ELSEVIER

Journal of Hydrology 158 (1994) 265–284

Journal of Hydrology

[1]

# Optimal use of the SCE-UA global optimization method for calibrating watershed models

Qingyun Duan[a], Soroosh Sorooshian*[,b], Vijai K. Gupta[b]

[a]*NWS/NOAA, 1325 East West Highway, Silver Spring, MD 20910, USA*

[b]*Department of Hydrology and Water Resources, The University of Arizona, Tucson, AZ 85721, USA*

(Received 24 September 1993; revision accepted 16 January 1994)

**Abstract**

The difficulties involved in calibrating conceptual watershed models have, in the past, been partly attributable to the lack of robust optimization tools. Recently, a global optimization method known as the SCE-UA (shuffled complex evolution method developed at The University of Arizona) has shown promise as an effective and efficient optimization technique for calibrating watershed models. Experience with the method has indicated that the effectiveness and efficiency of the algorithm are influenced by the choice of the algorithmic parameters. This paper first reviews the essential concepts of the SCE-UA method and then presents the results of several experimental studies in which the National Weather Service river forecast system–soil moisture accounting (NWSRFS-SMA) model, used by the National Weather Service for river and flood forecasting, was calibrated using different algorithmic parameter setups. On the basis of these results, the recommended values for the algorithmic parameters are given. These values should also help to provide guidelines for other users of the SCE-UA method.

## 1. Introduction

Computer-based hydrologic models have become popular with practicing hydrologists and water resources engineers for performing hydrologic forecasts and for managing water systems. One type of computer model which is often used is the conceptual watershed model. Such models predict the magnitude of streamflows generated by a precipitation event by simulating the physical processes influencing the movement of water over and through the soil. The accuracy of these predictions

* Corresponding author.


*SSDI* 0022-1694(94)02467-P

depends on how well the model structure is defined and how the model parameters are determined. Conceptual models generally have a large number of parameters which are not directly measurable and must therefore be estimated through model calibration, i.e. by fitting the simulated outputs of the model to the observed outputs of the watershed by adjusting the model parameters. A measure of the fit between the simulated and observed outputs is called a calibration criterion or objective function. The goal of calibration is to find those values for the model parameters that minimize (or maximize, if appropriate) the specified calibration criterion.

Despite the usefulness and popularity of conceptual watershed models, their performance has not always been acceptable because of uncertainty in the model parameter estimates. For example, Johnston and Pilgrim (1976) tried unsuccessfully for over 2 years to find a unique set of parameters for the Boughton model. The National Weather Service (NWS), which is responsible for providing river and flood forecasts for more than 3000 river basins throughout the USA, has expressed similar concerns with the hydrologic models within the NWS river forecast system (NWSRFS). As pointed out by Brazil and Hudlow (1980): 'One of the most difficult problems faced is the calibration of the models within NWSRFS for various parts of the country'.

During the last two decades, a great many studies on issues relating to conceptual model calibration have been published (e.g. Ibbitt, 1972; Johnston and Pilgrim, 1976; Sorooshian and Dracup, 1980; Restrepo-Posada, 1982; Kuczera, 1983a,b; Gupta and Sorooshian, 1983; Sorooshian and Gupta, 1983, 1985; Sorooshian et al., 1983; Troutman, 1985a,b; Duan et al., 1988). Although these efforts have helped to increase our understanding of the nature of the calibration problems, only limited success in alleviating the severity of the difficulties has been achieved (Sorooshian et al., 1983; Ibbitt and Hutchinson, 1984). Recently, Duan et al. (1992) conducted a detailed study of a simple six-parameter conceptual model (SIXPAR) using synthetic data to identify clearly the nature of the difficulties encountered in conceptual model calibration. The study found that, despite the simple model structure and the absence of model structural error or input data error, the parameter estimation problems are not trivial. Duan et al. summarized these problems as a list of five features (see Table 1).

The primary conclusion of the Duan et al. (1992) study was that the optimization techniques employed for parameter estimation are not powerful enough to deal with the response surface conditions encountered in model calibration. The calibration

Table 1
Summary of the five major characteristics complicating the optimization problem in CRR model calibration

| | | |
|---|---|---|
| 1. | Regions of attraction | More than one main convergence region |
| 2. | Minor local optima | Many small 'pits' in each region |
| 3. | Roughness | Rough response surface with discontinuous derivatives |
| 4. | Sensitivity | Poor and varying sensitivity of response surface in region of optimum and non-linear parameter interaction |
| 5. | Shape | Non-convex response surface with long curved ridges |

techniques commonly used rely on direct-search optimization algorithms such as the Simplex method of Nelder and Mead (1965) and the pattern search method of Hooke and Jeeves (1961) (see, e.g. Johnston and Pilgrim, 1976; Pickup, 1977; Hendrickson et al., 1988). These algorithms are designed to solve single-optimum problems and are not able to deal effectively with all of the problems listed in Table 1. Experienced hydrologists typically cope with the inadequacy of these optimization methods by use of a manual calibration stage; the optimization algorithms are only used for fine-tuning of selected parameters. However, manual calibration requires detailed understanding of the model, which can only be obtained through many years of calibration experience. Furthermore, it can be very tedious and time-consuming (Baffaut and Delleur, 1989). Recently, researchers have been exploring ways to incorporate 'expert knowledge' of conceptual watershed models into the automatic calibration procedures (Baffaut and Delleur, 1989; Wheater et al., 1989; Harlin, 1991). These schemes are highly model dependent and are difficult to generalize to other models.

In recent years, many researchers have begun to investigate the use of globally based optimization methods for model calibration. Brazil (1988) investigated the use of the adaptive random search (ARS) method (Pronzato et al., 1984) to calibrate the soil moisture accounting model of the NWSRFS (NWSRFS-SMA), and reported that the ARS method was capable of producing promising results when used as part of a multi-level calibration strategy. Wang (1991) reported that the genetic algorithm (Holland, 1975), with fine-tuning by a local search method, can provide an efficient and robust means for calibration of the Xinanjiang watershed model.

Duan et al. (1992, 1993) presented a new global optimization method known as the SCE-UA method (abbreviation for shuffled complex evolution method developed at The University of Arizona). This method is based on a synthesis of the best features from several existing methods, including the genetic algorithm, and introduces the new concept of complex shuffling. The method was designed specifically for the purpose of dealing with the peculiar problems encountered in conceptual watershed model calibration. Extensive testing on a simplified research version of the NWSRFS-SMA model (i.e. the SIXPAR model) and the full-scale NWSRFS-SMA model revealed that the SCE-UA method was both effective and efficient, compared with other existing global methods, including the ARS method and the multistart Simplex method (Duan et al., 1992; Sorooshian et al., 1993).

## 2. Scope of this paper

In the above-mentioned studies, it was found that the effectiveness and efficiency of the SCE-UA method are sensitive to the choice of algorithmic parameters (Duan et al., 1992, 1993). This paper presents a review of the essential concepts of the SCE-UA method and the results of a study conducted to establish guidelines on how to choose the algorithmic parameters of this method according to the degree of difficulty of the calibration problem.

## 3. The SCE-UA method

### 3.1. Desirable properties of a global optimization method

The five features listed in Table 1 that characterize the problems encountered in conceptual watershed model calibration are typical of many optimization problems faced by workers in various fields. An optimization algorithm that aims to deal with them must possess the following properties: (1) global convergence in the presence of multiple regions of attraction; (2) ability to avoid being trapped by small pits and bumps on the objective function surface; (3) robustness in the presence of differing parameter sensitivities and parameter interdependence; (4) non-reliance on the availability of an explicit expression for the objective function or the derivatives; (5) capability of handling high-parameter dimensionality.

### 3.2. Description of the SCE-UA method

The SCE-UA method embodies the desirable properties described above and is specifically designed to deal with the peculiarities encountered in conceptual watershed model calibration. The method is based on a synthesis of four concepts: (1) 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; (4) complex shuffling. The first three concepts are drawn from existing approaches that have been proven successful in the past (Holland, 1975; Price, 1978, 1983; Manetsch, 1990; Wang, 1991), and the last concept was recently introduced (Duan et al., 1992, 1993; Sorooshian et al., 1993). The synthesis of these elements makes the SCE-UA method effective and robust, and also flexible and efficient. A general description of the steps of the SCE-UA method is given below (a more detailed presentation of the theory underlying the SCE-UA algorithm has been given by Duan et al. (1992, 1993)):

(1) Generate sample—sample $s$ points randomly in the feasible parameter space and compute the criterion value at each point. In the absence of prior information on the approximate location of the global optimum, use a uniform probability distribution to generate a sample.

(2) Rank points—sort the $s$ points in order of increasing criterion value so that the first point represents the smallest criterion value and the last point represents the largest criterion value (assuming that the goal is to minimize the criterion value).

(3) Partition into complexes—partition the $s$ points into $p$ complexes, each containing $m$ points. The complexes are partitioned such that the first complex contains every $p(k-1)+1$ ranked point, the second complex contains every $p(k-1)+2$ ranked point, and so on, where $k = 1, 2, \ldots, m$.

(4) Evolve each complex—evolve each complex according to the competitive complex evolution (CCE) algorithm (which is elaborated below).

(5) Shuffle complexes—combine the points in the evolved complexes into a single sample population; sort the sample population in order of increasing criterion value;

shuffle (i.e. re-partition) the sample population into $p$ complexes according to the procedure specified in Step 3.

(6) Check convergence—if any of the pre-specified convergence criteria are satisfied, stop; otherwise, continue.

(7) Check the reduction in the number of complexes—if the minimum number of complexes required in the population, $p_{\min}$, is less than $p$, remove the complex with the lowest ranked points; set $p = p - 1$ and $s = pm$; return to Step 4. If $p_{\min} = p$, return to Step 4. (It should be noted that this step is a new feature and is added to the version presented by Duan et al. (1992, 1993).)

The initial random sampling of the parameter space provides the potential for locating the global optimum without being biased by pre-specified starting points. The partition of the population into several communities facilitates a freer and more extensive exploration of the feasible space in different directions, thereby allowing for the possibility that the problem has more than one region of attraction. The shuffling of communities enhances the survivability by a sharing of the information (about the search space) gained independently by each community.

One key component of the SCE-UA method is the CCE algorithm, as mentioned in Step 4. This algorithm, based on the Nelder and Mead (1965) Simplex downhill search scheme, is presented briefly as follows:

(I) Construct a subcomplex by randomly selecting $q$ points from the complex (community) according to a trapezoidal probability distribution. The probability distribution is specified such that the best point (i.e. the point with the best function value) has the highest chance of being chosen to form the subcomplex, and the worst point has the least chance.

(II) Identify the worst point of the subcomplex and compute the centroid of the subcomplex without including the worst point.

(III) Attempt a reflection step by reflecting the worst point through the centroid. If the newly generated point is within the feasible space, go to Step IV; otherwise, randomly generate a point within the feasible space and go to Step VI.

(IV) If the newly generated point is better than the worst point, replace the worst point by the new point. Go to Step VII. Otherwise, go to Step V.

(V) Attempt a contraction step by computing a point halfway between the centroid and the worst point. If the contraction point is better than the worst point, replace the worst point by the contraction point and go to Step VII. Otherwise, go to Step VI.

(VI) Randomly generate a point within the feasible space. Replace the worst point by the randomly generated point.

(VII) Repeat Steps II–VI $\alpha$ times, where $\alpha \geqslant 1$ is the number of consecutive offspring generated by the same subcomplex.

(VIII) Repeat Steps I–VII $\beta$ times, where $\beta \geqslant 1$ is the number of evolution steps taken by each complex before complexes are shuffled.

In the CCE algorithm, each point of a complex is a potential 'parent' with the ability to participate in the process of reproducing offspring. (The terms 'parent', 'reproduction', 'offspring', etc., are borrowed from the literature on the genetic method for global optimization.) A subcomplex functions like a pair of parents, except that it may comprise more than two members. Use of a stochastic scheme to

construct subcomplexes allows the parameter space to be searched more thoroughly. The idea of competitiveness is introduced in forming subcomplexes, based on the notion that the stronger survives better and breeds healthier offspring than the weaker. Inclusion of the competitive measure expedites the search towards promising regions. The measure of competitiveness in the reproductive process is realized by use of a trapezoidal probability distribution which favors better points over worse points. The Nelder–Mead Simplex downhill search scheme is utilized to generate offspring. This scheme is insensitive to nonsmoothness of the response surface and allows the algorithm to make use of response surface information to guide the search toward the improvement direction. In addition to the Simplex scheme, offspring are introduced at random locations of the feasible space under certain conditions to make sure the evolution process is not interrupted as a result of some unusual conditions encountered in the search space; this is somewhat analogous to mutation in response to stress

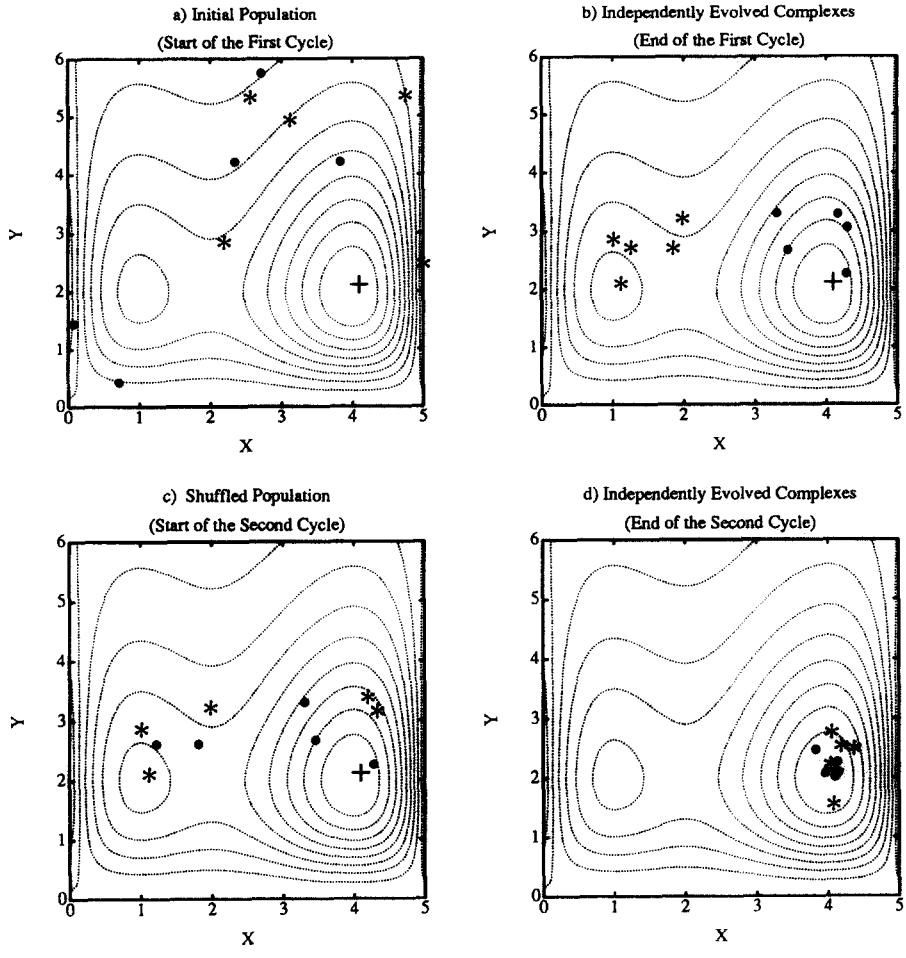


Fig. 1. Illustration of the shuffled complex evolution (SCE-UA) method.

in biological evolution. Every member in the population is given at least one chance to contribute to the reproduction process before being displaced or discarded. Thus, no information contained in the sample population is ignored.

The SCE-UA method is explained in Fig. 1, by use of a two-dimensional example. The contour lines represent a function surface with a global optimum located at (4,2) and a local optimum located at (1,2). Fig. 1(a) shows that a sample population containing $s$ (in this case, 10) points is divided into $p$ (two) communities (complexes), each containing $m$ (five) members, marked by • and $*$, respectively. As each community undergoes an independent evolution process, one community (marked by $*$) is converging toward the local optimum, whereas the other (marked by •) is converging toward the global optimum. The locations of the members in the two evolved communities at the end of the first evolution cycle are illustrated in Fig. 1(b) (to demonstrate clearly the scenario that the two complexes were converging toward two distinct optima, the number of evolution steps taken by each complex, $\beta$, was set to a relatively large value of 10). The two evolved communities are shuffled according to the procedure specified in Step 5. The new memberships of the two evolved communities after shuffling are displayed in Fig. 1(c), and the two communities at the end of the second evolution cycle are shown in Fig. 1(d). It is clear that both communities are now converging toward the global optimum.

The CCE algorithm is illustrated graphically in Fig. 2, where the black dots (•) indicate the locations of the points in a complex before the evolution step is taken. A subcomplex containing $q$ points (in this case, three, i.e. forming a triangle) is selected, according to a trapezoidal probability distribution, to initiate an evolution step. The symbol $*$ represents the new points generated by the evolution steps. The 'reflection' step, which is implemented by reflecting the worst point in a subcomplex through the centroid of the other points, is displayed in Figs. 2(a), 2(b), and 2(d). Because the reflected point has a lower criterion value than the worst point, the worst point is discarded and replaced by the new point. Thus, an evolution step is completed. In Fig. 2(c), the new point is generated by a 'contraction' step (the new point lies halfway between the worst point and the centroid of the other points), after rejecting a reflection step for not improving the criterion value. In Fig. 2(e), a 'mutation' step is taken by randomly selecting a point in the feasible parameter space to replace the worst point of the subcomplex. This is done after a reflection step is attempted, but results in a point outside of the feasible parameter space (another scenario in which a mutation step can be taken is when both the reflection step and the contraction step do not improve the criterion value). In this example, each subcomplex produces one offspring, i.e. $\alpha = 1$. The final complex after $\beta$ (five) evolution steps is illustrated in Fig. 2(f).

### 3.3. Selection of algorithmic parameters

The SCE-UA method contains many probabilistic and deterministic components that are controlled by some algorithmic parameters. For the method to perform optimally, these parameters must be chosen carefully. They are: $m$, the number of points in a complex; $q$, the number of points in a subcomplex; $p$, the number of

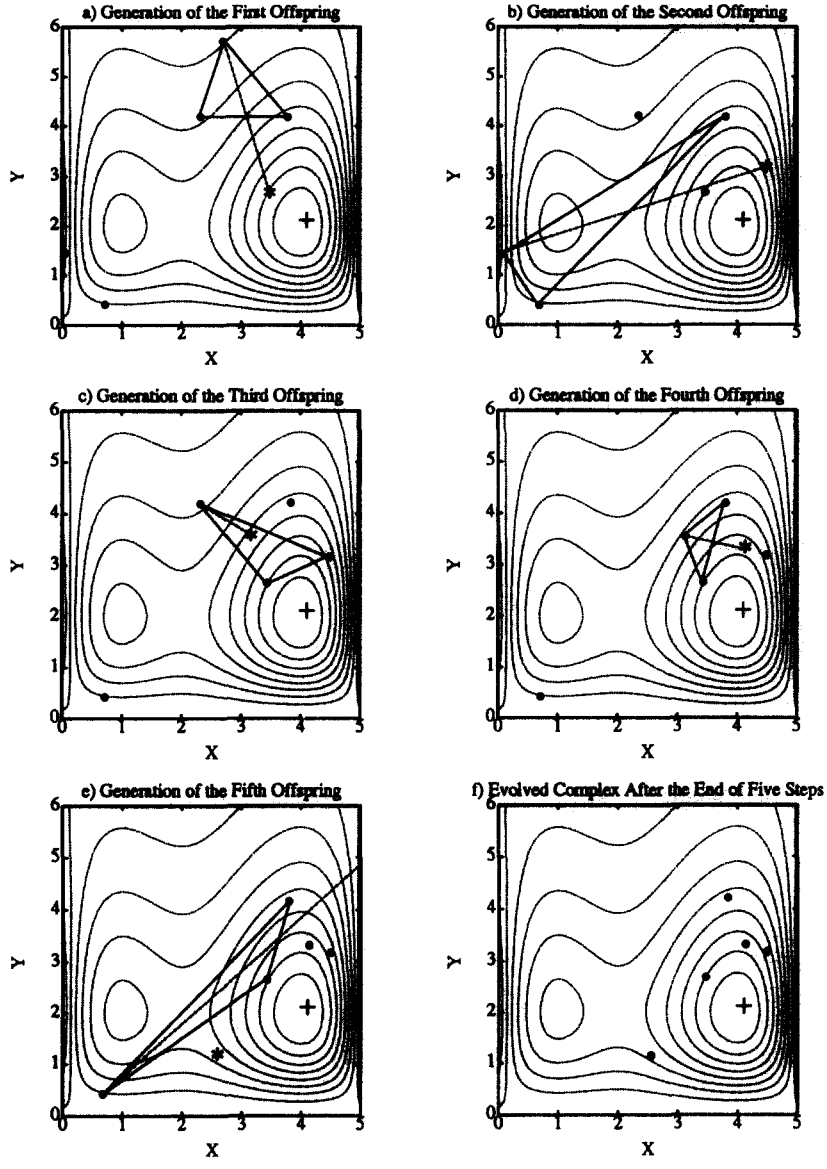


Fig. 2. Illustration of the evolution steps taken by each complex.

complexes; $p_{\min}$, the minimum number of complexes required in the population; $\alpha$, the number of consecutive offspring generated by each subcomplex; $\beta$, the number of evolution steps taken by each complex.

Theoretically, the number of points in each complex, $m$, may take any value greater than or equal to two. However, if there are too few points in each complex, the search would proceed in a manner similar to the ordinary Simplex procedure, and the global

search capability would be undermined. Conversely, if $m$ is chosen too large, it may result in excessive use of the computer processing time (CPU) with no certain gain in effectiveness. Our previous investigation (see, e.g. Duan et al., 1993) indicated that, by setting $m$ to $2n + 1$, where $n$ is the number of parameters to be optimized on, and varying the number of complexes, $p$, the SCE-UA algorithm provided better overall performance in coping with a wide range of optimization problems than by increasing the $m$ value alone.

The number of points in each subcomplex, $q$, may vary between two and $m$. By selecting the value of $n + 1$, the subcomplex is chosen as a Simplex; this defines a first-order approximation (hyperplane) to the objective function surface and will, therefore, give a reasonable estimate of the local improvement direction.

The number of offspring which each subcomplex generates before it is put back into the complex, $\alpha$, can be any number greater than or equal to one. If $\alpha$ is equal to one, only one of the original parents will be replaced. As $\alpha$ is increased, the search becomes more strongly biased in favor of local search of the parameter space.

The number of evolution steps taken by each complex before the complexes are shuffled, $\beta$, can be any positive integer. If $\beta$ is small, the complexes will be shuffled frequently, but will not be able to conduct much independent exploration of the parameter space; if $\beta$ is large, each complex will rapidly shrink into a small cluster, and global search effectiveness may be lost.

The required number of complexes, $p$, is strongly dependent on the nature of the

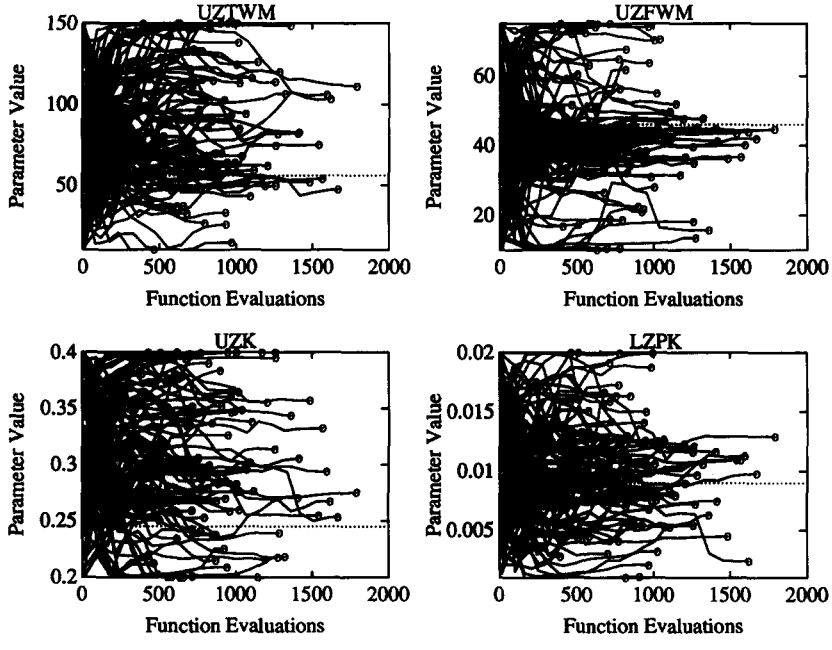


Fig. 3. Convergence behavior of four of the NWSRFS-SMA model parameters for 100 independent trials of the Simplex algorithm (synthetic data study); dotted lines represent the 'true' parameter values (from Sorooshian et al., 1993).

problem (Duan et al., 1992, 1993; Sorooshian et al., 1993). The higher the degree of difficulty, the larger the number of complexes required to locate the global optimum.

Parameter $p_{min}$, the minimum number of complexes (between one and $p$) required in the population, is introduced into the SCE-UA algorithm to improve efficiency. The justification is that as the search process proceeds, the population is converging into an increasingly smaller space, and fewer points are necessary to provide an adequate population density.

The SCE-UA method, along with the widely used multi-start Simplex (MSX) method, was evaluated on some standard test problems as well as on the calibration of two watershed models (see Duan et al., 1992, 1993; Sorooshian et al., 1993). Figs. 3 and 4 (from Sorooshian et al., 1993) demonstrated the capability of the SCE-UA method and the MSX method in finding the global optimum of the NWSRFS-SMA model parameters. Sorooshian et al. (1993) pointed out that in the 10 SCE-UA trials, the global optimum was located precisely with a 100% success rate, whereas none of the 100 Simplex trials was able to find the global optimum. The best estimated parameter set from the Simplex trials barely came within the 20% range of the global optimum. In conducting those evaluation studies, four of the algorithmic parameters of the SCE-UA method were assigned to some default values, as follows: (1) the number of points in a complex, $m = 2n + 1$; (2) the number of points in a subcomplex, $q = n + 1$; (3) the number of consecutive offspring generated by each subcomplex, $\alpha = 1$; (4) the number of evolution steps taken by each complex, $\beta = m$.

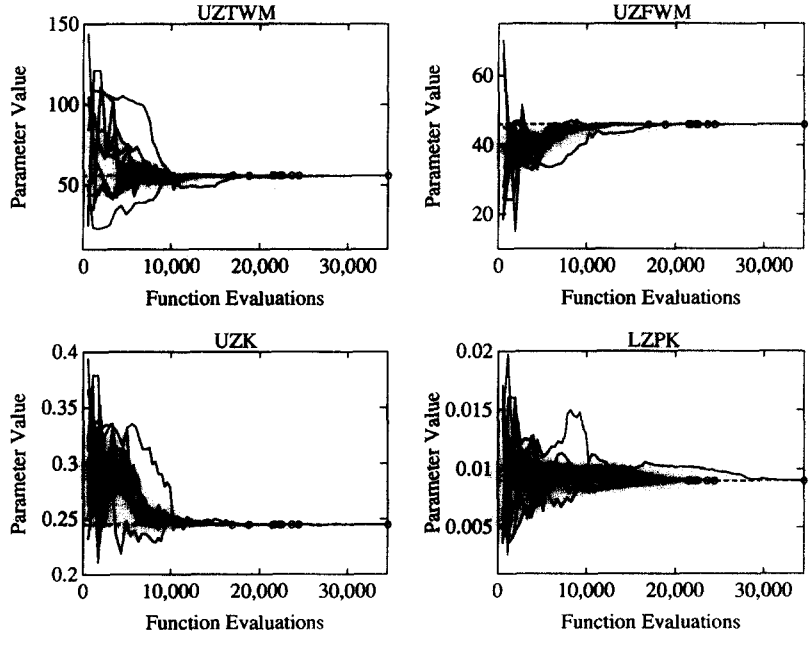


Fig. 4. Convergence behavior of four of the NWSRFS-SMA model parameters for 10 independent trials of the SCE-UA method (synthetic data study); dotted lines represent the 'true' parameter values (from Sorooshian et al., 1993).

Different values for the other two parameters—the number of complexes $p$ and the minimum number of complexes required in the population $p_{min}$—were tested. However, the merits of the default values for the four parameters listed above or the different values assigned to $p$ and $p_{min}$ were never confirmed. Further, our experience with the SCE-UA method has indicated that the effectiveness and efficiency of the algorithm are influenced by the choice of the algorithmic parameters. Therefore, the main purpose of this paper is to investigate the proper choices for algorithmic parameters. More specifically, a series of experimental studies will be conducted to examine the proper values for $p$, $p_{min}$, $\alpha$, and $\beta$. When a selected algorithm parameter is being tested, other parameters will take on either their default values or the values explicitly specified in the text.

## 4. Experimental studies for determining the algorithmic parameters of the SCE-UA method

### *4.1. Design of experimental studies*

The NWSRFS-SMA model embodies all the difficulties delineated in Table 1. It is used by the River Forecast Centers of the NWS to perform real-time river and flood

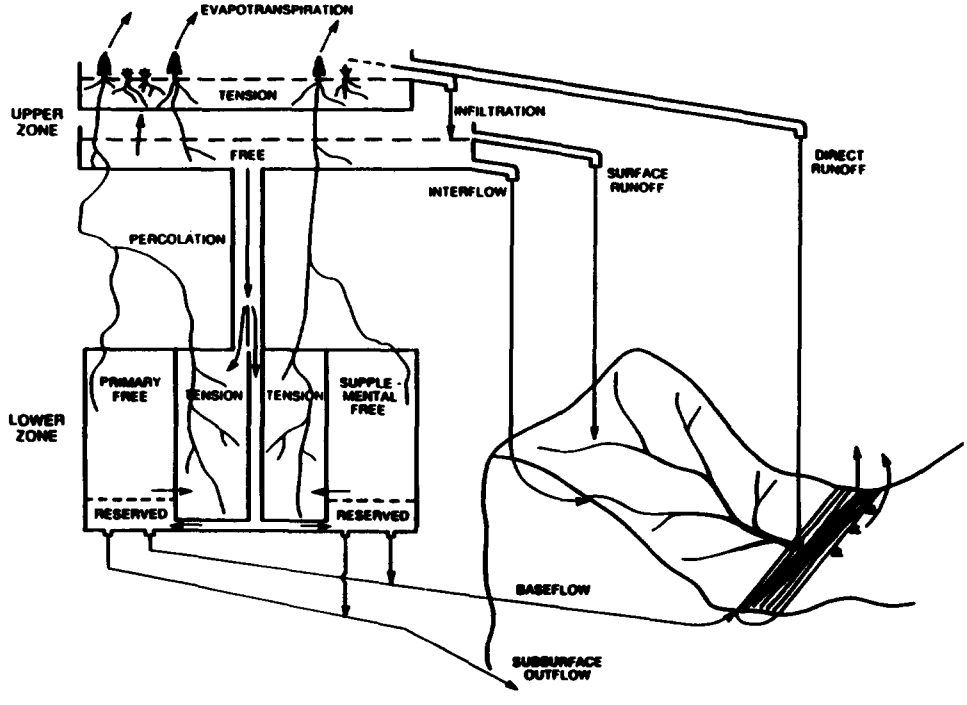


Fig. 5. Schematic description of the NWSRFS soil moisture accounting (NWSRFS-SMA) model (from Brazil, 1988).

Table 2
Parameters of the NWSRFS-SMA model

| Parameters | Description |
|---|---|
| UZTWM | Maximum capacity of the upper zone tension water storage (mm) |
| UZFWM | Maximum capacity of the upper zone free water storage (mm) |
| LZTWM | Maximum capacity of the lower zone tension water storage (mm) |
| LZFPM | Maximum capacity of the lower zone free water primary storage (mm) |
| LZFSM | Maximum capacity of the lower zone free water supplement storage (mm) |
| ADIMP | Additional impervious area (decimal fraction) |
| UZK | Upper zone free water lateral depletion rate ($day^{-1}$) |
| LZPK | Lower zone primary free water depletion rate ($day^{-1}$) |
| LZSK | Lower zone supplemental free water depletion rate ($day^{-1}$) |
| ZPERC | Maximum percolation rate (dimensionless) |
| REXP | Exponent of the percolation equation (dimensionless) |
| PCTIM | Impervious fraction of the watershed area (decimal fraction) |
| RIVA | Riparian vegetation area (decimal fraction) |
| PFREE | Fraction of water percolating from upper zone which goes directly to lower zone free water storage (decimal fraction) |
| SIDE | Ratio of deep recharge to channel baseflow (dimensionless) |
| RSERV | Fraction of lower zone free water not transferable to lower zone tension water (decimal fraction) |

forecasts as well as extended streamflow predictions. The NWSRFS-SMA model was originally developed by Burnash et al. (1973) and modified by the Hydrologic Research Laboratory of the NWS (Peck, 1976; Brazil, 1988). A detailed description of the model is available in the literature (e.g. Burnash et al., 1973; Peck, 1976) and will not be discussed here. In this study, the research version of the NWSRFS-SMA model maintained by the Department of Hydrology and Water Resources, The University of Arizona, was used.

Table 3
The true parameter values and the lower and upper parameter bounds used for the synthetic study

| Parameter | True value | Lower bound | Upper bound |
|---|---|---|---|
| UZTWM | 56.000 | 10.000 | 150.000 |
| UZFWM | 46.000 | 10.000 | 75.000 |
| LZTWM | 131.000 | 75.000 | 400.000 |
| LZFPM | 162.000 | 50.000 | 1000.000 |
| LZFSM | 23.000 | 10.000 | 300.000 |
| ADIMP | 0.173 | 0.000 | 0.200 |
| UZK | 0.245 | 0.200 | 0.400 |
| LZPK | 0.009 | 0.001 | 0.020 |
| LZDK | 0.043 | 0.020 | 0.250 |
| PCTIM | 0.043 | 0.000 | 0.100 |
| ZPERC | 226.000 | 5.000 | 250.000 |
| REXP | 3.650 | 1.100 | 4.000 |
| PFREE | 0.063 | 0.000 | 0.600 |

Parameters not optimized: RSERV = 0.3; RIVA = 0.0; SIDE = 0.0.

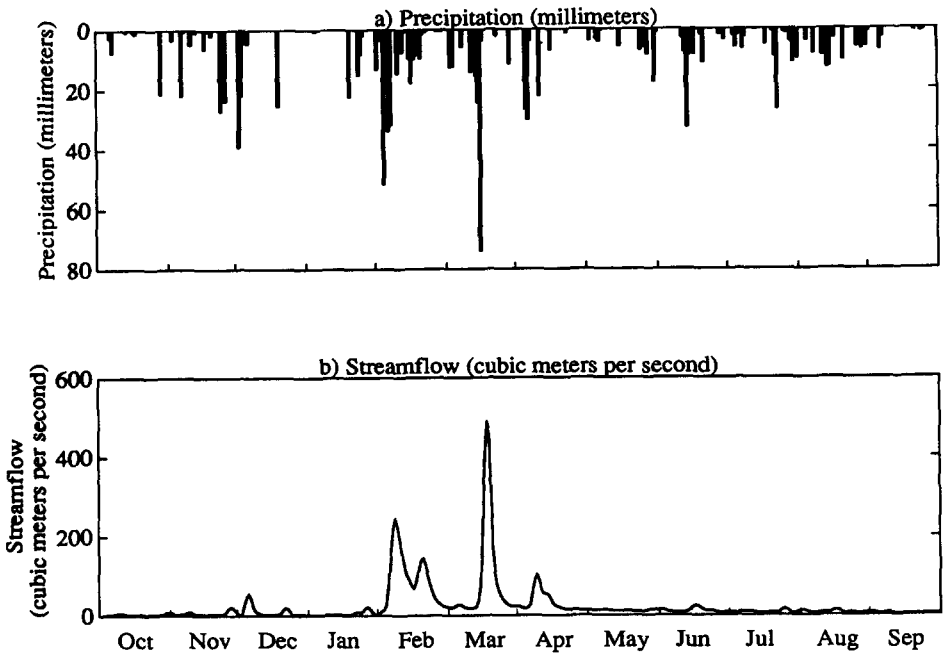


Fig. 6. Hydrologic data used for the calibration of the NWSRFS-SMA model.

A schematic description of the NWSRFS-SMA model is presented in Fig. 5. The model is controlled by 16 parameters (see Table 2). The inputs to the model are the mean areal precipitation (mm) and the potential evapotranspiration (mm). The outputs are the streamflow runoff (cms) at the outlet of the basin and the actual evapotranspiration (mm).

The precipitation and potential evapotranspiration data from the time period between 1 October 1955 and 30 September 1956 from the Leaf River Basin near Collins, Mississippi, were used as the model inputs. The parameter set obtained by Brazil (1988) was assumed as the 'true' parameter set (see Table 3), and using this and the precipitation and evapotranspiration inputs, a sequence of streamflows was generated. This sequence of streamflows was treated as the 'observed' streamflow data for the calibration period. The precipitation and 'observed' streamflow time series are displayed in Fig. 6. By using synthetically generated streamflow data, the precise 'true' global optimum is known beforehand. Hence, the experimental studies can be conducted without the influence of errors in the model structure.

The optimization studies reported here simulated attempts to calibrate selected parameters of the NWSRFS-SMA model by initiating the search algorithm at randomly selected points in the feasible parameter space, defined as the hypercube bounded by the limits of the optimizing parameters as specified in Table 3. Parameters not being optimized were fixed to their 'true' values. Up to 13 out of the 16 model parameters were included for optimization. Three parameters, RSERV, RIVA, and

Table 4
Optimization of various parameters: two-dimensional study results

| | | Parameter groups | | | | | | Average |
|---|---|---|---|---|---|---|---|---|
| | | A | B | C | D | E | F | |
| $p = 1$ | NS[a] | *10* | *10* | *10* | *10* | *10* | *10* | 10 |
| | AFE[b] | *99* | *120* | *108* | *92* | *89* | *112* | 103 |
| $p = 2$ | NS | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| | AFE | 193 | 220 | 188 | 187 | 184 | 168 | 190 |

[a] NS, the number of successful runs.
[b] AFE, the average number of function evaluations of the successful runs.

SIDE, were fixed at the following values: RSERV = 0.3, RIVA = 0.0, and SIDE = 0.0. These parameters are not optimizable and should not be changed unless hydrometeorological conditions indicate otherwise (Peck, 1976).

The objective function (or calibration criterion) used was the mean daily square root of the difference between the observed flows and simulated flows (DRMS). Because there were no errors in the synthetic data, the criterion value at the 'true' parameter set (the global optimum) is known to be 0.0. Two stopping criteria were used to terminate an optimization run. First, an optimization run would be stopped if a DRMS value of $10^{-3}$ or less was obtained; in this case, the run was considered a success. If, however, before that occurring, the population of points converged into a small space, the measure of which was less than $10^{-3}$% of the feasible space, the run was also stopped and was termed a failure.

### 4.2. Experimental Study I

The first study explored whether different combinations of model parameters chosen for optimization would influence the selection of the algorithmic parameter $p$. The first 12 parameters listed in Table 3 were divided into several groups which

Table 5
Optimization of various parameters: four-dimensional study results

| | | Parameter groups | | | Average |
|---|---|---|---|---|---|
| | | A | B | C | |
| $p = 1$ | NS | *10* | *10* | *10* | 10 |
| | AFE | *465* | *272* | *359* | 365 |
| $p = 2$ | NS | 10 | 10 | 10 | 10 |
| | AFE | 537 | 506 | 519 | 521 |
| $p = 4$ | NS | 10 | 10 | 10 | 10 |
| | AFE | 1038 | 957 | 957 | 984 |

Table 6
Optimization of various parameters: six-dimensional study results

| | | Parameter groups | | Average |
|---|---|---|---|---|
| | | A | B | |
| $p = 1$ | NS | 6 | 6 | 6 |
| | AFE | 675 | 712 | 694 |
| $p = 2$ | NS | *10* | *10* | 10 |
| | AFE | *973* | *880* | 927 |
| $p = 4$ | NS | 10 | 10 | 10 |
| | AFE | 1783 | 1676 | 1729 |
| $p = 8$ | NS | 10 | 10 | 10 |
| | AFE | 3517 | 3172 | 3344 |

contain predetermined numbers of parameters $n$. For example, if $n = 2$, the first parameter group would then consist of parameters UZTWM and UZFWM, the second group LZTWM and LZFPM, and so on. For a given parameter group, 10 independent optimization runs were conducted with $p$ set at different values. The algorithmic parameter $p_{\min}$ was set equal to $p$. For each set of runs, the number of successes (NS) and the average number of function evaluations (AFE) of the successful runs were recorded.

The results for the two-, four- and six-parameter optimization study cases are shown in Tables 4, 5, and 6, respectively. In the two- and four-parameter optimization cases (i.e. $n = 2$ and $n = 4$), $p = 1$ was sufficient to ensure that all optimization

Table 7
Test results on selection of $p_{\min}$; Scheme 1: $p_{\min} = p$

| | | Dimension, $n$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | 2 | 4 | 6 | 8 | 10 | 13 |
| $p = 1$ | NS | *10* | *10* | 6 | 0 | 0 | 0 |
| | AFE | *103* | *365* | 694 | 0 | 0 | 0 |
| $p = 2$ | NS | 10 | 10 | *10* | *10* | 1 | 0 |
| | AFE | 190 | 521 | *927* | *1682* | 4307 | 0 |
| $p = 4$ | NS | | 10 | 10 | 10 | *10* | *10* |
| | AFE | | 984 | 1730 | 2749 | *5949* | *13299* |
| $p = 8$ | NS | | | 10 | 10 | 10 | 10 |
| | AFE | | | 3344 | 5159 | 8367 | 14783 |
| $p = 12$ | NS | | | | 10 | 10 | 10 |
| | AFE | | | | 7466 | 11805 | 22248 |
| $p = 20$ | NS | | | | | | 10 |
| | AFE | | | | | | 33366 |

Table 8
Test results on selection of $p_{min}$; Scheme 2: $p_{min} = p/2$

| | | Dimension, $n$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | 2 | 4 | 6 | 8 | 10 | 13 |
| $p = 2$ | NS | *10* | 9.7 | 9 | 0 | 0 | 0 |
| | AFE | *108* | 315 | 688 | 0 | 0 | 0 |
| $p = 4$ | NS | | *10* | *10* | *10* | 1 | 0 |
| | AFE | | *568* | *1017* | *1664* | 3415 | 0 |
| $p = 8$ | NS | | | 10 | 10 | *10* | *10* |
| | AFE | | | 1920 | 2981 | *5168* | *13719* |
| $p = 12$ | NS | | | | 10 | 10 | 10 |
| | AFE | | | | 4446 | 7239 | 14596 |
| $p = 20$ | NS | | | | | | 10 |
| | AFE | | | | | | 20526 |

runs would successfully find the 'true' parameters. When $n = 6$, a $p$ value equal to two or larger was sufficient to achieve a 100% success rate.

The test results in this study did not seem to indicate that different combinations of parameters chosen for optimization influenced the effectiveness or efficiency of the optimization runs. The dimensionality was the primary factor determining the proper choice of algorithm parameter $p$.

Table 9
Test results on selection of $p_{min}$; Scheme 3: $p_{min} = 1$

| | | Dimension, $n$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | 2 | 4 | 6 | 8 | 10 | 13 |
| $p = 2$ | NS | *10* | 9.7 | 9 | 0 | 0 | 0 |
| | AFE | *108* | 315 | 688 | 0 | 0 | 0 |
| $p = 4$ | NS | | *10* | 8 | 0 | 0 | 0 |
| | AFE | | *398* | 691 | 0 | 0 | 0 |
| $p = 8$ | NS | | | 9.5 | 2 | 0 | 0 |
| | AFE | | | 1088 | 1789 | 0 | 0 |
| $p = 12$ | NS | | | *10* | 8 | 0 | 0 |
| | AFE | | | *1864* | 2650 | 0 | 0 |
| $p = 20$ | NS | | | | *10* | 6 | 0 |
| | AFE | | | | *5776* | 7410 | 0 |
| $p = 30$ | NS | | | | | *10* | 0 |
| | AFE | | | | | *15176* | 0 |
| $p = 40$ | NS | | | | | | *10* |
| | AFE | | | | | | *33147* |

### 4.3. Experimental Study II

Study II explored the selection of the minimum number of complexes required in the optimization search, $p_{min}$. Three schemes for selecting $p_{min}$ were investigated. In Scheme 1, $p_{min}$ was set to $p$. In Scheme 2, $p_{min}$ was set to the larger value of one and INT($p/2$), where INT is an operator truncating a real number to the nearest integer. In Scheme 3, $p_{min}$ was equal to one. Under each scheme, a series of optimization runs were carried out. The results are shown in Tables 7–9. In these tables, two values were recorded: the number of successes out of 10 runs and the average number of function evaluations of the successful runs. For problems where dimensionality was smaller than eight, the results were obtained by averaging the results of different groups. The italic values in the tables indicate the recommended values for $p$ if a 100% success rate is desired.

The results suggest that Scheme 1 should be the preferred scheme to follow because it gives the best overall performance in terms of effectiveness and efficiency. Scheme 2 generally fulfilled the promise to improve efficiency, but it was achieved at the cost of a slight decrease in effectiveness. Scheme 3 performed unsatisfactorily, compared with the other two schemes.

### 4.4. Experimental Study III

Study III explored the proper selection of the number of evolution steps taken by each complex before shuffling, $\beta$. The study was conducted using the recommended parameter settings from Scheme 1 in Study II as the benchmark. The results for $\beta$ values one, $n+1$, $2n+1$, $3n+1$, and $4n+1$, where $\beta = 2n+1$ is the default setting, are given in Table 10. The italic values indicate the best choices for $\beta$ given $n$. There is no clear indication of preferred strategy for selecting $\beta$ according to the results of this

Table 10
Test results on selection of $\beta$

| | | Dimension, $n$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | 2 | 4 | 6 | 8 | 10 | 13 |
| $\beta = 1$ | NS | *10* | 10 | *10* | 9 | *10* | 8 |
| | AFE | *95* | 335 | *906* | 1754 | *5027* | 14895 |
| $\beta = n+1$ | NS | 10 | 10 | 10 | *10* | 10 | 10 |
| | AFE | 96 | 348 | 924 | *1667* | 5199 | 14883 |
| $\beta = 2n+1$[a] | NS | 10 | 10 | 10 | 10 | 10 | *10* |
| | AFE | 103 | 365 | 927 | 1682 | 5949 | *13299* |
| $\beta = 3n+1$ | NS | 10 | *10* | 10 | 10 | 10 | 10 |
| | AFE | 98 | *297* | 1109 | 1892 | 5190 | 14856 |
| $\beta = 4n+2$ | NS | 10 | 10 | 10 | 10 | 10 | 10 |
| | AFE | 99 | 331 | 1001 | 2083 | 6159 | 15119 |

[a] Default value in the SCE-UA algorithm.

particular study. It seems clear, however, that, by setting $\beta$ to too small a value, there is a danger of compromising the effectiveness of the algorithm (see Table 10, the case for $\beta = 1$ and $n = 8, 13$). It was also obvious that if the value for $\beta$ was too large (e.g. $\beta = 4n + 1$), the efficiency is decreased. Therefore, we continue to recommend the use of default value of $\beta = 2n + 1$.

### 4.5. Experimental Study IV

This study examined the choice of the number of offspring produced by each subcomplex, $\alpha$. Again, the recommended settings in Scheme 1 of Study II were used as a benchmark for comparison purpose. The results for $\alpha = 1$ and $\alpha = 2$, where the former is the default setting, are shown in Table 11. The results clearly show that the default setting is far superior to the second setting.

### 4.6. Summary of the results

We have conducted extensive numerical studies to investigate the proper selection of four algorithmic parameters of the SCE-UA method—$p$, $p_{\min}$, $\alpha$, and $\beta$. The first experimental study found that, no matter what parameters were chosen for optimization, the SCE-UA method was consistently able to find the 'true' parameters, provided that a sufficiently large value for $p$, the number of complexes, was given. Experimental Study II investigated three schemes for selecting $p_{\min}$: (1) $p_{\min} = p$; (2) $p_{\min} = p/2$; (3) $p_{\min} = 1$. It was recommended that Scheme 1, which sets $p_{\min}$ equal to $p$, should be used. Study III examined the selection of $\beta$. The results did not give a clear indication of the preferred choice, but did, however, implicate the shortcomings of choosing a value that is too small or too large. Therefore, we continue to recommend that the default value of $2n + 1$ should be used. Study IV clearly showed that the value for $\alpha$ should be set to one, the default value.

## 5. Conclusions

The inability to find the global optima for parameters of conceptual watershed

Table 11
Test results on selection of $\alpha$

| | | Dimension, $n$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | 2 | 4 | 6 | 8 | 10 | 13 |
| $\alpha = 1$[a] | NS | *10* | *10* | *10* | *10* | *10* | *10* |
| | AFE | *103* | *365* | *927* | *1682* | *5949* | *13299* |
| $\alpha = 2$ | NS | 9 | 0 | 0 | 0 | 0 | 0 |
| | AFE | 254 | 0 | 0 | 0 | 0 | 0 |

[a] Default value in the SCE-UA algorithm.

models by conventional optimization methods has, in the past, caused deep frustration among model users and thus limited the usefulness of such models. With the advent of a newly developed global optimization method—the SCE-UA method—we can now produce reliable estimates of global optima for large complex optimization problems. The extensive experimental studies presented here illustrate how to use the SCE-UA method in an efficient and effective manner.

The experimental studies were carried out by using a complex watershed model—the NWSRFS-SMA model—as a test problem. It is our belief, as well as that of many other researchers and practitioners, that this model embodies many typical problems encountered in the calibration of watershed models, and that it is one of the more difficult watershed models to calibrate (Brazil and Hudlow, 1980; Duan et al., 1992). Therefore, the recommended values for the SCE-UA algorithmic parameters derived here can be construed as guidelines for most applications. However, owing to the diverse and peculiar nature of the modeling problems we face every day, it may be necessary that experimental procedures similar to those delineated in this paper be employed to derive the preferred algorithmic parameters for a particular problem.

## 6. Acknowledgments

Partial financial support for this research was provided by the National Science Foundation (Grants ECE-86105487 and BCS-8920851) and by the Hydrologic Research Laboratory of the National Weather Service (Grant NA90AA-H-HY505).

## 7. References

Baffaut, C. and Delleur, J.W., 1989. Expert system for calibrating SWMM. J. Water Resour. Planning Manage., 115(3): 278–298.

Brazil, L.E., 1988. Multilevel calibration strategy for complex hydrologic simulation models. Ph.D. Thesis, Department of Civil Engineering, Colorado State University, Fort Collins.

Brazil, L.E. and Hudlow, M.D., 1980. Calibration procedures used with the National Weather Service river forecast system. In: Y.Y. Haimes and J. Kindler (Editors), Water and Related Land Resources. Pergamon, New York, pp. 457–566.

Burnash, R.J.C., Ferral, R.L. and McGuire, R.A., 1973. A generalized streamflow simulation system: conceptual models for digital computers. Joint Federal–State River Forecast Center, Sacramento, CA.

Duan, Q., Sorooshian, S. and Ibbitt, R.P., 1988. A maximum likelihood criterion for use with data collected at unequal time intervals. Water Resour. Res., 24(7): 1163–1173.

Duan, Q., Sorooshian, S. and Gupta, V.K., 1992. Effective and efficient global optimization for conceptual rainfall–runoff models. Water Resour. Res., 28(4): 1015–1031.

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

Gupta, V.K. and Sorooshian, S., 1983. Uniqueness and observability of conceptual rainfall–runoff model parameters: the percolation process examined. Water Resour. Res., 19(1): 269–276.

Harlin, J., 1991. Development of a process oriented calibration scheme for the HBV hydrological model. Nordic Hydrol., 22(1): 15–36.

Hendrickson, J.D., Sorooshian, S. and Brazil, L., 1988. Comparison of Newton-type and direct search algorithms for calibration of conceptual rainfall–runoff models. Water Resour. Res., 24(5) 691–700.

Holland, J.H., 1975. Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor.

Hooke, R. and Jeeves, T.A., 1961. Direct search solutions of numerical and statistical problems. J. Assoc. Comput. Mach., 8(2): 212–229.

Ibbitt, R.P., 1972. Effects of random data errors on the parameter values for a conceptual model. Water Resour. Res., 8(1): 70–78.

Ibbitt, R.P. and Hutchinson, P.D., 1984. Model parameter consistency and fitting criteria. Proc. IFAC 9th Triennial World Congress, Budapest, 2–6 July, Vol. 4, pp. 153–157.

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

Kuczera, G., 1983a. Improved parameter inference in catchment models: 1. Evaluating parameter uncertainty. Water Resour. Res., 19(5): 1151–1162.

Kuczera, G., 1983b. Improved parameter inference in catchment models: 2. Combining different kinds of hydrologic data and testing their compatibility. Water Resour. Res., 19(5): 1163–1172.

Manetsch, T.J., 1990. Towards efficient global optimization in large dynamic systems—adaptive complex method. IEEE Trans. Syst. Man Cybernet., 20(1): 257–261.

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

Peck, E.L., 1976. Catchment modeling and initial parameter estimation for the National Weather Service river forecast system. NOAA Tech. Memo., NWS HYDRO-31, National Weather Service, Silver Spring, MD.

Pickup, G., 1977. Testing the efficiencies of algorithms and strategies for automatic calibration of rainfall–runoff models. Hydrogeol. Sci. Bull., 22(2): 257–274.

Price, W.L., 1978. A controlled random search procedure for global optimization. In: C.W.L. Dixon and G.P. Szego (Editors), Towards Global Optimization, 2. North-Holland, Amsterdam, pp. 71–84.

Price, W.L., 1983. Global optimization by controlled random search. J. Optimization Theory Appl., 40(3): 333–348.

Pronzato, L., Walter, E., Venot, A. and Lebruchec, J.F., 1984. A general purpose global optimizer: implementation and applications. Math. Comput. Simulation, 26: 412–422.

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

Sorooshian, S. and Dracup, J.A., 1980. Stochastic parameter estimation procedures for hydrologic rainfall–runoff models: correlated and heteroscedastic error cases. Water Resour. Res., 16(2): 430–442.

Sorooshian, S. and Gupta, V.K., 1983. Automatic calibration of conceptual rainfall–runoff models: the question of parameter observability and uniqueness. Water Resour. Res., 19(1): 251–259.

Sorooshian, S. and Gupta, V.K., 1985. The analysis of structural identifiability: theory and application to conceptual rainfall–runoff models. Water Resour. Res., 21(4): 487–495.

Sorooshian, S., Gupta, V.K. and Fulton, J.L., 1983. Evaluation of maximum likelihood parameter estimation techniques for conceptual rainfall–runoff models—influence of calibration data variability and length on model credibility. Water Resour. Res., 19(1): 251–259.

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

Troutman, B.M., 1985a. Errors and parameter estimation in precipitation–runoff modeling: I. Theory. Water Resour. Res., 21(8): 1195–1213.

Troutman, B.M., 1985b. Errors and parameter estimation in precipitation–runoff modeling: II. Case study. Water Resour. Res., 21(8): 1214–1222.

Wang, Q.J., 1991. The genetic algorithm and its application to calibrating conceptual rainfall–runoff models. Water Resour. Res., 27(9): 2467–2471.

Wheater, H.S., Bishop, K.H. and Beck, M.B., 1989. The identification of conceptual hydrological models for surface acidification. Hydrol. Process., 1(1): 89–109.