Hydrol. Earth Syst. Sci. Discuss., 11, 13745–13795, 2014
www.hydrol-earth-syst-sci-discuss.net/11/13745/2014/
doi:10.5194/hessd-11-13745-2014

Hydrology and Earth System Sciences
Open Access
Discussions

Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework

M. S. Raleigh[1], **J. D. Lundquist**[2], **and M. P. Clark**[1]

[1]National Center for Atmospheric Research, Boulder, Colorado, USA
[2]Civil and Environmental Engineering, University of Washington, Seattle, Washington, USA

Received: 24 October 2014 – Accepted: 21 November 2014 – Published: 16 December 2014

Correspondence to: M. S. Raleigh (raleigh@ucar.edu)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Physically based models provide insights into key hydrologic processes, but are associated with uncertainties due to deficiencies in forcing data, model parameters, and model structure. Forcing uncertainty is enhanced in snow-affected catchments, where weather stations are scarce and prone to measurement errors, and meteorological variables exhibit high variability. Hence, there is limited understanding of how forcing error characteristics affect simulations of cold region hydrology. Here we employ global sensitivity analysis to explore how different error types (i.e., bias, random errors), different error distributions, and different error magnitudes influence physically based simulations of four snow variables (snow water equivalent, ablation rates, snow disappearance, and sublimation). We use Sobol' global sensitivity analysis, which is typically used for model parameters, but adapted here for testing model sensitivity to co-existing errors in all forcings. We quantify the Utah Energy Balance model's sensitivity to forcing errors with 1 520 000 Monte Carlo simulations across four sites and four different scenarios. Model outputs were generally (1) more sensitive to forcing biases than random errors, (2) less sensitive to forcing error distributions, and (3) sensitive to different forcings depending on the relative magnitude of errors. For typical error magnitudes, precipitation bias was the most important factor for snow water equivalent, ablation rates, and snow disappearance timing, but other forcings had a significant impact depending on forcing error magnitudes. Additionally, the relative importance of forcing errors depended on the model output of interest. Sensitivity analysis can reveal which forcing error characteristics matter most for hydrologic modeling.

1 Introduction

Physically based models allow researchers to test hypotheses about the role of specific processes in hydrologic systems and how changes in environment (e.g., climate, land cover) may impact key hydrologic fluxes and states (Barnett et al., 2008; Deems et al.,

HESSD
11, 13745–13795, 2014

2013; Leavesley, 1994). Due to the complexity of processes represented, these models usually require numerous inputs consisting of (1) meteorological forcing variables and (2) model parameters. Most inputs are not measured at the locations of interest and require estimation; hence, large uncertainties may propagate from hydrologic model inputs to outputs. Despite ongoing efforts to quantify forcing uncertainties (e.g., Bohn et al., 2013; Flerchinger et al., 2009) and to develop methodologies for incorporating uncertainty into modeling efforts (e.g., Clark and Slater, 2006; He et al., 2011a; Kavetski et al., 2006a; Kuczera et al., 2010), many analyses continue to ignore uncertainty. These often assume either that all forcings, parameters, and structure are correct (Pappenberger and Beven, 2006) or that only parametric uncertainty is important (Vrugt et al., 2008b). Neglecting uncertainty in hydrologic modeling reduces confidence in hypothesis tests (Clark et al., 2011), thereby limiting the usefulness of physically based models.

There are fewer detailed studies focusing on forcing uncertainty relative to the number of parametric and structural uncertainty studies (Bastola et al., 2011; Benke et al., 2008; Beven and Binley, 1992; Butts et al., 2004; Clark et al., 2008, 2011; Essery et al., 2013; Georgakakos et al., 2004; Jackson et al., 2003; Kuczera and Parent, 1998; Liu and Gupta, 2007; Refsgaard et al., 2006; Slater et al., 2001; Smith et al., 2008; Vrugt et al., 2003a, b, 2005; Yilmaz et al., 2008). Di Baldassarre and Montanari (2009) suggest that forcing uncertainty has attracted less attention because it is "often considered negligible" relative to parametric and structural uncertainties. Nevertheless, forcing uncertainty merits more attention in some cases, such as in snow-affected watersheds where meteorological and energy balance measurements are scarce (Bales et al., 2006; Raleigh, 2013; Schmucki et al., 2014) and prone to errors (Huwald et al., 2009; Rasmussen et al., 2012). Forcing uncertainty is enhanced in complex terrain where meteorological variables exhibit high spatial variability (Feld et al., 2013; Flint and Childs, 1987; Herrero and Polo, 2012; Lundquist and Cayan, 2007). As a result, the choice of forcing data can yield substantial differences in calibrated model parameters (Elsner et al., 2014) and in modeled hydrologic processes, such as snowmelt and

HESSD
11, 13745–13795, 2014

evapotranspiration (Mizukami et al., 2014; Wayand et al., 2013). Thus, forcing uncertainty demands more attention in snow-affected watersheds.

Previous work on forcing uncertainty in snow-affected regions has yielded basic insights into how forcing errors propagate to model outputs and which forcings introduce the most uncertainty in specific outputs. However, these studies have typically been limited to: (1) empirical/conceptual models (He et al., 2011a, b; Raleigh and Lundquist, 2012; Shamir and Georgakakos, 2006; Slater and Clark, 2006), (2) errors for a subset of forcings (e.g., precipitation or temperature only) (Burles and Boon, 2011; Dadic et al., 2013; Durand and Margulis, 2008; Xia et al., 2005), (3) model sensitivity to choice of forcing parameterization (e.g., longwave) without considering uncertainty in parameterization inputs (e.g., temperature and humidity) (Guan et al., 2013), and (4) simple representations of forcing errors (e.g., Kavetski et al., 2006a, b). The last is evident in studies that only consider single types of forcing errors (e.g., bias) and single distributions (e.g., uniform), and examines errors separately (Burles and Boon, 2011; Koivusalo and Heikinheimo, 1999; Raleigh and Lundquist, 2012; Xia et al., 2005). Examining uncertainty in one factor at a time remains popular but fails to explore the uncertainty space adequately, ignoring potential interactions between forcing errors (Saltelli and Annoni, 2010; Saltelli, 1999). Global sensitivity analysis explores the uncertainty space more comprehensively by considering uncertainty in multiple factors at the same time.

The purpose of this paper is to assess how specific forcing error characteristics influence outputs of a physically based snow model. To our knowledge, no previous study has investigated this topic in snow-affected regions. It is unclear how (1) different error types (bias vs. random errors), (2) different error distributions, and (3) different error magnitudes across all forcings affect model output. The motivating research question is "how do assumptions regarding forcing error characteristics impact our understanding of uncertainty in physically based model output?" Using the (Sobol, 1990) global sensitivity analysis framework, we investigate how artificial errors introduced into high-quality observed forcings (temperature, precipitation, wind speed, humidity, shortwave

HESSD
11, 13745–13795, 2014

radiation, and longwave radiation) at four sites in contrasting snow climates propagate to four snow model outputs (peak snow water equivalent, ablation rates, snow disappearance timing, and sublimation) that are important to cold regions hydrology. We select a single model structure and set of parameters to clarify the impact of forcing uncertainty on model outputs. Specifically, we use the physically based Utah Energy Balance (UEB) snow model (Mahat and Tarboton, 2012; Tarboton and Luce, 1996) because it is computationally efficient. The presented framework could be extended to other models.

2 Study sites and data

We selected four seasonally snow covered study sites (Table 1) in distinct snow climates (Sturm et al., 1995; Trujillo and Molotch, 2014). The sites included (1) the tundra Imnavait Creek (IC, 930 m) site (Euskirchen et al., 2012; Kane et al., 1991; Sturm and Wagner, 2010), located north of the Brooks Range in Alaska, USA, (2) the maritime Col de Porte (CDP, 1330 m) site (Morin et al., 2012) in the Chartreuse Range in the Rhône-Alpes of France, (3) the intermountain Reynolds Mountain East (RME, 2060 m) sheltered site (Reba et al., 2011) in the Owyhee Range in Idaho, USA, and (4) the continental Swamp Angel Study Plot (SASP, 3370 m) site (Landry et al., 2014) in the San Juan Mountains of Colorado, USA.

The sites had high-quality observations of the model forcings at hourly time steps. Serially complete published datasets are available at CDP, RME, and SASP (see citations above). At IC, data were available from multiple co-located stations (Griffin et al., 2010; Bret-Harte et al., 2010a, b, 2011a, b, c; Sturm and Wagner, 2010); these data were quality controlled, and gaps in the data were filled as described in Raleigh (2013). We considered only one year for analysis at each site (Table 1) due to the high computational costs of the modeling experiment. Measured evaluation data (e.g., snow water equivalent, SWE) at daily resolution were used for qualitative assessment of model output. SWE was observed at snow pillows at IC, CDP, and RME. At SASP, acoustic

HESSD
11, 13745–13795, 2014

snow depth data were converted to daily SWE using density from nearby sites and local snow pit measurements (Raleigh, 2013).

We adjusted the available precipitation data at each site with a multiplicative factor to ensure the base model simulation with all observed forcings reasonably represented observed SWE before conducting the sensitivity analysis. Schmucki et al. (2014) demonstrated that precipitation adjustments are necessary for realistic SWE simulations even at well-instrumented sites. Precipitation adjustments were most necessary at IC, where windy conditions preclude effective measurements (Yang et al., 2000). In contrast, only modest adjustments were necessary at the other three sites because they were located in sheltered clearings and because the data already had some corrections applied in the published data. Precipitation data were increased by 60 % at IC and 15 % at SASP, and decreased by 10 % at CDP and RME. The initial discrepancies between modeled and observed SWE may have resulted from deficiencies in the measured forcings, model parameters, model structure, and measured verification data. It was beyond the scope of this study to optimize model parameters and unravel the relative contributions of uncertainty for factors other than the meteorological forcings.

3 Methods

3.1 Model and output metrics

The Utah Energy Balance (UEB) is a physically based, one-dimensional snow model (Mahat and Tarboton, 2012; Tarboton and Luce, 1996; You et al., 2013). UEB represents processes such as snow accumulation, snowmelt, albedo decay, surface temperature variation, liquid water retention, and sublimation. UEB has a single bulk snow layer and an infinitesimally thin surface layer for energy balance computations at the snow-atmosphere interface. UEB tracks state variables for snowpack energy content, SWE, and a dimensionless snow surface age (for albedo computations). We ran UEB at

HESSD
11, 13745–13795, 2014

hourly time steps with six forcings: air temperature (T_{air}), precipitation (P), wind speed (U), relative humidity (RH), incoming shortwave radiation (Q_{si}), and incoming longwave radiation (Q_{li}). We used fixed parameters across all scenarios (Table 2). We initialized UEB during the snow-free period; thus, model spin-up was unnecessary.

With each UEB simulation, we calculated four summary output metrics: (1) peak (i.e., maximum) SWE, (2) mean ablation rate, (3) snow disappearance date, and (4) total annual snow sublimation/frost. The first three metrics are important for the timing and magnitude of water availability and identification of snowpack regime (Trujillo and Molotch, 2014), while the fourth impacts the partitioning of annual P into runoff and evapotranspiration. We calculated the snow disappearance date as the first date when 90 % of peak SWE had ablated, similar to other studies that use a minimum SWE threshold for defining snow disappearance (e.g., Schmucki et al., 2014). The mean ablation rate was calculated in the period between peak SWE and snow disappearance, and was taken as the absolute value of the mean of all SWE decreases.

3.2 Forcing error scenarios

To test how error characteristics in forcings affect model outputs, we created four scenarios (Fig. 1 and Table 3) with different assumptions regarding error types, distributions, and magnitudes (i.e., error ranges). In the first scenario, only bias (normally or lognormally distributed) was introduced into all forcings at a level of high uncertainty (based on values observed in the field, see Sect. 3.2.3 below). This scenario was named "NB," where N denotes normal (or lognormal) error distributions and B denotes bias only. The remaining three scenarios were identical to NB except one aspect was changed: scenario NB + RE considered both bias and random errors (RE), scenario UB considered uniformly distributed biases, and scenario NB_lab considered error magnitudes at minimal values (i.e., specified instrument accuracy as found in a laboratory). Constructed in this way (Fig. 1), we could test model sensitivity to (1) bias vs. random errors by comparing NB and NB + RE, (2) error distributions by comparing NB and UB, and (3) error magnitudes by comparing NB and NB_lab.

HESSD
11, 13745–13795, 2014

3.2.1 Error types

Forcing data inevitably have some (unknown) combination of bias and random errors. However, hydrologic sensitivity analyses have tended to focus more on bias with little or no attention to random errors (Raleigh and Lundquist, 2012), and rarely any consideration of interactions between error types. Lapo et al. (2014) tested biases and random errors in Q_{si} and Q_{li} forcings, finding that biases generally introduced more variance in modeled SWE than random errors. Their experiment considered biases and random errors separately (i.e., no error interactions allowed), and examined only a subset of the required forcings (i.e., radiation). Here, we examined co-existing biases in all forcings in NB, UB, and NB_Lab, and co-existing biases and random errors in all forcings in NB + RE.

Table 3 shows the assignment of error types for the four scenarios. We relied on studies that assess errors in measurements or estimated forcings to identify typical characteristics of biases and random errors. Published bias values were more straightforward to interpret than random errors because common metrics, such as root mean squared error (RMSE) and mean absolute error (MAE), encapsulate both systematic and random errors. Hence, when defining random errors, published RMSE and MAE served as qualitative guidelines.

3.2.2 Error distributions

Error distributions (Table 3) were the same across scenarios NB, NB + RE, and NB_lab. The UB scenario adopted a naive hypothesis that the probability distribution of biases was uniform, a common assumption in sensitivity analyses. However, a uniform distribution implies that extreme and small biases are equally probable. It is likely that error distributions more closely resemble non-uniform distributions (e.g., normal distributions) in reality.

Unfortunately, error distributions are reported less frequently than error statistics (e.g., bias, RMSE) in the literature. T_{air} and RH errors likely follow normal distributions

HESSD
11, 13745–13795, 2014

(Mardikis et al., 2005; Phillips and Marks, 1996), as do Q_{si} and Q_{li} errors (T. Bohn, personal communication, 2014). Conflicting reports over the distribution of U indicated that errors may be approximated with a normal (Phillips and Marks, 1996), a lognormal (Mardikis et al., 2005), or a Weibull distribution (Jiménez et al., 2011). For simplicity, we assumed that U errors were normally distributed. Finally, we assumed P errors followed a lognormal distribution to account for snow redistribution due to wind drift/scour (Liston, 2004). Error distributions were truncated in cases when the introduced errors violated physical limits (e.g., negative U; see Sect. 3.3.5).

3.2.3 Error magnitudes

We considered two magnitudes of forcing uncertainty: levels of uncertainty found in the (1) field vs. (2) a controlled laboratory setting (Table 3). Field and laboratory cases were considered because they sampled realistic errors and minimum errors, respectively. We expected that the error ranges exerted a major control on model uncertainty and sensitivity.

NB, NB + RE, and UB considered field uncertainties. Field uncertainties depend on the source of forcing data and on local conditions (e.g., Flerchinger et al., 2009). To generalize the analysis, we chose error ranges that enveloped the reported uncertainty of different methods for acquiring forcing data. T_{air} error ranges spanned errors in measurements (Huwald et al., 2009) and commonly used models, such as lapse rates and statistical methods, (Bolstad et al., 1998; Chuanyan et al., 2005; Fridley, 2009; Hasenauer et al., 2003; Phillips and Marks, 1996). P error ranges spanned undercatch (e.g., Rasmussen et al., 2012) and drift/scour errors. Because UEB lacks dynamic wind redistribution, accumulation uncertainty was not linked to U but instead to P errors (e.g., drift factor, Luce et al., 1998). Results were thus most relevant to areas with prominent snow redistribution (e.g., alpine zone). U error ranges spanned errors in topographic drift models (Liston and Elder, 2006; Winstral et al., 2009) and numerical weather prediction (NWP) models (Cheng and Georgakakos, 2011). RH error ranges spanned errors in observations (Déry and Stieglitz, 2002) and empirical methods (e.g., Bohn et al.,

HESSD
11, 13745–13795, 2014

2013; Feld et al., 2013). Q_{si} error ranges spanned errors in empirical methods (Bohn et al., 2013), radiative transfer models (Jing and Cess, 1998), satellite-derived products (Jepsen et al., 2012), and NWP models (Niemelä et al., 2001b). Q_{li} error ranges spanned errors in empirical methods (Bohn et al., 2013; Flerchinger et al., 2009; Herrero and Polo, 2012) and NWP models (Niemelä et al., 2001a).

In contrast, scenario NB_lab assumed laboratory levels of uncertainty for each forcing. These uncertainty levels vary with the type and quality of sensors, as well as related accessories (e.g., wind shield for a *P* gauge), which we did not explicitly consider. We assumed that the manufacturers' specified accuracy of meteorological sensors at a typical SNOw TELemetry (SNOTEL) (Serreze et al., 1999) site in the western USA were representative of minimum uncertainties in forcings because of the widespread use of SNOTEL data in snow studies. While we used the specified accuracy for *P* measurements in NB_lab, we note that the instrument uncertainty of ±3 % was likely unrepresentative of errors likely to be encountered. For example, corrections applied to the *P* data (see Sect. 2) exceeded this uncertainty by factors of 3 to 20.

3.3 Sensitivity analysis

Numerous approaches that explore uncertainty in numerical models have been developed in the literature of statistics (Christopher Frey and Patil, 2002), environmental modeling (Matott et al., 2009), and hydrology model optimization/calibration (Beven and Binley, 1992; Duan et al., 1992; Kavetski et al., 2002, 2006a, b; Kuczera et al., 2010; Vrugt et al., 2008a, b). Among these, global sensitivity analysis is an elegant platform for testing the impact of input uncertainty on model outputs and for ranking the relative importance of inputs while considering co-existing sources of uncertainty. Global methods are ideal for non-linear models (e.g., snow models). The Sobol (1990) (hereafter Sobol') method is a robust global method based on the decomposition of variance (see below). We investigate Sobol', as it is often the baseline for testing sensitivity analysis methods (Herman et al., 2013; Li et al., 2013; Rakovec et al., 2014; Tang et al., 2007).

HESSD
11, 13745–13795, 2014

3.3.1 Overview

One can visualize any hydrology or snow model as:

$$\mathbf{Y} = M(\mathbf{F}, \theta) \quad (1)$$

where **Y** is a matrix of model outputs (e.g., SWE), $M(\,)$ is the model operator, **F** is a matrix of forcings (e.g., T_{air}, P, U, etc.), and θ is an array of model parameters (e.g., snow surface roughness). The goal of sensitivity analysis is to quantify how variance (i.e., uncertainty) in specific input factors (**F** and θ) influences variance in specific outputs (**Y**). Sensitivity analyses tend to focus more on the model parameter array (θ) than on the forcing matrix (Foglia et al., 2009; Herman et al., 2013; Li et al., 2013; Nossent et al., 2011; Rakovec et al., 2014; Rosero et al., 2010; Rosolem et al., 2012; Tang et al., 2007; van Werkhoven et al., 2008). Here, we extend the sensitivity analysis framework to forcing uncertainty by creating k new parameters ($\theta_1, \theta_2, \ldots \theta_k$) that specify forcing uncertainty characteristics (Vrugt et al., 2008b). By fixing the original model parameters (Table 2), we focus solely on the influence of forcing errors on model output (**Y**). Note it is possible to consider uncertainty in both forcings and parameters in this framework.

3.3.2 Sobol' sensitivity analysis

Sobol' sensitivity analysis uses variance decomposition to attribute output variance to input variance. First-order and higher-order sensitivities can be resolved; here, only the total-order sensitivity is examined (see below). The Sobol' method is advantageous in that it is model independent, can handle non-linear systems, and is among the most robust sensitivity methods (Saltelli and Annoni, 2010; Saltelli, 1999). The primary limitation of Sobol is that it is computationally intensive, requiring a large number of samples to account for variance across the full parameter space. Below, we describe the methodology but note Saltelli and Annoni (2010) provide an excellent overview for designing a Sobol' sensitivity analysis.

HESSD
11, 13745–13795, 2014

3.3.3 Sensitivity indices and sampling

Within the Sobol' global sensitivity analysis framework, the total-order sensitivity index (S_{Ti}) describes the variance in model outputs (**Y**) due to a specific parameter (θ_i), including both unique (i.e., first-order) effects and all interactions with all other parameters:

$$S_{Ti} = \frac{E[V(\mathbf{Y}|\theta_{\sim i})]}{V(\mathbf{Y})} = 1 - \frac{V[E(\mathbf{Y}|\theta_{\sim i})]}{V(\mathbf{Y})} \quad (2)$$

where E is the expectation (i.e., average) operator, V is the variance operator, and $\theta_{\sim i}$ signifies all parameters except θ_i. The latter expression defines S_{Ti} as the variance remaining in **Y** after accounting for variance due to all other parameters ($\theta_{\sim i}$). S_{Ti} values have a range of [0, 1]. Interpretation of S_{Ti} values was straightforward because they explicitly quantified the variance introduced to model output by each parameter (i.e., forcing errors). As an example, an S_{Ti} value of 0.7 for bias parameter θ_i on output Y_j indicates 70 % of the output variance was due bias in forcing i (including unique effects and interactions).

Selecting points in the k-dimensional space for calculating S_{Ti} was achieved using the quasi-random Sobol' sequence (Saltelli and Annoni, 2010). The sequence has a uniform distribution in the range [0, 1]. Figure 2a shows an example Sobol' sequence in two dimensions.

Evaluation of Eq. (2) requires two sampling matrices, which we built with the Sobol' sequence and refer to as matrices **A** and **B** (Fig. 2a). To construct **A** and **B**, we first specified the number of samples (N) in the parameter space and the number of parameters (k), depending on the error scenario (Table 3). For each scenario and site, we generated a ($N \times 2k$) Sobol' sequence matrix with quasi-random numbers in the [0, 1] range, and then divided it in two parts such that matrices **A** and **B** were each distinct (N x k) matrices. Calculation of S_{Ti} required perturbing parameters; therefore, a third Sobol' matrix ($\mathbf{A_B}$) was constructed from **A** and **B**. In matrix $\mathbf{A_B}$, all columns were from

HESSD
11, 13745–13795, 2014

A, except the *i*th column, which was from the *i*th column of **B**, resulting in a (kN x k) matrix (Fig. 2a). Section 3.3.5 provides specific examples of this implementation.

A number of numerical methods are available for evaluating sensitivity indices (Saltelli and Annoni, 2010). From Eq. (2), we compute S_{Ti} as (Jansen, 1999; Saltelli and Annoni, 2010):

$$S_{Ti} = \frac{\frac{1}{2N}\sum_{j=1}^{N}\left[f(\mathbf{A})_j - f\left(\mathbf{A_B}^{(i)}\right)_j\right]^2}{V(\mathbf{Y})} \quad (3)$$

where $f(\mathbf{A})$ is the model output evaluated on the **A** matrix, $f(\mathbf{A_B}^{(i)})$ is the model output evaluated on the $\mathbf{A_B}$ matrix where the *i*th column is from the **B** matrix, and *i* designates the parameter of interest. Evaluation of S_{Ti} required $N(k+2)$ simulations at each site and scenario.

3.3.4 Bootstrapping of sensitivity indices

To test the reliability of S_{Ti}, we used bootstrapping with replacement across the $N(k+2)$ outputs, similar to Nossent et al. (2011). The mean and 95 % confidence interval were calculated using the Archer et al. (1997) percentile method and 10 000 samples. For all cases, final S_{Ti} values were generally close to the mean bootstrapped values, suggesting convergence. Thus, we report only the mean and 95 % confidence intervals of the bootstrapped S_{Ti} values.

3.3.5 Workflow and error introduction

Figure 2 shows the workflow for creating the Sobol' **A**, **B**, and $\mathbf{A_B}$ matrices, converting Sobol' values to errors, applying errors to the original forcing data, executing the model and saving outputs, and calculating S_{Ti} values. The workflow was repeated at all sites and scenarios. Each step is described in more detail below:

HESSD
11, 13745–13795, 2014

Step 1) Generate an initial ($N \times 2k$) Sobol' matrix (with N and k values for each scenario, Table 3), separate into **A** and **B**, and construct $\mathbf{A_B}$ (Fig. 2a). NB + RE had $k = 12$ (six bias and six random error parameters). All other scenarios had $k = 6$ (all bias parameters).

Step 2) In each simulation, map the Sobol' value of each forcing error parameter (θ) to the specified error distribution and range (Fig. 2b, Table 3). For example, $\theta = 0.75$ would map to a Q_{si} bias of $+50\,W\,m^{-2}$ for a uniform distribution in the range $[-100\,W\,m^{-2}, +100\,W\,m^{-2}]$.

Step 3) In each simulation, perturb (i.e., introduce artificial errors) the observed time series of the ith forcing (F_i) with bias (all scenarios), or both bias and random errors (NB + RE only) (Fig. 2c):

$$F_i' = F_i\theta_{B,i}b_i + (F_i + \theta_{B,i})(1 - b_i) + \theta_{RE,i}Rc_i \quad (4)$$

where F_i' is the perturbed forcing time series, $\theta_{B,i}$ is the bias parameter for forcing i, b_i is a binary switch indicating multiplicative bias ($b_i = 1$) or additive bias ($b_i = 0$), $\theta_{RE,i}$ is the random error parameter for forcing i, R is a time series of randomly distributed noise (normal distribution, mean = 0) scaled in the [−1, 1] range, and c_i is a binary switch indicating whether random errors are introduced ($c_i = 1$ in scenario NB + RE and $c_i = 0$ in all other scenarios). For T_{air}, U, RH, Q_{si}, and Q_{li}, $b_i = 0$; for P, $b_i = 1$. For P, U, and Q_{si}, we restricted random errors to periods with positive values. We checked F_i' for non-physical values (e.g., negative Q_{si}) and set these to physical limits. This was most common when perturbing U, RH, and Q_{si}; negative values of perturbed P only occurred when random errors were considered (Eq. 4). Due to this resetting of non-physical errors, the error distribution was truncated (i.e., it was not always possible to impose extreme errors). Additional tests (not shown) suggested that distribution truncation changed sensitivity indices minimally (i.e., < 5 %) and did not alter the relative ranking of forcing errors.

Step 4) Input the $N(k + 2)$ perturbed forcing datasets into UEB (Fig. 2d). At each site, NB + RE required 140 000 simulations, whereas the other scenarios each required

HESSD
11, 13745–13795, 2014

80 000 simulations, for a total of 1 520 000 simulations in the analysis. The doubling of k in NB + RE did not result in twice as many simulations because the number of simulations scaled as $N(k+2)$.

Step 5) Save the model outputs for each simulation (Fig. 2e).

Step 6) Calculate S_{Ti} for each forcing error parameter and model output (Fig. 2f) based on Sects. 3.3.3–3.3.4. Prior to calculating S_{Ti}, we screened the model outputs for cases where UEB simulated too little or too much snow (which can occur with perturbed forcings). For a valid simulation, we required a minimum peak SWE of 50 mm, a minimum continuous snow duration of 15 d, and identifiable snow disappearance. We rejected samples that did not meet these criteria to avoid meaningless or undefined metrics (e.g., peak SWE in ephemeral snow or snow disappearance for a simulation that did not melt out). The number of rejected samples varied with site and scenario (Table 4). On average, 92 % passed the requirements. All cases had at least 86 % satisfactory samples, except in UB at SASP, where only 34 % met the requirements. Despite this attrition, S_{Ti} values still converged in all cases.

4 Results

4.1 Uncertainty propagation to model outputs

Figure 3 shows density plots of daily SWE from UEB at the four sites and four forcing error scenarios (Fig. 1, Table 3), while Fig. 4 summarizes the model outputs. As a reminder, NB assumed normal (or lognormal) biases at field level uncertainty. The other scenarios were the same as NB, except NB + RE considered both biases and random errors, UB considered uniform distributions, and NB_lab considered lower error magnitudes (i.e., laboratory level uncertainty).

Large uncertainties in SWE were evident, particularly in NB, NB + RE, and UB (Fig. 3a–l). The large range in modeled SWE within these three scenarios often translated to large ranges in mean ablation rates (Fig. 4e–h), snow disappearance dates

HESSD
11, 13745–13795, 2014

(Fig. 4i–l) and total sublimation (Fig. 4m–p). In contrast, SWE and output uncertainty in NB_lab was comparatively small (Figs. 3m–p and 4). The envelope of SWE simulations in NB_lab generally encompassed observed SWE at all sites, except during early winter at IC (Fig. 3m), which was possibly due to initial *P* data quality and redistribution of snow to the snow pillow site. NB_lab simulations were expected to encompass observed SWE due to the adjustments made to the original *P* data (see Sect. 2).

NB and NB + RE generally yielded similar SWE density plots (Fig. 3a–h), but NB + RE yielded a higher frequency of extreme SWE simulations. NB and NB + RE also had very similar (but not equivalent) mean outputs values and ensemble spreads at all sites except IC (Fig. 4). This initial observation suggested that forcing biases contributed more to model uncertainty than random errors at CDP, RME, and SASP. IC may have had higher sensitivity to random errors due to the low snow accumulation at that site and brief snowmelt season (less than 10 d).

NB and UB yielded generally very different model outputs (Figs. 3 and 4). The only difference in these two scenarios was the assumption regarding error distribution (Table 3). Uniformly distributed forcing biases (scenario UB) yielded a more uniform ensemble of SWE simulations (Fig. 3i–l), larger mean values of peak SWE and ablation rates, and later snow disappearance, as well as larger uncertainty ranges in all outputs. At some sites, UB also had a higher frequency of simulations where seasonal sublimation was negative.

Relative to NB, NB_lab had smaller uncertainty ranges in all model outputs (Figs. 3 and 4), an expected result given the lower magnitudes in forcing errors in NB_lab (Table 3). Likewise, NB_lab SWE simulations were generally less biased than NB, relative to observations (Fig. 3). NB_lab generally had higher mean peak SWE and ablation rates, and later mean snow disappearance timing than NB (Fig. 4).

HESSD
11, 13745–13795, 2014

4.2 Model sensitivity to forcing error characteristics

Total-order (S_{Ti}) sensitivity indices from the Sobol' sensitivity analysis are shown in Figs. 5–8 for the four scenarios. In these figures, "B" signifies bias and "RE" signifies random errors. Results from each scenario are described below.

Scenario NB showed that UEB peak SWE was most sensitive to P bias at all sites, with S_{Ti} values ranging from 0.90 to 1.00 (Fig. 5a–d). P bias was also the most important factor for ablation rates and snow disappearance at all sites in scenario NB (Fig. 5e–l). After P bias, T_{air} bias was the next most important factor at CDP while biases in Q_{si} and Q_{li} were secondarily important at RME (Fig. 5f and g). In NB, sublimation was most sensitive to RH bias at IC, CDP, and RME, while at SASP sublimation was most sensitive to U bias (Fig. 5m–p). Q_{si} and Q_{li} biases were secondarily important to sublimation at IC and CDP, while T_{air} bias had secondary importance at RME and SASP.

At all sites in NB + RE, peak SWE was most sensitive to P bias, with S_{Ti} ranging from 0.95 to 1.00 (Fig. 6a–d). At CDP, RME, and SASP in NB + RE, ablation rates and snow disappearance were also most sensitive to P bias, with S_{Ti} ranging from 0.94 to 1.00 (ablation rates) and 0.74 to 0.93 (snow disappearance). At IC, no single error emerged as a dominant control on ablation rates, while snow disappearance was most sensitive to Q_{li} bias (S_{Ti} = 0.75). Sublimation in NB + RE was most sensitive to different errors at each site, where the dominant factors were RH bias at IC, Q_{li} bias at CDP, T_{air} bias at RME, and U bias at SASP (Fig. 6m–p).

At all sites in UB, P bias was most important for peak SWE, ablation rates, and snow disappearance (Fig. 7a–l). The only exception was at IC, where ablation rates had similar sensitivity to P bias and U bias. Sublimation was most sensitive to RH bias at IC, CDP, and RME, and U bias as SASP (Fig. 7m–p). For sublimation in UB, Q_{li} bias was secondarily important at CDP, and U bias was secondarily important at IC and RME.

HESSD
11, 13745–13795, 2014

Relative to the other scenarios, NB_lab portrayed different model sensitivities to forcing errors (Fig. 8). Across all sites and outputs, Q_{li} bias was consistently the most important factor. This was surprising, given that the bias magnitudes were lower for Q_{li} than for Q_{si} (Table 3). Whereas P bias was often important for peak SWE, ablation rates, and snow disappearance in the other scenarios, P bias was seldom important in NB_lab (main exception was peak SWE at IC, Fig. 8a). This was due to the discrepancy between the specified accuracy for P gauges and typical errors encountered in the field (Rasmussen et al., 2012; Sieck et al., 2007).

4.3 Impact of error types

Figure 9 compares the mean S_{Ti} values (above) from NB and NB + RE to test how forcing error type affects model sensitivity. In this test, only the six bias parameters from NB + RE were compared, as these were found in both scenarios. Across sites and model outputs, S_{Ti} values were higher in NB + RE than NB. This suggested that random errors interact with bias, thereby increasing model sensitivity to bias. However, while the S_{Ti} values differed between these two scenarios, the overall importance ranking of forcing biases was generally not altered, and NB offered the same general conclusions regarding the relative impacts of biases in the forcings.

4.4 Impact of error distributions

Figure 10 compares mean S_{Ti} values from NB and UB to test how the distribution of bias influences model outputs. S_{Ti} values from the two scenarios generally plotted close to the 1:1 line, suggesting good correspondence in the sensitivity of UEB under different bias distributions. In other words, the error distribution had little impact on model sensitivity to forcing errors. With a few exceptions where sensitivities of less important terms were clustered (e.g., Fig. 10e), the hierarchy of forcing biases was similar between these two scenarios.

HESSD
11, 13745–13795, 2014

4.5 Impact of error magnitude

Figure 11 compares mean S_{T_i} values from NB and NB_lab to test the importance of forcing error magnitudes to model output. The results showed that the total sensitivity of model outputs to forcing biases depended substantially on the levels of forcing uncertainty considered. As a primary example, the scenarios did not agree whether P bias or Q_{li} bias was the most important factor for peak SWE, ablation rates, and snow disappearance dates at all four sites (Fig. 11a–l). At IC and SASP, peak SWE sensitivity to the secondary forcings (i.e., forcings most important after the most important factor) was greater in NB_lab than NB. In contrast, sublimation was more sensitive to the secondary forcings in NB than NB_lab.

5 Discussion

Here we have examined the sensitivity of physically based snow simulations to forcing error characteristics (i.e., types, distributions, and magnitudes) using Sobol' global sensitivity analysis. Among these characteristics, the magnitude of biases had the most significant impact on UEB simulations (Figs. 3–4) and on model sensitivity (Fig. 11). Random errors were important in that they introduced more interactions in the uncertainty space, as evident in the higher S_{T_i} values in scenario NB + RE vs. NB (Fig. 9), but they were rarely among the most important factors for the model outputs at the sites (Fig. 6). The assumed distribution of biases was important in that it increased the range of model outputs (compare NB and UB in Fig. 4), but this did not often translate to different model sensitivity to forcing errors (Fig. 10).

Our central argument at the onset was that forcing uncertainty may be comparable to parametric and structural uncertainty in snow-affected catchments. To support our argument, we compare our results at CDP in 2005–2006 to Essery et al. (2013), who assessed the impact of structural uncertainty on SWE simulations from 1701 physically based snow models at the same site/year. Comparing our SWE ensemble (Fig. 3b, f,

HESSD
11, 13745–13795, 2014

j, n) to the corresponding ensemble in Fig. 10 of Essery et al. (2013), it is clear that the forcing uncertainty considered in most scenarios here overwhelms the structural uncertainty at this site. Whereas the 1701 models in Essery et al. (2013) generally have peak SWE between 325–450 mm, the 95 % interval due to forcing uncertainty spans 100–565 mm in NB, 110–580 mm in NB + RE, 125–1880 mm in UB, and 370–430 mm in NB_lab (Fig. 4b). Spread in snow disappearance due to structural uncertainty spans mid-April to early-May in Essery et al. (2013), but the range of snow disappearance due to forcing uncertainty spans late-March to early-May in NB and NB + RE, late-March to early July in UB, and mid-April to early-May in NB_lab (Fig. 4j). Structural uncertainty is less impactful on model outputs at this site than the forcing uncertainty of NB, NB + RE, and UB, and is only marginally more impactful than the minimal forcing uncertainty tested in NB_lab. Thus, forcing uncertainty cannot always be discounted, and the magnitude of forcing uncertainty is a critical factor in how forcing uncertainty compares to parametric and structural uncertainty.

It could be argued that forcing uncertainty only appears greater than structural uncertainty in the CDP example because of the large P error magnitudes (Table 3), which are representative of barren areas with drifting snow (e.g., alpine areas, cold prairies) but perhaps not representative of sheltered areas (e.g., forest clearings). To check this, we conducted a separate test (no figures shown) replicating scenario NB with smaller P biases ranging from −10 to +10 %. This P uncertainty range was selected because Meyer et al. (2012) found 95 % of SNOTEL sites (often in forest clearings) had observations of accumulated P within 20 % of peak SWE. This test resulted in a 95 % peak SWE interval of 228–426 mm and a 95 % snow disappearance interval spanning early-April to early-May. These ranges were still larger than the ranges due to structural uncertainty (Essery et al., 2013), further demonstrating the importance of forcing uncertainty in snow-affected areas.

One resounding result in the field uncertainty scenarios was the dominant effect of P bias on modeled peak SWE, ablation rates, and snow disappearance. This confirmed previous reports that P uncertainty is a dominant control on snowpack dynamics (Du-

HESSD
11, 13745–13795, 2014

rand and Margulis, 2008; He et al., 2011a; Schmucki et al., 2014). However, we note wind uncertainty in snow-affected areas can also be important to snowpack dynamics through drift/scour processes (Mott and Lehning, 2010; Winstral et al., 2013), and our "drift factor" formulation (Luce et al., 1998) did not account for the role of wind in P uncertainty. Thus, U is likely more important than the results suggest. Future work could account for this by assigning P errors that are correlated with U.

It was surprising that P bias was often the most critical forcing error for ablation rates (Figs. 5–7). This is contrary to other studies that have suggested the most important factors for snowmelt are radiation, wind, humidity, and temperature (e.g., Zuzel and Cox, 1975). Ablation rates were highly sensitive to P bias because it controlled the timing and length of the ablation season. Positive (negative) P bias extends (truncates) the fraction of the ablation season in the warmest summer months when ablation rates and radiative energy approach maximum values. Trujillo and Molotch (2014) reported a similar result based on SNOTEL observations.

While peak SWE, ablation rates, and snow disappearance dates had similar sensitivities to forcing errors (particularly to P biases), sublimation exhibited notably different sensitivity to forcing errors. P bias was frequently the least important factor for sublimation, in contrast to the other model outputs. In only a few cases (e.g., all sites in NB + RE), P errors explained more than 5 % of uncertainty in modeled sublimation, and these cases were likely tied to the control of P on snowpack duration (when sublimation is possible). Biases in RH, U, and T_{air} were often the major controls on modeled sublimation in the field uncertainty scenarios (i.e., NB, NB + RE, and UB), while Q_{li} bias controlled modeled sublimation in the lab uncertainty scenario (i.e., NB_lab). These results partially agree with the sensitivity analysis of Lapp et al. (2005), who showed the most important forcings for sublimation in the Canadian Rockies were U and Q_{si}. These results suggest that no single forcing is important across all modeled variables, and model sensitivity strongly depends on the output of interest.

The question remains: "what can be done about forcing errors in hydrologic modeling?" First, the results suggest model-based hypothesis testing must account for un-

HESSD
11, 13745–13795, 2014

certainties in forcing data. The results also identify the need for continued research in constraining P uncertainty in snow-affected catchments. This may be achieved using advanced pathways for quantifying snowfall precipitation, such as NWP models (Rasmussen et al., 2011, 2014). However, in a broader sense, the hydrologic community should consider whether deterministic forcings (i.e., single time series for each forcing) are a reasonable practice for physically based models, given the large uncertainties in both future (e.g., climate change) and historical data (especially in poorly monitored catchments) and the complexities of hydrologic systems (Gupta et al., 2008). We suggest that probabilistic model forcings (e.g., Clark and Slater, 2006) present one potential path forward where measures of forcing uncertainty can be explicitly included in the forcing datasets. The challenges are (1) to ensure statistical reliability in our understanding of forcing errors and (2) to assess how best to input probabilistic forcings into current model architectures.

Limitations of the analysis are (1) we do not consider parametric and structural uncertainty and (2) we only consider a single model and sensitivity analysis method. We expect different snow models may yield different sensitivities to forcing uncertainty. For example, both Koivusalo and Heikinheimo (1999) and Lapo et al. (2014) found UEB (Tarboton and Luce, 1996) and the SNTHERM model (Jordan, 1991) exhibited significant differences in radiative and turbulent heat exchange. In other models, we expect P biases would still dominate peak SWE, ablation rates, and snow disappearance timing, but we might expect different sensitivities to other forcing errors. Likewise, it is possible that different sensitivity analysis methods might yield different results than the Sobol' method (Pappenberger et al., 2008).

Finally, while the Sobol' method is often considered the "baseline" method in global sensitivity analysis, we note that it comes at a relatively high computation cost (1 520 000 simulations for examining only a single year at four sites across four error scenarios) and may be prohibitive for many modeling applications. Ongoing research is developing new sensitivity analysis methods that compare well to Sobol' but with reduced computational demands (e.g., Rakovec et al., 2014). We expect that detailed

HESSD
11, 13745–13795, 2014

sensitivity analyses that concurrently consider uncertainty in forcings, parameters, and structure in a hydrologic model will be more feasible in the future with better computing resources and advances in sensitivity analysis methods.

6 Conclusions

Application of the Sobol' sensitivity analysis framework across sites in contrasting snow climates reveals that forcing uncertainty can significantly impact model behavior in snow-affected catchments. Model output uncertainty due to forcings can be comparable to or larger than model uncertainty due to model structure. Key considerations in model sensitivity to forcing errors are the magnitudes of forcing errors and the outputs of interest. For the sensitivity of the model tested, random errors in forcings are generally less important than biases, and the distribution of biases is relatively less important than the magnitude of biases.

The analysis shows how forcing uncertainty might be included in a formal sensitivity analysis framework through the introduction of new parameters that specify the characteristics of forcing uncertainty. The framework could be extended to other physically based models and sensitivity analysis methodologies, and could be used to quantify how uncertainties in model forcings and parameters interact. In future work, it would be interesting to assess the interplay between co-existing uncertainties in forcing errors, model parameters, and model structure, and to test how model sensitivity changes relative to all three sources of uncertainty.

Author contributions. The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Acknowledgements. M. Raleigh was supported by a post-doctoral fellowship in the Advanced Study Program at the National Center for Atmospheric Research (sponsored by NSF). J. Lundquist was supported by NSF (EAR-838166 and EAR-1215771). Thanks to M. Sturm, G. Shaver, S. Bret-Harte, and E. Euskirchen for assistance with Imnavait Creek data, S. Morin for assistance with Col de Porte data, D. Marks for assistance with Reynolds Mountain data, C.

HESSD
11, 13745–13795, 2014

Landry for assistance with Swamp Angel data, and E. Gutmann and P. Mendoza for feedback. For Imnavait Creek data, we acknowledge US Army Cold Regions Research and Engineering Laboratory, the NSF Arctic Observatory Network (AON) Carbon, Water, and Energy Flux monitoring project and the Marine Biological Laboratory, Woods Hole, and the University of Alaska, Fairbanks. The experiment was improved thanks to conversations with D. Slater.

References

Archer, G. E. B., Saltelli, A., and Sobol, I. M.: Sensitivity measures,anova-like techniques and the use of bootstrap, J. Stat. Comput. Sim., 58, 99–120, doi:10.1080/00949659708811825, 1997. 13757

Bales, R. C., Molotch, N. P., Painter, T. H., Dettinger, M. D., Rice, R., and Dozier, J.: Mountain hydrology of the western United States, Water Resour. Res., 42, W08432, doi:10.1029/2005WR004387, 2006. 13747

Barnett, T. P., Pierce, D. W., Hidalgo, H. G., Bonfils, C., Santer, B. D., Das, T., Bala, G., Wood, A. W., Nozawa, T., Mirin, A. A., Cayan, D. R., and Dettinger, M. D.: Human-induced changes in the hydrology of the western United States, Science, 319, 1080–1083, doi:10.1126/science.1152538, 2008. 13746

Bastola, S., Murphy, C., and Sweeney, J.: The role of hydrological modelling uncertainties in climate change impact assessments of Irish river catchments, Adv. Water Resour., 34, 562–576, doi:10.1016/j.advwatres.2011.01.008, 2011. 13747

Benke, K. K., Lowell, K. E., and Hamilton, A. J.: Parameter uncertainty, sensitivity analysis and prediction error in a water-balance hydrological model, Math. Comput. Model., 47, 1134–1149, doi:10.1016/j.mcm.2007.05.017, 2008. 13747

Beven, K. and Binley, A.: The future of distributed models: model calibration and uncertainty prediction, Hydrol. Process., 6, 279–298, doi:10.1002/hyp.3360060305, 1992. 13747, 13754

Bohn, T. J., Livneh, B., Oyler, J. W., Running, S. W., Nijssen, B., and Lettenmaier, D. P.: Global evaluation of MTCLIM and related algorithms for forcing of ecological and hydrological models, Agr. Forest Meteorol., 176, 38–49, doi:10.1016/j.agrformet.2013.03.003, 2013. 13747, 13753, 13754, 13783

HESSD
11, 13745–13795, 2014

Bolstad, P. V., Swift, L., Collins, F., and Régnière, J.: Measured and predicted air temperatures at basin to regional scales in the southern Appalachian mountains, Agr. Forest Meteorol., 91, 161–176, doi:10.1016/S0168-1923(98)00076-8, 1998. 13753, 13783

Bret-Harte, S., Shaver, G., and Euskirchen, E.: Eddy Flux Measurements, Ridge Station, Imnavait Creek, Alaska – 2010, Long Term Ecological Research Network, doi:10.6073/pasta/fb047eaa2c78d4a3254bba8369e6cee5, 2010a. 13749

Bret-Harte, S., Shaver, G., and Euskirchen, E.: Eddy Flux Measurements, Fen Station, Imnavait Creek, Alaska – 2010, Long Term Ecological Research Network, doi:10.6073/pasta/dde37e89dab096bea795f5b111786c8b, 2010b. 13749

Bret-Harte, S., Euskirchen, E., Griffin, K., and Shaver, G.: Eddy Flux Measurements, Tussock Station, Imnavait Creek, Alaska – 2011, Long Term Ecological Research Network, doi:10.6073/pasta/44a62e0c6741b3bd93c0a33e7b677d90, 2011a. 13749

Bret-Harte, S., Euskirchen, E., and Shaver, G.: Eddy Flux Measurements, Fen Station, Imnavait Creek, Alaska – 2011, Long Term Ecological Research Network, doi:10.6073/pasta/50e9676f29f44a8b6677f05f43268840, 2011b. 13749

Bret-Harte, S., Euskirchen, E., and Shaver, G.: Eddy Flux Measurements, Ridge Station, Imnavait Creek, Alaska – 2011, Long Term Ecological Research Network, doi:10.6073/pasta/5d603c3628f53f494f08f895875765e8, 2011c. 13749

Burles, K. and Boon, S.: Snowmelt energy balance in a burned forest plot, Crowsnest Pass, Alberta, Canada, Hydrol. Process., 25, 3012–3029, doi:10.1002/hyp.8067, 2011. 13748

Butts, M. B., Payne, J. T., Kristensen, M., and Madsen, H.: An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation, J. Hydrol., 298, 242–266, doi:10.1016/j.jhydrol.2004.03.042, 2004. 13747

Cheng, F.-Y. and Georgakakos, K. P.: Statistical analysis of observed and simulated hourly surface wind in the vicinity of the Panama Canal, Int. J. Climatol., 31, 770–782, doi:10.1002/joc.2123, 2011. 13753, 13783

Christopher Frey, H. and Patil, S. R.: Identification and review of sensitivity analysis methods, Risk Anal., 22, 553–578, doi:10.1111/0272-4332.00039, 2002. 13754

Chuanyan, Z., Zhongren, N., and Guodong, C.: Methods for modelling of temporal and spatial distribution of air temperature at landscape scale in the southern Qilian mountains, China, Ecol. Model., 189, 209–220, doi:10.1016/j.ecolmodel.2005.03.016, 2005. 13753, 13783

Clark, M. P. and Slater, A. G.: Probabilistic quantitative precipitation estimation in complex terrain, J. Hydrometeorol., 7, 3–22, doi:10.1175/JHM474.1, 2006. 13747, 13766

HESSD
11, 13745–13795, 2014

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T., and Hay, L. E.: Framework for Understanding Structural Errors (FUSE): a modular framework to diagnose differences between hydrological models, Water Resour. Res., 44, W00B02, doi:10.1029/2007WR006735, 2008. 13747

Clark, M. P., Kavetski, D., and Fenicia, F.: Pursuing the method of multiple working hypotheses for hydrological modeling, Water Resour. Res., 47, 1–16, doi:10.1029/2010WR009827, 2011. 13747

Dadic, R., Mott, R., Lehning, M., Carenzo, M., Anderson, B., and Mackintosh, A.: Sensitivity of turbulent fluxes to wind speed over snow surfaces in different climatic settings, Adv. Water Resour., 55, 178–189, doi:10.1016/j.advwatres.2012.06.010, 2013. 13748

Deems, J. S., Painter, T. H., Barsugli, J. J., Belnap, J., and Udall, B.: Combined impacts of current and future dust deposition and regional warming on Colorado River Basin snow dynamics and hydrology, Hydrol. Earth Syst. Sci., 17, 4401–4413, doi:10.5194/hess-17-4401-2013, 2013. 13746

Déry, S. and Stieglitz, M.: A note on surface humidity measurements in the cold Canadian environment, Bound.-Lay. Meteorol., 102, 491–497, doi:10.1023/A:1013890729982, 2002. 13753, 13783

Di Baldassarre, G. and Montanari, A.: Uncertainty in river discharge observations: a quantitative analysis, Hydrol. Earth Syst. Sci., 13, 913–921, doi:10.5194/hess-13-913-2009, 2009. 13747

Duan, Q., Sorooshian, S., and Gupta, V.: Effective and efficient global optimization for conceptual rainfall-runoff models, Water Resour. Res., 28, 1015–1031, doi:10.1029/91WR02985, 1992. 13754

Durand, M. and Margulis, S. A.: Effects of uncertainty magnitude and accuracy on assimilation of multiscale measurements for snowpack characterization, J. Geophys. Res., 113, D02105, doi:10.1029/2007JD008662, 2008. 13748, 13764

Elsner, M. M., Gangopadhyay, S., Pruitt, T., Brekke, L. D., Mizukami, N., and Clark, M. P.: How does the choice of distributed meteorological data affect hydrologic model calibration and streamflow simulations?, J. Hydrometeorol., 15, 1384–1403, doi:10.1175/JHM-D-13-083.1, 2014. 13747

Essery, R., Morin, S., Lejeune, Y., and B Ménard, C.: A comparison of 1701 snow models using observations from an alpine site, Adv. Water Resour., 55, 131–148, doi:10.1016/j.advwatres.2012.07.013, 2013. 13747, 13763, 13764

HESSD
11, 13745–13795, 2014

Euskirchen, E. S., Bret-Harte, M. S., Scott, G. J., Edgar, C., and Shaver, G. R.: Seasonal patterns of carbon dioxide and water fluxes in three representative tundra ecosystems in northern Alaska, Ecosphere, 3, 4, doi:10.1890/ES11-00202.1, 2012. 13749

Feld, S. I., Cristea, N. C., and Lundquist, J. D.: Representing atmospheric moisture content along mountain slopes: examination using distributed sensors in the Sierra Nevada, California, Water Resour. Res., 49, 4424–4441, doi:10.1002/wrcr.20318, 2013. 13747, 13754, 13783

Flerchinger, G. N., Xaio, W., Marks, D., Sauer, T. J., and Yu, Q.: Comparison of algorithms for incoming atmospheric long-wave radiation, Water Resour. Res., 45, 1–13, doi:10.1029/2008WR007394, 2009. 13747, 13753, 13754, 13783

Flint, A. L. and Childs, S. W.: Calculation of solar radiation in mountainous terrain, Agr. Forest Meteorol., 40, 233–249, doi:10.1016/0168-1923(87)90061-X, 1987. 13747

Foglia, L., Hill, M. C., Mehl, S. W., and Burlando, P.: Sensitivity analysis, calibration, and testing of a distributed hydrological model using error-based weighting and one objective function, Water Resour. Res., 45, W06427, doi:10.1029/2008WR007255, 2009. 13755

Fridley, J. D.: Downscaling climate over complex terrain: high finescale (< 1000 m) spatial variation of near-ground temperatures in a montane forested landscape (Great Smoky Mountains), J. Appl. Meteorol. Clim., 48, 1033–1049, doi:10.1175/2008JAMC2084.1, 2009. 13753, 13783

Georgakakos, K., Seo, D., Gupta, H., Schaake, J., and Butts, M.: Towards the characterization of streamflow simulation uncertainty through multimodel ensembles, J. Hydrol., 298, 222–241, doi:10.1016/j.jhydrol.2004.03.037, 2004. 13747

Goodison, B., Louie, P., and Yang, D.: WMO solid precipitation measurement intercomparison: final report, in: Instrum. Obs. Methods Rep. 67, vol. 67, World Meteorol. Organ., Geneva, Switzerland, p. 211, 1998. 13783

Griffin, K., Bret-Harte, S., Shaver, G., and Euskirchen, E.: Eddy Flux Measurements, Tussock Station, Imnavait Creek, Alaska – 2010, Long Term Ecological Research Network, doi:10.6073/pasta/7bba82256e0f5d9ec3d2bc9c25ab9bcf, 2010. 13749

Guan, B., Molotch, N. P., Waliser, D. E., Jepsen, S. M., Painter, T. H., and Dozier, J.: Snow water equivalent in the Sierra Nevada: blending snow sensor observations with snowmelt model simulations, Water Resour. Res., 49, 5029–5046, doi:10.1002/wrcr.20387, 2013. 13748

HESSD
11, 13745–13795, 2014

Guan, H., Wilson, J. L., and Makhnin, O.: Geostatistical mapping of mountain precipitation incorporating autosearched effects of terrain and climatic characteristics, J. Hydrometeorol., 6, 1018–1031, doi:10.1175/JHM448.1, 2005. 13783

Gupta, H. V., Wagener, T., and Liu, Y.: Reconciling theory with observations: elements of a diagnostic approach to model evaluation, Hydrol. Process., 22, 3802–3813, doi:10.1002/hyp.6989, 2008. 13766

Hasenauer, H., Merganicova, K., Petritsch, R., Pietsch, S. A., and Thornton, P. E.: Validating daily climate interpolations over complex terrain in Austria, Agr. Forest Meteorol., 119, 87–107, doi:10.1016/S0168-1923(03)00114-X, 2003. 13753, 13783

He, M., Hogue, T. S., Franz, K. J., Margulis, S. A., and Vrugt, J. A.: Corruption of parameter behavior and regionalization by model and forcing data errors: a Bayesian example using the SNOW17 model, Water Resour. Res., 47, 1–17, doi:10.1029/2010WR009753, 2011a. 13747, 13748, 13765

He, M., Hogue, T. S., Franz, K. J., Margulis, S. A., and Vrugt, J. A.: Characterizing parameter sensitivity and uncertainty for a snow model across hydroclimatic regimes, Adv. Water Resour., 34, 114–127, doi:10.1016/j.advwatres.2010.10.002, 2011b. 13748

Herman, J. D., Kollat, J. B., Reed, P. M., and Wagener, T.: Technical Note: Method of Morris effectively reduces the computational demands of global sensitivity analysis for distributed watershed models, Hydrol. Earth Syst. Sci., 17, 2893–2903, doi:10.5194/hess-17-2893-2013, 2013. 13754, 13755

Herrero, J. and Polo, M. J.: Parameterization of atmospheric longwave emissivity in a mountainous site for all sky conditions, Hydrol. Earth Syst. Sci., 16, 3139–3147, doi:10.5194/hess-16-3139-2012, 2012. 13747, 13754, 13783

Hutchinson, M. F., McKenney, D. W., Lawrence, K., Pedlar, J. H., Hopkinson, R. F., Milewska, E., and Papadopol, P.: Development and testing of Canada-wide interpolated spatial models of daily minimum–maximum temperature and precipitation for 1961–2003, J. Appl. Meteorol. Clim., 48, 725–741, doi:10.1175/2008JAMC1979.1, 2009. 13783

Huwald, H., Higgins, C. W., Boldi, M.-O., Bou-Zeid, E., Lehning, M., and Parlange, M. B.: Albedo effect on radiative errors in air temperature measurements, Water Resour. Res., 45, 1–13, doi:10.1029/2008WR007600, 2009. 13747, 13753, 13783

Jackson, C., Xia, Y., Sen, M. K., and Stoffa, P. L.: Optimal parameter and uncertainty estimation of a land surface model: a case study using data from Cabauw, Netherlands, J. Geophys. Res., 108, 4583, doi:10.1029/2002JD002991, 2003. 13747

HESSD
11, 13745–13795, 2014

Jansen, M. J.: Analysis of variance designs for model output, Comput. Phys. Commun., 117, 35–43, doi:10.1016/S0010-4655(98)00154-4, 1999. 13757

Jepsen, S. M., Molotch, N. P., Williams, M. W., Rittger, K. E., and Sickman, J. O.: Interannual variability of snowmelt in the Sierra Nevada and Rocky Mountains, United States: examples from two alpine watersheds, Water Resour. Res., 48, 1–15, doi:10.1029/2011WR011006, 2012. 13754, 13783

Jiménez, P. A., Dudhia, J., and Navarro, J.: On the surface wind speed probability density function over complex terrain, Geophys. Res. Lett., 38, L22803, doi:10.1029/2011GL049669, 2011. 13753

Jing, X. and Cess, R. D.: Comparison of atmospheric clear-sky shortwave radiation models to collocated satellite and surface measurements in Canada, J. Geophys. Res., 103, 28817, doi:10.1029/1998JD200012, 1998. 13754, 13783

Jordan, R.: A One-Dimensional Temperature Model for a Snow Cover: Technical Documentation for SNTHER M.89, p. 58, Special Report 91–16, US Army CRREL, Special Report 91–16, US Army CRREL, Hanover, NH, USA, 1991. 13766

Kane, D. L., Hinzman, L. D., Benson, C. S., and Liston, G. E.: Snow hydrology of a headwater Arctic basin: 1. Physical measurements and process studies, Water Resour. Res., 27, 1099–1109, doi:10.1029/91WR00262, 1991. 13749

Kavetski, D., Franks, S. W., and Kuczera, G.: Confronting input uncertainty in environmental modelling, in: Calibration of Watershed Models, edited by: Duan, Q., Gupta, H. V., Sorooshian, S., Roussea, A. N., and Turcotte, R., American Geophysical Union, Washington, D.C., doi:10.1029/WS006p0049, 49–68, 2002. 13754

Kavetski, D., Kuczera, G., and Franks, S. W.: Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, Water Resour. Res., 42, W03407, doi:10.1029/2005WR004368, 2006a. 13747, 13748, 13754

Kavetski, D., Kuczera, G., and Franks, S. W.: Bayesian analysis of input uncertainty in hydrological modeling: 2. Application, Water Resour. Res., 42, W03408, doi:10.1029/2005WR004376, 2006b. 13748, 13754

Koivusalo, H. and Heikinheimo, M.: Surface energy exchange over a boreal snowpack: comparison of two snow energy balance models, Hydrol. Process., 13, 2395–2408, doi:10.1002/(SICI)1099-1085(199910)13:14/15<2395::AID-HYP864>3.0.CO;2-G, 1999. 13748, 13766

HESSD
11, 13745–13795, 2014

Kuczera, G. and Parent, E.: Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm, J. Hydrol., 211, 69–85, doi:10.1016/S0022-1694(98)00198-X, 1998. 13747

Kuczera, G., Renard, B., Thyer, M., and Kavetski, D.: There are no hydrological monsters, just models and observations with large uncertainties!, Hydrolog. Sci. J., 55, 980–991, doi:10.1080/02626667.2010.504677, 2010. 13747, 13754

Landry, C. C., Buck, K. A., Raleigh, M. S., and Clark, M. P.: Mountain system monitoring at Senator Beck Basin, San Juan Mountains, Colorado: a new integrative data source to develop and evaluate models of snow and hydrologic processes, Water Resour. Res., 50, 1773–1788, doi:10.1002/2013WR013711, 2014. 13749

Lapo, K., Hinkelman, L., Raleigh, M., and Lundquist, J.: Impact of errors in the surface radiation balance on simulations of snow water equivalent and snow surface temperature, Water Resour. Res., in review, 2014. 13752, 13766

Lapp, S., Byrne, J., Townshend, I., and Kienzle, S.: Climate warming impacts on snowpack accumulation in an alpine watershed, Int. J. Climatol., 25, 521–536, doi:10.1002/joc.1140, 2005. 13765

Leavesley, G. H.: Modeling the effects of climate change on water resources – a review, Climatic Change, 28, 159–177, doi:10.1007/BF01094105, 1994. 13747

Li, J., Duan, Q. Y., Gong, W., Ye, A., Dai, Y., Miao, C., Di, Z., Tong, C., and Sun, Y.: Assessing parameter importance of the Common Land Model based on qualitative and quantitative sensitivity analysis, Hydrol. Earth Syst. Sci., 17, 3279–3293, doi:10.5194/hess-17-3279-2013, 2013. 13754, 13755

Liston, G. E.: Representing subgrid snow cover heterogeneities in regional and global models, J. Climate, 17, 1381–1397, doi:10.1175/1520-0442(2004)017<1381:RSSCHI>2.0.CO;2, 2004. 13753

Liston, G. E. and Elder, K.: A Meteorological Distribution System for High-Resolution Terrestrial Modeling (MicroMet), J. Hydrometeorol., 7, 217–234, doi:10.1175/JHM486.1, 2006. 13753, 13783

Liu, Y. and Gupta, H. V.: Uncertainty in hydrologic modeling: toward an integrated data assimilation framework, Water Resour. Res., 43, W07401, doi:10.1029/2006WR005756, 2007. 13747

HESSD
11, 13745–13795, 2014

Luce, C. H., Tarboton, D. G., and Cooley, K. R.: The influence of the spatial distribution of snow on basin-averaged snowmelt, Hydrol. Process., 12, 1671–1683, doi:10.1002/(SICI)1099-1085(199808/09)12:10/11<1671::AID-HYP688>3.0.CO;2-N, 1998. 13753, 13765, 13783

Lundquist, J. D. and Cayan, D. R.: Surface temperature patterns in complex terrain: daily variations and long-term change in the central Sierra Nevada, California, J. Geophys. Res., 112, D11124, doi:10.1029/2006JD007561, 2007. 13747

Luo, W., Taylor, M. C., and Parker, S. R.: A comparison of spatial interpolation methods to estimate continuous wind speed surfaces using irregularly distributed data from England and Wales, Int. J. Climatol., 28, 947–959, doi:10.1002/joc.1583, 2008. 13783

Mahat, V. and Tarboton, D. G.: Canopy radiation transmission for an energy balance snowmelt model, Water Resour. Res., 48, 1–16, doi:10.1029/2011WR010438, 2012. 13749, 13750

Mardikis, M. G., Kalivas, D. P., and Kollias, V. J.: Comparison of interpolation methods for the prediction of reference evapotranspiration an application in Greece, Water Resour. Manag., 19, 251–278, doi:10.1007/s11269-005-3179-2, 2005. 13753

Matott, L. S., Babendreier, J. E., and Purucker, S. T.: Evaluating uncertainty in integrated environmental models: a review of concepts and tools, Water Resour. Res., 45, W06421, doi:10.1029/2008WR007301, 2009. 13754

Meyer, J. D. D., Jin, J., and Wang, S.-Y.: Systematic patterns of the inconsistency between snow water equivalent and accumulated precipitation as reported by the snowpack telemetry network, J. Hydrometeorol., 13, 1970–1976, doi:10.1175/JHM-D-12-066.1, 2012. 13764

Mizukami, N., Clark, M. P., Slater, A. G., Brekke, L. D., Elsner, M. M., Arnold, J. R., and Gangopadhyay, S.: Hydrologic implications of different large-scale meteorological model forcing datasets in mountainous regions, J. Hydrometeorol., 15, 474–488, doi:10.1175/JHM-D-13-036.1, 2014. 13748

Morin, S., Lejeune, Y., Lesaffre, B., Panel, J.-M., Poncet, D., David, P., and Sudul, M.: An 18-yr long (1993–2011) snow and meteorological dataset from a mid-altitude mountain site (Col de Porte, France, 1325 m alt.) for driving and evaluating snowpack models, Earth Syst. Sci. Data, 4, 13–21, doi:10.5194/essd-4-13-2012, 2012. 13749

Mott, R. and Lehning, M.: Meteorological modeling of very high-resolution wind fields and snow deposition for mountains, J. Hydrometeorol., 11, 934–949, doi:10.1175/2010JHM1216.1, 2010. 13765

HESSD
11, 13745–13795, 2014

Niemelä, S., Räisänen, P., and Savijärvi, H.: Comparison of surface radiative flux parameterizations: Part I. Longwave radiation, Atmos. Res., 58, 1–18, doi:10.1016/S0169-8095(01)00084-9, 2001a. 13754, 13783

Niemelä, S., Räisänen, P., and Savijärvi, H.: Comparison of surface radiative flux parameterizations: Part II. Shortwave radiation, Atmos. Res., 58, 141–154, doi:10.1016/S0169-8095(01)00085-0, 2001b. 13754, 13783

Nossent, J., Elsen, P., and Bauwens, W.: Sobol' sensitivity analysis of a complex environmental model, Environ. Modell. Softw., 26, 1515–1525, doi:10.1016/j.envsoft.2011.08.010, 2011. 13755, 13757

Pappenberger, F. and Beven, K. J.: Ignorance is bliss: or seven reasons not to use uncertainty analysis, Water Resour. Res., 42, W05302, doi:10.1029/2005WR004820, 2006. 13747

Pappenberger, F., Beven, K. J., Ratto, M., and Matgen, P.: Multi-method global sensitivity analysis of flood inundation models, Adv. Water Resour., 31, 1–14, doi:10.1016/j.advwatres.2007.04.009, 2008. 13766

Phillips, D. and Marks, D.: Spatial uncertainty analysis: propagation of interpolation errors in spatially distributed models, Ecol. Model., 91, 213–229, doi:10.1016/0304-3800(95)00191-3, 1996. 13753, 13783

Rakovec, O., Hill, M. C., Clark, M. P., Weerts, A. H., Teuling, A. J., and Uijlenhoet, R.: Distributed Evaluation of Local Sensitivity Analysis (DELSA), with application to hydrologic models, Water Resour. Res., 50, 409–426, doi:10.1002/2013WR014063, 2014. 13754, 13755, 13766

Raleigh, M. S.: Quantification of uncertainties in snow accumulation, snowmelt, and snow disapperance dates, Ph.D. thesis, University of Washington, USA, 2013. 13747, 13749, 13750

Raleigh, M. S. and Lundquist, J. D.: Comparing and combining SWE estimates from the SNOW-17 model using PRISM and SWE reconstruction, Water Resour. Res., 48, 1–16, doi:10.1029/2011WR010542, 2012. 13748, 13752

Rasmussen, R., Liu, C., Ikeda, K., Gochis, D., Yates, D., Chen, F., Tewari, M., Barlage, M., Dudhia, J., Yu, W., Miller, K., Arsenault, K., Grubišić, V., Thompson, G., and Gutmann, E.: High-resolution coupled climate runoff simulations of seasonal snowfall over Colorado: a process study of current and warmer climate, J. Climate, 24, 3015–3048, doi:10.1175/2010JCLI3985.1, 2011. 13766

Rasmussen, R., Baker, B., Kochendorfer, J., Meyers, T., Landolt, S., Fischer, A. P., Black, J., Thériault, J. M., Kucera, P., Gochis, D., Smith, C., Nitu, R., Hall, M., Ikeda, K., and Gutmann, E.: How well are we measuring snow: the NOAA/FAA/NCAR winter precipitation test

HESSD
11, 13745–13795, 2014

bed, B. Am. Meteorol. Soc., 93, 811–829, doi:10.1175/BAMS-D-11-00052.1, 2012. 13747, 13753, 13762, 13783

Rasmussen, R., Ikeda, K., Liu, C., Gochis, D., Clark, M., Dai, A., Gutmann, E., Dudhia, J., Chen, F., Barlage, M., Yates, D., and Zhang, G.: Climate change impacts on the water balance of the Colorado headwaters: high-resolution regional climate model simulations, J. Hydrometeorol., 15, 1091–1116, doi:10.1175/JHM-D-13-0118.1, 2014. 13766

Reba, M. L., Marks, D., Seyfried, M., Winstral, A., Kumar, M., and Flerchinger, G.: A long-term data set for hydrologic modeling in a snow-dominated mountain catchment, Water Resour. Res., 47, W07702, doi:10.1029/2010WR010030, 2011. 13749

Refsgaard, J. C., van der Sluijs, J. P., Brown, J., and van der Keur, P.: A framework for dealing with uncertainty due to model structure error, Adv. Water Resour., 29, 1586–1597, doi:10.1016/j.advwatres.2005.11.013, 2006. 13747

Rosero, E., Yang, Z.-L., Wagener, T., Gulden, L. E., Yatheendradas, S., and Niu, G.-Y.: Quantifying parameter sensitivity, interaction, and transferability in hydrologically enhanced versions of the Noah land surface model over transition zones during the warm season, J. Geophys. Res., 115, D03106, doi:10.1029/2009JD012035, 2010. 13755

Rosolem, R., Gupta, H. V., Shuttleworth, W. J., Zeng, X., and de Gonçalves, L. G. G.: A fully multiple-criteria implementation of the Sobol' method for parameter sensitivity analysis, J. Geophys. Res.-Atmos., 117, D07103, doi:10.1029/2011JD016355, 2012. 13755

Saltelli, A.: Sensitivity analysis: Could better methods be used?, J. Geophys. Res., 104, 3789, doi:10.1029/1998JD100042, 1999. 13748, 13755

Saltelli, A. and Annoni, P.: How to avoid a perfunctory sensitivity analysis, Environ. Modell. Softw., 25, 1508–1517, doi:10.1016/j.envsoft.2010.04.012, 2010. 13748, 13755, 13756, 13757

Schmucki, E., Marty, C., Fierz, C., and Lehning, M.: Evaluation of modelled snow depth and snow water equivalent at three contrasting sites in Switzerland using SNOWPACK simulations driven by different meteorological data input, Cold Reg. Sci. Technol., 99, 27–37, doi:10.1016/j.coldregions.2013.12.004, 2014. 13747, 13750, 13751, 13765

Serreze, M. C., Clark, M. P., Armstrong, R. L., McGinnis, D. A., and Pulwarty, R. S.: Characteristics of the western United States snowpack from snowpack telemetry (SNOTEL) data, Water Resour. Res., 35, 2145–2160, doi:10.1029/1999WR900090, 1999. 13754

HESSD
11, 13745–13795, 2014

Shamir, E. and Georgakakos, K. P.: Distributed snow accumulation and ablation modeling in the American River basin, Adv. Water Resour., 29, 558–570, doi:10.1016/j.advwatres.2005.06.010, 2006. 13748
Sieck, L. C., Burges, S. J., and Steiner, M.: Challenges in obtaining reliable measurements of point rainfall, Water Resour. Res., 43, W01420, doi:10.1029/2005WR004519, 2007. 13762
Slater, A. G. and Clark, M. P.: Snow Data Assimilation via an Ensemble Kalman Filter, J. Hydrometeorol., 7, 478–493, doi:10.1175/JHM505.1, 2006. 13748
Slater, A. G., Schlosser, C. A., Desborough, C. E., Pitman, A. J., Henderson-Sellers, A., Robock, A., Vinnikov, K. Y., Entin, J., Mitchell, K., Chen, F., Boone, A., Etchevers, P., Habets, F., Noilhan, J., Braden, H., Cox, P. M., de Rosnay, P., Dickinson, R. E., Yang, Z.-L., Dai, Y.-J., Zeng, Q., Duan, Q., Koren, V., Schaake, S., Gedney, N., Gusev, Y. M., Nasonova, O. N., Kim, J., Kowalczyk, E. A., Shmakin, A. B., Smirnova, T. G., Verseghy, D., Wetzel, P., and Xue, Y.: The representation of snow in land surface schemes: results from PILPS 2(d), J. Hydrometeorol., 2, 7–25, doi:10.1175/1525-7541(2001)002<0007:TROSIL>2.0.CO;2, 2001. 13747
Smith, P. J., Beven, K. J., and Tawn, J. A.: Detection of structural inadequacy in process-based hydrological models: a particle-filtering approach, Water Resour. Res., 44, W01410, doi:10.1029/2006WR005205, 2008. 13747
Sobol', I.: On sensitivity estimation for nonlinear mathematcal models, Matematicheskoe Modelirovanie, 2, 112–118, 1990. 13748, 13754
Sturm, M. and Wagner, A. M.: Using repeated patterns in snow distribution modeling: an Arctic example, Water Resour. Res., 46, 1–15, doi:10.1029/2010WR009434, 2010. 13749
Sturm, M., Holmgren, J., and Liston, G. E.: A seasonal snow cover classification system for local to global applications, J. Climate, 8, 1261–1283, doi:10.1175/1520-0442(1995)008<1261:ASSCCS>2.0.CO;2, 1995. 13749
Tang, Y., Reed, P., Wagener, T., and van Werkhoven, K.: Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation, Hydrol. Earth Syst. Sci., 11, 793–817, doi:10.5194/hess-11-793-2007, 2007. 13754, 13755
Tarboton, D. and Luce, C.: Utah Energy Balance snow accumulation and melt model (UEB), in: Computer Model Technical Description Users Guide, Utah Water Res. Lab., and USDA For. Serv. Intermt. Res. Station, p. 64, Logan, UT, 1996. 13749, 13750, 13766
Thornton, P. E., Hasenauer, H., and White, M. A.: Simultaneous estimation of daily solar radiation and humidity from observed temperature and precipitation: an application over complex

HESSD
11, 13745–13795, 2014

terrain in Austria, Agr. Forest Meteorol., 104, 255–271, doi:10.1016/S0168-1923(00)00170-2, 2000. 13783

Trujillo, E. and Molotch, N. P.: Snowpack regimes of the Western United States, Water Resour. Res., 50, 5611–5623, doi:10.1002/2013WR014753, 2014. 13749, 13751, 13765

van Werkhoven, K., Wagener, T., Reed, P., and Tang, Y.: Characterization of watershed model behavior across a hydroclimatic gradient, Water Resour. Res., 44, W01429, doi:10.1029/2007WR006271, 2008. 13755

Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W., and Sorooshian, S.: Effective and efficient algorithm for multiobjective optimization of hydrologic models, Water Resour. Res., 39, 1214, doi:10.1029/2002WR001746, 2003a. 13747

Vrugt, J. A., Gupta, H. V., Bouten, W., and Sorooshian, S.: A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, Water Resour. Res., 39, 1201, doi:10.1029/2002WR001642, 2003b. 13747

Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W., and Verstraten, J. M.: Improved treatment of uncertainty in hydrologic modeling: combining the strengths of global optimization and data assimilation, Water Resour. Res., 41, W01017, doi:10.1029/2004WR003059, 2005. 13747

Vrugt, J. A., Braak, C. J. F., Gupta, H. V., and Robinson, B. A.: Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?, Stoch. Env. Res. Risk A., 23, 1011–1026, doi:10.1007/s00477-008-0274-y, 2008a. 13754

Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M., and Robinson, B. A.: Treatment of input uncertainty in hydrologic modeling: doing hydrology backward with Markov chain Monte Carlo simulation, Water Resour. Res., 44, doi:10.1029/2007WR006720, 2008b. 13747, 13754, 13755

Wayand, N. E., Hamlet, A. F., Hughes, M., Feld, S. I., and Lundquist, J. D.: Intercomparison of meteorological forcing data from empirical and mesoscale model sources in the N.F. American River Basin in northern Sierra Nevada, California, J. Hydrometeorol., 14, 677–699, doi:10.1175/JHM-D-12-0102.1, 2013. 13748

Winstral, A. and Marks, D.: Simulating wind fields and snow redistribution using terrain-based parameters to model snow accumulation and melt over a semi-arid mountain catchment, Hydrol. Process., 16, 3585–3603, doi:10.1002/hyp.1238, 2002. 13783

Winstral, A., Marks, D., and Gurney, R.: An efficient method for distributing wind speeds over heterogeneous terrain, Hydrol. Process., 23, 2526–2535, doi:10.1002/hyp.7141, 2009. 13753, 13783

HESSD
11, 13745–13795, 2014

Winstral, A., Marks, D., and Gurney, R.: Simulating wind-affected snow accumulations at catchment to basin scales, Adv. Water Resour., 55, 64–79, doi:10.1016/j.advwatres.2012.08.011, 2013. 13765

Xia, Y., Yang, Z.-L., Stoffa, P. L., and Sen, M. K.: Using different hydrological variables to assess the impacts of atmospheric forcing errors on optimization and uncertainty analysis of the CHASM surface model at a cold catchment, J. Geophys. Res., 110, D01101, doi:10.1029/2004JD005130, 2005. 13748

Yang, D., Kane, D. L., Hinzman, L. D., Goodison, B. E., Metcalfe, J. R., Louie, P. Y. T., Leavesley, G. H., Emerson, D. G., and Hanson, C. L.: An evaluation of the Wyoming Gauge System for snowfall measurement, Water Resour. Res., 36, 2665–2677, doi:10.1029/2000WR900158, 2000. 13750

Yilmaz, K. K., Gupta, H. V., and Wagener, T.: A process-based diagnostic approach to model evaluation: application to the NWS distributed hydrologic model, Water Resour. Res., 44, W09417, doi:10.1029/2007WR006716, 2008. 13747

You, J., Tarboton, D. G., and Luce, C. H.: Modeling the snow surface temperature with a one-layer energy balance snowmelt model, Hydrol. Earth Syst. Sci. Discuss., 10, 15071–15118, doi:10.5194/hessd-10-15071-2013, 2013. 13750

Zuzel, J. F. and Cox, L. M.: Relative importance of meteorological variables in snowmelt, Water Resour. Res., 11, 174–176, doi:10.1029/WR011i001p00174, 1975. 13765

HESSD
11, 13745–13795, 2014

Table 1. Basic characteristics of the snow study sites, ordered by increasing elevation.

Site Name	Site ID	Location	Elevation (m)	Study Period (Water Year)	Snow Climate	Sensors
Imnavait Creek	IC	N 68.62 W 149.30 Alaska, USA	930	2011	Tundra	T_{air}: Vaisala HMP45C P: Campbell Scientific TE 525 U: Met One 014A RH: Vaisala HMP45C Q_{si}: Kipp & Zonen CMA 6 Q_{li}: n/a*
Col de Porte	CDP	N 45.30 E 5.77 Rhône-Alpes, FR	1330	2006	Mountain (maritime)	T_{air}: PT 100/4 wires P: PG2000, GEONOR U: Chauvin Arnoux Tavid 87 RH: Vaisala HMP 45D Q_{si}: Kipp & Zonen CM14 Q_{li}: Eppley PIR
Reynolds Mountain East (sheltered site)	RME	N 43.07 W116.75 Idaho, USA	2060	2007	Mountain (intermountain)	T_{air}: Vaisala HMP 45 P: Belfort Universal Gages U: Met One 013/023 RH: Vaisala HMP 45 Q_{si}: Eppley Precision Pyranometer Q_{li}: Eppley PIR
Swamp Angel Study Plot	SASP	N 37.91 W 107.71 Colorado, USA	3370	2008	Mountain (continental)	T_{air}: Vaisala CS500 P: ETI Noah II U: RM Young Wind Monitor 05103-5 RH: Vaisala CS500 Q_{si}: Kipp & Zonen CM21 Q_{li}: Kipp & Zonen CG-4

* At IC, Q_{li} was taken as $Q_{li} = Q_{net} - (Q_{si} - Q_{so}) + (5.67 \times 10^{-8})\, T_{surf}^{4}$, where Q_{net} is measured net radiation ($W m^{-2}$), Q_{si} is measured incoming shortwave radiation ($W m^{-2}$), Q_{so} is measured reflected shortwave radiation ($W m^{-2}$), and T_{surf} is measured snow surface temperature (K).

HESSD
11, 13745–13795, 2014

Table 2. UEB model parameters used in all simulations and sites.

Description of parameter	Units	Value
Rain threshold temperature	°C	3
Snow threshold temperature	°C	−1
Snow emissivity	–	0.99
Bulk snow density	kg m^{-3}	300
Liquid water holding capacity	fraction	0.05
Snow saturated hydraulic conductivity	m hr^{-1}	20
Visual new snow albedo	–	0.85
Near infrared new snow albedo	–	0.65
New snow threshold depth to reset albedo	m	0.01
Snow surface roughness	m	0.005
Forest canopy fraction	fraction	0
Ground heat flux	W m^{-2}	0

HESSD
11, 13745–13795, 2014

Table 3. Details of error types, distributions, and uncertainty ranges for the four scenarios. Bold face in the error type, distribution, and uncertainty range indicates defining characteristics, relative to scenario NB.

Forcing	Error Type[a]	Distribution[b]	Range	Units	Citations and Notes
			Scenario NB (k = 6, N = 10 000)		
T_{air}	B	Normal	[−3.0, +3.0]	°C	Bolstad et al. (1998); Chuanyan et al. (2005) Fridley (2009); Hasenauer et al. (2003)
P	B	Lognormal	[−75, +300]	%	Goodison et al. (1998); Luce et al. (1998) Rasmussen et al. (2012); Winstral and Marks (2002)
U	B	Normal	[−3.0, +3.0]	m s^{-1}	Winstral et al. (2009)
RH	B	Normal	[−25, +25]	%	Bohn et al. (2013); Déry and Stieglitz (2002) Feld et al. (2013)
Q_{si}	B	Normal	[−100, +100]	W m^{-2}	Bohn et al. (2013); Jepsen et al. (2012) Jing and Cess (1998); Niemelä et al. (2001b)
Q_{li}	B	Normal	[−25, +25]	W m^{-2}	Bohn et al. (2013); Flerchinger et al. (2009) Herrero and Polo (2012); Niemelä et al. (2001a)
			Scenario NB + RE (k = 12, N = 10 000) This scenario has six bias parameters (identical to NB above), plus the following six random error parameters		
T_{air}	**RE**	Normal	|[0.0, 7.5]|	°C	Chuanyan et al. (2005); Fridley (2009) Hasenauer et al. (2003); Huwald et al. (2009) Phillips and Marks (1996)
P	**RE**	Lognormal	|[0.0, 25]|	%	Guan et al. (2005); Hasenauer et al. (2003) Hutchinson et al. (2009)
U	**RE**	Normal	|[0.0, 5]|	m s^{-1}	Cheng and Georgakakos (2011); Liston and Elder (2006); Luo et al. (2008); Winstral et al. (2009)
RH	**RE**	Normal	|[0.0, 15]|	%	Bohn et al. (2013); Liston and Elder (2006) Phillips and Marks (1996)
Q_{si}	**RE**	Normal	|[0.0, 160]|	W m^{-2}	Hasenauer et al. (2003); Jepsen et al. (2012) Liston and Elder (2006); Thornton et al. (2000)
Q_{li}	**RE**	Normal	|[0.0, 80]|	W m^{-2}	Bohn et al. (2013); Flerchinger et al. (2009) Liston and Elder (2006)
			Scenario UB (k = 6, N = 10 000) This scenario is identical to NB, except all probability distributions are **uniform**		
			Scenario NB_lab[c] (k = 6, N = 10 000)		
T_{air}	B	Normal	**[−0.30, +0.30]**	°C	Vaisala HMP45 specified accuracy
P	B	Lognormal	**[−3.0, +3.0]** [d]	%	RM Young 52202 specified accuracy
U	B	Normal	**[−0.30, +0.30]**	m s^{-1}	RM Young 05103 specified accuracy
RH	B	Normal	**[−3.0, +3.0]**	%	Vaisala HMP45 specified accuracy
Q_{si}	B	Normal	**[−25, +25]**	W m^{-2}	Li-Cor 200X specified accuracy of 5 %
Q_{li}	B	Normal	**[−15, +15]**	W m^{-2}	Assumed 5 % of mean intersite values

[a] B = bias, RE = random errors. Biases are additive (b_i = 0, Eq. 4) for all forcings except P, which has multiplicative bias (b_i = 1).

[b] Probability distributions were truncated in instances when introduction of errors caused non-physical forcing values (see Sect. 3.3.5).

[c] Uncertainty ranges in this scenario are based primarily on manufacturer's specified accuracy for typical sensors deployed at SNOTEL sites (NRCS Staff, personal communication, 2013). We assume the P storage gauge has the same accuracy as a typical tipping bucket gauge.

[d] We neglect P undercatch errors in the lab uncertainty scenario.

HESSD
11, 13745–13795, 2014

Table 4. Number of samples (model simulations) meeting the requirements for minimum peak SWE and snow duration and valid snow disappearance dates at each site in each scenario.

	IC	CDP	RME	SASP
NB	9898 (79 184)	9792 (78 336)	8799 (70 392)	9984 (79 872)
NB + RE	9943 (139 202)	9734 (136 276)	8648 (121 072)	9985 (139 790)
UB	8608 (68 864)	8925 (71 400)	9102 (72 816)	3399 (27 192)
NB_lab	10 000 (80 000)	10 000 (80 000)	10 000 (80 000)	10 000 (80 000)

HESSD
11, 13745–13795, 2014

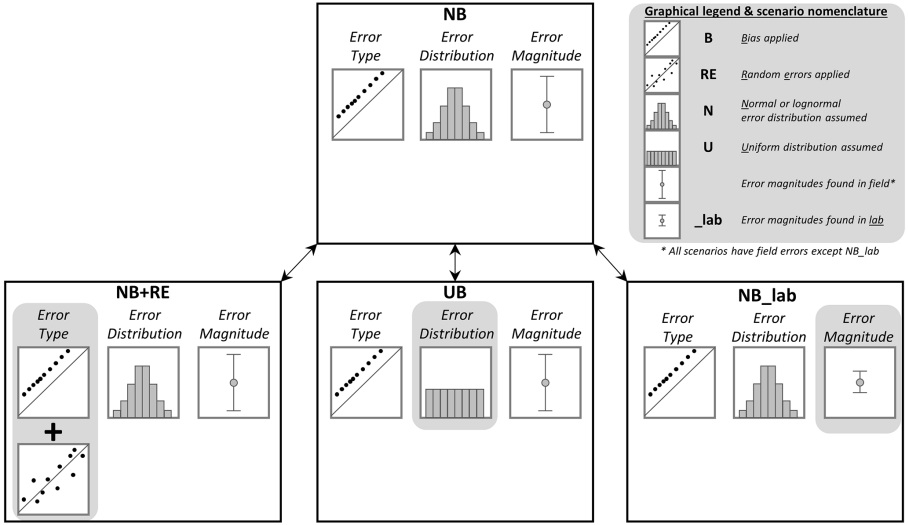

Figure 1. Scenarios of interest and the type, distribution, and magnitude of errors considered in each. NB considers normally (or lognormally) distributed biases with error magnitudes found in the field. NB + RE is the same as NB but also considers random errors. UB is the same as NB but considers uniformly distributed errors instead. NB_lab is the same as NB but considers laboratory error magnitudes.

HESSD
11, 13745–13795, 2014

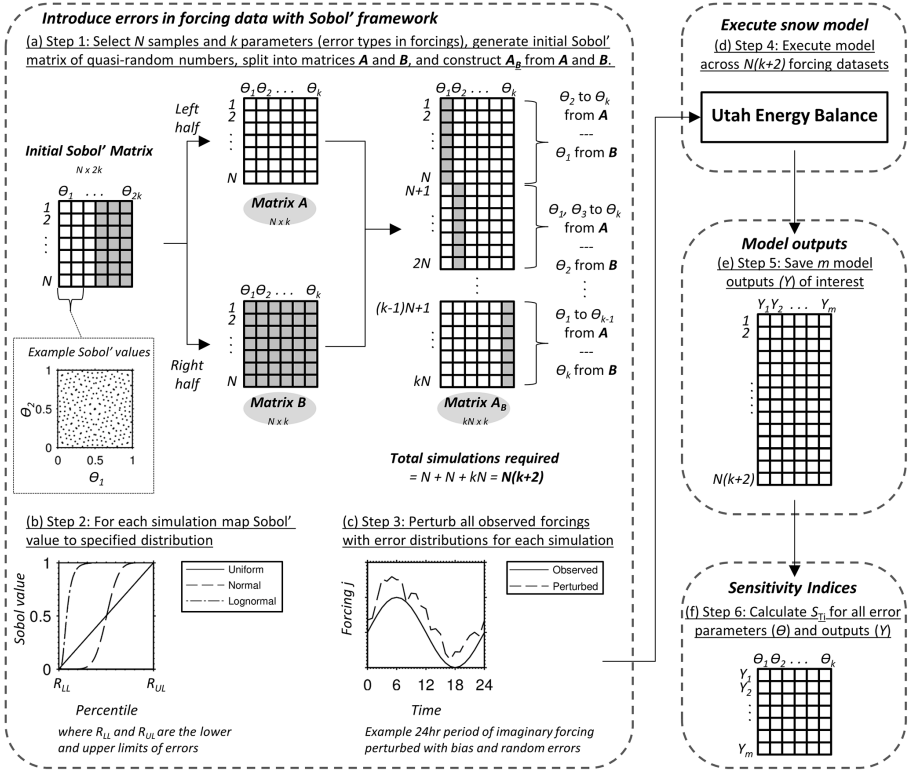

Figure 2. Conceptual diagram showing methodology for imposing errors on the forcings with error parameters (θ) within the Sobol' sensitivity analysis framework, and workflow for model execution and calculation of sensitivity indices on model outputs (Y).

HESSD
11, 13745–13795, 2014

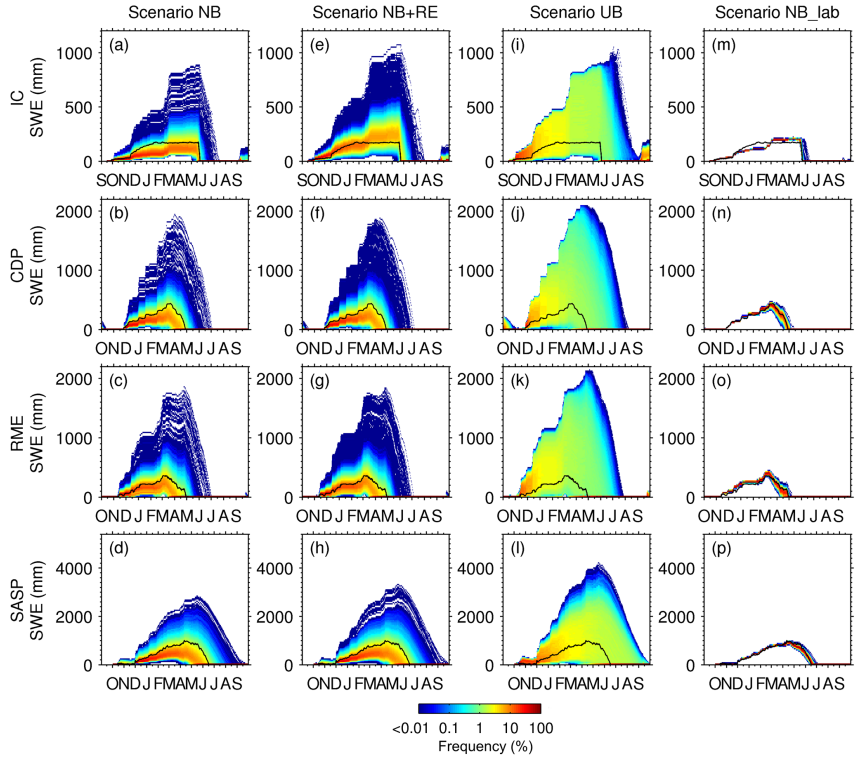

Figure 3. Observed (black line) and modeled SWE (color density plot) at the four sites across the four uncertainty scenarios (see Fig. 1 and Table 3). The number of model simulations in the density plots varies with the site and scenario (see Table 4). The density plots were constructed using 100 bins in the SWE dimension with relative frequency tabulated in each bin each day. Note the frequency colorbar is on a logarithmic scale. Sites are arranged from top to bottom in order of increasing elevation and decreasing latitude. Scenarios are defined as normally distributed bias (NB), normally distributed bias and random errors (NB + RE), uniformly distributed bias (UB), and normally distributed bias at laboratory error magnitudes (NB_lab).

HESSD
11, 13745–13795, 2014

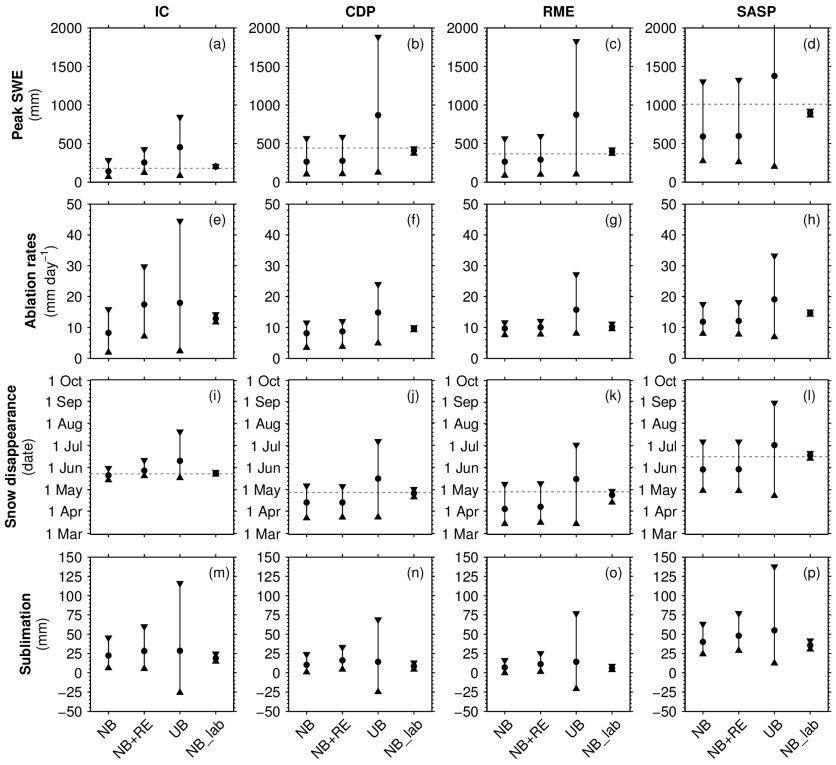

Figure 4. Distributions of model outputs (rows) at the four study sites (columns) arranged by scenario. For each scenario, the circle is the mean and the whiskers show the range encompassing 95 % of the simulations (see Table 4 for number of simulations for each site and scenario). The dashed lines in **(a–d)** and **(i–l)** are the observed values. Axes are matched between sites for a given model output; note that the range in scenario UB in **(d)** is truncated by the axes limits (upper value = 3030 mm). Scenarios are defined as normally distributed bias (NB), normally distributed bias and random errors (NB + RE), uniformly distributed bias (UB), and normally distributed bias at laboratory error magnitudes (NB_lab).

HESSD
11, 13745–13795, 2014

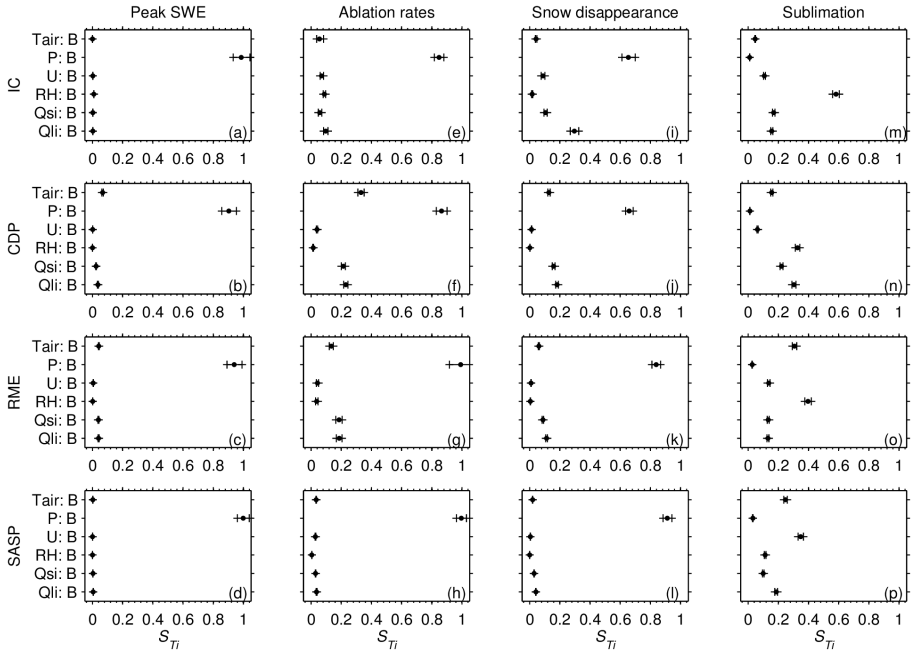

Figure 5. Total-order sensitivity indices (S_{Ti}) of four model response variables at the four sites in the NB scenario. Shown are the mean (bootstrapped) sensitivity indices and associated 95 % confidence intervals. The "B" after the forcing indicates bias.

HESSD
11, 13745–13795, 2014

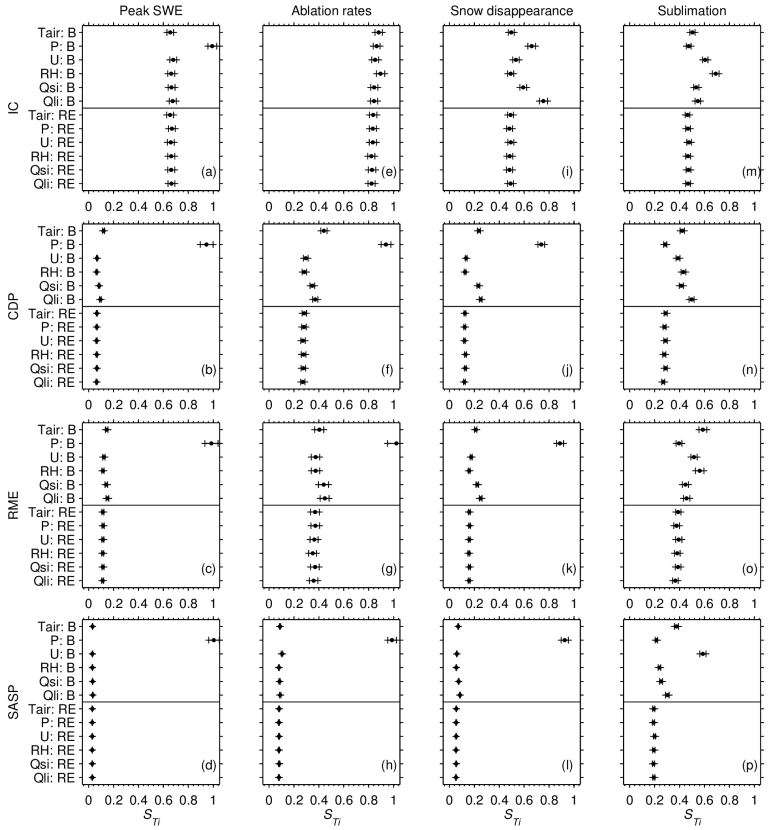

Figure 6. Same as Fig. 5, except in the NB + RE scenario. "B" indicates bias and "RE" indicates random error in the forcing. The horizontal line separates the biases (above the line) and random errors (below the line).

HESSD
11, 13745–13795, 2014

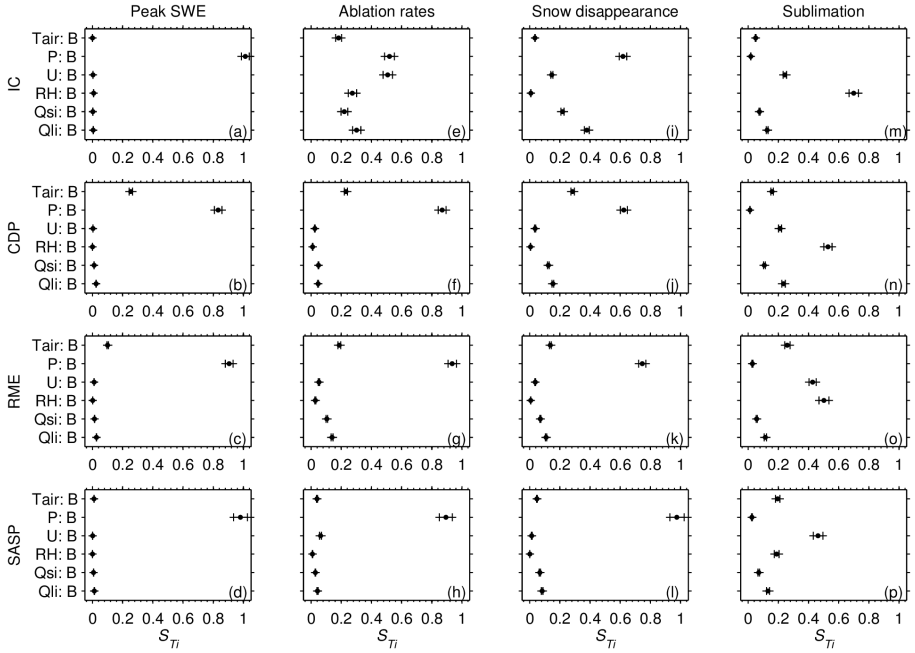

Figure 7. Same as Fig. 5, except in the UB scenario.

HESSD
11, 13745–13795, 2014

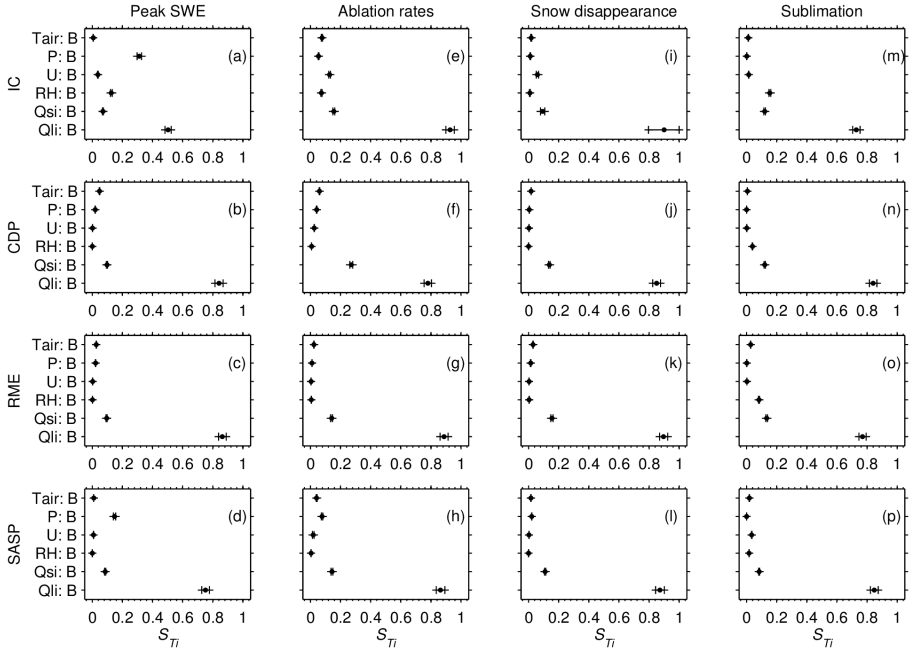

Figure 8. Same as Fig. 5, except in the NB_lab scenario.

HESSD
11, 13745–13795, 2014

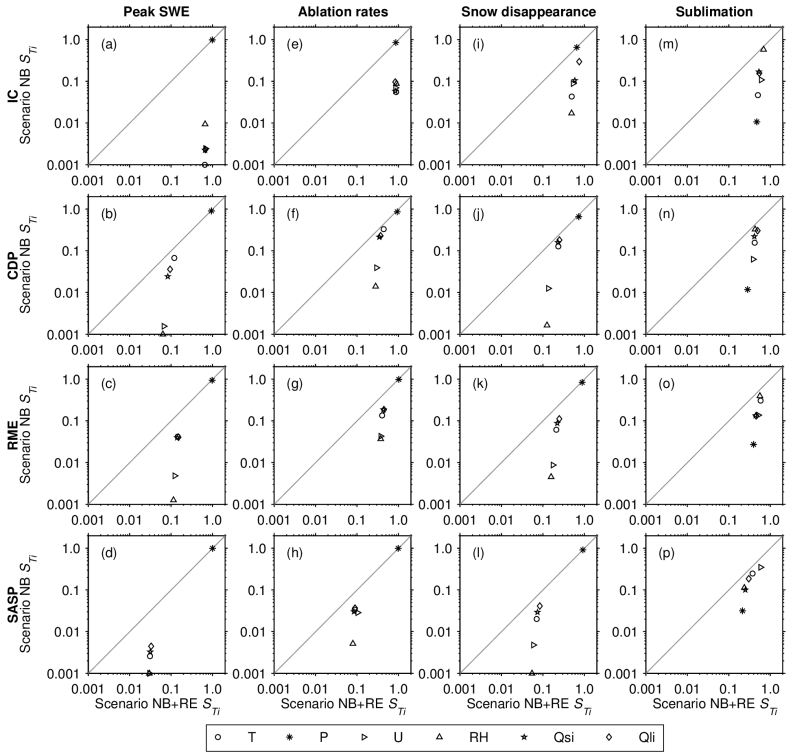

Figure 9. Test of model sensitivity as a function of forcing error type. Shown are the mean S_{Ti} values from scenario NB + RE (from Fig. 6) vs. scenario NB (from Fig. 5) for four model outputs at the four sites. Only bias parameters are shown (NB + RE yields different sensitivity due to interactions between bias and random errors). Forcing abbreviations are defined in the text. Note the plots are on log-log scale. S_{Ti} values less than 0.001 were set to 0.001 to display on the log-log plot. NB + RE considers normally distributed bias and random errors, while NB considers normally distributed bias only.

HESSD
11, 13745–13795, 2014

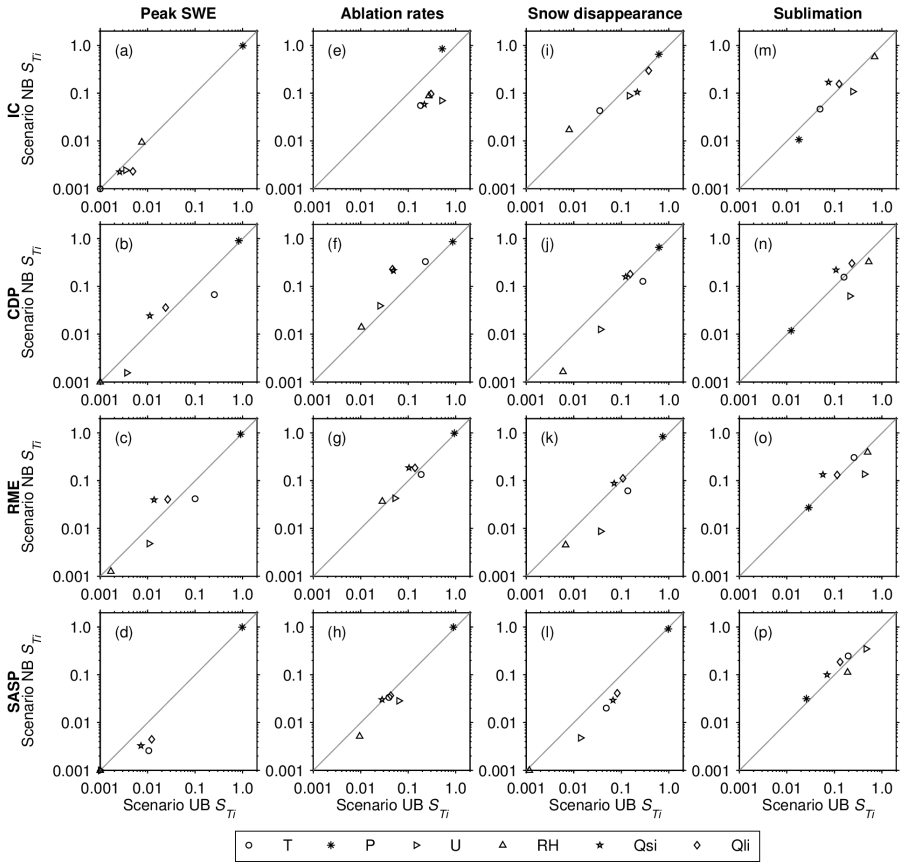

Figure 10. Same as Fig. 9, but comparing S_{Ti} values from scenarios UB (from Fig. 7) and NB (from Fig. 5) to test model sensitivity as a function of error distribution. UB considers uniformly distributed bias, while NB considers normally distributed bias.

HESSD
11, 13745–13795, 2014

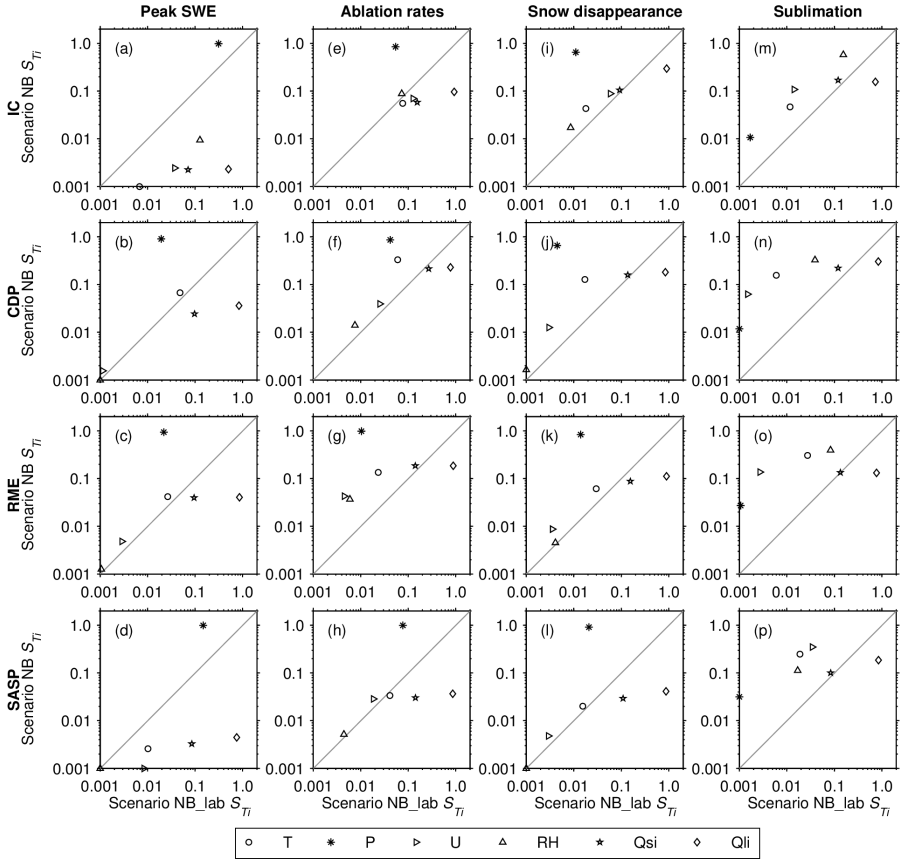

Figure 11. Same as Fig. 9, but comparing S_{Ti} values from scenarios NB_lab (from Fig. 8) and NB (from Fig. 5) to test model sensitivity as a function of error magnitudes. NB_lab considers normally distributed bias at error magnitudes found in the laboratory, while NB considers normally distributed bias at error magnitudes found in the field.

HESSD
11, 13745–13795, 2014