Sensitivity analysis of a land surface scheme using multicriteria methods

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

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

Abstract. Attempts to model surface-atmosphere interactions with greater physical realism have resulted in complex land surface schemes (LSS) with large numbers of parameters. A companion paper describes a multicriteria calibration procedure for extracting plot-scale estimates of the preferred ranges of these parameters from the various observational data sets that are now available. A complementary procedure is presented in this paper that provides an objective determination of the multicriteria sensitivity of the modeled variables to the parameters, thereby allowing the number of calibration parameters and hence the computational effort to be reduced. Two case studies are reported for the BATS model using data sets of typical quality but very different location and climatological regime (ARM-CART and Tucson). The sensitivity results were found to be consistent with the physical properties of the different environments, thereby supporting the reasonableness of the model formulation. Further, when the insensitive parameters are omitted from the calibration process, there is little degradation in the quality of the model description and little change in the preferred range of the remaining parameters.

1. Introduction and Scope

This paper is one of three that discuss the usefulness of multicriteria methods for the evaluation and improvement of land surface schemes (LSS). A companion paper [*Gupta et al.*, this issue] shows how multicriteria methods can be used to improve the estimates of LSS parameters by simultaneously constraining the model to measurements of several observed system responses such as heat fluxes, ground temperature, and surface soil moisture. This paper develops that study by introducing a robust multicriteria approach to parameter sensitivity analysis for LSS models and by showing how the methodology provides 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 the ARM-CART grassland site and the other from a semiarid site in the Sonoran Desert, Arizona. A third paper (in preparation) will discuss the power and applicability 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 methods to the generalized sensitivity analysis of LSS model parameters. In Section 4, the sensitive parameters of the BATS model for two study sites are identified, and it is shown that only the reduced set of sensitive parameters need to be calibrated to obtain good model performance. Section 5 discusses the results and future extensions.

2. Background

LSS models differ from classical hydrologic watershed models in that they are concerned with both water and energy balance, they are driven by multiple input variables (e.g., precipitation, shortwave and longwave radiation, wind speed, air temperature, and humidity), and they predict the evolution of several observable state variables (e.g., soil skin temperature, surface soil moisture) and output fluxes (e.g., latent heat, sensible heat, runoff). Such models often have a large number of parameters that must be specified. The Biosphere-Atmosphere Transfer Scheme (BATS) [*Dickinson et al.*, 1993], for example, has 27 parameters to be estimated, including 16 related to vegetation properties and eight related to soil properties, together with three initial moisture conditions. While the parameterizations of LSS models have been designed with the notion that it should be possible to estimate reasonable values for the parameter from measurable characteristics of the land surface, recent studies such as the PILPS 2c workshop [*Lettenmaier et al.*, 1996] have demonstrated that even simple manual (subjective) adjustment of a few model parameters can result in significant improvements in model performance. However, as long as the performance of an LSS model remains dependent on subjective human expertise for the specification of its parameters, it will remain difficult to conduct an objective evaluation of the relative merits of alternative parameterizations. It is useful therefore to explore the utility (objectivity and efficiency) of automated methods [*Sorooshian et al.*, 1993; *Gupta et al.*, 1998] for improving the parameter estimates used by LSS models; in a companion paper [*Gupta et al.*, this issue], we demonstrate how this can be done by applying multicriteria calibration methods.

As LSS models have become more complex, the number of parameters that must be estimated has significantly increased. While the typical conceptual watershed hydrology model may have only 10–15 parameters, the BATS 1e model has 27 parameters, and the current SiB2 model has 52 parameters that

must be specified. Many of these are "conceptual" parameters that are not directly measurable at the GCM-LSS scales and for which all that might reasonably be specified are the approximate ranges for the parameter values, based on some approximate understanding of the regional hydrogeology. If the input-state-output response of the LSS is not essentially sensitive to parameter variations within these ranges, it will be reasonable to use some nominal estimates (such as the midpoints of the parameter uncertainty ranges). However, if the LSS input-state-output response is sensitive to finer specification of these parameters, the only remaining recourse is to adjust the parameter so that model responses are constrained to closely match available observations via calibration [*Gupta et al.*, this issue]. It is important therefore to identify which of the parameters within the model are sensitive (for each hydroclimatic regime) and should be refined further via calibration.

A number of different methods have been proposed in the literature for the sensitivity analysis of LSS schemes. These include the most commonly used "one parameter at a time" approach [e.g., *Wilson et al.*, 1987a, b; *Pitman*, 1994; *Gao et al.*, 1996], the factorial method [e.g., *Henderson-Sellers*, 1992; *Lettenmaier et al.*, 1996], the Fourier amplitude sensitivity test (FAST) [*Collins and Avissar*, 1994], and the regionalized sensitivity analysis (RSA) methodology [*Franks et al.*, 1997]. Most studies assumed parameter independence and explored the individual impacts of each parameter on each system response taken one-at-a-time. None of the methods accounts for the joint multiparameter and multiresponse interactions. Although the factorial method allows (in principle) the exploration of multiparameter interactions, a simple two-level factorial experiment for a 25 parameter model would require an unreasonable number of model runs ($2^{25} = 33,554,432$). While this number can be reduced by a fractional factorial experiment that only considers second- or third-order interdependence, the effects of large numbers of possible interactions would then be ignored. The FAST method would only require about 6000 model runs for the same problem (extrapolating from *Cukier et al.* [1978]), but it again ignores interdependence among the parameters or requires an a priori estimate of the parameter covariance structure.

Another common feature of past studies is that they were directed not only at finding the influence of the model parameters but also at how this influence varies under different climatic forcing conditions. The studies were therefore conducted using artificially constructed atmospheric forcing data for different environments [e.g., *Henderson-Sellers*, 1992; *Gao et al.*, 1996]. Further, the impacts were typically evaluated in terms of integrated response statistics such as average or cumulative values over monthly or yearly time spans (e.g., total evapotranspiration, total runoff, and minimum upper layer soil temperature for a year [*Henderson-Sellers*, 1992]; mean of the annual heat fluxes, ground temperature, and soil moisture [*Pitman*, 1994]; annual cumulative monthly runoff, annual cumulative monthly accumulated absolute differences between soil moisture and porosity for perturbed and nominal parameters [*Lettenmaier et al.*, 1996]; and annual average variation of model responses, [*Gao et al.*, 1996]). To our knowledge, only *Franks et al.* [1997] used real system response data for their sensitivity analysis. However, they studied a time span of only 10 days and only considered the latent heat flux.

In general, these studies have reported that the latent and sensible heat fluxes are sensitive to the same parameters. One parameter that seems to be universally important for different LSS formulations is that related to the stomatal resistance of the vegetation. However, the results of different studies using BATS are inconsistent; for example, *Lettenmaier et al.* [1996] report high sensitivities to the Clapp and Hornberger parameter B, while *Gao et al.* [1996] and *Henderson-Sellers* [1992] did not. An important finding by *Gao et al.* [1996] is that the influence of errors in the initial soil layer water conditions may persist for several years beyond the end of the spin-up period.

The work presented here builds on the multicriteria framework presented by *Gupta et al.* [1998] and *Gupta et al.* [this issue]. The major aim of this paper is to demonstrate a robust, practical, and efficient approach for the parameter sensitivity analysis of LSS models, in a manner that properly accounts for the joint multiparameter/multiresponse interactions, and to illustrate the consequent benefits that can accrue. We show that the approach provides a parameter sensitivity classification which is consistent with physical understanding. We also show how this classification can be used effectively to reduce the dimensionality of the parameter estimation problem, thereby significantly reducing the computer time required to conduct a model calibration.

3. Multicriteria Sensitivity Analysis

3.1. Problem Formulation

A theoretical basis for the application of multicriteria theory to the calibration of conceptual multiparameter physically based models was presented by *Gupta et al.* [1998], and this is extended to LSS models by *Gupta et al.* [this issue]. The characteristic of LSS models is that they have a relatively large number (p) of parameters $\theta = \{\theta_1, \cdots \theta_p\}$, which influence the time evolution behavior of several (m) different modeled variables ($Z_j(\theta, t_j)$, $t_j = ta_j, \cdots, tb_j$, $j = 1, \cdots m$). We define a single criterion $f_j(\theta)$ for each separate (jth) model response to measure the distance between the set of modeled responses $Z_j(\theta)$ and some benchmark responses X_j (Notes: (1) it is trivial to extend the methodology to allow more than one complementary criterion to be defined on each modeled response; (2) without loss of generality, we assume that $f_j(\theta) > 0$). The benchmark responses could either be model simulations $Z_j(\theta^*)$ performed at some nominal point θ^* in the parameter space or a set of observations O_j made at a study site. The specification of the mathematical form of these criteria depends on the problem and the goals of the user; in this work we adopt the commonly used root-mean-square error measure of the deviation between two time series $(\text{RMSE}_j(\theta) = \text{sqrt}\{(1/n) \Sigma_{t=1, \ldots, n}(Z_j(\theta, t) - X_j(t))^2\})$. The multicriteria response of the model as a function of the parameters can then be defined on the feasible parameter space Θ as

$$F(\theta) = \{f_1(\theta), \cdots, f_m(\theta)\} \quad (1)$$

where $\theta \subset \Theta$. The primary objective of parameter sensitivity analysis is to determine which of the parameters θ give rise to significant variations in the multicriteria response function $F(\theta)$ as they are allowed to vary jointly over Θ and to determine the nature of those variations. A secondary objective is to classify the sensitivity of the parameters into some order of relative importance.

The above-stated goal of deriving multicriteria parameter sensitivity can be addressed by extending and improving the "regional sensitivity analysis" (RSA) single-criterion method

(a) Parameter Space   (b) Criterion Space
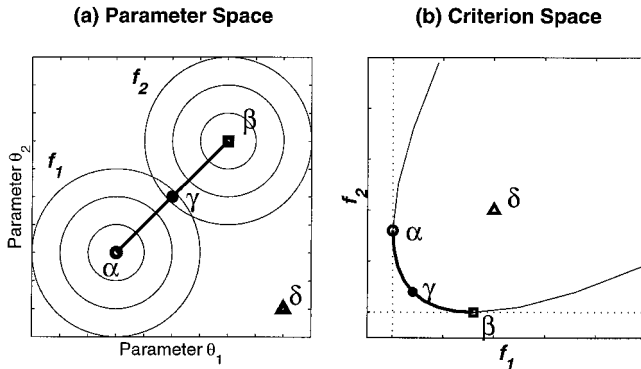
Figure 1. Example showing the Pareto solution set for a problem having two parameters (θ_1, θ_2) and two criteria (f_1, f_2): (a) parameter space and (b) criterion space. The points α and β indicate the solutions that minimize each of the individual criteria f_1 and f_2. The thick line indicates the set S of multicriteria minimizing points to the function $\{f_1, f_2\}$; γ is an element of the solution set S which is superior in the multicriteria sense to any point δ not contained in it.

[*Spear and Hornberger*, 1980; *Hornberger and Spear*, 1981]. The latter method begins with a uniform random sampling of *n* distinct points in the feasible parameter space Θ (we shall call this set of points *R*) and proceeds by partitioning Θ into a region *B* (called the "behavioral" region), which is characterized by a desirable modeled behavior, and its complement $\underline{B}$ (called the "nonbehavioral" region). The behavioral region can be defined either in terms of the parameter space that produces desired trajectories of the modeled responses or state variables [*Spear and Hornberger*, 1980], or in terms of the (subjectively assessed) desirable values of a summarizing criterion [*Hornberger et al.*, 1985; *Beven and Binley*, 1992; *Chang and Delleur*, 1992; *Spear et al.*, 1994; *Ohte and Bales*, 1995; *Franks et al.*, 1997]. In the context of (1) the behavioral region must be defined in terms of a multicriteria threshold applied to the function $F(\theta)$. A simple way to do this is to select a $T = \{T_1, \cdots, T_m\}$ and define $B(T)$ such that $F(\theta) < T$ (where $<$ denotes component-by-component superiority). However, this requires the subjective selection of *m* threshold values and, further, does not guarantee multicriteria superiority [*Bastidas*, 1998]. An alternative approach is to transform the multicriteria formulation (equation (1)) into a single-criterion form by some mapping such as

$$g(\theta) = \sum_{j=1,\cdots,m} w_j f_j(\theta) \qquad (2)$$

where the w_j are user-selected weights defining the relative importance of the individual criteria and to apply a single threshold to $g(\theta)$. However, this still requires the subjective selection of the *m* weights. These problems can be avoided largely by adopting the notion of Pareto Ranking, this being a scale-independent and truly multiobjective way to achieve a partition of the feasible space.

3.2. Pareto Ranking

A defining characteristic of the multicriteria response function $F(\theta)$ is that there is not, in general, a unique minimizing point in the parameter space. That is, because of errors in the model structure and other possible sources, it is usually impossible 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 of another. Figure 1a illustrates a simple case with two parameters (θ_1, θ_2) and a two-criteria response function $\{f_1, f_2\}$. Figure 1a shows the feasible parameter space Θ, and Figure 1b shows the projection of the parameter space into the function space. The points α and β indicate the solutions that minimize each of the individual criteria. The thick line indicates the set *S* of multicriteria minimizing points to the function $\{f_1, f_2\}$. If γ and δ are points arbitrarily selected from inside and outside *S*, respectively, then every point γ is superior to every point δ in a multicriteria sense because it has the property that $f_j(\gamma) < f_j(\delta)$, for $j = 1, 2$. However, it is not possible to find another point γ^* in *S* such that γ is superior to γ^*; instead, γ will be superior to γ^* for one criterion but inferior 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. However, we here call it the Pareto set.

The notion of Pareto Ranking [*Goldberg*, 1989] can be established using the set *R* as a sample approximation to Θ as follows: First, the Pareto set R_1 of the sample *R* is assigned rank 1 and set aside. Next the Pareto set R_2 of the remaining parameter space is identified, assigned rank 2, and also set aside. In this way the procedure is to assign progressively higher ranks until the entire sample *R* has been processed. Note that all the points belonging to a set R_i are not inferior to each other, but that each set R_i is superior to R_j ($j > i$) in a multicriteria sense. The above ranking procedure imposes an order such that lower-ranked points are closer to the Pareto set. Thus the requirement to partition Θ into behavioral and nonbehavioral regions reduces to the need to select an appropriate Pareto rank R^* as a threshold.

3.3. Testing for Sensitivity: MOGSA Algorithm

Once Θ has been partitioned into behavioral and nonbehavioral sets (i.e., *R* is partitioned into R_B and $R_{\underline{B}}$) by selecting a threshold Pareto rank, the degree of difference between those two sets of parameters forms the basis for conclusions regarding the relative sensitivities of the parameters. Following *Spear and Hornberger* [1980] and the nomenclature therein, we can test whether the individual a priori marginal parameter distributions of the two sets separate under the classification. If an individual marginal distribution does not separate (i.e., $p(\theta_i) = p(\theta_i/R_B) = p(\theta_i/R_{\underline{B}})$), this is because the parameter θ_i taken alone does not have a significant effect on the occurrence or nonoccurrence of the behavior; that is, the behavior is insensitive to θ_i over the multidimensional region of the parameter space.

In practice, and because they are simpler to compute, we actually use the Kolmogorov-Smirnov (*K-S*) two-sample test to establish whether the two cumulative marginal distributions are different. The *K-S* test is based on the statistic

$$d_{m,n} = \sup_x |P_B(x) - P_{\underline{B}}(x)| \qquad (3)$$

where P_B and $P_{\underline{B}}$ are the sample cumulative distributions functions corresponding to $p(\theta_i/B)$ and $p(\theta_i/\underline{B})$ for *n* behaviors and *m* nonbehaviors, respectively. The statistic represents the maximum distance between the two cumulative distribution function curves and can be associated with a particular significance level or probability value, thereby allowing a rela-
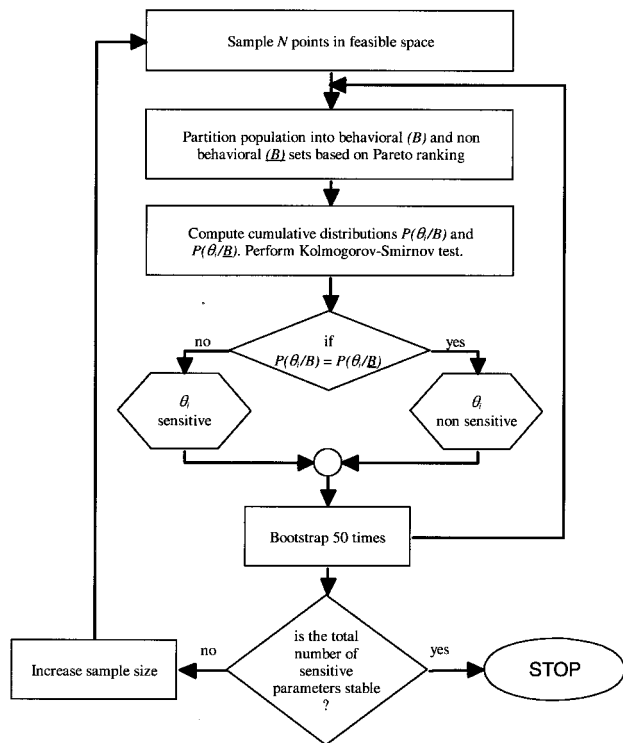
Figure 2. Flowchart of the MOGSA parameter sensitivity analysis method.

tive ranking of parameter sensitivities. In this study, significance levels below 1%, from 1 to 5%, and above 5% are used to define "high," "medium," and "low" parameter sensitivities, respectively.

The results of the sensitivity test described above will vary both with sample size and across samples, particularly for small sample sizes. Therefore we ensure statistical robustness by bootstrapping the sample (resampling 50–200 times with replacement) [*Efron*, 1979a, b; *Mooney and Duval*, 1993] and by using the median of the *K-S* statistics so obtained [*Rousseeuw*, 1991]. Further, we employ a procedure of successively increasing the sample size until the number of "sensitive" parameters stabilizes. The results could also be dependent on the choice of the thresholding Pareto rank R^*. Our computer implementation permits simultaneous analysis for several threshold values to test for such dependence. The methodology, named multiobjective generalized sensitivity analysis (MOGSA), is presented in Figure 2.

4. Case Studies

The above-described MOGSA procedure was used to estimate the parameter sensitivity of a typical, complex land surface scheme using two different data sets, one from a Sonoran Desert semiarid site near Tucson, Arizona [*Unland et al.*, 1996], and the other from the ARM-CART (atmospheric radiation measurement—cloud and radiation testbed) project E13 grassland site within the southern Great Plains GCIP study area. 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 (see Table 1). For a description of the BATS model and both

data sets, please refer to the companion paper [*Gupta et al.*, this issue].

The Tucson data were collected at 20 min time intervals and cover an 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. The observed model response variables are sensible heat (H), latent heat (λE), and ground temperature (T_g). The measured heat fluxes were flagged at certain times by *Unland et al.* [1996] as being of dubious quality, and these data were therefore not used in this study. The remaining observational data were available at 5219 time steps.

The ARM-CART data set covers the 5 month period from April 1 to August 25, 1995, with a sampling interval of 30 min. The observed forcing values are net radiation (R_n in W/m²), surface temperature (T_a in Kelvins (K)), atmospheric pressure (p_a in kPa), relative humidity (r_h in percent), wind velocity (V_a in m/s), and precipitation (P in millimeters (mm)). Several gaps in the data were filled via interpolation by taking account of the diurnal cycle and the occurrence of precipitation [see *Gupta et al.*, this issue]. The data set also contains time series observations for four variables that correspond with two model outputs, sensible heat (H in watts per square meter (W/m²)) and latent heat (λE in W/m²), and two model state variables (ground temperature (T_g in K) and soil moisture (S_w in mm)). Because of problems with the Bowen ratio system measurements, all H and λE data corresponding to Bowen ratio between 0.75 and 1.25 were discarded. The resulting number of time steps at which values of H, λE, T_g, and S_w are available is 4237.

4.1. Sensitivity Analysis Results

The MOGSA methodology was used to estimate the parameter sensitivity of BATS for each site using the RMSE between the observed and the model-simulated time series values of each observed system response as a different system response criterion: $\{H\}$, $\{\lambda E\}$, $\{T_g\}$, and $\{S_w\}$ for ARM-CART and $\{H\}$, $\{\lambda E\}$, and $\{T_g\}$ for Tucson. The parameters were allowed to vary between minimum and maximum values corresponding approximately to the full range of possible values for all locations on the Earth, except for the parameters rough and displa that do not include the rainforest vegetation type and parameter xmohyd for which the range was somewhat more restricted (see Table 1). Note that proper feasibility constraints were imposed on the allowable values for 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 are less than the corresponding soil layer depths, and the season variability of vegetation cover and leaf area index are smaller than their maximum values.

The algorithm stabilized for a minimum sample size of 750 for the Tucson data and for a minimum sample size of 3000 for the ARM-CART data. The Pareto rank 10 was chosen as the threshold value R^* because (1) it was found to achieve a higher number of sensitive parameters with a smaller sample size (for both the mean and the median of the K-S probability value and for both sensitivity thresholds 0.01 and 0.05); and (2) because it gave stable results regardless of sample sizes. The results are presented in Figure 3 for the Tucson site and Figure 4 for the ARM-CART site. The top subplots show the results for the 16 vegetation parameters and the bottom subplots show the re-

Table 1.   Parameters Considered in BATS Sensitivity Analysis

	Name	Description (units)	Range	Tucson Reasonable Range	ARM-CART Reasonable Range
Parameters Associated With Vegetation (18 Vegetation Types)					
1	vegc	vegetation cover	0.0–0.95	0.10–0.70	0.40–0.95
2	seasf	difference between vegc and fractional cover at 269 K	0.0–0.60	0.00–0.60	0.00–0.80
3	rough	aerodynamic roughness length (m)[a]	0.0024–1.0	0.01–0.84	0.01–0.35
4	displa	displacement height (m)[a]	0.0–5.0	0.05–1.50	0.05–1.50
5	rsmin	minimum stomatal resistance (s/m)	5.0–200.0	50.0–200.0	50.0–200.0
6	xla	maximum leaf area index	0.0–6.0	0.05–3.00	0.05–6.00
7	xlai0	minimum leaf area index	0.0–5.0	0.05–3.00	0.05–4.00
8	sai	stem area index	0.5–4.0	1.00–4.00	1.00–3.00
9	sqrtdi	inverse sqrt of leaf dimension (mm$^{-0.5}$)	5.0–10.0	5.00–10.00	5.00–10.00
10	fc	light dependence of stomatal resistance (m^2/w)	0.02–0.06	0.02–0.06	0.02–0.06
11	depuv	depth of top soil layer (m)	0.05–0.5	0.01–0.20	0.01–0.20
12	deprv	depth of root zone layer (m)	0.5–2.0	0.50–2.00	0.50–2.00
13	deptv	depth of total zone layer (m)	5.0–10.0	5.00–10.00	5.00–10.00
14	albvgs	vegetation albedo for shortwave < 0.7 μm	0.04–0.20	0.10–0.20	0.10–0.20
15	albvgl	vegetation albedo for longwave > 0.7 μm	0.18–0.40	0.20–0.40	0.20–0.40
16	rootf	ratio of roots in upper layer to roots in root layer	0.30–0.90	0.10–0.90	0.10–0.90
Parameters Associated With Soil Texture (12 Textures)					
17	xmopor	porosity	0.33–0.66	0.33–0.66	0.33–0.66
18	xmosuc	minimum soil suction (mm)	30.0–200.0	30.0–200.0	30.0–200.0
19	xmohyd	maximum hydraulic conductivity (mm/s)[b]	0.0008–0.01	0.0008–0.01	0.0008–0.01
20	xmowil	water content at which permanent wilting point occurs[c]	0.088–0.542		
21	xmofc	ratio of field capacity to saturated water content[c]	0.404–0.866		
22	bee	Clapp and Hornberger "*b*" parameter	3.5–10.8	3.5–10.8	3.5–10.8
23	skrat	ratio of soil thermal conductivity to that of loam	0.7–1.7	0.7–1.7	0.7–1.7
Parameters Associated With Soil Color (8 Colors)					
24	solour	soil albedo for different colored soils	0.05–0.12	0.05–0.12	0.05–0.12
Initial Conditions					
25	ssw	surface zone water content (m)	0.0–depuv	0.0–0.2	0.0–0.2
26	rsw	root zone water content (m)	0.0–deprv	0.0–5.0	0.0–5.0
27	tsw	total zone water content (m)	0.0–deptv	0.0–10.0	0.0–10.0

Global and reasonable ranges.
[a]Range does not include tropical rain forest.
[b]Range restricted.
[c]Parameters not considered in the analysis.

sults for the eight soil parameters and the three initial soil-moisture conditions.

The top part of each subplot shows the "global" sensitivity of the model, i.e., the sensitivity obtained using the Pareto ranking 10 as threshold. The bottom part of each subplot shows the single-criteria sensitivity of the model to each modeled variable. For consistency, the single-criterion thresholds were chosen to partition the sample populations into fractions similar to those achieved by the Pareto ranking threshold. In the bottom part of each subplot, each group of vertical bars corresponds to the sensitivity of a different parameter; the longer the bar, the more sensitive the parameter. Also indicated in the plots are the 1 and 5% significance levels. If a sensitivity bar crosses the 1% level, the parameter is considered to have high sensitivity; if it crosses the 5% level, it is considered to have medium sensitivity. Otherwise, it is considered to be relatively insensitive.

Considering first the Tucson site (Figure 3), the joint multicriteria analysis (top bars) indicates 15 sensitive parameters, specifically 12 parameters with high sensitivity, of which 5 are related to vegetation (vegc, rough, rsmin, depuv, and deprv), 5 are related to soil (xmopor, xmosuc, bee, and skrat); and 2 are the initial top and total water contents (ssw and tsw); and 3 parameters have medium sensitivity, of which 2 are related to vegetation (displa and xla), and 1 is related to soil (solour). The single-criterion analyses (bottom bars in Figure 3) reveal fur-

ther details about the relationships between different modeled variables and individual parameters. Note, for example, that the heat fluxes are sensitive to the roughness length and to the depths of the top and root soil layers but that the ground temperature is not. In particular, it is significant that none of the modeled variables are sensitive to the seasonal variation of the vegetation cover (seasf) and the minimum leaf area index (xlai0) which is consistent with the properties of the Tucson semiarid environment.

In the case of the ARM-CART site (Figure 4), the joint multicriteria analysis (top bars) indicates 16 sensitive parameters, specifically 14 parameters with high sensitivity, of which 8 are related to vegetation (vegc, seasf, rsmin, xla, xlai0, depuv, deprv, and rootf), 4 are related to soil (xmosuc, xmohyd, bee, and skrat), and 1 is the initial water content of the top soil layer (ssw); and 2 parameters with medium sensitivity, 1 related to soil porosity (xmopor), and 1 being the initial water content of the root layer (rsw). Note that the initial total soil layer water content (tsw) might also be considered of "medium" sensitivity. The single-criterion analyses (bottom bars in Figure 4) reveal that the heat fluxes are very sensitive to stomatal resistance but that the state variables (ground temperature and soil moisture) are not. Further, in contrast with the Tucson site, the modeled variables are all sensitive to the seasonal variation of the vegetation cover (seasf), and the heat fluxes are both sensitive to minimum leaf area index (xlai0). This is plausible,
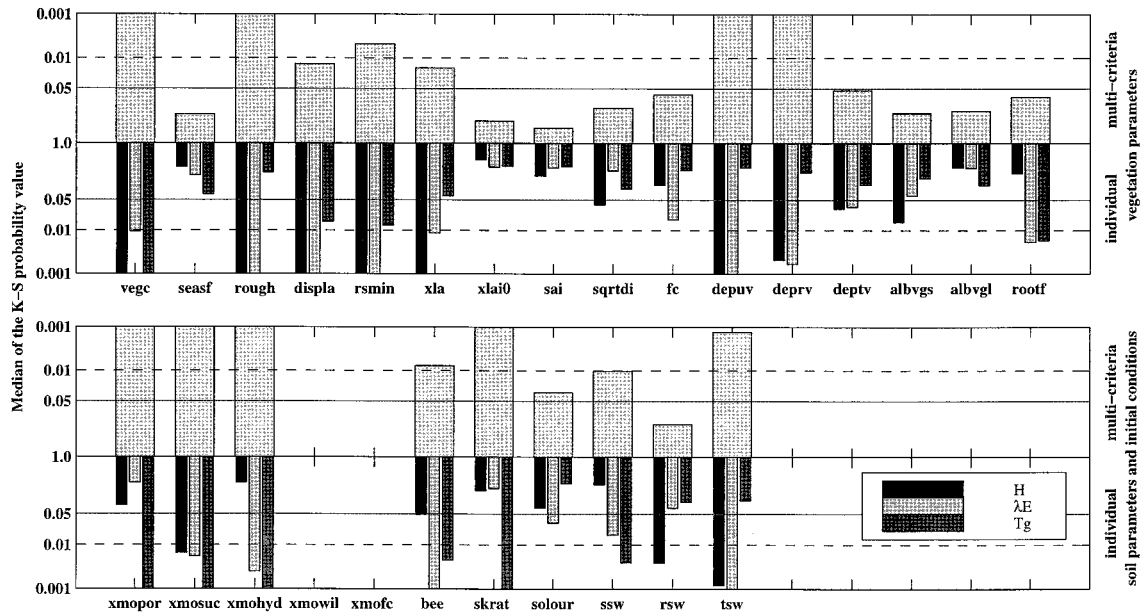
Figure 3. Parameter sensitivity analysis results for the BATS model applied to the Tucson semiarid site. Multicriteria results are indicated by upward pointing bars and individual-criterion results are indicated by downward pointing bars; the longer the bar, the more sensitive the parameter. Sensitivity bars crossing the 1% (0.01) and 5% (0.05) levels indicate parameters having high and medium sensitivity, respectively. The top subplot shows results for the vegetation-related parameters, and the bottom subplot shows results for the soil parameters and initial conditions. Parameters xmowil and xmofc were not included in the analysis.

considering that the period for which data are available is from April to August, and the vegetation over which measurements are made is irrigated crop.

Thus both sites show relatively high sensitivity to 11 parameters, 5 related to vegetation (vegc, rsmin, xla, depuv, and deprv), 6 related to soil (xmopor, xmosuc, xmohyd, bee, and skrat), and 1 being the initial water content of the surface soil layer (ssw). Both sites show relatively low sensitivity to 6 of the vegetation parameters, 3 related to stem and leaf properties (sai, sqrtdi, and fc), the depth of the total soil layer (deptv), and the vegetation albedos (albvgs, and albvgl).

However, the vegetation-related parameters (rough and displa) and the soil color parameter (solour) show high sensitivity at the Tucson site but low sensitivity at the ARM-CART site, while the vegetation-related parameters (seasf, xlaio, and rootf) and the soil parameter (xmotc) show low sensitivity at the Tucson site but high sensitivity for ARM-CART. Further, the vegetation-related roughness length (rough) is the most sensitive parameter for the Tucson site but is the least sensitive parameter for the ARM-CART site. In contrast, the difference between vegetation and fractional cover (seasf) is a very highly sensitive parameter for the ARM-CART site but a parameter with very low sensitivity at the Tucson site.

The sensitivity results for both sites are consistent with the model formulation and the physical properties of the different environments. For the semiarid Tucson site, 6 of the 8 most sensitive parameters are related to the soil storages and texture, while for the ARM-CART site, 5 of the 8 most sensitive parameters are related to vegetation properties. Moreover, the results generally agree with *Gao et al.* [1996], *Wilson et al.* [1987a, b], and *Lettenmaier et al.* [1996]. Note that at both sites in this study, there is low sensitivity to the light dependence of stomatal resistance (fc) and to the vegetation albedos (albvgs,

albvgl) in marked contrast to the results of *Henderson-Sellers* [1992] for tropical forest.

4.2. Application of the Sensitivity Results to Model Calibration

An important aim of the parameter sensitivity analysis is to allow the possible reduction in the number of parameters that must be estimated, thereby reducing the computational time required for model calibration. *Gupta et al.* [this issue] conducted a multicriteria calibration run for the Tucson site in which they estimated 25 parameters using the latent heat (H), sensible heat (λE), and ground temperature (T_g) system responses. This 25 parameter calibration run required ~20,000 iterations (model simulations) to converge to a solution. To illustrate the usefulness of parameter sensitivity analysis in reducing the computational effort associated with calibration, a similar multicriteria calibration run was made in which only 18 parameters were estimated, including the 15 sensitive parameters (vegc, rough, displa, rsmin, xla, depuv, deprv, xmopor, xmosuc, xmohyd, bee, skrat, solour, ssw, and tsw) and the three marginally sensitive parameters (fc, rootf, and rsw). (Note that in BATS the parameter xmowil is derived from other parameters and was therefore not included). In this run, the insensitive parameters (seasf, xlai0, sai, sqrtdi, deptv, albvgs, and albvgl) were prescribed to be at the BATS default values for semiarid regions [see *Gupta et al.*, this issue, Table 1]. The goal is to demonstrate that if the parameters diagnosed as being insensitive using the MOGSA methodology are not calibrated, there is only a marginal degradation in the quality of the calibrated model performance, and the preferred range of parameters is changed only slightly.

The results of these two runs are compared in Figure 5. The results shown in light shade correspond to the Pareto set ob-
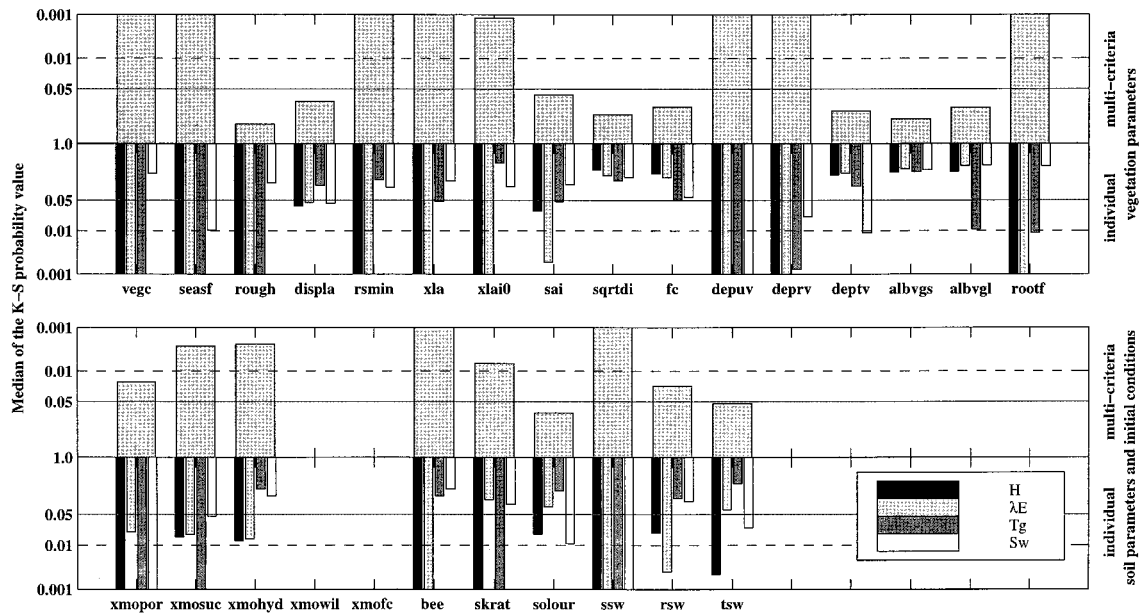
Figure 4. Parameter sensitivity analysis results for the BATS model applied to the ARM-CART grassland site E13. See explanation on caption for Figure 3.

tained by the 25 parameter multicriteria calibration run and the results shown in black correspond to the Pareto set obtained by the 18 parameter multicriteria calibration run. The parameter estimates are shown on a normalized parameter plot (Figure 5a); the parameters and the three initial soil moisture contents are listed along the x axis, and the y axis corresponds to the parameter values plotted on a range normalized by the allowable maximum and minimum values for the parameters (Table 1). Each line from left to right across the plot corresponds to a different parameter estimate. The lightly shaded area in the background corresponds to the parameter ranges considered to be reasonable for the Sonoran Desert semiarid Tucson site (see Table 1); these reasonable ranges were not used to constrain the calibration runs but were instead used as postcalibration checks on the reliability and power of the calibration procedure: that is, a successful calibration run should converge to estimates that fall within these ranges. Figure 5b shows the trade-offs between the different criteria in the multicriteria space; the three response criteria are listed along the x axis, while the y axis corresponds to the criterion values (RMSEs) with preferred 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 set of parameter values). Note that if a calibration result plots as a line, which 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 lines cross each other, then the results are noninferior to each other in a multicriteria sense.

It is important to note that all the solutions for both the 25 and the 18 parameter calibration runs fall within the reasonable bounds (Figure 5a). In the case of the 25 parameter calibration, some of the insensitive parameters (sqrtdi, deptv, albvgs, and albvgl) tend toward different locations from the fixed default values used in the 18 parameter calibration, and to be poorly constrained (i.e., they have a relatively wide spread). Highly sensitive parameters (e.g., vegc and xmopor)

tend to be tightly constrained and have similar values in both calibration runs, while the less sensitive parameters (e.g., bee and fc) tend to be less constrained and vary more. This is entirely consistent with the expectation that the less sensitive parameters will have a less well defined optimal region.

Figure 5b shows that the 25 parameter calibration run is superior to the 18 parameter calibration run in a multicriteria sense; but only marginally so, the improvement in the RMSEs for the three system responses is less than 10%. However, the 18 parameter calibration run required only 4000 iterations (model simulations) to converge to a solution, a fivefold reduction from the 25 parameter calibration. These results suggest that the parameter sensitivity analysis using MOGSA is a sensible and viable way to reduce the dimensionality of the parameter estimation problem, thereby reducing the computational time required for calibration. The left column of Figure 6 shows a typical 10 day portion of the time series match obtained by the 18 parameter calibration run. For each system response, the dots indicate the observed data and the lightly shaded area represents the range of variation at each time step corresponding to the Pareto set of parameter estimates. The relative uncertainty of the daytime λE trajectory is a consequence of the high degree of noise in the data. The statistics computed for each of the responses, as well as the scatterplots, are defined for the middle point of the trajectory set and are calculated for the entire year of data. The correlation between the simulated and the observed values is gratifying.

5. Summary and Conclusions

A companion paper [*Gupta et al.*, this issue] describes the development of a multicriteria calibration procedure for determining the preferred range of the many parameters required in a complex, present-day LSS and illustrates the application of that procedure using the BATS model with data from two field sites. In practice, it is desirable to restrict the number of parameters to be calibrated in order to reduce the
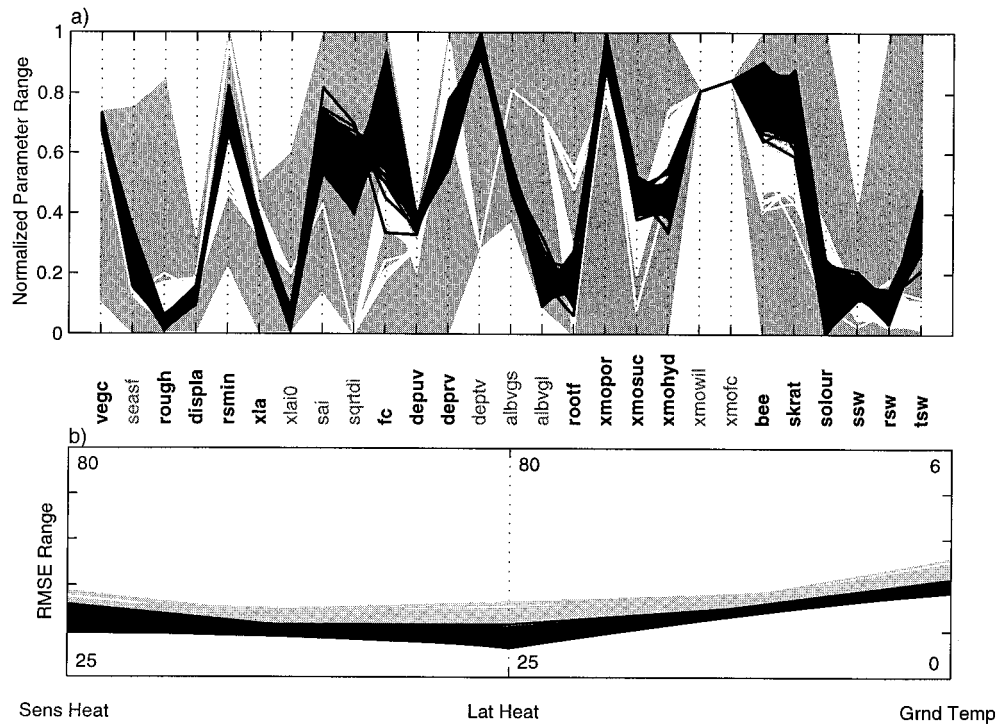
Figure 5. Results showing the benefits of using multicriteria parameter sensitivity analysis to reduce parameter dimensionality before calibration of the BATS model to the Tucson semiarid site. The full 25 parameter calibration results are shown in black and the reduced 18 parameter calibration results are shown in light shade. (a) Parameter estimates plot: each horizontal line represents a member of the Pareto solution set. The medium shade area indicates the reasonable range of parameter variability for semiarid regions. (b) Criterion values plot: each horizontal line represents a member of the Pareto solution set. Lines closer to the base of the plot indicate superior solutions.

computational effort. Further, the ability to make objective determination of the relative sensitivity of modeled variables to the specific parameters used in an LSS at a particular site has a broader value as a basis for specifying the relative importance of the physical processes represented in the model. Thus a methodology is required which is consistent with multicriteria parameter estimation methods and which is capable of determining the extent to which the modeled variables given by an LSS are sensitive to the value of the individual parameters used in it.

This need is addressed directly in this paper, by describing the development of a novel, multicriteria approach to evaluate parameter sensitivity, the MOGSA algorithm, which extends and improves the regional sensitivity analysis single-criterion method of *Spear and Hornberger* [1980] to multicriteria by sorting the possible parameter sets via the notion of Pareto ranking and by applying bootstrapping and sequential sampling to ensure statistical robustness of the results. The method enables the parameters to be ranked in terms of both absolute and relative sensitivity. This procedure is then applied in the context of the parameter estimation studies described in the companion paper. First, the parameters that have significant influence on the values of the modeled variables at the two sites are determined and are distinguished from the parameters that have little influence and which might therefore be safely omitted from the calibration run. The consequences of omitting these parameters from the multicriteria calibration were then investigated.

For both field sites considered in this study, the MOGSA

algorithm was effective in specifying the sensitivity of BATS-calculated variables to the parameters used in the model. The results provide insight into important parameters and associated processes in BATS at the two field sites. In both cases the modeled variables were found to be sensitive to 12 parameters. Five are related to vegetation, specifically vegetation cover, aerodynamic roughness, minimum stomatal resistance, depth of upper soil layer, and depth of root zone soil layer; six are related to soil, specifically porosity, minimum soil suction, maximum hydraulic conductivity, Clapp and Hornberger "b" parameter, and ratio of soil thermal conductivity to that of loam; and one is the initial water content of the surface layer. In both cases the modeled variables were found to be insensitive to six vegetation-related parameters, specifically stem area index, inverse square root of leaf dimension, light dependence of stomatal resistance, depth of the total soil layer, and the vegetation albedos for both shortwaves and longwaves.

However, important differences were found in the model sensitivities at the two sites. These presumably reflect the prevalent climate and the nature of the vegetation at the sites. Specifically, the parameters aerodynamic roughness, displacement height, and soil color show high sensitivity at the semiarid Tucson site but low sensitivity at the ARM-CART grassland site, while the parameters difference between vegc and fractional cover at 269 K, minimum leaf area index, ratio of roots in the upper layer to roots in the lower layer, and ratio of field capacity to saturated water content show low sensitivity at the Tucson site but high sensitivity for ARM-CART. In general, the sensitivity results for both sites are consistent with the
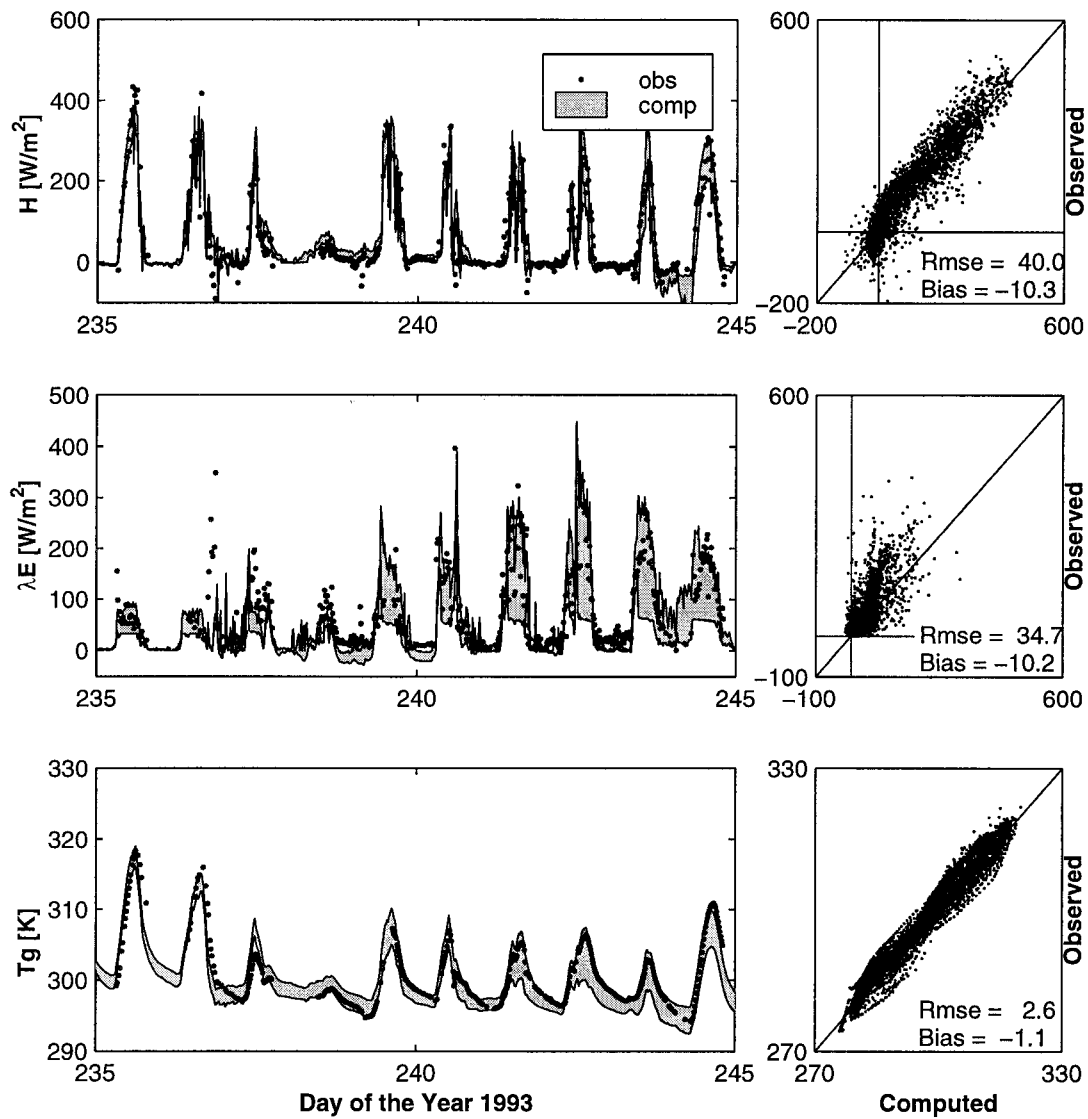
Figure 6. Model performance with the 18 parameter multicriteria calibration parameter estimates compared to the observed data. The time series plots on the left show a typical 10 day period; the lightly shaded region represents the range of output simulations corresponding to the range of the Pareto parameter estimates, and the dots represent the observed data. The scatterplots on the right compare the entire period of data with the midpoints of the model-simulated ranges.

model formulation and the physical properties of the different environments. For the Tucson site, six of the eight most sensitive parameters are related to soil properties, while for the ARM-CART site, five of the eight most sensitive parameters are related to vegetation properties.

Comparison between the results of calibration runs for the Tucson semiarid site, in which all the parameters were optimized and in which only a subset of sensitive parameters were optimized, was consistent with expectations. There was only limited degradation in the quality of the correspondence between modeled and observed values (the RMSE for the three criteria increased by less than 10%), and the ranges of preferred values for the parameters that were optimized in both runs were mutually consistent. However, the calibration run for the sensitive parameter subset required only 4000 iterations (model simulations) to converge to a solution, a fivefold reduction over the full parameter calibration.

Thus in summary, the multicriteria methodology MOGSA developed in this study proved successful when applied to evaluate the sensitivity of model-calculated variables to individual parameters for a complex LSS at two distinctly different field sites. The sensitivity results for both sites were found to be consistent with the physical properties of the different environments, thereby supporting the reasonableness of the model formulation. Further, the analysis provided evidence on the importance of parameters (and the associated processes) operating at the two sites, and when the parameters it identified as insensitive were omitted from the optimization process, there was little degradation in the quality of the model description and little change in the preferred range of the remaining parameters. The MOGSA 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*).

References

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.

Beven, K. J., and A. M. Binley, The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Process.*, 6, 29–44, 1992.

Chang, D. H., and J. W. Delleur, Systematic parameter estimation of watershed acidification model, *Hydrol. Process.*, 6, 29–44, 1992.

Collins, D. C., and R. Avissar, An evaluation with the Fourier amplitude sensitivity test (FAST) of which land surface parameters are of greatest importance in atmospheric modeling, *J. Clim.*, 7(5), 681–703, 1994.

Cukier, R. I., H. B. Levine, and K. E. Shuler, Nonlinear sensitivity analysis of multiparameter model systems, *J. Comput. Phys.*, 26(1), 1–42, 1978.

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.

Efron, B., Bootstrap methods: Another look at the jackknife, *Ann. Stat.*, 7, 1–26, 1979a.

Efron, B., Computers and the theory of statistics: Thinking the unthinkable, *Soc. Ind. Appl. Math.*, 21, 460–480, 1979b.

Franks, S. W., K. J. Beven, P. F. Quinn, and I. R. Wright, On the sensitivity of soil-vegetation-atmosphere transfer (SVAT) schemes: Equifinality and the problem of robust calibration, *Agric. For. Meteorol.*, 86, 63–75, 1997.

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

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

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.

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

Henderson-Sellers, A., Assessing the sensitivity of a land surface scheme to parameters used in tropical-deforestation experiments, *Q. J. R. Meteorol. Soc.*, 118, 1101–1116, 1992.

Hornberger, G. M., and R. C. Spear, An approach to the preliminary analysis of environmental systems, *J. Environ. Manage.*, 12(1), 7–18, 1981.

Hornberger, G. M., K. J. Beven, B. J. Cosby, and D. E. Sappington, Shenandoah watershed study: Calibration of a topography-based, variable contributing area hydrological model to a small forested catchment, *Water Resour. Res.*, 21(12), 1841–1850, 1985.

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

Mooney, C. Z., and R. D. Duval, *Bootstrapping: A Nonparametric Approach to Statistical Inference*, 75 pp., Sage, London, 1993.

Ohte, N., and R. C. Bales, Multidimensional parameter estimation of the integrated alpine hydrochemical model using Monte Carlo simulation, in *Biogeochemistry of Seasonally Snow Covered Catchments*, *IAHS Publ. 228*, pp. 175–183, Int. Assoc. of Hydrol. Sci., Gentbrugge, Belgium, 1995.

Pitman, A. J., Assessing the sensitivity of a land-surface scheme to the parameter values using a single column model, *J. Clim.*, 7(12), 1856–1869, 1994.

Rousseeuw, P. J., Tutorial to robust statistics, *J. Chemometrics*, 5, 1–20, 1991.

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.

Spear, R. C., T. M. Grieb, and N. Shang, Parameter uncertainty and interaction in complex environmental models, *Water Resour. Res.*, 30(11), 3159–3169, 1994.

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.

Wilson, M. F., A. Henderson-Sellers, R. E. Dickinson, and P. J. Kennedy, Sensitivity of the Biosphere-Atmosphere Transfer Scheme (BATS) to the inclusion of variable soil characteristics, *J. Clim. Appl. Meteorol.*, 26, 341–362, 1987a.

Wilson, M. F., A. Henderson-Sellers, R. E. Dickinson, and P. J. Kennedy, Investigation of the sensitivity of the land surface parameterization of the NCAR Community Climate Model in regions of tundra vegetation, *Climatology*, 7, 319–343, 1987b.

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).