WATER RESOURCES RESEARCH, VOL. 29, NO. 4, PAGES 1185–1194, APRIL 1993

Calibration of Rainfall-Runoff Models: Application of Global Optimization to the Sacramento Soil Moisture Accounting Model

SOROOSH SOROOSHIAN, QINGYUN DUAN,[1] AND VIJAI KUMAR GUPTA

Department of Hydrology and Water Resources, University of Arizona, Tucson

Conceptual rainfall-runoff models are difficult to calibrate by means of automatic methods; one major reason for this is the inability of conventional procedures to locate the globally optimal set of parameters. This paper investigates the consistency with which two global optimization methods, the shuffled complex evolution (SCE-UA) method (developed by the authors) and the multistart simplex (MSX) method, are able to find the optimal parameter set during calibration of the Sacramento soil moisture accounting model (SAC-SMA) of the National Weather Service River Forecast System (NWSRFS). In the first phase of this study, error-free synthetic data are used to conduct a comparative evaluation of the algorithms under "ideal" conditions. In 10 independent trials of each algorithm in which 13 parameters of the SAC-SMA model were optimized simultaneously, the SCE-UA method achieved a 100% success rate in locating the precise global optimum (i.e., the "true" parameter values) while the MSX method failed in all trials even with more than twice the number of function evaluations. In the second phase, historical data from the Leaf River watershed are used to conduct a comparative evaluation of the algorithms under "real" conditions, using two different estimation criteria, DRMS and HMLE; the SCE-UA algorithm obtained consistently lower function values and more closely grouped parameter estimates, while using one-third fewer function evaluations than the MSX algorithm.

INTRODUCTION AND SCOPE

The successful application of a conceptual rainfall-runoff (CRR) model depends in large measure on how well the model is calibrated. Attempts to calibrate CRR models, such as the Sacramento soil moisture accounting (SAC-SMA) model used by the National Weather Service (NWS) for river and flood forecasting, have been unable to obtain unique optimal parameter estimates using automatic calibration procedures [e.g., *Dawdy and O'Donnell*, 1965; *Nash and Sutcliffe*, 1970; *Chapman*, 1970; *Ibbitt*, 1970; *Monro*, 1971; *Johnston and Pilgrim*, 1976; *Pickup*, 1977; *Sorooshian*, 1978; *Larimore*, 1981; *Sorooshian and Gupta*, 1983; *Sorooshian et al.*, 1983; *Gupta and Sorooshian*, 1985; *Brazil and Krajewski*, 1987; *Hendrickson et al.*, 1988]. The inability to place a reasonable degree of confidence on the estimated parameter values translates into uncertainty regarding the accuracy of the model forecasts. Unless the best set of parameters associated with a given calibration data set can be found, it is difficult to determine how sensitive the parameter estimates (and hence the model forecasts) are to factors such as input and streamflow data error, model error, quantity and quality of data, estimation criterion used, and so on.

One reason that unique optimal parameter estimates are difficult to obtain is that the optimization procedures typically used are not powerful enough to do the job. In a previous paper [*Duan et al.*, 1992], we presented results that clearly established the nature of the problem of multiple optima in CRR models and showed that the optimization problem is more difficult than had been previously thought. The results revealed that in addition to the previously known problems of parameter interaction, nonconvexity of the response surface, and discontinuous derivatives, the problem of multiple optima occurs on at least two scales; at the "large" scale, we find that there is more than one broad "region of attraction" into which a search strategy may converge, while at the "small" scale, each major region of attraction contains numerous widely distributed local minima. The existence of large numbers of minor local optima on the response surface virtually guarantees that local-search optimization procedures will terminate prematurely; thus such procedures have a very low probability of successfully finding the optimal parameter set.

In the same paper, we evaluated the performance of four global search procedures on the research CRR model SIXPAR: the adaptive random search (ARS) method, a combined ARS/simplex method, the multistart simplex (MSX) method, and a new procedure entitled the shuffled complex evolution (SCE-UA) method developed at the University of Arizona by the authors. The results indicated that both the MSX method and the SCE-UA method are effective in consistently finding the globally optimal parameters of the SIXPAR model and that the SCE-UA method is three times more efficient. The ARS and ARS/simplex methods were both found to be inadequate.

In this paper, we apply the SCE-UA and MSX global optimization algorithms to the calibration of the Sacramento soil moisture accounting (SAC-SMA) model of the National Weather Service River Forecast System (NWSRFS). First, error-free synthetic data are used to conduct a comparative evaluation of the algorithms under "ideal" conditions. The model is then calibrated to historical data from the Leaf

[1] Now at NWS/NOAA Hydrologic Research Lab, Silver Spring, Maryland.

Paper number 92WR02617

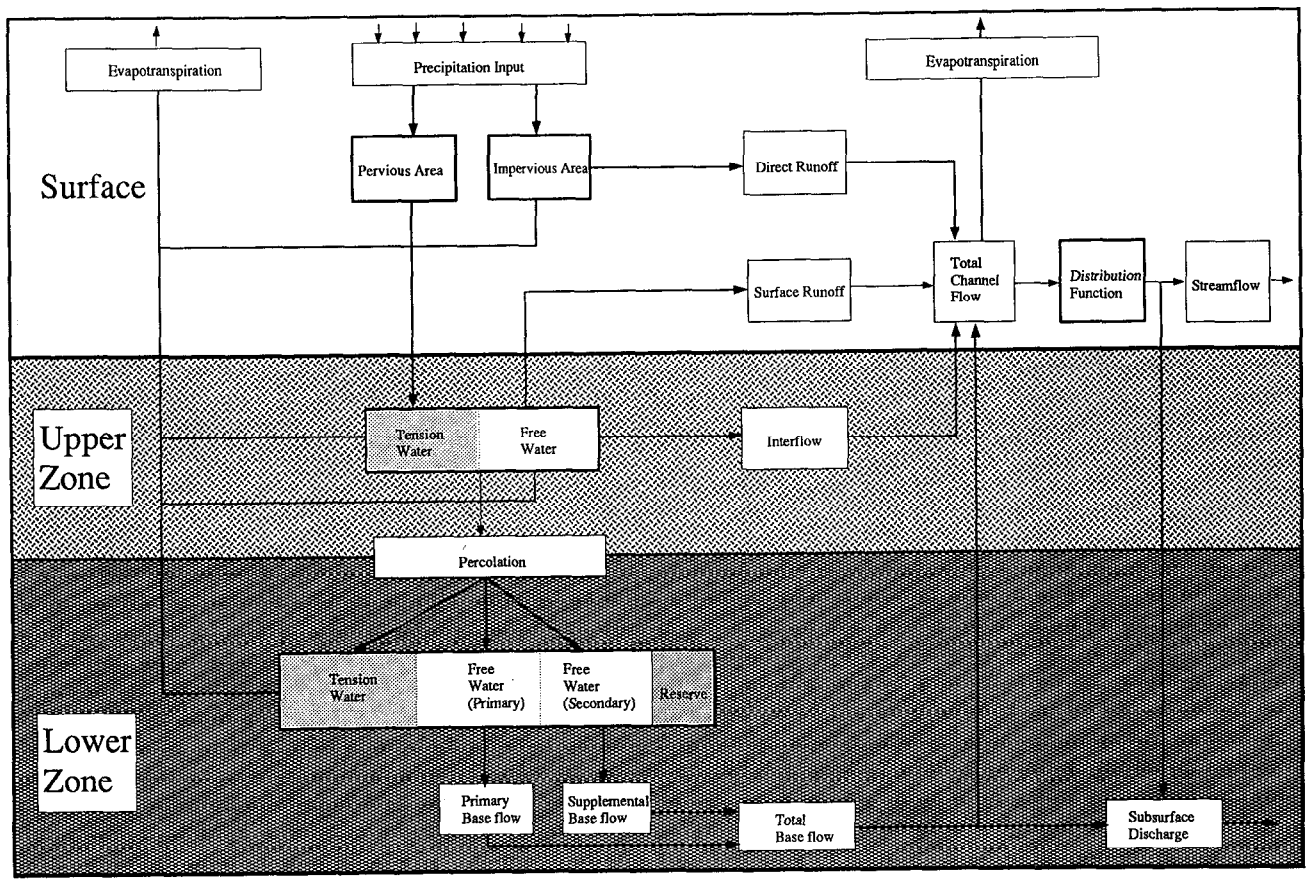

Fig. 1. Schematic representation of the SAC-SMA model.

THE SAC-SMA MODEL

The Sacramento soil moisture accounting (SAC-SMA) model (see Figure 1) is one of the components of the NWSRFS used to convert precipitation inputs into streamflow outputs [*Burnash et al.*, 1973; *Peck*, 1976; *Kitanidis and Bras*, 1980*a*, *b*; *Brazil and Hudlow*, 1981]. The inputs to the SAC-SMA model are precipitation and potential evapotranspiration. Precipitation is provided in the form of mean areal precipitation (average precipitation over the entire soil moisture accounting area). The outputs from the model are estimated evapotranspiration and channel inflow; the latter is converted into streamflow by means of a unit hydrograph. In this study, the research version of the SAC-SMA model maintained by the Department of Hydrology and Water Resources, the University of Arizona, was used.

THE STUDY BASIN AND HYDROLOGIC DATA

The Leaf River basin was chosen for this study. The basin is located north of Collins, Mississippi, with an area of approximately 1950 km^2. A reliable data set is available that represents a variety of hydrological conditions and phenomena. The watershed has been previously investigated by other researchers [e.g., *Brazil and Hudlow*, 1981; *Sorooshian et al.*, 1983; *Brazil*, 1988].

Fifteen water years of data spanning the time period October 1, 1954, to September 30, 1969, were obtained from the Hydrologic Research Laboratory of the National Weather Service (HRL-NWS). The data consist of mean daily streamflow (cubic meters per second), daily potential evapotranspiration estimates (millimeters per day), and 6-hourly mean areal precipitation totals (millimeters per 6 hours). The mean annual precipitation for this period is 1340 mm, and the mean annual runoff is 430 mm.

PARAMETERS OPTIMIZED

The SAC-SMA model has 16 parameters to be determined by the user (see Table 1). Following the recommendation of *Peck* [1976], the three parameters SIDE, RIVA, and RSERV were fixed at prespecified values (see Table 2). The remaining 13 parameters were selected for optimization, and the feasible parameter space was specified by fixing the upper and lower parameter bounds at the values recommended by *Brazil* [1988] (see Table 2). The unit hydrograph used for final routing was predetermined based on the observed hydrograph and was not included in optimization.

THE OPTIMIZATION METHODS

The Multistart Simplex (MSX) Method

The multi-start approach to dealing with multiple optima is to run several trials of any suitable local-search optimization method from different starting points in the feasible space. *Duan et al.* [1992] discussed this approach and showed that if the single-start failure probability $P_f(1)$ is less than approximately 0.8 (i.e., out of 100 trials of the method, started from randomly independent points in the parameter space, fewer than 80 will fail to find the global optimum), such a procedure is theoretically feasible.

The MSX method is based on the well-known simplex local-search procedure [*Nelder and Mead*, 1965]; for details see *Duan et al.* [1992]. The initial simplex for each run is created by randomly selecting $n + 1$ points in the feasible parameter space, where n is the number of parameters to be optimized. *Duan et al.* [1992] found that the single-start failure probability of the simplex method on the SIXPAR model was 0.65, and 12 restarts of a MSX procedure yielded a multistart failure probability of less than 1% (i.e., no more than one out of 100 random trials of the method failed to locate the global optimum).

The Shuffled Complex Evolution (SCE-UA) Method

The SCE-UA method is a general purpose global optimization strategy designed to handle the various response surface problems encountered in the calibration of nonlinear simulation models, particularly the multilevel or "nested" optima problem encountered with CRR models. Detailed descriptions and explanations of the method are given by *Duan et al.* [1992, 1993] and so will not be repeated here. In brief, the SCE-UA method involves the initial selection of a "population" of points distributed randomly throughout the feasible parameter space. In the absence of prior information on the approximate location of the global optimum, use a uniform sampling distribution. The population is partitioned into several "complexes," with each complex consisting of $2n + 1$ points. Each complex is then allowed to "evolve" so as to independently search the parameter space in a manner that is based on an extension of the simplex local-search algorithm. After a prescribed number of steps, the complexes are "shuffled" together and new complexes formed such that the information gained separately by each complex is shared. The evolution and shuffling procedures are repeated until prescribed stopping criteria are satisfied.

TABLE 1. Parameters of SAC-SMA Model

Parameter	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 supplemental 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 base flow, dimensionless
RSERV	fraction of lower zone free water not transferable to lower zone tension water, decimal fraction

TABLE 2. True Parameter Values and Lower and Upper Parameter Bounds Used for 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
LZSK	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.

In this study, a small modification to the SCE-UA method presented by *Duan et al.* [1992] was made to further improve the search efficiency. Preliminary testing indicated that once the algorithm has converged into a small region of the parameter space, it is not necessary to retain a large population of points to conduct the remaining local search. The SCE-UA algorithm was therefore modified as follows. Let p denote the number of complexes used to generate the initial sample population. After each shuffling operation, the number of complexes is reduced by one until the number of complexes reaches a prespecified minimum $p_{\min}$. Thereafter, the number of complexes remains fixed at $p_{\min}$. In the experimental study performed in this work, $p_{\min}$ is set to INT($p/2$), where INT is an operator truncating real numbers into integers. For example, for $p = 20$, the population after the first shuffling operation will be reduced to 19 complexes, and so on until the population after the 10th shuffling operation is reduced to 10 complexes; thereafter, the population size remains fixed at 10. This modification was found to result in approximately a one-third savings in the total number of iterations.

CALIBRATION TRIALS USING SYNTHETIC DATA

Calibration Data

One year of precipitation and evapotranspiration data (October 1, 1955, to September 30, 1956) from the Leaf River watershed were used for the synthetic data calibration studies. This subset of data was chosen because it closely reflects the average conditions of the entire available records and it activates all of the modes of model operation. The parameter set obtained by *Brazil* [1988] for this watershed was used as the "true" parameter set (see Table 2). With this "true" parameter set and the hydrologic input data, a sequence of streamflows was generated. This sequence of streamflows was treated as the "observed" streamflows for the calibration time period (see Figure 2).

Estimation Criterion

The daily root-mean-square (DRMS) estimation criterion used by the Hydrologic Research Lab of the National Weather Service was used for calibration of the model. The DRMS value is obtained by dividing the simple least squares (SLS) value by the number of data points used for calibration

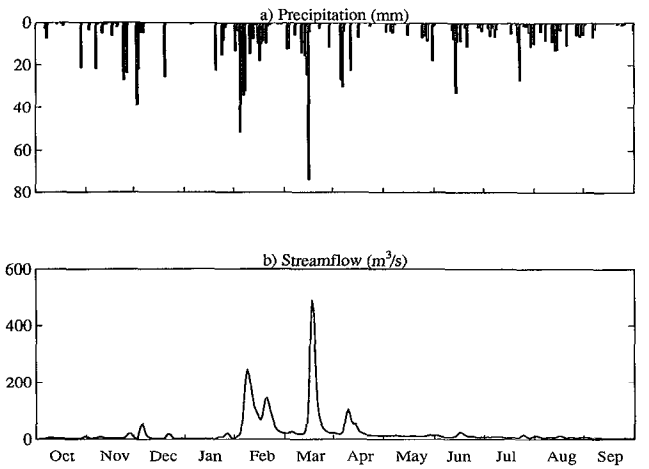

Fig. 2. The real precipitation and synthetic streamflow data (water year 1955–1956) used for the synthetic data study.

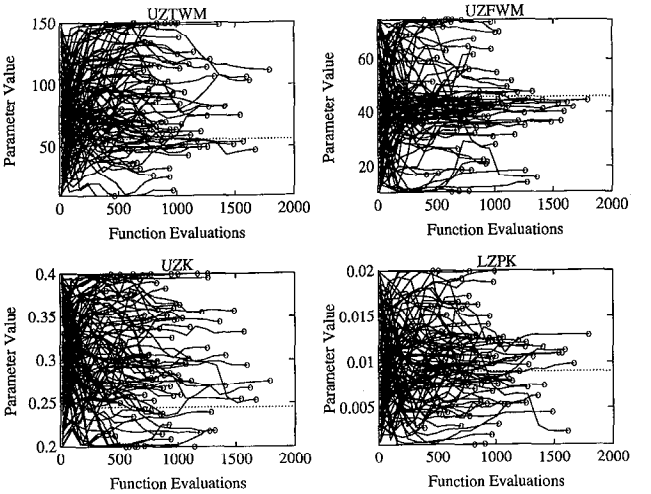

Fig. 3. Convergence behavior of four of the parameters for 100 independent trials of the simplex algorithm (synthetic data study); dotted lines represent the "true" parameter values.

and taking the square root (i.e., DRMS = $(SLS/N)^{0.5}$, where N is the number of data points).

Convergence Criteria

Two criteria were used for termination of the optimization runs. We know, from experience with the model, that a DRMS value of 0.003 or less guarantees that the parameter estimates are accurate to at least 3 significant digits of the "true" optimum. Therefore, an optimization run was considered a success as soon as a DRMS value of 0.003 or less was achieved (objective function convergence). If, however, prior to that occurring, the population of points converged into a small space such that the spread of the population in each parameter direction was smaller than one thousandth of the corresponding feasible parameter range, the run would be terminated because further search would not result in significant improvement of the parameter estimates (parameter convergence). In this case, the optimization run was considered to be a failure.

Testing of the MSX Method

The single-start failure probability of the MSX method on the SAC-SMA model was estimated by conducting 100 independent optimization trials of the simplex method. Figure 3 shows the convergence behavior of the 100 runs for four of the parameters; not one of the runs was able to locate the "true" parameter values, and the termination points were distributed throughout the feasible parameter space. Figure 4 plots the distance of the 100 different termination points from the "true" parameter set. Consider an n-dimensional rectangular box centered on the "true" parameter set, with each side length of the box equal to $\Delta\%$ of the corresponding parameter range. Figure 4 was constructed by counting how many of the termination points fall within the box as Δ varies from 0 to 100. It reveals that none of the 100 runs converged to locations which are close to the "true" parameter set. In fact, only one run converged inside the 20% n-dimensional rectangular box.

For the sake of discussion, suppose we relax our criterion and define a "successful" trial as one for which the final parameter set falls within 20% of the true parameter set. In this case the failure probability of a simplex trial is approximately equal to 0.99. Referring to the theoretical performance curve for a multistart algorithm (see Figure 5, reproduced from *Duan et al.* [1992]) we see that to achieve a 95% probability of success a minimum number of starts greater than 100 would be required.

Nevertheless, 10 independent trials of the MSX algorithm were run on the SAC-SMA model to evaluate the consistency of the algorithm performance. Limitations on computational resources prevented the larger number of trials that would have provided a better statistical representation of algorithm effectiveness. Each trial was run with 50 independent starts of the simplex algorithm; Figures 6*a*–6*d* show the best parameter sets obtained after 1, 10, 20, and 50 starts, respectively, for each trial. Each solid line indicates a parameter set obtained by one of the trials; the x axis of each figure indicates the 13 different parameters, and the y axis indicates the parameter value obtained, normalized to the feasible range (0.0 indicates the lower bound and 1.0 indicates the upper bound). The figures show that some of the parameters (e.g., UZTWM, UZFWM, LZTWM, and

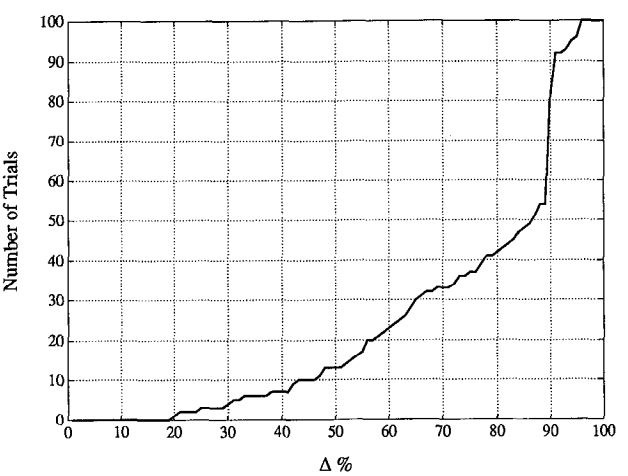

Fig. 4. The number of simplex trials that terminated inside a $\Delta\%n$-dimensional rectangular box centered around the "true" parameter set.

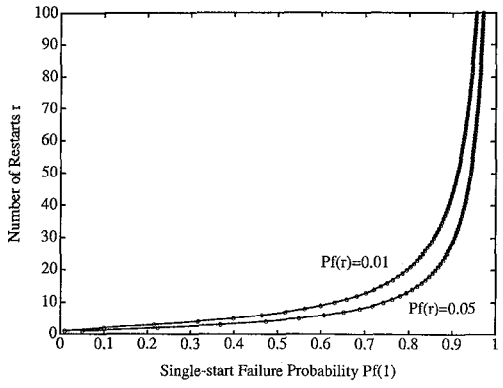

Fig. 5. Theoretical performance curve for any multistart algorithm.

LZFPM) have consistently converged to the same region of the feasible space while others (e.g., LZSK, ZPERC, and REXP) are quite spread out. The results have stabilized by about 20 starts with little improvement between 20 and 50 starts; the marginal benefit of further starts seems minimal. Table 3 gives the best function value and total number of function evaluations after 50 starts; the average number of function evaluations required was 45,887.

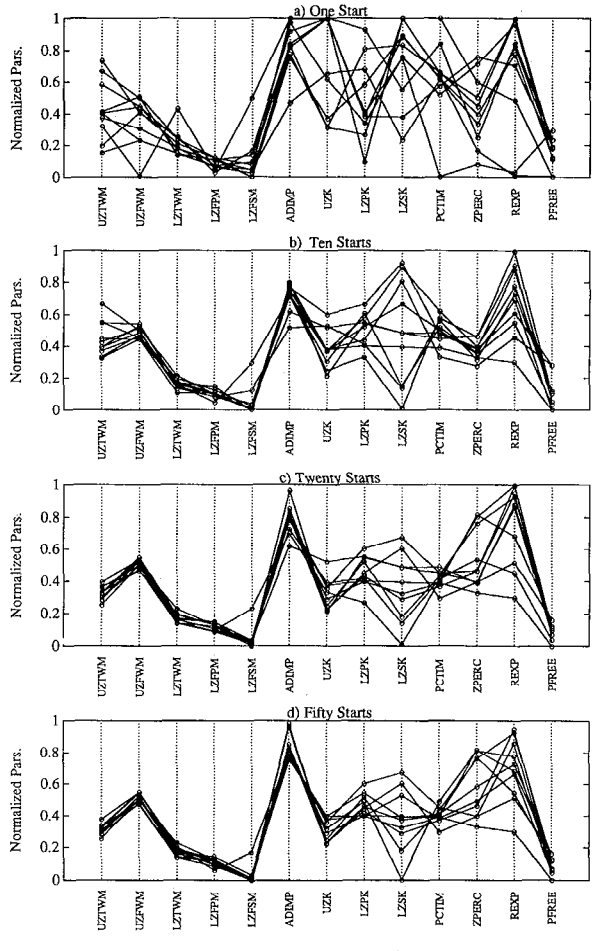

Fig. 6. The best parameter sets obtained after (*a*) 1, (*b*) 10, (*c*) 20, and (*d*) 50 starts for the 10 trials of the MSX algorithm (synthetic data study).

Testing of the SCE-UA Algorithm

The SCE-UA method has only one variable that must be specified by the user: the number of complexes of points in the parameter space to be randomly generated at the beginning of the search. Preliminary testing indicated that a suitable number for the initial number of complexes was 20. Ten independent trials of the SCE-UA algorithm were run on the SAC-SMA model to evaluate the consistency of the algorithm performance. For all 10 trials, the algorithm converged to the "true" values of the parameters; the parameter convergence behavior of the algorithm is displayed in Figure 7 for four of the parameters. Table 3 gives the best function value and total number of function evaluations for each trial; the average number of function evaluations to converge to the "true" parameter set was 23,024.

Discussion

The synthetic data study demonstrates a 100% success rate for the SCE-UA algorithm when calibrating the SAC-SMA model under "ideal" conditions, using 1 year of calibration data from the Leaf River watershed. Under identical conditions, none of the MSX runs was able to exactly locate the global optimum ("true" parameter values) even with more than twice the number of function evaluations; however, the MSX method was consistently able to obtain good estimates for some of the parameters while improving, to some extent, the estimates of the others.

CALIBRATION TRIALS USING HISTORICAL DATA

Calibration and Verification Data

The MSX and SCE-UA algorithms were next used to calibrate the SAC-SMA model using historical data from the Leaf River watershed. Data from the 7–water year period (October 1, 1955, to September 30, 1962) was used for calibration. The precipitation and streamflow data are displayed in Figure 8; for this period, the mean annual precipitation is 1407 mm and the mean annual runoff is 516 mm. To minimize the effects of errors in the estimates of the initial soil moisture storage contents, the 6-month period April 1 to September 30, 1955, was used as a buffer (or warm-up) period (i.e., the estimation criterion computed did not include this period). Data from the subsequent 7-year period (October 1, 1962, to September 30, 1969) was used in the verification phase of the study; the 6-month period April 1 to September 30, 1962 was used as a buffer.

Estimation Criteria

Due to the existence of model and data errors, it is important that an appropriate estimation criterion (measure of the match between the model output and the output behavior of the watershed; also called objective function) be chosen [*Sorooshian and Dracup*, 1980; *Sorooshian*, 1981; *Sorooshian et al.*, 1982, 1983; *Kuczera*, 1982, 1983; *Ibbitt and Hutchinson*, 1984; *Duan et al.*, 1988, *Gan and Burges*, 1990*a*, *b*]. The HRL-NWS commonly uses the DRMS function (presented above) as the estimation criterion. *Sorooshian et al.* [1983] suggest that heteroscedastic errors are frequently encountered in practice and that the heteroscedastic maximum likelihood estimator (HMLE) function (see appendix) that accounts for heteroscedasticity may be more

TABLE 3. Results of Synthetic Data Study

	Trial										
	1	2	3	4	5	6	7	8	9	10	Average
					MSX Method (50 Starts)						
Function evaluations	45,759	46,937	45,958	43,860	43,997	46,150	44,217	48,636	46,577	46,782	45,887
DRMS value	0.6426	0.7824	0.6602	0.7253	0.7157	1.056	0.6957	0.2733	0.6387	0.9433	0.7133
					SCE-UA Method						
Function evaluations	18,762	22,253	21,620	24,402	21,391	23,571	34,537	22,453	16,888	24,361	23,024
DRMS value	0.0024	0.0013	0.0011	0.0014	0.0011	0.0016	0.0016	0.0017	0.0028	0.0011	0.0016

suitable. Both the DRMS and the HMLE estimation criteria were used in the calibration trials reported here.

Stopping Criteria

Due to the existence of data and model errors, it is not appropriate to use an objective function convergence criterion (as in the synthetic data studies) as a basis for termination of the optimization process. In the case of the SCE-UA algorithm, experience with the algorithm indicated that after about 20 shuffling iterations the parameter estimates would stabilize in a region where search would subsequently terminate due to parameter convergence; therefore the SCE-UA calibration runs reported here were terminated when the algorithm completed 20 shuffling iterations. In the case of the MSX algorithm, the calibration runs were terminated after 20 starts.

Calibration Using the MSX Method

For each estimation criterion, 10 independent trials of the MSX algorithm were run on the SAC-SMA model to evaluate the consistency of the algorithm performance. Figures 9*a*–9*d* show the best parameter sets obtained after 1, 5, 10, and 20 starts, respectively, of each trial for the DRMS runs, and Figures 10*a*–10*d* show the same for the HMLE runs. The average number of function evaluations for 20 starts for the DRMS runs was 15,050, while the average for the HMLE runs was 18,266. As with the synthetic studies, some of the parameters seem to be better determined than the others. Also, the HMLE parameter estimates seem better grouped and of more reasonable value than the DRMS parameter estimates; some of the DRMS estimates are at or very close to their bounding values.

Calibration Using the SCE-UA Algorithm

For each estimation criterion, 10 independent trials of the SCE-UA algorithm with 20 initial complexes were run on the SAC-SMA model to evaluate the consistency of the algorithm performance. Figure 11*a* shows the best parameter sets obtained for each DRMS trial, after 20 shuffling iterations, and Figure 11*b* shows the same for the HMLE trials. Note that the HMLE parameter estimates are more tightly grouped than the DRMS estimates.

Discussion

When comparing the MSX and SCE-UA algorithm performance, we see that (1) the SCE-UA best function values are consistently lower (see Figure 12*a* for DRMS and Figure 12*b* for HMLE), (2) the SCE-UA parameter estimates are much more closely grouped together (compare Figure 11*a* with Figure 9*d* for DRMS, and Figure 11*b* with Figure 10*d* for HMLE), and (3) the SCE-UA required fewer function evaluations: 10,748 compared to 15,050 when the DRMS criterion is used, and 10,510 compared to 18,266 when the HMLE criterion is used. These results suggest that the SCE-UA algorithm is more consistent, effective, and efficient than the MSX algorithm in finding the parameter values that minimize the estimation criterion used for model calibration.

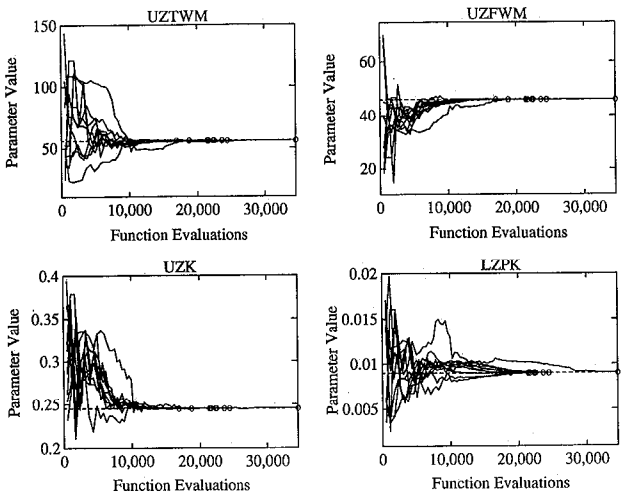

Fig. 7. Convergence behavior of four of the parameters for 10 independent trials of the SCE-UA algorithm (synthetic data study); dashed lines represent the "true" parameter values.

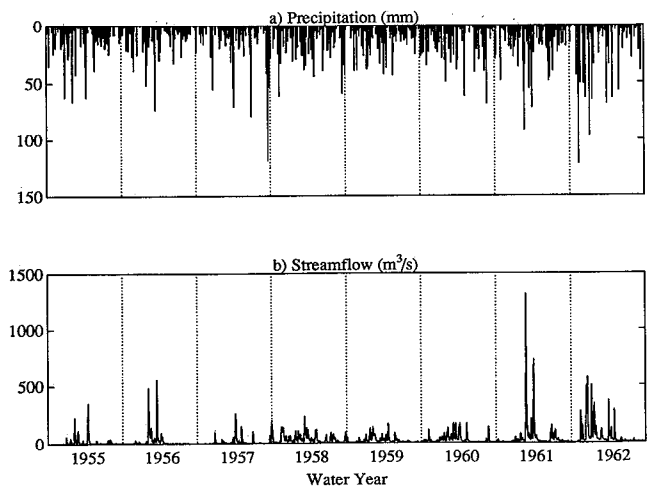

Fig. 8. The precipitation and streamflow data (water years 1955–1962) used for the historical data study.

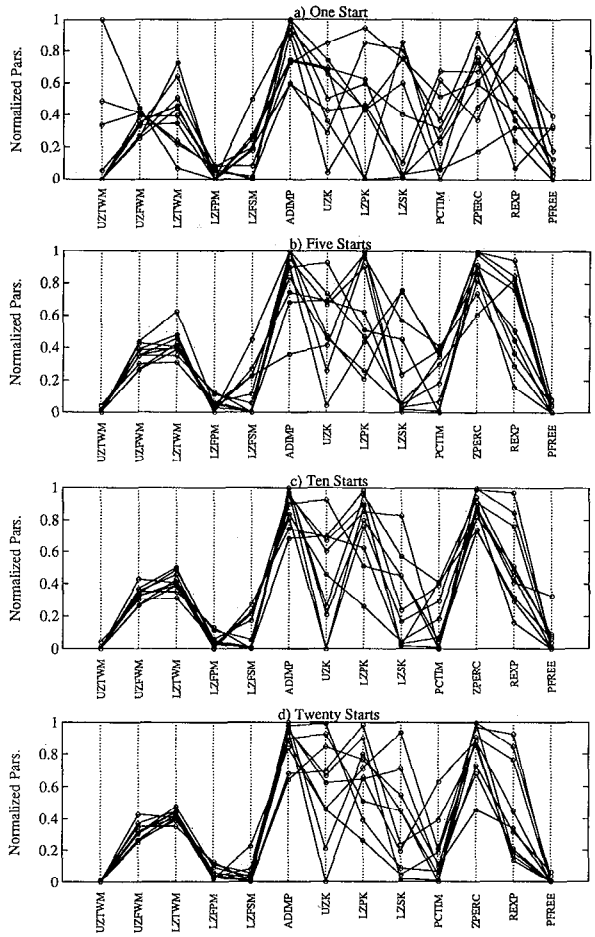

Fig. 9. The best parameter sets obtained after (*a*) 1, (*b*) 5, (*c*) 10 and (*d*) 20 starts for the 10 trials of the MSX algorithm using the DRMS estimation criterion (historical data study).

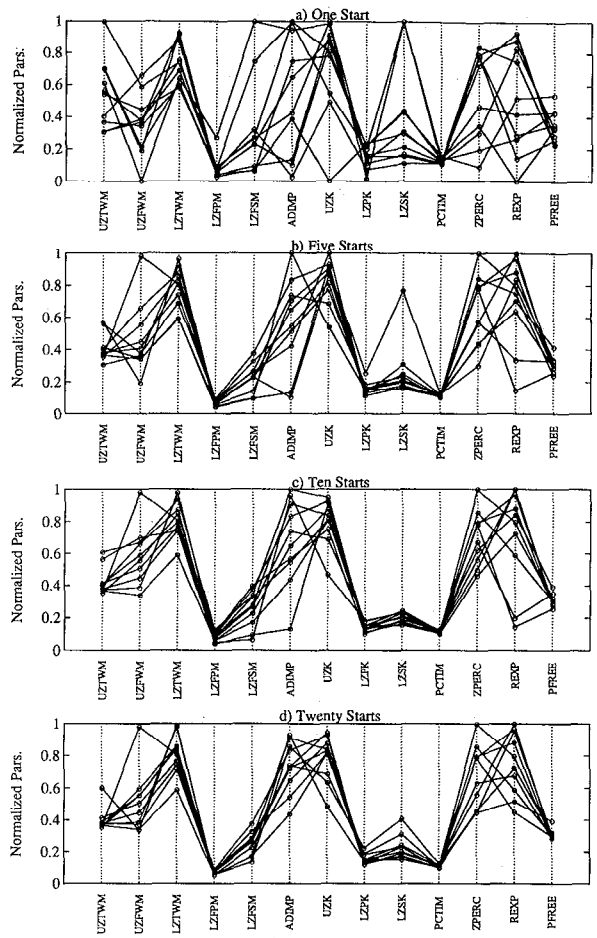

Fig. 10. The best parameter sets obtained after (*a*) 1, (*b*) 5, (*c*) 10 and (*d*) 20 starts for the 10 trials of the MSX algorithm using the HMLE estimation criterion (historical data study).

Notice that some discrepancies were observed in the final parameter estimates even in the SCE-UA trials, especially when DRMS was used as the calibration criterion (e.g., parameters UZK, LZSK, and REXP; see Figure 11*a*). This occurred despite the fact that the final criterion values of all trials were very similar (see Figure 12). This phenomenon is probably attributable to the fact that it is very difficult to define a calibration criterion that wholly captures all aspects of the stochastic nature of the errors in hydrologic data [*Sorooshian and Dracup*, 1980; *Sorooshian et al.*, 1983; *Kuczera*, 1983]. Another possible cause is that the global optimum is not unique because the model contains structural errors and the calibration data do not contain sufficient information. The importance of these issues has been identified in our earlier publications [*Sorooshian and Gupta*, 1983; *Gupta and Sorooshian*, 1983; *Sorooshian and Gupta*, 1985]; these issues require further study.

Verification

Although the main focus of this paper is the abilities and consistency of performance of the global optimization algorithms in finding the parameter values that minimize the estimation criterion, the results of verification studies are also presented here for completeness. The historical data calibration studies reported above constitute four sets of trials: MSX-DRMS, MSX-HMLE, SCE-UA-DRMS, and SCE-UA-HMLE. From each of these sets, the parameter set with the lowest estimation criterion value was selected. These four parameter sets (see Table 4) were used to

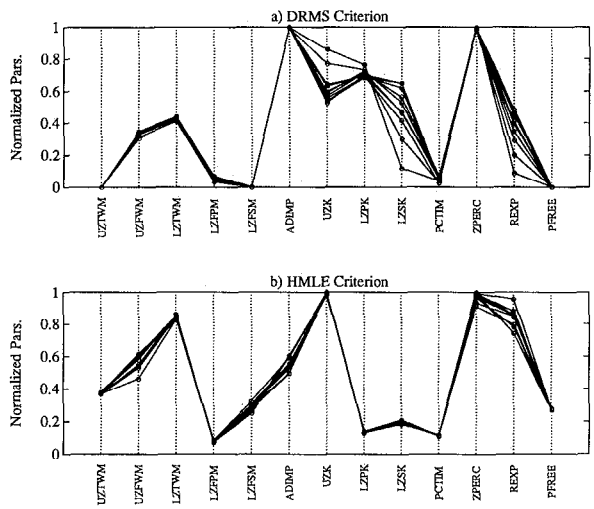

Fig. 11. The best parameter sets obtained for the 10 trials of the SCE-UA algorithm (historical data study): (*a*) DRMS criterion; (*b*) HMLE criterion.

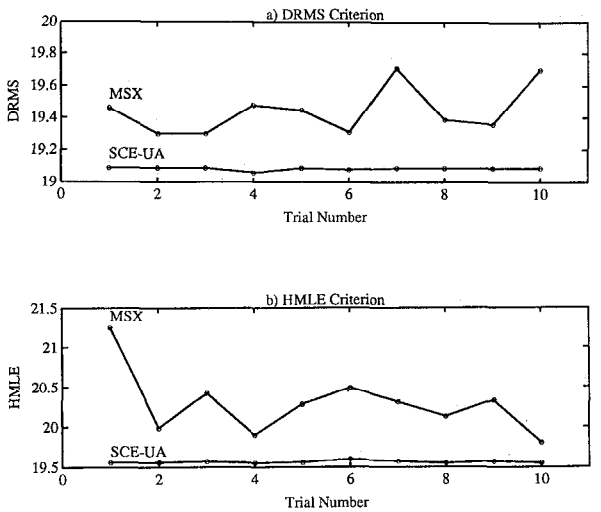

Fig. 12. Comparison of the best function values obtained using the MSX and SCE-UA algorithms: (*a*) DRMS criterion; (*b*) HMLE criterion.

simulate streamflows for the 7–water year verification data set period; three statistics are presented in Table 4: DRMS, HMLE, and percent bias (PBIAS) of the residuals. The percent bias is defined as follows:

$$\text{PBIAS} = 100 \sum_{t=1}^{n} [Q_{\text{sim}}(t) - Q_{\text{obs}}(t)/Q_{\text{obs}}(t)] \quad (1)$$

where $Q_{\text{sim}}(t)$ and $Q_{\text{obs}}(t)$ are the simulated and observed mean flows.

The results indicate that no one parameter set is significantly superior to the others. While the DRMS calibration parameter sets have lower DRMS statistics for the verification period than the HMLE, the DRMS parameter sets tend to overestimate the flows (positive PBIAS) and the HMLE parameter sets tend to underestimate the flows (negative PBIAS).

Summary and Discussion

This paper has investigated the consistency with which two global optimization algorithms are able to find the optimal parameter values during calibration of the SAC-SMA flood forecasting model. In the synthetic data study, a 100% success rate was obtained using the SCE-UA algorithm, while none of the MSX runs were able to exactly locate the global optimum ("true" parameter values) even with more than twice the number of function evaluations. However, the MSX algorithm was consistently able to obtain good estimates for some of the parameters while improving, to some extent, the estimates of the others. In the real data study, the SCE-UA algorithm obtained consistently lower function values and more closely grouped parameter estimates, while using one-third fewer function evaluations than the MSX algorithm.

We have shown that the SCE-UA algorithm is a relatively consistent, effective, and efficient optimization method capable of locating the global optimum in the parameter space. However, as we have discussed in our previous work, the location of the global optimum of CRR models is greatly influenced by a combination of factors such as imperfections in model structure, data errors, and the choice of estimation criteria, among others. Any additional improvements in the performance of models can only be achieved if the above mentioned factors contributing to model nonidentifiability problems are addressed. It is only then that the optimized parameters are accurate representations of their underlying physical values.

Appendix

The HMLE estimator [*Sorooshian* 1978, 1981; *Sorooshian and Dracup*, 1980] is the maximum likelihood, minimum variance, asymptotically unbiased estimator when the errors in the output data are Gaussian, zero mean, and uncorrelated and have nonstationary variance in time. The variance of the errors is assumed to be related to the level of the output (magnitude of the flows). Such errors are believed to be common in streamflow data. The estimator has the form

TABLE 4. Verification Results

Parameter	DRMS Results		HMLE Results	
	MSX	SCE-UA	MSX	SCE-UA
UZTWM	10.38	10.08	63.62	61.61
UZFWM	29.88	30.10	46.02	46.02
LZTWM	217.9	208.5	347.3	344.4
LZFPM	84.90	82.97	125.7	122.3
LZFSM	34.37	10.14	92.33	92.17
ADIMP	0.196	0.199	0.145	0.099
UZK	0.399	0.372	0.389	0.399
LZPK	0.008	0.016	0.003	0.004
LZSK	0.040	0.047	0.064	0.064
PCTIM	0.007	0.004	0.012	0.011
ZPERC	249.4	249.2	159.8	241.8
REXP	1.651	1.359	3.064	3.664
PFREE	0.000	0.001	0.171	0.165
	Statistics			
Calibration period (DRMS)	19.30	19.09		
Forecast period (DRMS)	15.49	16.13	17.72	17.30
Calibration period (HMLE)			19.81	19.55
Forecast period (HMLE)			21.58	20.60
Percent bias	14.67	9.65	−11.74	−10.50

$$\min_{\theta,\lambda} \text{HMLE} = \frac{\frac{1}{n}\sum_{t=1}^{n} w_t \varepsilon_t^2}{\left[\prod_{t=1}^{n} w_t\right]^{1/n}} \tag{A1}$$

where $\varepsilon_t = q_{t,\text{obs}} - q_{t,\text{sim}}$ is the model residual at time t; $q_{t,\text{obs}}$ and $q_{t,\text{sim}}$ are observed and simulated flows respectively; w_t is the weight assigned to time t, computed as

$$w_t = f_t^{2(\lambda-1)} \tag{A2}$$

where $f_t = q_{t,\text{true}}$ is the expected true flow at time t, n is the number of data points, and λ is the unknown transformation parameter which stabilizes the variance. The expected flow $q_{t,\text{true}}$ is approximated using either $q_{t,\text{obs}}$ or $q_{t,\text{sim}}$ [*Sorooshian et al.*, 1983]. *Fulton* [1982] showed that the estimator can become unstable when $q_{t,\text{sim}}$ is used to approximate f_t and recommended using $q_{t,\text{obs}}$. While this is the recommended procedure at present, it should be noted that use of $q_{t,\text{obs}}$ will cause some degree of bias in the estimate of λ [*Gupta*, 1984].

The HMLE is solved in two stages. First, given a set of model parameters, the residuals of the model are obtained. Next an estimate of λ must be obtained; *Sorooshian* [1978] showed that this could be done by solving the following implicit expression to obtain an estimate of the parameter λ using an iterative numerical procedure:

$$\left[\sum_{t=1}^{n} \ln(f_t) \sum_{t=1}^{n} w_t \varepsilon_t^2\right] - n\left[\sum_{t=1}^{n} w_t \ln(f_t) \varepsilon_t^2\right] = 0 \tag{A3}$$

The value of λ is substituted into (A1) and (A2) to compute the value of the HMLE function.

Duan [1991] has devised an equivalent, but more stable, procedure for estimating λ, by rearranging (A3) as follows:

$$R = \frac{R_d}{R_n} - 1 = 0 \tag{A4}$$

where

$$R_d = \sum_{t=1}^{n} w_t \varepsilon_t^2 \tag{A5}$$

$$R_n = \sum_{t=1}^{N} w_t \varepsilon_t^2 a_t \tag{A6}$$

$$a_t = \frac{\ln f_t}{a_d} \tag{A7}$$

$$a_d = \frac{1}{n}\sum_{t=1}^{n} \ln f_t \tag{A8}$$

With this arrangement of terms, the HMLE function value can be computed as

$$\text{HMLE} = \frac{(1/n)R_d}{\exp[2(\lambda-1)a_d]} \tag{A9}$$

The revised procedure for estimating λ and computing HMLE, given $f_{t,\text{obs}}$ and $f_{t,\text{sim}}$, is as follows: (1) Select $f_t = f_{t,\text{obs}}$ or $f_{t,\text{sim}}$ or $(\alpha f_{t,\text{obs}} + \beta f_{t,\text{sim}})$, where $\alpha + \beta = 1$, $\alpha, \beta \geq 0$, and $t = 1, 2, \cdots, n$. (2) Compute a_d, using (A8), and a_t (for $t = 1, 2, \cdots, n$) using (A7). (3) Use an iterative procedure (e.g., golden section, or Fibonacci method) to estimate λ such that $R = 0$ in (A4); if the procedure requires an initial value, use $\lambda = 1$. (4) Compute HMLE using (A9).

Acknowledgments. The authors gratefully acknowledge the constructive comments and suggestions of two anonymous reviewers and those of Bryson Bates of CSIRO Western Australia Laboratories, Perth, which resulted in great improvement of our paper. Technical support was provided by the Center for Computing Information and Technology at the University of Arizona. 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 (grants NA85AA-H-HY088 and NA90AA-H-HY505). Financial assistance provided to Q.D. by the Salt River Project Fellowship is gratefully acknowledged.

REFERENCES

Brazil, L. E., Multilevel calibration strategy for complex hydrologic simulation models, Ph.D. dissertation, 217 pp., Dep. of Civ. Eng., Colo. State Univ., Fort Collins, 1988.

Brazil, L. E., and M. D. Hudlow, Calibration procedures used with the National Weather Service River Forecast System, in *Water and Related Land Resource Systems*, edited by Y. Y. Haimes and J. Kindler, pp. 457–466, Pergamon, New York, 1981.

Brazil, L. E., and W. F. Krajewski, Optimization of complex hydrologic models using random search methods, paper presented at Conference on Engineering Hydrology, Hydraul. Div., Am. Soc. of Cir. Eng., Williamsburg, Va., Aug. 3–7, 1987.

Burnash, R. J. E., R. L. Ferral, and R. A. McGuire, A generalized streamflow simulation system, report, 204 pp., Jt. Fed.-State River Forecast Center, Sacramento, Calif., 1973.

Chapman, T. G., Optimization of a rainfall-runoff model for an arid zone catchment, *UNESCO Publ. 96*, pp. 126–143, U. N. Educ., Sci., and Cult. Organ., Wellington, New Zealand, 1970.

Dawdy, D. R., and T. O'Donnell, Mathematical models of catchment behavior, *J. Hydraul. Div. Am. Soc. Civ. Eng.*, *91*(HY4), 113–137, 1965.

Duan, Q., A global optimization strategy for efficient and effective calibration of hydrologic models, Ph.D thesis, 371 pp., Dep. of Hydrol. and Water Resour., Univ. of Ariz., Tucson, 1991.

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

Duan, Q., S. Sorooshian, and V. K. Gupta, 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. Optimiz. Theory Appl.*, *76*(3), 501–521, 1993.

Fulton, J., Discussion, modification, and extension of some maximum likelihood techniques for model calibration with application to rainfall-runoff models, M.S. thesis, Case Western Reserve Univ., Cleveland, Ohio, Aug. 1982.

Gan, T. Y., and S. J. Burges, An assessment of a conceptual rainfall-runoff model's ability to represent the dynamics of small hypothetical catchments, 1, Models, model properties, and experimental design, *Water Resour. Res.*, *26*(7), 1595–1604, 1990*a*.

Gan, T. Y., and S. J. Burges, An assessment of a conceptual rainfall-runoff model's ability to represent the dynamics of small hypothetical catchments, 2, Hydrologic responses for normal and extreme rainfall, *Water Resour. Res.*, *26*(7), 1605–1619, 1990*b*.

Gupta, V. K., The identification of conceptual watershed models,

Ph.D. dissertation, Case Western Reserve Univ., Cleveland, Ohio, Jan. 1984.

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

Gupta, V. K., and S. Sorooshian, The automatic calibration of conceptual catchment models using derivative-based optimization algorithms, *Water Resour. Res.*, *21*(4), 473–486, 1985.

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

Ibbitt, R. P., Systematic parameter fitting for conceptual models of catchment hydrology, Ph.D. dissertation, Imp. Coll. of Sci. and Technol., Univ. of London, 1970.

Ibbitt, R. P., and P. D. Hutchinson, Model parameter consistency and fitting criteria, paper presented at the IFAC 9th Triennial World Congress, Int. Fed. of Autom. Control, Budapest, July 2–6, 1984.

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

Kitanidis, P. K., and R. L. Bras, Adaptive filtering through detection of isolated transient errors in rainfall-runoff models, *Water Resour. Res.*, *16*(4), 740–748, 1980*a*.

Kitanidis, P. K., and R. L. Bras, Real-time forecasting with a conceptual hydrological model, 1, Analysis of uncertainty, *Water Resour. Res.*, *16*(6), 1025–1033, 1980*b*.

Kuczera, G., On the relationship of the reliability of parameter estimates and hydrologic time series data used in calibration, *Water Resour. Res.*, *18*(1), 146–154, 1982.

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

Larimore, W. E., Applications of the maximum likelihood parameter identification to hydrologic forecasting, *Rep. TR-1480-3*, The Anal. Sci. Corp., Reading, Mass., 1981.

Monro, J. C., Direct search optimization in mathematical modeling and a watershed model application, *NOAA Tech. Memo. NWS Hydro-12*, Natl. Oceanic and Atmos. Admin., U.S. Dep. of Commerce, Silver Spring, Md., 1971.

Nash, J. E., and J. V. Sutcliffe, River flow forecasting through conceptual models, I, A discussion of principles, *J. Hydrol.*, *10*(3), 282–290, 1970.

Nelder, J. A., and R. Mead, A simplex method for function minimization, *Comput. J.*, *7*(4), 308–313, 1965.

Peck, E. L., Catchment modeling and initial parameter estimation for the National Weather Service river forecast system, *NOAA Tech. Memo. NWS Hydro-31*, Natl. Oceanic and Atmos. Admin., U.S. Dep. of Commerce, Silver Spring, Md., 1976.

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

Sorooshian, S., Considerations of stochastic properties in parameter estimation of hydrologic rainfall-runoff models, Ph.D. dissertation, Univ. of Calif., Los Angeles, 1978.

Sorooshian, S., Parameter estimation of rainfall-runoff models with heteroscedastic streamflow errors—Noninformative data case, *J. Hydrol.*, *52*(1/2), 127–138, 1981.

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

Sorooshian, S., and V. K. Gupta, Automatic calibration of conceptual rainfall-runoff models: The question of parameter observability and uniqueness, *Water Resour. Res.*, *19*(1), 260–268, 1983.

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

Sorooshian, S., V. K. Gupta, and J. L. Fulton, Parameter estimation of conceptual rainfall-runoff models assuming autocorrelated data errors—A case study, in V. P. Singh (Editor), *Statistical Analysis of Rainfall and Runoff, Proceedings of the International Symposium on Rainfall-Runoff Modeling, Mississippi State, Mississippi, May 18–21, 1981*, edited by V. P. Singh, pp. 491–504, Water Resources Publications, Littleton, Colo., 1982.

Sorooshian, S., V. K. Gupta, and J. L. Fulton, 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, 1983.

Q. Duan, Hydrologic Research Lab, NWS/NOAA, 1325 East West Highway, Silver Spring, MD 20910.

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

(Received May 6, 1992;
revised October 27, 1992;
accepted November 2, 1992.)