Reliability of Semiarid Flash Flood Modeling Using Bayesian Framework

Article *in* Journal of Hydrologic Engineering · September 2016

CITATIONS
0

READS
9

4 authors, including:

Mohsen Pourreza Bilondi
University of Birjand
17 PUBLICATIONS **10** CITATIONS
SEE PROFILE

S. Zahra SAMADI
University of South Carolina
33 PUBLICATIONS **112** CITATIONS
SEE PROFILE

Bijan Ghahraman
Ferdowsi University Of Mashhad
57 PUBLICATIONS **320** CITATIONS
SEE PROFILE

Reliability of Semiarid Flash Flood Modeling Using Bayesian Framework

Mohsen Pourreza-Bilondi[1]; S. Zahra Samadi[2]; Ali-Mohammad Akhoond-Ali[3]; and Bijan Ghahraman[4]

Abstract: A case study examining Bayesian techniques for assessing parameter and predictive uncertainty of semiarid flash flood events is presented here. The focus is on testing a fully distributed rainfall-runoff model (i.e., AFFDEF) linked with Markov chain Monte Carlo (MCMC) samplers to simulate four semiarid flash flood events with varying rainfall durations (<24 h) and amounts (>20 mm). MCMC samplers showed consistent behaviors with the a priori assumption and successfully improved performances on complex and multivariate search problems of semiarid flood simulation over the Abol-Abbas watershed, Iran. Analysis suggests that parameters associated with infiltration and interception capacity along with the contributing area threshold for the digital river network were the key model parameters and were more influential on the shape and volume of the flood hydrograph. Model predictive uncertainty was heavily dominated by error and bias in the soil water storage capacity, which reflects inadequate representation of the upper soil zone processes in the AFFDEF distributed model. Overall, the modeling results revealed that a fat-tailed Gaussian distribution using the standard least-squares (SLS) error assumption yielded improved estimates of parameter and predictive uncertainty for the semiarid flood events. This case study emphasizes the importance of proper statistical representation of the residual error distribution as a basis to improve parameter and predictive uncertainty. **DOI: 10.1061/(ASCE)HE.1943-5584.0001482.**

Author keywords: Flash flood modeling; Parameter and predictive uncertainty; Markov chain Monte Carlo sampler; Semiarid watershed.

Introduction

The consequences and trends toward a warmer climate have increased the chances of flooding (Milly et al. 2002) and consequently have increased flood losses (e.g., Kundzewicz and Kaczmarek 2000; Sharifi et al. 2012). Arid and semiarid regions, i.e., areas where the annual rain is less than 250 and 250–500 mm/year, respectively, are particularly vulnerable to these changes in climate (e.g., Yatheendradas et al. 2008; Samadi et al. 2013a, b; Etemadi et al. 2014, 2015; Mosaedi et al. 2015). In these regions, where intense thunderstorms are the major characteristic of the climate system, storms typically intensify very rapidly, often leading to short-fused local flash floods. The resulting extreme floods are highly localized, heterogeneous and dominated by environmental gradients (e.g., Volkmann et al. 2010).

Arid and semiarid runoff originates predominantly from partial areas (Osborn 1964; Michaud 1992) and is prone to magnify various errors and uncertainty. In this type of system, epistemic uncertainty, arising from inadequate characterization of the system, must be treated explicitly in order to characterize what is well understood and what is not about the process, scale, and physics of the watershed (S. Samadi et al., "Estimating hydrologic model predictive uncertainty in the presence of complex residual error structures," submitted, *J. Hydrol. Eng.*, ASCE, Reston, Virginia). Nonetheless, the common challenge is that a few sets of processes control model behavior over limited ranges of scale and time. Consequently, uncertainty may rise quickly (e.g., Yatheendradas et al. 2008) and remains in the efficient allocation that typically requires a higher level of model fidelity to characterize uncertainty and spatial variability in hydrological processes.

In recent decades, various mathematical methods have been developed to treat uncertainty in hydrology. Methods to represent model parameter, state, and prediction uncertainty include classical Bayesian (Kuczera and Parent 1998; Thiemann et al. 2001; Tonkin and Doherty 2009; Moore et al. 2010; Vrugt et al. 2003b, 2009a, Schoups and Vrugt 2010; Laloy and Vrugt 2012), set-theoretic (Klepper et al. 1991; Vrugt et al. 2001), sequential data assimilation (Madsen et al. 2003; Vrugt et al. 2005; Moradkhani et al. 2005), multiobjective optimization methods (Vrugt et al. 2003a; SadeghiTabas et al. 2016; Rurui Zhou et al. 2016), and multimodel averaging methods (Georgekakos et al. 2004; Ajami et al. 2007; Vrugt and Robinson 2007). These methodologies differ in their mathematical rigor, the underlying assumptions about the residual error distribution, and the way those assumptions are stated explicitly in the modeling procedure.

Among these methods, formal Bayesian models or Markov chain Monte Carlo (MCMC) samplers are one of the most advanced techniques available in the hydrology modeling community. These algorithms enable calibration of complex multivariate distributions by casting them as the invariant distribution of a Markov chain (McMillan and Clark 2009). MCMC methods specify the functional form of the joint probability density function (PDF) of the residual errors a priori (e.g., Schoups and Vrugt 2010). Then they generate a random walk through the parameter space (Blasone et al. 2008) to improve search efficiency and converge to the posterior PDF in order to derive the appropriate form for the likelihood function (e.g., Box and Tiao 2011). Based on a specific likelihood function the probability that the observed data were generated

[1]Assistant Professor, Dept. of Water Engineering, Faculty of Agriculture, Univ. of Birjand, 97175615 Birjand, Iran (corresponding author). E-mail: Mohsen.pourreza@birjand.ac.ir

[2]Adjunct Faculty, Dept. of Civil and Environmental Engineering, Univ. of South Carolina, Columbia, SC 29208.

[3]Professor, Dept. of Hydrology and Water Resources, Faculty of Water Sciences Engineering, Shahid Chamran Univ., Ahvaz, Iran.

[4]Professor, Dept. of Water Engineering, Faculty of Agriculture, Ferdowsi Univ. of Mashhad, 9177948974 Mashhad, Iran.

Note. This manuscript was submitted on January 19, 2016; approved on September 8, 2016; published online on November 16, 2016. Discussion period open until April 16, 2017; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Hydrologic Engineering***, © ASCE, ISSN 1084-0699.

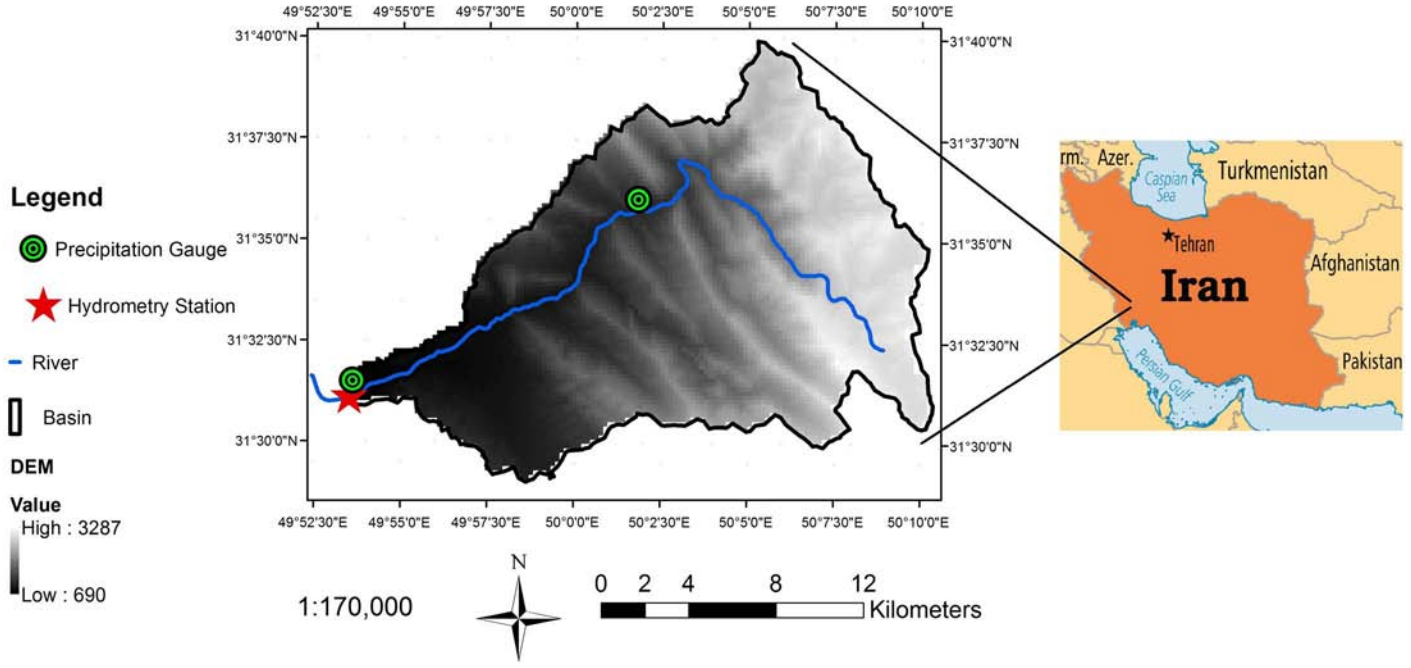

Fig. 1. Location map of the study area

by a particular parameter set is quantified by model parameter inferences. The mapping from parameter space to likelihood space identifies a range of plausible parameter sets and allows estimation of parameter and predictive uncertainty (Schoups and Vrugt 2010). Further discussion and information on formal Bayesian approaches has been given by Mantovan and Todini (2006), Vrugt et al. (2009b), Stedinger et al. (2008), McMillan and Clark (2009), Schoups and Vrugt (2010), and Nourali et al. (2016).

The main goal of this paper is to test the applicability of MCMC Bayesian approaches to adequately simulate flash flood events with varying degrees of magnitudes and duration. Our study presents the first attempt that linked a fully distributed event-based hydrology model, i.e., AFFDEF (Moretti and Montanari 2007), with different MCMC Bayesian algorithms to test the residual error distribution models [i.e., the standard least squares (SLS) versus the generalized log-likelihood function (GL) error functions] that allow for parameter inference, and outline a proper method for predictive simulation. The authors follow an approach similar to Schoups and Vrugt (2010), in that the complexity of Bayesian algorithms is gradually increased, starting from the simplest framework, i.e., the shuffled complex evolution Metropolis (SCEM-UA, Vrugt et al. 2003b), to a more complex approach, i.e., the differential evolution adaptive Metropolis (DREAM; Vrugt et al. 2008) and the DREAM-ZS (Schoups and Vrugt 2010), until satisfactory results are obtained by posterior checks and analysis of predictive uncertainty. The authors further discuss the difficulties and challenges posed by the statistical assumption of residual error distribution on estimation of parameter and predictive uncertainty and the ways in which different properties and factors may interrupt robust simulation in the semiarid flood modeling as our case study.

Case Study and Data

This case study was conducted in the Abol-Abbas watershed with a drainage area of 290 km^2, located in the northern part of the Marun River basin in southwest Iran. The river originates from the Zagros Mountains (from east to west) and discharges into the Persian Gulf. According to the De Martonne formula (De Martonne 1926), this area is classified as a semiarid region with average annual precipitation and daily temperature of 650 mm and 13°C, respectively. Rainfall data of Malagha and Bagh-e-Malek climate stations, located respectively in the central and downstream parts of the watershed, were used to generate a rainfall ASCII file [Fig. 1; the pixel size of the digital elevation model (DEM) is 100 × 100 m and the DEM layer was generated using a contour shapefile (1:25,000)]. Rainfall has been recorded in the Malagha station since 1983 and monitored at a downstream gauge since 1973. There is a hydrometery station (Pole-Manjanigh) in the downstream portion which is represented by red star in Fig. 1 with an average daily discharge of 4.2 m^3/s during 1980–2009 period. Daily flow varies considerably from 1.5 m^3/s during the dry season to 13 m^3/s during the wet season.

Spatial interpolation methods such as the Thiessen polygon approach and inverse distance weighting (IDW) method were used to reproduce the spatial continuity of rainfall fields based on rain gauge measurements. Both methods highlighted the fact that both rainfall gauges produced same values to the flood magnitudes and amount.

Four late fall and winter flood events (December 13, 1979; January 24, 1983; November 15, 1984; and January 31, 1993) were identified as potential dreadful events using measured hyetograph and hydrograph data. Those extreme events caused devastating floods with and high death totals and economic losses in Bagh-e-Malek city from the 1970s to 1990s. Flood events were selected based on date of occurrence, volume and duration, starting time, number of peaks, peak values at the outlet, and soil antecedent moisture conditions (Table 1). Soil moisture conditions [according to Chow et al. (1988)] prior to the first and the third events were dry while the soil in the second and the fourth events was moderately moisturized. The fourth event had the longest intense rainfall (23 h) and flow durations (24 h), while the first event had the least storm duration (6 h) and the third one had the least flow duration (11 h). The fourth flood was a multipeak (three-peak) hydrograph with a maximum peak of 40 m^3/s, while the rest of the events were

Table 1. Four Flash Floods and the Corresponding Rainfall Characteristics

Event	Date of occurrence	Flow starting time	Flow duration (h)	Number of peak	Rainfall (mm)	Rainfall duration (h)	Observed peak flow (m^3/s)	Soil antecedent moisture condition
1	December 13, 1979	4:00 a.m.	12	1	20	6	14.26	Dry
2	January 24, 1983	9:00 a.m.	13	1	43	8	78	Average
3	November 15, 1984	11:00 a.m.	11	1	37	7	25.5	Dry
4	January 31, 1993	1:00 a.m.	24	3	74	23	6, 15.8, 40 (multipeak)	Average

single-peak events. The main properties of the selected flood events are summarized in Table 1.

Methodology

MCMC Sampling Method

This study used three different MCMC algorithms for uncertainty assessment, which are briefly explained as follows:

SCEM-UA algorithm: Latin hypercube sampling (LHS) is used in the SCEM-UA method to produce initial population points (a number of parameter set) that are randomly distributed throughout the feasible parameter space. Posterior densities for each parameter set are then calculated using a Bayesian inference scheme (Box and Tiao 2011). The population is partitioned into q complexes, and in each complex k ($k = 1,2, \ldots, q$), a chain is assigned from a point that exhibits the highest posterior density. The Metropolis criterion is then used to test if the candidate point can be accepted as a new parameter set. Each sequence is evolved with the sequence evolution Metropolis (SEM) algorithm and the process continues many times to converge to a posterior distribution. A full, detailed description and discussion of SCEM-UA was given by Vrugt et al. (2003b).

DREAM algorithm: The DREAM sampling scheme is an adaptation of the SCEM-UA global optimization algorithm that uses a differential evolution as a genetic algorithm for population evolution. Within this technique a preset number (N) of Markov chains (a chain refers to a vector containing one set of parameter realizations) are simultaneously run in parallel. The chains are initialized by Latin hypercube sampling and the parameter space using uniform distribution. These chains form a population, conveniently stored as $N \times d$ matrix X, with d as the dimension of the parameter space. For each chain, $i \in \{1,2, \ldots, N\}$, a candidate point z_i (vector) is generated by taking a fixed multiple of the differences between randomly chosen pairs of chains (without replacement) of X_i (X without x_i) as shown in Eq. (1) (Vrugt et al. 2009a)

$$z_i = x_i + (1+e)\gamma(\delta, d_{\text{eff}})\left[\sum_{j=1}^{\delta} x_{r_1(j)} - \sum_{n=1}^{\delta} x_{r_2(n)}\right] + \varepsilon \qquad (1)$$

where δ = number of pairs used to generate the proposal; γ = jump rate; and $r_1(j)$, $r_2(n) \in \{1,2, \ldots, N\}$ but $r_1(j) \neq r_2(n) \neq i$. The value of e is drawn from U_d $(-b, b)$ with $|b| < 1$, and $\varepsilon \sim N_d(0, b^*)$ is a white noise term with b^* small compared to the width of the target distribution. The Metropolis ratio is used to decide if the candidate point can be acceptable as a new parameter set. If accepted, the chain moves from x_i to z_i; otherwise the location of the chain remains unchanged. From the guidelines of γ in random walk Metropolis (RWM), a good choice of $\gamma = 2.38/\sqrt{2\delta.d_{\text{eff}}}$ will be updated, where d_{eff} denotes the number of dimensions. Using this approach, a MC and a stationary distribution (or a posterior distribution) are obtained. After a so-called burn-in period, the convergence of a DREAM run can be monitored with the $\hat{R}$ statistic (Gelman and Rubin 1992), which determines convergence to the stationary posterior distribution. This statistic compares the between and within variance of different parallel chains. A value of $\hat{R}$ smaller than 1.2 for each parameter ($\hat{R}_k < 1.2$, $k = 1, 2, \ldots, d$) diagnoses convergence to a limiting distribution. The samples generated after convergence can be used to summarize the posterior distribution and to communicate parameter and predictive uncertainty. A number of steps in each chain are required to reach stationary (convergence), commonly called *burn-in*, and these samples are removed from the analysis (Dekker et al. 2012). Readers are referred to Vrugt et al. (2009a) for more detailed information on the DREAM algorithm.

DREAM-ZS algorithm: DREAM-ZS, proposed by Schoups and Vrugt (2010), is based on the original DREAM algorithm (Vrugt et al. 2009a). This newest approach uses sampling from an archive of past states to generate candidate points in each individual chain and requires only three parallel chains to summarize the posterior distribution. DREAM-ZS does not require outlier detection and removal; it maintains a detailed balance at every single step in each of the parallel chains (Schoups and Vrugt 2010). In addition, DREAM-ZS contains a snooker update to generate jumps beyond parallel direction updates (ter Braak and Vrugt 2008) in order to increase the diversity of the candidate points. The snooker axis runs through the states of two different chains, and the orientation of this jump is different from the parallel direction update, utilized in DREAM (Laloy and Vrugt 2012). The algorithmic implementation of the snooker update within the context of the differential evolution Markov chain (DE-MC) method was proposed by ter Braak (2006). More discussion on this algorithm is provided by Schoups and Vrugt (2010) and Laloy and Vrugt (2012).

A linkage between MCMC and AFFDEF is illustrated in Fig. 2. Briefly, the Abol-Abbas flash flood simulation is achieved using a modified version of the curve number (CN) method (Moretii and Montanari 2007) distinguishing differences between surface runoff and infiltration. AFFDEF then predicted runoff individually for each grid cell and routed flow in the channel to obtain flood volume. The output of the AFFDEF model was coupled with the MCMC samplers in order to compute parameter and predictive uncertainty of semiarid flood events.

Likelihood Function

Both DREAM and SCEM-UA assume that the residual errors from the predictions $[e(\theta^t) = \hat{y}(\theta^t) - y]$ are mutually independent and follow a Gaussian distribution with constant variance (e.g., Vrugt et al. 2008, 2009a, b). The likelihood of a parameter set θ^t for describing the observed data y can then be computed using the Bayesian inference method proposed by Box and Tiao (2011) [Eq. (2)]

$$L(\theta^t|\mathbf{y}) = \exp\left\{-\frac{1}{2}\sum_{i=1}^{N}\left[\frac{e(\theta^{(t)})_i}{\sigma}\right]^2\right\} \qquad (2)$$

where $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$ is a vector of n unknown parameters and e = vector of statistically independent measurement errors with zero expectation and constant variance σ^2, where y is an $N \times 1$ vector of the observed data. Assuming a noninformative

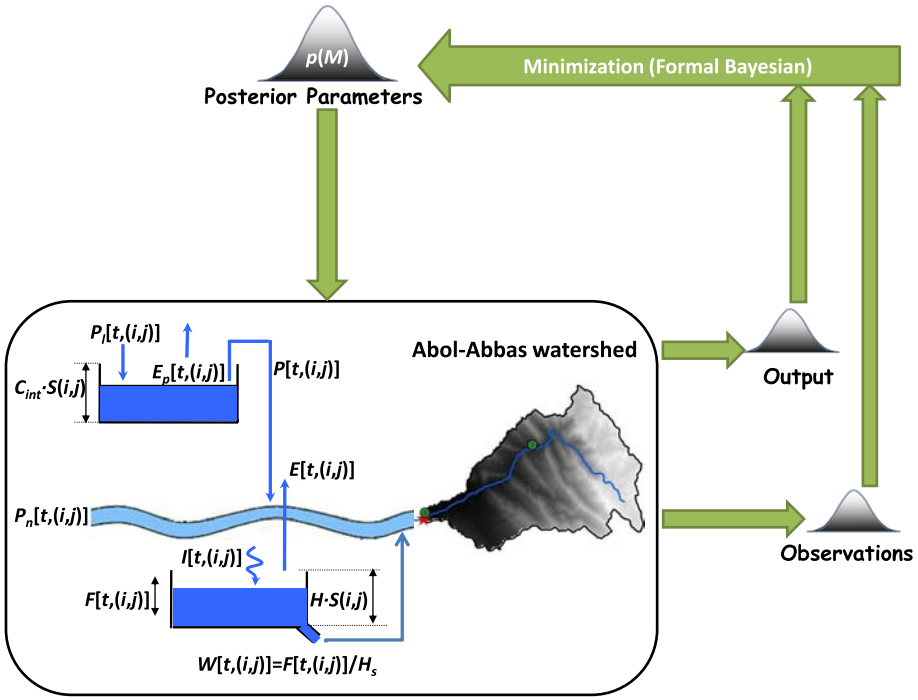

Fig. 2. MCMC algorithms linked to AFFDEF (data from Moretti and Montanari 2007)

prior of the form $p(\theta) \propto \sigma^{-1}$, Box and Tiao (2011) showed that the influence of σ can be integrated out, leading to the following form of the posterior density of $\theta^{(t)}$:

$$p(\theta^{(t)}|\mathbf{y}) \propto [M(\theta^{(t)})]^{-N/2} \tag{3}$$

where

$$M(\theta^{(t)}) = \sum_{i=1}^{N} e(\theta^{(t)})_i^2 \tag{4}$$

M is simply the familiar sum of the squared measurement error that is commonly used as the likelihood function in hydrology. For the application of Bayesian inference, one may adopt the approach of Box and Tiao (2011) or Thiemann et al. (2001). For instance, DREAM-ZS uses the generalized formal likelihood function [Eq. (5)], which embraces temporal correlation, heteroscedasticity, and the non-Gaussian nature of model residuals, while the standard least-squares [Eq. (4)] objective function is used in the DREAM model

$$L(\eta|Y) = n \log \frac{2\sigma_\xi \omega_\beta}{\xi + \xi^{-1}} - \sum_{t=1}^{n} \log \sigma_t - c_\beta \sum_{t=1}^{n} |a_t|^{2/(1+\beta)} \tag{5}$$

where η = parameters associated with the hydrology model and the residual error model $(\sigma_\xi, c_\beta, \omega_\beta, \mu_t)$, which are computed as a function of skewness ξ and kurtosis β parameters, as detailed by Schoups and Vrugt (2010); Y = observation; $a_{\xi,t} = \xi^{-\text{sign}(\mu_\xi + \sigma_\xi a_t)}(\mu_\xi + \sigma_\xi a_t)$; and a_t denotes a random error with zero mean and unit standard deviation, described by a skew exponential power (SEP) density with parameters ξ and β to account for nonnormality.

Model Evaluation

In this research, two indices were used to quantify the goodness of calibration/uncertainty performance: a P-factor, which is the percentage of data bracketed by a 95% prediction uncertainty band (95PPU; the maximum value is 100%), and an R-factor [or d-factor; Eqs. (7) and (8)], which is the average width of the uncertainty band divided by the standard deviation of the corresponding measured variable [the minimum value is zero (Abbaspour et al. 2004, 2007)]. Theoretically, the value for the P-factor ranges between 0 and 100%, while the R-factor ranges between 0 and infinity. A P-factor of 1 and R-factor of 0 is a simulation that exactly corresponded to measured data. The average thickness of the 95PPU band or the R-factor are estimated in every run and the best simulation can be judged based on the simulation with almost observed and modeled data located inside the 95% band with the highest P-factor and the least R-factor

$$\bar{d}_x = \frac{1}{k}\sum_{i=1}^{k}(X_U - X_L) \tag{7}$$

$$R\text{-factor} = \frac{\bar{d}_x}{\sigma_x} \tag{8}$$

where $\bar{d}_x$ = average distance between the upper and the lower 95PPU; X_U and X_L = upper and the lower boundaries of 95PPU; and σx = standard deviation of the measured data. Both the P-factor and R-factor were calculated based on flow rates that are greater than 50% of the peak values since the emphasis of this study is on flash flood forecasting. As all uncertainties in the conceptual model and inputs are reflected in the measurement, bracketing most of the measured data in the predictive 95PPU ensures that uncertainty is depicted by the parameter uncertainty.

Because of the inconsistent behavior of semiarid flood models, the total uncertainty index (TUI) is also calculated based on the P-factor and R-factor for each flood event [Eq. (9)]

$$\text{TUI} = \frac{\text{P}_{\text{Factor}}}{\text{R}_{\text{Factor}}} \tag{9}$$

Table 2. AFFDEF Parameters Used in This Study

Parameter's number	Parameter	Dimension and symbol	Estimation method	Estimated values	Lower bound	Upper bound
1	Contributing area threshold	A_0 (km^2)	Calibrated	—	0.3	1.5
2	Channel width/height ratio for the hillslope	w_v (-)	Calibrated	—	300	100,000
3	Strickler coefficients for the first class of roughness on the hillslope	$k_{sv}(1)$ (m$^{1/3}$ s^{-1})	Calibrated	—	0.05	10
4	Strickler coefficients for the second class of roughness on the hillslope	$k_{sv}(2)$ (m$^{1/3}$ s^{-1})	Calibrated	—	0.05	10
5	Strickler coefficients for the third class of roughness on the hillslope	$k_{sv}(3)$ (m$^{1/3}$ s^{-1})	Calibrated	—	0.5	20
6	Saturated hydraulic conductivity	K_{sat} (ms^{-1})	Calibrated	—	0.001	0.1
7	Bottom discharge parameter for the infiltration reservoir capacity	H_S (s)	Calibrated	—	20,000	800,000
8	Multiplying parameter for the infiltration reservoir capacity	H (-)	Calibrated	—	0.05	0.9
9	Multiplying parameter for the interception reservoir capacity	C_{int} (-)	Calibrated	—	0.05	0.7
10	Channel width/height ratio for the channel network	w_r (-)	Estimated	15	—	—
11	Maximum Strickler roughness for the channel network	k^0_{sr} (m$^{1/3}$ s^{-1})	Estimated	40	—	—
12	Minimum Strickler roughness for the channel network	k^1_{sr} (m$^{1/3}$ s^{-1})	Estimated	12	—	—
13	Width of the rectangular cross section of the subsurface water flow	B_p (m)	Estimated	0.5	—	—

Increasing TUI leads to performance improvement. For estimating the total uncertainty (i.e., parameter and all other sources), a constant Gaussian error in terms of the model residuals (computed by subtracting the predicted hydrograph with the highest posterior probability) was added to the model prediction (Feyen et al. 2007). This common approach should only be employed when the residuals are normally distributed and follow a normal distribution (Schoups and Vrugt 2010).

Rainfall-Runoff Model

The AFFDEF model is a spatially distributed raster-based rainfall-runoff model developed at the University of Bologna, Italy. This model discretizes the basin in square cells coinciding with the pixels of the digital elevation model, which describes the basin topography. By applying the D8 method (flow direction method), the river network can be extracted from the DEM itself (Band 1986) to estimate flow paths. As a grid-based model, each cell receives water from its upslope neighbor and discharges to the downslope neighbor. Distinction between hill slope rill and network channel cells is based on the concept of constant critical support areas proposed by Montgomery and Foufoula-Georgiou (1993). Accordingly, rill flow is assumed to occur in a cell where the upstream drainage area does not exceed the value of the critical support area (contributing area threshold) A_0, while channel flow occurs in a cell area with a higher A_0 value.

In the AFFDEE model, the interaction among soil, vegetation, and atmosphere is modeled using a modified CN approach in each cell (Soil Conservation Service 1972). This approach allows the user to compute infiltration and surface runoff continuously over time (Brath et al. 2003). Effective evapotranspiration is computed at a local scale using a radiation-based method (Doorenbos et al. 1984). The actual amount of loss from the soil-plant continuum is calculated by considering an actively growing plant or crop during the simulation procedure. However, in short-duration forecasts, ET variability has no significant influence on the simulation. The Muskingum-Cunge method is then used to route downstream surface and subsurface flows.

AFFDEF can be used for any kind of basin, but it is best suited for an area where the runoff follows the infiltration excess mechanism. Infiltration excess overland flow (or Horton flow), occurs when water enters a soil system faster than the soil can absorb or move it, such as when precipitation exceeds the infiltration capacity of the soil. This process often dominates in arid regions (Bloschl et al. 2013) where significant surface sealing occurs during rainfall events. In this study, the AFFDEF source code was modified by including the subsurface module (similar to a continuous simulation model) in order to adjust for semiarid rainfall-runoff processes. The modified version was efficiently capable of resolving subsurface impediments in a short-term forecast and it decreased bias and error by increasing the number of parameters involved in the simulation.

The upper and lower bounds that define the prior uncertainty ranges of the AFFDEF parameters are listed in Table 2. For example, saturated hydraulic conductivity ranges from 0.001 to 0.1 m/s while the Strickler coefficient for the first class of roughness on the hill slope varies between 0.05 and 10. Overall, 13 parameters were included in the semiarid modeling procedure (Table 2). The values of some parameters (Parameters 1–9) were considered for calibration while the rest of the low-sensitivity parameters (Parameters 10–13) were determined by either physical reasoning or in situ measurement. Calibration parameters and their absolute values were optimized during the uncertainty process by consuming a time of 20–30 h on a machine with a CPU intel 2.2-core i7 and 8 GB memory.

Results

Convergence of the MCMC Samplers

Figs. 3–6 illustrated the evolution of the R statistic (Gelman and Rubin 1992) for each individual model parameter. Most of the parameters in the first, third, and fourth events (Figs. 3, 5, and 6) met the convergence criteria after approximately 14,000 runs, while the second event encountered the criteria after a large number of runs (~45,000). In this study, both DREAM and SCEM-UA sampling algorithms generated 15 chains ($q = 15$), each with 3,000 iterations except for the second event, in which 5,000 evaluation runs were achieved (in total, 75,000 evaluation runs for the second event and 45,000 runs for the rest of the events). Also, 25,000 evaluation numbers were achieved for each of the three chains in the DREAM-ZS algorithm.

Based on the R statistic, the SCEM-UA sampler never met the convergence criterion especially for the first and fourth events, so its result proved to be far from convergence (Figs. 3 and 6). In other words, the convergence diagnostic criteria never reached a stationary posterior distribution in SCEM-UA even after performing 45,000 runs. DREAM-ZS, at the same time, converged to a very different posterior parameter distribution. Our results presented herein demonstrated substantial differences between the marginal parameter distributions inferred by DREAM-ZS and the rest of the algorithms (see the next section). Comparably, DREAM convergence was quicker than in the other models. DREAM

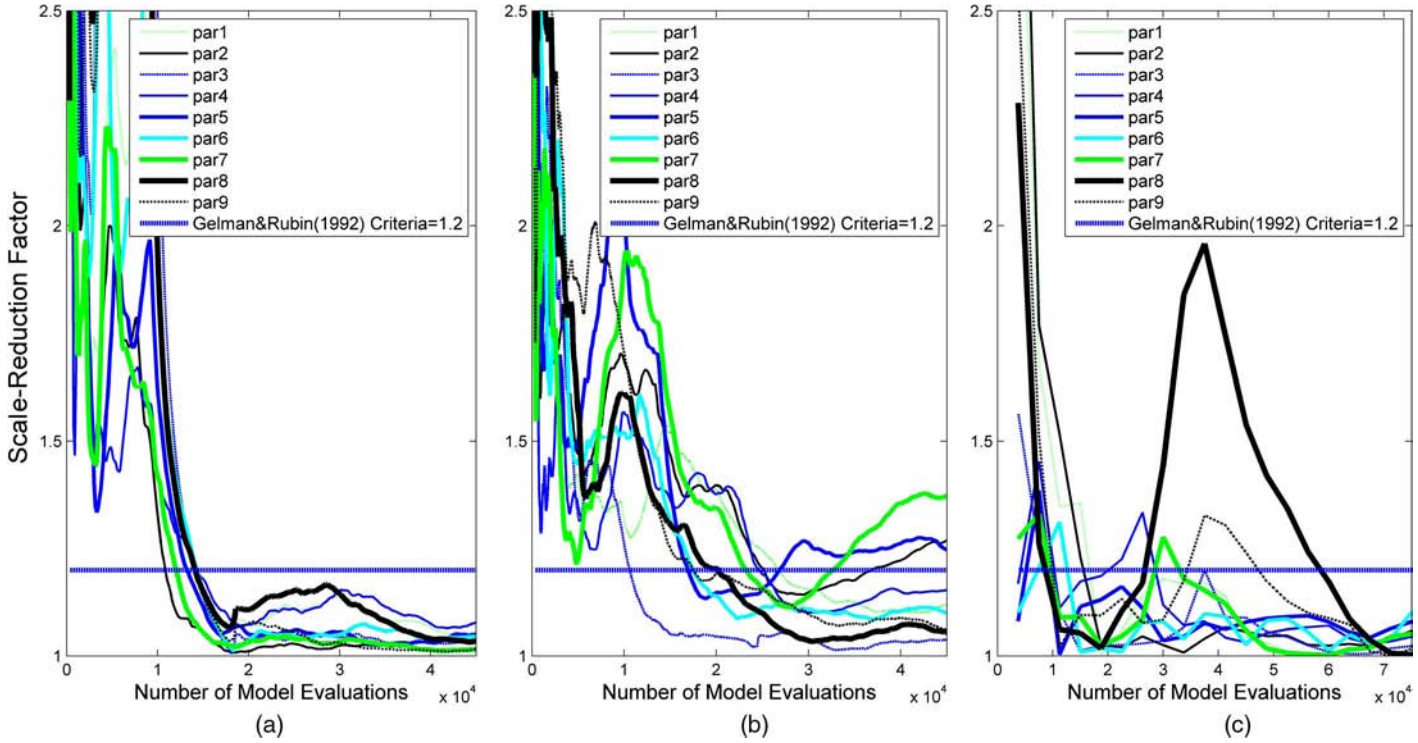

Fig. 3. Evolution of the Gelman and Rubin scale-reduction factor of the AFFDEF parameters using (a) DREAM; (b) SCEM-UA; and (c) DREAM-ZS for the first event

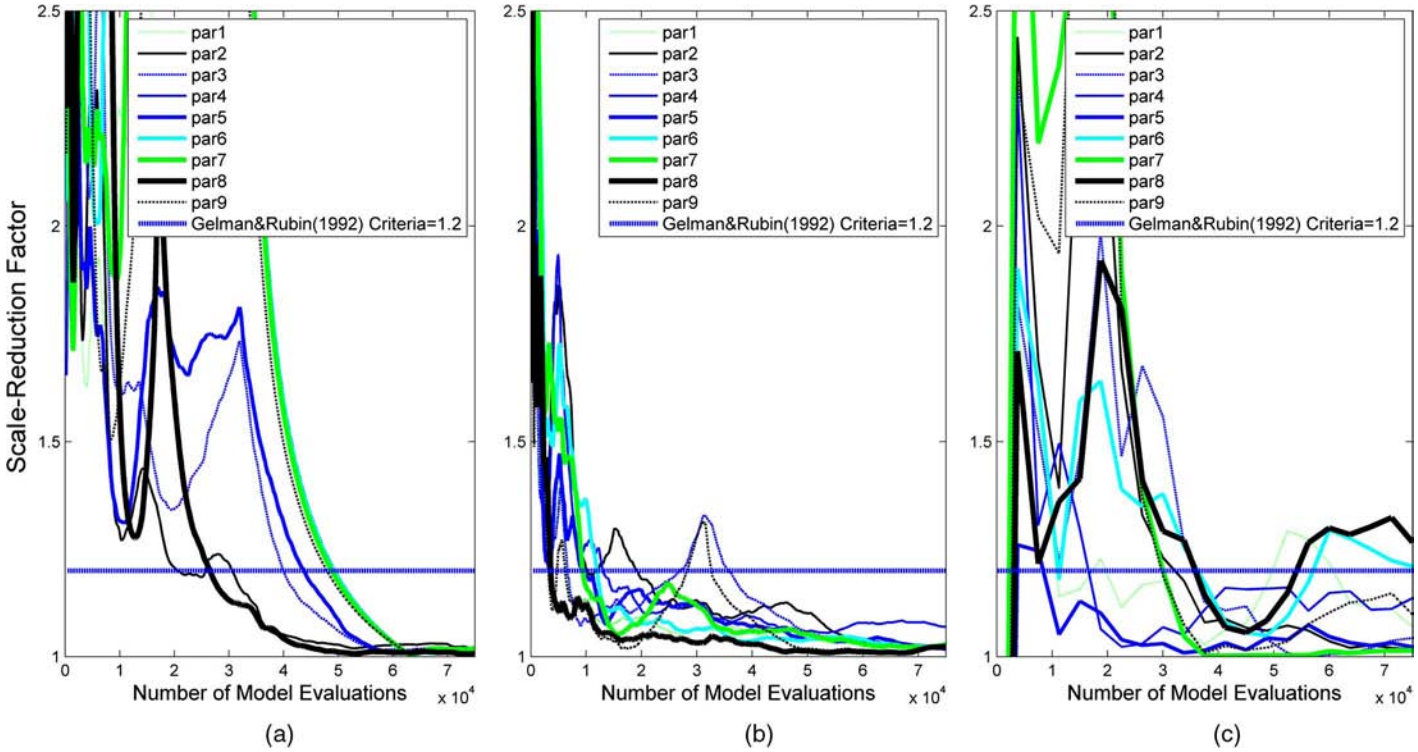

Fig. 4. Evolution of the Gelman and Rubin scale-reduction factor of the AFFDEF parameters using (a) DREAM; (b) SCEM-UA; and (c) DREAM-ZS for the second event

was more efficient at traversing the parameter space by quickly achieving convergence at a stationary posterior distribution ($\hat{R} < 1.2$). For instance, DREAM convergence for each of the AFFDEF parameters was achieved after approximately 14,000 iterations in the first event and up to 45,000 iterations for the second one.

Overall, modeling efforts proved that DREAM is most efficient in generating posterior samples, because it required the least

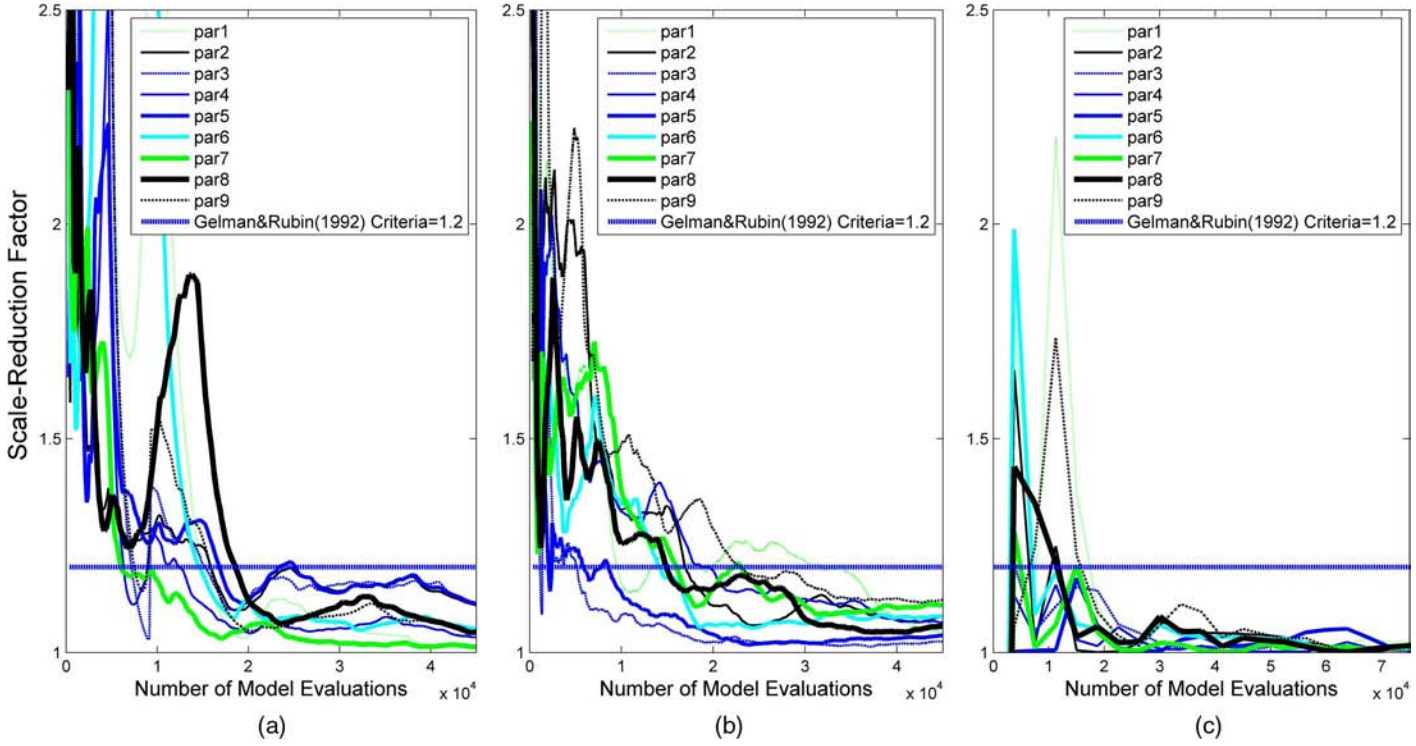

Fig. 5. Evolution of the Gelman and Rubin scale-reduction factor of the AFFDEF parameters using (a) DREAM; (b) SCEM-UA; and (c) DREAM-ZS for the third event

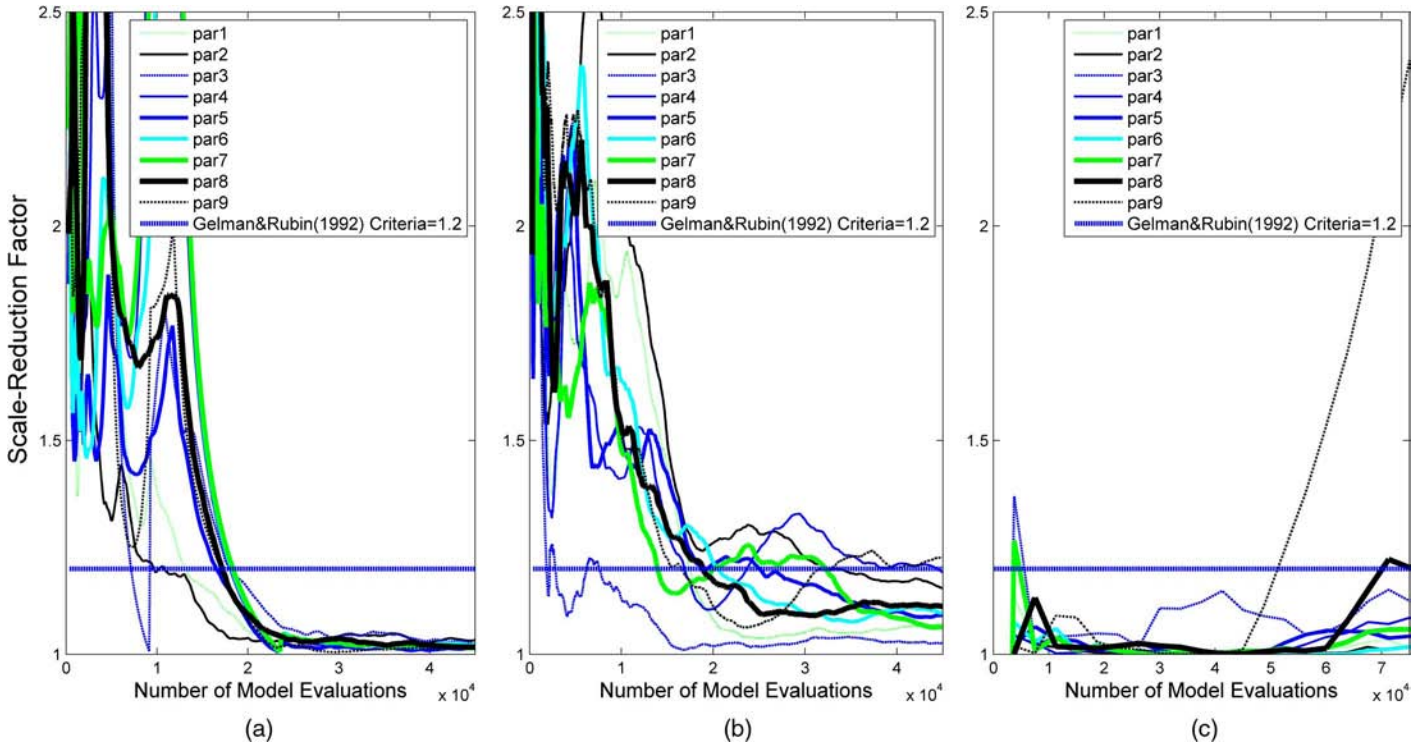

Fig. 6. Evolution of the Gelman and Rubin scale-reduction factor of the AFFDEF parameters using (a) DREAM; (b) SCEM-UA; and (c) DREAM-ZS for the fourth event

amount of computational effort and time to explore the posterior and was able to cope with difficult response surfaces as well. The result demonstrated by the DREAM algorithm not only converged most rapidly compared to the rest of the algorithms, but also provided samples that most closely represented the underlying target distribution, especially for the first and third events. Around 80% of the samples generated by DREAM were thrown away, and 20% of the last simulation was used to summarize the posterior

Table 3. Modeling Performances for Four Flood Events

Method	Criteria	Parameter uncertainty				Total uncertainty			
		Event 1	Event 2	Event 3	Event 4	Event 1	Event 2	Event 3	Event 4
DREAM	*R*-factor	0.75	0.11	0.45	0.26	1.64	1.72	2.61	2.26
	P-factor %	50	0	20	0	83.33	75	80	100
	TUI	67.02	0.00	44.30	0.00	50.88	43.71	30.63	44.27
SCEM-UA	*R*-factor	1.02	2.23	1.36	1.76	1.86	3.29	3.00	3.62
	P-factor %	50	50	40	20	83.33	75	80	100
	TUI	48.85	22.37	29.43	11.36	44.72	22.80	26.69	27.64
DREAM-ZS	*R*-factor	2.77	15.55	7.56	17.65	4.84	23.25	10.05	20.64
	P-factor %	66.67	0	0	40	100	75	20	100
	TUI	24.05	0.00	0.00	2.27	20.66	3.23	1.99	4.85

distribution. Although this study showed the proficiency of the DREAM algorithm in this study, the authors stress that the performance of uncertainty algorithms may be case/data dependent, especially in the case of semiarid simulation.

Predictive Uncertainty

A summary comparison among MCMC samplers along with their goodness of fit is provided in Table 3. SCEM-UA captured parameter uncertainty with higher performances (high *P*-factor value), but this model never met the convergence criteria; thus its samples could not represent the posterior probability distribution. Comparably, the *R*-factor and TUI values showed better results for the DREAM algorithm. In addition, the best performance was obtained when the rainfall amount and duration were higher than 20 mm and 6 h, correspondingly.

However, all floods occurred during late fall and winter, but modeling results revealed that the uncertainty associated with lower-magnitude rainfall was smaller than higher magnitudes. Among three MCMC algorithms, DREAM and somehow SCEM (for the first event) performed best. Total uncertainty analysis suggests that the first flood event was more skillfully predicted by DREAM (*R*-factor = 1.64, *P*-factor = 83%, and TUI = 50.8) and also SCEM (*R*-factor = 1.86, *P*-factor = 83%, and TUI = 44.7), while the fourth event (multipeak event) posed significant challenges for all algorithms.

Figs. 7–9 exhibited the observed data (gray dots), a 95% hydrograph predictive uncertainty band as a result of the posterior distribution of the parameter estimates (shaded area), and a 95% hydrograph predictive uncertainty associated with the total uncertainty (light shaded area) in the DREAM, the SCEM-UA, and the DREAM-ZS samplers (the *x*-axis represents time from 00:00 a.m. to 23:00 p.m.).

Clearly, there is good agreement between the estimates of forecast uncertainty derived by DREAM rather than by DREAM-ZS and SCEM-UA. The uncertainty ranges obtained by DREAM were slightly smaller and revealed better performances at various events. Among all algorithms, the 95% predictive uncertainty of DREAM-ZS varied largely among individual events. For instance, the 95% predictive uncertainty of the fourth event is wider and it seems that calibration was not particularly successful. Also, time to peak (T_p) was well predicted by DREAM and SCEM-UA, while DREAM-ZS presented a shorter T_p. DREAM-ZS, on the other

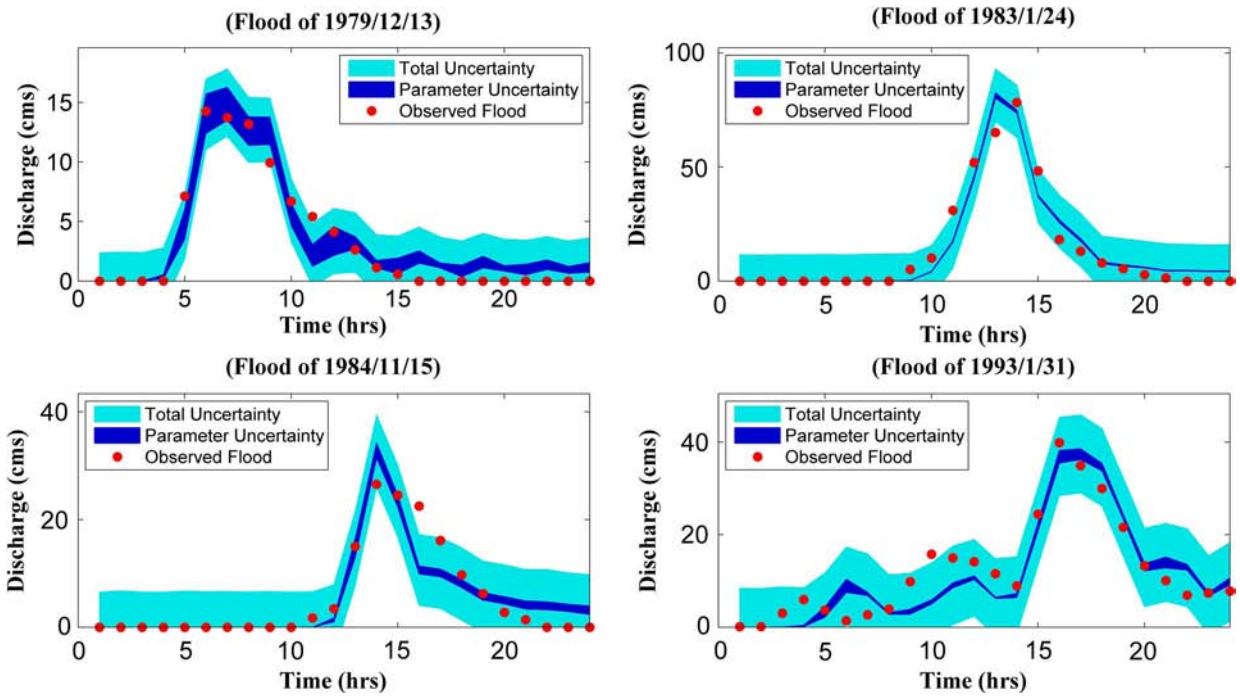

Fig. 7. 95% prediction uncertainty ranges of four flood events in the DREAM algorithm

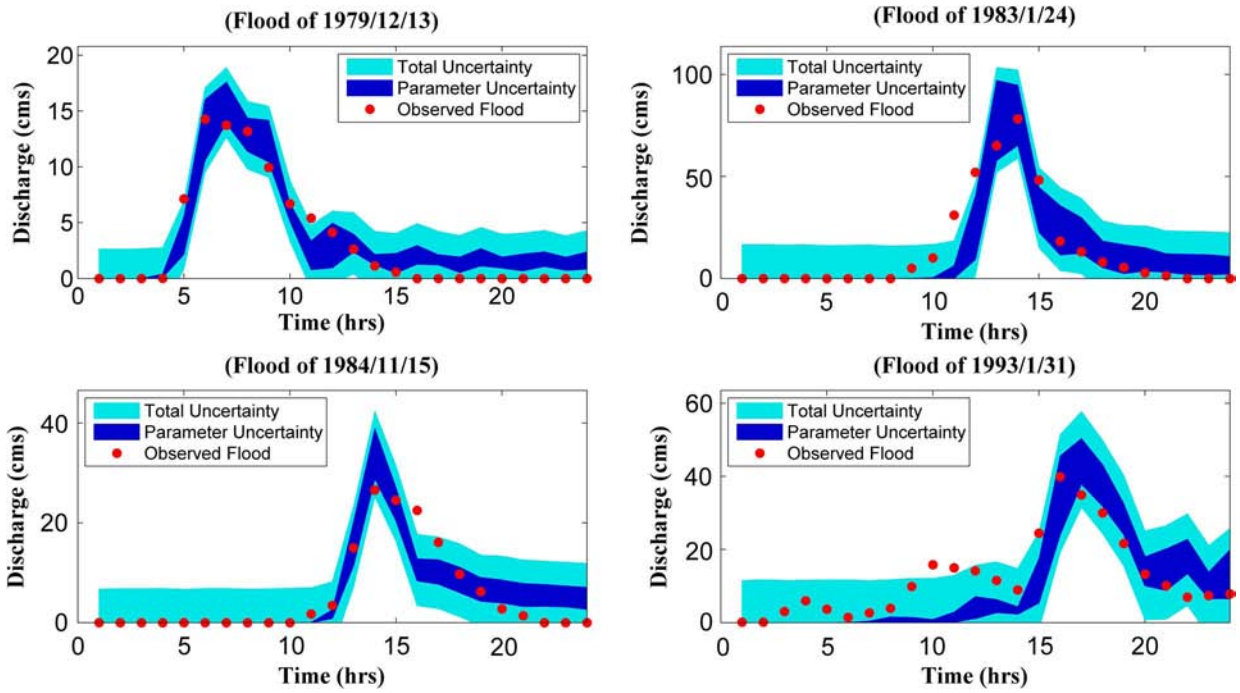

Fig. 8. 95% prediction uncertainty ranges of four flood events in SCEM-UA

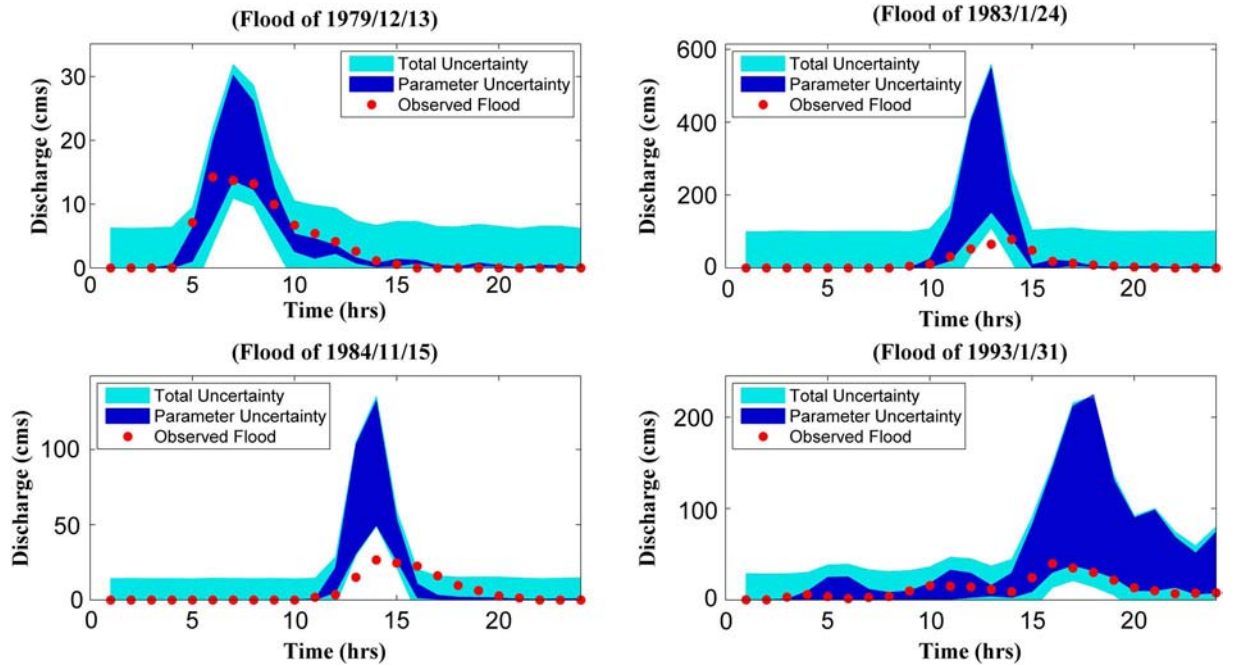

Fig. 9. 95% prediction uncertainty ranges of four flood events in the DREAM-ZS algorithm

hand, resulted in more peaked hydrographs with heavier tails compared to others.

In addition, during the winter flood event of January 24, 1983, error and uncertainty of rainfall-runoff processes contributed less weight to the total uncertainty. The flood event of December 15, 1984, showed the same behavior as the January 24, 1983 event. Further, for the extreme event on December 13, 1979, a wide parameter uncertainty bound was estimated, especially during peak flow. In results from DREAM, when the flood reached 20–30 m^3/s, the parameter uncertainty grew larger and then diminished when the flow rose over 50 m^3/s. In practice, the parameter uncertainty is maximized during a flow of 20–50 m^3/s and minimized when the flood is less than 20 m^3/s and greater than 50 m^3/s.

Model evaluation further revealed the proficiency of the DREAM model in total parameter uncertainty assessment. For instance, DREAM skillfully predicted the flood characteristics (peak, volume, and shape) of the first event, resulting in a strong relationship between simulated flood and observation [i.e., Nash-Sutcliffe efficiency $(NSE) > 0.84$ and root mean square error

Table 4. Performance Criteria for Flash Flood Events

Method	Event	NSE	RMSE
DREAM	Event 1	0.93	0.25
	Event 2	0.93	1.21
	Event 3	0.84	0.69
	Event 4	0.84	0.88
SCEM-UA	Event 1	0.92	0.27
	Event 2	0.86	1.70
	Event 3	0.83	0.71
	Event 4	0.69	1.21
DREAM-ZS	Event 1	0.56	0.65
	Event 2	−4.08	10.42
	Event 3	0.28	1.48
	Event 4	−0.89	2.99

(RMSE) < 1.21; Table 4]. However, DREAM and the other algorithms experienced significant challenges when simulating a multipeak flood event. This may reflect the heterogeneity of the semiarid hydroclimate system, which is often poorly characterized by direct measurement (e.g., Kirchner 2006), depending upon the nature of the storm.

Parameter Uncertainty and Posterior Distribution

The violin plots of the marginal posterior distribution of parameter ensemble widths are illustrated for nine AFFDEF parameters (Fig. 10), optimized by the DREAM algorithm. A violin plot is a box plot combined with kernel density plots, added on each side of the box plot to exhibit the probability distribution of the data set. These plots were created using the last 20% of the samples

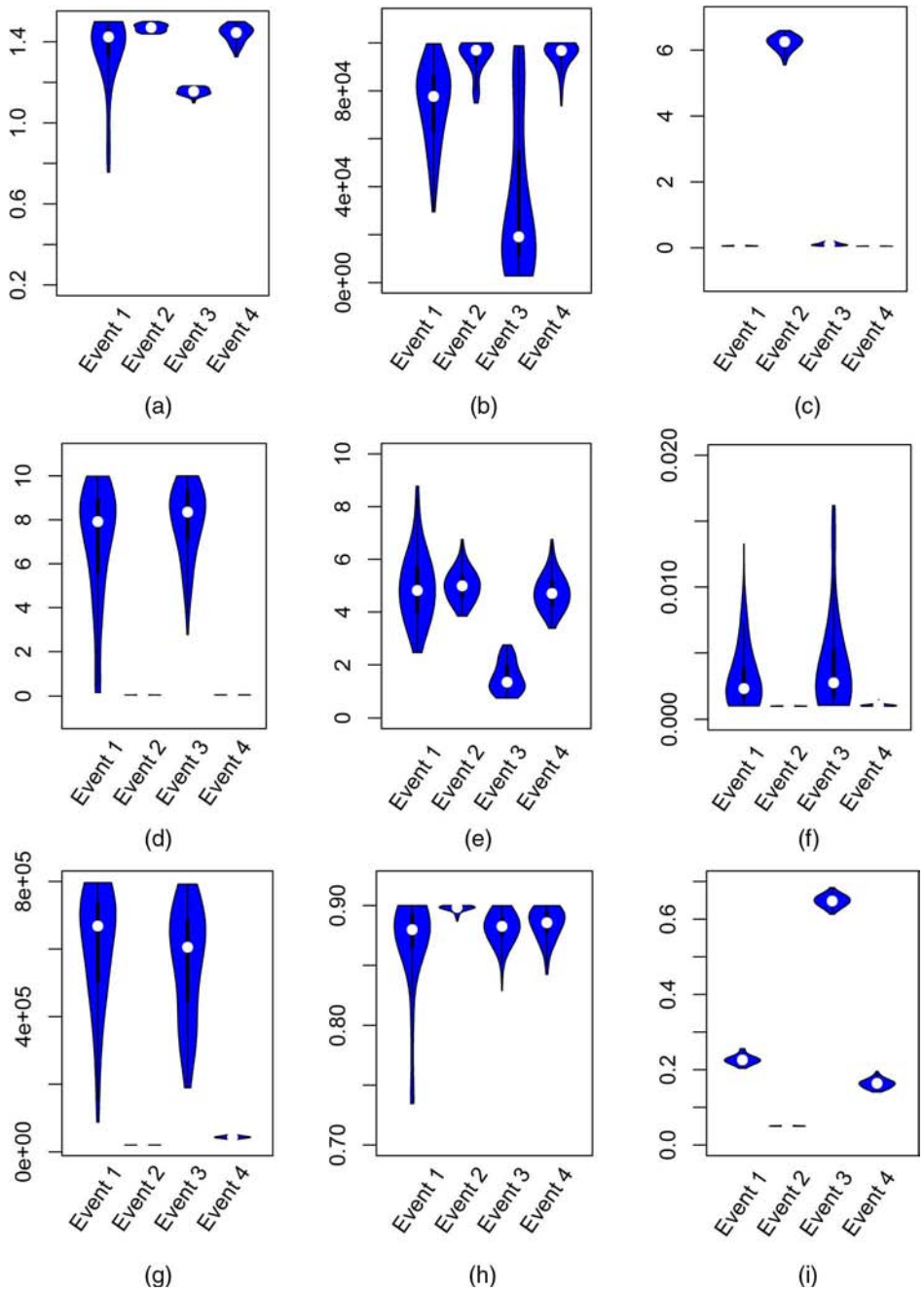

Fig. 10. Violin plots of the marginal posterior distributions of DREAM algorithm for four flood events

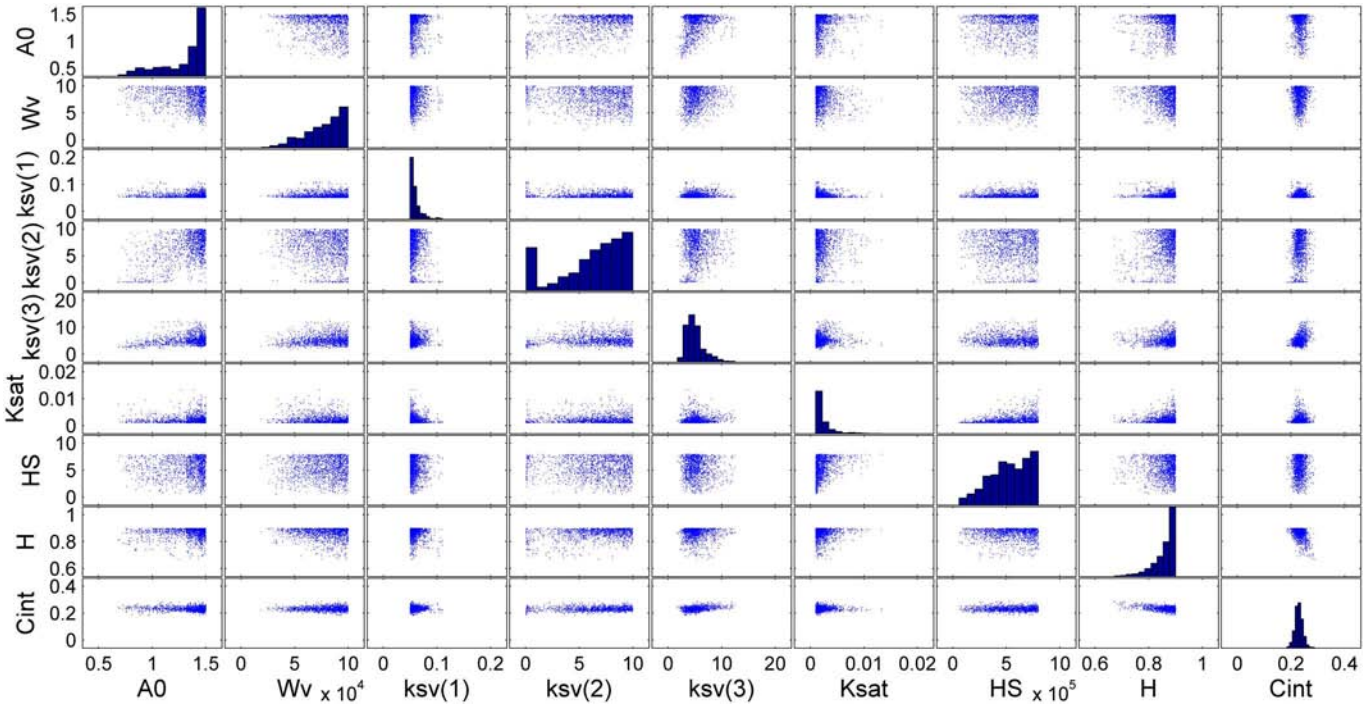

Fig. 11. Marginal distributions and two-dimensional correlation plot of posterior parameter samples of the DREAM model for the first event

generated by the models. In this figure, the thick black line and white dot show the 25th and 75th percentile ranges and median, respectively, and the thin black line shows the 5th and 95th percentile ranges of parameter space. Letters A–I represent Parameters 1–9 (Table 2).

DREAM computed approximately similar median values for most of the parameters, particularly for the second and fourth events. Moving from the first to the ninth parameters, more dispersion and skewness is noted in the median values of the ninth parameter (i.e., multiplying the parameter for the interception reservoir capacity). In addition, DREAM skillfully optimized all parameters (except for C_{int} and $K_{SV(1)}$) for the second and the fourth events, so that their posterior parameters have similar ranges. This might be related to having a similar soil antecedent moisture class (AMC) for these two events, since AFFDEF showed more sensitivity to initial soil conditions and properties. This might also reflect the fact that simulated soil water storage capacity of the hydrology model was inadequate to represent the upper soil zone processes.

The 25th and 75th percentile (thick black lines in Fig. 10) ranges of posterior distribution are effectively narrowed by DREAM, particularly for the second and fourth events. These two events had average soil moisture conditions and apparently their soil moisture conditions play an important role in the forecast scheme. In most cases, a low flood event (i.e., the first event) showed a more biased and more skewed probability by generating larger percentiles of posterior parameter ranges. It can be argued that the first and fourth events represent, respectively, the lowest and highest uncertainties in posterior parameter distribution. Perhaps low to moderate flood rates propagate a low degree of parameter uncertainty in modeling, whereas a high flood event (>40 m^3/s) contributes high bias and uncertainty in modeling.

For further investigation, marginal distributions and a two-dimensional correlation matrix of the posterior parameter samples are graphically illustrated in Figs. 11–14. The correlation matrix of posterior parameter samples of DREAM as the best algorithm indicated that the allocations of some parameters either arrived to a similar distribution or were at least very close to one another. For example, the fifth parameter (Strickler coefficient for Class 3 on the hill slope) shows approximately a normal distribution with a value of the mode that appears more physically reasonable. Also the first parameter (contributing area threshold) occupied only a relatively small region interior to the uniform prior distribution. It can be inferred that the observations contain sufficient information to estimate these parameters.

Arguably, histograms of the eighth and the ninth parameters show a normal distribution especially for the second and the fourth events. In addition, multiplying the parameter for the infiltration reservoir capacity seems to steadily sustain the value around 0.9 for all events, whereas the ninth parameter varies in a range from 0.05 to 0.7. The histogram of the bottom discharge parameter for the infiltration reservoir capacity differs from the normal distribution for all events and tends to concentrate most of the probability mass at the upper bounds.

Based on these results, there appears to be a nonlinear relationship between coefficients for the third class of roughness on the hillslope and the width/height ratio for the hillslope during the third flood event. The relatively small values for $k_{sv}(3)$ (1.3–1.4 $m^{1/3}\,s^{-1}$) and w_v (950) indicated that the model is very nonlinear around these optimal values. However, the posterior distribution of $k_{sv}(3)$ is very close to the normal distribution, indicating a well-estimated value for this parameter during the peak flow of 25 m^3/s. Arguably, when the channel area and depth are reduced (reduction in w_v), Manning's n-value increases in the formulation, which causes diminishing in the Strickler value. Parameter robustness was also quantified for all events by computing the variance of the posterior mean divided by total variance of the MCMC posterior samples (CV = SD/MV, where SD = standard deviation and MV = mean value). For each parameter, this ratio was smaller for the DREAM algorithm compared to others, so the authors simply presented the DREAM results in Table 5.

Based on Table 5, multiplying the parameter for the infiltration reservoir capacity (C_{int}), contributing to the area threshold, and multiplying the parameter for the interception reservoir capacity

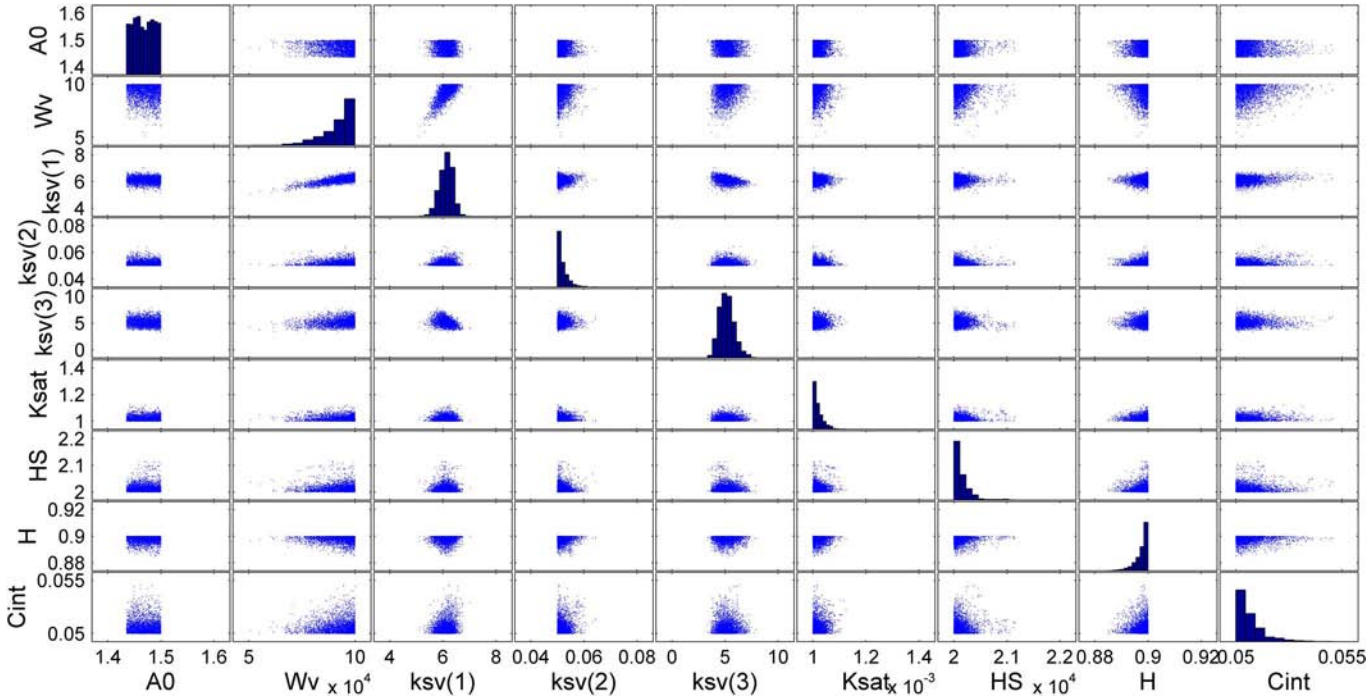

Fig. 12. Marginal distributions and two-dimensional correlation plot of posterior parameter samples of the DREAM model for the second event

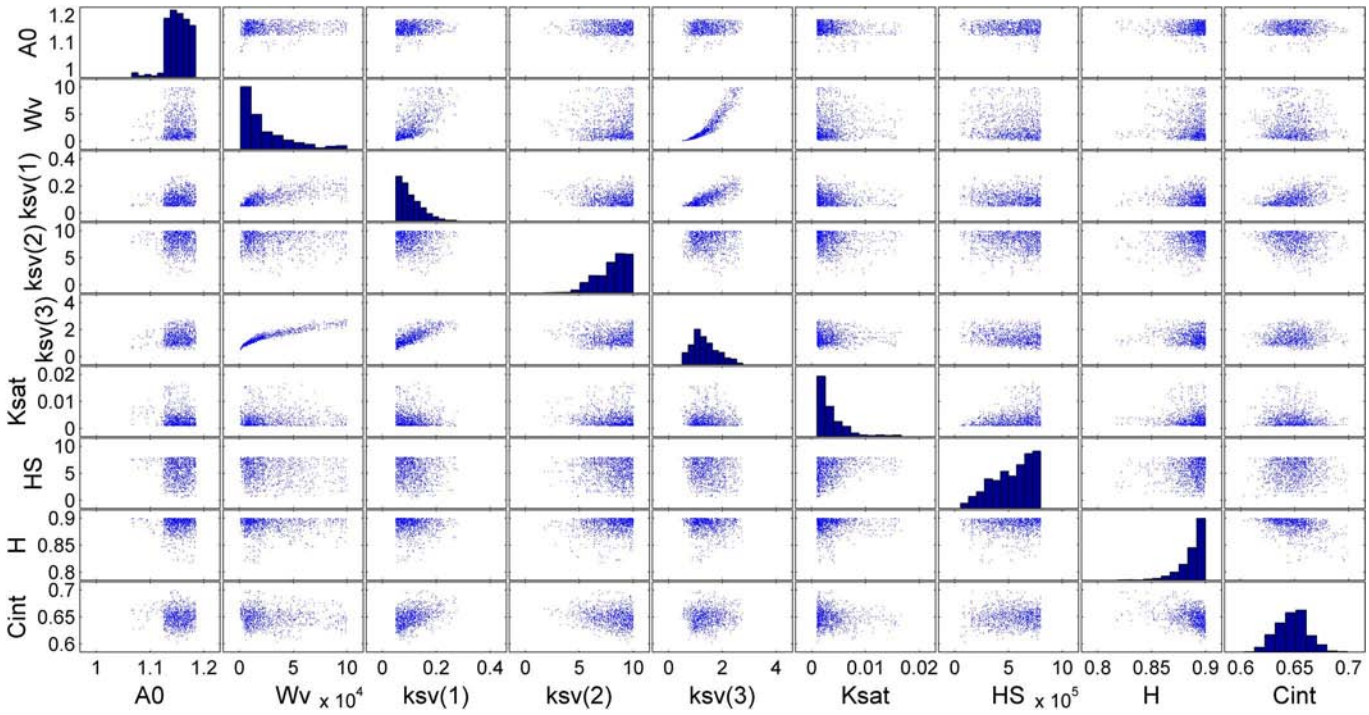

Fig. 13. Marginal distributions and two-dimensional correlation plot of posterior parameter samples of the DREAM model for the third event

(H) have the lowest CV values among the different events in the DREAM algorithm. So these parameters seem to be very sensitive and influential in the forecast model. Soil antecedent moisture and initial basin conditions are two key parameters in the model for separating effective rainfall into runoff and evapotranspiration (e.g., Samadi 2016). Together, these results indicate that the simulation of the standard least-squares parameter estimates is consistent with observations. In practice, a fat-tailed Gaussian distribution allowed for a combination of small to large errors, which induced robustness against outliers, and randomness in this semiarid flood modeling.

Conclusion and Future Work

This paper examined MCMC Bayesian algorithms to simulate flash flood events over a 290-km^2 semiarid watershed system in Iran.

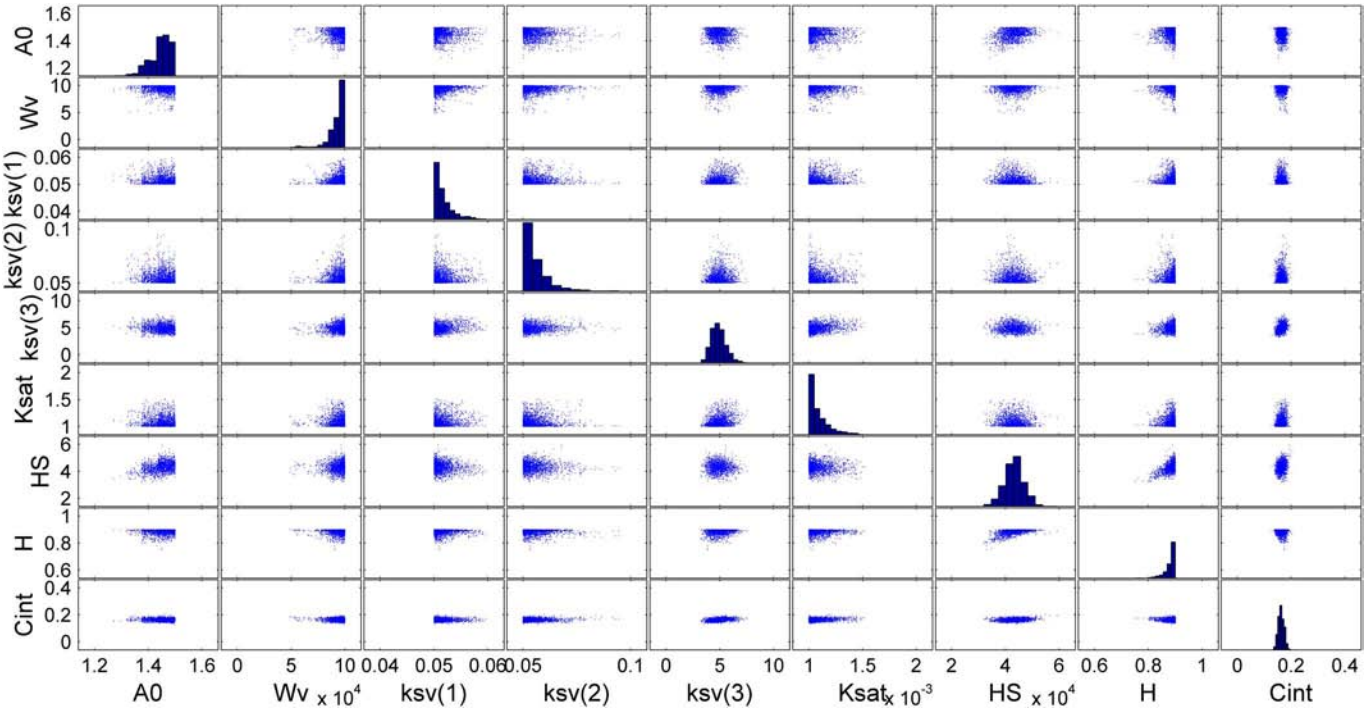

Fig. 14. Marginal distributions and two-dimensional correlation plot of posterior parameter samples of the DREAM model for the fourth event

Table 5. Coefficient of Variation for All Events Using the DREAM Algorithm

Parameter	Event 1	Event 2	Event 3	Event 4
A_0	16.09	1.28	1.83	2.57
W_v	22.31	7.79	95.18	7.88
K_{sv} (1)	17.46	3.99	41.43	3.23
K_{sv} (2)	48.95	3.61	17.89	10.34
K_{sv} (3)	33.89	13.39	35.05	13.72
K_{sat}	71.10	1.80	80.53	8.44
H_s	34.21	0.75	34.87	8.22
H	4.84	0.27	1.33	2.24
C_{int}	6.30	1.25	2.18	6.42

Bayesian algorithms were coupled with a fully distributed hydrology model, i.e., AFFDEF, to simulate four semiarid flood events with varying degrees of magnitudes and duration. The authors also modified the source code of AFFDEF by including a new subsurface module that was designed to resolve groundwater impediment in a short-term forecast model.

Analysis suggests that among three uncertainty algorithms DREAM showed more proficiency in inferring the underlying error structure and summarizing the posterior parameter distribution. The standard least-squares error model combined with a fat-tailed Gaussian distribution worked quite well for most of the events, especially for the first event. Because in semiarid flash flood modeling, peak flow dominates the overall flood values, it appears that SLS is more reliable in this case study since it puts more emphasis on fitting peak flows, whereas the GL error model (i.e., DREAM-ZS) puts less emphasis on this because of heteroscedastic errors (e.g., Schoups and Vrugt 2010). In other words, flood events with high peak rates and fewer near-zero flows can be best described by the Gaussian distribution. Such conditions may lead to a residual error distribution that is less skewed but more peaked. However, DREAM showed more sensitivity to the simulation of multipeak events (with extremely high peaks). Arguably, the sensitivity of parameters was extremely maximized during multipeak events, while model parameters showed less sensitivity during single-peak events. This reveals that addressing other sources of uncertainty may be equally important, particularly for multipeak events. Nonetheless, the performance of uncertainty algorithms may be data dependent and differs case by case.

In addition, parameter inference based on the GL error assumption yielded parameter estimates that deviated significantly from those obtained by the SLS error model. This is most notably the case for the interception reservoir capacity parameter (C_{int}), which is a key physics-based parameter in the AFFDEF model for separating effective rainfall into runoff, evaporation, and percolation. Parameter uncertainty, as measured by the width or spread of the posterior parameter distribution, is overestimated by GL (not shown here) for most of the parameters compared to SLS. This is, for example, clearly the case for K_{sat} and H_s.

Although the authors computed the residual error distribution on estimates of parameter uncertainty, these errors may also be caused by a combination of other sources of errors (e.g., measurement errors, model structural error, and error in rainfall spatial distribution and magnitudes). For instance, it could be hypothesized that runoff in this watershed is dominated by the infiltration-excess mechanism (Hortonian overland flow). This would imply that the runoff coefficient not only depends on soil antecedent moisture conditions, as accounted for in the hydrologic model, but also on rainfall intensity. As the intensity of semiarid rainfall varies in space and time even during a very short-term storm, rainfall data used in this study may not represent spatial variability and heterogeneity of semiarid rainfall magnitudes and dynamics. Hence, a switch to remote-sensed and/or radar-based rainfall data, as well as coupling additional modules to the Hortonian overland flow mechanism in the model, may be necessary to adequately simulate semiarid flood events.

Deficiency in flood modeling may be further attributed to the discrepancy in the structure of Bayesian algorithms (e.g., S. Samadi et al., "Estimating hydrologic model predictive uncertainty in the presence of complex residual error structures," submitted, *J. Hydrol.*

Eng., ASCE, Reston, Virginia). Bayes' law suffers from interaction between individual error sources (Vrugt et al. 2009b), which leads to structured and complex residual error distribution (S. Samadi et al., "Estimating hydrologic model predictive uncertainty in the presence of complex residual error structures," submitted, *J. Hydrol. Eng.*, ASCE, Reston, Virginia). In sum, when the MCMC Bayesian model is applied to a short-term forecast realm with high dimensionality and magnitudes, it faces two major challenges. First, the short-term nonlinear processes may make MCMC simulation computationally unaffordable to evaluate the posterior distribution function. Second, a high-dimensional parameter space makes exploration of the posterior parameter distribution especially difficult.

Combining the advanced MCMC algorithm with some dimensionality reduction techniques (Lu et al. 2014) and/or reparametrizing the original inverse problem using super and base parameters derived from principal component analysis (e.g., Tonkin and Doherty 2005) seems attractive to accelerate convergence to a limiting distribution (Laloy and Vrugt 2012; Schoups et al. 2010). In addition, evaluation of SLS and GL error assumptions is necessary to address the challenges posed by dimensionality and physics of the system with respect to different watershed systems (humid versus arid/semiarid); hydrological conditions [e.g., dry versus wet soil antecedent conditions (Clark et al. 2008)]; and diagnostic signatures that have explanatory power to identify dominant hydrological processes within a model (e.g., Gupta et al. 2008).

Apart from enhancing Bayesian techniques and comparing different hydrological diagnostic indices, improving the ability of hydrology models to explain the heterogeneity of biophysical and hydrological processes across a range of spatial scales is necessary to motivate modeling schemes to better suit local conditions. For instance, explicitly simulating all dominant biophysical and hydrological processes to represent spatial variability and hydrological connectivity (Clark et al. 2015); robust flux parameterizations over a hierarchy of spatial scales (McMillan et al. 2014); different modeling options for physical processes and numerical solvers (e.g., Clark et al. 2015); and advances in multiple model representations (hypotheses) of hydrologic processes (e.g., Pomeroy et al. 2007; Clark et al. 2008; Essery et al. 2013; Clark et al. 2015) are required to advance simulation, particularly over a short-term forecast scheme. Such capabilities help facilitate the diagnosis of model adequacy problems, determine the major sources of uncertainties (e.g., Moradkhani et al. 2005; Liu and Gupta 2007), classify dominant hydrological processes, and identify critical areas of future research.

The fundamental questions for future studies follow the key challenges: (1) How can researchers assimilate the understanding of environmental physics and processes in the model domain and define the major sources and nature of model deficiencies (Gupta et al. 2012)? (2) What is likely the interplay between model parameters and (physical) process parameterizations? (3) How can those interlinked physical processes be combined to produce the system-scale response across a hierarchy of spatial and temporal scales? The authors leave these approaches for future discussion.

Acknowledgments

The authors appreciate those persons and agencies that assisted in accessing research data. Particular acknowledgment is given to the SCEM-UA, DREAM, and DREAM-ZS developer, Dr. Jasper Vrugt, from the University of California at Irvine, for his advice throughout the study. Special thanks are owed to Dr. Karim Abbaspour at the Swiss Federal Institute of Aquatic Science and Technology (Eawag), for providing an opportunity to visit Eawag for the first author during this research. The authors wish to thank three anonymous reviewers and the associate editor for their constructive criticism and fruitful discussions. This project was funded by Khozestan Water and Power Authority (KWPA) for Shahid Chamran University (Grant No. 038-02-02-89). The modified source code of AFFDEF linked to MCMC samplers can be obtained from the first author upon request.

References

Abbaspour, K. C., et al. (2007). "Modelling hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT." *J. Hydrol.*, 333(2), 413–430.

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

Ajami, N. K., Duan, Q., and Sorooshian, S. (2007). "An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction." *Water Resour. Res.*, 43(1), W01403.

Band, L. E. (1986). "Topographic partition of watersheds with digital elevation models." *Water Resour. Res.*, 22(1), 15–24.

Blasone, R. S., Vrugt, J. A., Madsen, H., Rosbjerg, D., Robinson, B. A., and Zyvoloski, G. A. (2008). "Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov chain Monte Carlo sampling." *Adv. Water Resour.*, 31(4), 630–648.

Blöschl, G., Nester, T., Komma, J., Parajka, J., and Perdigão, R. A. P. (2013). "The June 2013 flood in the Upper Danube basin, and comparisons with the 2002, 1954 and 1899 floods." *Hydrol. Earth Syst. Sci.*, 17(12), 5197–5212.

Box, G. E., and Tiao, G. C. (2011). *Bayesian inference in statistical analysis*, Vol. 40, Wiley, Hoboken, NJ.

Brath, A., Montanari, A., and Moretti, G. (2003). "Assessing the effects on flood risk of land-use changes in the last five decades: An Italian case study." *Hydrology in Mediterranean and Semiarid Regions: Int. Conf.*, IAHS Press,Wallingford, U.K., 278, 435–441.

Chow, V. T., Maidment, D. R., and Mays, L. W. (1988). *Applied hydrology*, McGraw-Hill, New York, 572.

Clark, M. P., et al. (2008). "Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models." *Water Resour. Res.*, 44(12), W00B02.

Clark, M. P., et al. (2015). "Improving the representation of hydrologic processes in earth system models." *Water Resour. Res.*, 51(8), 5929–5956.

Dekker, S. C., Vrugt, J. A., and Elkington, R. J. (2012). "Significant variation in vegetation characteristics and dynamics from ecohydrological optimality of net carbon profit." *Ecohydrology*, 5(1), 1–18.

De Martonne, E. (1926). "L'indice d'aridité." *Bulletin de l'Association de géographes français*, 3(9), 3–5.

Doorenbos, J., et al. (1984). "Guidelines for predicting crop water requirements." FAO, Rome.

Essery, R., Morin, S., Lejeune, Y., and Menard, C. B. (2013). "A comparison of 1701 snow models using observations from an alpine site." *Adv. Water Resour.*, 55, 131–148.

Etemadi, H., Samadi, S., and Sharifikia, M. (2014). "Uncertainty analysis of statistical downscaling techniques in an arid region." *Clim. Dyn.*, 42(11–12), 2899–2920.

Etemadi, H., Samadi, S., Sharifikia, M., and Smoak, J. M. (2015). "Assessment of climate change downscaling and non-stationarity on the spatial pattern of a mangrove ecosystem in an arid coastal region of southern Iran." *Theor. Appl. Climatol.*, 1–15.

Feyen, L., Vrugt, J. A., Nualláin, B. Ó., van der Knijff, J., and De Roo, A. (2007). "Parameter optimisation and uncertainty assessment for large-scale streamflow simulation with the LISFLOOD model." *J. Hydrol.*, 332(3), 276–289.

Gelman, A., and Rubin, D. B. (1992). "Inference from iterative simulation using multiple sequences." *Stat. Sci.*, 7(4), 457–472.

Georgekakos, K. P., Seo, D. J., Gupta, H., Schaake, J., and Butts, M. B. (2004). "Characterizing streamflow simulation uncertainty through multimodel ensembles." *J. Hydrol.*, 298(1–4), 222–241.

Gupta, H. V., Clark, M. P., Vrugt, J. A., Abramowitz, G., and Ye, M. (2012). "Towards a comprehensive assessment of model structural adequacy." *Water Resour. Res.*, 48(8), 1–16.

Gupta, H. V., Sorooshian, S., and Yapo, P. O. (1998). "Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information." *Water Resour. Res.*, 34(4), 751–763.

Gupta, H. V., Wagener, T., and Liu, Y. (2008). "Reconciling theory with observations: Elements of a diagnostic approach to model evaluation." *Hydrol. Processes*, 22(18), 3802–3813.

Kirchner, J. W. (2006). "Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology." *Water Resour. Res.*, 42(3), 1–5.

Klepper, O., Scholten, H., and van de Kamer, J. P. G. (1991). "Prediction uncertainty in an ecological model of the Oosterschelde Estuary." *J. Forecasting*, 10(1–2), 191–209.

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

Kundzewicz, Z. W., and Kaczmarek, Z. (2000). "Coping with hydrological extremes." *Water Int.*, 25(1), 66–75.

Laloy, E., and Vrugt, J. A. (2012). "High-dimensional posterior exploration of hydrologic models using multiple-try DREAM (ZS) and high-performance computing." *Water Resour. Res.*, 48(1), 1–18.

Liu, Y., and Gupta, H. V. (2007). "Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework." *Water Resour. Res.*, 43(7), 1–18.

Lu, D., Ye, M., Hill, M. C., Poeter, E. P., Curtis, G. P. (2014). "A computer program for uncertainty analysis integrating regression and Bayesian methods." *Environ. Modell. Software*, 60, 45–56.

Madsen, H. E. N. R. I. K., Rosbjerg, D., Damgard, J., and Hansen, F. S. (2003). "Data assimilation in the MIKE 11 Flood Forecasting system using Kalman filtering." *Int. Assoc. Hydrol. Sci. Publ.*, 281, 75–81.

Mantovan, P., and Todini, E. (2006). "Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology." *J. Hydrol.*, 330(1), 368–381.

McMillan, H., and Clark, M. (2009). "Rainfall-runoff model calibration using informal likelihood measures within a Markov chain Monte Carlo sampling scheme." *Water Resour. Res.*, 45(4), 1–12.

McMillan, H., Gueguen, M., Grimon, E., Woods, R., Clark, M., and Rupp, D. E. (2014). "Spatial variability of hydrological processes and model structure diagnostics in a 50 km^2 catchment." *Hydrol. Processes*, 28(18), 4896–4913.

Michaud, J. D. (1992). "Distributed rainfall-runoff modeling of thunderstorm generated floods: A case study in a mid-sized, semi-arid watershed in Arizona." Ph.D. dissertation, Dept. of Hydrology and Water Resource, Univ. of Arizona, Tucson, AZ.

Milly, P. C. D., Wetherald, R., Dunne, K. A., and Delworth, T. L. (2002). "Increasing risk of great floods in a changing climate." *Nature*, 415(6871), 514–517.

Montgomery, D. R., and Foufoula-Georgiou, E. (1993). "Channel network source representation using digital elevation models." *Water Resour. Res.*, 29(12), 3925–3934.

Moore, C., Wöhling, T., and Doherty, J. (2010). "Efficient regularization and uncertainty analysis using a global optimization methodology." *Water Resour. Res.*, 46(8), 1–17.

Moradkhani, H., Hsu, K.-L., Gupta, H., and Sorooshian, S. (2005). "Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter." *Water Resour. Res.*, 41(5), 1–14.

Moretti, G., and Montanari, A. (2007). "AFFDEF: A spatially distributed grid based rainfall-runoff model for continuous time simulations of river discharge." *Environ. Modell. Software*, 22(6), 823–836.

Morse, B. S., Pohll, G., Huntington, J., and Rodrigues-Castillo, R. (2003). "Stochastic capture zone analysis of arsenic-contaminated well using the generalized likelihood uncertainty estimator (GLUE) methodology." *Water Resour. Res.*, 39(6), n/a.

Mosaedi, A., Zare Abyane, H., Ghabaei Sough, M., and Zahra Samadi, S. Z. (2015). "Long-lead drought forecasting using equiprobability transformation function for reconnaissance drought index." *Water Resour. Manage.*, 29(8), 2451–2469.

Nourali, M., Ghahraman, B., Pourreza-Bilondi, M., and Davary, K. (2016). "Effect of formal and informal likelihood functions on uncertainty assessment in a single event rainfall-runoff model." *J. Hydrol.*, 540, 549–564.

Osborn, H. B. (1964). "Effect of storm duration on runoff from rangeland watersheds in the semiarid southwestern United States." *Hydrol. Sci. J.*, 9(4), 40–47.

Pomeroy, J. W., et al. (2007). "The cold regions hydrological model: A platform for basing process representation and model structure on physical evidence." *Hydrol. Processes*, 21(19), 2650–2667.

Sadeghi-Tabas, S., Samadi, S. Z., Akbarpour, A., and Pourreza-Bilondi, M. (2016). "Sustainable groundwater modeling using single-and multi-objective optimization algorithms." *J. Hydroinform.*, jh2016006.

Samadi, S. (2016). "Assessing the sensitivity of SWAT physical parameters to potential evapotranspiration estimation methods over a coastal plain watershed in the southeastern United States." *Hydrol. Res.*, nh2016034.

Samadi, S., Carbone, G., Mahdavi, M., Sharifi, F., and Bihamta, M. R. (2013a). "Statistical downscaling of streamflow in a semi-arid catchment." *Water Resour. Manage.*, 27(1), 117–136.

Samadi, S., Wilson, C. A., and Moradkhani, H. (2013b). "Uncertainty analysis of statistical downscaling models using Hadley Centre Coupled Model." *Theor. Appl. Climatol.*, 114(3–4), 673–690.

Schoups, G., and Vrugt, J. A. (2010). "A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non- Gaussian errors." *Water Resour. Res.*, 46(10), 1–17.

Schoups, G., Vrugt, J. A., Fenicia, F., and Van de Giesen, N. C. (2010). "Corruption of accuracy and efficiency of Markov chain Monte Carlo simulation by inaccurate numerical implementation of conceptual hydrologic models." *Water Resour. Res.*, 46, W10530.

Sharifi, F., Samadi, S. Z., and Wilson, C. A. M. E. (2012). "Causes and consequences of recent floods in the Golestan catchments and Caspian Sea regions of Iran." *Nat. Hazards*, 61(2), 533–550.

Soil Conservation Service. (1972). "Storm rainfall depth and distribution." *National engineering handbook, section 4, hydrology*, U.S. Dept. of Agriculture,Washington, DC.

Stedinger, J. R., Vogel, R. M., Lee, S. U., and Batchelder, R. (2008). "Appraisal of the generalized likelihood uncertainty estimation (GLUE) method." *Water Resour. Res.*, 44(12), W00B06.

ter Braak, C. J. F. (2006). "A Markov chain Monte Carlo version of the genetic algorithm differential evolution: Easy Bayesian computing for real parameter space." *Stat. Comput.*, 16(3), 239–249.

ter Braak, C. J. F., and Vrugt, J. A. (2008). "Differential evolution Markov chain with snooker updater and fewer chains." *Stat. Comput.*, 18(4), 435–446.

Thiemann, M., Trosset, M., Gupta, H., and Sorooshian, S. (2001). "Bayesian recursive parameter estimation for hydrological models." *Water Resour. Res.*, 37(10), 2521–2535.

Tonkin, M., and Doherty, J. (2009). "Calibration- constrained Monte Carlo analysis of highly parameterized models using subspace techniques." *Water Resour. Res.*, 45(12), W00B10.

Tonkin, M. J., and Doherty, J. (2005). "A hybrid regularized inversion methodology for highly parameterized environmental models." *Water Resour. Res.*, 41(10), W10412.

Volkmann, T. H., Lyon, S. W., Gupta, H. V., and Troch, P. A. (2010). "Multicriteria design of rain gauge networks for flash flood prediction in semiarid catchments with complex terrain." *Water Resour. Res.*, 46(11), W11554.

Vrugt, J. A., Diks, C. G., Gupta, H. V., Bouten, W., and Verstraten, J. M. (2005). "Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation." *Water Resour. Res.*, 41(1), W01017.

Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W., and Sorooshian, S. (2003a). "Effective and efficient algorithm for multi-objective optimization of hydrologic models." *Water Resour. Res.*, 39(8), 1214.

Vrugt, J. A., Gupta, H. V., Bouten, W., and Sorooshian, S. (2003b). "A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters." *Water Resour. Res.*, 39(8), 1201.

Vrugt, J. A., and Robinson, B. A. (2007). "Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging." *Water Resour. Res.*, 43(1), W01411.

Vrugt, J. A., Ter Braak, C. J., Clark, M. P., Hyman, J. M., and Robinson, B. A. (2008). "Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation." *Water Resour. Res.*, 44(12), W00B09.

Vrugt, J. A., Ter Braak, C. J., Gupta, H. V., and Robinson, B. A. (2009a). "Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?" *Stochastic Environ. Res. Risk Assess.*, 23(7), 1011–1026.

Vrugt, J. A., Ter Braak, C. J. F., Diks, C. G. H., Robinson, B. A., Hyman, J. M., and Higdon, D. (2009b). "Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling." *Int. J. Nonlinear Sci. Numer. Simul.*, 10(3), 273–290.

Vrugt, J. A., Weerts, A. H., and Bouten, W. (2001). "Information content of data for identifying soil hydraulic parameters from outflow experiments." *Soil Sci. Soc. Am. J.*, 65(1), 19–27.

Yatheendradas, S., et al. (2008). "Understanding uncertainty in distributed flash flood forecasting for semiarid regions." *Water Resour. Res.*, 44(5), W05S19.

Zhou, R., Li, Y., Lu, D., Liu, H., and Zhou, H. (2016). "An optimization based sampling approach for multiple metrics uncertainty analysis using generalized likelihood uncertainty estimation." *J. Hydrol.*, 540, 274–286.