JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 104, NO. D16, PAGES 19,491–19,503, AUGUST 27, 1999

Parameter estimation of a land surface scheme using multicriteria methods

H. V. Gupta, L. A. Bastidas, S. Sorooshian, W. J. Shuttleworth, and Z. L. Yang

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

Abstract. Attempts to create models of surface-atmosphere interactions with greater physical realism have resulted in land surface schemes (LSS) with large numbers of parameters. The hope has been that these parameters can be assigned typical values by inspecting the literature. The potential for using the various observational data sets that are now available to extract plot-scale estimates for the parameters of a complex LSS via advanced parameter estimation methods developed for hydrological models is explored in this paper. Results are reported for two case studies using data sets of typical quality but very different location and climatological regime (ARM-CART and Tucson). The traditional single-criterion methods were found to be of limited value. However, a multicriteria approach was found to be effective in constraining the parameter estimates into physically plausible ranges when observations on at least one appropriate heat flux and one properly selected state variable are available.

1. Introduction and Scope

This paper is one of three papers that discuss the usefulness of multicriteria methods for the evaluation and improvement of a land surface scheme (LSS). The focus of this paper is to show how multicriteria methods can be used to improve the estimates of LSS parameters. A companion paper [*Bastidas et al*., this issue] introduces a multicriteria approach to parameter sensitivity analysis, thereby providing a way to reduce the dimensionality of the parameter estimation problem. In both papers, the methodology is illustrated using the Biosphere-Atmosphere Transfer Scheme (BATS) [*Dickinson et al*., 1993]) and two data sets, one from a grassland site in Oklahoma/Kansas, and the other from a semiarid site in the Sonoran Desert, Arizona. A third paper (in preparation) will discuss the use of multicriteria methods for the evaluation of model performance and for model intercomparison.

The paper is organized as follows: Section 2 discusses the background and context for this work and presents a review of the literature. Section 3 introduces the theoretical and practical basis for applying multicriteria calibration methods to the estimation of LSS model parameters. Calibration of the BATS model to the two study sites is discussed in section 4, and section 5 discusses the results and future extensions.

2. Background

In order for an LSS to simulate the input-state-output behavior of a portion of the land surface with minimal uncertainty, it is necessary to estimate (select) appropriate values for the model parameters. Here this process is referred to as parameter estimation. It is important to note that because an LSS is a (simplified) conceptual representation of the real system, the model parameters are also conceptual representations of physical properties of the system. Where the concordance between the model conceptualization and the real world is somehow close, it may be possible to obtain estimates for certain model parameters by direct measurements on the real system. However, where the concordance is less close, the model parameters can realistically only be viewed as abstract conceptual representations of physical quantities for which useful estimates cannot be found by direct measurement.

From the earliest simple LSSs such as the "bucket model" [*Manabe*, 1969] to the relatively complex contemporary LSSs such as the Biosphere-Atmosphere Transfer Scheme (BATS) [*Dickinson et al*., 1993], the simple biosphere model (SiB2) [*Sellers et al*., 1996] and the variable infiltration capacity (VIC-2L) model [*Liang et al*., 1994], a common characteristic is the approximate description of the large-scale integrated areal land surface response using system equations derived from the understanding of land surface physics at the relatively small patch scale. Further, because of the need to parameterize an LSS for each (large scale) GCM grid square across the globe, the goal has usually been to construct schemes whose large-scale parameters can be defined from observed (hence measurable) quantities. For example, the BATS model uses look-up tables for model parameters corresponding to a fixed number of discrete soil-vegetation-climate regimes, the values in the tables having being deduced from studies published in the literature.

Some LSS model parameters, such as albedo, percent vegetated cover, leaf area index, etc., can be measured (estimated) at both the patch scale and the large scale (via remote sensing), and the relationship between their area-averaged values at patch and large scales is linear. However, other parameters such as soil hydraulic conductivity, stomatal resistance, aerodynamic resistance, etc., are not easily measured at the relevant scales, and their relationship at different scales is less simple [*Arain et al*., 1996; *Shuttleworth et al*., 1997]. For example, saturated soil hydraulic conductivity is defined as the rate of flow of water through a unit cross section of soil caused by a unit hydraulic gradient. Even at the small scale, there is poor consistency in the estimates obtained using different measurement techniques [*Brewer and Wheatcraft*, 1994]. Consequently, at the GCM-LSS model scales, there is poor understanding of

Paper number 1999JD900154.
0148-0227/99/1999JD900154$09.00

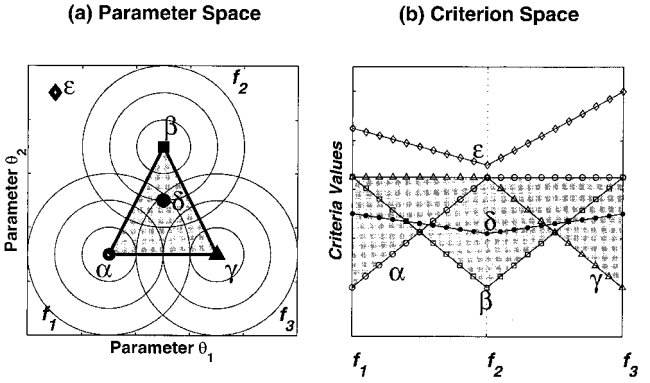


Figure 1. Example showing the Pareto solution set (shaded region) for a simple problem having two parameters (θ_1, θ_2) and three criteria (f_1, f_2, f_3): (a) parameter space plot showing contours of constant function value for each criterion. The points α, β, and γ minimize different criteria, δ is a multicriteria (Pareto) solution and ε is an inferior solution); (b) multidimensional criterion space projected onto a two-dimensional plot. Each vertical axis indicates a different criterion, and each line joining all three axes indicates a point in the parameter space.

how to obtain a valid and representative estimate. All that might reasonably be specified is the approximate range for the parameter value, based on some approximate understanding of the regional hydrogeology. (Note that the current practice of using parameter look-up tables based on approximate knowledge of climatic regime, vegetation, and soil types is arguably prone to error). If the input-state-output response of the LSS is not sensitive to parameter variations within this range [see *Bastidas et al*., this issue], it is presumably reasonable to use some nominal estimate (such as the midpoint of the parameter uncertainty range). However, if the LSS response is sensitive to finer specification of the parameter, the only remaining recourse is to adjust the value of the parameter so that the model responses are constrained to better match available observations. Here this process is referred to as model calibration or, more simply, calibration.

Although current LSSs involve much conceptual abstraction, it has not been usual to apply calibration procedures to them. However, a few recent publications have investigated the potential benefits of calibration. For example, *Sellers et al*. [1989] employed manual calibration on the SiB model to improve estimates of nine of its parameters by fitting to five 2 week periods of field data. *Franks and Beven* [1997] used the generalized likelihood uncertainty estimation approach to reduce the output flux uncertainty for the TOPUP model by constraining the model parameters using short periods of FIFE (10 days) and ABRACOS (2.5 months) data. Perhaps the most revealing results come from the PILPS2c workshop, during which a large number of LSS models were intercompared; a significant conclusion was that the models that employed parameter adjustment through calibration tended to provide better performance than those that did not [*Lettenmaier et al*., 1996].

During the past two decades, the issue of how to calibrate conceptual models to observed data has received substantial attention within the hydrologic community (for a review, see *Gupta et al*. [1998]). This has included the development of maximum likelihood techniques for measuring the closeness between a model output and the data [e.g., *Sorooshian and Dracup*, 1980; *James and Burges*, 1982], reliable optimization methods for solving the nonlinear parameter estimation problem [e.g., *Duan et al*., 1992; *Sorooshian et al*., 1993], insight into the appropriate quantity and most informative kind of data [e.g., *Kuczera*, 1982; *Gupta and Sorooshian*, 1985; *Yapo et al*., 1996], and methods for representing model uncertainty [e.g., *Spear and Hornberger*, 1980; *Kuczera*, 1988; *Franks et al*., 1998]. A framework for extending this knowledge and experience to the emerging generation of multi-input-output hydrologic models was recently presented by *Gupta et al*. [1998] and tested in preliminary fashion on the U.S. National Weather Service flood forecasting model [*Gupta et al*., 1998; *Yapo et al*., 1997] and a soil moisture assimilation model [*Houser*, 1996].

The work presented here uses the framework presented by *Gupta et al*. [1998]. The major aim of this paper is to demonstrate a systematic and comprehensive approach for the application of calibration procedures to LSS models, to illustrate the consequent benefits that can accrue, and to discuss some of the difficulties that may be encountered. We show that multicriteria calibration leads to conceptually realistic estimates for the model parameters with consequent improvements in model performance.

3. Multicriteria Parameter Estimation Methodology

A theoretical and practical basis for the application of multicriteria theory to the calibration of conceptual physically based models was presented by *Gupta et al*. [1998]. The method is summarized briefly as follows: Consider a model having parameters $\theta = \{\theta_1, \cdots, \theta_p\}$ which is to be calibrated using time series observations collected on m different simulated response variables ($Z_j(\theta, t_j)$, $t_j = ta_j, \cdots, tb_j$, $j = 1, \cdots, m$). A separate criterion $f_j(\theta)$ for each model response is defined to measure the distance between the model-simulated responses Z_j and the observations O_j. The specification of the mathematical form of these criteria depends on the problem and the goals of the user. However, it is common practice to use a measure of residual variance such as the root-mean-square error ($\mathrm{RMSE}_j(\theta) = \mathrm{sqrt}\ \{(1/n)\Sigma_{t=1,\cdots,n}\ (Z_j(\theta, t) - O_j(t))^2\}$); for a discussion of this, see *Gupta et al*. [1998]. The multicriteria model calibration problem can then be formally stated as the optimization problem:

$$\text{Minimize } F(\theta) = \{f_1(\theta), \cdots, f_m(\theta)\} \text{ subject to } \theta \subset \Theta \quad (1)$$

where the goal is to find the values for θ within the feasible set Θ that simultaneously minimize all of the m criteria.

An important characteristic of the multiobjective problem is that it does not, in general, have a unique solution. Because of errors in the model structure (and other possible sources), it is not usually possible to find a single point θ at which all the criteria have their minima. Instead, it is common to have a set of solutions, with the property that moving from one solution to another results in the improvement of one criterion while causing deterioration in another. A simple problem having two parameters (θ_1, θ_2) and three criteria (f_1, f_2, f_3) is illustrated in Figure 1. Figure 1a shows the feasible parameter space Θ, and Figure 1b is a projection of the positive quadrant of the multidimensional criteria space onto a two-dimensional plot; each vertical axis line indicates a different criterion, and each point in Θ is represented by a line joining all three axes. The

points α, β, and γ indicate the different solutions that minimize each of the individual criteria. The shaded region indicates the set S of multicriteria solutions to the problem. If δ and ε are points arbitrarily selected from inside and outside S, respectively, then, in a multicriteria sense, every point δ is superior to every point ε; that is, $f_j(\delta) < f_j(\varepsilon)$, for all $j = 1, \cdots, 3$. However, no point in S is superior to any other. A particular point may be superior to others for one (or more) criterion, but it is inferior to them for at least one other criterion. The set S of solutions is variously called the trade-off set, noninferior set, nondominated set, or efficient set, but here, we call it the Pareto set. The Pareto set represents the minimal uncertainty that can be achieved for the parameters via calibration, without subjectively assigning relative weights to the individual model responses. The size and properties of this set are related to errors in the model structure and data. Only when the model is a perfect representation of the system (and there are no systematic biases in the observation data) can S become a unique solution.

There are a number of different approaches to solving the multicriteria problem (equation (1)), and the relative merits of these is the subject of ongoing research [e.g., see *Harboe*, 1992]. Recently, *Yapo et al*. [1997] presented an efficient population-based optimization strategy that can provide an approximate representation of the Pareto set with a single optimization run. This algorithm, which is called the multiobjective complex evolution (*MOCOM-UA*) method, is based on extensions of the successful *SCE-UA* single-criterion method [*Duan et al*., 1992, 1993, 1994]. The MOCOM-UA method begins by uniformly sampling the feasible space Θ at a number of locations and then uses a multicriteria population evolution strategy to drive this population of sample points toward the Pareto region; for details, see [*Yapo et al*., 1997]. The final solution therefore consists of a set of randomly distributed points which approximately represent the Pareto set.

4. Case Studies

The multicriteria parameter estimation methodology described above was used to calibrate a land surface scheme to two different data sets, one from the ARM-CART (atmospheric radiation measurement—cloud and radiation test bed) project grassland site in Oklahoma/Kansas and the other from a Sonoran Desert semiarid site [*Unland et al*., 1996]. In both cases the data are representative of the local small-(patch) scale hydrometeorology (rather than the scale of a GCM grid), and thus a critical evaluation of the performance of the parameter estimation method is possible. The LSS used for this study was the off-line version of the Biosphere-Atmosphere Transfer Scheme (BATS 1e) [*Dickinson et al*., 1993], which has 24 independent parameters and three initial soil-moisture conditions that must be specified.

4.1. BATS Land Surface Scheme

BATS is a conceptual parameterization of the ecohydrological processes at the scale of individual plots of vegetation (50–1000 m). The model consists of six interacting hydrometeorological components (three layers of soil, a canopy air component, a canopy leaf-stem component, and a snow-covered portion). Together, these components simulate the various radiative and hydrological process at the land atmosphere interface, including the exchange of solar and longwave radiation, precipitation inputs (rain, snow, and dew), runoff, and the surface transfer of momentum and sensible and latent heat exchanges. The specification of parameters is conventionally made via a global land surface classification consisting of 18 land cover types (12 vegetation and 6 nonvegetation) and 12 global soil types. Each land cover type is characterized by 16 vegetation-related parameters, and each soil type is characterized by eight parameters that represent the hydraulic, thermal, and reflective properties of the soil (see Table 1). Therefore the model has a total of 24 parameters that must be estimated, but two of these, xmowil (the wilting point parameter) and xmofc (the ratio of the field capacity to the saturated water content) are not independent parameters. The parameter xmowil is computed as a function of two other parameters, namely the hydraulic conductivity (xmohyd) and the minimum soil suction (xmosuc), while xmofc is used only when the land cover is assigned to be semidesert [*Dickinson et al*., 1993].

In principle, the BATS model computes the evolution of 12 state variables, namely the temperature and water content for each of the six model components. However, two of these are not independent because the model assumes that the temperature of the lowest soil layer is constant and that when snow cover is present, it has the same temperature as the upper soil layer. Thus BATS uses 10 water-energy conservation equations to solve for the dynamical evolution of the 10 independent state variables. *Gao et al*. [1996] showed that errors in specification of initial values for these state variables tend to decay rapidly, with the notable exception being the initial moisture contents of the soil layers. In the work of *Bastidas et al*. [this issue] the results of a multicriteria parameter sensitivity analysis show that the sensitivity of the model response to the initial soil-moisture contents is significant. Therefore to avoid problems caused by poor specification of these variables, the initial soil-moisture contents of the three soil layers were included as parameters to be estimated.

4.2. Parameter Estimation for ARM-CART Grassland Study Site

4.2.1. Data set. The ARM-CART study area covers a wide region in the southern Great Plains and includes parts of the states of Kansas and Oklahoma. The spatial coverage of the project enables the study of the spatial variation of land surface characteristics over an area. Data from one centrally located site (E13) were used in this study. The data cover the 5-month period from April 1 to August 25, 1995, with a sampling interval of 30 min. The observed atmospheric variables are net radiation (R_n in W/m^2), surface temperature (T_a in K), atmospheric pressure (p_a in kPa), relative humidity (r_h in %), wind velocity (V_a in m/s), and precipitation (P in mm).

The available data contained many gaps and some obvious errors. Because an uninterrupted time series of inputs is required by the model, short gaps (2–3 hours) were filled by simple linear interpolation. One of the gaps spanned 3 days, and data from a nearby site (E15) were then used to complete the series.

The data set also included observed time series of variables that correspond with two model outputs (sensible heat (H in W/m^2) and latent heat (λE in W/m^2)), and two model state variables (ground temperature (T_g in K) and soil moisture (S_w in mm)). The data on H and λE are derived from a Bowen ratio system that is reported to have poor reliability when the value of the Bowen ratio is close to 1.0; therefore all values of H and λE in periods when the Bowen ratio was between 0.75

Table 1. Definition of BATS Parameters

				Tucson			ARM-CART		
	Name	Description (units)	Global Range	Reasonable Range	Default Value	Calibrated Range	Reasonable Range	Default Value	Calibrated Range
		Parameters Associated With Vegetation (18 Vegetation Types)							
1	vegc	vegetation cover	0.00–0.95	0.10–0.70	0.10	0.64–0.70	0.40–0.95	0.80	0.90–0.95
2	seasf	difference between vegc and fractional cover at 269 K	0.00–0.60	0.00–0.60	0.10	0.12–0.28	0.00–0.80	0.60	0.21–0.34
3	rough	aerodynamic roughness length (m)[a]	0.0024–1.00	0.01–0.84	0.10	0.11–0.06	0.01–0.35	0.06	0.01–0.21
4	displa	displacement height (m)[a]	0.00–5.00	0.05–1.50	0.00	0.47–0.78	0.05–1.50	0.00	1.47–2.54
5	rsmin	minimum stomatal resistance (s/m)	5.0–200.0	50.0–200.0	200.0	134.8–165.5	50.0–200.0	200.0	48.20–93.85
6	xla	maximum leaf area index	0.00–6.00	0.05–3.00	6.00	1.73–2.31	0.05–6.00	6.00	3.53–4.74
7	xlai0	minimum leaf area index	0.00–5.00	0.05–3.00	0.50	0.05–0.45	0.05–4.00	0.50	0.11–2.24
8	sai	stem area index	0.50–4.00	1.00–4.00	2.00	2.37–3.36	1.00–3.00	0.50	2.48–2.84
9	sqrtdi	inverse sqrt of leaf dimension (mm∧-0.5)	5.00–10.00	5.00–10.00	5.00	6.97–8.48	5.00–10.00	5.00	6.90–8.42
10	fc	light dependence of stomatal resistance (m^2/w)	0.02–0.06	0.02–0.06	0.02	0.03–0.06	0.02–0.06	0.02	0.03–0.05
11	depuv	depth of top soil layer (m)	0.05–0.50	0.01–0.20	0.10	0.17–0.20	0.01–0.20	0.10	0.09–0.13
12	deprv	depth of root zone layer (m)	0.50–2.00	0.50–2.00	1.00	1.30–1.67	0.50–2.00	1.00	1.43–1.82
13	deptv	depth of total zone layer (m)	5.00–10.00	5.00–10.00	10.00	9.42–10.00	5.00–10.00	10.00	7.69–8.54
14	albvgs	vegetation albedo for shortwave <0.7 μm	0.04–0.20	0.10–0.20	0.17	0.11–0.13	0.10–0.20	0.08	0.13–0.15
15	albvgl	vegetation albedo for longwave >0.7 μm	0.18–0.40	0.20–0.40	0.34	0.20–0.25	0.20–0.40	0.28	0.28–0.34
16	rootf	ratio of roots in upper layer to roots in root layer	0.30–0.90	0.10–0.90	0.80	0.15–0.32	0.10–0.90	0.30	0.38–0.49
		Parameters Associated With Soil Texture (12 Textures)							
17	xmopor	porosity	0.33–0.66	0.33–0.66	0.57	0.62–0.66	0.33–0.66	0.45	0.41–0.57
18	xmosuc	minimum soil suction (mm)	30.00–200.00	30.00–200.0	200	94.56–119.12	30.0–200.0	200	114.46–142.23
19	xmohyd	maximum hydraulic conductivity (mm/s)[b]	0.0008–0.01	0.0008–0.0100	0.0022	0.0039–0.0059	0.0008–0.0100	0.0089	0.0032–0.0047
20	xmowil	water content at which permanent wilting point occurs	0.088–0.542		0.455			0.3000	
21	xmofc	ratio of field capacity to saturated water content	0.404–0.866		0.794			0.653	
22	bee	Clapp and Hornberger "*b*" parameter	3.50–10.80	3.50–10.80	8.40	8.17–10.10	3.50–10.80	5.50	7.95–8.31
23	skrat	ratio of soil thermal conductivity to that of loam	0.70–1.70	0.7–1.7	0.85	1.29–1.58	0.70–1.70	1.10	1.25–1.64
		Parameter Associated With Soil Color (8 Colors)							
24	solour	soil albedo for different colored soils	0.05–0.12	0.05–0.12	0.10	0.05–0.07	0.05–0.12	0.10	0.08–0.10
		Initial Conditions							
25	ssw	surface zone water content (m)[c]	0.00–depuv	0.0–0.2	0.10	0.06–0.10	0.0–0.2	0.03	0.02–0.08
26	rsw	root zone water content [m][c]	0.00–deprv	0.0–5.0	0.43	0.08–0.30	0.0–5.0	0.45	0.63–1.05
27	tsw	total zone water content [m][c]	0.00–deptv	0.0–10.0	2.04	2.15–4.86	0.0–10.0	0.45	3.32–5.32

Default values, global and reasonable ranges, and final ranges obtained by multicriteria calibration.
[a]Range does not include tropical rain forest.
[b]Range constrained.
[c]Default value obtained using compromise optimization.

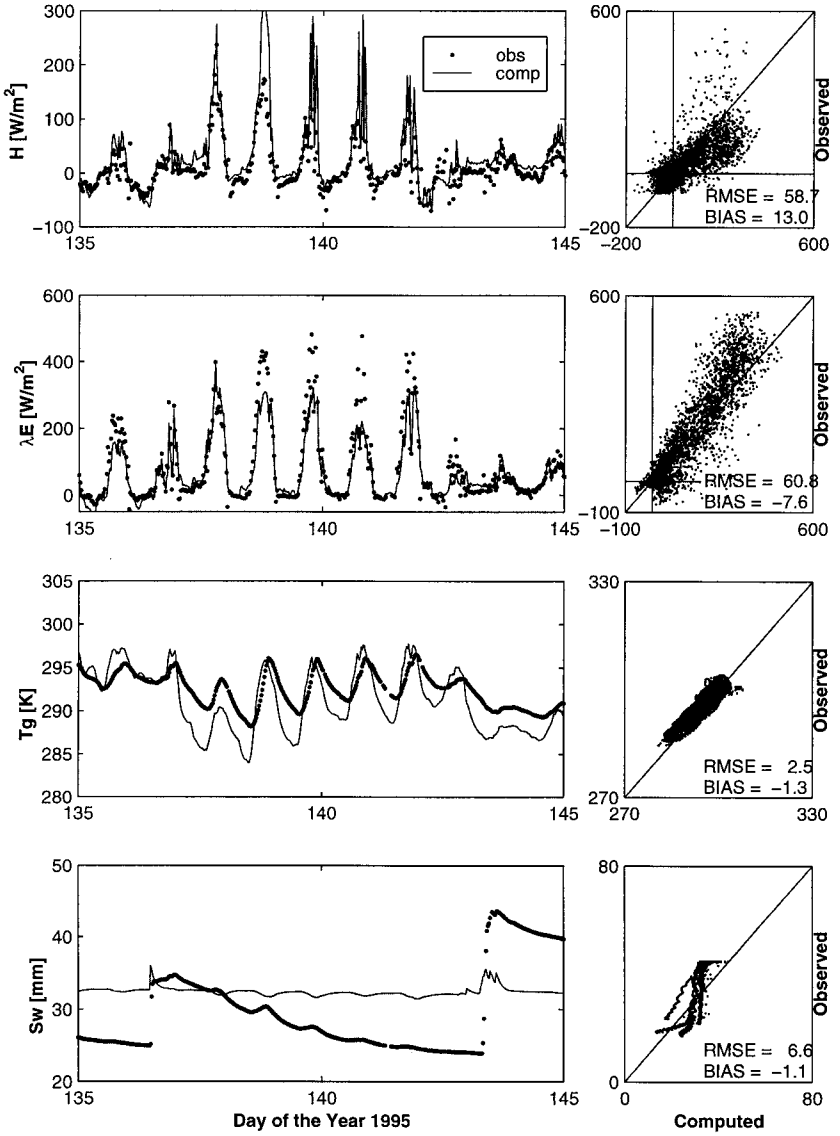


Figure 2. Comparison of modeled variables (solid line) with the observed data (dots) for the ARM-CART site default parameters. The time series plots on the left show a representative 10 day period during which rainfall occurred. The scatterplots on the right show the entire period of data.

and 1.25 were discarded. Further, a model spin-up period of 1 month was used to reduce the influence of the errors in the specification of initial values of the state variables. The resulting number of time steps for which values of H, λE, T_g, and S_w are available for constraining the model parameters is 4237. In this study of model calibration the root-mean-square error (RMSE) between the observed and model-simulated time series for each of the four observed model response values (H, λE, T_g, and S_w) was computed, and these were selected as the four criteria to be minimized.

4.2.2. Control run using BATS default parameters. As mentioned earlier, the standard method for parameter specification in BATS automatically assigns values for the 24 parameters based on user specifications of land cover, soil texture, and soil color. The values for these so-called default parameters for the ARM-CART grassland site are listed in Table 1. To provide a basis for comparison, the default BATS parameters for grassland sites were used to generate a control run simulation (see Figure 2). To minimize the effects of poor specification of initial values, the initial water contents in the three soil layers were estimated by optimization following *Bastidas* [1998] and a one month spin-up period was used.

In Figure 2, each row of plots corresponds to one of the four observed model responses. The four scatterplots on the right-hand side depict the correspondence between the entire set of observed and the simulated values. The time series plots on the left show the response dynamics for a representative 10 day period during which precipitation occurred (DOY 120–129, 1995) with the observations indicated by dots and the model simulations indicated by solid lines. Although the observed and simulated output energy fluxes H and λE for the 10 day period appear to agree quite well, the scatterplots for the entire data set show that many points are quite far from the 1:1 line. This scatter can arise due to the combined effects of model and observational errors, but the scatterplots show a tendency toward underestimation of H (bias = 13.01 W/m^2) and overes-

timation of λE (bias = -7.57 W/m^2), suggesting that there is a problem with the partitioning of energy within the model. In contrast, correspondence of the observed and simulated state variables T_g and S_w for the 10 day period is poor. In particular, the diurnal variation in simulated ground temperature tends to peak earlier and be larger than the observed diurnal variation, and the simulated soil moisture does not track the observed changes associated with precipitation events on days 122, 125, 126, and 127. The peculiar shape of the scatterplot for S_w suggests problems with the partitioning of moisture within the model; we later show that this can be improved by adjusting the parameters. However, we should not ignore the possibility of some incompatibility between the S_w that is measured and that which is simulated by the model.

The discrepancies between the observations and BATS model default parameter simulations are a consequence of the combined influence of data errors, structural inadequacies in the model, and parameter misspecification. Parameter misspecification can arise because the default estimates are only loosely constrained by field observations of vegetation and soil type. If discrepancies can be reduced by constraining the parameters by calibration against data, it might then be possible to determine whether the energy and moisture partitioning problems are associated with structural inadequacies in the model.

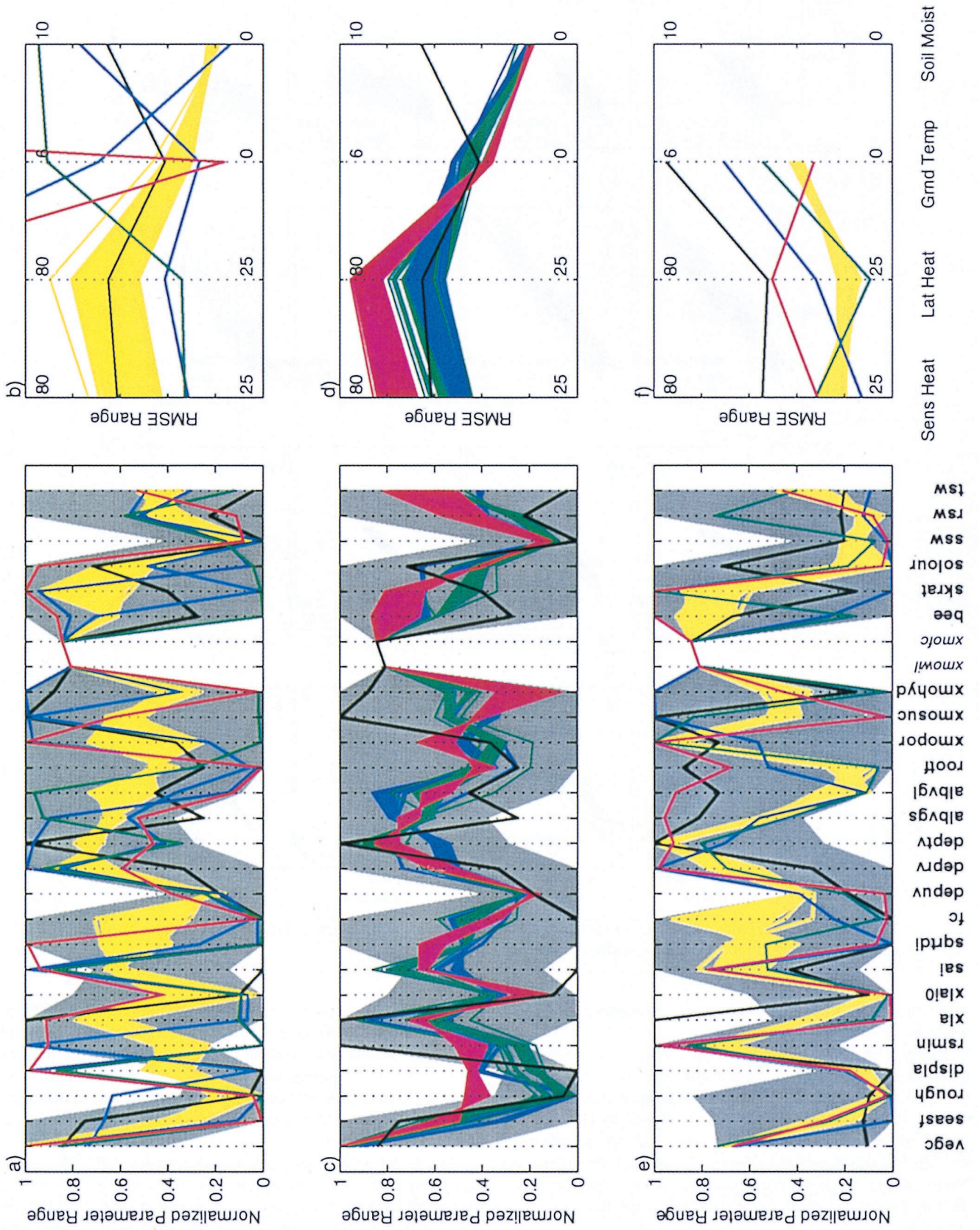
4.2.3. Parameter estimation by model calibration. To explore the relative value of different observational data sets for constraining the model parameters, several different calibration studies were performed. Each of the observed system responses was first used individually to calibrate the model as if the data on other model responses were not available; these are called single-criterion calibrations. Then the system responses were used in pairs, triples, and all at once; these are here referred to as two-criteria, three-criteria, and four-criteria calibrations, or collectively as multicriteria calibrations. For simplicity of notation, a closed pair of brackets will denote a calibration run; for example, $\{H\}$ denotes a single-criterion calibration using H as the calibration criterion, and $\{H, \lambda E, T_g\}$ denotes a multicriteria calibration using H, λE, and T_g as the calibration criteria. The SCE-UA global optimization method [*Duan et al*., 1993] and the MOCOM-UA global optimization method [*Yapo et al*., 1997] were used for the single-criterion and multicriteria calibrations, respectively.

To allow a relatively rigorous test of the power of multicriteria calibration, it was assumed that there was no information available for the study site. In addition, the following two conditions were imposed on each of the calibration runs. First, all 22 independently specifiable model parameters were estimated by calibration, despite the fact that several could have been specified using field observations. Thus after including the three initial soil-moisture contents, a total of 25 unknown variables were optimized. Second, the initial parameter uncertainty was specified in terms of minimum and maximum values that are believed to correspond approximately to the full range of possible values for any location on Earth (see Table 1). Table 1 also lists more restrictive upper and lower bounds on each parameter which specify the reasonable range of values for the ARM-CART grassland site. These reasonable ranges were not used to constrain the calibrations but were instead used as postcalibration checks on the reliability and credibility of the calibration procedure.

However, it is important to point out that proper feasibility constraints were imposed on the allowable values of the parameters to preserve the physical realism of the parameterization. For example, the thicknesses of the shallower soil layers are constrained to be less than the deeper soil layers, the initial water contents of each soil layer were constrained to be less than the depth of the corresponding soil layer, and the season variability of vegetation cover and leaf area index were constrained to be smaller than their prescribed maximum values.

4.2.3.1. Single-criterion calibrations: The results of the single-criterion calibrations are summarized in Figures 3, 4, and Plates 1a and 1b. Figure 3 presents the scatterplots showing the correspondence between the observed and the simulated values for each of the four single-criterion calibrations, and Figure 4 shows a detailed depiction of the time series behavior (for the chosen 10-day period). In both figures, each row corresponds to a different model response, and each column (from left to right) shows the results of the $\{H\}$, $\{\lambda E\}$, $\{T_g\}$, and $\{S_w\}$ calibrations, respectively. Plate 1a is a normalized parameter plot (as introduced by *Sorooshian et al*. [1993]) of the different parameter estimates; the 24 BATS parameters and the 3 initial soil-moisture contents are listed along the x axis (xmowil and xmofc are included for completeness), and the y axis corresponds to the parameter values plotted on a range normalized by the global maximum and minimum values for the parameters. Each line going from left to right across the plot corresponds to a different calibration result (blue = $\{H\}$, green = $\{\lambda E\}$, red = $\{T_g\}$, and cyan = $\{S_w\}$); the default parameter control is also included using the color black. The grey shadowed area corresponds to the reasonable range for each parameter value for the ARM-CART grassland site (see Table 1). Plate 1b shows the trade-offs between the different criteria in the multicriteria space; the four response criteria are listed along the x axis, and the y axis corresponds to the criterion values (RMSEs) with better values toward the bottom of the plot. Each line going from left to right across the plot corresponds to a calibration result (and hence a different parameter value). Note that if a calibration result plots as a line that falls entirely below (or above) that of a different result, the former can be said to be absolutely superior (inferior) in a multicriteria sense. However, if the two

Plate 1. (opposite) Comparison of parameter estimates and final criteria values for the different calibration runs: (a) normalized parameter estimates for the ARM-CART site (black = default, blue = $\{H\}$ calibration, green = $\{\lambda E\}$ calibration, red = $\{T_g\}$ calibration, cyan = $\{S_w\}$ calibration, yellow = $\{H, T_g, S_w\}$ calibration, light grey = reasonable ranges); (b) criterion values for the ARM-CART site (color code same as Plate 1a; (c) normalized parameter estimates for the ARM-CART site (black = default, blue = $\{H, S_w\}$ calibration, green = $\{\lambda E, S_w\}$ calibration, red = $\{T_g, S_w\}$ calibration, light grey = reasonable ranges); (d) criterion values for the ARM-CART site (color code same as Plate 1c); (e) normalized parameter estimates for the Tucson site (black = default, blue = $\{H\}$ calibration, green = $\{\lambda E\}$ calibration, red = $\{T_g\}$ calibration, yellow = $\{H, \lambda E, T_g\}$ calibration, light grey = reasonable ranges); (f) criterion values for the Tucson site (color code same as Plate 1b).


a)
b)
c)
d)
e)
f)
Normalized Parameter Range
RMSE Range
Sens Heat
Lat Heat
Grnd Temp
Soil Moist
vegc
seasf
rough
displa
rsmin
xla
xla0
sai
sqrtdi
fc
depuv
deprv
deptv
albvgs
albvgl
rootf
xmopor
xmosuc
xmohyd
xmowl
xmofc
bee
skrat
solour
ssw
rsw
tsw

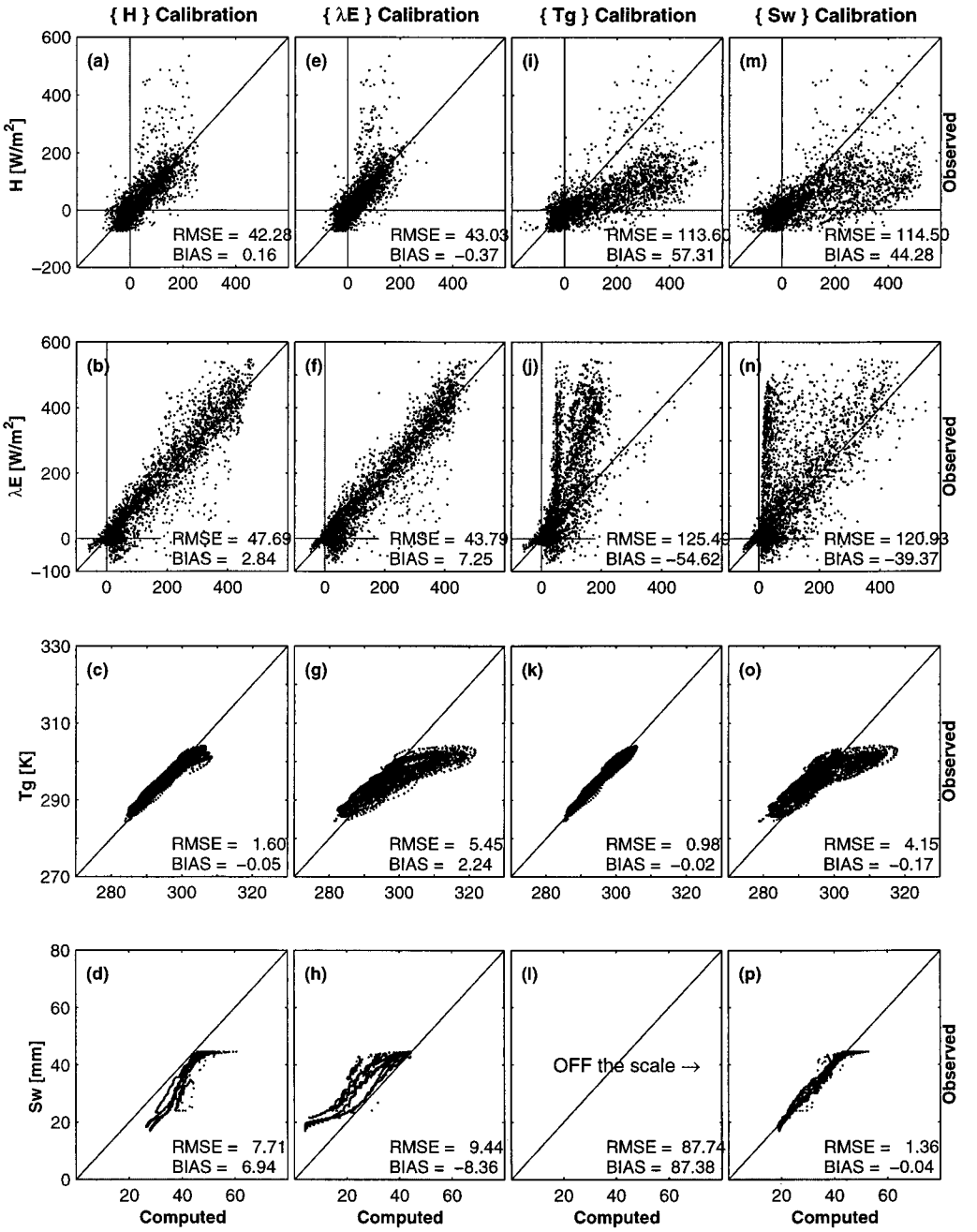


Figure 3. Scatterplots comparing modeled variables (solid line) and observed data (dots) for the parameter sets estimated obtained by different single-criterion calibration runs.

lines cross each other then the results are noninferior to each other in a multicriteria sense.

The results of the single-criterion calibrations can be summarized as follows: Each calibration tends to perform much better than the default run on its corresponding model response, giving smaller RMSE values as well as closer correspondence to the scatterplot 1:1 line while performing less well on the other model responses (Figures 3 and Plate 1b). None of the single-criterion calibrations is superior to the default parameter estimates (black) in a multicriteria sense (Plate 1b).

Only the results of the $\{H\}$ calibration can be considered as reasonable in comparison to the default parameter estimates. There is approximately a 25% RMSE improvement in matching the H, λE, and T_g responses and only a slight RMSE deterioration in matching S_w (20%) (see Plate 1b). There is also a significant improvement in matching the diurnal pattern of variation of T_g (compare Figures 4a and 2) and a qualitative improvement in the response of S_w to precipitation, although the actual S_w level is too high (bias = 6.94) (compare Figures 3d and 4a to Figure 2).

The $\{\lambda E\}$ calibration (green line) gives good RMSE results for H and λE but shows significant deterioration (70–100%) from the defaults for T_g and S_w. There is a significant deterioration in matching the amplitude of variation of T_g; which is also reflected in the shape of the scatterplot (compare Figures 3g and 4b to Figure 2). As with the $\{H\}$ calibration, there is a qualitative improvement in the response of S_w to precipitation,

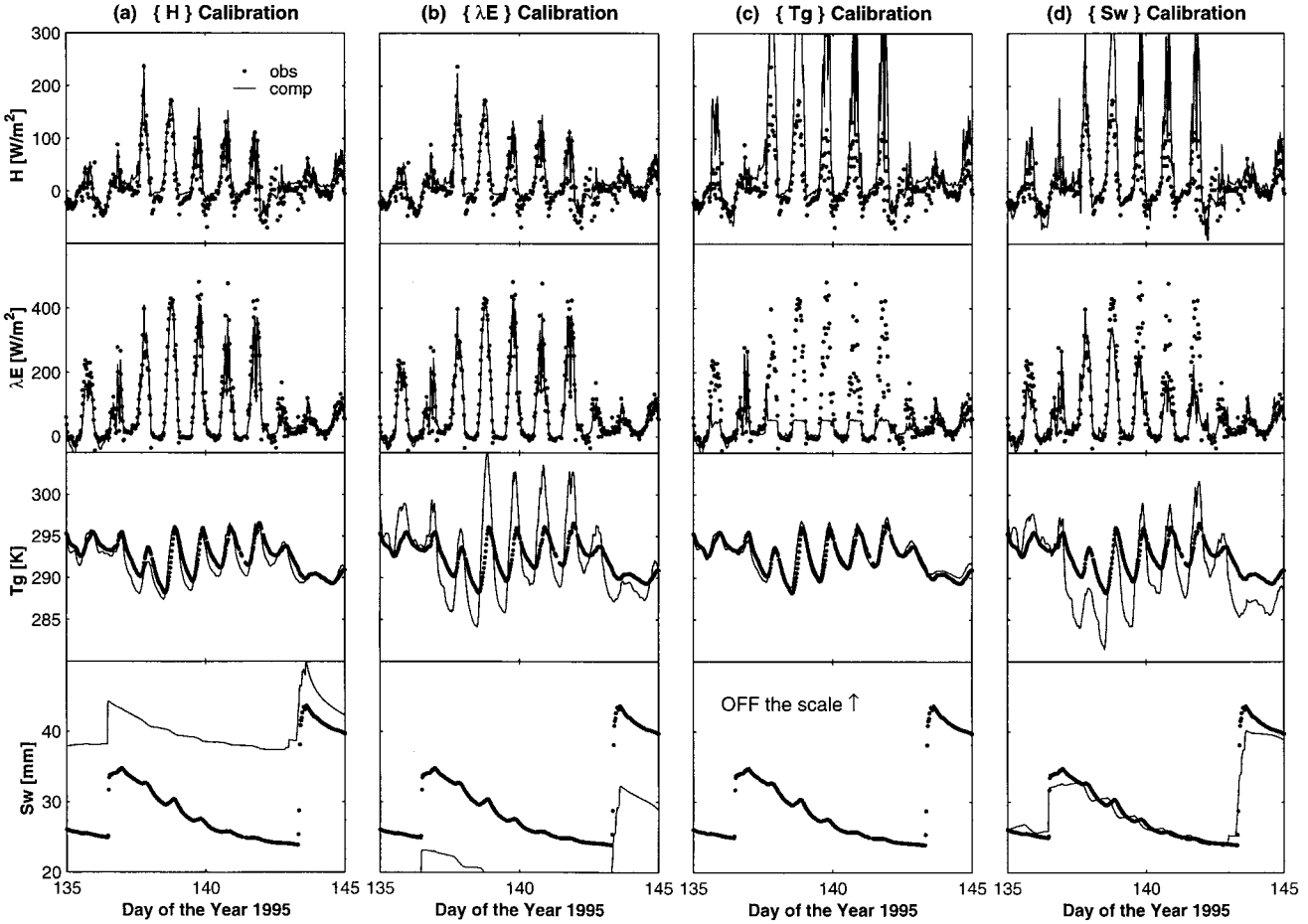


Figure 4. Comparison of modeled variables (solid line) with the observed data (dots) for the parameter sets estimated by different ARM-CART site single-criterion calibration runs; the time series plots show a representative 10 day period during which rainfall occurred.

although the actual S_w level is now too low (bias = −8.36) (Figures 3d and 4b).

The $\{T_g\}$ calibration gives good RMSE values for T_g but gives such poor RMSEs for the other responses that they are above the upper ends of the scales in Plate 1b. The visual match to T_g is excellent, but there is clearly a problem with the energy balance; the daily peak in H is too high and λE is clipped at a low value (Figures 3i–3k and 4c). In addition, the simulated S_w is much too high, being completely off the scales of the scatterplot and time series plot (Figures 31 and 4c).

The $\{S_w\}$ calibration gives good RMSE values for S_w but poor RMSEs for the other responses (Plate 1b). The visual match to the amplitude and pattern of variation of S_w is excellent (Figures 3p and 4d), but the energy balance is clearly incorrect. Note also the significant increase in scatter for H and λE (Figures 3m–3n).

In general, the different calibrations tend to provide somewhat different parameter values for most of the parameters (Plate 1a). However, some of the parameters tend to fall at opposite ends of their feasible ranges when calibrated using different response data. In particular, the aerodynamic parameters rough (roughness length) and displa (displacement height) tend to have opposite behavior when using $\{H\}$, $\{\lambda E\}$, and $\{T_g\}$ than when using $\{S_w\}$. However, the calibrated estimates for vegc (vegetation cover), sai (stem area index), fc (light dependence of stomatal resistance), rootf (ratio of roots in upper layer to lower layer), and ssw (surface zone water content) tend to be similar and relatively insensitive to the response used for calibration. The different calibrations tend to give parameters that fall within the reasonable value bounds (grey shadowed area) with the exception of rough (roughness length), displa (displacement length), rsmin (minimum stomatal resistance), and sai, suggesting that the reasonable bounds do not provide sufficiently tight constraints to guarantee good model performance.

4.2.3.2. Multicriteria calibrations: The results of the several multicriteria calibrations are presented in Plates 1a–1d and Figure 5. The results for the $\{H, S_w\}$, $\{\lambda E, S_w\}$, and $\{T_g, S_w\}$ two-criteria calibrations (blue = $\{H, S_w\}$, green = $\{\lambda E, S_w\}$, and red = $\{T_g, S_w\}$) and the $\{H, Tg, S_w\}$ calibration are shown (yellow) because these gave the best results. (Note that the four-criteria $\{H, \lambda E, T_g, S_w\}$ calibration failed to improve on the three-criteria calibration results). Remember that for multicriteria calibration the results consist of a Pareto set of solutions, reflecting a range of trade-offs in matching the associated criteria; no single member of the set is better than any other member in a multicriteria sense. Therefore instead of picking a single representative member for analysis of the results, the results for the entire set of Pareto solutions are shown.

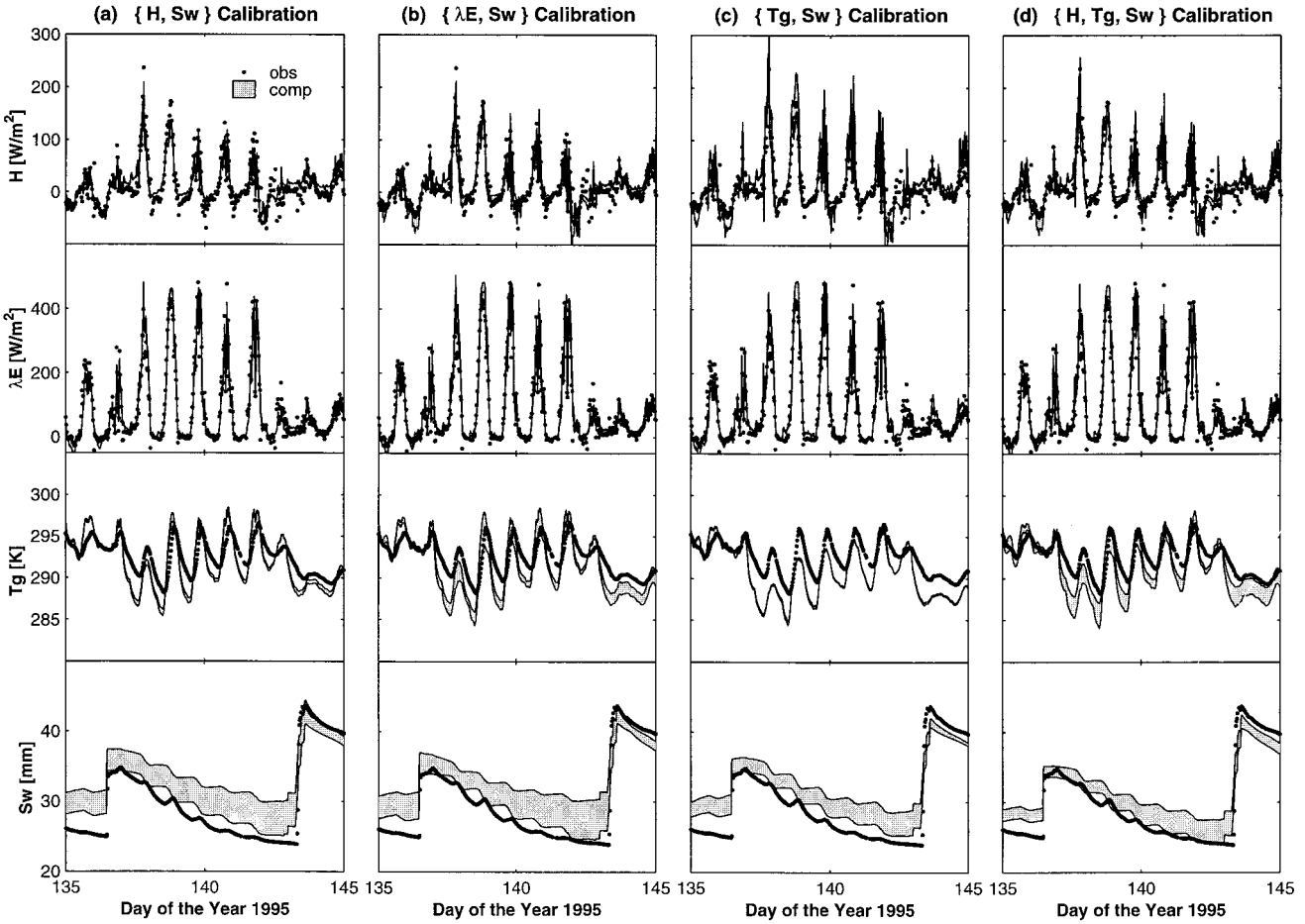


Figure 5. Comparison of modeled variables (light grey shaded region) with the observed data (dots) for the parameter sets estimated by different ARM-CART site multicriteria calibration runs. The time series plots show a representative 10 day period during which rainfall occurred.

In general, the results of the multicriteria calibrations can be summarized as follows: All of these multicriteria calibrations are generally quite good and are comparable to the default in a multicriteria sense (Plates 1b and 1d). Note that the tendency to give poor matching of the fluxes not included for calibration has been diminished if not removed (compare with Plate 1b). All the multicriteria calibrations show marked RMSE improvement in S_w compared to the default (black), which is not surprising, given that the S_w data are being used to constrain the model (Plates 1b and 1d). The $\{T_g, S_w\}$ calibration (red) tends to match the T_g and S_w fluxes well but at the expense of slightly poorer matching of the heat fluxes H and λE (Plate 1d).

The time series plots for all of the multicriteria calibrations appear to give reasonably good matching of the diurnal pattern of variability of all four observed system responses (Figures 5a–5d). However, there is a noticeable trade-off in the ability of the model to simultaneously match both T_g and S_w well. This can be seen even in the $\{T_g, S_w\}$ calibration (Figure 5c), which is not constrained to match H or λE and shows a tendency for the entire set of Pareto solutions (shaded area) to underestimate T_g while overestimating S_w. In general, however, the biases are small.

The degree of variability in the parameter values (Plates 1a and 1c), both within the Pareto spread of the multicriteria calibrations and from one multicriteria calibration to another, is reduced in comparison to the single-criterion calibrations (Plate 1a). Notice also that the parameter values tend to lie within the reasonable range (grey-shaded area); the most serious violations are the $\{T_g, S_w\}$ estimates (Plate 1c, red) for the parameters rough (roughness length) and displa (displacement height), which lie only slightly outside the range (and are much improved from the single-criterion calibrations). The three-criteria $\{H, Tg, S_w\}$ calibration provides the most consistency with the reasonable parameter ranges (Plate 1a, yellow).

The $\{H, Tg, S_w\}$ calibration gives the best overall results and is superior to the default in the multicriteria sense; the RMSEs are improved by 10.4% for H, 1.2% for λE, 12.6% for T_g, and 70.3% for S_w. The next best results are given by the $\{H, S_w\}$ calibration. Then come the $\{\lambda E, S_w\}$ and $\{T_g, S_w\}$ calibrations, which are not so good but are superior to the single-criterion calibrations.

4.3. Parameter Estimation for Tucson Semiarid Study Site

4.3.1. Data set. The Tucson hydrometeorological site is located in the semiarid Sonoran Desert in the southwestern United States. The data were collected as part of ongoing research by the Department of Hydrology and Water Resources of the University of Arizona into the hydrometeorology of semiarid environments [*Unland et al.*, 1996]. The data used here were collected at 20 min time intervals and covers an

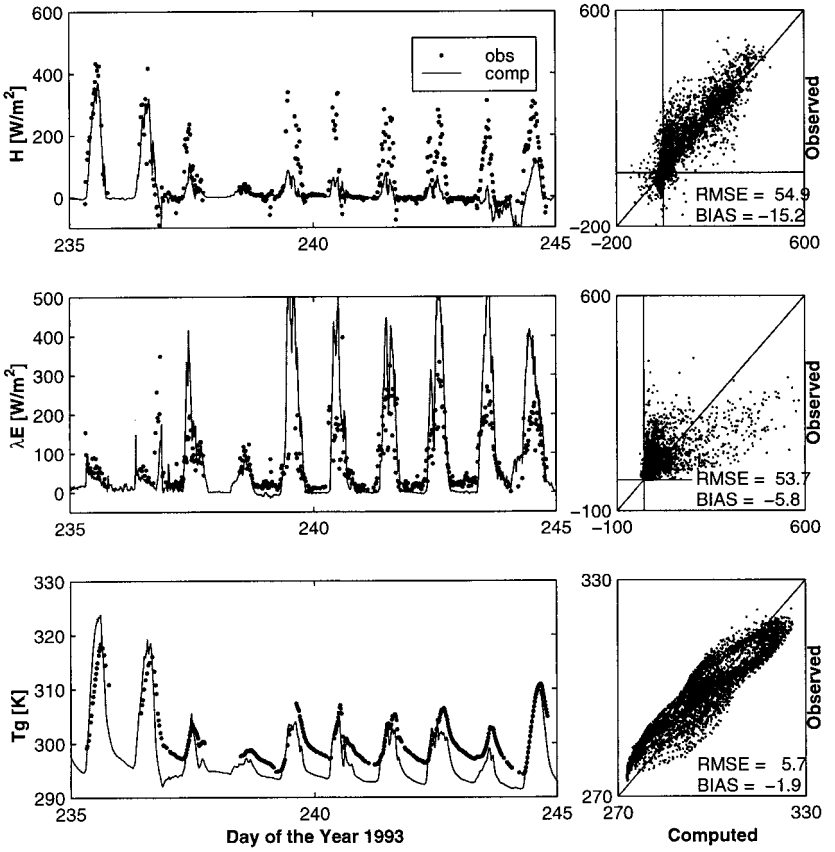


Figure 6. Comparison of modeled variables (solid line) with the observed data (dots) for the Tucson site default parameters. The time series plots on the left show a representative 10 day period during which rainfall occurred. The scatterplots on the right show the entire period of data.

entire year from May 1993 to April 1994. The measured atmospheric forcings are net radiation, incoming radiation, air temperature, precipitation, specific humidity, and wind speed. Because the atmospheric pressure was not recorded, a constant value of 91.29 kPa was used on the basis of the average air pressure measured at the nearby Tucson International Airport and corrected for the elevation. The observed system response variables are sensible heat (H), latent heat (λE), and ground temperature (T_g). The heat fluxes were measured by eddy correlation from July 19 to August 9, 1993, and by a Bowen ratio system from August 10, 1993 to March 26, 1994. *Unland et al.*, [1996] reported a problem with the reliability of the Bowen ratio system when the observed values of latent heat are small. These flagged values are considered to be of dubious quality and were therefore not used in this research; the remaining system response data contain measurements at 5219 time steps. In addition, an initial 2 month period of measurements for the atmospheric variables (but not the observed fluxes) was available; this period was used as a model "spin-up" period to diminish the influence of the incorrect specification of initial state variable values.

4.3.2. Multicriteria parameter estimation. A series of single-criterion and multicriteria calibration runs were conducted for the Tucson site following a similar procedure as for the ARM-CART site, and the major results are presented in Plates 1e–1f, and Figures 6 and 7. The default parameter performance of the BATS model is shown in Plate 1f; the default parameter values are plotted in Plate 1e (black). Notice that the default value for parameter xla (maximum leaf area index) is well above the reasonable range for the Sonoran Desert semiarid environment but that the other parameters are within the reasonable ranges. However, the time series plots and scatterplots (Figure 6) show that the partitioning of energy is incorrect and that the model is not matching the ground temperature well, especially after rainfall occurs on days 237 and 238. Plate 1f shows that all the single-criterion calibrations provide RMSEs ($\{H\}$ = blue, $\{\lambda E\}$ = green, $\{T_g\}$ = red) that are better than the default in a multicriteria sense. These calibrations also provide parameter estimates that are entirely within the reasonable ranges (grey-shaded area). As expected, fitting on a single criterion tends to emphasize matching of the associated response at the expense of the others, and there is a wide variation in the parameter estimates (Plate 1e); for some parameters the single-criterion calibrations give estimates on opposite sides of the reasonable ranges (e.g., albvgl (long-wave vegetation albedo), rootf (upper to lower layer root ratio), xmosuc (minimum soil suction), and skrat (soil conductivity ratio)). Plate 1f also shows that the three-criteria $\{H, \lambda E, T_g\}$ calibration results (yellow) are much superior to the default (the RMSEs are improved by 32% for H, 39% for λE, and 59% for T_g), and the set of Pareto solutions arguably gives better trade-offs in matching to the three response variables than the individual single-criterion calibrations. In addition, the parameter estimates (Plate 1e) are closely grouped, fall within the reasonable ranges for the Sonoran Desert semiarid environment, and appear to be very

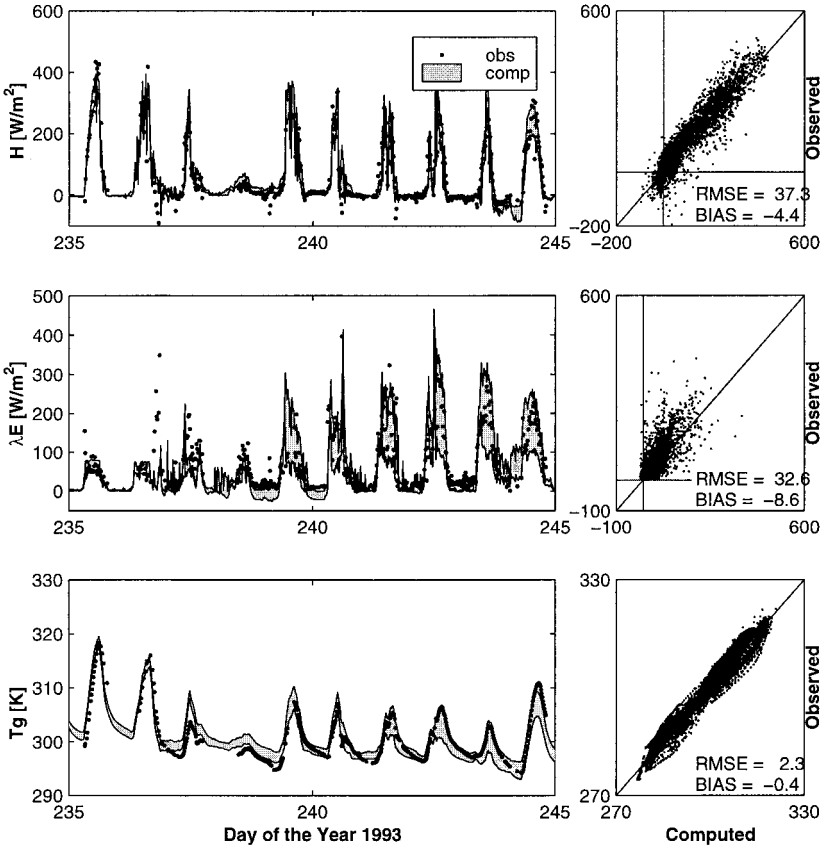


Figure 7. Comparison of modeled variables (lightly shaded region) with the observed data (dots) for the parameter sets estimated by the Tucson site $\{H, \lambda E, T_g\}$ multicriteria calibration run. The time series plots show a representative 10 day period during which rainfall occurred. The scatterplots on the right compare the entire period of data compared with the midpoints of the modeled ranges.

realistic for the Tucson study site. The time series plots and scatterplots (Figure 7) show that the partitioning of energy has been improved (the considerable scatter remaining in the latent heat flux is attributable to measurement difficulties), and the model is now matching the ground temperature well. This three-criteria calibration run required approximately 20,000 iterations (model simulations) to converge to the solution.

5. Summary and Conclusions

Over the last decade, attempts to create models of surface-atmosphere interactions with greater physical realism has resulted in a suite of land surface schemes that use many parameters (e.g., BATS). The hope has been that these parameters can be assigned typical values by inspecting the literature. Often, vegetation and soil types are grouped into broad classes to aid in assigning the many parameters. However, observational data are now available which allow testing of such models at the patch scale. Thus the opportunity to contribute to the available literature on appropriate land surface parameters by optimizing models against time series of measured variables now exists. However, the derivation of the required parameters from such data has before now proven problematic because of the high parametric content of the models. The research presented in this paper is motivated by the need to contribute to the availability of the required land surface parameters by finding an effective means for extracting values for the plot-scale parameters from typical field data. Specifically, the possibility of whether extensions into a multiobjective form of the advanced parameter estimation methods developed for use with hydrological models can be used for practical and efficient estimation of the parameters in a complex LSS is explored.

Both traditional single-criterion estimation techniques and the recently developed multicriteria methods were applied to the above problem using two data sets considered typical in terms of data quality but very different in terms of location and climatological regime (ARM-CART and Tucson). To the extent possible, the results were interpreted with a view to providing more general insight. In general, it was found that the traditional single-criterion methods are of limited value for estimating the preferred range of parameter estimates for a highly parameterized land surface scheme via calibration to field observations. However, the multicriteria approach was found to be effective at constraining the parameter estimates into physically plausible ranges when (and this is an important point) observations on certain combinations of outputs and state variables are available.

For the two case studies presented in this paper, successful calibration was achieved using observations of at least one heat flux (found here to be sensible heat) and one state variable. For the ARM-CART grassland site, where evaporation is the largest flux, soil moisture was the preferred state variable, and inclusion of soil temperature gave limited additional advan-

tage. For the Tucson semiarid site, where sensible heat is the dominant flux, soil temperature was effective as the observed state variable (however, soil-moisture values were not available for comparison). These results are suggestive that the preferred state variable to be measured should be related to the primary heat flux at a particular site. This remains a topic for future research. Nonetheless, techniques for establishing which parameters are sensitive to the available variables are certainly required, which is addressed in the companion paper [*Bastidas et al*., this issue].

Regardless, the central conclusions of this paper are clear. A multicriteria parameter optimization method that is efficient and simple to use with modern land surface schemes has been developed, and when used with at least two criteria (e.g., the sensible heat flux and an appropriate state variable), the new technique is effective in constraining the preferred range for the many land-surface-related parameters that must be defined in a typical LSS. The MOCOM algorithm has been written to be generally applicable to any model, and the code is available from the authors on request (e-mail: *hoshin@hwr.arizona.edu*).

Acknowledgments. 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 support provided to L. Bastidas by the Escuela Politecnica Nacional (Quito—Ecuador) and the Fulbright-Laspau Program is gratefully acknowledged. Special thanks are due to Helene Unland and Jim Washburne for providing the data, to Jim Broermann and Dan Braithwaite for providing assistance with computational resources, and to Corrie Thies for proofreading.

References

Arain, A. M., J. D. Michaud, W. J. Shuttleworth, and A. J. Dolman, Testing of vegetation parameter aggregation rules applicable to the Biosphere-Atmosphere Transfer Scheme (BATS) at the FIFE site, *J. Hydrol*., *177*, 1–22, 1996.

Bastidas, L. A., Parameter estimation for hydrometeorological models using multicriteria methods, Ph.D. dissertation, Dep. of Hydrol. and Water Resour., Univ. of Arizona, Tucson, 1998.

Bastidas, L. A., H. V. Gupta, S. Sorooshian, W. J. Shuttleworth, and Z. L. Yang, Sensitivity analysis of a land surface scheme using multicriteria methods, *J. Geophys. Res*., this issue.

Brewer, K. E., and S. W. Wheatcraft, Including multi-scale information in the characterization of hydraulic conductivity distributions, in *Wavelets in Geophysics*, edited by E. Foufoula-Georgiou and P. Kumar, pp. 213–248, Academic, San Diego, Calif., 1994.

Dickinson, R. E., A. Henderson-Sellers, and P. J. Kennedy, Biosphere Atmosphere Transfer Scheme (BATS) version 1e as coupled to the NCAR Community Climate Model, *NCAR Tech. Note*, *NCAR/TN−387+STR*, 72 pp., Natl. Cent. for Atmos. Res., Boulder, Colo., 1993.

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

Duan, Q., V. K. Gupta, and S. Sorooshian, A shuffled complex evolution approach for effective and efficient global minimization, *J. Optimization Theor. Appl*., *76*(3), 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.

Franks, S. W., and K. J. Beven, Bayesian estimation of uncertainty in land-surface-atmosphere flux predictions, *J. Geophys. Res*., *102*, 23,991–23,999, 1997.

Franks, S. W., P. Gineste, K. J. Beven, and P. Merot, On constraining the predictions of a distributed model: The incorporation of fuzzy estimates of saturated areas into the calibration process, *Water Resour. Res*., *34*(4), 787–797, 1998.

Gao, X., S. Sorooshian, and H. V. Gupta, Sensitivity analysis of the biosphere-atmosphere transfer scheme, *J. Geophys. Res*., *101*, 7279–7289, 1996.

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, H. V., S. Sorooshian, and P. O. Yapo, Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resour. Res*., *34*(4), 751–763, 1998.

Harboe, R., Multiobjective decision making techniques for reservoir operation, *Water Resour. Bull*., *28*(1), 103–110, 1992.

Houser, P., Remote-sensing soil moisture using 4DDA, Ph.D. dissertation, Dep. of Hydrol. and Water Resour., Univ. of Arizona, Tucson, 1996.

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. of Agric. Eng., St. Joseph, Mich., 1982.

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

Lettenmaier, D., D. Lohmann, E. F. Wood, and X. Liang, *PILPS-2c Workshop Report*, Princeton Univ., Princeton, N. J., October 28–31, 1996.

Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges, A simple hydrologically based model of land surface water and energy fluxes for general circulation models, *J. Geophys. Res*., *99*, 14,415–14,428, 1994.

Manabe, S., Climate and the ocean circulation, I, The atmospheric circulation and the hydrology of the Earth's surface, *Mon. Weather Rev*., *97*, 739–774, 1969.

Sellers, P. J., W. J. Shutleworth, J. L. Dorman, A. Dalcher, and J. M. Roberts, Calibrating the simple biosphere model (SiB) for amazonian tropical forest using field and remote sensing data, 1, Average calibration with field data, *J. Appl. Meteorol*., *28*, 728–756, 1989.

Sellers, P. J., D. A. Randall, G. J. Collatz, J. A. Berry, C. B. Field, D. A. Dazlich, C. Zhang, G. D. Collelo, and L. Bounoua, A revised land surface parameterization (SiB2) for atmospheric GCMs, I, Model formulation, *J. Clim*., *9*, 676–705, 1996.

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., 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*(4), 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.

Shuttleworth, W. J., Z. L. Yang, and M. A. Arain, Aggregation rules for surface parameters in global models, *Hydrol. Earth Syst. Sci*., *2*, 217–226, 1997.

Unland, H., P. Houser, W. J. Shuttleworth, and Z.-L. Yang, Surface flux measurement and modeling at a semi-arid Sonoran Desert site, *Agric. For. Meteorol*., *82*, 119–153, 1996.

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*., *204*, 83–97, 1997.

L. A. Bastidas, H. V. Gupta, W. J. Shuttleworth, S. Sorooshian, and Z. L. Yang, Department of Hydrology and Water Resources, University of Arizona, Tucson, AZ 85721. (hoshin@hwr.arizona.edu)

(Received September 2, 1998; revised February 8, 1999; accepted February 22, 1999.)