using the norm of the optimum parameter set. The norm was computed by first normalizing the four optimum parameter values by their respective means and standard deviations evaluated from all the 210 prediction cases. This transformed each parameter to a consistent 0–1 scale. The norm was then computed as the square root of the sum of squares of the normalized optimum parameter values. For the lower panel, the optimum quantile is essentially the prediction quantile with the highest a posteriori likelihood weight after application of Bayesian updating. For both schemes, we note convergence. The norm gradually minimizes and then stabilizes to a near-zero value as additional storm data is incorporated in the optimization process (upper panel). The error bars representing one standard deviation gradually reduce with successive prediction cases. This outcome suggests that the space of the parameter set that contains the optimum set is becoming increasingly constrained as more data is taken into account. For the probabilistic scheme, the mean optimum quantile appears to stabilize at the 70th percentile after the 14th Bayesian updating (upper panel). The error bars, however, do not decrease as they do in the deterministic scheme. This outcome indicates that there may be more than one region of high likelihood values in the parameter space, which may be an indicator of the nonstationarity of the hydrologic system and the inherent uncertainty.

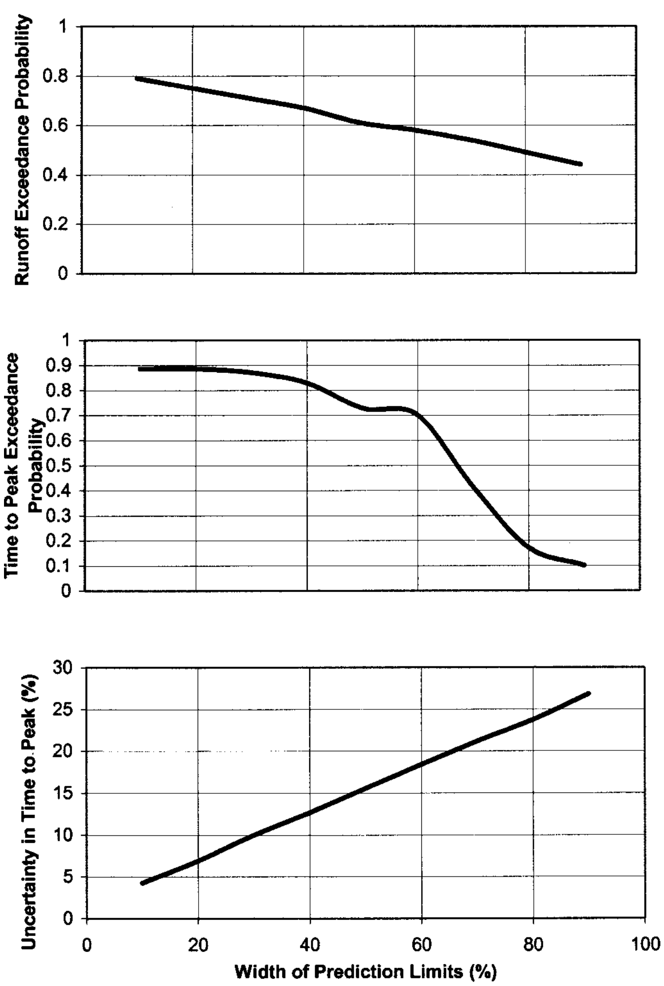

Fig. 6. Runoff and time-to-peak exceedance probability (upper and middle panels) versus width of prediction limit (upper-lower quantile range); lower panel: uncertainty in time-to-peak versus width of prediction limit

Conclusions

This study assessed a probabilistic discharge prediction scheme based on an uncertainty framework called GLUE. In prediction, the probabilistic scheme simulated the most probable hydrograph with the upper and lower uncertainty limits associated with a given confidence bound. Bayes' theorem was used to update the posterior likelihood weights of the parameter sets and prediction quantiles. Upon comparison with the conventional optimum parameter set deterministic prediction, it was observed that the probabilistic scheme was subject to nearly 50% less variability in time-to-peak prediction error. The probabilistic scheme has an added value to decision making and risk assessment because of the uncertainty predicted for the arrival time of peak runoff and magnitude of the flood wave. The procedure is simple in design, is model-independent, and can be easily implemented in a real-time operational scenario for computationally efficient rainfall-runoff models.

The findings are conditioned on the characteristics of a single watershed. Hence, as part of a natural extension of this study, multiple watersheds under different hydro-climatic regimes should be studied to understand the wider range of variability in performance of the prediction scheme. Furthermore, a major limitation of our scheme is its requirement for multiple model runs, which is prohibitive for use with physically based distributed rainfall-runoff models. However, such models are increasingly used to predict consequences of land use and climatic change in catchments. Their incorporation in a GLUE framework to support the probabilistic scheme formulated in this article would be inefficient because of the large number of Monte Carlo runs needed for both model prediction and parameter sampling. An improved parameter-sampling scheme that both accelerates the parameter search and reduces the total number of model runs is desirable. The uniform sampling procedure usually recommended for GLUE (Beven and Binley 1992; Freer et al. 1996; Beven and Freer 2001) is the aspect that requires special attention. Methods such as the guided Monte Carlo scheme of Shorter and Rabitz (1997), the tree-structured search of Spear et al. (1994), Latin hypercube sampling (McKay et al. 1979), and the Monte Carlo Markov chain method (Kuczera and Parent 1998; Bates and Campbell 2001) are worth considering as future extensions to this study.

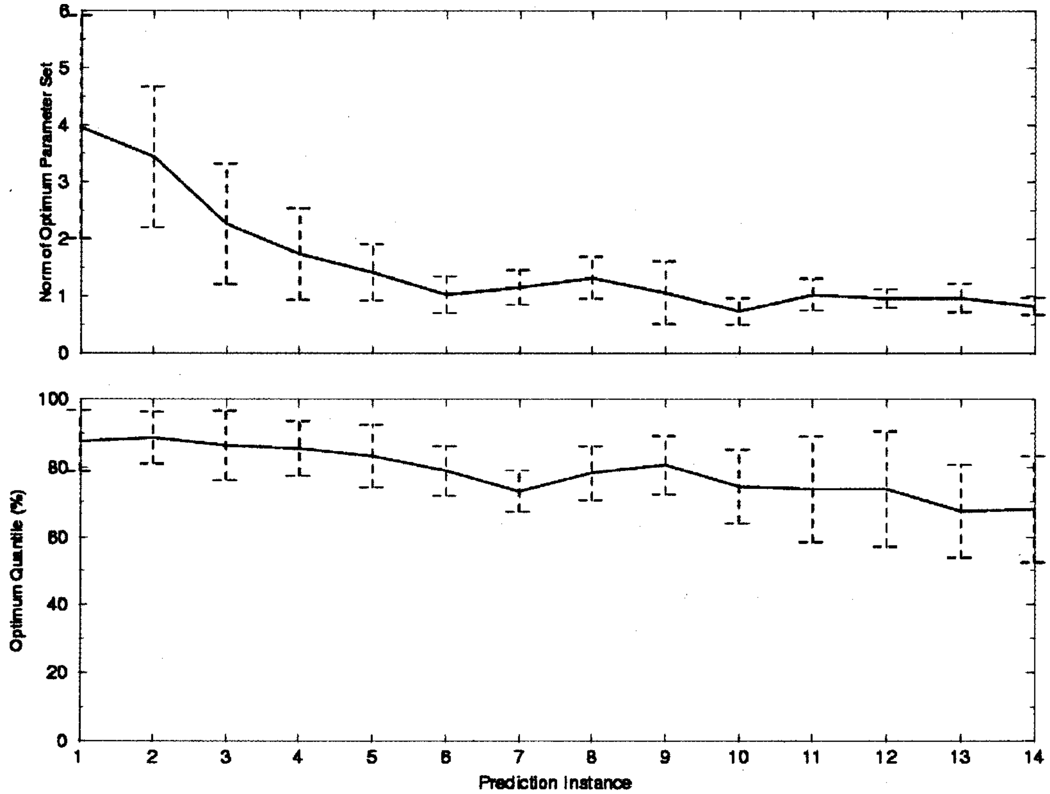

Fig. 7. Mean and one standard deviation (vertical bars) of the norm of optimum parameter set (upper panel) and optimum quantile (lower panel) versus prediction instance

Acknowledgments

The writers wish to thank Dr. Marco Borga of the University of Padua, Legarno, Italy, for providing the watershed and storm data for this study. This study was supported by a grant from NSF-Geosciences (EAR-0132942). The first writer was supported by a NASA Earth System Science Graduate Student Fellowship.

References

Bates, B. C., and Campbell, E. P. (2001). "A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling." *Water Resour. Res.*, 37(4), 937–947.

Beven, K. J., and Binley, A. (1992). "The future of distributed models: Model calibration and uncertainty prediction." *Hydrol. Proc.*, 6, 279–298.

Beven, K. J., and Kirkby, M. J. (1979). "A physically based variable contributing area model of basin hydrology." *Hydrol. Sci. Bull.*, 24(1), 43–69.

Beven, K. J., and Freer, J. (2001). "Equifinality, data assimilation, and uncertainty estimation in mechanistic modeling of complex environmental systems using the GLUE methodology." *J. Hydrol.*, 249, 11–29.

Beven, K. J., Lamb, R., Quinn, P., Romanowicz, R., and Freer, J. (1995). "TOPMODEL." *Computer models of watershed hydrology*, V. P. Singh, ed., Water Resource Publication, Fort Collins, Colo., 627–668.

Borga, M., Anagnostou, E. N., and Frank, E. (2000). "On the use of real-time radar rainfall estimates for flood prediction in mountainous basins." *J. Geophys. Res.*, 105(D2), 2269–2280.

Day, G. N. (1985). "Extended streamflow forecasting using NWSRFS." *J. Water Resour. Plan. Manage.*, 111(2), 157–170.

Duan, Q., Sorooshian, S., and Gupta, V. K. (1992). "Effective and efficient global optimization for conceptual rainfall-runoff models." *Water Resour. Res.*, 28(4), 1015–1031.

Fisher, J. I., and Beven, K. J. (1996). "Modelling of streamflow at Slapton Wood using TOPMODEL within an uncertainty estimation framework." *Field Studies*, 8, 577–584.

Franks, S. W., and Beven, K. J. (1997). "Bayesian estimation of uncertainty in land surface-atmospheric flux predictions." *J. Geophys. Res.*, 102(D20), 23991–23999.

Franks, S. W., Beven, K. J., and Gash, J. H. C. (1998). "On constraining the predictions of a distributed model: the incorporation of fuzzy estimates of saturated areas into the calibration process." *Water Resour. Res.*, 34, 787–797.

Freer, J., Beven, K. J., and Ambroise, B. (1996). "Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach." *Water Resour. Res.*, 32(7), 2161–2173.

Georgakakos, K. P. (1987). "Real-time flash flood prediction." *J. Geophys. Res.*, 92(D8), 94–158.

Hossain, F., Anagnostou, E. N., Borga, M., and Dinku, T. (2004a). "Hydrologic model sensitivity to parameter and radar rainfall estimation uncertainty." *Hydrolog. Process.*, 18(17), 3277–3291.

Hossain, F., Anagnostou, E. N., and Dinku, T. (2004b). "Sensitivity analyses of passive microwave and IR retrieval and sampling on flood prediction uncertainty for a medium sized watershed." *IEEE Trans. Geosci. Remote Sens.*, 42(1), 130–139.

Kelly, K. S., and Krzysztofowicz, R. (1994). "Probability distributions for flood warning systems." *Water Resour. Res.*, 30(4), 1145–1152.

Kitanidis, P. K., and Bras, R. L. (1980). "Real-time forecasting with a conceptual hydrologic model: 1. Analysis of uncertainty." *Water Resour. Res.*, 16(6), 1025–1033.

Krzysztofowicz, R. (1999). "Bayesian Theory of Probabilistic forecasting via deterministic hydrologic model." *Water Resour. Res.*, 35(9), 2739–2750.

Krzysztofowicz, R., and Reese, S. (1991). "Bayesian analyses of seasonal runoff forecasts." *Stochastic Hydrol. Hydr.*, 5(4), 295–322.

Kuczera, G., and Parent, E. (1998). “Monte Carlo assessment of parameter uncertainty in conceptual catchment models: The Metropolis algorithm.” *J. Hydrol.*, 211, 69–85.
McKay, M. D., Beckman, R. J., and Conover, W. J. (1979). “A comparison of three methods for selecting values of input variables in the analysis of output from a computer code.” *Technometrics*, 21(2), 239–245.
Murphy, A. H., and Carter, G. M. (1980). “On the comparative evaluation of objective and subjective precipitation probability forecasts in terms of economic value.” *8th Conf. on Weather Forecasting and Analysis*, *American Meteorological Society*, Boston, 478–187.
Quinn, P. F., Beven, K. J., Chevallier, P., and Planchon, O. (1991). “The prediction of hillslope flow paths for distributed hydrological modeling using digital terrain models.” *Hydrol. Proc.*, 5, 59–79.
Quinn, P. F., Beven, K. J., and Lamb, R. (1995). “The $\ln(a/\tan\beta)$ index: how to calculate it and how to use it in the TOPMODEL framework.” *Hydrol. Proc.*, 9, 161–182.
Romanowicz, R., Beven, K. J., and Tawn, J. (1994). “Evaluation of predictive uncertainty in nonlinear hydrological models using Bayesian approach.” *Statistics for the environment. II. Water related issues*, V. Barnett and K. F. Turkman, eds., Wiley, Chichester, U.K., 297–317.
Schaake, J. C., and Larson, L. (1998). “Ensemble streamflow prediction (ESP): Progress and research needs.” In preprints of *Special Symposium on Hydrology*, J19–J24, American Meteorological Society, Boston.
Schaake, J. C., Welles, E., and Graziano, T. (2001). “Comment of ‘Bayesian theory of probabilistic forecasting via deterministic model,’ by Roman Krzysztofowicz.” *Water Resour. Res.* 37(2), 439.
Seo, D. J., Perica, S., Welles, E., and Schaake, J. C. (2000). “Simulation of precipitation fields from probabilistic quantitative precipitation forecast.” *J. Hydrol.*, 239, 203–229.
Shorter, J. A., and Rabitz, H. A. (1997). “Risk analysis by the guided Monte Carlo technique.” *J. Stat. Comput. Simul.*, 57, 321–336.
Spear, R. C., Grieb, T. M., and Shang, N. (1994). “Parameter uncertainty and interaction in complex environmental models.” *Water Resour. Res.*, 30, 3159–3170.
Spear, R. C., and Hornberger, G. M. (1980). “Eutrophication in Peel Inlet, II, Identification of critical uncertainties via generalized sensitivity analysis.” *Water Resour. Res.*, 4, 43–49.