Hydrological model sensitivity to parameter and radar rainfall estimation uncertainty

Faisal Hossain,[1] Emmanouil N. Anagnostou,[1]* Tufa Dinku[1] and Marco Borga[2]

[1] *Civil and Environmental Engineering, University of Connecticut, Storrs, CT, USA*
[2] *Department of Land and Agroforest Environment, University of Padova, Agripolis, Legnaro, Italy*

Abstract:

Radar estimates of rainfall are being increasingly applied to flood forecasting applications. Errors are inherent both in the process of estimating rainfall from radar and in the modelling of the rainfall–runoff transformation. The study aims at building a framework for the assessment of uncertainty that is consistent with the limitations of the model and data available and that allows a direct quantitative comparison between model predictions obtained by using radar and raingauge rainfall inputs. The study uses radar data from a mountainous region in northern Italy where complex topography amplifies radar errors due to radar beam occlusion and variability of precipitation with height. These errors, together with other error sources, are adjusted by applying a radar rainfall estimation algorithm. Radar rainfall estimates, adjusted and not, are used as an input to TOPMODEL for flood simulation over the Posina catchment (116 km^2). Hydrological model parameter uncertainty is explicitly accounted for by use of the GLUE (Generalized Likelihood Uncertainty Estimation). Statistics are proposed to evaluate both the wideness of the uncertainty limits and the percentage of observations which fall within the uncertainty bounds. Results show the critical importance of proper adjustment of radar estimates and the use of radar estimates as close to ground as possible. Uncertainties affecting runoff predictions from adjusted radar data are close to those obtained by using a dense raingauge network, at least for the lowest radar observations available. Copyright © 2004 John Wiley & Sons, Ltd.

INTRODUCTION

Early warning with the use of radar rainfall observations and hydrologic models is crucial for minimizing flood and flash flood-related hazards. The potential benefit of using radar observations is particularly large in mountainous areas, where the rugged nature of the terrain and altitudinal effects impose significant limitations to the real-time operation of raingauge networks. Radar rainfall estimation is, though, subject to errors caused by various factors ranging from instrument issues (e.g., calibration, measurement noise) to the high complexity and variability in the relationship of the measurement to precipitation parameters (Austin, 1987; Joss and Lee, 1995; Andrieu *et al.*, 1997; Borga *et al.*, 2002). The different sources of errors which compound the radar rainfall uncertainty affect in various ways the accuracy of rainfall–runoff simulations (Borga, 2002). Meaningful hydrological applications of weather radar rainfall estimates require therefore the rigorous analysis of the propagation of radar rainfall uncertainties through rainfall–runoff modelling.

Many studies have focused on the application of radar rainfall estimates in flood forecasting applications (Schell *et al.*, 1992; James *et al.*, 1993; Georgakakos *et al.*, 1996; Bell and Moore, 1998; Vieux and Bedient, 1998; Winchell *et al.*, 1998; Sempere-Torres *et al.*, 1999; Borga *et al.*, 2000; Ogden *et al.*, 2000). Common to these studies is the use of rainfall–runoff models with a single optimal parameter set. These models were

---

* Correspondence to: Emmanouil N. Anagnostou, Dept. of Civil and Environmental Eng, The University of Connecticut, 261 Glenbrook Road, U 2037, Storrs CT 06269, USA. E-mail: manos@engr.uconn.edu

calibrated based on using a reference rainfall input (most often based on dense raingauge networks) and comparisons carried out with runoff simulations obtained by applying radar-based precipitation inputs. This allowed us to explore issues related to the impact of uncertainties due to (1) radar rainfall estimation errors and (2) the spatiotemporal sampling of precipitation fields on runoff simulation.

However, current research has shown that the concept of the optimum parameter set may be questioned for the case of hydrological modelling. Most such models are sufficiently complex that there may be many different sets of parameter values within a given model structure that may be compatible with the data available for calibration. Certainly, one of those parameter sets will be 'optimum' according to some measure of goodness of fit, but that optimum may not survive application to a different data set or different measure of goodness of fit. Parameter sets that give almost equally good fits may also be scattered throughout the parameter space. All these observations are at the heart of the equifinality concept, introduced by Beven and Binley (1992). In radar hydrology, the acceptance of the equifinality concept suggests that the results from sensitivity analyses carried out based on an 'optimum parameter set' concept may not be adequate to describe the influence of radar rainfall uncertainties on runoff simulation, since equally acceptable parameter sets may exhibit different sensitivity to rain inputs with varied uncertainty. It is the statistical characterization of runoff uncertainty associated with the input–parameter uncertainty interaction that is the key to understanding potential improvements in radar algorithms and investigating scenarios of rainfall–runoff models.

This study therefore aims at building a framework for the assessment of uncertainty that is consistent with the limitations of the model and data available and that allows a direct quantitative comparison between model predictions obtained by using radar and raingauge rainfall inputs. To achieve this purpose, an explicit attempt to account for the associated rainfall–runoff modelling uncertainties by use of the GLUE (Generalized Likelihood Uncertainty Estimation; Beven and Binley, 1992) is made. This is a Bayesian Monte Carlo simulation-based technique, developed as an extension of the Generalized Sensitivity Analysis (GSA) of Spear and Hornberger (1980). The method was outlined in concept by Beven and Binley (1992) and other applications using different types of likelihood measure have been demonstrated by a number of researchers (Romanowicz *et al*., 1994; Freer *et al*., 1996; Franks and Beven, 1997; Beven and Freer, 2001; among others). The uncertainty assessment is carried out here through application of radar-estimated precipitation to a lumped rainfall–runoff model for the Posina basin, a medium-sized watershed located in a mountainous region in northern Italy, where major error sources are represented by radar beam partial blocking and variability of reflectivity with altitude.

Results from this study are expected to provide insights on the use of radar rainfall estimates for hydrological forecasting and, more specifically, to provide an objective framework to improve the problem definition for questions such as: How confident is the prediction based on radar rainfall estimates? What are the principal sources of the uncertainty in runoff prediction? How can these uncertainties be reduced?

The paper is organized as follows. The next section presents the study region and data set. The third section describes the rainfall–runoff simulation study, and the fourth section illustrates the GLUE methodology and its application in this context. The final two sections complete the paper with discussion and conclusions, respectively.

STUDY AREA AND DATA

The region chosen for this study is located in northern Italy, close to Venice and Padua. Details about the study area, including its terrain characteristics and rain climatology, can be found in Borga *et al*. (2000). Radar data are from a C-band Doppler radar located on the Monte Grande hill, which is 60 km from the Posina catchment (116 km^2) and 476 m above sea level. The radar provides sweeps at 10 different elevation angles and covers a radius of 120 km (see Table I for technical details of the radar). In this study we used only data from the northwest quadrant (Figure 1, right panel) of the radar viewing area and up to a radius of 100 km. This radar sector is selected because it encompasses the most complex relief with peak elevations

Table I. Monte Grande weather radar characteristics

Parameter	Value
Location	45°21′46″N, 11°40′25″E
Wavelength	5·5 cm
Polarization	linear horizontal
3 dB beam width	0·9°
Peak power	250 kW
Quantization	256 level
Elevations (degrees)	0·5, 1·0, 1·5, 2·5, 3·5, 4·5, 6·0, 7·5, 10·0, 15·0
Range	120 km
Update time	15 min
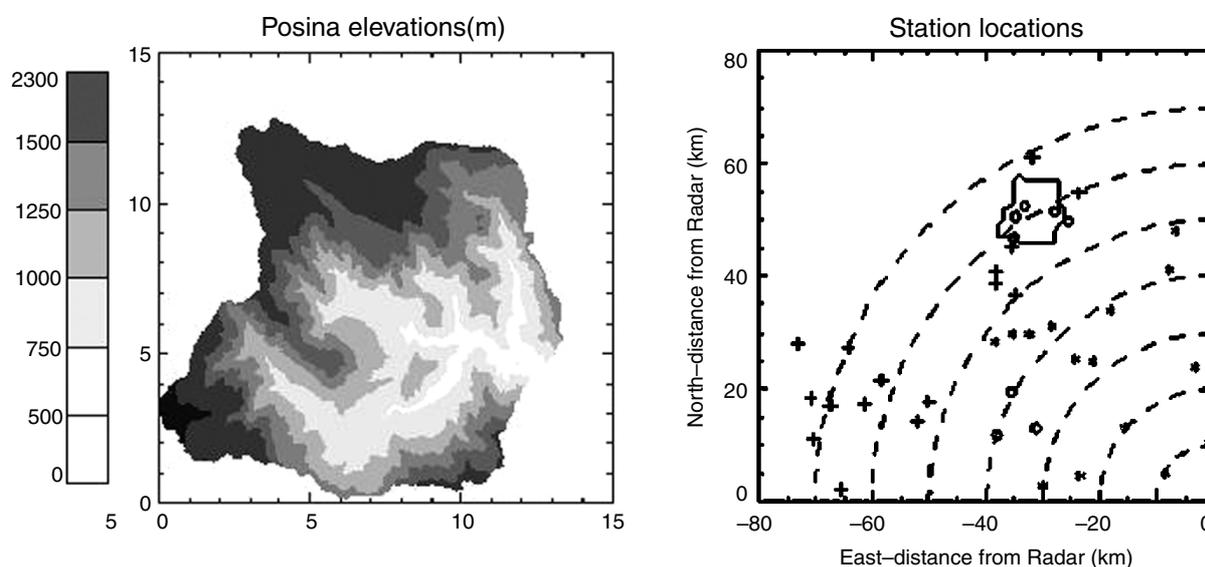
Figure 1. Left panel: relief map of the Posina basin. Right panel: locations, relative to radar position, of the raingauge stations used for mean-field radar bias estimation (*), error statistics computation during optimization of the radar rainfall algorithm (+), and independent stations close to the Posina basin used in this study (○). The dotted semicircles represent 10 km radar ranges

reaching 2500 m. In the Posina, altitude ranges from 2230 to 390 m above sea level at the outlet (Figure 1, left panel).

The investigation is performed for five storm events (see Table II for description of these events) that represent a typical meteorological situation associated with flooding in southern alpine regions. These storm events are characterized by cyclogenesis in the Lion Gulf or surrounding regions, which often establishes over the western Mediterranean in autumn months (Bacchi *et al.*, 1996). Cold fronts generated by this cyclonic circulation bring humid warm air from the south and/or the southwest, developing pre-frontal convective clouds and frontal stratiform clouds that impact the northern coastline.

Radar error sources and adjustment procedures

The major factors affecting radar rainfall estimation in the study area are non-uniform vertical profile of reflectivity (VPR), orographic enhancement of precipitation, ground clutter, wavelength attenuation, uncertainty in the reflectivity-to-rainrate ($Z-R$) conversion and radar calibration stability effects.

Table II. Statistical summary of the selected storms

Event	Beginning of event	Duration (h)	Cumulated areal rainfall (raingauge) (mm)	Max. gauge rain rate (mm h^{-1})	Peak discharge (m^3 s^{-1})
OCT92	2 Oct. 1992	120	440·3	18·0	192·50
NOV94	6 Nov. 1994	72	149·9	11·7	106·90
OCT96	14 Oct. 1996	96	299·8	12·9	156·50
NOV96	13 Nov. 1996	120	179·9	10·2	70·80
DEC97	18 Dec. 1997	84	126·1	9·3	65·90

The radar rainfall algorithm developed by Dinku *et al*. (2002) has been implemented here for adjustment of major radar error sources. This is a multi-component radar rainfall estimation algorithm that includes optimum parameter estimation and error correction schemes associated with radar operation over mountainous terrains. Algorithm pre-processing steps include correction for terrain beam blocking, adjustment for rain attenuation, and interpolation of reflectivity data from polar radar coordinates to a fixed three-dimensional Cartesian grid (hereafter named Constant Altitude Plan Position Indicator, CAPPI). The error correction schemes also include a simple but efficient approach to correct for the vertical variation of reflectivity at short–medium ranges and a stochastic filtering approach for mean field radar rainfall bias (MFB) adjustment (associated with systematic and drift errors in the radar calibration and biased $Z–R$ relationship). MFB adjustment is based on a uniform scaling factor considered representative of the whole region covered by the radar. It is defined as the ratio of the 'true' to the radar estimated mean area rainfall. For VPR adjustment, a correction procedure based on the following steps is devised. First, within a radius of 50 km of a certain Cartesian radar pixel location, neighbouring pixels with rain data and associated blocking level below an upper threshold are identified. Average rainfall is evaluated for the identified pixel values at a reference CAPPI level (e.g., 1 km) and for other CAPPI levels (2 km, 3 km, etc.). The VPR correction factor of each upper CAPPI level (i.e., 2 km, 3 km, etc.) is defined as the ratio of their averages to the lower CAPPI level averages. Finally, the rainfall values of the pixels at the 2 km and 3 km CAPPI levels are multiplied by the corresponding factors and are used in place of values of lower CAPPI levels if these are blocked by terrain features. One advantage of this correction scheme is that it is simple to implement. It also takes into account, to some extent, the spatial variability of VPR since for each pixel the correction factor is computed from neighbouring locations. Details about these correction procedures and assessment of their significance in radar rainfall estimation can be found in Dinku *et al*. (2002). Ground clutter removal was part of the raw radar data quality control system; hence, it was not part of the pre-processing steps described herein.

Thirty-nine raingauge stations within 80 km range from the radar (see Figure 1, right panel) are available. Fifteen of those located near the radar are used for the on-line mean-field bias estimation, while 16 raingauges located at ranges greater than 47 km had been used by Dinku *et al*. (2002) for calibration of their radar rainfall algorithm devised herein. Finally, the five independent raingauges (see Figure 1, right panel) located within and near the Posina catchment provide data for the reference basin-averaged rainfall estimates used in this study. The standard deviation of this reference rainfall estimation error, derived from Kriging, was found to be negligible (about 11%) with respect to the standard deviation of rainfall. This gave confidence that the raingauge measurements of these rain events are adequate to characterize the radar rainfall error structure (see Borga *et al*., 2000 for details of the Kriging technique used to derive basin-average rainfall). For consistency with the hydrological analysis constituting the second stage of the present study, hourly rainfall accumulations are considered for both adjustment and comparison with raingauge-derived estimates reported below.

We have considered two radar rainfall estimation scenarios to apply the radar rainfall estimation algorithm. In the first scenario the CAPPI ranged from 500 to 1500 m above the radar horizon level (476 m a.s.l.). This

implies dominant effects of radar beam occlusion and less impact of the variability of reflectivity with height. The second scenario was associated with a CAPPI ranged from 3000 to 4000 m above the radar horizon level. In this case VPR effects are dominant. Estimates obtained for Scenario 2 may be deemed representative of results that would have been obtained when lower scans were blocked or too contaminated by ground effects, or for a more distant catchment. Hereafter the results of the two scenarios will be named 'Scenario 1 or 2' and 'Adjusted or Unadjusted' (or Non-adjusted), representing whether or not we used the adjustment procedures incorporated into the radar rainfall estimation algorithm. Note that the unadjusted scenario is based on radar observations corrected for radar beam occlusion and that the main difference between unadjusted and adjusted scenarios is the adjustment for mean-field bias and VPR effects.

Figure 2 compares cumulative hyetographs obtained using raingauge and radar data for OCT92 flood event. The figure highlights relatively minor radar estimation biases for Scenario 1 (both unadjusted and adjusted) compared to the relatively large underestimation (more than 50%) that affects Scenario 2 unadjusted radar estimates. For Scenario 2, adjustment is characterized by 14% overestimation. For a statistical evaluation of the radar adjustment procedure, the following criteria are selected.
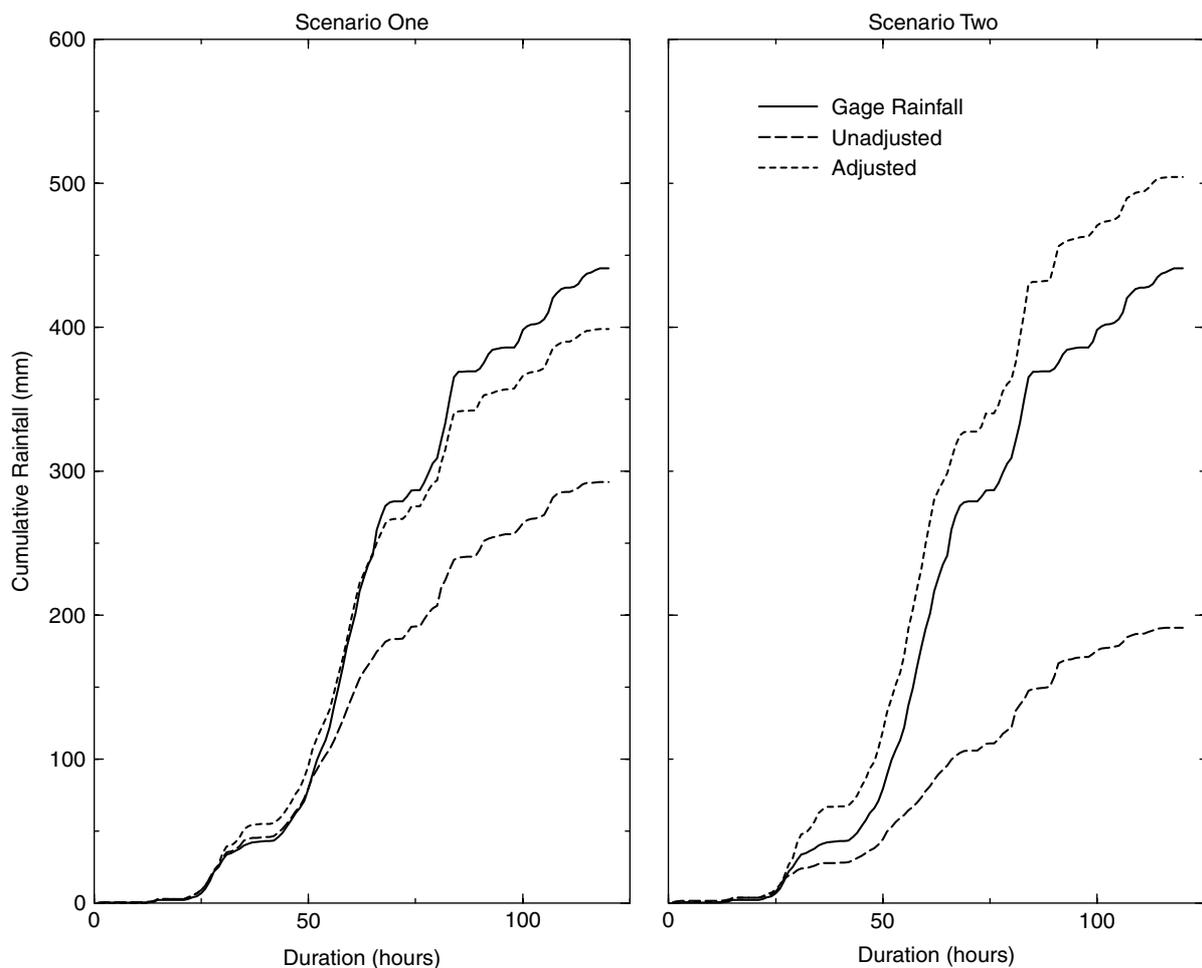
Figure 2. Cumulative raingauge and radar hyetographs for basin-averaged rainfall for OCT92 flood event

(1) The mean relative error (MRE):

$$\text{MRE} = \frac{\dfrac{1}{N_t} \sum_{i=1}^{N_t} \left(R_i^r - R_i^s \right)}{\dfrac{1}{N_t} \sum_{i=1}^{N_t} R_i^s} \tag{1}$$

where N_t is the number of hours in the storm event, R_i^s is the raingauge mean–areal rain rate at time i, and R_i^r is the corresponding radar mean–areal rain rate.

(2) The fractional standard error (FSE):

$$\text{FSE} = \frac{\left[\dfrac{1}{N_t} \sum_{i=1}^{N_t} \left(R_i^r - R_i^s \right)^2 \right]^{0.5}}{\dfrac{1}{N_t} \sum_{i=1}^{N_t} R_i^s} \tag{2}$$

The MRE and FSE values are representative of the systematic and random error components of the radar rainfall estimates, respectively. The values of MRE and FSE are reported in Table III for each radar rainfall estimation scenario and for each storm event, as well as for the ensemble (overall) storm events. Reasonable performances are obtained by applying the adjustment procedure, particularly for Scenario 1. The overall underestimation, which affects the unadjusted estimates (particularly for Scenario 2), is greatly reduced. Underestimation is reduced by almost 90% for Scenario 1 and overcompensated for Scenario 2 (with more than 20% overall overestimation). Adjustment allows reducing FSE by more than 30% for both scenarios. Among the factors adversely affecting the bias correction for Scenario 2, less-than-effective VPR correction is probably the most important. Indeed, the correction procedure is designed for optimal application at lower elevations than those included in Scenario 2, and it is tested here at its limits.

THE RAINFALL–RUNOFF MODEL

The rainfall–runoff model TOPMODEL (Beven and Kirkby, 1979) was chosen to simulate the rainfall–runoff processes of the Posina catchment. This model makes a number of simplifying assumptions about the runoff generation processes that are thought to be reasonably valid in this wet, humid catchment. TOPMODEL is a semi-distributed watershed model that can simulate the variable source area mechanism of storm runoff

Table III. MRE and FSE statistics for unadjusted and adjusted radar rainfall input scenarios

| | Scenario 1 | | | | Scenario 2 | | | |
| | Adjusted | | Unadjusted | | Adjusted | | Unadjusted | |
	FSE	MRE	FSE	MRE	FSE	MRE	FSE	MRE
OCT92	0·467	−0·096	0·775	−0·337	0·621	0·143	1·074	−0·565
NOV94	0·590	−0·097	1·040	−0·482	1·008	0·315	1·610	−0·769
OCT96	0·480	−0·079	0·746	−0·437	0·767	0·255	0·979	−0·641
NOV96	0·560	−0·121	0·910	−0·396	0·752	0·017	1·214	−0·618
DEC97	0·797	0·293	0·741	−0·364	1·202	0·372	1·460	−0·779
Overall	0·539	−0·058	0·854	−0·386	0·797	0·193	1·210	−0·629

generation and incorporates the effect of topography on flow paths. 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. Detailed background information on the model and applications can be found in Beven *et al.* (1995). The model has been applied in the study region by a previous work of Borga *et al.* (2000). As with many other TOPMODEL applications, the topographic index $\ln(a/\tan\beta)$ is used as an index of hydrological similarity, where a is the area draining through a point and $\tan\beta$ is the local surface slope. The use of this form of topographic index implies an effective transmissivity profile that declines exponentially with increasing storage deficits. In this study, the derivation of the topographic index from a 20 m grid size catchment digital terrain model utilized the multiple flow direction algorithm of Quinn *et al.* (1991, 1995). For the case of unsaturated zone drainage, a simple gravity-controlled approach is adopted in the TOPMODEL version used here, where a vertical drainage flux is calculated for each topographic index class using a time delay based on local storage deficit. The watershed was discretized into 32 sub-basins. The generated runoff is routed to the subcatchment outlet by using an overland flow delay function. Channel routing effects are considered using an approach based on an average channel flow velocity for the channel network.

The model parameters are: T0 ($\ln(m^2\ h^{-1})$), the mean catchment value of $\ln(T_o)$, where T_o is the lateral transmissivity when the soil is saturated to the surface; the exponential decay rate of transmissivity with depth, SZM (m); SRMAX (m), the maximum storage capacity of the root zone; XK0 ($m\ h^{-1}$), the surface hydraulic conductivity; TD ($h\ m^{-1}$), the time delay parameter used to simulate the vertical unsaturated drainage flux; RV ($m\ h^{-1}$), the overland flow velocity parameter; CHV ($m\ h^{-1}$), the channel flow velocity parameter. To initialize the saturated zone, the relationship between the saturated zone storage and the subsurface flow can be used if an initial discharge is known and can be assumed to be the result of drainage from the saturated zone only. This assumption is used here to derive the initial average subsurface storage deficit from the first discharge of each event which is still on a recession curve. The model was run at an hourly time step using basin-averaged rainfall input and considering homogeneous soils all over the catchment.

THE GLUE FRAMEWORK

The GLUE framework of Beven and Binley (1992) was used to assess the resulting uncertainty in the predictions. The procedure has been described by Romanowicz *et al.* (1994, p. 299) as 'in essence a Bayesian approach to uncertainty estimation for nonlinear hydrological models that recognises explicitly the equivalence, or near equivalence, of different parameter sets . . . in the representation of hydrological processes'. GLUE is based on Monte Carlo simulation: a large number of model runs are made, each with random parameter values selected from probability distributions for each parameter. The acceptability of each run is assessed by comparing predicted to observed discharges through some chosen likelihood measure. Runs that achieve a likelihood below a certain threshold may then be rejected as 'non-behavioural'. The likelihoods of these non-behavioural parameterizations are set to zero and are thereby removed from the subsequent analysis. Following the rejection of non-behavioural runs, the likelihood weights of the retained runs are rescaled so that their cumulative total is 1·0.

At each time step the predicted outputs from the retained runs are likelihood weighted and ranked to form a cumulative distribution of the output variable from which chosen quantiles can be selected to represent model uncertainty. In such a procedure the simulations contributing to a particular quantile interval may vary from time step to time step, reflecting the non-linearities and varying time delays in model responses. In this study discharge estimates of the 5th and 95th percentiles ($Q_{j,0.05}$ and $Q_{j,0.95}$, respectively, for the jth hour) were adopted as reference uncertainty bounds.

While GLUE is based on a Bayesian conditioning approach, the likelihood measure is achieved through a goodness of fit criterion as a substitute for a more traditional likelihood function. The likelihood value is associated with a particular set of parameter values within a given model structure. The likelihood associated

with a particular parameter value may therefore be expected to vary depending on the values of the other parameters, and there may be no clear optimum parameter set. The likelihood measure employed in this study is the classical index of efficiency, E (Nash and Sutcliffe, 1970):

$$E = \left[1 - \frac{\sigma^2_e}{\sigma^2_o} \right] \tag{3}$$

where σ^2_e is the variance of errors and σ^2_o the variance of observations (i.e, discharge). This likelihood measure is consistent with the requirements of the GLUE, as it increases monotonically as the similarity of behaviour increases.

There are a number of possible ways to define the behavioural threshold used to refine the likelihood distribution for a set of simulations. If it is possible to make some binary tests of model predictions against observed behaviour, then this can be used to distinguish behavioural and non-behavioural simulations. It is often impossible to make such a clear-cut decision. In the majority of cases it may be necessary to impose some essentially arbitrary behavioural threshold. This is not necessarily a drawback of the GLUE methodology. In setting a behavioural threshold, the modeller is able to make explicit his/her conditions for acceptance of a model. Furthermore, changing the behavioural threshold should result in a narrower or wider range of 'acceptable' behaviours, which would in turn change the estimated uncertainty about the model. This subjective element is likely to be implicit in any evaluation of a proposed model; the Bayesian methodology of GLUE helps to make this subjectivity explicit.

In previous studies, a behavioural rejection threshold has been chosen arbitrarily as a given value of the likelihood function (Freer *et al*., 1996, for example). Alternatively, a fixed proportion of the simulations can be retained according to their ranked likelihoods so that the best n solutions are considered behavioural (the best 50% for Beven and Binley, 1992). The last approach has been used below.

To implement the GLUE methodology, each parameter of TOPMODEL was specified a range of possible values. Table IV lists the four TOPMODEL parameters used for the GLUE and the ranges assigned to each. Constant (calibrated) values were used for three less sensitive parameters. While the possibility of correlation between parameters exists, we have no *a priori* reason to assume any correlation structure among parameters, so a uniform sampling strategy was adopted.

Following the methodology outlined above, uncertainty associated with model parameterization was assessed. A sequence of seven flood events, spanning the period of 1987–1995 and excluding the validation sequence for which radar data are available, was selected as the conditioning data set. Rainfall input for this conditioning data set was provided by raingauge data. Model predictions were carried out, and the model likelihood measure was calculated using the efficiency index. From the specified parameter ranges (Table IV), a large number of simulations were run that allowed us to select 20 000 parameter sets characterized by a simulation efficiency greater than 0·3. The maximum likelihood value (E) achieved on the conditioning data set was 0·81. Figure 3 shows the projections of the sample of points on the goodness of fit response surface onto each individual GLUE parameter dimension (dotty plots). It is shown that for each parameter there are good simulations across a wide range of parameters, as well as poor simulations.

Table IV. Parameter value ranges used for GLUE sampling

		Minimum value	Maximum value	Sampling strategy
SZM	(m)	0·001	0·25	Uniform
TD	(h m^{-1})	0·001	15·0	Uniform
T0	(ln(m^2 h^{-1}))	0·001	10·00	Uniform
RV	(m h^{-1})	50·0	2000·0	Uniform
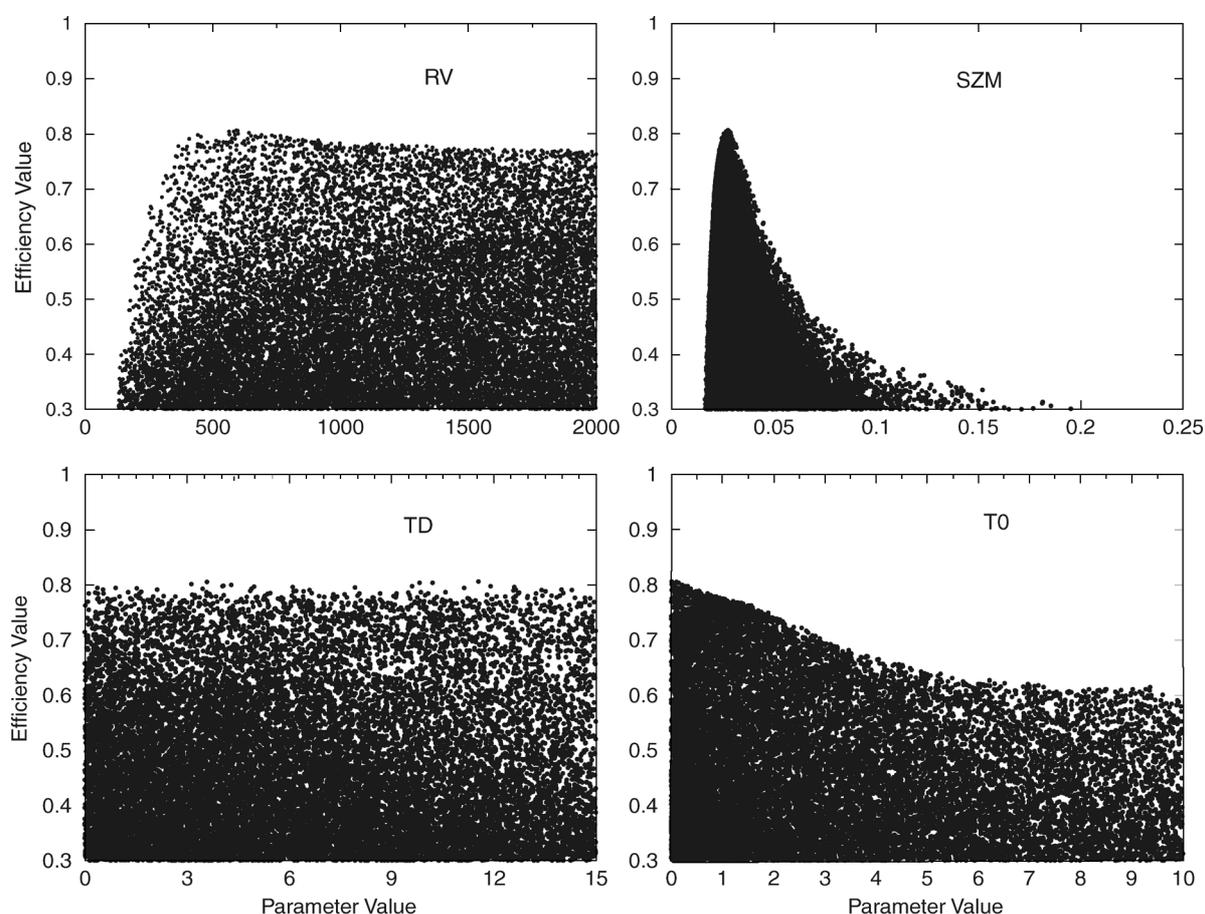
Figure 3. Dotty plots of the Nash–Sutcliffe model efficiency measure for the four GLUE parameters. Each dot represents one run of the model with parameter values chosen randomly by uniform sampling across the range of each parameter in Table IV

A number of fixed proportions of the simulations were retained according to their ranked likelihoods so that the best n solutions were considered behavioural. These fixed proportions ranged from the best 10% to the best 100% (i.e., all the 20 000 parameter sets with efficiency greater than 0·3), at a 10% increment. Uncertainty bounds conditioned on the calibration flood sequence using the best 10% to the best 100% of the realizations were then propagated to the (validation) flood sequence (see Table II) using rainfall estimates from the raingauge network and from the two radar rainfall estimation scenarios, both adjusted and unadjusted.

COMPARING PREDICTIVE UNCERTAINTY FOR DIFFERENT RAINFALL INPUTS

The GLUE procedure provides a range of parameter likelihood-weighted predictions that may be compared with discharge measurements. It may be found, as will be shown below, that the observations with which model predictions are to be compared may still fall outside the calculated uncertainty limits. If it is accepted that a sufficiently wide range of parameter values has been examined, and the deviation of the observations is greater than would be expected from measurement error, then this would suggest that the model structure, or the imposed boundary conditions (including the rainfall input), should be considered inadequate to describe the system under study. Two contrasting issues should be considered at this stage: if either the uncertainty limits

are drawn too narrowly or the whole simulation envelope is biased, then a comparison with observations will suggest that the model structure is invalid; if, on the other hand, they are drawn too widely, then it might be concluded that the model structure has little predictive capability. These observations outline the basic structure for a framework aimed at a direct quantitative comparison of predictive uncertainty between model responses obtained from competing rainfall estimates. In that framework, both the wideness of the uncertainty limits and the percentage of observations included in the limits should be evaluated. Furthermore, since any behavioural threshold is arbitrary in some sense, the comparisons should be carried out for a range of behavioural thresholds.

Given that the same conditions are applied throughout GLUE concerning: (1) the sample of *a priori* parameter set used; (2) the choice of likelihood measure; and (3) the discharge observations used in the calculation of the likelihood measure and for the comparison with the uncertainty limits, the analysis of results obtained by conditioning the model with a reference rainfall (derived from a dense raingauge network, for instance) and propagating the uncertainty bounds by using competing rainfall inputs offers a convenient way to quantify the impact of errors in rainfall input on runoff modelling uncertainty. This is carried out by comparing the range of likelihood-weighted predictions, obtained by using alternative competing algorithmic structures for radar rainfall estimation, with the observations. In accordance with a previous application of GLUE by Borga (2001), two statistics are defined as follows for a given behavioural threshold:

$$
\text{ER} = \frac{N_{\text{exceedances,radar}}}{N_{\text{exceedances,gauge}}}
$$

$$
\text{UR} = \frac{\sum_{j=1}^{N_T} \left(\text{qsim}^{\text{radar}}_{j,0\cdot95} - \text{qsim}^{\text{radar}}_{j,0\cdot05} \right)}{\sum_{j=1}^{N_T} \left(\text{qsim}^{\text{gauge}}_{j,0\cdot95} - \text{qsim}^{\text{gauge}}_{j,0\cdot05} \right)}
\tag{4}
$$

where $N_{\text{exceedance}}$ is the number of times the observed discharge falls outside the calculated uncertainty limits, ER is the exceedance ratio signifying the ratio of exceedances for radar input to gauge input, and UR is the uncertainty ratio signifying the ratio of the aggregate wideness in runoff simulation uncertainty for radar input to gauge input. The N_T is the total number of time steps for simulation of storm events, with j pertaining to a given time step, and qsim is the simulated discharge with its superscript signifying the type of rainfall input.

It is noted again that the two statistical measures presented herein need to be assessed jointly as individual interpretations can be erroneous due to their inherent competing characteristics. Also, if the gauge-based rainfall is assumed to be the reference value (i.e., closest representation of the true rainfall), then a closer look at the denominator of the ER and UR in equation (4) will reveal that it actually signifies the impact of the model's parameter uncertainty on runoff prediction uncertainty. The numerator can then be considered to be representative of the combined impact of model parameter uncertainty and rain input error on runoff simulation uncertainty. Hence the two ratios are essentially a normalization of combined input–parameter uncertainty to the corresponding model's parameter uncertainty, with ER aimed at comparing the relative structural validity and UR aimed at comparing the relative predictive capability. A ratio value nearing 1 would indicate the runoff predictive uncertainty for the given radar input mimicking that from a dense raingauge network.

Figures 4 and 5 show the uncertainty bounds (with *n* equal to 100%, i.e., with use of all the 20 000 parameter sets with efficiency greater than 0·3) propagated for the OCT 1992 flood event by using different rainfall inputs. Examination of the simulations obtained by using raingauge data (Figure 4) demonstrates that large uncertainties can be associated with model predictions even when using the reference rainfall. However, the uncertainty bounds enclose the observed time series relatively well, and only a few flow observations can be found outside of the uncertainty envelopes, indicating deficiencies in the data and/or
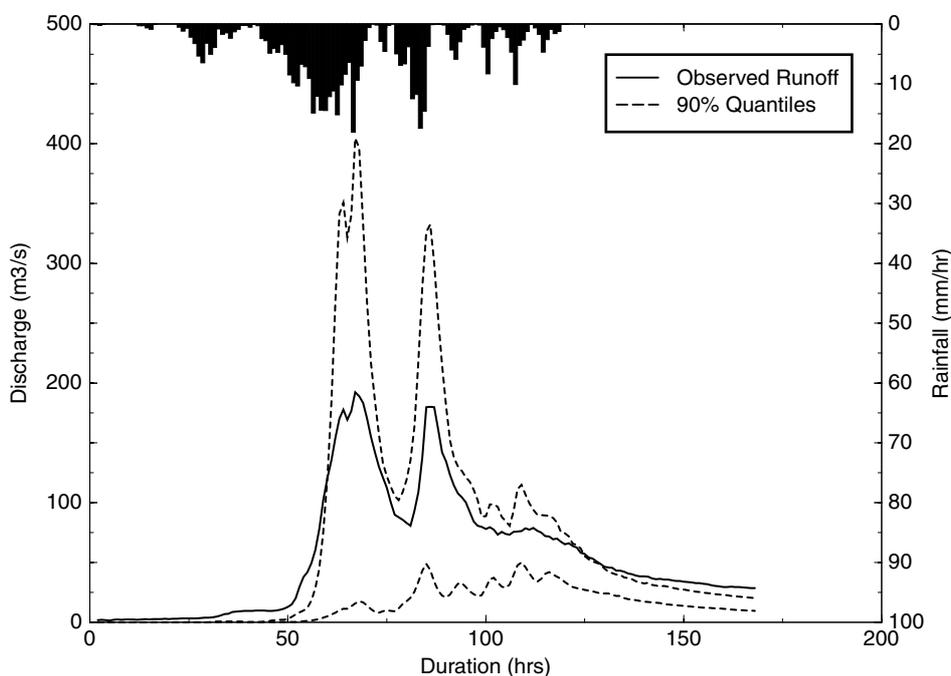
Figure 4. Propagated uncertainty bounds for OCT92 flood event with rainfall input from raingauge data

model structure. Comparison between the various uncertainty bounds for the radar scenarios shows that narrow uncertainty bounds are associated with negatively biased (underestimated) unadjusted radar rainfall input. Radar error adjustment results in a widening of the uncertainty bounds. This provides evidence that, contrary to conventional wisdom, appropriate adjustment of the radar estimates may (rightly) increase predictive runoff uncertainty, at least when adjustment is associated with the correction of a negative bias in raw radar estimates. The reduction of the negative bias and hence the increase in magnitude of the radar rainfall estimation through adjustment procedures magnifies further in the non-linear propagation as higher runoff values, resulting in wider runoff simulation uncertainty bounds.

For the adjusted case of Scenario 1, we observe that simulation uncertainty is very similar to that for gauges. This provides strong evidence that radar scans as close to the ground as possible and adjusted for errors can characterize the rainfall–runoff transformation as accurately as a relatively dense gauge network, even in a terrain characterized by rough orography. Note that, for this event, rainfall estimates from unadjusted Scenario 2 are affected by more than 50% negative bias (underestimation), while adjustment results in 14% overestimation (see Figure 2). Examination of the simulations obtained by using unadjusted Scenario 2 radar estimates shows that the uncertainty bounds are considerably narrower than those obtained for the case of Scenario 1. This is due to the global negative bias characterizing this scenario. For the same reason, all discharge observations fall outside of the uncertainty envelope. Radar error adjustment allows us to reduce the global bias and results in a widening of the runoff simulation uncertainty. This causes the bounds to bracket the storm hydrograph realistically.

Figure 6 shows the exceedance ratios (ER) and uncertainty ratios (UR) evaluated for different behavioural thresholds for the entire study flood sequence. Several features are worth noting in this figure. The patterns of the uncertainty structure associated with the use of radar rainfall estimates are relatively similar for the two different scenarios. The differences are summarized as follows: (1) larger values are found for ER in Scenario 2 (adjusted and unadjusted) with respect to Scenario 1 and (2) the difference in UR between adjusted and unadjusted scenarios is higher for Scenario 2. These effects are clearly due to the different rainfall input
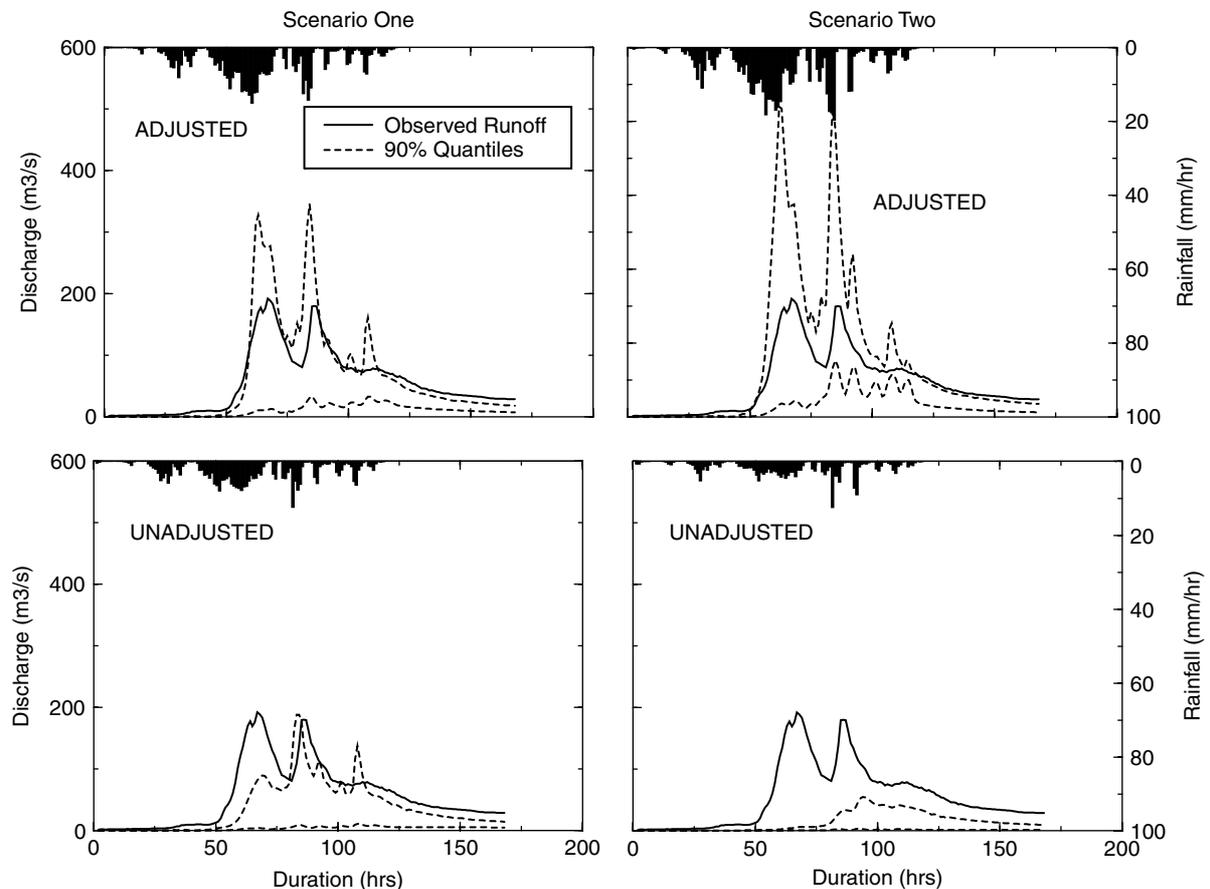
Figure 5. Propagated uncertainty bounds for OCT92 flood event with rainfall input from radar rainfall estimates

biases characterizing the two scenarios. Radar rainfall errors (particularly those associated with VPR effects) and their adjustment exhibit a more pronounced effect for Scenario 2, where the relatively small values of UR for unadjusted estimates and the relatively large values for adjusted estimates mirror the characteristics of the radar rainfall bias (negative and positive, respectively) and of its propagation through the rainfall–runoff transformation. It is noted also that beyond a 40% parameter uncertainty (best 40%), adjusted Scenario 1 radar rainfall yields very similar runoff simulation uncertainty bounds as gauges, even though there remains a 30–50% increase in ER.

Examination of Figure 6 also shows that there is a tendency for UR (ER) to decrease (increase) with increasing number of behavioural realizations retained in the analysis. The effect of varying the behavioural threshold is to modify the uncertainty bounds; by setting a stricter threshold, the uncertainty bounds are narrower. The slight decrease of UR with increasing number of behavioural realizations (i.e., with the parameter predictive uncertainty) indicates that the uncertainty bounds associated with using radar estimates (either adjusted and not) widen less than those associated with raingauge input. The increase of ER statistics with increasing number of behavioural realizations indicates that relatively less flow observations are falling within the predictive uncertainty bounds by using radar input than by using raingauge input. This suggests that the uncertainty bounds associated with the radar input become relatively more biased than those associated with gauge input, with increasing number of realizations retained as behavioural. In other words, the interaction of biased radar input with less representative model parameters results in more biased runoff predictions. This
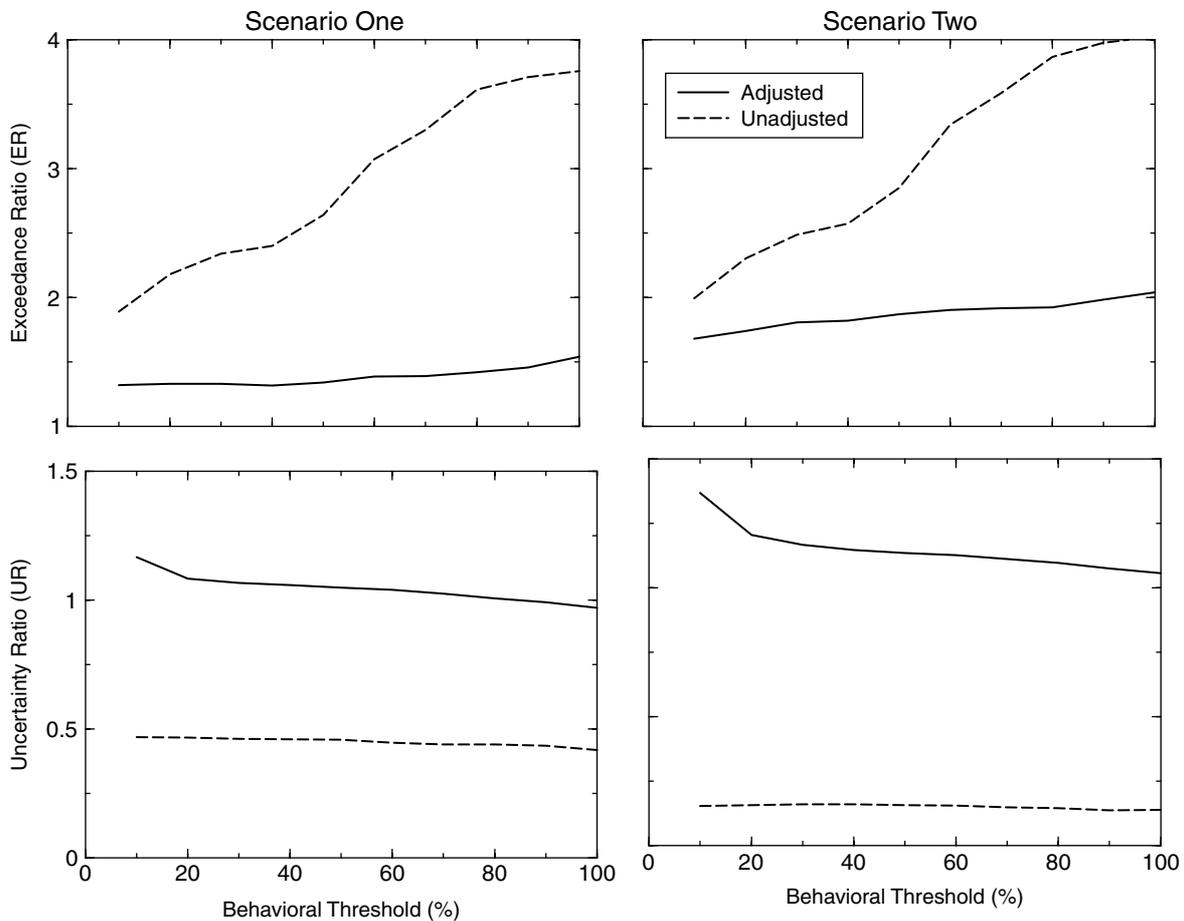
Figure 6. Runoff prediction uncertainty assessment in terms of uncertainty ratio and exceedance ratio for various behavioural thresholds

finding is confirmed by the comparison of ER statistics among Scenarios 1 and 2, which shows a quicker increase of ER for Scenario 2, characterized by more biased radar estimates.

The global picture emerging from this analysis demonstrates that runoff prediction uncertainty given radar rainfall input cannot be simply partitioned into two additive terms: the term related to model parameter uncertainty and the term related to the propagation of radar rainfall uncertainty. Rather, radar rainfall uncertainty—particularly when the rainfall bias term is important—acts in a highly non-linear sense on the model parameter uncertainty, by either magnifying or reducing it according to the nature of the rainfall estimation bias. The above considerations should not imply that unbiased radar rainfall estimates lead to unitary UR: random errors still play a role and indeed examination of the UR statistic for Scenario 1 shows (between the best 0% to 40%) that its value is greater than 1 even when using radar estimates only slightly biased.

CONCLUSIONS

The GLUE procedure has been used in this study as a means of hydrological model comparison using different rainfall inputs, provided by raingauge networks and by radar estimates according to various processing

scenarios. The proposed analysis framework allows evaluating both the wideness of the predictive uncertainty limits and the percentage of observations included in the limits, with varying the behavioural threshold. This helps to assess the impact of radar rainfall errors on the output of a hydrological model previously conditioned using rainfall data from a reasonably dense raingauge network. The evaluation is reported in terms of both structural validity and predictive capability of the resulting model output.

In the Introduction we posed three questions: (1) How confident are runoff predictions based on radar rainfall? (2) What are the causes of errors in runoff? (3) How can the errors be reduced? The investigation reported herein showed how it is possible to use the GLUE framework to provide an objective answer to these questions. We observed that the runoff model defined by using unadjusted radar estimates is structurally invalid due to poorly defined input data. The wideness of uncertainty bounds for runoff predictions obtained by using the adjusted radar rainfall (especially for Scenario 1, which is associated with moderate vertical reflectivity profile effect) is similar to those obtained based on basin averaged gauge rainfall. That is, use of radar scans closest to the ground offers runoff prediction as confident as that obtained from a dense gauge network. The principal sources of errors in radar rainfall estimation, such as mean field bias and VPR effects, magnify in the non-linear rainfall–runoff transformation as runoff prediction uncertainty. For mitigation of these error sources, combined adjustment procedures for MFB and VPR effect, such as those developed by Dinku *et al*. (2002), are effective means of improving accuracy of runoff predictions.

There are also other features worth summarizing here. Runoff simulations appear sensitive to the impact of errors related to variability of reflectivity with height, which dominate the radar error structure (particularly for high radar scans). However, structural errors due to the combination of radar errors and model parameter uncertainty result in 30–50% more discharge observations falling outside the uncertainty bounds with respect to the use of raingauge data. For Scenario 2 (adjusted) these percentages increase to 70–100%, with slightly larger uncertainty bounds.

Use of the type of analysis proposed here provides a clear view of the relative effects of input and parameter uncertainty upon model output and indeed is a valuable tool in analysing and ranking the sources of predictive uncertainty. It is hoped that, being explicit about the levels of uncertainty, limitations within the radar processing algorithm and the hydrological model can be improved upon, or additional data can be acquired in order to reduce the predictive uncertainty. Natural extensions of this work include: (i) the consideration of radar error structures and adjustment algorithms different from those used in this analysis; (ii) the implementation of the analysis framework for continuous hydrological simulation instead of event-based flood modelling described here; (iii) examining the effects of different likelihood functions and parameter sampling ranges within the GLUE methodology. Another interesting aspect worth studying is the impact of the quality of the gauge estimates on the characterization of the radar rainfall error propagation in runoff. Different scenarios of varying quality of gauge estimates could be studied by considering hypothetical combinations of gauge networks (e.g., excluding the ones closest to the basin in the analysis). All these extensions will further improve our understanding of runoff prediction uncertainty associated with the two main error sources: in rainfall input and in hydrologic modelling.

REFERENCES

Andrieu H, Creutin JD, Delrieu G, Faure D. 1997. Use of weather radar for the hydrology of a mountainous area. Part I: Radar measurement interpretation. *Journal of Hydrology* **34**: 225–239.

Austin PM. 1987. Relation between measured radar reflectivity and surface rainfall. *Monthly Weather Review* **115**: 1053–1070.

Bacchi B, Ranzi R, Borga M. 1996. Statistical characterization of spatial patterns of rainfall cells in extratropical cyclones. *Journal of Geophysical Research* **101**(D21): 26 277–26 286.

Bell VA, Moore RJ. 1998. A grid-based distributed flood forecasting model for use with weather radar data. 2. Case studies. *Hydrology and Earth System Sciences* **2**(2–3): 278–283.

Beven KJ, Binley AM. 1992. The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes* **6**: 279–298.

Beven KJ, Freer J. 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modeling of complex environmental systems using the GLUE methodology. *Journal of Hydrology* **249**: 11–29.

Beven KJ, Kirkby MJ. 1979. A physically-based variable contributing area model of basin hydrology. *Hydrological Sciences Journal* **24**(1): 43–69.

Beven KJ, Lamb R, Quinn P, Romanowicz R, Freer J. 1995. TOPMODEL. In *Computer Models of Watershed Hydrology*, Singh VP (ed). Water Resource Publication: Fort Collins, CO; 627–668.

Borga M. 2001. Use of radar rainfall estimates in rainfall–runoff modeling: an assessment of predictive uncertainty. In *Proceedings of Fifth International Symposium on Hydrological Applications of Weather Radar—Radar Hydrology*, Kyoto, Japan; 451–456.

Borga M. 2002. Accuracy of radar rainfall estimates for streamflow simulation. *Journal of Hydrology* **267**: 26–39.

Borga M, Anagnostou EN, Frank E. 2000. On the use of real-time radar rainfall estimates for flood prediction in mountainous basins. *Journal of Geophysical Research* **105**(D2): 2269–2280.

Borga M, Tonelli F, Moore RJ, Andrieu H. 2002. Long term assessment of bias adjustment in radar rainfall estimation. *Water Resources Research* **38**(11): 1226.

Dinku T, Anagnostou EN, Borga M. 2002. Improving radar-based estimation of rainfall over complex terrain. *Journal of Applied Meteorology* **41**(12): 1163–1178.

Franks SW, Beven KJ. 1997. Bayesian estimation of uncertainty in land surface–atmospheric flux predictions. *Journal of Geophysical Research* **102**(D20): 23 991–23 999.

Freer J, Ambroise B, Beven KJ. 1996. Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach. *Water Resources Research* **32**: 2161–2173.

Georgakakos KP, Sperfslage JA, Guetter AK. 1996. Operational GIS based models for NEXRAD radar data in the U.S. In *Proceedings of the International Conference on Water Resources and Environmental Research*, 29–31 October, 1996, Water Resources and Environmental Research Center, Kyoto University, Kyoto, Japan; 603–609.

James WP, Robinson CG, Bell JF. 1993. Radar-assisted real-time flood forecasting. *Journal of Water Resources Planning and Management* **119**(1): 32–44.

Joss J, Lee R. 1995. The application of radar–gage comparisons to operational precipitation profile corrections. *Journal of Applied Meteorology* **34**: 2612–2630.

Nash JE, Sutcliffe JV. 1970. River flow forecasting through conceptual models, 1. A discussion of principles. *Journal of Hydrology* **10**: 282–290.

Ogden FL, Sharif HO, Senarath SUS, Smith JA, Beck ML, Richardson JR. 2000. Hydrologic analysis of the Fort Collins, Colorado, flash flood of 1997. *Journal of Hydrology* **228**: 82–100.

Quinn PF, Beven KJ, Chevallier P, Planchon O. 1991. The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. *Hydrological Processes* **5**: 59–79.

Quinn PF, Beven KJ, Lamb R. 1995. The $\ln(a/\tan b)$ index: how to calculate it and how to use it in the TOPMODEL framework. *Hydrological Processes* **9**: 161–182.

Romanowicz R, Beven KJ, Tawn J. 1994. Evaluation of predictive uncertainty in non-linear hydrological models using Bayesian approach. In *Statistics for the Environment. II. Water Related Issues*, Barnett V, Turkman KF (eds). John Wiley & Sons: Chichester, UK; 297–317.

Schell GS, Madramootoo CA, Austin GL, Broughton RS. 1992. Use of radar measured rainfall for hydrologic modelling. *Canadian Agricultural Engineering* **34**(1): 41–48.

Sempere-Torres D, Corral C, Raso J, Malgrat P. 1999. Use of weather radar for combined sewer overflows monitoring and control. *Journal of Environmental Engineering (ASCE)* **125**(4): 372–380.

Spear RC, Hornberger GM. 1980. Eutrophication in Peel Inlet, II. Identification of critical uncertainties via Generalized Sensitivity Analysis. *Water Research* **4**: 43–49.

Vieux BE, Bedient PB. 1998. Estimation of rainfall for flood prediction from WSR-88D reflectivity: a case study, 17–18 October 1994. *Weather and Forecasting* **13**: 407–415.

Winchell M, Gupta HV, Sorooshian S. 1998. On the simulation of infiltration- and saturation-excess runoff using radar-based rainfall estimates: effects of algorithm uncertainty and pixel aggregation. *Water Resources Research* **34**(10): 2655–2670.