Hydrol. Earth Syst. Sci. Discuss., 10, 4739–4797, 2013
www.hydrol-earth-syst-sci-discuss.net/10/4739/2013/
doi:10.5194/hessd-10-4739-2013

Hydrology and Earth System Sciences
Open Access
Discussions

Approximate Bayesian Computation in hydrologic modeling: equifinality of formal and informal approaches

M. Sadegh[1] **and J. A. Vrugt**[1,2]

[1]Department of Civil and Environmental Engineering, University of California, Irvine, 4130 Engineering Gateway, Irvine, CA 92697-2175, USA
[2]Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, the Netherlands

Received: 12 February 2013 – Accepted: 2 April 2013 – Published: 12 April 2013

Correspondence to: J. A. Vrugt (jasper@uci.edu)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

In recent years, a strong debate has emerged in the hydrologic literature how to properly treat non-traditional error residual distributions and quantify parameter and predictive uncertainty. Particularly, there is strong disagreement whether such uncertainty framework should have its roots within a proper statistical (Bayesian) context using Markov chain Monte Carlo (MCMC) simulation techniques, or whether such a framework should be based on a quite different philosophy and implement informal likelihood functions and simplistic search methods to summarize parameter and predictive distributions. In this paper we introduce an alternative framework, called Approximate Bayesian Computation (ABC) that summarizes the differing viewpoints of formal and informal Bayesian approaches. This methodology has recently emerged in the fields of biology and population genetics and relaxes the need for an explicit likelihood function in favor of one or multiple different summary statistics that measure the distance of each model simulation to the data. This paper is a follow up of the recent publication of Nott et al. (2012) and further studies the theoretical and numerical equivalence of formal (DREAM) and informal (GLUE) Bayesian approaches using data from different watersheds in the United States. We demonstrate that the limits of acceptability approach of GLUE is a special variant of ABC in which each discharge observation of the calibration data set is used as a summary diagnostic.

1 Introduction

In a common inverse problem, we wish to estimate the parameters, $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_d\}$ of a model, $\mathcal{H}$, given observations of the system behavior, $\tilde{\mathbf{Y}} = \{\tilde{y}_1, \ldots, \tilde{y}_n\}$. The observations or data are linked to the unknown parameters $\boldsymbol{\theta}^*$ through some physical system, $\Im$

$$\tilde{\mathbf{Y}} \leftarrow \Im(\boldsymbol{\theta}^*) + \boldsymbol{\varepsilon}, \tag{1}$$

HESSD
10, 4739–4797, 2013

where $\boldsymbol{\varepsilon} = \{\varepsilon_1, \ldots, \varepsilon_n\}$ is a $n \times 1$ vector of measurement errors. Examples of such problems are widespread in many different fields of study including medical imaging (Kaipio et al., 2004), reservoir characterization (Stenerud et al., 2008) and cosmology (Jimenez et al., 2004). When a model hypothesis, or simulator, $\mathbf{Y} \leftarrow \mathcal{H}(\boldsymbol{\theta}^*, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{x}}_0)$ of the physical process is available, one can model the data

$$\tilde{\mathbf{Y}} \leftarrow \mathcal{H}(\boldsymbol{\theta}^*, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{x}}_0) + \boldsymbol{e}, \tag{2}$$

where $\tilde{\boldsymbol{u}} = \{\tilde{u}_1, \ldots, \tilde{u}_n\}$ denotes the forcing data, $\tilde{\boldsymbol{x}}_0$ signifies the initial states, and $\boldsymbol{e} = \{e_1, \ldots, e_n\}$ includes observation error (input and calibration data) as well as error due to the fact that the simulator, $\mathcal{H}(\boldsymbol{\theta}^*|\cdot)$ may be systematically different from reality, $\Im(\boldsymbol{\theta}^*)$ for the parameters $\boldsymbol{\theta}^*$. The latter may arise from, e.g. numerical error, spatial discretization and improper (conceptual) model formulation.

Figure 1 provides an overview of possible error sources that affect our ability to correctly describe the physical system, $\Im(\boldsymbol{\theta}^*)$ of interest. Forcing data, model parameter, model state, and calibration data error are represented with a probability density function (pdf), whose statistical properties are typically unknown. Errors in the modeled (5) output, $y_{t(t>0)}$ and (6) state, $\boldsymbol{x}_{t(t>0)}$ dynamics originate from a wide variety of different error sources, including (1) inadequate and/or incomplete knowledge of the model parameters, $\boldsymbol{\theta}^*$ (2) errors in the input (forcing) data, $\boldsymbol{u}$ and (3) initial states, $\boldsymbol{x}_0$ (4) structural inadequacies in the model equations, and/or improper dimensionality of the state space, and (7) errors in the calibration data, $\tilde{y}_{t(t>0)}$. The mathematical operator $\bigotimes$ (also called "likelihood function") is used to judge the distance between the model predictions and corresponding calibration data. This function should explicitly recognize the contribution and role of each individual error source in determining the error residual, but is very difficult to specify correctly with very weak prior information, and hence the pitchfork symbol is used.

Within the context of hydrologic modeling, measured rainfall depths, and estimates of (potential) evapotranspiration typically constitute the main forcing data. These two input variables strongly determine the simulated streamflow at interior points and the

HESSD
10, 4739–4797, 2013

catchment outlet, surface runoff, soil moisture fluxes and storage of water in the catchment. Examples of model states are soil moisture, groundwater table depth, and hydraulic heads (amongst others). Their knowledge is beneficial to adequately represent the storage of water in the variably saturated zone and groundwater, and hence ensure an adequate model calibration. Finally, calibration data often involves time series of (spatially-distributed) streamflow observations, or time-lapse measurements of tracer concentrations. Inevitably, each of these data sources is subject to uncertainty, which severely complicates parameter estimation and quantification of model structural errors.

During the past 4 decades much research has been devoted to the development of computer based methods for fitting hydrologic models to calibration data (e.g. streamflow, water chemistry, groundwater table depth, soil moisture, snow water equivalent). That research has primarily focused on six different issues: (1) the development of specialized objective functions that appropriately represent and summarize the errors between model predictions and observations, (2) the search for efficient optimization algorithms that can reliably solve the hydrologic model calibration problem, (3) the determination of the appropriate quantity and most informative kind of data, (4) the selection of an appropriate numerical solver for the partially structured differential and algebraic equation systems of hydrologic models, (5) the representation of uncertainty, and (6) the development of methods for inferring and refining the mathematical structure and process equations of hydrologic models.

Research into error residual distributions had led to the development of a suite of different (hierarchical) likelihood functions for measuring the "closeness" between the model simulations (predictions) and the corresponding data (Ibbitt and O'Donnell, 1974; Sorooshian and Dracup, 1980; Kuczera, 1983a; Bates and Campbell, 2001; Kavetski et al., 2006a; Marshall et al., 2007; Schoups and Vrugt, 2010a; Smith et al., 2010). Recent work by Schoups and Vrugt (2010a) has resulted in a generalized likelihood function that encapsulates many of the existing likelihood functions in the

HESSD
10, 4739–4797, 2013

hydrologic literature, but with additional flexibility to simultaneously account for correlated, heteroscedastic, and nontraditional error residual distributions.

Research into optimization methods has led to the development of a wide variety of different search methods. Whereas initial approaches utilized local search principles that seek iterative improvement of the objective function from a single starting point in the parameter space (Ibbitt, 1972; Johnston and Pilgrim, 1976; Sorooshian and Dracup, 1980; Restrepo, 1982; Kuczera, 1983a,b; Gupta and Sorooshian, 1983; Sorooshian et al., 1983b; Troutman, 1985a,b), problems with parameter insensitivity, curved ridges, local minima, and multiple different regions of attraction has stimulated the development of population based search algorithms that use multiple different points concurrently to locate the global optimum (Wang, 1991; Duan et al., 1992; Yapo et al., 1998; Seibert, 2000; Khu and Madsen, 2005; Chu et al., 2010). In this regard, the Shuffled Complex Evolution global optimization algorithm of Duan et al. (1992) has shown to be effective and efficient in calibrating conceptual watershed models. Recent developments include simple randomized adaptation (Mazi et al., 2004; Tolson and Shoemaker, 2007), multimethod ensemble (Vrugt and Robinson, 2007; Vrugt et al., 2009b), and filtering based (Pauwels, 2008) parameter estimation methods that further improve search efficiency and reliability.

Research into the information content of data has led to the understanding that it is not the length of the data that matters, but the variability of the observed discharge data (Kuczera, 1982; Sorooshian et al., 1983a; Gupta and Sorooshian, 1985; Yapo et al., 1996). Wet and dry periods are both required to make sure that all the different components of the watershed model are excited and the different parameters can be estimated from the calibration data. Post-audit simulations presented in Vrugt et al. (2002) using a Bayesian analysis, adaptive Random Walk Metropolis resampling, and value of information (VOI) framework has demonstrated that only a few (daily) streamflow data measurements are necessary to reliably calibrate a conceptual hydrologic model. The remaining data contain redundant information and could be used to evaluate the reliability of the actual model structure.

HESSD
10, 4739–4797, 2013

Research into numerical solvers has demonstrated that explicit (Euler based) time-stepping schemes introduce considerable streamflow simulation errors and spurious local minima, "pits" and irregularities in the objective function space (Kavetski et al., 2003, 2006c; Kavetski and Clark, 2010; Schoups et al., 2010b). These findings provide a deeper understanding of the convergence problems of local search methods, and demonstrate a need for implicit solvers that iteratively adjust the integration time step based on the state dynamics.

Research into the characterization of uncertainty has resulted in formal and informal statistical approaches. While initial attempts have focused primarily on methods to quantify parameter uncertainty (Beven and Binley, 1992; Freer et al., 1996; Gupta et al., 1998; Kuczera and Parent, 1998; Vrugt et al., 2002; Wagener et al., 2003; Beven, 2006; Vrugt and Robinson, 2007), emerging approaches include state-space filtering (Vrugt et al., 2005; Moradkhani et al., 2005a,b; Slater and Clark, 2006; Reichle, 2008; Salamon and Feyen, 2009; DeChant and Moradkhani, 2012; Vrugt et al., 2012), multi-model averaging (Butts et al., 2004; Georgakakos et al., 2004; Ajami et al., 2007), and various (non)Bayesian approaches to treat individual error sources and assess predictive uncertainty (Montanari and Brath, 2004; Vrugt et al., 2005; Kavetski et al., 2006a,b; Kuczera et al., 2006; Huard and Mailhot, 2006; Jacquin and Shamseldin, 2007; Fenicia et al., 2007; Marshall et al., 2007; Montanari and Grossi, 2008; Vrugt et al., 2008a,b; Reichert and Mieleitner, 2009; Solomatine and Shrestha, 2009; Kuczera et al., 2010; Renard et al., 2011; Rings et al., 2012). Much progress has also been made in the treatment of forcing data error (Clark and Slater, 2006; Kavetski et al., 2006a,b; Vrugt et al., 2008a), development of a formal hierarchical framework to formulate, build and test different watershed models (Clark et al., 2008), and algorithms for efficient sampling of parameter and predictive uncertainty distributions (Kuczera and Parent, 1998; Vrugt et al., 2008a; Kuczera et al., 2010; Laloy and Vrugt, 2012).

Finally, research into structural adequacy has resulted in data-based mechanistic (Young, 2002, 2013), data assimilation (Vrugt et al., 2005; Smith et al., 2008; Bulygina and Gupta, 2011), and other stochastic techniques (Reichert and Mieleitner, 2009) for

HESSD
10, 4739–4797, 2013

inference and iterative refinement of the mathematical structure of conceptual hydrologic models. This has led to the understanding that discharge data contain sufficient information to warrant the identification of a suitable model structure that mimics as closely and consistently as possible the observed watershed behavior at the temporal and spatial scale of measurement.

Most of these development assume input data and model structural errors to be "negligibly small" or to be somehow "absorbed" into the output error residuals. The residuals are then expected to behave statistically similar as the calibration data measurement error. These assumptions are statistically convenient but typically not borne out of the actual probabilistic properties of the residual errors which may show changing bias, variance (heteroscedasticity), skewness, and correlation structures under different hydrologic conditions (and for different parameter sets). This is in part due to the presence of model structural and forcing (input) data errors whose contribution may, in general, be substantially larger than the (calibration) data measurement error. These errors do not necessarily have any inherent probabilistic properties that can be exploited in the construction of an explicit likelihood function. For linear systems it is known that ignoring such errors will lead to bias in the estimates of parameter values. The strong and generally difficult to justify assumptions about the nature of the errors have led Beven and coworkers to advocate informal statistical approaches using the Generalized Likelihood Uncertainty Estimation (GLUE) methodology (Beven and Binley, 1992; Beven, 1993, 2006, 2008; Beven and Freer, 2001; Beven et al., 2008).

The origins of the GLUE method lie in trying to deal with uncertainty estimation problems for which simple explicit (theoretical) likelihood assumptions do not seem appropriate. The GLUE methodology rejects the traditional statistical basis for the likelihood function in favor of finding a set of representations (model inputs, model structures, model parameter sets, model errors) that are behavioral in the sense of being acceptably consistent with the (non-error-free) observations. An informal likelihood measure is used to avoid over conditioning and exclude parts of the model (parameter) space that might provide acceptable fits to the data and be useful in prediction. Since its

HESSD
10, 4739–4797, 2013

introduction in 1992, GLUE has found widespread application for uncertainty assessment in many fields of study, including modeling of the rainfall-runoff transformation (Beven and Binley, 1992; Freer et al., 1996; Lamb et al., 1998), soil erosion (Brazier et al., 2001), tracer dispersion in a river reach (Hankin et al., 2001), groundwater and well capture zone delineation (Feyen et al., 2001; Jensen, 2003), unsaturated zone (Mertens et al., 2004), flood inundation (Romanowicz et al., 1996; Aronica et al., 2002), land-surface-atmosphere interactions (Franks et al., 1997), soil freezing and thawing (Hanson and Lundin, 2006), crop yields and soil organic carbon (Wang et al., 2005), and ground radar-rainfall estimation (Tadesse and Anagnostou, 2005). Applications of GLUE are also found in distributed hydrologic modeling (McMichael et al., 2006; Muleta and Nicklow, 2005).

In recent years, a strong debate has emerged in the hydrologic community between those proponents that adhere strongly to the underlying philosophy of GLUE and believe that the method is a useful working methodology for assessing parameter and predictive uncertainty in non-ideal cases, and researchers and practitioners that strongly oppose incorrect usage of statistics in favor of coherent probabilistic approaches (Gupta et al., 1998; Beven and Young, 2003; Gupta et al., 2003; Christensen, 2004; Montanari, 2005; Mantovan and Todini, 2006; Stedinger et al., 2008; Beven et al., 2008; Beven, 2008; Vrugt et al., 2008b,c). In this paper we draw inspiration from recent developments in population and evolutionary genetics (Pritchard et al., 1999; Beaumont et al., 2002), and introduce "likelihood-free" inference to hydrologic modeling and uncertainty quantification. This approach was introduced in the statistical literature about three decades ago (Diggle and Gratton, 1984) for cases when an explicit likelihood (objective) function cannot be justified. This class of methods is also referred to as Approximate Bayesian Computation (ABC) (Marjoram et al., 2003; Sisson et al., 2007; Del Moral et al., 2008; Joyce and Marjoram, 2008; Grelaud et et al., 2009; Ratmann et al., 2009) and has many elements in common with the limits of acceptability approach of GLUE.

HESSD
10, 4739–4797, 2013

This paper is a follow up of our earlier work presented in Vrugt et al. (2008c) and demonstrates that the limits of acceptability approach of GLUE is a special case of ABC if each individual calibration observation is used as a summary metric. This work follows a different approach and line of reasoning than Nott et al. (2012) who demonstrated that GLUE corresponds to a certain approximate Bayesian procedure even when the "generalized likelihood" is not a true likelihood.

The remainder of this paper is organized as follows. Section 2 briefly summarizes the Bayesian approach to model parameter and predictive uncertainty estimation, and reiterates the problem which formulation of the likelihood function to use to summarize nontraditional error residual distributions. In Sect. 3 we subsequently introduce likelihood-free inference to hydrologic modeling and analysis, and demonstrate the main elements of the ABC procedure by application to a simple Nash–Cascade series of three linear reservoirs. This is followed by Sect. 4 in which the conceptual and statistical equivalence of ABC and the limits of acceptability approach of GLUE is demonstrated. Section 5 then proceeds with a comparison between GLUE and ABC using the Sacramento soil moisture accounting (SAC-SMA) model (Burnash et al., 1973) and hmodel (Schoups and Vrugt, 2010a) and discharge data form two contrasting watersheds in the US. Finally, in Sect. 6 we summarize our main results and discuss the main findings.

2 Bayesian inference

The classical Bayesian approach to model calibration considers the model parameters to be the only source of uncertainty and estimates their posterior probability density function (pdf) by maximizing $p(\boldsymbol{\theta}|\tilde{\mathbf{Y}})$ using Bayes theorem

$$p(\boldsymbol{\theta}|\tilde{\mathbf{Y}}) = \frac{p(\boldsymbol{\theta})p(\tilde{\mathbf{Y}}|\boldsymbol{\theta})}{p(\tilde{\mathbf{Y}})}, \quad (3)$$

HESSD
10, 4739–4797, 2013

where $p(\boldsymbol{\theta})$ signifies the prior parameter distribution, and $\mathcal{L}(\boldsymbol{\theta}|\tilde{\mathbf{Y}}) \equiv p(\tilde{\mathbf{Y}}|\boldsymbol{\theta})$ denotes the likelihood function. The normalization constant or evidence, $p(\tilde{\mathbf{Y}})$ is not required for the parameters as all our statistical inferences (mean, standard deviation, etc.) about them can be made from the unnormalized density. Explicit knowledge of $p(\tilde{\mathbf{Y}})$ is desired for Bayesian model selection and averaging.

Under ideal conditions with an adequate model and perfect forcing data, the error residuals follow a zero-mean Gaussian distribution

$$\mathcal{L}(\boldsymbol{\theta}|\tilde{\mathbf{Y}}, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{x}}_0) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\hat{\sigma}^2_{\tilde{\mathbf{Y}}}}} \exp\left[-\frac{1}{2}\hat{\sigma}^{-2}_{\tilde{\mathbf{Y}}}(\tilde{y}_t - y_t(\boldsymbol{\theta}, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{x}}_0))^2\right], \quad (4)$$

and $\boldsymbol{\theta}$ should converge to $\boldsymbol{\theta}^*$ where $\hat{\sigma}_{\tilde{\mathbf{Y}}}$ is an estimate of the standard deviation of the measurement error. The value of $\hat{\sigma}_{\tilde{\mathbf{Y}}}$ can be specified a-priori based on knowledge of the measurements errors, or alternatively its value can be inferred simultaneously with the values of $\boldsymbol{\theta}$ (Vrugt et al., 2008b; Bikowski et al., 2012; Laloy and Vrugt, 2012). It is worth noting that the data often come from only a single experiment. So while it is possible to quantify numerical errors, such as those due to discretization (see Kaipio et al., 2004; Nissinen et al., 2009), there is no opportunity to control the boundary conditions of (large-scale) natural systems to obtain data from additional experiments in which some controllable inputs have been varied.

The likelihood function, $\mathcal{L}(\cdot)$ in Eq. (4) is useful for simple regression problems, but the assumption of independent identically distributed Gaussian error residuals cannot be justified in environmental modeling. The presence of input data and model structural errors introduces complex error residual distributions whose probabilistic properties are difficult to describe accurately with classical likelihood functions. The choice of an adequate likelihood function, $L(\boldsymbol{\theta}|\tilde{\mathbf{Y}})$ has therefore been the subject of considerable debate in the hydrologic and statistical literature. In response to this, Schoups and Vrugt (2010a) has introduced a generalized likelihood function that better extends the applicability of commonly likelihood functions to situations where residual errors

HESSD
10, 4739–4797, 2013

are correlated, heteroscedastic, and non-Gaussian with varying degrees of kurtosis and skewness. Application to daily rainfall-runoff data from a dry and humid basin showed that (1) residual errors are much better described by a heteroscedastic, first-order, auto-correlated error model with a Laplacian distribution function characterized by heavier tails than a Gaussian distribution; and (2) compared to a standard least-squares approach, proper representation of the statistical distribution of residual errors yields tighter predictive uncertainty bands and different parameter uncertainty estimates that are less sensitive to the particular time period used for inference, (3) multiplicative bias factors improve the prediction of peak flow, and (4) near zero-flows are better described with a skewed error distribution.

The generalized likelihood function improves the statistical description of the error residuals, yet it does not separate out the effect of individual error sources. Another from the viewpoint of this paper less important deficiency is that the use of a single performance metric, $\mathcal{L}$ no matter how carefully chosen, is inadequate to extract all information from the available calibration data. The use of such "insufficient statistic" promotes equifinality, and makes it unnecessarily difficult to find the preferred parameter values. This is not desirable and explains why calibration of highly-parameterized models is often found to be time consuming and difficult.

3 Approximate Bayesian Computation

Whereas traditional Bayesian approaches require us to specify an explicit likelihood function, $\mathcal{L}(\boldsymbol{\theta}|\tilde{\mathbf{Y}})$, ABC approaches avoid explicit evaluation of the likelihood function in favor of (a set of) summary variables that better extract the information from the available data. The premise behind ABC is that $\boldsymbol{\theta}'$ should be a sample from the posterior distribution as long as the distance between the observed and simulated data, hereafter referred to as $\rho\left(\tilde{\mathbf{Y}}, \mathbf{Y}(\boldsymbol{\theta}')\right)$ is less than some small value, ϵ. For sufficiently complex models and large data sets the probability of happening upon a simulation run

HESSD
10, 4739–4797, 2013

that yields precisely the same data set as the one observed will be very small, often unacceptably so. So rather than considering the data, $\tilde{\mathbf{Y}}$ itself we consider a summary statistic of the data, $S(\tilde{\mathbf{Y}})$, and use a distance function (Marjoram et al., 2003; Sisson et al., 2007)

$$\rho\left(S(\tilde{\mathbf{Y}}), S(\mathbf{Y}(\boldsymbol{\theta}'))\right) \leq \epsilon, \tag{5}$$

to decide whether to accept the parameter values, $\boldsymbol{\theta}_i$ or not. A pseudo-code of the generic ABC approach is given below.

Algorithm 1 Rejection sampler (ABC-REJ)

for $i = 1, \ldots, N$ **do**
  repeat
    generate $\boldsymbol{\theta}'$ from the prior distribution, $p(\boldsymbol{\theta})$
    simulate $\mathbf{Y}$ from the model, $\mathbf{Y} \leftarrow \mathcal{H}(\boldsymbol{\theta}'|\cdot)$
  until $\rho\left(S(\tilde{\mathbf{Y}}), S(\mathbf{Y}(\boldsymbol{\theta}))\right) \leq \epsilon$
  set $\boldsymbol{\theta}_i = \boldsymbol{\theta}'$
end for

In words, the ABC algorithm proceeds as follows. First we sample a candidate point, $\boldsymbol{\theta}'$ from some prior distribution, $p(\boldsymbol{\theta})$. We then use this proposal to simulate the output of the model, $\mathbf{Y} \leftarrow \mathcal{H}(\boldsymbol{\theta}'|\cdot)$ and use this n-vector to calculate one or multiple summary metrics. A distance function, $\rho\left(S(\tilde{\mathbf{Y}}), S(\mathbf{Y}(\boldsymbol{\theta}'))\right)$ is then used to decide whether to accept $\boldsymbol{\theta}'$ or not. If this distance function is smaller than some pre-defined tolerance value, ϵ then the simulation is close enough to the observations that the candidate point, $\boldsymbol{\theta}'$ has some nonzero probability of being in the approximate posterior distribution, $\hat{p}\left(\boldsymbol{\theta}|\rho\left(S(\tilde{\mathbf{Y}}), S(\mathbf{Y})\right) \leq \epsilon\right)$. This algorithm converges to the true posterior $p(\boldsymbol{\theta}|\tilde{\mathbf{Y}})$ when $\epsilon \to 0$, provided that the summary statistic(s), $S(\cdot)$ is (are) near sufficient (Pritchard et al., 1999; Beaumont et al., 2002; Ratmann et al., 2009; Turner and van Zandt, 2012).

HESSD
10, 4739–4797, 2013

To illustrate the ABC methodology, we consider a Nash–Cascade instantaneous unit hydrograph. This model routes inflow (rainfall) through a series of linear reservoirs that all have the same recession coefficient. Mathematically, this cascade of m linear reservoirs with recession coefficient r can be written as follows (Nash, 1960)

$$\mathfrak{h}_t(r,m) = \frac{1}{r\Gamma(m)}\left(\frac{t}{r}\right)^{(m-1)} \exp\left(-\frac{t}{r}\right), \tag{6}$$

where t (days) denotes time, $\Gamma(\cdot)$ signifies the gamma function, and $\mathfrak{h}_t(\cdot)$ is the modeled response at time t. A 365 day period with synthetic daily streamflow data (in $m^3 s^{-1}$) was generated by driving the Nash–Cascade model of Eq. (6) with an artificial precipitation record. We assume $m = 3$ reservoirs, and a recession constant of $r = 2$ days. This artificial data set is subsequently perturbed with a heteroscedastic error (non-constant variance) with standard deviation equal to 20 % of the original simulated discharge values. Figure 2 plots the original simulated discharge time series (blue line) and the corrupted observations (red circles) used in the ABC analysis to derive the posterior distribution of the recession constant.

We are now left with a selection of the summary statistic, $S(\cdot)$ to decide whether a candidate point (model simulation) is behavioral or not. For illustrative purposes we start with the mean of the actual data,

$$\rho\left(S(\tilde{\mathbf{Y}}), S(\mathbf{Y}(\boldsymbol{\theta}))\right) = \left|\text{mean}(\tilde{\mathbf{Y}}) - \text{mean}(\mathbf{Y}(\boldsymbol{\theta}))\right| \tag{7}$$

to estimate the posterior distribution of the recession constraint. This metric is rather weak and cannot be considered "sufficient". Yet, it serves to show how the ABC methodology works in practice.

A uniform prior with $r \in [0,4]$ was used in all our calculations. To increase computational efficiency, we used an improved variant of the ABC population Monte Carlo (PMC) scheme of Turner and van Zandt (2012), the details of which appear in Appendix A. In short, the PMC sampler starts out as ABC-REJ during the first iteration,

HESSD
10, 4739–4797, 2013

$j = 1$, but using a much larger initial value for ϵ. During each successive next step, $j = \{2, \ldots, J\}$ the value of ϵ is decreased and the proposal distribution, $q_j(\boldsymbol{\theta}_k^{j-1}, \cdot) = N_d(\boldsymbol{\theta}_k^{j-1}, \sum^j)_{(j>1)}$ adapted using $\sum^j = \mathrm{Cov}(\boldsymbol{\theta}_1^{j-1}, \ldots, \boldsymbol{\theta}_N^{j-1})$ with $\boldsymbol{\theta}_k$ drawn from a discrete multinomial distribution, $\mathfrak{F}(\boldsymbol{\theta}_{1:N}^{j-1} | \boldsymbol{w}_{1:N}^{j-1})$ where $\boldsymbol{w}_{1:N}^{j-1}$ denote the posterior weights $(w_l^{j-1} \geq 0; \sum_{l=1}^N w_l^{j-1} = 1)$. Through a sequence of successive (multi)normal proposal distributions the prior sample is thus iteratively refined until a sample of the posterior distribution is obtained. This approach, similar in spirit as the adaptive Metropolis sampler of (Haario et al., 1999, 2001) receives a much higher sampling efficiency than ABC-REJ, particularly for cases where the prior sampling distribution, $p(\boldsymbol{\theta})$ is a poor approximation of the actual posterior distribution.

The PMC sampler of Turner and van Zandt (2012) assumes that the sequence of ϵ values is specified by the user. This does not necessarily lead to the most efficient search. Our sampler therefore adaptively determines the next value of $\epsilon_j; j > 1$ from the cumulative distribution function of the $\rho(\cdot)$ values of the N most recent accepted samples. Details of this procedure are given in Appendix B. For the present case study, an initial value of $\epsilon = 1$ is used, and this value is adaptively decreased until a value of $\epsilon = 0.05$ is reached. Lower values of ϵ provide similar posterior estimates, yet unnecessarily increase the computational burden of the ABC analysis (Vrugt and Sadegh, 2013).

Figure 3a presents a histogram of the posterior marginal distribution of r derived from the ABC-PMC analysis using the mean observed flow as summary statistic. The red square denotes the true parameter value used to create the synthetic data. For completeness we also present in the middle panel the results of DREAM (Vrugt et al., 2008a, 2009a) using the likelihood function of Eq. (4) but with a heteroscedastic measurement error, $\hat{\boldsymbol{\sigma}}_{\tilde{\mathbf{Y}}} = 0.2 \times [\tilde{y}_1, \ldots, \tilde{y}_n]$.

Perhaps not surprisingly, the ABC-derived posterior distribution is poorly defined by calibration against the mean observed discharge value. The behavioral recession constants extend a larger portion of its prior distribution. This suggests that the observed

HESSD
10, 4739–4797, 2013

(synthetic) discharge data do not contain information about the recession constant of the three reservoirs. This finding is perhaps not surprising. Many different values of the recession constant, r can be found with mean simulated discharge value similar to that of the observed data, but with poor accuracy of the simulated streamflow dynamics. Indeed, if a classical likelihood function is used (Fig. 3b) the recession constant is much better defined with maximum a-posteriori density equal to $r = 2$, and 95 % posterior parameter uncertainty ranges that vary between 1.95 and 2.05.

Fortunately, nothing prevents us from using more than one summary statistic in the ABC-analysis to measure different and complementary parts of model behavior. To be meaningful in practice, such statistics should preferably measure hydrologically relevant signatures of watershed behavior. Such approach was introduced in our previous work (Vrugt and Sadegh, 2013), and for simplicity we now augment the first metric (mean of the data) with another simple statistic (standard deviation of data)

$$\rho\left(S(\tilde{\mathbf{Y}}), S(\mathbf{Y}(\boldsymbol{\theta}))\right) = \max\left(\left|\text{mean}(\tilde{\mathbf{Y}}) - \text{mean}(\mathbf{Y}(\boldsymbol{\theta}))\right|, \left|\text{std}(\tilde{\mathbf{Y}}) - \text{std}(\mathbf{Y}(\boldsymbol{\theta}))\right|\right), \quad (8)$$

to decide whether the model simulation can be considered behavioral or not. The results of this analysis are shown in Fig. 3c using a minimum value of $\epsilon = 0.05$. The recession constant appears much better defined, but the width of the (marginal) posterior distribution is still considerably larger than what can be expected from a classical likelihood function using MCMC simulation with DREAM (Fig. 3b). This simply conveys that our two summary metrics are jointly insufficient, and that, if so desired, more powerful metrics should be used.

The ABC methodology allows the use of a wide arsenal of summary metrics and distance functions to judge the distance between the model simulation and observations. Common examples in genetics include the Canberra, Euclidean, and Manhattan distance. Those are readily applied in hydrology as well, including summary statistics such as the Nash–Sutcliffe Efficiency (NSE) (Nash and Sutcliffe, 1970), Mean Square Error (MSE), and others listed in Table 1 of Gupta et al. (1998). Temporal disaggregation of the data and model simulations would preserve the statistical moments of $\tilde{\mathbf{Y}}$

HESSD
10, 4739–4797, 2013

such as the mean, median, standard deviation, kurtosis, and skewness. The use of flow duration curves could be beneficial in this regard as characteristic of the watersheds response to rainfall (Vrugt and Sadegh, 2013).

4 Statistical equivalence of ABC and GLUE

Now the basic principles of ABC have been discussed in some detail using the simple one-parameter unit hydrograph toy problem, it is not difficult to see the many similarities of GLUE and ABC

1. The distance function specified in Eq. (5) has many elements in common with the triangular, trapezoidal or beta fuzzy-membership functions used in the limits of acceptability approach of GLUE. This is perhaps more obvious if we use the following notation

$$\rho\left(S(\tilde{\mathbf{Y}}), S(\mathbf{Y}(\boldsymbol{\theta}))\right) = \sum_{t=1}^{n} I\left(|\tilde{y}_t - y_t(\boldsymbol{\theta})| \leq \delta_t\right), \tag{9}$$

where $I(A)$ is a simple indicator function that is "1" if A is true, and "0" otherwise, and $\delta_t; t = \{1, \ldots, n\}$ constitutes the effective observation error that takes into account multiple sources of error (Beven, 2006). This value is defined a-priori by the user. The ABC approach can thus be made mathematically equivalent to the limits of acceptability approach of GLUE if each observation is used as summary statistic.

2. The ABC-REJ sampler is similar to the Latin Hypercube sampling strategy used in GLUE to find behavioral solutions. Both methods use a fixed proposal distribution to sample from the prior parameter distribution. If the corresponding simulation falls within the bounds specified by the effective observation error, then the parameter combination will be classified as behavioral, otherwise, the proposal will be rejected. Sampling continues until N behavioral solutions are found.

HESSD
10, 4739–4797, 2013

Following the first proposition, a solution is deemed behavioral with ABC if its simulated discharge time series falls within the interval, $[\tilde{y}_t - \delta_t, \tilde{y}_t + \delta_t]$ for $t = \{1, \ldots, n\}$. This is similar to the limits of acceptability approach of GLUE if a simple discrete (0/1) membership function is used. For the synthetic toy example used herein, we define the effective observation error to be $\delta_t = \alpha \times \hat{\boldsymbol{\sigma}}_{\tilde{y}_t}$ with $\alpha = 2$. This is equivalent to $\delta_t = 0.4\tilde{y}_t$.

The goal of the ABC analysis now becomes finding all those parameter combinations that consistently fall within the effective observation error of the discharge data, and hence receive a perfect score of Eq. (9) equal to n. This constitutes a maximization problem, and differs from a typical implementation of ABC where the distance to the summary statistics, and value of ϵ is being minimized. In our numerical implementation with the PMC sampler we therefore adapt Eq. (9) and calculate

$$\rho\left(S(\tilde{\mathbf{Y}}), S(\mathbf{Y}(\boldsymbol{\theta}))\right) = \frac{1}{n}\left(n - \sum_{t=1}^{n} I\left(|\tilde{y}_t - y_t(\boldsymbol{\theta})| \leq \delta_t\right)\right), \tag{10}$$

to decide whether a simulation is behavioral or not. For a perfect simulation, $\rho(\cdot)$ will be zero. But in most practical applications it is not possible to find a simulation that satisfies $\epsilon = 0$. For instance, for the present Nash-Cascade toy example with $\alpha = 2$ and thus $\delta_t = 0.4 \times \tilde{y}_t; t = \{1, \ldots, n\}$, a minimum value of $\rho(\cdot)$ of about 0.05 is to be expected. This follows directly from statistical theory (about 95 % of the observations are included in the interval of 2 times the standard deviation).

The adaptive updating strategy of ϵ in PMC not only guarantees a more efficient search strategy than ABC-REJ (GLUE), but also automatically determines the maximum attainable coverage of the discharge observations within the limits of acceptability. In the first iteration, we set $\epsilon = 0.75$ and thus define a behavioral solution as one that contains at least 25 % of the observed discharge data within the interval, $[y_t - \delta_t, y_t + \delta_t]_{(t=\{1,\ldots,n\})}$. During each successive next iteration the value of ϵ is sequentially reduced, and the PMC sampler terminates when the difference between two subsequent ϵ values is less than 0.02, or in mathematical notation $\epsilon_j - \epsilon_{j-1} < 0.02$. In

HESSD
10, 4739–4797, 2013

all our simulations presented herein we request PMC to create $N = 1000$ behavioral solutions at each different ϵ value (iteration). We report our results for $\epsilon \leq 0.10$.

Figure 4a presents the results of the ABC-PMC analysis and plots the 95 % streamflow simulation uncertainty ranges (dark grey region) using the ABC-PMC sampler. This result is derived from the $N = 1000$ posterior solutions using the 2.5 and 97.5 percentile of the simulated discharge values. The artificial discharge observations are indicated with red circles. The simulations nicely track the observed data with uncertainty intervals that appear relatively narrow and encompass about 90 % of the data. The upper panel plots the results for GLUE using the same limits of acceptability. Latin hypercube sampling was used to create 1000 behavioral solutions at $\epsilon = 0.10$ using an algorithm virtually identical to that of ABC-REJ. Perhaps not surprising, the results are identical to those presented for ABC. Although the numerical results are identical, the computational efficiency of both methods is not. The ABC-PMC sampler exhibits an acceptance rate of about 53.5 % whereas for GLUE (and hence ABC-REJ) this is about 17.0 %.

5 Case studies: hydrologic modeling

Now the ABC method has been discussed in some detail and the theoretical connection of this approach with GLUE has been demonstrated, we proceed with numerical simulation using five years of daily streamflow data from the French Broad river basin at Ashvile, North Carolina (1 January 1962 to 30 December 1966) and the Leaf River (1 October 1952 to 30 September 1957) north of Collins, Mississippi. These watersheds have been studied extensively in the literature and details of the data can be found in related publications. Two lumped conceptual hydrologic models are used to describe the rainfall-runoff transformation. This includes the 7-parameter hmodel described in detail in Schoups and Vrugt (2010a) and the 13-parameter SAC-SMA model (Burnash et al., 1973). Inputs to these models include mean areal precipitation (MAP) and potential evapotranspiration (PET) while the outputs are estimated evapotranspiration and

HESSD
10, 4739–4797, 2013

channel inflow. Numerical, conceptual, and computational details of both models can be found in the cited publications, and so will not be repeated herein. Tables 1 and 2 summarize the parameters of both models and their upper and lower bound values.

Implementation of the limits of acceptability approach requires knowledge of the effective observation error. This error varies dynamically with flow level and constitutes the combined effect of input data, model structural and calibration data measurement error. In practice, the user defines the limits of acceptability for each individual observation, but for convenience we follow a different approach and set the effective observation error as a multiple of the actual discharge measurement error. We follow Vrugt et al. (2005) and use consecutive differences of the calibration data to calculate the measurement data error

$$\hat{\sigma}_{\tilde{\mathbf{Y}}} = \sqrt{\binom{2/}{/}^{-1} (\Delta^{/} \tilde{y}_t)^2}, \tag{11}$$

where $\Delta^{/}$ denotes the difference operator applied / subsequent times (Rice, 1984; Hall et al., 1990; Seifert et al., 1993; Dette et al., 1998). This estimator was introduced in Vrugt et al. (2005) and shown to work well for daily and hourly discharge data. Heteroscedasticity is easily identified by applying the nonparametric estimator locally in the calibration data time series. This provides a n-vector of measurement errors, hereafter referred to as $\hat{\boldsymbol{\sigma}}_{\tilde{\mathbf{Y}}} = \{\hat{\sigma}_{\tilde{y}_1}, \ldots, \hat{\sigma}_{\tilde{y}_n}\}$. The limits of acceptability in Eq. (9) are now defined to be, $\delta_t = \alpha \times \hat{\sigma}_{\tilde{y}_t}$ for $t = \{1, \ldots, n\}$ using $\alpha = 2$. We now summarize the results of GLUE and ABC for both models and watersheds.

Figure 5 plots histograms of the behavioral solutions of an illustrative set of five SAC-SMA model parameters for the French Broad watershed. The PMC sampler terminated its search with $\epsilon \leq 0.06$ corresponding to a coverage of 94 % of the discharge data within the effective observation error. The top panel presents the results for GLUE (limits of acceptability) and the bottom panel shows the corresponding counterparts for ABC. To limit the computational burden, GLUE was terminated after 300 behavioral solutions were found. This is sufficient for comparative purposes. The marginal

HESSD
10, 4739–4797, 2013

distribution of the lower zone primary free water depletion rate (LZPK) follows a normal distribution, whereas the histograms of the other parameters deviate considerably from normality and tend to assign the highest probability mass at the lower (PCTIM, ADIMP and LZFSM) or upper bound (LZFPM). The posterior parameter uncertainty ranges appear rather large, and essentially cover the entire prior distribution defined previously in Table 2. This uncertainty is perhaps unrealistically large and much larger than what can be expected from an explicit likelihood function, but not surprising given the size of the effective observation error used to define the limits of acceptability. What is most important however is the finding that the GLUE and ABC derived posterior parameter distributions are essentially similar. This provides further support for our claim that the limits of acceptability approach of GLUE can be interpreted as a special case of formal Bayes. We will further elaborate on this equivalence in the Sect. 6 of this paper.

Now the posterior parameter uncertainty has been defined, we focus our attention on the actual discharge simulations. Figure 6 presents the outcome of this analysis, and presents the 95 % streamflow uncertainty ranges (gray region) of the GLUE (top panel) and ABC (bottom panel) derived posterior parameter distribution. The simulation uncertainty ranges appear rather large, but nicely cover approximately 74 % of the discharge observations. The simulation results of GLUE and ABC are in strong agreement, which is to be expected given the strong similarity of the behavioral samples derived with both methods.

Although the numerical results of GLUE and ABC are very similar, the PMC sampler requires only 1/30 (1/8) of the simulations of GLUE to locate $N = 1000$ posterior solutions for the SAC-SMA (hmodel) (see Table 3). The advantage of PMC is more and more apparent with increasing dimensionality of the parameter space. If the search space is relatively low-dimensional (hmodel) and the space of behavioral solutions relatively large in comparison to the prior parameter space, both sampling methods will rapidly sample $N = 1000$ behavioral solutions. If, on the contrary, the search space is of higher dimensions (SAC-SMA), or the behavioral solution space is made up of a small portion of the prior parameter space, Latin hypercube sampling (and ABC-REJ) will

HESSD
10, 4739–4797, 2013

be rather inefficient needing many thousands of draws from the prior distribution to find just a handful of good (behavioral) solutions. The PMC sampler achieves a higher sampling efficiency by iteratively reducing the value of ϵ during the search, and adaptively updating the scale and orientation of the proposal distribution. Note that the PMC and Latin Hypercube sampling strategies used herein vary all parameters at a time, and hence further efficiency improvements are to be expected in high-dimensional parameter spaces with the user of genetic operators such as crossover and mutation.

To provide more insights into the sampled parameter space of the French Broad river basin, please consider Fig. 7 that presents two-dimensional scatter plots of the posterior samples derived with GLUE (top panel) and ABC-PMC (bottom panel) for three selected parameter pairs. The bivariate sample plots appear very similar and confirm our previous findings in Fig. 5 and demonstrate significant scatter with behavioral samples that extend their entire uniform prior ranges. But this does not necessarily mean that the posterior parameter space is badly defined. Instead, large portions of the (A) LZSK-LZPK, (B) PFREE-ADIMP, and (C) ZPERC-LZPK subspaces are virtually empty and thus deemed non-behavioral. This suggests at least some level of correlation between the posterior parameter samples. The difference in sampling density between both panels is simply due to an insufficient computational budget for GLUE to create $N = 1000$ behavioral solutions. GLUE was terminated after 300 posterior samples were found.

We now proceed with out-of-sample prediction, and plot in Fig. 8 the streamflow uncertainty ranges (gray region) of the SAC-SMA model for a three year portion of the evaluation data set of the French Broad watershed. This period commences immediately after the last day of the calibration data set, and the initial states at $t = 0$ have been derived from the calibration ensemble. The top panel presents the results of GLUE whereas the bottom panel plots the corresponding results of ABC. Perhaps not surprisingly, both methods exhibit similar results and provide a discharge ensemble that envelops about 70 % of the observed discharge data (red circles). The strong similarity between the simulations results of the calibration and evaluation sample inspires

HESSD
10, 4739–4797, 2013

confidence in the ability of the behavioral parameter set to accurately describe the rainfall-runoff transformation of the French Broad river basin.

Table 4 summarizes the results of GLUE and ABC for the SAC-SMA model and French Broad watershed, and presents the root mean squared error (RMSE) of the posterior mean discharge simulation and associated coverage of the 95 % prediction intervals for the calibration and evaluation period. For completeness, we also list the results of the SAC-SMA model with a formal likelihood function, Eq. (4) using the heteroscedastic measurement error, $\hat{\sigma}_{\tilde{\mathbf{Y}}}$ derived from the nonparametric difference operator. The listed statistics summarize our main findings thus far. The results of GLUE and ABC are virtually identical and show a consistent performance during the calibration and evaluation period. The 95 % uncertainty ranges derived with both methods encompass about 70 % of the discharge observations. This coverage is significantly larger than the approximately 12–17 % derived from a classical likelihood function. This finding has important practical utility, for instance within the context of flood forecasting. The behavioral parameter distribution derived with ABC and GLUE provides a reasonable initial estimate of the total out-of-sample prediction uncertainty. On the contrary, the posterior uncertainty derived from a classical likelihood function exhibits an unrealistic small coverage and thus needs to be artificially inflated with a random prediction error to create a statistically meaningful streamflow uncertainty interval.

Our main focus thus far has been on the SAC-SMA model, without recourse to the simulation results of the hmodel. Figure 9 shows posterior histograms of five of the hmodel parameters derived with GLUE (top panel) and ABC (bottom panel) using the French Broad calibration data set. The PMC sampler determined a maximum possible coverage of 95 % of the discharge data within the uniform hypercube defined by the effective observation error. The results in Figs. 9–11 thus pertain to this coverage level. The marginal posterior parameter distributions derived with GLUE and ABC again demonstrate a strong agreement. Most of the hmodel parameters, with the exception of I_{max} and α_E are reasonably well defined by calibration against the observed discharge data. The parameter Q_{max} is particularly well resolved and favors values close to zero

HESSD
10, 4739–4797, 2013

– something that is physically rather unrealistic and likely due to errors in the model formulation and precipitation data.

Figure 10 presents two-dimensional scatter plots of the posterior samples of three selected parameter pairs. The top panel corresponds to GLUE and the bottom panel illustrates the same results for ABC. Each plus symbol depicts a behavioral solution. Because of sampling inefficiency the GLUE calculations were terminated after 100 behavioral samples were identified. This explains the apparent differences in sampling density. Nevertheless, the bivariate plots of the posterior samples derived with both methods are in strong agreement with each other with behavioral solutions that occupy only a relatively small part of the prior parameter space. This is particularly true for the $\alpha_F - Q_{max}$ subspace. The sampled parameter pairs appear rather uncorrelated which suggests that the different hmodel parameters each control a different part of the simulated watershed response. This simplifies posterior inference, and favors a hierarchical sampling approach in which parameters are estimated sequentially.

We now demonstrate how the hmodel parameter uncertainty translates into streamflow simulation uncertainty. We separately depict the results for the calibration (Fig. 11) and evaluation (Fig. 12) period. As expected, the simulation results derived with GLUE and ABC are in close agreement. The 95 % simulation uncertainty ranges encompass about 67 % of the calibration data observations (see Table 5). This is much higher than the 7 % coverage derived with a classical likelihood function using MCMC simulation with DREAM. Yet, between days 205 and 270 the posterior ensemble systematically over predicts the actual discharge observations. This positive bias is likely caused by an error in the measured rainfall data around day 205. This rainfall error accumulates in the simulated state variables and continues to persist until the next significant rainfall event around day 270. Rainfall data correction would seem appropriate Kavetski et al. (2006a,b); Vrugt et al. (2008a), but is beyond the scope of the present paper.

The evaluation data period again highlights the strong operational similarity of GLUE and ABC, but the average width of the 95 % streamflow simulation uncertainty ranges appears somewhat smaller. Indeed, the coverage has reduced to approximately 61 %

HESSD
10, 4739–4797, 2013

of the discharge observations. Note that the posterior ensemble systematically underestimates the peak flow events. This is simply the effect of an increased rainfall intensity during the evaluation period. For practical application it would seem most productive to extend the length of the calibration data period to include a number of larger storm events. This would certainly improve the fitting of the peak flow events, but not affect the main conclusions of this paper.

We now turn our attention on the Leaf River data set, and present in Fig. 13 histograms of the marginal posterior distributions of a representative group of five SAC-SMA parameters. The PMC sampler determined a maximum coverage of about 56 % of the discharge data within the effective observation error used herein. The top panel displays the results for GLUE for this coverage level, whereas the bottom panel shows the corresponding counterparts derived with ABC. The histograms derived with both method again are strikingly similar, yet the PMC sampler used in ABC is about 30 times more efficient (not tabulated) in sampling the $N = 1000$ behavioral solutions. Note that SAC-SMA parameters are not particularly well defined by calibration against the limits of acceptability. The marginal posterior uncertainty ranges are rather large, with the exception of the parameter LZPK that tends to favor values close to zero.

Finally, Fig. 14 illustrates the performance of the GLUE and ABC derived posterior parameter distributions for an independent evaluation period. The GLUE (top panel) and ABC (bottom panel) derived 95 % posterior streamflow uncertainty ranges again appear nearly equivalent (expected) and contain about 65 % of the observed discharge values. This is somewhat smaller than the 80 % coverage obtained during the calibration data period (not shown herein).

6 Discussion and conclusions

In the past two decades, the GLUE methodology of Beven and Binley (1992); Beven and Freer (2001); Beven (2006) has found widespread application and use for model parameter and predictive uncertainty analysis. This method rejects the formal Bayesian

HESSD
10, 4739–4797, 2013

paradigm in favor of finding a set of of behavioral solutions that are acceptably close to the non-error-free observations. This avoids over conditioning of the posterior parameter space in non-ideal cases with nontraditional error residual distributions. Indeed, Tables 4–6 demonstrate that the GLUE derived 95 % simulation uncertainty ranges encompass a much higher percentage of the discharge observations than the posterior parameter predictive uncertainty intervals derived from a classical likelihood function. Formal likelihood functions that do not adequately describe the probabilistic properties of the error residuals tend to overestimate the actual information content of the data, and provide estimates of the posterior parameter uncertainty that are overly optimistic.

Many have criticized the GLUE methodology for being subjective and lacking an appropriate mathematical underpinning. To help bridge the cap between informal and formal Bayesian approaches, this paper introduced likelihood-free inference to hydrologic modeling and uncertainty analysis. This approach was introduced in the statistical literature about three decades ago (Diggle and Gratton, 1984) for cases when an explicit likelihood (objective) function cannot be justified. Such approaches, also referred to as ABC, use one or multiple (sufficient) statistics to estimate the posterior parameter distribution. The premise behind ABC is that $\boldsymbol{\theta}'$ should be a sample from the posterior distribution as long as the distance between the observed and simulated summary statistics is smaller than some small value, ϵ. An example of this was given in Sect. 3 by calibration of the Nash–Cascade model against the mean and standard deviation of the discharge data. In the limit of ϵ going to zero, the behavioral solution space should converge to the actual posterior distribution, pending the assumption that the chosen summary statistic(s) is (are) near sufficient (Pritchard et al., 1999; Vrugt and Sadegh, 2013). But this was certainly not the case for the Nash–Cascade example. The mean and standard deviation are rather weak summary metrics and this explains why the marginal posterior distribution of the recession constant was too wide (see Fig. 3c) and did not converge to its expected distribution (Fig. 3b). Thus, there is a clear need for meaningful summary statistics with a compelling diagnostic power. Examples include the annual runoff and baseflow coefficient, and the flow duration curve as used

HESSD
10, 4739–4797, 2013

in Vrugt and Sadegh (2013). Note that the ABC approach differs from multiple objective calibration frameworks in that each summary metric is simultaneously minimized.

Numerical simulations presented in Figs. 4–14 have shown that, if each observation is treated as a summary variable, then the ABC approach obtains very similar results as the limits of acceptability approach of GLUE. A similar conclusion was drawn in previous work by Nott et al. (2012) but following a different line of reasoning and within the context of the more traditional GLUE methodology presented by Beven and Binley (1992). One issue deserves special attention, and that is that within the limit of acceptability framework, the value of ϵ needs to be taken much larger than what is deemed statistically adequate. Standard applications of likelihood free inference define a solution to be behavioral if the chosen summary statistics are within a small distance of their observed counterparts. Yet for hydrologic systems with many calibration observations the probability of happening upon a simulation run that yields exactly the same data set as the one observed will be extremely small. The effective observation error remedies this problem, but the magnitude of this value is typically much larger than the theoretical value of ϵ to guarantee converge to the true posterior parameter distribution. Thus, although the limits of acceptability approach of GLUE is a special variant of the more generic ABC approach, care should be exercised with interpretation of the posterior distribution.

Our current research efforts are directed towards improving the sampling efficiency of ABC using MCMC simulation with DREAM. Papers on this topic are expected in due course.

Appendix A

Suppose some measurement data $\tilde{\mathbf{Y}} = \{\tilde{y}_1, \ldots, \tilde{y}_n\}$, and a model that predicts $\mathbf{Y} \leftarrow \mathcal{H}(\theta|\cdot)$ with parameter values, $\boldsymbol{\theta} \in \boldsymbol{\Theta} \in \mathbb{R}^d$. We define a prior distribution, $p(\boldsymbol{\theta})$ and a vector with decreasing tolerance values, $\boldsymbol{\varepsilon} = \{\epsilon_1, \ldots, \epsilon_J\}$ so that

HESSD
10, 4739–4797, 2013

$\epsilon_{j+1} < \epsilon_j, \forall j \in \{2,\ldots,J\}$. The ABC population Monte Carlo method proceeds as follows (Turner and van Zandt, 2012)

Algorithm 2 ABC-PMC

At iteration $j = 1$,
for $i = 1,\ldots,N$ **do**
  while $\rho(S(\tilde{\mathbf{Y}}),S(\mathbf{Y})) > \epsilon_1$ **do**
    Sample $\boldsymbol{\theta}^{'}$ from the prior, $\boldsymbol{\theta}^{'} \sim p(\boldsymbol{\theta})$
    Simulate data $\mathbf{Y}$ using $\boldsymbol{\theta}^{'}$, $\mathbf{Y} \leftarrow \mathcal{H}(\boldsymbol{\theta}^{'}|\cdot)$
    Calculate $\rho(S(\tilde{\mathbf{Y}}),S(\mathbf{Y}(\boldsymbol{\theta}^{'})))$
  end while
  Set $\theta_i^1 \leftarrow \boldsymbol{\theta}^{'}$
  Set $w_i^1 \leftarrow \frac{1}{N}$
end for
Set $\sigma_1^2 \leftarrow 2\mathrm{Cov}(\boldsymbol{\theta}_{1:N}^1)$,
At iteration $j > 1$,
for $j = 2,\ldots,J$ **do**
  for $i = 1,\ldots,N$ **do**
    while $\rho\big(S(\tilde{\mathbf{Y}}),S(\mathbf{Y})\big) > \epsilon_j$ **do**
      Sample $\boldsymbol{\theta}^{'}$ from the previous iteration, $\boldsymbol{\theta}^{'} \sim \theta_{1:N,j-1}$ with probability $\mathbf{w}_{1:N}^{j-1}$
      Perturb $\boldsymbol{\theta}^{'}$ by sampling $\boldsymbol{\theta}^{''} \sim N(\boldsymbol{\theta}^{'},\sigma_{j-1}^2)$
      Simulate data $\mathbf{Y}$ using $\boldsymbol{\theta}^{''}$, $\mathbf{Y} \leftarrow \mathcal{H}(\boldsymbol{\theta}^{''}|\cdot)$
      Calculate $\rho\big(S(\tilde{\mathbf{Y}}),S(\mathbf{Y}(\boldsymbol{\theta}^{''}))\big)$
    end while
    Set $\boldsymbol{\theta}_i^j \leftarrow \boldsymbol{\theta}^{''}$
    Set $w_i^j \leftarrow \frac{p(\boldsymbol{\theta}_i^j)}{\sum_{k=1}^N w_k^{j-1} q(\boldsymbol{\theta}_k^{j-1}|\boldsymbol{\theta}_i^j,\sigma_{j-1}^2)}$
  end for
  Set $\sigma_j^2 \leftarrow 2\mathrm{Cov}(\boldsymbol{\theta}_{1:N}^j)$
end for

This concludes the pseudo-code of the population Monte Carlo sampler.

HESSD
10, 4739–4797, 2013

Appendix B

The PMC sampler of Turner and van Zandt (2012) assumes that the sequence of ϵ values is specified by the user. Practical experience suggests that a poor selection of $\boldsymbol{\varepsilon} = \{\epsilon_1, \ldots, \epsilon_J\}$ can lead to very low acceptance rates or even premature convergence if ϵ has been taken too small. We therefore implement an arguably more advanced strategy and let the sampler decide which values of $\epsilon_j (j > 1)$ to use.

This strategy is implemented in words as follows. The user defines ϵ_1. In practice, a large value will typically suffice. At the end of the first iteration (just after σ_1^2 has been calculated), the algorithm computes the cumulative distribution function (cdf) of the N accepted $\rho\left(S(\tilde{\mathbf{Y}}), S(\mathbf{Y})\right)$ values. This function ranges between 0 and 1 and describes the probability that a random variable X (in this case $\rho(\cdot)$) will be found at a value less than or equal to x. The value of ϵ_2 is then taken to be that value of $\rho(\cdot)$ at which the cdf is equal to 0.1. The PMC sampler proceeds with the next iteration, $j = 2$ and this recipe is continued during each successive next iteration until ϵ_j reaches some lower default value defined by the user (Sect. 3), or when the successive reduction in ϵ has become smaller than 0.02; $\epsilon_j - \epsilon_{j-1} < 0.02$ (Sect. 4).

Numerical simulation has shown that this adaptive updating mechanism of ϵ significantly enhances the search efficiency of the PMC sampler. Moreover, this implementation does not require the user to specify $\varepsilon = \{\epsilon_2, \ldots, \epsilon_J\}$. This constitutes an important practical advantage.

Acknowledgements. Both authors highly appreciate the support and funding from the UC-Lab Fees Research Program Award 237285. The ABC methodology developed herein is written in MATLAB and available upon request from the second author (jasper@uci.edu).

HESSD
10, 4739–4797, 2013

References

Ajami, N. K., Duan, Q., and Sorooshian, S.: An integrated hydrologic Bayesian multimodel combination framework: confronting input, parameter, and model structural uncertainty in hydrologic prediction, Water Resour. Res., 43, W01403, doi:10.1029/2005WR004745, 2007. 4744

Aronica, G., Bates, P. D., and Horritt, M. S.: Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE, Hydrol. Proccess., 16, 2001–2016, 2002. 4746

Bates, B. C. and Campbell, E. P.: A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling, Water Resour. Res., 37, 937–947, 2001. 4742

Beaumont, M. A., Zhang, W., and Balding, D. J.: Approximate Bayesian Computation in population genetics, Genetics, 162, 2025–2035, 2002. 4746, 4750

Beven, K. J.: Prophecy, reality and uncertainty in distributed hydrological modeling, Adv. Water Resour., 16, 41–51, 1993. 4745

Beven, K.: A manifesto for the equifinality thesis, J. Hydrol., 320, 18–36, 2006. 4744, 4745, 4754, 4762

Beven, K.: Comment on "Equifinality of formal.(DREAM) and informal.(GLUE) Bayesian approaches in hydrologic modeling?" by Vrugt et al., Stoch. Env. Res. Risk A., 23, 1059–1060, doi:10.1007/s00477-008-0283-x, 2008. 4745, 4746

Beven, K. J. and Binley, A. M.: The future of distributed models: Model calibration and uncertainty prediction, Hydrol. Proccess., 6, 279–298, 1992. 4744, 4745, 4746, 4762, 4764

Beven, K. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modeling of complex environmental systems using the GLUE methodology, J. Hydrol., 249, 11–29, 2001. 4745, 4762

Beven, K. and Young, P.: Comment on "Bayesian recursive parameter estimation for hydrologic models" by Thiemann et al., Water Resour. Res., 39, 1116, doi:10.1029/2001WR001183, 2003. 4746

Beven, K., Smith, P. J., and Freer, J. E.: So why would a modeller choose to be incoherent?, J. Hydrol., 354, 15–32, 2008. 4745, 4746

Bikowski, J., Huisman, J. A., Vrugt, J. A., Vereeken, H., and van der Kruk, J.: Inversion and sensitivity analysis of ground penetrating radar data with waveguide dispersion using de-

HESSD
10, 4739–4797, 2013

terministic and Markov chain Monte Carlo methods, Near Surf. Geophys., Special issue "Physics-based integrated characterization", 10, 641–652, doi:10.3997/1873-0604.2012041, 2012. 4748
Brazier, R.E., Beven, K. J., Anthony, S. G., and Rowan, J. S.: Implications of model uncertainty for the mapping of hillslope-scale soil erosion predictions, Earth Surf. Proc. Land., 26, 1333–1352, 2001. 4746
Bulygina, N. and Gupta, H.: Correcting the mathematical structure of a hydrological model via Bayesian data assimilation, Water Resour. Res., 47, W05514, doi:10.1029/2010WR009614, 2011. 4744
Burnash, R. J., Ferral, R. L., and McGuire, R. A.: A Generalized Streamflow Simulation System: Conceptual Modeling for Digital Computers, US Department of Commerce National Weather Service and State of California Department of Water Resources, 1973. 4747, 4756
Butts, M. B., Payne, J. T., Kristensen, M., and Madsen, H.: An evaluation of the impact of model structure on hydrological modeling uncertainty for streamflow simulation, J. Hydrol., 298, 242–266, 2004. 4744
Christensen, S.: A synthetic groundwater modeling study of the accuracy of GLUE uncertainty intervals, Nord. Hydrol., 35, 45–59, 2004. 4746
Chu, W., Gao, X., and Sorooshian, S.: Improving the shuffled complex evolution scheme for optimization of complex nonlinear hydrological systems: application to the calibration of the Sacramento soil-moisture accounting model, Water Resour. Res., 46, W09530, doi:10.1029/2010WR009224, 2010. 4743
Clark, M. P. and Slater, A. G.: Probabilistic quantitative precipitation estimation in complex terrain, J. Hydrometeorol., 7, 3–22, 2006. 4744
Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T., and Hay, L. E.: Framework for Understanding Structural Errors(FUSE): a modular framework to diagnose differences between hydrological models, Water Resour. Res., 44, W00B02, doi:10.1029/2007WR006735, 2008. 4744
DeChant, C. M. and Moradkhani, H.: Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting, Water Resour. Res., 48, W04518, doi:10.1029/2011WR011011, 2012. 4744
Del Moral, P., Doucet, A., and Jasra, A.: An adaptive sequential Monte Carlo method for approximate Bayesian Computation, Technical Report, Imperial College London, 2008. 4746

HESSD
10, 4739–4797, 2013

Dette, H., Munk, A., and Wagner, T.: Estimating the variance in nonparametric regression – What is a reasonable choice?, J. Roy. Stat. Soc. B, 60, 751–764, 1998. 4757
Diggle, P. J. and Gratton, R. J.: Monte Carlo methods of inference for implicit statistical models, J. Roy. Stat. Soc. B, 46, 193–227, 1984. 4746, 4763
Duan, Q., Gupta, V. K., and Sorooshian, S.: Effective and efficient global optimization for conceptual rainfall-runoff models, Water Resour. Res., 28, 1015–1031, 1992. 4743
Fenicia, F., Savenije, H. H. G., Matgen, P., and Pfister, L.: A comparison of alternative multiobjective calibration strategies for hydrological modeling, Water Resour. Res., 43, W03434, doi:10.1029/2006WR005098, 2007. 4744
Feyen, L., Beven, K. J., De Smedt, F., and Freer, J. E.: Stochastic capture zone delineation within the generalized likelihood uncertainty estimation methodology: conditioning on head observations, Water Resour. Res., 37, 625–638, 2001. 4746
Franks, S.W., Beven, K. J., Quinn, P. F., and Wright, I. R.: On the sensitivity of soil-vegetation-atmosphere transfer(SVAT) schemes: equifinality and the problem of robust calibration, Agr. Forest Meteorol., 86, 63–75, 1997. 4746
Freer, J., Beven, K., and Ambroise, B.: Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach, Water Resour. Res., 32, 2161–2173, 1996. 4744, 4746
Georgakakos, K. P., Seo, D. J., Gupta, H., Schaake, J., and Butts, M. B.: Towards the characterization of streamflow simulation uncertainty through multimodel ensembles, J. Hydrol., 298, 222–241, 2004. 4744
Grelaud, A., Robert, C., Marin, J., Rodolphe, F., and Taly, J.: ABC likelihood-free methods for model choice in Gibbs random fields, Bayesian Analysis, 4, 317–336, 2009. 4746
Gupta, H., Thiemann, M., Trosset, M., and Sorooshian, S.: Reply to comment by Beven and Young on "Bayesian recursive parameter estimation for hydrologic models" by Thiemann et al., Water Resour. Res., 39, 1117, doi:10.1029/2002WR001405, 2003. 4746
Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Toward improved calibration of hydrologic models: multiple and noncommensurable measures of information, Water Resour. Res., 34, 751–763, 1998. 4744, 4746, 4753
Gupta, V. K. and Sorooshian, S.: Uniqueness and observability of conceptual rainfall-runoff model parameters: the percolation process examined, Water Resour. Res., 19, 269–276, 1983. 4743

HESSD
10, 4739–4797, 2013

Gupta, V. K. and Sorooshian, S.: The relationship between data and the precision of parameter estimates of hydrological models, J. Hydrol., 81, 57–77, 1985. 4743
Hall, P., Kay, J. W., and Titterington, D. M.: Asymptotically optimal difference-based estimation of variance in nonparametric regression, Biometrika, 77, 521–528, 1990. 4757
Haario, H., Saksman, E., and Tamminen, J.: Adaptive proposal distribution for random walk Metropolis algorithm, Computation. Stat., 14, 375–395, 1999. 4752
Haario, H., Saksman, E., and Tamminen, J.: An adaptive Metropolis algorithm, Bernoulli, 7, 223–242, 2001. 4752
Hankin, B. G., Hardy, R., Kettle, H., and Beven, K. J.: Using CFD in a GLUE framework to model the flow and dispersion characteristics of a natural fluvial dead zone, Earth Surf. Proc. Land., 26, 667–687, 2001. 4746
Hansson, K. and Lundin, C.: Equifinality and sensitivity in freezing and thawing simulations of laboratory and in situ data, Cold Reg. Sci. Technol., 44, 20–37, 2006. 4746
Huard, D. and Mailhot, A.: A Bayesian perspective on input uncertainty in model calibration: application to hydrological model "abc", Water Resour. Res., 42, W07416, doi:10.1029/2005WR004661, 2006. 4744
Ibbitt, R. P.: Effects of random data errors on the parameter values for a conceptual model, Water Resour. Res., 8, 70–78, 1972. 4743
Ibbitt, R. P. and O'Donnell, T.: Designing conceptual catchment models for automatic fitting methods, Mathematical Models in Hydrology Symposium, IAHS-AISH, Publication No. 2, 461–475, 1974. 4742
Jacquin, A. P. and Shamseldin, A. Y.: Development of a possibilistic method for the evaluation of predictive uncertainty in rainfall-runoff modeling, Water Resour. Res., 43, W04425, doi:10.1029/2006WR005072, 2007. 4744
Jensen, J. B.: Parameter and Uncertainty Estimation in Groundwater Modelling, Ph. D. thesis, Department of Civil Engineering, Aalborg University, Series Paper No. 23, 2003. 4746
Jimenez, R., Verde, L., Peiris, H., and Kosowsky, A.: Fast cosmological parameter estimation from microwave background temperature and polarization power spectra, Phys. Rev. D, 70, 23005, doi:10.1103/PhysRevD.70.023005, 2004. 4741
Johnston, P. R. and Pilgrim, D.: Parameter optimization for watershed models, Water Resour. Res., 12, 477–486, 1976. 4743
Joyce, P. and Marjoram, P.: Approximately suficient statistics and Bayesian Computation, Stat. Appl. Genet. Mo. B., 7, ISSN (Online) 1544-6115, doi:10.2202/1544-6115.1389, 2008. 4746

HESSD
10, 4739–4797, 2013

Kaipio, J. P., Seppänen, A., Somersalo, E., and Haario, H.: Posterior covariance related optimal current patterns in electrical impedance tomography, Inverse Probl., 20, 919–936, 2004. 4741, 4748

Kavetski, D. and Clark, M. P.: Ancient numerical daemons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction, Water Resour. Res., 46, W10511, doi:10.1029/2009WR008896, 2010. 4744

Kavetski, D., Kuczera, G., and Franks, S. W.: Semi-distributed hydrological modeling: a "saturation path" perspective on TOPMODEL and VIC, Water Resour. Res., 39, 1246, doi:10.1029/2003WR002122, 2003. 4744

Kavetski, D., Kuczera, G., and Franks, S. W.: Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, Water Resour. Res., 42, W03407, doi:10.1029/2005WR004368, 2006a. 4742, 4744, 4761

Kavetski, D., Kuczera, G., and Franks, S. W.: Bayesian analysis of input uncertainty in hydrological modeling: 2. Application, Water Resour. Res., 42, W03408, doi:10.1029/2005WR004376, 2006b. 4744, 4761

Kavetski, D., Kuczera, G., and Franks, S. W.: Calibration of conceptual hydrological models revisited: 1. Overcoming numerical artifacts, J. Hydrol., 320, 173–186, doi:10.1016/j.jhydrol.2005.07.012, 2006c. 4744

Khu, S. T. and Madsen, H.: Multiobjective calibration with Pareto preference ordering: an application to rainfall-runoff model calibration, Water Resour. Res., 41, W03004, doi:10.1029/2004WR003041, 2005. 4743

Kuczera, G.: On the relationship between the reliability of parameter estimates and hydrologic time series used in calibration, Water Resour. Res., 18, 146–154, 1982. 4743

Kuczera, G.: Improved parameter inference in catchment models, 1. Evaluating parameter uncertainty, Water Resour. Res., 19, 1151–1162, doi:10.1029/WR019i005p01151, 1983a. 4742, 4743

Kuczera, G.: Improved parameter inference in catchment models: 2. Combining different kinds of hydrologic data and testing their compatibility, Water Resour. Res., 19, 1163–1172, 1983b. 4743

Kuczera, G. and Parent, E.: Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm, J. Hydrol., 211, 69–85, 1998. 4744

HESSD
10, 4739–4797, 2013

Kuczera, G., Kavetski, D., Franks, S., and Thyer, M.: Towards a Bayesian total error analysis of conceptual rainfall-runoff models: characterizing model error using storm-dependent parameters, J. Hydrol., 331, 161–177, doi:10.1016/j.jhydrol.2006.05.010, 2006. 4744

Kuczera, G., Kavetski, D., Renard, B., and Thyer, M.: A limited-memory acceleration strategy for MCMC sampling in hierarchical Bayesian calibration of hydrological models, Water Resour. Res., 46, W07602, doi:10.1029/2009WR008985, 2010. 4744

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, W01526, doi:10.1029/2011WR010608, 2012. 4744, 4748

Lamb, R., Beven, K., and Myrabø, S.: Use of spatially distributed water table observations to constrain uncertainty in a rainfall-runoff model, Adv. Water Resour., 22, 305–317, 1998. 4746

Mantovan, P. and Todini, E.: Hydrological forecasting uncertainty assessment: incoherence of the GLUE methodology, J. Hydrol., 330, 368–381, doi:10.1016/j.jhydrol.2006.04.046, 2006. 4746

Marjoram, P., Molitor, J., Plagnol, V., and Tavaré, S.: Markov chain Monte Carlo without likelihoods, P. Natl. Acad. Sci. USA, 100, 15324–15328, 2003. 4746, 4750

Marshall, L. A., Sharma, A., and Nott, D. J.: Towards dynamic catchment modelling: a Bayesian hierarchical mixtures of experts framework, Hydrol. Proccess., 21, 847–861, 2007. 4742, 4744

Mazi, K., Koussis, A. D., Restrepo, P. J., and Koutsoyiannis, D.: A groundwater-based, objective-heuristic parameter optimization method for a precipitation-runoff model and its application to a semi-arid basin, J. Hydrol., 290, 243–258, 2004. 4743

McMichael, C. E., Hope, A. S., and Loaiciga, H. A.: Distributed hydrological modeling in California semi-arid shrublands: MIKE SHE model calibration and uncertainty estimation, J. Hydrol., 317, 307–324, 2006. 4746

Mertens, J., Madsen, H., Feyen, L., Jacques, D., and Feyen, J.: Including prior information in the estimation of effective soil parameters in unsaturated zone modelling, J. Hydrol., 294, 251–269, 2004. 4746

Montanari, A.: Large sample behaviors of the generalized likelihood uncertainty estimation(GLUE) in assessing the uncertainty of rainfall-runoff simulations, Water Resour. Res., 41, W08406, doi:10.1029/2004WR003826, 2005. 4746

Montanari, A. and Brath, A.: A stocastic approach for assessing the uncertainty of rainfall-runoff simulations, Water Resour. Res., 40, W01106, doi:10.1029/2003WR002540, 2004. 4744

HESSD
10, 4739–4797, 2013

Montanari, A. and Grossi, G.: Estimating the uncertainty of hydrological forecasts: a statistical approach, Water Resour. Res., 44, W00B08, doi:10.1029/2008WR006897, 2008. 4744
Moradkani, H., Hsu, K., Gupta, H., and Sorooshian, S.: Uncertainty assessment of hydrologic model states and parameters: sequential data assimilation using the particle filter, Water Resour. Res., 41, W05012, doi:10.1029/2004WR003604, 2005a. 4744
Moradkhani, H., Sorooshian, S., Gupta, H. V., and Houser, P. R.: Dual state–parameter estimation of hydrological models using ensemble Kalman filter, Adv. Water Resour., 28, 135–147, 2005b. 4744
Muleta, M. K. and Nicklow, J. W.: Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model, J. Hydrol., 306, 127–145, 2005. 4746
Nash, J. E.: A unit hydrograph study with particular reference to British catchments, P. I. Civil. Eng., 17, 249–282, 1960. 4751
Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models, Part I: A discussion of principles, J. Hydrol., 10, 282–290, 1970. 4753
Nissinen, A., Heikkinen, L. M., Kolehmainen, V., and Kaipio, J. P.: Compensation of errors due to discretization, domain truncation and unknown contact impedances in electrical impedance tomography, Meas. Sci. Technol., 20, 105504, doi:10.1088/0957-0233/20/10/105504, 2009. 4748
Nott, D. J., Marshall, L., and Brown, J.: Generalized likelihood uncertainty estimation (GLUE) and approximate Bayesian Computation: what's the connection?, Water Resour. Res., 48, W12602, doi:10.1029/2011WR011128, 2012. 4740, 4747, 4764
Pauwels, V. R. N.: A multi-start weight-adaptive recursive parameter estimation method, Water Resour. Res., 44, W044156, doi:10.1029/2007WR005866, 2008. 4743
Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A., and Feldman, M. T.: Population growth of human Y chromosomes: a study of Y chromosome microsatellites, Mol. Biol. Evol., 16, 1791–1798, 1999. 4746, 4750, 4763
Ratmann, O., Andrieu, C., Wiuf, C., and Richardson, S.: Model criticism based on likelihood-free inference, with an application to protein network evolution, P. Natl. Acad. Sci. USA, 106, 1–6, 2009. 4746, 4750
Reichle, R. H.: Data assimilation methods in the earth sciences, Adv. Water Resour., 31, 1411–1418, 2008. 4744

HESSD
10, 4739–4797, 2013

Reichert, P. and Mieleitner, J.: Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters, Water Resour. Res., 45, W10402, doi:10.1029/2009WR007814, 2009. 4744

Renard, B., Kavetski, D., Leblois, E., Thyer, M., Kuczera, G., and Franks, S. W.: Toward a reliable decomposition of predictive uncertainty in hydrological modeling: characterizing rainfall errors using conditional simulation, Water Resour. Res., 47, W11516, doi:10.1029/2011WR010643, 2011. 4744

Restrepo-Posada, J. P.: Automatic parameter estimation of a large conceptual rainfall-runoff model: a maximum likelihood approach, Ph. D. thesis, Massachusetts Institute of Technology, Cambridge, MA, 330 pp., 1982. 4743

Rice, J.: Bandwidth choice for nonparametric kernel regression, Ann. Stat., 12, 1215–1230, 1984. 4757

Rings, J., Vrugt, J. A., Schoups, G., Huisman, J. A., and Vereecken, H.: Bayesian model averaging using particle filtering and Gaussian mixture modeling: theory, concepts, and simulation experiments, Water Resour. Res., 48, W05520, doi:10.1029/2011WR011607, 2012. 4744

Romanowicz, R. J., Beven, K. J., and Tawn, J.: Bayesian calibration of flood inundation models, in: Floodplain Processes, edited by: Anderson, M. G. and Walling, D. E., Wiley, Chichester, 333–360, 1996. 4746

Salamon, P. and Feyen, L.: Assessing parameter, precipitation, and predictive uncertainty in a distibuted hydrological model using sequential data assimilatiom with the particle filter, J. Hydrol., 376, 428–442, 2009. 4744

Schoups, G. and Vrugt, J. A.: A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic and non-Gaussian errors, Water Resour. Res., 46, W10531, doi:10.1029/2009WR008933, 2010a. 4742, 4747, 4748, 4756

Schoups, G., Vrugt, J. A., Fenicia, F., and van de Giesen, N. C.: Corruption of accuracy and efficiency of Markov Chain Monte Carlo simulation by inaccurate numerical implementation of conceptual hydrologic models, Water Resour. Res., 46, W10530, doi:10.1029/2009WR008648, 2010b. 4744

Seibert, J.: Multi-criteria calibration of a conceptual runoff model using a genetic algorithm, Hydrol. Earth Syst. Sci., 4, 215–224, doi:10.5194/hess-4-215-2000, 2000. 4743

Seifert, B., Gasser, T., and Wolf, A.: Nonparametric estimation of residual variance revisited, Biometrika, 80, 373–383, 1993. 4757

HESSD
10, 4739–4797, 2013

Sisson, S. A., Fan, Y., and Tanaka, M. M.: Sequential Monte Carlo without likelihoods, P. Natl. Acad. Sci. USA, 104, 1760–1765, 2007. 4746, 4750

Slater, A. G. and Clark, M. P.: Snow data assimilation via an ensemble Kalman filter, J. Hydrometeorol., 7, 478–493, 2006. 4744

Smith, P., Beven, K. J., and Tawn, J. A.: Detection of structural inadequacy in process-based hydrological models: a particle-filtering approach, Water Resour. Res., 44, W01410, doi:10.1029/2006WR005205, 2008. 4744

Smith, T., Sharma, A., Marshall, L., Mehrotra, R., and Sisson, S.: Development of a formal likelihood function for improved Bayesian inference of ephemeral catchments, Water Resour. Res., 46, W12551, doi:10.1029/2010WR009514, 2010. 4742

Solomatine, D. P. and Shrestha, D. L.: A novel method to estimate model uncertainty using machine learning techniques, Water Resour. Res., 45, W00B11, doi:10.1029/2008WR006839, 2009. 4744

Sorooshian, S. and Dracup, J. A.: Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: correlated and heteroscedastic error cases, Water Resour. Res., 16, 430–442, 1980. 4742, 4743

Sorooshian, S. and Gupta, V. K.: Automatic calibration of conceptual rainfall-runoff models: the question of parameter observability and uniqueness, Water Resour. Res., 19, 260–268, 1983. 4743

Sorooshian, S., Gupta, V. K., and Fulton, J. L.: Evaluation of maximum likelihood parameter estimation techniques for conceptual rainfall-runoff models: influence of calibration data variability and length on model credibility, Water Resour. Res., 19, 251–259, 1983. 4743

Stedinger, J. R., Vogel, R. M., Lee, S. U., and Batchelder, R.: Appraisal of Generalized Likelihood Uncertainty Estimation(GLUE) methodology, Water Resour. Res., 44, WOOB06, doi:10.1029/2008WR006822, 2008. 4746

Stenerud, V. R., Kippe, V., Lie, K., and Datta-Gupta, A.: Adaptive multiscale streamline simulation and inversion for high-resolution geomodels, Soc. Petrol. Eng. J., 13 , 99–111, 2008. 4741

Tadesse, A. and Anagnostou, E. N.: A statistical approach to ground radar-rainfall estimation, J. Atmos. Ocean. Tech., 22, 1055–1071, 2005. 4746

Tolson, B. A. and Shoemaker, C. A.: Dynamically dimensioned search algorithm for computationally efficient watershed model calibration, Water Resour. Res., 43, W01413, doi:10.1029/2005WR004723, 2007. 4743

HESSD

10, 4739–4797, 2013

Troutman, B. M.: Errors and parameter estimation in precipitation-runoff modeling: I. Theory, Water Resour. Res., 21, 1195–1213, 1985a. 4743
Troutman, B. M.: Errors and parameter estimation in precipitation-runoff modeling: II. Case study, Water Resour. Res., 21, 1214–1222, 1985b. 4743
Turner, B. M. and van Zandt, T.: A tutorial on approximate Bayesian Computation, J. Math. Psychol, 56, 69–85, 2012. 4750, 4751, 4752, 4765, 4766
Vrugt, J. A. and Robinson, B. A.: Improved evolutionary optimization from genetically adaptive multimethod search, P. Natl. Acad. Sci. USA, 104, 708–711, 2007. 4743, 4744
Vrugt, J. A. and Sadegh, M.: Toward diagnostic model calibration and evaluation: approximate Bayesian Computation, Water Resour. Res., in review, 2013. 4752, 4753, 4754, 4763, 4764
Vrugt, J. A., Bouten, W., Gupta, H. V., and Sorooshian, S.: Toward improved identifiability of hydrologic model parameters: the information content of experimental data, Water Resour. Res., 38, 1312, doi:10.1029/2001WR001118, 2002. 4743, 4744
Vrugt, J. A., Diks, C. G. H., Bouten, W., Gupta, H. V., and Verstraten, J. M.: Improved treatment of uncertainty in hydrologic modeling: combining the strengths of global optimization and data assimilation, Water Resour. Res., 41, W01017, doi:10.1029/2004WR003059, 2005. 4744, 4757
Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M., and Robinson, B. A.: Treatment of input uncertainty in hydrologic modeling: doing hydrology backward with Markov chain Monte Carlo simulation, Water Resour. Res., 44, W00B09, doi:10.1029/2007WR006720, 2008a. 4744, 4752, 4761
Vrugt, J. A., ter Braak, C. J. F., Gupta, H. V., and Robinson, B. A.: Equifinality of formal (DREAM) and informal.(GLUE) Bayesian approaches in hydrologic modeling?, Stoch. Env. Res. Risk A., 23, 1011–1026, 2008b. 4744, 4746, 4748
Vrugt, J. A., ter Braak, C. J. F., Gupta, H. V., and Robinson, B. A.: Response to comment by Keith Beven on "Equifinality of formal.(DREAM) and informal.(GLUE) Bayesian approaches in hydrologic modeling?" by Vrugt et al., Stoch. Env. Res. Risk A., doi:10.1007/s00477-008-0284-9, 2008c. 4746, 4747
Vrugt, J. A., ter Braak, C. J. F., Diks, C. G. H., Higdon, D., Robinson, B. A., and Hyman, J. M.: Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling, Int. J. Nonlin. Sci. Num., 10, 273–290, 2009a. 4752
Vrugt, J. A., Robinson, B. A., and Hyman, J. M.: Self-adaptive multimethod search for global optimization in real-parameter spaces, IEEE T. Evolut. Comput., 13, 243–259, 2009b. 4743

HESSD
10, 4739–4797, 2013

Vrugt, J. A., ter Braak, C. J. F., Diks, C. G. H., and Schoups, G.: Hydrologic data assimilation using particle Markov chain Monte Carlo simulation: theory, concepts and applications, Adv. Water Resour., 51, 457–478, doi:10.1016/j.advwatres.2012.04.002, 2012. 4744

Wagener, T., McIntyre, N., Lees, M. J., Wheater, H. S., and Gupta, H. V.: Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis, Hydrol. Proccess., 17, 455–476, 2003. 4744

Wang, Q.: The genetic algorithm and its application for calibrating conceptual rainfall-runoff models, Water Resour. Res., 27, 2467–2471, 1991. 4743

Wang, X., He, X., Williams, J. R., Izaurralde, R. C., and Atwood, J. D.: Sensitivity and uncertainty analyses of crop yields and soil organic carbon simulated with EPIC, T. ASAE, 48, 1041–1054, 2005. 4746

Yapo, P. O., Gupta, H. V., and Sorooshian, S.: Calibration of conceptual rainfall-runoff models: sensitivity to calibration data, J. Hydrol., 181, 23–48, 1996. 4743

Yapo, P. O., Gupta, H. V., and Sorooshian, S.: Multi-objective global optimization for hydrologic models. J. Hydrol., 204, 83–97, 1998. 4743

Young, P. C.: Advances in real-time flood-forecasting forecasting, Philos. T. R. Soc. A, 360, 1433–1450, 2002. 4744

Young, P. C.: Hypothetico-inductive data-based mechanistic modeling of hydrological systems, Water Resour. Res., 49, 915–935, doi:10.1002/wrcr.20068, 2013. 4744

HESSD
10, 4739–4797, 2013

Table 1. Prior uncertainty ranges of hmodel parameters.

Parameter	Symbol	Minimum	Maximum	Units
Maximum interception	I_{max}	0	10	mm
Soil water storage capacity	S_{max}	10	1000	mm
Maximum percolation rate	Q_{max}	0	100	$mm\,d^{-1}$
Evaporation parameter	α_E	0	100	–
Runoff parameter	α_F	−10	10	–
Time constant, fast reservoir	K_F	0	10	days
Time constant, slow reservoir	K_S	0	150	days

HESSD
10, 4739–4797, 2013

Table 2. Description of the SAC-SMA model parameters and their (uniform) prior uncertainty ranges.

Parameter	Symbol	Minimum	Maximum	Units
Upper zone tension water maximum storage	UZTWM	1.0	150.0	mm
Upper zone free water maximum storage	UZFWM	1.0	150.0	mm
Lower zone tension water maximum storage	LZTWM	1.0	500.0	mm
Lower zone free water primary maximum storage	LZFPM	1.0	1000.0	mm
Lower zone free water supplemental maximum storage	LZFSM	1.0	1000.0	mm
Additional impervious area	ADIMP	0.0	0.40	–
Upper zone free water lateral depletion rate	UZK	0.1	0.5	day^{-1}
Lower zone primary free water depletion rate	LZPK	0.0001	0.025	day^{-1}
Lower zone supplemental free water depletion rate	LZSK	0.01	0.25	day^{-1}
Maximum percolation rate	ZPERC	1.0	250.0	–
Exponent of the percolation equation	REXP	1.0	5.0	–
Impervious fraction of the watershed area	PCTIM	0.0	0.1	–
Fraction percolating from upper to lower zone free water storage	PFREE	0.0	0.6	–

HESSD
10, 4739–4797, 2013

Table 3. Computational efficiency of GLUE and ABC for the French Broad data set: acceptance rate (AR, %) and total number of SAC-SMA and hmodel function evaluations (FE) required to sample N = 1000 behavioral solutions.

	SAC-SMA		hmodel	
	ABC	GLUE	ABC	GLUE
AR, %	0.41	0.016	0.50	0.06
FE, –	242 004	6 110 640*	201 192	1 608 810*

* Derived from linear scaling of FE needed to sample 300 (SAC-SMA) and 100 (hmodel) behavioral solutions.

HESSD
10, 4739–4797, 2013

Table 4. Summary statistics of the GLUE, ABC and DREAM (formal likelihood function) derived posterior parameter distribution for the calibration and evaluation period of the French Broad river basin: root mean square error (RMSE) of the posterior mean SAC-SMA simulation and coverage of the associated 95 % streamflow simulation uncertainty ranges.

	ABC		DREAM		GLUE	
	Calibration	Evaluation	Calibration	Evaluation	Calibration	Evaluation
RMSE, $m^3 s^{-1}$	6.90	5.45	4.81	4.86	7.07	5.49
Coverage, %	73.89	69.89	17.35	11.59	76.74	71.08

HESSD
10, 4739–4797, 2013

Table 5. Summary statistics of the GLUE, ABC and DREAM (formal likelihood function) derived posterior parameter distribution for the calibration and evaluation period of the French Broad river basin: root mean square error (RMSE) of the posterior mean hmodel simulation and coverage (%) of the associated 95 % streamflow simulation uncertainty ranges.

	ABC		DREAM		GLUE	
	Calibration	Evaluation	Calibration	Evaluation	Calibration	Evaluation
RMSE, $m^3 s^{-1}$	6.53	6.00	5.24	5.09	6.55	5.98
Coverage, %	67.38	61.86	7.33	10.58	68.14	60.86

HESSD
10, 4739–4797, 2013

Table 6. Summary statistics of the GLUE, ABC and DREAM (formal likelihood function) derived posterior parameter distribution for the calibration and evaluation period of the Leaf River watershed: root mean square error (RMSE) of the posterior mean hmodel simulation and coverage (%) of the associated 95 % streamflow simulation uncertainty ranges.

	ABC		DREAM		GLUE	
	Calibration	Evaluation	Calibration	Evaluation	Calibration	Evaluation
RMSE, $m^3 s^{-1}$	24.87	23.23	16.45	19.28	25.02	23.42
Coverage, %	80.02	64.78	22.11	21.90	82.54	68.25

HESSD
10, 4739–4797, 2013

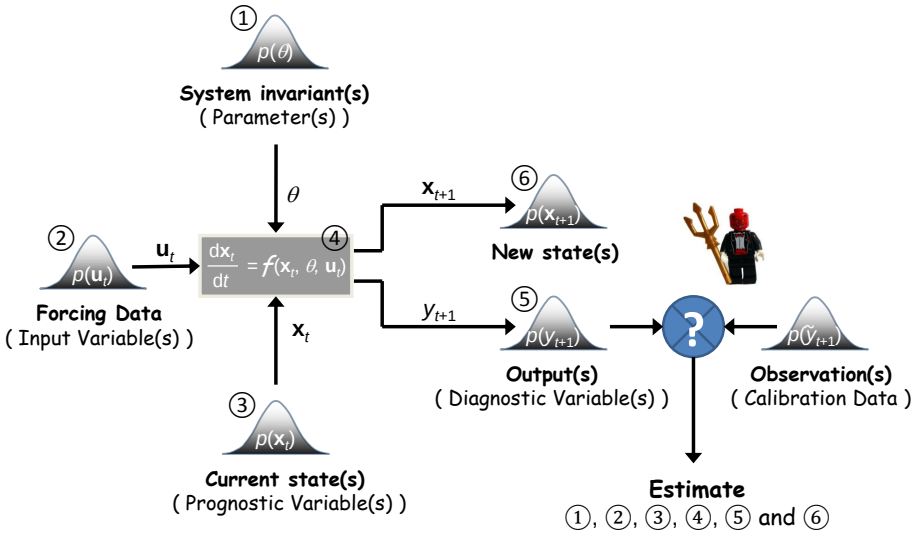

Fig. 1. Explicit recognition of the role of (1) parameter, (2) forcing data, (3) initial state, (4) model structural, (5) output, and (6) state uncertainty. The pitchfork symbol illuminates the difficulty which formulation of the likelihood function (and prior distribution/parameterization of individual error sources) to use to summarize the error residuals. Explicit treatment of individual error sources is required to increase the prospects of explaining the reasons for model inadequacy and learning from the experimental data.

HESSD
10, 4739–4797, 2013

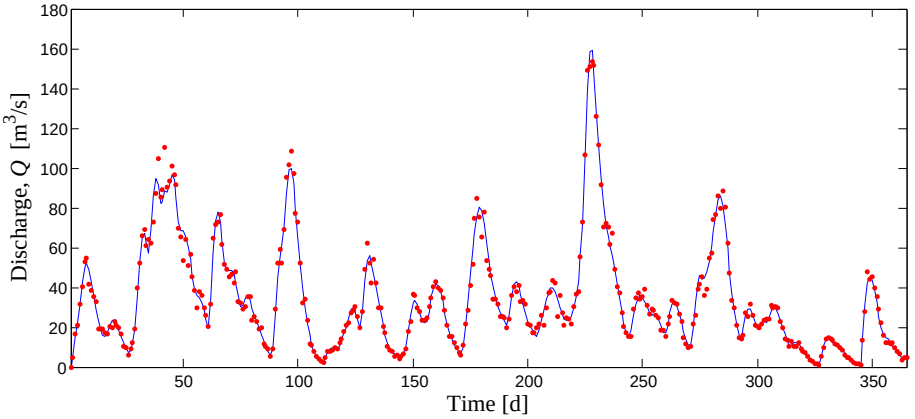

Fig. 2. Synthetic discharge time series (blue line) simulated with the Nash Cascade model, and the error corrupted observations (red points) used in the GLUE and ABC analysis.

HESSD
10, 4739–4797, 2013

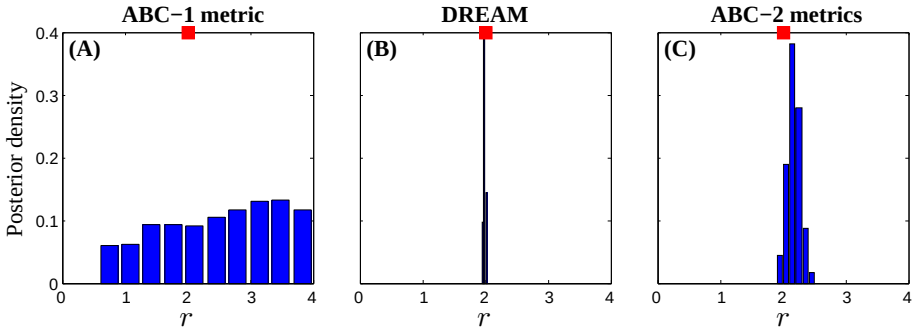

Fig. 3. Posterior distribution of the recession parameter derived from **(A)** ABC using the mean of the data as summary metric, **(B)** DREAM using a heteroscedastic measurement error, and **(C)** ABC using the mean and standard deviation of the data as summary statistics. The true parameter value is indicated with the symbol "■".

HESSD
10, 4739–4797, 2013

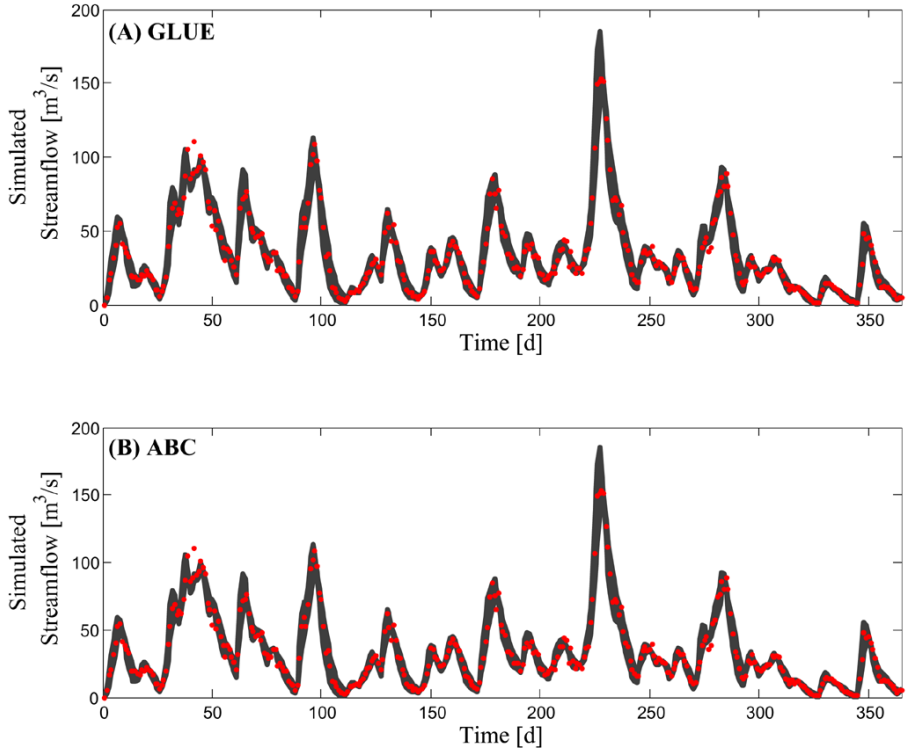

Fig. 4. 95 % streamflow uncertainty ranges (dark region) derived from GLUE **(A)** and ABC **(B)**. The red points mark the actual discharge observations. The prediction uncertainty ranges derived with both methods appear virtually identical and nicely capture the desired percentage of streamflow observations.

HESSD
10, 4739–4797, 2013

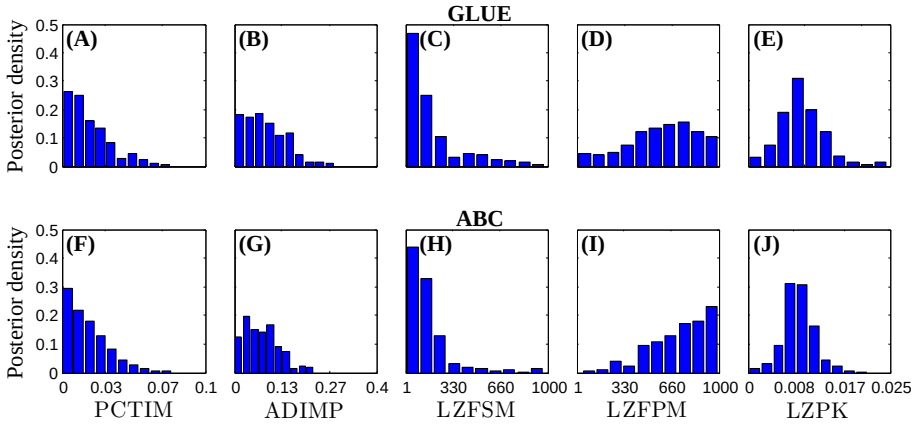

Fig. 5. Posterior distribution of five randomly chosen SAC-SMA model parameters derived from **(A–E)** GLUE, and **(F–J)** ABC using historical streamflow data from the French Broad river basin.

HESSD
10, 4739–4797, 2013

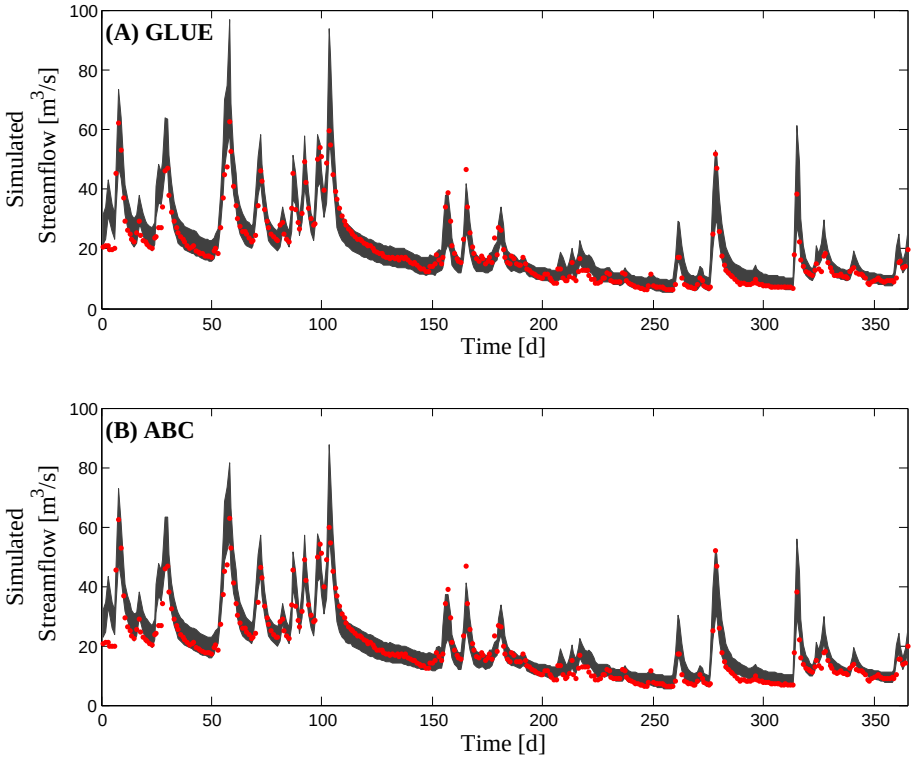

Fig. 6. SAC-SMA derived 95 % streamflow simulation uncertainty ranges (grey region) of the calibration period of the French Broad river basin using **(A)** GLUE and **(B)** ABC. The observed discharge data are indicated with the solid red dots. We limit our display to the first 365 days of the calibration data set to simplify graphical interpretation. The simulation uncertainty ranges appear very similar and nicely cover the observed discharge data.

HESSD
10, 4739–4797, 2013

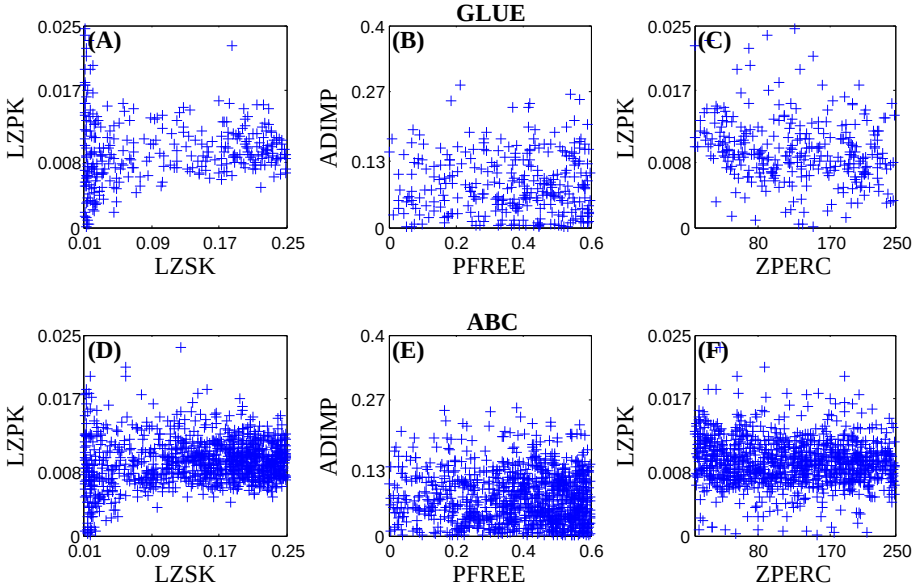

Fig. 7. Bivariate scatter plots of the behavioral (posterior) samples of three different (randomly selected) parameter pairs using GLUE (top panel) and ABC (bottom panel): **(A,D)** LZSK–LZPK, **(B,E)** PFREE–ADIMP, and **(C,F)** ZPERC–LZPK. The scatter plots derived with both methods are in close agreement but demonstrate an important difference in sampling density. The computational budget of GLUE was limited to approximately 2 days, and within this time frame the Latin hypercube sampling method located only 300 behavioral solutions.

HESSD
10, 4739–4797, 2013

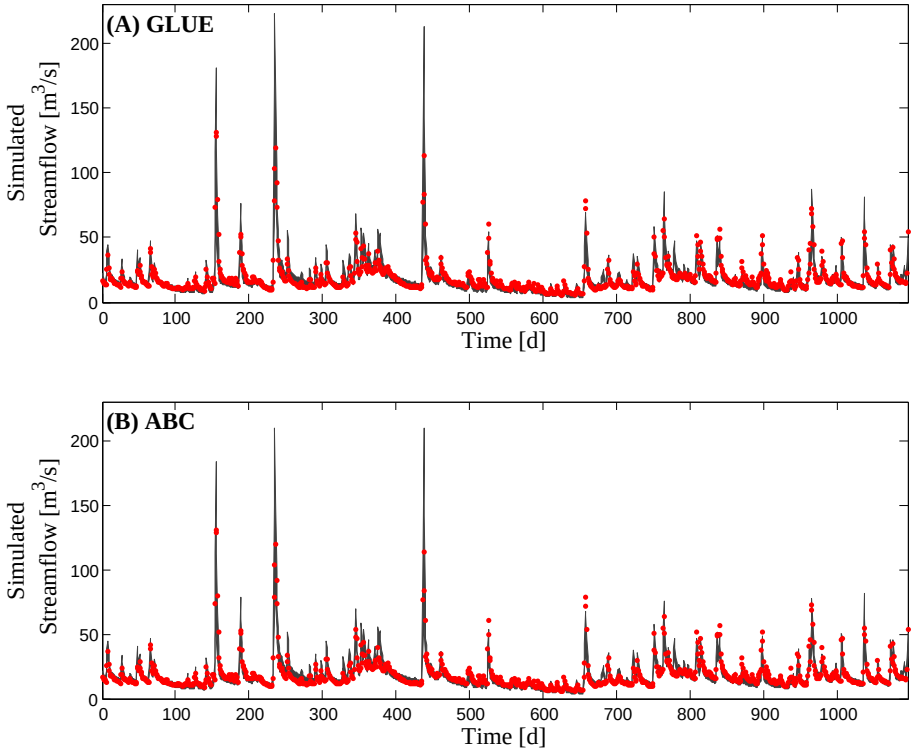

Fig. 8. SAC-SMA derived 95 % streamflow simulation uncertainty ranges (grey region) for a three-year portion of the evaluation period of the French Broad river basin using the **(A)** GLUE and **(B)** ABC derived posterior parameter distribution. The observed discharge data are indicated with the solid red dots.

HESSD
10, 4739–4797, 2013

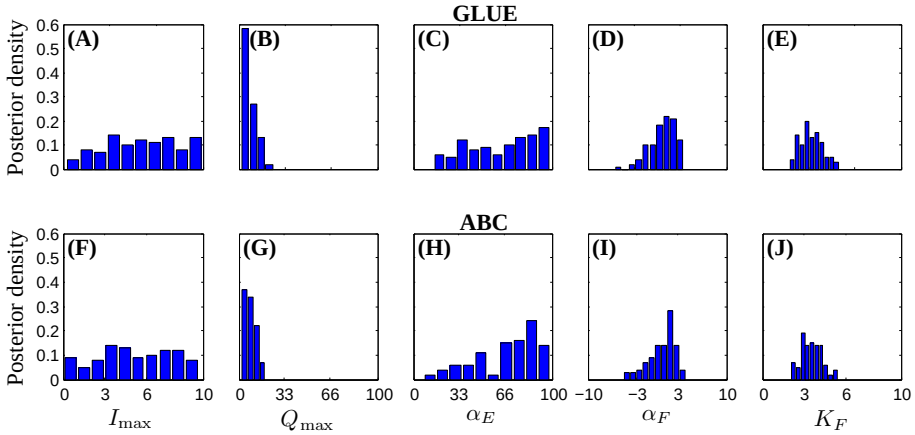

Fig. 9. Posterior distribution of five randomly selected hModel parameters derived from the French Broad calibration data set, **(A,F)** I_{max}, **(B,G)** Q_{max}, **(C,H)** α_E, **(D,I)** α_F, and **(E,J)** K_F.

HESSD
10, 4739–4797, 2013

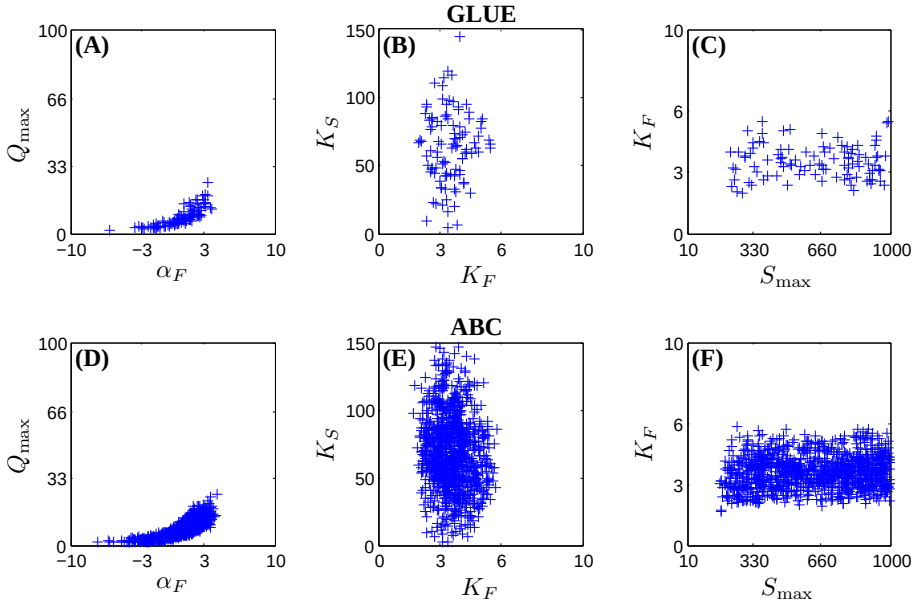

Fig. 10. Bivariate scatter plots of the behavioral (posterior) samples of three different (randomly selected) hmodel parameter pairs using GLUE (top panel) and ABC (bottom panel): **(A,D)** α_F – Q_{max}, **(B,E)** K_F–K_S, and **(C,F)** S_{max}–K_F. The scatter plots derived with both methods are in close agreement but demonstrate an important difference in sampling density. The computational budget of GLUE was limited to about 3 days, and this has resulted in 100 behavioral solutions.

HESSD
10, 4739–4797, 2013

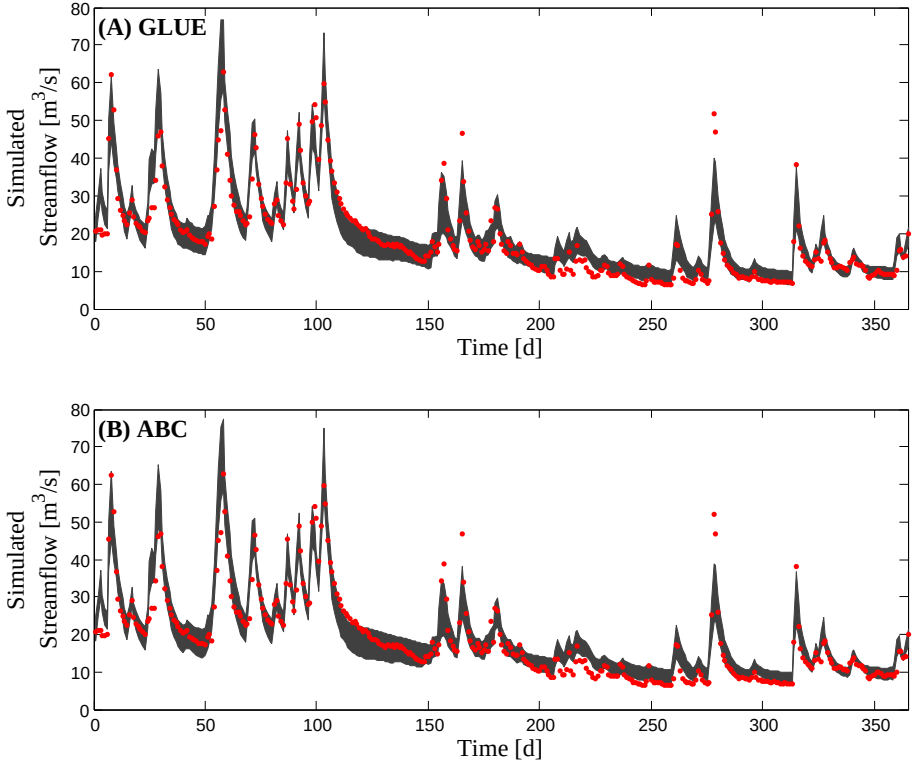

Fig. 11. hmodel derived 95 % streamflow simulation uncertainty ranges (grey region) of the calibration period of the French Broad river basin using **(A)** GLUE and **(B)** ABC. The observed discharge data are indicated with the solid red dots. We limit our display to the first 365 days of the calibration data set to facilitate graphical interpretation. The uncertainty ranges appear very similar and nicely cover the observed discharge data.

HESSD
10, 4739–4797, 2013

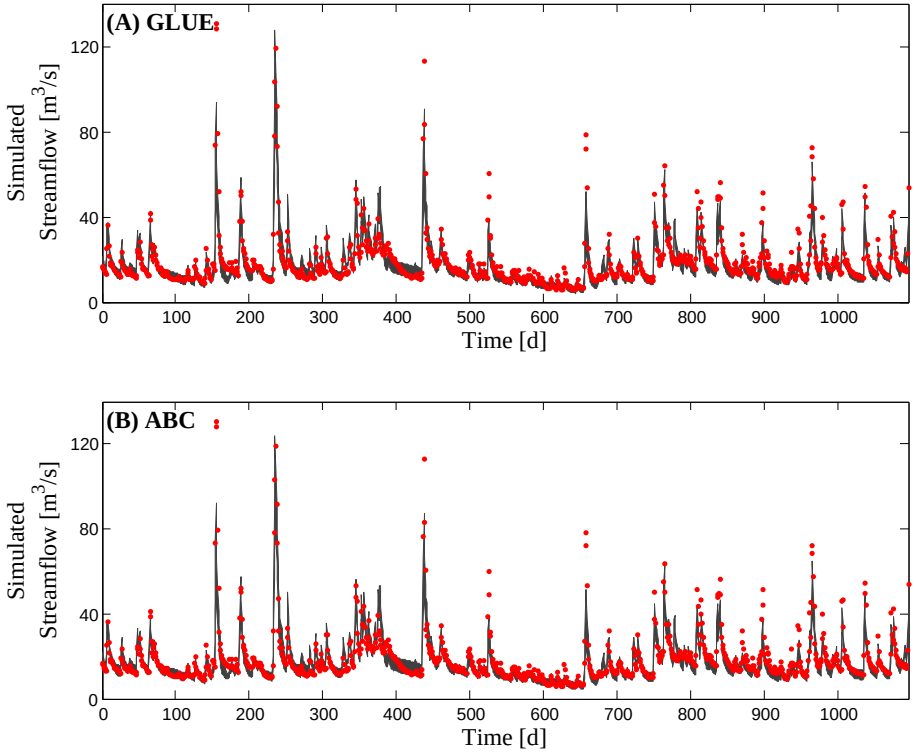

Fig. 12. hmodel derived 95 % streamflow simulation uncertainty ranges (grey region) for a three-year portion of the evaluation period of the French Broad river basin using the **(A)** GLUE and **(B)** ABC derived posterior parameter distribution. The observed discharge data are indicated with the solid red dots.

HESSD
10, 4739–4797, 2013

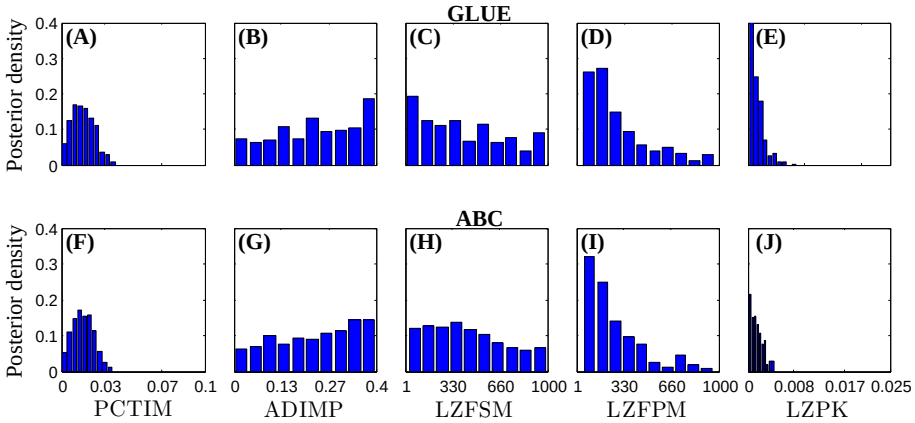

Fig. 13. Posterior distribution of five randomly chosen SAC-SMA model parameters derived from GLUE (top panel), and ABC (bottom panel) using historical streamflow data from the Leaf River watershed in Mississippi.

HESSD
10, 4739–4797, 2013

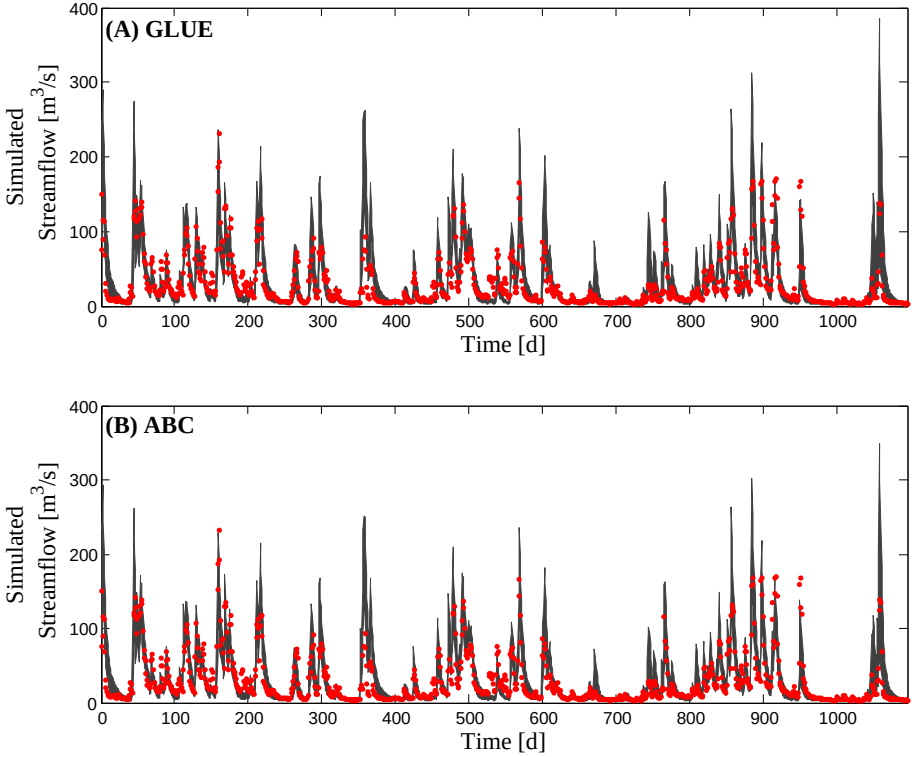

Fig. 14. SAC-SMA derived 95 % streamflow simulation uncertainty ranges (grey region) for a three-year portion of the evaluation period of the Leaf River watershed using the **(A)** GLUE and **(B)** ABC derived posterior parameter distribution. The observed discharge data are indicated with the solid red dots.

HESSD
10, 4739–4797, 2013