JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 115, D03106, doi:10.1029/2009JD012035, 2010

Quantifying parameter sensitivity, interaction, and transferability in hydrologically enhanced versions of the Noah land surface model over transition zones during the warm season

Enrique Rosero,[1,2] Zong-Liang Yang,[1] Thorsten Wagener,[3] Lindsey E. Gulden,[1,2] Soni Yatheendradas,[4,5] and Guo-Yue Niu[6]

Received 10 March 2009; revised 24 June 2009; accepted 20 October 2009; published 3 February 2010.

[1] We use sensitivity analysis to identify the parameters that are most responsible for controlling land surface model (LSM) simulations and to understand complex parameter interactions in three versions of the Noah LSM: the standard version (STD), a version enhanced with a simple groundwater module (GW), and version augmented by a dynamic phenology module (DV). We use warm season, high-frequency, near-surface states and turbulent fluxes collected over nine sites in the U.S. Southern Great Plains. We quantify changes in the pattern of sensitive parameters, the amount and nature of the interaction between parameters, and the covariance structure of the distribution of behavioral parameter sets. Using Sobol′'s total and first-order sensitivity indexes, we show that few parameters directly control the variance of the model response. Significant parameter interaction occurs. Optimal parameter values differ between models, and the relationships between parameters also change. GW decreases unwarranted parameter interaction and appears to improve model realism, especially at wetter study sites. DV increases parameter interaction and decreases identifiability, implying it is overparameterized and/or underconstrained. At a wet site, GW has two functional modes: one that mimics STD and a second in which GW improves model function by decoupling direct evaporation and base flow. Unsupervised classification of the posterior distributions of behavioral parameter sets cannot group similar sites based solely on soil or vegetation type, helping to explain why transferability between sites and models is not straightforward. Our results suggest that the a priori assignment of parameters should also consider the climatic conditions of a study location.

Citation: Rosero, E., Z.-L. Yang, T. Wagener, L. E. Gulden, S. Yatheendradas, and G.-Y. Niu (2010), Quantifying parameter sensitivity, interaction, and transferability in hydrologically enhanced versions of the Noah land surface model over transition zones during the warm season, *J. Geophys. Res.*, *115*, D03106, doi:10.1029/2009JD012035.

1. Introduction

[2] Like other environmental models built to support scientific reasoning and to test hypotheses to improve our understanding of the Earth system, land surface models (LSMs) have grown in sophistication and complexity [*Pitman*, 2003; G.-Y. Niu et al., The community Noah land surface model with multiphysics options, unpublished manuscript, 2009]. The evaluation of LSM simulations is consequently nontrivial and, especially when LSMs are to be used in predictive mode for operational forecasting, policy assessments, or decision making, demands more powerful methods for the analysis of their behavior [*Saltelli*, 1999; *Jakeman et al.*, 2006; *Wagener and Gupta*, 2005; *Randall et al.*, 2007; *Gupta et al.*, 2008; *Abramowitz et al.*, 2008]. One powerful method in this context is sensitivity analysis (SA). In this article, we inform LSM development by using sophisticated SA to guide the ongoing development of the commonly used Noah LSM [*Ek et al.*, 2003].

[3] SA is the process of investigating the role of the various assumptions, simplifications and other factors (including input data and parameters) in controlling the simulations made by a model. SA is a tool that enables the exploration of high-dimensional parameter spaces of complex environmental models to better understand what controls model performance [*Saltelli et al.*, 2008]. Monte Carlo-based SA uses multiple model realizations to evaluate the range of model outcomes and identifies the input

[1]Department of Geological Sciences, Jackson School of Geosciences, University of Texas at Austin, Austin, Texas, USA.
[2]Now at ExxonMobil Upstream Research Company, Houston, Texas, USA.
[3]Department of Civil and Environmental Engineering, Pennsylvania State University, University Park, Pennsylvania, USA.
[4]Hydrological Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.
[5]Also at Earth System Science Interdisciplinary Center, University of Maryland, College Park, Maryland, USA.
[6]Biosphere 2 Earth Science, University of Arizona, Tucson, Arizona, USA.

0148-0227/10/2009JD012035$09.00

parameters that give rise to this uncertainty [*Wagener et al.*, 2001; *Wagener and Kollat*, 2007]. Used to its full potential, SA weighs model adequacy and relevance, identifies critical regions in the space of the inputs, unravels parameter interactions, establishes priorities for research, and, through an interactive process of revising the model structure, leads to simplified models and increased understanding of the natural system [*Saltelli et al.*, 2006].

[4] Detailed SA has so far been underutilized in LSM development and evaluation. If SA has been performed, then approaches to quantify "sensitivity" (the rate of change in model response with respect to a factor) are very frequently restricted to a simple exploratory analysis of the effects of factors taken one at a time (OAT), without regard for their interactions and only in the neighborhood of an initial reference set of factors. Although OAT is only justified for linear models [*Saltelli*, 1999; *Bastidas et al.*, 1999; *Saltelli et al.*, 2006], it has been used to explore the effects of parameters [e.g., *Pitman*, 1994; *Gao et al.*, 1996; *Chen and Dudhia*, 2001; *Trier et al.*, 2008], meteorological forcing, and ancillary data sets [e.g., *Kato et al.*, 2007; *Gulden et al.*, 2008a]. Global sensitivity analysis, the evaluation of sensitivity across the full feasible factor space, is considered a more powerful and sophisticated approach, though one with higher computational demands. One of the earliest approaches to global SA is the regionalized sensitivity analysis (RSA) [*Hornberger and Spear*, 1981]. RSA samples the entire parameter space and provides a robust assessment of the way parameter distributions change between subjectively defined "good" and "bad" (i.e., behavioral and nonbehavioral) model simulations [e.g., *Bastidas et al.*, 2006; *Prihodko et al.*, 2008] or within the behavioral ranges of different models [e.g., *Gulden et al.*, 2007; *Demaria et al.*, 2007]. By not explicitly accounting for interactions between parameters, RSA is prone to type II errors (nonidentification of an influential parameter) [*Saltelli et al.*, 2008]. RSA does not quantify the extent to which a parameter affects the variance of the model output, and it is typically applied with the sole purpose of identifying parameters that merit calibration [e.g., *Bastidas et al.*, 1999] (see the auxiliary material).[1] Another common approach to SA is the factorial method, a global variance-based SA (VSA) that explicitly accounts for parameter interactions. It uses a set of model runs whose parameters have been perturbed from an arbitrary reference value (default) to identify parameters that affect the variance of model output. Because accounting for higher-order interactions requires a prohibitive number of model runs, factorial analyses in LSM research have been limited to two factor interactions of few selected parameters [e.g., *Henderson-Sellers*, 1993; *Liang and Guo*, 2003; *Oleson et al.*, 2008] and have therefore not fully characterized parameter space. When RSA and VSA are used separately, both the dearth of firm conclusions regarding the effect of dominant parameters (and their interactions) on the model variance [e.g., *Bastidas et al.*, 2006] and the inability to draw cause-effect relationships between parameter regions and model responses [e.g., *Liang and Guo*, 2003] have precluded SA findings from being widely used in LSM development.

[5] We employ SA to compare the performance and physical realism of three versions of the Noah LSM: the standard Noah (STD), a version augmented with a simple groundwater model (GW) [*Niu et al.*, 2007], and a version augmented with an interactive canopy model (DV) [*Dickinson et al.*, 1998] simulate the land surface states and fluxes at nine sites in a transition zone between wet and dry climates using the data sets of International H_2O Project 2002 (IHOP_2002) [*LeMone et al.*, 2007]. Because of the strength of the land-atmosphere coupling in transition zones [*Koster et al.*, 2004], we focus on warm season climates of the U.S. Southern Great Plains. Neglecting uncertainty in the meteorological forcing, we document how parameter interaction and sensitivity varies with model, site, soil, vegetation, and climate.

[6] We use the Monte Carlo-based VSA method of Sobol′ to quantify total and first-order sensitivity indexes. The method of Sobol′ is more robust (it employs a representative sample of the parameter space) and efficient than factorial analysis [*Saltelli*, 2002], and it bypasses the perceived complexities (e.g., the design of the calculation matrix) often associated with factorial analysis. Note that because LSM developers have attempted to use physical principles when designing their models, the parameters of such physically based models are assumed to correspond to unchanging physical characteristics of a system. Consequently, the level of parameter interaction can be treated as an indirect measure of the physical realism of LSMs. That is, it is assumed that physically based models with less undesirable parameter interactions are better (i.e., more physically realistic) [*Beck*, 1987; *Spear et al.*, 1994; *Gupta et al.*, 2005]. We show that only a few parameters directly control model variance and that parameter interaction is significant.

[7] We look at the marginal distributions of behavioral parameters to investigate the ways in which "physically meaningful" LSM parameters function within alternate model structures. We focus on selected dominant parameter interactions that dictate model response. Because LSM parameter are assumed to be physically meaningful values [e.g., *Dickinson et al.*, 1986] that can either be measured in the field (e.g., porosity) or be inferred from (remotely sensed) observations (e.g., leaf area index (LAI)), their values should not change between models for a given site. We show that the distributions of the behavioral parameters differ between models and that the relationships between parameters change.

[8] A priori assignment of parameters based on soil texture and vegetation type is standard practice in the application of LSMs, justified by the assumption that physically meaningful parameters can be transferred between locations that share the same physical characteristics [e.g., *Sellers et al.*, 1996]. As a consequence of our SA-enabled model evaluation, we observe that LSM parameters are highly interactive and change between models and between sites, which implies that a priori assignment of parameters may not be justified. We use unsupervised classification to test parameter transferability. The similarity of estimated multivariate posterior distributions of behavioral parameters and their sensitivity for each site are compared to those obtained at other sites. We show that the changes between sites are not solely controlled by soil

[1] Auxiliary materials are available in the HTML. doi:10.1029/2009JD012035.

texture or vegetation types but appear to be strongly related to the climatic gradient.

[9] This paper is organized as follows. Experimental design and driving questions are formulated in section 2. Data sets, models, and methods are described in section 3. Section 4 presents the patterns of sensitivity obtained by the global variance-based method of Sobol′. Section 5 presents a case study demonstrating the use of SA to understand the functional relationships between behavioral parameters, whose interaction serves to characterize model structure and test hypotheses that regard the formulation of model. Section 6 discusses implications of the results for the transferability of parameters between locations with similar physical characteristics. Conclusions are summarized in section 7.

2. Driving Questions and Experimental Design

[10] Our research is guided by three questions that define our experimental design.

[11] 1. What are the dominant model parameters across the study region? We run a suite of Monte Carlo simulations to identify parameters that exert the greatest control on the variability of simulated fluxes and states at each IHOP site for all three models (STD, GW, and DV). We quantify sensitivity using the method of Sobol′. The SA results guide our subsequent investigation.

[12] 2. How do the dominant parameters' interactions change between models? With our focus toward model development, we investigate the relationships between behavioral model parameters and quantify how they change between models using the estimates of the total order sensitivity, the multivariate posterior parameter distributions, and the covariance structures.

[13] 3. How do behavioral parameters change with dominant physical characteristics of the study site? We summarize the relationships between model parameters and physical characteristics by classifying the multivariate posterior parameter distributions according to the sites' soil and vegetation types. Our classification provides insights into how parameters can be transferred to ungauged locations.

3. Models, Data, and Methods

3.1. Hydrologically Enhanced Versions of the Noah LSM

[14] We compared the standard Noah LSM release 2.7 (STD) to one that couples a lumped, unconfined aquifer model to the model soil column (GW) and a version that we equipped with a short-term phenology module (DV).

3.1.1. Noah Standard Release 2.7

[15] Noah [*Ek et al.*, 2003; *Mitchell et al.*, 2004] is a one-dimensional, medium complexity LSM used in operational weather and climate forecasting. The model is forced by incoming short- and long-wave radiation, precipitation, surface pressure, relative humidity, wind speed and air temperature. The computed state variables include soil moisture and temperature, water stored on the canopy and snow on the ground. Prognostic variables include turbulent heat fluxes, and fluxes of moisture and momentum. Noah has a single canopy layer with climatologically prescribed albedo and vegetation greenness fraction. The soil profile of Noah is partitioned into 4 layers (lower boundaries at 0.1, 0.4, 1.0 and 2.0 m below the surface). The vertical movement of water is governed by mass conservation and a diffusive form of the Richard's equation. Infiltration is represented by a conceptual parameterization for the subgrid treatment of precipitation and soil moisture. Drainage at the bottom of the soil layer is controlled only by gravitational forces; and the percolation process neglects hydraulic diffusivity. Direct evaporation from the topsoil layer, from water intercepted by the canopy and adjusted potential Penman-Monteith transpiration are combined to represent total evapotranspiration. The surface energy balance determines the skin temperature of the combined ground vegetation surface. Soil layer temperature is resolved with a Crank-Nicholson numerical scheme. Diffusion equations for the soil temperature determine ground heat fluxes. The Noah LSM uses soil and vegetation lookup tables for static soil and vegetation parameters such as porosity, hydraulic conductivity, minimum canopy resistance, roughness length, leaf area index.

3.1.2. Noah Augmented With a Simple Groundwater Model

[16] GW couples a lumped unconfined aquifer model [*Niu et al.*, 2007] to the lower boundary of the STD soil column. In GW, water flows vertically in both directions between the aquifer and the soil column. The modeled hydraulic potential is the sum of the soil matric and gravitational potentials. The relative water head between the bottom soil layer and the water table determines either gravitational drainage or upward diffusion of water driven by capillary forces. Aquifer specific yield is used to convert the water stored in the aquifer to water table depth. When water is plentiful, the water table is within the model's soil column; if water is insufficient to maintain a near-surface aquifer, the water table falls below the soil column. An exponential function of water table depth modifies the maximum rate of subsurface runoff (for computation of base flow) and determines the fraction of the grid cell that is saturated at the land surface (for calculation of surface runoff) [*Niu et al.*, 2005]. Observed moderate recharge rates for nonirrigated agricultural ecosystems in the U.S. Southern Great Plains [*Scanlon et al.*, 2005] warrant the simple representation of an aquifer for the simulation of surface-to-atmosphere fluxes in the region.

3.1.3. Noah Augmented With a Short-Term Dynamic Phenology Module

[17] We coupled the canopy module of *Dickinson et al.* [1998] to STD in order to compute changes in vegetation greenness fraction that result from environmental perturbations. The module allocates carbon assimilated during photosynthesis to leaves, roots, and stems; the fraction of photosynthate allocated to each reservoir is a function of, among other things, the existing biomass density. The model also tracks growth and maintenance respiration and represents carbon storage. Unlike STD, which computes greenness fraction by linear interpolation between monthly climatological values, DV represents short-term phenological variation by allowing leaf area to vary as a function of soil moisture, soil temperature, canopy temperature, and vegetation type. DV makes vegetation fraction an exponential function of LAI [*Yang and Niu*, 2003]. Because DV links vegetation fraction to dynamic LAI, DV makes direct

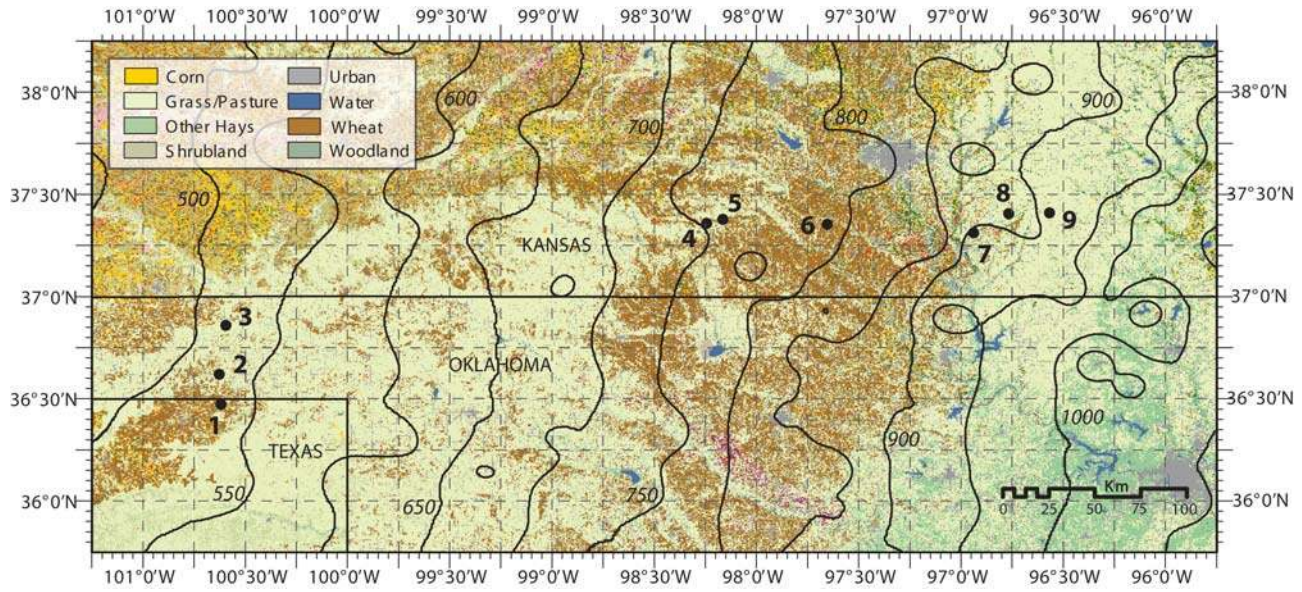

Figure 1. IHOP_2002 near-surface state and flux stations. The contours show the strong east–west mean annual precipitation (MAP) gradient. The nine sites were sited in representative land covers (see Table 1): six on grassland of varying thickness, two on winter wheat, one on bare ground, and one on shrubland. The surface temperature of the dry (MAP = 550 mm), sparsely vegetated sites (1–3) is mainly linked to the soil moisture. In contrast, the green, lush vegetation of the wet sites (7–9) (MAP = 900 mm) controls the surface temperature. In sites 4–6 (MAP = 750 mm), a mix of winter wheat and grassland, the surface temperature is influenced by both soil moisture and vegetation.

soil evaporation, canopy evaporation, and transpiration more responsive to environmental conditions than STD. Unlike *Dickinson et al.* [1998], we parameterized the effect of water stress on stomatal conductance as a function of soil moisture deficit, not as a function of soil matric potential.

3.2. IHOP_2002 Sites and Data Sets

[18] We used data sets from the IHOP_2002 field campaign (available at http://www.rap.ucar.edu/research/land/observations/ihop/) [*LeMone et al.*, 2007] to evaluate the three versions of the Noah LSM at nine sites located along the Kansas-Oklahoma border and in northern Texas (Figure 1). The nine stations were sited to obtain a representative sample of the region, which spans a strong west–east (east–west) gradient of rainfall (topography and the Bowen ratio). We used 45 days of high-frequency, multisensor measurements of meteorological forcing, surface-to-atmosphere fluxes, and near-surface soil moisture and temperature. Site characteristics, soil and vegetation classes, mean meteorological values, average heat fluxes and near-surface states for the observation period are summarized in Table 1.

3.3. Model Initialization and Spin-up

[19] Following *Rodell et al.* [2005], we initialized each of the four soil layers at 50% saturation and at the multiannual mean temperature. To drive the spin-up (between 1 January 2000 and 13 May 2002), we used downscaled North American Land Data Assimilation System [*Cosgrove et al.*, 2003] meteorological forcing, interpolated from a 60 min to a 30 min time step. The models were subsequently driven by IHOP_2002 meteorological forcing from 13 May 2002 to 25 June 2002 (day of year 130 to 176). For GW, water table depth was initialized assuming equilibrium of gravitational and capillary forces in the soil profile [*Niu et al.*, 2007].

3.4. Evaluation Data Sets

[20] To evaluate the models, we used sensible heat flux (H), latent heat flux (LE), ground heat flux (G), ground temperature (Tg), and first layer soil moisture (SMC_{5cm}). All data was recorded at a 30 min time step. In situ, high-frequency flux and near-surface state measurements are an integrated response of the land surface and therefore provide useful data for examining model soundness at a specific location [*Bastidas et al.*, 2001; *Stöckli et al.*, 2008]. To assess model performance, we used the root mean square error (RMSE) measure.

3.5. Parameters Considered in the Sensitivity Analysis

[21] We studied all 10 soil and 10 vegetation parameters of STD, assigned a priori via look-up tables. We included eight parameters responsible for the phenology module and four that control the groundwater module to analyze a total of 28 and 24 parameters for DV and GW, respectively. All other coefficients in the models were kept constant at the recommended values. Default values and feasible ranges (Table 2) for all parameters were taken from the literature [e.g., *Chen and Dudhia*, 2001; *Hogue et al.*, 2006].

3.6. Sobol′ Indices for Global Variance-Based Sensitivity Analysis

[22] We used the variance-based method of *Sobol′* [1993, 2001] to efficiently identify the factors that contribute most to the variance of a model's response. The method of Sobol′ deals explicitly with parameter interaction and has recently been used to quantify model sensitivity and parameter

Table 1. Average Meteorology, Near-Surface States, and Turbulent Fluxes Observed During the Calibration Period at the Nine IHOP_2002 Sites[a]

	Site								
	1	2	3	4	5	6	7	8	9
Lat (°N)	36.4728	36.6221	36.8610	37.3579	37.3781	37.3545	37.3132	37.4070	37.4103
Lon (°W)	100.6179	100.6270	100.5945	98.2447	98.1636	97.6533	96.9387	96.7656	96.5671
Vegetation type	bare ground (1)	grassland (7)	sagebrush (9)	pasture (7)	wheat (12)	wheat (12)	pasture (7)	grassland (7)	pasture (7)
Soil texture	sandy clay loam (7)	sandy clay loam (7)	sandy loam (4)	loam (8)	loam (8)	clay loam (6)	silty clay loam (2)	silty clay loam (2)	silty clay loam (2)
Rain (mm)	154.5	69.1	72.4	164.5	173.6	203.6	175.4	296.6	250.8
MAP (mm)	530	540	560	740	750	800	900	880	900
Elevation (m)	872	859	780	509	506	417	382	430	447
Ta (°C)	21.4	21.7	22.5	20.7	20.7	21.0	20.7	20.1	19.9
β	1.08	0.92	1.11	0.41	0.46	0.63	0.20	0.14	0.24
H (W m^{-2})	70.5	70.7	75.7	43.9	51.9	61.4	25.9	17.1	27.9
LE (W m^{-2})	65.1	76.1	68.2	106.2	111.2	97.1	126.4	122.8	115.3
G (W m^{-2})	−10.4	−6.4	−9.3	−2.7	−5.1	−7.5	−5.6	−12.1	−10.5
Tg (°C)	24.1	24.1	25.8	23.2	21.9	22.9	22.3	22.4	22.7
SMC_{5cm} (%)	15.4	18.0	7.0	18.0	18.1	19.0	33.2	32.8	34.0

[a]Calibration period is 13 May to 25 June. See Figure 1. Indices of vegetation types and soil classes are in parenthesis. Rainfall is cumulative over the observation period. Dry, sparsely vegetated sites (1–3) receive almost half of the amount of mean annual precipitation (MAP) than wet sites (7–9), with lush vegetation. Mean 2 m air temperature (Ta), Bowen ratio (β), sensible (H), latent (LE) and ground (G) heat flux, ground temperature (Tg) and soil moisture content at 5 cm (SMC_{5cm}).

interactions in hydrology [e.g., *Tang et al.*, 2006; *Bois et al.*, 2008; *Ratto et al.*, 2007; *Yatheendradas et al.*, 2008; *van Werkhoven et al.*, 2008]. Our review of the literature shows that it has not yet been used for LSM SA.

[23] Sobol′ indices enable researchers to distinguish the subset of independent input factors (such as the model parameters) $X = \{x_1, \ldots, x_i, \ldots, x_k\}$ that account for most of the variance of the model's response $Y = f(X)$ either by themselves (first-order) or due to interaction with other parameters (higher-order). For completeness, we briefly summarize the Monte Carlo-based scheme presented by

Table 2. Feasible Ranges of Noah-LSM Parameters Considered in the Sensitivity Analysis

Parameter	Description	Units	Min	Max
Soil parameters				
maxsmc	Maximum volumetric soil moisture	$m^3\ m^{-3}$	0.35	0.55
psisat	Saturated soil matric potential	$m\ m^{-1}$	0.1	0.65
satdk	Saturated soil hydraulic conductivity	$m\ s^{-1}$	1E-6	1E-5
b	Clapp-Hornberger *b* parameter	-	4	10
quartz	Quartz content	-	0.1	0.82
refdk	Used with refkdt to compute runoff parameter kdt		0.05	3
fxexp	Bare soil evaporation exponent	-	0.2	4
refkdt	Surface runoff parameter		0.1	10
czil	Zilintikevich parameter	-	0.05	8
csoil	Soil heat capacity	$J\ m^{-3}\ K^{-1}$	1.26	3.5
Vegetation parameters				
vrcmin	Minimal stomatal resistance	$s\ m^{-1}$	40	400
rgl	Radiation stress parameter used in F1 term of canopy resistance		30	100
hs	Coefficient of vapor pressure deficit term F2 in canopy resistance		36	47
z0	Roughness length	m	0.01	0.1
lai	Leaf area index	-	0.1	5
cfactr	Exponent in canopy water evaporation function	-	0.4	0.95
cmcmax	Maximum canopy water capacity used in canopy evaporation	m	0.1	2.0
sbeta	Used to compute canopy effect on ground heat flux	-	−4	−1
rsmax	Maximum stomatal resistance	$s\ m^{-1}$	2,000	10,000
topt	Optimum air temperature for transpiration	K	293	303
Dynamic phenology parameters (Noah-DV only)				
fragr	Fraction of carbon into growth respiration	-	0.1	0.5
gl	Conversion between greenness fraction and LAI	-	0.1	1.0
rssoil	Soil respiration coefficient	$s^{-1} \times 10^{-6}$	0.005	0.5
tauhf	Average inverse optical depth for 1/e decay of light	-	0.1	0.4
bf	Parameter for present wood allocation		0.4	1.3
wstrc	Water stress parameter		10	400
xlaimin	Minimum leaf area index	-	0.05	0.5
sla	Specific leaf area	-	5	70
Groundwater parameters (Noah-GW only)				
rous	Specific yield	$m^3\ m^{-3}$	0.01	0.5
fff	*e*-folding depth of saturated hydraulic capacity	m^{-1}	0.5	10
fsatmx	Maximum saturated fraction	%	0	90
rsbmx	Maximum rate of subsurface runoff	$m\ s^{-1} \times 10^{-3}$	0.01	1

Saltelli [2002] to compute first-order and total Sobol′ sensitivity indices.

[24] The first-order sensitivity index (S_i) represents a measure of the sensitivity of $Y = f(x_1, x_2, \ldots, x_k)$ (the RMSE of a model realization evaluated against observations in our case) to variations in the parameter (or factor) x_i. S_i is defined as the ratio of the variance of Y conditioned on the ith factor (V_i) to the total unconditional variance (V):

$$S_i = \frac{V_i}{V(Y)} = \frac{V(E(Y|x_i))}{V(Y)} = \frac{\hat{U}_i - \hat{E}^2(Y)}{\hat{V}(Y)} \quad (1)$$

where

$$\hat{U}_i = \frac{1}{n-1}\sum_{r=1}^{n} f(x_{r1}, x_{r2}, \ldots, x_{rk}) \cdot f\left(x'_{r1}, x'_{r2}, \ldots, x'_{r(i-1)}, x_{ri}, x'_{r(i+1)}, \ldots, x'_{rk}\right), \quad (2)$$

is obtained from products of values of f computed from the sample matrix (n model realizations long) times values of f computed from another n realizations matrix where all k parameters except x_i are resampled.

[25] The estimates of the mean squared and the total variance are computed as

$$\hat{E}^2(Y) = \frac{1}{n}\sum_{r=1}^{n} f(x_{r1}, x_{r2}, \ldots, x_{rk}) f\left(x'_{r1}, x'_{r2}, \ldots, x'_{rk}\right) \quad (3)$$

$$\hat{V}(Y) = \frac{1}{n}\sum_{r=1}^{n} f(x_{r1}, x_{r2}, \ldots, x_{rk})^2 - \hat{E}^2(Y). \quad (4)$$

[26] Instead of computing all $2^k - 1$ terms of the variance decomposition

$$V(Y) = \sum_{i}^{k} V_i + \sum_{i}\sum_{j>i} V_{ij} + \ldots.. + V_{12..k}, \quad (5)$$

(which would require as many as $n2^k$ model runs), in addition to estimating S_i, it is customary to estimate only the total sensitivity index (S_{Ti}) associated with parameter x_i. S_{Ti} encompasses the effect that of all the terms in the variance decomposition that include the factor x_i have on the variance of the model's response. S_{Ti} is estimated by the difference between the global unconditional variance of Y and the total contribution to the variance of Y that is caused by factors other than x_i, divided by the unconditional variance:

$$S_{Ti} = \frac{V(Y) - V(E(Y|x_{-i}))}{V(Y)} = 1 - \frac{\hat{U}_{-i} - \hat{E}^2(Y)}{\hat{V}(Y)}, \quad (6)$$

where

$$\hat{U}_{-i} = \frac{1}{n-1}\sum_{r=1}^{n} f(x_{r1}, x_{r2}, \ldots, x_{rk}) \cdot f\left(x_{r1}, x_{r2}, \ldots x_{r(i-1)}, x'_{ri}, x_{r(i+1)}, \ldots x_{rk}\right) \quad (7)$$

is obtained from products of values of f computed from the sample matrix times the values of f computed from another matrix where only x_i is resampled.

[27] A significant difference between S_{Ti} and S_i points to an important role of the interactions of the ith factor (at all orders) in affecting Y [*Saltelli et al.*, 2006]. Identification of such parameter interactions can help guide model development. S_{Ti} are also useful to identify input factors that are noninfluential, which can help reduce the dimensionality of the parameter estimation problem. If an S_{Ti} is negligible, then it is reasonable to fix that factor to a reasonable value within its range of uncertainty, and the dimensionality of the space of input factors or model parameters can be reduced accordingly [*van Werkhoven et al.*, 2009].

3.7. Sampling Strategies for Sensitivity Analysis

[28] We generated samples of model parameters using Latin hypercube sampling (LH) and of the behavioral parameter sets through multiobjective calibration.

3.7.1. Latin Hypercube Monte Carlo Sampling

[29] We ran a total of 405,000 Monte Carlo simulations sampling random parameter sets (15,000 samples for each model and site) to obtain a representation of the range of model responses that was sufficiently detailed yet that also balanced computational constraints. We assumed that our conservative sampling was adequate. We used LH because it combines the strengths of stratified and random sampling to ensure that all regions of the parameter space are represented in the sample [*McKay et al.*, 1979; *Helton and Davis*, 2003]. LH divides each parameter range into disjoint intervals of equal probability. From each hypercube, one sample value is randomly taken. We sampled uniformly within feasible bounds (Table 2). For each sample, we recorded the RMSE of 5 criteria: H, LE, G, Tg, and SMC_{5cm}.To create all the matrices involved in the computation of the Sobol′ indices, we used a modified LH that enables replication [*Tang et al.*, 2007].

3.7.2. Multiobjective Markov Chain Monte Carlo Parameter Estimation Technique

[30] We used the efficient Markov Chain Monte Carlo sampling strategy of *Vrugt et al.* [2003] to approximate the joint posterior distribution of optimal parameters. The simultaneous minimization of the RMSE of multiple criteria {H, LE, G, Tg, SMC_{5cm}} allowed us to constrain the models to be consistent with several types of observations and facilitated the identification of the underlying posterior distribution of physically meaningful behavioral parameter sets. It was assumed that sets from the posterior distribution cause the model to mimic the processes it was designed to represent [*Gupta et al.*, 1999; *Bastidas et al.*, 2001; *Leplastrier et al.*, 2002; *Hogue et al.*, 2006]. The calibration algorithm runs, in parallel, multiple chains of evolving parameter distributions to provide a robust exploration of the parameter space. These chains communicate with each other through an external population of points, which are used to continuously update the size and shape of the proposal distribution in each chain. This procedure allows an initial population of parameter sets (uniformly sampled within preestablished, feasible ranges) to converge to a stationary sample, which maximizes the likelihood function and fairly approximates the Pareto set (PS). The Pareto set represents the multiobjective trade-off: no member of the PS

can perform better with respect to one objective without simultaneously performing worse with respect to another competing objective [*Gupta et al.*, 1998]. We used a sample of 150 parameter sets to represent the posterior distribution of "behavioral" parameter sets.

3.8. Hierarchical Clustering for Comparisons of Parameter Distributions

[31] Unsupervised classification of the behavioral parameter distributions allowed us to understand data similarities across locations, with specific focus on the relationships between types of parameters and sites. We used a clustering strategy to classify the marginal posterior distributions of calibrated parameters sets into groups. Agglomerative hierarchical clustering methods start with n groups (one object per group) and successively merge the two most similar groups until a single group is left. We used MATLAB's complete linkage algorithm to implement the clustering, in which the maximum distance between objects, one coming from each cluster, represents the smallest sphere that can enclose all objects in the two groups within a single cluster [*Hair et al.*, 1995]. Because the distance measures (e.g., Manhattan, Euclidean) used to measure dissimilarity between observations may influence the membership of samples to groups, we used the cophenetic correlation coefficient to assess the quality of the linkage [*Martinez and Martinez*, 2002]. We used dendrograms to visualize the links between the objects as inverted U-shaped lines, whose height represents the distance between the objects.

4. Which Parameters Are Sensitive?

[32] VSA showed that there are only a few parameters that, by themselves, exert significant influence on the model predictions. In contrast, parameter interaction dominates and is hence the principal mechanism for sensitivity. Figures 2, 3, and 4 present, for all sites, all considered parameters, and all models, the Sobol′ first-order sensitivity indexes (S_i, which is the fraction of the total variance of RMSE that can be solely attributed to the ith parameter) and the residual between Sobol′'s total and first-order sensitivity index ($S_{\mathrm{Ti}} - S_i$, which is the fraction of total variance that results from the interaction of the ith parameter with other parameters at all orders). When the influence of parameters changed as we would physically expect, we interpret the results as consistent with our hypothesis that, to a first order, a model adequately represents the site-to-site variation in the water and energy cycles. Site-to-site variation in the most sensitive parameters is not chiefly governed by soil or vegetation type but, similar to other studies [e.g., *Liang and Guo*, 2003; *Demaria et al.*, 2007; *van Werkhoven et al.*, 2008], appears to be of secondary importance when compared to the influence of the predominant climatic gradient. Although we cannot rule out the potential importance of other east–west gradients (e.g., the topographic or hydrogeologic gradient), in section 4.1 we provide explanations for the observed patterns that are consistent with the climatological change between sites.

4.1. First-Order Sensitivity (S_i)

[33] For several key parameters, a pattern of first-order sensitivity can be linked to the hydrology of the sites. For most sites and models, the greatest first-order control on simulated top layer soil moisture is porosity (maxsmc) (Figure 4a). At dry sites 1–3, where direct evaporation is presumably a major component of LE flux, for STD and GW, the bare soil evaporation exponent (fxexp) exerts the highest first-order control on soil moisture. The LE flux simulated by GW at dry sites is controlled by fxexp and specific yield (rous), which partially controls the depth to the water table. Parameter lai directly controls transpiration and hence the surface energy budget; at the most vegetated sites (7–9), lai consequently shapes most of the variance of H and LE for both STD and GW (Figures 2a and 3a). The initial value of lai is not important to DV's simulated H and LE because DV allows lai to change over time. Instead, minimum stomatal resistance (rcmin) exerts the most control on DV-simulated LE. Two new parameters associated with DV, gl and sla, which control the calculation of lai, also exert first-order control on the simulated energy fluxes. In the sparsely vegetated sites (1–3), the Zilintikevich coefficient (czil) plays a significant role in the variance of H.

[34] The specific parameters that control model variance change between models and between sites. In STD, as the mean annual precipitation (MAP) increases, fxexp becomes less important to top layer soil moisture ($\mathrm{SMC_{5cm}}$) and refkdt, a parameter involved in determining maximum rates of infiltration, becomes more important (Figure 4a). This pattern changes for GW, in which surface runoff is relatively deemphasized and subsurface runoff is relatively emphasized (see discussion about GW's preferred modes of operation, section 5). In GW, although fxexp still exerts first-order control on $\mathrm{SMC_{5cm}}$ at dry sites, refkdt has little direct influence on $\mathrm{SMC_{5cm}}$ at wet sites. The most sensitive parameter for $\mathrm{SMC_{5cm}}$ at sites 1–3 is rous, which controls whether aquifer water is accessible to the near-surface soil. Consistent with our expectations, soil suction (psisat), which in GW controls upward movement of water from the aquifer to the soil, has significant control on $\mathrm{SMC_{5cm}}$ within GW but not within STD, in which psisat plays a less dominant role in shaping soil hydraulic behavior (Figure 4a).

[35] Especially in the case of STD and DV, as sites get wetter, the surface exchange coefficient czil exerts progressively less influence and rcmin progressively more influence on H (Figure 2a). The shift is consistent with our expectation that at more vegetated sites, stomatal resistance should be important to determining the surface energy balance. As a site's MAP increases, rcmin and lai increasingly shape simulated LE, and fxexp becomes less influential (Figure 3a). Even at the dry sites (1–3), DV favors larger values of vegetation fraction (shdfac) than are prescribed by STD and GW. As a consequence, DV stands apart from GW and STD in that fxexp does not directly contribute to variance of any objective at the three driest sites (with the exception of unvegetated site 1, at which LE is controlled by fxexp).

[36] Examinations of S_i that are not in line with expectations may be used to help modelers diagnose likely problems with conceptualization, forcing data, and/or model structure. For instance, in STD, fxexp has the highest S_i of simulated H and LE at site 6. We do not expect direct evaporation to be a relatively more significant component of the LE flux at site 6 than at climatically similar sites 4 and 5 or at the semiarid sites 1–3. The discrepancy implies that

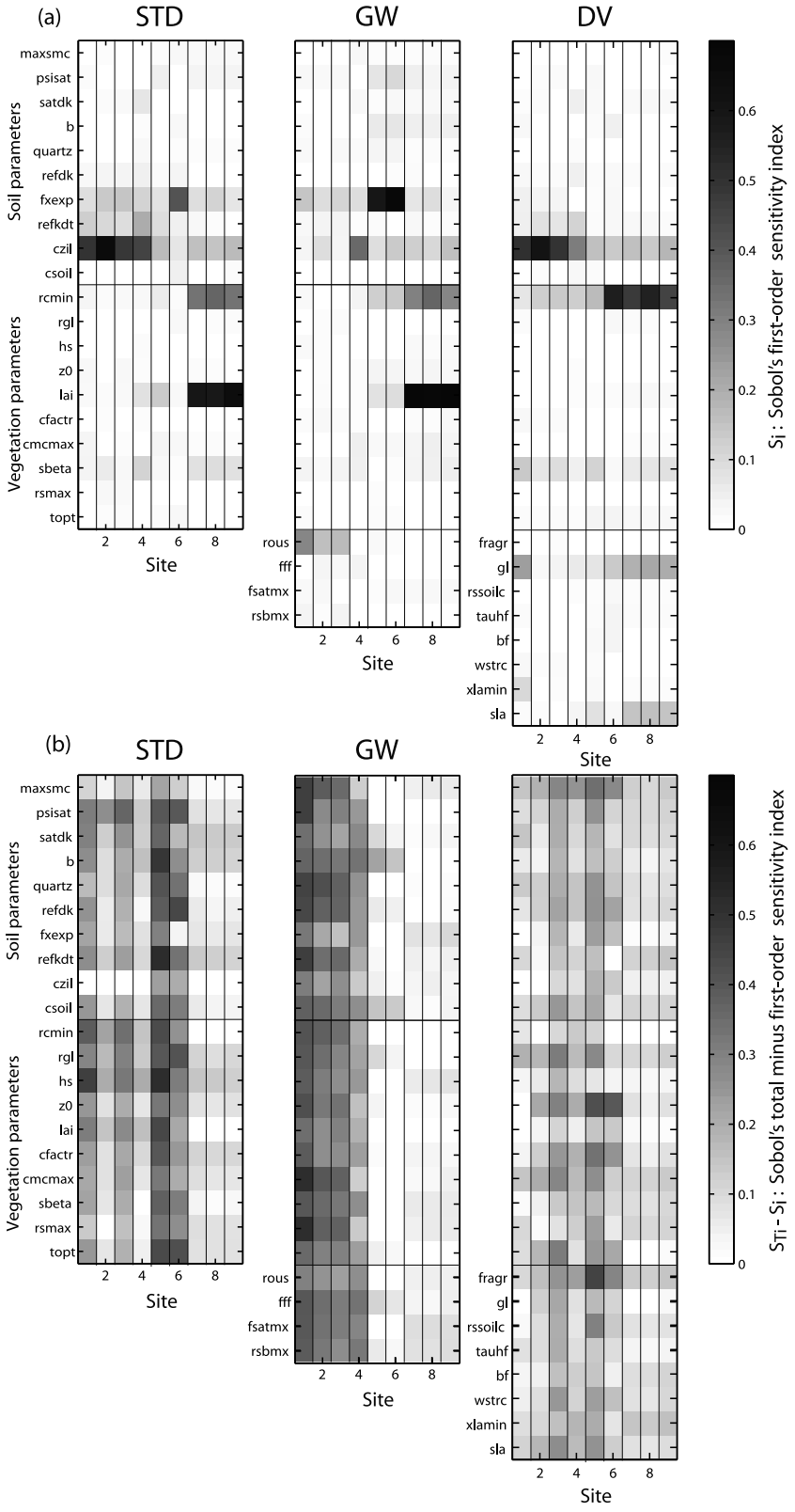

Figure 2. (a) First-order Sobol′ sensitivity indices for the parameters of STD, GW, and DV at all sites. S_i stands for the individual contribution of a parameter to the variance of the RMSE of H. (b) Difference between Sobol′'s total sensitivity index and S_i. S_{Ti} - S_i is the contribution to the variance through interactions with other parameters. Parameters grouped by soil and vegetation. Regional sensitivity patterns from semiarid (MAP = 550 mm), sparsely vegetated sites (1–3) to semihumid (MAP = 900 mm) sites (7–9) with green, lush vegetation, are easily distinguishable.

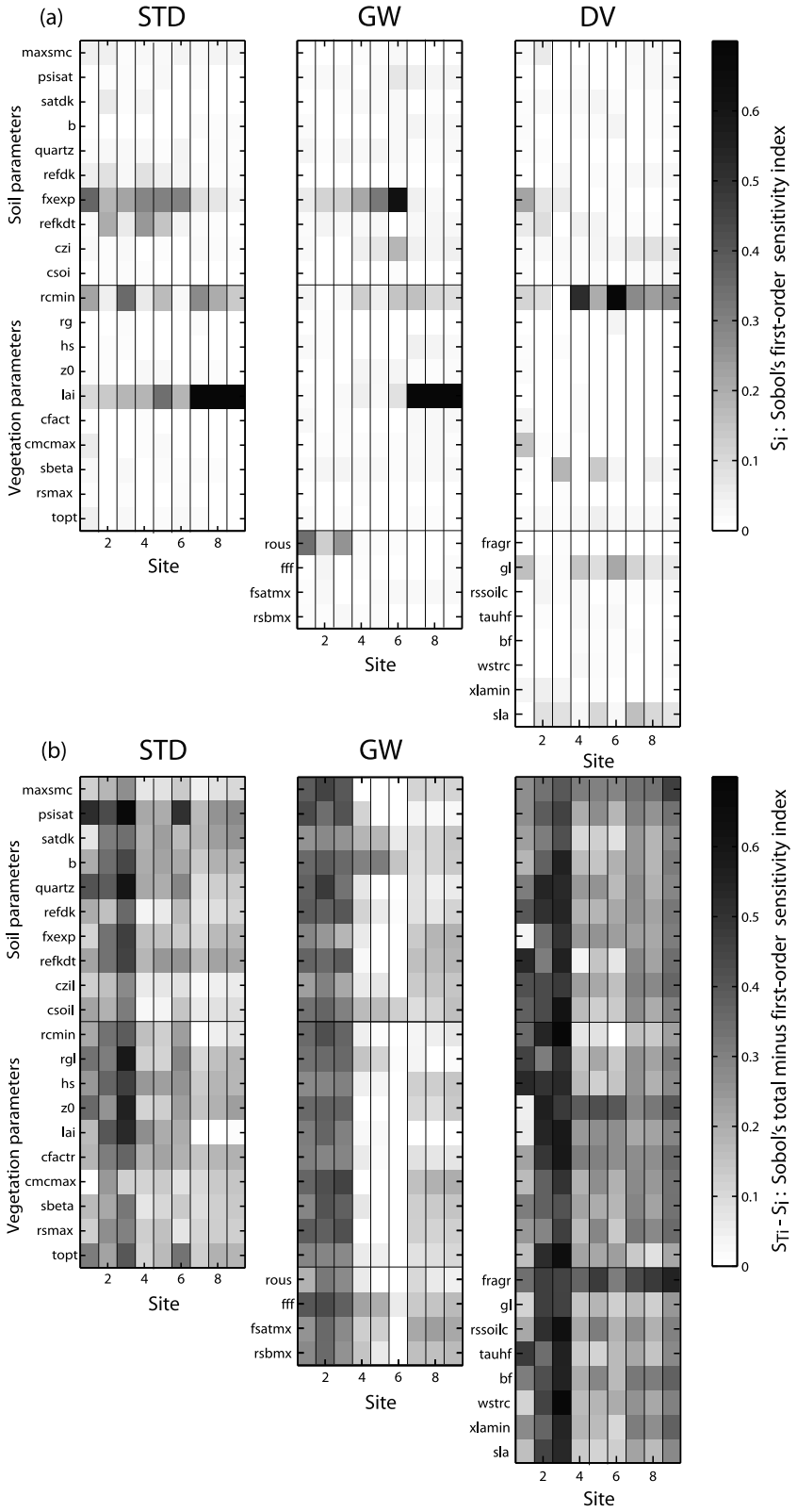

Figure 3. Same as Figure 2 but for LE.

either our conceptual understanding of the physical processes at site 6 is incorrect, that the model does not adequately represent the physical processes, and/or that our forcing and/or evaluation data are faulty at one or more of the sites.

4.2. Sensitivity Through Interactions (S_i–S_{Ti})

[37] Interactions between parameters are responsible for most of the variance in the models' predicted H, LE, and SMC_{5cm} (Figures 2b, 3b, and 4b). If we assume that the parameterizations are correct, then the significant parameter

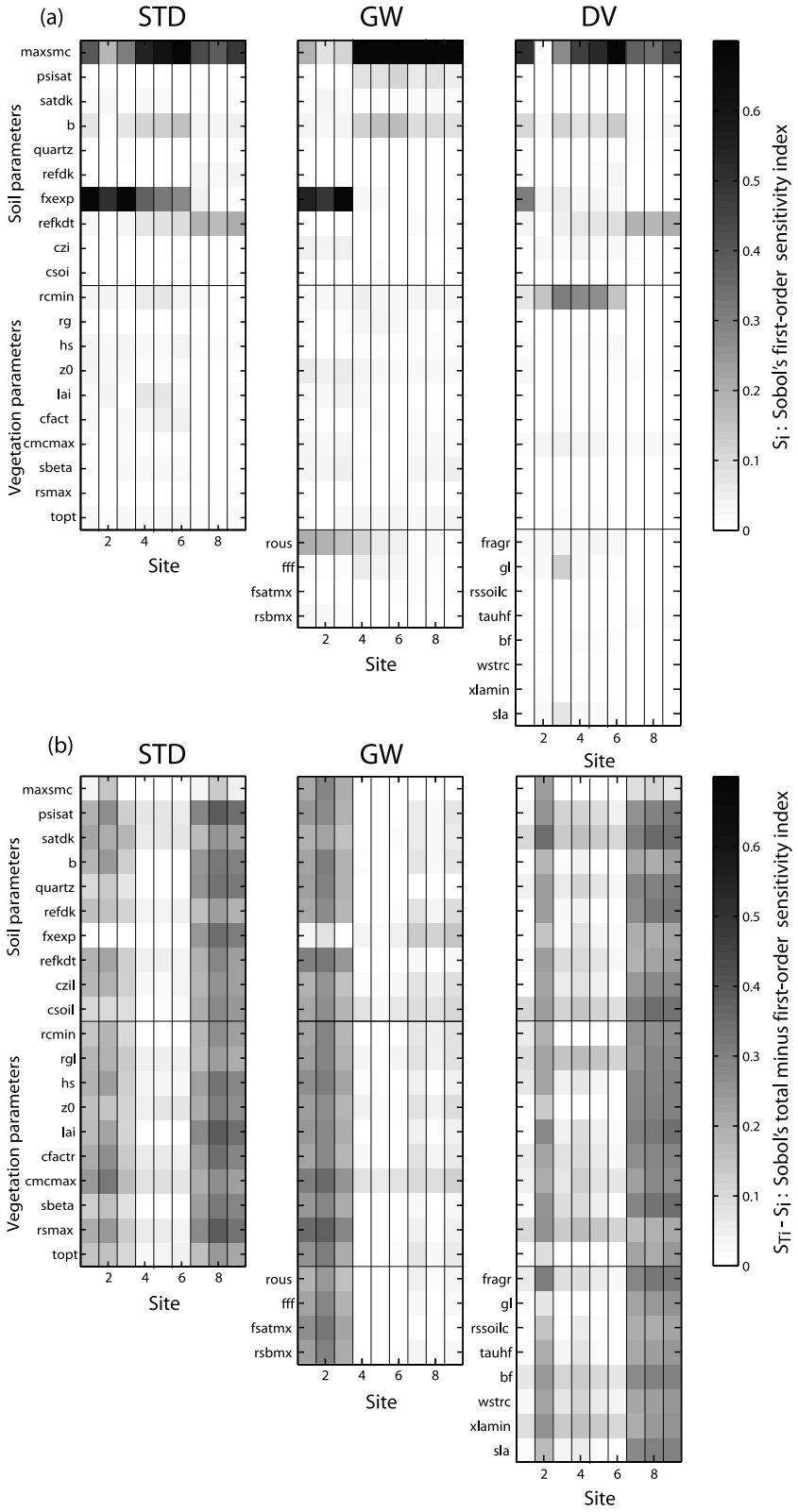

Figure 4. Same as Figure 2 but for SMC_{5cm}.

interaction indicates model overparameterization [*Saltelli et al.*, 2008; *Bastidas et al.*, 2006; *Yatheendradas et al.*, 2008]. Arguably, it is also possible that the observed parameter interaction results from models that are either too simplistic and/or incorrect. Although parameter interaction may not be an inherently negative trait (e.g., in porous media, we expect hydraulic conductivity and porosity to be functionally related), when there are no known functional relationships between the physical quantities that two parameters represent, interaction is likely to be indication that the model

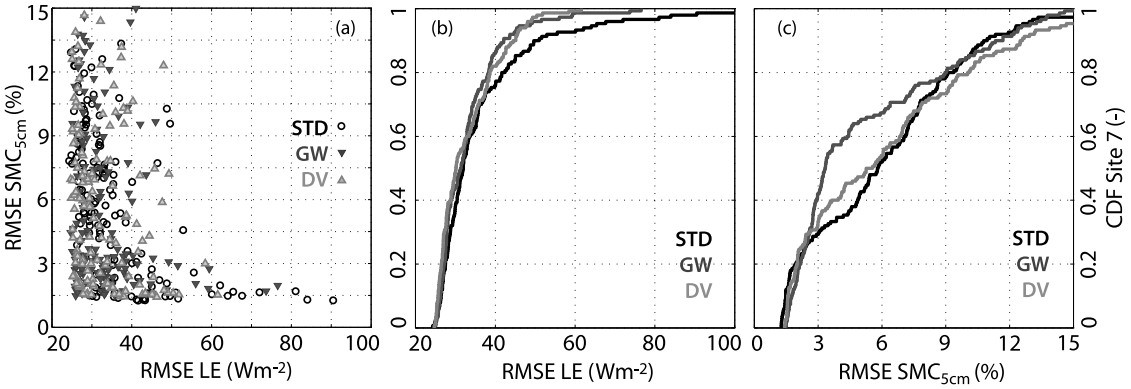

Figure 5. Trade-off LE-SMC_{5cm} and cumulative distribution functions (CDF) of scores of behavioral STD, GW, and DV at site 7. (a) Scatterplot in objective function space of parameter sets that maximize the likelihood function after multiobjective calibration against {H, LE, G, Tg, SMC_{5cm}}. CDF of root-mean-square errors (RMSE) of behavioral runs evaluated against observed (b) LE and (c) SMC_{5cm}. GW (dark gray) and DV (light gray) perform as good as or better than STD (black).

works in a way that is not consistent with the conceptual model from which the parameterizations were built.

[38] All models exhibit the most parameter interaction at the driest sites, consistent with the findings of *Liang and Guo* [2003] and suggesting the need to revise the formulation of all three models for semiarid regions [*Hogue et al.*, 2005; *Rosero and Bastidas*, 2007]. Especially for H and SMC, GW reduces parameter interaction at the middling moisture (4–6) and semihumid sites (7–9) (e.g., Figure 5b). GW's reduction of parameter interaction is evidence (although by no means conclusive) that GW is more realistic than STD at sites 4–9. This result is consistent with foregoing observations on the robustness of GW [*Gulden et al.*, 2007]. Conversely, GW appears to increase parameter interaction at the driest sites (1–3), indicating STD better represents semiarid processes than GW. DV parameters are much more interactive than those of STD and GW, especially at the wettest sites when simulating LE and SMC_{5cm}. The increased interaction between the DV-specific parameters and the rest of the conceptually unrelated STD parameters suggests DV is not functioning as its developers intended. The significant parameter interaction is consistent with the poor robustness of DV [*Rosero et al.*, 2009].

[39] Looked at in full, the models best represent the surface water and energy balances at the intermediate moisture and wet sites, where parameter interaction tends, within a given model, to be lowest. Because it reduces parameter interaction, GW is most likely of any of the three models to be representing the key physical processes with the most realism.

5. How Do Sensitive Parameters Interact and Shape Model Behavior? Case Study at Site 7

[40] Toward our objective of thoroughly evaluating the physical realism of the three models presented, we performed a case study in which sensitivity analysis linked model identification and model development. We followed the impact of shifted preferred values of three physically meaningful parameters that made considerable contributions to the output variance: porosity (smcmax), the muting factor for vegetation's effect on thermal conductivity (sbeta), and minimum stomatal resistance (rcmin). We examined the model structures at site 7 because at that site STD, DV, and GW show nearly equivalent performance when using their behavioral parameter sets (Figure 5). Such "equifinality" occurs frequently in hydrologic modeling [*Beven and Freer*, 2001]. When equifinality occurs, distinguishing a "best" model is not trivial. It requires us not only to confront the simulations with observed behavior to test for consistency [*Rosero et al.*, 2009] but also to understand the underlying model structures (the relationship between parameters) that make the models perform equally well. We show below how sensitivity analysis offers the power and the ability to discriminate between the structures of STD, GW and DV that do and do not conform to our physical understanding of the systems.

5.1. Relationships Between Behavioral Parameters Change Between Models

[41] The models have distinctly different optimal parameter distributions for the same physical parameters (Figure 6), implying not only that the parameters cannot be transferred between models but that the relationships between them are different. Even the direction of "sensitivity" (understood as the rate of change of RMSE with parameter value along the range of possible values of the parameter) changes between models (e.g., Figure 6a). The simulation of SMC_{5cm} by STD and DV degrades as porosity increases, while GW improves. We also note that, along the possible range, the response can be enhanced (Figure 6d) or become relatively insensitive to changes in parameter value (Figure 6c). The identifiability of parameters (when parameters have a clearly defined local minimum) changes between models. For example, in DV, there is a clear low point of the RMSE of LE along the range of values of the maximum water-holding capacity of the canopy (cmcmax), but STD and GW have less of a preference (Figure 6c). The interquartile range of rcmin of STD is smaller than that of GW or DV (Figure 6b). The fundamental implication of our observations is that although the different optimal values of parameters are important (as found during model identification), the change in the functional relationship between the parameters (the information contained in

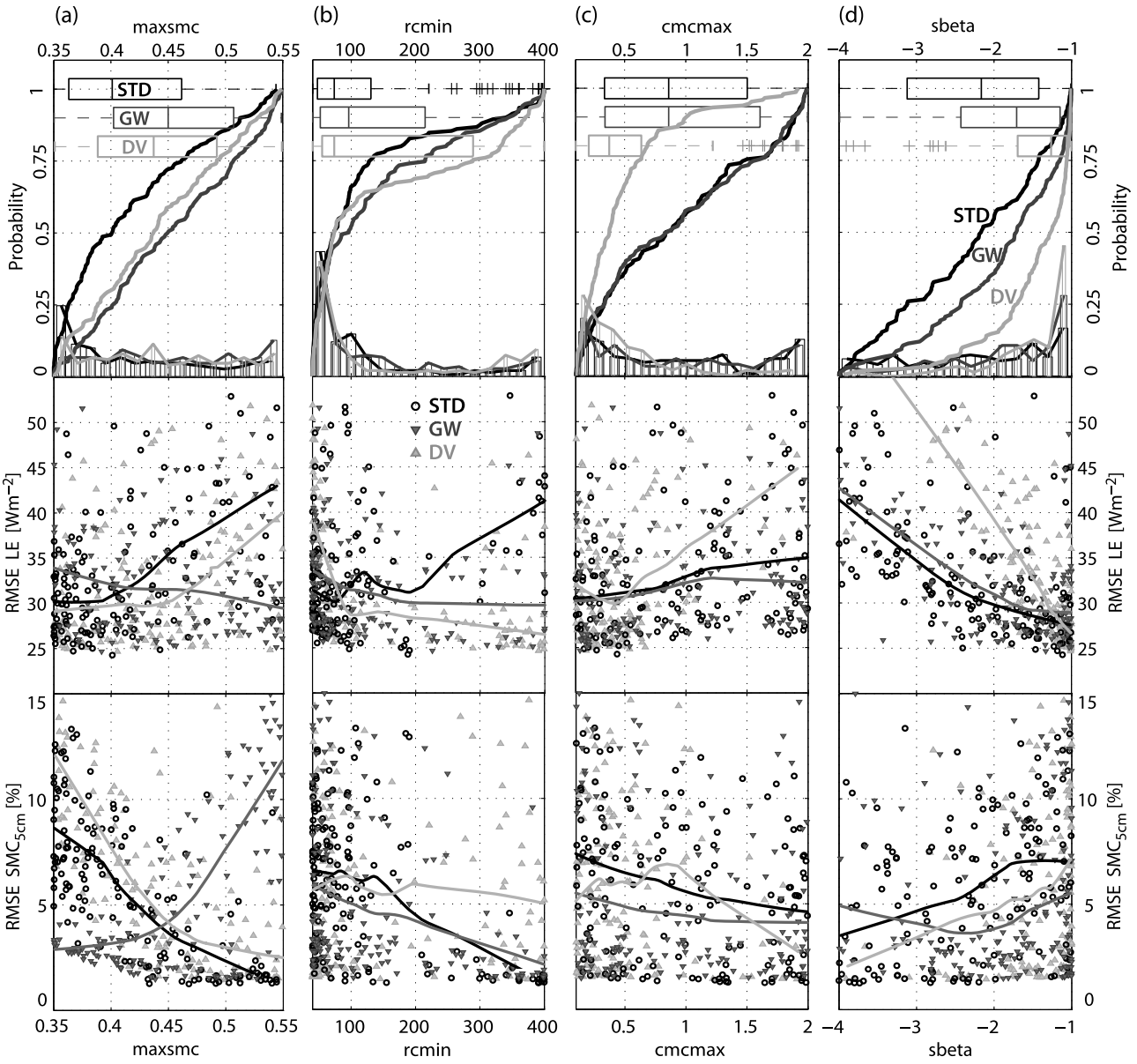

Figure 6. Marginal cumulative distribution functions (CDF) of the posterior distribution of selected behavioral parameter sets at site 7. (a) Porosity (maxsmc), (b) minimum stomatal resistance (rcmin), (c) maximum water holding capacity of the canopy (cmcmax), and (d) effect of the vegetation on ground heat flux (sbeta). Along with the CDFs, the histograms and interquatile ranges are shown. The trend in the scatterplots of RMSE of LE and SMC_{5cm} is shown by fitting a minimum complexity polynomial. Note that GW (dark gray), DV (light gray), and STD (black) are shown.

the interactions) is most relevant for purposes of model development.

5.1.1. Role of Porosity (maxsmc)

[42] In all three versions of the Noah LSM, higher values of maxsmc tend to decrease direct evaporation from the first soil layer (E_{dir}). E_{dir} is estimated as the product of Penman's potential evaporation (ET_{pot}), the complement of the vegetated fraction (shdfac), and the ratio of top layer volumetric soil moisture (SMC_1) to maxsmc:

$$E_{dir} = ET_{pot}(1 - \text{shdfac})\left(\frac{SMC_1 - SMC_{dry}}{\text{maxsmc} - SMC_{dry}}\right)^{\text{fx exp}}. \quad (8)$$

SMC_{dry} is the lowest possible volumetric water content of the topsoil layer, and fxexp is a parameter ranging from 0.2 to 4.

[43] In both STD and DV, the error in simulated LE tends to be relatively small when maxsmc is low and relatively large when maxsmc is high (Figure 6a). However, GW better simulates LE as maxsmc increases. The tendency of STD and DV to simulate LE well when maxsmc is low (and direct evaporation from the soil consequently tends to be high) implies that STD and DV often underestimate direct evaporation at site 7. The tendency of STD to underestimate direct evaporation was also suggested by *Peters-Lidard et al.* [2008], who improved results by changing the value of

fxexp from 2 to 1. Given the same maxsmc, GW more easily simulates sufficient direct evaporation, perhaps because of wetter soil [*Rosero et al.*, 2009].

[44] In STD and DV, parameter maxsmc controls both surface and subsurface runoff. Hydraulic conductivity (wcnd) is computed by scaling saturated hydraulic conductivity (satdk) by wetness (SMC/maxsmc), raised to an exponent containing the Clapp and Hornberger parameter (b):

$$\text{wcnd} = \text{dksat}\left(\frac{\text{SMC}}{\text{maxsmc}}\right)^{2b+3}. \quad (9)$$

Lower maxsmc yields higher wcnd, which means water moves through the soil more quickly. For subsurface runoff (Runoff2), wcnd controls lateral water movement through the soil. In STD and DV, Runoff2 is wcnd times the slope of the grid cell. Consequently, higher maxsmc decreases Runoff2. Higher maxsmc also decreases surface runoff (Runoff1) by increasing the maximum rate of infiltration. Both changes increase soil wetness.

[45] GW changes the way runoff is computed; maxsmc does not control surface or subsurface runoff in GW, which eliminates two of the three ways that maxsmc controls soil moisture. Runoff2 is represented as an exponential function of depth to water [*Niu et al.*, 2007]:

$$\text{Runoff2} = \text{rsbmx}\, e^{-\text{fff}*Z_{WT}}, \quad (10)$$

where rsbmx is the maximum rate of subsurface runoff, fff is the e-folding depth of saturated hydraulic conductivity, and Z_{WT} is the depth to the water table, which is computed by the model. Runoff1 is computed using a version of the function used to compute Runoff2 [*Niu et al.*, 2005]:

$$\text{Runoff1} = \text{pcpdrp}*\left(\text{fsatmx}\, e^{-0.5*\text{fff}*Z_{WT}}\right), \quad (11)$$

where pcpdrp is the effective incident water and the second term is the fraction of unfrozen grid cell that is saturated.

[46] In STD (and DV), maxsmc couples two physically unrelated (or very weakly related) processes (direct soil evaporation and lateral surface and subsurface runoff). GW decouples these processes by eliminating the dependence of parameterized lateral runoff on maxsmc. This decoupling reduces the spurious parameter interaction of maxsmc and, within GW, nearly eliminates the trade-off between good simulation of LE and $\text{SMC}_{5\text{cm}}$. GW is, in this regard, a better model for simulating fluxes at site 7.

[47] The question remains – why does GW poorly simulate $\text{SMC}_{5\text{cm}}$ when maxsmc increases? maxsmc is used to compute vertical hydraulic conductivity (using the same function as STD). GW uses vertical hydraulic conductivity to regulate the flow of water between the aquifer and soil down a hydraulic gradient. Higher maxsmc yields lower hydraulic conductivity, which, in addition to decreasing the transfer of water between layers within the soil column, decreases the communication between the aquifer and the soil profile (that is, it decreases the flow of water between the two, increasing the potential for water to be retained near the surface). At site 7, GW best simulates $\text{SMC}_{5\text{cm}}$ when high vertical hydraulic conductivity connects the aquifer and soil.

[48] Consistent with the work of others [e.g., *Demaria et al.*, 2007], parameter values and model sensitivity to maxsmc are not consistent between sites along a climatic gradient or even within a set of sites with similar characteristics. Conclusions about model performance are therefore difficult to generalize. This lack of continuity of behavior between sites is consistent with at least one of the following possibilities: (1) model parameterizations do not represent key aspects of the system and/or (2) our multiobjective calibration provided insufficient constraint for the estimation of behavioral parameters. We suggest the use of observed infiltration and/or runoff to increase the strength of conclusions drawn regarding the physical realism of runoff-related processes in GW.

5.1.2. Role of the Thermal Conductivity Muting Factor (sbeta)

[49] All three models compute ground heat flux (G) using a flux-gradient relationship:

$$G = \text{DF}_1 \frac{\text{STC}_1 - T_1}{0.5 * \text{ZSOIL}_{(1)}}, \quad (12)$$

in which STC_1 is the temperature at the center of the first soil layer ($0.5*\text{ZSOIL}_{(1)}$) and T_1 is the surface temperature. DF_1 is the heat conductivity of the topsoil layer.

[50] Noah assumes that, as vegetation cover increases, heat flux into the ground decreases. sbeta and the vegetated fraction (shdfac) mute DF_1:

$$\text{DF}_1 = \text{DF}_1 e^{\text{sbeta}*\text{shdfac}}. \quad (13)$$

[51] At site 7, the mode of the posterior probability distribution of all three models is near the bound of the explored parameter range (−1) (Figure 6d). The preference for near-bound values is more pronounced in DV, which at site 7 tends to have shdfac values near 1.0 (putting downward pressure on the value of sbeta). The skewed posterior parameter distributions suggest that an even less negative value of sbeta may have yielded better results at site 7.

[52] The assumption that vegetation necessarily decreases the thermal conductivity of the top layer of the soil may be incorrect. If the "vegetation effect" on thermal conductivity is real, then the model underestimates the top layer soil thermal conductivity. At site 7 (and at several other sites), there is a clear trade-off between H and G that is mediated by the thermal conductivity. The trade-off suggest a need for revised process representation.

[53] When comparing site 7 simulations to those of the other two wet sites (8 and 9), we see a roughly consistent preference for near zero values of sbeta. At the drier sites (1–6), the model's strong preference for near zero values of sbeta is less obvious; however, shdfac is closer to zero at these sites, which lowers the value of the muting factor equation (13).

5.1.3. Role of Minimum Stomatal Resistance (rcmin)

[54] Parameter rcmin controls much of the variance in H and LE, especially at the wetter sites. As rcmin increases, the ratio of actual to potential evapotranspiration decreases.

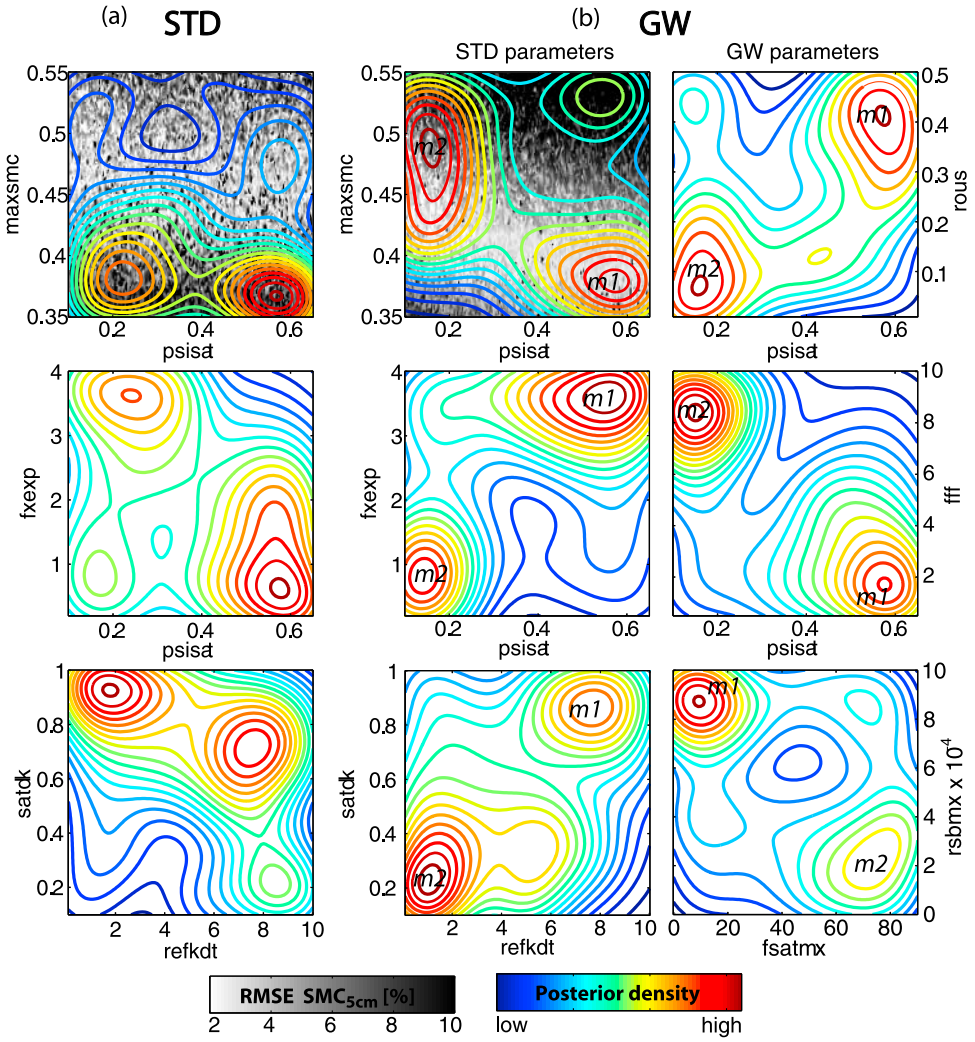

Figure 7. Multivariate posterior distribution of the behavioral parameters of STD and GW at site 7 shown for selected parameter combinations in bivariate plots. Higher density of parameter values are indicated with increasingly redder contours. The response surface of SMC_{5cm} is shown in the back; darker regions have higher errors. The bimodal behavior of GW is signaled by m1 and m2. See text for explanation.

Overall, rcmin has a more consistent influence on the variance of H than on that of LE.

[55] At site 7, all three models perform best with low values for rcmin (Figure 6b), which increases LE for a given potential evapotranspiration; however, rcmin is less identifiable in GW and DV. The mode of the rcmin distribution is higher for GW than for STD, perhaps because GW tends to have a wetter soil and a more robust simulation of LE. The spread of the posterior parameter distribution of rcmin for DV is significantly larger than that for STD, although both models share the same mode. This decrease in identifiability of parameters functionally related to lai (as is rcmin) is consistent with the added degrees of freedom available in DV (DV parameters gl and sla are most important in predicting lai) (Figure 2). Because DV simulations include a wider spread of lai states, they also have a wider spread of "good" rcmin values.

5.2. How Does GW Have to Be Adjusted to Make It Work Better Than or as Well as STD at Site 7?

[56] The response surface of RMSE SMC_{5cm} changes between STD and GW (Figure 7; e.g., see maxsmc versus psisat). For GW, the shape of the bivariate posterior distributions of soil parameters that are shared with STD is significantly different, presumably because of interaction of the GW parameters and module with those of STD. Such shifts in model function affect the model covariance structure as shown in Table 3.

[57] After multiobjective parameter estimation at site 7, GW functions in one of two preferred modes (Figure 7b). In the slightly preferred first mode (m1), the parameters work together to help GW function as the developers likely intended. Strong communication between the aquifer and the soil column is supported by relatively high values of saturated hydraulic conductivity (satdk), low values of the reciprocal of the *e*-folding depth of hydraulic conductivity

Table 3. Spearman Rank Correlation Coefficients Between Parameter Sets Belonging to the Behavioral Set for STD and GW[a]

	STD						
GW	maxsmc	psisat	satdk	fxexp	rous	fff	fsatmx
maxsmc		−0.10	−0.40	0.29			
psisat	−0.33		−0.14	−0.32			
satdk	−0.09	0.49		0.22			
fxexp	−0.26	0.41	0.23				
rous	−0.01	0.26	0.24	0.14			
fff	0.11	−0.46	−0.45	−0.49	−0.37		
fsatmx	−0.22	−0.04	−0.17	0.09	−0.37	0.17	
rsbmx	0.11	−0.25	−0.13	−0.21	0.32	0.08	−0.24

[a]Note the change in the covariance structure in Figure 7. See Table 1 for abbreviations of parameter names.

(fff), and low porosity (maxsmc). A relatively low surface runoff scaling factor (fsatmx) and a relatively high subsurface runoff scaling factor (rsbmx) ensure that subsurface runoff dominates surface runoff. Consistent with natural processes, high soil suction (psisat) pulls water upward. A high aquifer specific yield (rous) deepens the water table (weakening the direct influence of the saturated zone on the model soil column) and transforms more water to runoff rather than to recharge.

[58] In the second mode (m2), GW adopts parameter values that make the model work as one would expect STD to function (i.e., the model operates with parameters that render GW nonfunctional) (Figure 7b). Relatively high values of fff effectively seal the bottom of the soil column, limiting communication between the aquifer and the soil column; high maxsmc decreases the vertical conductivity, further inhibiting the already poor communication. High maxsmc favors decreased direct evaporation. Surface runoff is augmented by a relatively high fsatmx; subsurface runoff is lessened by the relatively low rsbmx.

[59] These alternative behaviors are a possible explanation for the issue identified by *Rosero et al.* [2009], who showed that despite very good performance of calibrated GW, the model suffered from low robustness (i.e., a high sensitivity to unmeasurable parameters).

5.3. How Does DV Have to be Adjusted to Make It Work Better Than or as Well as STD at Site 7?

[60] STD and DV functionally differ in two ways: (1) STD prescribes shdfac using monthly climatological values (∼0.7 at site 7), while DV predicts it as a function of environmental variation in moisture and radiation availability and (2) STD treats lai as a parameter, while DV uses shdfac to predict lai variation using a functional relationship:

$$\mathrm{lai} = \max\left(\mathrm{xlai}_{\min}, \frac{-1}{\mathrm{gl}} \log^{-1}(1 - \mathrm{shdfac})\right). \qquad (14)$$

[61] Vegetation affects all components of LE flux (via shdfac): (1) vegetation shades the soil, modulating direct evaporation (E_{dir}); (2) vegetation retains water above the soil, contributing to evaporation from the canopy (E_c); and (3) vegetation fuels transpiration (E_{transp}). In DV, a high value of conversion parameter gl fixes shdfac near 1 and yields a regime in which E_c and E_{transp} are strongly favored over E_{dir}. Low values of gl fix shdfac near zero and promote a regime in which E_{dir} is the dominant component of LE. When shdfac is near zero, both E_c and E_{transp} are minimized. At sites with sufficient vegetation, DV enables the model to correctly give more weight to E_{transp}. STD, unable to change the value of shdfac to shift the balance of components of LE, favors higher lai (which decreases stomatal resistance and increases E_{transp}) as means for increasing total LE.

[62] When compared to STD, DV can achieve "good" model performance using a wider range of values for shdfac and lai. We see this decreased identifiability of DV parameters when comparing the bivariate posterior parameter distributions of STD to those of DV at site 7 (Figure 8). The identifiability in the response surface of RMSE LE has changed (e.g., lai versus rcmin) (Figure 8). The decrease in identifiability of parameters that are functionally related to shdfac and/or lai can be seen across the IHOP sites (results not shown). The interplay of the parameters of the DV module also leads to changes in parameter densities of STD and DV (Figure 8). We see additional evidence for increased interaction between parameters in DV when we note that the models' covariance structure has been altered (Table 4). For example, rcmin and maxsmc are positively correlated in STD, but in DV they have a very slight negative correlation.

[63] Although the increased flexibility of lai and shdfac values may improve the model's simulation of seasonal and interannual variation in surface fluxes, over time scales examined here, DV does not appear to improve the model. The constraints imposed by the turbulent and near-surface states may be insufficient for the complexity of the model and/or DV may need to be constrained with observations of carbon fluxes and plant growth. When there is little vegetation (e.g., at sites 1–3), DV may be failing to consider special water use features associated with the semiarid vegetation [*Unland et al.*, 1996]. The function of the DV module may also be hindered by Noah's lack of a separate canopy layer [*Rosero et al.*, 2009] and/or by the absence of a more complex Ball-Berry type of stomatal conductance formulation (Niu et al., unpublished manuscript, 2009).

6. What Are the Implications of Our Sensitivity Analysis Results for Parameter Transferability?

[64] Our foregoing assessments have shown that parameter interaction is a significant contributor to model variance (section 4) and that the behavioral posterior parameter distributions for a given site change between models (section 5) and for a given model between sites (not shown; see Figure 9). These observations challenge the long-standing assumption of land surface modeling that LSM parameters are physically meaningful quantities. Because developers have attempted to use physical principles when designing LSMs, physically based model parameters have been assumed to correspond to physical characteristics of a system [e.g., *Dickinson et al.*, 1986], which can be either measured in the field (e.g., porosity) or inferred from (remotely sensed) observations (e.g., LAI). Identical LSM parameters are used in locations that share the same physical characteristics [e.g., *Sellers et al.*, 1996]. "Parameter transferability," the a priori assignment of parameter values based on a site's physical characteristics (e.g., soil and vegetation type), depends on the appropriateness of the above stated assumption. By making sets of vegetation-related (soil-related)

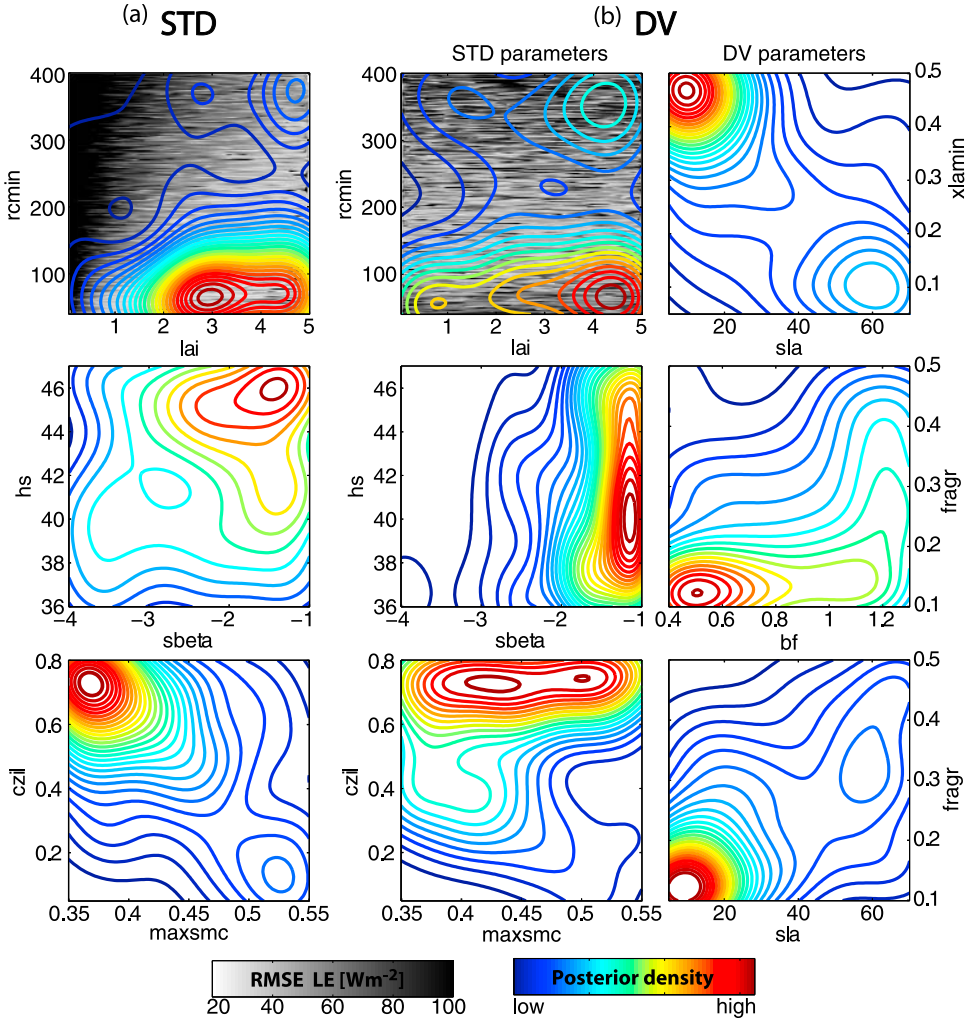

Figure 8. Bivariate depiction of the posterior distribution of behavioral parameters of STD and DV at site 7. Higher densities of parameter values are indicated with red contours. The response surface of LE is shown in the back; darker regions have higher errors. Note the significant change in the identifiability of hs and maxsmc.

parameters functions of vegetation (soil) type, LSMs contain the implicit assumption that vegetation (soil) type solely determines the ideal values of vegetation (soil) parameters.

[65] The joint multivariate posterior distribution summarizes much of the information regarding the relationships between model parameters (i.e., the model structure) at a particular location given observed data sets. To test the assumption that parameters and parameter relationships directly relate to physical characteristics, we compare the similarity of the marginal posterior distributions of the behavioral parameter sets between sites. We also evaluate the extent to which climate determines the similarity of parameter distributions between locations.

6.1. Testing Parameter Transferability Between Sites Using Soil Textures and Vegetation Types

[66] If parameters were readily "transferable" between sites solely based on the sites' vegetation type, we would expect the distributions of the vegetation parameters at two sites with the same vegetation type but different climatic regime (e.g., sites 2 and 8) to be more similar than the distributions of the same parameters at two sites with different vegetation but similar climate (e.g., sites 2 and 1). This expectation is in general not supported by our evidence. The distributions of rcmin and lai (Figures 9a and 9b) and rsmax and z0 are more similar between sites with similar

Table 4. Spearman Rank Correlation Coefficients Between Parameter Sets Belonging to the Behavioral Set for STD and DV[a]

	STD						
DV	rcmin	hs	maxsmc	psisat	fragr	bf	xlaimin
rcmin		−0.35	0.44	0.02			
hs	0.30		−0.15	−0.36			
maxsmc	−0.16	−0.29		−0.10			
psisat	0.50	0.36	−0.21				
fragr	0.58	0.24	−0.02	0.10			
bf	0.61	0.30	−0.19	0.59	0.40		
xlaimin	−0.72	−0.31	0.10	−0.31	−0.62	−0.54	
sla	0.80	0.21	−0.15	0.35	0.66	0.45	−0.67

[a]Note the change in the covariance structure in Figure 8. See Table 1 for abbreviations of parameter names.

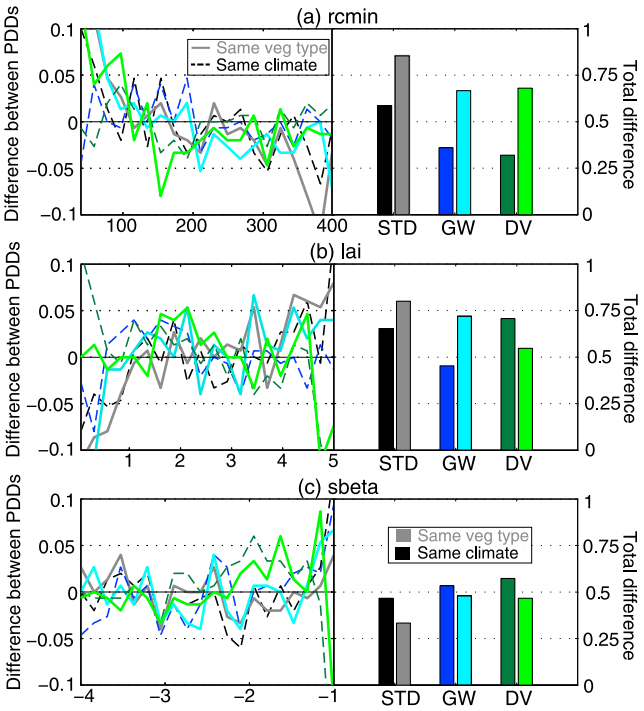

Figure 9. For the selected, sensitive vegetation parameters (a) rcmin, (b) lai, and (c) sbeta, (left) the difference between the marginal posterior parameter distributions (PPD) obtained at sites with the same vegetation but different climate (sites 2 and 8) (continuous, bright lines) and the difference between the marginal posterior parameter distributions obtained at sites with similar climate but different vegetation (sites 1 and 2) (dashed, dark lines) are shown. (right) Bar graphs; note that the total difference between parameter distributions at sites with the same vegetation but different climate (brightly colored bars) is generally not smaller than the difference of distributions of the same parameters between contiguous sites with similar climate but different vegetation (dark colored bars).

climate (dry) than they are between sites with the same vegetation (grass). The parameters hs and cmcmax show a similar lack of transferability. Only sbeta shows "transferability" (i.e., there are smaller differences between the distributions from sites with the same vegetation cover) for all models (Figure 9c). Parameter cfactr is transferable, but only for STD. Parameter rgl could be considered transferable, but only for DV and GW. The IHOP data set does not enable us to test parameter transferability using two sites with the same soil texture but different climatology.

[67] The case studies above are by no means conclusive, but they do not support the hypothesis that parameters are transferable solely based on vegetation type. The results instead suggest that LSM parameters are more sensitive to climatic forcing than to a specific land cover classification. Our results are consistent with similar observations for other hydrologic models [*Demaria et al.*, 2007; *van Werkhoven et al.*, 2008], and they are also consistent with observations made using single optimal parameter sets for the Noah LSM [*Hogue et al.*, 2005; *Rosero and Bastidas*, 2007; *Gutmann and Small*, 2007].

6.2. Synthesizing Sensitivity to Site, Soil, and Vegetation Classes by Means of Clustering

[68] In order to more quantitatively synthesize knowledge gained through sensitivity analysis for use at ungauged locations, we built upon the aforementioned idea of comparing the similarity of parameter distributions across sites by complementing the approach with unsupervised, agglomerative hierarchical clustering methods [*Rosero and Bastidas*, 2007].

[69] For each IHOP site, we obtained a stable, multivariate posterior probability distribution χ of behavioral parameter sets $X = \{x_1, x_2, \ldots, x_i, \ldots x_k\}$ using multiobjective parameter estimation. The marginal probability distribution for the ith parameter is χ_i. To circumvent comparing each site to every other, two at a time (similar to section 6.1), we define a triangular probability distribution D_i as a reference distribution for each parameter. $D_i = 1$ when the value of parameter x_i is the "default" for the site. $D_i = 0$ when x_i is at either edge of the feasible range. This

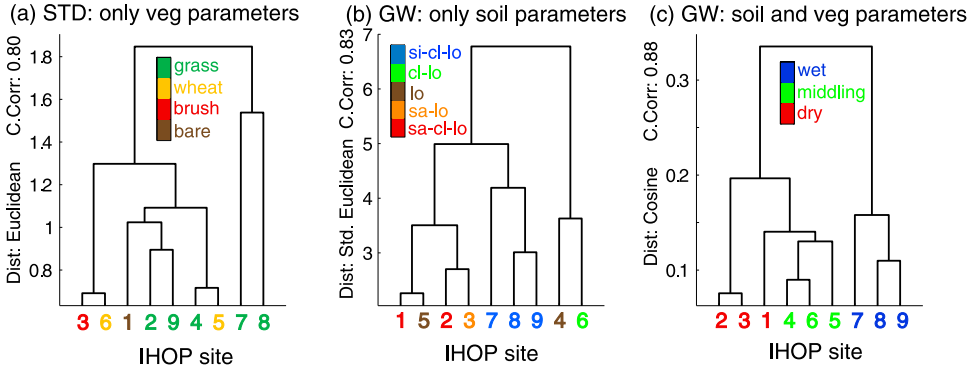

Figure 10. Clustering of sites using (a) only the vegetation parameters of STD, (b) only the soil parameters of GW, and (c) both soil and vegetation parameters of GW. The similarity between marginal distributions of behavioral parameters at all sites is compared using different distances. Figures 10a–10c report the distance that maximizes the cophenetic correlation coefficient of the linkage. Note that neither soil nor vegetation parameters render groups solely based on soil or vegetation type. The clusters of all parameters seem to have a strong relationship with the three climatic zones.

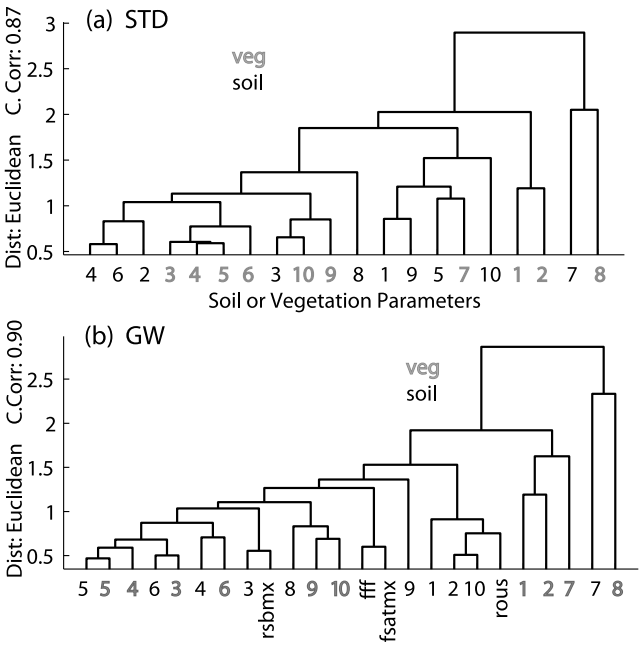

Figure 11. Clustering of soil (black), vegetation (gray), and GW parameters for the behavioral, marginal posterior distributions of (a) STD and (b) GW at all sites. The cophenetic correlation coefficient for the complete linkage for the parameters of STD and GW is 0.87 and 0.90, respectively. GW parameters seem to behave in a similar way as the soil parameters do.

step allows us to introduce the assumption that the parameters relate to soil and vegetation types.

[70] For each parameter, and at each site, we quantify the closeness between the cumulative distribution of the ''optimal'' values of x_i (i.e., the marginal χ_i) and the reference distribution D_i. We use the Hausdorff norm to quantify the difference $\chi_i - D_i$. For each model, the matrix of similarity of the marginal distributions of k parameters at all the n evaluation sites is

$$S = \begin{bmatrix} \chi_{11} - D_{11} & \dots & \chi_{k1} - D_{k1} \\ \dots & \dots & \dots \\ \chi_{k1} - D_{k1} & \dots & \chi_{kn} - D_{kn} \end{bmatrix}. \quad (15)$$

S can be used to identify groups of parameters that are similar between locations or to identify locations where groups of parameters behave alike. We use the unsupervised, agglomerative hierarchical clustering algorithm (described in section 3.8) to find these groups without making any further assumptions about the number of groups.

[71] If the previously described assumption of parameter transferability based on site characteristics holds (and if IHOP vegetation classifications are correct), then, given the set of similarity vectors created using the set of vegetation parameter distributions $S(x_{veg,1..n})$, a clustering procedure should be able to classify similar sites in groups that resemble the IHOP vegetation type groupings (Table 1). Similarly, clustering of $S(x_{soil,1..n})$ would result in sites grouped according to the IHOP soil texture classification (Table 1).

[72] Applying a suite of distance metrics (e.g., Manhattan, Euclidean, Cosine), neither soil nor vegetation parameters render groups of sites that partition solely based on the expected soil or vegetation classifications. Figure 10a shows the classification tree (dendrogram) for STD using the Euclidean distance, which maximizes the cophenetic correlation coefficient of the linkage (also shown). None of the distance metrics allowed us to classify $S(x_{veg,1..n})$ by location in a way that matched the IHOP vegetation classifications. Given the subset $S(x_{soil,1..n})$, composed of the similarity vectors of the 10 soil parameters at all sites, classification of the IHOP sites according to soil characteristics was also not feasible (Figure 10b). Using similarity vectors for STD, GW, and DV, some (but not all) of the distances identified sites 7, 8, and 9 as having the same soil and same vegetation type (although, because they also share the same climate type, we are unable to definitively attribute such classification to shared vegetation type). The rest of the sites do not strongly coalesce according to physical properties. For example, the pasture sites are not distinctively grouped; sites 5 and 6 (wheat crops) are never grouped according to vegetation (Figure 10a). Sites 1 and 2 (sandy clay loam) and sites 4 and 5 (loam) do not cluster together using soil parameters (Figure 10b). These results are consistent with earlier findings presented here, which suggest that interaction between soil and vegetation parameters is significant (section 4), to the point that it shapes the posterior parameter distributions (section 5). These results also suggest that soil and vegetation type are not, by themselves, good physical characteristics by which to transfer parameters.

[73] To account for interdependence between soil and vegetation parameters, we classified the entire matrix $S(x_{soil},x_{veg})$. If parameters can be transferred based on shared vegetation and soil type, then the clustering of the entire matrix should identify groups of sites with the same vegetation and soil type (e.g., sites 7–9). Figure 10c shows a pattern (found with several distances) that is consistent across models: sites 7–9 cluster together. Sites 7, 8, and 9 also have similar climates, and the classification of the sites shows strong resemblance to the climatic gradient. Given this data set, we cannot disprove the contention that parameters can be transferred between sites that have both the same vegetation and soil type.

[74] If we instead cluster S looking for groups of parameters, we expect that x_{soil} will as a whole behave in a similar way across sites. In other words, one can produce a map of sensitivity to characterize which parameters are most similar to their default (prior distribution) and which are not. Figure 11 shows representative groupings of the behavioral, marginal posterior distributions of STD and GW parameters at all sites. Using a suite of distance measures, we were unable to identify definitive clusters of soil and vegetation parameters within the set of similarity vectors S, meaning that individual parameters are not sensitive in groups that primarily relate to soil alone or to vegetation alone. The new GW parameters do behave in a way that is similar to other soil parameters (Figure 11b), which informs the estimation of GW parameters for distributed applications.

[75] We conclude that the primary site-to-site control on parameters values is not a site's soil or vegetation type alone. This result is consistent with the notion that LSM

parameters, which must represent physical processes across multiple scales, are "effective" values rather than physically derived quantities [*Wagener and Gupta*, 2005]. It is also consistent with the assertion that interaction between classes of parameters (e.g., "soil" parameters and "vegetation" parameters) is very important. Our clustering analysis suggests that climate is a major control of site-to-site variation in parameter values and supports recommendations that climate be considered when transferring parameter values between sites [*Liang and Guo*, 2003; *Demaria et al.*, 2007; *van Werkhoven et al.*, 2008].

7. Summary and Conclusion

[76] Sensitivity analysis allows us to draw conclusions regarding land surface model (LSM) development and model assessment practices, the functioning of three versions of the widely used versions of the Noah LSM, and the a priori estimation of parameter values. Our work yields several conclusions that can be generalized to LSM and other environmental models in general, and several conclusions that are specific to the Noah LSM.

[77] We show that the clear patterns of parameter importance identified by variance-based sensitivity analysis (VSA) are consistent with site-to-site variation in climate and with model-to-model changes in physical parameterization. VSA shows that parameter interactions within models exert significant control on the model output variance. Shifts in parametric control on variance and covariance hint at whether a model represents the water and energy cycles in a way that is consistent with the assumptions underlying the models. Although the optimal value of a parameter is useful information, the change in the functional relationship between parameters is more likely to be relevant for model development and hypothesis testing.

[78] Transfer of parameters based solely on similarity in vegetation type or soil texture is not a viable method for a priori parameter estimation. The work presented here shows that vegetation type and soil texture are not the most significant contributors to site-to-site variation in optimal parameter values. Interaction between soil and vegetation parameters is significant and varies between sites; and explains at least partially why the transfer of parameters based solely on shared vegetation type or soil texture does not work. The primary factor controlling site-to-site variation in parameters is likely to be climate, although, given the relatively small data set used here, the combination of a site's vegetation type and soil texture or some unidentified factor cannot be ruled out as the dominant controlling factor. The lack of viability of parameter transfer based solely on soil and vegetation type is a conclusion that has significant implications for the field of regional and global land surface modeling, which depends on parameter transfer based on stand-alone vegetation type and soil texture as a means for a priori parameter estimation.

[79] Looking specifically at the performance of the three versions of the Noah LSM used in this study (STD, GW, and DV), we make several nonsite-specific conclusions regarding model behavior. All three models exhibit significant parameter interaction, indicating that the models are overparameterized and/or underconstrained. All three show the least parameter interaction at the middling moisture and wet sites and the most parameter interaction at the three driest sites. This difference suggests a need for reformulation of Noah LSM such that semiarid regions are more realistically represented. On the whole, GW has less parameter interaction than STD (except at dry sites), indicating that it represents land surface system with the most realism of any of the three models. GW is also least sensitive to errant parameters at the wettest sites (where groundwater is likely the most influential). DV has much more parameter interaction than STD, which provides evidence that the model is not performing as its developers intended, does not add value to STD, and/or requires additional constraint. Specific to site 7, we make the following observations: (1) STD and DV tend to underestimate direct evaporation from the soil and (2) GW does not (maybe because of wet soil). The assumption that vegetation decreases the thermal conductivity of the top layer of the soil is not well supported by the data (this conclusion can be roughly generalized to other sites, especially the wet sites). At site 7, GW functions in one of two modes – the slightly preferred mode works in a way that mirrors what the developers likely intended; the second mode makes GW function as one might expect STD to work. Constraining runoff may isolate the more realistic mode. GW has less spurious parameter interaction in part because it decouples direct evaporation and subsurface runoff (which are coupled via porosity in STD and DV). This decoupling appears to make the model function more realistically, with less trade-off between the simulation of soil moisture and LE. Adding modules (DV, GW) decreases the identifiability of minimum stomatal resistance, although all three models prefer low minimum stomatal resistance (thus increasing LE for a given set of conditions). Across several sites, DV functions in one of two modes (1) the first emphasizes direct soil and canopy evaporation over transpiration and (2) the second emphasizes transpiration over direct evaporation from the soil and canopy.

[80] Our approach to sensitivity analysis complements new methods for characterizing typical modes of LSM behavior [*Gulden et al.*, 2008b; *Rosero et al.*, 2009] within a model diagnostic framework [*Gupta et al.*, 2008] that helps bridge the gap between model identification and development. We encourage other modeling groups to perform similar analyses with their models as a way to ensure rapid, continued improvement of our understanding and modeling of environmental processes.

[81] **Acknowledgments.** We thank Pedro Restrepo at OHD/NWS, Dave Gochis at NCAR and Ken Mitchell at NCEP for their insight. We appreciated suggestions by M. Bayani Cardenas, Charles S. Jackson, and the comments of two anonymous reviewers. We acknowledge the International H_2O Project for the data sets. We benefited from the computational resources at the Texas Advanced Computing Center (TACC). The first author was supported by the Graduate Fellowship of the Hydrology Training Program of the OHD/NWS. This project was also funded by the NOAA grant NA07OAR4310216, NSF, and the Jackson School of Geosciences.

References

Abramowitz, G., R. Leuning, M. Clark, and A. Pitman (2008), Evaluating the performance of land surface models, *J. Clim.*, *21*, 5468–5481, doi:10.1175/2008JCLI2378.1.

Bastidas, L. A., H. V. Gupta, S. Sorooshian, W. J. Shuttleworth, and Z. L. Yang (1999), Sensitivity analysis of a land surface scheme using multicriteria methods, *J. Geophys. Res.*, *104*(D16), 19,481–19,490, doi:10.1029/1999JD900155.

Bastidas, L. A., H. V. Gupta, and S. Sorooshian (2001), Bounding the parameters of land-surface schemes using observational data, in *Land Surface Hydrology, Meteorology, and Climate: Observations and Modeling*, *Water Sci. Appl.*, vol. 3, edited by V. Lakshmi et al., pp. 65–76, AGU, Washington, D. C.

Bastidas, L. A., T. S. Hogue, S. Sorooshian, H. V. Gupta, and W. J. Shuttleworth (2006), Parameter sensitivity analysis for different complexity land surface models using multicriteria methods, *J. Geophys. Res.*, *111*, D20101, doi:10.1029/2005JD006377.

Beck, M. B. (1987), Water quality modeling: A review of the analysis of uncertainty, *Water Resour. Res.*, *23*, 1393–1442, doi:10.1029/WR023i008p01393.

Beven, K. J., and J. Freer (2001), Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, *J. Hydrol.*, *249*, 11–29, doi:10.1016/S0022-1694(01)00421-8.

Bois, B., et al. (2008), Using remotely sensed solar radiation data for reference evapotranspiration estimation at a daily time step, *Agric. For. Meteorol.*, *148*, 619–630, doi:10.1016/j.agrformet.2007.11.005.

Chen, F., and J. Dudhia (2001), Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 Modeling System. Part I: Model implementation and sensitivity, *Mon. Weather Rev.*, *129*, 569–585, doi:10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2.

Cosgrove, B. A., et al. (2003), Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project, *J. Geophys. Res.*, *108*(D22), 8842, doi:10.1029/2002JD003118.

Demaria, E. M., B. Nijssen, and T. Wagener (2007), Monte Carlo sensitivity analysis of land surface parameters using the Variable Infiltration Capacity model, *J. Geophys. Res.*, *112*, D11113, doi:10.1029/2006JD007534.

Dickinson, R. E., A. Henderson-Sellers, P. J. Kennedy, and M. F. Wilson (1986), Biosphere-Atmosphere Transfer Scheme (BATS) for the NCAR Community Climate Model, *NCAR Tech. Note NCAR/TN-275+STR*, 69 pp., Natl. Cent. for Atmos. Res., Boulder, Colo.

Dickinson, R. E., M. Shaikh, R. Bryant, and L. Graumlich (1998), Interactive canopies for a climate model, *J. Clim.*, *11*, 2823–2836, doi:10.1175/1520-0442(1998)011<2823:ICFACM>2.0.CO;2.

Ek, M. B., et al. (2003), Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model, *J. Geophys. Res.*, *108*(D22), 8851, doi:10.1029/2002JD003296.

Gao, X., S. Sorooshian, and H. V. Gupta (1996), Sensitivity analysis of the biosphere-atmosphere transfer scheme, *J. Geophys. Res.*, *101*(D3), 7279–7289, doi:10.1029/95JD03161.

Gulden, L. E., E. Rosero, Z.-L. Yang, M. Rodell, C. S. Jackson, G.-Y. Niu, P. J.-F. Yeh, and J. Famiglietti (2007), Improving land-surface model hydrology: Is an explicit aquifer model better than a deeper soil profile?, *Geophys. Res. Lett.*, *34*, L09402, doi:10.1029/2007GL029804.

Gulden, L. E., Z.-L. Yang, and G.-Y. Niu (2008a), Sensitivity of biogenic emissions simulated by a land-surface model to land-cover representations, *Atmos. Environ.*, *42*, 4185–4197, doi:10.1016/j.atmosenv.2008.01.045.

Gulden, L. E., E. Rosero, Z.-L. Yang, T. Wagener, and G. Niu (2008b), Model performance, model robustness, and model fitness scores: A new method for identifying good land-surface models, *Geophys. Res. Lett.*, *35*, L11404, doi:10.1029/2008GL033721.

Gupta, H. V., S. Sorooshian, and P. O. Yapo (1998), Toward improved calibration of hydrologic models: Multiple and non-commensurable measures of information, *Water Resour. Res.*, *34*, 751–763, doi:10.1029/97WR03495.

Gupta, H. V., L. A. Bastidas, S. Sorooshian, W. J. Shuttleworth, and Z. L. Yang (1999), Parameter estimation of a land surface scheme using multicriteria methods, *J. Geophys. Res.*, *104*(D16), 19,491–19,504, doi:10.1029/1999JD900154.

Gupta, H. V., K. J. Beven, and T. Wagener (2005), Model calibration and uncertainty estimation, in *Encyclopedia of Hydrological Sciences*, edited by M. G. Anderson et al., pp. 1–17, John Wiley, Chichester, U. K.

Gupta, H. V., T. Wagener, and Y. Liu (2008), Reconciling theory with observations: Elements of a diagnostic approach to model evaluation, *Hydrol. Processes*, *22*, doi:10.1002/hyp.6989.

Gutmann, E. D., and E. E. Small (2007), A comparison of land surface model soil hydraulic properties estimated by inverse modeling and pedotransfer functions, *Water Resour. Res.*, *43*, W05418, doi:10.1029/2006WR005135.

Hair, J. F., R. E. Anderson, R. L. Tatham, and W. C. Black (1995), *Multivariate Data Analysis With Readings*, Prentice-Hall, Englewood Cliffs, N. J.

Helton, J., and F. Davis (2003), Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems, *Reliab. Eng. Syst. Safety*, *81*(1), 23–69, doi:10.1016/S0951-8320(03)00058-9.

Henderson-Sellers, A. (1993), A factorial assessment of the sensitivity of the BATS land-surface parameterization scheme, *J. Clim.*, *6*, 227–247, doi:10.1175/1520-0442(1993)006<0227:AFAOTS>2.0.CO;2.

Hogue, T. S., L. A. Bastidas, H. V. Gupta, S. Sorooshian, K. Mitchell, and W. Emmerich (2005), Evaluation and transferability of the Noah land surface model in semiarid environments, *J. Hydrometeorol.*, *6*, 68–84, doi:10.1175/JHM-402.1.

Hogue, T. S., L. A. Bastidas, H. V. Gupta, and S. Sorooshian (2006), Evaluating model performance and parameter behavior for varying levels of land surface model complexity, *Water Resour. Res.*, *42*, W08430, doi:10.1029/2005WR004440.

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

Jakeman, A. J., R. A. Letcher, and J. P. Norton (2006), Ten iterative steps in development and evaluation of environmental models, *Environ. Modell. Software*, *21*, 602–614, doi:10.1016/j.envsoft.2006.01.004.

Kato, H., M. Rodell, F. Beyrich, H. Cleugh, E. van Gorsel, H. Liu, and T. P. Meyers (2007), Sensitivity of land surface simulations to model physics, land characteristics, and forcings, at four CEOP sites, *J. Meteorol. Soc. Jpn.*, *85A*, 187–204, doi:10.2151/jmsj.85A.187.

Koster, R. D., et al. (2004), Regions of strong coupling between soil moisture and precipitation, *Science*, *305*(5687), 1138, doi:10.1126/science.1100217.

LeMone, M. A., et al. (2007), NCAR/CU surface, soil, and vegetation observations during the International H_2O Project 2002 field campaign, *Bull. Am. Meteorol. Soc.*, *88*, 65–81, doi:10.1175/BAMS-88-1-65.

Leplastrier, M., A. J. Pitman, H. Gupta, and Y. Xia (2002), Exploring the relationship between complexity and performance in a land surface model using the multicriteria method, *J. Geophys. Res.*, *107*(D20), 4443, doi:10.1029/2001JD000931.

Liang, X., and J. Guo (2003), Intercomparison of land surface parameterization schemes: Sensitivity of surface energy and water fluxes to model parameters, *J. Hydrol.*, *279*, 182–209, doi:10.1016/S0022-1694(03)00168-9.

Martinez, W., and A. Martinez (2002), *Computational Statistics Handbook With MATLAB*, 2nd ed., 767 pp., Chapman and Hall, Boca Raton, Fla.

McKay, M., R. Beckman, and W. Conover (1979), A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics*, *21*(2), 239–245, doi:10.2307/1268522.

Mitchell, K. E., et al. (2004), The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system, *J. Geophys. Res.*, *109*, D07S90, doi:10.1029/2003JD003823.

Niu, G.-Y., Z.-L. Yang, R. E. Dickinson, and L. E. Gulden (2005), A simple TOPMODEL-based runoff parameterization (SIMTOP) for use in global climate models, *J. Geophys. Res.*, *110*, D21106, doi:10.1029/2005JD006111.

Niu, G.-Y., Z.-L. Yang, R. E. Dickinson, L. E. Gulden, and H. Su (2007), Development of a simple groundwater model for use in climate models and evaluation with Gravity Recovery and Climate Experiment data, *J. Geophys. Res.*, *112*, D07103, doi:10.1029/2006JD007522.

Oleson, K. W., G. B. Bonan, J. Feddema, and M. Vertenstein (2008), An urban parameterization for a global climate model. Part II: Sensitivity to input parameters and the simulated urban heat island in offline simulations, *J. Appl. Meteorol. Climatol.*, *47*, 1061–1076, doi:10.1175/2007JAMC1598.1.

Peters-Lidard, C. D., D. M. Mocko, M. Garcia, J. A. Santanello Jr., M. A. Tischler, M. S. Moran, and Y. Wu (2008), Role of precipitation uncertainty in the estimation of hydrologic soil properties using remotely sensed soil moisture in a semiarid environment, *Water Resour. Res.*, *44*, W05S18, doi:10.1029/2007WR005884.

Pitman, A. (1994), Assessing the sensitivity of a land-surface scheme to the parameter values using a single column model, *J. Clim.*, *7*, 1856–1869, doi:10.1175/1520-0442(1994)007<1856:ATSOAL>2.0.CO;2.

Pitman, A. J. (2003), Review: The evolution of, and revolution in, land surface schemes designed for climate models, *Int. J. Climatol.*, *23*, 479–510, doi:10.1002/joc.893.

Prihodko, L., A. S. Denning, N. P. Hanan, I. Baker, and K. Davis (2008), Sensitivity, uncertainty and time dependence of parameters in a complex land surface model, *Agric. For. Meteorol.*, *148*, 268–287, doi:10.1016/j.agrformet.2007.08.006.

Randall, D. A., et al. (2007), Climate models and their evaluation, in *Climate Change 2007: The Physical Science Basis: Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon et al., pp. 561–648, Cambridge Univ. Press, Cambridge, U. K.

Ratto, M., P. C. Young, R. Romanowicz, F. Pappenberger, A. Saltelli, and A. Pagano (2007), Uncertainty, sensitivity analysis and the role of data

based mechanistic modeling in hydrology, *Hydrol. Earth Syst. Sci.*, *11*, 1249–1266.

Rodell, M., P. R. Houser, A. A. Berg, and J. S. Famiglietti (2005), Evaluation of 10 methods for initializing a land surface model, *J. Hydrometeorol.*, *6*, 146–155, doi:10.1175/JHM414.1.

Rosero, E., and L. A. Bastidas (2007), Evaluation of parameter transferability for land surface models across semi-arid environments, paper presented at 21st Conference on Hydrology, Am. Meteorol. Soc., San Antonio, Tex., 15–18 Jan.

Rosero, E., Z.-L. Yang, L. E. Gulden, G.-Y. Niu, and D. J. Gochis (2009), Evaluating enhanced hydrological representations in Noah LSM over transition zones: Implications for model development, *J. Hydrometeorol.*, *10*, 600–622, doi:10.1175/2009JHM1029.1.

Saltelli, A. (1999), Sensitivity analysis: Could better methods be used?, *J. Geophys. Res.*, *104*(D3), 3789–3793, doi:10.1029/1998JD100042.

Saltelli, A. (2002), Making best use of model evaluations to compute sensitivity indices, *Comput. Phys. Commun.*, *145*, 280–297, doi:10.1016/S0010-4655(02)00280-1.

Saltelli, A., M. Ratto, S. Tarantola, and F. Campolongo (2006), Sensitivity analysis practices: Strategies for model-based inference, *Reliab. Eng. Syst. Safety*, *91*(10–11), 1109–1125, doi:10.1016/j.ress.2005.11.014.

Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola (2008), *Global Sensitivity Analysis: The Primer*, John Wiley, Chichester, U. K.

Scanlon, B. R., R. C. Reedy, D. A. Stonestrom, D. E. Prudic, and K. F. Dennehy (2005), Impact of land use and land cover change on groundwater recharge and quality in the southwestern US, *Global Change Biol.*, *11*, 1577–1593, doi:10.1111/j.1365-2486.2005.01026.x.

Sellers, P., S. Los, C. Tucker, C. Justice, D. Dazlich, G. Collatz, and D. Randall (1996), A revised land surface parameterization (SiB2) for atmospheric GCMs. Part II: The generation of global fields of terrestrial biophysical parameters from satellite data, *J. Clim.*, *9*, 706–737, doi:10.1175/1520-0442(1996)009<0706:ARLSPF>2.0.CO;2.

Sobol′, I. M. (1993), Sensitivity analysis for nonlinear mathematical models, *Math. Modell. Comput. Exp.*, *1*, 407–414.

Sobol′, I. M. (2001), Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Math. Comput. Simul.*, *55*, 271–280, doi:10.1016/S0378-4754(00)00270-6.

Spear, R. C., T. M. Grieb, and N. Shang (1994), Parameter uncertainty and interaction in complex environmental models, *Water Resour. Res.*, *30*, 3159–3169, doi:10.1029/94WR01732.

Stöckli, R., D. M. Lawrence, G.-Y. Niu, K. W. Oleson, P. E. Thornton, Z.-L. Yang, G. B. Bonan, A. S. Denning, and S. W. Running (2008), Use of FLUXNET in the Community Land Model development, *J. Geophys. Res.*, *113*, G01025, doi:10.1029/2007JG000562.

Tang, Y., P. Reed, T. Wagener, and K. van Werkhoven (2006), Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation, *Hydrol. Earth Syst. Sci. Discuss.*, *3*, 3333–3395.

Tang, Y., P. Reed, K. van Werkhoven, and T. Wagener (2007), Advancing the identification and evaluation of distributed rainfall-runoff models using global sensitivity analysis, *Water Resour. Res.*, *43*, W06415, doi:10.1029/2006WR005813.

Trier, S. B., F. Chen, K. W. Manning, M. A. LeMone, and C. A. Davis (2008), Sensitivity of the PBL and precipitation in 12-day simulations of warm-season convection using different land surface models and soil wetness conditions, *Mon. Weather Rev.*, *136*, 2321–2343, doi:10.1175/2007MWR2289.1.

Unland, H., P. Houser, W. J. Shuttleworth, and Z.-L. Yang (1996), Surface flux measurement and modeling at a semi-arid Sonoran Desert site, *Agric. For. Meteorol.*, *82*, 119–153, doi:10.1016/0168-1923(96)02330-1.

van Werkhoven, K., T. Wagener, P. Reed, and Y. Tang (2008), Characterization of watershed model behavior across a hydroclimatic gradient, *Water Resour. Res.*, *44*, W01429, doi:10.1029/2007WR006271.

van Werkhoven, K., T. Wagener, P. Reed, and Y. Tang (2009), Sensitivity-guided reduction of parametric dimensionality for multiobjective calibration of watershed models, *Adv. Water Resour.*, *32*(8), 1154–1169.

Vrugt, J. A., H. V. Gupta, L. A. Bastidas, W. Bouten, and S. Sorooshian (2003), Effective and efficient algorithm for multiobjective optimization of hydrologic models, *Water Resour. Res.*, *39*(8), 1214, doi:10.1029/2002WR001746.

Wagener, T., and H. V. Gupta (2005), Model identification for hydrological forecasting under uncertainty, *Stochastic Environ. Res. Risk Assess.*, *19*(6), 378–387, doi:10.1007/s00477-005-0006-5.

Wagener, T., and J. Kollat (2007), Numerical and visual evaluation of hydrological and environmental models using the Monte Carlo analysis toolbox, *Environ. Modell. Software*, *22*, 1021–1033, doi:10.1016/j.envsoft.2006.06.017.

Wagener, T., D. P. Boyle, M. J. Lees, H. S. Wheater, H. V. Gupta, and S. Sorooshian (2001), A framework for development and application of hydrological models, *Hydrol. Earth Syst. Sci.*, *5*, 13–26.

Weckwerth, T. M., et al. (2004), An overview of the International H_2O Project (IHOP_2002) and some preliminary highlights, *Bull. Am. Meteorol. Soc.*, *85*, 253–277, doi:10.1175/BAMS-85-2-253.

Yang, Z.-L., and G.-Y. Niu (2003), Versatile integrator of surface and atmosphere processes: Part 1. Model description, *Global Planet. Change*, *38*, 175–189, doi:10.1016/S0921-8181(03)00028-6.

Yatheendradas, S., T. Wagener, H. Gupta, C. Unkrich, D. Goodrich, M. Schaffner, and A. Stewart (2008), Understanding uncertainty in distributed flash flood forecasting for semiarid regions, *Water Resour. Res.*, *44*, W05S19, doi:10.1029/2007WR005940.

L. E. Gulden and E. Rosero, ExxonMobil Upstream Research Company, PO Box 2189, Houston, TX 77252, USA. (erosero@mail.utexas.edu)

G.-Y. Niu, Biosphere 2 Earthscience, University of Arizona, PO Box 8746, Tucson, AZ 85738, USA.

T. Wagener, Department of Civil and Environmental Engineering, Pennsylvania State University, 226B Sackett Bldg., University Park, PA 16802, USA.

Z.-L. Yang, Department of Geological Sciences, Jackson School of Geosciences, University of Texas at Austin, 1 University Station C1100, Austin, TX 78712, USA. (liang@jsg.utexas.edu)

S. Yatheendradas, Hydrological Sciences Branch, NASA Goddard Space Flight Center, Code 614.3, Greenbelt, MD 20771, USA.