WATER RESOURCES RESEARCH, VOL. 35, NO. 3, PAGES 767–773, MARCH 1999

Probabilistic optimization for conceptual rainfall-runoff models: A comparison of the shuffled complex evolution and simulated annealing algorithms

Mark Thyer and George Kuczera

Department of Civil, Surveying and Environmental Engineering, University of Newcastle, Callaghan, New South Wales, Australia

Bryson C. Bates

Cooperative Research Centre for Catchment Hydrology and CSIRO Land and Water, Wembley, Western Australia, Australia

Abstract. Automatic optimization algorithms are used routinely to calibrate conceptual rainfall-runoff (CRR) models. The goal of calibration is to estimate a feasible and unique (global) set of parameter estimates that best fit the observed runoff data. Most if not all optimization algorithms have difficulty in locating the global optimum because of response surfaces that contain multiple local optima with regions of attraction of differing size, discontinuities, and long ridges and valleys. Extensive research has been undertaken to develop efficient and robust global optimization algorithms over the last 10 years. This study compares the performance of two probabilistic global optimization methods: the shuffled complex evolution algorithm SCE-UA, and the three-phase simulated annealing algorithm SA-SX. Both algorithms are used to calibrate two parameter sets of a modified version of *Boughton*'s [1984] SFB model using data from two Australian catchments that have low and high runoff yields. For the reduced, well-identified parameter set the algorithms have a similar efficiency for the low-yielding catchment, but SCE-UA is almost twice as robust. Although the robustness of the algorithms is similar for the high-yielding catchment, SCE-UA is six times more efficient than SA-SX. When fitting the full parameter set the performance of SA-SX deteriorated markedly for both catchments. These results indicated that SCE-UA's use of multiple complexes and shuffling provided a more effective search of the parameter space than SA-SX's single simplex with stochastic step acceptance criterion, especially when the level of parameterization is increased. Examination of the response surface for the low-yielding catchment revealed some reasons why SCE-UA outperformed SA-SX and why probabilistic optimization algorithms can experience difficulty in locating the global optimum.

1. Introduction

Conceptual rainfall-runoff (CRR) models simulate the physical processes that comprise the land phase of the hydrologic cycle. The rainfall-runoff transformation is modeled by a set of transfer functions that links several interconnected conceptual water stores. In most instances, estimates for the model parameters must be obtained by fitting computed to observed hydrographs as direct physical measurement is impossible.

The calibration of CRR models has been researched extensively over the last two decades. Many studies have reported difficulties in finding unique (global) parameter estimates [*Johnston and Pilgrim*, 1976; *Duan et al.*, 1992; *Gan and Biftu*, 1996; *Sorooshian et al.*, 1993]. *Duan et al.* [1992] argue that the majority of automatic calibration algorithms are affected by five key problems: (1) multiple regions of attraction, where the local optimum found depends on the parameter values used to start the algorithm (a region of attraction is a subregion of the parameter space surrounding a local minimum for which application of a deterministic hill-climbing algorithm starting at any point in the subregion will lead to convergence at that minimum); (2) minor local optima, where there are small pits in the response surface (the map of the objective function in the parameter space) within each region of attraction; (3) roughness, when the response surface contains points with discontinuous derivatives; (4) sensitivity, where there is poor and varying model sensitivity to the parameters in the vicinity of an optimum and nonlinear interaction between parameters; and (5) shape, when the response surface is nonconvex and contains long curved ridges.

Paper number 1998WR900058.
0043-1397/99/1998WR900058$09.00

Deterministic algorithms such as the simplex method of *Nelder and Mead* [1965] require manual intervention during the initialization and calibration stages to overcome these problems. This requires a good knowledge of the characteristics of the CRR model [*Gan and Biftu*, 1996]. These problems have led to the use and refinement of probabilistic search algorithms that reduce the need for manual intervention during parameter estimation. These methods can almost guarantee asymptotic convergence to the global optimum as the number of sample points in the parameter space increases [*Corana et al.*, 1987; *Dekkers and Aarts*, 1991].

This study compares the performances of two promising

probabilistic optimization algorithms on the basis of their robustness and efficiency: the shuffled complex evolution algorithm (SCE-UA) of *Duan et al.* [1992, 1994] and the simulated-annealing algorithm (SA-SX) developed by *Bates* [1994]. *Duan et al.* [1992] claim that SCE-UA produces reliable global estimates for large and complex optimization problems. *Gan and Biftu* [1996] compared the performances of SCE-UA, the Nelder-Mead simplex (SX), and the multistart simplex (MSX) methods. They found that MSX was computationally inefficient and that although the SX method was robust enough, it required multiple runs to be effective. In contrast, SCE-UA could consistently locate the global optimum with synthetic "error-free" data, but this finding did not hold for real data. *Kuczera* [1997] found that SCE-UA was more robust and efficient than traditional genetic algorithms. SA-SX is based on the approach described by *Press and Teukolsky* [1991]. *Bates* [1994] showed that the performance of the SA-SX algorithm was superior to that of the MSX algorithm for both error-free and real data. Although *Bates* [1994] and *Sumner et al.* [1997] indicated that SA-SX showed promise, *Bates* [1994] cautioned that it had a high computational cost and that care was needed in the specification of the algorithmic parameters.

The SCE-UA and SA-SX algorithms are compared by calibrating the modified SFB model [*Sumner et al.*, 1997] to the same data sets with the same objective function. The modified SFB model is a relatively simple CRR model with eight parameters and has been applied to many Australian catchments [*Sumner et al.*, 1997; *Ye et al.*, 1997]. Despite the model's simplicity there is considerable evidence that its response surface contains multiple optima. The model is fitted to data for two Australian catchments: Scott Creek at Scott's Bottom (503502), and Allyn River at Halton (210022). These catchments have low and high runoff yields, respectively, and the performance of optimization algorithms will be assessed for both cases and contrasted with each other. Consideration of the search technique used by each of the algorithms will provide possible reasons for any difference in performance. Investigation of the response surface for Scott Creek will be undertaken to provide further insight into the problems faced by probabilistic optimization algorithms.

2. Probabilistic Optimization Algorithms

2.1. Simulated-Annealing Algorithm (SA-SX)

SA-SX is a three-phase algorithm that combines simulated annealing with the simplex method of *Nelder and Mead* [1965]. Simulated annealing was originally proposed by *Kirkpatrick et al.* [1983] as method for minimizing multivariate functions. SA-SX conducts a thorough exploration of the parameter space by randomly perturbing the objective function values at the simplex vertices, thereby enabling the acceptance of uphill steps on the response surface as well as downhill steps. The magnitude of the random perturbation and therefore the probability of accepting an uphill step is determined by a control parameter, T (analogous to temperature in physical annealing processes). An annealing schedule controls the rate of decrement of T. By using a stochastic step acceptance criterion it is possible to escape from local minima when searching for a candidate global optimum. In the course of the optimization process T is decreased slowly, causing a gradual reduction in the probability of accepting an uphill move.

During the first phase, starting values for the simplex are found by uniform sampling from a hypercube that defines the feasible parameter space. Simulated annealing commences in the second phase with an initially large T that is reduced gradually according to a user-specified annealing schedule. When T is reduced by a factor of 3, an occasional restart occurs in which a vertex of the simplex is replaced by the best solution encountered when that solution is not in the simplex. Without this modification the algorithm tends to converge to local optima. As T decreases to its final value (generally $T \rightarrow 0$), the algorithm enters its third phase and reduces to the conventional simplex method to facilitate convergence. More detailed explanations of the SA-SX algorithm are given by *Bates* [1994] and *Sumner et al.* [1997].

2.2. Shuffled Complex Evolution Algorithm (SCE-UA)

The SCE-UA algorithm of *Duan et al.* [1992, 1994] is designed to deal with the peculiarities of parameter estimation in CRR models. SCE-UA is based on four main concepts: (1) a combination of deterministic and probabilistic approaches; (2) systematic evolution of a "complex" of points spanning the parameter space, in the direction of global improvement; (3) competitive evolution; and (4) complex shuffling. These four features represent a synthesis of the best features of several methods [*Gan and Biftu*, 1996].

Initially a random set of points (a "population") is sampled from the parameter space and partitioned into a number of complexes. Each of these complexes is then allowed to evolve using competitive evolution techniques that are based largely on the simplex method. At periodic stages in the evolution, the entire set of points is shuffled and reassigned to new complexes to enable information sharing. As the search progresses, the entire population should move toward the neighborhood of the global optimum, provided the initial population size is sufficiently large.

The combination of competitive evolution and complex shuffling ensures that the information on the parameter space gained by each of the individual complexes is shared throughout the entire population. This results in a robust optimization algorithm that conducts an efficient search of the parameter space.

2.3. Comparison Issues

The performance of an optimization algorithm can be measured by its robustness and efficiency. Here robustness is interpreted as the probability of finding the same (and hopefully global) optimum from a series of independent trials. Efficiency is determined by the number of function evaluations (model runs per optimization trial) required by the algorithm to satisfy prescribed convergence criteria. It is important when undertaking a comparison that the basis be fair and equitable.

SCE-UA and SA-SX have a number of algorithmic parameters that must be tuned to ensure optimal robustness and efficiency. *Duan et al.* [1994] report recommended values for the SCE-UA algorithmic parameters, the most important being the number of complexes that must be set according to the number of fitted parameters. The recommended values are used in this study. For the SA-SX algorithm it is important to choose an annealing schedule that enables the algorithm to make a broad search of the parameter space by accepting a large number of uphill steps but not so large as to produce a totally random walk. The annealing schedules reported by *Sumner et al.* [1997] were used for the model and data sets at hand.

The selection of convergence criteria also affects the fairness of the comparison. SA-SX uses two criteria that must be satisfied simultaneously:

$$\frac{2|S(\theta)^H - S(\theta)^L|}{|S(\theta)^L| + |S(\theta)^H|} < \sqrt{\varepsilon} \tag{1}$$

and

$$\frac{|\theta_j^{(k)} - \theta_j^L|}{\max(|\theta_j^L|, \theta_j^0)} < \sqrt[3]{\varepsilon} \quad j = 1, \cdots, p;\ k = 1, \cdots, p+1 \tag{2}$$

where $S(\theta)$ denotes the objective function value associated with the $p \times 1$ parameter vector θ, the superscripts H and L denote the vertices of the simplex with the highest and lowest $S(\theta)$ values, θ_j^0 is the initial estimated for the jth parameter, $\theta_j^{(k)}$ is the jth parameter for the kth vertex, and ε is the machine epsilon (the smallest positive number such that $1 + \varepsilon > 1$). Criterion (1) ensures that the vertices in the simplex have similar $S(\theta)$ values, while criterion (2) ensures that all the vertices are in the same subregion of the parameter space. Thus (1) and (2) are relatively strict criteria. A single and commonly used convergence criterion is used for SCE-UA:

$$\frac{|S(\theta)_{i+1}^L - S(\theta)_i^L|}{|S(\theta)_i^L|} < \text{tol} \tag{3}$$

where i denotes the iteration number and "tol" is some specified small value (generally 1×10^{-6}). Thus criterion (3) is not as strict as criteria (1) and (2).

It would be ideal to use the same convergence criteria for both algorithms. However, this presents some problems. Application of criteria (1) and (2) to SCE-UA is not practical. SA-SX has only one simplex whereas SCE-UA has a number of complexes, and to run SCE-UA until all complexes have reached a single optimum would greatly increase the number of function evaluations required and compromise the efficiency of the algorithm. Application of criterion (3) to SA-SX may cause the algorithm to stop prematurely during its second phase. Consequently, the comparison is based on the results obtained with the convergence criteria that provide the best performance for each algorithm as well as the results obtained from an exchange of convergence criteria.

Five hundred independent trials were used to compare the performances of the algorithms. Empirical quantile plots of the $S(\theta)$ values and the number of function evaluations required for convergence were used to assess robustness and efficiency.

3. Modified SFB Model

The CRR model used in this study is the modified SFB model described by *Sumner et al.* [1997]. This model, and the original SFB model developed by *Boughton* [1984], have been applied to many Australian catchments [e.g., *Nathan and McMahon*, 1990; *Chiew et al.*, 1993; *Sumner et al.*, 1997; *Ye et al.*, 1997]. Both models require daily rainfall and potential evapotranspiration data to compute monthly streamflow. Eight parameters require specification in the modified SFB model: the surface storage capacity (S); the daily infiltration capacity (F); the base flow factor ($0 \le B \le 1$), which determines the portion of the daily depletion of water in the lower store; the fraction of surface storage capacity that does not drain to the lower store (NDC); the maximum limiting rate of evapotranspiration (E_{MAX}); the lower store depletion factor (DPF); the base flow

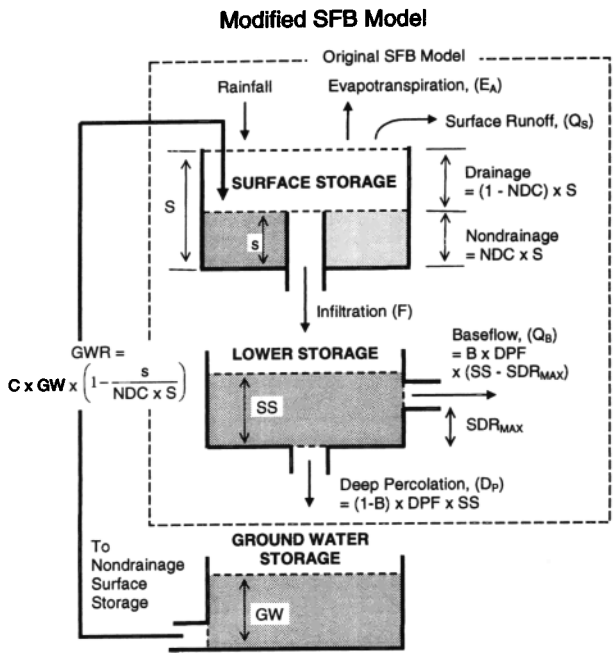

Figure 1. Schematic of the modified SFB model structure.

threshold for the lower store (SDR_{MAX}); and a drainage coefficient (C) to regulate the use of deep percolation to meet evapotranspirative demand in the nondrainage component of the surface store. A schematic of the model is given in Figure 1.

4. Model Calibration

The study catchments were chosen to gauge the performance of the SCE-UA and SA-SX algorithms when different characteristics within the model structure are forced to varying degrees. A warm-up period of 3 months was used to minimize the effects of the initial store contents on the parameter estimates. The objective function used is defined by

$$S(\theta) = \min_{\theta} \sum_{t=4}^{n} z_t^2 \tag{4}$$

where the disturbance z_t is

$$z_t = (Q_t^\lambda - \hat{Q}_t^\lambda) - \phi(Q_{t-1}^\lambda - \hat{Q}_{t-1}^\lambda) \qquad \lambda \neq 0 \tag{5}$$

in which Q_t and $\hat{Q}_t$ denote observed and computed monthly runoff at time t, λ is a transformation constant [*Box and Cox*, 1964], and ϕ denotes the parameter of a first-order autoregressive process.

Two sets of calibrations were considered herein: (1) estimation of the parameter subset used by *Sumner et al.* [1997] and (2) estimation of the entire parameter set. An F test was used to compare their residual sum of squares (RSS_i, $i = 1, 2$). The test statistic for the null hypothesis that the second calibration contributes nothing to the model fit above that contributed by the first calibration is:

$$F = \frac{(RSS_1 - RSS_2)}{(8 - m)\dfrac{RSS_2}{(N - 8 - 3)}} \tag{6}$$

Table 1. Parameter Hypercube Limits for the Modified SFB Model: Scott Creek at Scott's Bottom, Catchment 503502

Parameter	Lower Bound	Upper Bound
S^*	40.0	100.0
F^*	4.0	10.0
B^*	0.0	0.5
NDC	0.0	1.0
E_{MAX}	0.0	10.0
DPF*	0.0	0.1
SDR^*_{MAX}	0.0	100.0
C^*	0.0	2.0

*Parameters estimated by *Sumner et al.* [1997].

where m is the number of parameters estimated by *Sumner et al.* [1997], N is the number of observations and F has an F distribution with $8 - m$ and $N - 8 - 3$ degrees of freedom. This is an approximate test in that the models are nonlinear in a statistical sense and are not nested. However, use of the test facilitated an assessment of the performance of the algorithms when different levels of parameterization are used.

4.1. Scott Creek at Scott's Bottom (503502)

4.1.1. Description of catchment and data. Scott Creek is located within the Onkaparinga River Basin, some 20 km south east of Adelaide, South Australia. The data set covers the period from January 1970 to December 1985 [*Chiew and McMahon*, 1993]. The catchment has a drainage area of 27 km^2, mean annual rainfall of 950 mm, and mean annual runoff of 130 mm. The rainfall/runoff ratio (14%) indicates a low runoff yield. This presents a real challenge in parameter estimation as such catchments have few streamflow events for a given period of record, and hence their data sets contain little information on CRR model parameters [*Ye et al.*, 1997]. Previous studies by *Sumner et al.* [1997] and *Bates* [1994] found many different solutions with $S(\theta)$ values close to a candidate global optimum. Initially, six model parameters (S, F, B, SDR_{MAX}, DPF, and C) were fitted with the remainder set to *Boughton*'s [1984] default values (E_{MAX} = 8.9 mm d^{-1} and NDC = 0.5). For all trials ϕ was estimated during the optimization process and $\lambda = 0.5$ [*Sumner et al.*, 1997, Table 1]. The hypercube limits for the feasible parameter space are reported in Table 1.

4.1.2. Results. Figure 2 shows quantile plots of the $S(\theta)$ values obtained from the SCE-UA and SA-SX runs for the parameter subset used by *Sumner et al.* [1997]. In 99.2% of the trials the SCE-UA algorithm converged to various optima with $S(\hat{\theta}) \approx 700$ where $\hat{\theta}$ denotes the estimated optimum parameter vector. The SA-SX algorithm converged to a far greater range of $S(\theta)$ values, with 44% of the trials converging to optima with higher $S(\theta)$ values than those found by SCE-UA. This indicates that SCE-UA is almost twice as likely to converge to a lower optimum than SA-SX. Thus SCE-UA is much more robust than SA-SX for the Scott Creek case.

Figure 3 shows quantile plots for the number of function evaluations. In 99.2% of the trials SCE-UA required fewer than 3700 function evaluations to converge. By comparison SA-SX required up to 5900 function evaluations to achieve convergence in 99.2% of the trials. Thus SCE-UA is 37% more efficient than SA-SX for the Scott Creek case.

When convergence criterion (3) was applied to SA-SX, there was little change in the spread of the $S(\theta)$ values, and the number of function evaluations decreased by approximately 20%. Thus, even when the convergence criterion is not as strict, SCE-UA remains more efficient and unequivocally more robust. When the stricter criteria (1) and (2) were applied to the SCE-UA method, 99.6% of the trials converged to optima with similar objective function values ($S(\hat{\theta})$ in the range 699.6 to 700.1). The lowest of these optima will hereafter be denoted by $\hat{\theta}^+$. A very small number of the trials located a lower optimum (denoted by $\hat{\theta}^-$) with $S(\hat{\theta}^-) \approx 693$. The maximum number of evaluations (10,000) was reached when converging to $\hat{\theta}^-$. Overall the number of function evaluations increased by 50%. Although the SA-SX was more efficient in this instance, SCE-UA was more robust in that it consistently located optima with lower $S(\theta)$ values than those found by SA-SX in the 44% of its trials.

The modified SFB model was calibrated using the entire parameter set. The increase in explained variance due to the use of the entire parameter set rather than the subset used by *Sumner et al.* [1997] is significant at the 0.05 level (p value = 0.014). Thus *Sumner et al.*'s [1997] claim of parsimony is not convincing. Figure 4 shows the empirical quantile plots of the $S(\theta)$ values. SCE-UA is undeniably more robust, converging to optima with $S(\theta) < 680$ in 94.5% of the trials, whereas SA-SX converged to optima with $S(\theta) > 680$ in 91.8% of its trials. SA-SX was approximately twice as efficient, though its lack of robustness overshadows any efficiency benefits.

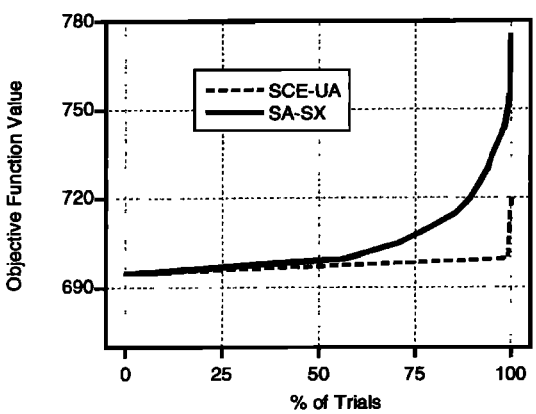

Figure 2. Empirical quantile plot of objective function values for parameter subset used by *Sumner et al.* [1997], Scott Creek catchment.

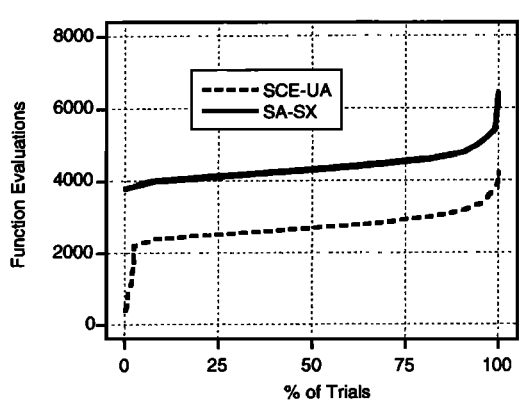

Figure 3. Empirical quantile plot of function evaluations for parameter subset used by *Sumner et al.* [1997], Scott Creek catchment.

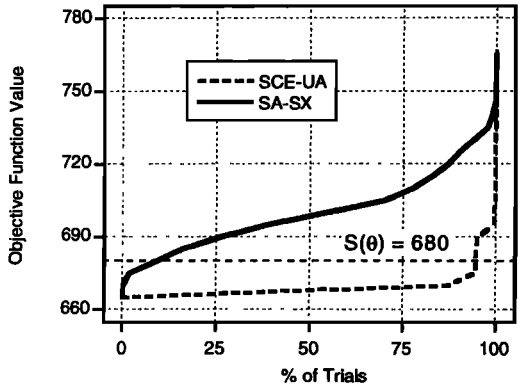

Figure 4. Empirical quantile plot of objective function values for full parameter set, Scott Creek catchment.

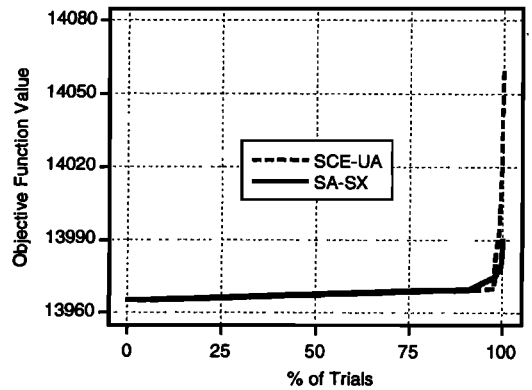

Figure 5. Empirical quantile plot of objective function values for parameter subset used by *Sumner et al.* [1997], Allyn River catchment.

4.2. Allyn River at Halton (210022)

4.2.1. Description of catchment and data. The Allyn River is a tributary of the Hunter River in eastern New South Wales, Australia. The upper 40% of the catchment is heavily timbered, with the balance open forest, cleared grazing land, and some arable land close to the river. The data set covers the period from January 1977 to December 1984 [*Chiew and McMahon*, 1993]. The catchment has an area of 205 km^2, mean annual rainfall of 1200 mm, and mean annual runoff of 350 mm. The rainfall data had to be scaled by a factor of 1.2 because of the position of the rainfall gauge and known rainfall gradients over the catchment. Although the Allyn River catchment has a much larger percentage runoff yield (runoff/rainfall ratio $\approx$ 30%) than Scott Creek, *Sumner et al.* [1997] had to use significant intervention to arrive at a set of realistic parameter estimates. Initially, four model parameters (*S, B,* DPF, and *C*) were fitted and the remaining parameters were fixed: E_{MAX} = 8.9 mm d^{-1}, F = 8.9 mm d^{-1}, NDC = 0.2, and SDR_{MAX} = 0 mm. For all trials $\lambda = 1$ and $\phi = 0$ [*Sumner et al.*, 1997, Table 1]. The hypercube limits for the feasible parameter space are reported in Table 2.

4.2.2. Results. Figure 5 shows quantile plots for the $S(\theta)$ values obtained from the SCE-UA and SA-SX runs for the parameter subset used by *Sumner et al.* [1997] There is little difference between the robustness of the algorithms, with SA-SX converging in 99.8% of the trials to the same optimum as SCE-UA in 97.6% of its trials. However, perusal of the function evaluation quantile plots in Figure 6 reveals a major difference in efficiency. In all of the trials SCE-UA required fewer than 1300 function evaluations to converge, whereas in 99% of the SA-SX trials the number of function evaluations required for convergence ranged from 7100 to 7900. Thus the SA-SX algorithm required approximately six times the amount of computational effort as SCE-UA for roughly the same level of robustness for the Allyn River case.

Applying the stricter convergence criteria (1) and (2) to SCE-UA resulted in convergence to the same optimum in every trial, but the number of function evaluations increased to between 1900 and 2500. This represents a marginal 2.4% increase in robustness and a halving of efficiency. When criterion (3) was applied to the SA-SX algorithm its reliability was comprised slightly, with only 91% of the trials converging to the same optimum. The efficiency increased slightly with all trials requiring fewer than 7400 function evaluations. However, in 1% of these trials SA-SX terminated prior to entering its third phase and the vertices of the final simplex were not in the vicinity of the candidate global optimum. Thus the stricter convergence criteria (1) and (2) were needed to maintain the robustness of SA-SX for the Allyn River case.

The modified SFB model was calibrated using the entire parameter set. The increase in explained variance due to the use of the entire parameter set rather than the subset used by *Sumner et al.* [1997] is significant at the 0.05 level (p value = 0.009). Thus *Sumner et al*'s [1997] claim of parsimony is not convincing. Figure 7 shows the $S(\theta)$ quantile plots. In 84% of the trials SCE-UA converged to optima with lower $S(\theta)$ than

Table 2. Parameter Hypercube Limits for the Modified SFB Model: Allyn River at Halton, Catchment 210022

Parameter	Lower Bound	Upper Bound
S*	50.0	250.0
F	3.0	20.0
B*	0.2	0.8
NDC	0.0	1.0
E_{max}	0.0	10.0
DPF*	0.0	0.4
SDR_{max}	0.0	100.0
C*	0.0	3.0

*Parameters estimated by *Sumner et al.* [1997].

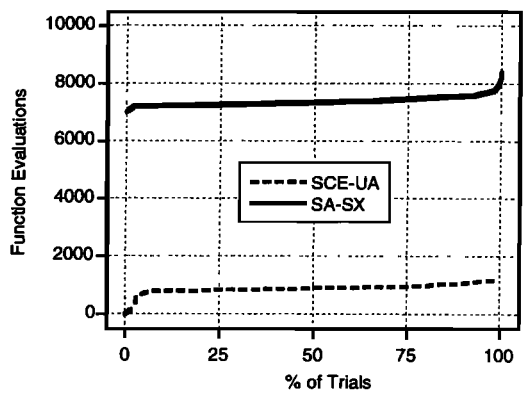

Figure 6. Empirical quantile plot of function evaluations for parameter subset used by *Sumner et al.* [1997], Allyn River catchment.

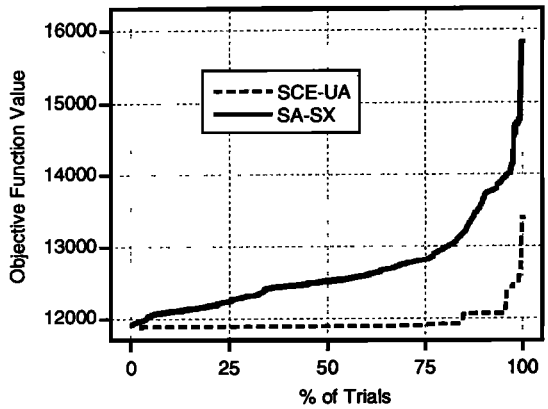

Figure 7. Empirical quantile plot of objective function values for full parameter set, Allyn River catchment.

SA-SX could locate in any of its trials. Although the algorithms had similar efficiencies, comparison of Figures 5 and 7 indicates that the robustness of SA-SX deteriorated markedly when the level of parameterization was increased although the SA-SX was not tuned to the full parameter set. This result is consistent with that for Scott Creek.

4.3. Response Surface Analysis: Scott Creek

The Scott Creek case is examined further because of the problems experienced during calibration. The response surface plots provide insight into the reasons why both optimization algorithms were not able to consistently locate the global optimum. Figure 8 shows a two-dimensional cross section of the response surface for Scott Creek in the vicinity of the estimates $\hat{\theta}^-$ and $\hat{\theta}^+$. The C-SDR_{MAX} plane was chosen as the estimates of these parameters varied the most between $\hat{\theta}^-$ and $\hat{\theta}^+$. The remaining parameters were fixed at their $\hat{\theta}$+ values. The plot reveals that the region of attraction for $\hat{\theta}^+$ is much larger than that for $\hat{\theta}^-$. Other obvious features are the presence of a high ridge with a narrow saddle that separates $\hat{\theta}^-$ and $\hat{\theta}^+$. Considering that the upper limit for SDR_{MAX} was set to 100 (see Table 1), the region of attraction for $\hat{\theta}^+$ is exceptionally large. This explains why the SCE-UA converged to $\hat{\theta}^+$ in 99.6% of the trials when convergence criteria (1) and (2) were used. The other optimum $\hat{\theta}^-$ is located on the edge of the feasible parameter space. The relatively small attractor region explains why $\hat{\theta}^-$ was found in only very few of the SCE-UA trials. The chance of selecting initial parameter estimates in the region of attraction for $\hat{\theta}^-$ is low. If an initial simplex is not seeded with parameter estimates from this region, there is little chance of the simplex projecting into the region during a probabilistic search.

Attention is also drawn to the multiple optima having $S(\theta)$ very close to $S(\theta^+)$ yet with distinctly different values of the C parameter. Such features highlight the need to focus on parameter uncertainty rather than merely finding the global optimum.

5. Summary and Conclusions

This paper has compared the robustness and efficiency of two probabilistic optimization algorithms (SA-SX and SCE-UA) using the same conceptual rainfall-runoff (CRR) model, the same data sets, and the same objective function. Two levels of parameterization were considered: subsets of the CRR model parameters that were deemed to be parsimonious in a previous study and the entire parameter set. The differences between the performances of the algorithms for the study catchments (Scott Creek and Allyn River) and the chosen parameterization levels are striking. The Allyn River catchment has a higher runoff yield, and the initial level of parameterization of the CRR model was lower than that for the Scott Creek. For calibration to the reduced parameter set, both algorithms had a similar level of robustness for the Allyn River case, but the efficiency of SCE-UA was more than six times that of SA-SX. However, for the Scott Creek case SCE-UA was almost twice as robust as SA-SX even though it had a similar efficiency.

When calibrating the full parameter set, the robustness of SA-SX deteriorated markedly relative to that of SCE-UA. This may be because the annealing schedule was not tuned for the larger parameter set.

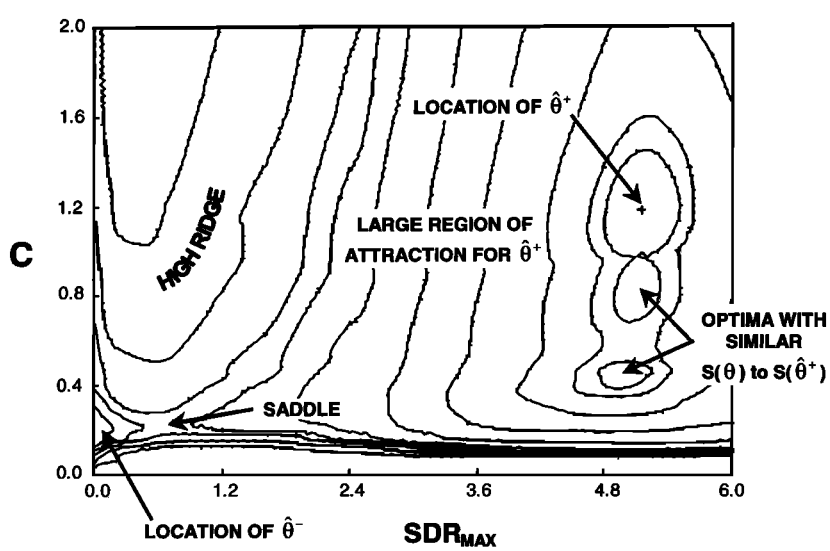

Figure 8. Two dimensional cross section of the response surface in the vicinity of the apparent global optimum, Scott Creek catchment.

The superiority of the SCE-UA can be attributed to its use of multiple complexes rather than a single simplex. Although SA-SX's stochastic step acceptance criterion allows the simplex to escape from local optima, SCE-UA has more information about the response surface compared to SA-SX at any iteration during an optimization run. By sharing this information through complex shuffling SCE-UA conducts a far more effective search of the entire parameter space. Implementation of SCE-UA is much simpler because the recommended values for its algorithmic parameters are generic. In contrast, it appears that the annealing schedule for SA-SX must be tuned for the catchment and the parameterization level at hand.

Nevertheless, the results suggest that SCE-UA should not be viewed as a panacea capable of satisfactorily calibrating a poor CRR model with unrepresentative and error-laden data. Even the most powerful optimization algorithm will experience difficulties under such circumstances. If a candidate global optimum cannot be found consistently, emphasis would be better placed on assessing parameter feasibility and uncertainty and on evaluating the precision of the model predictions.

Acknowledgments. The authors would like to gratefully acknowledge the comments provided by the reviewers, Qingyun Duan and Hoshin Gupta.

References

Bates, B. C., Calibration of the SFB model using a simulated annealing approach, *Proceedings of the Water Down Under 94 Conference, Natl. Conf. Publ. 94/15*, vol. 3, pp. 1–6, Inst. of Eng., Canberra, A. C. T., Australia, 1994.

Boughton, W. C., A simple model for estimating the water yield of ungauged catchments, *Aust. Civ. Eng. Trans., Inst. Eng. Aust.*, *26*(2), 83–88, 1984.

Box, G. E. P., and D. R. Cox, An analysis of transformations, *J. R. Stat. Soc. Ser. B*, *26*(2), 211–252, 1964.

Chiew, F. H. S., and T. A. McMahon, Complete set of daily rainfall, potential evapotranspiration and streamflow data for twenty eight unregulated Australian catchments, 53 pp., Cent. for Environ. and Appl. Hydrol., Univ. of Melbourne, Melbourne, Australia, 1993.

Chiew F. H. S., M. J. Stewardson, and T. A. McMahon, Comparison of six rainfall-runoff modelling approaches, *J. Hydrol.*, *147*, 1–36, 1993.

Corana, A., M. Marchesi, C. Martini, and S. Ridella, Minimizing multimodal functions of continuous variables with the "simulated annealing" algorithm, *Assoc. Comput. Mach. Trans. Math. Softw.*, *13*(3), 262–280, 1987.

Dekkers, A., and E. Aarts, Global optimization and simulated annealing, *Math. Prog.*, *50*, 367–393, 1991.

Duan, Q., S. Sorooshian, and V. K. Gupta, Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, *24*(7), 1163–1173, 1992.

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.

Gan, Y. G., 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.

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

Kirkpatrick, S., C. D. Gelatt, and M. P. Vecchi, Optimization by simulated annealing, *Science*, *220*, 671–680, 1983.

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

Nathan, R. J., and T. A. McMahon, The SFB Model Part I-validation of fixed model parameters, *Aust. Civ. Eng. Trans., Inst. Eng. Aust.*, *32*(3), 166–162, 1990.

Nelder, J. A., and R. Mead, A simplex method for function minimization, *Comput. J.*, *7*, 308–313, 1965.

Press, W. H. and S. A. Teukolsky, Simulated annealing optimization over continuous spaces, *Comput. Phys.*, *5*(4), 426–429, 1991.

Sorooshian, S., Q. Duan, and V. K. Gupta, Calibration of conceptual rainfall-runoff models using global optimization: Application to the Sacramento soil moisture accounting model, *Water Resour. Res.*, *29*(4), 1185–1194, 1993.

Sumner, N. R., P. M. Fleming, and B. C. Bates, Calibration of a modified SFB model for twenty-five Australian catchments using simulated annealing, *J. Hydrol.*, *197*, 166–188, 1997.

Ye, W., B. C. Bates, N. R. Viney, M. Sivapalan, and A. J. Jakeman, Performance of conceptual rainfall-runoff models in low-yielding ephemeral catchments, *Water Resour. Res.*, *33*(1), 153–166, 1997.

B. C. Bates, CSIRO Land and Water, Private Bag, PO Wembley, WA 6014, Australia. (bryson.bates@per.clw.csiro.au)

G. Kuczera and M. Thyer, Department of Civil, Surveying and Environmental Engineering, University of Newcastle, Callaghan, NSW 2308, Australia. (cegak@cc.newcastle.edu.au; cemat@cc.newcastle.edu.au)

(Received April 2, 1998; revised October 5, 1998,
accepted October 14, 1998.)