

Effective and efficient algorithm for multiobjective optimization of hydrologic models

Jasper A. Vrugt,¹ Hoshin V. Gupta,² Luis A. Bastidas,³ Willem Bouten,¹ and Soroosh Sorooshian²

Received 2 October 2002; accepted 7 April 2003; published 20 August 2003.

[1] Practical experience with the calibration of hydrologic models suggests that any single-objective function, no matter how carefully chosen, is often inadequate to properly measure all of the characteristics of the observed data deemed to be important. One strategy to circumvent this problem is to define several optimization criteria (objective functions) that measure different (complementary) aspects of the system behavior and to use multicriteria optimization to identify the set of nondominated, efficient, or Pareto optimal solutions. In this paper, we present an efficient and effective Markov Chain Monte Carlo sampler, entitled the Multiobjective Shuffled Complex Evolution Metropolis (MOSCEM) algorithm, which is capable of solving the multiobjective optimization problem for hydrologic models. MOSCEM is an improvement over the Shuffled Complex Evolution Metropolis (SCEM-UA) global optimization algorithm, using the concept of Pareto dominance (rather than direct single-objective function evaluation) to evolve the initial population of points toward a set of solutions stemming from a stable distribution (Pareto set). The efficacy of the MOSCEM-UA algorithm is compared with the original MOCOM-UA algorithm for three hydrologic modeling case studies of increasing complexity. *INDEX TERMS*: 1869 Hydrology: Stochastic processes; 1836 Hydrology: Hydrologic budget (1655); 1894 Hydrology: Instruments and techniques; *KEYWORDS*: parameter optimization, Markov chain Monte Carlo, multicriteria calibration, population diversity, Pareto ranking, hydrologic models

Citation: Vrugt, J. A., H. V. Gupta, L. A. Bastidas, W. Bouten, and S. Sorooshian, Effective and efficient algorithm for multiobjective optimization of hydrologic models, *Water Resour. Res.*, 39(8), 1214, doi:10.1029/2002WR001746, 2003.

1. Introduction and Scope

[2] Many hydrologic models contain parameters that cannot be measured directly, but must be estimated through indirect methods such as calibration. In the process of calibration, the hydrologist adjusts the values of the model parameters such that the model is able to closely match the behavior of the real system it is intended to represent. In its most elementary form, this calibration is performed by manually adjusting the parameters while visually inspecting the agreement between observations and model predictions [Janssen and Heuberger, 1995]. In this approach, the “closeness” of the model with the measurements is typically evaluated in terms of several subjective visual measures, and a semi-intuitive trial-and-error process is used to perform the parameter adjustments [Boyle *et al.*, 2000]. Because of the subjectivity and time-consuming nature of manual trial-and-error calibration, there has been a great deal of research into the development of automatic methods for calibration of hydrologic

models [e.g., Gupta and Sorooshian, 1994]. Automatic methods seek to take advantage of the speed and power of digital computers, while being objective and relatively easy to implement.

[3] Research into the development of automatic calibration methods has focused mainly on the selection of a single-objective measure and the selection of an automatic optimization strategy that can reliably optimize (maximize or minimize, as appropriate) that measure. In this regard, the Shuffled Complex Evolution (SCE-UA) global optimization algorithm developed by Duan *et al.* [1992, 1993] has proved to be consistent, effective, and efficient in locating the parameter values of a hydrologic model that optimize a given objective function. However, practical experience with the calibration of hydrologic models suggests that single-objective functions, no matter how carefully chosen, are often inadequate to properly measure all of the characteristics of the observed data deemed to be important. Consequently, single-objective calibration methods do not usually provide parameter estimates that are considered acceptable by practicing hydrologists. Another emerging problem is that many of the latest hydrologic watershed or land-surface models simulate several output fluxes (e.g., water, energy, chemical constituents, etc.) for which measurement data are available, and all these data must be correctly utilized to ensure proper model calibration [Beven and Kirkby, 1979; De Grosbois *et al.*, 1988; Kuczera, 1982, 1983; Woolhiser *et al.*, 1990; Kuczera and Mroczkowski, 1998; Gupta *et al.*, 1998]. One strategy to

¹Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, Netherlands.

²Department of Hydrology and Water Resources, University of Arizona, Tucson, Arizona, USA.

³Department of Civil and Environmental Engineering, Utah State University, Logan, Utah, USA.

explicitly recognize the multiobjective nature of the calibration problem is to define several optimization criteria (objective functions) that measure different (complementary) aspects of the system behavior and to use a multicriteria optimization method to identify the set of nondominated, efficient, or Pareto optimal solutions [Gupta *et al.*, 1998; Yapo *et al.*, 1998; Boyle *et al.*, 2000]. The Pareto solutions represent tradeoffs among the different incommensurable and often conflicting objectives, having the property that moving from one solution to another results in the improvement of one objective while causing deterioration in one or more others.

[4] A simple way to obtain a crude approximation of the Pareto solution set is to weigh the different criteria into a single aggregated scalar and to run a large number of independent single-criteria optimization runs using different values for the weights. Popular aggregation methods include the weighted-sum approach, target vector optimization, and the method of goal attainment [Srinivas and Deb, 1994; Fonseca and Fleming, 1995]. Recently, Madsen [2000] used an aggregation approach, in combination with the single-criterion SCE-UA global optimization algorithm, to construct an estimate of the Pareto front for a rainfall-runoff model application. Although the aggregation method is (in principle) simple to implement, a complete single-objective optimization must be solved to obtain each discrete Pareto solution, making this approach inefficient, cumbersome, and time-consuming. Moreover, there is arguably a significant advantage to maintaining the independence of the various criteria, because a full multicriteria optimization will allow an analysis of the tradeoffs among the different criteria and enable hydrologists to better understand the limitations of the current hydrologic model structure, thereby gaining insights into possible model improvements [Gupta *et al.*, 1998].

[5] Recently, a variety of evolutionary algorithms have become available which are designed to evolve multiple nondominated solutions concurrently in a single optimization run, thereby guiding the search in several directions at the same time. These algorithms use the concepts of Pareto dominance, rather than single-objective function evaluations, to construct a uniform estimate of the Pareto solution set. In the context of hydrologic modeling, Yapo *et al.* [1998] developed the Multiobjective Complex Evolution (MOCOM-UA) global optimization method, which solves the multiobjective calibration problem by combining the strengths of the complex shuffling strategy and downhill Simplex evolution (adapted from the SCE-UA global optimization algorithm) with the concepts of Pareto dominance. Various hydrologic and hydrometeorologic calibration and evaluation studies have demonstrated that the MOCOM algorithm can provide an efficient estimate of the Pareto set of solutions [Gupta *et al.*, 1998, 1999; Yapo *et al.*, 1998; Bastidas *et al.*, 1999; Boyle *et al.*, 2000, 2001; Wagener *et al.*, 2001; Xia *et al.*, 2002; Leplastrier *et al.*, 2002]. However, during the course of these investigations, we became aware that the current procedure has several weaknesses [Gupta *et al.*, 2003], including the tendency of the MOCOM algorithm to cluster the Pareto solutions into a central compromise region of the Pareto set, and a tendency to converge prematurely for case

studies involving larger numbers of parameters and strongly correlated performance criteria.

[6] In this paper, we present an effective and efficient Markov Chain Monte Carlo (MCMC) sampler, entitled the Multiobjective Shuffled Complex Evolution Metropolis (MOSCEM-UA) algorithm, which is capable of generating a fairly uniform approximation of the “true” Pareto frontier within a single optimization run. The algorithm is closely related to the SCEM-UA algorithm [Vrugt *et al.*, 2003], recently developed to infer the probabilistic uncertainty associated with the use of a single-objective function, and uses a newly developed, improved concept of Pareto dominance to generate a fairly uniform approximation of the “true” Pareto frontier (thereby also containing the single-criteria solutions at the extremes of the Pareto solution set). The features and capabilities of the MOSCEM-UA algorithm are illustrated using three hydrologic modeling case studies of increasing complexity, and the results are compared with the original MOCOM-UA algorithm.

2. Multiobjective Optimization

[7] To facilitate the description of the multiobjective optimization approach, let us write the hydrologic model η as follows:

$$\hat{y} = \eta(\xi|\theta) \quad (1)$$

where \hat{y} denotes the model output, ξ is the input data, and θ is a vector with n unknown parameters. We assume that the model structure specified by equation (1) is predetermined and fixed, and that realistic upper and lower bounds for each of the model parameters can be specified, thereby defining Θ , the feasible parameter space:

$$\theta \in \Theta \subseteq \mathcal{R}^n \quad (2)$$

Here \mathcal{R}^n denotes the n -dimensional Euclidean space. If Θ is not the entire domain space \mathcal{R}^n , the identification problem is said to be constrained.

[8] We first consider the situation in which the hydrological model is required to simulate only one aspect of the system and for which z directly observed output values, $\{y_i, i = 1, \dots, z\}$, are available. The difference between the model-simulated output and the observed data can be represented by the residual vector:

$$E(\theta) = G[\hat{y}(\theta)] - G[y] = \{e_1(\theta), e_2(\theta), \dots, e_z(\theta)\} \quad (3)$$

where the function $G(\cdot)$ allows for linear or nonlinear transformations of the simulated and observed data. The development of a measure $F(\theta)$, hereafter referred to as objective function, that mathematically measures the “size” of $E(\theta)$ is typically based on assumptions regarding the distributions of the measurement errors presented in the data. By far the most popular objective function is the Simple Least Squares (SLS), which is also the maximum likelihood estimator when the measurement errors are known to be Gaussian, homoscedastic, and uncorrelated.

[9] For single-objective problems (where $F(\theta)$ is a scalar), the Shuffled Complex Evolution (SCE-UA) global optimization algorithm developed by Duan *et al.* [1992, 1993] has

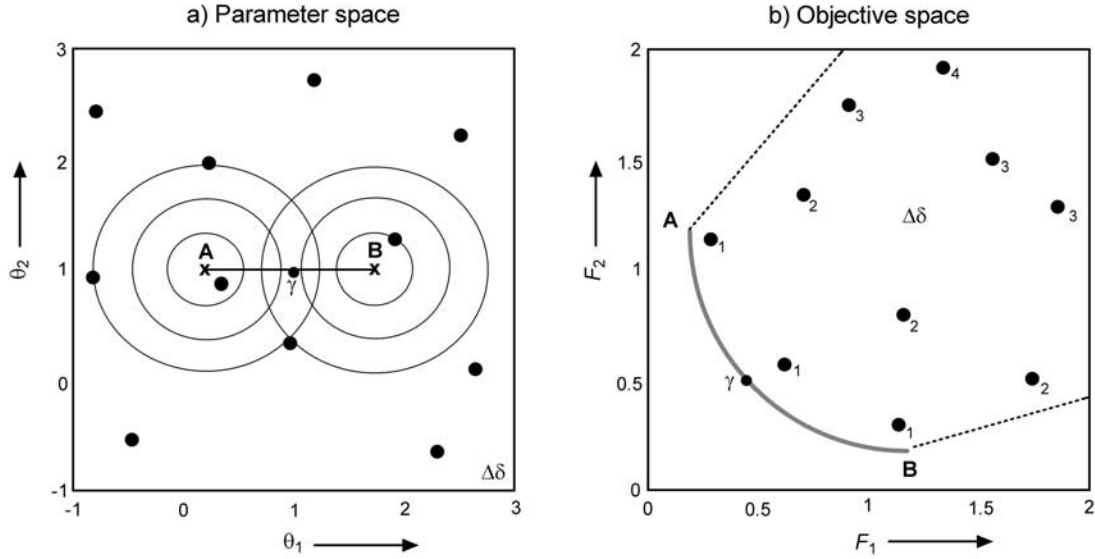


Figure 1. Illustration of the concept of Pareto optimality for a problem having two parameters (θ_1, θ_2) and two criteria (F_1, F_2), in the (a) parameter and (b) objective space. The points A and B indicate the solutions that minimize each of the individual criteria F_1 and F_2 . The thick line joining A and B corresponds to the Pareto set of solutions; γ is an element of the solution set, which is superior in the multicriteria sense to any other point in δ .

proved to be consistent, effective, and efficient in locating the values of the hydrologic model parameters that minimize (or maximize) the objective function. However, to quote *Kuczera and Parent* [1998] “. . .no hydrologist should be naïve enough to rely on a uniquely determined value for each of the model parameters θ , whatever the skill and imagination of the modeler might be.” In fact, it is typical that the vicinity of the global optimum contains several behavioral parameter sets with similar performance in reproducing the observed data. This issue prompted the design of a new methodology, entitled the Shuffled Complex Evolution Metropolis (SCEM-UA) algorithm that infers the best attainable parameter values and simultaneously estimates its underlying posterior distribution within a single optimization run [Vrugt *et al.*, 2003]. The uncertainty in the parameter estimates, based on probabilistic arguments, can subsequently be used to summarize the uncertainties in the output predictions of the hydrologic model.

[10] The classical single-objective optimization approach operates under the central assumption that a single-objective function is able to properly extract all of the information contained in the time series of observations. However, practical experience with the calibration of hydrologic models suggests that the magnitude of structural error in the model for some portions of the model response may, in general, be equivalent to or even substantially larger than the measurement error and that these structural or model errors do not necessarily have any inherent probabilistic property that can be exploited in the construction of an objective function [Gupta *et al.*, 1998]. Because of the presence of these structural inadequacies in the hydrologic model, any single (scalar) objective function, no matter how carefully chosen, is inadequate to properly measure all of the characteristics of the observed data deemed to be important.

[11] These considerations imply the design of a calibration strategy that has the ability to simultaneously incorporate several objective functions. A strategy that can address this challenge is multiobjective optimization, which has its roots in late nineteenth century welfare economics, in the work of *Edgeworth* [1881], and can be stated as follows:

$$\min_{\theta \in \Theta} F(\theta) = \begin{bmatrix} F_1(\theta) \\ F_2(\theta) \\ \vdots \\ F_M(\theta) \end{bmatrix} \quad (4)$$

where $F_i(\theta)$ is the i th of M objective functions. The solution to this problem will in general, no longer be a single “best” parameter set but will consist of a Pareto set $P(\Theta)$ of solutions in the feasible parameter space Θ corresponding to various trade-offs among the objectives. The Pareto set of solutions defines the minimum uncertainty in the parameters that can be achieved without stating a subjective relative preference for minimizing one specific component of $F(\theta)$ at the expense of another. Figure 1 illustrates the Pareto solution set for a simple problem where the aim is to simultaneously minimize two objectives (F_1, F_2) with respect to two parameters (θ_1, θ_2).

[12] The individual points A and B minimize objectives F_1 and F_2 , respectively, whereas the solid line joining A and B represents the theoretical Pareto set of solutions. The black dots indicate an initial set of parameter estimates, while the number in subscript denotes their corresponding Pareto rank. Moving along the line from A to B results in the improvement of F_2 while successively causing deterioration in F_1 . The points falling on the line AB represent trade-offs between the objectives and are called nondominated, noninferior, or efficient solutions. Put simply, the

feasible parameter space (shaded region) can be partitioned into “good” or Pareto solutions and “bad” or “inferior” solutions. In the absence of additional information, it is impossible to distinguish any of the Pareto solutions (rank 1 points) as being objectively better than any of the other Pareto solutions. Furthermore, every member of the Pareto set will match some characteristic of the observed data better than any other member of the Pareto set, but the trade-off will be that some other characteristic of the observed data will not be as well-matched [Yapo *et al.*, 1998]. Because of errors in the model structure (and other possible sources), it is usually not possible to find a single point θ at which all of the criteria have their minima. Note that this multiobjective equivalence of parameter sets is different from the probabilistic representation of parameter uncertainty, estimated using the SCEM-UA algorithm.

3. Effective and Efficient Algorithm to Solve the Multiobjective Optimization Problem

[13] While it may be relatively simple to pose the optimization problem into a multicriteria framework, solving this problem to identify the Pareto set of solutions is not easy and has been the subject of much research. Ideally, the multiobjective optimization algorithm should find the set of all nondominated solutions, which will constitute the global trade-off surface. However, because computational resources are finite, multiobjective solution algorithms typically approximate the Pareto set using a number of representative solutions.

[14] The Vector Evaluated Genetic Algorithm (VEGA), developed by Schaffer [1985], was the first algorithm able to cope with multiple objectives simultaneously, without resorting to a strategy of scalarization by aggregation in order to solve a single-objective surrogate problem instead. However, case studies have demonstrated that the VEGA algorithm has a tendency to ignore the most compromising points among the objectives on the trade-off curve, and that the algorithm eventually converges to a single point on the Pareto set. The first problem is an artifact of the VEGA strategy in which the evolution tends to favor individuals with extremely good performance among one of the objectives. To avoid these problems, Ritzel *et al.* [1994] modified the evolution process in VEGA by selecting parents based on a nondominance ranking procedure called “Pareto ranking” [Goldberg, 1989]. Performance testing demonstrated that the resulting Pareto genetic algorithm (GA) was superior to VEGA. However, the Pareto GA has an inherent tendency to converge too quickly, especially when the algorithmic parameters are not set properly, thereby yielding indistinguishable solutions that do not necessarily belong to the Pareto set.

[15] In response to these issues, Yapo *et al.* [1998] developed the Multiobjective Complex evolution (MOCOM-UA) method, a general-purpose multiobjective global optimization algorithm designed to efficiently generate a fairly uniform approximation of the Pareto set for a broad class of problems. The MOCOM-UA algorithm is an extension of the successful SCE-UA single-objective global optimization algorithm developed by Duan *et al.* [1992, 1993] and merges the strengths of controlled random search [Price, 1987] with a competitive evolution [Holland, 1975], Pareto ranking [Goldberg, 1989], and multiobjective downhill Simplex

strategy. For a detailed description and explanation of the algorithm, please refer to Yapo *et al.* [1998]. Various applications of the MOCOM algorithm in hydrologic and hydro-meteorologic calibration and evaluation studies [Gupta *et al.*, 1998, 1999; Yapo *et al.*, 1998; Bastidas *et al.*, 1999; Boyle *et al.*, 2000, 2001; Wagener *et al.*, 2001; Xia *et al.*, 2002; Leplastrier *et al.*, 2002] have demonstrated the usefulness of the MOCOM algorithm. However, during the course of these investigations, it has become apparent that the current methodology has some serious weaknesses that need to be resolved. In further investigations, we discovered that these weaknesses are typical of the evolutionary algorithms, which are currently available for solving the multiobjective optimization problem.

[16] The first failing of the MOCOM-UA algorithm is that it does not consistently generate a uniform approximation to the Pareto front, but tends to cluster the solutions in the compromise region among the objectives (e.g., see points on the Pareto set of Figure 1b), thereby leaving the ends of the Pareto frontier unrepresented. Consequently, the Pareto set of solutions does not contain the individual single-criterion (SCE-UA) solutions, which represent the theoretical extreme ends of the Pareto frontier. The second, perhaps more important, failure is the inability of the evolution strategy in the MOCOM algorithm to converge to solutions within the “true” Pareto set for case studies involving large numbers of parameters and highly correlated performance criteria (e.g., typical of soil-vegetation-atmosphere transfer scheme (SVATS) models, also known as land-surface models (LSMs)). The algorithm tends, instead, to converge to a fuzzy region surrounding the Pareto set and, in some cases, does not converge at all. Note, that the phenomenon of genetic drift, where the members of the population drift to a single solution, is a characteristic typical of many evolutionary search algorithms. To prevent the collapse of the algorithm to a single region of highest attraction, the evolutionary algorithm incorporates a strategy that preserves the diversity of the sampled population.

3.1. Preservation of Diversity in Population

[17] To find a set of nondominated solutions, rather than a single-point solution, a multiobjective evolutionary algorithm must perform a multimodal search that samples the Pareto-optimal set uniformly [Zitzler and Thiele, 1999]. We believe that there are two main reasons why current strategies for solving the multiobjective calibration problem do not preserve the diversity in the population, thereby tending to converge toward a compromise solution among the objectives instead.

[18] 1. The first is replacement strategy. Replacement of a member of the existing population occurs only if the generated offspring has a higher fitness than its parent. Although this evolution strategy essentially causes the algorithm to converge to a set of Pareto solutions, in the case of complex-shaped response surfaces with different regions of attraction (i.e., in the case of hydrologic models), this might cause the algorithm to prematurely converge to a single region of highest attraction surrounding the Pareto set.

[19] 2. The second is fitness assignment. In a multi-objective problem, several objectives are to be considered simultaneously, and ordered ranking of the population by conventional scalar sorting is therefore not possible. Fortu-

nately, *Goldberg* [1989] suggested an elegant superiority-inferiority method for the ranking of a population of criteria, based on their mutual dominance relations. However, this widely used concept of Pareto ranking for fitness assignment in the case of multiple objectives does not distinguish between members having an identical rank. The solutions at the extreme ends of the Pareto frontier are assigned an identical “fitness” as the members of the Pareto set located in the most compromised region; however a greater number of solutions are usually found in this compromise region or “niche” of the parameter space.

[20] With regard to the “replacement strategy” problem, the clumping tendency might be overcome by using a large initial population size, but this will significantly increase the number of required function evaluations and adversely affect the efficiency of the search algorithm. In examining ways to preserve the diversity in the sampled population and therefore avoid clumping of the solutions and premature convergence of the MOCOM algorithm, it seems natural to consider the evolution strategy employed in the SCEM-UA algorithm [*Vrugt et al.*, 2003], which is also designed to converge to a distribution of points, rather than a single “best” parameter set. The stochastic nature of the Metropolis-annealing scheme in the SCEM-UA algorithm counters any tendency to collapse to a single region of attraction, thereby making possible the simultaneous identification of the “best” parameter set as well as its underlying posterior distribution. With regard to the “fitness assignment” problem, we propose a new (improved) concept of Pareto dominance that enables an evolutionary algorithm to preserve the diversity in the population.

3.2. Fitness Assignment Based on the Number of External Nondominated Points

[21] The rank fitness assignment procedure begins by identifying all of the nondominated individuals in the population and assigning them rank “one”. While the original Pareto ranking concept now proceeds by peeling off these points and identifies the nondominated points of the remaining population (assigned rank “two”), the proposed fitness assignment by *Zitzler and Thiele* [1999] proceeds as follows.

[22] 1. Store all of the rank “one” points in an external nondominated set P' and the remaining dominated points of the population in a set entitled P .

[23] 2. Each solution $i \in P'$ is assigned a real value $r_i \in [0, 1)$, called strength. The strength is proportional to the number of population members $j \in P$ for which $i \succeq j$. Let N be the number of individuals in P that are covered by i and s is the population size ($P + P'$). The strength is now defined as, $r_i = \frac{N}{s}$. For each member i of P' , the fitness (f_i) is identical to its computed strength (r_i).

[24] 3. The fitness of the remaining dominated individuals $j \in P$ is calculated by summing the strengths of all external nondominated solutions $i \in P'$ that cover j :

$$f_j = 1 + \sum_{i=1}^{i \leq j} r_i \quad \text{where } f_j \in [1, s) \quad (5)$$

To ensure that the members of P have a lower fitness than the members of P' , the number one is added to the total sum. The closer the computed f value (in equation (5)) is to zero, the higher the fitness of the sampled point.

[25] To illustrate the proposed fitness assignment method and the difference with the conventional ranking method for a two-objective (F_1, F_2) problem, consider Figure 2 which presents computed fitness values for each ranking method at two different stages during the optimization.

[26] In the case of conventional Pareto ranking (Figures 2a and 2b), points having an identical rank number are not distinguishable, even though the solutions at the extreme end are in some sense much more unique than other solutions having the same rank. In the case of the proposed fitness assignment (Figures 2c and 2d), nondominated individuals at the extreme end of the Pareto cluster are preferred, and individuals having many neighbors in their niche are penalized due to the high strength value of the associated nondominated point (see shaded area in Figure 2c). Both of these principles of the proposed fitness assignment method preserve the diversity of the population and therefore favor uniform spacing of the solutions along the Pareto frontier (see Figure 2c), thereby further reducing the chances of clumping of the solutions in the most compromised region (e.g., Figure 2b) and of premature convergence.

[27] Although the fitness assignment method by *Zitzler and Thiele* [1999] aims to preserve the diversity in the population, our attempts to apply the method to hydrologic models containing large numbers of parameters and with correlated objectives has revealed a major drawback of the method. When the nondominated external set (P') contains only one member, which tends to occur when evaluating a population of points having strongly correlated objectives, the total set of dominated points will then appear to have identical fitness, thereby making it difficult for the MO algorithm to find a direction of improvement. To circumvent this problem, we modified the fitness assignment of the members of the dominated set, previously defined in step 3, by adding the Pareto rank of each of the members of P computed using the traditional ranking concept of the dominated set [*Goldberg*, 1989] to the sum of strengths calculated in equation (5). We found that this modification further improves the convergence properties of the MOSCEM algorithm.

3.3. Multiobjective Shuffled Complex Evolution Metropolis (MOSCEM) Algorithm

[28] In this section, we present the newly developed Multiobjective Shuffled Complex Evolution Metropolis (MOSCEM-UA developed in collaboration between the University of Amsterdam and University of Arizona) algorithm. The evolution strategy employed in the MOSCEM-UA algorithm is identical to the strategy utilized in the SCEM-UA algorithm [*Vrugt et al.*, 2003], with the exception that, to evolve the initial population of points toward a set of solutions stemming from a stable distribution, the probability ratio concept in the SCEM-UA algorithm is replaced with a multiobjective fitness assignment concept. The MOSCEM-UA algorithm is presented below and illustrated in the Figures 3 and 4.

Setup (see Figure 3)

1. To initialize the process, choose the population size s and the number of complexes q .
2. Generate s samples $\{\theta_1, \theta_2, \dots, \theta_s\}$ from the prior distribution and compute the multiobjective vector $F(\theta_i)$ at each point θ_i .

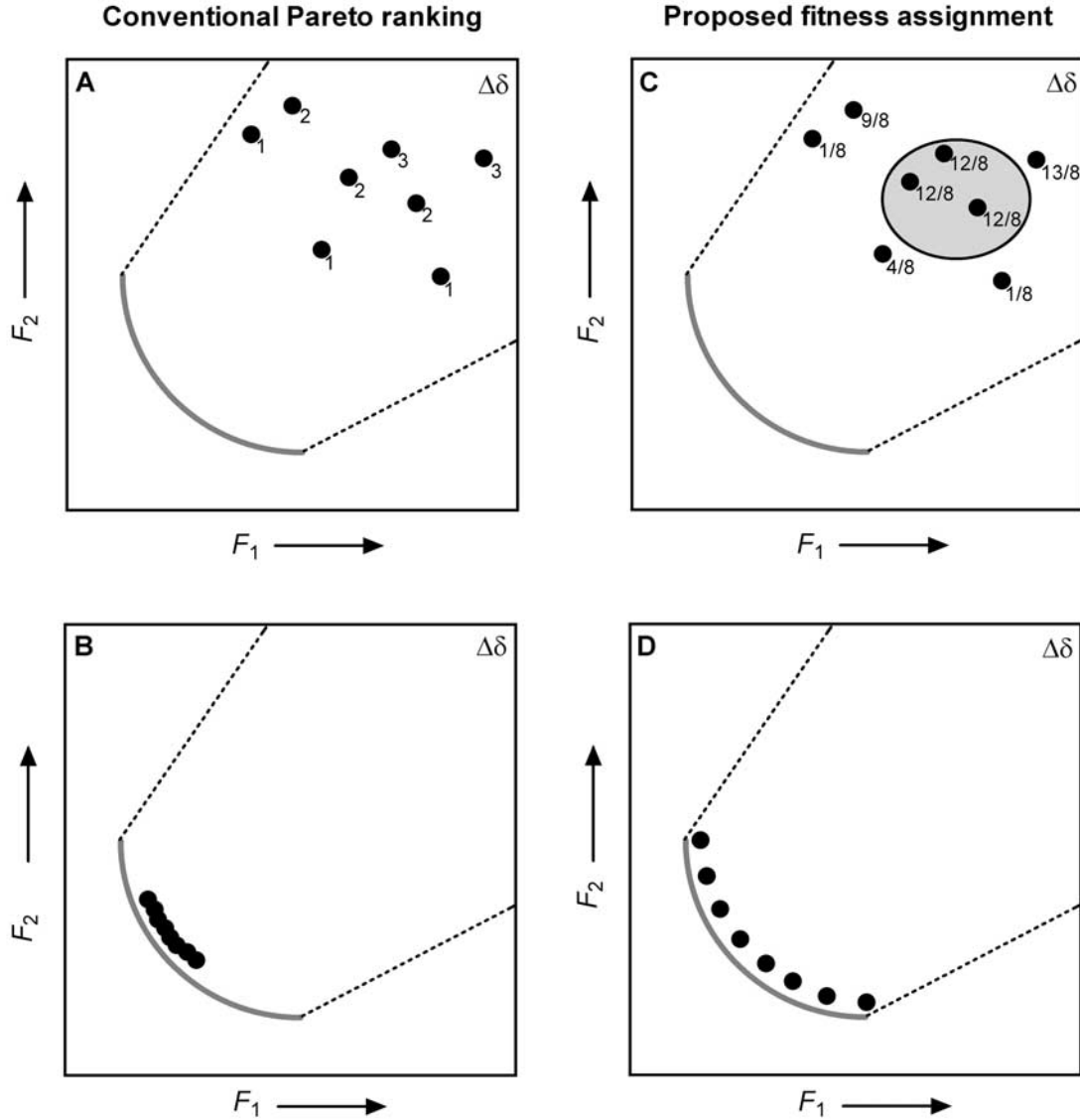


Figure 2. (a and b) Illustration of the conventional Pareto ranking concept and (c and d) the proposed fitness assignment concept for a two-objective (F_1 , F_2) problem. For more explanation, see text.

3. Compute the fitness f_i for each individual of the sample (section 3.2), sort the s individuals by decreasing fitness value, and store them in an array $D [1:s, 1:n + M + 1]$, where n is the number of parameters, so that the first row of D represents the point with the “best” fitness. The extra columns in D are used to store the multiobjective vector and the fitness values.
4. Initialize the starting points of the parallel sequences, S^1, S^2, \dots, S^q , such that S^k is $D[k, 1:n + 1]$, where $k = 1, 2, \dots, q$.
5. Partition D into q complexes C^1, C^2, \dots, C^q , each containing m points, such that the first complex contains every $q(j - 1)$ sorted point of D , the second complex contains every $q(j - 1) + 2$ sorted point, and so on, where $j = 1, 2, \dots, m$.

Sequence Evolution (see Figure 4): **For $k = 1$ to $k = q$ Do Begin**

For $\beta = 1$ to $\beta = L$ Do Begin

Compute the covariance structure Σ^k of the parameters of C^k .

Randomly draw a uniform label Z over interval $[0, 1]$.

While Not Drawn Feasible Offspring Do Begin

Generate offspring according to:

$$\theta^{(t+1)} = N\left(\theta^{(t)}, \Sigma^k\right) \quad (6)$$

where $\theta^{(t)}$ is the current draw of S^k .

End

Begin Metropolis Step

[I] Compute f_{t+1} using the points in C^k and the current draw of S^k .

[II] Compute the ratio, $\alpha = \left(\frac{f_t}{f_{t+1}}\right)^{\beta(r_{t+1})}$

where β is a scaling factor and f_t is the fitness associated with the current draw of S^k .

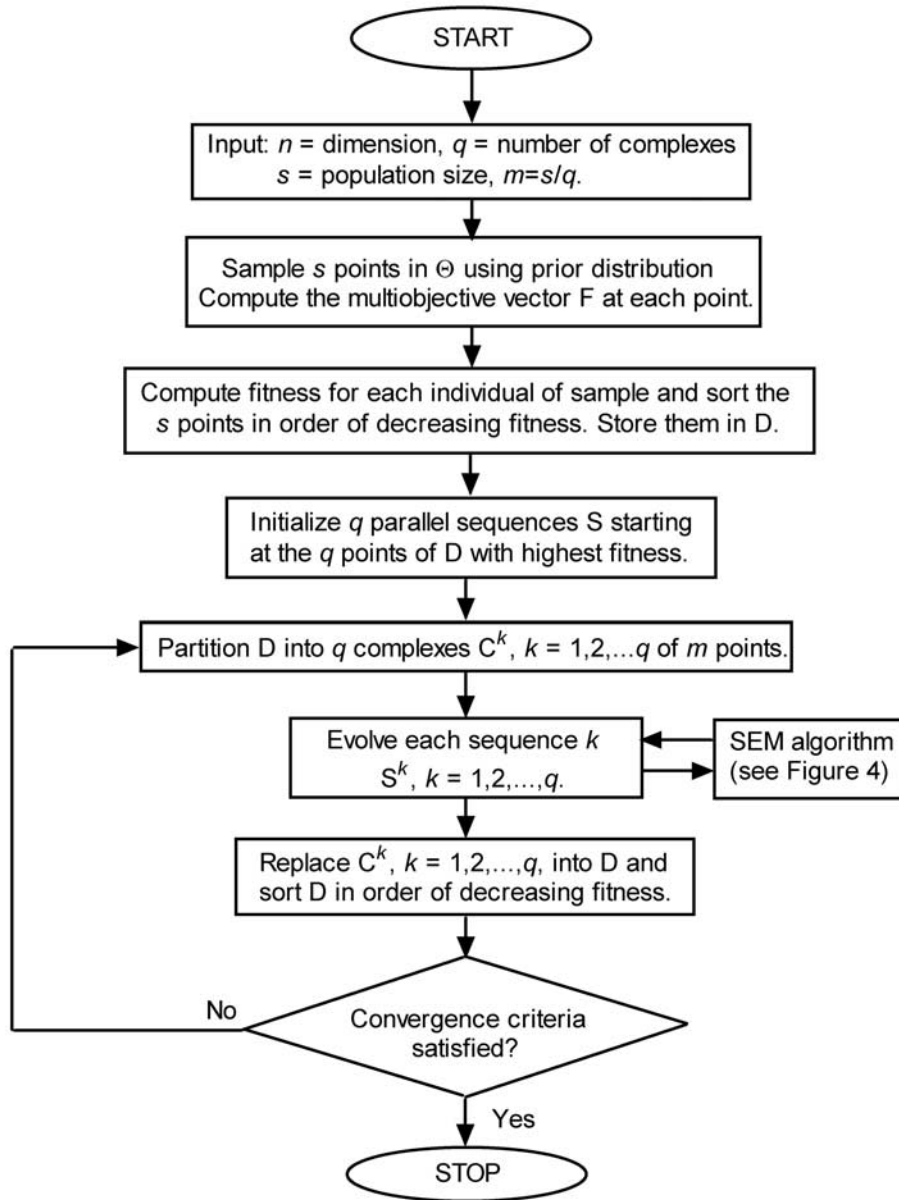


Figure 3. Flowchart of the MOSCEM-UA algorithm.

[III] If $\alpha \geq Z$, then accept the offspring. However, if $\alpha < Z$, then reject the offspring and remain at the current position, that is, $\theta^{(t+1)} = \theta^{(t)}$.

[IV] Add the point $\theta^{(t+1)}$ to the sequence S^k .
[V] Replace the worst point of C^k with $\theta^{(t+1)}$.

End Metropolis Step

End

End

6. Unpack all complexes C back into D and sort the points in order of increasing fitness value.
7. Check convergence statistic. If convergence criteria are satisfied, stop; otherwise, return to step 5.

The MOSCEM-UA algorithm combines the strengths of (1) the complex shuffling employed in the SCE-UA algorithm [Duan *et al.*, 1992, 1993], (2) the probabilistic

covariance-annealing search procedure of the SCEM-UA algorithm [Vrugt *et al.*, 2003], and (3) our improved version of the fitness assignment concept of Zitzler and Thiele [1999] to construct an efficient and uniform estimate of the Pareto solution set.

[29] To summarize, the MOSCEM algorithm takes an initial population of points, randomly spread out in the feasible parameter space. For each individual of the population the multiobjective vector F is computed and the population is ranked and sorted using an improved version of the fitness assignment concept developed by Zitzler and Thiele [1999]. The population is partitioned into several complexes and, in each complex k ($k = 1, 2, \dots, q$), a parallel sequence is launched starting from the point that exhibits the highest fitness. A new candidate point in each sequence k is generated using a multivariate normal distribution centered around the current draw of sequence (k) augmented

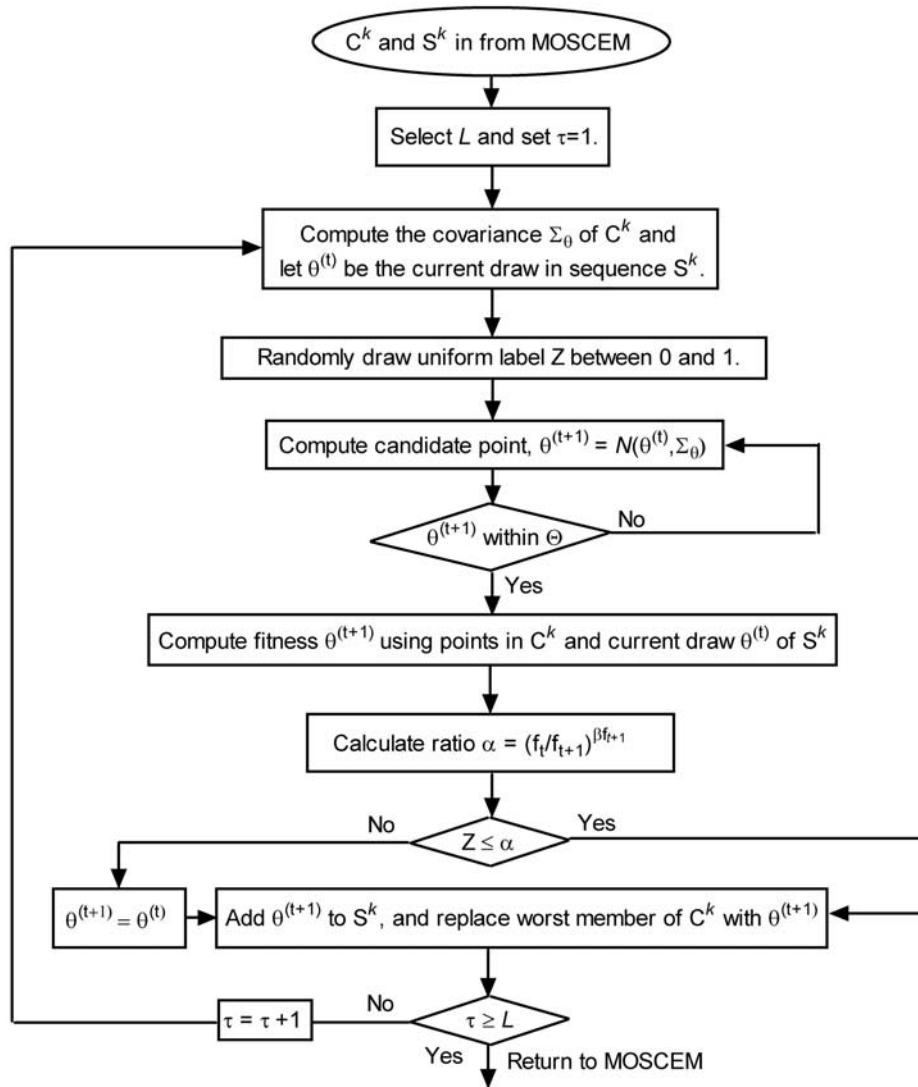


Figure 4. Flowchart of the Sequence Evolution Metropolis (SEM) algorithm employed in the MOSCEM-UA algorithm.

with the covariance structure induced between the points in complex k . A Metropolis-type of acceptance rule is used to test whether the offspring should be added to the current sequence or not. If the offspring (candidate point) is accepted, it replaces the worst member of the current complex k . However, if the candidate point is rejected, the worst member of C^k is replaced with the last member of S^k . Finally, after a prescribed number of iterations, the complexes are replaced into the fixed population of points and new complexes are formed through a process of shuffling. Iterative application of the various algorithmic steps causes the population to converge toward the Pareto set of solutions.

[30] The newly developed MOSCEM algorithm differs from the original MOCOM algorithm in three essential ways. These modifications prevent premature convergence of the algorithm to an indistinguishable region surrounding the Pareto set and help to avoid clustering of solutions in the most compromised region among the objectives. In the first place, the MOSCEM algorithm uses an improved fitness assignment method, which preserves the diversity of the population, whereas the MOCOM algorithm uses the stan-

dard Pareto ranking concept introduced by *Goldberg* [1989]. In the second place, the multiobjective downhill simplex method used by the MOCOM algorithm is replaced with a probabilistic covariance-annealing search method, which is well-suited to deal with the strong correlation structures between the parameters in the Pareto set that are typically encountered in hydrologic modeling. Moreover, the stochastic nature of the annealing scheme prevents the collapse of the MOSCEM algorithm into a relatively small region of some single “best” parameter set, thereby further preserving diversity of the sampled population and enabling the algorithm to generate a fairly uniform approximation of the Pareto front. Finally, the MOSCEM algorithm uses the strengths of the shuffling procedure and complex partitioning employed in the single-objective SCE-UA global optimization algorithm [*Duan et al.*, 1992, 1993] to conduct an efficient search of the parameter space.

[31] The MOSCEM algorithm has four algorithmic parameters that must be specified by the user: the population size (s), the number of complexes-sequences (q), which in turn also determine the number of points within each

complex ($m = s/q$), the number of evolutionary steps in each complex before reshuffling (L), and the scaling factor (β) that directly determines the acceptance probability of the generated offspring. The version of the MOSCEM-UA algorithm used for the optimizations reported in this paper used the values of $L = (m/4)$ and $\beta = (1/2)$. Therefore the only variables that need to be specified by the user are the population size s and the number of complexes p . In the first case study, we will pay special attention to the sensitivity of the performance of the MOSCEM algorithm to the algorithmic parameters s and p .

3.4. Performance Criteria

[32] In multiobjective optimization, the definition of performance is substantially more complex than for single-objective optimization problems, because the optimization goal itself consists of various subgoals: (1) The distance of the nondominated solution set to the Pareto-optimal front should be minimized. (2) A uniform distribution of the solutions along the Pareto front is desirable. (3) The Pareto solution set should cover the full trade-off range of the various objectives, thereby including the single-objective solutions, which represent the theoretical ends of the Pareto frontier. In the literature, attempts can be found that try to formalize the above objectives by means of quantitative measures. However, in this study, we used a visual comparison of the results obtained using the MOCOM and MOSCEM algorithms to evaluate performance in terms of the three factors mentioned above.

3.5. Initial Sampling Distribution

[33] The purpose of the initial (prior) sampling distribution is to quantify the knowledge which is available before collecting and processing any data about the location of the Pareto solution set in the parameter space. If the initial sampling distribution is selected in order to closely approximate the true joint distribution of the parameters associated with the “true” Pareto set of solutions, the MOSCEM algorithm will very rapidly generate a set of nondominated solutions that closely approximates this “true” Pareto set. However, in the case of hydrologic models, usually very little a priori knowledge is available about the location of the Pareto set in the parameter space. Consequently, if the initial sampling distribution is chosen to express this high level of initial uncertainty (for example, *Beven and Binley [1992]* suggested a uniform distribution over a large rectangle of parameter values), the rate of convergence of the algorithm to the final Pareto set of solutions will tend to be slow.

[34] To explicitly address the influence of the prior sampling distribution on the effectiveness and computational efficiency of the MOSCEM algorithm for constructing an estimate of the Pareto solution set, we conducted two different experiments (case studies 2 and 3). In the first experiment, we assumed that there was no prior information available about the location of the Pareto solution set in the parameter space. Accordingly, a uniform prior distribution over the pre-specified upper and lower bounds for each of the model parameters was used to initialize the MOSCEM algorithm (step 2). In the second experiment, we initialized the MOSCEM algorithm using approximate prior information about the location and structure induced in the joint distribution of the parameters in the Pareto set of solutions.

Such a prior sampling distribution was approximated with the following steps.

[35] 1. Use each of the M objective functions, F_i , involved in the multicriteria framework ($i = 1, 2, \dots, M$) to separately locate the best attainable parameter values ($\theta_{i,opt}$) using the SCE-UA global optimization algorithm [*Duan et al., 1992, 1993*].

[36] 2. Use a traditional first-order approximation to estimate the multivariate posterior joint probability density function, $p(\theta_i|\mathbf{y})$, at each of the solutions i :

$$p(\theta_i|\mathbf{y}) \propto \exp\left[-\frac{1}{2\sigma^2}(\theta - \theta_{i,opt})^T X^T X (\theta - \theta_{i,opt})\right] \quad (7)$$

where σ is the root mean square error (RMSE) of the fit at the final solution, and X is the Jacobian or sensitivity matrix evaluated at $\theta_{i,opt}$.

[37] 3. For each of the objectives under consideration, generate s/M points using the multivariate posterior joint probability distribution specified in equation (7). The initial population of points for the MOSCEM algorithm now constitutes the total of s/M generated samples corresponding to each objective i .

[38] Empirical investigations reported in this paper reveal that the latter approach, of first approximating the structure induced in the joint distributions of the Pareto solutions followed by initializing the MOSCEM algorithm with this distribution, is computationally very efficient and, because the theoretical ends of the Pareto frontier are reflected in the prior, helps to reduce the typical MO algorithm problems reported above.

4. Case Studies

[39] We compare the power and applicability of the original MOCOM and MOSCEM algorithm for three case studies with increasing complexity. The first is a standard mathematical case study using a simple two-dimensional multiobjective (MO) test problem. This illustrates the concepts of indistinguishability and demonstrates the ability of each algorithm to infer the known location of the Pareto optimal set. The second and third case studies explore the relative effectiveness and efficiency of the MOCOM and MOSCEM algorithms for two multicriteria calibration hydrological model applications; the Sacramento Soil Moisture Accounting model (SAC-SMA) conceptual watershed model, and the Biosphere Atmosphere Transfer Scheme (BATS) land surface model. In these case studies, we are especially concerned with the robustness of the MOCOM and MOSCEM algorithms by comparing the results of the multiobjective optimization with individual single-criterion solutions obtained using the SCE-UA global optimization algorithm developed by *Duan et al. [1992, 1993]*. We also explicitly examine the influence of the initial sampling distribution on the effectiveness and efficiency of the MOSCEM algorithm.

4.1. Case Study I: A Simple Two-Dimensional MO Problem

[40] The applicability of the MOCOM and MOSCEM algorithms for generating an approximation of the Pareto set was examined using a simple mathematical test problem for which the exact location and shape of the Pareto set is

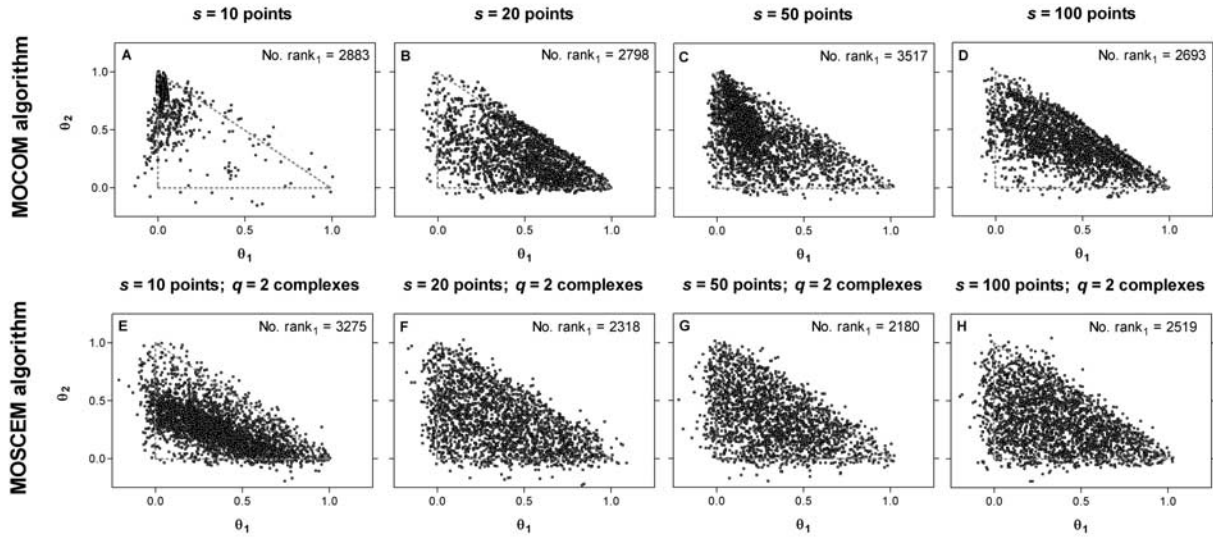


Figure 5. Scatterplots of the final set of rank 1 points after performing 5000 function evaluations with the (a–d) MOCOM and (e–g) MOSCEM algorithm using a population size (s) of 10, 20, 50, and 100 points for a two-dimensional mathematical test problem having three objectives.

known and can be easily computed using geometry. Consider the following two-dimensional MO test problem:

$$\min_{\theta \in \Theta} F(\theta) \left\{ \theta_1^2 + \theta_2^2, (\theta_1 - 1)^2 + \theta_2^2, \theta_1^2 + (\theta_2 - 1)^2 \right\} \quad (8)$$

The Pareto solution set corresponding to equation (8) consists of a triangular-shaped area in the parameter space, having the corner points (0, 0)–(0, 1) and (1, 0) for θ_1 and θ_2 , respectively.

[41] Figure 5 presents scatterplots of the final rank 1 points after performing 5000 function evaluations with the MOCOM (Figures 5a–5d) and MOSCEM (Figures 5e–5g) algorithms using population sizes (s) of 10, 20, 50, and 100 points, respectively. In general, the MOCOM- and MOSCEM-generated points are consistent with the Pareto target distribution of points defined in equation (8) and are indicated by the dashed gray lines in Figure 5. Note, however, that the MOCOM algorithm does not generate a uniform sampling of the Pareto set of solutions, but has the tendency to cluster the sampled points around a single mode, even with increasing population size. On the contrary, the MOSCEM algorithm maintains a uniform sampling density within the Pareto set of solutions, and the results are relatively insensitive to the specified population size and number of complexes (see also Table 1). Hence a more diverse population yields better estimates of the final statistical moments of the Pareto distribution of points and as such is an additional advantage of the MOSCEM algorithm over the MOCOM algorithm.

4.2. Case Study II: The Sacramento Soil Moisture Accounting Model

[42] We compare the effectiveness and efficiency of the MOCOM and MOSCEM algorithms by means of a case study involving calibration of the Sacramento soil moisture accounting (SAC-SMA) model using data from the Leaf River watershed (1950 km²) near Collins, Mississippi. The

SAC-SMA model is used by the National Weather Service (NWS) for flood forecasting throughout the United States and has 16 parameters that need to be specified by the user (see Table 2). While a few of these parameters might be estimated by relating them to observable watershed characteristics, most of the parameters are abstract conceptual representations of the watershed and must be estimated through calibration. On the basis of a recommendation by Peck [1976], the parameters SIDE, RIVA, and RSERV were fixed at prespecified values. The remaining 13 parameters were selected for the multicriteria optimization, and the feasible parameter space was defined by fixing the upper and lower bounds at their “level zero” estimates presented by Boyle *et al.* [2000].

[43] The data, obtained from the National Weather Service Hydrology Laboratory (HL), consist of mean areal precipitation (mm/day), potential evapotranspiration (mm/day), and streamflow (m³/s). Because the SAC-SMA and Leaf River data have been discussed extensively in previous work [see, e.g., Sorooshian *et al.*, 1993; Duan *et al.*, 1993, 1994; Yapo *et al.*, 1998], the details of these will not be reported here. In keeping with previous multicriteria studies [Boyle *et al.*, 2000], approximately 10 years (28 July 1952 to

Table 1. Total Number of Rank 1 Points After Performing 5000 Function Evaluations With the MOSCEM Algorithm as a Function of the Population size s and the Number of Complexes q

Number of Complexes q	Population Size s				
	10	20	50	100	250
1	1789	1970	2282	2728	2515
2	2931	1934	2047	2327	2484
5		1915	1740	1846	2085
10			1582	1856	1885
20				1572	
50					1384

Table 2. Parameters of the SAC-SMA Model, With Their Initial Uncertainty and Multicriteria Calibrated Range Using the MOCOM and MOSCEM Algorithm

Parameter	Description	Units	“Level Zero” Range	MOCOM	MOSCEM
<i>Capacity Thresholds</i>					
1	UZTWM upper zone tension water maximum storage	[mm]	1.0–150.0	4.01–5.20	3.52–9.74
2	UZFWM upper zone free water maximum storage	[mm]	1.0–150.0	25.37–36.53	4.37–34.02
3	LZTWM lower zone tension water maximum storage	[mm]	1.0–500.0	284.32–302.42	241.90–302.48
4	LZFPM lower zone free water primary maximum storage	[mm]	1.0–1000.0	56.05–101.69	59.26–115.15
5	LZFSM lower zone free water supplemental maximum storage	[mm]	1.0–1000.0	13.18–19.10	12.68–27.44
6	ADIMP additional impervious area	[-]	0.0–0.40	0.35–0.36	0.31–0.40
<i>Recession Parameters</i>					
7	UZK upper zone free water lateral depletion rate	[day ⁻¹]	0.1–0.5	0.15–0.24	0.10–0.45
8	LZPK lower zone primary free water depletion rate	[day ⁻¹]	0.0001–0.025	0.01–0.02	0.01–0.02
9	LZSK lower zone supplemental free water depletion rate	[day ⁻¹]	0.01–0.25	0.22–0.25	0.22–0.25
<i>Percolation and Other</i>					
10	ZPERC maximum percolation rate	[-]	1.0–250.0	222.75–240.98	58.46–242.59
11	REXP exponent of the percolation equation	[-]	0.0–5.0	1.08–1.66	0.07–1.62
12	PCTIM impervious fraction of the watershed area	[-]	0.0–0.1	0.00–0.00	0.00–0.01
13	PFREE fraction percolating from upper to lower zone free water storage	[-]	0.0–0.1	0.00–0.00	0.00–0.02
<i>Not Optimized</i>					
14	RIVA riparian vegetation area	[-]	0.0		
15	SIDE ratio of deep recharge to channel base flow	[-]	0.0		
16	RSERV fraction lower zone free water not transferable to tension water	[-]	0.3		

30 September 1962) of historical hydrological data were used for model calibration. To reduce sensitivity to state-value initialization errors, a 65-day warm-up period was used.

[44] Because any conceptual watershed model will, in general, be unable to match all of the different aspects of the watershed’s behavior observed in the measured hydrograph, we follow a method similar to *Boyle et al.* [2000] and partition the hydrograph into a driven (D) and nondriven (ND) part, based on information from the measured hydrograph. A pair of root mean square error (RMSE) objective functions were computed, F_D to measure the ability of the model to simulate the driven portion of the hydrograph, and F_{ND} to measure the ability of the model to simulate the nondriven part of the hydrograph. The Pareto optimal solution space for the two criteria was estimated using a population size of 500 points and 100,000 trials with the MOCOM and MOSCEM algorithms. The results of this two-criteria $\{F_{ND}, F_D\}$ calibration are summarized in Figures 6, 7, and 8.

[45] Figure 6 presents normalized parameter plots for each of the parameters of the SAC-SMA model using either MOCOM algorithm (Figure 6a), the MOSCEM algorithm with uniform prior sampling on the feasible parameter space (Figure 6b), or the MOSCEM algorithm utilizing prior sampling information (Figure 6c). The 13 SAC-SMA model parameters are listed along the x axis, while the y axis corresponds to the parameter values scaled according to their prior uncertainty ranges (defined in Table 2) to yield normalized ranges. Each line across the graph represents one parameter set. The solid and dashed black lines going from left to right across the plots correspond to the single-objective solutions of F_D and F_{ND} obtained by separately fitting to each criterion using the SCE-UA global optimization algorithm [Duan et al., 1992], while the gray lines denote members of the Pareto set of solutions. The objective function plots on the right-hand side in Figure 6 depict two-dimensional projections of the bicriterion trade-off

surfaces represented by the Pareto set of solutions. Additionally, Table 2 lists the Pareto uncertainty intervals of the parameters estimated with the MOCOM and MOSCEM algorithms in the nontransformed parameter space. Figure 6a clearly illustrates that the MOCOM algorithm has generated a fairly uniform approximation of the Pareto frontier only in the most compromised region among the objectives and does not represent the extreme points of the Pareto frontier well. This clustering of solutions in the middle region of the Pareto frontier is also demonstrated in the estimated Pareto uncertainty intervals of the parameters, which for seven of the 13 SAC-SMA parameters (UZFWM, UZK, ADIMP, ZPERC, REXP, LZTWM, and LZPK) does not bracket the single-objective solutions (Figure 6a). On the contrary, the MOSCEM algorithm generates a fairly uniform approximation of the “true” Pareto solution set, which contains the single-criterion solutions at the extreme ends of the Pareto frontier. This is especially true when using prior information about the location of the Pareto solution set in the parameter space in the sampling strategy (Figure 6c). As a consequence, the estimated Pareto uncertainty intervals of the SAC-SMA parameters contain the single-criterion SCE solutions. Notice that, for most of the parameters, the Pareto solution set tends to cluster closely in the parameter space for the two objectives. However, there is considerable uncertainty associated with the percolation parameter ZPERC and the recession parameters UZK and LZPK in the SAC-SMA model, which play a major role in determining the shape of the hydrograph during recession periods. Also notice the close match between the Pareto uncertainty intervals estimated with the MOSCEM algorithm as illustrated in Figures 6b and 6c.

[46] To further illustrate the advantage of using prior information about the location of the Pareto solution set in the parameter space in the initial sampling with the MOSCEM algorithm, consider Figure 7, which presents the evolution of the bicriterion trade-off surface in the two-

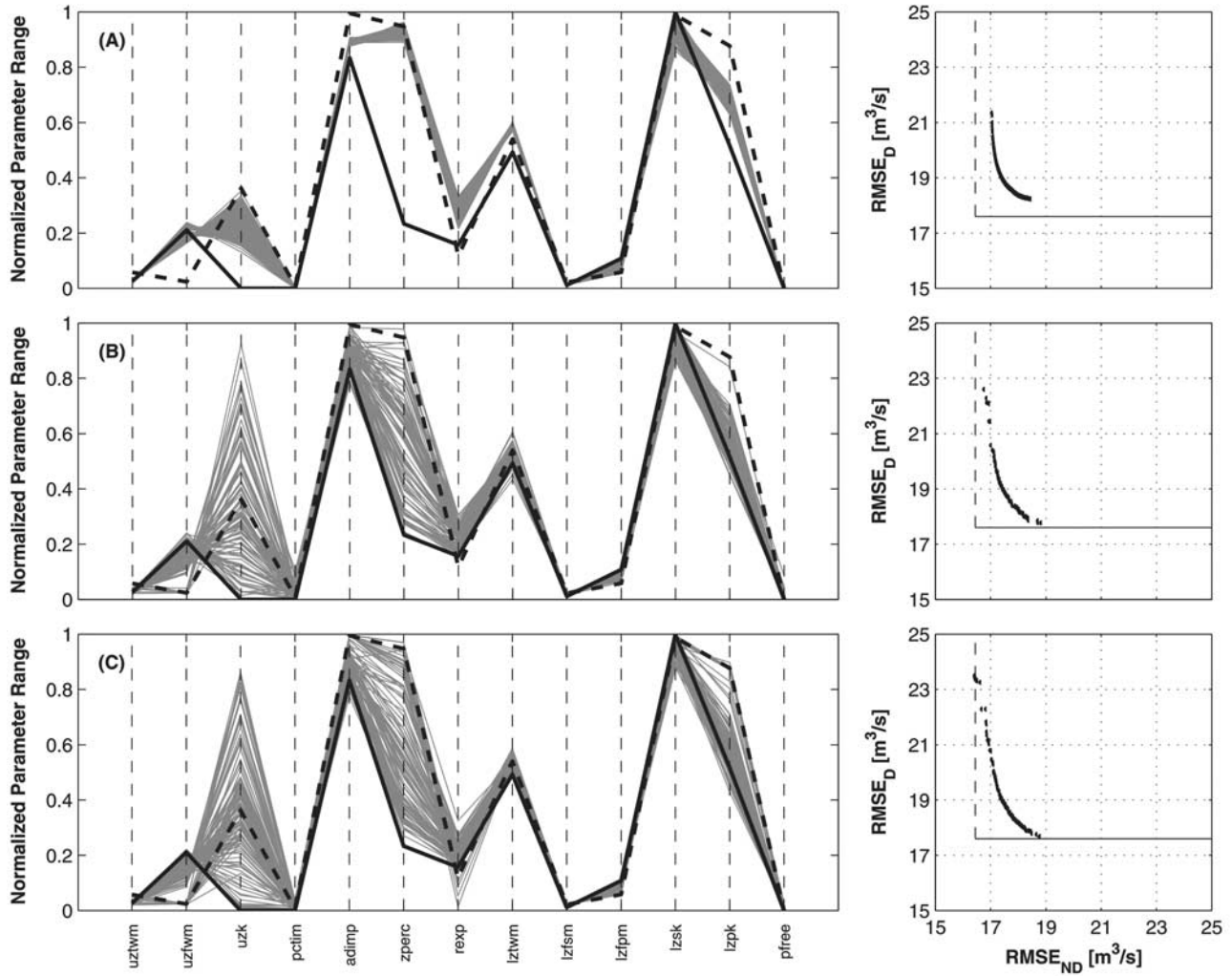


Figure 6. Normalized parameter plots for each of the SAC-SMA model parameters using a two-criteria $\{F_{ND}, F_D\}$ calibration with the (a) MOCOM, (b) MOSCEM with no prior information, and (c) MOSCEM with prior information algorithm. Each line across the graph denotes a single parameter set: shaded is Pareto solution set; solid and dashed lines are single-criterion solutions of F_D and F_{ND} , respectively. The squared plots at the right-hand side are two-dimensional projections of the objective space of the Pareto set of solutions. See color version of this figure in the HTML.

dimensional objective space as a function of the number of SAC-SMA model evaluations.

[47] When utilizing a uniform initial sampling of the feasible parameter space (no prior information) with the MOSCEM algorithm, typically 30,000 SAC-SMA model evaluations are needed to construct an estimate of the Pareto solution set. The benefit of going from 30,000 to 50,000 model evaluations can be considered to be marginal given the extra cost in terms of model runs. When prior information is used, 10,000 independent model evaluations with the SCE-UA algorithm are first needed (separately) for each of the two objectives to identify the single-criterion ends of the Pareto frontier. Notice that the initial population of points, stemming from this approach (outlined in section 3.5) using the single-criterion ends of the Pareto cluster, directly approximates the Pareto solution set as depicted with the upper triangular symbols in the two-dimensional bicriterion plot in Figure 7b. This suggests that the structure induced in the joint distribution of the parameters in the Pareto solution

set can be well approximated using information from the single-criterion ends of the Pareto frontier. Clearly, the use of prior information in the initial sampling with the MOSCEM algorithm is not only computationally more efficient, but also generates a more uniform estimate of the Pareto frontier which includes the single-criterion SCE solutions.

[48] The hydrograph uncertainty ranges (shaded area) associated with the Pareto solution set estimated with the MOSCEM algorithm for a 500-day portion of the calibration period are displayed in Figure 8 using a logarithmic transformation of the streamflows.

[49] The observed streamflows are indicated with dots, while the single-criterion solutions for the driven and nondriven portions of the hydrograph are indicated with the solid and dashed black lines, respectively. Note that the streamflow prediction uncertainty ranges match the medium- and high-flow events very well, but do not bracket the observations and display bias (systematic error) on the long recessions, suggesting that the model structure may be

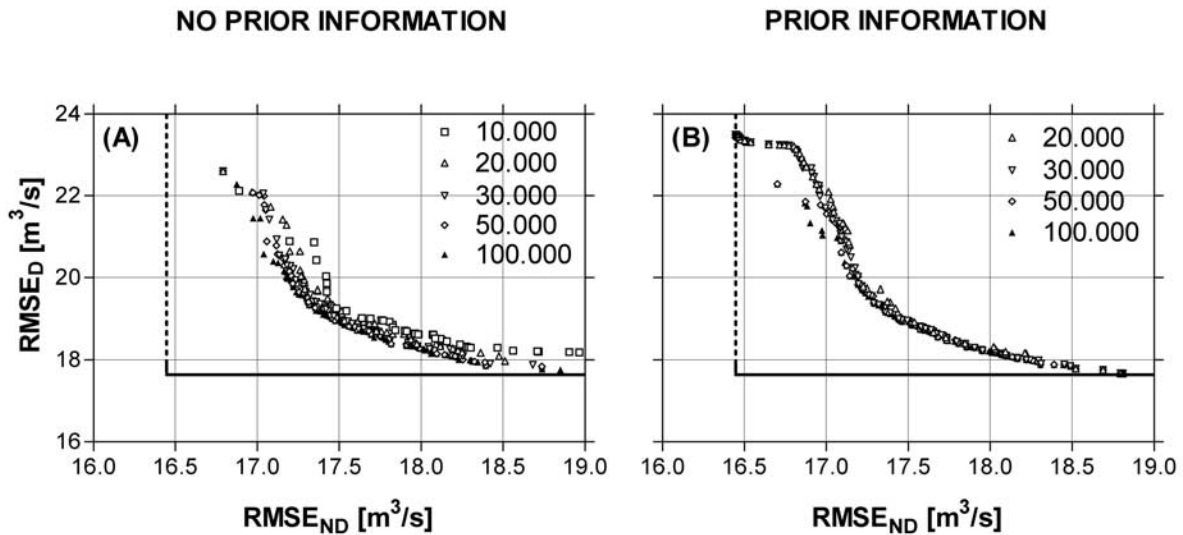


Figure 7. Evolution of the bicriterion trade-off surface in the two-dimensional objective space as a function of the number of SAC-SMA model evaluations using the MOSCEM algorithm (a) without and (b) with prior information about the location of the Pareto solution set in the parameter space.

in need of further improvement. Notice that the relatively large uncertainty found during low flow and recession periods is consistent with the relatively large uncertainty in the UZK and LZPK parameters.

4.3. Case Study III: The Biosphere Atmosphere Transfer Scheme (BATS) Model

[50] The third case study illustrates the power of the MOSCEM algorithm to perform a multicriteria $\{H, \lambda E\}$

calibration of the biosphere-atmosphere transfer scheme (BATS) land-surface model [Dickinson *et al.*, 1993] using measured sensible (H) and latent heat (λE) fluxes from the Oklahoma ARM-CART site. BATS is a conceptual parameterization of the ecohydrological processes at the scale of individual plots of vegetation (50–1000 m). The model consists of six interacting hydrometeorological components (three layers of soil, a canopy air component, a canopy leaf-stem component, and a snow covered portion)

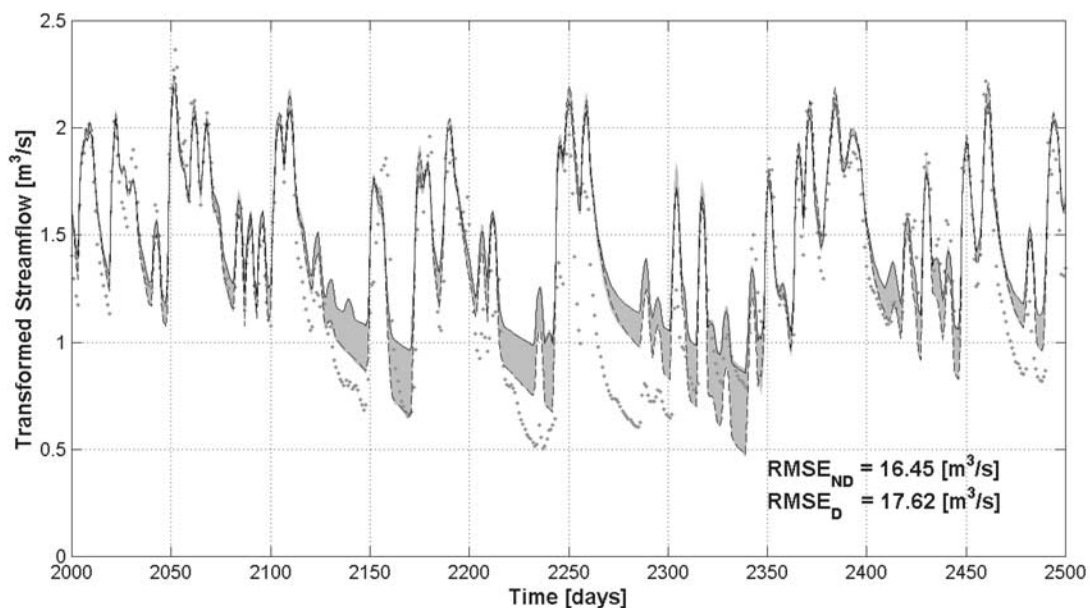


Figure 8. Hydrograph prediction uncertainty ranges (shaded area) associated with the Pareto solution set estimated with the MOSCEM algorithm for a 500-day portion of the calibration period. The solid circles correspond to the observed streamflow data; the solid line corresponds to the minimal F_D solution, and the dashed curve corresponds to the minimal F_{ND} solution. See color version of this figure in the HTML.

Table 3. Parameters of the BATS Model, Including Their Initial Uncertainty and Final Multicriteria Calibrated Range Obtained With the MOSCEM Algorithm

Parameter	Description	Units	Initial Reasonable Range	Multicriteria Range	
<i>Parameters Associated With Vegetation (18 Vegetation Types)</i>					
1	veg	vegetation cover	[-]	0.40–0.95	0.40–0.95
2	seasf	difference between vegc and fractional cover at 269 °K	[-]	0.00–0.80	0.00–0.79
3	rough	aerodynamic roughness length	[m]	0.01–0.35	0.01–0.03
4	displa	displacement height	[m]	0.05–1.50	0.06–1.48
5	rsmin	minimum stomatal resistance	[s/m]	50.0–200.0	50.01–199.94
6	0xla	maximum leaf area index	[-]	0.05–6.00	0.06–5.94
7	xlai0	minimum leaf area index	[-]	0.05–4.00	0.05–3.99
8	sai	stem area index	[-]	1.00–3.00	1.11–2.98
9	aqrt	inverse square root of leaf dimension	[mm ^{-0.5}]	5.00–10.00	5.00–9.99
10	fc	light dependence of stomatal resistance	[m ² /W]	0.02–0.06	0.02–0.06
11	depuv	depth of topsoil layer	[m]	0.01–0.20	0.01–0.20
12	deprv	depth of root zone layer	[m]	0.50–2.00	0.50–1.98
13	deptv	depth of total zone layer	[m]	5.00–10.00	5.00–9.88
14	albvg	vegetation albedo for shortwave <0.7 μm	[-]	0.10–0.20	0.10–0.19
15	albvg	vegetation albedo for longwave >0.7 μm	[-]	0.20–0.40	0.20–0.40
16	rootf	ratio of roots in upper layer to roots in root layer	[-]	0.10–0.90	0.10–0.89
<i>Parameters Associated With Soil Texture (12 Textures)</i>					
17	xmopor	porosity	[-]	0.33–0.66	0.33–0.66
18	xmosuc	minimum soil suction	[mm]	30.0–200.0	30.08–196.25
19	xmohyd	maximum hydraulic conductivity	[mm/s]	0.0008–0.0100	0.00–0.01
20	xmowil	water content at which permanent wilting point occurs	[-] ^a		
21	xmofc	ratio of field capacity to saturated water content	[-] ^a		
22	bee	Clapp and Hornberger “b” parameter	[-]	3.50–10.80	3.50–10.79
23	skrat	ratio of soil thermal conductivity to that of loam	[-]	0.70–1.70	0.70–1.70
<i>Parameters Associated With Soil Color (8 Colors)</i>					
24	solour	soil albedo for different colored soils	[-]	0.05–0.12	0.05–0.12
<i>Initial Conditions</i>					
25	ssw	surface zone water content	[m]	0.0–0.2	0.00–0.20
26	rsw	root zone water content	[m]	0.0–5.0	0.00–4.97
27	tsw	total zone water content	[m]	0.0–10.0	0.17–9.99

^aHere *xmowil* is function of *xmohyd* and *xmosuc*, while *xmofc* is only used when land cover is assigned semidesert.

and has 27 parameters to be estimated, including 16 related to vegetation properties and eight related to soil properties, together with three initial soil moisture conditions (see Table 3). Two of the parameters, *xmowil* (the wilting point) and *xmofc* (the ratio of field capacity to the saturated water content) are actually not independent parameters. The parameter *xmowil* is computed as a function of the hydraulic conductivity (*xmohyd*) and the minimum soil suction (*xmosuc*), while *xmofc* is used only when the land cover is assigned to be semidesert [Dickinson et al., 1993; Gupta et al., 1999].

[51] Land Surface Models (LSMs), like BATS, differ from hydrologic watershed models, such as the SAC-SMA model used in the previous case study, in that they are concerned with both water and energy balance (and more recently carbon and other fluxes): they are driven by multiple input variables (precipitation, short-wave and long-wave radiation, wind speed, air temperature, humidity, etc.), and they present the evolution of several observable state variables (soil temperature, surface soil moisture content, etc.) and output fluxes (latent heat, sensible heat, runoff, etc.) [Bastidas et al., 1999]. The hydrometeorological data set used in this study correspond to station E13 of the Atmospheric Radiation Measurement Cloud and Radiation Testbeds (ARM-CART) program in the Southern Great Plains site (SGP) in Oklahoma. The data cover the 5-month period from 1 April to 25 August 1995, with a sampling interval of 30 minutes, and include all of the necessary

atmospheric forcing for the model and observational information on sensible heat (H in W/m^2) and latent heat fluxes (λE in W/m^2). It has been previously established that the MOCOM algorithm fails to converge for this data set with this particular combination $\{H, \lambda E\}$ of objectives [Bastidas, 1998]. For each of the two criteria, the simulation error was measured using the RMSE statistic. For more information about the BATS model, the hydrometeorological data, and multicriteria calibration approaches applied to the BATS model, please refer to Dickinson et al. [1993], Bastidas [1998], Bastidas et al. [1999], and Gupta et al. [1999]. The Pareto optimal solution space for the two criteria was estimated using a population size of 2,000 points in combination with 100,000 trials with the MOSCEM algorithm. The results of the two-criteria $\{H, \lambda E\}$ calibration are summarized in Figures 9 and 10 and discussed below.

[52] Figure 9 presents the results for the two-criteria $\{H, \lambda E\}$ calibration with the MOSCEM algorithm in the normalized parameter and objective space for three different cases. In the first case (Figure 9a), a uniform prior sampling of the feasible parameter space was used, whereas in the second and third cases, each end ($\{H\}$ and $\{\lambda E\}$) of the Pareto frontier was first identified by single-objective optimization using the SCE-UA algorithm [Duan et al., 1992] (second case, Figure 9b) or the SCEM-UA algorithm [Vrugt et al., 2003] (third case, Figure 9c), and used to initialize the prior distribution for the MOSCEM algorithm.

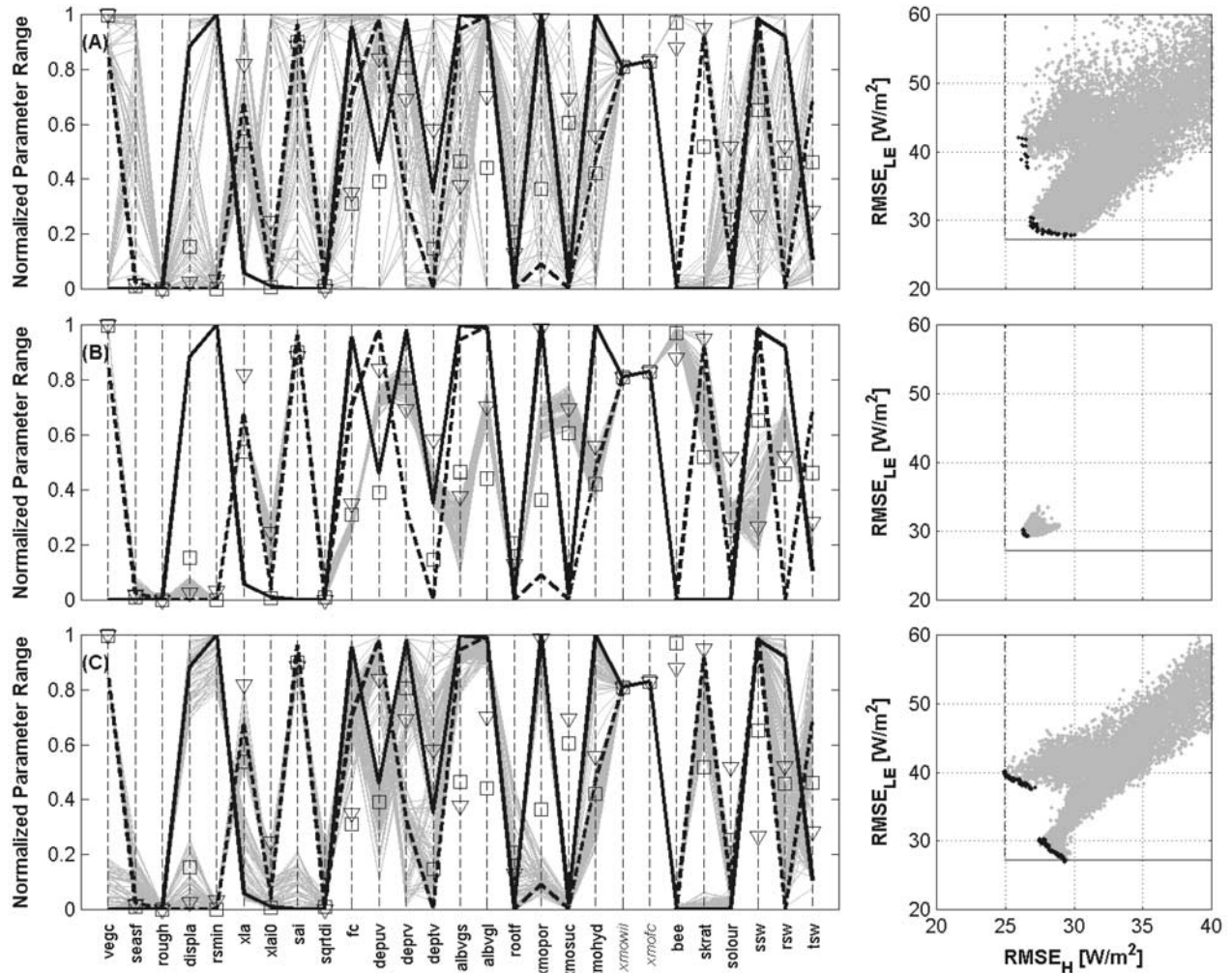


Figure 9. Normalized parameter plots for the ARM-CART site using a two-criteria $\{H, \lambda E\}$ calibration with the (a) MOSCEM with no prior information, (b) MOSCEM with prior SCE information, and (c) MOSCEM with prior SCEM information algorithm. Each line across the graph denotes a single parameter set: solid and dashed lines are single-criterion SCEM solutions of λE and H , respectively, and shaded lines are Pareto solution set. Squares and triangles denote single-criterion SCE solutions of λE and H , respectively. The squared plots at the right-hand side denote two-dimensional projections of the objective space of the Pareto set of solutions. In these plots, rank 1 solutions are indicated by solid dots. See color version of this figure in the HTML.

Each line going from left to right across the normalized parameter plots corresponds to a different parameter set (the solid and dashed black line denote the $\{\lambda E\}$ and $\{H\}$ single-criterion SCEM solutions, respectively); each gray line denotes a member of the $\{H, \lambda E\}$ Pareto set of solutions, and the square and triangular symbols denote the SCE solutions for $\{H\}$ and $\{\lambda E\}$, respectively. The 24 BATS parameters and three initial soil moisture parameters are listed along the x axis, and the y axis corresponds to the parameter values, normalized by their initial uncertainty ranges, as defined in Table 3. The linear-linear squared shaped plots at the right-hand side in Figure 9 depict two-dimensional projections of the bicriterion trade-off surface represented by the total set of points sampled with the MOSCEM algorithm. The Pareto rank one solutions in these plots are indicated by the black dots.

[53] The results presented in Figure 9 emphasize several important observations. In the first place, notice that the SCEM and MOSCEM single-criterion ($\{H\}$ and $\{\lambda E\}$) solutions are significantly better in terms of their RMSE performance measures (typically 10%) and occupy a different part of the parameter space than their counterparts obtained with the SCE-UA algorithm. To verify the consistency of the results with the SCE-UA algorithm, we ran the algorithm ten different times with an increasing number of complexes. Indeed, the algorithm consistently converged to the same region in the parameter space, well removed from the SCEM-identified global minimum. Consequently, when initialized with prior information obtained with the SCE-UA algorithm, the MOSCEM algorithm fails to converge to the “true” Pareto set as depicted in Figure 9b. In the other cases, however, the

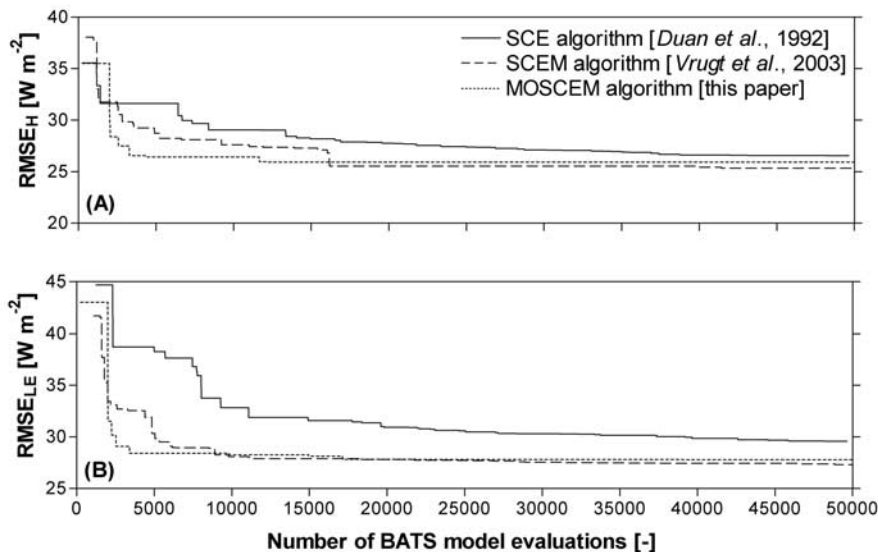


Figure 10. Evolution of the best root mean square error values for (a) sensible and (b) latent heat fluxes as a function of the number of BATS model evaluations with the SCE-UA, SCEM-UA, and MOSCEM-UA optimization algorithms.

MOSCEM algorithm generated a fairly uniform approximation of the Pareto frontier, thereby containing the best attainable single-objective $\{H\}$ and $\{\lambda E\}$ SCEM solutions, denoted with the solid and dashed lines in the squared plots.

[54] The second interesting observation is that the Pareto solution set is discontinuous in the objective space with clusters of solutions close to the single-criterion ends of the Pareto frontier, but with no solutions in the most compromised region among these objectives. This discontinuity is also observed in the normalized parameter plots where, for some of the BATS model parameters (VEGC, RSMIN, XLA, SAI, DEPTV, and SKRAT), two separate well-defined clusters of Pareto solutions can be found close to the single-criterion $\{H\}$ and $\{\lambda E\}$ SCE solutions, while no Pareto solutions are found in the parameter space in between these extreme ends.

[55] Although beyond the scope of this paper, we believe that the convergence problems of the SCE-UA algorithm are caused by the large number of interacting parameters in the BATS model and the highly complex, nonconvex shape of the response surface mapped out in the parameter space. At one end of the spectrum, deterministic search methods, like the Simplex algorithm implemented in the SCE-UA complex evolution strategy, are especially designed for response surfaces that exhibit a well-defined global minimum. At the opposite end, however, probabilistic search methods do not impose constraints on the shape of the response surface and are especially suited to deal with a high degree of randomness in the response surface. We posit that the calibration of the parameters in the BATS model involves a high degree of randomness and nonconvexity in the response surface, thereby causing problems in the identification of the global minimum for classical deterministic search algorithms.

[56] This explanation is also supported by Figure 10, which presents the evolution of the best RMSE values for the $\{H\}$ and $\{\lambda E\}$ criteria, as functions of the number of

BATS model evaluations with the SCE-UA [Duan et al., 1992], SCEM-UA [Vrugt et al., 2003], and MOSCEM-UA (this study) optimization algorithms. The results depicted in Figure 10 show that the SCEM-UA and MOSCEM-UA algorithms converge more quickly and to smaller function values, indicating that the probabilistic covariance-based search method has superior search capabilities over the Simplex search strategy implemented in the SCE-UA global optimization algorithm. While the SCE-UA algorithm requires 75,000 model evaluations to converge to a *sub-optimal* solution, only approximately 15,000 trials with the BATS model are needed with the SCEM-UA and MOSCEM-UA algorithms to identify the minimal $\{H\}$ and $\{\lambda E\}$ objective RMSE solutions. Although not explicitly illustrated here, using this prior information, the pattern of the resulting population evaluated in the two-dimensional $\{H, \lambda E\}$ objective space showed a striking similarity to the “true” Pareto set of solutions illustrated in Figure 9c. This suggests again that the joint distribution of the parameters in the Pareto solution set can be estimated using information obtained from the single-criterion ends of the Pareto frontier. This approach to estimating the prior distribution seems to be both robust and efficient and, because the theoretical ends of the Pareto frontier are computed beforehand, should help to minimize pitfalls that may arise in a wide variety of multiobjective calibration problem applications.

[57] A third significant and interesting observation (demonstrated in Figure 10) is that the multicriteria calibration approach seems to provide superior convergence speed compared to the single-objective (compare the dashed lines in Figures 10a and 10b) approaches. This suggests that minimizing several objectives simultaneously can increase the identifiability of the global minimum in the parameter space, an observation that deserves further investigation in future work.

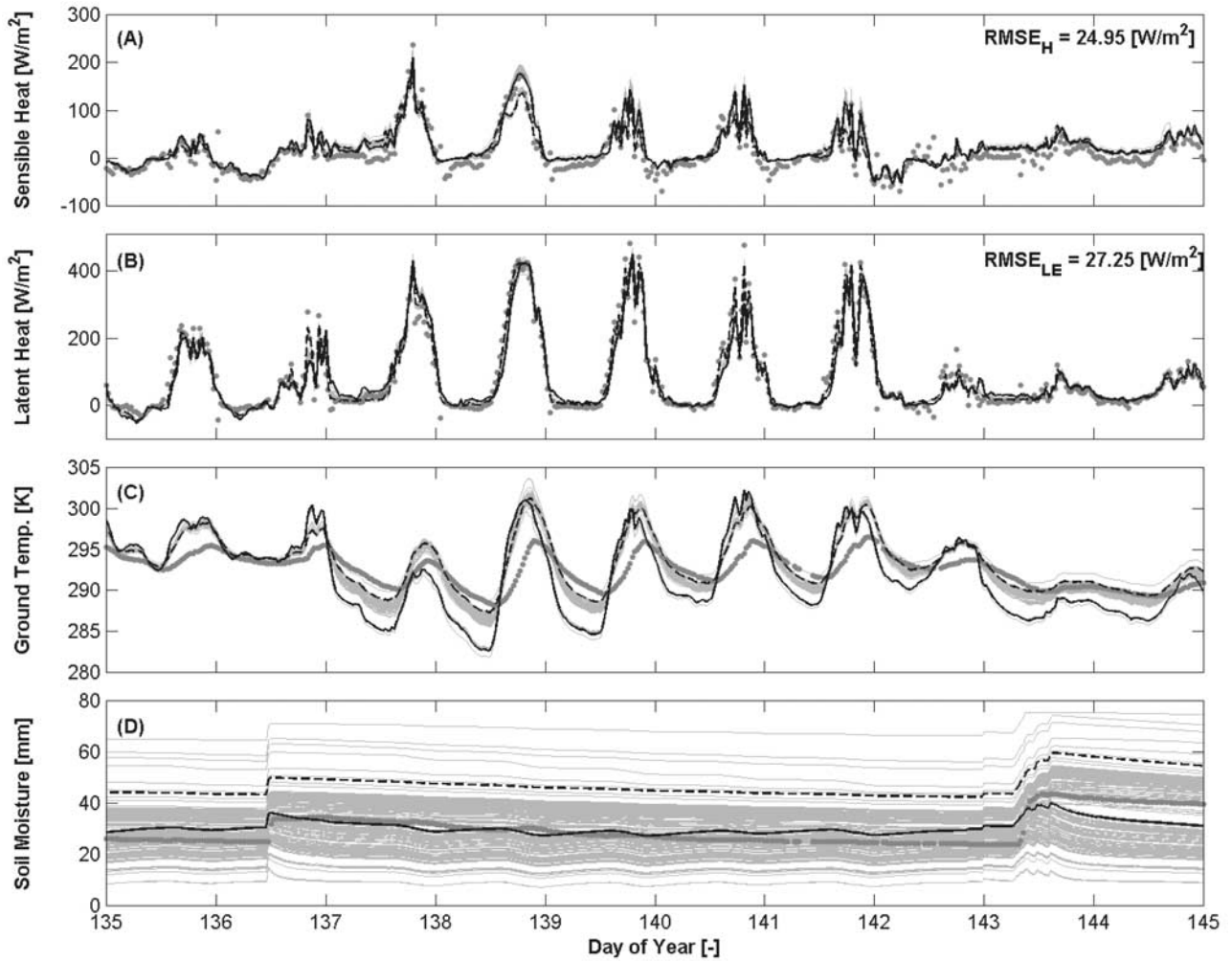


Figure 11. Time series plots of modeled sensible heat flux (W/m^2), latent heat flux (W/m^2), ground temperature (K) and soil moisture content (mm), with the BATS model for a representative 10-day period of the calibration period. The circles denote observed data, the shaded lines represent the Pareto $\{H, \lambda E\}$ set of solutions, and the single-criterion $\{H\}$ and $\{\lambda E\}$ SCEM solutions are indicated by the solid and dashed lines, respectively. See color version of this figure in the HTML.

[58] Finally, Figure 11 shows time series plots of modeled sensible heat flux (W/m^2), latent heat flux (W/m^2), ground temperature ($^{\circ}\text{K}$), and soil moisture content (mm) with the BATS model against the observed data (denoted with circles) for a representative 10-day period of the calibration period, during which rainfall occurred. Each of the Pareto set of solutions corresponding to the $\{H, \lambda E\}$ two-criteria calibration is indicated by a gray line, while the single-criterion $\{H\}$ and $\{\lambda E\}$ SCEM solutions are indicated by the solid and dotted black lines, respectively. The Pareto prediction uncertainty ranges bracket the sensible and latent heat fluxes during most of the time, but do not match the ground temperature (not included in the multicriteria calibration) very well. It is possible that a better match to this state variable could be obtained by including the ground temperature observations in the multicriteria optimization framework. Notice that the BATS model response to the observed sensible and latent heat fluxes is quite similar for each of the two single criterion $\{H\}$ and $\{\lambda E\}$ SCEM solutions (indicated with the solid and dashed lines), while two disconnected regions of model response are found for

the ground temperature, which was not included in the calibration. This clearly implies that, although there are two disconnected Pareto solution set clusters in the parameter space associated with each of the two calibration criteria (see Figure 9), both clusters generate similar model responses in terms of sensible and latent heat fluxes, indicating an interesting model structural issue that deserves further investigation.

5. Conclusions

[59] This paper has presented a Markov Chain Monte Carlo sampler, which is well suited for solving the multicriteria optimization problem for hydrologic models. The sampler, entitled the Multiobjective Shuffled Complex Evolution Metropolis (MOSCEM-UA developed in collaboration between the University of Amsterdam and the University of Arizona), merges the strengths of complex shuffling employed in the SCE-UA algorithm [Duan *et al.*, 1992, 1993] with the probabilistic covariance-based search methodology of the Metropolis algorithm [Metropolis *et al.*,

1953] and an improved fitness assignment concept of Zitzler and Thiele [1999] to construct an efficient and uniform estimate of the Pareto solution set. The MOSCEM algorithm is a multiobjective relative of the SCEM-UA algorithm [Vrugt et al., 2003], originally developed to infer the probabilistic uncertainty associated with the use of a single-objective function, but uses an innovative concept of Pareto dominance rather than direct-objective function evaluations to generate a fairly uniform approximation of the “true” Pareto frontier which includes the single-criteria end points of the Pareto solution set.

[60] The efficiency and effectiveness of the newly developed MOSCEM-UA algorithm for constructing an estimate of the Pareto solution set was compared with the MOCOM-UA algorithm developed by Yapo et al. [1998] for three case studies of increasing complexity. The first case study considered a simple two-dimensional mathematical test problem, while the second and third case studies explored the effectiveness and efficiency of the MOSCEM algorithm for a two-criteria calibration of the Sacramento Soil Moisture Accounting (SAC-SMA) conceptual watershed model and the Biosphere Atmosphere Transfer scheme (BATS) land-surface model. The three case studies clearly demonstrated that the MOCOM algorithm has the tendency to cluster the Pareto solutions in the most compromised region among the objectives, in the third case study, the MOCOM failed to converge. In contrast, the MOSCEM algorithm generates a fairly uniform approximation of the entire Pareto front, which includes the single-criterion end points in the estimated Pareto uncertainty intervals of the parameters. Furthermore, empirical investigations reported in this paper revealed that a strategy of first locating the single-criterion end points of the Pareto frontier, and using this information as a prior estimate of the structure induced in the Pareto solution set of the parameters, is computationally more efficient than imposing a uniform prior distribution on the model parameters during the initialization of the MOSCEM algorithm.

[61] Research aimed at further improvements of the Multiobjective Shuffled Complex Evolution Metropolis approach 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 MOSCEM-UA algorithm is available from the first author at j.vrugt@science.uva.nl.

[62] **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 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, and by NASA grant NAG5-8503-HyDIS (Hydrologic Data and Information System). Special thanks are due to Corrie Thies for proofreading of the manuscript.

References

- Bastidas, L. A., Parameter estimation for hydrometeorological models using multi-criteria methods, Ph.D. dissertation, Dep. of Hydrol. and Water Resour., Univ. of Ariz., Tucson, 1998.
- Bastidas, L. A., H. V. Gupta, S. Sorooshian, W. J. Shuttleworth, and Z. L. Yang, Sensitivity analysis of a land surface scheme using multi-criteria methods, *J. Geophys. Res.*, 104(D16), 19,481–19,490, 1999.
- Beven, K. J., and A. M. Binley, The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Processes*, 6, 279–298, 1992.
- Beven, K. J., and M. Kirkby, A physically based variable contributing area model of basin hydrology, *Hydrol. Sci. Bull.*, 24, 43–69, 1979.
- 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.
- Boyle, D. P., H. V. Gupta, and S. Sorooshian, Toward improved streamflow forecasts: Value of semidistributed modeling, *Water Resour. Res.*, 37(11), 2749–2759, 2001.
- De Grosbois, E., R. P. Hooper, and N. Christopherson, A multi-signal automatic calibration methodology for hydrochemical models: A case study of the Birkenes model, *Water Resour. Res.*, 24(8), 1299–1307, 1988.
- Dickinson, R. E., A. Henderson-Sellers, and P. J. Kennedy, Biosphere-atmosphere transfer scheme (BATS) version 1e as coupled to the NCAR Community Climate Model, *Tech. Note NCAR/TN-387+STR*, 72 pp., Natl. Cent. for Atmos. Res., Boulder, Colo., 1993.
- Duan, Q., V. K. Gupta, and S. Sorooshian, Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, 28(4), 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., S. Sorooshian, and V. K. Gupta, Optimal use of the SCE-UA global optimization method for calibrating watershed models, *J. Hydrol.*, 158, 265–284, 1994.
- Edgeworth, F. Y., *Mathematical Physics*, C. Kegan Paul, London, 1881.
- Fonseca, C. M., and P. J. Fleming, An overview of evolutionary algorithms in multi-objective optimization, *Evol. Comput.*, 3(1), 1–16, 1995.
- Goldberg, D. E., *Genetic Algorithms in Search, Optimization and Machine Learning*, 412 pp., Addison-Wesley-Longman, Reading, Mass., 1989.
- Gupta, V. K., and S. Sorooshian, Calibration of conceptual hydrologic models: Past, present and future, in *Trends in Hydrology: Research Trends*, pp. 329–346, Counc. of Sci. Res. Integration, Trivandrum, India, 1994.
- 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(4), 751–763, 1998.
- Gupta, H. V., L. A. Bastidas, S. Sorooshian, W. J. Shuttleworth, and Z. L. Yang, Parameter estimation of a land surface scheme using multi-criteria methods, *J. Geophys. Res.*, 104(D16), 19,491–19,503, 1999.
- Gupta, H. V., L. Bastidas, J. A. Vrugt, and S. Sorooshian, Multiple criteria global optimization for watershed model calibration, in *Calibration of Watershed Models*, *Water Sci. Appl. Ser.*, vol. 6, edited by Q. Duan et al., pp. 125–132, AGU, Washington, D. C., 2003.
- Holland, J., *Adaptation in Natural and Artificial Systems*, Univ. of Mich. Press, Ann Arbor, 1975.
- Janssen, P. H. M., and P. S. C. Heuberger, Calibration of process-oriented models, *Ecol. Modell.*, 83, 55–66, 1995.
- Kuczera, G., On the relationship between the reliability of parameter estimates and hydrologic time series used in calibration, *Water Resour. Res.*, 18(1), 146–154, 1982.
- Kuczera, G., Improved parameter inference in catchment models: 2. combining different kinds of hydrologic data and testing their compatibility, *Water Resour. Res.*, 19(5), 1163–1172, 1983.
- 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.
- Leplastrier, M., A. J. Pitman, H. V. Gupta, and Y. Xia, Exploring the relationship between complexity and performance in a land surface model using the multi-criteria method, *J. Geophys. Res.*, 107(D20), 4443, doi:10.1029/2001JD000931, 2002.
- Madsen, H., Automatic calibration of a conceptual rainfall-runoff model using multiple objectives, *J. Hydrol.*, 235, 276–288, 2000.
- 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.
- Peck, E. L., Catchment modeling and initial parameter estimation for the National Weather Service river forecast system, *Tech. Memo. NWS Hydro-31*, Natl. Oceanic and Atmos. Admin., Silver Spring, Md., 1976.
- Price, W. L., Global optimization algorithms for a CAD workstation, *J. Optim. Theory Appl.*, 55(1), 133–146, 1987.
- Ritzel, B. J., E. J. Wayland, and S. Ranjithan, Using genetic algorithms to solve a multiple objective groundwater pollution containment problem, *Water Resour. Res.*, 30(5), 1589–1603, 1994.

- Schaffer, J. D., Multiple objective optimization with vector evaluated genetic algorithms, in *Genetic Algorithms and Their Applications: Proceedings of the Third International Conference on Genetic Algorithms*, edited by J. J. Grefenstette, pp. 93–100, Lawrence Erlbaum, Hillsdale, N. J., 1985.
- 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(4), 1185–1194, 1993.
- Srinivas, N., and K. Deb, Multi-objective optimization using non-dominated sorting in genetic algorithm, *Evol. Comput.*, 2(3), 221–248, 1994.
- Vrugt, J. A., H. V. Gupta, W. Bouten, and S. Sorooshian, A shuffled complex evolution Metropolis algorithm for optimization and uncertainty assessment of hydrological model parameters, *Water Resour. Res.*, 39, doi:10.1029/2002WR001642, in press, 2003.
- Wagener, T., D. P. Boyle, M. J. Lees, H. S. Wheeler, H. V. Gupta, and S. Sorooshian, A framework for development and application of hydrological models, *Hydrol. Earth Syst. Sci.*, 5(1), 13–26, 2001.
- Woolhiser, D. A., R. E. Smith, and D. C. Goodrich, A kinematic runoff and erosion manual: Documentation and user manual, *ARS 77*, U.S. Dep. of Agric., Tucson, Ariz., 1990.
- Xia, Y., A. J. Pittman, H. V. Gupta, M. Leplastrier, A. Henderson-Sellers, and L. A. Bastidas, Calibrating a land surface model of varying complexity using multi-criteria methods and the Cabauw dataset, *J. Hydrometeorol.*, 3(2), 181–194, 2002.
- Yapo, P. O., H. V. Gupta, and S. Sorooshian, Multi-objective global optimization for hydrologic models, *J. Hydrol.*, 204, 83–97, 1998.
- Zitzler, E., and L. Thiele, Multi-objective evolutionary algorithms: A comparative case study and the strength Pareto approach, *IEEE Trans. Evol. Comput.*, 3(4), 257–271, 1999.

L. A. Bastidas, Department of Civil and Environmental Engineering, Utah State University, Logan, UT 84322, USA.

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.