WATER RESOURCES RESEARCH, VOL. 24, NO. 5, PAGES 691–700, MAY 1988

Comparison of Newton-Type and Direct Search Algorithms for Calibration of Conceptual Rainfall-Runoff Models

JENE' D. HENDRICKSON AND SOROOSH SOROOSHIAN

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

LARRY E. BRAZIL

Hydrologic Research Laboratory, U.S. National Weather Service, Silver Spring, Maryland

An examination of the calibration aspect of conceptual rainfall-runoff models was undertaken using the Sacramento soil moisture accounting model and a study comparing the performance of a Newton-type optimization algorithm with that of a direct search algorithm. Results indicate that the direct search algorithm is the more robust of the two because the Newton-type algorithm is more susceptible to poor conditioning of the response surface. Graphical studies of the response surface of the model's parameter space confirmed the presence of discontinuities and a rough-textured surface, particularly in the derivatives.

INTRODUCTION

Mathematical models have become an important tool in the study of hydrology. They are used for scientific study of watershed processes, engineering problem solving, and forecasting and predicting hydrologic phenomena. Model development typically occurs in several stages: problem definition, selection of a suitable type or structure of model, model calibration, and ideally, model verification. Successful completion of this process requires, at a minimum, (1) selection of a model type suitable to the problem at hand, (2) adequate calibration data, and (3) estimation of unique model parameters.

There are two basic approaches to estimation of model parameters: manual and automatic. This paper deals with the automatic approach, which has two major components: (1) the estimation criterion and (2) the optimization algorithm. The choice of estimation criterion has been discussed extensively in recent years [e.g., *Sorooshian and Dracup*, 1980; *Sorooshian et al.*, 1981, 1983; *James and Burges*, 1982; *Sefe and Boughton*, 1982; *Lemmer and Rao*, 1983; *Ibbitt and Hutchinson*, 1984; *Sorooshian and Gupta*, 1983]. Issues related to the choice of optimization algorithm are the topic of this paper. Discussion is restricted to the calibration of conceptual rainfall-runoff (CRR) models such as the Stanford [*Crawford and Linsley*, 1966], Sacramento soil moisture accounting [*Brazil and Hudlow*, 1981], or Boughton [*Boughton*, 1965] models. Much of the discussion, however, may be relevant to other types of nonlinear models.

HISTORICAL PERSPECTIVE

One technique often employed during model calibration is the use of an automated optimization algorithm which systematically searches the parameter space for the extremum of an estimation criterion which in some fashion measures the agreement between observed and simulated flows. Such an approach is especially useful in the latter stage of model "fine tuning" [*Brazil and Hudlow*, 1981]. Use of the automated approach implies a lessened reliance on the subjective judgement of the hydrologist performing the calibration, particularly an advantage when experienced and skilled model calibrators are in short supply. Automated optimization algorithms also speed the calibration process significantly, although they often require more computer time [*James and Burges*, 1982]. However, with recent advances in computer technology, computational restrictions are becoming less severe. While automated parameter estimates are in some sense more objective and reproducible than manual estimates, automated techniques suffer from a lack of mechanisms for maintaining conceptually realistic parameter values. Nevertheless, a review of the literature reveals that the use of automated algorithms is widespread [*Dawdy and O'Donnell*, 1965; *Nash and Sutcliffe*, 1970; *Monro*, 1971; *Clarke*, 1973; *Sorooshian*, 1983; *Isabel and Villeneuve*, 1986; *Wheater et al.*, 1986].

The calibration of a CRR model is often a nonlinear, unconstrained optimization problem. A number of different algorithms have been applied to such problems. Systematic algorithms may be divided into three major classes: direct search, gradient, and second derivative. Random methods [*Karnopp*, 1963; *Bekey and Masri*, 1983; *Pronzato et al.*, 1984] search the parameter space in a random fashion. Research in progress by the third author is exploring the use of these algorithms. Preliminary results seem to indicate that they are best suited for the early stages of estimation. (Terminology in the following paragraphs is largely taken from *Bard* [1974].)

Direct search methods sample the value of the estimation criterion in a systematic manner, without utilizing derivatives of the estimation criterion with respect to parameters. Popular examples of direct search methods include the simplex method [*Nelder and Mead*, 1965], pattern search algorithm [*Hooke and Jeeves*, 1961], and rotating directions method of *Rosenbrock* [1960]. For instance, the pattern search algorithm, which is used in this study, conducts a series of exploratory searches followed by pattern searches. In an attempt to decrease the function value, the exploratory search increments each parameter value in turn. The pattern search travels along the vector defined by the exploratory search.

The difference between gradient and second derivative methods is that the former uses only first derivatives, although this distinction may blur. The use of the term "gradient" used

Paper number 7W5049.
0043-1397/88/007W-5049$05.00

here differs from that of *Bard* [1974], who uses the term for both first and second derivative methods. The most basic of the gradient algorithms is that of steepest descent, which searches along the gradient direction. This method is quite inefficient, however, in most cases, including most quadratic response surfaces.

Newton-type methods are the most widely used of second derivative algorithms. The Newton method, also known as the Newton-Raphson method, uses the supplied values of the first and second derivatives at the current parameter point to construct a quadratic surface and then solves analytically for the minimum (in a minimization problem). The Marquardt version of the Newton method [*Levenberg*, 1944; *Marquardt*, 1963] restricts the search to the space of dominant eigenvectors to prevent long steps in the direction of poorly identifiable parameters, thus optimally interpolating between the steepest descent and Newton methods. The Gauss approximation to the Hessian, or matrix of second derivatives, may be utilized in implementing the Newton method. The Gauss method omits terms containing derivatives higher than first order from the analytical expression of the Hessian, as is described in the appendix.

The Davidon-Fletcher-Powell method [*Fletcher and Powell*, 1963] is a popular method which uses finite difference schemes to recursively approximate the inverse of the Hessian. This method, however, is not strictly a Newton-type method and is considered by some to be a gradient method.

Newton-type methods obtain more information about the response surface at each iteration than do direct search methods and thus should be expected to converge faster, although greater computational effort is required at each iteration to compute the derivatives. The experience of *Bard* [1974] has been that "[Gradient and second derivative] methods, even using finite difference approximations, have outperformed direct search methods on all but the most trivial parameter estimation problems, both in reliability and speed of convergence." On the other hand, *Himmelblau* [1972] states that

> As a general rule in solving unconstrained nonlinear programing problems, gradient and second-derivative methods converge faster than direct search methods. However, in practice, the derivative-type methods have two main barriers to their implementation. First, in problems with a modestly large number of variables, it is laborious or impossible to provide analytical functions for the derivatives needed in a gradient or second-derivative algorithm. Although evaluation of the derivative by difference schemes can be substituted for evaluation of analytical derivatives..., the numerical error introduced, particularly in the vicinity of the extremum, can impair the use of such substitutions.

The nonlinear optimization literature contains examples of comparative algorithm studies involving problems other than hydrologic modeling. This body of literature [*Leon*, 1966; *Wortman*, 1969; *Bard*, 1970; *Himmelblau*, 1972] indicates that gradient and second derivative methods are usually preferable to direct search methods, while, as illustrated below, the experience of those calibrating hydrologic models suggests otherwise. In the optimization literature, use of nonlinear algorithms on well-known test problems almost always yields the correct solution. In such a situation efficient algorithms, those that require the least computational time, are desired. However, in hydrologic optimization problems (as will be illustrated later in this paper), the correct parameter estimates frequently cannot be obtained. Hence hydrologists are also concerned with algorithm robustness, the ability to find the correct solution under a wide variety of conditions.

Little theoretical information is available to assist the hydrologist in selection of an effective algorithm; comparison of algorithms must generally be accomplished through experimentation. A number of researchers have performed comparative studies of algorithms using CRR models and a summary of their findings is given below. However, the results of comparative studies should be considered as a general guideline only. Details of algorithm implementation, termination criteria, heuristic logic introduced into the algorithm on the basis of experience, and characteristics of the test problem can have a significant impact on results.

Gupta and Sorooshian [1985] computed analytic derivatives of a simplified version of the soil moisture accounting model. They then compared the performance of the (direct search) simplex algorithm with the Marquardt-Gauss-Newton algorithm used in this study. Results from synthetic data indicated that both algorithms had similar abilities to find, or not find, optimal parameters, but that the simplex algorithm used significantly more computer time. *Johnston and Pilgrim* [1976] used both the simplex algorithm and the Davidon-Fletcher-Powell method in calibrating the Boughton model on historical data. They found that "Although both methods were reasonably satisfactory, the Simplex method (a direct search method) appeared to be less susceptible to irregularity of the response surface than the Davidon method (a descent method) and was more efficient in the early stages of optimization."

The work of *Ibbitt* [1970] and *Ibbitt and O'Donnell* [1971] is particularly interesting. They investigated the performance of five direct search, one gradient, two Newton-type, and one random search algorithms using the Stanford Watershed model and synthetic data. Ibbitt and O'Donnell found that the "rotating coordinate method of Rosenbrock, after suitable modification for dealing with hydrologic models... is the most [robust] of the nine methods for fitting the model..." However, the Davidon-Fletcher-Powell method was nearly as robust as the Rosenbrock method. Ibbitt and O'Donnell also noted convergence problems associated with methods which assume a quadratic response surface. Some of the problems were attributed to the violation of assumptions regarding continuity of derivatives. *Pickup* [1977] compared the performance of four optimization algorithms on a CRR model using synthetic data. He found the simplex method of Nelder and Mead to be most successful in finding correct parameter values. The Davidon-Fletcher-Powell method performed poorly due to becoming "trapped" on a local minima. It should be noted that the results of Pickup are based on only one calibration run; those of Ibbitt and O'Donnell are based on six calibration runs. However, as results presented later in this paper will indicate, many calibration runs with different initial parameter values may be needed to reliably compare the performance of several algorithms.

MOTIVATION FOR PRESENT WORK

Motivation for the current research stems from the apparent difficulties associated with the development of successful hydrologic models, particularly, the experiences reported in the literature of unsuccessful attempts to find reliable parameter estimates [e.g., *Johnston and Pilgrim*, 1976].

A number of factors may be responsible for poor performance of hydrologic models: inappropriate model structure, error in calibration and operational data, failure of point data to represent mean basin processes, and poorly estimated parameters. Some researchers feel that an inability to obtain truly optimal and unique parameters is a serious obstacle to

TABLE 1. Parameters of the Soil Moisture Accounting Model

Parameter	Explanation
UZTWM	maximum capacity of upper-zone tension storage, mm
UZFWM	maximum capacity of upper-zone free storage, mm
LZTWM	maximum capacity of lower-zone tension storage, mm
LZFPM	maximum capacity of lower-zone primary free storage, mm
LZFSM	maximum capacity of lower-zone secondary free storage, mm
ADIMP	fraction of basin which becomes impervious as all tension storage is met
UZK	lateral drainage rate of upper-zone free storage, fraction/day
LZPK	lateral drainage rate of lower-zone primary storage, fraction/day
LZSK	lateral drainage rate of lower-zone secondary storage, fraction/day
ZPERC	percolation parameter which indicates, when used with other parameters, the maximum possible percolation rate, dimensionless
REXP	percolation parameter, an exponent, determining the rate of change of the percolation rate as the lower-zone moisture varies from full to dry, dimensionless
PCTIM	fraction of basin which is impervious and contiguous with stream channels
RIVA	fraction of basin covered by streams, lakes, and riparian vegetation
PFREE	fraction of percolation water entering free storages, regardless of tension water deficiency
SIDE	ratio of groundwater flow entering channel to that bypassing channel
SAVED	fraction of lower-zone free water unavailable for evapotranspiration

rainfall-runoff modeling. For example, *Johnston and Pilgrim* [1976] state that "Until greater confidence can be placed in the estimation of appropriate parameter values for a particular watersheds, it is unlikely that the potential usefulness of watershed models will be fulfilled." Therefore the current study attempts to improve parameter estimates through selection of a robust optimization algorithm. A robust algorithm is one which is able to reliably obtain correct solutions under a variety of conditions.

In the past, Newton-type algorithms have not been a popular choice for the calibration of CRR models, possibly because of the difficulty of evaluating required derivatives. Since the estimation criterion of a CRR model is not available in closed form, it was thought that analytical derivatives were not available [*Moore and Clark*, 1981]. Finite difference derivatives were regarded with skepticism due to possible inaccuracies associated with numerical approximations. Recently, a method was proposed by *Gupta and Sorooshian* [1985] to analytically compute derivatives of a CRR model.

The primary goal of the current research therefore is to compare the performance of Newton-type and direct search optimization algorithms. Use of analytic first derivatives such as those suggested by *Gupta and Sorooshian* [1985] will avoid the possible disadvantages of numeric derivatives which were present in previous studies.

The remainder of the paper presents the results of the following: (1) response surface study; since gradient methods are particularly sensitive to response surface characteristics, response surface studies were first conducted in order to examine the response surface for smoothness, discontinuities, and convexity, and (2) comparative calibration study; the performance of the pattern search algorithm was compared to that of the Marquardt-Gauss-Newton algorithm using the Soil Moisture Accounting model.

METHODS

The model used here is the version of the Sacramento soil moisture accounting model used by the National Weather Service (NWS) in their National Weather Service River Forecast System (NWSRFS). A description of the model is found elsewhere in the literature [*Brazil and Hudlow*, 1981; *Peck*, 1976; *Kitanidas and Bras*, 1980] and will not be repeated here. A description of model parameters is found in Table 1, however. This model was selected because it is a general purpose CRR model which has received widespread application on watersheds of different sizes and types [*Brazil and Hudlow*, 1981]. The model is part of the NWSRFS and is linked with a unit hydrograph and various other hydrologic routing procedures to provide forecast information for rivers throughout the United States. The model currently is calibrated using manual techniques and a program utilizing the pattern search algorithm. Calibration data typically consist of 6-hour precipitation totals and mean daily discharges. Instantaneous discharges are available for selected events. For the purposes of the calibration study, a unit hydrograph was linked to the model, and mean daily flows were averaged from 6-hour discharges.

For the purpose of facilitating comparison of algorithms, error-free synthetic data were used in both the calibration and response surface studies. When using synthetic data, the location of the global minimum is known, so that determination of the correctness of estimated parameter values is straightforward. This is not the case when using observed data, since true parameter values are unknown. Bird Creek parameter estimates were utilized as "true parameter values." Bird Creek is a 2344-km^2 catchment near Sperry, Oklahoma, characterized by rolling terrain, subhumid climate, and a quickly responding stream. One month of synthetic true flows were created by running the model with Bird Creek parameter estimates. Precipitation data were selected to activate all modes of model behavior. Modes of model behavior were defined by the various combinations of runoff processes such as overland flow, base flow, and impervious area runoff. Precipitation data were varied until the model output contained periods of each possible mode. The precipitation data for the calibration study were changed slightly, while retaining activation of all modes, so that precipitation amounts were more consistent with actual Bird Creek climate. The model code and parameter estimates were supplied by the NWS.

A simple least squares estimator was used in both response surface and calibration studies. This estimator is the sum of the squares of the difference between model and observed streamflow at each time step. According to maximum likelihood (ML) theory, the appropriate form of the estimation

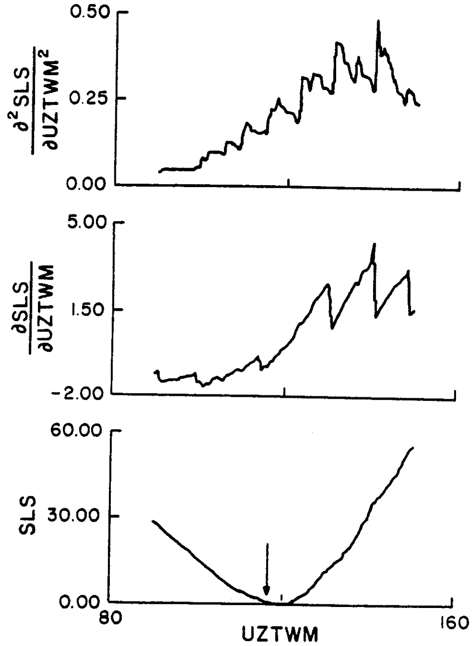

Fig. 1. Cross section and derivatives of SLS function along axis of parameter UZTWM. Arrow marks the location of the discontinuity shown at a smaller scale in Figure 4.

criterion depends on the error structure of the data. *Sorooshian and Gupta* [1983] have shown that for error or noise free synthetic data the difference between a ML and a simple least squares (SLS) estimator is negligible. The appropriate estimator for the current study is therefore SLS, and results and conclusions reported herein would not be changed by the use of a ML estimator.

Analytical first derivatives of the estimation criterion with respect to parameters were computed in accordance with the principles of analytical calculus as described in the work by *Gupta and Sorooshian* [1985]. Since this is a recursive method, a closed-form expression for the derivatives cannot be obtained. To the authors' knowledge, this is the first time that exact analytical first derivatives have been obtained for a complex CRR model. Since calculation of analytical second derivatives is computationally prohibitive, the Gauss method was employed to obtain a first-order approximation of the Hessian. Details on the methods used to calculate derivatives are found in the appendix.

In the course of implementing analytical derivatives, it was necessary to compute numerical derivatives to check the analytical derivatives. When using central finite differences, a 0.1% finite difference step size and a 64-bit computer word, numerical and analytical derivatives were found to be identical to at least the sixth decimal place. It appears that finite difference derivatives are sufficiently accurate for many applications, providing that an appropriate step size is used. Numerical derivatives have the advantage of being easier to implement than analytical derivatives, requiring less computer storage, and resulting in code which is easier to maintain. Comparison of computational efficiency would depend on details of implementation.

RESPONSE SURFACE STUDY

It is generally accepted that the nature of the response surface has a profound impact on optimization. This is particularly true when using a Newton-type algorithm because, for best results, both the response surface itself and its derivatives (with respect to parameters) should be smooth and continuous and the surface should be approximately quadratic. In addition, the "ideal" response surface is convex (in the case of minimization), having a single extremum, and is influenced little by parameter interaction. The form of the estimation criterion should therefore be carefully selected to achieve a well-behaved response surface. For example, it has been shown that use of maximum likelihood estimators can lead to reduced parameter interaction when using historical data [*Sorooshian and Gupta*, 1983] and that the square root of the SLS estimator has poorly behaved derivatives [*Pickup*, 1977].

Many researchers have noted the existence of discontinuities in a CRR estimation criterion or its derivatives [*Restrepo-Pasada and Bras*, 1982; *Pickup*, 1977; *Johnston and Pilgrim*, 1976; *Ibbitt and O'Donnell*, 1971; *Gupta and Sorooshian*, 1985]. The response surfaces of CRR models also tend to suffer from a high degree of nonlinearity, ridges and valleys resulting from parameter interaction and contain multiple optima ("potholes"). These features conspire to make the calibration of CRR models an extremely difficult task.

The response surface of the soil moisture accounting model was examined by plotting one-dimensional cross sections of the response surface and its first and second derivatives along the parameter axes of interest. Response surface plots for three parameters are shown in Figures 1–3. The bottom plot is of the SLS criterion, the middle plot is of analytical first derivatives, and the top plot is of the (approximated) second derivatives. In each case, the parameter value is represented along the x axis. Since neither the estimation criterion nor its derivatives are available in closed form, the plots were generated by sampling the criterion at 100 closely spaced points. Therefore the existence of discontinuities must be inferred from abrupt changes in value.

In Figure 1, parameter UZTWM (upper-zone tension water maximum) represents the amount of moisture that the upper

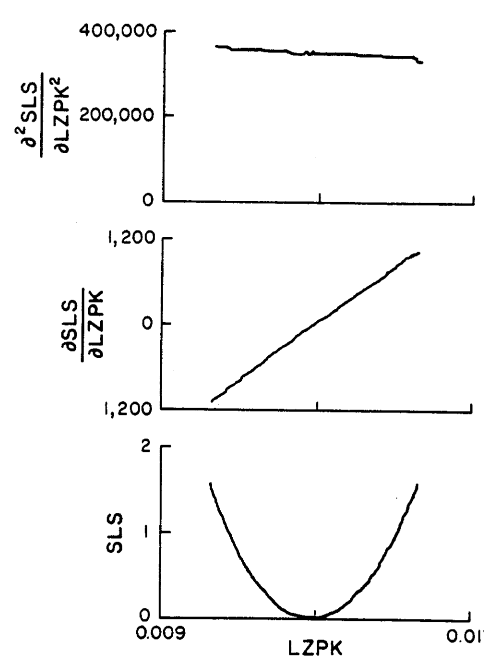

Fig. 2. Cross section and derivatives of SLS fu iction along axis of parameter LZPK.

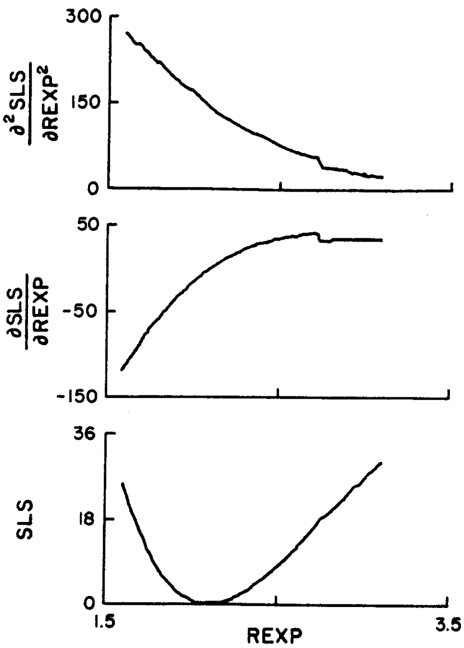

Fig. 3. Cross section and derivatives of SLS function along axis of parameter REXP.

soil zone will absorb before surface runoff occurs and is the parameter associated with the most abrupt threshold in the model. LZPK (lower-zone free primary recession), which is represented in Figure 2, is a recession parameter associated with the production of baseflow. REXP, which is a parameter associated with the complex, nonlinear percolation function that is a special feature of the model, is represented in Figure 3.

Plots at increasingly finer scales failed to discover the exact locations of the discontinuities; they were always located somewhere between two sampling points. Also, it was discovered that discontinuities exist at many different scales. The apparently smooth SLS function in Figure 1, in fact, contains the small discontinuity shown at a finer scale in Figure 4. The arrow in Figure 1 marks the location of the discontinuity in Figure 4.

Plots were made of 16 model parameters, although only three are shown here (Figures 1–3). They reveal that the response surface is convex, at least in the vicinity of the minimum, a condition necessary for the convergence of most algorithms. However, discontinuities and a rough or bumpy texture are also present, particularly in the derivatives. The fact that the derivatives have a rougher texture than the estimation criterion may explain why the Newton-type algorithm performed more poorly than the direct search algorithm in the study presented below. Recall that direct search algorithms base the next "step" on the SLS values at a number of points; Newton-type algorithms typically base the next step on the SLS value and derivatives at a single point. If that single point is located at an irregularity in the response surface, the algorithm will act on misleading information.

Extensive sampling of the soil moisture accounting model's estimation criterion and its derivatives using both synthetic and historical data has always yielded finite derivatives. This, coupled with the inability to exactly locate the discontinuities, leads to the conclusion that both the criterion and its first and second derivatives are piecewise continuous and everywhere finite. It is to be expected that the discontinuities may adversely affect the performance of gradient and second derivative methods. They will not, however, prevent their use. As a rule, these algorithms require that the response surface be continuous and differentiable and that the derivatives be continuous. A more appropriate requirement would be that the criterion and derivatives be piecewise continuous and everywhere finite. While large-scale discontinuities would be a problem for any algorithm, small-scale discontinuities such as shown in Figure 4 would probably have little effect on a Newton-type algorithm.

An investigation was undertaken to discover the cause of the discontinuities. Using a version of the soil moisture accounting model in which threshold structures were replaced with smoothing functions, *Restrepo-Posada and Bras* [1982] found discontinuities in a log likelihood criterion. They suggested that the discontinuities were due to the model's variable internal time step which is a function of the amount of moisture percolating from the upper to lower soil zones. This variable time step results in a more exact integration in time than a fixed time step provides. A discontinuity could occur if the number of internal time steps changes as parameter values change. Another possible cause of discontinuities is the existence of thresholds in the structure of the original model, in conjunction with a discrete time step. The dominant threshold structure in the soil moisture accounting model is the upper-zone tension reservoir, whose magnitude is described by parameter UZTWM. Precipitation must fill the upper-zone tension reservoir, which may only be depleted by evapotranspiration, before it is available for surface runoff or percolation. A discontinuity could occur if surface runoff does not occur for some value of UZTWM, but does occur for a smaller value of UZTWM.

The following procedures were used to determine if discontinuities were being caused by either of the above two mechanisms. The variable time step was removed from the

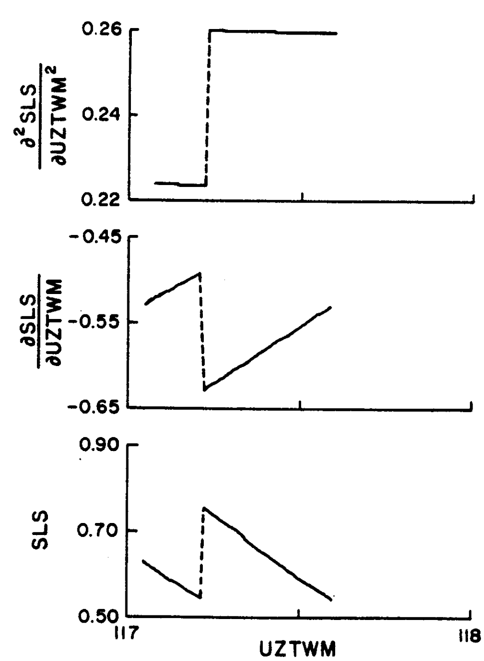

Fig. 4. Example of a discontinuity in a SLS function and its derivatives. This figure is a close-up of a very small discontinuity shown at full scale in Figure 1.

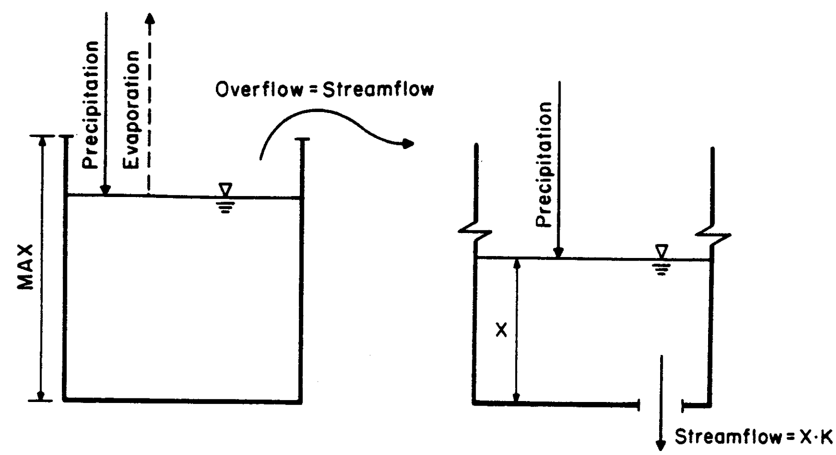

Fig. 5. Diagrams of models BOX (left) and HOLE (right) which were used to study response surface discontinuities.

model, and resulting response surface plots were examined. There was no noticeable decrease in discontinuities. Next, two extremely simple CRR models, shown in Figure 5, were formulated. Model BOX generates streamflow only when cumulative rainfall minus evaporation exceeds the threshold, causing the reservoir to overflow. Model HOLE has a variable internal time step identical to the one found in the soil moisture accounting model and produces runoff through recession parameter K [flow = K (reservoir contents)]. The primary feature of model BOX is its threshold behavior, and the primary feature of model HOLE is the variable time step.

Response surface plots were then produced using models BOX and HOLE. The thresholds of model BOX resulted in discontinuities, while the variable time step of model HOLE did not. From these results, the authors conclude that the major cause of discontinuities for the soil moisture accounting model is the existence of threshold structures such as the tension reservoirs and not the variable time step. Although the work of *Restrepo-Posada and Bras* [1982] suggests that there may be other causes of discontinuities, the complexity of the soil moisture accounting model precludes an exhaustive search for them. It is likely that thresholds are the major cause of discontinuities associated with other CRR models.

Kitanidas and Bras [1980] introduced a version of the soil moisture accounting model (used by *Restrepo-Posada and Bras* [1982]) in which abrupt thresholds were replaced with "*S*" curves. Such a change would be expected to replace discontinuities with smooth S-shaped jumps in the value. However, such smooth jumps would introduce perturbations in the derivatives and not completely solve the problem of nonsmoothness.

Comparative Calibration Studies

The calibration studies described here compared the performance of the (direct search) pattern search algorithm of *Hooke and Jeeves* [1961] with that of a Marquardt-Gauss-Newton method, hereafter referred to as the Newton algorithm for purposes of brevity. The pattern search algorithm was chosen because it is currently used by the NWS to calibrate the soil moisture accounting model. These two algorithms were felt to be reasonably representative of commonly used direct search and Newton-type methods, respectively. It is doubtful that use of different algorithms would have changed the final conclusions that direct search methods tend to be more robust and that Newton-type methods more computationally efficient. Code for the pattern search method was supplied by the NWS, and code for the Newton method was supplied by *Gupta* [1984]. Readers interested in details of algorithm implementation should contact the authors.

Care was taken to ensure that similar convergence criteria were used for both algorithms. Due to the differing structure of each algorithm, however, it was not possible to make convergence criteria identical. A calibration was terminated if the estimation criterion fell below a very small value (0.05 (m^3/s^2)), if the function value changed less than 1% in one iteration (Newton routine) or one pattern (pattern routine), or if a maximum number of iterations were exceeded. The maximum number of iterations were proportional to the number of parameters being optimized and were 2.4 times greater for pattern search than for Newton runs. The maximum iteration criterion did not cause convergence very often, and function convergence and minimum function value were most frequently the cause of termination. An additional pattern search convergence criterion (maximum number of unsuccessful patterns) caused termination a few times.

Initial calibrations were performed using the root mean square criterion ($[\text{SLS/number data points}]^{1/2}$), which is usually used by the NWS to calibrate the model. It was found that the Newton algorithm performed very poorly using the root mean square criterion. Plots of the response surface revealed that the SLS estimator was roughly quadratic as required by Newton-type algorithms, at least near the minimum, while the root mean square estimator was not. This finding agrees with that of Pickup [1977]. SLS was therefore used for the calibration studies presented here. The performance of the pattern search algorithm was virtually the same for both estimators.

Three kinds of calibration runs were performed: single-, two-, and four-parameter runs. Single-parameter optimization with error-free synthetic data is a straightforward problem, and success should be expected. All 14 model parameters which are manually or automatically estimated were included in the single-parameter study. For each run, the parameter of interest was perturbed 35%, and both algorithms were used in an attempt to recover the true value. As can be seen from Table 2, the pattern search algorithm was successful in finding all but 1 of the 14 correct values, while the Newton algorithm failed for 5 of the 14 parameters.

Two measures of the success of an optimization run were considered: (1) how close the final estimation criterion value

TABLE 2. Ending Values for One-Parameter Calibrations

Parameter	Initial Value	True Value	Newton Ending Value	Pattern Ending Value
UZTWM	52.0	80.0	60.6*	80.8
UZFWM	10.0	15.0	15.0	15.0
LZTWM	107.0	160.0	107.0*	159.6
LZFPM	189.0	140.0	140.0	139.9
LZFSM	19.0	14.0	14.0	14.0
ADIMP	0.11	0.17	0.17	0.17
UZK	0.20	0.30	0.30	0.30
LZPK	0.018	0.013	0.013	0.013
LZSK	0.170	0.126	0.126	0.126
ZPERC	31.0	48.0	48.0	47.9
REXP	1.37	2.10	2.39*	2.10
PCTIM	0.002	0.001	0.001	0.001
PFREE	0.027	0.020	0.811*	0.026*
SIDE	2.31	3.55	2.31*	3.55
Average Newton CPU = 25.2/run				
Average pattern search CPU = 45.6/run				

*Unsuccessful run.

was to zero and (2) how close the final parameter values were to the true values. In order to easily interpret the results of individual calibration runs, an overall performance index was devised. Final estimation criterion values were grouped into five ranges which were designated ratings of excellent, good, fair, poor, and very poor. Similar ratings were assigned to the percent closeness of the final parameter values to the true values, averaged between the calibrated parameters. The two ratings were then averaged. The resulting index emphasizes robustness, or the ability to obtain correct solutions, and does not consider computational efficiency.

Four sets of two parameters each were selected for the two-parameter calibrations. The ending parameter values of the two-parameter calibration runs are listed in Table 3. Table 4 sets forth the overall performance index, ending estimation criterion value, and CPU for each run. A 35% perturbation was used for the two-parameter calibrations.

A single-parameter set, ADIMP, UZK, LZPK, and REXP, was used for the four-parameter calibrations. All of these parameters were perturbed 15, 35, and 50% in separate runs. The ending parameter values for the four-parameter runs are contained in Table 5. Each run's overall performance index, ending estimation criterion value, and CPU are found in Table 6.

TABLE 3. Ending Parameter Values for Two-Parameter Calibration Runs

Parameter	Initial Value	True Value	Newton Ending Value	Pattern Ending Value
		Set 1		
ADIMP	0.110	0.170	0.125	0.167
PCTIM	0.002	0.001	0.000	0.003
		Set 2		
UZK	0.200	0.300	0.344	0.300
LZPK	0.018	0.013	0.013	0.013
		Set 3		
ADIMP	0.110	0.170	0.174	0.170
UZK	0.200	0.300	0.323	0.300
		Set 4		
LZPK	0.018	0.013	0.013	0.013
PCTIM	0.002	0.001	0.000	0.002

TABLE 4. Statistics for Two-Parameter Calibration Runs

Parameter Set	Overall Performance Index	Estimation Criterion Value, (m^3/s^2)	CPU
1 (Newton)	very poor	752.	25
2 (Newton)	poor	190.	27
3 (Newton)	fair	92.4	33
4 (Newton)	very poor	169.	29
1 (pattern)	poor	0.26	64
2 (pattern)	excellent	0.03	124
3 (pattern)	excellent	0.00	173
4 (pattern)	fair	0.33	257

The results of the above calibration runs indicate that the pattern search algorithm is more robust than the Newton algorithm when calibrating the soil moisture accounting model. It is significant that for over half the multiparameter sets, neither algorithm was successful in estimating the correct parameters, even under the ideal conditions of error-free synthetic data. Clearly, parameter estimation for the soil moisture accounting model is not a trivial task. A detailed analysis of the reasons for the poor performance of the algorithms was not made. Undoubtedly, response surface characteristics such as roughness, discontinuities, multiple optima, and ridges and valleys associated with parameter interaction play an important role. In some cases the Newton algorithm may have failed to recover from misleading derivatives at the initial parameter point due to an irregularity in the response surface. It is possible that in cases where the Newton algorithm completely failed that the quadratic assumption was poor.

The pattern search calibration runs used slightly more than two and one half (2 1/2) times as much total computer run time than did the Newton runs. This is probably due to the Newton algorithm obtaining more information at each iteration and therefore requiring fewer iterations. Apparently, the additional computer run time required to compute the derivatives is more than offset by the savings associated with fewer iterations.

Other researchers [*Johnston and Pilgrim*, 1976; *Ibbitt and*

TABLE 5. Ending Parameter Values for Four-Parameter Synthetic Calibrations

Parameter	Initial Value	True Value	Newton Ending Value	Pattern Ending Value
		15% Perturbation		
ADIMP	0.140	0.170	0.170	0.170
UZK	0.260	0.300	0.299	0.300
LZSK	0.150	0.126	0.150	0.129
REXP	1.790	2.100	2.255	2.119
		35% Perturbation		
ADIMP	0.110	0.170	0.169	0.169
UZK	0.200	0.300	0.300	0.300
LZSK	0.170	0.126	0.101	0.109
REXP	1.370	2.100	1.924	1.985
		50% Perturbation		
ADIMP	0.090	0.170	0.170	0.166
UZK	0.150	0.300	0.301	0.298
LZSK	0.190	0.126	0.176	0.084
REXP	1.050	2.100	2.489	1.830

TABLE 6. Statistics of Four-Parameter Calibration Runs

Algorithm	Overall Performance Index	Estimation Criterion Value, (m^3/s^2)	CPU
	15% Perturbation		
Newton	fair	6.46	27
Pattern	excellent	0.13	141
	35% Perturbation		
Newton	good	0.35	27
Pattern	good	0.59	142
	50% Perturbation		
Newton	fair	11.9	30
Pattern	fair	6.11	73

O'Donnell, 1971] have suggested that the sequential use of several optimization algorithms may produce better results than using one algorithm alone. It is possible that an apparent local minima which causes one algorithm to converge can be more readily escaped by another. *Isabel and Villeneuve* [1986] caution, however, that changes in the strictness of convergence criterion can effect the apparent robustness of various algorithms. Nevertheless, attempts were made to improve parameter estimates through sequential use of both algorithms. On unsuccessful multiparameter Newton runs, a pattern search calibration was made with initial parameter estimates equal to final Newton values. The reverse was done for unsuccessful pattern search runs, and results are set forth in Table 7. This procedure led to improved parameter estimates for 7 of 11 unsuccessful runs. In two of the seven multiparameter sets, sequential use of both algorithms produced excellent results, while use of either algorithm alone did not. Excellent results were obtained by a single algorithm for three sets, but excellent results could not be obtained by any method or combinations of methods for the final two sets.

The results of the calibration study indicated that the outcome of any calibration run was highly dependent on characteristics of the response surface in the vicinity of the initial parameter point and other factors. It appears that more than a few calibration runs using different parameters and different initial parameter points are necessary before comparison between optimization algorithms becomes meaningful.

A four-parameter calibration was performed using 3 years of historical Bird Creek data. The two algorithms converged to markedly different points in parameter space, although the final estimation criteria were similar: 341,000 and 304,000 (m^3/s^2). Three year split sample verification runs were also made. The fact that both parameter sets produced similar sequences of simulated flows for both calibration and verification data highlights the degree of parameter interaction within the model.

CONCLUSIONS

The method proposed by *Gupta and Sorooshian* [1985] provides the capability of computing analytic derivatives for CRR models. The main purpose of the current research was therefore to compare the use of a direct search and a Newton-type algorithm (using analytical first derivatives) for use in calibrating CRR models. The pattern search and Marquardt-Gauss-Newton algorithms, error-free synthetic data, and the soil moisture accounting model of the NWS were used for the investigation.

The primary finding is that the direct search algorithm was more likely than the Newton-type algorithm to find accurate parameter estimates, although the latter used less computer time. The primary reason for the lack of robustness of the Newton-type algorithm appears to be poor conditioning of the response surface due to discontinuities and, most importantly, lack of smoothness in the estimation criterion and its derivatives. It appears that Newton-type algorithms are not well-suited to the characteristics of CRR response surfaces. The authors believe that results would be similar for other algorithms or other CRR models. Moreover, the authors believe that none of the currently popular systematic search algorithms, used alone, are sufficiently robust for the difficult response surfaces associated with the calibration of CRR models. This belief is supported by the failure of either algorithm, used alone, to find correct parameter estimates in a significant number of cases in the current study. Additional findings are as follows.

1. The least squares response surface of the soil moisture accounting model was found to be rough textured and contain discontinuities. The first and second derivatives of the response surface (with respect to parameters) were more roughly textured and contained more numerous and more severe discontinuities than the surface itself, a factor which contributed to the relatively poor robustness of the Newton-type algorithm. Derivatives were found to exist on many scales.

2. The discontinuous estimation criterion and its derivatives were found to be piecewise continuous and everywhere finite. It is to be expected that the discontinuities will have some adverse impact on Newton-type algorithms. They will not, however, prevent their use.

3. The major cause of discontinuities in a CRR response surface was found to be the existence of thresholds in the

TABLE 7. Overall Performance Indices for Calibration Runs Using Both Algorithms Sequentially

Parameter Set	Newton	Pattern	Newton-Pattern	Pattern-Newton
Two-parameter 1	very poor	poor	fair	excellent
Two-parameter 2	poor	excellent	excellent	NP
Two-parameter 3	fair	excellent	excellent	NP
Two-parameter 4	very poor	fair	poor	fair
Four-parameter (15%)	fair	excellent	good	NP
Four-parameter (35%)	good	good	good	good
Four-parameter (50%)	fair	fair	excellent	fair
Average CPU	28	144	154	158

NP, calibration was not performed because initial calibration with first algorithm produced excellent results.

model, in conjunction with a discrete time step. For example, the worst discontinuities were associated with the tension reservoir which must fill before surface runoff can occur.

4. Analytic and finite difference first derivatives were found to be identical to at least the sixth decimal place. It appears that numerical derivatives are sufficiently accurate for many applications, providing that an appropriate finite difference step size is used.

5. The accuracy of parameter estimates should be enhanced by any technique which would result in better conditioning of the response surface such as the use of maximum likelihood estimators. When using error-free synthetic data, a least squares estimator approximately satisfied the Newton assumption of a quadratic response surface, at least in the vicinity of the optimum, while a root mean square estimator did not.

The authors believe that future research should focus on the roles of model structure and calibration data in the optimization problem, and on methods to alleviate the problems of rough response surface texture, discontinuities, multiple optima, parameter interaction, and parameter nonidentifiability.

APPENDIX

The first and second derivatives of a SLS function are as follows:

$$\frac{\partial SLS}{\partial \theta_i} = 2 \sum_{t=1}^{n} (qs_t(\boldsymbol{\theta}) - qo_t) \frac{\partial qs_t(\boldsymbol{\theta})}{\partial \theta_i}$$

$$H_{ij} = 2 \sum_{t=1}^{n} \frac{\partial qs_t(\boldsymbol{\theta})}{\partial \theta_i} \frac{\partial qs_t(\boldsymbol{\theta})}{\partial \theta_j} + (qs_t(\boldsymbol{\theta}) - qo_t) \frac{\partial^2 qs_t(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}$$

where

- $\boldsymbol{\theta}$ model parameters;
- i, j parameters of interest;
- $qs_t(\boldsymbol{\theta})$ simulated flows at time t;
- qo_t observed flows at time t;
- H_{ij} element of Hessian;
- n the number of data points.

The Gauss approximation to the Hessian omits the last term containing second derivatives of model flow, resulting in a first-order approximation utilizing only first derivatives.

Details of the methodology used to analytically compute first derivatives of model flow are given in the work by *Gupta and Sorooshian* [1985]. As each model state and output variable is updated, the derivative of that variable is computed recursively. In practice, one line of derivative code must be inserted after each line of model code for each parameter for which derivatives are desired. Using the chain rule of calculus, the derivatives are computing using

$$\frac{\partial X_t}{\partial \theta_i} = \frac{\partial f(\boldsymbol{\theta}, X_{t-1}, \mathbf{u}_t)}{\partial \theta_i} + \frac{\partial f(\boldsymbol{\theta}, X_{t-1}, \mathbf{u}_t)}{\partial X_{t-1}} \frac{\partial X_{t-1}}{\partial \theta_i}$$

$$\frac{\partial qs_t}{\partial \theta_i} = \frac{\partial g(\boldsymbol{\theta}, X_{t-1}, \mathbf{u}_t)}{\partial \theta_i} + \frac{\partial g(\boldsymbol{\theta}, X_{t-1}, \mathbf{u}_t)}{\partial X_{t-1}} \frac{\partial X_{t-1}}{\partial \theta_i}$$

where

- X_t model state at time t;
- $\mathbf{u}_t$ model inputs;
- $f(\ )$ model equation for updating state;
- $g(\ )$ model equation for computing flow.

Derivatives are initialized to zero at time zero.

Acknowledgments. The authors wish to extend thanks to Vijai Gupta for his help and assistance. They also wish to acknowledge the technical support provided by the Center for Computing Information and Technology at the University of Arizona. Support for this research was provided by the Hydrologic Research Laboratory of the National Weather Service (project NA85AA-H-HY088) and the National Science Foundation (grant ECE8610584).

REFERENCES

Bard, Y., Comparison of gradient methods for the solution of nonlinear parameter estimation problems, *SIAM J. Numer. Anal.*, *7*, 157, 1970.

Bard, Y., *Nonlinear Parameter Estimation*, pp. 83–120, Academic, Orlando, Fla., 1974.

Bekey, G. A., and S. F. Masri, Random search techniques for optimization of nonlinear systems with many parameters, *Math. Comput. Simul.*, *2*(3), 210–213, 1983.

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

Boughton, W. C., A new simulation technique for estimation catchment yield, *Rep. 78*, Water Res. Lab., Univ. of New South Wales, Manly Vale, Aust., 1965.

Clarke, R. T., A review of some mathematical models used in hydrology, with observations on their calibration and use, *J. Hydrol.*, *19*, 1–20, 1973.

Crawford, N. H., and R. K. Linsley, Digital simulation in hydrology: Stanford watershed model IV, *Stanford Univ. Dep. Civ. Eng. Tech. Rep. 39*, Stanford, Calif., 1966.

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

Fletcher, R., and M. J. Powell, A rapidly convergent descent method for minimization. *Comput. J.*, *6*, 163–168, 1963.

Gupta, V. K., The identification of conceptual watershed models, Ph.D. dissertation, Case Western Reserve Univ., Cleveland, Ohio, 1984.

Gupta, V. K., and S. Sorooshian, The automatic calibration of conceptual catchment models using derivative-based optimization algorithms, *Water Resour. Res.*, *21*(4), 473–485, 1985.

Himmelblau, D. M., *Applied Nonlinear Programming*, pp. 73–217, McGraw-Hill, New York, 1972.

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

Ibbitt, R. P., Systematic parameter fitting for conceptual models of catchment hydrology, Ph.D. thesis, Univ. of London, 1970.

Ibbitt, R. P., and P. D. Hutchinson, Model parameter consistency and fitting criteria, paper presented at International Federation of Automatic Control (IFAC) 9th World Congress, IFAC, Budapest, Hungary, July 2–6, 1984.

Ibbitt, R. P., and T. O'Donnell, Fitting methods for conceptual catchment models, *J. Hydraul. Eng.*, *97*(HY9), 1331–1342, 1971.

Isabel, D., and J. P. Villeneuve, Importance of the convergence criterion in the automatic calibration of hydrological models, *Water Resour. Res.*, *22*(10), 1367–1370, 1986.

James, L. D., and S. J. Burges, Selection, calibration and testing of hydrologic models, in *Hydrologic Testing of Small Watersheds*, edited by C. T. Haan, H. P. Johnson, and D. L. Brakensiek, pp. 437–471, American Society of Agricultural Engineers, St. Joseph, Mich., 1982.

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

Karnopp, D. C., Random search technique for optimization problems, *Automatica*, *1*, 111–121, 1963.

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

Lemmer, H. R., and A. R. Rao, Critical duration analysis and parameter estimation in ILLUDAS, *Tech. Rep. 153*, Water Resour. Res. Cent., Purdue Univ., West Lafayette, Ind., 1983.

Leon, A., A comparison among eight known optimizing procedures, in *Recent Advances in Optimization Techniques*, edited by A. Lavi and T. P. Vogl, pp. 23–42, John Wiley, New York, 1966.

Levenberg, K., A method for the solution of certain nonlinear problems in least squares, *Q. Appl. Math.*, *2*, 164–168, 1944.

Marquardt, D. W., An algorithm for the least squares estimation of nonlinear parameters, *SIAM J.*, *11*, 431–441, 1963.

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

Moore, R. J., and R. T. Clarke, A distribution function approach to rainfall runoff modeling, *Water Resour. Res.*, *17*(5), 1367–1382, 1981.

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 functional 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 Hydrol. 31*, U.S. Dep. of Commerce, Silver Spring, Md., 1976.

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

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

Restrepo-Posada, P. J., and R. L. Bras, Automatic parameter estimation of a large conceptual rainfall-runoff model: A maximum likelihood approach, *Rep. 267*, Dep. of Civ. Eng., Mass. Inst. Technol., Cambridge, Mass., 1982.

Rosenbrock, H. H., An automatic method of finding the greatest of least value of a function, *Comput. J.*, *3*, 175–184, 1960.

Sefe, F. T., and W. C. Boughton, Variation of model parameter values and sensitivity with type of objective function, *J. Hydrol. (NZ)*, *21*(1), 117–132, 1982.

Sorooshian, S., Surface water hydrology: On line estimation, *Rev. Geophys.*, *21*(3), 706–721, 1983.

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 model parameters: The question of parameter observability and uniqueness, *Water Resour. Res.*, *21*(4), 260–268, 1983.

Sorooshian, S., V. K. Gupta, and J. L. Fulton, Modifications to the automatic calibration procedures of the NWSRFS soil moisture accounting model: The CIT version, technical report to the Hydrol. Res. Lab., U.S. Nat. Weather Serv. Nat. Oceanic and Atmos. Admin., U.S. Dep. of Commerce, Boulder, Colo., 1981.

Sorooshian, S., V. K. Gupta, and J. L. Fulton, Evaluation of maximum likelihood estimation techniques for conceptual rainfall-runoff models: Influence of calibration data variability and length on model credibility, *Water Resour. Res.*, *10*(1), 251–259, 1983.

Wheater, H. S., K. H. Bishop, and M. B. Beck, The identification of conceptual hydrological models for surface water acidification, *Hydrol. Processes*, *1*, 89–109, 1986.

Wortman, J. D., NLPROG, *Ballistic Res. Lab. Mem. Rep. 1958*, Aberdeen Proving Grounds, Md., 1969.

L. E. Brazil, Hydrologic Research Laboratory, U.S. National Weather Service, Silver Spring, MD 20910.

J. D. Hendrickson and S. Sorooshian, Department of Hydrology and Water Resources, University of Arizona, Tucson, AZ 85721.

(Received July 28, 1987;
revised January 8, 1988;
accepted January 8, 1988.)