Case Study

Comparison of Likelihood-Free Inference Approach and a Formal Bayesian Method in Parameter Uncertainty Assessment: Case Study with a Single-Event Rainfall–Runoff Model

Mahrouz Nourali, Ph.D.[1]

Abstract: In the present study, DREAM$_{(ZS)}$ and DREAM$_{(ABC)}$ (which stands for differential evolution adaptive metropolis) algorithms were applied to determine the parameters' uncertainty in a single-event rainfall–runoff model, and rainfall multipliers were also used to correct rainfall forcing errors. Moreover, DREAM$_{(ZS)}$, based on the original DREAM algorithm, and the DREAM$_{(ABC)}$ algorithm, as a likelihood-free inference approach, were both used to explore the posterior parameters in high-dimensional inference problems. Before comparing DREAM$_{(ZS)}$ with DREAM$_{(ABC)}$, some underlying assumptions of residual distribution were analyzed and then fulfilled to obtain a suitable likelihood function and also to provide a more reliable estimation of the parameters. Despite the use of an acceptable likelihood function in the DREAM$_{(ZS)}$ algorithm, the results confirm the advantage of the DREAM$_{(ABC)}$ for assessing the uncertainty in a single-event model and high-dimensional parameter spaces. Moreover, an acceptable distance function used in DREAM$_{(ABC)}$ is suggested to assess the uncertainty in a single-event rainfall-runoff model (HEC-HMS). Occasional flash floods occur in this study region and in large parts of Iran. The results of this study can, therefore, be useful for achieving more accurate predictions and planning for flood control management. **DOI: 10.1061/(ASCE)HE.1943-5584.0002048.**

Author keywords: Acceptable likelihood function; Parameter uncertainty; DREAM$_{(ABC)}$; DREAM$_{(ZS)}$; Single-event rainfall–runoff model.

Introduction

Hydrological models represent the simplified characterizations of a real-world system (Moradkhani and Sorooshian 2009). These models have been applied in various fields of study, such as real-time flood predicting, the evaluation of the influence of land use and climate change, water balance estimations, water resource scheduling, and managing river basins, as well as the operation of water-related services and hydraulic energy systems (Alazzy et al. 2015). The rainfall–runoff model, as an example of hydrological models, summarizes the response of a catchment to forcing variables and predicts simulated runoff. The runoff predictions with these models have been recognized to be associated with a certain degree of uncertainty.

It has been argued that natural randomness, the uncertainty of parameters, model structure inaccuracies, and errors in the observational outcomes and rainfall forcing all cause the inability of a model to generate accurate predictions (Blasone 2007; Beven et al. 2011; Willems 2012; Sadegh and Vrugt 2013). These uncertainties occur as input data; for example, forcing data to contain significant errors because of systematic and random measurement errors related to gauges and scaling issues (McMillan et al. 2011). Streamflow data uncertainty may result from discharge gauging errors and errors of the rating curve (Renard et al. 2011), and model structural uncertainty is due to simplifications in model formulation. Deficiencies of model structure and errors in the streamflow observations and input data may result in parametric uncertainty (Renard et al. 2010). The estimation of model parameters involves significant uncertainties, meaning that values of model parameters cannot be exactly specified. For instance, it is possible to estimate these parameters by calibration; however, errors in this process are also inevitable. The calibrated model parameters are assigned unrealistic values to compensate for deficiencies in model structures and input and output measurements (Clark and Vrugt 2006). Therefore, uncertainty analysis is an indispensable element to quantify the uncertainty associated with the predictions for any hydrologic modeling study (Beven 2006; Blasone et al. 2008).

Among various approaches presented for uncertainty estimation in the literature (Vrugt and Bouten 2002; Vrugt and Robinson 2007; Reichert and Mieleitner 2009; Salamon and Feyen 2009), the Bayesian approach has been extensively applied for uncertainty assessment in hydrologic modeling (Jin et al. 2010; Li et al. 2010; Sikorska 2012; Nourali et al. 2016). In Bayesian inference, the prior knowledge of parameters and likelihood function are used to derive the posterior parameters. In most hydrological models, it is difficult to estimate the analytical solution for the posterior distribution. Consequently, the Markov chain Monte Carlo (MCMC) methodology is used to get the posterior distribution of the model parameters under a Bayesian framework (Smith and Marshall 2008; ter Braak and Vrugt 2008).

The reliability of the posterior parameters is dependent on the realism in the likelihood measure used to yield a fit between the predicted and observed outputs. In the Bayesian inference procedure, it is necessary to describe the distribution of the model errors reasonably to select an appropriate likelihood function (Li et al. 2011). Moreover, to achieve more accurate predictions, a likelihood function has to be selected, which considers various sources of

[1]Researcher, Dept. of Water Engineering, Faculty of Agriculture, Ferdowsi Univ. of Mashhad, International Campus, Mashhad 9177948974, Iran. ORCID: https://orcid.org/0000-0002-1918-6058. Email: mahrouznourali@yahoo.com

Note. This manuscript was submitted on July 16, 2020; approved on October 16, 2020; published online on December 18, 2020. Discussion period open until May 18, 2021; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Hydrologic Engineering***, © ASCE, ISSN 1084-0699.

error [such as errors in input data (forcing), model structures, and observed outcomes] separately, and provides reliable parameters of the model. Derived parameters and prediction uncertainties will not rely upon the arbitrary application of the formulation of the likelihood function (Yang et al. 2007).

Many hydrologic studies have used the formal Bayesian likelihood functions for describing the closeness of the simulated and observed flows and analyzing parameter uncertainty (Engeland et al. 2005; Jin et al. 2010; Li et al. 2010; Laloy et al. 2010; Schoups and Vrugt 2010). Formal likelihood functions that are based on statistical assumptions, such as the assumption of normally distributed residuals, homoscedasticity (constancy of variance), and independence of residual errors, provide unreliable parameters of the model because in many cases, these statistical assumptions of the residuals are obviously violated (Nourali et al. 2016). Residuals of the model often include errors in model structure and input data, and these epistemic errors result in autocorrelation and heteroscedasticity of the residuals (Beven et al. 2012). Violation of the assumptions of residual error reveals that a formal likelihood function used in the hydrologic studies must be based on a suitable statistical error model for input (forcing) data and model structural insufficiencies to allow for correlation, heteroscedasticity, the nonnormality of the residuals, and so forth (Beven et al. 2012). In addition, different approaches can be used to fulfill common assumptions of the errors. For instance, the first-order auto-regressive [AR(1)] scheme of the residuals is integrated into the likelihood function, which could be used to account for autocorrelations and eliminate a correlation of the errors (Li et al. 2011; Lu et al. 2013), and the Box–Cox transformation method is also applied to remove heteroscedasticity of model residuals (Cheng et al. 2014). Moreover, the latent variables (rainfall multipliers) can also be assigned to each individual storm event in order to account for rainfall forcing uncertainty (Kavetski et al. 2006a, b). The first-order autocorrelation coefficient (ρ) for the correlated error case, the additional latent variables (for example, rainfall multipliers for forcing error), and the model parameters can be estimated simultaneously during calibration. However, the number of parameters needed to estimate it is increased in this approach. To resolve this problem successfully, the DREAM$_{(ZS)}$ (differential evolution adaptive metropolis) algorithm (based upon the MCMC scheme) is applied for estimating the posterior parameters in high-dimensional inference problems (Minasny et al. 2011). DREAM$_{(ZS)}$, based on the original DREAM algorithm (Vrugt et al. 2009a), is efficient to estimate the posterior parameters in the hydrologic models (Laloy and Vrugt 2012). Accordingly, this algorithm applies the likelihood function to define the uncertainty of the posterior parameter and increases the convergence speed for exploring the posterior distribution of high-dimensional search problems.

Given the difficulty mentioned previously to configure the form of an appropriate likelihood function applied in the DREAM$_{(ZS)}$ algorithm, it would be possible to use the likelihood-free inference for estimating the posterior parameter distribution. This inference methodology is applied for cases in which the evaluation of the likelihood is computationally difficult or prohibitive, and the likelihood function could not be formulated explicitly (Gourieroux et al. 1993; Fu and Li 1997). The likelihood-free method, named the approximate Bayesian computation (ABC) (Beaumont 2010; Sadegh and Vrugt 2013, 2014; Chiachio et al. 2014; Abdessalem et al. 2018), uses a set of summary statistics to infer the posterior distributions. Literature reviews reveal the application of the ABC method in hydrology, including, signature-domain calibration of hydrological models (Fenicia et al. 2018), a connection between ABC and GLUE-based approaches (Nott et al. 2012), and the performance comparison between ABC and MCMC methods (Romero-Cuellar et al. 2019).

Besides, this method is useful for diagnostic model calibration and evaluation. The idea behind a diagnostic evaluation is to diagnose faulty components of a model that need improvement to explain the discrepancy between the computed and observed streamflow (Gupta et al. 2008). Ratmann et al. (2009) investigated the application of the ABC algorithm to diagnose problems in the model specification and demonstrated how this method diagnoses the model mismatch. Vrugt and Sadegh (2013) showed, in part, the capability of the likelihood-free approach using hydrological signatures for significant diagnosis, detection, and the resolution of model structural deficiencies and illuminated how the model should be enhanced.

ABC algorithms can take many forms (Abdessalem et al. 2018). The ABC rejection sampling (ABC-REJ) algorithm (Pritchard et al. 1999; Marjoram et al. 2003) and the ABC population Monte Carlo sampling (ABC-PMC) algorithm (Beaumont et al. 2009; Turner and van Zandt 2012) work well in low-dimensional parameter spaces (e.g., $d < 10$) due to the use of boxcar fitness kernel (Sadegh and Vrugt 2014). To search the high-dimensional parameter spaces efficiently, the DREAM$_{(ABC)}$ algorithm, which is the MCMC implementation of the ABC, has been proposed by Sadegh and Vrugt (2014). Using DREAM$_{(ABC)}$, the ABC sampling effectiveness is increased, and a continuous fitness function is applied to explore the posterior parameters efficiently. Moreover, the DREAM$_{(ABC)}$ algorithm is also used for solving the diagnostic assessment problem of a complex system model involving high-dimensional parameter spaces.

Several studies have used DREAM$_{(zs)}$ and various ABC algorithms [such as DREAM$_{(ABC)}$] for estimating the parameter uncertainty of the continuous models rather than the single-event models in long-term runoff forecasting (Schoups and Vrugt 2010; Koskela et al. 2012; Vrugt and Sadegh 2013; Sadegh et al. 2015). In a study, Sadegh and Vrugt (2014) made a comparison between DREAM and DREAM$_{(ABC)}$ and used a residual-based Gaussian likelihood function and three hydrologic signatures in DREAM and DREAM$_{(ABC)}$, respectively. There is no previous study in the hydrology literature to consider rainfall uncertainty and analyze and fulfill underlying assumptions of Gaussian distributed error residuals in the likelihood function before comparison between DREAM$_{(ZS)}$ and DREAM$_{(ABC)}$, despite the violation of traditional error assumptions. Rainfall uncertainty and the analysis of the assumption of residual distribution are also necessary to obtain a suitable likelihood function corresponding to statistical properties of the model residuals for the accurate estimation of posterior parameters and predictive uncertainty. The question is whether the DREAM$_{(ZS)}$, through supposing a proper likelihood function and considering rainfall uncertainty, can relatively and accurately estimate predictive uncertainty and posterior parameters in comparison with a likelihood-free inference approach, entitled DREAM$_{(ABC)}$. This will be answered by the results of the present study.

In this paper, we seek to continue our previous work (Nourali et al. 2016). In our previous work (Nourali et al. 2016), several formal and informal likelihood functions were applied to investigate the uncertainty of the parameters of a single-event rainfall–runoff model using DREAM$_{(ZS)}$. In this regard, a proper likelihood function based on the AR(1) model represented a better performance among the other likelihood functions. Notably, rainfall data error was also disregarded in this study. In the current study, the DREAM$_{(ZS)}$ algorithm and a likelihood-free inference approach entitled DREAM$_{(ABC)}$ are applied to determine the uncertainty of the parameter in a single-event rainfall–runoff model as well as comparing the performance of these algorithms with each other.

Rainfall multipliers are also used for each individual storm event in both DREAM$_{(ZS)}$ and DREAM$_{(ABC)}$ algorithms in order to

correct rainfall forcing errors that can yield a more precise estimation of the parameters and help to diagnose model structural errors (Vrugt et al. 2008; Koskela et al. 2012). Moreover, parallel computing is applied to decrease the high calculation time. Before comparing DREAM$_{(ZS)}$ with DREAM$_{(ABC)}$, some underlying assumptions of residual distribution were analyzed and then fulfilled, and a suitable likelihood function based on an appropriate autoregressive error model [elaborated on in the section "Formal Bayesian inference by DREAM$_{(ZS)}$ Algorithm"] was used in the DREAM$_{(ZS)}$ algorithm after analyzing the assumption of the residual distribution.

In the application of the DREAM$_{(ABC)}$, a proper distance function should be used considering reliable signature measures. For this purpose, three different distance functions [described in detail in the section "Approximate Bayesian Computation by DREAM$_{(ABC)}$ Algorithm"] are examined to assess their influence on posterior distributions of parameters, prediction uncertainty, computational efficiency, and the performance of the algorithms. The results are analyzed, and a proper distance function can be proposed, which is suitable for single flood events applied in the present case study. After considering rainfall uncertainty, the appropriate likelihood and distance functions are obtained for DREAM$_{(zs)}$ and DREAM$_{(ABC)}$, respectively, and the results obtained from running these algorithms can be compared to each other to evaluate the parameters' uncertainty of a single-event hydrologic model (HEC-HMS). Then, the DREAM$_{(ABC)}$ algorithm, having the best performance among the other algorithms, is applied to solve the diagnostic assessment problem of the HEC-HMS model used in the paper. There is no previous study to use the DREAM$_{(ABC)}$ algorithm in order to analyze the parameter uncertainty of the single-event rainfall–runoff model and compare it to the results of the DREAM$_{(ZS)}$ algorithm in Iran.

In the present study, the Tamar basin situated in the Gorganroud River basin, Golestan Province, Iran, is selected as the study area. This basin is a flood-prone region by considering the flood hazard map in the Gorganroud subwatersheds (Safaripour et al. 2012). Furthermore, occasional flash and disastrous floods occur in this study region as well as the provinces of Iran (Asgharpour and Ajdari 2011; Kamali and Mousavi 2014; Vaghefi et al. 2019). From 1990 to 2019, Golestan Province, and particularly the Gorganroud River basin in the north of Iran, has faced occasional flash floods that caused considerable damages (Safaripour et al. 2012; Sharifi et al. 2012; Gharibreza 2019). Considering the occurrence of occasional flash floods in the large parts of Iran, it is necessary to forecast floods and plan for flood control management accurately. The results of this study can, therefore, be useful for achieving more accurate predictions of the single-event modeling.

This paper is structured as follows; in the next section, a detailed description of the DREAM$_{(ABC)}$ and DREAM$_{(ZS)}$ algorithms and procedures for their implementation are fully explained. The uncertainty analysis of parameters and the model prediction are described to compare the algorithms. Next, the study area is explained in brief, and the hydrologic model is introduced. Ultimately, in the "Results and Discussion" section, the main conclusions of this study are presented, followed by a discussion of the main findings.

Methodology

The detailed description of two uncertainty analysis methods, including a formal Bayesian method and likelihood-free inference approach, is provided in the following sections.

Formal Bayesian Inference by DREAM$_{(ZS)}$ *Algorithm*

As mentioned in the "Introduction" section of this paper, parameter uncertainty of the hydrologic model can be assessed using a formal Bayesian likelihood in the DREAM$_{(ZS)}$ algorithm. DREAM$_{(ZS)}$ (Laloy and Vrugt 2012) is based on the original DREAM algorithm [see more details in the study by Vrugt et al. (2009a, b)], and the convergence of the chain to a stable posterior is accelerated by this algorithm for high-dimensional search problems. For monitoring the convergence of a DREAM$_{(ZS)}$ run, the R criterion of Gelman-Rubin (Gelman and Rubin 1992) of less than 1.2 is acceptable for all parameters; thus, each of the chains converges to the stable probability distribution. Following the convergence, the last 20% of the parameters are applied to form the posterior parameter densities.

To obtain a suitable likelihood function to be used in DREAM$_{(ZS)}$, first, assumptions of residuals are analyzed for flood events of the calibration phase [see more details on the applied method in the study by Nourali et al. (2016)]. When common assumptions of residuals (e.g., assumptions of Gaussian and independent errors) are violated, Box-Cox transformations (Box and Cox 1964) of the streamflow data and a p-order auto-regressive [AR(p)] scheme of the residuals incorporated into the formulation of the likelihood function can be applied to relax the explicit assumption of the residual distribution. Furthermore, the rainfall multipliers (μ) are assigned to each individual storm event, and these values and the first-order autocorrelation coefficient (ρ) (Table 3) are added to the optimization problem. Therefore, the number of dimensions will considerably increase, and parallel computing is applied for decreasing the high calculation time and to speed up the effectiveness of the posterior exploration (Laloy et al. 2010; Laloy and Vrugt 2012).

The formal likelihood function [$L(\theta|O)$] used in this study is as follows (Bates and Campbell 2001; Smith et al. 2015):

$$L(\theta|O) = -\frac{M-p}{2}\ln(2\pi\sigma^2) - \sum_{t=p+1}^{M}\frac{(\varepsilon_t^*(\theta,\mu) - \sum_{j=1}^{p}\emptyset_j\varepsilon_{t-j}^*(\theta,\mu))^2}{2\sigma^2} \quad (1)$$

where θ = vector of model parameters; σ^2 = model errors variance; $\emptyset$ = p-order the autoregressive coefficient; (μ) shows the rainfall multiplier; M = number of observed data; and $\varepsilon^*(\theta)$ shows the residuals vector after applying the Box–Cox transformation. The transformation parameter λ is fixed at the value of 0.3 (Vrugt et al. 2009b).

The same flood events used in the study by Nourali et al. (2016) and a new flood event (Table 2) are applied in the present study. In the "Results and Discussion" section, it will be explained that residual error assumptions for flood events of the calibration phase are violated. Firstly, the AR(1) model is applied to remove autocorrelation from the residuals, and then the results are analyzed. When the AR(1) model is unable to remove the autocorrelation from the errors, a higher-order AR model [e.g., AR(2) or AR (3)] is applied to remove the autocorrelation from residuals and obtain more realistic predictions.

Approximate Bayesian Computation by DREAM$_{(ABC)}$ *Algorithm*

The approximate Bayesian computation (ABC) sampling method is the class of likelihood-free methods applied in genetics (Tavaré et al. 1997; Foll et al. 2008) and ecology (Csilléry et al. 2010). This approach is used to obtain the posterior estimates and solve the problem of the Bayesian analysis when the likelihood function is difficult or impossible to be evaluated (Turner and Sederberg 2012). The ABC can only be implemented in the stochastic models (Wilkinson 2013; Burr and Skurikhin 2013; Fenicia et al. 2018). The random variable is not considered in the HEC-HMS model

as a deterministic model (USACE 2000), and, for a given input, the model will always produce the same output. However, the deterministic models in hydrology are never accurate, and these models cause uncertainty in the prediction (Montanari and Koutsoyiannis 2012). To apply the ABC methodology to a deterministic model (for example with the HEC-HMS model), the output of the model should be corrupted by a random measurement error. It means that the stochastic simulation is simply derived through the addition of a random measurement error to the simulated streamflow as follows (Sadegh and Vrugt 2013, 2014):

$$Q_{\text{comp}}(\theta) \leftarrow H(\theta|\cdot) + N(0, \hat{\sigma}_{Q_{obsj}}) \tag{2}$$

where Q_{obs} = actual observed values; Q_{comp} = predicted output; and $\hat{\sigma}_{Q_{obsj}}$ = error deviation of the streamflow measurement [Eq. (16)].

The idea of the ABC is that a candidate parameter (θ^*) from the prior distribution can be used for simulating the output of the model. Uniform prior distributions are used in all calculations using prior ranges (Table 3). The candidate prior value is accepted if the model prediction and observation are close enough [i.e., the distance between them, $\rho(Y(Q_{obs}), Y(Q_{\text{comp}}(\theta)))$ is less than a small tolerance (ε)] (Burr and Skurikhin 2013; Olson and Kleiber 2017). For adequately complex models in high-dimensional problems, the probability of providing a precise fit to the streamflow observations will not be high, and these models will frequently provide unacceptable fits to the data (Sadegh and Vrugt 2013, 2014). So, it may become necessary to reduce the dimensionality through the use of sufficient summary statistics, S (.), and a distance measure, $\rho(S(Q_{\text{obs}}), S(Q_{\text{comp}}(\theta)))$ (Beaumont 2010; Olson and Kleiber 2017.

A boxcar kernel (0/1) used in common ABC algorithms (e.g., ABC-REJ and ABC-PMC) cannot explore the parameter space efficiently to track the target distribution in high-dimensional parameter spaces (e.g., $d \geq 10$) (Turner and Sederberg 2012; Sadegh and Vrugt 2014). Thus, the DREAM$_{(\text{ABC})}$ algorithm [developed by Sadegh and Vrugt (2014)] can be applied to infer the parameters of the complex models. This algorithm substitutes the boxcar fitness kernel of the population Monte Carlo (PMC) simulation with a continuous kernel to find the posterior solutions efficiently. Other studies have presented the details of the DREAM$_{(\text{ABC})}$ algorithm (Sadegh and Vrugt 2014; Sadegh et al. 2015), and so they will not be considered in this study. In the current study, the total number of calibration parameters is 24 ($d \geq 10$) (Table 3), and so the use of the DREAM$_{(\text{ABC})}$ algorithm would be expected to yield more reliable results.

In the DREAM$_{(\text{ABC})}$ algorithm, the fitness of θ is calculated as follows (Sadegh and Vrugt 2014):

$$f(\theta, \varepsilon) = \varepsilon - \rho(S(Q_{obs}), S(Q_{\text{comp}}(\theta))) \tag{3}$$

Many summary statistics [e.g., the Nash-Sutcliffe coefficient (NS) (Nash and Sutcliffe 1970) and the other measures presented in the work of Gupta et al. (1998)] can be used in the ABC methodology (Sadegh and Vrugt 2013). As discussed by Gupta et al. (1998, 2005) and Wagener and Gupta (2005), a single measure of performance (e.g., the NS measure) is, in general, a weak statistic to make a meaningful evaluation of the model. Moreover, classical likelihood-based fitting methods [e.g., a classic Gaussian AR(1) model used in DREAM$_{(\text{ZS})}$] dilute and mix available information into an index having little remaining correspondence to specific behaviors of the system (Gupta et al. 2008). These approaches are not able to detect the structural deficiency of the model and cannot represent the malfunction of the model. Regarding solving the problem of the diagnostic model evaluation and identifying model structural deficiencies, Gupta et al. (2008) recommended applying multiple diagnostic measures as signature indices of system behavior. The DREAM$_{(\text{ABC})}$ algorithm has also been proposed as a likelihood-free method for the detection of model structural deficiencies (Vrugt and Sadegh 2013; Sadegh and Vrugt 2014). The advantage of DREAM$_{(\text{ABC})}$ is, therefore, obvious because diagnostic signature measures (summary statistics) used in the DREAM$_{(\text{ABC})}$ algorithm have a stronger diagnostic capability than some residual-based likelihood functions (Sadegh and Vrugt 2014).

The summary metrics should be adequate for approximating the posterior parameter distribution (Marjoram and Tavare´ 2006; Csilléry et al. 2010; Burr and Skurikhin 2013; Lintusaari et al. 2017). It is necessary to select these metrics carefully to draw significant information from the data correctly; thus, these criteria should signify hydrological signatures of relevant system behaviors. The appropriate summary statistics can be used for describing the model error and shedding light on the way to refine the model (Wilkinson 2013; Sadegh et al. 2015).

In the present case study, a set of ε values in an interval of [0.1, 1] is considered to obtain suitable ε values, and these values are separately used in each run of the DREAM$_{(\text{ABC})}$ algor ithm. The values of the final epsilon (ε) in Table 4 ar e determined by investigating a convergence to suitable stationary distribution for each run of the DREAM$_{(\text{ABC})}$. Thre edistance functions $\rho(\cdot)$ are used in the DREAM$_{(\text{ABC})}$ algorithm, and the rainfall-runoff model is run separately for each distance function. The first and second distance functions include different signatures that are selected based on characteristics of a runoff hydrograph that resulted from running the HEC-HMS model, whereas the third distance function uses the residual-based summary statistic. In the first distance function, three signature measures are used, including the peak flow error, the runoff volume error, and the peak time error (Green and Stephenson 1986; Tewolde 2005). These criteria are used in single-event modeling studies and are the main important features of a runoff hydrograph (Kamali et al. 2013). A comparison of the predicted and observed peak flow rates can be expressed as the absolute value of the error in the peak flow [Eq. (4)] (Green and Stephenson 1986; Tewolde 2005). The absolute value of the error in the volume of the runoff can be applied as the volume criterion [Eq. (5)]. The peak flow timing can be used as a hydrograph component in the absolute value of the peak time error, as shown in Eq. (6) (Tewolde 2005).

Three summary metrics, S_1–S_3, namely, the peak flow error [E_{peak}; Eq. (4)], the runoff volume error [E_{volum}; Eq. (5)], and the peak time error [E_{time}; Eq. (6)] are respectively defined

$$S_1 = E_{\text{peak}} = \frac{|Q_{p-\text{comp}}(\theta_i) - Q_{p-obs}|}{Q_{p-obs}} \tag{4}$$

$$S_2 = E_{\text{volum}} = \frac{|V_{\text{comp}}(\theta_i) - V_{\text{obs}}|}{V_{\text{obs}}} \tag{5}$$

$$S_3 = E_{\text{time}} = \frac{|t_{p-cp}(\theta_i) - t_{p-obs}|}{t_{p-obs}} \tag{6}$$

where $Q_{p-\text{comp}}(\theta) = \{Q_{p-\text{comp}}(\theta_1), \ldots, Q_{p-\text{comp}}(\theta_n)\}$ represents the vector of the estimated peak flow using the parameter set θ; Q_{p-obs} denotes the observed peak flow; $V_{\text{comp}}(\theta) = \{V_{\text{comp}}(\theta_1), \ldots, V_{\text{comp}}(\theta_n)\}$ is the vector of the computed total volume using the parameter set θ; V_{obs} denotes the observed total volume; $t_{p-\text{comp}}(\theta_i)$ is the time when $Q_{p-\text{comp}}$ occurs; and t_{p-obs} is the time when Q_{p-obs} happens.

The first distance function used in the DREAM$_{(\text{ABC})}$ algorithm is described

$$\rho(S(Q_{\text{obs}}), S(Q_{\text{comp}}(\theta)) = \max(S_k); \qquad k = \{1, 2, 3\} \tag{7}$$

In the second distance function, two signature measures are used, including high-flow and low-flow signatures defined as the normalized root-mean-square error [NRMSE, Eq. (9)], and the normalized and transformed root-mean-square error [NTRMSE, Eq. (11)], respectively (Corzo and Solomatine 2007; Dawson et al. 2007; Wagener et al. 2009; Tian et al. 2016). The RMSE [Eq. (8)] tends to put a stronger emphasis on fitting the high flows and can be normalized by the standard deviation of the observed value as follows:

$$\text{RMSE} = \sqrt{\frac{SSE}{M}} \tag{8}$$

$$S1 = \text{NRMSE} = \frac{\text{RMSE}}{\sigma_{Q_{obs}}} \tag{9}$$

$$SSE = \sum_{j=1}^{M} (Q_{\text{comp}j} - Q_{obsj})^2 \tag{10}$$

where $Q_{obsj} = j$th observation of the observed discharge; $Q_{\text{comp}j} = j$th type of the simulated discharge; M = number of observations; and $\sigma_{Q_{\text{obs}}}$ = standard deviation of the observed value.

The normalized and transformed root-mean-square error [NTRMSE, Eq. (11)] which emphasizes the low flows, is defined

$$S2 = \text{NTRMSE} = \frac{\sqrt{\frac{SSE^*}{M}}}{\sigma_{Q_{obs}}} \tag{11}$$

$$SSE^* = \sum_{j=1}^{M} (Q^*_{\text{comp}j} - Q^*_{obsj})^2 \tag{12}$$

$$Q^* = \frac{(1+Q)^{\lambda} - 1}{\lambda} \tag{13}$$

where $Q^*_{\text{comp}j}$ = Box-Cox transformed simulated flow (Box and Cox 1964) at time j; Q^*_{obsj} = Box-Cox transformation of the observed flow at time j; and λ is assigned a value of 0.3 (Shamir et al. 2005; Wagener et al. 2009).

For these summary metrics, the distance function is as follows:

$$\rho(S(Q_{obs}), S(Q_{\text{comp}}(\theta)) = \max(S_k); \qquad k = \{1, 2\} \tag{14}$$

In the third distance function, the distance measure between the predicted and observed discharge is applied in the formulation as follows (Sadegh and Vrugt 2013):

$$\rho(S(Q_{obs}), S(Q_{\text{comp}}(\theta))) = \frac{1}{M}\left(M - \sum_{j=1}^{M} I(|Q_{opj}(\theta) - Q_{obsj}| \le \delta_j)\right) \tag{15}$$

where $I(|Q_{\text{comp}j}(\theta) - Q_{obsj}| \le \delta_j)$ is an indicator function; δ_j = effective observation error equivalent to $\delta_j = \alpha \times \hat{\sigma}_{Q_{obsj}}$ with $\alpha = 2$; and $\hat{\sigma}_{Q_{obsj}}$ = error deviation of the measurements and is calculated as follows (Vrugt et al. 2005):

$$\hat{\sigma}_{Q_{obsj}} = \sqrt{\binom{2l}{l}^{-1} (\Delta^l Q_{obsj})^2} \tag{16}$$

where Δ^l = difference sequence of order l. This estimator works well for daily and hourly discharge data.

To estimate posterior parameters and the model output uncertainty with the DREAM$_{(\text{ABC})}$ algorithm, the multipliers (μ) are also introduced to rainfall events, and parallel computing is performed to reduce the computational time of the algorithms. The perception of detecting model structural deficiencies increases by considering the rainfall errors and using rainfall multipliers for uncertainty assessment with the DREAM$_{(\text{ABC})}$ algorithm (Vrugt and Sadegh 2013). In this paper, the DREAM$_{(\text{ABC})}$ algorithms based on the first distance function to the third distance function are shown as DREAM$_{(\text{ABC})-1}$, DREAM$_{(\text{ABC})-2}$, and DREAM$_{(\text{ABC})-3}$, respectively.

Criteria for Comparison of Performance of the Algorithms

Three categories of criteria introduced in the following sections, including computational efficiency of the algorithms, the uncertainty of posterior parameters, and model prediction uncertainty, are used to compare performances of the uncertainty analysis algorithms.

Computational Efficiency of the Algorithms

The final small value (ε), acceptance rate, AR (%), and the number of function evaluations (FE) required to identify the posterior parameters can be used to obtain the computational effectiveness of the algorithms. The results obtained for each algorithm are described in the "Results and Discussion" section.

Uncertainty of the Posterior Parameter

In this paper, the parameter uncertainty of the model is assessed separately for each flood event of the calibration period. The HEC-HMS model is separately run for each flood event and algorithm. Following the search convergence to a stationary distribution, the Kolmogorov–Smirnov criterion (D) (Massey 1951), coefficient of variation (CV), and the shapes of the posterior distributions of the parameters can be used for evaluating the uncertainty and defining the magnitude of the sensitivity of the parameters in each of the flood events and algorithms. The coefficient of variation (CV) criterion represents the extent of the variability of data in a sample in relation to the mean value. Smaller CVs indicate less variability of data, meaning that the data are scattered around the mean of the data and have less uncertainty. This confirms a higher sensitivity of parameters to the performance of the model. Parameters with smaller CVs can be set as effortlessly recognizable parameters (Wang et al. 2005; Shafiei et al. 2014).

The Kolmogorov–Smirnov (K-S) statistic quantifies a distance between cumulative distribution functions (CDFs) of the posterior parameters, $F_1(x)$, and the prior CDFs of parameters, $F_2(x)$.

Statistics of the K-S test are defined as:

$$D = \max(|F_1(x) - F_2(x)|) \tag{17}$$

In the K-S test, a higher D implies that the posterior parameters are more different from uniform prior parameters and the corresponding parameter is less uncertain. To apply the K-S test in the present study, after reaching the convergence, the last 20% of the posterior parameters with less uncertainty in the calibration period are obtained for each algorithm. Then, prior ranges were used to generate the prior CDFs of the parameters (Table 3). A detailed description of the K-S test is given by Nourali et al. (2016).

Uncertainty of the Model Prediction

The last 20% of the posterior parameters are used to create output predictions for estimating the uncertainty bounds of parameters. The results are then put for analysis, and 95% of the uncertainty bands are calculated at the 2.5 and 97.5 percentiles of the simulated data (Nourali et al. 2016). The indicator values, called P-factors

[the percentage of observations covered in the 95% prediction intervals (95PPU)], and R-factors (the relative average width of 95PPU) are used for evaluating prediction uncertainty (Szcześniak and Piniewski 2015; Camargos et al. 2018; Yang et al. 2018). The P-factor of about 100% and the R-factor close to zero indicate ideal parameter ranges and the best performance of the algorithms (Abbaspour 2011; Nourali et al. 2016). Better performance of the algorithms occurs when a realistic estimate is achieved, and the uncertainty band (R-factor) is as narrow as possible. The percentage of observations bracketed by the prediction uncertainty ranges (P-factor) confirms that the predictions are realistic and reasonable (Yang et al. 2008).

Mean posterior values of the RMSE criterion, Nash-Sutcliffe efficiency index (Nash and Sutcliff 1970), and Kling-Gupta efficiency (KG) (Gupta et al. 2009) are also applied for analyzing the difference in the observed flows and simulated flows obtained from the posterior parameter sets and to compare the algorithms [Eqs. (18)–(20)]

$$\text{RMSE} = \sqrt{\frac{\sum_{j=1}^{M} (Q_{\text{comp}j} - Q_{obsj})^2}{n}} \tag{18}$$

$$NS = 1 - \frac{\sum_{j=1}^{M} (Q_{\text{comp}j} - Q_{\text{obsj}})^2}{\sum_{j=1}^{M} (Q_{\text{obsj}} - \bar{Q}_{\text{obs}})^2} \tag{19}$$

$$KG = 1 - \sqrt{(cc-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \tag{20}$$

$$\alpha = \frac{\bar{Q}_{\text{comp}}}{\bar{Q}_{obs}} \tag{21}$$

$$\beta = \frac{\sigma_{Q_{\text{comp}}}}{\sigma_{Q_{obs}}} \tag{22}$$

where $\bar{Q}_{\text{comp}}$ and $\bar{Q}_{obs}$ = mean simulated and observed values, respectively; $\sigma_{Q_{\text{comp}}}$ = standard deviation of the simulated values; $\sigma_{Q_{\text{obs}}}$ = standard deviation of the observed values; and cc = Pearson correlation coefficient between the observed and simulated discharge. Satisfactory results of the simulation can be obtained when the KGE and the NS tend to 1 and the RMSE is close to 0.

Study Area and Hydrologic Modeling

In this study, a representative basin, Tamar (between 55°30′−56°04′E longitude and 37°24′−37°47′N latitude) situated in the Gorganroud River basin, Golestan Province, Iran, is used as a case study. This basin is selected as the study area. The climate of the region is semiarid, experiencing occasional flash floods that inflict significant damages, which necessitates appropriate planning for flood control management in the basin. The basin has a total area of 1,530 km^2, and more reliable data are readily available [see more detailed information of the basin in the study by Nourali et al. (2016)]. The mean areal rainfall is determined by Thiessen polygons for each storm event based on a sufficient number of rain gauges [Fig. 1(a), Table 1]. The discharge gauging station is situated at the outlet of the basin (see the hydrometric station in Fig. 1). Among the data of all observed stations, five flood events are chosen in this research (Table 2). A calibration is carried out for the first four events, and the last event is used for validation.

In the present study, the HEC-HMS hydrologic model is used, which is described in detail by Nourali et al. (2016). The model alters rainfall to runoff at the watershed outlet through the application of explicit process descriptions of the SCS-CN loss model and the Clark hydrograph method (to estimate the losses and transform rainfall to runoff, respectively) and the Muskingum method for flow routing. Among the parameters of the hydrologic model, the curve number (CN), initial loss coefficient (a), regional value (Cs), and Muskingum method parameter (Xm) are used in the prediction. Table 3 shows the parameters of the hydrologic model (24 parameters for 7 subbasins and three reaches) with their range of possible values. The correlation parameter (ρ) and the rainfall multiplier (μ) are also added to the hydrologic model parameters. As shown in Table 3, the range of parameter ρ is considered to be between 0 and 1, and the range of 0.25–0.5 is also utilized for the rainfall multipliers (μ) (Vrugt et al. 2009b; Nourali et al. 2016). For running the HEC-HMS model, the parameters and data of the rainfall-runoff model and each of the algorithms are imported, and each algorithm is linked to the HEC-HMS model in the MATLAB version 9.2 (R2017a) software. The posterior parameters with less uncertainty derived from the calibration period are used in the validation phase that will be described in the "Computational Efficiency Analysis" section. To use the posterior values of the rainfall multipliers (μ) in the validation phase, the model is run and calibrated using the prior uncertainty ranges (Table 3) for the fifth flood event (the flood event used in the validation phase). After confirming the convergence, the last 20% of the obtained posterior values of the rainfall multipliers (μ) are applied in the validation phase for each of the algorithms.

Results and Discussion

Evaluation of Residual Error Assumptions

Analysis of assumptions of Gaussian and independent errors for the first three flood events [see figure 2 in the study by Nourali et al. (2016)] and the fourth flood event (Fig. S1) revealed a violation of the residual error assumptions. In the present study, Fig. 2 shows the autocorrelation function plot of residuals for flood events used in the calibration phase before using the AR model and the Box–Cox transformation.

A step-by-step approach, beginning with AR(1), is applied to remove the autocorrelation of residuals, and the AR(1) model is incorporated into a formulation of the likelihood function after the Box–Cox transformation for the flood events used in the calibration phase [Eq. (1)]. The homoscedasticity of residual errors is also induced using the Box–Cox transformation.

As shown in Fig. 3, temporal autocorrelation of residuals at lag-1 is smaller compared to those of Fig. 2 for the first and third flood events. These results confirm that DREAM$_{(ZS)}$ has worked well with the AR(1) model, and this likelihood function is suitable to remove the autocorrelation of the errors for these mentioned flood events. Visual assessment of the autocorrelation function plot showed that the AR(1) model is unable to remove autocorrelation from the errors for the second and fourth flood events (Fig. S2); therefore, the AR(2) model is applied for further analysis. The trials conducted using this model demonstrated that the autocorrelation of the residuals had been well removed by the AR(2) model for the second and fourth flood events (Fig. 3), and this model is suitable to be used in the likelihood function of the DREAM$_{(ZS)}$ algorithm. After choosing appropriate AR models described previously for the flood events, these models are used as the likelihood functions applied in the DREAM$_{(ZS)}$ algorithm, and the results obtained from running the algorithm are analyzed, which will be described in the next sections.

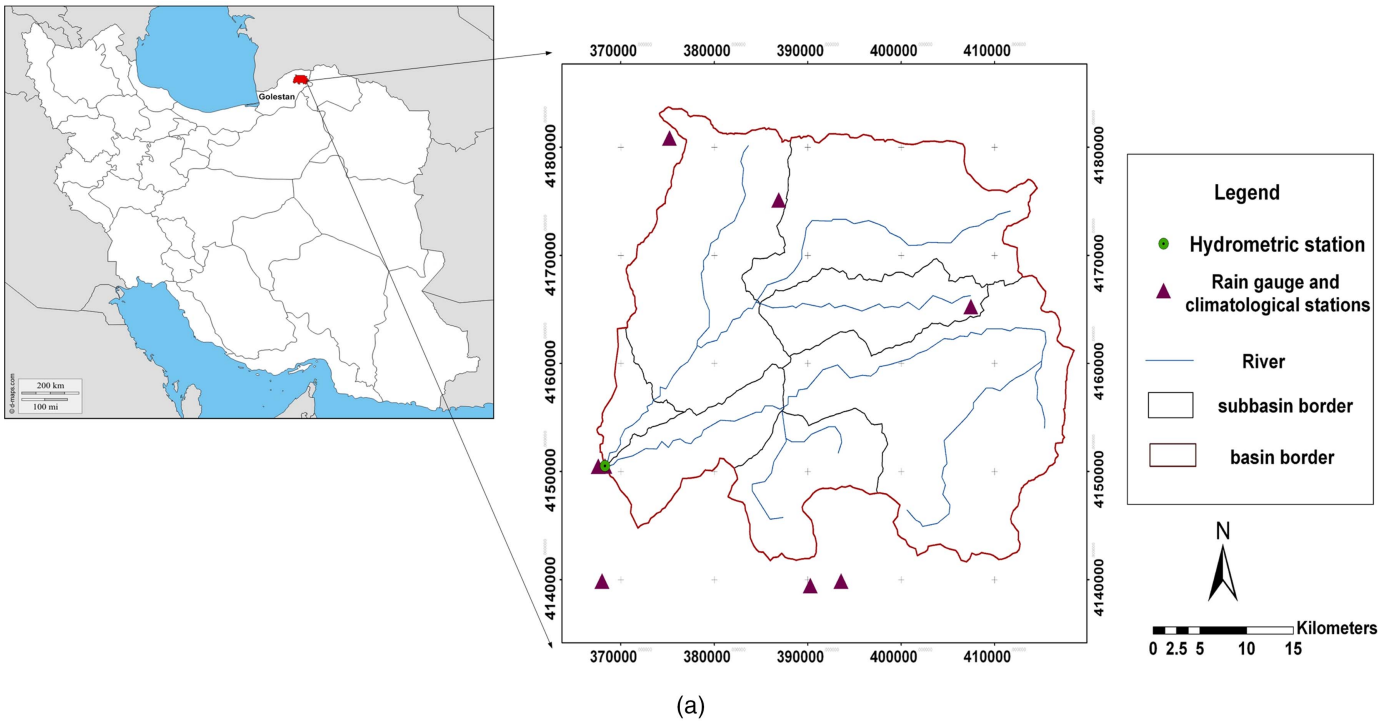

Fig. 1. (a) Study area and its subbasin (reprinted with permission from D-maps 2020); and (b) schematic representation of Tamar basin model in HEC-HMS (reprinted from Journal of Hydrology, Vol. 540, M. Nourali, B. Ghahraman, M. Pourreza-Bilondi, and K. Davary, "Effect of formal and informal likelihood functions on uncertainty assessment in a single event rainfall-runoff model," pp. 549–564, © 2016, with permission from Elsevier).

Computational Efficiency Analysis

Table 4 presents results obtained regarding the analysis of the final small value (ε), the acceptance rate, AR (%), and the number of function evaluations for uncertainty analysis algorithms. According to Table 4, the values of tolerance (ε), AR, and FE are different in various flood events for each algorithm, indicating that the computational efficiency of the algorithms depends on flood event characteristics and the distance function used in the algorithm. To obtain the results in Table 4 accurately, various $\text{DREAM}_{(\text{ABC})}$ algorithms are compared to each other, and the best algorithm is selected, which is compared to the $\text{DREAM}_{(\text{ZS})}$ algorithm. In the first step, the various $\text{DREAM}_{(\text{ABC})}$ algorithms are compared to each other as follows.

A comparison of $\text{DREAM}_{(\text{ABC})}$ algorithms presented in Table 4 illustrates that the values of the final epsilon (ε) used in $\text{DREAM}_{(\text{ABC})-3}$ are higher compared to those of $\text{DREAM}_{(\text{ABC})-1}$ and $\text{DREAM}_{(\text{ABC})-2}$ for each flood event. According to this result,

Table 1. General information of meteorological and hydrometric stations

Station	Coordinate of stations			Station type
	Longitude (UTM)	Latitude (UTM)	Elevation (m)	
Tamar	367584	4150504	132	Hydrometeorology
Golidagh	407429	4165341	1,000	Rain gauge
Ghavijigh	375224	4180871	500	Rain gauge
Gharnagh	386892	4175154	500	Rain gauge
Ghoochmaz	367997	4139898	160	Rain gauge
Tangrah	390288	4139471	330	Rain gauge
Golestan national park	393568	4139903	460	Hydrometeorology
Tamar	367911	4149683	132	Hydrometry

Table 2. Characteristics of selected flood events in Tamar basin

Event	Date	Period	Peak flow (m^3/s)	Duration (h)
1	September 19, 2004	Calibration	128	20
2	May 6, 2005	Calibration	299	30
3	August 9, 2005	Calibration	783	19
4	November 8, 2006	Calibration	68.4	23
5	October 8, 2005	Validation	120	13

Table 3. Uniform prior uncertainty ranges of the HEC-HMS model parameters, the first-order correlation coefficient (ρ), and rainfall multipliers (μ)

Parameter	Location	Lower limit	Upper limit
Curve number (CN)	Subbasin-1[a] (S1)	65	89
	Subbasin-2 (S2)	68	93
	Subbasin-3 (S3)	68	93
	Subbasin-4 (S4)	66	90
	Subbasin-5 (S5)	64	87
	Subbasin-6 (S6)	69	93
	Subbasin-7 (S7)	71	96
Loss coefficient (a)	7 subbasins	0.035	0.45
Regional value (Cs)	7 subbasins	0.2	0.65
Muskingum routing (Xm)	3 reaches	0.2	0.5
First-order correlation coefficient (ρ)	—	0	1
Rainfall multipliers (μ)	7 subbasins	0.25	2.5

[a]For the definition of subbasins, see Fig. 1.

applying the signature measures in the $DREAM_{(ABC)-1}$ and $DREAM_{(ABC)-2}$ algorithms causes more of an effect of the $DREAM_{(ABC)}$ algorithm on sampling. The number of function evaluations of $DREAM_{(ABC)-3}$ required to sample the posterior parameters is more than the $DREAM_{(ABC)-1}$ and $DREAM_{(ABC)-2}$ algorithms for most of the flood events. This result shows that $DREAM_{(ABC)-3}$, which uses only the residual-based summary statistic, has a lower convergence speed to explore the posterior distribution than the other $DREAM_{(ABC)}$ algorithms. In contrast, the $DREAM_{(ABC)-1}$ and $DREAM_{(ABC)-2}$ algorithms, which use the signatures based on the characteristic of a runoff hydrograph, need a smaller number of FE for most of the flood events. Moreover, the $DREAM_{(ABC)-1}$ and $DREAM_{(ABC)-2}$ algorithms exhibit a higher value of AR than those of the $DREAM_{(ABC)-3}$ algorithm for most of the flood events in the calibration phase. This result shows that the signatures used in $DREAM_{(ABC)-1}$ and $DREAM_{(ABC)-2}$ are relatively powerful statistics and, thus, display sufficient diagnostic capability to sufficiently constrain the model parameter space. Sadegh and Vrugt (2014) also pointed out that the summary metrics should be chosen properly to exhibit useful diagnostic power.

In the next step, the $DREAM_{(ZS)}$ is compared to $DREAM_{(ABC)-1}$ and $DREAM_{(ABC)-2}$ as suitable $DREAM_{(ABC)}$ algorithms in terms of computational efficiency. The results in Table 4 demonstrate that $DREAM_{(ZS)}$, which uses the AR model, has a lower acceptance rate (AR, %) for most of the flood events and a smaller convergence speed than $DREAM_{(ABC)-1}$ and $DREAM_{(ABC)-2}$ algorithms to explore the posterior distribution. The $DREAM_{(ZS)}$ increases the computational budget because of a larger number of FE.

The results of this section show that $DREAM_{(ABC)-1}$ and $DREAM_{(ABC)-2}$ algorithms are more efficient than $DREAM_{(ZS)}$ in sampling the target distribution. This result shows that the use of the $DREAM_{(ABC)}$ with proper summary statistics is preferred to the $DREAM_{(ZS)}$ algorithm for high-dimensional parameter spaces.

Analyzing Uncertainty of the Posterior Parameters

After running the model in the calibration period, the shapes and coefficient of variation (CV) of the last 20% of posterior parameters after convergence are inspected. The results showed that the posterior parameters obtained from $DREAM_{(ABC)-1}$ and $DREAM_{(ABC)-3}$

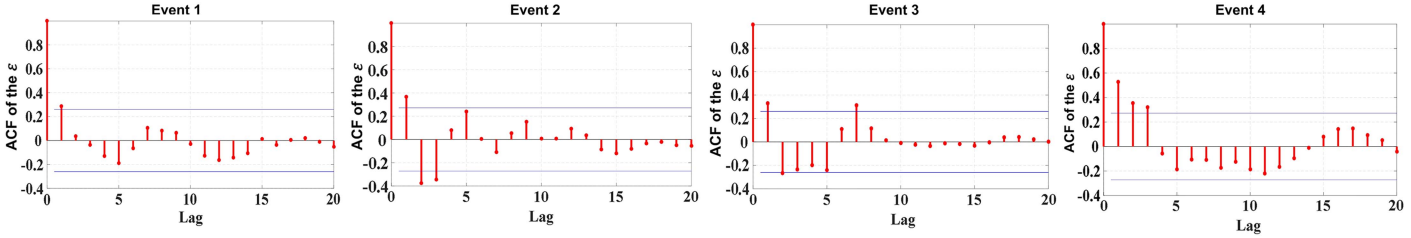

Fig. 2. Auto-correlation function (ACF) plot of residuals for the first four flood events employed in calibration period under the likelihood function obtained from assumptions of Gaussian and independent errors [likelihood function L5 in the study by Nourali et al. (2016)]. The horizontal lines in each of the plots represent the 95% confidence intervals of the time series of uncorrelated and normally distributed residuals.

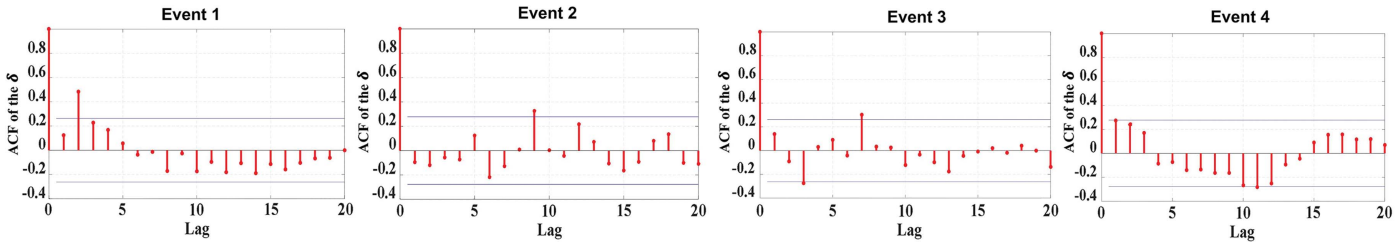

Fig. 3. Auto-correlation function (ACF) plot of residuals after Box–Cox and AR transformation for the first four flood events employed in the calibration period under the likelihood function [$L(\theta|O)$, Eq. (1)] of this study. The horizontal lines in each of the plots represent the 95% confidence intervals of the time series of uncorrelated and normally distributed residuals.

Table 4. Computational efficiency of $DREAM_{(ABC)-1}$, $DREAM_{(ABC)-2}$, $DREAM_{(ABC)-3}$, and $DREAM_{(ZS)}$

	Flood event											
	1			2			3			4		
Sampling algorithms	ε	AR (%)	Function evaluations	ε	AR (%)	Function evaluations	ε	AR (%)	Function evaluations	ε	AR (%)	Function evaluations
$DREAM_{(ABC)-1}$	0.1	26.67	7,920	0.1	45.42	5,520	0.15	77.92	2,160	0.1	51.25	8,880
$DREAM_{(ABC)-2}$	0.15	88.75	9,600	0.3	61.25	5,100	0.35	56.67	3,600	0.3	80.33	6,000
$DREAM_{(ABC)-3}$	0.6	65.83	6,960	0.5	78.75	12,000	0.6	40.83	5,400	0.8	45	12,000
$DREAM_{(ZS)}$	—	28.57	20,000	—	20.33	13,500	—	21.26	21,000	—	31.6	72,000

Note: This includes the final epsilon value, acceptance rate, AR (%), and the total number of function evaluations needed for posterior exploration.

have less uncertainty for the fourth event (Table 2). The posterior parameters obtained from the $DREAM_{(ABC)-2}$ algorithm for the third event (Table 2), and the posterior parameters obtained from the $DREAM_{(ZS)}$ algorithm for the first event, are also less uncertain (not shown because of the limited space). The posterior histograms, the results of the Kolmogorov–Smirnov test statistic (D), and the coefficient of variation (CV) of the last 20% of the parameters mentioned previously, which have less uncertainty, are presented in Figs. 4–6. Moreover, these posterior parameters are also used in the validation phase. The results of the running and a prediction of the model in the calibration and validation phases are presented in Fig. 7, which will be discussed in the "Assessing Model Prediction Uncertainty." According to the results of the posterior histograms, Kolmogorov–Smirnov test statistic (D), and CV value of the posterior parameters (Figs. 4–6), the sensitivity of the parameters is different for different sampling algorithms. Therefore, a careful selection is required to define fitness functions and sampling algorithms to explore the behavioral model space efficiently. A comparison between the posterior histograms derived from the algorithms (Fig. 4) shows that the shape of the distributions and the posterior ranges are not the same (normally distributed or a tendency to ascribe the highest probability mass at the lower and higher bounds) in $DREAM_{(ABC)-1}$ and $DREAM_{(ABC)-2}$ for most of the parameters. The width of the posterior distribution derived from the $DREAM_{(ABC)-1}$ and $DREAM_{(ABC)-2}$ algorithms is still considerably smaller than the posterior ranges obtained from the $DREAM_{(ABC)-3}$ and $DREAM_{(ZS)}$ for most of the parameters (Fig. 4), and their posterior parameters are more sensitive and have less uncertainty in most of the parameters. Moreover, this result confirms that the distance function used in $DREAM_{(ABC)-3}$ is a relatively weak metric having insufficient diagnostic power and extracting limited information from discharge data.

The Kolmogorov–Smirnov test statistic (D) shown in Fig. 5 confirms this result, and thus, the posterior parameters corresponding to $DREAM_{(ABC)-1}$ and $DREAM_{(ABC)-2}$ algorithms show a higher D and less uncertainty than those obtained from the other algorithms for most of the parameters. Table 5 and Fig. 6 show the mean value and CV of the posterior parameter sets used in Fig. 4, respectively. As Fig. 6 shows, most of the parameters have smaller values of the coefficient of variation in $DREAM_{(ABC)-1}$ and $DREAM_{(ABC)-2}$ algorithms. Therefore, most of the parameters obtained from $DREAM_{(ABC)-1}$ and $DREAM_{(ABC)-2}$ are well defined through calibration. These parameters are more reliable and have less uncertainty.

These results show that creating parameters with less uncertainty depends on the distance function and signature measures used in the $DREAM_{(ABC)}$ algorithm. The summary metrics used in $DREAM_{(ABC)-1}$ and $DREAM_{(ABC)-2}$ are relatively powerful statistics and so display sufficient diagnostic capability to constrain the model parameter space adequately. These signature measures are based on the characteristic of a runoff hydrograph, resulting from running a single-event model. However, diagnostic capability significantly increases using more signatures related to characteristic behaviors of the basin in the $DREAM_{(ABC)}$ analysis (Sadegh and Vrugt 2014). Smaller values of CV less than 10% for the curve number (CN) parameter derived from $DREAM_{(ZS)}$, and the various $DREAM_{(ABC)}$ algorithms presented in Fig. 6, imply that the posterior limits are smaller than the uniform prior ranges; as such, this parameter is the most sensitive parameter to use objective measures. A similar result has also been reported by Nourali et al. (2016).

To inspect the posterior parameters more, correlation coefficients between the parameters are considered. A relatively small correlation between posterior parameters (results not shown) confirms that observed flows involve adequate information for estimating these parameters. The posterior value of the correlation coefficient between the curve number (CN) and loss coefficient (a) (not shown) is more than 0.75 in most of the subbasins, implying that one of these parameters could be fixed prior to the hydrologic model calibration.

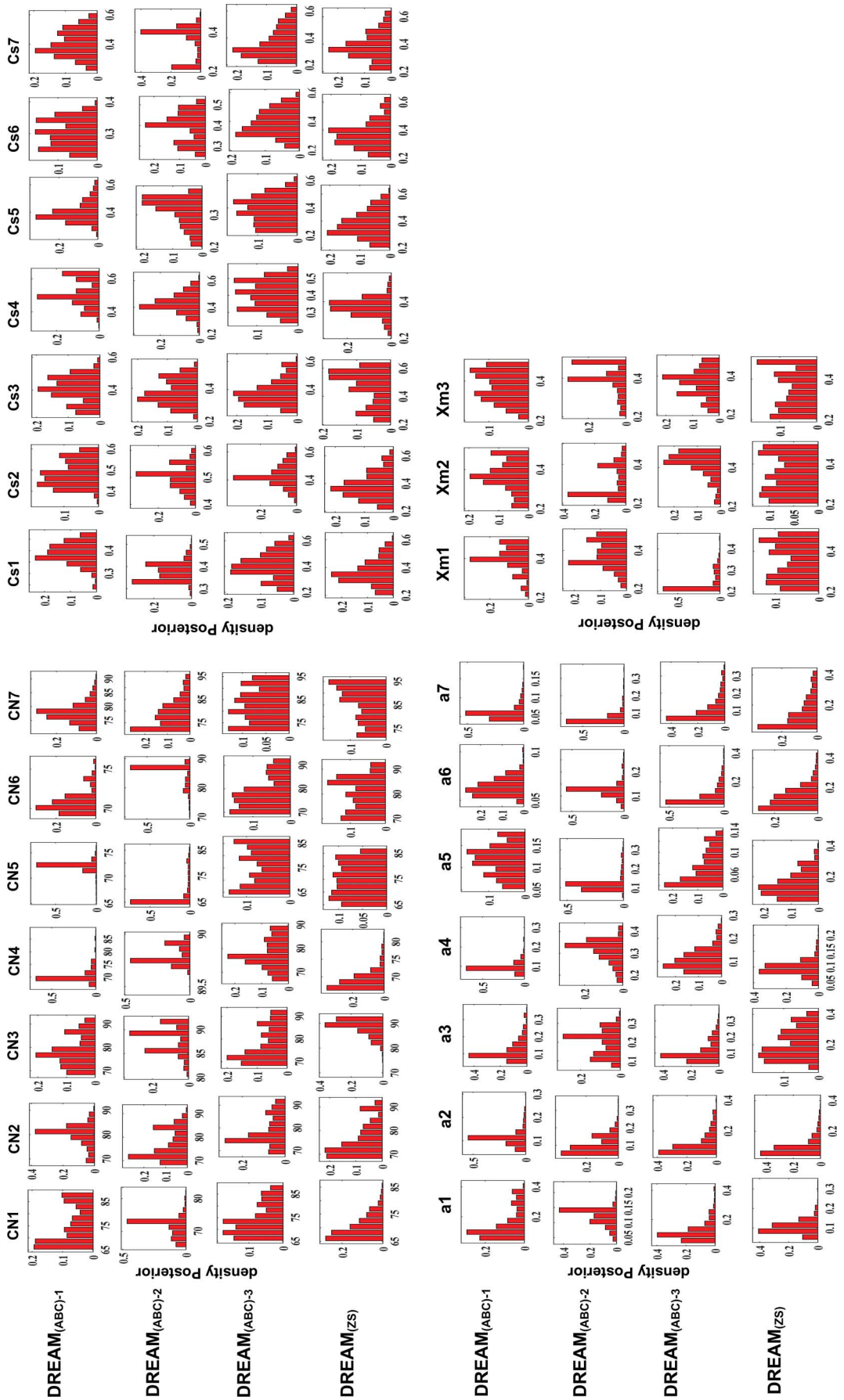

Fig. 4. Posterior histograms of hydrologic model parameters using $\text{DREAM}_{(\text{ABC})}$ and $\text{DREAM}_{(\text{ZS})}$ sampling algorithms.

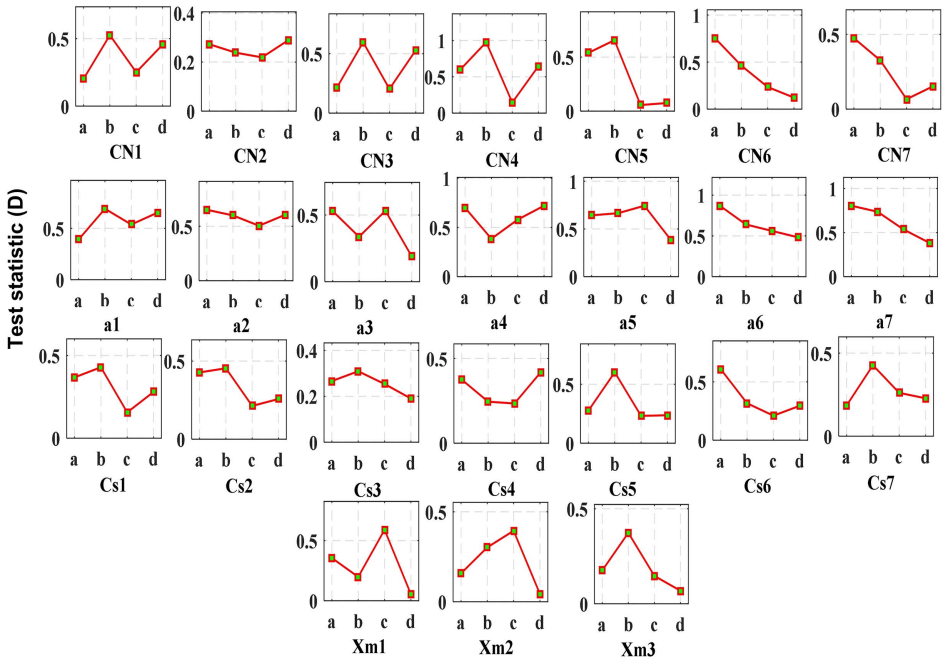

Fig. 5. Kolmogorov–Smirnov test statistic (D) of hydrologic model parameters under different sampling algorithms. The $\text{DREAM}_{(\text{ABC})-1}$, $\text{DREAM}_{(\text{ABC})-2}$, $\text{DREAM}_{(\text{ABC})-3}$, and $\text{DREAM}_{(\text{ZS})}$ sampling algorithms are shown as a, b, c, and d, respectively.

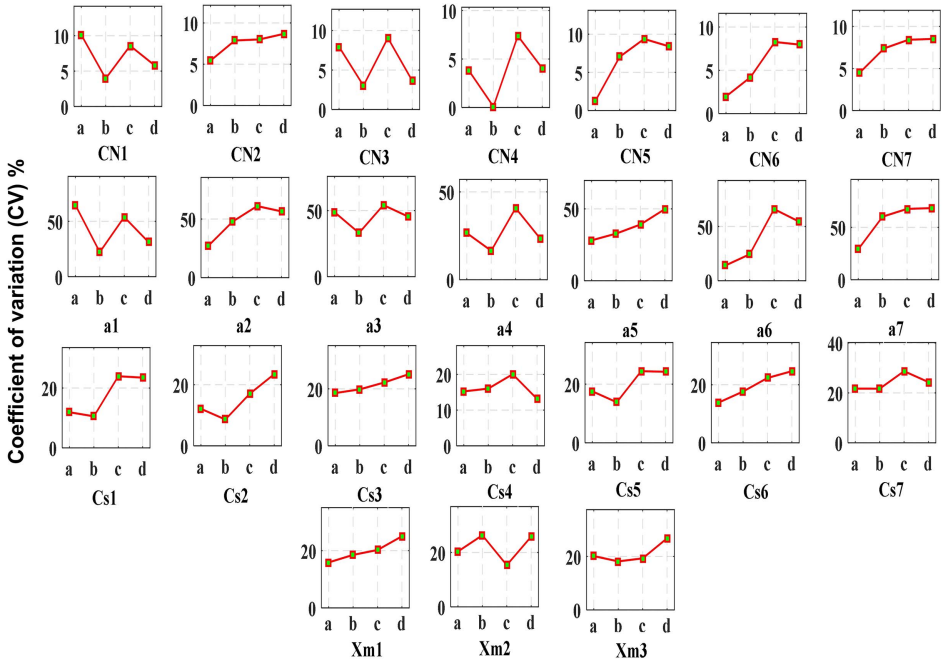

Fig. 6. Coefficient of variation of posterior distributions derived from different sampling algorithms. The $\text{DREAM}_{(\text{ABC})-1}$, $\text{DREAM}_{(\text{ABC})-2}$, $\text{DREAM}_{(\text{ABC})-3}$, and $\text{DREAM}_{(\text{ZS})}$ sampling algorithms are shown as a, b, c, and d, respectively.

In this research, the rainfall forcing error is considered in the parameter estimation problem by $\text{DREAM}_{(\text{ABC})}$ and $\text{DREAM}_{(\text{ZS})}$ samplers, and the result illustrates that rainfall uncertainty affects the streamflow prediction. The mean of the posterior parameters of the multipliers (μ) obtained from $\text{DREAM}_{(\text{ABC})}$ and $\text{DREAM}_{(\text{ZS})}$ samplers ranged between 0.3 and 2.3 in the subbasins, indicating a difference between observed rainfall and the rainfall obtained from the streamflow data.

The mean posterior values of $\emptyset_1$ (ranging between 0.3 and 0.6) and $\emptyset_2$ (ranging between 0.15 and 0.33) obtained from $\text{DREAM}_{(\text{ZS})}$ confirmed the existence of the autocorrelation of the model residuals; hence, AR(1) and AR(2) models were used in $\text{DREAM}_{(\text{ZS})}$ to remove the correlation of the errors.

Assessing Model Prediction Uncertainty

The investigation results of the model prediction uncertainty using P-factor and R-factor values, and the RMSE, KG, and NS criteria, are presented in this section. Through an inspection of these criteria, the performance of the different $\text{DREAM}_{(\text{ABC})}$ algorithms,

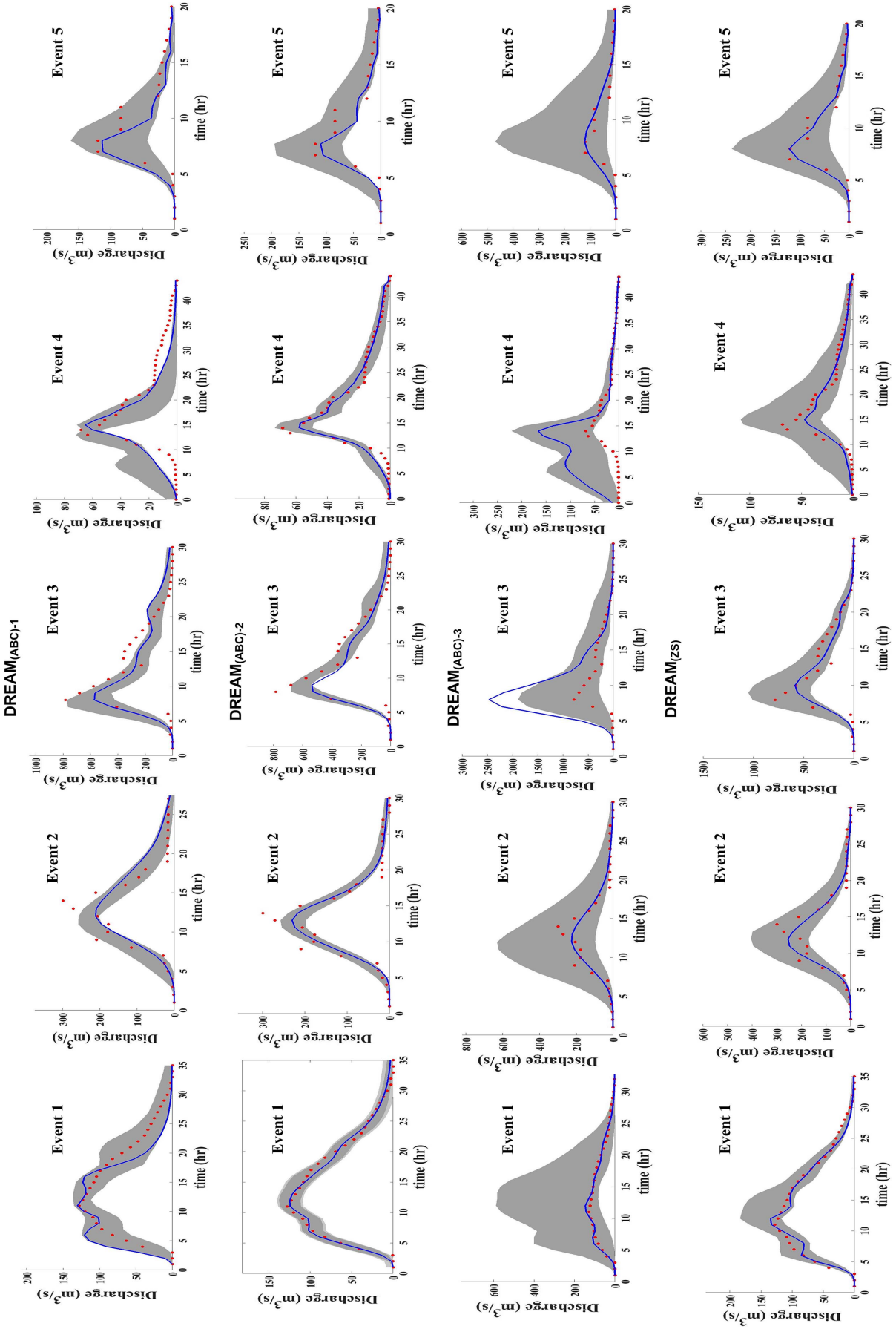

Fig. 7. Observed stream flows (dots), best simulation (solid line), and 95% model prediction uncertainty bands associated with the posterior distribution of the parameter estimates (dark shaded region).

Table 5. Mean values of posterior parameters derived from different sampling algorithms

Parameters	Sampling algorithms			
	$DREAM_{(ABC)-1}$	$DREAM_{(ABC)-2}$	$DREAM_{(ABC)-3}$	$DREAM_{(ZS)}$
CN1	74.82	71.42	74.07	70.08
CN2	80.01	76.62	78.45	76.12
CN3	78.31	87.97	77.72	88.60
CN4	70.59	89.79	77.35	69.07
CN5	72.40	67.46	75.44	74.90
CN6	70.81	85.89	78.12	80.00
CN7	77.66	78.64	82.76	85.60
a1	0.15	0.12	0.12	0.10
a2	0.11	0.10	0.11	0.10
a3	0.12	0.19	0.11	0.21
a4	0.10	0.31	0.11	0.10
a5	0.12	0.12	0.07	0.15
a6	0.06	0.13	0.10	0.13
a7	0.06	0.07	0.10	0.14
Cs1	0.39	0.37	0.40	0.37
Cs2	0.48	0.50	0.43	0.38
Cs3	0.41	0.38	0.38	0.46
Cs4	0.51	0.43	0.40	0.37
Cs5	0.40	0.31	0.39	0.38
Cs6	0.30	0.39	0.39	0.37
Cs7	0.40	0.36	0.35	0.38
Xm1	0.40	0.38	0.23	0.35
Xm2	0.37	0.31	0.41	0.35
Xm3	0.37	0.40	0.36	0.35

including $DREAM_{(ABC)-1}$, $DREAM_{(ABC)-2}$, and $DREAM_{(ABC)-3}$, is compared to each other. The results presented in Fig. 7 demonstrate that the 95% uncertainty ranges of posterior parameters derived from $DREAM_{(ABC)-3}$, which are indicated by the dark gray region, cover a higher ratio of the observations than those obtained from the other $DREAM_{(ABC)}$ algorithms (between 40% and 100%) for most of the flood events. However, the 95% posterior uncertainty ranges obtained from the $DREAM_{(ABC)-3}$ algorithm have a much higher *R*-factor value than those of the other $DREAM_{(ABC)}$ algorithms, and these uncertainty ranges indicate a large parameter uncertainty, as shown in Figs. 7 and 8. Therefore the distance function used in the $DREAM_{(ABC)-3}$ algorithm is a weak metric for model prediction uncertainty purposes, and the $DREAM_{(ABC)-3}$ does not represent a proper performance.

A comparison of the *P*-factor and *R*-factor values of the $DREAM_{(ABC)-1}$ and $DREAM_{(ABC)-2}$ algorithms in Fig. 8 shows that the ratio of the observations captured by the 95% confidence interval (*P*-factor) is almost the same (between 23% to 89%) for $DREAM_{(ABC)-1}$ and $DREAM_{(ABC)-2}$ algorithms. But $DREAM_{(ABC)-2}$ represents a lower R-factor value for most of the flood events (Fig. 8). A low thickness of the confidence interval for $DREAM_{(ABC)-2}$ shown in Fig. 7 confirms this result; therefore, $DREAM_{(ABC)-2}$ represents a better performance than the $DREAM_{(ABC)-1}$ algorithm. This finding shows that the summary metrics used in $DREAM_{(ABC)-2}$ are proper metrics for single-event modeling.

Now the $DREAM_{(ABC)-2}$, which represents a better performance than the other $DREAM_{(ABC)}$ algorithms, is compared to the $DREAM_{(ZS)}$ algorithm. According to the results, reasonable values of the *P*-factor were obtained from $DREAM_{(ABC)-2}$ (higher than 50%) for most of the flood events. The performance of the $DREAM_{(ABC)-2}$ algorithm can also be improved by considering the observed discharge data error and refinement of the underlying equations of the model. These improvements may cause an increase in the *P*-factor values. The 95% prediction uncertainty bounds derived from $DREAM_{(ZS)}$ cover a higher percentage of the measured data in comparison with the $DREAM_{(ABC)-2}$ algorithm (between 63% and 87%) for most of the flood events. This means that the 95% confidence intervals derived from $DREAM_{(ZS)}$ track the observed data better. This is because of the AR model applied in $DREAM_{(ZS)}$ for removing the time dependence of residuals caused by the input error and/or model structure error. But according to the results presented in the "Analyzing Uncertainty of the Posterior Parameters" section, the posterior distributions obtained from $DREAM_{(ZS)}$ are rather large and cover a wide region of their uniform prior ranges (Table 2). This means that the likelihood function used in $DREAM_{(ZS)}$ displays weak diagnostic capability to constrain parameter space sufficiently, whereas the metrics used in $DREAM_{(ABC)-2}$ are sufficient and extract much information from the observed flows. This result is confirmed by comparing *R*-factors of $DREAM_{(ZS)}$ with those of $DREAM_{(ABC)-2}$. The *R*-factor values obtained from the $DREAM_{(ZS)}$ algorithm are higher than those of the $DREAM_{(ABC)-2}$ algorithm (Fig. 8); therefore, $DREAM_{(ZS)}$ does not represent a suitable performance regarding model prediction uncertainty compared to $DREAM_{(ABC)-2}$ to achieve the desired parameter uncertainty ranges.

In contrast to the results of this paper described previously, a conclusion drawn in the work by Sadegh and Vrugt (2014) showed that most posterior histograms obtained from DREAM cover a small part of their prior range. This is not a reliable result to compare to the results of $DREAM_{(ABC)}$ because a residual-based Gaussian likelihood function was applied in their study without analyzing and fulfilling common assumptions of the residual distribution.

To assess the model prediction uncertainty and to quantify the goodness-of-fit, the RMSE, KG, and NS criteria could also be applied.

A comparison of the statistical indicators of the different $DREAM_{(ABC)}$ algorithms shows that the values of KGE and NS obtained from $DREAM_{(ABC)-3}$ are close to zero and negative (not shown). Moreover, this algorithm generates the RMSE

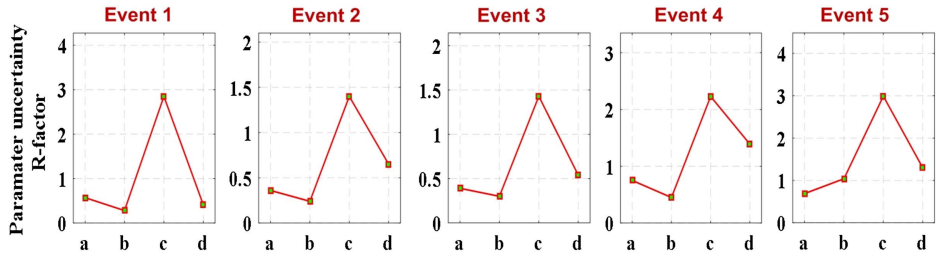

Fig. 8. *R*-factor value for parameter uncertainty assessment corresponding to different sampling algorithms and for flood events in calibration and validation phases. The $DREAM_{(ABC)-1}$, $DREAM_{(ABC)-2}$, $DREAM_{(ABC)-3}$, and $DREAM_{(ZS)}$ sampling algorithms are shown as a, b, c, and d, respectively.

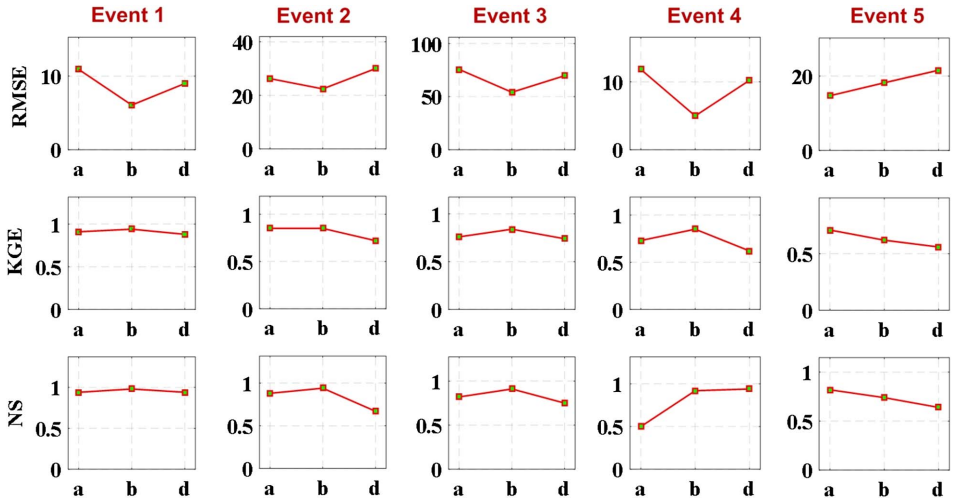

Fig. 9. Comparison of sampling methods performance. The $DREAM_{(ABC)-1}$, $DREAM_{(ABC)-2}$, $DREAM_{(ABC)-3}$, and $DREAM_{(ZS)}$ sampling algorithms are shown as a, b, c, and d, respectively.

value higher than its counterpart from $DREAM_{(ABC)-1}$ and $DREAM_{(ABC)-2}$ (not shown), indicating that the distance function used in $DREAM_{(ABC)-3}$ only extracts partial information from the observed flows (Sadegh and Vrugt 2014). Therefore, $DREAM_{(ABC)-3}$ does not represent a reliable result, whereas $DREAM_{(ABC)-2}$, with a lower RMSE and higher KGE and NS, represents a better performance for most of the flood events in calibration and validation phases (Fig. 9). The performance of the $DREAM_{(ABC)-2}$ and $DREAM_{(ZS)}$ algorithms is depicted in Fig. 9, indicating that the $DREAM_{(ABC)-2}$ algorithm presents a better performance for most of the flood events in the calibration and validation phases. The posterior mean RMSE value obtained from $DREAM_{(ABC)-2}$ was smaller compared to its counterpart from $DREAM_{(ZS)}$ in which the classical Gaussian likelihood function was used with heteroscedastic and independent residual errors, showing that $DREAM_{(ABC)-2}$ extracts much information from observed flows and can provide a better fit of the model to streamflow data.

The results obtained from $DREAM_{(ZS)}$ using the classical Gaussian likelihood function that is based on the assumptions of residual error (homoscedasticity, normality, and independence of the residuals) were reported in our previous work (Nourali et al. 2016). The results from our previous work showed that 95% parameter uncertainty bounds have a small width and cover fewer observations. Moreover, the posterior mean RMSE obtained from $DREAM_{(ZS)}$ had a lower value. According to the results presented in our previous work (Nourali et al. 2016), residual error assumptions were violated, and therefore, the results were unreliable to be used for comparison with the $DREAM_{(ABC)}$ algorithm. Vrugt and Sadegh (2013) and Sadegh and Vrugt (2014) also reported the same results using the DREAM and compared the derived results to their counterpart from ABC and $DREAM_{(ABC)}$. It is noteworthy that analyzing and fulfilling underlying assumptions of the residual distribution and using a suitable likelihood function in $DREAM_{(ZS)}$ results in reliable results that can be applied for comparison with the $DREAM_{(ABC)}$ algorithm.

The results of the present paper showed that the assumptions of residuals were violated, and a likelihood function based on the AR model and the Box–Cox transformations were used in the $DREAM_{(ZS)}$ algorithm. Moreover, the rainfall multipliers were also used in order to consider rainfall errors in the uncertainty assessment of the model. Therefore, the results obtained from the $DREAM_{(ZS)}$ are more reasonable and can be used to compare with those obtained from $DREAM_{(ABC)}$. The results presented in this section convincingly demonstrated the utility of summary statistics used in the $DREAM_{(ABC)-2}$ algorithm for purposes of calibration and model prediction uncertainty.

Diagnostic Model Evaluation

As discussed in the introduction, the $DREAM_{(ABC)}$ algorithm is not only applied for model calibration purposes but also partly provides insights into model malfunctioning, distinguishing which part of

the model needs improvement. Residual-based likelihood functions used in $DREAM_{(ZS)}$ have a limited ability to detect model malfunctioning, as Sadegh and Vrugt (2014) pointed out. The attributes, such as recession characteristics, peak flow, and time to peak, can be used in a diagnostic evaluation (Gupta et al. 2008).

In the present case study, $DREAM_{(ABC)-2}$, which represented a better performance among the other $DREAM_{(ABC)}$ algorithms, was considered to inspect model structural deficiencies. The results presented in the previous section show that two summary metrics used in $DREAM_{(ABC)-2}$, the NRMSE and NTRMSE, seem jointly sufficient for diagnostic analysis of a single-event model. These two metrics tend to put a stronger emphasis on fitting both the high and low flows. However, the 95% streamflow uncertainty ranges of parameters obtained from $DREAM_{(ABC)-2}$ (Fig. 7) do not contain the observed peak flows, and the peak flows are almost underestimated for the second and third flood events. In addition, the observed flows do not appear to be very well tracked by the 95% posterior uncertainty ranges in the recession periods (especially for the second and third flood events). The plots obtained from $DREAM_{(ABC)-1}$ (Fig. 7) also showed similar results in which the signatures were used based on characteristics of a runoff hydrograph.

This finding could be described through the structural insufficiencies in the model equations and the measurement errors in the output data (Taye et al. 2011; Chen et al. 2013); thus, the model is unable to achieve a fit for the computed hydrograph with respect to the observed record in both the recession limb and peak flow concurrently. Hence, it is possible to enhance the model structure and consider output measurement errors in uncertainty assessment to cover a higher ratio of the observations (Laloy et al. 2010; Vrugt and Sadegh 2013). As described in the "Study Area and Hydrologic Modeling" section, the SCS-CN and Clark methods are used in this paper. Other studies indicated that simulated runoff and peak flow were underestimated by the SCS method, and therefore, modified versions of the SCS method are needed to improve runoff predictions (Ponce and Hawkins 1996; Mishra and Singh 1999; Jacobs and Srinivasan 2005; Paudel et al. 2009).

The other finding presented in Fig. 7 is that the simulations in the recession period are unable to track the observed data that is likely caused by the base flow separation technique used in this research. The base flow is not considered in this study because of having an insignificant influence on the flood events (Cunderlik and Simonovic 2004). The straight-line method is used to separate the base flow from the hydrograph; hence, the other baseflow separation methods are proposed to analyze the recession limb and compare results with those obtained from the method used in this study. Besides, more summary statistics are required for achieving more accurate predictions and increasing the *P*-factor value (Vrugt and Sadegh 2013; Sadegh and Vrugt 2014).

This study ignores the measurement errors in output data, and thus, it would be necessary to consider the output data error in the uncertainty assessment of the HMS model for achieving more accurate predictions in future works.

Summary and Conclusions

In the classical likelihood-based fitting methods, which are based on the Bayesian inference procedure, a proper likelihood function is needed to infer parameters and predict the output of hydrological models accurately. A proper likelihood function can be obtained by analyzing and fulfilling the assumption of residual distribution and considering rainfall uncertainty. These reasons point out the difficulty with formulating the likelihood function. Moreover, the likelihood-based fitting method [e.g., the $DREAM_{(ZS)}$ method] lacks the power to detect model structural deficiencies. Therefore, Sadegh and Vrugt (2014), in a study, suggested the $DREAM_{(ABC)}$ algorithm as a likelihood-free inference approach that is capable of estimating the high-dimensional posterior distributions and also partly distinguishes which part of the model is needed to be improved. The continuous models and a residual-based Gaussian likelihood function were used to estimate parameter uncertainty using the DREAM algorithm in the study by Sadegh and Vrugt (2014) without fulfilling common assumptions of the errors, and the results were compared to its counterpart from $DREAM_{(ABC)}$. There is no previous study to consider rainfall uncertainty and analyze and fulfill underlying assumptions of Gaussian distributed error residuals in the likelihood function before comparison between $DREAM_{(ZS)}$ and $DREAM_{(ABC)}$, despite a violation of traditional error assumptions.

In the present study, two uncertainty analysis algorithms, namely, $DREAM_{(ZS)}$ and $DREAM_{(ABC)}$, were applied to analyze the uncertainties of the parameter in a single-event rainfall–runoff model. A likelihood function was used based on the AR model and the Box–Cox transformations in the $DREAM_{(ZS)}$ method due to a violation of the assumptions of the residual error, and three distance functions were also applied in the $DREAM_{(ABC)}$ algorithm. A comparison of the results obtained from the $DREAM_{(ZS)}$ and $DREAM_{(ABC)}$ algorithms can be described as follows:

- The computational efficiency of $DREAM_{(ABC)}$ and $DREAM_{(ZS)}$ algorithms indicates that the values of tolerance (ε), the number of function evaluations (FE), and the acceptance rate (AR) required for sampling the posterior parameters are dependent on the specifications of each event and the distance function applied in the algorithm. Moreover, the $DREAM_{(ABC)-1}$ and $DREAM_{(ABC)-2}$ algorithms, in which the signatures were used, based on characteristics of a runoff hydrograph, are more efficient than the other algorithms due to a higher AR and a smaller number of FE for most of the events.
- The results of the posterior histograms, the Kolmogorov–Smirnov test statistic (D), and the coefficient of variation value of the posterior parameters show that the posterior parameters derived from the $DREAM_{(ABC)-1}$ and $DREAM_{(ABC)-2}$ algorithms are more sensitive and have less uncertainty than those obtained from the other algorithms for most of the parameters. These parameters can be considered as reliable posterior parameters. Therefore, creating parameters with less uncertainty depends on the distance function and signature measures used in the $DREAM_{(ABC)}$ algorithm.
- The previously described results show that $DREAM_{(ABC)-1}$ and $DREAM_{(ABC)-2}$ algorithms represent almost the same performance in terms of computational efficiency and uncertainty of the posterior parameters.
- The results of the model prediction uncertainty derived from $DREAM_{(ABC)}$ and $DREAM_{(ZS)}$ algorithms show that $DREAM_{(ZS)}$ covers a higher percentage of the measured data in comparison with the $DREAM_{(ABC)-2}$ algorithm, but the width of the posterior distribution derived from $DREAM_{(ZS)}$ is larger, and their posterior parameters have more uncertainty in most of the parameters. Moreover, $DREAM_{(ABC)-2}$, with a lower RMSE and higher KGE and NS, represents a better performance for most of the flood events, and the computational efficiency of $DREAM_{(ABC)-2}$ is higher than the $DREAM_{(ZS)}$ algorithm. Therefore, $DREAM_{(ABC)-2}$ is more suitable than the $DREAM_{(ZS)}$ algorithm for the uncertainty assessment of the parameters and calibration purposes of the model, and $DREAM_{(ABC)-2}$ yields reliable parameter estimates. This result also shows that summary metrics used in $DREAM_{(ABC)-2}$ that

tend to put a stronger emphasis on fitting high and low flows are proper metrics, which can be suggested for the uncertainty assessment of single-event models and high-dimensional problems.

- Among $DREAM_{(ABC)}$ algorithms used in this paper, $DREAM_{(ABC)-3}$ does not represent a proper performance. This shows that the distance function applied in the $DREAM_{(ABC)-3}$ algorithm that uses the distance measure between predicted and observed discharge is a weak metric for purposes of model prediction uncertainty and uncertainty assessment of the parameters.
- The results of this paper showed that it is necessary to use appropriate likelihood and distance functions for $DREAM_{(zs)}$ and $DREAM_{(ABC)}$ algorithms, respectively, to compare the performance of these algorithms regarding uncertainties of the parameter and prediction. As a result of our previous work (Nourali et al. 2016), it was found that the AR(1) scheme of error residuals incorporated into a formal likelihood function applied in $DREAM_{(ZS)}$ could provide reliable model parameters. However, according to the results presented in this paper, and despite the use of an acceptable likelihood function in the $DREAM_{(ZS)}$ algorithm, the advantage of the $DREAM_{(ABC)-2}$ based on the likelihood-free inference approach is confirmed to assess the uncertainty in a single-event model and high-dimensional parameter spaces.
- According to the results, summary metrics used in $DREAM_{(ABC)-2}$ seem jointly suitable for the diagnostic analysis of a single-event model malfunctioning. A visual inspection of the uncertainty ranges of the parameter obtained from $DREAM_{(ABC)-2}$ shows that peak flows are almost underestimated for the second and third flood events, and the observed flows have not been covered by uncertainty bands in the recession periods. An improvement of the model structure causes more coverage of observations within the parameter uncertainty range. Therefore, modified versions of the SCS method are needed to improve runoff predictions of the HEC-HMS model, and the other baseflow separation methods are also proposed to be used for an analysis of the recession limb in future works. Moreover, output measurement errors should be considered in an uncertainty assessment of the HMS model. These modifications may yield accurate parameter estimates and more reliable predictions of the $DREAM_{(ABC)-2}$, and the results of their analysis can be compared to those obtained from the methods used in this study for uncertainty assessment of a single-event model.

The methods presented in this paper can also be suggested for performing future research with different climates, and the results obtained from them can be compared with the results of the present study.

Data Availability Statement

The data used in the paper were obtained from the Iran Water Research Institute (IWRI) and Natural Resources and Watershed Management Administration of Golestan, Iran. Some or all data, models, or code generated or used during the study are available from the corresponding author upon reasonable request.

Acknowledgments

The author would like to thank Dr. Jasper A. Vrugt profusely for providing the code for the $DREAM_{(ZS)}$ and $DREAM_{(ABC)}$ algorithms. The author is grateful for the data and information received on the case study from Dr. Mohsen Pourreza-Bilondi.

Supplemental Materials

Fig. S1 and S2 are available online in the ASCE Library (www.ascelibrary.org).

References

Abbaspour, K. C. 2011. *User manual for SWAT-CUP4, SWAT calibration and uncertainty programs*. Duebendorf, Switzerland: Swiss Federal Institute of Aquatic Science and Technology.

Abdessalem, A. B., N. Dervilis, D. Wagg, and K. Worden. 2018. "Model selection and parameter estimation in structural dynamics using approximate Bayesian computation." *Mech. Syst. Signal Process.* 99 (Sep): 306–325. https://doi.org/10.1016/j.ymssp.2017.06.017.

Alazzy, A. A., H. Lü, and Y. Zhu. 2015. "Assessing the uncertainty of the Xinanjiang rainfall-runoff model: Effect of the likelihood function choice on the GLUE method." *J. Hydrol. Eng.* 20 (10): 04015016. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001174.

Asgharpour, S. E., and B. Ajdari. 2011. "A case study on seasonal floods in Iran, watershed of Ghotour Chai Basin." *Procedia Social Behav. Sci.* 19 (Jan): 556–566. https://doi.org/10.1016/j.sbspro.2011.05.169.

Bates, B. C., and E. P. Campbell. 2001. "A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling." *Water Resour. Res.* 37 (4): 937–947. https://doi.org/10.1029/2000WR900363.

Beaumont, M. A. 2010. "Approximate Bayesian computation in evolution and ecology." *Annu. Rev. Ecol. Evol. Syst.* 41 (Dec): 379–406. https://doi.org/10.1146/annurev-ecolsys-102209-144621.

Beaumont, M. A., J. M. Cornuet, J. M. Marin, and C. P. Robert. 2009. "Adaptive approximate Bayesian computation." *Biometrika* 96 (4): 983–990. https://doi.org/10.1093/biomet/asp052.

Beven, K., P. Smith, I. Westerberg, and J. Freer. 2012. "Comment on: Pursuing the method of multiple working hypotheses for hydrological modeling." *Water Resour. Res.* 48 (11): W11801. https://doi.org/10.1029/2012WR012282.

Beven, K., P. J. Smith, and A. Wood. 2011. "On the colour and spin of epistemic error (and what we might do about it)." *Hydrol. Earth Syst. Sci.* 15 (10): 3123. https://doi.org/10.5194/hess-15-3123-2011.

Beven, K. J. 2006. "A manifesto for the enquiringly thesis." *J. Hydrol.* 320 (1): 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007.

Blasone, R. S. 2007. "Parameter estimation and uncertainty assessment in hydrological modelling." Ph.D. thesis. Institute of Environment and Resources, Technical Univ. of Denmark.

Blasone, R. S., J. A. Vrugt, H. Madsen, D. Rosbjerg, B. A. Robinson, and G. A. Zyvoloski. 2008. "Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov chain Monte Carlo sampling." *Adv. Water Resour.* 31 (4): 630–648. https://doi.org/10.1016/j.advwatres.2007.12.003.

Box, G., and D. Cox. 1964. "An analysis of transformations." *J. R. Stat. Soc. B* 26 (2): 211–252. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x.

Burr, T., and A. Skurikhin. 2013. "Selecting summary statistics in approximate Bayesian computation for calibrating stochastic models." *Biomed. Res. Int.* 2013 (Jan): 1–10. https://doi.org/10.1155/2013/210646.

Camargos, C., S. Julich, T. Houska, M. Bach, and L. Breuer. 2018. "Effects of input data content on the uncertainty of simulating water resources." *Water* 10 (5): 621. https://doi.org/10.3390/w10050621.

Chen, X., T. Yang, X. Wang, C. Xu, and Z. Yu. 2013. "Uncertainty intercomparison of different hydrological models in simulating extreme flows." *Water Resour. Manage.* 27 (5): 1393–1409. https://doi.org/10.1007/s11269-012-0244-5.

Cheng, Q. B., X. Chen, C. Y. Xu, C. Reinhardt-Imjela, and A. Schulte. 2014. "Improvement and comparison of likelihood functions for model calibration and parameter uncertainty analysis within a Markov chain Monte Carlo scheme." *J. Hydrol.* 519 (Nov): 2202–2214. https://doi.org/10.1016/j.jhydrol.2014.10.008.

Chiachio, M., J. L. Beck, J. Chiachio, and G. Rus. 2014. "Approximate Bayesian computation by subset simulation." *J. Sci. Comput.* 36 (3): 1339–1358. https://doi.org/10.1137/130932831.

Clark, M. P., and J. A. Vrugt. 2006. "Unraveling uncertainties in hydrologic model calibration: Addressing the problem of compensatory parameters." *Geophys. Res. Lett.* 33 (6): L06406. https://doi.org/10.1029/2005GL025604.

Corzo, G., and D. Solomatine. 2007. "Baseflow separation techniques for modular artificial neural network modelling in flow forecasting." *Hydrol. Sci. J.* 52 (3): 491–507. https://doi.org/10.1623/hysj.52.3.491.

Csilléry, K., M. Blum, O. Gaggiotti, and O. François. 2010. "Approximate Bayesian computation (ABC) in practice." *Trends Ecol. Evol.* 25 (7): 410–418. https://doi.org/10.1016/j.tree.2010.04.001.

Cunderlik, J. M., and S. P. Simonovic. 2004. *Calibration, verification, and sensitivity analysis of the HEC-HMS hydrologic model*. London, ON: Western Univ.

Dawson, C. W., R. J. Abrahart, and L. M. See. 2007. "HydroTest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts." *Environ. Modell. Software* 22 (7): 1034–1052. https://doi.org/10.1016/j.envsoft.2006.06.008.

D-maps. 2020. "Iran map, boundaries, provinces." Accessed April 28, 2016. https://d-maps.com/carte.php?num_car=5494&lang=en.

Engeland, K., C. Y. Xu, and L. Gottschalk. 2005. "Assessing uncertainties in a conceptual water balance model using Bayesian methodology." *Hydrol. Sci. J.* 50 (1): 45–63. https://doi.org/10.1623/hysj.50.1.45.56334.

Fenicia, F., D. Kavetski, P. Reichert, and C. Albert. 2018. "Signature-domain calibration of hydrological models using approximate Bayesian computation: Empirical analysis of fundamental properties." *Water Resour. Res.* 54: (6): 3958–3987. https://doi.org/10.1002/2017WR021616.

Foll, M., M. A. Beaumont, and O. Gaggiotti. 2008. "An approximate Bayesian computation approach to overcome biases that arise when using amplified fragment length polymorphism markers to study population structure." *Genetics* 179 (2): 927–939. https://doi.org/10.1534/genetics.107.084541.

Fu, Y. X., and W. H. Li. 1997. "Estimating the age of the common ancestor of a sample of DNA sequences." *Mol. Biol. Evol.* 14 (2): 195–199. https://doi.org/10.1093/oxfordjournals.molbev.a025753.

Gelman, A., and D. B. Rubin. 1992. "Inference from iterative simulation using multiple sequences." *Stat. Sci.* 7 (4): 457–472. https://doi.org/10.1214/ss/1177011136.

Gharibreza, M. 2019. "Huge inundation (March 2019) of Golestan Province, Iran, lessons that we learned." *OAJESS* 4 (3): 507–510.

Gourieroux, C., A. Monfort, and E. Renault. 1993. "Indirect inference." *J. Appl. Econ.* 8 (1): 85–118. https://doi.org/10.1002/jae.3950080507.

Green, I. R. A., and D. Stephenson. 1986. "Criteria for comparison of single event models." *Hydrol. Sci. J.* 31 (3): 395–411. https://doi.org/10.1080/02626668609491056.

Gupta, H. V., K. J. Beven, and T. Wagener. 2005. "Model calibration and uncertainty estimation." In *Encyclopedia of hydrological sciences*. Hoboken, NJ: Wiley.

Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez. 2009. "Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modeling." *J. Hydrol.* 377 (1): 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003.

Gupta, H. V., S. Sorooshian, and P. O. Yapo. 1998. "Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information." *Water Resour. Res.* 34 (4): 751–763. https://doi.org/10.1029/97WR03495.

Gupta, H. V., T. Wagener, and Y. Liu. 2008. "Reconciling theory with observations: Elements of a diagnostic approach to model evaluation." *Hydrol. Process.* 22 (18): 3802–3813. https://doi.org/10.1002/hyp.6989.

Jacobs, J. H., and R. Srinivasan. 2005. "Effects of curve number modification on runoff estimation using WSR-88D rainfall data in Texas watersheds." *J. Soil Water Conserv.* 60 (5): 274–278.

Jin, X. L., C. Y. Xu, Q. Zhang, and V. P. Singh. 2010. "Parameter and modeling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model." *J. Hydrol.* 383 (3–4): 147–155. https://doi.org/10.1016/j.jhydrol.2009.12.028.

Kamali, B., and S. J. Mousavi. 2014. "Automatic calibration of HEC-HMS model using multi-objective fuzzy optimal models." *Civ. Eng. Infrastruct. J.* 47 (1): 1–12.

Kamali, B., S. J. Mousavi, and K. C. Abbaspour. 2013. "Automatic calibration of HEC-HMS using single-objective and multi-objective PSO algorithms." *Hydrol. Process.* 27 (26): 4028–4042. https://doi.org/10.1002/hyp.9510.

Kavetski, D., G. Kuczera, and S. W. Franks. 2006a. "Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory." *Water Resour. Res.* 42 (3):W03407. https://doi.org/10.1029/2005WR004368.

Kavetski, D., G. Kuczera, and S. W. Franks. 2006b. "Bayesian analysis of input uncertainty in hydrological modeling: 2. Application." *Water Resour. Res.* 42 (4): W03408. https://doi.org/10.1029/2005WR004376.

Koskela, J. J., B. W. F. Croke, H. Koivusalo, A. J. Jakeman, and T. Kokkonen. 2012. "Bayesian inference of uncertainties in precipitation-streamflow modeling in a snow affected catchment." *Water Resour. Res.* 48 (11): W11513. https://doi.org/10.1029/2011WR011773.

Laloy, E., D. Fasbender, and C. L. Bielders. 2010. "Parameter optimization and uncertainty analysis for plot-scale continuous modeling of runoff using a formal Bayesian approach." *J. Hydrol.* 380 (1): 82–93. https://doi.org/10.1016/j.jhydrol.2009.10.025.

Laloy, E., and J. A. Vrugt. 2012. "High-dimensional posterior exploration of hydrologic models using multiple-try DREAM(ZS) and high-performance computing." *Water Resour. Res.* 48 (1): W01526. https://doi.org/10.1029/2011WR010608.

Li, L., J. Xia, C. Y. Xu, and V. P. Singh. 2010. "Evaluation of the subjective factors of the GLUE method and comparison with the formal Bayesian method in uncertainty assessment of hydrological models." *J. Hydrol.* 390 (3–4): 210–221. https://doi.org/10.1016/j.jhydrol.2010.06.044.

Li, L., C. Y. Xu, J. Xia, K. Engeland, and P. Reggiani. 2011. "Uncertainty estimates by Bayesian method with likelihood of AR (1) & normal model and AR (1) & multinormal model in different time-scales hydrological models." *J. Hydrol.* 406 (1–2): 54–65. https://doi.org/10.1016/j.jhydrol.2011.05.052.

Lintusaari, J., M. U. Gutmann, R. Dutta, S. Kaski, and J. Corander. 2017. "Fundamentals and recent developments in approximate Bayesian computation." *Syst. Biol.* 66 (1): 66–82. https://doi.org/10.1093/sysbio/syw077.

Lu, D., M. Ye, P. D. Meyer, G. P. Curtis, X. Shi, X. F. Niu, and S. B. Yabusaki. 2013. "Effects of error covariance structure on estimation of model averaging weights and predictive performance." *Water Resour. Res.* 49 (9): 6029–6047. https://doi.org/10.1002/wrcr.20441.

Marjoram, P., J. Molitor, V. Plagnol, and S. Tavare. 2003. "Markov chain Monte Carlo without likelihoods." *Proc. National Acad. Sci.* 100 (26): 15324–15328. https://doi.org/10.1073/pnas.0306899100.

Marjoram, P., and S. Tavare. 2006. "Modern computational approaches for analysing molecular genetic variation data." *Nat. Rev. Genet.* 7 (10): 759–770. https://doi.org/10.1038/nrg1961.

Massey, F. J. 1951. "The Kolmogorov–Smirnov test for goodness of fit." *J. Am. Stat. Assoc.* 46 (253): 68–78. https://doi.org/10.1080/01621459.1951.10500769.

McMillan, H., B. Jackson, M. Clark, D. Kavetski, and R. Woods. 2011. "Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models." *J. Hydrol.* 400 (1–2): 83–94. https://doi.org/10.1016/j.jhydrol.2011.01.026.

Minasny, B., J. A. Vrugt, and A. B. McBratney. 2011. "Confronting uncertainty in model-based geostatistics using Markov Chain Monte Carlo simulation." *Geoderma* 163 (3–4): 150–162. https://doi.org/10.1016/j.geoderma.2011.03.011.

Mishra, S. K., and V. P. Singh. 1999. "Another look at SCS-CN method." *J. Hydrol. Eng.* 4 (3): 257–264. https://doi.org/10.1061/(ASCE)1084-0699(1999)4:3(257).

Montanari, A., and D. Koutsoyiannis. 2012. "A blueprint for process-based modeling of uncertain hydrological systems." *Water Resour. Res.* 48 (9): W09555. https://doi.org/10.1029/2011WR011412.

Moradkhani, H., and S. Sorooshian. 2009. "General review of rainfall-runoff modeling: Model calibration, data assimilation, and uncertainty analysis." In *Hydrological modelling and the water cycle*. Berlin: Springer.

Nash, J. E., and J. V. Sutcliffe. 1970. "River flow forecasting through the conceptual model. Part 1: A discussion of principles." *J. Hydrol.* 10 (3): 282–290. https://doi.org/10.1016/0022-1694(70)90255-6.

Nott, D. J., L. Marshall, and J. Brown. 2012. "Generalized likelihood uncertainty estimation (GLUE) and approximate Bayesian computation:

What's the connection?" *Water Resour. Res.* 48 (12): W12602. https://doi.org/10.1029/2011WR011128.

Nourali, M., B. Ghahraman, M. Pourreza-Bilondi, and K. Davary. 2016. "Effect of formal and informal likelihood functions on uncertainty assessment in a single event rainfall-runoff model." *J. Hydrol.* 540 (Sep): 549–564. https://doi.org/10.1016/j.jhydrol.2016.06.022.

Olson, B., and W. Kleiber. 2017. "Approximate Bayesian computation methods for daily spatiotemporal precipitation occurrence simulation." *Water Resour. Res.* 53 (4): 3352–3372. https://doi.org/10.1002/2016WR019741.

Paudel, M., E. J. Nelson, and W. Scharffenberg. 2009. "Comparison of lumped and quasi-distributed Clark runoff models using the SCS curve number equation." *J. Hydrol. Eng.* 14 (10): 1098–1106. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000100.

Ponce, V. M., and R. H. Hawkins. 1996. "Runoff curve number: Has it reached maturity?" *J. Hydrol. Eng.* 1 (1): 11–19. https://doi.org/10.1061/(ASCE)1084-0699(1996)1:1(11).

Pritchard, J. K., M. T. Seielstad, A. Perez-Lezaun, and M. T. Feldman. 1999. "Population growth of human Y chromosomes: A study of Y chromosome microsatellites." *Mol. Biol. Evol.* 16 (12): 1791–1798. https://doi.org/10.1093/oxfordjournals.molbev.a026091.

Ratmann, O., C. Andrieu, C. Wiuf, and S. Richardson. 2009. "Model criticism based on likelihood-free inference, with an application to protein network evolution." *Proc. National Acad. Sci.* 106 (26): 10576–10581. https://doi.org/10.1073/pnas.0807882106.

Reichert, P., and J. Mieleitner. 2009. "Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters." *Water Resour. Res.* 45 (2): W10402. https://doi.org/10.1029/2009WR007814.

Renard, B., D. Kavetski, G. Kuczera, M. Thyer, and S. W. Franks. 2010. "Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors." *Water Resour. Res.* 46 (5): W05521. https://doi.org/10.1029/2009WR008328.

Renard, B., D. Kavetski, E. Leblois, M. Thyer, G. Kuczera, and S. W. Franks. 2011. "Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation." *Water Resour. Res.* 47 (11):W11516. https://doi.org/10.1029/2011WR010643.

Romero-Cuellar, J., A. Abbruzzo, G. Adelfio, and F. Francés. 2019. "Hydrological postprocessing based on approximate Bayesian computation (ABC)." *Stochastic Environ. Res. Risk Assess.* 33 (7): 1361–1373. https://doi.org/10.1007/s00477-019-01694-y.

Sadegh, M., and J. A. Vrugt. 2013. "Bridging the gap between GLUE and formal statistical approaches: Approximate Bayesian computation." *Hydrol. Earth Syst. Sci.* 17 (12): 4831–4850. https://doi.org/10.5194/hess-17-4831-2013.

Sadegh, M., and J. A. Vrugt. 2014. "Approximate Bayesian computation using Markov chain monte Carlo simulation: DREAM$_{(ABC)}$." *Water Resour. Res.* 50 (8): 6767–6787. https://doi.org/10.1002/2014WR015386.

Sadegh, M., J. A. Vrugt, C. Xu, and E. Volpi. 2015. "The stationarity paradigm revisited: Hypothesis testing using diagnostics, summary metrics, and DREAM$_{(ABC)}$." *Water Resour. Res.* 51 (11): 9207–9231. https://doi.org/10.1002/2014WR016805.

Safaripour, M., M. Monavari, M. Zare, Z. Abedi, and A. Gharagozlou. 2012. "Flood risk assessment using GIS (Case study: Golestan Province, Iran)." *Pol. J. Environ. Stud.* 21 (6): 1817–1824.

Salamon, P., and L. Feyen. 2009. "Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter." *J. Hydrol.* 376 (3–4): 428–442. https://doi.org/10.1016/j.jhydrol.2009.07.051.

Schoups, G., and J. A. Vrugt. 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): W10531. https://doi.org/10.1029/2009WR008933.

Shafiei, M., B. Ghahraman, B. Saghafian, K. Davary, S. Pande, and M. Vazifedoust. 2014. "Uncertainty assessment of the agro-hydrological SWAP model application at field scale: A case study in a dry region." *Agric. Water Manage.* 146 (1): 324–334. https://doi.org/10.1016/j.agwat.2014.09.008.

Shamir, E., B. Imam, E. Morin, H. V. Gupta, and S. Sorooshian. 2005. "The role of hydrograph indices in parameter estimation of rainfall–runoff models." *Hydrol. Process.* 19 (11): 2187–2207. https://doi.org/10.1002/hyp.5676.

Sharifi, F., V. Samadi, and C. A. M. E. Wilson. 2012. "Causes and consequences of recent floods in the Golestan catchments and Caspian Sea regions of Iran." *Nat. Hazards* 61 (2): 533–550. https://doi.org/10.1007/s11069-011-9934-1.

Sikorska, A. E. 2012. "Interactive comment on Bayesian uncertainty assessment of flood predictions in ungauged urban basins for conceptual rainfall-runoff models." *Hydrol. Earth Syst. Sci. Discuss.* 16 (4): 6284–6310. https://doi.org/10.5194/hess-16-1221-2012.

Smith, T., L. Marshall, and A. Sharma. 2015. "Modeling residual hydrologic errors with Bayesian inference." *J. Hydrol.* 528 (Sep): 29–37. https://doi.org/10.1016/j.jhydrol.2015.05.051.

Smith, T. J., and L. A. Marshall. 2008. "Bayesian methods in hydrology: A study of recent advancements in Markov chain Monte Carlo techniques." *Water Resour. Res.* 44 (12): W00B05. https://doi.org/10.1029/2007WR006705.

Szcześniak, M., and M. Piniewski. 2015. "Improvement of hydrological simulations by applying daily precipitation interpolation schemes in meso-scale catchments." *Water* 7 (12): 747–779. https://doi.org/10.3390/w7020747.

Tavaré, S., D. J. Balding, R. C. Griffiths, and P. Donnelly. 1997. "Inferring coalescence times from DNA sequence data." *Genetics* 145 (2): 505–518.

Taye, M. T., V. Ntegeka, N. P. Ogiramoi, and P. Willems. 2011. "Assessment of climate change impact on hydrological extremes in two source regions of the Nile River basin." *Hydrol. Earth Syst. Sci.* 15 (1): 209–222. https://doi.org/10.5194/hess-15-209-2011.

ter Braak, C. J. F., and J. A. Vrugt. 2008. "Differential evolution Markov chain with snooker updater and fewer chains." *Stat. Comput.* 18 (4): 435–446. https://doi.org/10.1007/s11222-008-9104-9.

Tewolde, M. H. 2005. "Flood routing in ungauged catchments using Muskingum methods." M.Sc. dissertation. School of Bioresources Engineering and Environmental Hydrology, Univ. of KwaZulu-Natal.

Tian, F., Y. Sun, H. Hu, and H. Li. 2016. "Searching for an optimized single-objective function matching multiple objectives with automatic calibration of hydrological models." *Hydrol. Earth Syst. Sci. Discuss.* 29 (6): 934–948. https://doi.org/10.5194/hess-2016-88.

Turner, B. M., and P. B. Sederberg. 2012. "Approximate Bayesian computation with differential evolution." *J. Math. Psychol.* 56 (5): 375–385. https://doi.org/10.1016/j.jmp.2012.06.004.

Turner, B. M., and T. Van Zandt. 2012. "A tutorial on approximate Bayesian computation." *J. Math. Psychol.* 56 (2): 69–85. https://doi.org/10.1016/j.jmp.2012.02.005.

USACE. 2000. *HEC-HMS technical reference manual*. Davis, CA: USACE.

Vaghefi, S. A., M. Keykhai, F. Jahanbakhshi, J. Sheikholeslami, A. Ahmadi, H. Yang, and K. C. Abbaspour. 2019. "The future of extreme climate in Iran." *Sci. Rep.* 9 (1): 1464. https://doi.org/10.1038/s41598-018-38071-8.

Vrugt, J. A., and W. Bouten. 2002. "Validity of first-order approximations to describe parameter uncertainty in soil hydrologic models." *Soil Sci. Soc. Am. J.* 66 (6): 1740–1751. https://doi.org/10.2136/sssaj2002.1740.

Vrugt, J. A., C. G. H. Diks, W. Bouten, H. V. Gupta, and J. M. Verstraten. 2005. "Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation." *Water Resour. Res.* 41 (1): W01017. https://doi.org/10.1029/2004WR003059.

Vrugt, J. A., and B. A. Robinson. 2007. "Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging." *Water Resour. Res.* 43 (1): W01411. https://doi.org/10.1029/2005WR004838.

Vrugt, J. A., and M. Sadegh. 2013. "Toward diagnostic model calibration and evaluation: Approximate Bayesian computation." *Water Resour. Res.* 49 (7): 4335–4345. https://doi.org/10.1002/wrcr.20354.

Vrugt, J. A., C. J. F. ter Braak, M. P. Clark, J. M. Hyman, and B. A. Robinson. 2008. "Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo

simulation." *Water Resour. Res.* 44 (12): W00B09. https://doi.org/10.1029/2007WR006720.

Vrugt, J. A., C. J. F. ter Braak, C. G. H. Diks, B. A. Robinson, J. M. Hyman, and D. Higdon. 2009a. "Accelerating Markov chain Monte Carlo simulation using self-adaptative differential evolution with randomized subspace sampling." *Int. J. Nonlinear Sci. Numer. Simul.* 10 (3): 273–290. https://doi.org/10.1515/IJNSNS.2009.10.3.273.

Vrugt, J. A., C. J. F. ter Braak, H. V. Gupta, and B. A. Robinson. 2009b. "Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?" *Stochastic Environ. Res. Risk Assess.* 23 (7): 1011–1026. https://doi.org/10.1007/s00477-008-0274-y.

Wagener, T., and H. V. Gupta. 2005. "Model identification for hydrological forecasting under uncertainty." *Stochastic Environ. Res. Risk. Assess.* 19 (6): 378–387. https://doi.org/10.1007/s00477-005-0006-5.

Wagener, T., K. V. Werkhoven, P. Reed, and Y. Tang. 2009. "Multiobjective sensitivity analysis to understand the information content in streamflow observations for distributed watershed modeling." *Water Resour. Res.* 45 (2): W02501. https://doi.org/10.1029/2008WR007347.

Wang, X., X. He, J. R. Williams, and R. C. Izaurralde. 2005. "Sensitivity and uncertainty analyses of crop yields and soil organic carbon simulated with EPIC." *Trans. ASAE* 48 (3): 1041–1054. https://doi.org/10.13031/2013.18515.

Wilkinson, R. D. 2013. "Approximate Bayesian computation (ABC) gives exact results under the assumption of model error." *Stat. Appl. Genet. Mol. Biol.* 12 (2): 129–141. https://doi.org/10.1515/sagmb-2013-0010.

Willems, P. 2012. "Model uncertainty analysis by variance decomposition." *Phys. Chem. Earth.* 42 (Jan): 21–30. https://doi.org/10.1016/j.pce.2011.07.003.

Yang, J., A. Jakeman, G. Fang, and X. Chen. 2018. "Uncertainty analysis of a semi-distributed hydrologic model based on a Gaussian process emulator." *Environ. Modell. Software* 101 (Mar): 289–300. https://doi.org/10.1016/j.envsoft.2017.11.037.

Yang, J., P. Reichert, and K. C. Abbaspour. 2007. "Bayesian uncertainty analysis in distributed hydrological modelling: A case study in the Thur River basin (Switzerland)." *Water Resour. Res.* 43 (10):W10401. https://doi.org/10.1029/2006WR005497.

Yang, J., P. Reichert, K. C. Abbaspour, J. Xia, and H. Yang. 2008. "Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China." *J. Hydrol.* 358 (1–2): 1–23. https://doi.org/10.1016/j.jhydrol.2008.05.012.