A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters

Jasper A. Vrugt,[1] Hoshin V. Gupta,[2] Willem Bouten,[1] and Soroosh Sorooshian[2]

Received 8 August 2002; revised 20 February 2003; accepted 2 April 2003; published 1 August 2003.

[1] Markov Chain Monte Carlo (MCMC) methods have become increasingly popular for estimating the posterior probability distribution of parameters in hydrologic models. However, MCMC methods require the a priori definition of a proposal or sampling distribution, which determines the explorative capabilities and efficiency of the sampler and therefore the statistical properties of the Markov Chain and its rate of convergence. In this paper we present an MCMC sampler entitled the Shuffled Complex Evolution Metropolis algorithm (SCEM-UA), which is well suited to infer the posterior distribution of hydrologic model parameters. The SCEM-UA algorithm is a modified version of the original SCE-UA global optimization algorithm developed by *Duan et al.* [1992]. The SCEM-UA algorithm operates by merging the strengths of the Metropolis algorithm, controlled random search, competitive evolution, and complex shuffling in order to continuously update the proposal distribution and evolve the sampler to the posterior target distribution. Three case studies demonstrate that the adaptive capability of the SCEM-UA algorithm significantly reduces the number of model simulations needed to infer the posterior distribution of the parameters when compared with the traditional Metropolis-Hastings samplers. *INDEX TERMS:* 1894 Hydrology: Instruments and techniques; 1836 Hydrology: Hydrologic budget (1655); 1821 Hydrology: Floods; 1860 Hydrology: Runoff and streamflow; 1869 Hydrology: Stochastic processes; *KEYWORDS:* parameter optimization, uncertainty assessment, Markov Chain Monte Carlo, automatic calibration, proposal distribution, hydrologic models

Citation: Vrugt, J. A., H. V. Gupta, W. Bouten, and S. Sorooshian, A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resour. Res.*, *39*(8), 1201, doi:10.1029/2002WR001642, 2003.

1. Introduction and Scope

[2] Hydrologic models often contain parameters that cannot be measured directly but which can only be inferred by a trial-and-error (calibration) process that adjusts the parameter values to closely match the input-output behavior of the model to the real system it represents. Traditional calibration procedures, which involve "manual" adjustment of the parameter values, are labor-intensive, and their success is strongly dependent on the experience of the modeler. Automatic methods for model calibration, which seek to take advantage of the speed and power of computers while being objective and relatively easy to implement, have therefore become more popular [e.g., *Boyle et al.*, 2000]. Since the early work reported by *Dawdy and O'Donnell* [1965], automatic calibration procedures have evolved significantly. However, many studies using such methods have reported difficulties in finding unique (global) parameter estimates [*Johnston and Pilgrim*, 1976; *Duan et al.*, 1992; *Sorooshian et al.*, 1993; *Gan and Biftu*, 1996].

[3] Regardless of the methodology used, most hydrologic models suffer from similar difficulties, including the existence of multiple local optima in the parameter space with both small and large domains of attraction (a subregion of the parameter space surrounding a local minimum), discontinuous first derivatives, and curving multidimensional ridges. These considerations inspired *Duan et al* [1992] to develop a powerful robust and efficient global optimization procedure, entitled the shuffled complex evolution (SCE-UA) global optimization algorithm. Numerous case studies have demonstrated that the SCE-UA algorithm is consistent, effective, and efficient in locating the optimal model parameters of a hydrological model [e.g., *Duan et al.*, 1992, 1993; *Sorooshian et al.*, 1993; *Luce and Cundy*, 1994; *Gan and Biftu*, 1996; *Tanakamaru*, 1995; *Kuczera*, 1997; *Hogue et al.*, 2000; *Boyle et al.*, 2000].

[4] While considerable attention has been given to the development of automatic calibration methods which aim to successfully find a single best set of parameter values, much less attention has been given to a realistic assessment of parameter uncertainty in hydrologic models. Estimates of hydrologic model parameters are generally error-prone, because the data used for calibration contain measurement errors and because the model never perfectly represents the system or exactly fits the data. Consequently, it is generally impossible to find a single point in the parameter space associated with good simulations; indeed, there may not even

[1]Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, Netherlands.
[2]Department of Hydrology and Water Resources, University of Arizona, Tucson, Arizona, USA.

0043-1397/03/2002WR001642$09.00

exist a well-defined region in the sense of a compact region interior to the prior parameter space. Although the SCE-UA global optimization algorithm can reliably find the global minimum in the parameter space, it still remains typically difficult, if not impossible, to find a unique "best" parameter set, whose performance measure differs significantly from other feasible parameter sets within this region. Such poor parameter identifiability may result in considerable uncertainty in the model output and, perhaps more important, make it virtually impossible to relate these parameter values to easily measurable soil or land-surface characteristics [*Schaap et al.*, 1998; *Duan et al.*, 2001; *Vrugt et al.*, 2002].

[5] Only recently have methods for realistic assessment of parameter uncertainty in hydrologic models begun to appear in the literature. These include the use of a multinormal approximation to parameter uncertainty [*Kuczera and Mroczkowski*, 1998], evaluation of likelihood ratios [*Beven and Binley*, 1992], parametric bootstrapping and Markov Chain Monte Carlo (MCMC) methods [e.g., *Tarantola*, 1987; *Kuczera and Parent*, 1998]. Because traditional statistical theory based on first-order approximations and multinormal distributions is typically unable to cope with the nonlinearity of complex models, MCMC algorithms have become increasingly popular as a class of general purpose approximation methods for problems involving complex inference, search, and optimization [*Gilks et al.*, 1996]. An MCMC method is a stochastic simulation that successively visits solutions in the parameter space with stable frequencies stemming from a fixed probability distribution. A variety of MCMC samplers can be constructed for any given problem by varying the sampling or proposal distribution subject to conditions that ensure convergence to the posterior target distribution. These algorithms originally arose from the field of statistical physics where they were used as models of physical systems that seek a state of minimal free energy. More recently, MCMC algorithms have been used in statistical inference and artificial intelligence [*Geman and Geman*, 1984; *Neal*, 1993].

[6] Recently, *Kuczera and Parent* [1998] used the Metropolis-Hastings algorithm [*Metropolis et al.*, 1953; *Hastings*, 1970], the earliest and most general class of MCMC samplers, in a Bayesian inference framework to describe parameter uncertainty in conceptual catchment models. The Metropolis-Hastings algorithm is the basic building block of classical MCMC methods and requires the choice of a proposal distribution to generate transitions in the Markov Chain. The choice of the proposal distribution determines the explorative capabilities of the sampler and therefore the statistical properties of the Markov Chain and its rate of convergence. If the selected proposal distribution closely approximates the posterior target distribution, the Markov Chain that is sampled will rapidly explore the parameter space, and it will not take long to obtain samples that can be treated as independent realizations of the target distribution of interest. However, a poor choice of the proposal distribution will result in slow convergence of the Markov Chain and an inability to recognize when convergence to a limiting distribution has been achieved. For complex hydrologic models, there is usually very little a priori knowledge available about the location of the high-probability density region within the parameter space. The proposal distribution should therefore express a great deal of initial uncertainty, thereby resulting in slow convergence to the final posterior target distribution (for example, *Beven and Binley* [1992] suggested imposing a uniform distribution over a large rectangle of parameter values). An important challenge therefore is to design MCMC samplers that exhibit fast convergence to the global optimum in the parameter space, while maintaining adequate occupation of the lower posterior probability regions of the parameter space.

[7] To improve the search efficiency of MCMC samplers, it seems natural to tune the proposal distribution during the evolution to the posterior target distribution, using information inferred from the sampling history induced by the transitions of the Markov Chain. This paper describes an adaptive MCMC sampler, entitled the Shuffled Complex Evolution Metropolis algorithm (SCEM-UA), which is an effective and efficient evolutionary MCMC sampler. The algorithm, a modification of the original SCE-UA global optimization algorithm developed by *Duan et al.* [1992], operates by merging the strengths of the Metropolis algorithm [*Metropolis et al.*, 1953], controlled random search [*Price*, 1987], competitive evolution [*Holland*, 1975], and complex shuffling [*Duan et al.*, 1992] to continuously update the proposal distribution and evolve the sampler to the posterior target distribution. The stochastic nature of the Metropolis-annealing scheme avoids the tendency of the SCE-UA algorithm to collapse to a single region of attraction (i.e., the global minimum), while information exchange (shuffling) allows biasing the search in favor of better solutions.

[8] This paper is organized as follows. In section 2, we describe the Metropolis-Hastings algorithm and the new SCEM-UA algorithm for estimating the posterior probability distribution of hydrologic model parameters. In section 3 we illustrate the power of both algorithms by means of three case studies with increasing complexity; here we are especially concerned with algorithm efficiency (particularly the number of simulations needed to converge to the stationary posterior distribution). Finally, in section 4 we summarize the methodology and discuss the results.

2. Search Algorithms for Assessment of Parameter Uncertainty

[9] We are interested in hydrologic models that predict outputs from inputs. These models are indexed by parameters, which may (or may not) be physically interpretable. We assume that the mathematical structure of the model is essentially predetermined and fixed. Following *Troutmann* [1985], the hydrologic model η can be written as

$$\hat{y} = \eta(\xi|\theta) + e \tag{1}$$

where $\hat{y}$ is $N \times 1$ vector of model outputs, $\xi = (\xi_1, \xi_2, \ldots, \xi_{1r})$ is an $N \times r$ matrix of input values, $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$ is a vector with n unknown parameters, and e is a vector of statistically independent errors with zero expectation and constant variance σ^2. In the classical approach to model calibration, the goal is to find the best attainable values of the parameters θ such that the vector of error terms, $E(\theta) = \{e(\theta)_1, e(\theta)_2, \ldots, e(\theta)_N\}$, is in some sense forced to be as close to "zero" as possible [*Gupta et al.*, 1998]. For this, the SCE-UA global optimization algorithm developed by *Duan et al.* [1992] has proven to be consistent, efficient, and effective.

[10] The goal of searching for a single optimal representation of equation (1) is, however, questionable. For instance, to quote *Kuczera and Parent* [1998, pp. 70–71], "...as hydrological models can be viewed as the result of a combination of conceptual and/or physically based transfer functions, no hydrologist should be naive enough to rely on a uniquely determined value for each of the model parameters θ, whatever the skill and imagination of the modeler may be." Most likely, a search conducted on the feasible parameter space close to the global optimum will reveal many behavioral parameter sets with quite similar performance in reproducing the observed data. If we want to be able to regionalize or relate model parameters to easily measurable land or soil-surface characteristics, a prerequisite is that the parameters be unique, preferably having a small variance. From this perspective, it is necessary to infer the parameter uncertainty resulting from calibration studies.

[11] While classical statistics consider the model parameters θ in equation (1) to be fixed but unknown, the Bayesian statistics treat them as probabilistic variables having a joint posterior probability density function (pdf), which captures the probabilistic beliefs about the parameters θ in the light of the observed data $\mathbf{y}$. The posterior pdf $p(\theta|\mathbf{y})$ is proportional to the product of the likelihood function and the prior pdf. The prior pdf with probability density (or mass) function $p(\theta)$ summarizes information about θ before any data are collected. This prior information usually consists of realistic lower and upper bounds on each of the parameters, thereby defining the feasible parameter space, $\theta \in \Theta \subset \Re^n$, and imposing a uniform (noninformative) prior distribution on this rectangle.

[12] Assuming that the residuals are mutually independent, Gaussian distributed, with constant variance, the likelihood of a parameter set $\theta^{(t)}$ for describing the observed data $\mathbf{y}$ can be computed using [*Box and Tiao*, 1973]

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

Assuming a noninformative prior of the form $p(\theta) \propto \sigma^{-1}$, *Box and Tiao* [1973] showed that the influence of σ can be integrated out, leading to the following form of the posterior density of $\theta^{(t)}$:

$$p\left(\theta^{(t)}|\mathbf{y}\right) \propto \left[M\left(\theta^{(t)}\right)\right]^{-\frac{1}{2}N} \tag{3}$$

where

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

For more information about the Bayesian inference scheme, please refer to *Box and Tiao* [1973] and, more recently, to *Thiemann et al.* [2001].

2.1. Traditional First-Order Approximation

[13] The classical approximation to obtain the posterior probability density function from equation (2) is to use a first-order Taylor series expansion of the nonlinear model equations evaluated at the globally optimal parameter estimates θ_{opt}. The estimated multivariate posterior joint probability density function of θ is then expressed as [*Box and Tiao*, 1973]

$$p(\theta|\mathbf{y}) \propto \exp\left[-\frac{1}{2\sigma^2}\left(\theta-\theta_{opt}\right)^T X^T X\left(\theta-\theta_{opt}\right)\right] \tag{5}$$

where X is the Jacobian or sensitivity matrix evaluated at θ_{opt}. The posterior marginal probability density function of $\theta^{(t)}$ is therefore approximated by the normal distribution, $N_i(\theta_{i,opt}, \sigma^2\sum_{ii})$, where $\sum_{ii}$ is the ith diagonal element of the covariance matrix computed as $\sqrt{(X^TX)^{-1}}$.

[14] If the hydrologic model is linear (or very nearly linear) in its parameters, the posterior probability density region estimated by equation (5) can give a good approximation of the actual parameter uncertainty. However, for nonlinear models (e.g., hydrologic models), this approximation can be quite poor [*Kuczera and Parent*, 1998; *Vrugt and Bouten*, 2002]. Besides exhibiting strong and nonlinear parameter interdependence, the surface of $p(\theta|\mathbf{y})$ can deviate significantly from the multinormal distribution. It may also have multiple local optima and discontinuous derivatives [*Duan et al.*, 1992]. In view of these considerations, it is evident that an explicit expression of the joint and marginal probability density functions is often not possible. Fortunately, MCMC samplers are very well suited to dealing with the peculiarities encountered in the posterior pdf of hydrologic model parameters.

2.2. Monte Carlo Sampling of Posterior Distribution: The SCEM-UA Algorithm

[15] Markov Chain schemes represent a general approach for sampling from the posterior probability distribution $p(\theta|\mathbf{y})$. A Markov Chain is generated by sampling $\theta^{(t+1)} \sim z(\theta|\theta^{(t)})$. This z is called the transition kernel or proposal distribution of the Markov Chain. Consequently, $\theta^{(t+1)}$ depends only on $\theta^{(t)}$ and not on $\theta^{(0)}$, $\theta^{(1)}$, ..., $\theta^{(t-1)}$. Ergodicity and convergence properties of MCMC algorithms to the posterior distribution have been intensively studied in recent literature, and conditions have been given for geometric convergence [*Mengersen and Tweedie*, 1996; *Roberts and Tweedie*, 1996]. In practice, this means that if one looks at the values generated by the Markov Chain, which are sufficiently far from the starting value, the successively generated parameter sets will be distributed with stable frequencies stemming from the posterior target distribution. Any statistical quantity of interest such as the probability density function and the various posterior moments can be evaluated from the generated pseudo sample. The most general and earliest MCMC algorithm, known as the Metropolis-Hastings (MH) algorithm [*Metropolis et al.*, 1953; *Hastings*, 1970], is given as follows:

[16] 1. Randomly start at a location in the feasible parameter space, $\theta^{(t)}$, and compute the posterior density, $p(\theta^{(t)}|\mathbf{y})$, relevant to this point according to equation (2) or (3).

[17] 2. Generate a new configuration $\theta^{(t+1)}$ from $z(\theta|\theta^{(t)})$, where $\theta^{(t+1)}$ is called a candidate point and z ( ) is called the proposal distribution.

[18] 3. Evaluate $p(\theta^{(t+1)}|\mathbf{y})$ using equation (2) or (3) and compute $\Omega = p(\theta^{(t+1)}|\mathbf{y})/p(\theta^{(t)}|\mathbf{y})$.

[19] 4. Randomly sample a uniform label Z over the interval 0 to 1.

[20] 5. If $Z \leq \Omega$, then accept the new configuration. However, if $Z > \Omega$, then reject the candidate point and remain at the current position, that is, $\theta^{(t+1)} = \theta^{(t)}$.

[21] 6. Increment t. If t is less than a prespecified number of draws, then return to step 2.

[22] Note that the MH algorithm will always accept candidate points (jumps) into a region of higher posterior probability but will also explore regions with lower posterior probability with probability Z. Indeed, this algorithm is a MCMC sampler generating a sequence of parameter sets, $\{\theta^{(0)}, \theta^{(1)}, \ldots, \theta^{(t)}\}$, that converges to the posterior probability distribution, $p(\theta|\mathbf{y})$, for large t [*Gelman et al.*, 1995]. However, the shape and the size of the proposal distribution $z(\ )$ is known to be very crucial for the convergence properties of the Markov Chain [see, for example, *Gilks et al.*, 1995, 1998]. When the proposal distribution is too large, too many candidate points are rejected, and therefore the chain slowly covers the target distribution. On the other hand, when the proposal distribution is too small, too many candidate points are accepted, and the chain traverses slowly through the parameter space, thereby resulting in slow convergence.

[23] Various approaches have been suggested to improve the efficacy and efficiency of MCMC samplers. For instance, the proposal distribution can be updated during the evolution to the posterior target distribution using information from the sampling history induced in the transitions of the Markov Chain [*Gilks et al.*, 1996; *Haario et al.*, 1999, 2001]. However, care must be taken to ensure that the adaptation process does not destroy the overall ergodicity of the Markov Chain or diminish the convergence rate. Some authors have suggested that performance might be improved by exchanging information among multiple samplers running in parallel [e.g., *Geyer*, 1991; *Kass and Raftery*, 1995]. One possibility, recently implemented in a hydrologic application by *Kuczera and Parent* [1998], is to periodically update the covariance structure of the proposal-jump distribution by selecting a sample of points generated by each multiple sequence. However, this approach is subject to the same difficulty encountered by any adaptive sampler: how to use information in a way that ensures convergence to the stationary target distribution and desirable asymptotic properties of kernel density estimates derived from the sampler. An important challenge is therefore to design samplers that rapidly converge to the global minimum in the parameters space, while maintaining sufficient occupation of the lower posterior probability regions of the parameter space.

[24] In examining ways to increase information exchange between the sampled candidate points in the Markov Chain, it seems natural to consider the SCE-UA global optimization strategy developed by *Duan et al.* [1992]. Recently, *Thyer et al.* [1999] examined the use of simulated annealing, a probabilistic optimization technique, which is intimately related to the MH algorithm, in combination with a Simplex downhill search method (SA-SX) to calibrate parameters in a conceptual catchment model. Although the stochastic nature of their SA-SX algorithm avoided getting trapped in local minima in the parameter space, the SCE-UA algorithm, which operates with a population of points divided into subcomplexes spread out over the feasible parameter space, was found to be more effective in searching the parameter space, especially when the dimension of the problem was increased. The SCE-UA approach has the desirable characteristic that it explicitly uses information about the nature of the response surface, extracted using the deterministic Simplex geometric shape, to direct the search into regions with higher posterior probability. Moreover, periodic shuffling of the population enhances survivability and performance by a sharing of the information gained independently by each community.

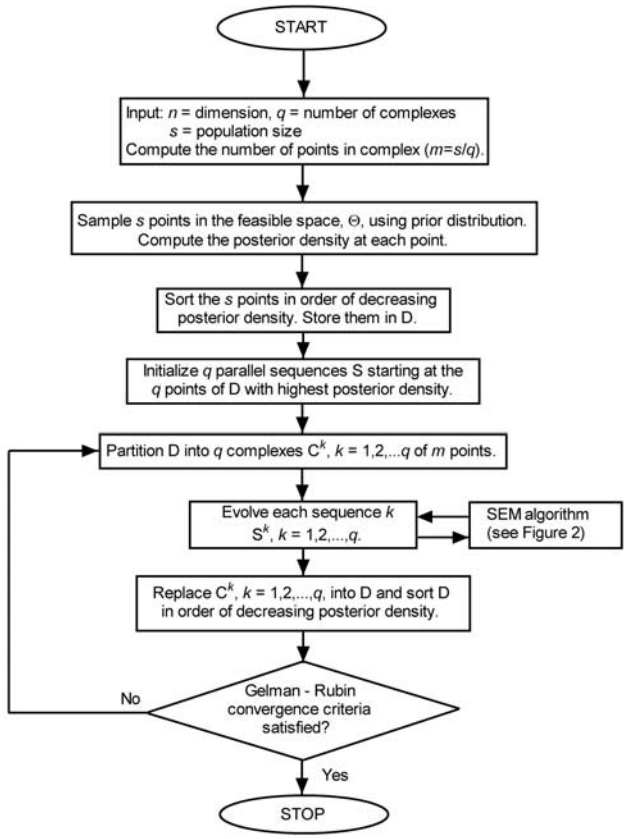

Figure 1. Flowchart of the Shuffled Complex Evolution Metropolis (SCEM-UA) algorithm.

2.3. The Shuffled Complex Evolution Metropolis Algorithm

[25] The goal of the SCE-UA algorithm developed by *Duan et al.* [1992] is to find a single best parameter set in the feasible space. Therefore it continuously evolves the population toward better solutions in the search space, relinquishing occupation of the regions of the parameter space with lower posterior density. This genetic drift, where the majority of the population converges toward a single mode, is typical of many evolutionary search algorithms. To prevent the collapse of the algorithm into the relatively small region of a single best parameter set, we have modified the SCE-UA algorithm. The new algorithm, entitled the Shuffled Complex Evolution Metropolis (SCEM-UA; developed in collaboration between the University of Amsterdam and the University of Arizona) algorithm is given below and is illustrated in Figures 1 and 2.

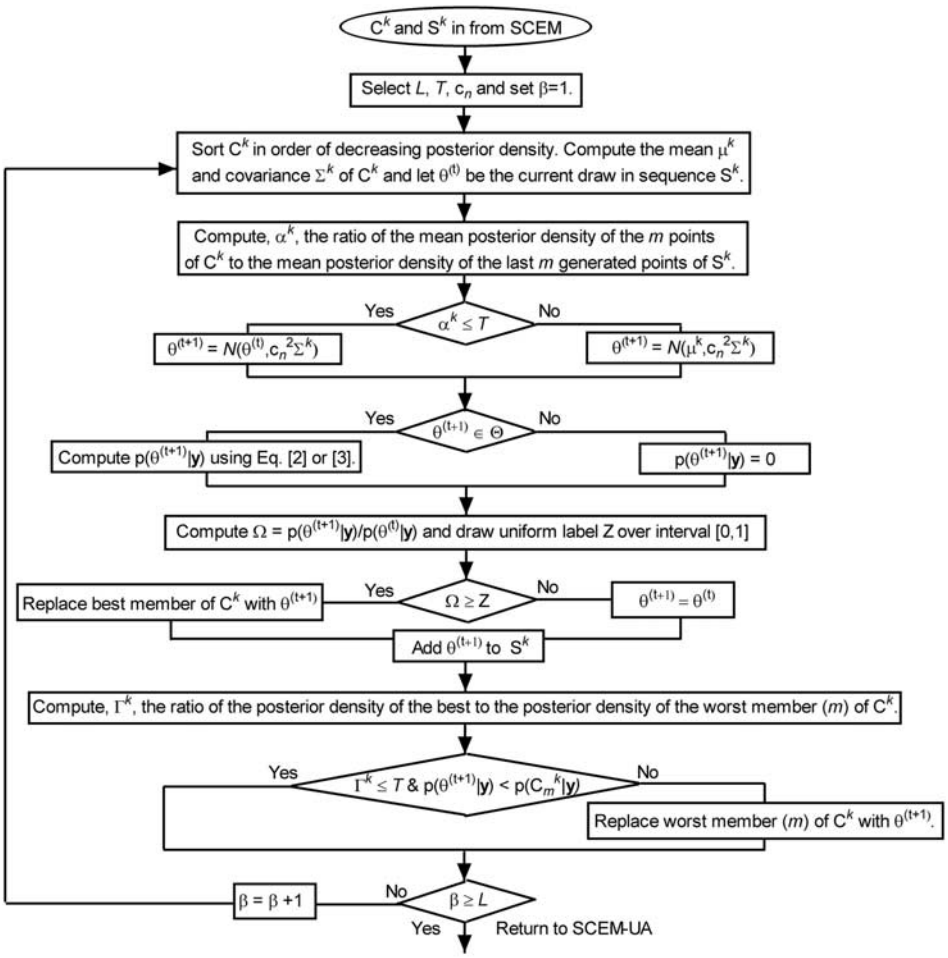

Figure 2. Flowchart of the (SEM) strategy employed in the SCEM-UA algorithm.

[26] 1. Generate sample. Sample s points $\{\theta_1, \theta_2, \ldots, \theta_s\}$ randomly from the prior distribution and compute the posterior density $\{p(\theta^{(1)}|\mathbf{y}), p(\theta^{(2)}|\mathbf{y}), \ldots, p(\theta^{(s)}|\mathbf{y})\}$ of each point using equation (2) or (3).

[27] 2. Rank points. Sort the s points in order of decreasing posterior density and store them in array D[1:s, 1:n + 1], where n is the number of parameters, so that the first row of D represents the point with the highest posterior density.

[28] 3. Initialize parallel sequences. Initialize the starting points of the parallel sequences, $S^1, S^2, \ldots, S^q$, such that S^k is D[k, 1:n + 1], where $k = 1, 2, \ldots, q$.

[29] 4. Partition into complexes. Partition the s points of D into q complexes $C^1, C^2, \ldots, C^q$, each containing m points, such that the first complex contains every $q(j-1)+1$ ranked point, the second complex contains every $q(j-1)+2$ ranked point of D, and so on, where $j = 1, 2, \ldots, m$.

[30] 5. Evolve each sequence. Evolve each of the parallel sequences according to the Sequence Evolution Metropolis algorithm outlined below.

[31] 6. Shuffle complexes. Unpack all complexes C back into D, rank the points in order of decreasing posterior density, and reshuffle the s points into q complexes according to the procedure specified in step 4.

[32] 7. Check convergence. Check the Gelman and Rubin (GR) convergence statistic. If convergence criteria are satisfied, stop; otherwise return to step 5. The definition of the GR convergence statistic appears in section 2.6.

[33] The use of a large initial random sample provides an extensive exploration of the parameter space, thereby increasing the chance of finding the global optimum of the prescribed density function. The use of a number of parallel sequences with different starting points enables an independent exploration of the search space, thereby allowing that the optimization problem has more than one region of attraction, and enables the use of heuristic tests to judge whether convergence of the sequences to a limiting distribution has been achieved. The use of complexes enables the collection of information gained about the search space by each individual sequence during the evolution process. The shuffling of these complexes enhances the survivability of the sequences by a global sharing of the information gained independently by each parallel sequence. This series of operations results in a robust MCMC sampler that conducts a robust and efficient search of the parameter space.

[34] One of the key components of the SCEM-UA algorithm is the Sequence Evolution Metropolis (SEM) algorithm, as mentioned in step 5. This algorithm produces new candidate points in each of the parallel sequences S^k by generating draws from an adaptive proposal distribution by using the information induced in the m samples of C^k. An

outline of the SEM algorithm is given below (see also Figure 2).

[35] I. Compute the mean μ^k, and covariance structure $\sum^k$ of the parameters of C^k. Sort the m point in complex C^k in order of decreasing posterior density and compute ^{k}, the ratio of the posterior density of the first (''best'') to the posterior density of the last (''worst'') member of C^k.

[36] II. Compute α^k, the ratio of the mean posterior density of the m points in C^k to the mean posterior density of the last m generated points in S^k.

[37] III. If α^k is smaller than a predefined likelihood ratio, T, generate a candidate point, $\theta^{(t+1)}$, by using a multinormal distribution centered on the last draw, $\theta^{(t)}$, of the sequence S^k, and covariance structure $c_n{}^2\sum^k$, where c_n is a predefined jumprate. Go to step V, otherwise continue with step IV.

[38] IV. Generate offspring, $\theta^{(t+1)}$, by using a multinormal distribution with mean μ^k and covariance structure $c_n{}^2\sum^k$, and go to step V.

[39] V. Compute the posterior density, $p(\theta^{(t+1)}|\mathbf{y})$, of $\theta^{(t+1)}$ using equation (2) or (3). If the generated candidate point is outside the feasible parameter space, set $p(\theta^{(t+1)}|\mathbf{y})$ to zero.

[40] VI. Compute the ratio $\Omega = p(\theta^{(t+1)}|\mathbf{y})/p(\theta^{(t)}|\mathbf{y})$ and randomly sample a uniform label Z over the interval 0 to 1.

[41] VII. If Z is smaller than or identical to Ω, then accept the new candidate point. However, if Z is larger than Ω, reject the candidate point and remain at the current position in the sequence, that is, $\theta^{(t+1)} = \theta^{(t)}$.

[42] VIII. Add the point $\theta^{(t+1)}$ to the sequence S^k.

[43] IX. If the candidate point is accepted, replace the best member of C^k with $\theta^{(t+1)}$, and go to step X; otherwise replace the worst member (m) of C^k with $\theta^{(t+1)}$, provided that Γ^k is larger than the predefined likelihood ratio, T, and $p(\theta^{(t+1)}|\mathbf{y})$ is higher than the posterior density of the worst member of C^k.

[44] X. Repeat the steps I–VIII L times, where L is the number of evolution steps taken by each sequence before complexes are shuffled.

[45] In the SEM algorithm, candidate points are generated using an adaptive multinormal proposal distribution with mean identical to the current draw in the sequence and covariance matrix corresponding to the structure induced in the m points of complex k. However, in situations where the mean posterior density of the last m generated points in sequence k is significantly smaller than the mean posterior density of the m points in the corresponding complex k, the center of the proposal distribution is temporarily switched to the mean of the points in the complex. This particular feature in the SEM algorithm (step IV) significantly reduces the chance that individual sequences get stuck in a local non-productive region of attraction, thereby further improving the mixing of the sequences. After generating a new candidate point, the posterior density relevant to this point is computed and the Metropolis-annealing [*Metropolis et al.*, 1953] criterion is used to judge whether the candidate point should be added to the current sequence or not. Finally, the last step (step IX) in the SEM algorithm considers which member of the current complex k should be replaced with the point $\theta^{(t+1)}$. When the candidate point is accepted, $\theta^{(t+1)}$ automatically replaces the best member of the complex. However, when the candidate point is rejected, $\theta^{(t+1)}$ replaces the worst point in complex k provided that Γ^k is larger than the predefined likelihood ratio, T, and the posterior density relevant to $\theta^{(t+1)}$ is higher than the posterior density corresponding to the worst point of complex k. Hence, when Γ^k is larger than some prior defined large number (i.e., $T > 10^5$), there is sufficient reason to believe that the covariance of the proposal distribution is specified too big as members with a too low probability are still present in C^k. Replacement of the worst member of C^k in this particular situation will facilitate convergence to a limiting distribution.

[46] In contrast with traditional MCMC samplers, the SCEM-UA algorithm is an adaptive sampler, where the covariance of the proposal or sampling distribution is periodically updated in each complex during the evolution to the posterior target distribution using information from the sampling history induced in the transitions of the generated sequences. Of course, it is not clear from the algorithm presented above whether the proposed procedure for updating the proposal distribution in view of the past history of the chains will result in an ergodic chain with desirable asymptotic properties of the kernel density estimates derived from the sampler [*Haario et al.*, 1999, 2001]. However, an empirical (experimental) investigation of the ergodicity of the SCEM-UA strategy has revealed that the algorithm performs very well, as demonstrated by different case studies presented in this paper.

[47] The SCEM-UA algorithm is different from the original SCE-UA algorithm presented by *Duan et al.* [1992] in two important ways. Both modifications are necessary to prevent the search from becoming mired in a small basin of attraction and thus to arrive at the correct posterior target distribution. First, the downhill Simplex method in the competitive complex evolution algorithm outlined by *Duan et al.* [1992] is replaced by a Metropolis-annealing covariance based offspring approach, thereby avoiding a deterministic drift toward a single mode. Second, the SCEM-UA algorithm does not further subdivide the complex into subcomplexes during the generation of the offspring (candidate points) and uses a different replacement procedure, to counter any tendency of the search to terminate occupations in the lower posterior density region of the parameter space.

2.4. Selection of Algorithmic Parameters in the SCEM-UA Algorithm

[48] The SCEM-UA algorithm contains two algorithmic parameters that need to be specified by the user. These are (1) the number of complexes/sequences, q, and (2) the population size, s, which in turn also determine the number of points within each complex ($m = s/q$). For simple problems with an uncorrelated or correlated Gaussian target distribution, relatively small population sizes ($s \leq 100$) and a small number of generated sequences/complexes ($q \leq 5$) will usually suffice. However, in the case of complex-shaped posterior probability density distributions, like highly nonlinear banana-shaped distributions, we recommend the use of larger population sizes ($s \geq 250$) and a larger number of parallel sequences ($q \geq 10$) to be able to precisely capture the complex shape of the covariance structure. Specific information about the

number of sequences/complexes and the population size can be found in the case study section of this paper. Additionally, the SEM algorithm contains three algorithmic parameters the values of which values must be chosen carefully. The SCEM-UA algorithm employed for the different case studies reported in this paper used the values $L = (m/10)$ and $T = 10^6$. As a basic choice, the value of the jump rate, c_n, was set to $2.4/\sqrt{n}$ [*Gelman et al.*, 1995]. Preliminary sensitivity analyses of the SCEM-UA algorithm indicated that these values for the algorithmic parameters work well for a broad range of applications.

2.5. Comparison of SCEM-UA Algorithm Against Traditional MH Samplers

[49] To enable a direct comparison in performance between the Metropolis-Hastings and SCEM-UA algorithms, the q parallel sequences in the MH sampler were initialized using the q points that exhibited the highest posterior density in the original s points of the population. Moreover, the proposal distribution for the traditional MH sampler was set identical to the covariance structure of the random initialized population of points in the feasible parameter space. We argue that this is fair, considering the fact that a uniform prior over the predefined feasible parameter space is usually the only information we have about the location of the high-posterior probability density region. To benchmark against more modern MCMC techniques, case study 2 also contains the results for other stable state-of-the-art MH samplers, which are known to maintain ergodicity.

2.6. Convergence of MCMC Samplers

[50] An important issue in MCMC sampling is convergence of the sampler to the stationary posterior distribution (step 7 of the SCEM-UA algorithm outlined in Figure 1). Theoretically, a homogeneous sampler converges in the limit as $t \rightarrow \infty$, but in any applied problem one must determine how many draws to make with the sampler. *Gelman and Rubin* [1992] developed a quantitative convergence diagnostic, $\sqrt{SR}$, which they call the scale reduction score, based on the within and between chain (sequence) variances:

$$\sqrt{SR} = \sqrt{\frac{g-1}{g} + \frac{q+1}{q \cdot g}\frac{B}{W}} \tag{6}$$

where g is the number of iterations within each sequence, B is the variance between the q sequence means, and W is the average of the q within-sequence variances for the parameter under consideration respectively. Note that the product of q and g is identical to the total number of draws, t, with the MH- or SCEM-UA sampler. A score close to 1 for $\sqrt{SR}$ for each of the parameters indicates convergence. However, because a score of unity is difficult to achieve, *Gelman and Rubin* [1992] recommend using values less than 1.2 to declare convergence to a stationary distribution.

3. Case Studies

[51] We compare the power and applicability of the Metropolis-Hastings and Shuffled Complex Evolution Metropolis algorithms for three case studies with increasing complexity. The first is a synthetic study using a simple one-parameter bimodal posterior probability distribution. This illustrates the ability of both search algorithms to infer the known posterior target distribution. The second case study explores the effectiveness and efficiency of the MH and SCEM-UA samplers for approximating a strongly nonlinear banana-shaped posterior probability distribution and investigates the ergodicity of the SCEM-UA sampler. Finally, the third case study involves assessment of parameter uncertainty using a five-parameter conceptual rainfall-runoff model. In case studies 2 and 3, we are especially concerned with algorithm efficiency, particularly the number of simulations needed to converge to the stationary posterior distribution.

3.1. Case Study I: A Simple Bimodal Probability Distribution

[52] We investigate the applicability of the Metropolis-Hastings algorithm and the Shuffled Complex Evolution Metropolis algorithm for assessment of parameter uncertainty in the presence of bimodality. Consider the following bimodal probability density function,

$$p(\theta) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\theta^2\right] \frac{2}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(2\theta - 8)^2\right] \tag{7}$$

which is the sum of two Gaussian probability distributions, each having a different well-defined optimum as demonstrated in Figure 3.

[53] The plot shows the bimodal character of the posterior distribution, with a high probability region in the area corresponding to the most probable parameter value ($\theta = 4$) and another high-density region at ($\theta = 0$) well separated from the first cluster.

[54] Figure 3 presents a histogram of 4000 samples generated using the Metropolis-Hastings and Shuffled Complex Evolution Metropolis algorithm outlined in section 2. Both algorithms generated five parallel sequences ($q = 5$), each with 1000 samples using a population size of 50 points. The first 200 samples of each sequence were discarded, because it is unlikely that the initial draws come from the stationary distribution needed to construct the posterior estimates. An acceptable scale reduction score of less than 1.2 indicated approximate convergence. For both algorithms, it is evident that they are able to successfully infer the posterior target distribution defined in equation (7).

3.2. Case Study II: A Two-Dimensional Banana-Shaped Posterior Target Distribution

[55] This case study explores the effectiveness and efficiency of the MH and SCEM-UA algorithms for inferring a two-dimensional highly nonlinear banana-shaped posterior target distribution. In this study, we are especially concerned with the ergodic properties of the SCEM-UA sampler. If the proposed SCEM-UA algorithm is able to generate a useful approximation of the highly complex nonlinear banana-shaped test distribution in this study, it seems reasonable to conjecture that the sampler is suited to construct accurate posterior estimates for the parameters in hydrologic models.

[56] The nonlinear banana-shaped distribution is constructed from the standard multivariate Gaussian distribution as follows. Let f be the density of the multivariate normal distribution, $N(0, \sum)$ with the covariance matrix

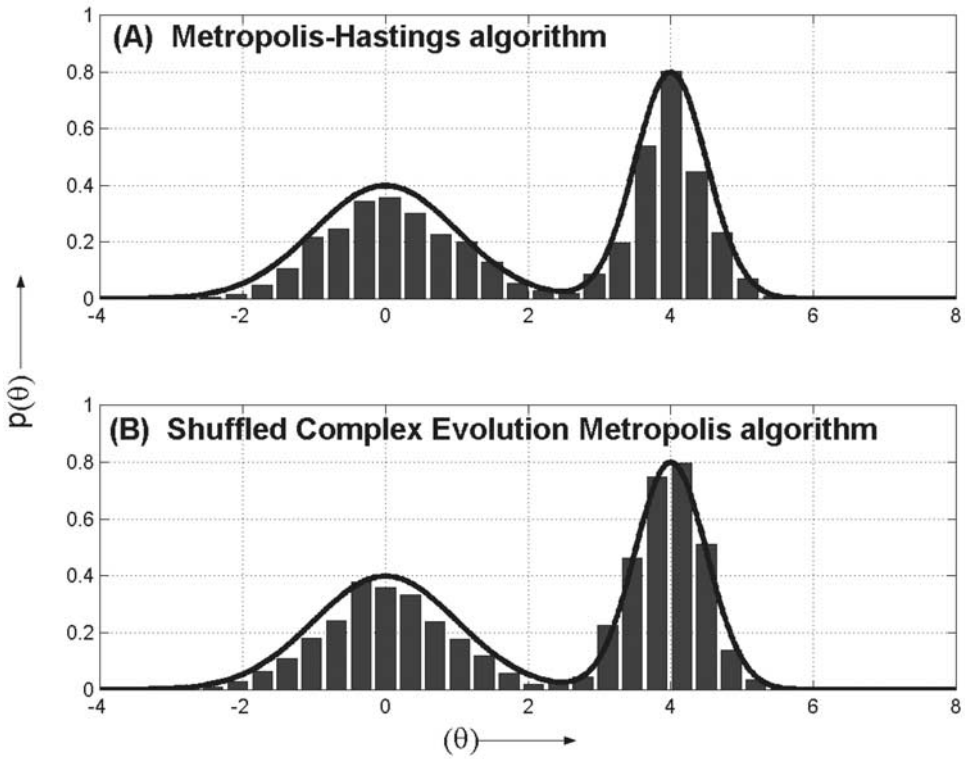

Figure 3. Bimodal probability distribution and histogram of 4000 samples generated using (a) the Metropolis and (b) the Shuffled Complex Evolution Metropolis algorithms. See color version of this figure at back of this issue.

given by $\sum$ = diag(100, 1, …, 1). The twisted Gaussian density function with nonlinearity parameter $b > 0$ is given by

$$f_b = f \circ \phi_b \tag{8}$$

where the function ϕ_b is

$$\phi_b(\theta) = \left(\theta_1, \theta_2 + b\theta_1^2 - 100b, \theta_3, \ldots, \theta_n\right) \tag{9}$$

The nonlinearity of function ϕ_b increases with b. In our test, we applied the value $b = 0.1$ to yield a strongly twisted banana-shaped target distribution. Owing to the complexity of the posterior density surface of this test distribution, the population size in the SCEM-UA algorithm was set to 1000 points, and the number of parallel sequences was set to 10. The feasible parameter space was taken to be a uniform distribution between −100 and 100 for each parameter. The test cases reported in this paper have been performed in the dimensions 2 and 8.

[57] In the two-dimensional case, the evolution of the Gelman-Rubin convergence diagnostic for the parameters θ_1 and θ_2 using the MH and SCEM-UA algorithms is illustrated in Figures 4a and 4b, respectively.

[58] Owing to random initializations of the starting points of the parallel sequences in the feasible parameters space, the scale reduction factor for the first 10,000 simulations is quite large using the MH sampler (Figure 4a). Thereafter, the convergence diagnostic for both parameters narrows down quickly and continues to widen and narrow

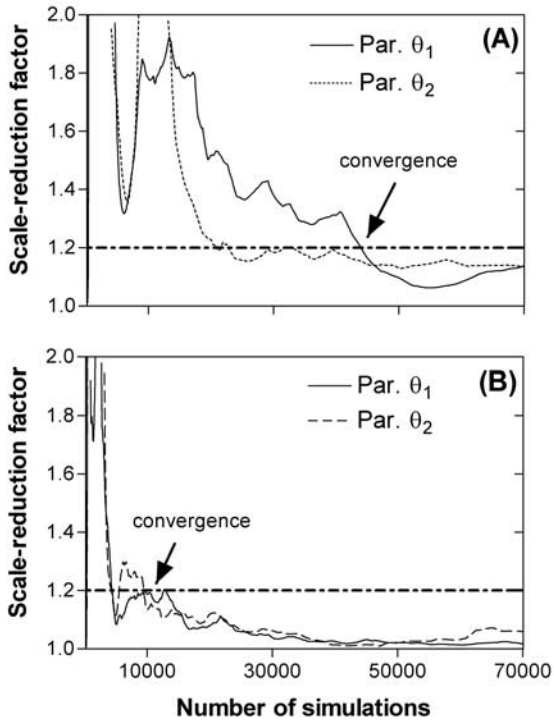

Figure 4. Evolution of the Gelman and Rubin scale-reduction factor for the nonlinear banana-shaped posterior test distribution, using (a) the MH sampler and (b) the SCEM-UA algorithm.

Table 1. Statistical Properties of the Two-Dimensional Banana-Shaped Posterior Distribution as a Function of the Number of Function Evaluations With the MH and SCEM-UA Samplers[a]

Number of Evaluations	MH		SCEM-UA	
	Mean(‖E‖)	SD(‖E‖)	Mean(‖E‖)	SD(‖E‖)
10,000	7.85	9.26	1.59	3.47
20,000	2.54	4.32	1.00	1.61
30,000	2.94	3.99	1.05	1.10
40,000	2.51	2.72	0.84	0.84
50,000	2.28	2.92	0.96	0.84
60,000	1.84	2.68	0.95	0.72
70,000	1.58	2.36	0.87	0.63
80,000	1.68	2.20	0.81	0.61

[a]Mean(‖E‖) denotes the average Euclidean distance of the SCEM-UA derived mean values from their true values at origo; thus mean ($\|E\|$) = $(\Sigma_{i=1}^{n}(E_t)^2)^{1/2}$. In a similar way, SD(‖E‖) denotes the average Euclidean distance of the SCEM-UA derived standard deviations from their true values at origo. To reduce the influence of sampling variability, the presented statistics denote averages over 100 independent runs.

intermittently. Finally, after approximately 45,000 simulations, the plot suggests that the parallel sequences have converged to a stationary posterior distribution for both parameters ($\sqrt{SR} < 1.2$). In the case of the SCEM-UA algorithm, the periodic updating of the covariance structure of the sampling-proposal distribution significantly improves the explorative capabilities of the sampler and the mixing of the sequences. Consequently, far fewer iterations are needed with the SCEM-UA algorithm than with the traditional MH sampler to achieve convergence to a stationary posterior distribution. Indeed, the evolution of the convergence diagnostic depicted in Figure 4b demonstrates convergence to a posterior distribution after approximately 10,000 simulations.

[59] Although the results in Figure 4 demonstrate a faster convergence rate of the SCEM-UA sampler over the traditional MH sampler, it is important to check whether the former sampler has converged to the true prior defined two-dimensional banana-shaped posterior test distribution and hence has the right ergodic properties. Table 1 summarizes the evolution of the average Euclidean distance of the "true" means and standard deviations of the prior defined probability distribution from those respective values estimated with the SCEM-UA algorithm as a function of the number of function evaluations. Each number in the table denotes an average over 100 independent runs. The results in Table 1 demonstrate that the average Euclidean distance of the SCEM-UA estimated mean values and standard deviations from the true values at origo slowly approaches to zero, suggesting that the sampler provides a correct estimation of the posterior target distribution and hence does not collapse to a small region of highest posterior density. More significantly, the posterior moments derived with the adaptive SCEM-UA sampler compare favorably well with identical counterparts derived using the MH sampler. This serves as numerical evidence that the SCEM-UA algorithm has the right ergodic properties and hence provides correct simulation of the target distribution.

[60] Note, however, that both samplers generate slightly biased estimates of the posterior moments as compared to their "true" values of the predefined banana-shaped test distribution. This is also demonstrated in Figures 5a and 5b, which present scatterplots of the (θ_1, θ_2) sampled MH and SCEM-UA points, respectively, that were generated after convergence of the sequences to a stationary posterior distribution has been achieved.

[61] The dark contour line refers to the theoretical 68.3 and 95% confidence regions of the predefined banana-

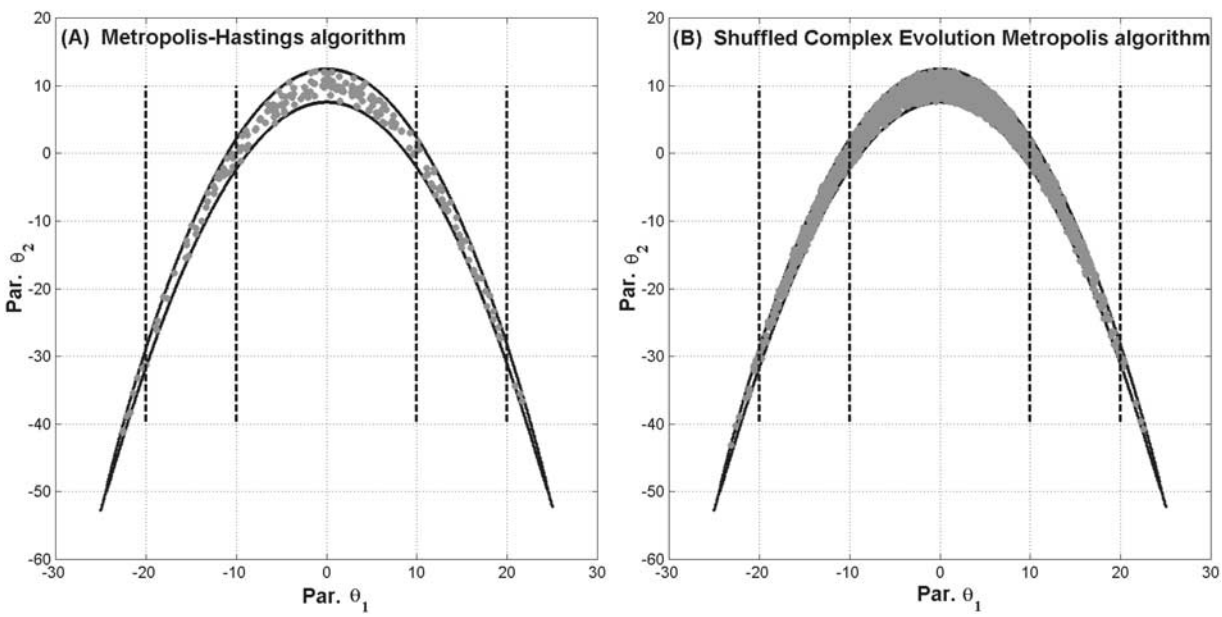

Figure 5. A scatterplot of the (θ_1,θ_2) samples generated after convergence to a stationary posterior distribution has been achieved using (a) the Metropolis, and (b) the Shuffled Complex Evolution Metropolis algorithms. The lines indicate the one-dimensional 68.3 and 95% confidence regions of the parameters. See color version of this figure at back of this issue.

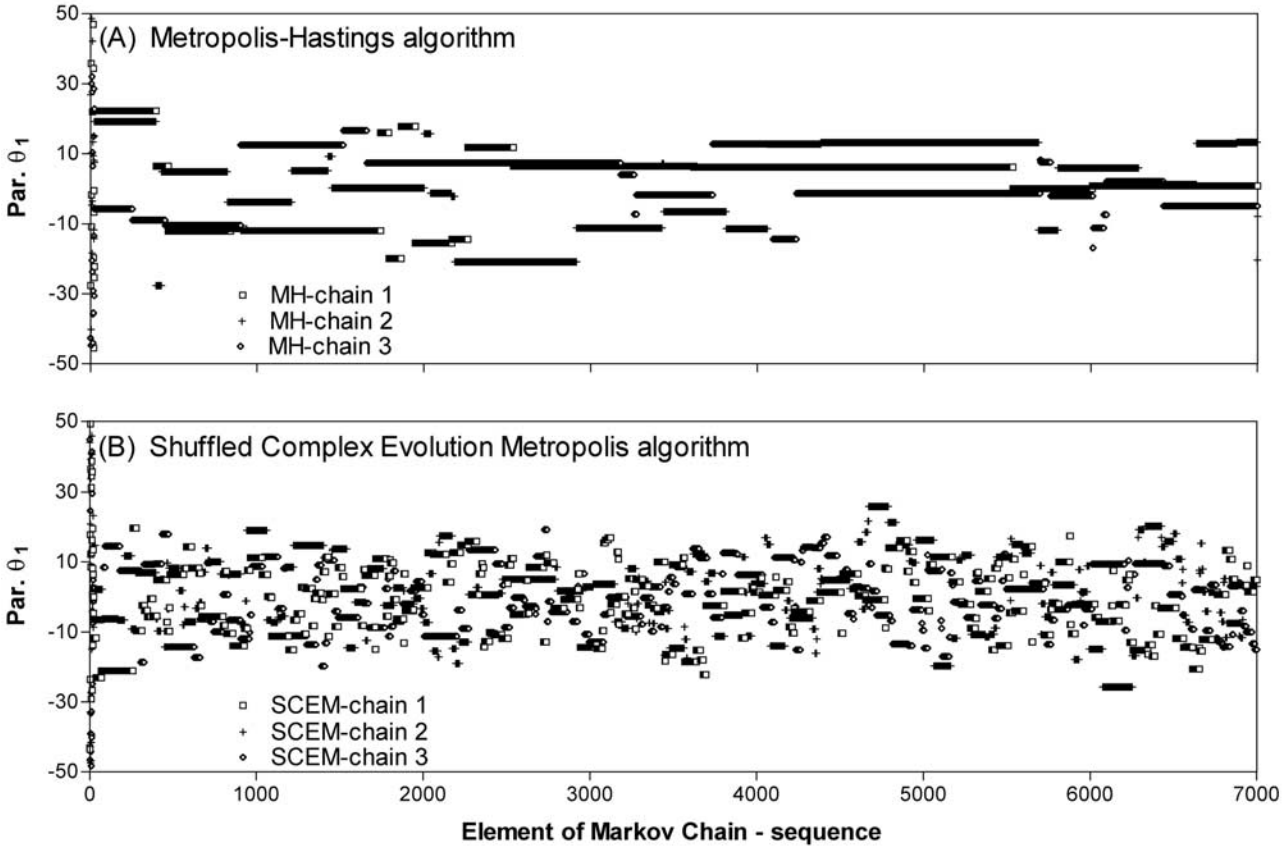

Figure 6. Transitions of the parameter θ_1 in three of the five parallel generated Markov Chains during the evolution to the banana-shaped posterior target distribution using the (a) the MH algorithm and (b) the SCEM-UA algorithm. For more explanation, please refer to the text.

shaped posterior target distribution. Both the MH and SCEM-UA sampled points and posterior moments are consistent with the test target distribution. The scatterplots presented in Figures 5a and 5b demonstrate that with an identical number of iterations, the population sampled using the SCEM-UA algorithm is much more diverse than the population sampled using the traditional MH sampler. However, especially in the case of the MH sampler, the sampling density at the extreme tails of the banana-shaped distribution is rather sparse, suggesting that the algorithms experience problems in exploring the lower posterior density region of the parameter space. This explains the slightly biased posterior moments presented in Table 1. Clearly, periodically updating of the covariance structure of the proposal distribution in view of the history induced in the transitions of the Markov Chains not only improves the convergence rate of the sampler but also significantly increases the diversity of the final sampled population. Hence a more diverse population yields better estimates of the final statistical moments of the posterior distribution and as such is an additional advantage of the SCEM-UA algorithm over traditional MH-samplers.

[62] The transitions of parameter θ_1 in three of the five parallel sequences (Markov Chains) during the evolution of the MH and SCEM-UA samplers to the stationary posterior distribution is illustrated in Figures 6a and 6b, respectively.

[63] For clarity, the three different parallel sequences are coded with different symbols. The 1-D scatterplots of the sampled parameter space demonstrate that at early stages during the evolution, the individual sequences tend to occupy different regions of the posterior surface. This low mixing of the paths, especially in the case of the MH sampler (Figure 6a), is associated with a relatively high value for the scale reduction factor (see Figure 4), indicating poor convergence. At a later stage during the evolution, all of the individual sequences have been able to fully explore the banana-shaped posterior target distribution, thereby resulting in a scale reduction factor smaller than 1.2, indicating convergence to a stationary distribution. Note, however, that the mixing of the MH-generated Markov Chains is quite poor, suggesting that the proposal distribution used to sample with the MH sampler was too large. Hence the transitions in the Markov Chain reveal a low diversity; too many candidate points are rejected and therefore the chain slowly covers the posterior target distribution. Graphical examination of the transitions and mixing of the different symbolic coded paths yields a similar picture about the convergence status of the sampler as the value of the convergence diagnostic presented in Figure 4. Practical experience with other case studies also suggests that the GR convergence diagnostic is useful for testing whether convergence to a limiting posterior distribution has been achieved even when the parallel chains/sequences are not fully independent, as in the case of the SCEM-UA sampler. Although the SCEM-UA algorithm is an adaptive sampler, which continuously updates the proposal distribution based

Table 2. Statistical Properties of the Eight-Dimensional Banana-Shaped Posterior Distribution as Estimated With the SCEM-UA Algorithm[a]

Method	AP	AM	SCEM-UA
Mean(‖E‖)	4.85	2.41	1.78
SD(‖E‖)	4.20	1.15	1.42

[a]Mean(‖E‖) and SD(‖E‖) denote the average distance of the SCEM-UA derived mean and standard deviations from their true values at origo calculated over 100 repetitions. For more explanation, please refer to the text.

on the information induced in the history of the sampled points, the sampler does not collapse to a single region of highest attraction. Graphical examination of the sampled parameter space demonstrates that the SCEM-UA sampler maintains occupation at the extreme tails of the hyperbolic banana-shaped distribution during the evolution. This ensures an asymptotic convergence to the desirable kernel density estimates and again serves as empirical evidence that the generated Markov Chain-sequences are ergodic.

[64] To test whether the SCEM-UA algorithm also provides a correction simulation of the target distribution in higher dimensions, the algorithm was run in various dimensions up to $n = 500$. The results of extensive tests in dimension $n = 8$ are summarized in Table 2. To benchmark against other modern adaptive MH samplers, the results of the adaptive proposal (AP) and adaptive Metropolis (AM) algorithms, as developed by *Haario et al.* [1999, 2001], are also included. These algorithms are known to maintain ergodicity. The results in Table 2 demonstrate that the SCEM-UA algorithm also generates a correct simulation of the banana shaped posterior target distribution in the eight-dimensional test case. The SCEM-UA algorithm has a consistent better performance than the AP algorithm and a quite similar performance as the AM algorithm. However, the exploration of the parameter space with a number of parallel sequences, rather than a single sequence search strategy as employed in the AP and AM algorithms, enables the SCEM-UA algorithm to deal with optimization problems that contain more than one region of attraction. Additionally, the shuffling procedure implemented in the SCEM-UA algorithm enhances the survivability F of the sequences by a global sharing of the information gained independently by each parallel sequence. Consequently, as opposed to the AP and AM algorithms developed by *Haario et al.* [1999, 2001] the SCEM-UA algorithm does not require an optimization technique to first locate the high probability density region in the parameter space. Although not explicitly demonstrated, we have successfully applied the SCEM-UA algorithm up to $n = 500$ dimensions.

[65] The empirical results presented here illustrate three important findings. First, in the case of a strong banana-shaped curvature of the posterior distribution of the parameters, the SCEM-UA algorithm is successfully able to infer the posterior target distribution. Second, graphical examination of the sampled parameter space and numerical analyses of the statistical properties of the Markov Chain demonstrate that the generated sequences are ergodic, thereby ensuring asymptotic convergence to the desirable kernel density estimates of the posterior target distribution. These results give confidence that the SCEM-UA algorithm will likely yield accurate estimates of the posterior moments for complex problems such as are usually found in hydrologic modeling. Third, periodic tuning of the sampling-proposal distribution by use of a local search direction (sequences), constructed using global information exchange (shuffling), can help make improvements in mixing and therefore in the efficiency (speed of convergence) of MCMC samplers.

3.3. Case Study III: The HYMOD Model

[66] This case study illustrates the usefulness of the Metropolis-Hastings and shuffled complex evolution Metropolis algorithms to hydrologists who are especially concerned with a realistic assessment of prediction uncertainty on hydrological responses. For this purpose, we used HYMOD, a five-parameter conceptual rainfall-runoff model (see Figure 7), introduced by *Boyle* [2000] and recently used by *Wagener et al.* [2001].

[67] The HYMOD model consists of a relatively simple rainfall excess model, described in detail by *Moore* [1985], connected with two series of linear reservoirs (three identical quick and a single reservoir for the slow response) and requires the optimization of five parameters to observed streamflow data: the maximum storage capacity in the catchment, C_{max} (L), the degree of spatial variability of the soil moisture capacity within the catchment, b_{exp} ( ), the factor distributing the flow between the two series of reservoirs, Alpha ( ), and the residence time of the linear quick and slow reservoirs, R_q (T) and R_s (T), respectively.

[68] In keeping with previous studies [e.g., *Thiemann et al.*, 2001], approximately 11 years (28 July 1952 to

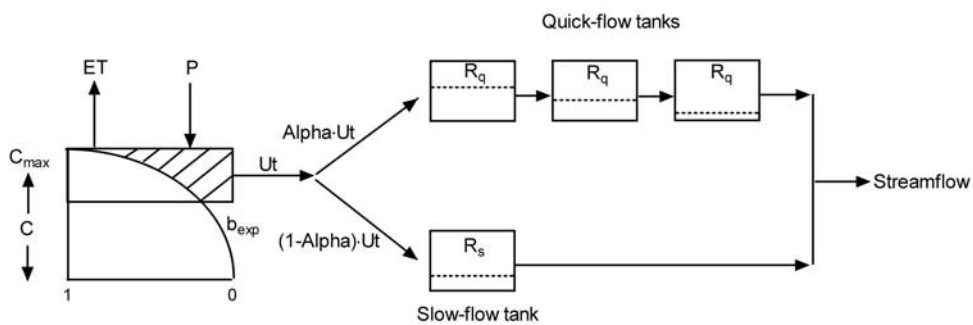

Figure 7. Schematic representation of the HYMOD model.

Table 3. Prior Uncertainty Ranges of the HYMOD Model Parameters

	Minimum	Maximum	Unit
C_{max}	1.000	500.000	mm
b_{exp}	0.100	2.000	[ ]
Alpha	0.100	0.990	[ ]
R_s	0.000	0.100	[d]
R_q	0.100	0.990	[d]

30 September 1962) of hydrologic data from the Leaf River watershed were used for model calibration. This humid 1944 km^2 watershed, located north of Collins, Mississippi, has been investigated intensively [e.g., *Brazil*, 1988; *Sorooshian et al.*, 1993; *Boyle et al.*, 2000; *Thiemann et al.*, 2001]. The data, obtained from the National Weather Service Hydrology Laboratory (HL), consist of mean areal precipitation (mm/d), potential evapotranspiration (mm/d), and streamflow (m^3/s). This data set was used to test the efficiency and effectiveness of the MH and SCEM-UA algorithms for estimating the posterior distribution of the model parameters and to assess prediction uncertainty on the hydrologic responses. The prior uncertainty ranges of the parameters are defined in Table 3. To reduce sensitivity to state value initialization, a 65-day warm-up period was used, during which no updating of the posterior density defined in equation (3) was performed. Moreover, we used a population size of 250 points and assumed that the output errors have a heteroscedastic (nonconstant) variance that is related to flow level and which can be stabilized using the transformation, $z = [(y + 1)^{\lambda} - 1]/\lambda$ with $\lambda = 0.3$ [*Misirli et al.*, 2003].

[69] The evolution of the Gelman and Rubin scale-reduction convergence diagnostic for each of the model parameters using the MH and SCEM-UA algorithms is illustrated in Figures 8a and 8b.

[70] Both sampling algorithms generated five parallel sequences ($q = 5$), each with 6000 samples. Note the different scales on the y-axes of the two plots. Clearly, the SCEM-UA algorithm is more efficient in traversing the parameter space, with convergence to a stationary posterior distribution ($\sqrt{SR} < 1.2$) for each of the HYMOD model parameters achieved after approximately 4000 simulations. In contrast, the MH sampler is far from convergence to a stationary distribution, with the scale-reduction factor for each of the parameters larger than 2, even after performing 30,000 simulations. To understand why, consider Figures 9a and 9b, which present the evolution of samples generated in three parallel launched sequences, using either the MH or the SCEM-UA algorithm.

[71] Because no information about the location of the highest probability density region in the parameter space is exchanged between parallel sequences launched using the MH sampler, it remains difficult for the sequences (chains)

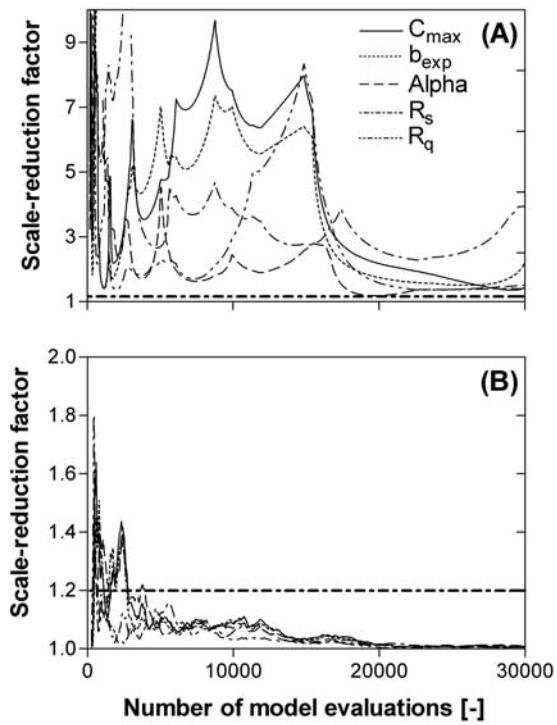

Figure 8. Evolution of the Gelman and Rubin scale-reduction factor for the parameters in the HYMOD model using 11 years of runoff data (1952–1962) for the Leaf River watershed for (a) the Metropolis-Hastings algorithm and (b) the Shuffled Complex Evolution Metropolis algorithm.

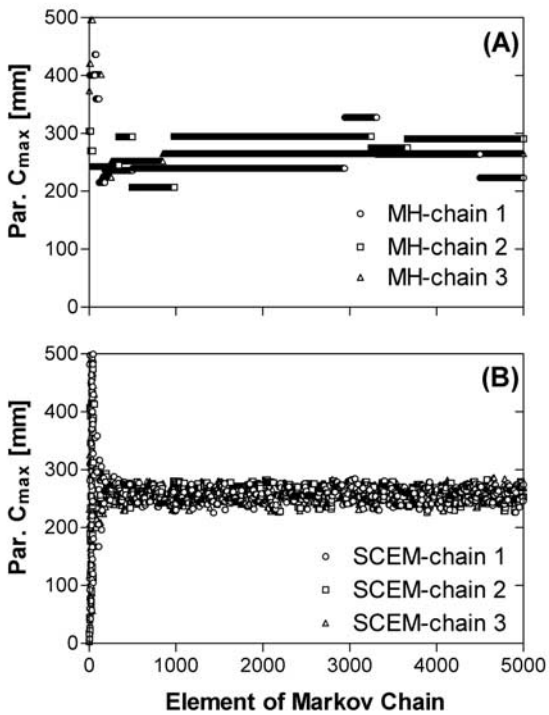

Figure 9. Markov Chain Monte Carlo C_{max} samples generated in three parallel sequences using either (a) the Metropolis-Hastings algorithm or (b) the Shuffled Complex Evolution Metropolis algorithm.

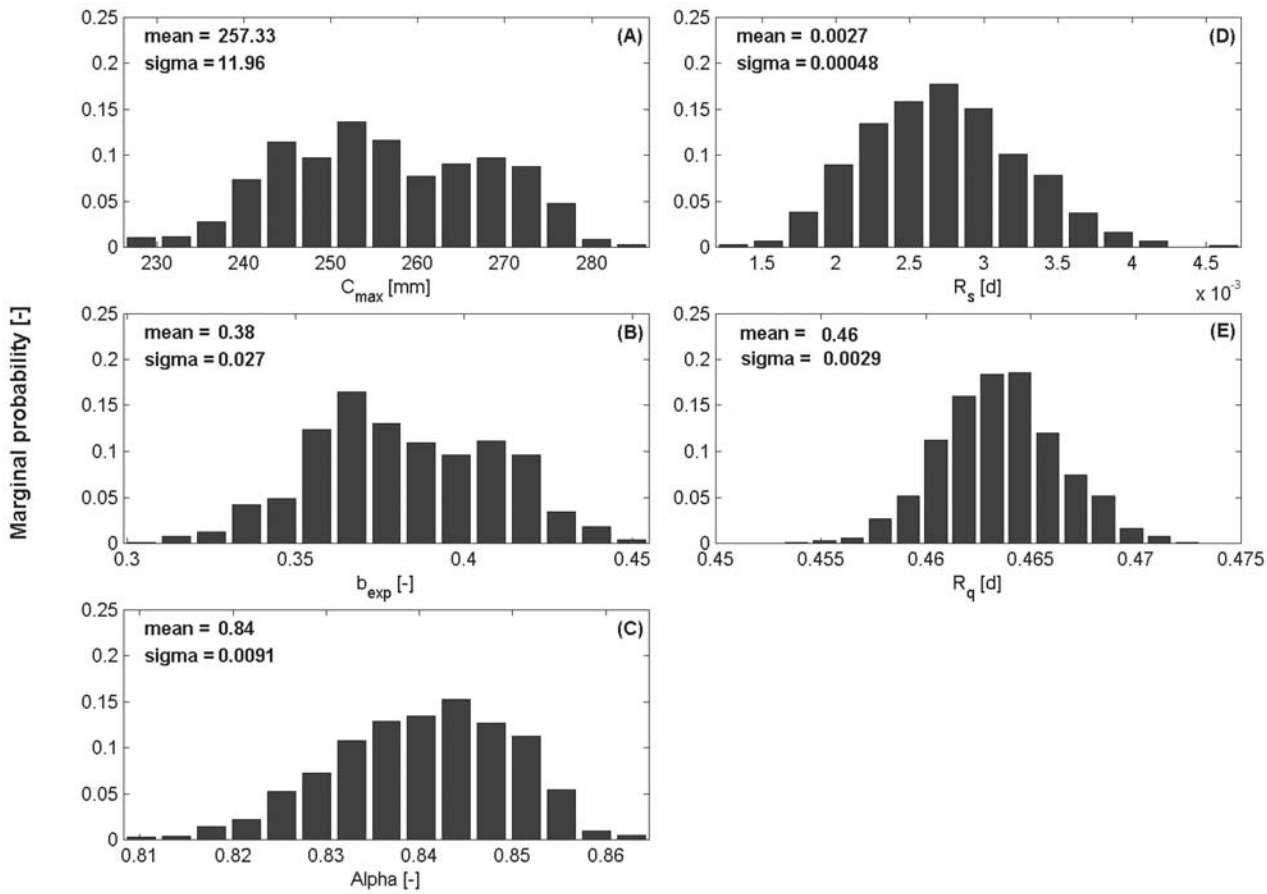

Figure 10. Marginal posterior probability distributions of the HYMOD model parameters (C_{max}, b_{exp}, Alpha, R_s, R_q) constructed using 10,000 samples generated after convergence to a posterior distribution has been achieved with the Shuffled Complex Evolution Metropolis algorithm.

to mix and terminate their occupation in regions of the parameter space with a low posterior density. Owing to this slow mixing rate, the posterior moments derived from the samples generated in each of the parallel chains differ appreciably between the different chains, and the scale-reduction factor remains far above the threshold value of 1.2, indicating lack of convergence. Given the ability of the SCEM-UA algorithm to exchange information about the search space gained by the different parallel launched sequences, this increases the explorative capabilities of the sampler and therefore the traversing speed of the chains through the feasible parameter space. This behavior is evident in Figure 9b. Consequently, the population sampled using the SCEM-UA algorithm is more diverse. Periodic shuffling of the complexes in the SCEM-UA algorithm ensures sharing of information gained independently by each community about the nature of the posterior distribution and therefore increases the traversal through the parameter space. This allows us to make more sound inferences about the nature of the posterior probability density function.

[72] Although not presented in this paper, we also performed a variety of experiments with the MH sampler to speed up convergence to the posterior distribution. One of those experiments was to periodically update the covariance structure of the jump-proposal distribution using a sample within each sequence, and to use this covariance structure to generate new candidate points in each parallel sequence. In this way, the information gained by each individual local sampler is more thoroughly exploited. However, using this approach, a significantly larger number of simulations were needed, as compared to the SCEM-UA algorithm, to achieve convergence (typically 10,000 simulations).

[73] Figure 10 presents the posterior marginal probability density distributions for each of the HYMOD model parameters inferred for the Leaf River watershed using the samples generated with the SCEM-UA algorithm.

[74] While the histograms of Alpha and the slow- and quick-tank recession parameters, R_s and R_q, exhibit an approximately Gaussian distribution, the posterior probability distributions for the other model parameters reveal the existence of several modes. This multimodality suggests the presence of multiple regions of attraction in the posterior surface and illustrates the severity of the optimization problem, even in the case of this simple and parsimonious five-parameter conceptual rainfall-runoff model using more than 10 years of daily streamflow data. Note, however, that the parameter sets sampled using the SCEM-UA algorithm

Table 4. Shuffled Complex Evolution Metropolis Posterior Mean (Mean), Standard Deviation (SD), Coefficient of Variation (CV), and Correlation Coefficients Between the Generated Samples for the HYMOD Model Parameters Using 11 Years of Runoff Data (1952–1962) for the Leaf River Watershed[a]

Parameter	Mean	SD	CV	C_{max}	b_{exp}	Alpha	R_s	R_q	SCEM-UA	SCE-UA
C_{max}	256.67	11.36	4.72	1.00	0.95	0.85	−0.44	−0.40	253.41	253.63
b_{exp}	0.38	0.026	7.24	...	1.00	0.85	−0.37	−0.38	0.38	0.38
Alpha	0.84	0.0091	1.18	...	...	1.00	−0.38	−0.45	0.84	0.84
R_s	0.0027	0.00051	19.44	...	...	...	1.00	0.27	0.0029	0.0030
R_q	0.46	0.0028	0.61	...	...	...	...	1.00	0.46	0.46

[a]Included is the parameter set found with highest posterior probability using the SCEM-UA algorithm and the original SCE-UA global optimization algorithm.

occupy a relatively small range, interior to their uniform prior distributions (e.g., Table 3), which indicates that the HYMOD model parameters are reasonable well identifiable from measured time series of runoff.

[75] In Table 4, we present the final posterior moments derived using those samples that were generated with the SCEM-UA algorithm after convergence to the stationary target distribution has been achieved. For this, the first 600 simulations of each parallel sequence were discarded (i.e., $\sqrt{SR} > 1.2$). Also included are the most likely parameter values, describing the approximately 11 years of runoff data of the Leaf River watershed, identified using the SCE-UA global optimization algorithm [*Duan et al.*, 1992]. As stated earlier, the posterior standard deviations and correlation coefficients between the sampled parameters depict that the parameters of the HYMOD model are well identifiable using measured runoff data. A direct comparison between the optimal parameter values derived using the original SCE-UA global optimization algorithm and the SCEM-UA algorithm demonstrates that the latter algorithm is not only able to conveniently derive the posterior distribution of the model parameters, but also successfully identifies the most likely parameter values within this high-density region. Clearly, this feature is an additional benefit of the SCEM-UA algorithm because it makes superfluous the two-step procedure in which the global optimum in the parameter space is first identified, followed by launching parallel MH samplers from this starting point to identify parameter uncertainty.

[76] Finally, Figure 11 illustrates how the results of the SCEM-UA algorithm can be translated into estimates of hydrograph prediction uncertainty, using data from Water Year 1953.

[77] Figure 11a shows the observed streamflows (dots), the 95% hydrograph prediction uncertainty associated only with the posterior distribution of the parameter estimates

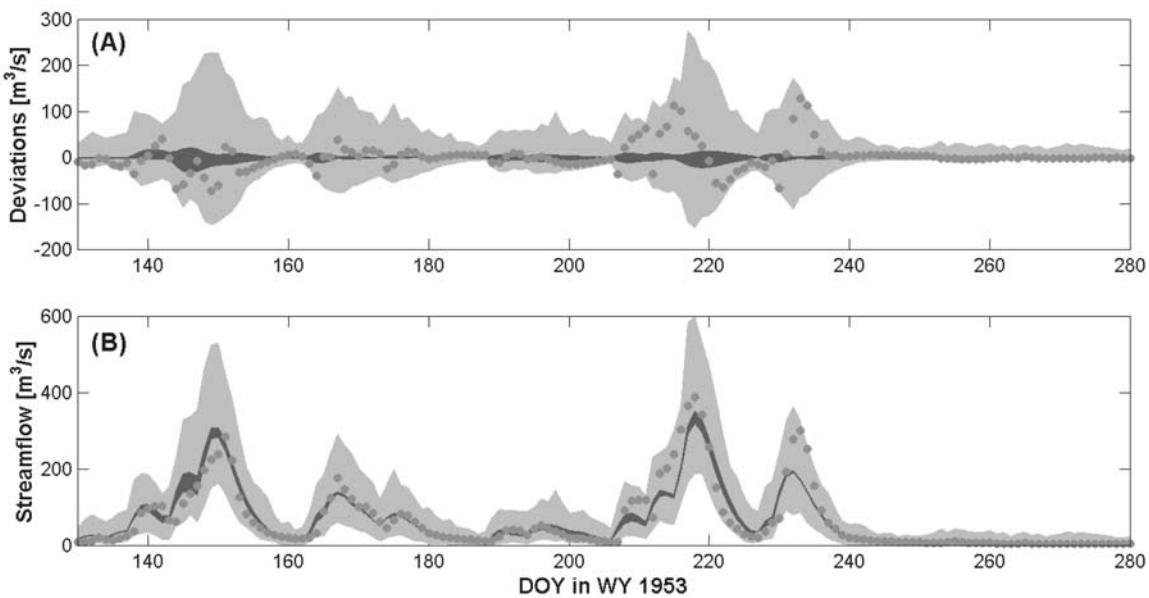

Figure 11. (a) Hydrograph prediction uncertainty associated with the most probable set derived using the SCEM-UA algorithm. The lighter shaded region denotes model uncertainty, whereas parameter uncertainty is indicated with the darker shaded region. The dots correspond to the observed streamflow data. (b) Hydrograph prediction uncertainty associated with the uncertainty in the model (lighter shading) and parameter estimates (darker shading) for the Water Year 1953. The dots correspond to the observed streamflow data. See color version of this figure at back of this issue.

(darker shaded region), and the 95% hydrograph prediction uncertainty associated with the total error (lighter shaded region) in terms of the model residuals (computed by subtracting the hydrograph prediction having the highest posterior probability). Figure 11b shows the same information in the streamflow space. Note that the 95% total prediction uncertainty ranges bracket the observed flows during the period, but are quite large, indicating considerable uncertainty in the predictions given the current model structure and the data used to condition the model. Further, the hydrograph prediction uncertainty associated only with the posterior distribution of the parameter estimates (darker shading) does not include the observations and displays bias on the long recessions, suggesting that the model structure may be in need of further improvement.

4. Summary

[78] This paper has presented a Markov Chain Monte Carlo sampler, which is well suited for the practical assessment of parameter uncertainty in hydrological models. The sampler, entitled the Shuffled Complex Evolution Metropolis algorithm, merges the strengths of the Metropolis-Hastings algorithm, controlled random search, competitive evolution, and complex shuffling to evolve a population of sampled points to an approximation of the stationary posterior distribution of the parameters. There are two differences between the SCEM-UA algorithm and the original SCE-UA algorithm presented by *Duan et al.* [1992]. These modifications prevent the search from becoming mired in a small basin of attraction and facilitate convergence to a stationary posterior target distribution. The first modification involves replacement of the downhill Simplex method by a Metropolis-annealing covariance-based offspring approach, thereby avoiding a deterministic drift toward a single mode. Second, the SCEM-UA algorithm does not further divide the complex into subcomplexes during the generation of the offspring and uses a different replacement procedure, to counter any tendency of the search to terminate occupations in the lower posterior density region of the parameter space.

[79] The efficiency and effectiveness of the newly developed SCEM-UA algorithm for estimating the posterior distribution of the parameters was compared with the traditional Metropolis-Hastings sampler for three case studies of increasing complexity. The first case study, a simple bimodal probability distribution, showed that the SCEM-UA algorithm does indeed successfully infer a known posterior target distribution. The second and third case study explored the effectiveness and efficiency of the algorithm for assessing parameter uncertainty in a highly nonlinear banana-shaped posterior test distribution and a conceptual watershed model. Both studies clearly demonstrated that the SCEM-UA is efficient, needing smaller numbers of simulations than the MH algorithm for realistic assessment of parameter uncertainty. The ability of the SCEM-UA algorithm to exchange information between parallel launched sequences increases the traversing speed of the chains through the feasible parameter space. Besides inferring the posterior distribution of the model parameters, an additional advantage of the SCEM-UA algorithm is that it simultaneously identifies the most likely parameter values within this high-density region. This makes superfluous the two-step procedure in which the global optimum in the parameter space is first identified, followed by launching parallel MH samplers from this starting point to estimate parameter uncertainty.

[80] Research aimed at further improvements of the Shuffled Complex Evolution Metropolis approach, including extensions to multicriteria problems, is ongoing. The results of this work will be reported in due course. As always, we invite dialog with others interested in these topics. The code for the SCEM-UA algorithm is available from the first author.

[81] **Acknowledgments.** The Earth Life Sciences and Research Council (ALW) partly supported the investigations of the first author with financial aid from the Netherlands Organization for Scientific Research (NWO). This material is also based upon work supported in part by SAHRA (Sustainability of Semi-Arid Hydrology and Riparian Areas) under the STC Program of the National Science Foundation, agreement EAR-9876800. The first author also wishes to thank SAHRA for making possible his extended visit with the Department of Hydrology and Water Resources of the University of Arizona, Tucson. Special thanks are due to Corrie Thies for proofreading of the manuscript. The quality of this paper has been greatly improved by the constructive comments of the anonymous reviewers and associate editor handling this manuscript.

References

Beven, K. J., and A. M. Binley, The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Processes*, *6*, 279–298, 1992.

Box, G. E. P., and G. C. Tiao, *Bayesian Inference in Statistical Analyses*, Addison-Wesley-Longman, Reading, Mass., 1973.

Boyle, D. P., Multicriteria calibration of hydrological models, Ph.D. dissertation, Dep. of Hydrol. and Water Resour., Univ. of Ariz., Tucson, 2000.

Boyle, D. P., H. V. Gupta, and S. Sorooshian, Toward improved calibration of hydrological models: Combining the strengths of manual and automatic methods, *Water. Resour. Res.*, *36*(12), 3663–3674, 2000.

Brazil, L. E., Multilevel calibration strategy for complex hydrological simulation models, Ph.D. dissertation, 217 pp., Colo. State Univ., Fort Collins, Colo., 1988.

Dawdy, D. R., and T. O'Donnell, Mathematical models of catchment behavior, *J. Hydraul. Div. Am. Soc. Civ. Eng.*, *91*, 113–137, 1965.

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

Duan, Q., V. K. Gupta, and S. Sorooshian, A shuffled complex evolution approach for effective and efficient global minimization, *J. Optim. Theory Appl.*, *76*(3), 501–521, 1993.

Duan, Q., J. Schaake, and V. Koren, A priori estimation of land surface model parameters, in *Land Surface Hydrology, Meteorology, and Climate*, *Water Sci. Appl. Ser.*, vol. 3, edited by V. Lakshmi, J. Albertson, and J. Schaake, pp. 77–94, AGU, Washington, D. C., 2001.

Gan, T. Y., and G. F. Biftu, Automatic calibration of conceptual rainfall-runoff models: Optimization algorithms, catchment conditions, and model structure, *Water Resour. Res.*, *32*(12), 3513–3524, 1996.

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

Gelman, A., J. B. Carlin, H. S. Stren, and D. B. Rubin, *Bayesian Data Analysis*, Chapmann and Hall, New York, 1995.

Geman, S., and D. Geman, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Mach. Intel.*, *6*, 721–741, 1984.

Geyer, C. J., Markov Chain Monte Carlo maximum likelihood, in *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, edited by E. M. Keramidas, pp. 156–163, Interface Found., Fairfax Station, Va., 1991.

Gilks, W. R., S. Richardson, and D. J. Spiegelharter, Introducing Markov Chain Monte Carlo, in *Markov Chain Monte Carlo in Practice*, edited by W. R. Gilks, S. Richardson, and D. J. Spiegelharter, pp. 1–19, Chapman and Hall, New York, 1995.

Gilks, W. R., S. Richardson, and D. Spiegelhalter (Eds.), *Practical Markov Chain Monte Carlo*, Chapman and Hall, New York, 1996.

Gilks, W., G. Roberts, and S. Sahu, Adaptive Markov Chain Monte Carlo through regeneration, *J. Am. Stat. Assoc.*, *443*(93), 1045–1054, 1998.

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

Haario, H., E. Saksman, and J. Tamminen, Adaptive proposal distribution for random walk Metropolis algorithm, *Comput. Stat.*, *14*(3), 375–395, 1999.

Haario, H., E. Saksman, and J. Tamminen, An adaptive Metropolis algorithm, *Bernoulli*, *7*(2), 223–242, 2001.

Hastings, W. K., Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, *57*, 97–109, 1970.

Hogue, T. S., S. Sorooshian, H. V. Gupta, A. Holz, and D. Braatz, A multistep automatic calibration scheme for river forecasting models, *J. Hydrometeorol.*, *1*, 524–542, 2000.

Holland, J., *Adaptation in Natural and Artificial Systems*, Univ. of Mich. Press, Ann Arbor, Mich., 1975.

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

Kass, R., and A. Raftery, Bayes factors, *J. Am. Stat. Assoc.*, *90*, 773–795, 1995.

Kuczera, G., Efficient subspace probablistic parameter optimization for catchment models, *Water Resour. Res.*, *33*(1), 177–185, 1997.

Kuczera, G., and M. Mroczkowski, Assessment of hydrological parameter uncertainty and the worth of multiresponse data, *Water Resour. Res.*, *34*(6), 1481–1489, 1998.

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

Luce, C. H., and T. W. Cundy, Parameter identification for a runoff model for forest roads, *Water Resour. Res.*, *30*(4), 1057–1069, 1994.

Mengersen, K. L., and R. L. Tweedie, Rates of convergence of the Hastings and Metropolis algorithms, *Ann. Stat.*, *24*, 101–121, 1996.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equations of state calculations by fast computing machines, *J. Chem. Phys.*, *21*, 1087–1091, 1953.

Misirli, F., H. V. Gupta, S. Sorooshian, and M. Thiemann, Bayesian recursive estimation of parameter and output uncertainty for watershed models, in *Calibration of Watershed Models*, *Water Sci. Appl. Ser.*, vol. 6, edited by Q. Duan et al., pp. 113–124, AGU, Washington, D. C., 2003.

Moore, R. J., The probability-distributed principle and runoff production at point and basin scales, *Hydrol. Sci. J.*, *30*(2), 273–297, 1985.

Neal, R., Probablistic inference using Markov Chain Monte Carlo methods, *Tech. Rep. CRG-TR-93-1*, Dep. of Comput. Sci., Univ. of Toronto, Toronto, Ont., Canada, 1993.

Price, W. L., Global optimization algorithms for a CAD workstation, *J. Optim. Theory Appl.*, *55*(1), 133–146, 1987.

Roberts, G. O., and R. L. Tweedie, Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms, *Biometrika*, *83*, 95–110, 1996.

Schaap, M. G., F. J. Leij, and M. T. van Genuchten, Neural network analysis for hierarchical prediction of soil hydraulic properties, *Soil Sci. Soc. Am. J.*, *62*, 847–855, 1998.

Sorooshian, S., Q. Duan, and V. K. Gupta, Calibration of rainfall-runoff models: Application of global optimization to the Sacramento Soil Moisture accounting model, *Water Resour. Res.*, *29*, 1185–1194, 1993.

Tanakamaru, H., Parameter estimation for the tank model using global optimization, *Trans. Jpn. Soc. Irrig. Drain. Reclam. Eng.*, *178*, 103–112, 1995.

Tarantola, A., *Inverse Problem Theory*, Elsevier-Sci., New York, 1987.

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

Thyer, M., G. Kuczera, and B. C. Bates, Probabilistic optimization for conceptual rainfall-runoff models: A comparison of the shuffled complex evolution and simulated annealing algorithms, *Water Resour. Res.*, *35*(3), 767–773, 1999.

Troutmann, B., Errors and parameter estimation in precipitation runoff modeling: 1. Theory, *Water Resour. Res.*, *21*, 1195–1213, 1985.

Vrugt, J. A., and W. Bouten, Validity of first-order approximations to describe parameter uncertainty in soil hydrologic models, *Soil Sci. Soc. Am. J.*, *66*, 1740–1751, 2002.

Vrugt, J. A., W. Bouten, H. V. Gupta, and S. Sorooshian, Toward improved identifiability of hydrologic model parameters: The information content of experimental data, *Water Resour. Res.*, *38*(12), 1312, doi:10.1029/2001WR001118, 2002.

Wagener, T., D. P. Boyle, M. J. Lees, H. S. Wheater, H. V. Gupta, and S. Sorooshian, A framework for development and application of hydrological models, *Hydrol. Earth Syst. Sci.*, *5*(1), 13–26, 2001.

W. Bouten and J. A. Vrugt, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Nieuwe Achtergracht 166, Amsterdam, 1018 WV, Netherlands. (j.vrugt@science.uva.nl)

H. V. Gupta and S. Sorooshian, Department of Hydrology and Water Resources, University of Arizona, Tucson, AZ 85721, USA.

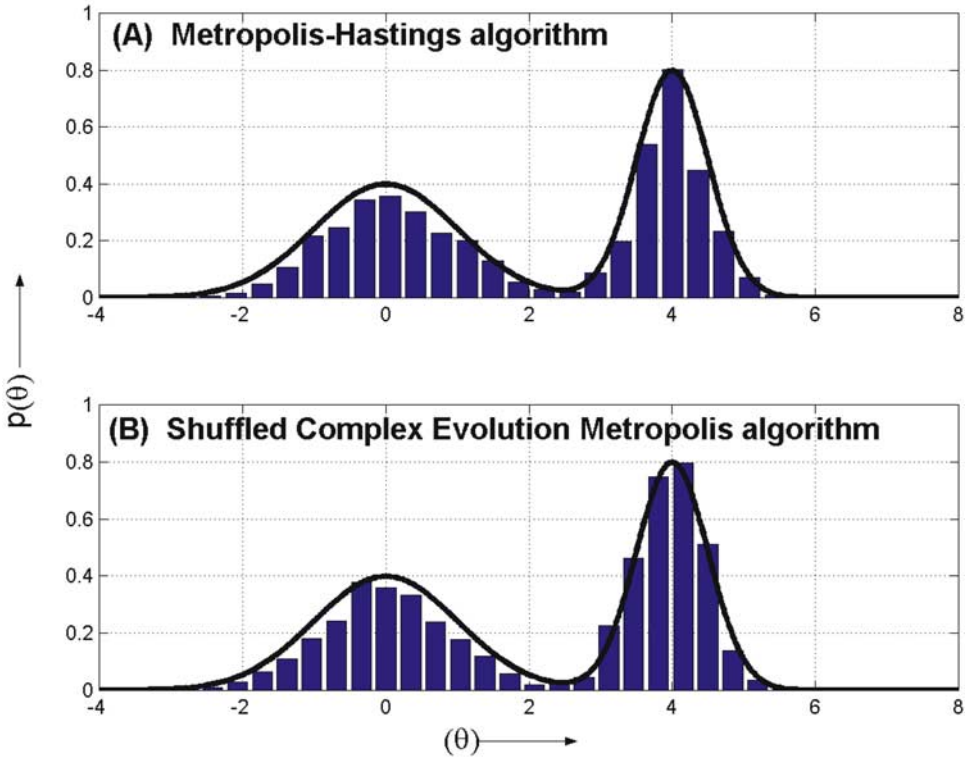

Figure 3. Bimodal probability distribution and histogram of 4000 samples generated using (a) the Metropolis and (b) the Shuffled Complex Evolution Metropolis algorithms.

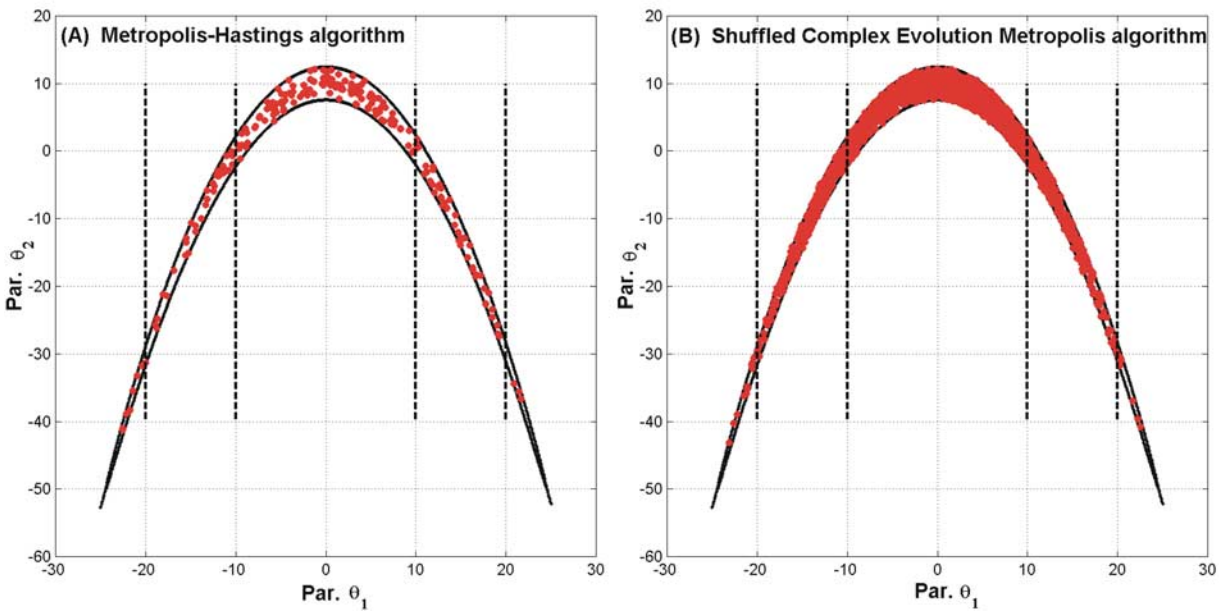

Figure 5. A scatterplot of the (θ_1,θ_2) samples generated after convergence to a stationary posterior distribution has been achieved using (a) the Metropolis, and (b) the Shuffled Complex Evolution Metropolis algorithms. The lines indicate the one-dimensional 68.3 and 95% confidence regions of the parameters.

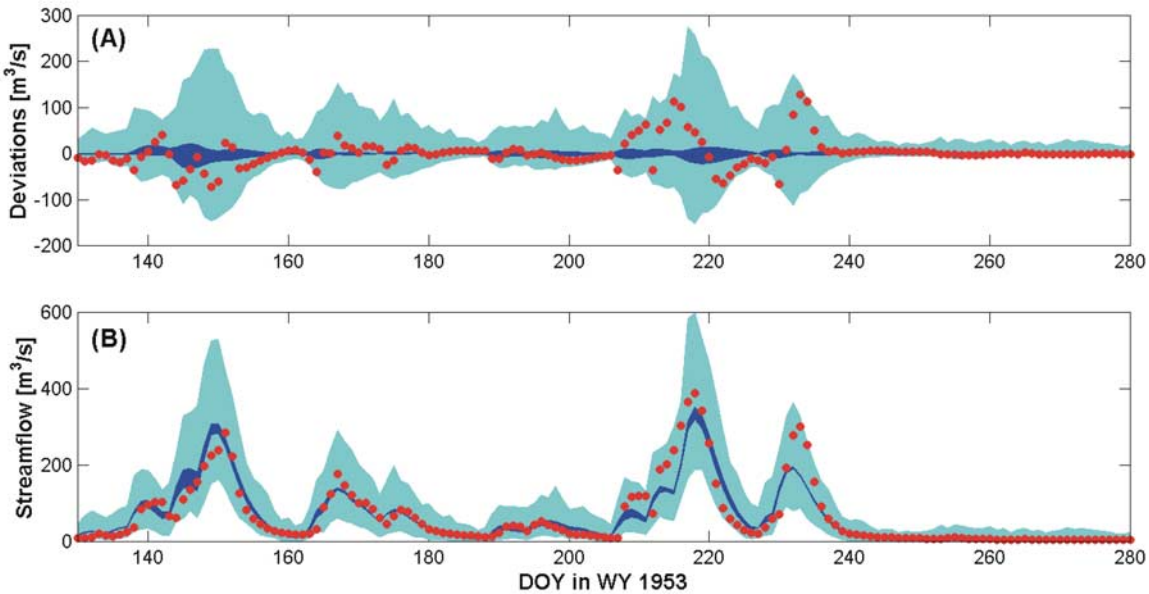

Figure 11. (a) Hydrograph prediction uncertainty associated with the most probable set derived using the SCEM-UA algorithm. The lighter shaded region denotes model uncertainty, whereas parameter uncertainty is indicated with the darker shaded region. The dots correspond to the observed streamflow data. (b) Hydrograph prediction uncertainty associated with the uncertainty in the model (lighter shading) and parameter estimates (darker shading) for the Water Year 1953. The dots correspond to the observed streamflow data.