https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

An improved Approximate Bayesian Computation approach for high-dimensional posterior exploration of hydrological models

Song Liu[1], Dunxian She[1,2], Liping Zhang[1,2], Jun Xia[1,2]

[1] State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, P. R. China

[2] Hubei Key Laboratory of Water System Science for Sponge City Construction, Wuhan University, Wuhan 430072, P. R. China

Corresponding authors:

Dr. Dunxian She

Email: shedunxian@whu.edu.cn

Dr. Liping Zhang

Email: zhanglp@whu.edu.cn

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023


Hydrology and Earth System Sciences
Discussions

Abstract

The Approximate Bayesian computation (ABC) methods provide a powerful tool for sampling from Bayesian posteriors for cases where we can simulate samples, but we have no access to an explicit expression of the likelihood function. The Simulated Annealing ABC (SABC) algorithm has been proposed to achieve a fast convergence to an unbiased approximation to the posterior by adaptively decreasing an initially coarse tolerance value. However, this algorithm uses a rather simplistic random walk Metropolis (RWM) sampler to generate trial moves in a Markov chain and always requires an excessive number of model evaluations for approximating the posterior, which inevitably lowers the sampling efficiency and limits its applications in more complex hydrologic modelling practices. Inspired by the advances made in Markov Chain Monte Carlo (MCMC) methods, we incorporated an adaptive Differential Evolution scheme to enhance the efficiency of SABC sampling. This scheme has its roots within Differential Evolution Markov Chains (DE-MC), and additionally utilizes a self-adaptive randomized subspace sampling strategy to optimally select the dimensions of parameters to be updated each time a proposal is generated. The superiority of the modified SABC (mSABC) over the original SABC algorithm was demonstrated through a SAC-SMA application to the Danjiangkou Reservoir region (DRR). The case study results showed that mSABC was far more efficient with lower computation costs and higher acceptance rates, and achieved higher numerical accuracy than the original SABC algorithm. mSABC also resulted in a better overall prediction

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

of streamflow time series and signatures. The introduction of more advanced MCMC sampler into SABC helps to speed up convergence to the approximate posterior while achieving better model performance, which significantly widens the applicability of SABC to complex posterior exploration problems.

Keywords: Bayesian inference, Approximate Bayesian computation, Differential Evolution, Subspace sampling, Hydrological modelling, Hydrological signatures

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023


Hydrology and Earth System Sciences
Discussions

1 Introduction

The Bayesian methods provide a statistically convenient vehicle for probabilistic uncertainty quantification of hydrological models (Evin et al., 2014; Mcinerney et al., 2017; Schoups and Vrugt, 2010). According to Bayes' theorem, the posterior distribution of the parameters of a model can be derived from the prior distribution of estimated parameters $\boldsymbol{\theta}$ and measurements of observed system behavior $\widetilde{\mathbf{Y}}$ as

$$p\left(\boldsymbol{\theta}\middle|\widetilde{\mathbf{Y}}\right) \propto p(\boldsymbol{\theta}) L\left(\widetilde{\mathbf{Y}}\middle|\boldsymbol{\theta}\right) \tag{1}$$

where $p(\boldsymbol{\theta})$ and $p\left(\boldsymbol{\theta}\middle|\widetilde{\mathbf{Y}}\right)$ signify the prior and posterior parameter distribution, respectively, and $L\left(\widetilde{\mathbf{Y}}\middle|\boldsymbol{\theta}\right)$ denotes the likelihood function of the model, i.e., the probability density of model outputs for given parameters $\boldsymbol{\theta}$ evaluated at the measurements $\widetilde{\mathbf{Y}}$. Challenges lie in the formulation of an appropriate likelihood function. If the statistical assumptions for formulating the likelihood are violated, the results of Bayesian methods are unreliable (Beven and Binley, 2014). Simple likelihood functions that assume independent identically distributed Gaussian error residuals are statistically convenient, but this oversimplified assumption cannot be justified in real-world applications. The presence of the observational data errors (Mcmillan et al., 2012) and model inadequacies (Gupta et al., 2012) introduces complex error residual structure whose probabilistic properties are difficult to characterize precisely with classical likelihood functions (Sadegh and Vrugt, 2013). The residuals might depend on the catchment and in most cases, they might be correlated in both time and space (Engeland

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023


Hydrology and Earth System Sciences
Discussions

and Gottschalk, 2002). Gupta et al. (1998) also stated that no objective, statistically correct likelihood function that takes into account all these aspects might exist. This therefore has been the focus of ongoing debate in hydrology literature (Beven and Binley, 2014).

In contrast, sampling from Bayesian posteriors is relatively straightforward (Kavetski et al., 2018). Likelihood-free inference has been suggested that simulates samples by sampling model outputs from an approximation to the posterior and compares them with the observational data without evaluating the likelihood function, which is nowadays referred to as Approximate Bayesian Computation (ABC) (e.g., Beaumont et al., 2002; Tavare et al., 1997). ABC methods originate in the statistical literature (Diggle and Gratton, 1984), and are especially useful for cases where we can generate samples, but do not have access to an explicit expression for the actual likelihood. The first application of ABC in the hydrological literature was found in Nott et al. (2012), where the theoretical correspondence between ABC and a variant of Bayesian methods known as the Generalized Likelihood Uncertainty Estimation (GLUE) (Beven and Binley, 1992) is clarified. ABC inference was then introduced as a possible vehicle for hydrologic modelling and uncertainty quantification by Vrugt and Sadegh (2013). More recent research into ABC generalized as a class of numerical algorithms for sampling from conditional distributions such as Bayesian posteriors was provided by Kavetski et al. (2018) and Fenicia et al. (2018). A thorough review of ABC methods is given by Beaumont (2019).

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023

The basic ABC algorithm replaces the outputs of a probability model with one or multiple summary statistics and introduces some prespecified tolerance value for its distance from the observed summary statistics (Tavare et al., 1997; Weiss and Von Haeseler, 1998). This method adopts a rather simplistic rejection sampling technique to simulate samples from an approximate posterior. Implementation of the rejection sampler is straightforward and very easy, but is not likely to result in robust estimates of posterior parameter distribution for complex inference problems (Sadegh and Vrugt, 2014). When the prior parameter distribution is just a poor approximation to the (unknown) actual posterior, which is always the case in realistic case studies, it requires excessive number of model evaluations and CPU times to acquire an adequate number of acceptable samples. To compensate for this inefficiency, a group of population Monte Carlo (PMC) algorithms based on sequential importance sampling is developed (Beaumont et al., 2009; Sisson et al., 2007; Turner and Van Zandt, 2012; Toni et al., 2009). The rationale of the ABC-PMC sampler is to use a sequence of monotonically decreasing tolerance values and iteratively evolve an ensemble of constant size towards an approximate posterior based on the accepted proposals (Beaumont, 2019). Each iteration consists of drawing a new particle from the old one with weights and resampling. A sequence of multi-normal proposal distributions derived from the adapted particle is used to successively search the parameter-output space and approximate the posterior. This algorithm has been demonstrated to have a significantly higher sampling efficiency than the basic ABC-REJ sampler for situations in which an

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023

107 uninformative flat prior extending far beyond the posterior distribution is chosen
108 (Sadegh and Vrugt, 2014). The ABC-PMC sampler of Beaumont et al. (2009) and
109 Turner and Van Zandt (2012) requires that the sequence of tolerance values be specified
110 a priori by the users. A poor selection of the tolerance values can lead to premature
111 convergence and provide misleading results. Inspired by Simulated Annealing, Albert
112 et al. (2015) presented an adaptive scheme that decreases the tolerance according to the
113 particles' distance from the target. The key question of how fast the tolerance should be
114 reduced in pursuit of a fast convergence speed to the correct posterior is pleasantly
115 answered by interpreting the tolerance parameter as the temperature of the environment
116 using non-equilibrium thermodynamics. The tolerance is adapted in such a way that the
117 entropy production is minimized (Albert et al., 2015). This class of particle algorithms
118 for ABC is known collectively as Simulated Annealing ABC, or SABC. Previous work
119 (Fenicia et al., 2018; Kavetski et al., 2018) has demonstrated the effectiveness and
120 efficiency of SABC for probabilistic uncertainty quantification in a few lumped
121 modelling practices.

122 Another algorithmic enhancement embedded in the SABC algorithm is that it uses
123 the Markov Chain Monte Carlo (MCMC) scheme to simulate samples from an
124 approximation to the posterior (Marjoram et al., 2003). The building block of this
125 method is a Markov Chain, which generates a random walk through the parameter space
126 and trial jumps from the current state of the chain to a new state. The most common
127 MCMC method is the random walk Metropolis (RWM) sampler. It works with a single

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023


Hydrology and Earth System Sciences
Discussions
Open Access
EGU

128 trajectory (chain), and a symmetric normal jump distribution $N_d(0,\Sigma)$ whose
129 covariance Σ is adapted using the accepted proposals of the chain according to
130 $\Sigma = \beta\Sigma + s\mathbf{I}_d$. Here, s is a small constant preventing the covariance matrix from
131 degenerating, and β is a scaling factor that depends only on the parameter dimension d.
132 This method may be adequate for simple inference problems involving just a handful
133 of parameters, but is not likely to achieve an adequate sampling efficiency and provide
134 accurate posterior estimates when $p\left(\boldsymbol{\theta}\middle|\widetilde{\mathbf{Y}}\right)$ is high-dimensional with complex
135 posterior surfaces that contain numerous local optima and multiple regions of attraction
136 (Laloy and Vrugt, 2012; Ter Braak, 2006; Ter Braak and Vrugt, 2008; Vrugt, 2016;
137 Vrugt et al., 2008; Vrugt et al., 2009). Consequently, it always requires an excessive
138 number of model evaluations to sample from the approximate posteriors, which limits
139 its use in hydrological models with high computation costs (Shafii and Tolson, 2015).

140 In a separate line of research, variants of the MCMC methods have been developed
141 for exploring the posteriors. To improve efficiency for high-dimensional problems, Ter
142 Braak and Vrugt (2008) has proposed an adaptive RWM method entitled Differential
143 Evolution Markov Chain (DE-MC). It uses multiple different chains running
144 sequentially or in parallel for sampling from the posterior distribution. DE-MC directly
145 utilizes the current states of the chains to generate the proposals to allow for direct
146 jumps between disconnected modes of complex posterior surfaces. This is a significant
147 strength of DE-MC acting as a multi-chain method compared to single chain methods
148 (Vrugt, 2016). Previous work (Vrugt et al., 2008; Vrugt et al., 2009) has shown that the

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023


Hydrology and Earth System Sciences
Discussions

efficiency of DE-MC can be further enhanced by combining self-adaptive randomized subspace sampling. For high-dimensional problems it is rather inefficient to update all d dimensions of parameters simultaneously (Haario et al., 2005), especially when parameters have vastly different scales. Subspace sampling is implemented by only updating randomly selected dimensions of parameters each time a candidate point is generated. By "self-adaptive" we mean that the dimensions of parameters that participate in the candidate generation are tuned adaptively during burn-in by favoring large jumps over small ones in each of the chains (Vrugt et al., 2009). In addition, this method includes higher-order chain pairs for candidate generation to increase diversity in the candidates, and outlier chain correction techniques to speed up convergence. This method, entitled DiffeRential Evolution Adaptive Metropolis (DREAM), maintains detailed balance and overall ergodicity of the Markov chains (Vrugt, 2016).

In this paper, we examine the use of an adaptive MCMC sampling within the SABC algorithm to improve the sampling efficiency and accelerate the chain convergence. The adopted MCMC sampler is part of the DREAM algorithm and intentionally ignores the outlier chain correction module as it destroys detailed balance of the Markov chains and can only be used during burn-in (Vrugt et al., 2009). A few papers do investigate the use of MCMC simulation in ABC to enhance the ABC sampling efficiency (Sadegh and Vrugt, 2014), but this approach uses a static tolerance value for guiding the search in the parameter space, which is a significant difference from SABC. The concept of the present paper is to generate a proposal using the

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023


Hydrology and Earth System Sciences
Discussions

adaptive MCMC sampler and derive the associated probabilistic predictions using the SABC method. For convenience we refer to the modified SABC algorithm based on the adaptive MCMC sampler as mSABC, in contrast to the original SABC algorithm implementing the simple RWM sampler. The superiority of mSABC over the original SABC is to be demonstrated using a calibration of the SAC-SMA model, which has been suggested to be a challenging task due to complex posterior surfaces (Duan et al., 1992) and thus frequently utilized as a benchmark hydrologic modelling experiment for validation of algorithmic enhancements (e.g., Laloy and Vrugt, 2012; Sadegh and Vrugt, 2014; Vrugt et al., 2009).

2 Methodology

2.1 Approximate Bayesian Computation

The Approximate Bayesian Computation (ABC) approaches provide an attractive solution to a Bayesian inference of a hydrological model where the likelihood function is impossible to formulate, or computationally expensive to evaluate. The conceptual basis of ABC is that we can always approximate the probability density function (pdf) of a probabilistic model by sampling from this probabilistic model (Kavetski et al., 2018). Given the observed streamflow data $\widetilde{\mathbf{Y}}$ and a sample from a probabilistic model $\mathbf{Y}(\boldsymbol{\theta})$, as the distance between the observed and simulated data $\rho\left(\widetilde{\mathbf{Y}}, \mathbf{Y}(\boldsymbol{\theta})\right)$ is lower than some tolerance value τ ($\tau \to 0$), $\boldsymbol{\theta}$ should be a sample from an approximation of the correct posterior (Marjoram et al., 2003; Sisson et al., 2007). For

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023


Hydrology and Earth System Sciences
Discussions

high-dimensional datasets, it is computationally efficient to consider the criterion $\rho\left(\mathbf{s}\left(\widetilde{\mathbf{Y}}\right),\mathbf{s}\left(\mathbf{Y}\left(\boldsymbol{\theta}\right)\right)\right)$ to determine whether to accept the candidate sample or not, where $\mathbf{s}\left(\cdot\right)$ is a vector of summary statistics (hydrological signatures) computed from the data. Typical signatures include the flow duration curve (FDC), baseflow index, and other streamflow characteristics (e.g., Addor et al., 2018; Clausen and Biggs, 2000; Olden and Poff, 2003; Westerberg and Mcmillan, 2015; Yadav et al., 2007).

Implementation of ABC using simple reject sampling scheme (ABC-REJ) consists of the following steps:

(1) Draw a sample $\boldsymbol{\theta}_0$ from the prior $p\left(\boldsymbol{\theta}\right)$;

(2) Generate simulated data $\mathbf{Y}\left(\boldsymbol{\theta}_0\right)$ from the probability model $\mathbf{Y}\left(\boldsymbol{\theta}\right)$ using $\boldsymbol{\theta}_0$;

(3) Accept the candidate $\boldsymbol{\theta}_0$ if the distance $\rho\left(\mathbf{s}\left(\widetilde{\mathbf{Y}}\right),\mathbf{s}\left(\mathbf{Y}\left(\boldsymbol{\theta}_0\right)\right)\right)$ is lower than some prespecified tolerance value τ;

(4) Repeat steps (1-3) until an adequate number of accepted samples is obtained.

Additional algorithmic advances have been proposed to enhance the sampling efficiency of the basic ABC-REJ algorithm. One common approach is to adaptively tighten an initially coarse tolerance as the sampling progresses (Toni et al., 2009). This allows for a sufficiently low final value of τ and high numerical accuracy in approximating the posterior $p\left(\boldsymbol{\theta}\middle|\widetilde{\mathbf{Y}}\right)$. The SABC algorithm of Albert et al. (2015) presents an adaptive annealing schedule that attempts a fast convergence to an unbiased

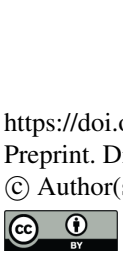
https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023


Hydrology and Earth System Sciences
Discussions

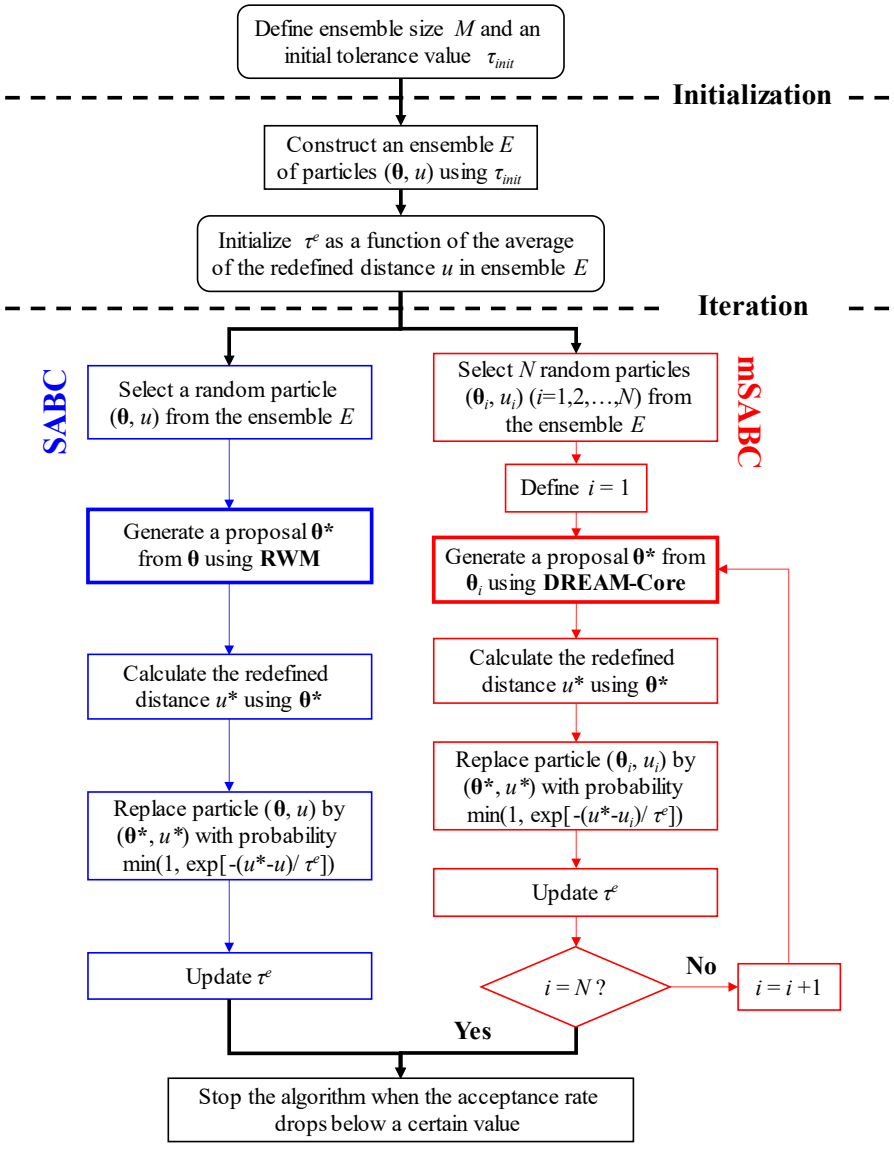
result. SABC starts with an ensemble of particles $\left(\boldsymbol{\theta};\mathbf{Y}\left(\boldsymbol{\theta}\right)\right)$ drawn from the prior in the parameter-output space which evolves according to a family of Metropolis kernels defined as $Z^{-1}\left(\tau\right)p\left(\boldsymbol{\theta}\right)L\left(\widetilde{\mathbf{Y}}\middle|\boldsymbol{\theta}\right)\exp\left(-\rho\left(\mathbf{s}\left(\widetilde{\mathbf{Y}}\right),\mathbf{s}\left(\mathbf{Y}\left(\boldsymbol{\theta}\right)\right)\right)\middle/\tau\right)$ for a decreasing sequence of tolerances. The adaptive schedule of τ stems from the non-equilibrium thermodynamics. The tolerance $\tau\left(t\right)$ is interpreted as the temperature of a gas (*system*) at time *t*, which is in connection with a heat bath (*environment*) whose temperature $\tau^{e}\left(t\right)$ can be controlled. The cooling of the system is pursued by lowering the temperature of the environment. The cooling schedule $\tau^{e}\left(t\right)$ is tuned according to the mean distance $U\left(t\right)$ of the particles from the data $\widetilde{\mathbf{Y}}$ using non-equilibrium thermodynamics of entropy production to minimize the required simulations from the likelihood (Albert et al., 2015). The SABC algorithm has been successfully applied to probabilistic uncertainty quantification of hydrological models (Fenicia et al., 2018; Kavetski et al., 2018).

The SABC algorithm is run in a sequence of steps. In the "Initialization" step, we construct the prior ensemble *E* according to an initial value of τ. An initial value of the optimal annealing schedule τ^{e} is estimated as a function of mean fields *U* of the prior ensemble. In the "Iteration" step, we update the ensemble *E* according to a sequence of Metropolis kernels where the proposal is generated using a random walk Metropolis (RWM) sampler. The value of τ^{e} is adaptively lowered as the iteration proceeds. A schematic description of the SABC algorithm is presented in **Fig. 1**.

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023



Hydrology and Earth System Sciences
Discussions



231

232 **Fig. 1** Schematic overview of the original SABC and mSABC algorithms. The original
233 SABC algorithm utilizes a simple random walk Metropolis (RWM) sampler for
234 proposal generation, while mSABC implements more advanced DREAM-Core sampler
235 to generate a proposal during the "Iteration" process.

236

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023

237 The original SABC algorithm utilizes a simple RWM sampler to generate a
238 proposal in the "Iteration" process. It also allows for adaptively updating the covariance
239 matrix of a multivariate proposal distribution using all previously accepted samples of
240 the Markov Chain to improve the sampling efficiency of the algorithm (Albert et al.,
241 2015). This scheme works well for simple unimodal inference problems, but becomes
242 inefficient when confronted with complex posterior surface with multiple different
243 regions of attraction and numerous local optima (Ter Braak and Vrugt, 2008; Ter Braak,
244 2006; Vrugt et al., 2008; Vrugt et al., 2009; Vrugt, 2016). Besides, simultaneous
245 updating of all parameter dimensions results in low acceptance probability for high-
246 dimensional problems (Vrugt, 2016; Sadegh and Vrugt, 2014; Vrugt et al., 2009). These
247 issues inevitably reduces the efficiency and numerical accuracy of SABC in an
248 approximation to the correct posterior.

249 **2.2 Adaptive Markov Chain Monte Carlo sampling**

250 The modified SABC (mSABC) algorithm we propose is motivated from the
251 development of DREAM$_{(ABC)}$ algorithm which uses the DiffeRential Evolution
252 Adaptive Metropolis (DREAM) method to speed up the sampling of ABC posterior
253 distributions (Sadegh and Vrugt, 2014). The mSABC algorithm replaces the simple
254 RWM sampling in the original SABC algorithm with an adaptive MCMC simulation in
255 pursuit of higher sampling efficiency. The proposed MCMC sampler constitutes the
256 core of a family of DREAM algorithms (Vrugt, 2016; Vrugt and Beven, 2018); We here
257 refer to it as DREAM-Core sampling. DREAM-Core stems from DE-MC (Ter Braak,

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023


Hydrology and Earth System Sciences
Discussions
Open Access
EGU

2006), and introduces a self-adaptive randomized subspace sampling strategy to accelerate convergence to the posterior (Vrugt et al., 2009; Sadegh and Vrugt, 2014).

Compared to RWM within SABC where a particle is randomly selected from the ensemble each time a candidate point is generated, mSABC selects N random particles simultaneously, each serving as the starting point of one independent chain in DREAM-Core (**Fig. 1**). It utilized the information about the scale and orientation of the proposal distribution contained in the remaining N-1 chains to generate a candidate in the current chain (Vrugt et al., 2009). Let $\mathbf{x}^i$ be the current state of chain $i=\{1,2,\ldots,N\}$, a candidate point in chain i, $\mathbf{z}^i$ is given by (Storn and Price, 1997)

$$\mathbf{z}^i = \mathbf{x}^i + \left(\mathbf{1}_{d^*} + \boldsymbol{\lambda}_{d^*}\right)\gamma\left(\delta, d^*\right)\sum_{j=1}^{\delta}\left(x_A^{\mathbf{a}_j} - x_A^{\mathbf{b}_j}\right) + \boldsymbol{\varepsilon}_{d^*} \quad (2)$$

where $\gamma = 2.38/\sqrt{2\delta d^*}$ is the jump rate, δ is the number of chain pairs used to generate the jump, and d^* is the number of dimensions to be updated jointly, stretching the parameter subspace A. $\mathbf{a}_j$ and $\mathbf{b}_j$ are vectors with δ integers drawn without replacement from $\{1,\ldots,i-1,i+1,\ldots,N\}$. The value of N should at least be equivalent to 2δ+1, or $N = 3$ for the default of $\delta = 1$ (Vrugt, 2016). The values of $\boldsymbol{\lambda}$ and $\boldsymbol{\varepsilon}$ are drawn from $U_{d^*}[-c,c]$ and $N_{d^*}(0,c_*)$, with $c = 0.1$ and c_* small compared to the width of the target distribution, $c_* = 10^{-12}$ say. we set γ = 1 at every 5th generation, or $p(\gamma = 1) = 0.2$ to enable jumping between disconnected modes of the posterior (Ter Braak, 2006).

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023


Hydrology and Earth System Sciences
Discussions

277 The subspace A spanned by randomly selected d^* dimensions of $\mathbf{x}^i$ is
278 constructed in DREAM-Core using a geometric sequence of crossover values
279 $\left\{\frac{1}{n_{\mathrm{CR}}},\frac{2}{n_{\mathrm{CR}}},\ldots,1\right\}$ with selection probabilities $\mathbf{p}_{\mathrm{CR}}$. A good choice of n_{CR} = 3 has
280 shown to work well in practice (Sadegh and Vrugt, 2014). To speed up the sampling of
281 the ABC posteriors, the selection probabilities $\mathbf{p}_{\mathrm{CR}}$ are tuned adaptively during burn-
282 in by maximizing the jumping distance in each of the N chains. This procedure is
283 described in detail in Vrugt et al. (2009).

284 A schematic description of the proposed mSABC algorithm is presented in **Fig. 1**.
285 The chain evolution of mSABC differs from classical Markov Chain methods. Each
286 time a proposal is generated, mSABC accepts randomly selected N particles from the
287 ensemble E as the starting points of chain evolution for next iteration. The aim of the
288 proposal generation in mSABC is to evolve the prior ensemble, instead of acquiring an
289 entire sequence of the Markov Chains in pursuit of convergence to the posterior.

290 **2.3 Criteria for the comparison**

291 Several frequently used uncertainty evaluation measures are used to
292 comprehensively quantify the predictive performance obtained with SABC and
293 mSABC respectively. The indices for assessing the 95% prediction limits include the
294 containing ratio (*CR*), average relative band-width (*RB*) and average relative deviation
295 amplitude (*RD*) (Xiong et al., 2009). All three metrics represent desirable characteristics
296 for the prediction limits. *CR* is computed as the ratio of the number of the observations

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023


Hydrology and Earth System Sciences
Discussions

297 enveloped by the 95% prediction bound across all time steps. *RB* is used to quantify the
298 average relative band width of the predictions. *RD* is used to quantify the actual
299 discrepancy between the trajectory consisting of the middle points of the prediction
300 bound and the observations. A higher value of *CR* and lower values of *RB* and *RD*
301 indicate better predictive performance. The *RB* and *RD* are calculated as

302 $$RB = \frac{1}{N_t}\sum_{i=1}^{N_t}\frac{q_i^u - q_i^l}{Q_i} \quad (3)$$

303 $$RD = \frac{1}{N_t}\sum_{i=1}^{N_t}\left|\frac{q_i^u + q_i^l}{2} - Q_i\right| \quad (4)$$

304 where Q_i is the observed flow at time *i*, q_i^u and q_i^l are the upper and lower limits
305 of the 95% prediction band respectively, and N_t is the number of time steps.

306 In addition, the reliability of probabilistic predictions is graphically evaluated
307 using the predictive quantile-quantile (PQQ) plot. A deviation from the diagonal line
308 (1:1 line) indicates the inconsistencies between the measurements and the model
309 predictions (Laio and Tamea, 2007; Thyer et al., 2009).

310 We also compared the performance of the trajectory consisting of the middle points
311 of the 95% prediction limits against the observed hydrograph. These performance
312 metrics include the root mean squared error (*RMSE*), correlation coefficient (*CC*), and
313 percent bias (*PBIAS*). The equations for computing *RMSE*, *CC* and *PBIAS* are presented
314 as follows:

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023


Hydrology and Earth System Sciences
Discussions

$$RMSE = \sqrt{\frac{1}{N_t}\sum_{i=1}^{N_t}\left(Q_i - q_i^m\right)^2} \tag{5}$$

$$CC = \frac{\sum_{i=1}^{N_t}\left(Q_i - \overline{Q}\right)\left(q_i^m - \overline{q^m}\right)}{\sqrt{\sum_{i=1}^{N_t}\left(Q_i - \overline{Q}\right)^2}\sqrt{\sum_{i=1}^{N_t}\left(q_i^m - \overline{q^m}\right)^2}} \tag{6}$$

$$PBIAS = \sum_{i=1}^{N_t}\left(Q_i - q_i^m\right) \Big/ \sum_{i=1}^{N_t} Q_i \times 100 \tag{7}$$

where q_i^m is the middle points of the 95% prediction limits at time *i*. $\overline{Q}$ and $\overline{q^m}$ are the average of the observed flows and the middle points of the 95% prediction limits respectively. Lower values of *RMSE* and absolute *PBIAS* and a higher value of *CC* correspond to better model performance.

3 Case study

A realistic case study is used to illustrate the advantages of mSABC over the original SABC algorithm in hydrologic modelling practice. We consider simulation of the rainfall-runoff relationship in the Danjiangkou Reservoir region (DRR) of China using the SAC-SMA hydrological model. The SAC-SMA model is a continuous conceptual rainfall-runoff model with spatially lumped parameters that represents the soil column as thin upper and thicker lower layers of multiple storages (Burnash et al., 1973), and has been extensively used for modelling of the rainfall-runoff process in literature (e.g., Gupta et al., 1998; Sadegh and Vrugt, 2014; Vrugt et al., 2009). The estimated daily reservoir inflows from 1998 to 2007 are collected for model calibration

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023


Hydrology and Earth System Sciences
Discussions

and validation. The first two years of data are used as burn-in to acquire stable and reliable estimates of initial states. Five years of daily hydrologic data (2000-2004) are used for calibration, and three more years (2005-2007) as the validation period. SAC-SMA is here applied at a 6-hourly time step with 14 parameters to be inferred during calibration (**Table 1**). Details on the case study area and model used in the present experiment are described in Liu et al. (2022b).

338 **Table 1** Description of the SAC-SMA parameters and residual error model parameters.

Parameter	Description	Lower bound	Upper bound
SAC-SMA parameters			
UZTWM	Upper zone tension water capacity (mm)	1.0	150.0
UZFWM	Upper zone free water capacity (mm)	1.0	150.0
LZTWM	Lower zone tension water capacity (mm)	1.0	500.0
LZFPM	Lower zone primary free water capacity (mm)	1.0	1000.0
LZFSM	Lower zone supplementary free water capacity (mm)	1.0	1000.0
UZK	Upper zone free water lateral depletion rate (day^{-1})	0.1	0.5
LZPK	Lower zone primary free water depletion rate (day^{-1})	0.0001	0.025
LZSK	Lower zone supplementary free water depletion rate (day^{-1})	0.01	0.25
ZPERC	Percolation demand scale parameter	1.0	250.0
REXP	Percolation demand shape parameter	0.0	5.0
PFREE	Percolating water split parameter	0.0	0.1
PCTIM	Impervious fraction of the watershed area	0.0	0.1
ADIMP	Additional impervious area	0.0	0.4
RRC	Retention coefficient of routing linear reservoirs (day^{-1})	0.0	1.0
Residual error model parameters			
α	Auto-regressive parameter of the AR1 process	0.0	1.0
σ	Standard deviation of the AR1 innovations	0.0	1.0

339

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

The probabilistic model $\mathbf{Y}(\boldsymbol{\theta})$ is specified in transformed space where the output of the SAC-SMA hydrological model $h(\cdot)$ is corrupted with a random residual error term $\varepsilon(\cdot)$. Specifically, the truncated Gaussian AR1 process given in Eq. (8) is used:

$$\mathbf{Y}(\boldsymbol{\theta}) = z^{-1}\left[z\left[h(\boldsymbol{\theta}_h);\lambda\right] + \varepsilon(\boldsymbol{\theta}_\varepsilon);\lambda\right] \tag{8}$$

where $z(q;\lambda) = \left(q^{\lambda} - 1\right)/\lambda$ is the Box-Cox transformation with fixed parameter λ = 0.2 (Mcinerney et al., 2017). The residual error model $\varepsilon(\cdot)$ is characterized by a first-order autoregressive (AR1) process, $\varepsilon_{t+1} = \alpha\varepsilon_t + \xi_t$ with truncated Gaussian innovations $\xi_t \sim TN\left(0,\sigma^2\right)$ (Fenicia et al., 2018). The SAC-SMA parameters $\boldsymbol{\theta}_h$ and residual error model parameter $\boldsymbol{\theta}_\varepsilon = \{\alpha,\sigma\}$ constitute the parameters $\boldsymbol{\theta}$ to be jointly inferred.

We select a vector of hydrological signatures as summary statistics of the ABC sampling algorithms. A detailed description of each signature is provided in **Appendix A**. The distance metric in ABC is then computed as the average (by magnitude) relative error across all N_s signatures ($N_s = 8$ in this case study):

$$\rho\left(\mathbf{s}\left(\widetilde{\mathbf{Y}}\right),\mathbf{s}\left(\mathbf{Y}(\boldsymbol{\theta})\right)\right) = \frac{1}{N_s}\sum_{j=1}^{N_s}\frac{\left|\mathbf{s}_j\left(\widetilde{\mathbf{Y}}\right) - \mathbf{s}_j\left(\mathbf{Y}(\boldsymbol{\theta})\right)\right|}{\mathbf{s}_j\left(\widetilde{\mathbf{Y}}\right)} \tag{9}$$

This is different from the settings of Kavetski et al. (2018) and also Fenicia et al. (2018). However, as a sufficient number of iterations is implemented, the choice of the distance metric and its tolerance (here, with an initial value of 0.3) has no significant impact on fair comparison of SABC and mSABC.

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023


Hydrology and Earth System Sciences
Discussions

In our case study, following the practice of Fenicia et al. (2018), the original SABC algorithm is configured to return 5000 posterior samples from a total of 2×10^6 iterations. The number of iterations needed by mSABC to achieve the computational convergence of Markov Chains is determined by plotting the posterior parameter distributions as a function of the number of iterations (see **Sect. 4.1**). The size of acceptable solutions by mSABC is set identical to that of SABC. The DREAM-Core sampler is executed using the default settings of the algorithmic variables specified previously. To preserve detailed balance and reversibility of the Markov Chains (Ter Braak and Vrugt, 2008; Vrugt et al., 2009), the $N = 3$ independent chains are run sequentially.

4 Results and discussion

4.1 Computational convergence

We used boxplots of the marginal posterior parameter distributions to examine the convergence of mSABC over the course of 7×10^5 iterations. **Figure 2** presented how the posterior parameter distributions derived with mSABC changed as a function of the number of iterations. We displayed the results of a representative set of five SAC-SMA parameters, i.e., PCTIM, ADIMP, LZFSM, LZFPM, and LZPK, and two residual error model parameters (α and σ). For comparison, the posterior parameter distributions derived with the original SABC were plotted as a benchmark. The posteriors derived with mSABC showed a converging pattern and stabilized after approximately 5×10^5 iterations. Therefore, in the present case study, a total number of 5×10^5 iterations is deemed as sufficient for mSABC to converge to the correct posterior.

We believe that in the absence of a formal convergence proof of the mSABC algorithm theoretically, the computational convergence of mSABC needs to be benchmarked on the original SABC algorithm which has already been proved to

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023

converge to the correct target distribution in previous applications (Fenicia et al., 2018). The converging pattern of the posteriors derived with mSABC implies that replacement of a simplistic RWM sampler with DREAM-Core sampling in SABC exerts no significant impact on the convergence to the correct posterior. The posteriors derived with mSABC, after approximately 2×10^5 iterations, achieve almost identical results to those derived with the original SABC algorithm. For parameters like LZPK and σ, there is a remarkable distinction between the posteriors derived with mSABC after 5×10^5 iterations and SABC, respectively. In the current calibration problem involving excessive number of parameters, the SABC algorithm implemented using a simplistic RWM sampler introduces additional bias to the posterior parameter distributions and fails to correctly infer the target distribution.

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023

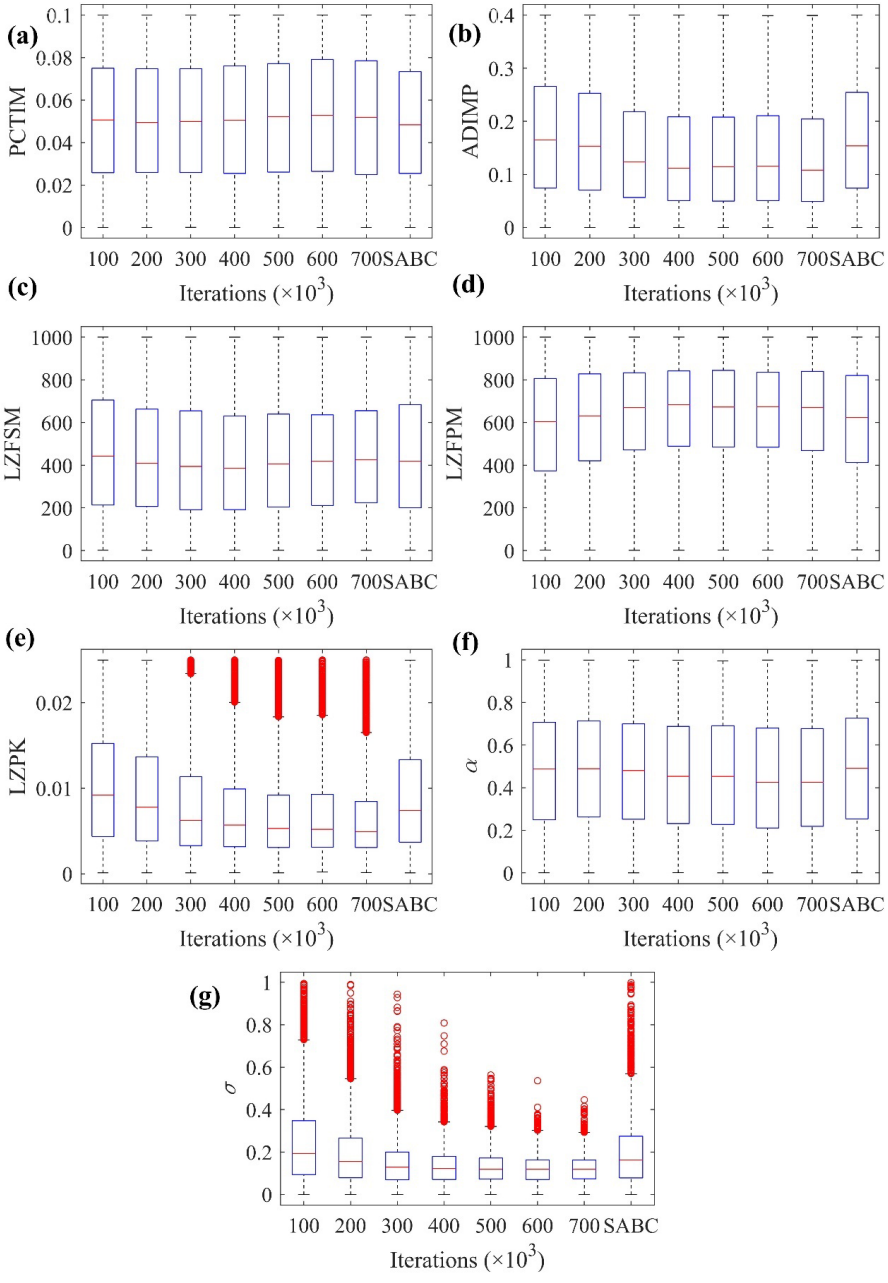


Fig. 2 Evolution of the posterior parameter distributions derived with mSABC as a function of iterations. The posteriors derived with the original SABC algorithm are provided as a benchmark.

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023


Hydrology and Earth System Sciences
Discussions

4.2 Sampling efficiency and cost comparison

The sampling efficiency of SABC and mSABC was compared in terms of the final value of τ (τ^e), acceptance rate, AR (%) and number of function evaluations, FEs needed for posterior exploration (**Table 2**). The number of FEs is calculated as: FEs = $T_{init} + T_{iter}*N$, where N is the number of Markov Chains ($N = 1$ for SABC, and $N = 3$ for mSABC), T_{init} and T_{iter} are the number of iterations used for the "Initialization" and "Iteration" step, respectively. The SABC algorithm has an AR value of 0.96%, and requires around 2 million FEs to generate 5000 posterior samples. The mSABC algorithm is far more efficient (AR = 7.26%), and needs about 70% of SABC FEs to obtain identical number of posterior samples. This constitutes a more than 7.5 times difference in sampling efficiency, and favors the use of mSABC for uncertainty quantification of complex, computationally expensive models. This finding confirms the superiority of DREAM over the optimal RWM sampler in previous work (Vrugt et al., 2009; Laloy and Vrugt, 2012). The advantage of DREAM-Core over RWM in enhancing the sampling efficiency still holds when incorporated into proposal generation in the SABC algorithms.

The final value of $\tau^e(t)$ provides valuable information about the bias from an approximation to the posterior (Albert et al., 2015). Given an identical initial value of $\tau^e(t)$ ($\tau = 0.3$), mSABC obtained a final value of $\tau^e(t)$ close to zero, significantly lower than that of SABC. Therefore, mSABC achieved higher numerical accuracy in approximating the posterior. A vivid description of how fast $\tau^e(t)$ was decreased in the "Iteration" step was presented in **Fig. 3**, where the value of $\tau^e(t)$ was plotted as a function of the number of iterations and FEs, respectively. mSABC showed a fast convergence to an approximation of the posterior. However, SABC maintained a slow

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023

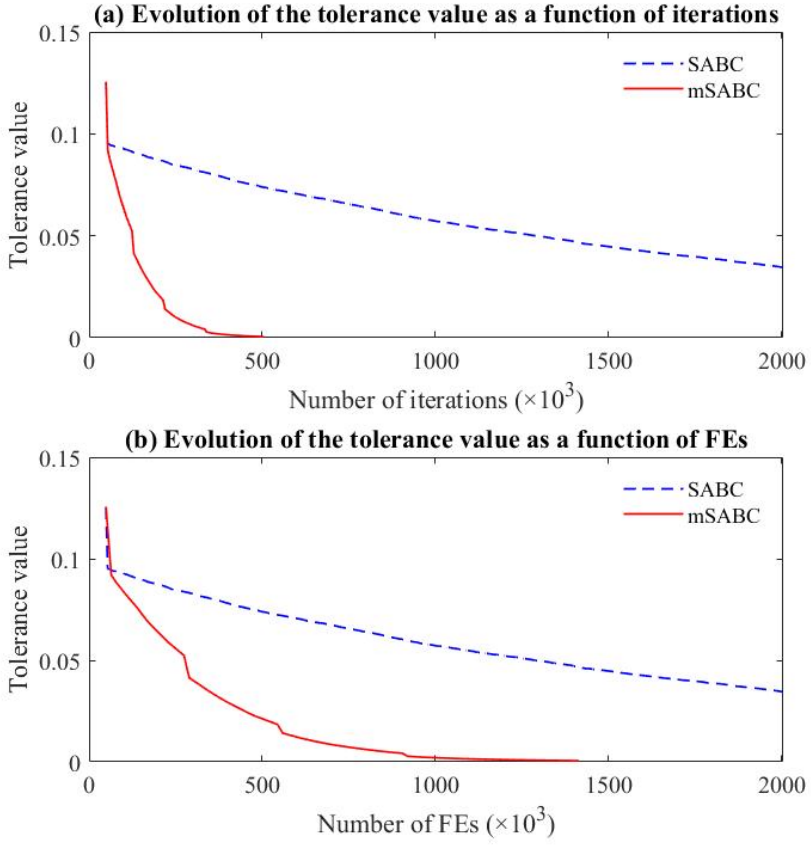
convergence speed throughout the "Iteration" step. A total number of approximately 1.5×10^5 iterations or 3.5×10^5 FEs for mSABC leads to a more relaxed final value of $\tau^e(t) = 0.035$, close to that achieved by the original SABC algorithm. This agrees with the findings of **Fig. 2**, where mSABC acquired almost identical posteriors to SABC after around 2×10^5 iterations. We concluded that mSABC helped to accelerate convergence to an approximate of the posterior at the same time not introduce additional bias with lower computational costs.

Table 2 Comparison of the sampling efficiency of SABC and mSABC in terms of the final value of $\tau^e(t)$, acceptance rate, AR (%) and number of function evaluations, FEs needed for posterior exploration.

Algorithm	Initial $\tau(t)$	Initial $\tau^e(t)$	Final $\tau^e(t)$	AR (%)	FEs
SABC	0.3	0.126	0.035	0.96	2.0×10^6
mSABC	0.3	0.126	0.00054	7.26	1.4×10^6

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023

Hydrology and Earth System Sciences
Discussions



435

436 **Fig. 3** Evolution of the tolerance value $\tau^e(t)$ as a function of **(a)** the number of
437 iterations and **(b)** the number of function evaluations (FEs) throughout the "Iteration"
438 step of SABC and mSABC.

439

440 **4.3 Parameter inference**

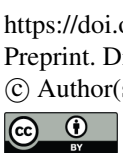
441 **Figure 4** presented histograms of the posterior parameter distributions derived with
442 SABC and mSABC respectively. We displayed the results of a representative set of five
443 SAC-SMA parameters, i.e., PCTIM, ADIMP, LZFSM, LZFPM, and LZPK, and two
444 residual error model parameters (α and σ). For most parameters, mSABC exhibited
445 sharper functional shapes and lower parameter uncertainties than the original SABC

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023

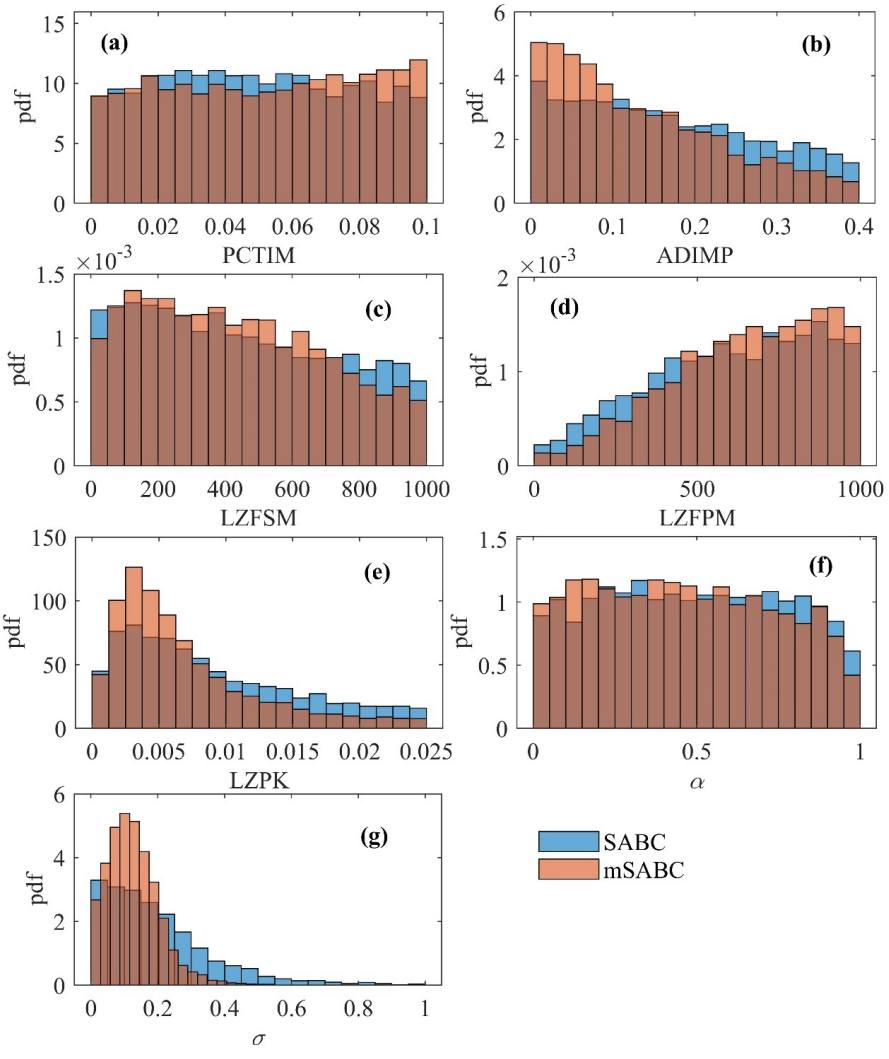

Hydrology and Earth System Sciences
Discussions

algorithm. The use of adaptive MCMC sampler (i.e., DREAM-Core sampling) in SABC helped to locate the high probability density region of the parameter space efficiently, which confirmed the findings of previous studies (Blasone et al., 2008). Not Surprisingly, most histograms extended a large part of the prior parameter ranges. For parameters like PCTIM and α, the posteriors showed no evident differences from the uniform priors. The parameters were poorly defined by both SABC and mSABC algorithms implementing using a vector of subjectively chosen signature indices. It is likely that these summary metrics are not sufficient, at least for the present case study. Ideally, the summary statistics of ABC should contain as much information as the original calibration dataset (Sadegh and Vrugt, 2013; Sadegh and Vrugt, 2014). These eight metrics are expected to extract only a limited portion of available information in the discharge time series. This has direct impact on constraining the parameter space, resulting in poor parameter inference. Unfortunately, there is still no common practice in identifying a set of (approximately) sufficient summary statistics in ABC applications (Liu et al., 2022a). The sufficiency issue of ABC is beyond the scope and aim of our study, and a further discussion on this topic is provided in literature (Fenicia et al., 2018; Kavetski et al., 2018).

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023



Hydrology and Earth System Sciences
Discussions



464

465 **Fig. 4** Comparison of probability density function (pdf) of the posterior parameter
466 distributions derived with SABC and mSABC. We presented the results of a variety of
467 five SAC-SMA parameters (**a-e**) and two residual error model parameters (**f-g**).

468

469 **4.4 Signature predictions**

470 We compared the predictive distributions of signatures derived with the original
471 SABC and mSABC algorithm over the validation period (**Fig. 5**). The performance of
472 signature predictions was reported in a relative error sense. The marginal distributions

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023

473 of predicted signatures generally center around zero with the exception of the heavily
474 skewed predictive distributions of RR, RLD and ACF obtained by both SABC and
475 mSABC. The systematic errors in RR suggest potential model deficiencies and
476 measurement errors, yet this requires more detailed analysis. The overestimation of
477 ACF is largely attributed to the introduction of AR1 process, which over-conditions the
478 autocorrelation structure. Overall, both algorithms suggested to provide an acceptable
479 reproduction of signatures, increasing the confidence in the application of these two
480 algorithms to the present case study.

481 Compared to SABC, mSABC achieved a generally better capture of the signatures,
482 with a median value close to zero and thinner upper and lower tails. Examples were
483 illustrated for Q_{10} and HPC, where a narrower predictive distribution centered on the
484 zero level was obtained. But counterexample signatures do exist. For example, with
485 respect to BFI, SABC generated better signature predictions in terms of the median of
486 the predictive distribution. A significantly larger deviation from the observed BFI was
487 achieved by mSABC. This finding was perhaps not surprising, since the ABC distance
488 metric to be minimized was formulated as the average of signature deviations for all
489 eight signatures (see Eq. (9)). The conflict between the predictive performances of Q_{10},
490 HPC and BFI could be possibly attributed to the difference in the specific features of
491 hydrological behavior these signatures aim to characterize. Previous work (Shafii et al.,
492 2015) has shown that there exists a clear, strong tradeoff between the reproduction of
493 high-flow regime (here, represented by Q_{10} and HPC) and low-flow regime (here,
494 represented by BFI) simultaneously. The improvement in the reproduction of Q_{10} and
495 HPC was achieved at the sacrifice of the accuracy in reproducing BFI.

496 We also conducted a comparison of the distribution of the ABC distances
497 associated with the 5000 posterior samples over the validation period. The distance

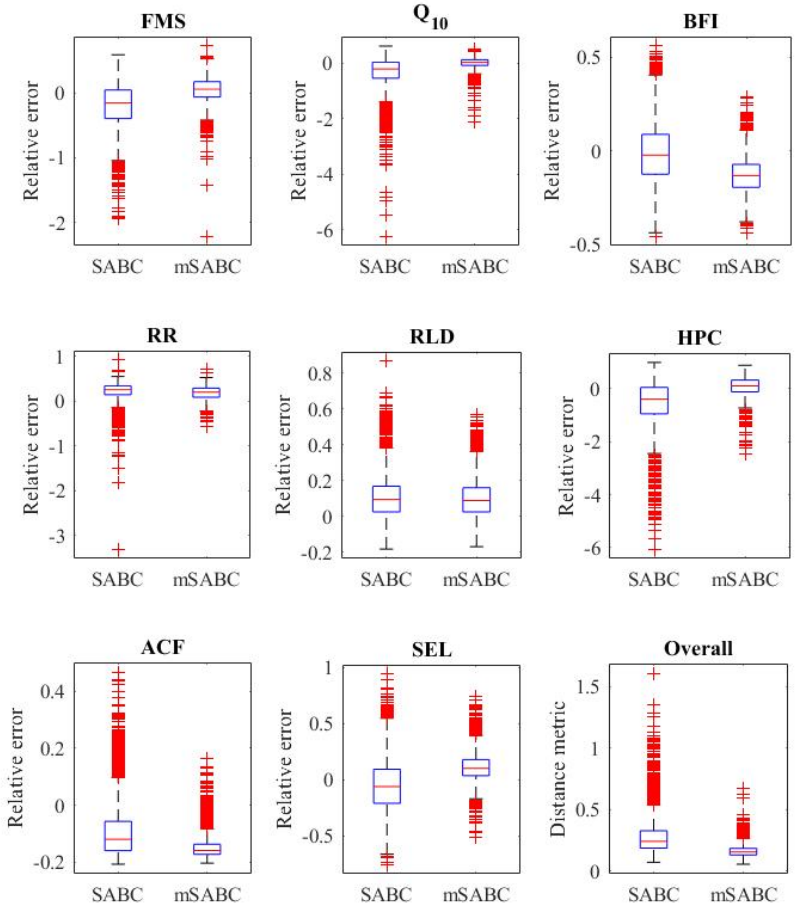
https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023

metric reflects the average relative error across all signatures. Compared to SABC, mSABC achieved higher numerical accuracy with a thinner upper tail of the predictive distribution. The median also decreased from 0.24 for SABC to 0.16 for mSABC. The skewness of the distance distributions confirmed the findings of Kavetski et al. (2018) and Fenicia et al. (2018), where the convergence of SABC (mSABC) to the approximate posterior does not necessarily imply that the achieved ABC distances are negligible for all posterior samples. For example, the mSABC distances fell in the narrow range from 0.06 to 0.74, whereas the SABC distances ranged from as low as 0.07 to as high as 1.43. These discrepancies primarily relate to the difficulties in matching all signatures simultaneously, especially in the presence of strong conflicts among them.

Both SABC and mSABC algorithms require a series of decreasing tolerance values for the ABC acceptance test. In the early process of lowering the tolerance, the model with a larger ABC distance has greater chance to be pooled in the acceptable models. This explains why there exist a small portion of "ill-posed" solutions in the last 5000 samples, suggested by heavy upper and lower tails in signature predictions. Compared to the original SABC algorithm, mSABC achieved a faster speed of lowering the tolerance (see **Fig. 3**), which resulted in a lower tolerance value and stricter acceptance criteria for signature deviations. The portion of unrealistic signature predictions was significantly reduced by mSABC.

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023

Hydrology and Earth System Sciences
Discussions



518

519 **Fig. 5** Predictions of streamflow signatures over the validation period. The distribution
520 of the ABC distance, computed as the average of the absolute relative errors for all
521 signatures, is also provided in the "Overall" plot.

522

523 **4.5 Streamflow predictions**

524 A comparison of daily streamflow predictions during the validation period
525 obtained by the original SABC and mSABC algorithm respectively was presented in
526 **Fig. 6**, where the 95% prediction limits and predictive quantile-quantile (PQQ) plots

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023


Hydrology and Earth System Sciences
Discussions

were compared. Both algorithms achieved a generally satisfying capture of streamflow time series over the validation period. mSABC produced a narrower uncertainty band at the cost of underestimating the peak flows across the high-flow period. This was suggested by a significantly lower *RB* value and also a reduction of *CR*. This can be largely attributed to (i) inaccurate observational datasets and (ii) the choice of hydrological signatures.

The former reason is inherent in almost all lumped hydrological models including SAC-SMA which only accept mean areal estimates as model drivers. The information contained in local measurements of uncommon large rainstorms is valuable for simulation of the highest flows, but is inevitably lost when averaged across the whole catchment. Besides, the estimated reservoir inflows used for model calibration and validation introduces additional errors into the model predictions. The data-related uncertainties contribute to the discrepancies between the observed and predicted flows regardless of the ABC sampling algorithms used.

For the latter reason, we intentionally use Q_{10}, the 10% flow exceedance values of streamflow to extract relevant information about high-flow regime in the DRR, although signatures such as the maximum (peak) flows can clearly improve the predictions in terms of the highest flows. However, they are more likely to be influenced by the flow errors (Westerberg and Mcmillan, 2015), resulting in a biased prediction. According to the findings of signature predictions in **Fig. 5**, mSABC decreased the proportion of solutions with large overestimation of high flows in the final set of posterior samples. This directly lowers the upper limits of uncertainty band. In this context, a larger number of high-flow records enveloped by SABC is merely the consequence of unrealistic wide uncertainty bands but not the benefit of SABC itself.

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023


Hydrology and Earth System Sciences
Discussions
Open Access
EGU

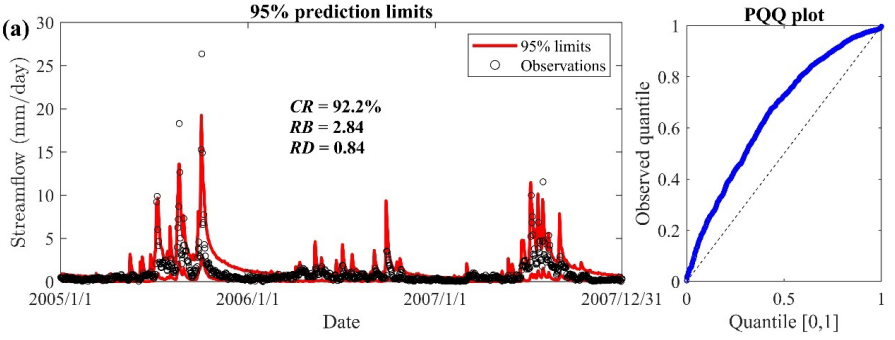
551 This can also be confirmed by better performance of PQQ plots with smaller deviation
552 from the 1:1 line and better reliability of probabilistic predictions acquired by mSABC.

553 **Figure 6** also provided a quantitative assessment of 95% prediction limits over the
554 validation period. Higher values of *CR* and lower values of *RB* and *RD* correspond to
555 better model predictions. Compared to the original SABC algorithm, mSABC achieved
556 lower values of *RB* and *RD*, accompanied by undesirable decrease of *CR*. On the whole,
557 mSABC suggested a better overall performance in predicting streamflow time series.

558 We have noticed that a clear, strong tradeoff exists between the performances of
559 *CR* and *RB* (*RD*). A high containing ratio, a narrow uncertainty band, and a small
560 deviation from the observed hydrograph, which represent three competing yet desirable
561 characteristics for the prediction limits, cannot be realized simultaneously (Xiong et al.,
562 2009). This phenomenon is prevalent in the hydrological literature (Xiong et al., 2009;
563 Sadegh and Vrugt, 2014; Zhou et al., 2016), which complicates direct comparison of
564 these techniques.

565

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023



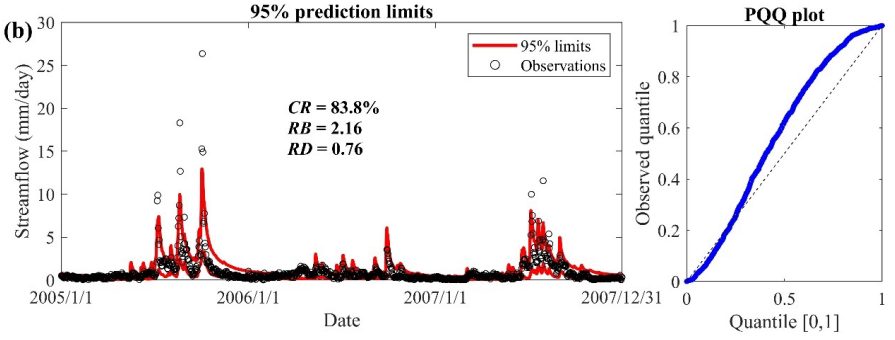


Fig. 6 Predictions of daily streamflow time series over the validation period obtained using (**a**) the original SABC algorithm and (**b**) mSABC algorithm. We display the results of 95% prediction limits (left) and Predictive Quantile-Quantile (PQQ) plots (right). The performance metrics of 95% prediction limits include the Containing Ratio (*CR*), Relative Band-width (*RB*), and Relative Deviation amplitude (*RD*).

We also assessed the performance of the trajectory consisting of the middle points of the 95% prediction limits against the observed hydrograph (**Table 3**). The performance metrics include the *RMSE*, correlation coefficient, CC and percent bias, *PBIAS*. Lower values of *RMSE* and *PBIAS* in the absolute sense and a higher value of *CC* indicate better model performance. Compared to SABC, the mSABC algorithm acquired a worse performance of *RMSE*, a slight decline of *CC*, and better performance of *PBIAS*. A negative *PBIAS* value for SABC suggested an overestimation of the observed hydrograph, which confirmed the finding that the 95% prediction limits derived with SABC was unrealistic wide in **Fig. 6**. The prediction produced by mSABC,

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023

on the contrary, suggested an underestimation of the observation, but with smaller magnitude of error in the absolute sense. On the whole, mSABC does not significantly deteriorate jeopardize the model performance with respect to the middle points of the 95% prediction limits.

Table 3 Evaluation of the middle points of the prediction limits over the validation period. *RMSE* – Root Mean Squared Error (mm/day); *CC* – Correlation Coefficient; *PBIAS* – Percent Bias (%).

Algorithm	*RMSE* (mm/day)	*CC*	*PBIAS* (%)
SABC	0.89	0.86	-5.23
mSABC	0.97	0.85	4.07

5 Conclusions

The original SABC algorithm implements a rather simplistic RWM sampler to generate a proposal from a single Markov Chain in the process of adaptively tightening an initial loose tolerance value (Albert et al., 2015). This scheme may be adequate for relatively simple low-dimensional inference problems, but is not likely to achieve a fast convergence and high numerical accuracy for more complex posterior exploration problems (Ter Braak, 2006; Ter Braak and Vrugt, 2008; Vrugt et al., 2008; Vrugt et al., 2009; Laloy and Vrugt, 2012). In this paper, we have demonstrated the potential of improving the original SABC algorithm by implementing an adaptive Differential Evolution algorithm with self-adaptive randomized subspace sampling (Vrugt et al., 2009), here referred to as DREAM-Core for speeding up convergence to an approximation to the posterior while maintaining equivalent or better predictive abilities. Through a comparison of the inference results using RWM and DREAM-Core for proposal generation, we demonstrated the following conclusions:

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023


Hydrology and Earth System Sciences
Discussions

(1) The use of DREAM-Core sampling in mSABC has little impact on the computational convergence of the sampled Markov Chains. It requires around 5×10^5 iterations for mSABC to converge to an approximation of the correct posterior by benchmarking it against the original SABC algorithm.

(2) The modified SABC (mSABC) algorithm is far more efficient with higher acceptance probability, and requires a lower number of function evaluations (FEs) to achieve a much lower final tolerance value.

(3) The mSABC algorithm acquires sharper functional shapes of the posterior parameter distributions, and helps to locate the high probability density region of the parameter space efficiently.

(4) The mSABC algorithm achieves a generally better capture of signature predictions over the validation period. The ABC distances associated with posterior samples are largely reduced in terms of both median values and overall distributions.

(5) The mSABC algorithm achieves a better overall prediction of streamflow time series over the validation period. A quantitative assessment of streamflow predictions favors mSABC for reliable probabilistic predictions and tighter uncertainty band with an undesirable decrease of the Containing Ratio (*CR*).

Future work is to further investigate its comparison with other state-of-the-art ABC sampling algorithm, e.g., DREAM$_{(ABC)}$ algorithm (Sadegh and Vrugt, 2014), in terms of numerical accuracy and efficiency in real-world applications.

Appendix A: Description of Hydrological Signatures

Table A1 in this appendix presents a description of a vector of eight hydrological signatures employed in the case study.

630 **Table A1** Hydrological signatures employed in the case study.

Group	Signature	Name	Description	Unit
Flow duration curve (FDC)	FMS	FDC midsegment slope	Slope of the FDC between the log-transformed 20% and 70% flow exceedance values of streamflow (Yilmaz et al., 2008)	[-]
	Q_{10}	Flow percentile	10% flow exceedance values of streamflow divided by median flow (Clausen and Biggs, 2000)	[-]
Flow dynamics	BFI	Base flow index	Fraction of base flow within the total flow (Eckhardt, 2008)	[-]
	ACF	Lag-1 autocorrelation coefficient	Lag-1 autocorrelation coefficient of daily streamflow time series	[-]
Frequency of flow events	RLD	Rising limb density	Number of peaks divided by the total time the hydrograph is rising (Shamir et al., 2005)	[day^{-1}]
	HPC	High pulse count	Number of occurrences during which flow remains 3 times median daily flow (Clausen and Biggs, 2000)	[-]
Rainfall-Runoff	RR	Runoff ratio	Ratio of runoff volume to areal precipitation	[-]
	SEL	Streamflow elasticity	Sensitivity of streamflow to change of precipitation (Addor et al., 2018)	[-]

631

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023


Hydrology and Earth System Sciences
Discussions

Code and data availability

The Matlab source code of the original SABC is available on request from Dr. Dmitri Kavetski. The data that support the findings of this study, along with the code of the modified SABC program, are available from the corresponding authors upon reasonable request.

Author contribution

S. Liu designed the experiments, and prepare the original draft with contributions from all co-authors. D. X. She and L. P. Zhang contributed to the validation of the overall reproductivity of the experiments and the revision of the original draft. J. Xia and D. X. She were responsible for resources and fund acquisition.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

This research is sponsored by National Key Research and Development Program of China (Grant No. 2016YFC0402709). We appreciate Dr. Dmitri Kavetski for sharing the Matlab implementation of the original SABC algorithm. The source codes of 6-hourly SAC-SMA model are retrieved from the DREAM suite, which is available on request from Dr. Jasper A. Vrugt.

References

Addor, N., Nearing, G., Prieto, C., Newman, A. J., Le Vine, N., and Clark, M. P.: A ranking of hydrological signatures based on their predictability in space, Water Resour. Res., 54, 8792-8812, https://doi.org/10.1029/2018wr022606, 2018.

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023

654 Albert, C., Künsch, H. R., and Scheidegger, A.: A simulated annealing approach to
655 approximate Bayes computations, Stat. Comput., 25, 1217-1232,
656 https://doi.org/10.1007/s11222-014-9507-8, 2015.
657 Beaumont, M. A.: Approximate Bayesian Computation, Annu. Rev. Stat. Appl., 6, 379-
658 403, https://doi.org/10.1146/annurev-statistics-030718-105212, 2019.
659 Beaumont, M. A., Zhang, W. Y., and Balding, D. J.: Approximate Bayesian computation
660 in population genetics, Genetics, 162, 2025-2035, 2002.
661 Beaumont, M. A., Cornuet, J.-M., Marin, J.-M., and Robert, C. P.: Adaptive
662 approximate Bayesian computation, Biometrika, 96, 983-990,
663 https://doi.org/10.1093/biomet/asp052, 2009.
664 Beven, K. and Binley, A.: The future of distributed models: Model calibration and
665 uncertainty prediction, Hydrol. Process., 6, 279-298,
666 https://doi.org/10.1002/hyp.3360060305, 1992.
667 Beven, K. and Binley, A.: GLUE: 20 years on, Hydrol. Process., 28, 5897-5918,
668 https://doi.org/10.1002/hyp.10082, 2014.
669 Blasone, R.-S., Vrugt, J. A., Madsen, H., Rosbjerg, D., Robinson, B. A., and Zyvoloski,
670 G. A.: Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov
671 Chain Monte Carlo sampling, Adv. Water Resour., 31, 630-648,
672 https://doi.org/10.1016/j.advwatres.2007.12.003, 2008.
673 Burnash, R. J., Ferral, R. L., and McGuire, R. A.: A Generalized Streamflow Simulation
674 System: Conceptual Modeling for Digital Computers, Joint Fed.-State River Forecast
675 Cent., Sacramento, Calif.1973.
676 Clausen, B. and Biggs, B. J. F.: Flow variables for ecological studies in temperate
677 streams: groupings based on covariance, J. Hydrol., 237, 184-197,
678 https://doi.org/10.1016/S0022-1694(00)00306-1, 2000.

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023


Hydrology and Earth System Sciences
Discussions

679 Diggle, P. J. and Gratton, R. J.: Monte Carlo methods of inference for implicit statistical
680 models, J. Roy. Stat. Soc. B Met., 46, 193-227, https://doi.org/10.1111/j.2517-
681 6161.1984.tb01290.x, 1984.
682 Duan, Q. Y., Sorooshian, S., and Gupta, V.: Effective and efficient global optimization
683 for conceptual ranfall-runoff models, Water Resour. Res., 28, 1015-1031,
684 https://doi.org/10.1029/91WR02985, 1992.
685 Eckhardt, K.: A comparison of baseflow indices, which were calculated with seven
686 different baseflow separation methods, J. Hydrol., 352, 168-173,
687 https://doi.org/10.1016/j.jhydrol.2008.01.005, 2008.
688 Engeland, K. and Gottschalk, L.: Bayesian estimation of parameters in a regional
689 hydrological model, Hydrol. Earth Syst. Sci., 6, 883-898, https://doi.org/10.5194/hess-
690 6-883-2002, 2002.
691 Evin, G., Thyer, M., Kavetski, D., McInerney, D., and Kuczera, G.: Comparison of joint
692 versus postprocessor approaches for hydrological uncertainty estimation accounting for
693 error autocorrelation and heteroscedasticity, Water Resour. Res., 50, 2350-2375,
694 https://doi.org/10.1002/2013WR014185, 2014.
695 Fenicia, F., Kavetski, D., Reichert, P., and Albert, C.: Signature‐domain calibration of
696 hydrological models using Approximate Bayesian Computation: Empirical analysis of
697 fundamental properties, Water Resour. Res., 54, 3958-3987,
698 https://doi.org/10.1002/2017WR021616, 2018.
699 Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Toward improved calibration of
700 hydrologic models: Multiple and noncommensurable measures of information, Water
701 Resour. Res., 34, 751-763, https://doi.org/10.1029/97wr03495, 1998.

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023


Hydrology and Earth System Sciences
Discussions

Gupta, H. V., Clark, M. P., Vrugt, J. A., Abramowitz, G., and Ye, M.: Towards a comprehensive assessment of model structural adequacy, Water Resour. Res., 48, https://doi.org/10.1029/2011wr011044, 2012.

Haario, H., Saksman, E., and Tamminen, J.: Componentwise adaptation for high dimensional MCMC, Computation. Stat., 20, 265-273, https://doi.org/10.1007/BF02789703, 2005.

Kavetski, D., Fenicia, F., Reichert, P., and Albert, C.: Signature - domain calibration of hydrological models using Approximate Bayesian Computation: Theory and comparison to existing applications, Water Resour. Res., 54, 4059-4083, https://doi.org/10.1002/2017WR020528, 2018.

Laio, F. and Tamea, S.: Verification tools for probabilistic forecasts of continuous hydrological variables, Hydrol. Earth Syst. Sci., 11, 1267-1277, https://doi.org/10.5194/hess-11-1267-2007, 2007.

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

Liu, S., She, D., Zhang, L., and Xia, J.: A hybrid time- and signature-domain Bayesian inference framework for calibration of hydrological models: a case study in the Ren River basin in China, Stoch. Environ. Res. Risk A., https://doi.org/10.1007/s00477-022-02282-3, 2022a.

Liu, S., She, D., Zhang, L., Xia, J., Chen, S., and Wang, G.: Quantifying and reducing the uncertainty in multi-source precipitation products using Bayesian total error analysis: a case study in the Danjiangkou Reservoir region in China, J. Hydrol., https://doi.org/10.1016/j.jhydrol.2022.128557, 2022b.

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023

726 Marjoram, P., Molitor, J., Plagnol, V., and Tavare, S.: Markov Chain Monte Carlo
727 without Likelihoods, P. Nati. Acad. Sci. USA, 100, 15324-15328,
728 https://doi.org/10.1073/pnas.0306899100, 2003.
729 McInerney, D., Thyer, M., Kavetski, D., Lerat, J., and Kuczera, G.: Improving
730 probabilistic prediction of daily streamflow by identifying Pareto optimal approaches
731 for modeling heteroscedastic residual errors, Water Resour. Res., 53, 2199-2239,
732 https://doi.org/10.1002/2016WR019168, 2017.
733 McMillan, H., Krueger, T., and Freer, J.: Benchmarking observational uncertainties for
734 hydrology: rainfall, river discharge and water quality, Hydrol. Process., 26, 4078-4111,
735 https://doi.org/10.1002/hyp.9384, 2012.
736 Nott, D. J., Marshall, L., and Brown, J.: Generalized likelihood uncertainty estimation
737 (GLUE) and approximate Bayesian computation: What's the connection?, Water Resour.
738 Res., 48, https://doi.org/10.1029/2011WR011128, 2012.
739 Olden, J. D. and Poff, N. L.: Redundancy and the choice of hydrologic indices for
740 characterizing streamflow regimes, River Res. Appl,, 19, 101-121,
741 https://doi.org/10.1002/rra.700, 2003.
742 Sadegh, M. and Vrugt, J. A.: Bridging the gap between GLUE and formal statistical
743 approaches: approximate Bayesian computation, Hydrol. Earth Syst. Sci., 17, 4831-
744 4850, https://doi.org/10.5194/hess-17-4831-2013, 2013.
745 Sadegh, M. and Vrugt, J. A.: Approximate Bayesian Computation using Markov Chain
746 Monte Carlo simulation: DREAM(ABC), Water Resour. Res., 50, 6767-6787,
747 https://doi.org/10.1002/2014WR015386, 2014.
748 Schoups, G. and Vrugt, J. A.: A formal likelihood function for parameter and predictive
749 inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian
750 errors, Water Resour. Res., 46, https://doi.org/10.1029/2009WR008933, 2010.

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023


Hydrology and Earth System Sciences
Discussions

751 Shafii, M. and Tolson, B. A.: Optimizing hydrological consistency by incorporating
752 hydrological signatures into model calibration objectives, Water Resour. Res., 51, 3796-
753 3814, https://doi.org/10.1002/2014WR016520, 2015.
754 Shafii, M., Tolson, B., and Shawn Matott, L.: Addressing subjective decision-making
755 inherent in GLUE-based multi-criteria rainfall–runoff model calibration, J. Hydrol.,
756 523, 693-705, https://doi.org/10.1016/j.jhydrol.2015.01.051, 2015.
757 Shamir, E., Imam, B., Gupta, H. V., and Sorooshian, S.: Application of temporal
758 streamflow descriptors in hydrologic model parameter estimation, Water Resour. Res.,
759 41, https://doi.org/10.1029/2004wr003409, 2005.
760 Sisson, S. A., Fan, Y., and Tanaka, M. M.: Sequential Monte Carlo without likelihoods,
761 P. Nati. Acad. Sci. USA, 104, 1760-1765, https://doi.org/10.1073/pnas.0607208104,
762 2007.
763 Storn, R. and Price, K.: Differential evolution - A simple and efficient heuristic for
764 global optimization over continuous spaces, J. Global Optim., 11, 341-359,
765 https://doi.org/10.1023/A:1008202821328, 1997.
766 Tavare, S., Balding, D. J., Griffiths, R. C., and Donnelly, P.: Inferring coalescence times
767 from DNA sequence data, Genetics, 145, 505-518, 1997.
768 ter Braak, C. J. F.: A Markov Chain Monte Carlo version of the genetic algorithm
769 Differential Evolution: Easy Bayesian computing for real parameter spaces, Stat.
770 Comput., 16, 239-249, https://doi.org/10.1007/s11222-006-8769-1, 2006.
771 ter Braak, C. J. F. and Vrugt, J. A.: Differential Evolution Markov Chain with snooker
772 updater and fewer chains, Stat. Comput., 18, 435-446, https://doi.org/10.1007/s11222-
773 008-9104-9, 2008.
774 Thyer, M., Renard, B., Kavetski, D., Kuczera, G., Franks, S. W., and Srikanthan, S.:
775 Critical evaluation of parameter consistency and predictive uncertainty in hydrological

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023

776 modeling: A case study using Bayesian total error analysis, Water Resour. Res., 45,
777 https://doi.org/10.1029/2008WR006825, 2009.
778 Toni, T., Welch, D., Strelkowa, N., Ipsen, A., and Stumpf, M. P. H.: Approximate
779 Bayesian computation scheme for parameter inference and model selection in
780 dynamical systems, J. Roy. Soc. Interface, 6, 187-202,
781 https://doi.org/10.1098/rsif.2008.0172, 2009.
782 Turner, B. M. and Van Zandt, T.: A tutorial on approximate Bayesian computation, J.
783 Math. Psychol., 56, 69-85, https://doi.org/10.1016/j.jmp.2012.02.005, 2012.
784 Vrugt, J. A.: Markov chain Monte Carlo simulation using the DREAM software
785 package: Theory, concepts, and MATLAB implementation, Environ. Modell. Softw.,
786 75, 273-316, https://doi.org/10.1016/j.envsoft.2015.08.013, 2016.
787 Vrugt, J. A. and Beven, K. J.: Embracing equifinality with efficiency: Limits of
788 Acceptability sampling using the DREAM(LOA) algorithm, J. Hydrol., 559, 954-971,
789 https://doi.org/10.1016/j.jhydrol.2018.02.026, 2018.
790 Vrugt, J. A. and Sadegh, M.: Toward diagnostic model calibration and evaluation:
791 Approximate Bayesian computation, Water Resour. Res., 49, 4335-4345,
792 https://doi.org/10.1002/wrcr.20354, 2013.
793 Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M., and Robinson, B. A.:
794 Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward
795 with Markov chain Monte Carlo simulation, Water Resour. Res., 44,
796 https://doi.org/10.1029/2007WR006720, 2008.
797 Vrugt, J. A., ter Braak, C. J. F., Diks, C. G. H., Robinson, B. A., Hyman, J. M., and
798 Higdon, D.: Accelerating Markov Chain Monte Carlo simulation by Differential
799 Evolution with Self-Adaptive Randomized Subspace Sampling, Int. J. Nonlin. Sci.
800 Num., 10, 273-290, https://doi.org/10.1515/IJNSNS.2009.10.3.273, 2009.

https://doi.org/10.5194/hess-2022-414
Preprint. Discussion started: 31 January 2023

801 Weiss, G. and von Haeseler, A.: Inference of population history using a likelihood
802 approach, Genetics, 149, 1539-1546, 1998.
803 Westerberg, I. K. and McMillan, H. K.: Uncertainty in hydrological signatures, Hydrol.
804 Earth Syst. Sci., 19, 3951-3968, https://doi.org/10.5194/hess-19-3951-2015, 2015.
805 Xiong, L., Wan, M., Wei, X., and O'Connor, K. M.: Indices for assessing the prediction
806 bounds of hydrological models and application by generalised likelihood uncertainty
807 estimation, Hydrol. Sci. J., 54, 852-871, https://doi.org/10.1623/hysj.54.5.852, 2009.
808 Yadav, M., Wagener, T., and Gupta, H.: Regionalization of constraints on expected
809 watershed response behavior for improved predictions in ungauged basins, Adv. Water
810 Resour., 30, 1756-1774, https://doi.org/10.1016/j.advwatres.2007.01.005, 2007.
811 Yilmaz, K. K., Gupta, H. V., and Wagener, T.: A process-based diagnostic approach to
812 model evaluation: Application to the NWS distributed hydrologic model, Water Resour.
813 Res., 44, https://doi.org/10.1029/2007WR006716, 2008.
814 Zhou, R., Li, Y., Lu, D., Liu, H., and Zhou, H.: An optimization based sampling
815 approach for multiple metrics uncertainty analysis using generalized likelihood
816 uncertainty estimation, J. Hydrol., 540, 274-286,
817 https://doi.org/10.1016/j.jhydrol.2016.06.030, 2016.

818