Uncertainty-based automatic calibration of HEC-HMS model using sequential uncertainty fitting approach

S. Jamshid Mousavi, K. C. Abbaspour, B. Kamali, M. Amini and H. Yang

ABSTRACT

This study presents the application of an uncertainty-based technique for automatic calibration of the well-known Hydrologic Engineering Center-Hydrologic Modelling System (HEC-HMS) model. Sequential uncertainty fitting (SUFI2) approach has been used in calibration of the HEC-HMS model built for Tamar basin located in north of Iran. The basin was divided into seven sub-basins and three routing reaches with 24 parameters to be estimated. From the four events, three were used for calibration and one for verification. Each event was initially calibrated separately. As there was no unique parameter set identified, all events were then calibrated jointly. Based on the scenarios of separately and jointly calibrated events, different candidate parameter sets were inputted to the model verification stage where recalibration of initial abstraction parameters commenced. Some of the candidate parameter sets with no physically meaningful parameter values were withdrawn after recalibration. Then new ranges of parameters were identified based on minimum and maximum values of the remaining parameter sets. The new parameter ranges were used in an uncertainty analysis using SUFI2 technique resulting in much narrower parameter intervals that can simulate both verification and calibration events satisfactorily in a probabilistic sense. Results show that the SUFI2 technique linked to HEC-HMS as a simulation–optimization model can provide a basis for performing uncertainty-based automatic calibration of event-based hydrologic models.

Key words | automatic calibration, HEC-HMS, rainfall–runoff modelling, SUFI2

S. Jamshid Mousavi (corresponding author)
B. Kamali
School of Civil Engineering,
Amirkabir University of Technology,
Tehran,
Iran
E-mail: *jmosavi@aut.ac.ir*

K. C. Abbaspour
M. Amini
H. Yang
Swiss Federal Institute of Aquatic Science and Technology (Eawag),
Duebendorf,
Switzerland

INTRODUCTION

Conceptual hydrologic models play a significant role in predicting a basin's response to different climatic and meteorological processes within natural systems; however, these models require a number of estimated parameters. Model calibration is the procedure of adjusting the parameter values until the model predictions match observed data. Manual calibration of high-fidelity hydrologic (simulation) models is tedious, time consuming and sometimes impractical, especially when the number of parameters is large; moreover, the high degrees of nonlinearity involved in different hydrologic processes and non-uniqueness of inverse-type calibration problems make it difficult to find a single set of parameter values. There is a large body of literature on the automatic calibration of hydrologic models (e.g. Johnston & Pilgrim 1976; Gupta & Sorooshian 1985; Duan *et al*. 1992; Ibrahim & Liong 1993; Eckhardt & Arnold 2001; Ebtehaj *et al*. 2010). Nicklow *et al*. (2010) have provided a broad discussion on the state-of-the-art use of global optimization techniques and their application to hydrologic model calibration.

There are a number of challenging issues in the automatic calibration of hydrologic models such as: model structure and identifiability (e.g. Gan *et al*. 1996; Vrugt *et al*. 2002), parameters non-uniqueness (e.g. Sorooshian & Gupta 1983), type of objective function (e.g. Gan *et al*. 1996), multi-objective optimization (e.g. Yapo *et al*. 1998; Madsen 2000, 2003) and prediction uncertainty (Abbaspour *et al*. 2004, 2007).

Input parameters of hydrologic models are seldom known with certainty. Therefore, they are not capable of

doi: 10.2166/hydro.2011.071

describing the exact hydrologic processes. Input data and structural uncertainties related to scale and approximations in system processes are different sources of uncertainty that make it difficult to model exact hydrologic phenomena. Modelling uncertainty has been addressed in several studies (e.g. Muleta & Nicklow 2004; Moradkhani *et al.* 2005; Tolson & Shoemaker 2008). There are a number of algorithms dealing with how to include different types of uncertainty in calibration procedures such as maximum likelihood-based models (e.g. Sorooshian *et al.* 1983), Bayesian approach (e.g. Kuczera 1983a, b; Romanowicz *et al.* 1994; Kavetski *et al.* 2006a, b; Ajami *et al.* 2007; Thyer *et al.* 2009) and its extension to the generalized likelihood uncertainty estimation (GLUE) approach (Beven & Binley 1992; Freer *et al.* 1996; Beven & Freer 2001), parameter solution (ParaSol) (Van Griensven & Meixner 2006), sequential uncertainty fitting (SUFI2) (Abbaspour *et al.* 2004) and Markov chain Monte Carlo (MCMC) (e.g. Kuczera & Parent 1998). Yang *et al.* (2008) compared GLUE, ParaSol, SUFI2 and MCMC methods in an application to a watershed in China and found that the different methods each may converge to different solutions at different locations of the parameters space with more or less the same discharge results.

This literature review illustrates that most applications of uncertainty techniques are for the calibration of continuous hydrologic models. A continuous model simulates a longer period, predicting watershed response both during and between precipitation events. The present study is about the automatic calibration of the event-based HEC-HMS (Hydrologic Modelling System of Hydrologic Engineering Center of US Army Corps of Engineers) hydrologic model (USACE 2002) using the SUFI2 algorithm (Abbaspour *et al.* 2004). The algorithm has been used in basin-wide (Abbaspour *et al.* 2007), country-wide (Faramarzi *et al.* 2008) and continent-wide (Schuol *et al.* 2008) large-scale Soil and Water Application Tool (SWAT) (Arnold *et al.* 1998) calibration problems. In this study, the SUFI2 technique is linked to the HEC-HMS model and a three-stage procedure is proposed for uncertainty-based calibration of the combined SUFI2-HMS model. The application of the proposed model is presented through testing it in calibration of the HEC-HMS model built for the Tamar basin located in north-east Iran.

THE HEC-HMS MODEL

The HEC-HMS model (USACE 2002) was developed as a replacement for HEC-1, which has long been considered a standard model for hydrologic simulation. Most of the hydrologic models employed in HEC-HMS are event-based models simulating a single storm requiring the specification of all conditions at the start of the simulation. The soil moisture accounting model in the HEC-HMS is the only continuous model that simulates both wet and dry weather behaviour.

The basin processes in the HEC-HMS are modelled by six main components. The meteorological component is the first computational unit by which rainfall input is spatially and temporally distributed over the basin. Rainfall is subject to losses modelled by the rainfall loss component. The resulting effective rainfall contributes either to direct runoff, modelled by a direct runoff component where it is transferred to overland flow, or to groundwater aquifers, modelled by a baseflow component. For flood events, the baseflow may not be significant. Both overland flow and baseflow enter river channels where transition and attenuation of streamflow are simulated by the river routing component. Finally, the effect of reservoirs, detention basins and natural depressions (lakes, ponds, wetlands) is considered by the reservoir component. Each of the components may have a number of parameters that should be estimated through calibration.

In automatic calibration, the parameter values depend on the objective function of the search or optimization algorithm. In characterizing a runoff hydrograph, three characteristics of time-to-peak, peak of discharge and total runoff volume are of the most importance. It is therefore important that simulated and observed hydrographs match as much as possible in terms of those characteristics. We have used the weighted root mean squared error (RMSE) as the minimization objective function expressed as follows:

$$g(i) = \mathrm{RMSE} = \sqrt{\frac{\sum_i C_i (Q_{\mathrm{obs},i} - Q_{\mathrm{sim},i})^2}{k}} \quad (1)$$

where $Q_{\mathrm{obs},i}$ and $Q_{\mathrm{sim},i}$ are respectively observed and simulated discharges, k is the number of data, and C_i is the weight of each discharge value. Some desired discharges

may receive larger weights such as the peak flow. In this study, event-based models of HEC-HMS library are used and linked to the SUFI2 algorithm as an uncertainty based automatic calibration technique.

SUFI2 APPROACH

The problems with calibration render uncertainty analysis absolutely essential in hydrologic models. SUFI2 combines calibration and uncertainty analysis to find parameter uncertainties that result in prediction uncertainties bracketing most of the measured data, while producing the smallest possible prediction uncertainty band. In the SUFI2 approach, parameters are assumed to be stochastic. In theory, the parameters can have any distribution yet their cumulative distribution is uniform with an interval of 0–1. In application, however, SUFI2 uses uniform distribution in Latin Hypercube Sampling (LHS). This fits well with the HMS definition of each parameter as they all have a physically meaningful range; however, when the 95 per cent confidence interval of a parameter is calculated, SUFI2 uses a *t*-type distribution around the best parameter value (Equations (6) and (7)). These are sampled parameters taken from an LHS.

Starting with physically meaningful intervals for the parameters, the iterative LHS based calibration with SUFI2 aims to minimize the objective function and to minimize the parameter ranges while bracketing a reasonable percentage of the measured data in the 95 per cent prediction uncertainty (95 PPU) band. This index is a measure of the uncertainties being accounted for in the model and is referred to as the *P-factor* or the percentage of measured data bracketed by the 95 PPU.

Another measure quantifying the strength of a calibration/uncertainty analysis is the *R-factor*, which is the average thickness of the 95 PPU band divided by the standard deviation of the measured data. SUFI2 hence seeks to bracket most of the measured data with the smallest possible uncertainty band. SUFI2 starts by assuming a large parameter uncertainty and then decreases this uncertainty in steps while monitoring the *P-factor* and the *R-factor*. A *P-factor* of 1 and *R-factor* of zero is a simulation that exactly corresponds to the measured data. A balance must be often reached between the two factors. A short step-by-step description of SUFI2 algorithm is as follows (Abbaspour 2008).

Step 1. An objective function should be defined. There are 7 different options (multiplicative and summation form of RMSE, Chi-square, R^2, Nash–Sutcliffe, bR^2, and ranked sum of square error). In this application we used RMSE as defined in Equation (1).

Step 2. Physically meaningful absolute minimum and maximum ranges are defined for the parameters as

$$b_j : b_{j,\text{abs\_min}} \le b_j \le b_{j,\text{abs\_max}} \quad j = 1, \ldots, m \tag{2}$$

where b_j is the jth parameter and m is the number of parameters to be optimized.

Step 3. A sensitivity analysis is performed by keeping all parameters constant to realistic values, while varying each parameter within ranges assigned in the second step.

Step 4. Initial ranges are assigned to parameters for the first round of LHS

$$b_j : b_{j,\min} \le b_j \le b_{j,\max} \quad j = 1, \ldots, m \tag{3}$$

These ranges are smaller than the absolute ranges, are subjective and depend upon experience. The sensitivity analysis in step 3 provides a valuable guide for selecting appropriate ranges.

Step 5. LHS is used next to draw n sets of parameter combinations where n is the number of desired simulations. In this study, we have used a dynamic procedure in which n is larger in the first few iterations. The simulation program (HEC-HMS) is then run n times in each sampling round and the simulated output variable(s) of interest (discharge), corresponding to the measurements is saved.

Step 6. As a first step in assessing the simulations, the objective function (g) is calculated.

Step 7. A series of measures are calculated to evaluate each sampling round. First, the sensitivity matrix, J, of $g(b)$ is computed using

$$J = \Delta g_i / \Delta b_j^i \quad i = 1, \ldots, C_2^n, \; j = 1, \ldots, m \tag{4}$$

where Δg_i is the difference in the objective function between any two parameter sets and the number of comparisons

equals all possible combinations of parameter sets calculated by $C_2^n = n!/[2!*(n-2)!]$, which is the number of rows in the sensitivity matrix J. The relative difference of objective values with respect to parameter j ($\Delta g_i/\Delta b_j^i$) constitutes an element of J with m columns where m is the number of parameters. Note that Δb_j^i is the difference of jth parameter values for combination i.

An equivalent of a Hessian matrix H, is then calculated by following the Gauss–Newton method as $H = J^{\mathrm{T}}J$. Based on the Cramer–Rao theorem (Press *et al*. 1992), an estimate of the lower bound of the parameter covariance matrix, C, is calculated from $C = s_g^2(J^{\mathrm{T}}J)^{-1}$ where s_g^2 is the variance of the objective function values resulting from n runs. The estimated standard deviation and 95 per cent confidence interval of a parameter b_j is calculated from the diagonal elements of C as

$$s_j = \sqrt{C_{jj}} \tag{5}$$

$$b_{j,\mathrm{lower}} = b_j^* - \mathrm{abs}(t_{v,0.025})s_j \tag{6}$$

$$b_{j,\mathrm{upper}} = b_j^* + \mathrm{abs}(t_{v,0.025})s_j \tag{7}$$

where b_j^* is the parameter b_j for one of the best sets and v is the degree of freedom $(n-m)$. Parameter correlations can then be assessed as follows

$$r_{lj} = \frac{C_{lj}}{\sqrt{C_{ll}}\sqrt{C_{jj}}} \tag{8}$$

The correlation matrix r quantifies the change in the objective function as a result of a change in parameter l, relative to changes in the other parameters j.

In the calculations of this step, on one hand, n should be large enough in order to increase the chance of generating good solutions in each sampling round. On the other hand, when the parameter ranges become quite narrow after a few iterations, Δb_j for some js in J may tend to equal almost zero depending upon round-off and truncation errors and computation precision. This may make estimation of C elements numerically unstable, due to the $(J^{\mathrm{T}}J)$ term as it may become singular. In this case, we set those elements of J to zero based on the following interpretation: suppose $\mathrm{d}b_r \approx 0$ after LHS, the result would be

$$\frac{\mathrm{d}g}{\mathrm{d}b_r} = \sum_{j=1}^{b} \frac{\partial g}{\partial b_j}\mathrm{d}b_j/\mathrm{d}b_r$$

For $j = r$, we have $\partial g/\partial b_j = 0$ and for $j \neq r$ one can set $\mathrm{d}b_j/\mathrm{d}b_r = 0$ as the parameters are assumed to be independent leading to $\Delta g_i/\Delta b_r \approx 0$ in case $\mathrm{d}b_r \approx 0$.

Step 8. The 95 PPU is calculated by the 2.5th (X_{L}) and 97.5th (X_{U}) percentiles of the cumulative distribution of every simulated point. The average distance d between the upper and the lower 95 PPU is determined by

$$\bar{d}_x = \sum_{z=1}^{k} (X_{\mathrm{U},z} - X_{\mathrm{L},z})/k \tag{9}$$

where k is the number of observed data points. *R-factor* is subsequently expressed as

$$R\text{-}factor = d_x/\sigma_X \tag{10}$$

where σ_X is the standard deviation of the measured variable X (discharge). A value of less than 1 is a desirable measure for the *R-factor* (Abbaspour 2008).

Step 9. The value of d tends to be quite large during the first sampling round; hence, further sampling rounds are needed with updated parameter ranges calculated by

$$\begin{aligned} b'_{j,\min} &= b_{j,\mathrm{lower}} - \max\left\{\frac{(b_{j,\mathrm{lower}} - b_{j,\min})}{2}, \frac{(b_{j,\max} - b_{j,\mathrm{upper}})}{2}\right\} \\ b'_{j,\max} &= b_{j,\mathrm{upper}} + \max\left\{\frac{(b_{j,\mathrm{lower}} - b_{j,\min})}{2}, \frac{(b_{j,\max} - b_{j,\mathrm{upper}})}{2}\right\} \end{aligned} \tag{11}$$

where b' indicates updated values. Parameters of the best simulation is used to calculate $b_{j,\mathrm{lower}}$ and $b_{j,\mathrm{upper}}$. More details on SUFI2 approach may be found in Abbaspour (2008).

CASE STUDY AND THE MODEL SET UP

The study area is the Gorganroud river basin in Iran extending from the north-west of Khorasan province to the Caspian Sea in the eastern side. The basin is divided into three sub-basins named Tamar, Tangrah and Galikesh. Due to flash floods and subsequent damages, there is an urgent need for flood control management plans for the basin. The most severe flood event in Gorganrood occurred in August 2001 with a peak flow of 3,000 m^3/s. Annual distribution of rainfall depth in the basin varies between 200 and 850 mm (IWRI 2009). Data on measured flood events is lacking, hence one may have difficulty in building a calibrated model of the basin. In this study, we have focused on the Tamar basin as data for this region is more reliable. The Tamar basin has an area of about 1,530 km^2 with only one hydrometric station used for data compilation. Among all measured flood events, four have more reliable data: three events are used for calibration and one for verification in this study. The date and amount of peak flows of the events are given in Table 1. The hydrographs and interpolated hyetographs of each event are also shown in Figure 1.

Table 1 | Date, peak discharge and duration of flood events at the Tamar basin

Event	Date	Peak flow (m^3/s)	Duration (hr)
1	19 September 2004	128	20
2	6 May 2005	299	30
3	9 August 2005	783	19
4	8 October 2005	120	13

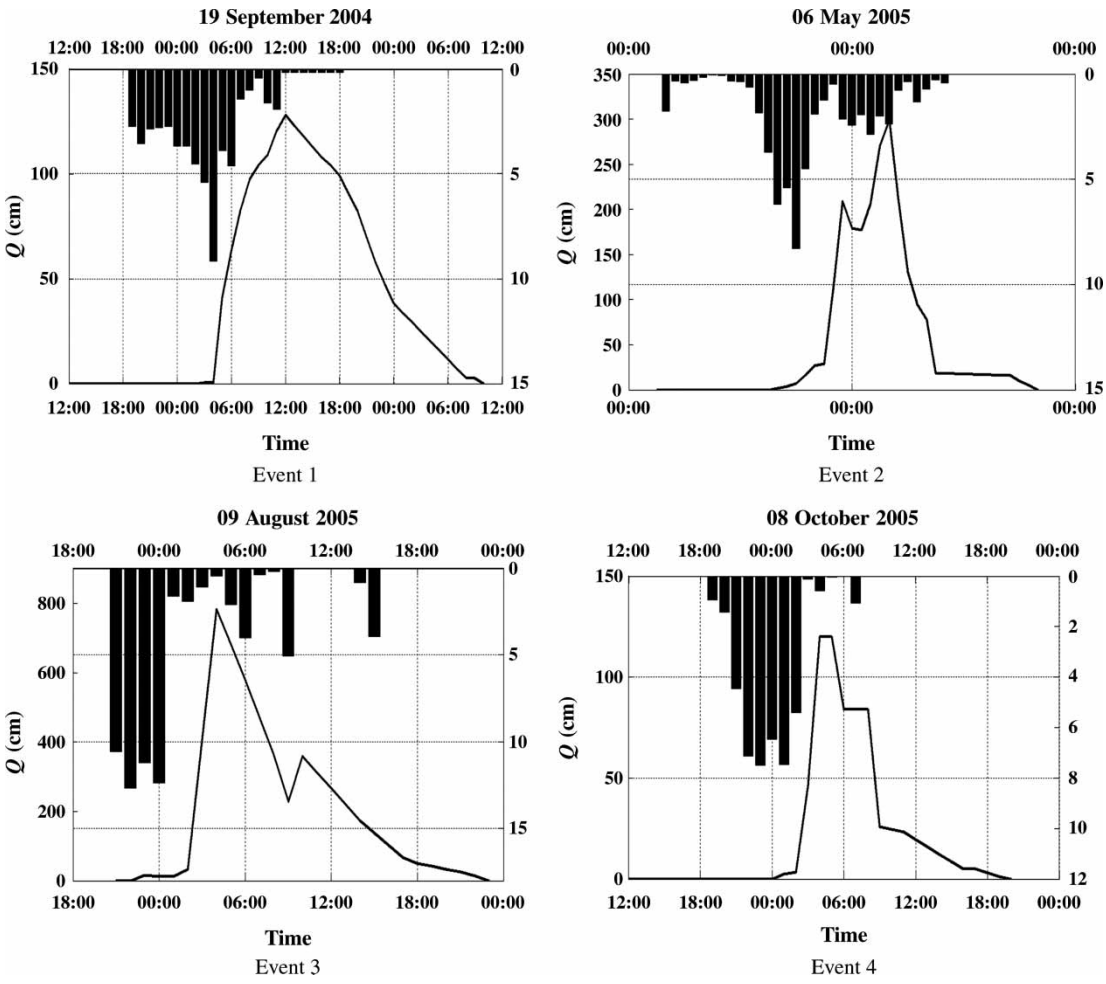

Figure 1 | Observed hydrographs and hyetographs of flood events at the Tamar basin.

The Tamar basin was divided into seven sub-basins with three routing reaches and four junctions (Figure 2). There exists only one hydrometric station with reliable measured discharge data located at the basin outlet (junction 3). Junctions are used to represent river confluence points or other important points of interest such as hydrometric stations. They are also needed for characterizing routing reaches as an important element of hydrologic models.

The Soil Conservation Service-Curve Number (SCS-CN) and Clark methods were used for estimating hydrologic losses and transforming rainfall to runoff. The Muskingum method was used for flow routing in the reaches. The SCS-CN method has two parameters which are curve number and initial abstraction coefficient. Upper and lower limits for the curve numbers were defined using physiographic maps of the sub-basins and soil conservation service table ± 5 (IWRI 2009). Initial abstractions were assumed to be between $0.15S$ and $0.25S$ (IWRI 2009) where S is the maximum retention of the basin. The Clark hydrograph method also has two parameters including time of concentration and storage coefficient. For the time of concentration, t_c, estimated from the basin's lag time, t_{lag}, the following equations were used.

$$t_{lag} = \frac{(L \times 3.28)^{0.8} \times (1,000/CN - 9)^{0.7}}{1,900y^{0.5}} \quad (12)$$

$$t_c = 1.67 * t_{lag} \quad (13)$$

where L is the length of river (m), y is the slope of sub-basin and S (cm) was already defined. As all of the above parameters were known, there was no need to calibrate the time of concentration. However, another equation has been presented relating storage coefficient, R, and time of concentration as follows

$$\frac{R}{R + t_c} = \text{const} \quad (14)$$

with suggested values between 0.2 and 0.6 for the constant number. We calibrated the constant number and consequently the storage coefficient. Therefore, each sub-basin has three parameters resulting in 21 parameters to which three routing parameters (one for each reach) are added. Of the two parameters of the Muskingum method, K was estimated by the *const* in Equation (14). Initial parameter ranges and values in each sub-basin were based on preliminary manual calibration and are given in Table 2 (IWRI 2009).

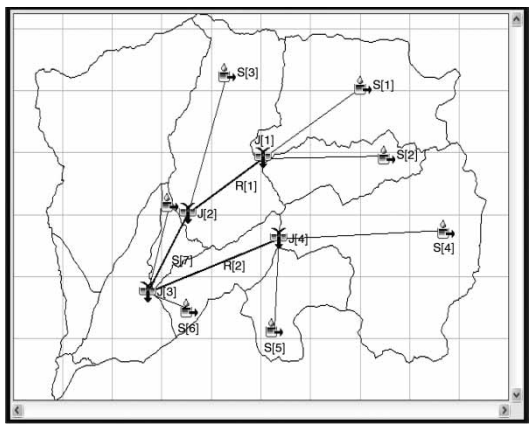

Figure 2 | Schematic representation of the Tamar basin model in HEC-HMS.

ANALYSIS AND RESULTS

Analyses with synthetic data

Before using SUFI2 in the case study and real data, it would be helpful to test how it performs against synthetically generated error-free data. The procedure was conducted in two stages: first, a number of error-free synthetic hydrographs were constructed using parameter values sampled from some assumed narrow intervals for parameters; second, the SUFI2 algorithm was used to test if the best parameter sets are within the assumed ranges and whether the simulated hydrographs based on the identified parameters are close to the error-free target hydrographs aforementioned.

LHS was used to generate the synthetic parameter sets. In this regard, the range of each parameter (i.e. its uncertainty) was taken as 10 per cent of its upper bound. Four different parameter sets were sampled from these selected ranges. Then runoff hydrographs at the basin outlet were simulated using the sampled parameter sets as synthetic target hydrographs. Subsequently, SUFI2 was used as an inverse approach for estimating the best parameter sets associated with each of the target hydrographs. Note that

Table 2 | Manually calibrated parameter values for each event and their upper and lower bounds (suggested by IWRI 2009)

		Upper limit	Lower limit	Event 1	Event 2	Event 3	Event 4			Upper limit	Lower limit	Event 1	Event 2	Event 3	Event 4
Curve number	Subbasin 1	91	60	66.1	86.3	80.6	91	Constant	Subbasin 1	0.65	0.2	0.34	0.47	0.53	0.28
	Subbasin 2	91	61	63.2	85.5	76.3	91		Subbasin 2	0.65	0.2	0.50	0.15	0.52	0.58
	Subbasin 3	87	58	80.1	82.1	70.2	87		Subbasin 3	0.65	0.2	0.53	0.53	0.63	0.49
	Subbasin 4	85	60	78.1	64.85	60.6	85		Subbasin 4	0.65	0.2	0.25	0.49	0.61	0.64
	Subbasin 5	84	50	51	84	84	82.9		Subbasin 5	0.65	0.2	0.25	0.32	0.64	0.58
	Subbasin 6	91	70	86.3	80.2	79.9	91		Subbasin 6	0.65	0.2	0.43	0.53	0.63	0.47
	Subbasin 7	91	70	77.2	72.1	80.0	91		Subbasin 7	0.65	0.2	0.43	0.51	0.64	0.61
Initial abstraction	Subbasin 1	0.25S	0.15S	0.24S	0.16S	0.22S	0.15S	Muskingum X	Reach 1	0.5	0.2	0.47	0.48	0.49	0.42
	Subbasin 2	0.25S	0.15S	0.18S	0.2S	0.20S	0.24S								
	Subbasin 3	0.25S	0.15S	0.17S	0.24S	0.22S	0.13S		Reach 2	0.5	0.2	0.21	0.41	0.40	0.42
	Subbasin 4	0.25S	0.15S	0.25S	0.22S	0.21S	0.24S								
	Subbasin 5	0.25S	0.15S	0.21S	0.23S	0.20S	0.21S		Reach 3	0.5	0.2	0.25	0.46	0.47	0.48
	Subbasin 6	0.25S	0.15S	0.19S	0.24S	0.23S	0.16S								
	Subbasin 7	0.25S	0.15S	0.22S	0.13S	0.19S	0.17S								

initial parameter ranges in the first SUFI2 iteration were much wider than the narrow bounds used to generate samples. All other basin and rainfall characteristics were similar to those considered for the Tamar basin case study. Table 3 reports the bounds and the sampled parameter values as well as parameter sets identified by SUFI2 and Figure 3 compares the hydrographs simulated using the sampled and the identified parameter values.

Figure 3 illustrates that most of parameter values identified by SUFI2 are within or close to the narrow bounds selected for the parameters. However, a number of parameter values are outside of the bounds and are not correctly identified, while the simulated hydrographs based on the identified parameter values are almost the same as the sampled error-free target hydrographs (Figure 3). This is an issue reported in several other studies (e.g. Sorooshian & Gupta 1983; Yang *et al*. 2008) showing the non-uniqueness feature of calibration problems as there could be different parameter values located at different regions of the parameters' space all resulting in more or less similar simulated hydrographs. Also, the response surface (objective function) of a calibration problem might be insensitive with respect to some of the parameters. These difficulties may cause non-identifiability problems for any automatic calibration technique.

We tested whether the non-identifiability issue with some of the parameters in our case is something related to SUFI2 only or if it could be found in other calibration approaches as well. This was done using two other global optimization techniques to identify parameter sets associated with each of the sampled error-free target hydrographs. These two techniques are particle swarm optimization (PSO) and genetic algorithms (GAs). The results comparing SUFI2 with the other two techniques are also presented in Table 3 and Figure 3. The objective function values obtained by different techniques are reported in the last row of the table.

Comparison of the objective values associated with the best solutions of the last iteration of SUFI2 with those of PSO and GAs show that, in all cases, SUFI2 has arrived at objective values better than or at least very close to, those of other global optimization techniques. Furthermore, some of the parameter values identified by SUFI2, PSO and GAs are different while their simulated hydrographs

Table 3 | Testing SUFI2 performance against synthetic parameter sets and hydrographs compared to PSO and GA

			Sample 1				Sample 2				Sample 3				Sample 4			
				Identified				Identified				Identified				Identified		
	Upper bound	Lower bound	Sampled	SUFI2	PSO	GA	Sampled	SUFI2	PSO	GA	Sampled	SUFI2	PSO	GA	Sampled	SUFI2	PSO	GA
CN_1	91	82	84.9	89.9	85.2	75.5	88.8	90.8	88.6	92.6	84.3	90.9	87.6	86.3	84.5	89.9	87.0	87.0
CN_2	91	82	84.2	80.7	82.3	82.3	82.6	86.9	88.7	88.4	84.3	62.8	88.5	66.9	83.2	73.8	86.6	83.6
CN_3	91	82	84.3	73.8	83.5	86.9	87.8	85.1	84.7	80.4	82.1	84.0	81.04	63.7	87.3	80.7	84.3	83.7
CN_4	91	82	87.5	84.9	84.8	88.8	89.2	84.9	84.8	84.9	90.6	84.9	84.8	90.3	87.47	84.9	83.9	85.0
CN_5	91	82	85.1	80.7	83.1	79.1	88.7	83.2	83.9	79.5	90.8	83.9	83.8	89.3	85.1	80.6	83.4	80.5
CN_6	91	82	89.5	88.1	90.4	88.6	85.5	71.7	89.1	94.8	89.3	86.5	87.8	85.9	89.4	88.1	85.7	84.7
CN_7	91	82	84.6	86.6	83.6	73.7	86.7	79.6	88.2	89.8	85.3	88.6	89.2	75.5	84.6	86.5	90.3	83.8
Ia_1	0.25	0.225	0.25	0.16	0.22	0.15	0.24	0.19	0.15	0.18	0.23	0.19	0.16	0.2	0.25	0.16	0.19	0.19
Ia_2	0.25	0.225	0.24	0.22	0.18	0.24	0.23	0.24	0.15	0.22	0.23	0.19	0.16	0.21	0.24	0.22	0.19	0.18
Ia_3	0.25	0.225	0.25	0.18	0.24	0.23	0.23	0.17	0.17	0.19	0.23	0.17	0.15	0.26	0.25	0.18	0.19	0.18
Ia_4	0.25	0.225	0.25	0.2	0.18	0.22	0.24	0.19	0.15	0.14	0.23	0.16	0.15	0.17	0.25	0.20	0.18	0.18
Ia_5	0.25	0.225	0.23	0.16	0.18	0.14	0.23	0.18	0.15	0.21	0.24	0.15	0.15	0.18	0.23	0.16	0.19	0.19
Ia_6	0.25	0.225	0.23	0.23	0.18	0.21	0.24	0.25	0.17	0.15	.24	0.19	0.24	0.16	0.23	0.23	0.17	0.23
Ia_7	0.25	0.225	0.24	0.17	0.17	0.19	0.23	0.17	0.17	0.18	0.23	0.24	0.16	0.16	0.24	0.18	0.24	0.22
$Cons_1$	0.6	0.54	0.54	0.59	0.58	0.39	0.55	0.57	0.58	0.6	0.56	0.59	0.56	0.46	0.54	0.59	0.56	0.56
$Cons_2$	0.6	0.54	0.59	0.26	0.51	0.32	0.58	0.52	0.21	0.6	0.54	0.59	0.59	0.26	0.59	0.25	0.59	0.5
$Cons_3$	0.6	0.54	0.57	0.59	0.54	0.67	0.57	0.44	0.57	0.49	0.54	0.49	0.23	0.38	0.58	0.59	0.57	0.56
$Cons_4$	0.6	0.54	0.58	0.51	0.51	0.58	0.59	0.44	0.51	0.51	0.56	0.45	0.46	0.53	0.58	0.51	0.54	0.53
$Cons_5$	0.6	0.54	0.55	0.38	0.56	0.57	0.59	0.35	0.50	0.39	0.56	0.27	0.33	0.51	0.55	0.38	0.48	0.37
$Cons_6$	0.6	0.54	0.55	0.42	0.44	0.28	0.57	0.55	0.58	0.6	0.59	0.27	0.58	0.50	0.55	0.42	0.33	0.23
$Cons_7$	0.6	0.54	0.59	0.42	0.29	0.41	0.55	0.52	0.56	0.5	0.56	0.58	0.57	0.47	0.58	0.42	0.5	0.49
$Xmusk_{,1}$	0.5	0.45	0.47	0.29	0.47	0.42	0.45	0.42	0.38	0.47	0.47	0.24	0.48	0. 29	0.47	0.29	0.43	0.41
$Xmusk_{,2}$	0.5	0.45	0.48	0.47	0.42	0.4	0.49	0.48	0.46	0.34	0.49	0.33	0.46	0.25	0.48	0.47	0.48	0.33
$Xmusk_{,3}$	0.5	0.45	0.47	0.28	0.37	0.4	0.47	0.48	0.33	0.32	0.46	0.21	0.45	0.26	0.46	0.28	0.48	0.46
RMSE	–	–	–	9.2	6.5	14.5	–	25	17.7	17.11	–	19.2	23.8	20.17	–	9.2	4.67	5.17

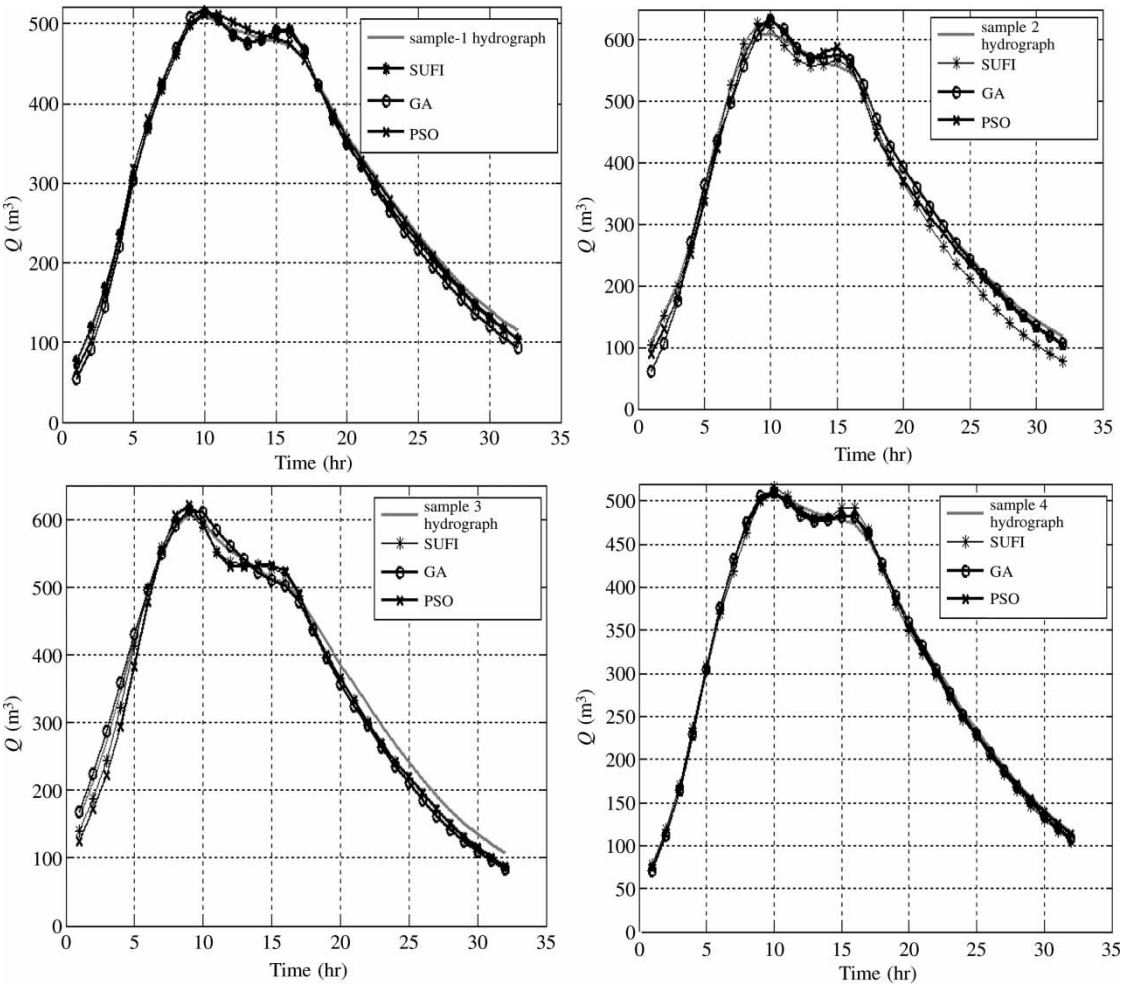

Figure 3 | Simulated hydrographs using sampled and identified parameter values by SUFI2, GA and PSO.

(Figure 3) are almost identical. This reveals the well-known non-uniqueness characteristic of inverse calibration problems reported in other studies such as Yang *et al.* (2008), wherein SUFI2 has been compared with three other uncertainty based techniques (i.e. GLUE, MCMC and ParaSol).

Calibrating the hydrologic model of the Tamar basin using SUFI2 approach

Of the four flood events, the first three events were used for calibration and the last for verification. Figures 4 and 5(a and b) show the parameter and simulated discharge sets including the best solution in the first and last iterations of the SUFI2 approach, respectively. Figure 5(c) depicts the objective value of the best solution against the iteration number for event 1. It is seen that the best simulated discharge of the last iteration is close to the observed flood hydrograph, especially the rising limb; however, a slight divergence is seen at terminal hours of the recession curve. Attempts taken to improve the simulated recession curve showed that divergence can be removed at the expense of worsening peak discharge and time-to-peak of the simulated hydrograph.

Figures 6(a) and 7(a) show the last-iteration's simulated discharges and convergence curve of the best objective value for the second event. For this event and event 3, the variation of parameter ranges and discharges in different iterations are not shown due to a lack of space and because they were the same as that of Figure 4 for the first event. We found no parameter sets that could accurately simulate the

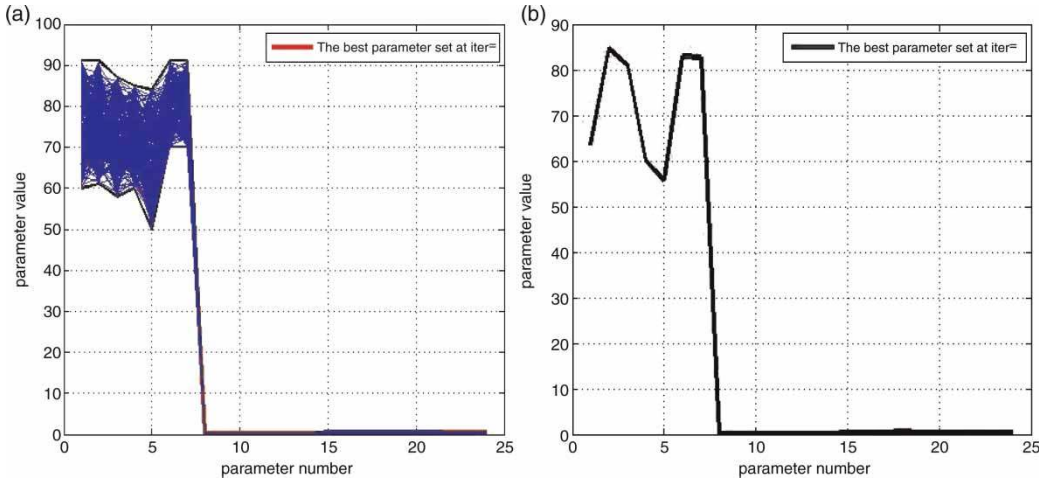

Figure 4 | Best parameter set, parameter bounds and 200/25 simulated parameter sets in the first (plot a)/last (plot b) iteration for event 1. Note: the best set may not be observable due to large number of simulations.

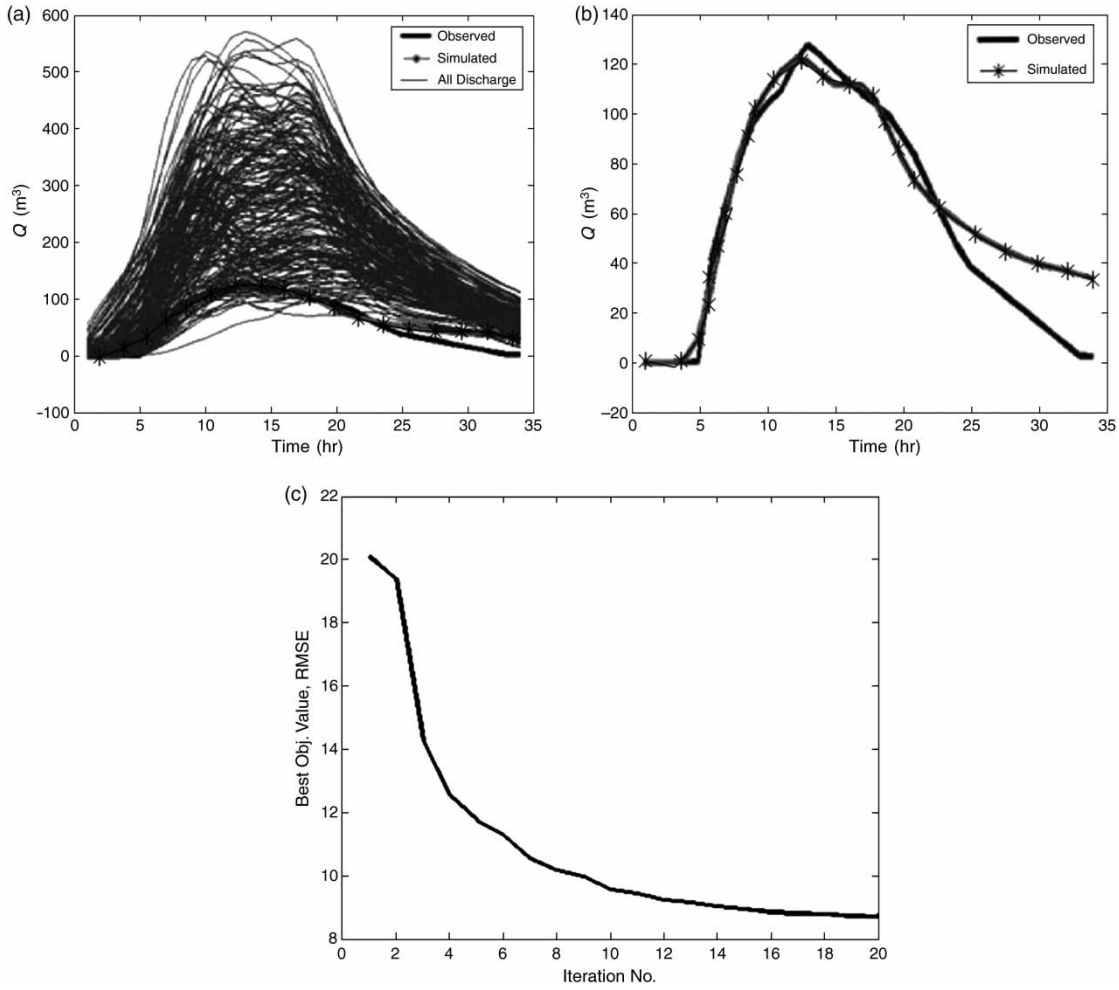

Figure 5 | Best and 200/25 simulated discharge hydrographs at basin outlet in the first (plot a)/last (plot b) iteration and convergence of objective function (plot c), event 1.

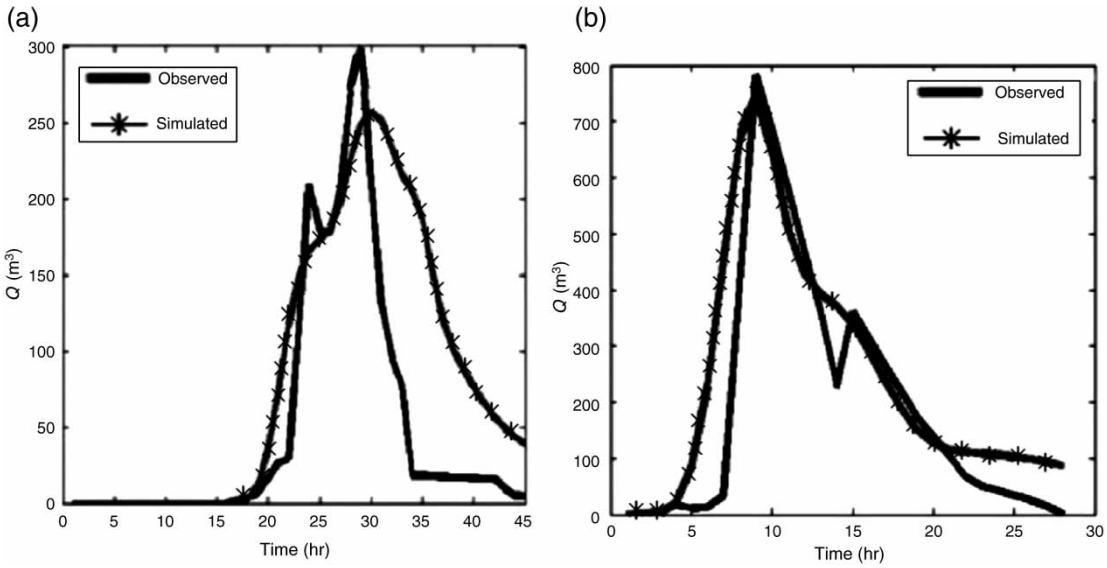

Figure 6 | Best and 25 simulated discharge hydrographs at basin outlet in the last iteration, events 2 (plot a) and 3 (plot b).

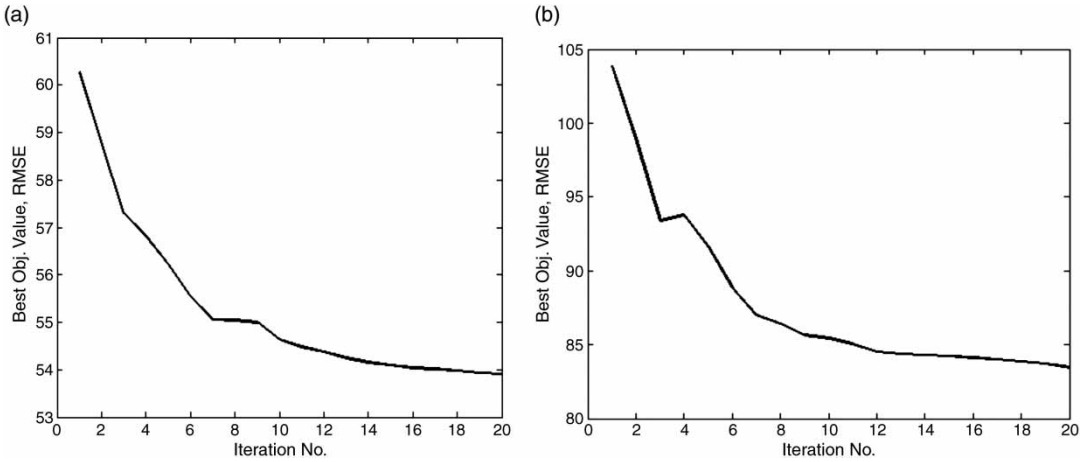

Figure 7 | Convergence curve of the best objective function, events 2 (plot a) and 3 (plot b).

second event, particularly its recession limb. A significant effort was put into improving the calibration of this event by running the model with a larger number of simulations in each iteration and also with a larger iteration number as well as widening parameter ranges. The results, however, were almost the same as the one shown in Figure 6(a) indicating that this was not a calibration issue.

Calibrated hydrograph of event 3 is illustrated in Figure 6(b) with its objective function convergence curve for the best simulation in Figure 7(b).

As mentioned before, the number of simulations was dynamically tuned. In the first iteration, around 200 were chosen in order to increase the chance of generating a good or near global optimal solution. Results showed that this many simulations were unnecessary and thus we used 100 in the second iteration and 25 in subsequent iterations. This strategy performed quite satisfactorily and the results were similar to the case where 200 simulations were made in each iteration. Moreover, as the iterations proceed and the parameter bounds become narrower in subsequent iterations, generating a large number of parameter sets within much narrower parameter intervals may cause numerical instabilities in calculating elements of the Hessian matrix.

After inspecting the calibrated parameter values of events 1 to 3 in different iterations, we found that each event may be simulated by either a single- or interval-

valued (as SUFI2 estimates parameter bounds associated with different uncertainty levels) sets of parameters. No unique parameter values (intervals) were found by which all three events could be reasonably simulated. This is of course understandable for those parameters related directly to the initial moisture condition of the basin such as initial abstraction coefficients, although it has been reported (IWRI 2009) that the anticipated moisture condition for all three events before starting the flood storms had been the average condition. Moreover, the non-uniqueness feature of parameter values were tested by comparing parameter intervals in iterations with acceptable *P-factor*, *R-factor* and RMSE values to test the width of intersection of the intervals of separately calibrated events. It was found that overlaps of the intervals were small. We will test this further in the next step when joint calibration of events is performed.

Figure 8 illustrates the last iteration's discharges and the best objective function's convergence curve in which events 1, 2 and 3 are calibrated jointly. This shall be called scenario-0 hereafter. Figure 8 shows that although events 2 (Figure 8(b)) and 3 (Figure 8(c)) are simulated as well as they were in the single-event calibration cases, event 1 (Figure 8(a)) has not been calibrated adequately as its volume and peak discharge are significantly overestimated. Moreover, increasing the number of simulations and iterations as well as enlarging the parameter ranges did not significantly improve the situation. In another attempt named scenario-0W, a larger weight (say 5) was given to the RMSE of event 1 in the objective function in

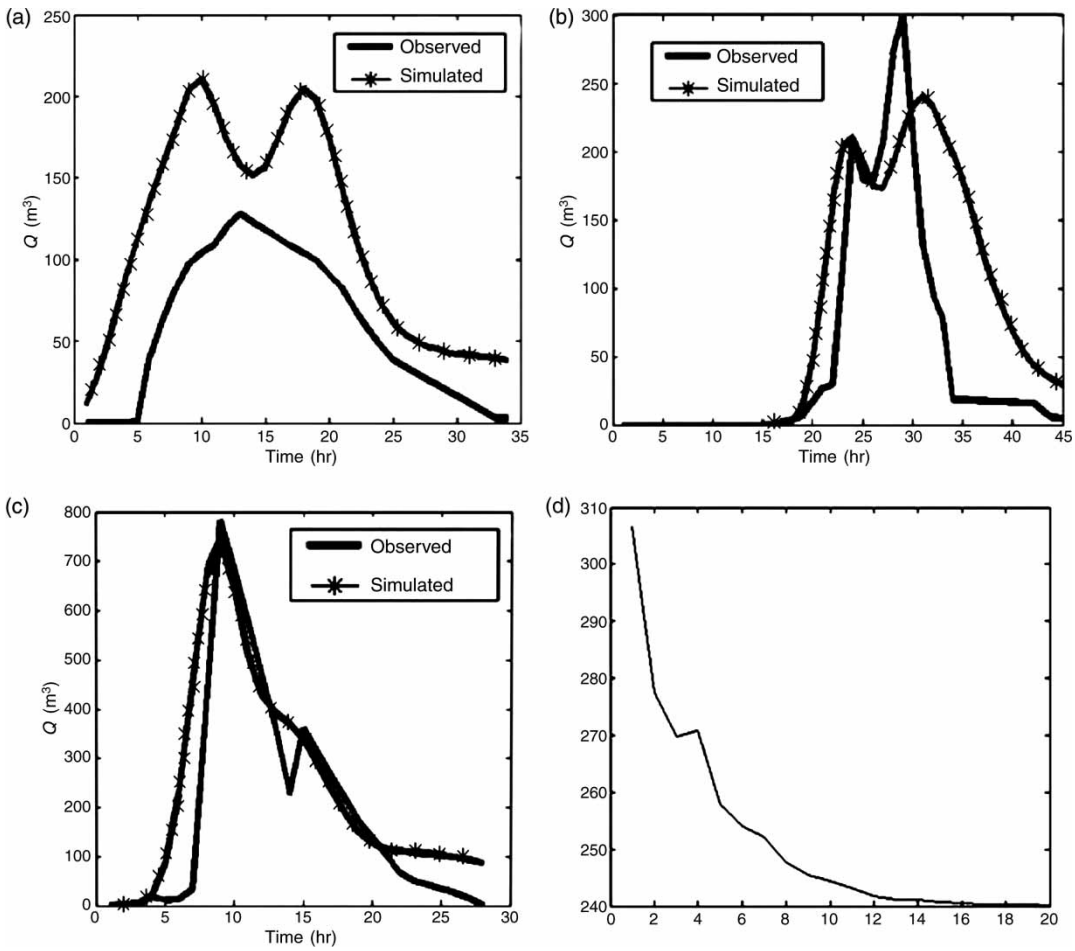

Figure 8 | Best and 25 simulated discharge hydrographs of events 1 (plot a), 2 (plot b) and 3 (plot c) at basin outlet in the last iteration and the convergence curve of the best objective function (plot d), jointly calibrated events, scenario-0.

order to encourage the model to find a parameter set by which event 1 can be better simulated. The simulated discharges (not shown because of lack of space) showed those of event 3 will be significantly under-estimated, even though flow and runoff volume of event 1 were simulated much better.

Results of scenarios-0 and 0W support this hypothesis that there is no unique set of parameters in which events 1 and 3 can simultaneously be calibrated. As a result, we decided to increase the number of calibration parameters, especially those directly related to the basin's initial condition. To keep the number of parameters as small as possible, only seven parameters including either initial abstraction parameters of event 1 or event 3 were considered to be different. Thus, the number of independent calibration parameters was increased from 24 to 31 without widening parameter ranges. As a result, flood hydrographs were better simulated but still with larger RMSE values compared with those obtained from single-event calibration scenarios. Therefore, the upper bounds of initial abstraction coefficients of event 1 were increased from 0.25 to 0.45, while keeping those of events 2 and 3 at 0.25. This scenario, referred to as scenario-1, allows the model to decrease peak flow and flood volume of the first event's simulated hydrograph. The last-iteration's simulated hydrographs compared with the observed floods including the variation of the best objective value in different iterations are depicted in Figure 9.

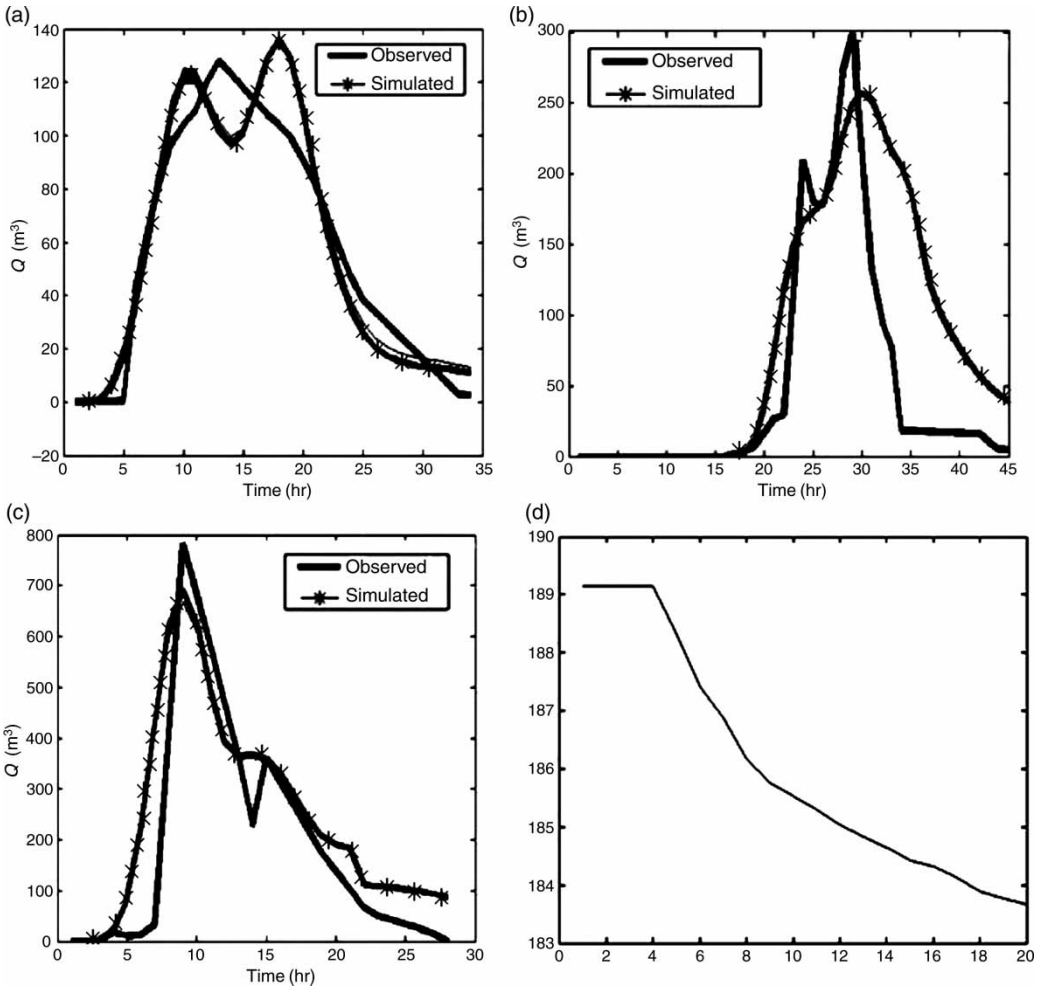

Figure 9 | Best and 25 simulated discharge hydrographs of events 1 (plot a), 2 (plot b) and 3 (plot c) at basin outlet in the last iteration and the convergence curve of the best objective function (plot d), jointly calibrated events with 31 parameters, scenario-1.

Figure 9 shows that all events have been reasonably simulated, although event 1's hydrograph is bi-modal. The better obtained performance can be verified by comparing Figures 9(d) and 7(d) where the best simulated objective function of scenario-1 (183.7) is better than scenario-0 (about 240). Inspecting the best parameter values in scenario 1 implies that initial abstraction coefficients of two sub-basins, out of seven, were about 0.4, which is not physically justified. More physically meaningful coefficients could have been achieved if we had added another set of seven parameters to treat the initial abstraction coefficients of all events differently, or consider separate time of concentrations at least for one of the events. However, we believe this may not be helpful, because the number of parameters should be kept as small as possible as long as reasonable simulated discharges are achieved and the price paid is only widening the parameter ranges slightly beyond their initially selected values. This is certainly an issue needing more attention because the uncertainty in parameters that have physical values that are not meaningful may come from different sources such as: gauge precipitations, observed discharges, neglecting spatial distribution of precipitation, and/or incompleteness of the adopted rainfall–runoff model, and the functions representing physical processes taking place in the basin.

Another way to improve the model fit is to decrease the lower bound of initial abstractions of event 3 and keep their upper bounds unchanged (say 0.25) for all events. Therefore, in another scenario (scenario-2) the lower bounds of initial abstractions of event 3 were decreased from 0.15 to zero in order to let the model increase the peak flow and flood volume of event 3. Note that in this case the number of parameters was again 31 assuming that the initial abstraction parameters of events 1 and 2 can be different from event 3. Furthermore, (not shown because of lack of space) all three discharge hydrographs were reasonably simulated while event 1's discharge hydrograph remained bi-modal. Inspecting the best parameter values of event 3 obtained in this scenario (scenario 2) revealed that the initial abstractions of two sub-basins were less than 0.05. In conclusion, one should have either increased the initial abstraction coefficients of event 1 or decreased those of event 3 in order to have all three events jointly calibrated.

Because in both scenarios 1 and 2, the initial abstractions at least for some sub-basins were either larger (scenario 1) or smaller (scenario 2) than the initially selected upper and lower bounds, we were interested to test if it was possible to find another set of parameters in which initial abstractions were within, or close to, the initially selected bounds. For this reason, the 31-parameter calibration problem was re-run with the upper bounds of initial abstraction coefficients of event 1 set to 0.35 (instead of 0.45 in scenario 1) and the lower bounds of the coefficients of other events (2 and 3) set to 0.05 (instead of zero in scenario 2). This scenario, named scenario 3 resulted in the same results as those of scenarios 1 and 2 in terms of simulated discharges and the best objective function.

There are now three parameter sets (scenarios 1–3) simulating all calibration events more or less properly but with different initial abstraction coefficients. We hoped that the other 17 parameters of these three sets could be almost the same. If it was so, we would say that there is a unique set of parameters by which all calibration events and hopefully validation events could be simulated. However, this was not the case, which we believe could be due to the non-uniqueness feature of the calibration problem. So the question of which parameter set is better or closer to the true parameter values still remains difficult to answer. On the other hand, the answer could also depend upon the criterion or criteria characterizing the meaning of 'better' and 'true' in the context of uncertainty-based modelling. We will discuss this issue later. Although the set with initial abstractions closer to the physically meaningful values (scenario 3) may be preferred, this is something that needs to be evaluated in the verification stage.

VERIFICATION ANALYSES

There are nine sets of calibrated candidate parameter sets: three sets resulting from separately calibrated events, one set as the mean values of the three sets, and five sets obtained from joint calibrations (i.e. 0, 0W, 1, 2 and 3). Figure 10(a) illustrates the simulated flood hydrograph of event 4, as the verification event, by the parameters of the above nine sets.

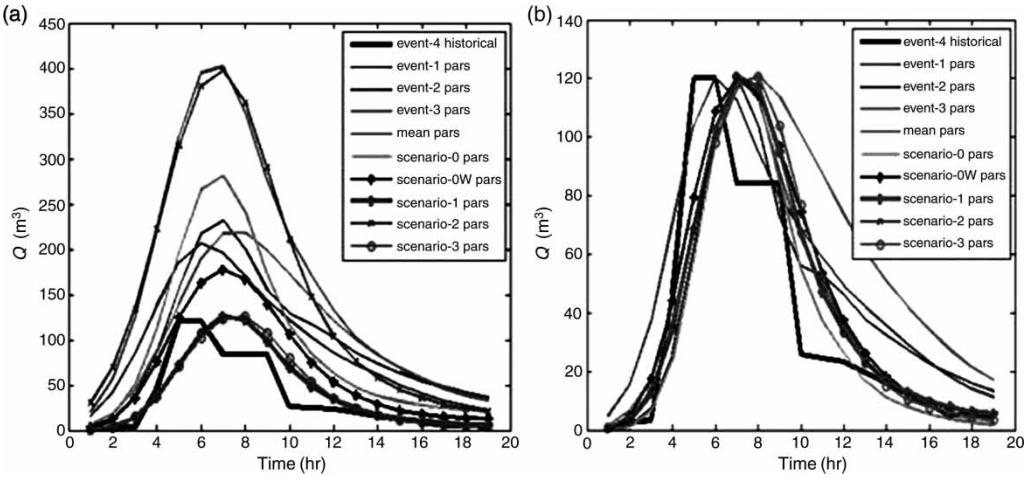

Figure 10 | Simulating event 4 by the parameter sets obtained in calibration stage without (plot a) and with (plot b) recalibrating initial abstractions.

Results show that none of the parameter sets can simulate event 4's hydrograph, although two of them (i.e. those of scenarios 1 and 3 of jointly calibrated events) perform better than others. However, because seven calibrated parameters deal with initial abstraction, these parameters need to be calibrated for event 4. Calibration here means finding a constant coefficient, which upon multiplication by initial abstractions makes the peak discharge of each set almost equal to that of the observed ones. This procedure guarantees relative ratios of initial abstractions of sub-basins to remain unchanged which makes sense from a physical point of view. Figure 10(b) shows all the simulated hydrographs with recalibrated initial abstraction parameters.

Generally, a model should not be recalibrated in the verification stage as verification is for testing a calibrated model to check how well that model performs against data not used in the calibration stage. However, recalibration has been done only for the initial abstraction parameters and not the basin parameters. This is a feature that should be recognized in event-based hydrologic models because parameters representing the initial basin conditions should in reality not be calibrated. If parameters related to initial conditions of the basin are not properly identified for each specific flood event, other basin parameters will not be correctly identified. On one hand, we can not use all of the data and events in the calibration stage; as it is necessary to test how the calibrated model and the parameters suggested will perform against future events and to make sure the calibrated model has not been overfitted. On the other hand, the parameter values related to the basin initial conditions obtained based on calibration events may not be representative for basin initial conditions of other flood events such as verification events. Therefore, they need to be adjusted according to the other events' initial conditions.

Figure 10(b) shows that most of the simulated hydrographs are similar because they have reconstructed peak discharges with a slight error in the time-to-peak value, except in the first solution set. Moreover, parameter sets have difficulty with simulating the atypical recession curves of the observed hydrograph; however, jointly calibrated event scenarios (the last five scenarios) perform better than others (first four sets). Consequently, it would be hard to select the best parameter set among those nine sets. However, comparison of their simulated objective function (RMSE) and initial abstraction (I_a) ranges could be helpful (Table 4 gives such additional information).

As shown in Table 4, the set obtained by averaging the parameter values of the single-event calibration scenarios has not performed as well as the other sets. Moreover, parameter sets whose initial abstractions are either less than 0.05 or greater than 0.45 were considered to be less physically meaningful; hence, sets three and eight and the one with averaged parameter values were removed. There remained six parameter sets whose RMSE were close

Table 4 | Simulated objective function and initial abstraction values of all recalibrated parameters

Parameter sets	Event 1	Event 2	Event 3	Mean Events	Scenario 0	Scenario 0W	Scenario 1	Scenario 2	Scenario 3
Simulated RMSE	21.6431	22.8237	21.4624	38.0544	21.4112	21.6312	21.5938	21.5938	24.6156
Sub-basin's 1 I_a coefficient	0.4005	0.2989	0.5201	0.3008	0.4567	0.2938	0.4070	0.95	0.2205
Sub-basin's 2 I_a coefficient	0.3579	0.2586	0.6050	0.3013	0.4545	0.2941	0.4227	0.95	0.3092
Sub-basin's 3 I_a coefficient	0.3380	0.2894	0.4394	0.2631	0.4584	0.2100	0.4538	0.9500	0.2309
Sub-basin's 4 I_a coefficient	0.4052	0.2747	0.5936	0.3141	0.3902	0.3200	0.4263	0.9500	0.3570
Sub-basin's 5 I_a coefficient	0.3563	0.3628	0.4826	0.2964	0.3606	0.2710	0.3968	0.0054	0.3581
Sub-basin's 6 I_a coefficient	0.3846	0.3113	0.4698	0.2875	0.3474	0.2312	0.2258	0.6586	0.2564
Sub-basin's 7 I_a coefficient	0.3155	0.3906	0.4749	0.2913	0.4426	0.1979	0.1796	0.8876	0.1246
Min I_a	0.3155	0.2586	0.4749	0.2631	0.3474	0.1979	0.1796	0.0054	0.1246
Max I_a	0.4052	0.3906	0.6050	0.3141	0.4584	0.3200	0.4538	0.95	0.3581

while their parameter values were within acceptable ranges. The simulated RMSEs of the five remaining sets range between 21.41 and 22.82, which are statistically identical. The last solution (scenario 3) has a slightly larger RMSE value of 24.62, but it was not removed from consideration because its ranges of initial abstractions were closer to the initially selected lower and upper bounds. The parameter values of those six solutions, other than initial abstractions, are presented in Table 5.

Table 5 shows that the ranges of parameter values are still wide and non-unique; hence, a unique set cannot be selected. So far we have not addressed the indices of uncertainty in SUFI2, mainly *P-factor* and *R-factor*. SUFI2 does not search for single-valued parameters but intervals

Table 5 | Comparison of parameter values, other than initial abstractions, of six screened-out parameter sets

Parameter values	Event 1	Event 2	Scenario 0	Scenario 0W	Scenario 1	Scenario 3
CN1	65.97	78.69	82.57	77.70	83.60	78.10
CN2	84.34	72.56	83.59	82.72	78.76	76.31
CN3	76.03	67.82	67.28	60.91	68.66	58.30
CN4	60.05	60.84	62.11	60.01	61.09	60.96
CN5	58.39	79.92	77.11	63.42	72.94	71.47
CN6	88.88	76.43	74.35	76.37	73.76	70.19
CN7	83.61	75.07	72.34	70.32	79.52	71.04
SC1	0.5199	0.4171	0.2785	0.2232	0.2879	0.2355
SC2	0.2984	0.5614	0.4742	0.4091	0.3182	0.3575
SC3	0.3095	0.2776	0.5406	0.5968	0.4700	0.4476
SC4	0.5987	0.5660	0.5961	0.5988	0.2696	0.4446
SC5	0.4968	0.2005	0.2640	0.2302	0.3039	0.3102
SC6	0.5213	0.3188	0.5529	0.2593	0.5114	0.5380
SC7	0.4201	0.3144	0.3512	0.2559	0.3877	0.2502
X1	0.4464	0.3245	0.3226	0.2012	0.2381	0.4232
X2	0.2540	0.4333	0.4564	0.3781	0.3563	0.3831
X3	0.4063	0.3692	0.3345	0.3912	0.3186	0.4711

within which different randomly generated samples produce acceptable hydrographs.

DEALING WITH UNCERTAINTY

At this stage, the minimum and maximum values of each parameter of the six solutions presented in Table 5 were selected to form the new uncertainty bounds of parameters. Then SUFI2 was run with the new bounds to calculate uncertainty indicators, that is, *P-factors* and *R-factors*. Table 6 presents the results obtained by running SUFI2 in the above-mentioned condition. The indicators in the first iteration show the performance of the new parameter intervals in terms of model prediction uncertainty. It is seen that *P-factor* in this iteration is very good (0.895). However, although the newly-estimated parameter ranges are much narrower than those initially selected at the start of calibration procedure, they result in a significant model prediction uncertainty (*R-factor* = 1.656).

Fortunately, better parameter ranges with a lesser total prediction uncertainty (smaller *R-factor*) with acceptable *P-factors* (*P-factor* > 65%) have been achieved in some iterations. One can see by inspecting Table 6 that the parameter ranges in iteration three could be a good solution in which the *P-factor* is equal to 0.737, while significant reduction in model prediction uncertainty has been obtained with an *R-factor* equal to 0.807. Figure 11(a) shows the new parameter bounds (first iteration's values) and the bounds in iteration three (Figure 11(b)) and Figure 12 illustrates simulated discharges including the best hydrograph compared

Table 6 | Uncertainty indicators and the best solution's RMSE obtained from running SUFI2 with new parameter bounds for verification event 4 (values in parentheses are for smoothed hydrograph)

Iteration	1	2	3	4	5	6	7	8	9	10
P-factor %	89.47 (94.74)	84.21 (78.95)	73.68 (84.21)	68.42 (68.42)	47.3 (68.42)	31.58 (47.37)	36.84 (42.11)	26.32 (36.84)	15.79 (26.32)	15.78 (26.32)
R-factor	1.6559 (1.6538)	1.0795 (0.813)	0.8849 (0.4887)	0.7457 (0.3653)	0.5234 (0.3430)	0.2096 (0.2267)	0.1767 (0.1676)	0.1296 (0.1140)	0.1129 (0.1110)	0.0704 (0.0837)
Par-factor	0.2047 (0.2047)	0.1285 (0.1098)	0.0783 (0.0741)	0.0571 (0.0566)	0.0395 (0.0433)	0.0289 (0.0319)	0.0227 (0.0225)	0.0149 (0.0154)	0.0117 (0.0125)	0.0085 (0.0090)
Best sol's RMSE	14.45 (9.1706)	13.62 (9.2093)	13.74 (8.7700)	13.36 (7.6496)	13.22 (7.4637)	13.05 (7.2927)	12.78 (7.1373)	12.69 (7.0989)	12.62 (7.0065)	12.56 (6.9803)

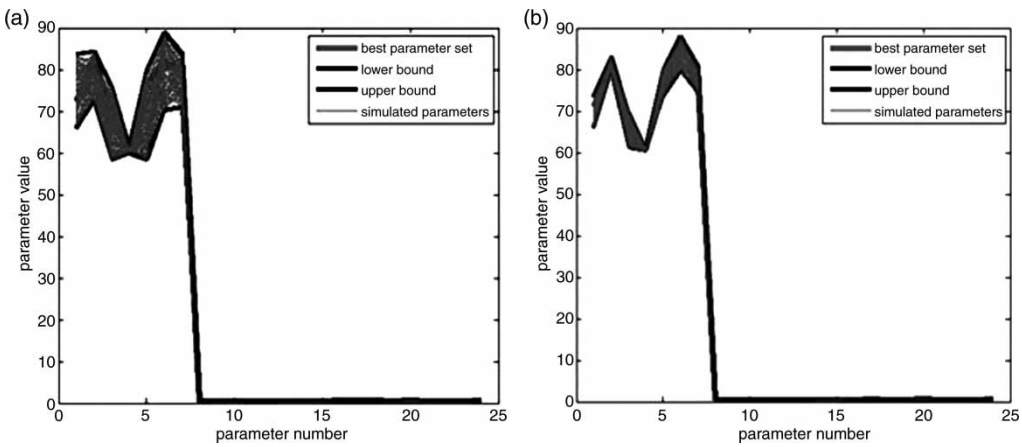

Figure 11 | Simulated parameter sets including the best set and the new parameter bounds in the first (plot a) and third (plot b) iterations of SUFI2 for verification event 4.

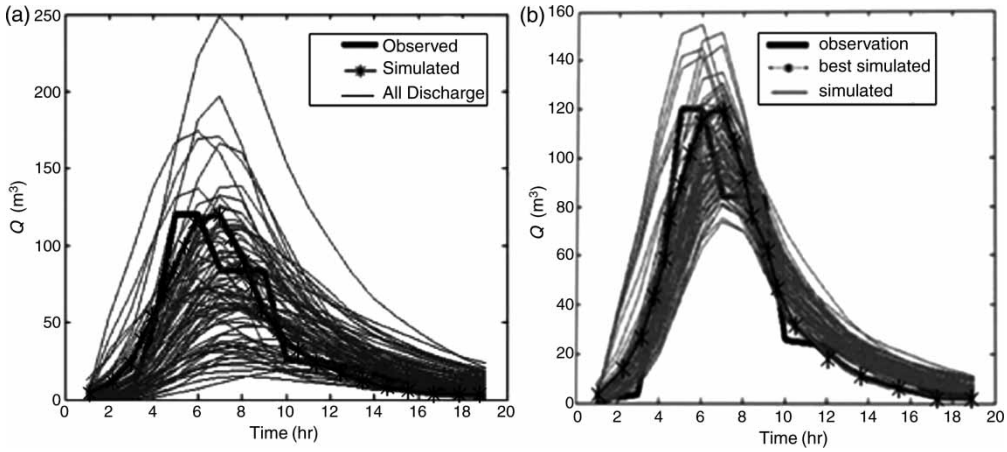

Figure 12 | Simulated discharges with the new parameter bounds compared with the observed flood in the first (plot a) and third (plot b) iterations of SUFI2 for verification event 4.

with the observed flood in the first (Figure 11(a)) and the third (Figure 11(b)) iterations of SUFI2.

Comparison of plots (a) and (b) in Figure 11 clearly shows the advantage of iteration 3's solution in which the model parameter uncertainty (width of parameter ranges) has been significantly decreased. To quantify the model parameter uncertainty, we define another factor, *Par-factor*, in SUFI2 algorithm which is a normalized sum of parameter ranges as follows

$$Par\text{-}factor = \frac{\sum_{j=1}^{m}(b_{\max_j} - b_{\min_j})}{\sum_{j=1}^{m}(b_{\max_j} + b_{\min_j})/2} \quad (15)$$

where $b_{\max_j}$ and $b_{\min_j}$ are respectively the upper and lower bounds of parameter *j*. *Par-factor* compared with *R-factor* shows how a certain degree of uncertainty inputted to the model propagates through the model relationships resulting in model prediction uncertainty. In fact, if the basin model was a linear system with respect to its parameters, the relation between *Par-factor* and *R-factor* would be linear. However, hydrologic models are often highly nonlinear. Therefore, a small amount of model parameter uncertainty (small *Par-factor*), when propagated through model nonlinear functions and equations, may still cause significant model prediction uncertainty (large *R-factor*). Figure 13 shows the variations of *Par-factor* and *R-factor* against iteration number. This represents the dynamics of model prediction uncertainty, or in other words the rate of model prediction uncertainty reduction with respect to the rate of model parameter uncertainty reduction. Consequently, break points at which slope of the

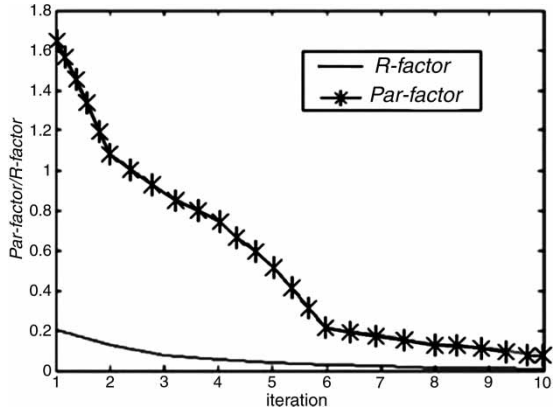

Figure 13 | Variation of *Par-factor* versus *R-factor* in the last run of SUFI2 with the new initial parameter ranges for event 4.

plot changes could help one choose the final parameter ranges of the model. It can be seen in the figure that the rate of reduction of model prediction uncertainty (*R-factors*) in the first six iterations are larger than that in succeeding iterations.

It is noteworthy to mention here that sometimes *P-factor* is dominated by some uncharacteristic and perhaps less important ordinates of the predicted variable (discharge), such as an unusual recession curve of the observed flood hydrograph of event 4 and also less important terminating ordinates of the curve. For example, the terminating segment of the recession curve hydrograph of event 4 contains a significant number of discharge ordinates, prediction of which is not as important as the ones close-to-peak discharges. This might result in an unreal and under-estimated *P-factor* exaggerating undesirable fitting capability of model

predictions. Depending upon the shape of the observed predicted signals, it is our belief that estimating the *P-factor* and *R-factor* using a smoothed hydrograph can result in more realistic values. Due to the atypical shape of the observed hydrograph of event 4 and especially its recession curve, we repeated our analysis by smoothing out the hydrograph of the event. The results are reported in Table 6 within parenthesis. The *P-factors* and also RMSE values are better than those associated with non-smoothed hydrograph.

According to the above explanation, the final parameter ranges of the model were selected as those of the fourth iteration with respectively *P-factor* = 0.684, *R-factor* = 0.365, *Par-factor* = 0.057 and RMSE = 7.65. Keeping these parameters ranges fixed, we ran SUFI2 one more time to calibrate the initial abstraction parameters. It is crucial to emphasize that we must not allow SUFI2 to change and re-adjust the selected parameter ranges any more in this model execution; hence, the SUFI2 algorithm was revised in this stage to keep all parameter ranges fixed all over iterations except those of initial abstractions. The indicator values for each of the events and also the simulated hydrographs in iteration 4, compared to observed floods, as well as the best objective values are presented in Table 7 and Figure 14. Here we see that the RMSE of the best objective function values are quite close to those already obtained in the previous calibration stage, which shows the fact that the proposed parameter ranges include such sets of parameters resulting in the best RMSE values ever found. Also, it was verified that the *P-factor* and *R-factor* values after a few iterations of the model, when initial abstractions were stabilized, are close to the values obtained in the calibration stage. The results indicate that the proposed parameter ranges were properly selected. Figure 14 also confirms that the selected parameter ranges perform satisfactorily. The *P-factors* of events 2 and 3 are not large enough but this is mostly related to atypical shapes of the hydrographs making it impossible to have large values of *P-factor*s and small values of *R-factor*s, simultaneously. We checked this issue in the scenarios of jointly calibrated events with 31 parameters and realized that the same outcome had been obtained in those scenarios. Moreover, inspecting plots (b) and (c) in Figure 14 implies that smaller *P-factor*s of events 2 and 3 are partly due to less-important discharge ordinates in terminating segment of the recession curves of the events. Overall, the results and analyses presented indicate that the final parameter ranges and the proposed procedure to locate them performed satisfactorily considering the very limited number and atypical shapes of the flood hydrographs available in the studied basin.

COMPARING SUFI2 WITH NELDER AND MEAD SEARCH ALGORITHM

In this section and before concluding the study, we explore why it is justified to calibrate HEC-HMS using SUFI2, while there are other available automatic techniques built into the software.

There are two methods of automatic calibration available in HMS: the first is the univariate gradient, which evaluates and adjusts one parameter at a time while holding other parameters constant; the second method is the Nelder and Mead (NM) which uses a downhill simplex search

Table 7 | Results of first 10 iterations obtained by SUFI2 for the problem of jointly calibrated events with fixed parameter intervals except initial abstraction ranges

Iteration		1	2	3	4	5	6	7	8	9	10
Best objective Function		223.3	207.5	196.7	191.7	188.2	189.5	188.8	186.7	189.0	188.4
Par-factor		0.089	0.084	0.082	0.081	0.080	0.079	0.079	0.079	0.078	0.078
Event 1	*P-factor* %	79.41	91.17	58.82	73.52	64.70	73.52	70.58	67.64	67.64	73.52
	R-factor	2.617	1.594	1.356	1.110	0.955	0.885	0.820	0.788	0.768	0.736
Event 2	*P-factor*	60.00	53.33	48.88	37.77	22.22	20.00	17.77	20.00	20.00	20.00
	R-factor	0.858	0.497	0.403	0.364	0.317	0.312	0.286	0.271	0.268	0.271
Event 3	*P-factor*	46.42	28.57	28.57	32.14	28.57	21.42	25.00	25.00	21.42	21.42
	R-factor	0.559	0.357	0.291	0.256	0.244	0.253	0.255	0.232	0.226	0.229

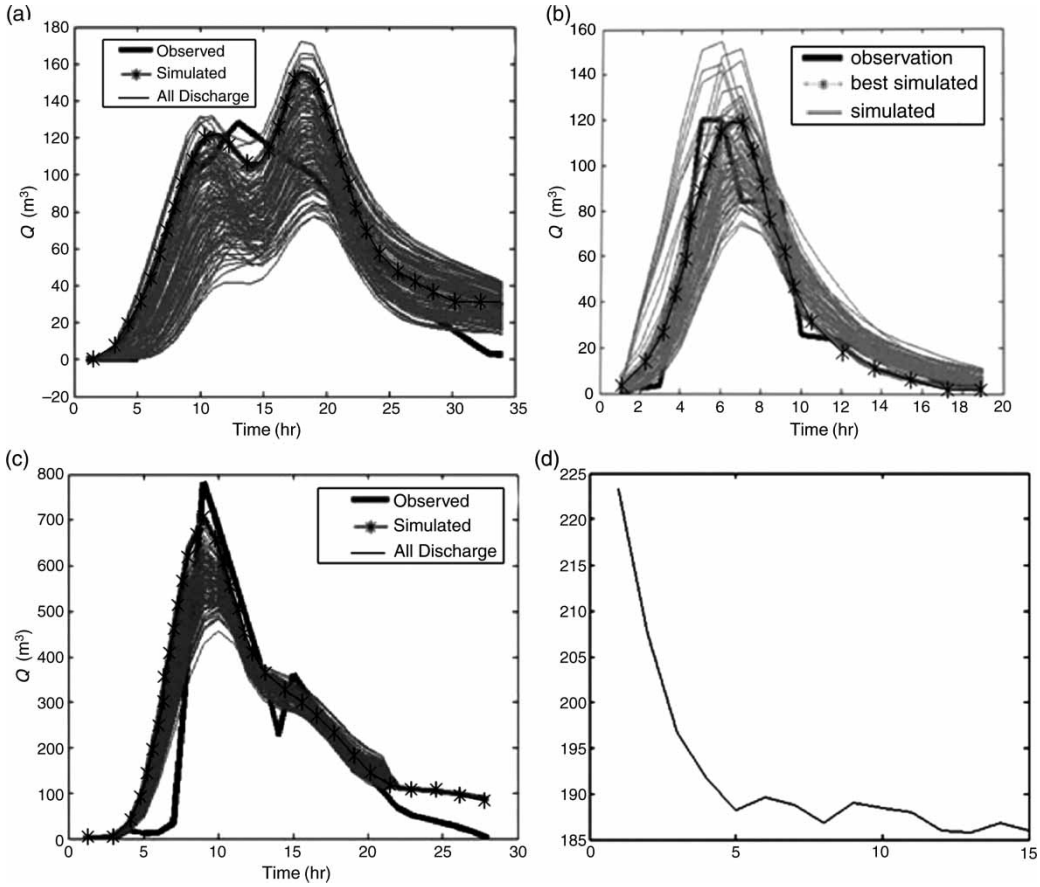

Figure 14 | Iteration 4's simulated discharges of events 1 (plot a), 2 (plot b) and 3 (plot c) and the convergence curve of the best objective value (plot d) obtained by SUFI2 for the problem of jointly calibrated events with fixed parameter intervals except initial abstractions.

algorithm allowing all the parameters to vary simultaneously and determines which parameter should be adjusted. The default method is the univariate gradient.

The SUFI2 technique is essentially an uncertainty based technique. Although it is capable of finding the best parameter values within the proposed parameter ranges, SUFI2 does not look for single-valued parameter sets. None of the two available automatic calibration techniques in HMS are uncertainty based, while the proposed three-step procedure is related to how proper ranges of parameter values should be determined in event-based calibration problems considering the non-uniqueness feature of the problems – an impossibility by any deterministic calibration approach like the ones available in HMS.

Another issue is that the modeller can choose any performance criterion as the objective function of the SUFI2-HMS model where different weights can be assigned to different discharge values such as the weights used in the objective function of the proposed model. Although there are a number of alternatives for the selection of objective function, they are limited and there is no possibility for assigning different weights to different discharge ordinates. In other words, the weighted RMSE, as considered in HMS-SUFI2, is not among the alternative objective functions of the HMS calibration techniques.

In HMS, the parameters that can be selected as calibration parameters of the available techniques are limited and pre-defined. Therefore, the user is not allowed to select any parameter that they want to calibrate while this limitation does not exist in the methodology proposed in SUFI2-HMS. Fixing the parameters to be calibrated could be useful for a user with less expertise; while flexibility in selecting calibration parameters may be useful for those having more understanding on how different parameters

may affect the response of the hydrologic model being calibrated.

Besides the above points, there are other serious limitations with the available calibration methods in HMS with respect to the proposed methodology. An important limitation is that calibrating more than one event jointly, as done for joint-type calibration scenarios, is not possible using the available automatic techniques in HMS – one can only do single-event type calibration. In this regard, we attempted to do joint-type calibration by available calibration schemes in HMS through introducing multiple events sequentially as one synthetic single event. This was done by changing starting time of rainfall hyetographs of the events, synthetically. However, it was not possible to consider different initial conditions for the events like what was done in our analysis using SUFI2. This is an important limitation as we clearly showed that single event calibration scenarios may not be adequate and the first tree events cannot be properly calibrated jointly unless separating initial abstraction parameters of events 1 and 3. Also, some of the variables of the SCS hydrologic loss model like cumulative rainfall depth could not be introduced correctly to the model when using the mentioned approach, for example, doing joint-type calibration through simulating multiple single events as one synthetic event.

As the most important reason, it is now tested how well the available techniques will perform compared with the SUFI2 approach in single event scenarios. To do so, it is important to select first which technique is better to be used for calibration. In this regard, the NM algorithm could be more preferable than the univariate gradient as the latter performs poorly for non-convex and multimodal response surfaces; because it focuses on searching for only one parameter at a time that can result in convergence to a local optimum solution, especially when the number of parameters is large. Therefore, the NM technique was selected for comparison with the best parameter set obtained by the SUFI2 approach.

Figure 15 shows the results obtained by the NM method. Note that the results are associated with the model outputs

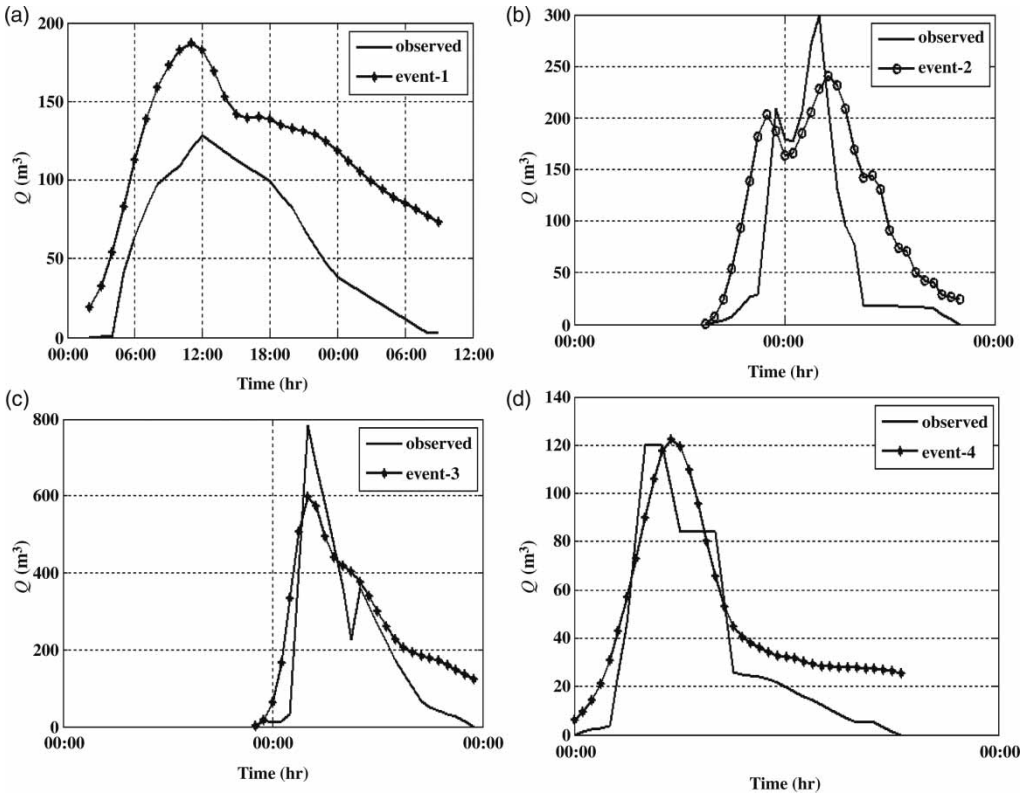

Figure 15 | Single event calibration using Nelder and Mead search algorithm available in HEC-HMS.

after a number of trials as the first set of optimized parameters performed poorly. Therefore, they were improved iteratively by using solution of a trial as the initial solution of the next one. Comparing the results presented in Figure 15 with those illustrated in Figures 5(b) and 6 shows that NM has not performed as satisfactorily as SUFI2 approach for single event calibration.

SUMMARY AND CONCLUSIONS

A three stage analysis was proposed to find the parameter ranges of the HEC-HMS hydrologic model built for the Tamar basin in Iran. In the first stage, a number of candidate parameter sets were identified using the sequential uncertainty fitting (SUFI2) technique based on scenarios of both separately and jointly calibrated events. After recalibrating the basin's initial condition parameters of the candidate sets and then performing a screening procedure through which some of the candidate sets were withdrawn, a new set of parameter ranges were selected according to the minimum and maximum values of the screened parameter sets. In the second stage, the SUFI2 algorithm was run with the new parameter ranges, which were much narrower than those of initially selected at the start of calibration. The outcome of the run was the final parameter ranges simulating verification event as adequately as possible in a probabilistic sense. Finally, it was tested in the last stage if the set of final selected ranges can simulate all events satisfactorily in terms of the uncertainty indicators, that is, *P-factor* and *R-factor*.

The above analysis revealed the existence of different parameter sets that can altogether simulate verification events quite well, which shows the non-uniqueness feature of the calibration problem under study. However, the methodology has benefited from that feature by finding new parameter intervals that should be fine tuned further in order to decrease input and model prediction uncertainties. A quantified measure of parameter uncertainty was introduced as *par-factor*, which along with *P-factor* and *R-factor* measures of SUFI2 gives insights to how the parameter or input-to-model uncertainty will propagate through a hydrologic model structure and how SUFI2 can reach to a balanced solution in terms of input and output model uncertainties.

The SUFI2-HMS model, from another point of view, may be considered as a simulation-optimization procedure in which the simulation model is the HEC-HMS model linked to the SUFI2 algorithm as an optimization algorithm. In this regard, SUFI2 is able to search for a deterministic solution by sequentially bracketing of the parameter ranges over 15–20 iterations, when the parameter ranges become extremely narrow. In this case, SUFI2 acts as a probabilistic search technique like other population-based evolutionary algorithms to locate the best deterministic parameter values with respect to a defined objective function. From this point of view, SUFI2 was compared with PSO and GA as two global optimization techniques, wherein SUFI2 performed comparably to the other optimization techniques. This is important as the Nelder and Mead algorithm, available in HEC-HMS, neither performed satisfactorily for single-event calibration scenarios nor can it be used for joint-event calibration scenarios.

The proposed methodology performed well in the automated calibration of an event-based hydrologic model; however, the authors are aware of a drawback of the presented analysis – this undertaking was not a completely fair validation procedure. It is because validation events represent possible future scenarios and thus are not available at the time of model calibration. Hence, an event being selected as a validation event should not be used to receive any more feedback for adjusting parameter values and ranges. However, this remark was not fully taken into consideration, mostly because of being seriously short of enough observed events in this calibration study. Therefore, the proposed methodology, although sound and useful, should be validated in other case studies with more observed flood events.

ACKNOWLEDGEMENTS

This work was carried out during the sabbatical period of the first author who was supported partially by Amirkabir University of Technology and partially by Eawag. The ideas provided by Peter Riechart and Monireh Faramarzi of Eawag are greatly acknowledged. Also, the help of M. Akhtari and F. Kouhiyan of IWRI in providing data and information on the case study and B. Saghafian for his

comments on the first draft of the manuscript are highly appreciated.

REFERENCES

Abbaspour, K. C. 2008 *Swat-Cup2: SWAT Calibration and Uncertainty Programs Manual Version 2, Department of Systems Analysis, Integrated Assessment and Modelling (SIAM), Eawag*. Swiss Federal Institute of Aquatic Science and Technology, Duebendorf, Switzerland.

Abbaspour, K. C., Johnson, A. & van Genuchten, M. Th. 2004 Estimating uncertain flow and transport parameters using a sequential uncertainty fitting procedure. *Vadose Zone Journal* **3** (4), 1340–1352.

Abbaspour, K. C., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J., Zobrist, J. & Srinivasan, R. 2007 Modelling hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. *J. Hydrol.* **333** (2–4), 413–430.

Ajami, N. K., Duan, Q. & Sorooshian, S. 2007 An integrated hydrologic Bayesian multimodel combination framework: confronting input, parameter, and model structural uncertainty in hydrologic prediction. *Water Resour. Res.* **43**, W01403.

Arnold, J. G., Srinivasan, R., Muttiah, R. S. & Williams, J. R. 1998 Large area hydrologic modeling and assessment part I: model development. *J. Am. Water Res. Assoc.* **34** (1), 73–89.

Beven, K. & Binley, A. 1992 The future of distributed models – model calibration and uncertainty prediction. *Hydrological Processes* **6** (3), 279–298.

Beven, K. & Freer, J. 2001 Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *J. Hydrol.* **249** (1–4), 11–29.

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

Ebtehaj, M., Moradkhani, H. & Gupta, H. V. 2010 Improving robustness of hydrologic parameter estimation by the use of moving block bootstrap resampling. *Water Resour. Res.* **46**, W07515, 14 PP.

Eckhardt, K. & Arnold, J. G. 2001 Automatic calibration of a distributed catchment. *J. Hydrol.* **251** (1–2), 103–109.

Faramarzi, M., Abbaspour, K. C., Yang, H. & Schulin, R. 2008 Application of SWAT to quantify internal renewable water resources in Iran. *J. Hydrol. Sci.* **48** (1), 65–77.

Freer, J., Beven, K. & 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.

Gan, T. Y., Dlamini, E. M. & Biftu, G. F. 1996 Effects of model complexity and structure, data quality and objective functions on hydrologic modeling. *J. Hydrol.* **192** (1–4), 81–103.

Gupta, H. V. & Sorooshian, S. 1985 The automatic calibration of conceptual catchment models using derivative-based optimization algorithms. *Water Resour. Res.* **21** (4), 473–485.

Ibrahim, Y. & Liong, S.-Y. 1993 A method of estimating optimal catchment model parameters. *Water Resour. Res.* **29** (9), 3049–3058.

Iran Water Research Institute, Water Resources Department (IWRI) 2009 Report on Hydrologic Model Calibration: Gorganroud Flood Warning System Project. Tehran, Iran.

Johnston, P. & Pilgrim, D. 1976 Parameter optimization for watershed models. *Water Resour. Res.* **12** (3), 477–486.

Kavetski, D., Kuczera, G. & Franks, S. W. 2006a Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. *Water Resour. Res.* **42**, W03407.

Kavetski, D., Kuczera, G. & Franks, S. W. 2006b Bayesian analysis of input uncertainty in hydrological modeling: 2. Application. *Water Resour. Res.* **42**, W03408.

Kuczera, G. 1983a Improved parameter inference in catchment models 1. Evaluating parameter uncertainty. *Wat. Resour. Res.* **19** (5), 1151–1162.

Kuczera, G. 1983b Improved parameter inference in catchment models 2. Combining different kinds of hydrologic data and testing their compatibility. *Water Resour. Res.* **19** (5), 1163–1172.

Kuczera, G. & Parent, E. 1998 Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. *J. Hydrol.* **211** (1–4), 69–85.

Madsen, H. 2000 Automatic calibration of conceptual rainfall-runoff model using multiple objectives. *J. Hydrol.* **235** (3–4), 276–288.

Madsen, H. 2003 Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. *Adv. Water Res.* **26** (2), 205–216.

Moradkhani, H., Sorooshian, S., Gupta, H. V. & Houser, P. 2005 Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Adv. Water Res.* **28** (2), 135–147.

Muleta, M. K. & Nicklow, J. W. 2004 Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. *J. Hydrol.* **306** (1–4), 127–145.

Nicklow, J., Reed, P. M., Savic, D., Dessalegne, T., Harrell, L., Chan-Hilton, A., Karamouz, Minsker, B., Ostfeld, A., Singh, A. & Zechman, E. 2010 State of the art for genetic algorithms and beyond in water resources planning and management. *ASCE Journal of Water Resources Planning & Management* **136** (4), 412–432.

Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. 1992 *Numerical Recipe, The Art of Scientific Computation*, 2nd edition. Cambridge University Press, Cambridge, UK.

Romanowicz, R. J., Beven, K. & Tawn, J. 1994 Evaluation of predictive uncertainty in nonlinear hydrological models using a Bayesian approach. In *Statistics for the Environment 2, Water Related Issues* (V. Barnett & K. F. Turkman, eds.). Wiley, Chichester, pp. 297–315.

Schuol, J., Abbaspour, K. C., Yang, H., Srinivasan, R. & Zehnder, A. 2008 Modeling blue and green water availability in Africa. *Water Resour. Res.* **44**, W07406.

Sorooshian, S. & Gupta, H. V. 1983 Automatic calibration of conceptual rainfall-runoff models: the question of parameter observability and uniqueness. *Water Resour. Res.* **19** (1), 260–268.

Sorooshian, S., Gupta, H. V. & Fulton, J. 1983 Evaluation of maximum likelihood parameter estimation techniques for conceptual rainfall-runoff models: influence of calibration data variability and length on model credibility. *Wat. Resour. Res.* **19** (1), 251–259.

Thyer, M., Renard, B., Kavetski, D., Kuczera, G., Franks, S. W. & Srikanthan, S. 2009 Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis. *Water Resour. Res.* **45**, W00B14.

Tolson, B. A. & Shoemaker, C. A. 2008 Efficient prediction uncertainty approximation in the calibration of environmental simulation models. *Water Resour. Res.* **44**, W04411.

USACE 2002 *Hydrologic Modeling System HEC-HMS, Applications Guide*. US Army Corps of Engineers, Hydrologic Engineering Center, Washington, DC.

Van Griensven, A. & Meixner, T. 2006 Methods to quantify and identify the sources of uncertainty for river basin water quality models. *Water Sci. Tech.* **53** (1), 51–59.

Vrugt, J. A., Bouten, W., Gupta, H. V. & Sorooshian, S. 2002 Toward improved identifiability of hydrologic model parameters: The information content of experimental data. *Water Resour. Res.* **38** (12), 1312.

Yang, J., Abbaspour, K. C., Reichert, P. & Yang, H. 2008 Comparing uncertainty analysis techniques for a SWAT application to Chaohe Basin in China. *J. Hydrol.* **358** (1–2), 1–23.

Yapo, P. O., Gupta, H. V. & Sorooshian, S. 1998 Multi-objective global optimization for hydrologic models. *J. Hydrol.* **204** (1–4), 83–97.

First received 8 May 2010; accepted in revised form 17 January 2011. Available online 10 May 2011