

1 DRAFT FOR THE JOURNAL OF GEOPHYSICAL RESEARCH

2
3
4
5 **Quantifying Parameter Sensitivity, Interaction and Transferability in**
6 **Hydrologically Enhanced Versions of Noah-LSM over Transition Zones**

7
8
9 ENRIQUE ROSERO,¹ ZONG-LIANG YANG,¹ THORSTEN WAGENER,² LINDSEY E.
10 GULDEN,¹ SONI YATHEENDRADAS,^{3,4} and GUO-YUE NIU¹

11
12
13 ¹*Department of Geological Sciences, Jackson School of Geosciences, The University of Texas at*
14 *Austin, Austin, Texas, USA*

15 ²*Department of Civil and Environmental Engineering, The Pennsylvania State University,*
16 *University Park, Pennsylvania, USA*

17 ³*NASA Goddard Space Flight Center, Hydrological Sciences Branch, Greenbelt, Maryland, USA*

18 ⁴*Earth System Science Interdisciplinary Center, University of Maryland, College Park,*
19 *Maryland, USA*

20
21 Manuscript submitted 10 March 2009
22

23
24 *Corresponding author address:*
25 *Zong-Liang Yang,*
26 *1 University Station #C1100,*
27 *The University of Texas at Austin,*
28 *Austin, TX 78712-0254.*
29 *E-mail: liang@jsg.utexas.edu*
30

ABSTRACT

1 We use sensitivity analysis to identify the parameters that are most responsible for shaping land
2 surface model (LSM) simulations and to understand the complex interactions in three versions of
3 the Noah LSM: the standard version (STD), a version enhanced with a simple groundwater
4 module (GW), and version augmented by a dynamic phenology module (DV). We use warm
5 season, high-frequency, near-surface states and turbulent fluxes collected over nine sites in the
6 US Southern Great Plains. We quantify changes in the pattern of sensitive parameters, the
7 amount and nature of the interaction between parameters, and the covariance structure of the
8 distribution of behavioral parameter sets. Using Sobol's total and first-order sensitivity indexes,
9 we show that very few parameters directly control the variance of the model output. Significant
10 parameter interaction occurs so that not only the optimal parameter values differ between
11 models, but the relationships between parameters change. GW decreases parameter interaction
12 and appears to improve model realism, especially at wetter sites. DV increases parameter
13 interaction and decreases identifiability, implying it is overparameterized and/or
14 underconstrained. A case study at a wet site shows GW has two functional modes: one that
15 mimics STD and a second in which GW improves model function by decoupling direct
16 evaporation and baseflow. Unsupervised classification of the posterior distributions of behavioral
17 parameter sets cannot group similar sites based solely on soil or vegetation type, helping to
18 explain why transferability between sites and models is not straightforward. This evidence
19 suggests a priori assignment of parameters should also consider climatic differences.

21 **Key words:** sensitivity analysis, land surface models, Noah LSM, model identification

22 **Index terms:** 1843 Land/atmosphere interactions, 1805 Computational hydrology, 0550 Model
23 verification and validation, 1873 Uncertainty assessment, 1847 Modeling

1 1. Introduction

2 Like other environmental models built to support scientific reasoning and testable
3 hypotheses to improve our understanding of the Earth system, land-surface models (LSMs) have
4 grown in sophistication and complexity (Pitman, 2003; Niu et al., 2009). The evaluation of LSM
5 simulations is consequently non-trivial and, especially when LSMs are to be used in predictive
6 mode for operational forecasting, policy assessments, or decision making, demands more
7 powerful methods for the analysis of their behavior (Saltelli, 1999; Jakeman et al., 2006; Randall
8 et al., 2007; Gupta et al., 2008; Abramowitz et al., 2009). One such method is sensitivity analysis
9 (SA), which is the process of investigating the role of the various assumptions, simplifications
10 and other input (parameter) uncertainties in shaping the simulations made by a model. SA is a
11 tool that enables exploring high-dimensional parameter spaces of complex environmental models
12 to better understand what controls model performance (Saltelli et al., 2008). Monte Carlo-based
13 uncertainty and SA uses the results of multiple model realizations to evaluate the range of model
14 outcomes and identifies the input parameters that give rise to this uncertainty (Wagener et al.,
15 2001; Wagener and Kollat, 2007). Used to its full potential, SA weighs model adequacy and
16 relevance, identifies critical regions in the space of the inputs, unravels parameter interactions,
17 establishes priorities for research, and, through an interactive process of revising the model
18 structure, leads to simplified models and increased understanding of the natural system (Saltelli
19 et al., 2006). In this article, we inform LSM development by using sophisticated SA to guide the
20 development of the commonly used Noah LSM (Ek et al., 2003).

21 SA has been an underutilized resource in LSM development. Approaches to quantify
22 ‘sensitivity’ (the rate of change in model response with respect to a factor) have very frequently
23 been restricted to a simple exploratory analysis of the effects of factors taken one-at-a-time

1 (OAT), without regard for interactions between factors. Such factors can be parameters (e.g.,
2 Pitman, 1994; Gao et al., 1996; Chen and Dudhia, 2001; Trier et al., 2008), meteorological
3 forcing, or ancillary data sets (e.g., Kato et al., 2007; Gulden et al., 2008a). OAT is illicit and
4 unjustified for non-linear models (Saltelli, 1999; Bastidas et al., 1999; Saltelli et al., 2006). More
5 powerful and sophisticated approaches that account, explicitly or implicitly, for parameter
6 interactions include the factorial method (e.g., Liang and Guo, 2003; Oleson et al., 2008) and
7 regionalized sensitivity analysis (RSA) (e.g., Bastidas et al., 2006, Demaria et al., 2007;
8 Prihodko et al., 2008). The factorial method is a global variance-based sensitivity analysis (VSA)
9 in which a selected number of model runs whose parameters have been perturbed by an arbitrary
10 percent from an arbitrary reference value (default) are evaluated to identify parameters that affect
11 the variance of the model output and to calculate a rough estimate of the low-order interactions.
12 RSA, on the other hand, representatively samples the entire parameter space and provides a more
13 robust assessment of the way in which parameter distributions change between subjectively
14 defined ‘good’ and ‘bad’ (i.e., behavioral and non-behavioral) model simulations. RSA does not
15 explicitly account for interactions between parameters and is typically applied with the sole
16 purpose of identifying sensitive parameters to reduce the dimensionality of the calibration
17 problem (e.g., Bastidas et al., 1999). When RSA and VSA are used separately, both the lack of
18 firm conclusions regarding the effect of dominant parameters (and their interactions) on the
19 output (e.g., Bastidas et al., 2006) and the inability to draw cause-effect relationships between
20 parameter regions and model responses (e.g., Liang and Guo, 2003) have precluded SA findings
21 from being widely used in model development. Our approach complements the information
22 about the posterior distributions of behavioral parameters (RSA) with information about their
23 contributions to model variance (VSA). The Monte Carlo-based VSA that we use (the method of

1 Sobol') is more robust than factorial analysis because it employs a representative sample of
2 parameter space and bypasses the perceived 'complexities' often associated with powerful SA
3 methods.

4 Our SA guides the investigation of the performance and physical realism of three
5 versions of the Noah LSM: the standard Noah (STD), a version augmented with a simple
6 groundwater model (Niu et al., 2007) (GW), and a version augmented with an interactive canopy
7 model (Dickinson et al., 1998) (DV). We evaluate the performance of Noah LSM in simulating
8 the land-surface states and fluxes at nine sites in a transition zone between wet and dry climates
9 using the datasets of IHOP_2002 (LeMone et al., 2007). Because of the strength of the land-
10 atmosphere coupling in transition zones (Koster et al., 2004), our work focuses on warm-season
11 climates of the US Southern Great Plains. We document how parameter interaction and
12 sensitivity varies with model, site, soil, vegetation, and climate. We investigate the ways in
13 which alternate model structures affect the behavior of 'physically meaningful' LSM parameters,
14 focusing on the dominant parameter interactions that dictate model response. We show that only
15 a few parameters directly control model variance and that parameter interaction is significant.
16 We further compare the similarity of estimated multivariate posterior distributions of behavioral
17 parameters and their sensitivity against those obtained at other sites. We show that the changes
18 between sites are not solely controlled by soil or vegetation types but appear to be strongly
19 related to the climatic gradient.

20 This paper is organized as follows. Datasets, models, and methods are described in
21 section 2. Experimental design and driving questions are formulated in section 3. Section 4
22 presents the patterns of sensitivity obtained via RSA and the global variance-based method of
23 Sobol'. Section 5 presents a case study demonstrating the use of SA to understand the functional

1 relationships between behavioral parameters, whose interaction serves to characterize model
2 structure and test hypotheses that regard the formulation of model. Section 6 discusses
3 implications of the results for the transferability of parameters between locations with similar
4 physical characteristics. Conclusions are summarized in section 7.

6 **2. Models, data and methods**

7 **2.1. Hydrologically enhanced versions of Noah LSM**

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

11 **2.1.1. Noah standard release 2.7 (STD)**

12 Noah (Ek et al., 2003; Mitchell et al., 2004) is a one-dimensional, medium complexity
13 LSM used in operational weather and climate forecasting. The model is forced by incoming
14 short- and longwave radiation, precipitation, surface pressure, relative humidity, wind speed and
15 air temperature. The computed state variables include soil moisture and temperature, water
16 stored on the canopy and snow on the ground. Prognostic variables include turbulent heat fluxes,
17 and fluxes of moisture and momentum. Noah has a single canopy layer with climatologically
18 prescribed albedo and vegetation greenness fraction. The soil profile of Noah is partitioned into 4
19 layers (lower boundaries at 0.1, 0.4, 1.0 and 2.0 m below the surface). The vertical movement of
20 water is governed by mass conservation and a diffusive form of the Richard's equation.
21 Infiltration is represented by a conceptual parameterization for the subgrid treatment of
22 precipitation and soil moisture. Drainage at the bottom is controlled only by gravitational forces;
23 percolation neglects hydraulic diffusivity. Direct evaporation from the top soil layer, from water

1 intercepted by the canopy and adjusted potential Penman-Monteith transpiration are combined to
2 represent total evapotranspiration. The surface energy balance determines the skin temperature of
3 the combined ground-vegetation surface. Soil-layer temperature is resolved with a Crank-
4 Nicholson numerical scheme. Diffusion equations for the soil temperature determine ground heat
5 flux. The Noah LSM uses soil and vegetation lookup tables for static soil and vegetation
6 parameters such as porosity, hydraulic conductivity, minimum canopy resistance, roughness
7 length, leaf area index, etc.

8 **2.1.2. Noah augmented with a short-term dynamic phenology module (DV)**

9 We coupled the canopy module of Dickinson et al. (1998) to STD in order to compute
10 changes in vegetation greenness fraction that result from environmental perturbations. The
11 module allocates carbon assimilated during photosynthesis to leaves, roots, and stems; the
12 fraction of photosynthate allocated to each reservoir is a function of, among other things, the
13 existing biomass density. The model also tracks growth and maintenance respiration and
14 represents carbon storage. Unlike STD, which computes greenness fraction by linear
15 interpolation between monthly climatological values, DV represents short-term phenological
16 variation by allowing leaf area to vary as a function of soil moisture, soil temperature, canopy
17 temperature, and vegetation type. DV makes vegetation fraction an exponential function of leaf
18 area index (LAI) (Yang and Niu, 2003). Because DV links vegetation fraction to dynamic LAI,
19 DV makes direct soil evaporation, canopy evaporation, and transpiration more responsive to
20 environmental conditions than STD. Unlike Dickinson et al. (1998), we parameterize the effect
21 of water stress on stomatal conductance as a function of soil moisture deficit, not as a function of
22 soil matric potential.

23 **2.1.3. Noah augmented with a simple groundwater model (GW)**

1 GW couples a lumped unconfined aquifer model (Niu et al., 2007) to the lower boundary
2 of the STD soil column. In GW, water flows vertically in both directions between the aquifer and
3 the soil column. The modeled hydraulic potential is the sum of the soil matric and gravitational
4 potentials. The relative water head between the bottom soil layer and the water table determines
5 either gravitational drainage or upward diffusion of water driven by capillary forces. Aquifer
6 specific yield is used to convert the water stored in the aquifer to water table depth. When water
7 is plentiful, the water table is within the model's soil column; if water is insufficient to maintain
8 a near-surface aquifer, the water table falls below the soil column. An exponential function of
9 water table depth modifies the maximum rate of subsurface runoff (for computation of baseflow)
10 and determines the fraction of the grid cell that is saturated at the land surface (for calculation of
11 surface runoff) (Niu et al., 2005). The model does not explicitly consider lateral transport of
12 groundwater between grid cells.

13 **2.2. IHOP_2002 sites and datasets**

14 We used datasets available at <http://www.rap.ucar.edu/research/land/observations/ihop/>
15 from the IHOP_2002 field campaign (Weckwerth et al. 2004; LeMone et al., 2007) to evaluate
16 the three versions of Noah LSM at nine sites along the Kansas-Oklahoma border and in northern
17 Texas (Fig. 1). The nine stations were sited to obtain a representative sample of the region,
18 which spans a strong east–west rainfall gradient and Bowen ratio. We used 45 days of high-
19 frequency, multi-sensor measurements of meteorological forcing, surface-to-atmosphere fluxes,
20 and near-surface soil moisture and temperature. Site characteristics, soil and vegetation classes,
21 mean meteorological values, average heat fluxes and near-surface states for the observation
22 period are summarized in Table 1.

23 **2.3. Model initialization and spin-up**

1 Following Rodell et al. (2005), we initialized each of the four soil layers at 50%
2 saturation and at the multi-annual-mean temperature. To drive the spin-up (between January 1,
3 2000, and May 13, 2002), we used downscaled North American Land Data Assimilation System
4 (NLDAS) (Cosgrove et al., 2003) meteorological forcing, interpolated from a 60-minute to a 30-
5 minute time step. The models were subsequently driven by IHOP_2002 meteorological forcing
6 from May 13, 2002, to June 25, 2002 (DOY 130 to 176). For GW, water table depth was
7 initialized assuming equilibrium of gravitational and capillary forces in the soil profile (Niu et
8 al., 2007).

9 **2.4. Evaluation datasets**

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

16 **2.5. Parameters considered in the sensitivity analysis**

17 We selected 10 soil and 10 vegetation parameters of STD that have been deemed
18 important at similar locations (Demarty et al. 2004; Bastidas et al., 2006). We included eight
19 parameters responsible for the phenology module and four that control the groundwater module
20 to analyze a total of 28 and 24 parameters for DV and GW, respectively. All other coefficients in
21 the models were kept constant at the recommended values. Default values and feasible ranges
22 (Table 2) for all parameters were taken from the literature (e.g., Chen and Dudhia, 2001; Hogue
23 et al., 2006).

1 2.6. Methods for sensitivity analysis

2 We use the variance-based method of Sobol' (Sobol', 1993; 2001) to efficiently identify
3 the factors that contribute most to the variance of a model's response. Unlike RSA, the method
4 of Sobol' deals explicitly with parameter interaction and has recently been used to quantify
5 model sensitivity and parameter interactions in hydrology (e.g., Ratto et al., 2007; Tang et al.,
6 2006, 2007; Yatheendradas et al., 2008; van Werkhoven et al., 2008). To our knowledge, it has
7 not yet been used for LSM SA.

8 Since variance-based SA (VSA) does not allow a mapping of the sensitivity back to the
9 parameter space, we complemented our evaluation with regionalized sensitivity analysis (RSA).

10 2.6.1. Sobol' indices for global variance-based sensitivity analysis (VSA)

11 Sobol' indices enable researchers to distinguish the subset of independent input factors
12 $X = \{x_1, \dots, x_i, \dots, x_k\}$ that account for most of the variance of the model's response $Y = f(X)$ either by
13 themselves (first-order) or due to interaction with other parameters (higher-order). For
14 completeness, here we summarize the efficient Monte Carlo-based scheme presented by Saltelli
15 (2002) to compute first-order and total Sobol' sensitivity indices.

16 The first-order sensitivity index (S_i) represents a measure of the sensitivity of
17 $Y = f(x_1, x_2, \dots, x_k)$ (the RMSE of a model realization evaluated against observations) to
18 variations in parameter x_i . S_i is defined as the ratio of the variance of Y conditioned on the i^{th}
19 factor (V_i) to the total unconditional variance (V):

$$20 \quad S_i = \frac{V_i}{V(Y)} = \frac{V(E(Y | x_i))}{V(Y)} = \frac{\hat{U}_i - \hat{E}^2(Y)}{\hat{V}(Y)} \quad (1)$$

$$21 \quad \text{Where} \quad \hat{U}_i = \frac{1}{n-1} \sum_{r=1}^n f(x_{r1}, x_{r2}, \dots, x_{rk}) f(x'_{r1}, x'_{r2}, \dots, x'_{r(i-1)}, x_{ri}, x'_{r(i+1)}, \dots, x'_{rk}) \quad (2)$$

1 is obtained from products of values of f computed from the sample matrix (n model realizations
 2 long) times values of f computed from another n -realizations matrix where all k parameters
 3 except x_i are re-sampled.

4 The estimates of the mean squared and the total variance are computed as:

$$5 \quad \hat{E}^2(Y) = \frac{1}{n} \sum_{r=1}^n f(x_{r1}, x_{r2}, \dots, x_{rk}) f(x'_{r1}, x'_{r2}, \dots, x'_{rk}) \quad (3)$$

$$6 \quad \hat{V}(Y) = \frac{1}{n} \sum_{r=1}^n f(x_{r1}, x_{r2}, \dots, x_{rk})^2 - \hat{E}^2(Y) \quad (4)$$

7 Instead of computing all $2^k - 1$ terms of the variance decomposition:

$$8 \quad V(Y) = \sum_i^k V_i + \sum_i \sum_{j>i} V_{ij} + \dots + V_{12..k} \quad (5)$$

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

$$15 \quad S_{Ti} = \frac{V(Y) - V(E(Y|x_{-i}))}{V(Y)} = 1 - \frac{\hat{U}_{-i} - \hat{E}^2(Y)}{V(Y)} \quad (6)$$

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

17 is obtained from products of values of f computed from the sample matrix times the values of f
 18 computed from another matrix where only x_i is re-sampled. To spare nk simulations for each
 19 model and each site (and to show that is possible to mine regular Monte Carlo runs, customary

1 used in land-surface modelling), we estimated the second part of this term by interpolating the
2 response surface Y (as explained in Section 2.7.3).

3 A significant difference between S_{Ti} and S_i points to an important role of the interactions
4 of the i^{th} factor (at all orders) in affecting Y (Saltelli et al., 2006). Identification of such parameter
5 interactions can help guide model development. S_{Ti} are also useful to identify input factors that
6 are non-influential, which can help reduce the dimensionality of the parameter estimation
7 problem. If an S_{Ti} is negligible, then it is reasonable to fix that factor to any value within its range
8 of uncertainty, and the dimensionality of the space of input factors or model parameters can be
9 reduced accordingly (van Werkhoven et al., 2009).

10 **2.6.2. Kolmogorov-Smirnov test for regionalized sensitivity analysis (RSA)**

11 In RSA (Hornberger and Spear, 1981; Spear et al., 1994; Bastidas et al., 1999, 2006), the
12 Kolmogorov-Smirnov (K-S) test helps identify which parameters differ when model
13 performance is ‘behavioral’ (B) (‘good’, ‘acceptable’) versus ‘non-behavioral’ (\bar{B}). Note that
14 RSA does not give any information regarding the extent to which a parameter affects the
15 variance of the model output. It instead focuses on which factors are believed to affect the
16 occurrence of behavioral model realizations.

17 The two-sample K-S test (two-sided version) is independently applied to the univariate
18 marginal cumulative density functions (CDF) of each parameter to test whether the distributions
19 of B and \bar{B} are different at a particular significance level. The maximum distance between the
20 two CDFs serves to evaluate the null hypothesis that the parameter values come from the same
21 probability distribution. It is intuitive that, if the CDFs are dissimilar from one another (as well
22 as different from the a priori marginal distribution of that parameter), then such parameter is
23 ‘sensitive’ (Saltelli, 2006).

1 Generally, a set of subjective constraints or thresholds provides a qualitative definition
 2 for B , and model realizations are classified into either B or \bar{B} . In the research presented here, we
 3 considered the calibrated parameter sets (which belong to the joint posterior probability
 4 distribution that maximizes the likelihood function) as representative of B . The \bar{B} sample was
 5 defined by parameter sets whose RMSE is half a standard deviation or more away from the mode
 6 of the cost function. For robustness, for each parameter, several different samples from \bar{B} were
 7 taken and compared to the one of B . We used 5% as the significance level for our RSA.

8 **2.7 Sampling strategies for sensitivity analysis**

9 We generated representative samples of model parameters using Latin Hypercube
 10 Sampling (LHS) and of the behavioral parameter sets through multi-objective calibration. We
 11 also sampled from the response surfaces obtained via LHS using inexpensive Support Vector
 12 Machines (SVM).

13 **2.7.1 Latin Hypercube Monte Carlo sampling (LHS)**

14 We ran a total of 405,000 Monte Carlo simulations sampling random parameter sets
 15 (15,000 samples for each model and site) to obtain a detailed representation of the range of
 16 model responses. We used LHS because it combines the strengths of stratified and random
 17 sampling to ensure that all regions of the parameter space are represented in the sample (McKay
 18 et al., 1979; Helton and Davis, 2003). LHS divides each parameter range into disjoint intervals of
 19 equal probability. From each hypercube, one sample value is randomly taken. We sampled
 20 uniformly within feasible bounds (Table 2). For each sample, we recorded the RMSE of 5
 21 criteria: H, LE, G, Tg, and SMC_{5cm} . To create all the matrices involved in the computation of the
 22 S_i index of the Sobol' method (Eq. 1), we used a modified LHS that enables replication. We

1 sequentially sampled the same parameter value twice from a given hypercube, while all other
2 parameters in the set were chosen according to the LHS.

3 **2.7.2 Multi-objective Markov Chain Monte Carlo parameter estimation technique**

4 We used the efficient Markov Chain Monte Carlo sampling strategy of Vrugt et al.
5 (2003) to approximate the joint posterior distribution of optimal parameters. The simultaneous
6 minimization of the RMSE of multiple criteria $\{H, LE, G, Tg, SMC_{5cm}\}$ allowed us to constrain
7 the models to be consistent with several types of observations and facilitated the identification of
8 the underlying posterior distribution of physically meaningful behavioral parameter sets. It is
9 hoped that sets from the posterior distribution cause the model to mimic the processes it was
10 designed to represent (Gupta et al., 1999; Bastidas et al., 2001; Leplastrier et al., 2002; Hogue et
11 al., 2006). The calibration algorithm runs, in parallel, multiple chains of evolving parameter
12 distributions to provide a robust exploration of the parameter space. These chains communicate
13 with each other through an external population of points, which are used to continuously update
14 the size and shape of the proposal distribution in each chain. This procedure allows an initial
15 population of parameter sets (uniformly sampled within pre-established, feasible ranges) to
16 converge to a stationary sample, which maximizes the likelihood function and fairly
17 approximates the Pareto set. The Pareto set (PS) represents the multi-objective tradeoff: no
18 member of the PS can perform better with respect to one objective without simultaneously
19 performing worse with respect to another competing objective (Gupta et al., 1998). We used a
20 sample of 150 parameter sets to represent the posterior distribution of ‘behavioral’ parameter sets
21 (*B*).

22 **2.7.3 Response surface sampling (meta-modeling) using Support Vector Machines**

23 **(SVM)**

1 To inexpensively compute the second component of the U_i term (Eq. 7) for the Sobol'
2 method (Section 2.6.1), we approximated the response surfaces of the LSMs using SVM trained
3 on the Monte Carlo samples. We used the SVMlight libraries of Joachims (1999) available at
4 <http://svmlight.joachims.org/>.

5 SVM are a robust learning strategy that efficiently generalizes a non-linear functional
6 dependence between a set of multi-dimensional inputs and a vector of outputs, $Y=f(X)$. SVM use
7 kernels to map the complex data into high-dimensional feature spaces in which linear
8 relationships can describe separation planes. Through a sparse set of training examples that
9 'support' the hyperplane that best fits the observed data, SVM create a model to approximate Y
10 (Vapnik, 1998). SVM have been successfully applied in a variety of hydrological applications
11 (e.g., Asefa et al., 2004, 2005, 2006; Khalil et al., 2005; Kaheil et al., 2008a, 2008b). SVM
12 perform better than artificial neural networks (Dibike et al., 2001; Gill et al., 2008), with lower
13 complexity (only 3 parameters), less computational cost (faster training), and reduced sensitivity
14 to noisy inputs and sparse training samples.

15 Based on statistical learning theory (Smola and Schölkopf, 2004), SVM are characterized
16 by a mechanism that avoids overfitting and that results in good generalization. A coefficient C
17 determines the tradeoff between the complexity (or flatness) of the function f and the amount to
18 which deviations larger than empirical error ε are tolerated. In the SVM training, prediction error
19 and model complexity are simultaneously minimized in a quadratic optimization problem, which
20 results in a uniquely global optimum (Vapnik, 1998). We used Gaussian radial basis function
21 kernels and determined near-optimal values of the 3 SVM hyper-parameters (the capacity C , the
22 soft margin ε and the kernel width γ) using the SCE calibration algorithm (Duan et al., 1992).

1 For detailed explanation on the estimation procedure for SVM parameters, we refer interested
2 readers to Khalil et al. (2006).

3 The error in the estimation of the U_i term is small. For example, an SVM ($C=97.12$, $\epsilon=$
4 0.0025 and $\gamma=0.0323$) trained on 5,000 random samples of GW [$Y=RMSE(SMC_{5cm})$] at Site 9
5 (and calibrated against other 5,000 testing samples) predicted the remaining independent
6 validation samples from the Y surface with a RMSE of 0.93 [%]. The estimated variance of Y
7 (8.857) is only 3% different than the ‘true’ variance (8.584) obtained by the Monte Carlo runs.

8 **2.8. Hierarchical clustering for comparisons of parameter distributions**

9 Unsupervised classification of behavioral parameter distributions allowed us to gain
10 understanding about the similarities of the data across locations, specifically about the
11 relationships between types of parameters and sites. We used clustering methods to separate the
12 collection of marginal posterior distributions of calibrated parameters sets into groups.
13 Agglomerative hierarchical clustering methods start with n groups (one object per group) and
14 successively merge the two most similar groups until a single group is left. We used MATLAB’s
15 complete linkage algorithm, in which the maximum distance between objects, one coming from
16 each cluster, represents the smallest sphere that can enclose all objects in the two groups into a
17 single cluster (Hair et al., 1995). Since the distance used to measure dissimilarity between
18 observations (e.g., Manhattan, Euclidean, etc.) may influence the membership of samples to
19 groups, we used the cophenetic correlation coefficient to assess the quality of the linkage
20 (Martinez and Martinez, 2002). We used dendrograms to show the links between the objects as
21 inverted U-shaped lines, whose height represents the distance between the objects.

22

23 **3. Driving questions and experimental design**

1 We first ask: What are the dominant model parameters across the region? We ran a suite
2 of Monte Carlo simulations to identify parameters that exert the greatest control on the
3 variability of simulated fluxes and states at each IHOP site for all 3 models (STD, GW and DV).
4 We quantify sensitivity using RSA and the method of Sobol'. Our SA guides our further
5 investigation.

6 We then address the question: What are the dominant interactions between model
7 parameters, and how do these change between models? With our focus toward model
8 development, we investigate the relationships between behavioral model parameters and quantify
9 how they change between models using the estimates of the total-order sensitivity, the
10 multivariate posterior parameter distributions, and the covariance structures.

11 We finally ask: How do behavioral parameters change with dominant physical
12 characteristics of the land? We summarize the relationships between model parameters and
13 physical characteristics by classifying the multivariate posterior distributions according to types
14 of soil and vegetation. Our classification provides insights into how parameters can be
15 transferred to ungauged locations.

16

17 **4. What parameters are sensitive?**

18 We report the results of our regionalized and variance-based sensitivity analyses, which
19 we use in the next section to help evaluate model behavior.

20 **4.1. Regionalized sensitivity analysis (RSA)**

21 The sensitivity of a given parameter changes between models and varies by location,
22 without an easily recognizable pattern. Table 3 presents the model parameters deemed sensitive
23 at the 5% significance level according to a Kolmogorov-Smirnov (K-S) test between samples

1 that drive behavioral (multi-objectively calibrated) and non-behavioral simulations. A value of 1
2 means the parameter is sensitive, 0 means insensitive. The number of sensitive parameters is
3 tabulated by site and by class. In all versions of Noah LSM, soil parameters tend to be more
4 sensitive than vegetation parameters.

5 At five of the nine sites, GW has the fewest number of sensitive parameters of any of the
6 three models, despite having more total parameters than STD. At the dry sites (1-3), the addition
7 of GW reduces the number of vegetation parameters that are sensitive, perhaps because when
8 GW processes are not included, the vegetation parameters are given more weight in controlling
9 the water balance of the soil. At site 9, despite decreasing the number of sensitive vegetation and
10 soil parameters with respect to STD, GW does not even make the parameters associated with
11 GW sensitive.

12 DV most often has the highest number of sensitive parameters. The number of vegetation
13 parameters at dry sites that are sensitive in DV is significantly higher than at wet sites. Not
14 surprisingly, the conversion factor (*gl*) and the specific leaf area (*sla*) are sensitive everywhere.
15 The fraction of carbon into growth (*fragr*) and the water stress parameter (*wtsr*) are sensitive at
16 eight sites, while the coefficient for soil respiration (*rssoil*), the optical depth (*tauhf*) and the
17 wood allocation parameter (*bf*) are the least often sensitive. The variability of parameter
18 sensitivity between even similar sites perhaps points to a lack of constraint of DV with respect to
19 the partitioning of the carbon budget or suggests that interactions with other parameters
20 decreases the separation of the CDFs, consequently hindering the identification of sensitive
21 parameters (Yatheendradas et al., 2008).

22 The widespread use of RSA in LSM development has been precluded by what has been
23 RSA's lack of firm conclusions because of the relatively poor agreement between the

1 conclusions of RSA studies. Even though we and Bastidas et al. (2006) used roughly the same
2 version of Noah (STD) and allowed many of the same parameters to vary, we found that our
3 RSA is at best partially consistent with the assessment of Bastidas et al. (2006). One possible
4 reason for our divergent assessments is that we did not analyze exactly the same parameters and
5 did not sample from exactly the same ranges as did Bastidas et al. A single parameter's
6 sensitivity, as defined in RSA approaches, is likely a function of the values of other model
7 parameters; alteration of the sampled range of a single parameter can influence the sensitivity of
8 many other parameters (Spear and Hornberger, 1980). Results presented elsewhere (Rosero et
9 al., 2009) have shown that adjusting only a partial set of parameters alters model structure (i.e.,
10 the relationships between parameters); it follows that, with changed model structure,
11 regionalized sensitivity would also vary.

12 Perhaps the most important reason why information obtained through RSA has not been
13 used to inform LSM development is that RSA provides neither information regarding the
14 magnitude of a parameter's influence on model output nor information about the influence that
15 the parameter exerts through interactions. RSA only identifies a subset of parameter sets that
16 improves (if only slightly) model output. A K-S statistic that corresponds to 'high sensitivity'
17 means the parameters used to generate the behavioral model realizations are 'a lot different' from
18 those used to generate the nonbehavioral realizations. 'High sensitivity' does not mean that a
19 parameter is highly important to (exerts significant control over) model output. Furthermore,
20 having invariant parameter distributions in the transition from \bar{B} to B is a necessary but not
21 sufficient condition for insensitivity (Spear and Hornberger, 1980). Quantifying regionalized
22 sensitivity has the potential to allow a mapping of the model sensitivity to parameters back to the

1 parameter space, but without information on the change in model's variance, RSA appears only
 2 useful for limiting the dimensionality of the calibration problem at a specific site.

3 RSA is distinct from the variance-based analysis of sensitivity in model output to
 4 parameter values. The method of Sobol' identifies parameters that exert significant control on the
 5 variance of model output. Sobol' indices are arguably a more useful tool for model developers
 6 and the model's operational users.

7 **4.2 Global variance-based sensitivity analysis (VSA)**

8 Our VSA shows that there are only a few parameters that, by themselves, exert
 9 significant influence on model predictions. In contrast, parameter interaction predominates and is
 10 hence the principal mechanism for sensitivity. Figures 2, 3, and 4 present, for all sites, all
 11 considered parameters, and all models, the Sobol' first-order sensitivity indexes (S_i , which is the
 12 fraction of the total variance of RMSE that can be solely attributed to the i^{th} parameter) and the
 13 residual between Sobol''s total and first-order sensitivity index ($S_{Ti} - S_i$, which is the fraction of
 14 total variance that results from the interaction of the i^{th} parameter with other parameters at all
 15 orders). The easily recognizable pattern of sensitivity can be linked to physical characteristics of
 16 the sites.

17 **4.2.1 First-order sensitivity (S_i)**

18 The identified sensitive parameters (Figs. 2-4) are consistent with our expectations. When
 19 parameter influences changes as we would physically expect, we interpret these results as
 20 consistent with our hypothesis that to a first order, the models are adequately representing the
 21 site-to-site variation in these key components of the water and energy cycles. We observe that,
 22 for most sites and models, the greatest first-order control on simulated top-layer soil moisture is
 23 porosity (*maxsmc*) (Fig.4a). At dry sites 1-3, where direct evaporation is presumably a major

1 component of LE flux, for STD and GW, the bare soil evaporation exponent (*fxexp*) exerts the
 2 most first-order control on soil moisture. The LE flux simulated by GW at dry sites is controlled
 3 by *fxexp* and specific yield (*rous*), which controls depth to the water table. *lai* directly exerts
 4 control on transpiration and hence on the surface energy budget; at the most vegetated sites (7-9),
 5 *lai* consequently shapes the most the variance of H and LE for both STD and GW (Fig 2a, 3a).
 6 The initial value *lai* is not important to determining DV's simulated H and LE because, unlike
 7 the other two models, DV allows the initial value of *lai* to change. Instead, minimum stomatal
 8 resistance (*rcmin*) exerts the most control on DV-simulated LE flux. Two new parameters
 9 associated with DV, *gl* and *sla*, which control the calculation of *lai*, also exert first-order control
 10 on the simulated energy balance. In sparsely vegetated sites (1-3), the Zilintikevich coefficient
 11 (*czil*) plays a significant role in the variance of H.

12 The control of parameters on model variance changes between models and between sites.
 13 The shift in dominant parameters gives clues about how the models are representing the water
 14 and energy cycles from semi-arid (1-3) to semi-humid (7-9) sites. Looking only at STD, we see
 15 that as the mean annual precipitation (MAP) at the sites increases, *fxexp* becomes less important
 16 to determining the top-layer soil moisture, and *refkdt*, a parameter involved in determining
 17 maximum rates of infiltration, becomes more important (Fig. 4a). Expanding our view to include
 18 GW, in which surface runoff is de-emphasized and subsurface runoff is emphasized (see
 19 discussion about GW's preferred modes of operation, Section 5), we see that although *fxexp* is
 20 still a key parameter exerting first-order control on SMC at the dry sites, *refkdt* has little direct
 21 influence on soil moisture at the wet sites. The most sensitive parameter for SMC_{5cm} at sites 1-3
 22 is the aquifer's specific yield (*rous*), which effectively controls whether aquifer water will be
 23 accessible to the near-surface soil. Consistent with our expectations, soil suction (*psisat*), which

1 in GW controls upward movement of water from the aquifer to the soil, has significant control
2 on SMC within GW but not within STD, in which *psisat* exerts less control over soil hydraulic
3 behavior (Fig. 4a).

4 As sites get progressively wetter, surface exchange coefficient *czil* exerts progressively
5 less influence on simulated H, and *rcmin* exerts progressively more influence on H (Fig. 2a). The
6 shift in control is consistent with our expectation that at more vegetated sites, stomatal resistance
7 will be more important to determining the surface energy balance. As a site's MAP increases,
8 *rcmin* and *lai* have increasing control over simulated LE, while *fxexp* becomes less influential
9 (Fig. 3a). Even at dry sites (1-3), DV tends to favor larger values of vegetation fraction (*shdfac*)
10 than are prescribed by STD and GW. As a consequence, DV stands apart from GW and STD in
11 that *fxexp* does not directly contribute to variance of any objective at the three driest sites (with
12 the exception of the unvegetated, bare-soil site 1, where LE is controlled by *fxexp*).

13 Examinations of S_i that are not in line with expectations may be used to help modelers
14 diagnose problems with forcing data and/or model structure. For instance, *fxexp* has the highest
15 S_i of simulated H and LE at site 6. Site 6 receives 800 mm of precipitation each year. We do not
16 expect direct evaporation to be a more significant component of the LE flux at site 6 than at
17 climatically similar sites 4 and 5 or at the semi-arid sites 1-3. The discrepancy implies that our
18 conceptual understanding of the physical processes at site 6 is incorrect, that the model does not
19 adequately represent the physical processes, and/or that our forcing and/or evaluation data are
20 faulty.

21 In contrast to what was observed with RSA, the VSA shows a clear pattern of sensitivity
22 across the sites. Site-to-site variation in the most sensitive parameters is not chiefly governed by
23 soil or vegetation type but, similar to other studies (e.g, Liang and Guo, 2003; Demaria et al.

1 2007; van Werkhoven et al., 2008), appears to be of secondary importance when compared to the
 2 influence of climatic gradient.

3 4.2.2 Sensitivity through interactions (S_I-S_{TI})

4 Figures 2b, 3b, and 4b show that interactions between parameters are responsible for
 5 most of the variance in the models' predicted H, LE, and SMC. If we assume that the
 6 parameterizations are correct, then the significant parameter interaction indicates model
 7 overparameterization (Saltelli et al., 2008). Overparameterization causes increased parameter
 8 interaction (observed here) and decreases the percentage of parameters that are deemed sensitive
 9 by RSA (Bastidas et al., 2006; Yatheendradas et al., 2008) (observed in the RSA described
 10 above). VSA enables the quantification of the significance of parameter interaction and the
 11 identification of the most interactive parameters. Although parameter interaction may not be an
 12 inherently negative trait (e.g., in porous media, we expect hydraulic conductivity and porosity to
 13 be functionally related), especially when there are no known functional relationships between the
 14 physical quantities that developers believe two parameters represent, parameter interaction is
 15 likely to be indication that the model is working in ways that are not consistent with the
 16 conceptual model from which the parameterizations were built.

17 All models exhibit the most parameter interaction at the driest sites, consistent with the
 18 findings of Liang and Guo (2003) and suggesting the need to revise the formulation of all three
 19 models for semi-arid regions (Hogue et al., 2005; Rosero and Bastidas, 2007). Especially for H
 20 and SMC, GW reduces parameter interaction at the middling moisture (4-6) and semi-humid
 21 sites (7-9) (e.g., Fig.5b). GW's reduction of parameter interaction is evidence (although by no
 22 means conclusive) that GW is more realistic than STD at sites 4-9. This is consistent with
 23 observations on the robustness of GW (Gulden et al., 2007). Conversely, GW appears to produce

1 significant parameter interaction at semiarid sites 1-3, indicating a need for an improved
2 parameterization of groundwater processes at semi-arid sites. DV parameters are much more
3 interactive than those of STD and GW, especially at the wettest sites when simulating LE and
4 SMC. The increased interaction between the DV-specific parameters and the rest of the
5 conceptually unrelated STD parameters suggests DV is not functioning as its developers
6 intended. The significant parameter interaction is consistent with the poor robustness of DV
7 (Rosero et al., 2009).

8 There is significant interaction between the vegetation parameters and the sensitivity of
9 soil parameters. These results fully support the findings of Demaria et al. (2007) and Liang and
10 Guo (2003). The significant interplay between soil and vegetation parameters implies that
11 parameter interaction between different model parts changes both the optimal value of a given
12 parameter and the relationships between parameters. In the next section, we use the multivariate
13 posterior distribution of behavioral parameters to characterize the impact of adding new modules
14 on parameter interactions.

15 Looked at in full, we conclude that the models best represent the surface water and
16 energy balance at the intermediate moisture and wet sites, where parameter interaction tends,
17 within a given model, to be lowest. Because it reduces parameter interaction, GW is most likely
18 of any of the three models to be representing the key physical processes with the most realism.

19

20 **5. How do sensitive parameters interact and shape model behavior?**

21 We present a case study in which sensitivity analysis (SA) links model identification and
22 model development. SA identifies parameters whose behavior merits further examination: VSA

1 indicates which parameters are most influential in determining model output; RSA helps us
2 identify modes in parameter distributions that yield distinctive model behavior.

3 **5.1 A case study at Site 7**

4 At site 7, STD, DV, and GW show nearly equivalent performance when using their
5 behavioral parameter sets (Figs. 5, 6). Such ‘equifinality’ is well documented in the hydrologic
6 modeling literature (e.g. Beven and Freer, 2001; Rosero et al., 2009). In this case, distinguishing
7 a ‘best’ model is not trivial. It requires us not only to confront the simulations with observed
8 behavior to test for consistency but to fully understand the underlying model structures (the
9 relationship between parameters) that make STD, GW and DV perform equally well. We show
10 how sensitivity analysis offers the power and the ability to discriminate between model
11 structures that conform to our physical understanding of the systems.

12 **5.2 Focus on sensitive parameters to better understand model function**

13 We follow the impact of shifted preferred values of three ‘physically meaningful’
14 parameters that made considerable contributions to first-order variance: porosity (*smcmax*), the
15 muting factor for vegetation’s effect on thermal conductivity (*sbeta*), and minimum stomatal
16 resistance (*rcmin*). The fundamental implication of our observations is that although the different
17 optimal values of parameters are important (as found during model identification), the change in
18 functional relationship between the parameters (the information contained in the interactions) is
19 most relevant for purposes of model development.

20 Figure 7 shows the marginal CDF of the posterior multivariate distribution for selected
21 sensitive parameters in the behavioral range at Site 7. Along with the CDFs, box plots show the
22 median and interquartile ranges of the skewed parameter distributions. The scatterplots show the
23 variation in RMSE of LE and SMC_{5cm} along the range of plausible values of the parameter. The

1 trend is emphasized by fitting the points with a polynomial of minimal complexity. Even a
 2 cursory analysis of Figure 7 shows that the direction of ‘sensitivity’ (understood as the rate of
 3 change of score with parameter value) changes between models (e.g., Fig. 7a). The simulation of
 4 SMC_{5cm} by STD and DV degrades as porosity increases, while GW improves. We also note that,
 5 along the plausible range, the response can be enhanced (Fig. 7d) or become relatively
 6 insensitive to changes in parameter value (Fig. 7c). The identifiability of parameters (understood
 7 as having a clearly defined local minimum) changes between models. For example, in DV, there
 8 is a clear low point of the RMSE of LE along the range of values of the maximum water-holding
 9 capacity of the canopy (cm_{max}), but STD and GW have less of a preference (Fig. 7c). The
 10 interquartile range of rc_{min} of STD is smaller than that of GW or DV (Fig. 7b). We can clearly
 11 see that different models have distinct optimal parameter values for the same physical parameter,
 12 implying not only that the parameters cannot be transferred between models but that the
 13 relationships between them are different.

14 **5.2.1 The role of porosity ($maxsmc$)**

15 In all three versions of Noah, higher values of $maxsmc$ tend to decrease direct
 16 evaporation from the first soil layer (E_{DIR}). E_{DIR} is estimated as the product of Penman’s potential
 17 evaporation (ET_{pot}), the complement of the vegetated fraction ($shdfac$), and the ratio of top-layer
 18 volumetric soil moisture (SMC_1) to $maxsmc$:

$$19 \quad E_{DIR} = ET_{pot} (1 - shdfac) \left(\frac{SMC_1 - SMC_{dry}}{maxsmc - SMC_{dry}} \right)^{fxexp} \quad (8)$$

20 SMC_{dry} is the lowest volumetric water content that can exist in the top soil layer, and $fxexp$ is a
 21 parameter ranging from 0.2 to 4.

1 In STD and DV, the error in simulating LE tends to be relatively small when *maxsmc* is
 2 low and relatively large when *maxsmc* is high (Fig. 7a). However, GW shows slight
 3 improvement in simulating LE as *maxsmc* increases. The tendency of STD and DV to simulate
 4 LE well when *maxsmc* is low (and direct evaporation from the soil consequently tends to be
 5 high) implies that STD and DV often underestimate direct evaporation at site 7. Given the same
 6 *maxsmc* value, GW more easily simulates sufficient direct evaporation, perhaps because of
 7 wetter soil (Rosero et al., 2009).

8 In STD and DV, *maxsmc* controls both surface and subsurface runoff. Hydraulic
 9 conductivity of the soil layer (*wcnd*) is computed using Clapp and Hornberger's method, which
 10 scales saturated hydraulic conductivity (*dksat*) by wetness ($SMC / maxsmc$), raised to an
 11 exponent determined by the Clapp and Hornberger parameter (*b*):

$$12 \quad wcnd = dksat \left(\frac{SMC}{maxsmc} \right)^{2b+3} \quad (9)$$

13 Lower *maxsmc* yields higher *wcnd*, which means water moves through the soil more quickly.
 14 Note that, in the case of parameterizing subsurface runoff, hydraulic conductivity is a
 15 representation of the speed at which the water can move laterally through the soil. In STD and
 16 DV, subsurface runoff (*Runoff2*) is *wcnd* times the slope. Consequently, higher values of *maxsmc*
 17 decrease *Runoff2*. Higher values of *maxsmc* also decrease surface runoff (*Runoff1*) by increasing
 18 the maximum rate of infiltration. Both changes increase the wetness of the surface soil.

19 GW changes the way runoff is computed; *maxsmc* does not control surface or subsurface
 20 runoff in GW, eliminating two of the three ways in which *maxsmc* controls soil moisture.

21 *Runoff2* is represented as an exponential function of depth to water (Niu et al., 2007):

$$22 \quad Runoff2 = rsbm x e^{-ff * Z_{wr}} \quad (10)$$

1 Where $rsbmx$ is the maximum rate of subsurface runoff, fff is the e-folding depth of saturated
 2 hydraulic conductivity, and Z_{WT} is a variable that represents the depth to the water table. $Runoff1$
 3 is a modified version of the same function used to compute subsurface runoff (Niu et al., 2005):

$$4 \quad Runoff1 = pcpdrp * (fsatmx e^{-0.5 * fff * Z_{WT}}) \quad (11)$$

5 Where $pcpdrp$ is the effective incident water and the second term is the fraction of unfrozen grid
 6 cell that is saturated.

7 In STD (and DV), $maxsmc$ couples two physically unrelated (or very weakly related)
 8 processes (direct soil evaporation and lateral surface and subsurface runoff). GW decouples these
 9 processes by eliminating the dependence of parameterized lateral runoff on $maxsmc$. This
 10 decoupling reduces the spurious parameter interaction of $maxsmc$ and, within GW, virtually
 11 eliminates the tradeoff between good simulation of LE and SMC5cm. GW is, in this regard, a
 12 better model for simulating fluxes at site 7.

13 The question remains – why does GW poorly simulate SMC when $maxsmc$ increases?
 14 $maxsmc$ is used to compute vertical hydraulic conductivity (using the same function described
 15 above), which GW uses to control the flow of water between the aquifer and soil down a
 16 hydraulic gradient. Higher $maxsmc$ yields lower hydraulic conductivity, which, in addition to
 17 decreasing the transfer of water between layers within the soil column, decreases the
 18 communication between the aquifer and the soil profile (that is, it decreases the flow of water
 19 between the two, increasing the potential for water to be retained near the surface). At site 7, GW
 20 best simulates SMC when high vertical hydraulic conductivity connects the aquifer and soil.

21 Consistent with the work of others (e.g., Demaria et al., 2007), parameter values and
 22 model sensitivity to $maxsmc$ are not consistent between sites along a climatic gradient or even
 23 within a set of sites with similar characteristics. Conclusions about model performance are

1 therefore difficult to generalize. This lack of continuity of behavior between sites is consistent
 2 with at least one of the following possibilities: (1) model parameterizations do not represent key
 3 aspects of the system and/or (2) our multi-objective calibration provided insufficient constraint
 4 for the estimation of behavioral parameters. We recommend use of runoff constraints before
 5 drawing firm conclusions regarding the physical reality of the runoff-related processes in GW.

6 **5.2.2 The role of SBETA**

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

$$8 \quad G = DF_1 \frac{STC_1 - T_1}{0.5 * ZSOIL_{(1)}} \quad (12)$$

9 In which STC_1 is the temperature at the center of the first soil layer ($0.5 * ZSOIL_{(1)}$) and T_1 is the
 10 surface temperature. DF_1 is the heat conductivity of the top soil layer.

11 Noah assumes that, as vegetation cover increases, heat flux into the ground decreases. It
 12 uses parameter $sbeta$ and the vegetated fraction ($shdfac$) to mute the thermal conductivity of the
 13 top layer of the soil (DF_1) by:

$$14 \quad DF_1 = DF_1 * e^{sbeta * shdfac} \quad (13)$$

15 At site 7, the mode of the posterior probability distribution of all three models is near the
 16 bound of the explored parameter range (-1) (Fig. 7d). The preference for near-bound values is
 17 more pronounced in DV, which at site 7 tends to have $shdfac$ values near 1.0, which puts
 18 downward pressure on the value of $sbeta$. The skewed posterior parameter distribution for all
 19 suggests that an even-less-negative value of $sbeta$ would have yielded better results at site 7.

20 The assumption that vegetation necessarily decreases the thermal conductivity of the top
 21 layer of the soil may be incorrect. If the ‘vegetation effect’ on thermal conductivity is real, the
 22 model underestimates the top-layer soil thermal conductivity. At site 7 (and at several other

1 sites), there is a clear tradeoff between H and G that is mediated by the thermal conductivity
 2 which hints at the need for revised process understanding.

3 When comparing site 7 simulations to those of the other two ‘wet’ sites (8 and 9), we see
 4 a roughly consistent preference for near-zero values of *sbeta*. At the drier sites (1-6), the model’s
 5 strong preference for near-zero values of *sbeta* is less obvious; however, *shdfac* is closer to zero
 6 at these sites, which lowers the value of the muting factor described by Eq.13.

7 **5.2.3 The role of minimum stomatal resistance (*rcmin*)**

8 *rcmin* controls much of the variance in H and LE, especially at wetter sites. As *rcmin*
 9 increases, the ratio of actual to potential evapotranspiration decreases. *rcmin* has a more
 10 consistent influence on the variance of H than on LE.

11 At site 7, all three models prefer a low *rcmin* (Fig. 7b), which increases LE for a given
 12 potential evapotranspiration; however, GW and DV show decreased identifiability of *rcmin*. The
 13 mode of the *rcmin* distribution is higher for GW than for STD, perhaps because GW tends to
 14 have a wetter soil and more robust LE. The spread of the posterior distribution of *rcmin* for DV
 15 is significantly larger than that for STD, although both models share the same mode. This
 16 decrease in identifiability of parameters functionally related to *lai* (as is *rcmin*) is consistent with
 17 the added degrees of freedom allowed by DV (DV parameters *gl* and *sla* are most important in
 18 predicting *lai* [Fig. 2]). Because DV simulations include a wider spread of *lai* states, they also
 19 have a wider spread of ‘good’ *rcmin* values.

20 **5.3 What changes in GW to make it work better than or as well as STD at Site 7?**

21 Figure 8 shows the bivariate posterior distribution of selected behavioral parameters of
 22 STD and GW. The response surface of RMSE SMC_{5cm} changes between STD and GW (e.g., see
 23 *maxsmc* vs. *psisat*). For GW, the shape of the posterior distributions of soil parameters that are

1 shared with STD is significantly different because of interaction with the GW parameters and
 2 module. Such shifts in model function affect the model covariance structure (Table 4).

3 After multi-objective parameter estimation, at site 7, GW functions in one of two
 4 preferred modes (Fig. 8b). In the first, slightly preferred mode (*m1*), the parameters work
 5 together to help GW function as the developers likely intended. Strong communication between
 6 the aquifer and the soil column is supported by relatively high values of saturated hydraulic
 7 conductivity (*satdk*), low values of the reciprocal of the e-folding depth of hydraulic conductivity
 8 (*fff*), and low porosity (*maxsmc*). A relatively low surface runoff scaling factor (*fsatmx*) and
 9 relatively high subsurface runoff scaling factor (*rsbmx*) ensure that subsurface runoff dominates
 10 surface runoff. Mimicking nature, high soil suction (*psisat*) pulls water upward. A high aquifer
 11 specific yield (*rous*) deepens the water table (weakening the direct influence of the saturated
 12 zone on the model soil column) and transforms more water to runoff rather than to recharge.

13 In the second mode (*m2*), GW adopts parameter values that make the model work as one
 14 would expect STD to function (i.e., the model functions with parameters that render GW
 15 nonfunctional) (Fig. 8b). Relatively high values of the reciprocal of the e-folding depth of
 16 hydraulic conductivity (*fff*) effectively seal the bottom of the soil column, limiting
 17 communication between the aquifer and the soil column; high porosity (*maxsmc*) decreases the
 18 vertical conductivity, further inhibiting the already poor communication between the soil and
 19 aquifer. High porosity favors decreased direct evaporation. Surface runoff is augmented by a
 20 relatively high surface runoff scaling factor *fsatmx*; subsurface runoff is lessened by the
 21 relatively low subsurface runoff scaling factor (*rsbmx*).

22 These alternative behaviors are a possible explanation for the issue identified by Rosero
 23 et al. (2009), who showed that despite very good performance of calibrated GW, the model

1 suffered from low robustness (high sensitivity to unmeasurable, errant parameters). The bimodal
 2 behavior of GW highlights the need to constrain GW with runoff observations, which will help
 3 identify a more consistent, better model.

4 **5.4 What changes in DV to make it work better than or as well as STD at site 7?**

5 DV and STD differ functionally in two ways: 1) DV predicts (rather than prescribes)
 6 *shdfac* as a function of environmental variation in moisture and radiation availability, and 2)
 7 rather than use a temporally constant, prescribed *lai*, DV uses *shdfac* to predict *lai* variation
 8 using a functional relationship:

$$9 \quad lai = \max \left(xlai_{\min}, \frac{-1}{gl} \log^{-1}(1 - shdfac) \right) \quad (14)$$

10 The vegetated fraction affects all three components of LE: vegetation shades the soil,
 11 modulating direct evaporation (E_{dir}); vegetation retains water above the soil, contributing to
 12 evaporation from the canopy (E_{c}); vegetation fuels transpiration (E_{transp}). Because of *shdfac*'s role
 13 in scaling all components of evapotranspiration, a high value of conversion parameter *gl* yields a
 14 regime in which E_{c} and E_{transp} are strongly favored over E_{dir} . Low values of *gl* fix *shdfac* near
 15 zero and promote a regime in which E_{dir} is the dominant component of LE. When *shdfac* is near
 16 zero, both E_{c} and E_{transp} are minimized. At sites with sufficient vegetation, DV enables the model
 17 to correctly give more weight to E_{transp} . When there is little vegetation (e.g., at sites 1-3), the
 18 coupling to DV may be failing to consider special water use features associated with the semi-
 19 arid vegetation (Unland et al., 1996).

20 STD keeps *shdfac* fixed at monthly climatological values (~0.7 at site 7) and does not
 21 relate *shdfac* to *lai*. STD, unable to change the value of *shdfac* to shift the balance of components

1 of LE, favors higher *lai* (which decreases stomatal resistance and increases E_{transp}) as means for
 2 increasing total LE.

3 When compared to STD, DV can achieve ‘good’ model performance using a wider range
 4 of values for *shdfac* and *lai*. We see this decreased identifiability of DV parameters when
 5 comparing the bivariate posterior parameter distributions of STD to those of DV at site 7 (Fig.
 6 9). The identifiability in the response surface of RMSE LE has changed (e.g. *lai* vs. *rcmin*). The
 7 decrease in identifiability of parameters that are functionally related to *shdfac* and/or *lai* can be
 8 seen across the IHOP sites (results not shown).

9 The interplay of the parameters of the DV module also leads to changes in parameter
 10 densities of STD and DV (Fig. 9). We see additional evidence for increased interaction between
 11 parameters in DV when we note that the models’ covariance structure has been altered (Table 5).
 12 For example, *rcmin* and *maxsmc* are positively correlated in STD, but in DV they have a very
 13 slight negative correlation.

14 Although the increased flexibility of *lai* and *shdfac* values may improve the model’s
 15 simulation of seasonal and interannual variation in surface fluxes, over time scales examined
 16 here, DV does not appear to improve the model. The constraints imposed by the turbulent and
 17 near-surface states may be insufficient for the complexity of the model and/or DV may need to
 18 be constrained with observations of carbon fluxes and plant growth. The function of the DV
 19 module may be hindered by Noah’s lack of a separate canopy layer (Rosero et al., 2009) or the
 20 absence of a more complex Ball-Berry type of stomatal conductance formulation (Niu et al.,
 21 2009).

22

23 **6. What does sensitivity analysis tell us about transferability?**

1 LSM parameters are frequently assumed to be physical quantities, which can be
2 measured and which have strong relationships with physical properties of the system. These
3 assumptions imply that optimal values should not vary between models and that relationships
4 between model parameters should be constant between models and between sites with similar
5 physical properties. By making sets of vegetation-related parameters functions of vegetation
6 type, LSMs contain the implicit assumption that vegetation type solely determines the ideal
7 values of vegetation parameters. A priori assignment of parameter values based on a site's
8 physical characteristics ('parameter transferability') depends on the above assumption. Our VSA
9 shows that interaction between parameters contributes most to a LSM's total variance causing
10 optimal parameter values to change significantly between models and sites. If parameters are
11 transferable between similar sites, then the relationships between parameters should also be
12 transferable between similar sites. The joint multivariate posterior distribution summarizes much
13 of the information regarding the relationships between model parameters (model structure) at a
14 particular location given observed datasets. We use the stable posterior distributions of the
15 behavioral parameter sets to test the assumption that parameters and parameter relationships
16 relate directly to physical characteristics. We also evaluate the extent to which climate
17 determines the similarity of parameters between locations.

18 **6.1 Testing parameter transferability between sites using soil and vegetation types**

19 If parameters were readily 'transferable' between sites solely based on the vegetation
20 class, we would expect the distributions of the vegetation parameters at two sites with the same
21 vegetation type but different climatic regime (e.g., sites 2 and 8) to be more similar than the
22 distributions of the same parameters at two sites with different vegetation but similar climate
23 (e.g., sites 2 and 1). Using evidence from the marginal posterior parameter distributions of

1 behavioral, sensitive, ‘physically meaningful’ vegetation parameters (*r_{min}*, *lai*, and *sbeta*),
 2 Figure 10 shows that this expectation does not hold. Figure 10 contrasts the vegetation
 3 parameters’ posterior distributions for sites 2 and 8 (grassland) against the posterior distributions
 4 for contiguous sites 1 and 2 (dry). The distributions of *r_{min}*, *r_{smax}*, *lai* and *z₀* are more similar
 5 between sites with similar climate than they are between sites with the same vegetation (Fig. 10a,
 6 10b). *hs* and *cm_{cmax}* show a similar lack of transferability. Only *sbeta* shows ‘transferability’
 7 (i.e., there are smaller differences between the distributions from sites with the same vegetation
 8 cover) for all models (Fig. 10c). Parameter *cfactr* does show transferability, but only for STD.
 9 *rgl* could be considered ‘transferable,’ but only for DV and GW.

10 Using the IHOP dataset, we cannot test parameter transferability using two sites with the
 11 same soil type but different climatology. We instead tested whether the distributions of soil
 12 parameters transfer along with soil or vegetation characteristics between these sites. Only *fxexp*
 13 appears ‘transferable’ between dry sites (1 and 2) with the same soil type (sandy clay loam) in
 14 STD and GW (Fig. 11b). Some parameters have more similar distributions between sites 2 and 8
 15 than between sites 1 and 2: *psisat* and *satdk* only for GW, Clapp and Hornberger’s *b* only for
 16 STD, *maxsmc* for STD and GW (Fig. 11a), and *czil* for STD and DV (Fig. 11a). Perhaps this is
 17 because, despite sharing the same soil type, site 1 received more than twice the amount of
 18 precipitation observed at site 2. We also hypothesize that transferability based on soil or
 19 vegetation type is not uniformly viable because of the interactions between soil and vegetation
 20 parameters, which condition the shape of the distributions. The interactions between the soil and
 21 vegetation parameters at grassland site 2 are significantly different than the interactions at
 22 unvegetated site 1.

1 We recognize that the case studies above are by no means conclusive, but they do not
 2 support the hypothesis that parameters are transferable solely based on vegetation type. The
 3 results instead suggest that land-surface model parameters are more sensitive to climatic forcing
 4 than to a specific soil or land-cover classification. Such results provide support to similar
 5 observations made about other hydrologic models (Demaria et al., 2007; van Werkhoven et al.,
 6 2008) and made about the Noah LSM and using single optimal parameter sets (Hogue et al.,
 7 2005; Rosero and Bastidas, 2007) .

8 **6.2 Synthesizing sensitivity to site, soil and vegetation classes by means of clustering**

9 In order to more quantitatively synthesize knowledge gained through sensitivity analysis
 10 for use at ungauged locations, we build upon the aforementioned idea of comparing the
 11 similarity of parameter distributions across sites (Rosero and Bastidas, 2007) by complementing
 12 the approach with unsupervised, agglomerative hierarchical clustering methods.

13 For each IHOP site, we have obtained a stable, multivariate probability distribution χ of
 14 behavioral parameter sets $X = \{x_1, x_2, \dots, x_i, \dots, x_k\}$ using multi-objective parameter estimation.
 15 The marginal probability distribution for the i^{th} parameter is χ_i . To circumvent comparing
 16 between sites against each other (two at a time, as done above), we define a triangular probability
 17 distribution D_i as a reference distribution for each parameter. $D_i=1$ when the value of parameter
 18 x_i is the “default” for the site. $D_i=0$ when x_i is at either edge of the feasible range. This step
 19 allows us to introduce the assumption that the parameters relate to soil and vegetation types.

20 For each parameter, and at each site, we quantify the closeness between the cumulative
 21 distribution of the ‘optimal’ values of x_i (i.e, the marginal χ_i) and the reference D_i . We use the
 22 Hausdorff norm to quantify the difference $\chi_i - D_i$. For each model, the matrix of ‘signatures’ of
 23 the marginal distributions of k parameters at all the n evaluation sites is:

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

2 S can be used to identify groups of parameters that are similar between locations or to identify
 3 locations where groups of parameters behave alike. We use the unsupervised, agglomerative
 4 hierarchical clustering algorithm (described in Section 2.8) to find these groups without making
 5 any further assumptions about the number of groups.

6 The Noah LSM works with three classes of parameters: parameters associated with soil
 7 type (soil parameters), parameters associated with vegetation cover type (vegetation parameters),
 8 and general parameters. If the previously described assumption of parameter transferability based
 9 on site characteristics holds (and if IHOP vegetation classifications are correct), then, given the
 10 set of signature vectors created using the set of vegetation parameter distributions $S(x_{veg,1..n})$, a
 11 clustering procedure should be able to classify similar sites in groups that resemble the IHOP
 12 vegetation type groupings (Table 1). Similarly, clustering of $S(x_{soil,1..n})$ would result in sites
 13 grouped according to the IHOP soil type classification (Table 1).

14 Applying a suite of distance metrics (e.g., Manhattan, Euclidean, Cosine, etc), neither soil
 15 nor vegetation parameters render groups of sites that partition solely based on the expected soil
 16 or vegetation classifications. Figure 12a shows the classification tree (dendrogram) for STD
 17 using the Euclidean distance, which maximizes the cophenetic correlation coefficient of the
 18 linkage (also shown). None of the distance metrics allowed us to classify $S(x_{veg,1..n})$ by location
 19 in a way that matched the IHOP vegetation classifications. Given the subset $S(x_{soil,1..n})$, composed
 20 of the signature vectors of the 10 soil parameters at all sites, classification of the IHOP sites
 21 according to soil characteristics was also not feasible (Fig. 12b). Using signature vectors for
 22 STD, GW, and DV, some (but not all) of the distances related sites 7, 8, and 9 as having the

1 same soil and same vegetation type (although, because they also share the same climate type, we
2 are unable to definitively attribute such classification to shared vegetation type). The rest of the
3 sites do not strongly coalesce according to physical properties. For example, the pasture sites are
4 not distinctively together; sites 5 and 6 (wheat crops) are never classified together according to
5 vegetation (Fig 12a). Sites 1 and 2 (sandy clay loam) and sites 4 and 5 (loam) do not cluster
6 together using soil parameters (Fig 12b). These results support the earlier findings presented
7 here, which suggest that interaction between soil and vegetation parameters is significant, to the
8 point that it shapes the posterior parameter distributions. These results also suggest that soil and
9 vegetation type are not, by themselves, good physical characteristics by which to transfer
10 parameters.

11 Next, to account for the interdependence between soil and vegetation parameters, we
12 classified the entire matrix $S(x_{\text{soil}}, x_{\text{veg}})$. If parameters can be transferred based on shared
13 vegetation and soil type, then the clustering of the entire matrix should identify groups of sites
14 with the same vegetation and soil type (e.g, sites 7-9). Figure 12c shows a pattern (found with
15 several distances) that is consistent across models: sites 7-9 cluster together. Sites 7, 8, and 9
16 have also similar climates, and the classification of the sites shows strong resemblance to the
17 climatic gradient. Given this dataset, we cannot disprove the contention that parameters can be
18 transferred between sites that have both the same vegetation type and the same soil type.

19 If we instead cluster S looking for groups of parameters, the expectation is that x_{soil} will
20 behave as a whole in a similar way across sites. In other words, one can produce a map of
21 sensitivity to characterize which parameters are most similar to their default (prior distribution)
22 and which are not. Figure 13 shows representative groupings of the behavioral, marginal
23 posterior distributions of STD and GW parameters at all sites. Again, using a suite of distances,

1 we were unable to identify definitive clusters of soil and vegetation parameters within the set of
2 signature vectors S , meaning that individual parameters are not sensitive in groups that relate
3 primarily to soil or vegetation, but to a combination of both. The new GW parameters do seem to
4 behave in a way that is similar to other soil parameters (Fig. 13b), which informs the estimation
5 of these parameters for distributed applications.

6 We conclude that the primary site-to-site control on the parameters identified as
7 ‘vegetation’ or ‘soil’ parameters is not a site’s type of soil or vegetation used in isolation. This is
8 consistent with the notion that given that parameters used by LSMs must represent physical
9 processes across multiple scales, the parameters become “effective” rather than physically
10 derived quantities (Wagener and Gupta, 2005) and that interaction between classes of parameters
11 are very important. Our clustering analysis suggests that climate is a major control of site-to-site
12 variation in parameter values and supports recommendations that mean climate be considered
13 when transferring parameter values between sites (Liang and Guo, 2003; Demaria et al., 2007;
14 vanWerkhoven et al., 2008).

15

16 7. Summary and Conclusions

17 We combine two powerful sensitivity analysis methods, regionalized sensitivity analysis
18 (RSA) and global variance-based sensitivity analysis (VSA), to inform model identification and
19 development. We draw conclusions regarding LSM development and model assessment
20 practices, the functioning of three versions of the widely used Noah LSM, and the a priori
21 assignment of parameter values. Our work yields several conclusions that can be generalized to
22 all LSM and to other environmental models and several others that are specific to the Noah
23 LSM.

1 We show that VSA complements RSA for the purposes of model development. We
2 perform VSA and RSA on a single set of Monte Carlo runs. When used together, VSA and RSA
3 help developers understand model behavior (i.e., how and when different modules interact, how
4 a model varies in performance between climate regimes, how a model performs differently
5 between sites, how different models perform at the same site, etc.). SA identifies parameters
6 whose behavior merits further examination: VSA identifies parameters responsible for model
7 uncertainty; RSA identifies modes in parameter distributions that yield distinctive behavior.

8 We show that the clear patterns of parameter importance identified by VSA are consistent
9 with site-to-site variation in climate and with model-to-model changes in physical
10 parameterization. VSA shows that parameter interactions within models exert significant control
11 on model variance, and we show that the interactions can be traced using RSA. Shifts in
12 parametric control on variance and covariance hint at whether a model represents the water and
13 energy cycles in a way that is consistent with expectations. Although the optimal value of a
14 parameter is useful information, the change in the functional relationship between parameters is
15 more relevant for model development and hypothesis testing. We assert that to advance model
16 development, VSA should be used in conjunction with RSA.

17 Transfer of parameters based solely on shared vegetation type or on shared soil type is
18 not a viable method for a priori parameter assignment. The work presented here shows that
19 vegetation type and soil type are not the most significant contributors to site-to-site variance in
20 optimal parameter values. Interaction between soil and vegetation parameters is significant and
21 varies between sites; parameter interaction at least partially explains why transfer of parameters
22 based solely on shared vegetation or soil type does not work. The primary factor controlling site-
23 to-site variation in parameters is likely climate, although, given the dataset used here, the

1 combination of a site's vegetation and soil type or some unidentified factor cannot be ruled out
2 as the dominant controlling factor. The lack of viability of parameter transfer based solely on
3 soil and vegetation type is a conclusion that has significant implications for the field of regional
4 and global land surface modeling, which depends on parameter transfer based on stand-alone
5 vegetation type and soil type as a means for a priori parameter assignment.

6 Looking specifically at the performance of the three versions of the Noah LSM used here
7 (STD, GW, and DV), we make several non-site-specific conclusions regarding model behavior.
8 All three models exhibit significant parameter interaction, indicating that the models are
9 overparameterized and/or underconstrained. All three show the least parameter interaction at the
10 middling-moisture and wet sites and the most parameter interaction at the three driest sites. This
11 difference suggests a need for reformulation of Noah LSM such that semi-arid regions are more
12 realistically represented. On the whole, GW has less parameter interaction than STD (except at
13 dry sites), indicating that it represents land-surface system with the most realism of any of the
14 three models. GW is also least sensitive to errant parameters at the wettest sites (where
15 groundwater is likely the most influential). DV has much more parameter interaction than STD,
16 which provides evidence that the model is not performing as its developers intended, does not
17 add value to STD, and/or requires additional constraint. Specific to site 7, we make the following
18 observations: STD and DV tend to underestimate direct evaporation from the soil; GW does not
19 (maybe because of wet soil). The assumption that vegetation decreases the thermal conductivity
20 of the top layer of the soil is not well supported by data (this conclusion can be roughly
21 generalized to other sites, especially the wet sites). At site 7, GW functions in one of two modes
22 – the slightly preferred mode works in a way that mirrors what the developers likely intended;
23 the second mode makes GW function as one might expect STD to work. Constraining runoff

1 may isolate the more realistic mode. GW has less spurious parameter interaction in part because
2 it decouples direct evaporation and subsurface runoff (which are coupled via porosity in STD
3 and DV). This decoupling appears to make the model function more realistically, with less
4 tradeoff between the simulation of soil moisture and LE. Adding modules (DV, GW) decreases
5 the identifiability of minimum stomatal resistance, although all three models prefer low
6 minimum stomatal resistance (thus increasing LE for a given set of conditions). Across several
7 sites, DV functions in one of two modes: the first emphasizes direct soil and canopy evaporation
8 over transpiration; the second emphasizes transpiration over direct evaporation from the soil and
9 canopy.

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

16
17 ***Acknowledgements*** We thank Pedro Restrepo at OHD/NWS, Dave Gochis at NCAR
18 and Ken Mitchell at NCEP for their insight. We appreciated suggestions by M. Bayani Cardenas,
19 Charles S. Jackson and Yasir H. Kaheil. We acknowledge the International H₂O Project for the
20 datasets. We benefited from the computational resources at the Texas Advanced Computing
21 Center (TACC). This project was funded by the Graduate Fellowship of the Hydrology Training
22 Program of the OHD/NWS, the NOAA grant no. NA07OAR4310216, NSF, and the Jackson
23 School of Geosciences.

24

1 REFERENCES

- 2 Abramowitz, G., Pitman, A. J., Gupta, H., Kowalczyk, E. and Wang, Y., (2009). On the need for a
3 biophysically based benchmark for land surface models, *J. Hydrometeor.*, accepted.
- 4 Asefa, T., M. W. Kemblowski, G. Urroz, M. McKee, and A. Khalil (2004), Support vectors–based
5 groundwater head observation networks design, *Water Resour. Res.*, *40*(11), W11509,
6 doi:10.1029/2004WR003304
- 7 Asefa, T., M. Kemblowski, U. Lall, and G. Urroz (2005), Support vector machines for nonlinear state
8 space reconstruction: Application to the Great Salt Lake time series, *Water Resour. Res.*, *41*,
9 W12422, doi:10.1029/2004WR003785.
- 10 Asefa, T., Kemblowski, M. W., McKee, M., & Khalil, A. (2006). Multi-time scale stream flow
11 prediction: The support vector machines approach. *J. Hydrol.*, *318*, 7–16.
- 12 Bastidas, L.A., H.V. Gupta, S. Sorooshian., W. J. Shuttleworth, and Z. L. Yang, 1999: Sensitivity analysis
13 of a land surface scheme using multicriteria methods, *J. Geophys. Res.*, *104*(D16), 19,481–19,490
- 14 Bastidas, L.A., H.V. Gupta, and S. Sorooshian., 2001: Bounding the parameters of land-surface schemes
15 using observational data. *Land surface hydrology, meteorology, and climate: Observations and*
16 *Modeling*. V. Lakshmi et al. (ed.), Water Science and Application, Vol.3, AGU, 65–76.
- 17 Bastidas, L.A., T. S. Hogue, S. Sorooshian, H. V. Gupta, and W. J. Shuttleworth, 2006: Parameter
18 sensitivity analysis for different complexity land surface models using multicriteria methods, *J.*
19 *Geophys. Res.*, *111*, D20101, doi:10.1029/2005JD006377.
- 20 Beven, K. J., and Freer, J., 2001: Equifinality, data assimilation, and uncertainty estimation in mechanistic
21 modelling of complex environmental systems, *J. Hydrol.*, *249*, 11–29
- 22 Beven, K. J. and Freer, J., 2001: Equifinality, data assimilation, and uncertainty estimation in mechanistic
23 modelling of complex environmental systems, *J. Hydrol.*, *249*, 11-29.
- 24 Chen, F., and J. Dudhia, 2001: Coupling an Advanced Land Surface–Hydrology Model with the Penn
25 State–NCAR MM5 Modeling System. Part I: Model Implementation and Sensitivity. *Mon. Wea.*
26 *Rev.*, *129*, 569–585.
- 27 Cosgrove, B. A., et al. (2003), Real-time and retrospective forcing in the North American Land Data
28 Assimilation System (NLDAS) project, *J. Geophys. Res.*, *108*(D22), 8842, doi:10.1029/2002JD003118.
- 29 Demaria, E. M., B. Nijssen, and T. Wagener, 2007: Monte Carlo sensitivity analysis of land surface
30 parameters using the Variable Infiltration Capacity model, *J. Geophys. Res.*, *112*, D11113,
31 doi:10.1029/2006JD00
- 32 Demarty, J., C. Otlé, I. Braud, A. Olioso, J. P. Frangi, L. Bastidas, and H. V. Gupta, 2004: Using a
33 multiobjective approach to retrieve information on surface properties used in a SVAT model, *J. Hydrol.*,
34 *287*, 214–236
- 35 Dickinson, R.E., M. Shaikh, R. Bryant, and L. Graumlich, 1998: Interactive Canopies for a Climate Model.
36 *J. Climate*, *11*, 2823–2836.
- 37 Dibike, B. Y., S. Velickov, D. Solomatine, and B. M. Abbot (2001), Model induction with support vector
38 machines: Introduction and applications, *J. Comput. Civ. Eng.*, *15*, 208–216
- 39 Duan, Q., V. K. Gupta, and S. Sorooshian (1992), Effective and efficient global optimization for
40 conceptual rainfall-runoff models, *Water Resour. Res.*, *28*(4), 1015–1031
- 41 Ek, M.B., et al., 2003: Implementation of Noah land surface model advances in the National Centers for
42 Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res.*, *108*(D22), 8851,
43 doi:10.1029/2002JD003296.
- 44 Gill, M. K., T. Asefa, Y. Kaheil, and M. McKee (2007), Effect of missing data on performance of
45 learning algorithms for hydrologic predictions: Implications to an imputation technique, *Water*
46 *Resour. Res.*, *43*, W07416, doi:10.1029/2006WR005298
- 47 Gulden, L.E., E. Rosero, Z.-L. Yang, M. Rodell, C.S. Jackson, G.-Y. Niu, P. J.-F. Yeh, and J. Famiglietti,
48 2007: Improving land-surface model hydrology: Is an explicit aquifer model better than a deeper soil
49 profile? *Geophys. Res. Lett.*, *34*, L09402, doi:10.1029/2007GL029804

- 1 Gulden, L. E., Z.-L. Yang, and G.-Y. Niu, (2008a), Sensitivity of biogenic emissions simulated by a
 2 land-surface model to land-cover representations, *Atmos. Env.*, doi:10.1016/j.atmosenv.2008.01.045
- 3 Gulden, L. E., E. Rosero, Z.-L. Yang, T. Wagener, and G. Niu (2008b), Model performance, model
 4 robustness, and model fitness scores: A new method for identifying good land-surface models,
 5 *Geophys. Res. Lett.*, 35, L11404, doi:10.1029/2008GL033721
- 6 Gupta, H.V., S. Sorooshian, and P.O. Yapo, Toward improved calibration of hydrologic models: Multiple
 7 and non-commensurable measures of information, *Water Resour. Res.*, 34(4), 751-763, 1998
- 8 Gupta, H.V., L.A. Bastidas, S. Sorooshian, W.J. Shuttleworth and Z.L. Yang, 1999: Parameter estimation of
 9 a land surface scheme using multicriteria methods. *J. Geophys. Res.*, 104(D16), 19491-19504.
- 10 Gupta, H.V., T. Wagener, Y. Liu (2008), Reconciling theory with observations: Elements of a diagnostic
 11 approach to model evaluation, *Hydrological Processes*, 22, doi:10.1002/hyp.6989.
- 12 Gao, X., S. Sorooshian, and H. V. Gupta (1996), Sensitivity analysis of the biosphere-atmosphere transfer
 13 scheme, *J. Geophys. Res.*, 101(D3), 7279–7289.
- 14 Helton, J., and F. Davis, 2003, Latin hypercube sampling and the propagation of uncertainty in analyses
 15 of complex systems, *Reliab. Eng. Syst. Safety*, 81(1), 23– 69.
- 16 Hair, J. F., Anderson, R. E., Tatham, R. L., Tatham, R. L. & Black, W. C. (1995). *Multivariate Data
 17 Analysis with Readings*. Prentice-Hall: Upper Saddle River, NJ.
- 18 Hornberger, G. M., and R. C. Spear (1981), An approach to the preliminary analysis of environmental
 19 systems, *J. Environ. Manage.*, 12, 7–18.
- 20 Hogue T.S., L. A. Bastidas, H.V. Gupta, S. Sorooshian, K. Mitchell and W. Emmerich, 2005: Evaluation
 21 and Transferability of the Noah Land-surface Model in Semi-arid Environments, *J. Hydrometeor.*, 6(1),
 22 68–84.
- 23 Hogue T.S., L. A. Bastidas, H. V. Gupta, and S. Sorooshian, 2006: Evaluating model performance and
 24 parameter behavior for varying levels of land surface model complexity, *Water Resour. Res.*, 42,
 25 W08430, doi:10.1029/2005WR004440
- 26 Jakeman, A.J., R.A. Letcher and J. P. Norton, 2006: Ten iterative steps in development and evaluation of
 27 environmental models. *Env. Modelling & Software*, 21, 602-614 ,doi:10.1016/j.envsoft.2006.01.004
- 28 Joachims T. (1999), Making large-Scale SVM Learning Practical. In *Advances in Kernel Methods -
 29 Support Vector Learning*, B. Schölkopf and C. Burges and A. Smola (ed.), MIT-Press.
- 30 Kato, H., M. Rodell, F. Beyrich, H. Cleugh, E. van Gorsel, H. Liu, and T.P. Meyers, (2007): Sensitivity
 31 of Land Surface Simulations to Model Physics, Land Characteristics, and Forcings, at Four CEOP
 32 Sites, *J. Meteor. Soc. Japan*, 85A, 187-204.
- 33 Khalil, A., M. N. Almasri, M. McKee, and J. J. Kaluarachchi (2005), Applicability of statistical learning
 34 algorithms in groundwater quality modeling, *Water Resour. Res.*, 41, W05010,
 35 doi:10.1029/2004WR003608.
- 36 Khalil, A., McKee, M., Kemblowski, M. W., Asefa, T., and Bastidas, L.A. (2006). Multiobjective
 37 analysis of chaotic dynamics systems with sparse learning machines. *Advances in Water Resources*,
 38 29, 72–88. doi:10.1016/j.advwatres.2005.05.011.
- 39 Kaheil, Y. H., M. K. Gill, M. McKee, L. A. Bastidas, and E. Rosero (2008a), “Downscaling and
 40 assimilation of surface soil moisture using ground truth measurements,” *IEEE Trans. Geosci. Remote
 41 Sens.*, vol. 46, no. 5, pp. 1375–1384, May 2008. DOI: 10.1109/TGRS.2008.916086
- 42 Kaheil, Y. H., Rosero, E., Gill, M. McKee, and L. A. Bastidas (2008b), Downscaling and Forecasting of
 43 Evapotranspiration Using a Synthetic Model of Wavelets and Support Vector Machines. *IEEE Trans.
 44 Geosci. Remote Sens.*, vol.46, no.9, pp. 2692-2707, Sep. 2008. DOI: 10.1109/TGRS.2008.919819
- 45 Koster et al., 2004: Regions of Strong Coupling Between Soil Moisture and Precipitation, *Science*, 305
 46 (5687), 1138.
- 47 LeMone, M. A., et al. (2007), NCAR/CU surface, soil, and vegetation observations during the
 48 International H2O Project 2002 field campaign, *Bull. Amer. Meteor. Soc.*, 88, 65-81.
- 49 Leplastrier, M., A. J. Pitman, H. Gupta, and Y. Xia, 2002: Exploring the relationship between complexity
 50 and performance in a land surface model using the multicriteria method, *J. Geophys. Res.*, 107(D20),
 51 4443, doi:10.1029/2001JD000931.

- 1 Liang, X., and J. Guo, 2003: Intercomparison of land surface parameterization schemes: Sensitivity of
2 surface energy and water fluxes to model parameters, *J. Hydrol.*, 279, 182-209
- 3 Martinez, W and Martinez A. (2002), Computational Statistics Handbook with MATLAB. Chapman and
4 Hall/CRD. ISBN:1584885661 2nd ed. - Boca Raton, Fla. : Chapman & Hall/CRC, 767 p
- 5 McKay, M., R. Beckman, and W. Conover (1979), A comparison of three methods for selecting values of
6 input variables in the analysis of output from a computer code, *Technometrics*, 21(2), 239–245
7 doi:10.2307/1268522
- 8 Mitchell, K. E., et al. (2004), The multi-institution North American Land Data Assimilation System
9 (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological
10 modeling system, *J. Geophys. Res.*, 109, D07S90, doi:10.1029/2003JD003823.
- 11 Niu, G.-Y., Z.-L. Yang, R.E. Dickinson, and L.E. Gulden, 2005: A simple TOPMODEL-based runoff
12 parameterization (SIMTOP) for use in GCMs, *J. Geophys. Res.*, 110, D21106,
13 doi:10.1029/2005JD006111.
- 14 Niu, G.-Y., Z.-L. Yang, R.E. Dickinson, L.E. Gulden, and H. Su, 2007: Development of a simple
15 groundwater model for use in climate models and evaluation with GRACE data, *J. Geophys. Res.*, 112,
16 D07103, doi:10.1029/2006JD007522
- 17 Niu, G.-Y., Yang, Z.-L., et al. (2009) The Community Noah Land-Surface Model with Multi-Physics
18 Options, *J. Geophys. Res.*, (submitted)
- 19 Oleson, K. W., G. B. Bonan, J. Feddema, M. Vertenstein, (2008) An urban parameterization for a global
20 climate model. Part II: Sensitivity to input parameters and the simulated urban heat island in offline
21 simulations. *J. Appl. Meteor. Climat.*, 47, 1061-1076, doi: 10.1175/2007JAMC1598.1.2
- 22 Pitman, A., 1994: Assessing the Sensitivity of a Land-Surface Scheme to the Parameter Values Using a
23 Single Column Model. *J. Climate*, 7, 1856–1869.
- 24 Pitman, A. J., (2003) Review: the evolution of, and revolution in, land surface schemes designed for
25 climate models, *Int. J. Climatol.* 23, 479–510, doi:10.1002/joc.893.
- 26 Prihodko L., A.S. Denning, N.P. Hanan, I. Baker, K. Davis, 2008: Sensitivity, uncertainty and time
27 dependence of parameters in a complex land surface model, *Agric. Forest Meteorol.*, 148(2), 268-
28 287, doi: 10.1016/j.agrformet.2007.08.006.
- 29 Randall, D.A., R.A. Wood, S. Bony, R. Colman, T. Fiechfet, J. Fyfe, V. Kattsov, A. Pitman, J. Shukla, J.
30 Srinivasan, R.J. Stouffer, A. Sumi and K.E. Taylor, 2007: Climate Models and Their Evaluation. In:
31 *Climate Change 2007: The Physical Science Basis*. Contribution of Working Group I to the Fourth
32 Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M.
33 Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.)]. Cambridge University
34 Press, Cambridge, United Kingdom and New York, NY, USA.
- 35 Ratto, M., P. C. Young, R. Romanowicz, F. Pappenberger, A. Saltelli, and A. Pagano, 2007: Uncertainty,
36 sensitivity analysis and the role of data based mechanistic modeling in hydrology. *Hydrol. Earth Syst.*
37 *Sci.*, 11, 1249–1266.
- 38 Rodell, M., P. R. Houser, A. A. Berg, and J. S. Famiglietti, 2005: Evaluation of 10 methods for initializing a
39 land surface model, *J. Hydrometeorol.*, 6, 146–155.
- 40 Rosero, E., and L. A. Bastidas (2007). Evaluation of LSM Parameter Transferability Across Semi-Arid
41 Environments, in Proceedings of the 21st Conference on Hydrology, AMS Meeting, San Antonio,
42 Texas, January 14-18. <http://ams.confex.com/ams/pdfpapers/117116.pdf>
- 43 Rosero E, Yang Z.-L., Gulden L.E., Niu G.-Y., Gochis D.J. (2009) Evaluating Enhanced Hydrological
44 Representations in Noah-LSM over Transition Zones: Implications for Model Development, *J.*
45 *Hydromet.*, doi: 10.1175/2009JHM1029.1. In Press
- 46 Spear R.C. and G.M. Hornberger, (1980). Eutrophication in Peel Inlet-II: Identification of critical
47 uncertainties via generalized sensitivity analysis, *Water Resour. Res.*, 14, pp. 43–99.
- 48 Spear, R. C., T. M. Grieb, and N. Shang, 1994. Parameter uncertainty and interaction in complex
49 environmental models, *Water Resour. Res.*, 30(11), 3159-3169.
- 50 Smola, A. J., and Schölkopf, B. (2004). A tutorial on support vector regression, *Statistics and Computing*,
51 14, 199–222.

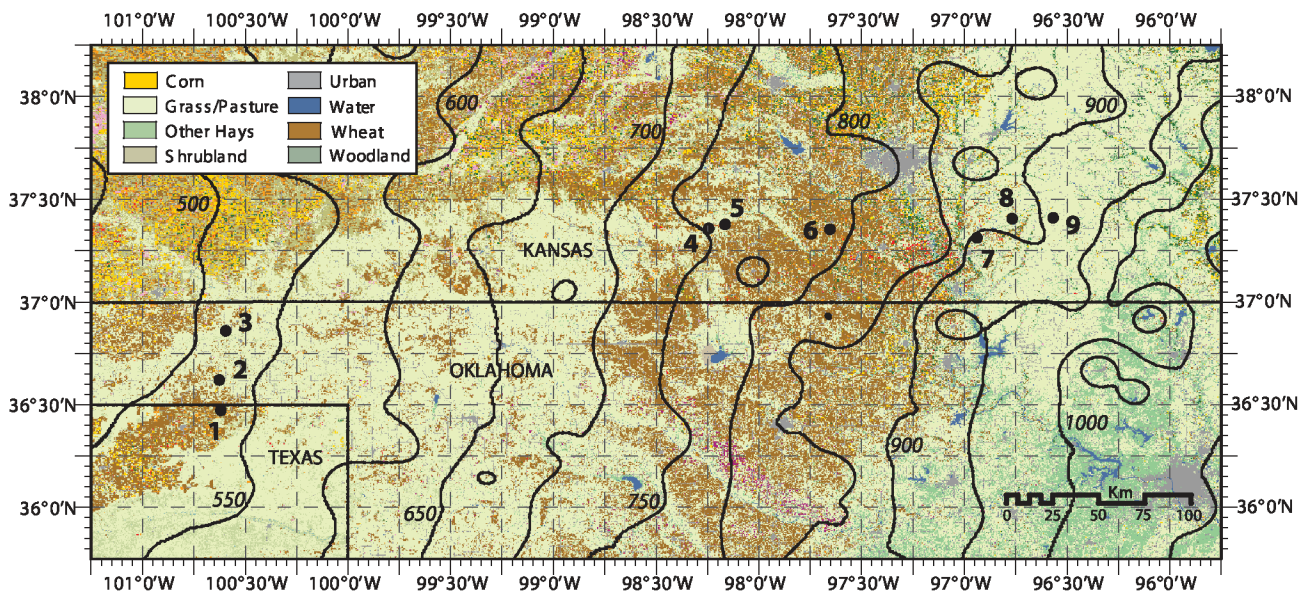
- 1 Stöckli, R., D. M. Lawrence, G.-Y. Niu, K. W. Oleson, P. E. Thornton, Z.-L. Yang, G. B. Bonan, A. S.
 2 Denning, and S. W. Running (2008), Use of FLUXNET in the Community Land Model development,
 3 *J. Geophys. Res.*, **113**, G01025, doi:10.1029/2007JG000562.
- 4 Sobol', I. M. (1993), Sensitivity analysis for non-linear mathematical models, *Math. Modelling Comput.*
 5 *Exp.*, **1**, 407–414.
- 6 Sobol', I. M. (2001), Global sensitivity indices for nonlinear mathematical models and their Monte Carlo
 7 estimates, *Math. Comput. Simul.*, **55**, 271–280, doi:10.1016/S0378-4754(00)00270-6
- 8 Saltelli, A. (1999), Sensitivity analysis: Could better methods be used?, *J. Geophys. Res.*, **104**(D3), 3789–
 9 3793
- 10 Saltelli, A., 2002, Making best use of model evaluations to compute sensitivity indices. *Computer Physics*
 11 *Communication*, **145**, 580–297
- 12 Saltelli, A., M. Ratto, S. Tarantola, F. Campolongo, 2006, Sensitivity analysis practices: Strategies for
 13 model-based inference, *Reliability Engineering & System Safety*, **91**(10-11), 1109-1125
- 14 Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola, (2008)
 15 Global sensitivity analysis: The primer. Wiley-Interscience. ISBN-13: 978-0470059975
- 16 Tang, Y., P. Reed, T. Wagener, and K. van Werkhoven (2006), Comparing sensitivity analysis methods
 17 to advance lumped watershed model identification and evaluation, *Hydrol. Earth Syst. Sci. Discuss.*,
 18 **3**, 3333–3395.
- 19 Tang, Y., P. Reed, K. van Werkhoven, and T. Wagener (2007), Advancing the identification and
 20 evaluation of distributed rainfall-runoff models using global sensitivity analysis, *Water Resour. Res.*,
 21 **43**, W06415, doi:10.1029/2006WR005813
- 22 Trier, S.B., F. Chen, K.W. Manning, M.A. LeMone, and C.A. Davis, 2008: Sensitivity of the PBL and
 23 Precipitation in 12-Day Simulations of Warm-Season Convection Using Different Land Surface
 24 Models and Soil Wetness Conditions. *Mon. Wea. Rev.*, **136**, 2321–2343.
- 25 Unland, H., P. Houser, W. J. Shuttleworth, and Z.-L. Yang, 1996, Surface flux measurement and
 26 modeling at a semi-arid Sonoran Desert site, *Agric. For. Meteorol.*, **82**, 119–153.
- 27 van Werkhoven, K., T. Wagener, P. Reed, and Y. Tang (2008), Characterization of watershed model
 28 behavior across a hydroclimatic gradient, *Water Resour. Res.*, **44**, W01429,
 29 doi:10.1029/2007WR006271
- 30 van Werkhoven, K., T. Wagener, P. Reed, and Y. Tang (2009), Sensitivity-guided reduction of parametric
 31 dimensionality for multiobjective calibration of watershed models, *Adv. Water Resour.*, in press
- 32 Vapnik, V. N. (1998), *Statistical Learning Theory*, John Wiley, Hoboken, N. J.
- 33 Vrugt, J. A., H. V. Gupta, L. A. Bastidas, W. Bouten, and S. Sorooshian, 2003: Effective and efficient
 34 algorithm for multiobjective optimization of hydrologic models, *Water Resour. Res.*, **39**(8), 1214,
 35 doi:10.1029/2002WR001746.
- 36 Wagener, T., Boyle, D. P., Lees, M. J., Wheeler, H. S., Gupta, H. V. and Sorooshian, S. 2001. A
 37 framework for development and application of hydrological models, *Hydrol. Earth Syst. Sci.*, **5**, 13-
 38 26.
- 39 Wagener, T., and H. V. Gupta, 2005: Model identification for hydrological forecasting under uncertainty,
 40 *Stoch. Environ. Res. Risk Assess.*, **19**(6), 378–387, doi:10.1007/s00477-005-0006-5.
- 41 Wagener, T., and J. Kollat, 2007: Numerical and visual evaluation of hydrological and environmental
 42 models using the Monte Carlo Analysis Toolbox, *Environ. Modell. Softw.*, **2**, 1021–1033
- 43 Weckwerth, et al., 2004: An overview of the International H2O Project (IHOP_2002) and some preliminary
 44 highlights. *Bull. Amer. Meteor. Soc.*, **85**, 253–277.
- 45 Yatheendradas, S., T. Wagener, H. Gupta, C. Unkrich, D. Goodrich, M. Schaffner, and A. Stewart (2008),
 46 Understanding uncertainty in distributed flash flood forecasting for semiarid regions, *Water Resour.*
 47 *Res.*, **44**, W05S19, doi:10.1029/2007WR005940
- 48 Yang, Z.-L., and G.-Y. Niu, 2003: Versatile integrator of surface and atmosphere processes (VISA) Part 1:
 49 Model description. *Glob. Planet. Change* **38**, 175–189
- 50

1 **FIGURES**

2

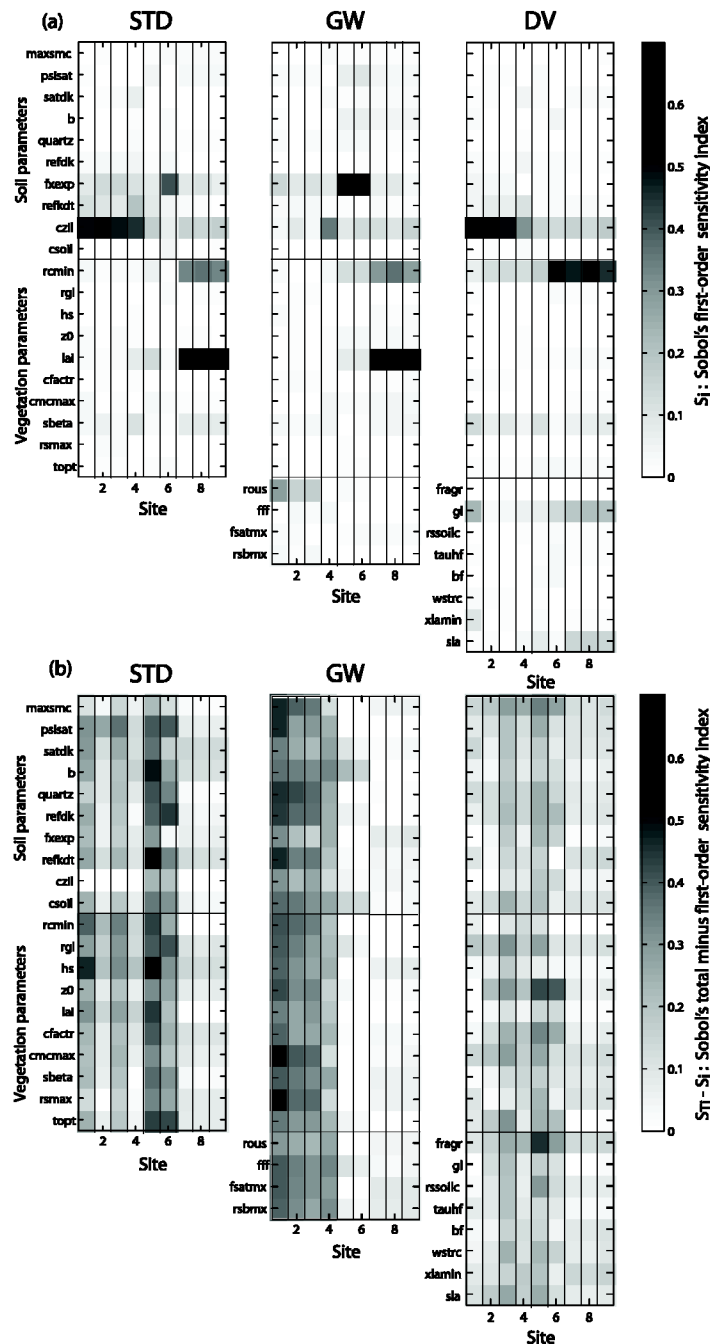
3 **Figure 1.** IHOP_2002 near-surface state and flux stations. The contours show the strong east - west mean
 4 annual precipitation (MAP) gradient. The nine sites were located in representative land covers (see Table
 5 1): six on grassland of varying thickness, two on winter wheat, one on bare ground,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

10

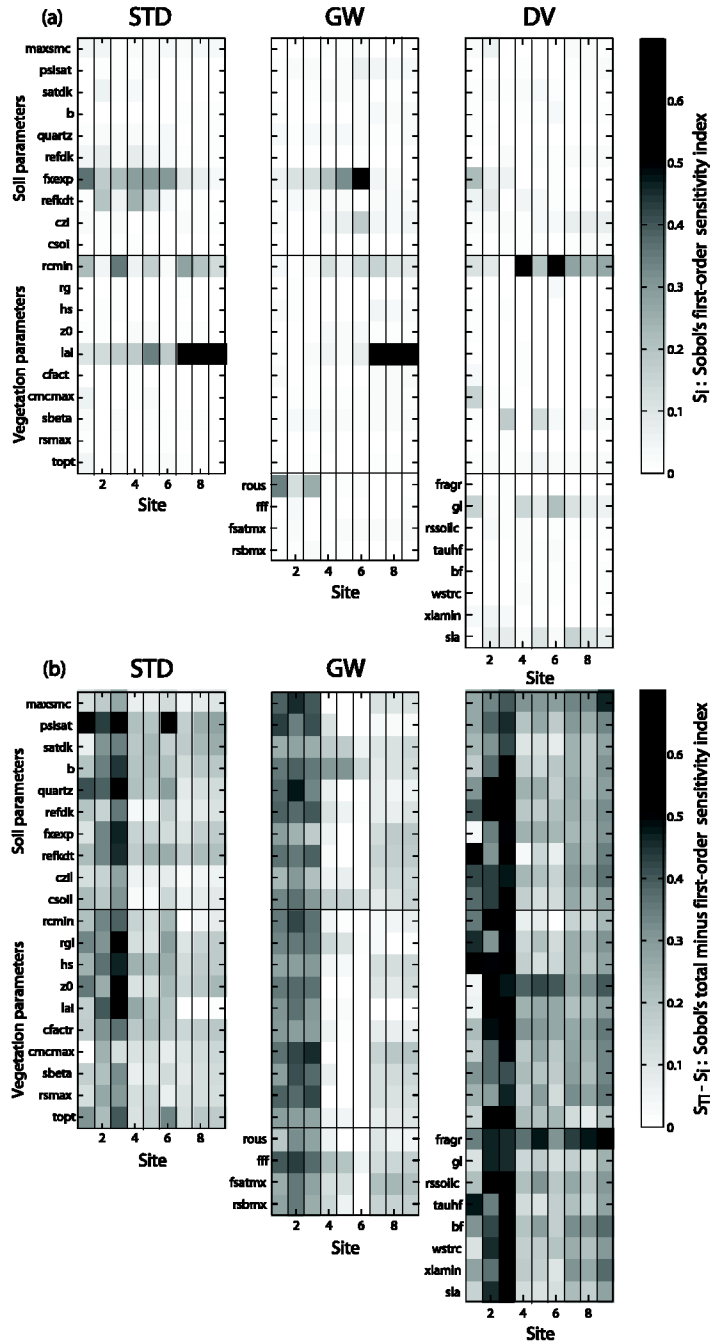


11
12

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



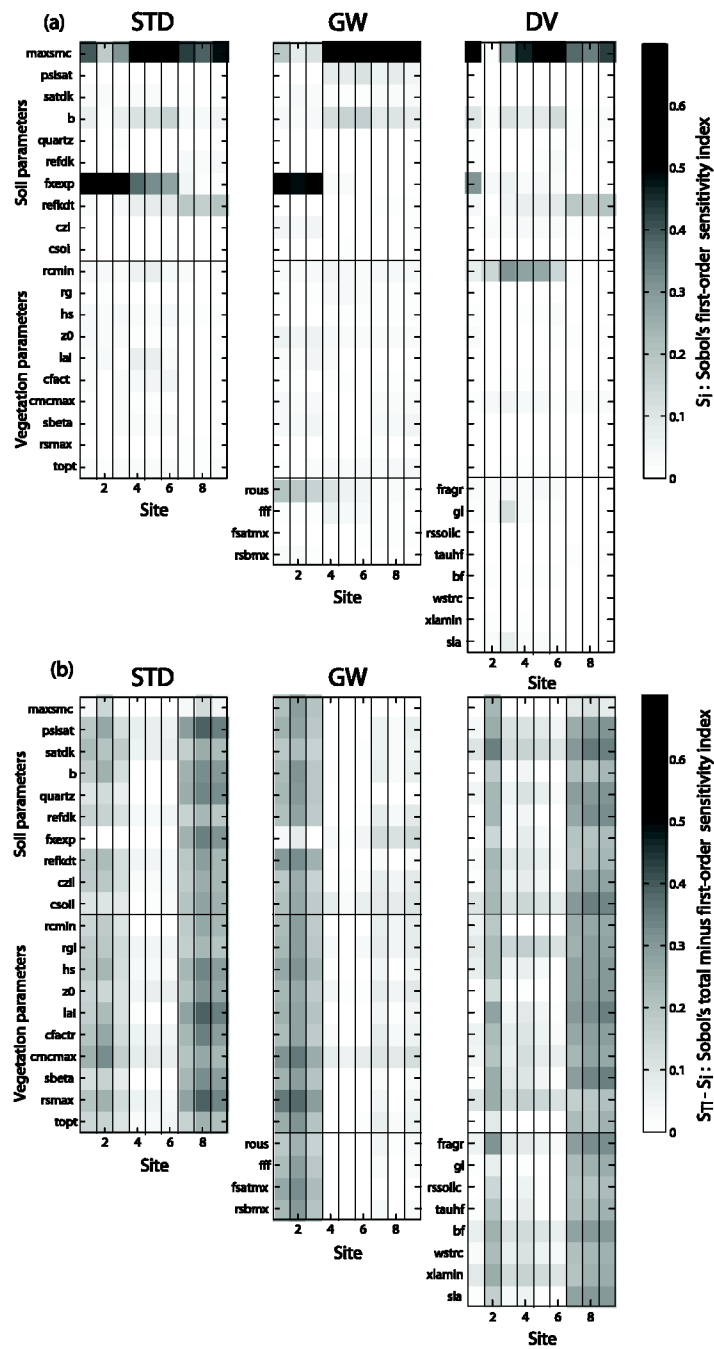
1 **Figure 3.** Same as Figure 2 but for LE.
 2



3
 4

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

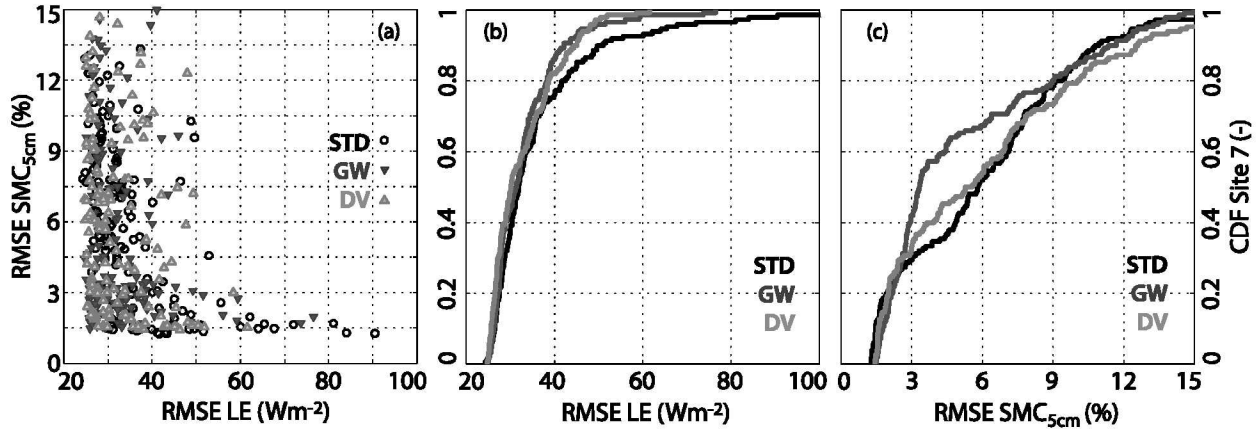
2



3

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

6

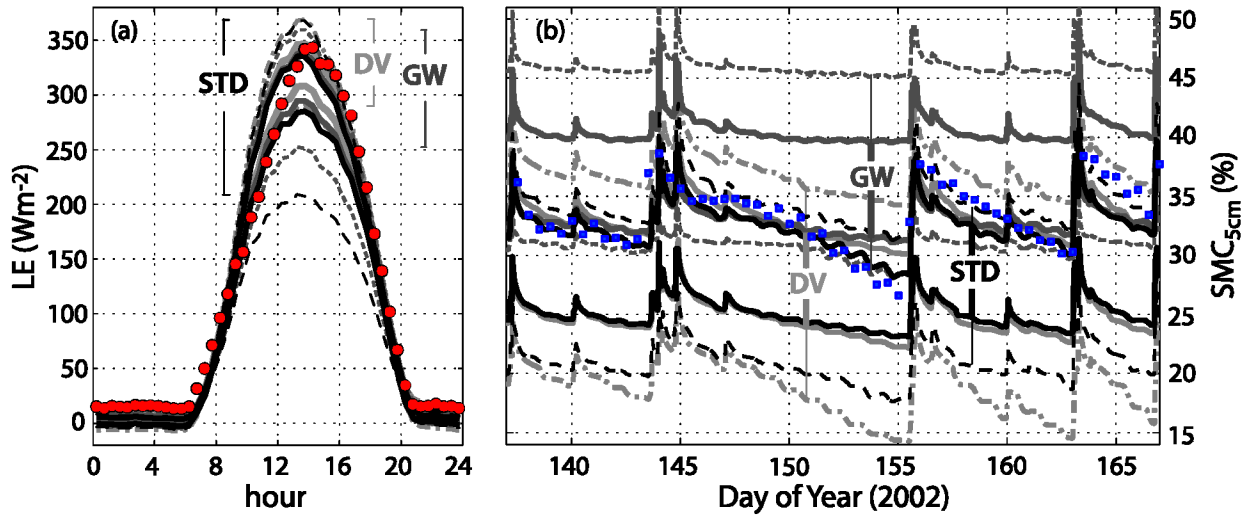


7

8

1 **Figure 6.** Uncertainty ranges of the 150 ensemble members of behavioral runs of STD (black), GW (dark
 2 gray) and DV (light gray) at Site 7. (a) Average diurnal LE. (b) Hourly SMC_{5cm} . Plots show the
 3 interquartile range IQR (50% of the runs) in continuous lines and the 90% confidence bounds (5% to 95%
 4 quantile) in dashed lines. Observations are shown with symbols. The spread of LE by STD is larger than
 5 that of DV, GW. However, IQR shows a very similar LE envelope. The ensemble of GW shows wetter
 6 SMC_{5cm} than the rest. The IQR of STD and DV are very similar, but the 90% confidence of STD has
 7 lower spread than DV.

8

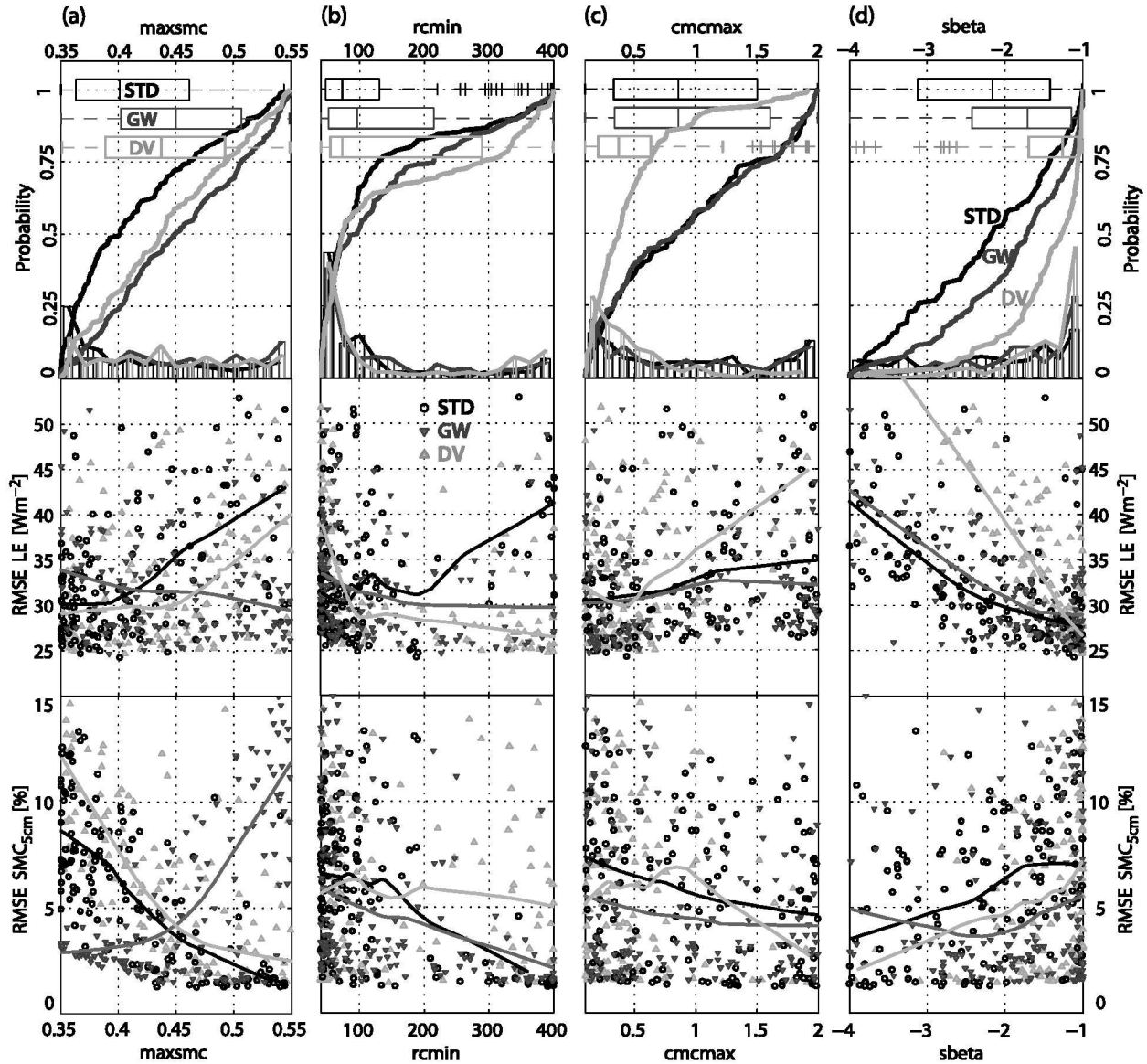


9

10

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

7



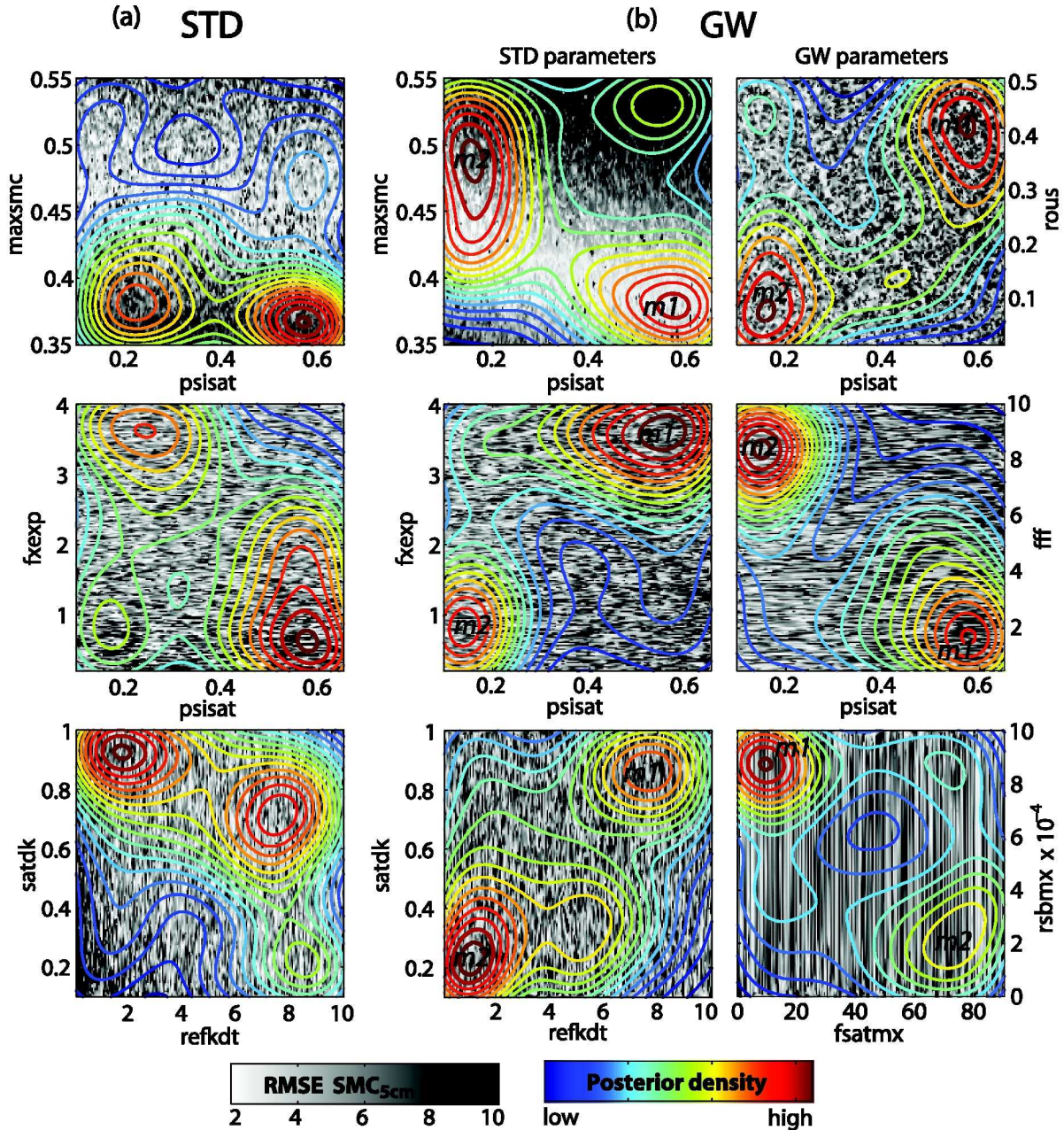
8
9

10

11

1 **Figure 8.** Multivariate posterior distribution of the behavioral parameters of STD and GW at site 7 shown
 2 for selected parameter combinations in bivariate plots. Higher density of parameter values are indicated
 3 with increasingly redder contours. The response surface of SMC_{5cm} is shown in the back; darker regions
 4 have higher errors. The bi-modal behavior of GW is signaled by $m1$ and $m2$. See text for explanation.

5



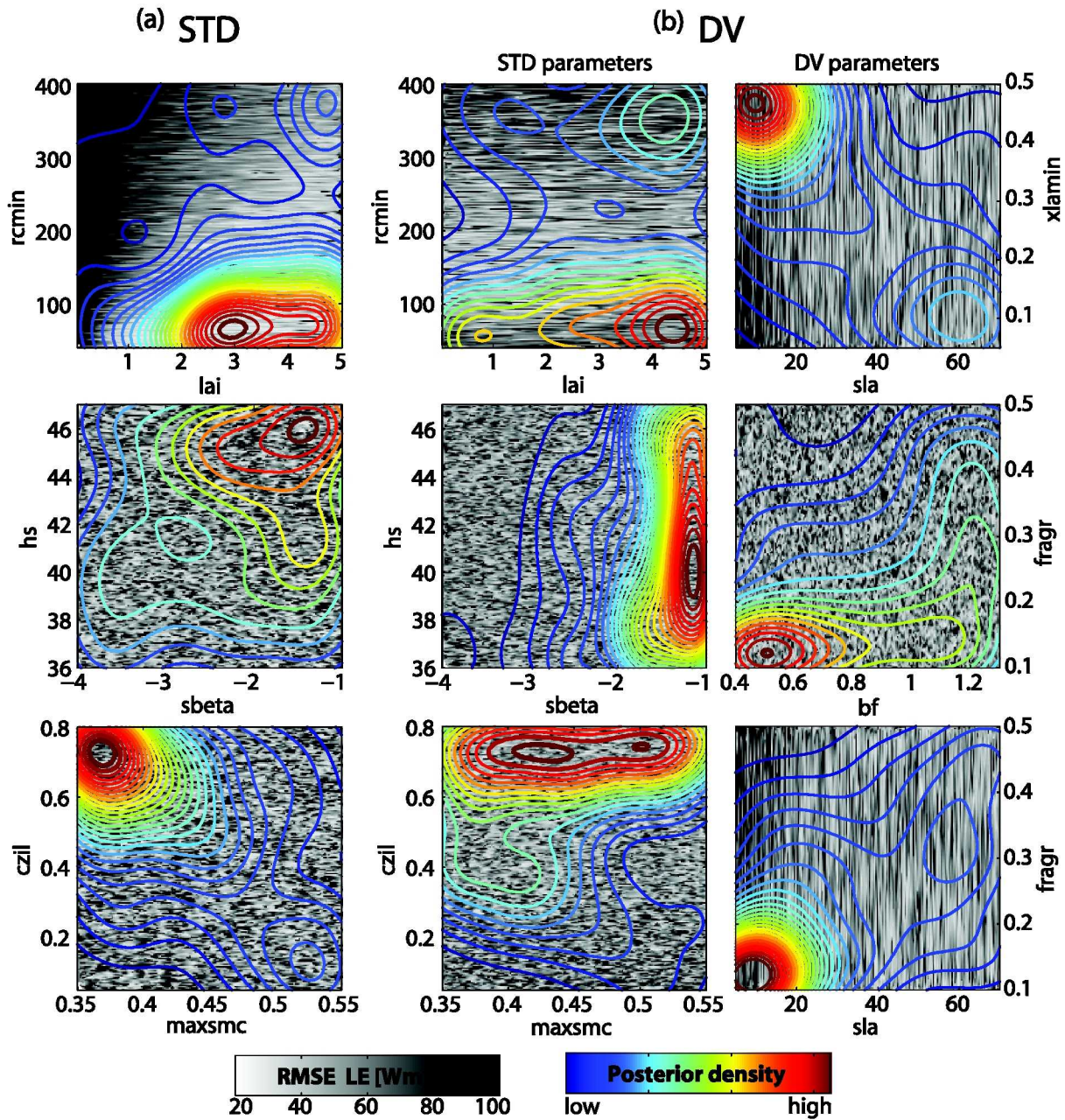
6

7

8

1 **Figure 9.** Bivariate depiction of the posterior distribution of behavioral parameters of STD and DV at Site
 2 7. Higher density of parameter values are indicated with red contours. The response surface of LE is
 3 shown in the back; darker regions have higher errors. Note the significant change in the identifiability of
 4 hs and maxsmc.

5

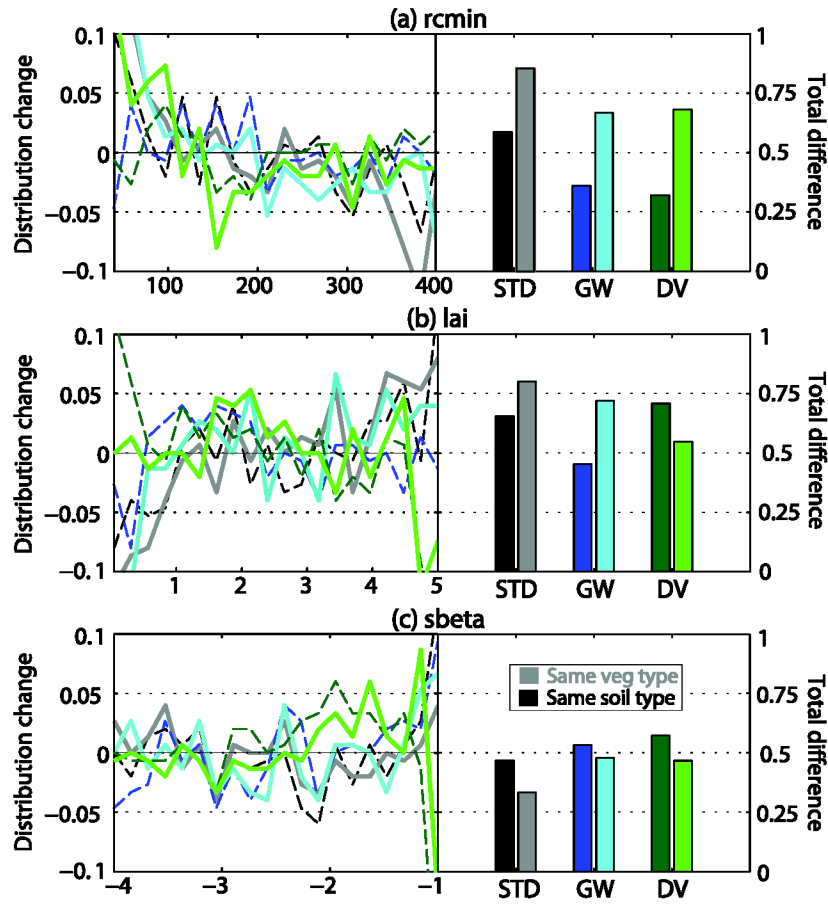


6

7

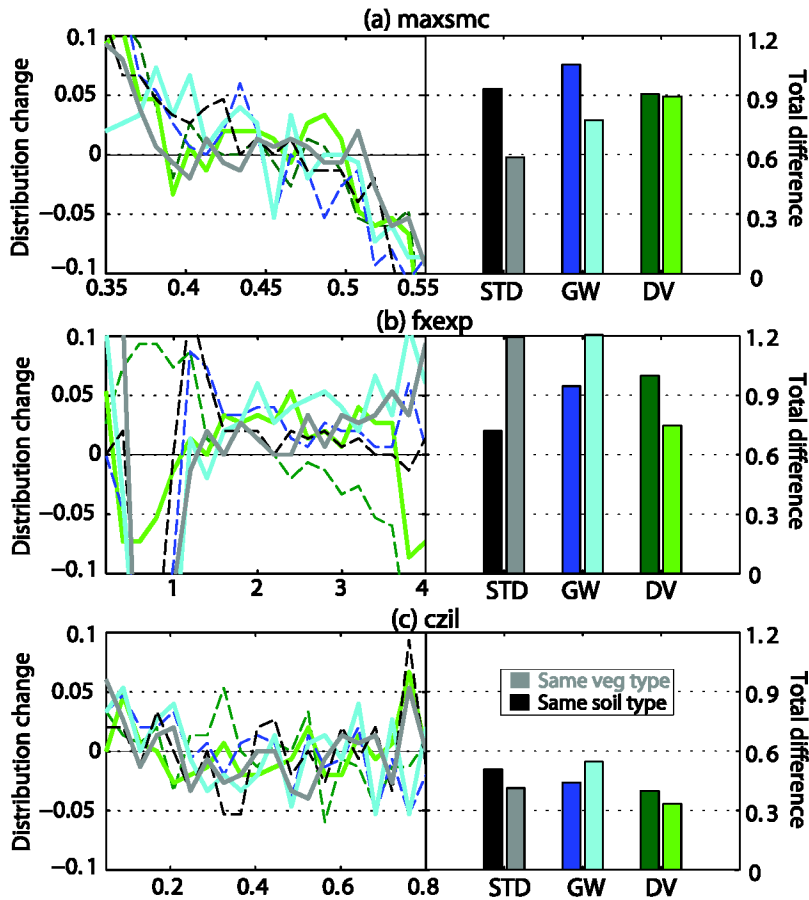
8

1 **Figure 10.** Comparison of the marginal posterior parameter distributions of selected, sensitive vegetation
 2 parameters: (a) rcmin, (b) lai, and (c) sbeta. The total difference between parameter distributions at sites
 3 with the same vegetation cover type (Site 2 and 8) –grassland– (continuous, bright lines) is not smaller
 4 than the difference of distributions of the same parameters between contiguous sites (Site 2 and 1)
 5 (dashed, dark lines), which share the same soil type (sandy clay loam).
 6



7
8
9

1 **Figure 11.** Comparison of the marginal posterior parameter distributions of selected, sensitive soil
 2 parameters: (a) maxsmc, (b) fxexp, and (c) czil. The total difference between parameter distributions at
 3 contiguous sites with the same soil type (Site 1 and 2) –sandy clay loam– (dashed, dark lines) is only
 4 sometimes smaller than the difference of distributions of the same parameters between sites with different
 5 meteorology (Site 2 and 8) (continuous, bright lines), which share the same vegetation cover type
 6 (grassland).
 7



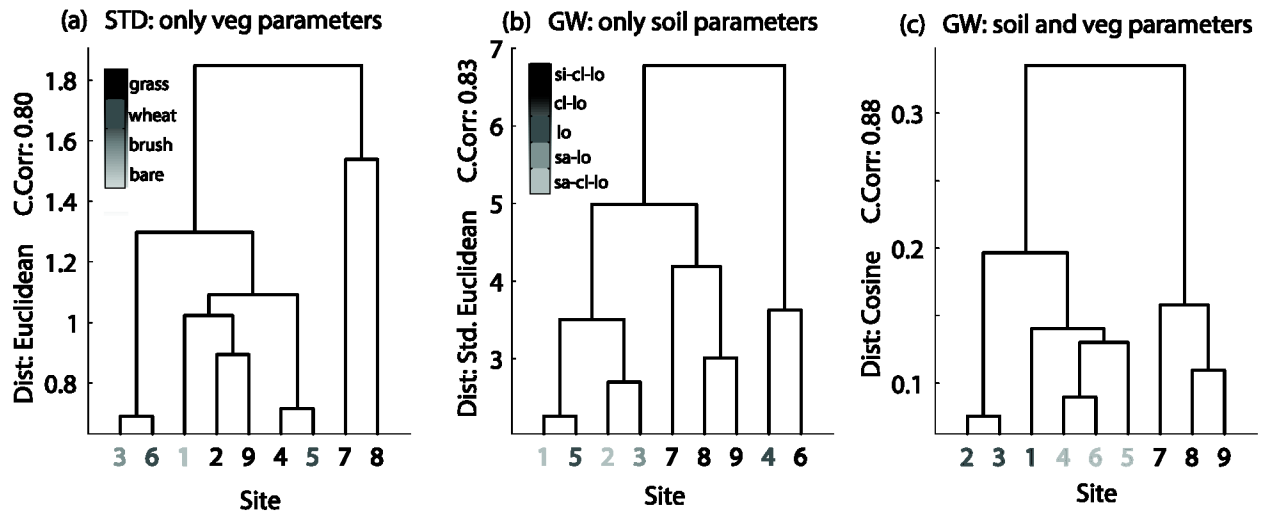
8

9

10

1 **Figure 12.** Clustering of sites using: (a) only the vegetation parameters of STD, (b) only the soil
 2 parameters of GW, and (c) both soil and vegetation parameters of GW. The similarity between marginal
 3 distributions of behavioral parameters at all sites is compared using different distances. The plots report
 4 the distance that maximizes the cophenetic correlation coefficient of the linkage. Note that neither soil nor
 5 vegetation parameters render groups solely based on soil or vegetation type. The clusters of all parameters
 6 seem to have a strong relationship with the 3 climatic zones: (1-3) semi-arid, (4-6) middling, and (7-9)
 7 semi-humid.

8

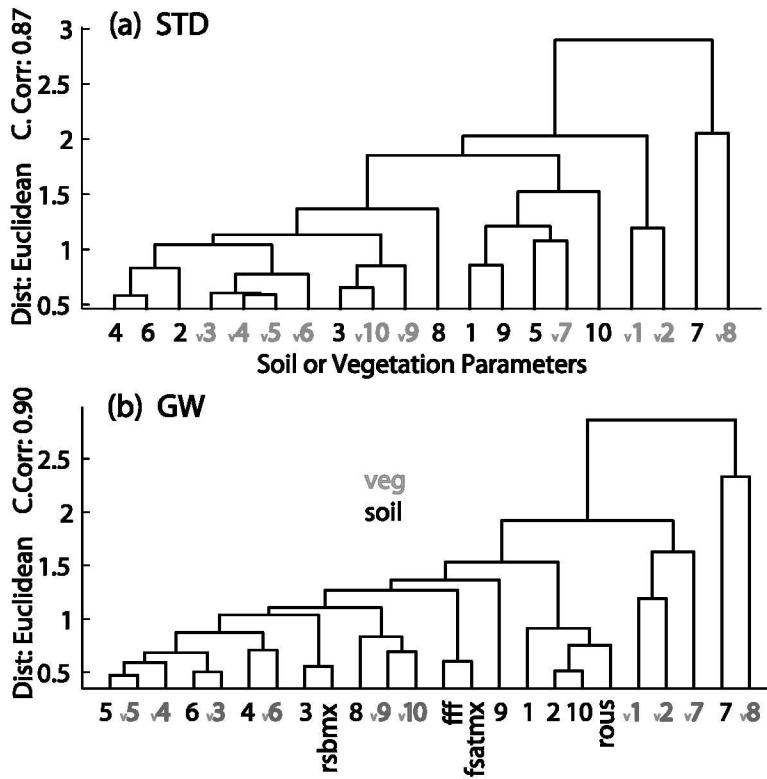


9

10

1 **Figure 13.** Clustering of soil (black), vegetation (gray) and GW-only parameters for the behavioral,
 2 marginal posterior distributions of (a) STD and (b) GW at all sites. The cophenetic correlation coefficient
 3 for the complete linkage for the parameters of STD and GW is 0.87 and 0.90, respectively. GW
 4 parameters seem to behave in a similar way as the soil parameters do.

5



6

7

8

1 **TABLES**

2 **Table 1.** Average meteorology, near-surface states and turbulent fluxes observed during the calibration
3 period (13 May - 25 Jun) at the nine IHOP_2002 sites. Indices of vegetation and soil classes are in
4 parenthesis. Rainfall is cumulative over the observation period. Dry, sparsely vegetated sites (1-3) receive
5 almost half of the amount of mean annual precipitation (MAP) than wet sites (7-9), with lush vegetation.
6 Mean 2-m air temperature (T_a), sensible (H), latent (LE) and ground (G) heat flux, ground temperature
7 (T_g) and soil moisture content at 5-cm (SMC_{5cm}).

8

Site	1	2	3	4	5	6	7	8	9
Lat ($^{\circ}$ N)	36.4728	36.6221	36.8610	37.3579	37.3781	37.3545	37.3132	37.4070	37.4103
Lon ($^{\circ}$ 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 type	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
T_a ($^{\circ}$ C)	21.4	21.7	22.5	20.7	20.7	21.0	20.7	20.1	19.9
H (Wm^{-2})	70.5	70.7	75.7	43.9	51.9	61.4	25.9	17.1	27.9
LE (Wm^{-2})	65.1	76.1	68.2	106.2	111.2	97.1	126.4	122.8	115.3
G (Wm^{-2})	-10.4	-6.4	-9.3	-2.7	-5.1	-7.5	-5.6	-12.1	-10.5
T_g ($^{\circ}$ 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

9

10

1 **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^3m^{-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	$Jm^{-3}K^{-1}$	1.26	3.5
Vegetation parameters				
rcmin	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
cmcmx	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	$s^{-1}\ x1E-6$	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^3m^{-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	$ms^{-1}\ 1E-3$	0.01	1

2

1 **Table 3.** Sensitive parameters according to a Kolmogorov-Smirnov test between samples that drive
 2 behavioral (multi-objectively calibrated) and non-behavioral simulations. 1 stands for sensitive, 0 for
 3 insensitive. The number of sensitive parameter is tabulated by class (soil, vegetation, and new GW or
 4 DV). See Table 1 for parameter names. No clear regional pattern of sensitivity can be readily discerned.

Site	1			2			3			4			5			6			7			8			9				
	No. Par.	STD	GW	DV	STD	GW	DV	STD	GW	DV	STD	GW	DV	STD	GW	DV	STD	GW	DV	STD	GW	DV	STD	GW	DV	STD	GW	DV	
soil parameters	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	2	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	0	1	1	0	1	1	1	0	1	0	0	0	0
	3	1	0	1	0	0	1	0	1	1	1	0	1	1	1	1	0	1	1	1	0	0	1	1	0	0	1	0	0
	4	1	0	0	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	0	0	1	0	1	1	1	0	0
	5	1	1	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1
	6	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	0	0	0	1	0	1	1
	7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	1	1	0	1	0	1
	8	1	1	1	1	0	0	1	0	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	0	0	1	0	1
	9	1	0	1	1	1	1	1	0	1	1	1	1	1	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1
	10	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0
#	10	7	8	7	8	8	8	7	10	8	9	10	7	8	9	7	9	10	8	5	8	8	7	6	7	4	7	7	
vegetation parameters	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	2	0	0	0	0	0	1	0	1	0	1	1	0	0	0	0	0	1	0	1	1	1	1	0	1	1	1	1	
	3	0	0	1	1	0	1	1	0	1	1	1	1	0	0	1	0	1	1	0	0	0	0	0	1	0	0	0	
	4	1	0	0	0	0	1	1	0	1	1	0	1	1	1	1	0	0	1	1	0	0	1	0	1	1	0	1	
	5	1	1	1	1	1	1	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	
	6	0	0	0	0	0	1	1	0	1	1	1	1	0	1	1	0	1	1	1	0	0	0	0	1	0	0	1	
	7	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	0	1	1	
	8	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	9	0	1	0	1	0	0	0	0	1	1	1	0	0	1	1	0	1	1	1	0	1	0	1	0	1	0	1	1
	10	1	0	1	0	0	1	0	0	1	0	0	1	1	1	1	0	1	1	1	1	1	0	1	0	1	1	1	1
#	6	3	6	5	3	9	5	3	9	8	8	8	6	8	9	4	9	9	8	6	7	5	6	6	8	6	9	9	
new GW or DV parameters	1		1	1		1	0		1	1		1	1		1	1		1	1		0	1		1	1		0	1	
	2		1	1		0	1		1	1		1	1		1	1		1	1		0	1		1	1		0	1	
	3		1	0		0	1		0	0		1	1		1	1		1	1		1	0		0	0		0	0	
	4		1	1		0	0		0	0		1	0		1	0		1	0		1	0		1	1		0	1	
	5			0			1			1			1			0			0			0			1			1	
	6			1			1			1			1			1			1			0			1			1	
	7			1			1			1			1			1			1			1			0			0	
	8			1			1			1			1			1			1			1			1			1	
#		4	6		1	6		2	6		4	7		4	6		4	6		2	4		3	6		0	6		
Total	16	14	20	12	12	23	13	12	25	16	21	25	13	20	24	11	22	25	16	13	19	13	16	18	15	10	22		

5
6

1 **Table 4.** Spearman rank correlation coefficients between parameter sets belonging to the behavioral set
 2 for STD (up the diagonal) and GW (below the diagonal). Note the change in the covariance structure in
 3 Fig. 8. See Table 1 for abbreviations of parameter names.

	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	
rsbm	0.11	-0.25	-0.13	-0.21	0.32	0.08	-0.24

4
 5
 6
 7
 8
 9
 10
 11

1 **Table 5.** Spearman rank correlation coefficients between parameter sets belonging to the behavioral set
 2 for STD (up the diagonal) and DV (below the diagonal). Note the change in the covariance structure in
 3 Fig 9. See Table 1 for abbreviations of parameter names.

DV	STD						
	rmin	hs	maxsmc	psisat	fragr	bf	xlaimin
rmin		-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

4

5

6

7

1 LIST OF FIGURES

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

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

15 FIG 3. Same as Figure 2 but for LE.

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

17 FIG 5. Tradeoff LE- SMC_{5cm} and cumulative distribution functions (CDF) of scores of behavioral STD,
 18 GW and DV at Site 7. (a) Scatterplot in objective function space of parameter sets that maximize the
 19 likelihood function after multi-objective calibration against $\{H, LE, G, T_g, SMC_{5cm}\}$. CDF of root mean
 20 squared errors (RMSE) of behavioral runs evaluated against observed (b) LE, and (c) SMC_{5cm} . GW (dark
 21 grey), DV (light gray) perform as good as or better than STD (black).

22 FIG 6. Uncertainty ranges of the 150 ensemble members of behavioral runs of STD (black), GW (dark
 23 gray) and DV (light gray) at Site 7. (a) Average diurnal LE. (b) Hourly SMC_{5cm} . Plots show the
 24 interquartile range (50% of the runs) in continuous lines and the 90% confidence bounds (5% to 95%
 25 quantile) in dashed lines. Observations are shown with symbols. The spread of LE by STD is larger than
 26 that of DV, GW. The ensemble of GW shows wetter SMC_{5cm} than the rest.

27 FIG 7. Marginal cumulative distribution functions (CDF) of the posterior distribution of selected
 28 behavioral parameter sets at Site 7. (a) Porosity [$maxsmc$], (b) minimum stomatal resistance [$rcmin$], (c)
 29 maximum water holding capacity of the canopy [$cmcmmax$], and (d) effect of the vegetation on ground heat
 30 flux [$sbeta$]. Along with the CDFs, the histograms and interquartile ranges are shown. The trend in the

1 scatterplots of RMSE of LE and SMC_{5cm} is shown by fitting a minimum complexity polynomial. Note
2 that in all subpanels GW (dark grey), DV (light gray) and STD (black) are shown.

3 FIG 8. Multivariate posterior distribution of the behavioral parameters of STD and GW at site 7 shown
4 for selected parameter combinations in bivariate plots. Higher density of parameter values are indicated
5 with increasingly redder contours. The response surface of SMC_{5cm} is shown in the back; darker regions
6 have higher errors. The bi-modal behavior of GW is signaled by $m1$ and $m2$. See text for explanation.

7 FIG 9. Bivariate depiction of the posterior distribution of behavioral parameters of STD and DV at Site 7.
8 Higher density of parameter values are indicated with red contours. The response surface of LE is shown
9 in the back; darker regions have higher errors. Note the significant change in the identifiability of hs and
10 $maxsmc$.

11 FIG 10. Comparison of the marginal posterior parameter distributions of selected, sensitive vegetation
12 parameters: (a) $rcmin$, (b) lai , and (c) $sbeta$. The total difference between parameter distributions at sites
13 with the same vegetation cover type (Site 2 and 8) (continuous, bright lines) is not smaller than the
14 difference of distributions of the same parameters between contiguous sites (Site 2 and 1) (dashed, dark
15 lines), which share the same soil type.

16 FIG 11. Comparison of the marginal posterior parameter distributions of selected, sensitive soil
17 parameters: (a) $maxsmc$, (b) $fxexp$, and (c) $czil$. The total difference between parameter distributions at
18 sites with the same soil type (Site 5 and 7) (continuous, bright lines) is in general not smaller than the
19 difference of distributions of the same parameters between contiguous sites (Site 5 and 6) (dashed, dark
20 lines), which share the same vegetation cover type.

21 FIG 12. Clustering of sites using: (a) only the vegetation parameters of STD, (b) only the soil parameters
22 of GW, and (c) both soil and vegetation parameters of GW. The similarity between marginal distributions
23 of behavioral parameters at all sites is compared using different distances. The plots report the distance
24 that maximizes the cophenetic correlation coefficient of the linkage. Note that neither soil nor vegetation
25 parameters render groups solely based on soil or vegetation type. The clusters of all parameters seem to
26 have a strong relationship with the 3 climatic zones.

27 FIG 13. Clustering of soil (black), vegetation (gray) and GW parameters for the behavioral, marginal
28 posterior distributions of (a) STD and (b) GW at all sites. The cophenetic correlation coefficient for the
29 complete linkage for the parameters of STD and GW is 0.87 and 0.90, respectively. GW parameters seem
30 to behave in a similar way as the soil parameters do.

31

1 **LIST OF TABLES**

2 TABLE 1. Average meteorology, near-surface states and turbulent fluxes observed during the calibration
3 period (13 May - 25 Jun) at the nine IHOP_2002 sites. Indices of vegetation and soil classes are in
4 parenthesis. Rainfall is cumulative over the observation period. Dry, sparsely vegetated sites (1-3) receive
5 almost half of the amount of mean annual precipitation (MAP) than wet sites (7-9), with lush vegetation.
6 Mean 2-m air temperature (T_a), sensible (H), latent (LE) and ground (G) heat flux, ground temperature
7 (T_g) and soil moisture content at 5-cm (SMC_{5cm}).

8 TABLE 2. Feasible ranges of Noah-LSM parameters considered in the sensitivity analysis.

9 TABLE 3. Sensitive parameters according to a Kolmogorov-Smirnov test between samples that drive
10 behavioral (multi-objectively calibrated) and non-behavioral simulations. 1 stands for sensitive, 0 for
11 insensitive. The number of sensitive parameter is tabulated by class (soil, vegetation, and new GW or
12 DV). No clear regional pattern of sensitivity can be readily discerned.

13 TABLE 4. Spearman rank correlation coefficients between parameter sets belonging to the behavioral set
14 for STD (up the diagonal) and GW (below the diagonal). Note the change in the covariance structure in
15 Fig. 8.

16 TABLE 5. Spearman rank correlation coefficients between parameter sets belonging to the behavioral set
17 for STD (up the diagonal) and DV (below the diagonal). Note the change in the covariance structure in
18 Fig. 9.