WATER RESOURCES RESEARCH, VOL. 34, NO. 4, PAGES 751–763, APRIL 1998

Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information

Hoshin Vijai Gupta and Soroosh Sorooshian
Department of Hydrology and Water Resources, University of Arizona, Tucson

Patrice Ogou Yapo
Department of Systems and Industrial Engineering, University of Arizona, Tucson

Abstract. Several contributions to the hydrological literature have brought into question the continued usefulness of the classical paradigm for hydrologic model calibration. With the growing popularity of sophisticated "physically based" watershed models (e.g., land-surface hydrology and hydrochemical models) the complexity of the calibration problem has been multiplied many fold. We disagree with the seemingly widespread conviction that the model calibration problem will simply disappear with the availability of more and better field measurements. This paper suggests that the emergence of a new and more powerful model calibration paradigm must include recognition of the inherent multiobjective nature of the problem and must explicitly recognize the role of model error. The results of our preliminary studies are presented. Through an illustrative case study we show that the multiobjective approach is not only practical and relatively simple to implement but can also provide useful information about the limitations of a model.

1. Introduction and Scope

Many hydrologic models must be "calibrated" to be useful for the solution of practical problems. By calibration we mean that the hydrologist must estimate values for the model "parameters" that enable the model to closely match the behavior of the real system it represents. In some cases the appropriate values for a model parameter can be determined through direct measurements conducted on the real system. However, in many situations the model parameters are conceptual representations of abstract watershed characteristics and must be determined through a trial-and-error process that adjusts the parameter values to match the model response to historical input-output data.

During the past 2 decades a great deal of research has been devoted to the development of automated (computer-based) methods for the estimation of model parameters by fitting them to historical data. That research has focused primarily on four issues: (1) the development of specialized techniques for handling the kinds of errors present in the measured data, (2) the search for an optimization strategy that can reliably solve the parameter estimation problem, (3) the determination of the appropriate quantity and most informative kind of data, and (4) the efficient representation of the uncertainty of the calibrated model (structure and parameters) and translation of that uncertainty into uncertainty in the model response. Research into techniques for accounting for data error has led to the development of maximum likelihood functions for measuring the "closeness" of the model and the data [e.g., *Sorooshian and Dracup*, 1980; *Sorooshian*, 1981; *Sorooshian et al.*, 1982, 1983; *James and Burges*, 1982; *Sefe and Boughton*, 1982; *Lemmer and Rao*, 1983; *Kuczera*, 1983a, b; *Ibbitt and Hutchinson*, 1984]. Research into optimization methods has led to the use of population-evolution-based search strategies [e.g., *Brazil and Krajewski*, 1987; *Brazil*, 1988; *Wang*, 1991; *Duan et al.*, 1992, 1993; *Sorooshian et al.*, 1993]. In this regard the shuffled complex evolution (SCE-UA) global optimization algorithm has proved to be consistent, effective, and efficient in locating the globally optimal model parameters of a hydrologic model [e.g., *Duan et al.*, 1992, 1993; *Sorooshian et al.*, 1993; *Luce and Cundy*, 1994; *Gan and Biftu*, 1996; *Tanakamaru*, 1995; *Tanakamaru and Burges*, 1997; *Kuczera*, 1997]. Research into data requirements has led to the understanding that the informativeness of the data is far more important than the amount used for model calibration [e.g., *Kuczera*, 1982; *Sorooshian et al.*, 1983; *Gupta and Sorooshian*, 1985; *Yapo et al.*, 1996]. Finally, research into representation of model uncertainty has led to practical procedures for rigorous statistical analysis of model parameter uncertainty [e.g., *Spear and Hornberger*, 1980; *Jones*, 1983; *Kuczera*, 1988]. With these developments the capabilities and limitations of the classical strategy, rooted in statistical approaches, for calibrating single-output hydrologic models with up to 10–15 parameters may be considered to be reasonably well understood.

Notwithstanding the progress mentioned above, we share the concern surfacing in the hydrological literature that the "classical" approach to model calibration has some serious limitations that necessitate the emergence of a new and more powerful paradigm [see, e.g., *Klepper et al.*, 1991; *van Straten and Keesman*, 1991; *Beven and Binley*, 1992; *Yapo et al.*, 1996]. One of these limitations is the fact that it is typically difficult, if not impossible, to find a unique "best" parameter set that obviates consideration of other feasible parameter sets. It is our opinion that the various classical attempts (including our earlier work) to locate unique model parameters are based on a philosophy that involves some longstanding and questionable assumptions that arise from the adoption of classical statistical techniques for the fitting of empirical models to data. A second and, from the point of view of this paper, perhaps more serious

Paper number 97WR03495.
0043-1397/98/97WR-03495$09.00

limitation is the woeful inadequacy of current strategies in the face of the emerging generation of multi-input-output hydrologic models [e.g., *Beven and Kirkby*, 1979; *Kuczera*, 1982, 1983a, b; *De Grosbois et al.*, 1988; *Woolhiser et al.*, 1990; *Yan and Haan*, 1991a, b; *Gupta and Sorooshian*, 1994a, b; *Yapo et al.*, 1996].

Some interesting (and somewhat similar) methods for addressing these limitations, particularly in the context of prediction uncertainty, have begun to appear in the literature. These include the generalized likelihood uncertainty estimation (GLUE) framework for representing model parameter and prediction uncertainty within the context of Monte Carlo analysis coupled with Bayesian estimation and propagation of uncertainty [see, e.g., *Beven and Binley*, 1992; *Freer et al.*, 1996], the Monte Carlo set membership procedure (MCSM) [see, e.g., *Keesman*, 1990; *van Straten and Keesman*, 1991], and the prediction uncertainty method (PU) [see, e.g., *Klepper et al.*, 1991]. All of these approaches are directly or indirectly related to the generalized sensitivity analysis (GSA) method of G. M. Hornberger and colleagues at the University of Virginia [see *Spear and Hornberger*, 1980]. Although these approaches still have weaknesses that need to be addressed (e.g., the GLUE technique requires subjective decisions in the selection of prior parameter distributions, the "likelihood" criterion, and the cutoff thresholds), we share the opinion voiced by *Clarke* [1994, p. 345] that such approaches represent "... a bold attempt to introduce some much-needed new thinking into a field that is in grave danger of becoming intellectually sterile."

The ideas presented in this paper have similarities to and also some notable differences from the GSA-based GLUE, MCSM, and PU statistical uncertainty analysis methods. The major similarity is the use of an initial Monte Carlo sampling of the feasible parameter space to approximate prior parameter and prediction uncertainty. The major difference is our focus on the inherent multiobjective nature of the model calibration problem. To be clear, we state at the outset that we do not consider either the statistical uncertainty analysis methods discussed in the literature or the multiobjective approach presented here to be complete in themselves. We consider these approaches to be complementary. Indeed, to be satisfactory, the emerging paradigm for model calibration will need to recognize and incorporate treatment of both the statistical representation of uncertainty and the multiplicity and noncommensurate nature of measures for extracting useful information from the data. In this paper we will focus largely on the latter issue. The next two sections of this paper present our point of view and the basis for our reasoning and (where relevant) compare and contrast the two approaches. The final sections illustrate the validity and efficacy of our arguments by presenting and discussing the practical results of our initial studies to date.

2. Toward a Multiobjective View

Consider a system $\mathcal{S}$ for which a hydrologic model $\mathcal{H}$ is to be calibrated. We assume that the mathematical structure of the model is essentially predetermined and fixed and that physically realistic upper and lower bounds on each of the model parameters can be specified a priori (thereby defining the feasible parameter space, i.e., the initial uncertainty in the parameters). To begin, let us consider the simple formulation in which the model is required to simulate only one aspect of the system, say the time evolution of streamflow. Let $D = \{d_1, \cdots, d_n\}$ represent the vector of streamflow measurement data available at time steps $1, \cdots, n$ and let $O(\theta) = \{o(\theta)_1, \cdots, o(\theta)_n\}$ represent the corresponding vector of estimated model output fluxes generated using the parameter values θ. The difference between the model-simulated fluxes and the measurement data can be represented by the residual vector $E(\theta) = G[O(\theta)] - G(D) = \{e(\theta)_1, \cdots, e(\theta)_n\}$, where the function $G(\ )$ allows for linear or nonlinear transformations (such as log transformations, power transformations, weightings, etc.) of the streamflow measurement data and the corresponding estimated fluxes. In the classical approach to model calibration the goal is to find the best values for the parameters θ such that E is in some sense made as close to "zero" as possible. The standard approach is to define some measure L of the "length" of vector E and then to attempt to find the values of the model parameters θ that minimize L. However, there is no unambiguously "correct" way in which to define this measure of length (commonly called the objective function). By far the most popular measure is the mean squared-error estimator (MSE), appropriate when the measurement errors are known to be uncorrelated and homoscedastic (having constant variance) or when the properties of the measurement errors are unknown. When the measurement errors are believed to be heteroscedastic (nonconstant variance), the heteroscedastic maximum likelihood estimator (HMLE) can be used [*Sorooshian and Dracup*, 1980]. Table 1 lists some of the other objective functions that are commonly referenced in the literature; this particular list consists of the various measures used by the Hydrologic Research Laboratory of the National Weather Service for manual and stage-wise semiautomated calibration of the Sacramento soil moisture accounting (SAC-SMA) model of the National Weather Service River Forecast System (NWSRFS) [*Brazil*, 1988].

In over 2 decades of investigation it has not proved possible to clearly demonstrate that a particular objective function is better suited for calibration of a model than some other [e.g., *Chapman*, 1970; *Diskin and Simon*, 1977; *Sorooshian et al.*, 1983; *Yan and Haan*, 1991a, b; *Yapo et al.*, 1996a]. Further, even when a particular objective function is chosen, it has proved impossible to find any best parameter set for a given watershed. To illustrate this problem, a 4-month portion of the measured and simulated hydrographs for the Leaf River Basin in Mississippi is displayed in Figure 1. The SAC-SMA model was calibrated to this watershed using the SCE-UA algorithm and 8 water years of calibration data. Separate calibration runs were made using the daily root mean square estimator (DRMS = $MSE^{1/2}$) and the HMLE. Notice that the best DRMS parameter set matches the early portion of the hydrograph extremely well, and the best HMLE parameter set systematically underestimates it, while for the latter portion of the hydrograph the HMLE parameter set performs well and the DRMS parameter set overestimates it. The results of calibrating the model separately to each of the 40 water years of available data using the SCE-UA algorithm and the DRMS objective function are shown in Figures 2a and 2b. The best DRMS fit obtained for each year is shown in Figure 2a, and the variation in the 40 corresponding globally optimal parameter sets is shown in Figure 2b. Clearly, to choose a single best parameter set would be difficult; a parameter set that gives excellent forecasts for 1 year might perform very poorly on another.

Now, we certainly expect that a computer-based model of a watershed, being an imperfect representation of a physical

Table 1. Objective Functions Used by the National Weather Service for Calibration of the SAC-SMA Model

Name	Description	Formula*
DRMS	Daily Root Mean Squared Error	$\sqrt{\frac{1}{n}\sum_{t=1}^{n}(d_t - o_t(\theta))^2}$
TMVOL	Total Mean Monthly Volume Squared Error	$\sum_{i=1}^{n_{\text{month}}}\left(\frac{1}{n_{\text{day}}(i)}\sum_{t=1}^{n_{\text{day}}(i)}(d_t - o_t(\theta))\right)^2$
ABSERR	Mean Absolute Error	$\frac{1}{n}\sum_{t=1}^{n}\lvert d_t - o_t(\theta)\rvert$
ABSMAX	Maximum Absolute Error	$\max_{1\le t\le n}\lvert d_t - o_t(\theta)\rvert$
NS	Nash-Sutcliffe Measure	$1 - \frac{\frac{1}{n}\sum_{t=1}^{n}[d_t - o_t(\theta)]^2}{\frac{1}{n}\sum_{t=1}^{n}(d_t - \bar{d})^2}$
BIAS	Bias (mean daily error)	$\frac{1}{n}\sum_{t=1}^{n}(d_t - o_t(\theta))$
PDIFF	Peak Difference	$\max_{1\le t\le n}\{d_t\} - \max_{1\le t\le n}\{o_t(\theta)\}$
RCOEF	First Lag Autocorrelation	$\frac{\frac{1}{n}\sum_{t=1}^{n}(d_t - o_t(\theta))(d_{t+1} - o_{t+1}(\theta))}{\sigma_d\sigma_{o(\theta)}}$
NSC	Number of Sign Changes	(Count the number of times the sequence of residuals changes sign)

*Minimize with respect to θ.

system, will be unable to provide a perfect match to the data. This inability may be due to the presence of errors in both the data and the model (by model error we mean error arising because of simplified or otherwise imperfect representations of the structure of the system). The common approach to dealing with this is to make some assumptions regarding the statistical distribution of the output data errors (input data errors are typically ignored; for exceptions, see *Troutman* [1985a, b] and *Kitanidis and Bras* [1980a] among others) and either consider the "model error" to be "small" or to be somehow "absorbed" into the output error residual (for exceptions, see *Kitanidis and Bras* [1980b, c]). The residual is then expected to behave statistically in the same manner as the output measurement error. On the basis of our study of the problems of calibrating hydrologic models, we suggest that

1. The magnitude of the model error for some portions of the model response may, in general, be equivalent to or even substantially larger than the output measurement error.

2. The model errors do not necessarily have any inherent probabilistic properties that can be exploited in the construction of an objective function (e.g., if a nonlinear relationship is approximated with a linear one, the approximation errors that arise are not random in the probabilistic sense). While we can assume a probability structure for model error (as do *Kitanidis and Bras* [1980c]), this will be purely for the sake of mathematical convenience.

On the basis of these propositions we are faced with the possibility that there may not exist an objective "statistically correct" choice for the objective function and therefore no statistically correct "optimal" choice for the model parameters. In fact, we are left with the intuitively reasonable concept that the hydrologist may choose among several possible parameter sets (and indeed model structures), each of which closely matches the hydrograph in different ways; for example, one parameter set (model) may better match the peak flows, while another may give more emphasis to matching the recessions (this is similar to what happens when different experts are called upon to "manually" calibrate the same model). Note that this rationale for multiobjective equivalence of several parameter sets (models) is different from the rationale for what *Beven and Binley* [1992] call "equifinality" of parameter sets (models), what *van Straten and Keesman* [1991] call "equally probable (or characteristic)" parameter sets, or what *Klepper et al.* [1991] call "acceptable" parameter sets. Those authors base their arguments on the probabilistic representation of parameter (model) uncertainty. Our arguments, however, are based on the multiple ways in which the best fit of a model to the data can be defined. The multiobjective equivalence of parameter sets is more commonly referred to as "pareto optimal" or "nondominant" in the literature; in this paper we shall adopt the terminology pareto optimal. The equifinal (GLUE), equally probable (MCSM), acceptable (PU), and

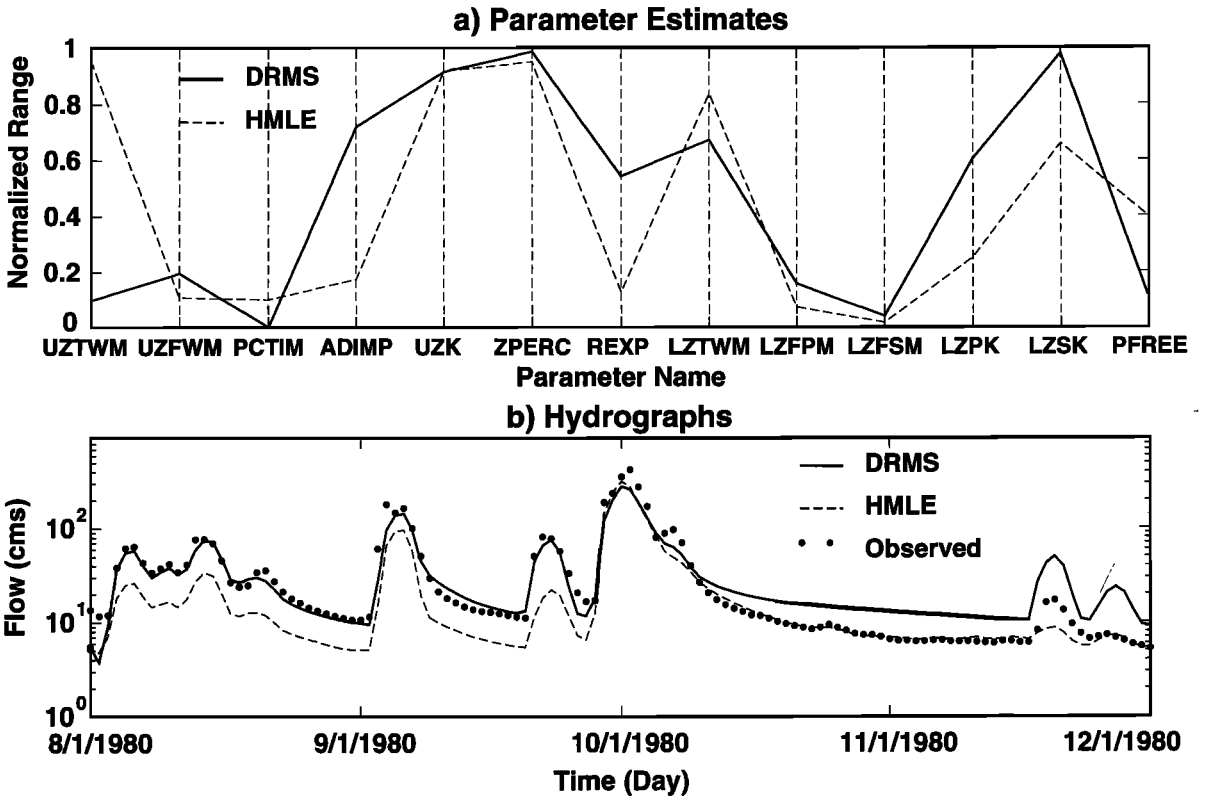

Figure 1. Example showing how different "optimal" parameter sets are better at matching different portions of the hydrograph.

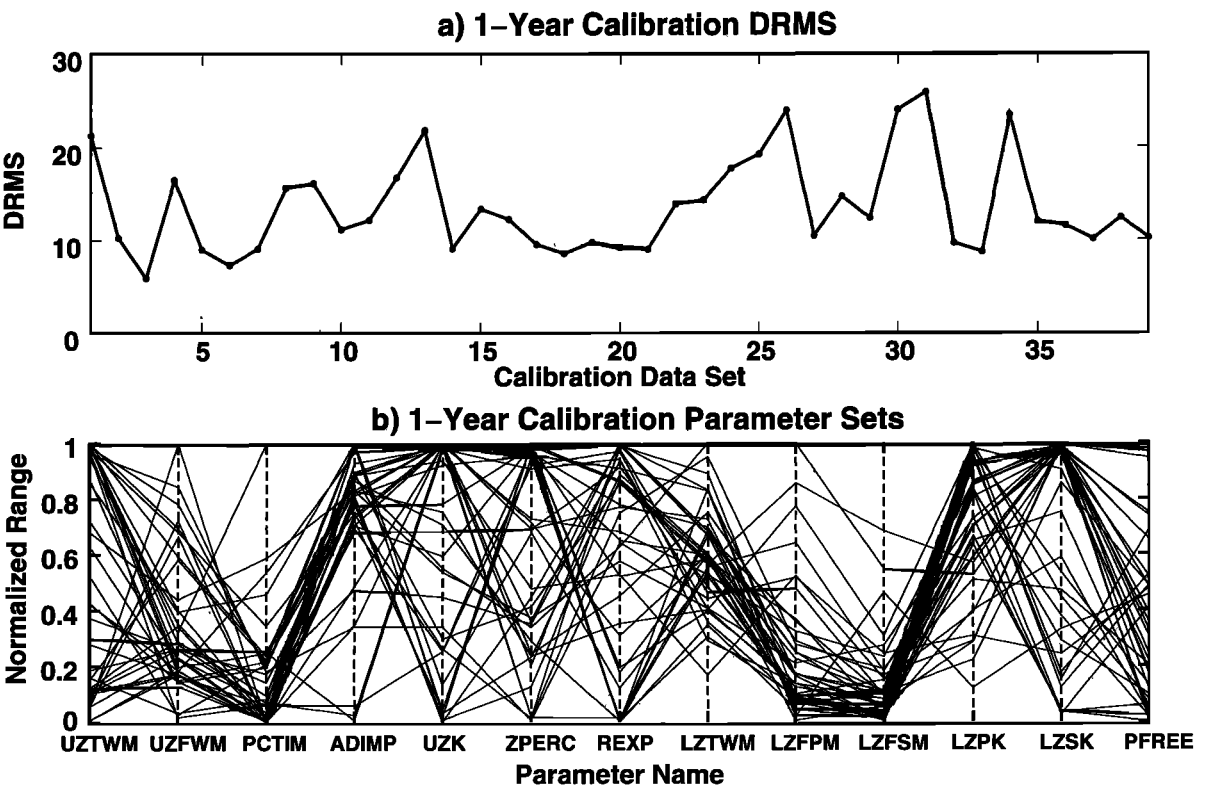

Figure 2. Example showing the variation in results obtained by calibration of the Sacramento soil moisture accounting (SAC-SMA) model to 40 different water years for the Leaf River watershed.

pareto optimal parameter sets (models) may overlap but will not, in general, be equivalent.

We are therefore faced with the question, "Do any objective guidelines exist that take model error into consideration and can aid in proper calibration of the model?" To explore this issue, we examine the assumptions that go into the classical formulation. Each component of E can be written as $e_j(\theta) = em_j(\theta) + ed_j$, where $em_j(\theta)$ is the model error and ed_j is the additive data measurement error. Unless $ed_j = 0$, it will be impossible to drive $E(\theta)$ to zero, even if the model were perfect. Because of this, and because $E(\theta)$ is a vector of n components, the goal of model calibration might be more correctly stated as that of finding θ such that $e_j(\theta) = ed_j$ for all $j = 1, \cdots, n$. This condition is met only if $em_j(\theta) = 0$, and the model output therefore perfectly matches the true system output. Only if the model were a perfect representation of the system might such a goal be achievable. We must make some assumptions in order to proceed. For example,

1. We can focus on the measurement errors (the classical regression approach), make some assumptions about the statistical distribution of those errors, and apply maximum likelihood or Bayesian theory to arrive at an objective function that computes some appropriately weighted sum of the model residuals [e.g., *Sorooshian and Dracup*, 1980; *Kuczera*, 1983a, b; *Yan and Haan*, 1991a, b]. By optimizing this function with respect to θ we will obtain a solution such that the statistical distribution of the model output residuals approximates the assumed error distribution as closely as is permitted by the structure of the model. Examples of this approach are the MSE (assumes homoscedastic and independent errors) and the HMLE (assumes heteroscedastic and independent errors).

2. We can ignore the statistical properties of the measurement errors and decide instead to find some approach that will tend to drive each and every one of the elements of $E(\theta)$ as close to zero as possible. However, this leaves us with the multiobjective optimization problem:

$$\min \text{ (with respect to } \theta) \; |E(\theta)| = \{|e_1(\theta)|, \cdots, |e_n(\theta)|\} \quad (1)$$

(where $|e_j(\theta)|$ is the absolute value of e_j) which may, in general, have a unique solution only if we assume that the model and data errors are nonexistent.

The properties of the first approach have been extensively explored in the literature, and so we will focus primarily on the properties of the second. The multiobjective optimization problem stated in (1) is considerably more difficult to solve than a single-objective one. Further, the solution to this problem will, by its very nature, not be unique (we discuss this in more detail below).

One classical way to approach the problem of multiple measures is to make some assumption that permits "... combining them into a single index..." [*Beven*, 1993] that works well in practice (see e.g., the suggestions given by *van Straten* [1983], *Klepper et al.* [1991], and *Beven and Binley* [1992]). For example, one typical approach is to find θ in order to minimize some measure of the dispersion of the model residuals around zero in whatever manner is permitted by the structure of the model. For instance, by assigning equal weights to the absolute magnitude of each residual and computing the average we obtain the mean absolute-error estimator (MAE), or by weighting the magnitude of each squared residual equally and computing the average we obtain the mean squared-error estimator (MSE). In the absence of a compelling and reasonable basis for the assignment of the weights we are faced with the unavoidable fact that the model calibration problem is inherently multiobjective and that any attempt to convert it into a single-objective problem must necessarily involve some degree of subjectivity.

If we proceed with the multiobjective optimization problem as stated, the first issue that must be addressed is that the problem stated in (1) is not properly defined in multiobjective terms. This is because the magnitudes of the individual model residuals are directly related through the structural dynamics of the model. Therefore it is necessary to find some transformation F of $|E|$ that provides us with only a set of (relatively) unrelated measures that preserve the information content of the data. Actually, more general forms of F can also be derived by directly applying various transformations (such as max{ }, min{ }, median{ }, Var{ }, etc.) directly to $O(\theta)$ and D). In practice the dimension m of F will generally be significantly less than the dimension n of E. We discuss the problem of finding the set of relatively unrelated measures contained in F later. Having found this set of measures, we can correctly state the calibration problem as

$$\min \text{ (with respect to } \theta) \; F(\theta) = \{f_1(\theta), \cdots, f_m(\theta)\} \quad (2)$$

Second, with this formulation the solution will no longer, in general, be a single unique parameter set (model) but will consist of a Pareto set $P(\Theta)$ of solutions in the feasible parameter space Θ corresponding to various trade-offs among the objectives. The definition of this Pareto set is such that any member θ_i of the set has the properties: (1) For all nonmembers θ_j there exists at least one member θ_i such that $F(\theta_i)$ is strictly less than $F(\theta_j)$, and (2) it is not possible to find θ_j within the Pareto set such that $F(\theta_j)$ is strictly less than $F(\theta_i)$ (i.e., by "strictly less than" we mean $f_k(\theta_j) < f_k(\theta_i)$ for all $k = 1, \cdots, m$).

According to the first of these statements the feasible parameter space can be partitioned into "good" solutions (Pareto solutions) and "bad" solutions. According to the second, in the absence of additional information it is not possible to distinguish any of the good (Pareto) solutions as being objectively better than any of the other good solutions (i.e., there is no uniquely best solution). Every member θ_i of the Pareto set will match some characteristics of the hydrograph better than every other member of the Pareto set, but the trade-off will be that some other characteristics of the hydrograph will not be as well matched.

When compared to the classical single-objective formulation, several interesting advantages of the multi-objective approach become immediately apparent (see the hypothetical example illustrated in Figure 3). We begin with an initial model uncertainty $U^o(\Theta)$ represented by the size of the feasible parameter space Θ to be searched (lighter shaded portion of Figure 3a) and by the corresponding size of the space $U^o(Q)$ of possible hydrograph solutions Q (shown here in terms of $\log(Q)$; see lighter shaded portion of Figure 3b). As pointed out by *Beven and Binley* [1992], these spaces can be properly constrained using prior knowledge about what parameter combinations and watershed responses are reasonable. The goal of the single-objective approach is to find a single unique solution Θ^* thereby reducing the final size of the parameter uncertainty $U^f(\Theta)$ and, hence, $U^f(Q)$ to zero. However, given the existence of model and data error, $U^f(Q)$ must remain nonzero, and this solution is therefore unreasonable. In contrast, the multiobjective approach finds that because of the existence of

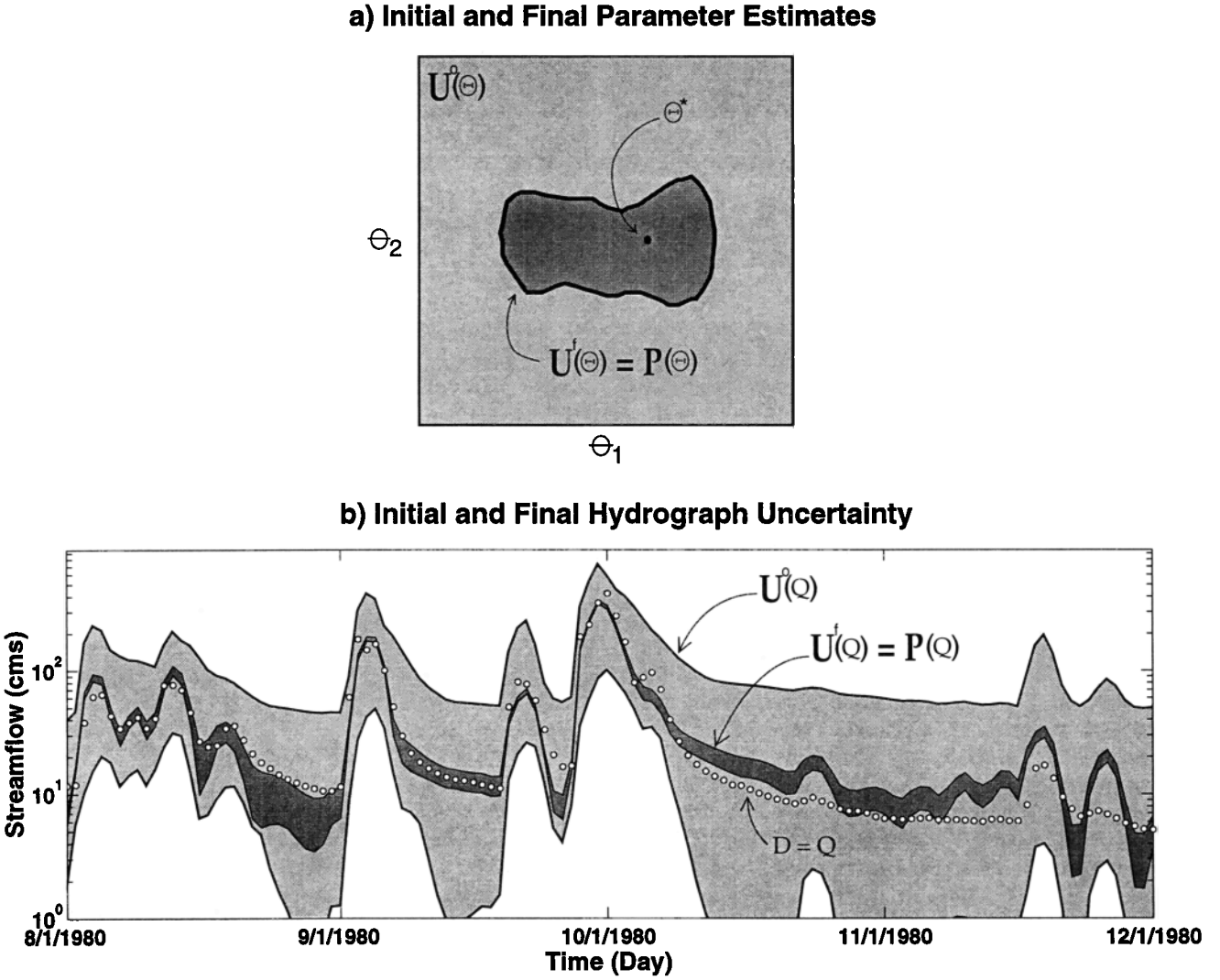

Figure 3. Illustration of parameter and hydrograph estimates obtained using the multiobjective calibration approach.

model error, the minimal value for $U^f(\Theta)$ is the Pareto parameter space $P(\Theta)$ (darker shaded portion of Figure 3a), and the minimal value for $U^f(Q)$ is the associated Pareto hydrograph space $P(Q)$ (darker shaded portion of Figure 3b). Given the existence of data errors in addition to the model errors, the actual parameter and hydrograph uncertainties must be larger than these minimum bounds. Further, only under the conditions that the data and model errors are nonexistent can the uncertainty in the parameters and hydrograph be reduced to zero.

Note that any fuzziness in the specification of the Pareto parameter spaces ($P(\Theta)$ and $P(Q)$) in the multiobjective approach arises from only two factors: (a) subjectivity in the selection of the measures in F and (b) the statistical uncertainty in the computation of each measure $f_j(\theta)$ arising from sampling considerations. Unlike the GLUE, MCSM, and PU approaches, an arbitrary threshold of acceptability in the value of the measure is not required. Similar to these approaches, however, the size and characteristics of $P(\Theta)$ and $P(Q)$ provide very useful information about the limitations of the model. In the hypothetical illustration we see that the Pareto hydrograph space does not bracket significant portions of the observed data, indicating either sizeable measurement error or some potential deficiencies in the model structure. The systematic nature of the discrepancy might suggest that the culprit is deficiency in the model structure. A large Pareto range in some of the parameters (e.g., θ_1 in Figure 3a) might suggest that the deficiency lies primarily in the structural representations associated with those parameters. The classical single-objective approach provides little or no such guidance.

3. Dealing With Multiple Output Fluxes

The formulation presented above argues that even in the case of a model with just one output flux to be simulated, the calibration problem is inherently multiobjective. However, many of today's hydrologic and environmental models are designed to simulate not only streamflow but also various components of stream chemistry, sediment load, latent and sensible heat flux, soil moisture, and so on. Measurement data on several of these fluxes (say D^1–D^k) may be available that can be used to help calibrate the parameters of the model. The goal of model calibration now becomes that of finding values for the model parameters θ such that the model-simulated

fluxes match all k of these (noncommensurable) measurement data fluxes as closely as possible. To apply the multiobjective approach, we construct the extended data vector $D = \{D^1 \cdots, D^k\}$, find the set of relatively unrelated and noncommensurable measures that preserve the information content of this extended data, and proceed in the manner described earlier.

4. Effective and Efficient Multiobjective Optimization

To solve the model calibration problem stated in (2), we must do two things. First, the functions F^j must be specified; that is, we must find a set of relatively unrelated measures of the differences between the model simulations and the observations that preserves the information contained in the data. This issue will be addressed in the next section. Second, we must find a method to effectively and efficiently (inexpensively) estimate the location of the Pareto solution set $P(\Theta)$. We will tackle the second problem in this section.

The field of optimization theory has studied the multiobjective optimization problem quite extensively [*Goicoechea et al.*, 1982]. Because the Pareto set seldom consists of a finite number of solutions, most multiobjective techniques attempt to identify a countable number of distinct solutions distributed within the Pareto region. The classical methods for obtaining such solutions can be categorized as a posteriori methods, a priori methods, and interactive methods. Examples of a posteriori methods (also called generating techniques) include the weighing method [*Zadeh*, 1963], the ε-constraint method [*Marglin*, 1967], and the goal attainment method [*Gembicki*, 1974]. Examples of a priori methods include the goal programming and the compromise programming methods [*Zeleny*, 1973]. Examples of interactive techniques include the surrogate worth trade-off method (SWT) [*Haimes and Hall*, 1974] and the trade-off development method (TRADE) [*Goicoechea et al.*, 1976]. Presentations and discussions of these methods can be found in textbooks [*Goicoechea et al.*, 1982; *Szidarovsky et al.*, 1986] and in review papers [*Hipel*, 1992; *Szidarovsky and Szenteleki*, 1987; *Yapo et al.*, 1992].

The overriding characteristic of classical multiobjective optimization methods is the sequential generation of Pareto solutions. As an illustration, we shall consider the weighing method and assume there are 5 objectives $f_1(\theta), \ldots, f_5(\theta)$ to be minimized. In this approach, each objective is allocated a weight, and the multiobjective optimization problem is converted into a single-objective optimization problem as

$$\min \text{ (with respect to } \theta) \; F(\theta) = w_1 f_1(\theta) + \cdots + w_5 f_5(\theta) \qquad (3)$$

where $w_1 + \cdots + w_5 = 1$. This problem can be easily solved using standard single-objective global optimization algorithms such as the SCE-UA. By randomly (or in some other fashion) selecting different values for the weights allocated to the five objectives we can generate as many discrete Pareto solutions as necessary to obtain an acceptable approximation of the continuous Pareto space. Alternatively, we can interactively guide the selection of weights until a "satisfactory" solution point is discovered. For example, *Yan and Haan* [1991a, b] used three objectives to generate a limited number of Pareto solutions while calibrating the U.S. Geological Survey (USGS) precipitation-runoff modeling system (PRMS) [*Leavesley et al.*, 1983]. Although they did not attempt to approximate the entire Pareto space, their results indicated generally improved model performance in comparison to single-objective calibration.

Although the classical approach is seemingly simple to implement, it carries a heavy price: for each discrete Pareto solution we must solve a complete single-objective optimization problem. If, for example, we need 100 discrete solutions to approximate the continuous solution space, we would have to reinitialize and rerun the optimization procedure 100 separate times. If, as in *Sorooshian et al.* [1993], each single-objective optimization run requires as many as 5,000–10,000 function evaluations, we are faced with a potential cost exceeding of the order of half-a-million to a million function evaluations, not a very heartening prospect.

Fortunately, an effective and efficient nonclassical method for solving the multiobjective problem in its original form has been developed. The method, entitled multiobjective complex evolution (MOCOM-UA), is a general purpose global multiobjective optimization algorithm that provides an effective and efficient estimate of the Pareto solution space with only a single optimization run and does not require subjective weighting of the objectives. MOCOM-UA is based on an extension of the SCE-UA population evolution method reported by *Duan et al.* [1993]. A detailed description and explanation of the method are given by *Yapo et al.* [1997a, b] and so will not be repeated here.

In brief the MOCOM-UA method involves the initial selection of a "population" of p points distributed randomly throughout the s-dimensional feasible parameter space $U^o(\theta)$. In the absence of prior information about the location of the Pareto optimum a uniform sampling distribution is used. For each point the multiobjective vector $F(\theta)$ is computed, and the population is ranked and sorted using a Pareto-ranking procedure suggested by *Goldberg* [1989]. Simplexes of $s + 1$ points are then selected from the population according to a robust rank-based selection method [*Whitley*, 1989]. A multiobjective extension of the downhill simplex method is used to evolve each simplex in a multiobjective improvement direction. Iterative application of the ranking and evolution procedures causes the entire population to converge toward the Pareto optimum. The procedure terminates automatically when all points in the population become nondominated. Experiments conducted using standard synthetic multiobjective test problems have shown that the final population provides a fairly uniform approximation to the Pareto solution space $P(\Theta)$ [*Yapo et al.*, 1997a, b].

5. Selection of the Objective Functions

To implement the multiobjective procedure outlined above, it is necessary to specify a set of relatively unrelated functions F ("unrelated" in the sense that they measure different important aspects of the differences between the observed data D and the model simulations $O(\theta)$) that can be used to extract the useful information contained in the data and transform it into estimates for the parameters (models). In the systems theoretic sense, useful "information" can be viewed as that which enables one to test a hypothesis. There are two important issues to be addressed here.

First, it should be noted that the hypothesis to be tested is always a subjective consequence of the interaction between the context of the problem and what the modeler considers to be important. In the context of watershed modeling the modeler

must decide what are the important characteristics of watershed behavior to be reproduced by the calibrated model and what constitutes an effective measure of that behavior. For example, during manual calibration of the SAC-SMA model the HRL-NWS hydrologist may examine the values of several, if not all, of the measures listed in Table 1. Note that these include measures of the daily and monthly residual variance (DRMS and total monthly volume (TMVOL)), the mean daily error (bias), the error in matching of peak flow (peak difference (PDIFF)), and the measures of the "nonwhiteness" or systematic nature of the residuals (R coefficient (RCOEF) and number of sign changes (NSC)). The final outcome of manual calibration is a result of the attempt to strike a balance in optimizing (minimizing or maximizing, as appropriate) all of these measures. The hypothesis (rarely explicitly stated) is, of course, that it is possible to find values for the model parameters that can achieve acceptable optimal values for each of the measures under consideration. In the standard automatic calibration strategy, only a single measure (typically some measure of the dispersion of the residuals around zero, e.g., MSE) or some combination of measures into a single measure, is used.

Second and equally important is the fact that a hypothesis typically involves several underlying assumptions that must be tested as part of the hypothesis testing procedure. In the context of watershed modeling this might involve a rigorous analysis of the residuals to verify that they belong to some a priori assumed distribution, are unbiased, are homogenous, and have no systematic components, etc. [e.g., see *Yapo et al.*, 1996]. For example, in the context of applying the maximum likelihood theory to the calibration of a watershed model the hypothesis might be that it is possible to find a set of parameters such that the variance of the residuals can be minimized to some acceptable value while assuming that the underlying errors belong to some distribution, typically having zero mean and insignificant autocorrelation. In the process of residual analysis it will be necessary to construct measures (tests) to detect any deviation of the model behavior from the a priori assumptions (i.e., regarding the distribution, bias, and correlation of the residuals). From this perspective it should be clear that each assumption underlying a hypothesis can actually be viewed as a measure of model performance that could and (in our opinion) should (if possible) be explicitly included in the set of measures F employed in the statement of the hypothesis. Of course, it may not be easy or possible to formalize all assumptions as quantitative measures, and such assumptions may still need to be treated in the postcalibration evaluation. We illustrate these issues in the context of the simple case study presented below.

6. Case Study

6.1. The Model and Data

The following case study illustrates that the multiobjective calibration procedure is relatively simple to implement and provides useful information which improves our understanding of the problem. We apply the methods described above to the calibration of the SAC-SMA model using historical data from the Leaf River watershed (1950 km^2) located north of Collins, Mississippi. A reliable 40-water-year data set that represents a variety of hydrological conditions and phenomena is available for this watershed. The SAC-SMA model has 16 parameters to be determined by the user (see Table 2). It is typical for three of these parameters to be fixed at prespecified values, while the remaining 13 must be determined by the process of calibration. The upper and lower bounds that define the initial uncertainty in the parameter estimates for this watershed are listed in Table 2. Following the recommendations of *Yapo et al.* [1996], 8 consecutive water years of data spanning the wettest years on record were selected for model calibration. Because the SAC-SMA model and the Leaf River data have been discussed extensively in previous work [e.g., see *Burnash et al.*, 1973; *Peck*, 1976; *Kitanidis and Bras*, 1980a, b, c; *Brazil and Hudlow*, 1981; *Brazil*, 1988; *Sorooshian and Gupta*, 1983; *Sorooshian et al.*, 1982, 1983, 1993; *Duan et al.*, 1993, 1994; and *Yapo et al.*, 1996], we will not describe the details of these here. Note that the model has only a single output flux to be matched; therefore the case study provides a relatively simple illustrative test of the multiobjective calibration method.

6.2. Statement of the Hypothesis/Selection of the Objective Functions

Because the purpose of this case study is illustrative, we shall employ a rather simple calibration hypothesis, that it is possible to find values for the model parameters that can provide acceptable optimal (minimal) values for the residual standard deviation (measured by DRMS), the residual bias (measured by BIAS), and the residual whiteness (measured by the negative of NSC). Further, we will not make any assumptions regarding the underlying distribution of the errors. Note that in the classical context of single-objective calibration [e.g., *Yapo et al.*, 1996] the measure chosen to be minimized might typically be the residual variance, while bias and whiteness would remain as part of the underlying assumptions to be tested via postcalibration residual analysis.

These three measures have been selected from Table 1, giving consideration to the fact that DRMS, TMVOL, absolute error (ABSERR), absolute maximum error (ABSMAX), and Nash-Sutcliffe coefficient (NS) are all measures of dispersion of the model residual around zero and cannot therefore reasonably be considered as unrelated (in the sense mentioned earlier); note, in particular, that NS = 1-DRMS^2/Var(d). In fact, a crude test in which all nine of the measures listed in Table 1 were computed at 500 parameter locations randomly and uniformly sampled from the entire initial parameter space indicated that the five measures of dispersion listed above (and also PDIFF and RCOEFF) were correlated to a degree exceeding ± 0.89, suggesting that they tended to measure very similar characteristics of the hydrograph (e.g., corr{DRMS-TMVOL} = +0.99; see Figure 4d). Further, those seven measures tended to be much less correlated with bias and NSC (e.g., corr{DRMS-bias} = +0.09, corr{DRMS-NSC} = −0.48, corr{bias-NSC} = −0.39}; see Figures 4a–4c).

6.3. Calibration Using the Three Objectives

The MOCOM-UA algorithm was used to estimate the Pareto solution space for the three measures DRMS, BIAS, and NSC. A search population size of 500 points was selected on the basis of experimental evidence that larger population sizes gave only marginal improvements in the approximation of the Pareto solution space [*Yapo et al.*, 1997a, b]. The procedure used 25,702 function evaluations to converge to an estimate of the Pareto set. Notice the relative efficiency of the MOCOM-UA method: while just one single-objective SCE-UA calibration would require ~5,000–10,000 function evaluations, the MOCOM-UA algorithm has generated 500 Pareto solutions with only 3–5 times as many function evalu-

Table 2. Parameters and State Variables of the Sacramento Soil Moisture Accounting (SAC-SMA) Model

Parameters Optimized	Description	Lower Bound	Upper Bound
Maximum capacity thresholds			
UZTWM	upper zone tension water maximum storage (mm)	1.0	150.0
UZFWM	upper zone free water maximum storage (mm)	1.0	150.0
LZTWM	lower zone tension water maximum storage (mm)	1.0	1000.0
LZFPM	lower zone free water primary maximum storage (mm)	1.0	1000.0
LZFSM	lower zone free water supplemental maximum storage (mm)	1.0	1000.0
ADIMP	additional impervious area (decimal fraction)	0.0	0.4
Recession parameters			
UZK	upper zone free water lateral depletion rate (day^{-1})	0.1	0.5
LZPK	lower zone primary free water depletion rate (day^{-1})	0.0001	0.025
LZSK	lower zone supplemental free water depletion rate (day^{-1})	0.01	0.25
Percolation and other parameters			
ZPERC	maximum percolation rate (dimensionless)	1.0	250.0
REXP	exponent of the percolation equation (dimensionless)	1.0	5.0
PCTIM	impervious fraction of the watershed area (decimal fraction)	0.0	0.1
PFREE	fraction of water percolating from upper zone directly to lower zone free water storage (decimal fraction)	0.0	0.6

Parameters Not Optimized	Description	Fixed Value
RIVA	riparian vegetation area (decimal fraction)	0.0
SIDE	ratio of deep recharge to channel baseflow (dimensionless)	0.0
RSERV	fraction of lower zone free water not transferrable to lower zone tension water (decimal fraction)	0.3

State Variables	Description
UZTWC	upper zone tension water storage content (mm)
UZFWC	upper zone free water storage content (mm)
LZTWC	lower zone tension water storage content (mm)
LZFPC	lower zone free primary water storage content (mm)
LZFSC	lower zone free secondary water storage content (mm)
ADIMC	additional impervious area content (mm)

ations. The estimated parameter trade-off region $P(\Theta)$ is shown in Figure 5a (each line on the plot represents one of the 500 estimates), and the final hydrograph trade-off region $P(Q)$ (indicated simply as the range between minimum and maximum) is compared with the observed data (in log space) in Figure 5b. Note that the parameter trade-off region (Pareto solution space $P(\Theta)$) is quite small compared to the initial parameter uncertainty $U(\Theta)$). Any parameter set chosen from within this Pareto space is a good solution in the sense that it provides a certain trade-off in the minimization of the three objectives. Any parameter set chosen from outside this Pareto space is a bad solution in the sense that it will have worse values for all three objectives than any points within the Pareto space. The user who requires a best parameter set (or sets) will need to decide on which kind of trade-off(s) among the objectives is acceptable for the model application on hand.

The hydrograph trade-off space plot (Figure 5b) indicates that any of the Pareto parameter sets provides extremely good matching of the medium- to high-flow regions of the hydrograph; this is indicated by the narrow hydrograph bounds in these regions. In contrast, the model seems unable to match (in a relative error sense) the recession portions of the hydrograph as well as the flood events; this is indicated by relatively wide hydrograph bounds and portions where the bounds do not bracket the observed data. If the model were to be used for applications requiring accurate estimates of moderate and low flows (e.g., as input to a streamflow chemistry model or to estimate biochemical oxygen demand), we suspect that the model structure may need to be refined. All in all, however, we can conclude that the model calibration conducted using the multiobjective procedure has been very successful and that the calibrated SAC-SMA model can be used for flood forecasting (its intended role) with a considerable degree of confidence.

6.4. Streamflow Forecasting Using the Calibrated Model

The classical approach to model calibration and flow forecasting relies on the selection of a single best parameter estimate and forecast. However, as has been suggested in the literature [see, e.g., *Beven and Binley*, 1992], given the relative abundance of computational power, there is no reason why the user cannot generate several plausible model forecasts or simulations associated with the entire set of estimated parameters. In the context of the multiobjective approach presented here these plausible model forecasts can be based on a sampling of parameter estimates from within the Pareto solution space as illustrated in Figure 5b using range forecasts indicating upper and lower limits. In addition, sample trajectories related to minimum DRMS, minimum bias, and minimum NSC, etc. (as determined through the calibration procedure), could be displayed as additional information (not shown in Figure 5 because the uncertainty in estimation of the peaks flows is already quite small).

If, however, it is necessary (for practical and computational reasons) to forecast or simulate a single "most likely" flow value, the user is faced with the task of selecting from within the Pareto solution space a specific parameter estimate. This, of course, will require the subjective assignment of the relative importance to the various measures to arrive at a best compromise solution. To facilitate this process, we are currently developing graphic visualization techniques that display the

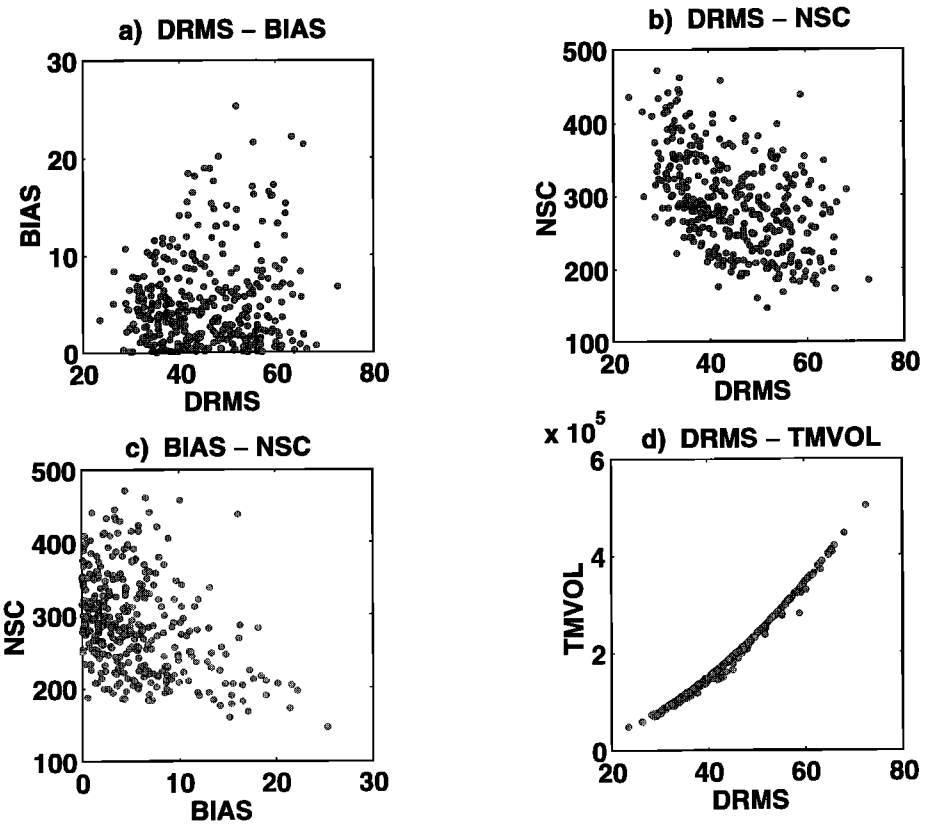

Figure 4. Four selected plots showing the 500 randomly generated points projected in two-objective subspaces.

decision variables (parameter estimates, associated objective function values, and hydrographs) in a convenient manner, allowing the user to "mouse click" on a particular selection (parameter estimate or objective function weighting) and observe the simulated hydrograph associated with it. This subjective procedure will enable the user to apply additional information and personal experience to the decision process in an efficient manner. Many of the systems theoretical developments in the field of multiobjective decision making can also be used to advantage [e.g., *Laabs and Schultz*, 1992; *Goicoechea et al.*, 1976; *Haimes et al.*, 1975]. A powerful advantage of this approach is that it can include the classical solution (e.g., best DRMS parameter estimates) among the several alternatives provided to the decision maker but does not mask the fact that on the basis of the available information none of the solutions is inherently superior to any other.

7. Conclusion

Field measurements, prior information, and manual and automated techniques for calibration are three techniques used in parameter estimation for hydrologic models. With the growing popularity of complex physically based distributed watershed models (e.g., land-surface hydrology and hydrochemical models) the use of more and better field measurements for specifying model parameters has gained in importance and attention. However, the requirement for automated parameter estimation techniques is not going to simply disappear. Our research experiences suggest that the classical model calibration paradigm needs serious review and that further progress will only come about through the implementation of a new and more powerful paradigm based in part on the ideas advanced in this paper. In particular, it is necessary to recognize that (1) the structural errors, arising from the fact that any model is only an approximation (hopefully reasonable) of reality, cannot be ignored or treated only as stochastic variables to be lumped into some output residual, (2) the problem of model identification and calibration is inherently multiobjective, even in the case of only one output time series (the application of least squares and other statistical techniques for model fitting is largely an attempt to bypass the difficulties inherent in multiobjective approaches), and (3) there is a real need to be able to judge the reliability of a model, not as some overall approximate measure but in terms of each model prediction. A further issue not addressed in this paper is that the errors in the input data cannot be ignored.

When these facts are faced head on, it becomes apparent that there is no objective way in which a unique model solution can be obtained. Rather, the best that one can obtain using objective procedures is a model set, specifiable as a region of the parameter space. In the context of multiple measures of model performance this model set defines the Pareto solution set (which is also a minimal estimate of the parameter uncertainty) in which it is not possible to objectively select a specific parameter set (model) as being superior to any other parameter set (model). This Pareto solution space translates into a trade-off range in the model predictions (the model is only capable of, at best, indicating the range in which the field observation might be observed). The size and properties of this

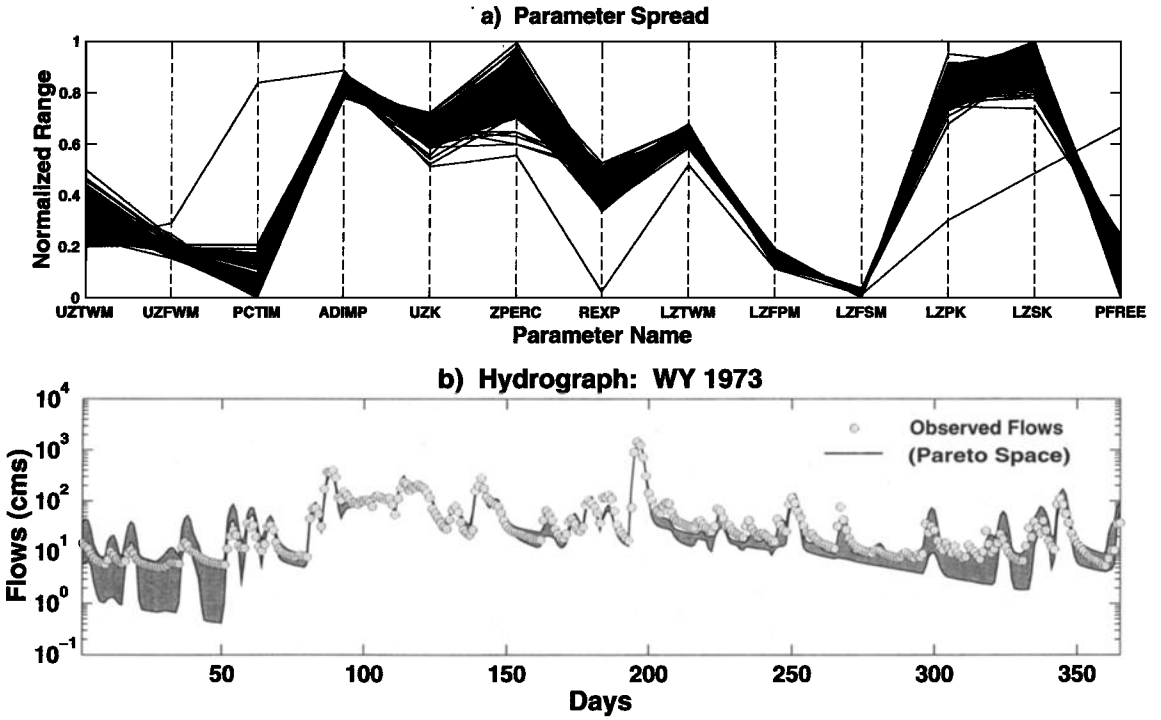

Figure 5. The final parameter and hydrograph ranges obtained by multiobjective calibration of the SAC-SMA model using three objectives.

model set and the sizes and properties of the trade-off range in the model predictions are characteristics which will help in the evaluation of the adequacy or inadequacy of the model. Analysis of these features will provide insight into the manner in which the model needs to be improved and into the confidence that can be ascribed to the model predictions.

The results of our preliminary investigation of this new model calibration approach have been presented in this paper. Through a case study we have shown that the multiobjective calibration approach is practical and relatively simple to implement and can also provide useful information that helps to understand better the limitations of a model. Research into a number of theoretical and experimental issues related to this work is ongoing. This includes (1) the proper manner for selecting the set of measures of model performance, (2) the sensitivity of the results to the number of measures and the amount of data, and (3) the extension of the multiobjective theory to account for stochastic uncertainties in the observation data, thereby providing more than a minimal estimate of model uncertainty. In collaboration with colleagues the multiobjective calibration approach is currently also being applied to some of the more sophisticated physically based hydrologic models such as soil-vegetation transfer schemes (SVATS) and hydrochemical watershed models. The results of these studies will be reported in due course. We welcome dialog on these and other ideas related to hydrologic model calibration. The code for the MOCOM multiobjective optimization algorithm is available from the first author by request (send email to hoshin@hwr.arizona.edu).

Acknowledgments. The current version of this manuscript has been greatly improved as a result of constructive and critical review comments provided by R. Clarke, D. A. Jones, and K. Beven, as well as enthusiastic discussions with the members of the HWR Surface Water Modeling and Calibration Group: K. Hsu, L. Bastidas, M. Winchell, M. Thiemann, and A. Vizzaccaro. Partial financial support for this research was provided by the National Science Foundation (grants EAR-9415347 and EAR-9418147), the Hydrologic Research Laboratory of the National Weather Service (grants NA47WH0408 and NA57WH0575), and the National Aeronautics and Space Administration (NASA-EOS grant NAGW2425). Financial assistance provided to P. O. Yapo by The University of Arizona Graduate College is gratefully acknowledged.

References

Beven, K. J., Prophecy, reality and uncertainty in distributed hydrological modeling, *Adv. Water Resour.*, *16*, 41–51, 1993.

Beven, K. J., and A. M. Binley, The future of distributed models: Model calibration and uncertainty prediction, *Hydrological Processes*, *6*, 279–298, 1992.

Beven, K. J., and M. Kirkby, A physically based variable contributing area model of basin hydrology, *Hydrol. Sci. Bull.*, *24*, 43–69, 1979.

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

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.

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

Burnash, R. J. E., R. L. Ferral, and R. A. McGuire, *A Generalized Streamflow Simulation System*, 204 pp., Joint Fed.-State River Forecast Cent., Sacramento, Calif., 1973.

Chapman, T. G., Optimization of a rainfall-runoff model for an arid-zone catchment, *Symposium on the Results of Research on Representative and Experimental Basins*, vol. *96*, pp. 126–144, Int. Assoc. of Sci. and Hydrol., 1970.

Clarke, R. C., *Statistical Modeling in Hydrology*, John Wiley, New York, 1994.

De Grosbois, E., R. P. Hooper, and N. Christopherson, A multisignal automatic calibration methodology for hydrochemical models: A

case study of the Birkenes Model, *Water Resour. Res.*, *24*, 1299–1307, 1988.

Diskin, M. H., and E. Simon, A procedure for the selection of objective function for hydrologic simulation models, *J. Hydrol.*, *34*, 129–149, 1977.

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

Duan, Q., V. K. Gupta, and S. Sorooshian, A shuffled complex evolution approach for effective and efficient global minimization, *J. Optim. Theory Appl.*, *76*, 501–521, 1993.

Duan, Q., S. Sorooshian, and V. K. Gupta, Optimal use of the SCE-UA global optimization method for calibrating watershed models, *J. Hydrol.*, *158*, 265–284, 1994.

Freer, J., A. M. Beven, and B. Ambroise, Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach, *Water Resour. Res.*, *32*, 2161–2173, 1996.

Gan, T. Y., and G. F. Biftu, Automatic calibration of conceptual rainfall-runoff models: Optimization algorithms, catchment conditions, and model structure, *Water Resour. Res.*, *32*, 3513–3524, 1996.

Gembicki, F. W., Vector optimization for control with performance and parameter sensitivity indices, Ph.D. thesis, Case Western Reserve Univ., Cleveland, Ohio, 1974.

Goicoechea, A., L. Duckstein, and M. M. Fogel, Multiobjective programming in watershed management: A study of the Charleston watershed, *Water Resour. Res.*, *12*, 1085–1092, 1976.

Goicoechea, A., D. R. Hasen, and L. Duckstein, *Multi-Objective Decision Analysis with Engineering and Business Applications*, 511 pp., John Wiley, New York, 1982.

Goldberg, D. E., *Genetic Algorithms in Search, Optimization, and Machine Learning*, 412 pp., Addison-Wesley, Reading, Mass., 1989.

Gupta, V. K., and S. Sorooshian, The relationship between data and the precision of parameter estimates of hydrologic models, *J. Hydrol.*, *81*, 57–77, 1985.

Gupta, V. K., and S. Sorooshian, Calibration of conceptual hydrologic models: Past, present and future, in *Trends in Hydrology*, edited by Council of Scientific Research Integration, pp. 329–346, Research Trends, Kaithamukku, India, 1994a.

Gupta, V. K., and S. Sorooshian, A new optimization strategy for global inverse solution of hydrologic models, in *Numerical Methods in Water Resources*, edited by A. Peters et al., Kluwer Acad., Norwell, Mass., 1994b.

Haimes, Y. Y., and W. A. Hall, Multiobjectives in water resource systems analysis: The surrogate worth trade off method, *Water Resour. Res.*, *10*, 615–624, 1974.

Haimes, Y. Y., W. A. Halls, and H. T. Freedman, *Multiobjective Optimization in Water Resources Systems: The Surrogate Worth Trade-off Method*, 200 pp., Elsevier Sci., New York, 1975.

Hipel, K. W., Multiple objective decision making in water resources, *Water Resour. Bull.*, *28*, 3–12, 1992.

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

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, Am. Soc. Agric. Eng., 1982.

Jones, D. A., Statistical analysis of empirical models fitted by optimization, *Biometrika*, *70*, 67–88, 1983.

Keesman, K. J., Set-theoretic parameter estimation using random scanning and principal component analysis, *Math. Comput. Simul.*, *32*, 535–543, 1990.

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

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

Kitanidis, P. K., and R. L. Bras, Real-Time forecasting with a conceptual hydrological model, 2, Applications and results, *Water Resour. Res.*, *16*, 1034–1044, 1980c.

Klepper, O., H. Scholten, and J. P. G. van de Kamer, Prediction uncertainty in an ecological model of the Oosterschelde Estuary, *J. Forecasting*, *10*, 191–209, 1991.

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

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

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

Kuczera, G., On validity of first-order prediction limits for conceptual hydrological models, *J. Hydrol.*, *103*, 229–247, 1988.

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

Laabs, H., and G. A. Schultz, Reservoir management rules derived with the aid of multiple objective decision making techniques, *Water Resour. Bull.*, *28*, 211–222, 1992.

Leavesley, G. H., R. W. Lichty, B. M. Troutman, and L. G. Saindon, Precipitation-runoff modeling system: User's manual, *Water Resour. Invest. Rep. 83-4238*, U.S. Geol. Surv., Denver, Colo., 1983.

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, Indiana, 1983.

Luce, C. H., and T. W. Cundy, Parameter identification for a runoff model for forest roads, *Water Resour. Res.*, *30*, 1057–1069, 1994.

Marglin, S. A., *Public Investment Criteria*, MIT Press, Cambridge, Mass., 1967.

Peck, E. L., Catchment modeling and initial parameter estimation for the National Weather Service River Forecast System, *NOAA Tech. Mem. NWS Hydro-31*, 60 pp., U.S. Dep. of Commer., Silver Spring, Md., 1976.

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

Sorooshian, S., Parameter estimation of rainfall-runoff models with heteroscedastic streamflow errors: The noninformative data case, *J. Hydrol.*, *52*, 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*, 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*, 251–259, 1983.

Sorooshian, S., V. K. Gupta, and J. L. Fulton, Parameter estimation of conceptual rainfall-runoff models assuming autocorrelated data errors: A case study, in *Statistical Analysis of Rainfall and Runoff*, edited by V. P. Singh, pp. 491–504, WRP, 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*, 251–259, 1983.

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

Spear, R. C., and G. M. Hornberger, Eutrophication in Peel Inlet, II, Identification of critical uncertainties via generalized sensitivity analysis, *Water Res.*, *14*, 43–49, 1980.

Szidarovszky, F., and K. Szenteleki, A multiobjective optimization model for wine production, *Appl. Math. Comput.*, *22*, 255–275, 1987.

Szidarovszky, F., M. E. Gershon, and L. Duckstein, *Techniques for Multiobjective Decision Making in Systems Management, Advances in Industrial Engineering*, vol. 2, 506 pp., Elsevier, New York, 1986.

Tanakamaru, H., Parameter estimation for the tank model using global optimization, *Trans. Jap. Soc. Irrig. Drainage Reclam. Eng.*, *178*, 103–112, 1995.

Tanakamaru, H., and S. J. Burges, Application of global optimization to parameter estimation of the tank model, paper presented at International Conference on Water Resources and Environment Research: Towards the 21st Century, Kyoto, Japan, October 29–31, 1996.

Troutman, B. M., Errors and parameter estimation in precipitation-runoff modeling, 1, Theory, *Water Resour. Res.*, *21*, 1195–1213, 1985a.

Troutman, B. M., Errors and parameter estimation in precipitation-runoff modeling, 2, Case study, *Water Resour. Res. 21*, 1195–1213, 1985b.

van Straten, G., Maximum likelihood estimation of parameters and

uncertainty in phytoplankton models, in *Uncertainty and Forecasting of Water Quality*, edited by M. B. Beck and G. van Straten, pp. 151–171, Springer-Verlag, New York, 1983.

van Straten, G., and K. J. Keesman, Uncertainty propagation and speculation in projective forecasts of environmental change: A lake-eutrophication example, *J. Forecasting, 10*, 163–190, 1991.

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

Whitley, D., The genitor algorithm and selection pressure: Why rank-based allocation of reproductive trials is best, paper presented at Third International Conference on Genetic Algorithms, 1989.

Woolhiser, D. A., R. E. Smith, and D. C. Goodrich, A kinematic runoff and erosion manual: Documentation and user manual, *ARS 77*, Agric. Res. Serv., U.S. Dep. of Agric., Washington, D. C., 1990.

Yan, J., and C. T. Haan, Multiobjective parameter estimation for hydrologic models—Weighting of errors, *Trans. Am. Soc. Agric. Eng., 34*, 135–141, 1991a.

Yan, J., and C. T. Haan, Multiobjective parameter estimation for hydrologic models—multiobjective programming, *Trans. Am. Soc. Agric. Eng., 34*, 848–856, 1991b.

Yapo, P. O., N. Buras, and F. Szidarovszky, Applications of multiobjective decision making in water resources management, *Pure Math. Appl., Ser. C, 3*, 77–112, 1992.

Yapo, P., H. V. Gupta, and S. Sorooshian, Calibration of conceptual rainfall-runoff models: Sensitivity to calibration data, *J. Hydrol., 181*, 23–48, 1996.

Yapo, P. O., H. V. Gupta, and S. Sorooshian, Multi-objective global optimization for hydrologic models, *J. Hydrol.*, in press, 1997a.

Yapo, P. O., H. V. Gupta, and S. Sorooshian, A multiobjective global optimization algorithm with application to calibration of hydrologic models, *HWR Tech. Rep. 97-050*, Dep. of Hydrol. and Water Resour., Univ. of Ariz., Tucson, 1997b.

Zadeh, L. A., Optimality and non-scalar valued performance criteria, *IEEE Trans., C-8*, 59–60, 1963.

Zeleny, M., *Compromise Programming in Multiple Criteria Decision Making*, edited by J. L. Cocchrane and M. Zeleny, pp. 262–301, Univ. S. C. Press, Columbia, 1973.

H. V. Gupta and S. Sorooshian, Department of Hydrology and Water Resources, University of Arizona, Box 210011, Tucson, AZ 85721-0011. (e-mail: hoshin@hwr.arizona.edu)

P. O. Yapo, Department of Systems and Industrial Engineering, University of Arizona, Tucson, AZ 85721.

(Received April 1, 1997; revised December 1, 1997;
accepted December 1, 1997.)