June 17, 2015

Dr. Ross Woods
Editor
Hydrology and Earth System Sciences

Dear Dr. Woods,

We have completed the minor revisions to our *Hydrology and Earth System Sciences* manuscript "Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework" (hess-2014-461) and are now resubmitting it.

In the pages below, we include our point-by-point response to the reviewer's comments and a marked-up version of the manuscript to show all modifications made through the course of the review process. We appreciate your time and the reviewers' time in the review process and look forward to hearing your decision. Thank you again for considering our manuscript for publication.

Sincerely,

Mark S. Raleigh

Mark S. Raleigh, PhD
Email: raleigh@ucar.edu
Phone: +1 (303) 497-8381

Response to second review by F. Pianosi (Referee)

Note: reviewer comments are in italics and the authors' responses are in normal face.

Comment: *The paper has been substantially revised after the first round of reviews and I think it is significantly improved. In particular, with respect to the comments raised in my review, the discussion of GSA results and their interpretation is much more solid and convincing.*

The authors found some inadequacies in their code implementing the error models and after fixing them they were able to produce GSA results that look much more reliable. They also enriched their discussion of GSA results with more insights on their physical meaning and implications, and more interpretation of ``surprising'' or unexpected results.

In my opinion this revised manuscript is suitable for publication and could constitute an interesting contribution to the readership of HESS, and a nice demonstration of how GSA methods can be used to learn about the behavior of a hydrological model and its sources of uncertainty.

I have few minor suggestions (mainly again relating to methodological aspects and references of GSA), that the authors may consider in a final revision of their manuscript.

Response: We thank you for your time and thoughtful comments in this second review.

Comment: *Lines 376-378: ``Sensitivity analyses tend to focus more on the model parameter ... than on the forcing matrix''. True, however recently there have been an increasing number of applications considering other input factors than parameters, for instance*

Baroni, G., Tarantola, S., 2014. A general probabilistic framework for uncertainty and global sensitivity analysis of deterministic models: A hydrological case study. Environmental Modelling \& Software 51, 26–34.

and references therein. It would be good to cite some of these works, also to stress that GSA does not need to focus on parameters only.

Response: Thank you for making us aware of this work. We have added a sentence here stating that other sensitivity analyses have considered other sources of model uncertainty (with citations).

Comment: *Lines 387-389: ``Sobol' sensitivity analysis uses variance decomposition to attribute output variance to input variance''. I would say ``to input variability'' or ``to input uncertainty'' (see similar comment in my previous review)*

Response: We changed the wording to "input uncertainty".

Comment: *Lines 396-399: ``A key assumption to the Sobol' approach is that the factors are independent'': I know the authors introduced this sentence to address one of my previous comments, however the sentence is not technically completely correct (I was a bit sloppy about this in my review too). There are some recent papers that show how to apply the Sobol' approach in the case of correlated/dependent inputs (Lilburne &Tarantola, 2009. Sensitivity analysis of spatial models.; Kucherenko et al, 2012. Estimation of global sensitivity indices for models with dependent variables; Castaing and Gratiet, 2015, ANOVA decomposition of conditional Gaussian processes for sensitivity analysis with dependent inputs), however they require a sampling and approximation strategy different from the one implemented in this paper and described in Sec. 3.3.3. So maybe better rephrase the sentence like:*
``A key assumption to the implementation of the Sobol' approach used in this paper (see Sec. 3.3.3) is that the factors are independent''

Response: We agree that this is important to distinguish. We have changed the wording based on your recommendation. We also add a new sentence stating "Frameworks have been proposed for the case of correlated factors (e.g., forcing errors) in a sensitivity analysis (e.g., Kucherenko et al., 2012), but we leave those applications for future work."

Comment: *Line 436: ``Final S_{Ti} values'': what does ``final'' mean here?*

Response: By "final" we mean the S_Ti values computed using all samples. We have changed the wording to reflect this convention.

Comment: *Line 442 and following (and Figure 2): I think the term ``Sobol' values'' is a bit confusing, one could think it refers to values of the sensitivity indices rather than input samples, or suggest that there is a link between the quasi-random sampling strategy and the variance decomposition idea, while there is not (apart from the fact that Sobol' contributed to both). Maybe better use ``input factor samples'' or ``parameter samples''?*

Response: Thank you for pointing out this potentially confusing issue. We have renamed "Sobol' values" as "input factor samples" in three cases in Section 3.3.5 and in Figure 2.

Comment: *Lines 679-680: ``we calculated daily sensitivity indices of modeled SWE to forcing biases at each site and scenario (Fig. 8).'' I understand that here the output variable whose sensitivity is being assessed is the simulated SWE on given day (and sensitivity index is re-estimated at each day in the simulation horizon). If so, this is conceptually quite different from the output definitions used so far (aggregate indicators from the entire time series of simulated variables). Maybe good to add one sentence to further clarify the point? Readers not particularly familiar with GSA may struggle here.*

Response: We added sentences to help clarify how this analysis is different: "This final analysis was conceptually different than the previous analyses (Fig 5-7) in terms of the model output considered. Whereas the previous analyses computed sensitivity indices for summative model outputs (e.g., peak SWE, total sublimation), the final analysis re-calculated sensitivity indices for SWE each day. This approach allowed us to examine how SWE model sensitivity changed as a function of time within the snow season."

Comment: *Table 4: I understand the number outside parenthesis is the number of ``behavioural'' samples in the initial Sobol' matrix and the one in parenthesis is model simulations, the relation being the usual $N_{sim} = N_{samp}\times (k+2)$ where $k=8$ in your case study - but it took a while to figure out - maybe better just say it in the caption? Also, this seems to suggest that you checked output samples associated to the initial Sobol' matrix only, while I suppose you should also check the feasibility after resampling - in other words, check the final matrix A_B?*

Response: We can see how the relationship between number of samples and model simulations might not be obvious based on the table and caption, so clarification is warranted. The reported figures include all simulations from the initial matrices A, B, and A_B. Recall that A and B each require N simulations while A_B requires kN simulations, hence the total number of simulations is N+N+kN = Nx(k+2). If a sample is rejected in one of these the matrices, the corresponding simulations in all three matrices (sampling and resampling matrices) are also rejected. So in effect we are checking the simulations from the final matrix A_B, as well as the A and B matrices.

We have revised the caption to read: "Number of samples (N) and model simulations (in parentheses) meeting the requirements for minimum peak SWE and snow duration and valid snow disappearance dates at each site in each scenario. The number of model simulations scaled as Nx(k+2), where k=12 in scenario NB+RE and k=6 in all other scenarios. When a simulation was rejected, all related simulations (based on resampling) were also rejected."

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

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 and which error characteristics are most important. Here we employ global sensitivity analysis to explore how (1) different error types (i.e., bias, random errors), (2) different error probability distributions, and (3) 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 1 840 000 Monte Carlo simulations across four sites and five different scenarios. Model outputs were (1) consistently more sensitive to forcing biases than random errors, (2) generally less sensitive to forcing error distributions, and (3) critically sensitive to different forcings depending on the relative magnitude of errors. For typical error magnitudes found in areas with drifting snow, precipitation bias was the most important factor for snow water equivalent, ablation rates, and snow disappearance timing, but other forcings had a more dominant impact when precipitation uncertainty was due solely to gauge undercatch. 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; Clark et al., 2011b; Deems et al., 2013; Leavesley, 1994). Due to the complexity of processes represented, these models usually require numerous meteorological forcing inputs and 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; Clark and Slater, 2006; Flerchinger et al., 2009) and to develop methodologies for incorporating uncertainty into modeling efforts (e.g., He et al., 2011b; Kavetski et al., 2006a; Kuczera et al., 2010; Slater and Clark, 2006), 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., 2011b), 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, 2011b, 2015a, 2015b; Essery et al., 2013; Georgakakos et al., 2004; Jackson et al., 2003; Kelleher et al., 2015; Kuczera and Parent, 1998; Liu and Gupta, 2007; Refsgaard et al., 2006; Slater et al., 2001; Smith et al., 2008; Vrugt et al., 2003a, 2003b, 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 due to environmental or instrumental factors (Huwald et al., 2009; Lundquist et al., 2015; 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 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, 2011b; 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; Lapo et al., 2015; 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, 2006b). 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). Lapo et al. (2015) show that biases have a greater impact than random errors on modeled snow water equivalent and surface temperature, but this analysis only considers longwave and shortwave forcings and considers errors separately. 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). In contrast, 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 use global sensitivity analysis to assess how specific forcing error characteristics influence outputs of a physically based snow model. To our knowledge, no previously published 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 impact of forcing errors on models can be tested by corrupting forcings with specified characteristics (e.g., artificial biases and random errors) and quantifying the impact on model outputs (e.g., Oudin et al., 2006; Spank et al., 2013), but we are unaware of any detailed studies that have done this type of experiment for all meteorological forcings commonly required for physically based snow models. We hypothesize that (1) model outputs are more sensitive to biases than random errors in forcing variables, (2) the assumed probability distribution for biases will alter the relative ranking of importance in forcing errors, and (3) the magnitude of forcing biases will have a strong influence on which forcing errors are most important.

In our view, it is important to clarify the relative impact of specific error characteristics on modeling applications, so as to prioritize future research directions, improve understanding of model sensitivity, and to address questions related to network design. For example, given budget constraints, is it better to invest in a heating apparatus for a radiometer (to minimize bias due to frost formation on the radiometer dome) or in a higher quality radiometer (to minimize random errors associated with measurement precision)? Additionally, it is important to contextualize different meteorological data errors, as these errors are usually studied independently of each other (e.g., longwave radiation, Flerchinger et al., 2009; air temperature, Huwald et al., 2009), and it is unclear how they compare in terms of model sensitivity.

The overarching 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 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 region 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. We selected these sites because of the quality and completeness of the forcing data, and because they spanned contrasting climates (Table 1), allowing us to check for potential climate-dependencies in sensitivity to forcing errors. Generalization of the results with climate was not possible due to the low sample size of sites.

The sites had high-quality observations of 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 (Bret-Harte et al., 2010a, 2010b, 2011a, 2011b, 2011c; Griffin et al., 2010; 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 experiment. Measured evaluation data (e.g., snow water equivalent, SWE) at daily resolution were used only for qualitative assessment of model output. SWE was observed at snow pillows at IC and RME. At CDP, a cosmic ray detector collected SWE data. At SASP, acoustic snow depth data were converted to daily SWE using density inferred from nearby SNOw TELemetry (SNOTEL) (Serreze et al., 1999) sites and local snow pit measurements (Raleigh, 2013).

We adjusted the available precipitation data at each site with a multiplicative factor to correct for potential undercatch errors (e.g., Goodison et al., 1998; Rasmussen et al., 2012; Yang et al., 2000) and to ensure the base model simulation with all observed forcings reasonably represented observed SWE before conducting the sensitivity analysis. Several studies have demonstrated the necessity of precipitation adjustments for realistic SWE simulations, even at well-instrumented sites (e.g., Hiemstra et al., 2006; Reba et al., 2011; Schmucki et al., 2014). 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. We considered adjustment multipliers ranging from 0.5 to 2.5 (increments of 0.05) and selected the multiplier that yielded the lowest root mean squared error between observed and modeled SWE. Precipitation multipliers were 1.6 at IC and 1.15 at SASP, and 0.9 at CDP and RME. The undercatch errors at IC were consistent with the 61-68% undercatch errors found by Yang et al. (2000) for Wyoming-type gauges in wind-blown regions.

The initial discrepancies between modeled and observed SWE (prior to applying the above precipitation multipliers) may have resulted from deficiencies in the measured forcings, model parameters, model structure, and measured verification data, and justification of our decision to apply precipitation multipliers was warranted. Manual observations of SWE (e.g., snow surveys, snow pits) generally supported the automatically collected SWE observations (no figures shown), and thus differences between observed and modeled SWE did not likely stem from issues in the verification data. Sites where we decreased the precipitation data (CDP and RME) were also the warmer sites and experienced more mixed rain-snow events in the winter. Hence, we considered multiple hypotheses to explain the SWE differences at these sites: (1) the choice of rain-snow parameterization, (2) the choice of parameters (e.g., threshold temperatures) for the rain-snow parameterization, and (3) the quality of the forcing data (e.g., precipitation). For these warmer sites, an exploratory analysis revealed that either (1) or (3) could explain the SWE differences, but auxiliary data (e.g., precipitation phase data) were not available to discriminate these hypotheses. Choosing a different rain-snow parameterization might minimize the SWE differences at the warmer sites but would not rectify the SWE differences at the colder sites (IC and SASP) where most winter precipitation falls as snow. Therefore, the most straightforward and consistent approach was to adjust the precipitation data and to leave the native UEB parameterizations intact. 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. Nevertheless, we suggest these precipitation adjustments minimally affected the sensitivity analysis, as we did not quantitatively compare the model outputs to the observed response variables (e.g., SWE).

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 refreezing, and sublimation. Due to the one-dimensional structure of the model, UEB does not account for lateral mass transfer of snow (e.g., wind-induced snow drifting), and therefore these processes must be represented in other model components (e.g., precipitation uncertainty, see Sec. 3.2.3). 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 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. 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 examined five 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 distributed for additive errors or lognormally distributed for multiplicative precipitation errors) was introduced into all forcings at a level of high uncertainty (based on values observed in the field, see Sec. 3.2.3 below). This scenario was named “NB,” where N denotes normal (or lognormal) error distributions and B denotes bias only. The remaining scenarios were identical to NB except one aspect was changed: scenario NB+RE considered both bias and random errors (RE) in all forcings, scenario UB considered uniformly distributed biases in all forcings, scenario NB_gauge considered precipitation error magnitudes associated with gauge undercatch, and scenario NB_lab considered error magnitudes for all forcings 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 (high forcing uncertainty) to both NB_gauge (moderate uncertainty in precipitation but high uncertainty for all other forcings) and NB_lab (low forcing uncertainty).

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), whereas data assimilation methods often focus on random errors but assume bias does not exist (e.g., Dee, 2005). Rarely is there any consideration of interactions between these error types. As a recent example, Lapo et al. (2015) 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, NB_gauge, 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 five 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

In their recent review of global sensitivity analysis applications in hydrological modeling, Song et al. (2015) identified the selection of probability distributions (this section) and ranges (Sec. 3.2.3) as among the most important considerations. While it is common practice in sensitivity analysis to assume a uniform distribution when sampling model parameters (e.g., Campolongo et al., 2011; Rosero et al., 2010), this may fail to represent the real distribution of errors in meteorological forcing data, as the uniform distribution implies that extreme and small biases are equally probable. It is more likely that real error distributions more closely resemble non-uniform distributions, with higher probability of smaller biases and lower probability of more extreme biases (e.g., normal distributions). Investigators in other fields (e.g., Foscarini et al., 2010; Touhami et al., 2013) have tested how distribution assumptions (uniform vs. normal) change their computed measures of model sensitivity. These studies broadly suggest that the grouping of most important factors may be similar under different distribution assumptions, particularly in cases when interactions are minimal, but the relative ranking of factors within those groups may vary depending on the distribution. Here we test how the assumed probability distribution influences the sensitivity of a snow model to forcing errors.

We designed the UB scenario with the naive hypothesis that the probability distribution of biases was uniform for all six meteorological variables. In contrast, error distributions (Table 3) were assumed non-uniform (described below) in scenarios NB, NB+RE, NB_gauge, and NB_lab. Unfortunately, error distributions are reported less frequently than error statistics (e.g., bias, RMSE) in the literature. We assumed that T_{air} and *RH* errors follow normal distributions (Mardikis et al., 2005; Phillips and Marks, 1996), as do Q_{si} and Q_{li} errors. 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) or to account for precipitation gauge undercatch (Durand and Margulis, 2007). Error distributions were truncated in cases when the introduced errors violated physical limits (e.g., negative *U*; see Sec. 3.3.5).

3.2.3. Error magnitudes

We considered three magnitudes of forcing uncertainty (Table 3): levels of uncertainty found (1) in the field for all forcings (i.e., NB), (2) in the field for all forcings except precipitation (which has uncertainty due to precipitation gauge undercatch, i.e., NB_gauge), and (3) in a controlled laboratory setting (i.e., NB_lab). These cases were considered because they sampled realistic errors (NB and NB_gauge) and minimum errors (NB_lab). We expected that the error ranges exerted a major control on model uncertainty and sensitivity, as demonstrated in several prior sensitivity analyses (see review of Song et al., 2015).

Consideration of error magnitudes was achieved in each scenario by assigning a range to each error probability distribution (see Sec. 3.2.2 and Table 3). While non-uniform distributions (e.g., normal) are typically described by measures other than the range (e.g., mean and variance), we scaled these distributions (see Sec. 3.3.5 for details) such that they were bounded within a specified range. This convention was necessary to ensure that differences between scenarios NB and UB were due solely to the shape of the error probability distributions, and not due to differences in both distribution shape and the domain. Additionally, this followed the typical practice of sensitivity analysis where the range specifies the domain of the distribution.

We considered field uncertainties in all forcings in NB, NB+RE, and UB, and in all forcings except precipitation in NB_gauge. Field uncertainties depend on the source of forcing data and on local conditions (e.g., Flerchinger et al., 2009; Lundquist et al., 2015). To generalize the analysis, we chose error ranges for the field uncertainty 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). *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., 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).

P error ranges spanned both undercatch (e.g., Rasmussen et al., 2012) and wind drift/scour errors in NB, NB+RE, and UB, but only undercatch errors in NB_gauge. We assumed that *P* biases due to gauge undercatch in NB_gauge ranged from -10% to +10% because Meyer et al. (2012) found 95% of SNOTEL sites (often in forest clearings) had observations of accumulated *P* within 20% of peak SWE. Results of NB, NB+RE, and UB were thus most relevant to areas with prominent snow redistribution (e.g., alpine zone), whereas NB_gauge results were more relevant to areas with minimal wind drift errors. It could be argued that uncertainty due to snow drift processes is a structural issue and not a source of forcing error; however, this distinction depends strongly on what type of model is considered. This process is clearly a structural component for snow models with explicit (e.g., three dimensional models with dynamic wind transport, Lehning et al., 2006) or implicit (one dimensional models with probabilistic subgrid snow variability routines, e.g., Clark et al., 2011a) treatment of snow redistribution. However, when a one dimensional snow model is applied at length scales shorter than drift process length scales (as assumed here with UEB), then it is not possible to account for snow drift in a structural sense. Therefore, we treat drifting snow as a form of precipitation error in NB, NB+RE, and UB. Because UEB lacks dynamic wind redistribution, accumulation uncertainty was not linked to *U* errors but instead to *P* errors (e.g., drift factor, Luce et al., 1998).

In contrast, scenario NB_lab assumed laboratory levels of uncertainty (i.e., measurement accuracy) for each forcing. Skiles et al. (2012) considered a similar scenario in their sensitivity analysis of the SNOBAL model (Marks and Dozier, 1992; Marks et al., 1992) to instrument accuracy at SASP, finding a 5 day range in uncertainty in modeled snow disappearance, with longwave uncertainty having the greatest impact. An emerging sensitivity analysis (Sauter and Obleitner, 2015) with the CROCUS model (Brun et al., 1992) applied on the Kongsvegen Glacier (Svalbard) indicates that longwave measurement uncertainty has an approximately comparable effect on modeled snow depth as ±25% precipitation uncertainty, but is the most dominant influence on the modeled energy balance and turbulent heat flux (relative to the measurement uncertainty of other forcings). Here we build on these efforts to examine how instrument accuracy impacts modeled snow variables in a variety of seasonal snow climates. In general, laboratory uncertainty levels vary with the type and quality of sensors, as well as related accessories (e.g., radiation shield for the temperature sensor), which we did not explicitly consider. Because the actual sensors available varied between sites (Table 1) and we needed consistent errors across sites within scenario NB_lab, we assumed that the manufacturers' specified accuracy of meteorological sensors at a typical SNOTEL site 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 idealized *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 Sec. 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 optimization/calibration of hydrology and earth systems models (Beven and Binley, 1992; Duan et al., 1992; Kavetski et al., 2002, 2006a, 2006b; Kuczera et al., 2010; Razavi and Gupta, 2015; Song et al., 2015; Vrugt et al., 2008a, 2008b). 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).

3.3.1. Overview: model conceptualization and sensitivity

One can visualize any hydrology or snow model (e.g., UEB) as:

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

where $\boldsymbol{Y}$ is a matrix of model outputs (e.g., SWE), M( ) is the model operator, $\boldsymbol{F}$ is a matrix of forcings (e.g., T_{air}, P, U, etc.), and $\boldsymbol{\theta}$ is an array of model parameters (e.g., Table 2). The goal of sensitivity analysis is to determine which input factors ($\boldsymbol{F}$ and $\boldsymbol{\theta}$) are most important to specific outputs ($\boldsymbol{Y}$) (Matott et al., 2009). Sensitivity analyses often focus more on the model parameter array ($\boldsymbol{\theta}$) 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). However, recent analyses have considered other input factors and sources of uncertainty (e.g., Baroni and Tarantola, 2014; Schoups and Hopmans, 2006). Here, we extend the sensitivity analysis framework to forcing uncertainty by creating k new parameters (ϕ_1, ϕ_2, … ϕ_k) that specify forcing uncertainty characteristics (Vrugt et al., 2008b) and reformulate equation 1 as:

$$\mathbf{Y} = M(\mathbf{F},\boldsymbol{\theta},\phi) \tag{2}$$

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 uncertainty. First-order and higher-order sensitivities can be resolved; here, only the total-order sensitivities were examined (see below) for clarity and because the resulting first-order sensitivity indices were typically comparable to the total-order sensitivity indices (e.g., 83% of all cases had total-order and first-order indices within 10% of each other), suggesting minimal error interactions. 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. A key assumption to the Sobol' approach used in this paper (see Sec. 3.3.3) is that the factors are independent; hence, our analysis does not consider cases of correlated errors (e.g., a positive measurement bias in T_{air} that causes a negative *RH* bias). Frameworks have been proposed for the case of correlated factors (e.g., forcing errors) in a sensitivity analysis (e.g., Kucherenko et al., 2012), but we leave those applications for future work. Below, we provide a brief summary of the Sobol' sensitivity analysis methodology implemented here but note that further details can be found in Saltelli et al. (2010).

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 forcing error (ϕ_i), including both unique (i.e., first-order) effects and all interactions with all other parameters:

$$S_{Ti} = \frac{E[V(Y \mid \phi_{\sim i})]}{V(Y)} = 1 - \frac{V[(E(Y \mid \phi_{\sim i}))]}{V(Y)} \quad (3)$$

where *E* is the expectation (i.e., average) operator, *V* is the variance operator, and $\phi_{\sim i}$ signifies all parameters except ϕ_i. The latter expression defines S_{Ti} as the variance remaining in *Y* after accounting for variance due to all other parameters ($\phi_{\sim i}$). 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 ϕ_i on output Y_j indicates 70% of the output variance was due to bias in forcing *i* (including unique effects and interactions).

A number of numerical methods are available for evaluating sensitivity indices, and most adopt a Monte-Carlo approach (Saltelli et al., 2010). Evaluation of Eq. (3) requires two sampling matrices, which we 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). Selecting input factor samples for these two matrices was achieved using the quasi-random Sobol' sequence (Saltelli and Annoni, 2010). The sequence can be approximated as a uniform distribution in the range [0, 1]. Figure 2a shows input factor samples from an example Sobol' sequence in two dimensions. For each scenario and site, we generated a (*N* x *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 factors; therefore, a third Sobol' matrix ($\boldsymbol{A_B}$) was constructed from ***A*** and ***B***. In matrix $\boldsymbol{A_B}$, all columns were from ***A***, except the *ith* column, which was from the *ith* column of ***B***, resulting in a (*kN* x *k*) matrix (Fig. 2a). Sec. 3.3.5 provides specific examples of this implementation. From Eq. (3), we compute S_{Ti} as (Jansen, 1999; Saltelli et al., 2010):

$$S_{Ti} = \frac{\frac{1}{2N}\sum_{j=1}^{N}\left(f(\boldsymbol{A})_j - f(\boldsymbol{A}_{\boldsymbol{B}}^{(i)})_j\right)^2}{V(Y)} \qquad (4)$$

where f(***A***) is the model output evaluated on the ***A*** matrix, $f(\boldsymbol{A}^{(i)}_{\boldsymbol{B}})$ is the model output evaluated on the $\boldsymbol{A_B}$ matrix where the *ith* 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 (i.e., computed sensitivity indices with all samples considered) were close to the mean bootstrapped values (i.e., 99% had a difference less than 0.001 and no difference was greater than 0.003), 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 $\boldsymbol{A}_{\mathbf{B}}$ matrices, mapping input factor samples 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:

Step 1) Generate an initial (N x $2k$) Sobol' matrix (with N and k values for each scenario, Table 3), separate into ***A*** and ***B***, and construct $\boldsymbol{A}_{\mathbf{B}}$ (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 input factor sample of each forcing error parameter (ϕ) to the specified error distribution and range (Fig. 2b, Table 3). Here we treat the input factor samples as quantiles, which allows us to map these to errors via different probability distributions. For a uniform distribution, the quantile value scales linearly between the specified lower and upper error ranges (Fig. 2b). This linear scaling is not possible for normal (or lognormal) distributions (due to differences in distribution shape) and we therefore map the quantile values to normal (or lognormal) distributions scaled within the specified range. We begin by generating a probability distribution of random numbers with specified mean=0 and standard deviation of 1 for the case of a normal distribution, and with specified mean=20 and standard deviation of 0.5 for the case of a lognormal distribution. The random numbers of the distribution are normalized in the [0, 1] range by subtracting the minimum value and dividing by the maximum value, and then quantiles of these normalized values are computed. The final step of the mapping is to multiply the normalized quantile by the specified range of uncertainty and adding the lower bound value. For example, a Q_{si} bias parameter of ϕ=0.75 (quantile value) in the [-100 W m^{-2}, +100 W m^{-2}] range would map to a Q_{si} bias of +50 W m^{-2} when assuming a uniform probability distribution but only +14 W m^{-2} when assuming a normal distribution. For context, a bias parameter of +50 W m^{-2} or higher has about a 25% probability of occurring in the uniform distribution but only 2% in the normal distribution.

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

$$F'_i = F_i \phi_{B,i} b_i + \left(F_i + \phi_{B,i}\right)\left(1 - b_i\right) + \phi_{RE,i} R c_i \qquad (5)$$

where F'_i is the perturbed forcing time series, $\phi_{B,i}$ 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), $\phi_{RE,i}$ 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*. The decision to treat biases as multiplicative for *P* but additive for all other forcings was made based on practical considerations (e.g., multiplicative bias in T_{air} are difficult to interpret) and on convention of past studies that report forcing errors. However, we note this is somewhat subjective, as errors in some forcings (e.g. radiation) have been reported in both conventions. For *P*, *U*, and Q_{si}, we restricted random errors to periods with positive values. Similar to other sensitivity analyses (e.g., Baroni and Tarantola, 2014), 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. 5). 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 thus we assumed this truncation 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 four scenarios each required 80 000 simulations, for a total of 1 840 000 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). The outputs included daily time series of SWE, and four summary outputs including peak SWE, mean ablation rate, snow disappearance date, and total snow sublimation.

Step 6) Calculate S_{Ti} for each forcing error parameter and model output (Fig. 2f) based on Sect. 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); this was an essential step to ensure meaningful results. Other studies (e.g., Pappenberger et al., 2008) have also applied screening methods to model output prior to calculating sensitivity indices. For a valid simulation, we required a minimum peak SWE of 50 mm, a minimum continuous snow duration of 15 days, 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, 94% passed the requirements. All cases had at least 86% satisfactory samples, except in UB at SASP, where only ~34% met the requirements. In this case, the most common reason for rejecting a simulation was that too much snow was simulated, such that it never disappeared by the end of the model run. The rejected runs were characterized by high (positive) precipitation biases and low (negative) biases in T_{air}, Q_{si}, and Q_{li}. Despite this attrition, S_{Ti} values still converged in all cases.

4. Results

4.1. Propagation of forcing uncertainty to model outputs

Figure 3 shows density plots of daily SWE from UEB at the four sites and five 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, NB_gauge considered gauge undercatch biases in precipitation, and NB_lab considered lower error magnitudes in all forcings (i.e., laboratory level uncertainty).

Large uncertainties in SWE were evident, particularly in NB, NB+RE, and UB (Fig 3.a-l). The large range in modeled SWE within these three scenarios often translated to large ranges in mean ablation rates (Fig 4.e-h), snow disappearance dates (Fig 4.i-l) and total sublimation (Fig 4.m-p). In contrast, SWE and output uncertainties in NB_gauge and NB_lab were comparatively small (Fig. 3m-t and Fig. 4). Model output ranges were generally larger in NB_gauge than NB_lab. The envelope of SWE simulations in NB_lab more tightly 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 and NB+RE generally yielded similar SWE density plots (Fig. 3a-h), but NB+RE yielded a slightly 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 random errors in the forcings had minimal impact on model behavior at CDP, RME, and SASP. NB+RE and NB model outputs were slightly different at IC (particularly for the ablation rates), indicating that random errors had some influence there, and this was possibly due to the low snow accumulation (~200 mm peak SWE observed) at that site and brief snowmelt season (less than 10 days in the observations).

NB and UB yielded generally very different model outputs (Fig. 3 and Fig. 4). The only difference in these two scenarios was the assumption regarding error distribution (Table 3). Uniformly distributed forcing biases (scenario UB) yielded a relatively 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 (i.e., condensation).

Contrasting NB and NB_gauge, NB_gauge had a lower uncertainty range in SWE and slightly higher mean peak SWE at all sites (Fig. 3 and Fig. 4). With the exception of RME, the ranges in ablation rates in NB_gauge were at least 50% smaller than in NB (Fig. 4 e-h). Snow disappearance ranges were marginally smaller in NB_gauge relative to NB (Fig. 4i-l). Finally, sublimation ranges were very similar between NB and NB_gauge (Fig. 4m-p).

Relative to NB, NB_lab had smaller uncertainty ranges in all model outputs (Fig. 3 and Fig. 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).

4.2. Model sensitivity to forcing error characteristics

Total-order sensitivity indices (S_{Ti}) were calculated for four summary variables of model output (peak SWE, mean ablation rates, snow disappearance dates, and total sublimation) and for daily SWE output at all sites and error scenarios. Examination of the total-order indices with sample size indicated that most indices stabilized after evaluating the model at 3 000 to 5 000 samples (no figures shown). Below we sequentially compare sensitivity indices from different scenarios to scenario NB to test the impact of differences in error characteristics (type, probability distribution, and magnitudes).

4.2.1. Impact of error types

We first focus on sensitivity to forcing bias, as this error type was common to scenarios NB and NB+RE. Figure 5 shows the computed total-order sensitivity indices from the two scenarios (with sensitivities to biases and random errors shown separately in NB+RE). Both NB and

NB+RE showed that UEB peak SWE was most sensitive to *P* bias at all sites (Fig.5a-d). In both scenarios, *P* bias was also the most important factor for ablation rates and snow disappearance at all sites (Fig. 5e-l). For ablation rates in NB, T_{air} bias was the next most important factor (after *P* bias) at CDP, while biases in Q_{si} and Q_{li} were secondarily important at RME (Fig.5f-g). For ablation rates at IC in NB+RE, most types of errors had some baseline influence (i.e., S_{Ti}≥0.5) on model sensitivity (Fig. 5e). In both NB and NB+RE, biases in the radiation terms were of secondary importance to snow disappearance timing (Fig. 5i-k). In contrast to the other three model outputs, sublimation in NB and NB+RE was insensitive to *P* bias and the most important factors varied somewhat between sites and scenarios (Fig. 5m-p). In both scenarios, sublimation was most sensitive to *RH* bias at IC and *U* bias at SASP. At CDP and RME, sublimation was most sensitive to *RH* bias in NB; however, in NB+RE, sublimation was most sensitive to Q_{li} bias at CDP and to T_{air} bias at RME (Fig. 5n-o). In both scenarios, biases in T_{air}, Q_{si}, or Q_{li} were generally of secondary importance for sublimation.

We hypothesized that the snow model outputs would have higher sensitivity to biases than to random errors in the forcings. The results of our analysis generally supported this hypothesis. Across all outputs and sites, S_{Ti} values for random errors were always less than or comparable to the smallest S_{Ti} bias values, and the most important factor was always a bias term (Figure 5). Furthermore, there was typically high correspondence between NB and NB+RE (bias terms only) in terms of identifying the most important forcing error (e.g., *P* bias in peak SWE and ablation rates at all sites, Fig. 5a-h). The main exceptions were snow disappearance at IC (Fig. 5i), and sublimation at CDP and RME (Fig. 5n-o), where the two scenarios identified different errors as the most important factor. However, even in these exceptional cases, the two scenarios yielded similar groupings of more important vs. least important errors. For example, biases in T_{air} and *RH* were important to sublimation at RME in both scenarios (Fig. 5o), though they distinguished these sensitivities differently (i.e., NB found *RH* bias was more important whereas NB+RE found T_{air} bias was more important).

While there was general correspondence between NB and NB+RE (bias terms), sensitivity indices were not identical across cases, due to interactions between biases and random errors in

NB+RE. Random errors changed model sensitivity to biases, and the change in sensitivity was more notable (i.e., absolute change exceeding 0.10) for ablation rates and snow disappearance at IC (Fig. 5e,i) and sublimation at all sites (Fig. 5m-p). Random errors amplified model sensitivity to biases in some cases (e.g., *U* bias in all sublimation scenarios) but diminished model sensitivity to biases in other cases (e.g., *RH* bias in all sublimation scenarios). Because consideration of second-order sensitivity indices was beyond the scope of the study, we were unable to determine which specific interactions were important in terms of error types, and leave this topic for future work.

4.2.2. Impact of probability distribution of errors

We hypothesized that the assumed probability distribution of errors would alter the relative hierarchy of forcing biases. However, the results did not consistently support this hypothesis (Fig. 6). In all cases, scenarios NB and UB identified the same factor as the most important and similar factors as the least important at all sites. Specifically, *P* bias was most important for peak SWE, ablation rates, and snow disappearance at all sites in both scenarios (Fig. 6a-l). The only exception was in scenario UB at IC, where ablation rates had similar sensitivity to *P* bias and *U* bias. In both scenarios, T_{air} bias was the second most important factor for peak SWE and ablation rates at the warmest site, CDP. Both scenarios showed that *RH* bias was the least important factor to snow disappearance at all four sites (Fig. 6i-l). Finally, both NB and UB showed that *P* bias was least important for sublimation (in contrast to the other model outputs) and that *RH* and *U* biases were among the most sensitive factors for sublimation (Fig. 6m-p). More specifically, sublimation was most sensitive to *RH* bias at IC, CDP, and RME, and *U* bias as SASP (Fig. 6m-p).

For a few specific forcings and outputs, the selected probability distribution played a role in model sensitivity to that type of forcing bias. For example, assumption of a uniform probability distribution (UB) for forcing errors enhanced the sensitivity of sublimation to *U* and *RH* biases but reduced sublimation sensitivity to Q_{si} and Q_{li} biases at all sites (Fig. 6m-p). In contrast, assuming a normal distribution (NB) of biases yielded the opposite results. Additionally, modeled ablation rates at IC were notably more sensitive to forcing biases (precipitation excluded) in scenario UB than in NB.

4.2.3. Impact of error magnitude

We hypothesized that the relative magnitude of forcing errors would exert a strong control on model sensitivity. Comparing NB to NB_gauge and to NB_lab generally supported this hypothesis (Fig. 7). The contrast in S_{Ti} values between scenarios NB, NB_gauge, and NB_lab implied that the specified ranges of forcing errors was a critical determinant of model sensitivity.

While *P* bias was the most important factor at all sites in NB for peak SWE, ablation rates, and snow disappearance, *P* bias was never the most important factor for these model outputs in NB_gauge, and in many cases was among the least important errors (Fig. 7a-l). In NB_gauge, peak SWE was most sensitive to *RH* bias at IC, T_{air} bias at CDP and RME, and Q_{li} bias at SASP (Fig. 7a-d). Ablation rates in NB_gauge were most sensitive to T_{air} bias at CDP and to Q_{li} bias at IC, RME, and SASP (Fig. 7e-h). Snow disappearance was also most sensitive to Q_{li} bias at all four sites in NB_gauge (Fig. 7i-l). However, for sublimation at all sites, NB and NB_gauge yielded very similar sensitivities to forcing biases (Fig. 7m-p). Specifically, in both NB and NB_gauge, modeled sublimation was most sensitive to *RH* bias at IC, CDP, and RME and to *U* bias at SASP (Fig. 7m-p). The similarity in sublimation sensitivity indices between NB and NB_gauge emerged because these scenarios only differed in terms of *P* uncertainty (Table 3) and because *P* bias was not important to modeled sublimation. The contrast between sensitivity indices in these two scenarios and for these four outputs illustrated that model sensitivity may depend on both the magnitudes of uncertainty for specific forcings and on the output of interest.

Whereas NB_gauge demonstrated that reducing the magnitude of forcing uncertainty in one factor (i.e., precipitation) was sufficient to change which factors were most and least important, NB_lab showed that changing the magnitude of forcing uncertainty in all terms could yield a substantially different pattern of model sensitivity (Fig. 7). As a primary example, scenarios NB and NB_lab 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. 7a-l). For sublimation, NB_lab sensitivity indices indicated that Q_{li} bias was most important, whereas *RH* bias (IC, CDP, and RME) and *U* bias (SASP) were most important in NB (Fig. 7m-p). Across all sites and outputs in NB_lab, Q_{li} bias was consistently the most important factor (Fig. 7). In one sense, this was surprising, given that the bias magnitudes were lower for Q_{li} than for Q_{si} (Table 3). However, the albedo of snow minimizes the amount of energy transmitted to the snowpack from Q_{si}, thereby rendering Q_{si} errors less important than Q_{li} errors. Additionally, the non-linear nature of the model may enhance the role of Q_{li} through interactions with other factors. The general lack of importance in *P* bias in NB_lab (main exception was peak SWE at IC, Fig. 7a) was due to the discrepancy between the laboratory specified accuracy for *P* gauges and typical errors encountered in the field.

4.2.4. Relative controls of forcing error characteristics on SWE sensitivity

The above results sequentially compared sensitivity indices from different error scenarios to NB in order to ascertain how different assumptions regarding error types, distributions, and magnitudes translated to changes in model sensitivity. To summarize the relative controls of these three forcing error characteristics on model sensitivity, we calculated daily sensitivity indices of modeled SWE to forcing biases at each site and scenario (Fig. 8). This final analysis was conceptually different than the previous analyses (Fig 5-7) in terms of the model output considered. Whereas the previous analyses computed sensitivity indices for summative model outputs (e.g., peak SWE, total sublimation), the final analysis re-calculated sensitivity indices for SWE each day. This approach allowed us to examine how SWE model sensitivity changed as a function of time within the snow season.

Comparing the broad patterns in the time varying S_{Ti} values across the five scenarios, it was evident that error magnitudes were the greatest determinant in model sensitivity to forcing errors through the snow season (compare Fig. 8a-l with Fig. 8m-t). NB, NB+RE, and UB exhibited similar patterns, with high S_{Ti} in *P* bias throughout the year and with the other forcing biases yielding low S_{Ti} values in the winter and increasing S_{Ti} values in the spring and early summer for some forcings (Fig. 8a-l). In contrast, NB_gauge and NB_lab (Fig. 8m-t) had lower S_{Ti} values for P bias, and more coherent changes in S_{Ti} values that were more synchronized with the specific part of the snow season.

After error magnitudes, the next most important determinant to model sensitivity was the probabilistic distribution of forcing errors (compare Fig. 8a-d and Fig. 8i-l). Relative to NB, UB tended to yield lower S_{Ti} values for P bias. UB also had higher S_{Ti} values for biases in T_{air}, Q_{li}, and Q_{si} as time progressed at IC, CDP, and RME (Fig. 8i-k). Finally, the addition of random errors was least important to model sensitivity, as the evolution of S_{Ti} bias values was very similar between NB and NB+RE at most sites (compare Fig. 8a-d and Fig. 8e-h). Random errors mattered the most to modeled SWE at IC, but random errors only changed S_{Ti} values (on average) by less than 10%.

5. Discussion

Here we examined the sensitivity of physically-based snow simulations to forcing error characteristics (i.e., types, probability distributions, and magnitudes) using Sobol' global sensitivity analysis. A key result is that among these three characteristics, the magnitude of biases had the most significant impact on UEB simulations (Figs. 3-4) and on model sensitivity (Figs. 7-8). The assumed probability distribution of biases was important in that it increased the range of model outputs (compare NB and UB in Fig. 4), but surprisingly, this usually translated to only modest changes in model sensitivity to forcing errors (Figs. 6 and 8). Random errors were usually less important than biases. Although random errors changed model sensitivity to biases through error interactions, this effect was only large in specific conditions (e.g., ablation rates at IC, Fig. 5e), and the snow model was never more sensitive to random errors than to biases (Fig. 5). Below we discuss these three error characteristics (in order of importance, as suggested by the results), place forcing errors in the context of structural uncertainty, and identify limitations of the analysis and future research directions.

5.1. Ranges of error magnitudes

The results supported our hypothesis that the magnitude of biases strongly influences the relative importance of forcing errors. The three magnitudes of uncertainty considered (NB, NB_gauge, and NB_lab) all resulted in different patterns in model sensitivity to forcing biases, and these patterns also varied with the output of interest (Fig. 7). Modeled peak SWE, ablation rates, and snow disappearance were consistently sensitive to *P* bias in scenario NB and to Q_{li} bias in scenario NB_lab, but there was less consistency in the dominant forcing errors across these three outputs in scenario NB_gauge. 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. Biases in *RH*, *U*, and T_{air} were often the major controls on modeled sublimation in NB, NB+RE, UB, and NB_gauge, while Q_{li} bias controlled modeled sublimation in NB_lab. These field 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}. However, they did not consider Q_{li} in their sensitivity analysis and so the experiments are not exactly comparable. These results suggest that no single forcing is important across all modeled variables, and model sensitivity strongly depends on the output of interest.

The dominant effect of *P* bias on modeled peak SWE, ablation rates, and snow disappearance in the field scenarios (e.g., NB) confirmed previous reports that *P* uncertainty is a major control on snowpack dynamics (Durand and Margulis, 2008; He et al., 2011b; Schmucki et al., 2014). It was surprising that *P* bias was often the most critical forcing error for ablation rates in these scenarios (Fig. 5-6). Prior investigations into the relative importance of forcings to ablation were typically framed for a snowpack at the end of winter, such that *P* uncertainty was not considered (e.g., Zuzel and Cox, 1975). The results here showed that ablation rates were highly sensitive to *P* bias and this is likely because it controlled the timing and length of the ablation season. Positive *P* bias extends the fraction of the ablation season in the warmest summer months when ablation rates and radiative energy approach maximum values, whereas negative *P* bias truncates the fraction of ablation in the warm season. Trujillo and Molotch (2014) reported a similar result based on SNOTEL observations.

The contrast between scenarios NB, NB_gauge, and NB_lab highlights that selection of the error ranges is a critical step in sensitivity analysis. However, we recognize that there is some subjectivity in the specification of these ranges. Quantification of errors in forcing estimation methods is best achieved through comparisons with surface observations (e.g., Bohn et al., 2013; Flerchinger et al., 2009), but it remains challenging to specify error ranges with confidence (Song et al., 2015). Key considerations controlling the ranges and impacts of forcing errors include the representativeness of the forcing data (e.g., reanalysis, numerical weather model output, extrapolated surface measurements, etc.) in the study area, the length scale of dominant processes (e.g., snow drifting), and the configuration of the snow model (e.g., spatial scale, complexity). Here we selected ranges in the field scenarios to encompass errors encountered across a variety of possible forcing data sources (Table 3), but ultimately the appropriate ranges must be tailored to the specific application. This supports the need for continual evaluation of forcing datasets across a variety of climates and environmental conditions.

5.2. Probability distribution of errors

The results did not universally support our hypothesis that the assumed probability distribution of biases was important to the relative ranking of forcing errors. The relative consistency in the dominant forcing errors between NB and UB may have emerged because the probability distributions of all six forcing biases varied together between these two scenarios (i.e., all forcing biases were uniform in UB and either normal or lognormal in NB). While we did not conduct additional tests, we suspect that changing the probability distribution of just a single forcing error (e.g., T_{air} bias) from normal to uniform would have uniquely enhanced model sensitivity to that particular forcing error (Touhami et al., 2013).

The similarity of results between scenarios NB and UB conform to findings in previous studies (e.g., Foscarini et al., 2010; Touhami et al., 2013) where uniform and normal distributions identified similar factors as the most important. These previous studies imply that greater differences in sensitivity indices (as a function of distribution) will emerge when factor interactions are more prominent. The case with the strongest error interactions here (i.e., ablation rates at IC) also yielded the largest differences in sensitivity indices between scenarios NB and UB, which is consistent with the prevailing logic.

5.3. Error types

The results were consistent with our hypothesis that the snow model is more sensitive to biases than to random errors in the forcings. While previous investigations supported this idea for shortwave and longwave forcings in physically based snow models (i.e., Lapo et al., 2015), the current study showed that biases are more important than random errors for all commonly required meteorological forcings (and not just irradiances). The model was more sensitive to biases and less sensitive to random errors due to the systematic nature of biases. In contrast, the effect of random errors tended to cancel out when integrating model outputs over long periods. Our selected model outputs were generally a function of several months of mass and energy exchange in the snowpack, thereby ensuring minimization of effects from random errors. Random errors only had a greater impact on ablation rates at IC (Fig. 5e), and this was because the relatively brief snowmelt period presented an opportunity for the random errors to not cancel out. Hence, the model may have greater sensitivity to random errors for other model outputs not considered here that integrate over relatively short time scales (e.g., snowmelt over a single day).

5.4. Contextualizing forcing and structural uncertainties

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 and to place our results in context, we compare our results at CDP in 2005-2006 to Essery et al. (2013), who assessed the impact of structural uncertainty in a suite of local snowpack processes (i.e., snow compaction, fresh snow density, snow albedo evolution, surface heat and moisture fluxes, snow cover fraction, snow hydrology, and thermal conductivity) on SWE simulations from 1701 physically based snow models at the same site/year. Figure 9 compares the 95% uncertainty ranges in peak SWE, ablation rates, and snow disappearance in NB, NB_gauge, and NB_lab to the ranges found across the 1701 snow models of Essery et al. (2013). From the comparisons at this site, it is clear that the uncertainty associated with drifting snow (i.e., scenario NB) overwhelms the structural uncertainty in local snowpack processes for all three model outputs. As discussed previously, it could be argued that the uncertainty due to drifting snow is a structural issue (not a forcing issue) and that this does not represent the uncertainty of sheltered areas where drifting snow less important. Hence, NB_gauge may be a better determinant of the level of uncertainty that can be attributed unambiguously to errors in forcing data. In that case, the output uncertainty range due to model forcing is still larger than that due to the structural uncertainty (as considered by Essery et al., 2013) in the cases of peak SWE and snow disappearance but is smaller for ablation rates (Fig. 9). As expected, the case of forcing uncertainty in NB_lab yields the lowest range in model outputs at CDP (Fig. 9), though it is interesting to note that the uncertainty in peak SWE due to structural uncertainty (90 mm) is only marginally larger than that due to the specified instrument accuracy (60 mm). These comparisons illustrate that forcing uncertainty cannot be discounted, and the magnitude of forcing uncertainty is a critical factor in how forcing uncertainty compares to other sources of uncertainty (e.g., structural). This resonates with the recent work of Magnusson et al. (2015) who found that uncertainty in the *P* forcing was a greater determinant of model performance than structural considerations.

5.5. Caveats and future research

Limitations of the analysis are that the impact of forcing error characteristics on model behavior is evaluated through the lens of a single sensitivity analysis method and a single snow model. It is possible that alternative sensitivity analysis methods might yield different results than the Sobol' method, as suggested in previous studies (e.g., Pappenberger et al., 2008). Likewise, we recognize it is possible that different snow models may yield different sensitivities to forcing uncertainty. As one example, both Koivusalo and Heikinheimo (1999) and Lapo et al. (2015) found UEB (Tarboton and Luce, 1996) and the SNTHERM model (Jordan, 1991) exhibited significant differences in radiative and turbulent heat exchange. As another example, the role of *U* bias on snowpack formation may vary strongly depending on the snow model configuration. Because of the lack of wind transport in UEB, we lumped snow drift uncertainty into *P* uncertainty via a "drift factor" formulation (Luce et al., 1998) and this could not account for the role of wind in snow drift/scour processes (Mott and Lehning, 2010; Winstral et al., 2013). This convention would be unnecessary for a model that explicitly models this process (e.g., the SNOWPACK model, Lehning et al., 2006), and for this type of model we would expect the role of *U* bias to be enhanced (relative to UEB) for outputs such as peak SWE and snow disappearance timing. While sensitivity may vary with model selection in these examples, there is also evidence suggesting that similar results may emerge when using different snow models for a similar type of error scenario. Despite using different models, a somewhat different suite of forcing variables, and slightly different error ranges, our NB_lab experiment corroborated independent reports that Q_{li} measurement uncertainty was most important to both modeled snow disappearance (Skiles et al., 2012) and sublimation/latent heat exchange (Sauter and Obleitner, 2015). Our analysis demonstrated this result was consistent across four snow climates and this result was apparent in four different model outputs (Fig. 7). The implication here is that more work is needed to better understand how different snow models respond to forcing uncertainty.

Generalizing the relationship between model sensitivity and site climate is a research topic of high interest. Although we found similarities in model sensitivity to specific forcing errors across sites (e.g., high sensitivity to *P* bias in peak SWE, ablation rates, and snow disappearance in NB, NB+RE, and UB, Fig. 8a-l), we note that the sites exhibited some differences in sensitivity when *P* uncertainty was reduced to gauge levels (Fig. 8m-p). Additionally, the sites exhibited differences in the relative importance of secondary forcing errors (Fig. 6-7). There may be interesting linkages between climate and model sensitivity, but we were unable to generalize relationships between site geo-characteristics and sensitivity indices because of the relatively low number of sites represented here (n=4 sites, 1 year each) and the confounding number of differences between sites. A much larger population of snow measurement sites is required in order to test relationships between sensitivity indices and site characteristics, and this is an important avenue of future research. A successful example of relating climate characteristics to sensitivity indices when many study sites and years are available can be found in van Werkhoven et al. (2008).

While the Sobol' method is often considered the "baseline" method in global sensitivity analysis, we note the limitation is that it comes at a relatively high computation cost (1 840 000 simulations across four sites and five error scenarios) and it may be prohibitive for many modeling applications (e.g., for models of higher complexity and dimensionality). For context, the typical time required for a single simulation was 1.4 seconds, resulting in a total computational expense of 720 hours (30 days) across all scenarios. Examination of the convergence rates indicated that most sensitivity indices stabilized after one-third of the simulations completed, and hence the same results could have been found using significantly fewer simulations (no figures shown). Ongoing research is developing new sensitivity analysis methods that compare well to Sobol' but with reduced computational demands (e.g., FAST, Cukier, 1973; method of Morris, 1991; DELSA, Rakovec et al., 2014), and is comparing how different methods classify sensitive factors differently (Pappenberger et al., 2008; Tang et al., 2007). We expect that detailed 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.

The question remains: "what can be done about forcing errors in hydrologic modeling?" First, the results suggest model-based hypothesis testing must account for uncertainties in forcing data. The results also highlight the need for continued research in constraining *P* uncertainty in snow-affected catchments. Progress is being achieved with advanced pathways for quantifying snowfall precipitation, such as NWP models (Rasmussen et al., 2011, 2014) and through systematic intercomparisons of precipitation and snow gauges (e.g., Solid Precipitation Intercomparison Experiment, http://www.rap.ucar.edu/projects/SPICE/). However, in a broader sense, the hydrologic community should also 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), which have a legacy in data assimilation methods (e.g., precipitation uncertainty, Durand and Margulis, 2007), 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.

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. Furthermore, this work demonstrates that sensitivity analysis can be applied to understand the role of specific error characteristics in model behavior. Key considerations in model sensitivity to forcing errors are the magnitudes of forcing errors and the outputs of interest. For the physically-based snow model tested, random errors in forcings are generally less important than biases, and the probability distribution of biases is relatively less important to model sensitivity 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. Based on this framework, 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.

Acknowledgements

M. Raleigh was supported by a post-doctoral fellowship in the Advanced Study Program at the National Center for Atmospheric Research (NCAR). NCAR is sponsored by the National Science Foundation. J. Lundquist was supported by NSF (EAR-838166 and EAR-1215771). We thank F. Pianosi, J. Li, R. Essery, R. Rosolem, A. Winstral, and one anonymous reviewer for their thoughtful reviews and R. Woods for serving as editor. 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. Landry for assistance with Swamp Angel data, and E. Gutmann and P. Mendoza for feedback. For Imnavait Creek data, we acknowledge U.S. 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. Simul., 58(2), 99–120, doi:10.1080/00949659708811825, 1997.

Di Baldassarre, G. and Montanari, A.: Uncertainty in river discharge observations: a quantitative analysis, Hydrol. Earth Syst. Sci., 13(6), 913–921, doi:10.5194/hess-13-913-2009, 2009.

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.

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 (80-.)., 319(5866), 1080–1083, doi:10.1126/science.1152538, 2008.

Baroni, G. and Tarantola, S.: A General Probabilistic Framework for uncertainty and global sensitivity analysis of deterministic models: A hydrological case study, Environ. Model. Softw., 51, 26–34, doi:10.1016/j.envsoft.2013.09.022, 2014.

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(5), 562–576, doi:10.1016/j.advwatres.2011.01.008, 2011.

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(11-12), 1134–1149, doi:10.1016/j.mcm.2007.05.017, 2008.

Beven, K. and Binley, A.: The future of distributed models: model calibration and uncertainty prediction, Hydrol. Process., 6(3), 279–298, doi:10.1002/hyp.3360060305, 1992.

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, Agric. For. Meteorol., 176, 38–49, doi:10.1016/j.agrformet.2013.03.003, 2013.

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, Agric. For. Meteorol., 91(3-4), 161–176, doi:10.1016/S0168-1923(98)00076-8, 1998.

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.

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.

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.

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, 2010a.

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, 2010b.

Brun, E., David, P., Sudul, M. and Brunot, G.: A numerical model to simulate snow-cover stratigraphy for operational avalanche forecasting, J. Glaciol., 38(128), 13–22, 1992.

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

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(1-4), 242–266, doi:10.1016/j.jhydrol.2004.03.042, 2004.

Campolongo, F., Saltelli, A. and Cariboni, J.: From screening to quantitative sensitivity analysis. A unified approach, Comput. Phys. Commun., 182(4), 978–988, doi:10.1016/j.cpc.2010.12.039, 2011.

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(5), 770–782, doi:10.1002/joc.2123, 2011.

Christopher Frey, H. and Patil, S. R.: Identification and Review of Sensitivity Analysis Methods, Risk Anal., 22(3), 553–578, doi:10.1111/0272-4332.00039, 2002.

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. Modell., 189(1-2), 209–220, doi:10.1016/j.ecolmodel.2005.03.016, 2005.

Clark, M. P., Hendrikx, J., Slater, A. G., Kavetski, D., Anderson, B., Cullen, N. J., Kerr, T., Örn Hreinsson, E. and Woods, R. A.: Representing spatial variability of snow water equivalent in hydrologic and land-surface models: A review, Water Resour. Res., 47(7), doi:10.1029/2011WR010745, 2011a.

Clark, M. P., Kavetski, D. and Fenicia, F.: Pursuing the method of multiple working hypotheses for hydrological modeling, Water Resour. Res., 47(9), 1–16, doi:10.1029/2010WR009827, 2011b.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Brekke, L. D., Arnold, J. R., Gochis, D. J. and Rasmussen, R. M.: A unified approach for process-based hydrologic modeling: 1. Modeling concept, Water Resour. Res., 51, doi:10.1002/2015WR017198, 2015a.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Gochis, D. J., Rasmussen, R. M., Tarboton, D. G., Mahat, V., Flerchinger, G. N. and Marks, D. G.: A unified approach for process-based hydrologic modeling: 2. Model implementation and case studies, Water Resour. Res., 51, doi:10.1002/2015WR017200, 2015b.

Clark, M. P. and Slater, A. G.: Probabilistic Quantitative Precipitation Estimation in Complex Terrain, J. Hydrometeorol., 7(1), 3–22, doi:10.1175/JHM474.1, 2006.

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(12), doi:10.1029/2007WR006735, 2008.

Cukier, R. I.: Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory, J. Chem. Phys., 59(8), 3873, doi:10.1063/1.1680571, 1973.

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.

Dee, D. P.: Bias and data assimilation, Q. J. R. Meteorol. Soc., 131(613), 3323–3343, doi:10.1256/qj.05.137, 2005.

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(11), 4401–4413, doi:10.5194/hess-17-4401-2013, 2013.

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

Duan, Q., Sorooshian, S. and Gupta, V.: Effective and efficient global optimization for conceptual rainfall-runoff models, Water Resour. Res., 28(4), 1015–1031, doi:10.1029/91WR02985, 1992.

Durand, M. and Margulis, S. A.: Correcting first-order errors in snow water equivalent estimates using a multifrequency, multiscale radiometric data assimilation scheme, J. Geophys. Res., 112(D13), 1–15, doi:10.1029/2006JD008067, 2007.

Durand, M. and Margulis, S. A.: Effects of uncertainty magnitude and accuracy on assimilation of multiscale measurements for snowpack characterization, J. Geophys. Res., 113(D2), D02105, doi:10.1029/2007JD008662, 2008.

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(4), 1384–1403, doi:10.1175/JHM-D-13-083.1, 2014.

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.

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(1), doi:10.1890/ES11-00202.1, 2012.

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, doi:10.1002/wrcr.20318, 2013.

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(3), 1–13, doi:10.1029/2008WR007394, 2009.

Flint, A. L. and Childs, S. W.: Calculation of solar radiation in mountainous terrain, Agric. For. Meteorol., 40(3), 233–249, doi:10.1016/0168-1923(87)90061-X, 1987.

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(6), doi:10.1029/2008WR007255, 2009.

Foscarini, F., Bellocchi, G., Confalonieri, R., Savini, C. and Van den Eede, G.: Sensitivity analysis in fuzzy systems: Integration of SimLab and DANA, Environ. Model. Softw., 25(10), 1256–1260, doi:10.1016/j.envsoft.2010.03.024, 2010.

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. Climatol., 48(5), 1033–1049, doi:10.1175/2008JAMC2084.1, 2009.

Georgakakos, K., Seo, D., Gupta, H., Schaake, J. and Butts, M.: Towards the characterization of streamflow simulation uncertainty through multimodel ensembles, J. Hydrol., 298(1-4), 222–241, doi:10.1016/j.jhydrol.2004.03.037, 2004.

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

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.

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(8), 5029–5046, doi:10.1002/wrcr.20387, 2013.

Guan, H., Wilson, J. L. and Makhnin, O.: Geostatistical Mapping of Mountain Precipitation Incorporating Autosearched Effects of Terrain and Climatic Characteristics, J. Hydrometeorol., 6(6), 1018–1031, doi:10.1175/JHM448.1, 2005.

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.

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

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(1), 114–127, doi:10.1016/j.advwatres.2010.10.002, 2011a.

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(7), 1–17, doi:10.1029/2010WR009753, 2011b.

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(7), 2893–2903, doi:10.5194/hess-17-2893-2013, 2013.

Herrero, J. and Polo, M. J.: Parameterization of atmospheric longwave emissivity in a mountainous site for all sky conditions, Hydrol. Earth Syst. Sci., 16(9), 3139–3147, doi:10.5194/hess-16-3139-2012, 2012.

Hiemstra, C. A., Liston, G. E. and Reiners, W. A.: Observing, modelling, and validating snow redistribution by wind in a Wyoming upper treeline landscape, Ecol. Modell., 197(1-2), 35–51, doi:10.1016/j.ecolmodel.2006.03.005, 2006.

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. Climatol., 48(4), 725–741, doi:10.1175/2008JAMC1979.1, 2009.

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(8), 1–13, doi:10.1029/2008WR007600, 2009.

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(D18), 4583, doi:10.1029/2002JD002991, 2003.

Jansen, M. J. W.: Analysis of variance designs for model output, Comput. Phys. Commun., 117(1-2), 35–43, doi:10.1016/S0010-4655(98)00154-4, 1999.

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(2), 1–15, doi:10.1029/2011WR011006, 2012.

Jiménez, P. A., Dudhia, J. and Navarro, J.: On the surface wind speed probability density function over complex terrain, Geophys. Res. Lett., 38(22), doi:10.1029/2011GL049669, 2011.

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(D22), 28817, doi:10.1029/1998JD200012, 1998.

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

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(6), 1099–1109, doi:10.1029/91WR00262, 1991.

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

Kavetski, D., Kuczera, G. and Franks, S. W.: Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, Water Resour. Res., 42(3), W03407, doi:10.1029/2005WR004368, 2006a.

Kavetski, D., Kuczera, G. and Franks, S. W.: Bayesian analysis of input uncertainty in hydrological modeling: 2. Application, Water Resour. Res., 42(3), W03408, doi:10.1029/2005WR004376, 2006b.

Kelleher, C., Wagener, T. and McGlynn, B.: Model-based analysis of the influence of catchment properties on hydrologic partitioning across five mountain headwater subcatchments, Water Resour. Res., 51, doi:10.1002/2014WR016147, 2015.

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

Kucherenko, S., Tarantola, S. and Annoni, P.: Estimation of global sensitivity indices for models with dependent variables, Comput. Phys. Commun., 183(4), 937–946, doi:10.1016/j.cpc.2011.12.020, 2012.

Kuczera, G. and Parent, E.: Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm, J. Hydrol., 211(1-4), 69–85, doi:10.1016/S0022-1694(98)00198-X, 1998.

Kuczera, G., Renard, B., Thyer, M. and Kavetski, D.: There are no hydrological monsters, just models and observations with large uncertainties!, Hydrol. Sci. J., 55(6), 980–991, doi:10.1080/02626667.2010.504677, 2010.

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, doi:10.1002/2013WR013711, 2014.

Lapo, K. E., Hinkelman, L. M., Raleigh, M. S. and Lundquist, J. D.: Impact of errors in the downwelling irradiances on simulations of snow water equivalent, snow surface temperature, and the snow energy balance, Water Resour. Res., 6(4), doi:10.1002/2014WR016259, 2015.

Lapp, S., Byrne, J., Townshend, I. and Kienzle, S.: Climate warming impacts on snowpack accumulation in an alpine watershed, Int. J. Climatol., 25(4), 521–536, doi:10.1002/joc.1140, 2005.

Leavesley, G. H.: Modeling the effects of climate change on water resources - a review, Clim. Change, 28(1-2), 159–177, doi:10.1007/BF01094105, 1994.

Lehning, M., Völksch, I., Gustafsson, D., Nguyen, T. A., Stähli, M. and Zappa, M.: ALPINE3D: a detailed model of mountain surface processes and its application to snow hydrology, Hydrol. Process., 20, 2111–2128, doi:10.1002/hyp.6204, 2006.

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(8), 3279–3293, doi:10.5194/hess-17-3279-2013, 2013.

Liston, G. E.: Representing Subgrid Snow Cover Heterogeneities in Regional and Global Models, J. Clim., 17(6), 1381–1397, doi:10.1175/1520-0442(2004)017<1381:RSSCHI>2.0.CO;2, 2004.

Liston, G. E. and Elder, K.: A Meteorological Distribution System for High-Resolution Terrestrial Modeling (MicroMet), J. Hydrometeorol., 7(2), 217–234, doi:10.1175/JHM486.1, 2006.

Liu, Y. and Gupta, H. V.: Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, Water Resour. Res., 43(7), doi:10.1029/2006WR005756, 2007.

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(10-11), 1671–1683, doi:10.1002/(SICI)1099-1085(199808/09)12:10/11<1671::AID-HYP688>3.0.CO;2-N, 1998.

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.

Lundquist, J. D., Wayand, N. E., Massmann, A., Clark, M. P., Lott, F. and Cristea, N. C.: Diagnosis of insidious data disasters, Water Resour. Res., 51, doi:10.1002/2014WR016585, 2015.

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(7), 947–959, doi:10.1002/joc.1583, 2008.

Magnusson, J., Wever, N., Essery, R., Helbig, N., Winstral, A. and Jonas, T.: Evaluating snow models with varying process representations for hydrological applications, Water Resour. Res., 51, doi:10.1002/2014WR016498, 2015.

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

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(3), 251–278, doi:10.1007/s11269-005-3179-2, 2005.

Marks, D. and Dozier, J.: Climate and energy exchange at the snow surface in the Alpine Region of the Sierra Nevada: 2. Snow cover energy balance, Water Resour. Res., 28(11), 3043–3054, doi:10.1029/92WR01483, 1992.

Marks, D., Dozier, J. and Davis, R. E.: Climate and energy exchange at the snow surface in the Alpine Region of the Sierra Nevada: 1. Meteorological measurements and monitoring, Water Resour. Res., 28(11), 3029–3042, doi:10.1029/92WR01482, 1992.

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(6), doi:10.1029/2008WR007301, 2009.

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(6), 1970–1976, doi:10.1175/JHM-D-12-066.1, 2012.

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(1), 474–488, doi:10.1175/JHM-D-13-036.1, 2014.

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(1), 13–21, doi:10.5194/essd-4-13-2012, 2012.

Morris, M. D.: Factorial Sampling Plans for Preliminary Computational Experiments, Technometrics, 33(2), 161–174, doi:10.1080/00401706.1991.10484804, 1991.

Mott, R. and Lehning, M.: Meteorological Modeling of Very High-Resolution Wind Fields and Snow Deposition for Mountains, J. Hydrometeorol., 11(4), 934–949, doi:10.1175/2010JHM1216.1, 2010.

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

Niemelä, S., Räisänen, P. and Savijärvi, H.: Comparison of surface radiative flux parameterizations: Part II. Shortwave radiation, Atmos. Res., 58(2), 141–154, doi:10.1016/S0169-8095(01)00085-0, 2001b.

Nossent, J., Elsen, P. and Bauwens, W.: Sobol’ sensitivity analysis of a complex environmental model, Environ. Model. Softw., 26(12), 1515–1525, doi:10.1016/j.envsoft.2011.08.010, 2011.

Oudin, L., Perrin, C., Mathevet, T., Andréassian, V. and Michel, C.: Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models, J. Hydrol., 320(1-2), 62–83, doi:10.1016/j.jhydrol.2005.07.016, 2006.

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.

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

Phillips, D. and Marks, D.: Spatial uncertainty analysis: propagation of interpolation errors in spatially distributed models, Ecol. Modell., 91(1-3), 213–229, doi:10.1016/0304-3800(95)00191-3, 1996.

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, doi:10.1002/2013WR014063, 2014.

Raleigh, M. S.: Quantification of uncertainties in snow accumulation, snowmelt, and snow disappearance dates, University of Washington., 2013.

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), doi:10.1029/2011WR010542, 2012.

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 Bed, Bull. Am. Meteorol. Soc., 93(6), 811–829, doi:10.1175/BAMS-D-11-00052.1, 2012.

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(3), 1091–1116, doi:10.1175/JHM-D-13-0118.1, 2014.

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. Clim., 24(12), 3015–3048, doi:10.1175/2010JCLI3985.1, 2011.

Razavi, S. and Gupta, H. V.: What do we mean by sensitivity analysis? The need for comprehensive characterization of "Global" sensitivity in Earth and Environmental Systems Models, Water Resour. Res., 51, doi:10.1002/2014WR016527, 2015.

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(7), W07702, doi:10.1029/2010WR010030, 2011.

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(11), 1586–1597, doi:10.1016/j.advwatres.2005.11.013, 2006.

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.

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.

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

Saltelli, A. and Annoni, P.: How to avoid a perfunctory sensitivity analysis, Environ. Model. Softw., 25(12), 1508–1517, doi:10.1016/j.envsoft.2010.04.012, 2010.

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M. and Tarantola, S.: Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index, Comput. Phys. Commun., 181(2), 259–270, doi:10.1016/j.cpc.2009.09.018, 2010.

Sauter, T. and Obleitner, F.: Assessment of the uncertainty of snowpack simulations based on variance decomposition, Geosci. Model Dev. Discuss., 8(3), 2807–2845, doi:10.5194/gmdd-8-2807-2015, 2015.

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.

Schoups, G. and Hopmans, J. W.: Evaluation of Model Complexity and Input Uncertainty of Field-Scale Water Flow and Salt Transport, Vadose Zo. J., 5(3), 951, doi:10.2136/vzj2005.0130, 2006.

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(7), 2145–2160, doi:10.1029/1999WR900090, 1999.

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.

Skiles, S. M., Painter, T. H., Deems, J. S., Bryant, A. C. and Landry, C. C.: Dust radiative forcing in snow of the Upper Colorado River Basin: 2. Interannual variability in radiative forcing and snowmelt rates, Water Resour. Res., 48(7), doi:10.1029/2012WR011986, 2012.

Slater, A. G. and Clark, M. P.: Snow Data Assimilation via an Ensemble Kalman Filter, J. Hydrometeorol., 7(3), 478–493, doi:10.1175/JHM505.1, 2006.

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(1), 7–25, doi:10.1175/1525-7541(2001)002<0007:TROSIL>2.0.CO;2, 2001.

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(1), W01410, doi:10.1029/2006WR005205, 2008.

Sobol′, I.: On sensitivity estimation for nonlinear mathematical models, Mat. Model., 2(1), 112–118, 1990.

Song, X., Zhang, J., Zhan, C., Xuan, Y., Ye, M. and Xu, C.: Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications, J. Hydrol., 523(225), 739–757, doi:10.1016/j.jhydrol.2015.02.013, 2015.

Spank, U., Schwärzel, K., Renner, M., Moderow, U. and Bernhofer, C.: Effects of measurement uncertainties of meteorological data on estimates of site water balance components, J. Hydrol., 492, 176–189, doi:10.1016/j.jhydrol.2013.03.047, 2013.

Sturm, M., Holmgren, J. and Liston, G. E.: A Seasonal Snow Cover Classification System for Local to Global Applications, J. Clim., 8(5), 1261–1283, doi:10.1175/1520-0442(1995)008<1261:ASSCCS>2.0.CO;2, 1995.

Sturm, M. and Wagner, A. M.: Using repeated patterns in snow distribution modeling: An Arctic example, Water Resour. Res., 46(12), 1–15, doi:10.1029/2010WR009434, 2010.

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(2), 793–817, doi:10.5194/hess-11-793-2007, 2007.

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.

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 terrain in Austria, Agric. For. Meteorol., 104(4), 255–271, doi:10.1016/S0168-1923(00)00170-2, 2000.

Touhami, H. Ben, Lardy, R., Barra, V. and Bellocchi, G.: Screening parameters in the Pasture Simulation model using the Morris method, Ecol. Modell., 266(1), 42–57, doi:10.1016/j.ecolmodel.2013.07.005, 2013.

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

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. Environ. Res. Risk Assess., 23(7), 1011–1026, doi:10.1007/s00477-008-0274-y, 2008a.

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(12), doi:10.1029/2007WR006720, 2008b.

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(1), W01017, doi:10.1029/2004WR003059, 2005.

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(8), doi:10.1029/2002WR001746, 2003a.

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(8), doi:10.1029/2002WR001642, 2003b.

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(3), 677–699, doi:10.1175/JHM-D-12-0102.1, 2013.

Van Werkhoven, K., Wagener, T., Reed, P. and Tang, Y.: Characterization of watershed model behavior across a hydroclimatic gradient, Water Resour. Res., 44(1), W01429, doi:10.1029/2007WR006271, 2008.

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(18), 3585–3603, doi:10.1002/hyp.1238, 2002.

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.

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.

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(D1), D01101, doi:10.1029/2004JD005130, 2005.

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(9), 2665–2677, doi:10.1029/2000WR900158, 2000.

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.

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(12), 15071–15118, doi:10.5194/hessd-10-15071-2013, 2013.

Zuzel, J. F. and Cox, L. M.: Relative importance of meteorological variables in snowmelt, Water Resour. Res., 11(1), 174–176, doi:10.1029/WR011i001p00174, 1975.

7. Tables

Table 1 Basic characteristics of the snow study sites, ordered from left-to-right by increasing elevation.

Site Name	Imnavait Creek	Col de Porte	Reynolds Mountain East (sheltered site)	Swamp Angel Study Plot
Site ID	IC	CDP	RME	SASP
Location	Alaska, USA	Rhône-Alpes, France	Idaho, USA	Colorado, USA
Latitude (N)	68.62	45.30	43.07	37.91
Longitude (E)	-149.30	5.77	-116.75	-107.71
Elevation (m)	930	1330	2060	3370
Study Period (WY)	2011	2006	2007	2008
Snow Climate	Tundra	Mountain (maritime)	Mountain (intermountain)	Mountain (continental)
Sensors	T_{air}: Vaisala HMP45C P: Campbell Scientific TE 525 U: Met One 014A RH: Vaisala HMP45C Q_{si}: Kipp & Zonen CMA 6 Q_{li}: none (taken as residual from measurements of all other radiation components[A])	T_{air}: PT 100/4 wires P: PG2000, GEONOR U: Chauvin Arnoux Tavid 87 – non-heated RH: Vaisala HMP 45D Q_{si}: Kipp & Zonen CM14 Q_{li}: Eppley PIR	T_{air}: Vaisala HMP 45 P: Belfort Universal Gages U: Met One 013/023 RH: Vaisala HMP 45 Q_{si}: Eppley Precision Spectral Pyranometer Q_{li}: Eppley PIR	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
Operators	NRCS, CRREL, Ameriflux	Météo-France	Northwest Watershed Research Center, Agricultural Research Service	Center for Snow and Avalanche Studies
Oct-Dec T_{air} (°C)	-16.1	2.0	0.2	-3.7
Jan-Mar T_{air} (°C)	-14.7	-1.6	-2.0	-8.7
Apr-Jun T_{air} (°C)	-1.4	8.9	8.4	2.7
Oct-Mar P[B] (mm)	200	690	480	1000
Mean annual U (m s^{-1})	2.2	1.0	1.6	1.1

[A] 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 (°C).

[B] Note that precipitation data were adjusted with a multiplier (see Section 2) prior to conducting the sensitivity analysis.

Table 2 UEB model parameters used across all simulations and sites

Description of parameter	**Units**	**Value**
Rain threshold temperature	°C	+3.0
Snow threshold temperature	°C	-1.0
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

Table 3 Details of error types, distributions, and uncertainty ranges for the five 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=10000)					
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][C]	%	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=10000)					
This scenario has six bias parameters (identical to NB above), plus the following six random error parameters					
$\boldsymbol{T_{air}}$	**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)
$\boldsymbol{P}$	**RE**	Lognormal	|[0.0, 25]|	%	Guan et al. (2005); Hasenauer et al. (2003); Hutchinson et al. (2009)
$\boldsymbol{U}$	**RE**	Normal	|[0.0, 5]|	m s^{-1}	Cheng and Georgakakos (2011); Liston and Elder (2006); Luo et al. (2008); Winstral et al. (2009)
$\boldsymbol{RH}$	**RE**	Normal	|[0.0, 15]|	%	Bohn et al. (2013); Liston and Elder (2006); Phillips and Marks (1996)
$\boldsymbol{Q_{si}}$	**RE**	Normal	|[0.0, 160]|	W m^{-2}	Hasenauer et al. (2003); Jepsen et al. (2012); Liston and Elder (2006); Thornton et al. (2000)
$\boldsymbol{Q_{li}}$	**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=10000)					
Identical to NB, except all probability distributions are **uniform**					
Scenario NB_gauge (k=6, N=10000)					
Identical to NB, except P uncertainty mimics documented differences between P and SWE at SNOTEL sites					
P	B	Lognormal	**[-10, +10]**	%	Meyer et al. (2012)
Scenario NB_lab[D] ***(k=6, N=10000)***					
T_{air}	B	Normal	**[-0.30, +0.30]**	°C	Vaisala HMP45 specified accuracy
P	B	Lognormal	**[-3.0, +3.0]**[E]	%	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. 5) 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 Sec. 3.3.5).
[C] The high upper P bias (300%) mimics cases where snowfall data collected in an area of drift deposition are assumed (incorrectly) to represent other basin locations.
[D] 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.
[E] We neglect P undercatch errors in the lab uncertainty scenario.

Table 4 Number of samples (N) and model simulations (in parentheses) meeting the requirements for minimum peak SWE and snow duration and valid snow disappearance dates at each site in each scenario. The number of model simulations scaled as Nx(k+2), where k=12 in scenario NB+RE and k=6 in all other scenarios. When a simulation was rejected, all related simulations (based on resampling) were also rejected.

	Scenario NB	**Scenario NB+RE**	**Scenario UB**	**Scenario NB_gauge**	**Scenario NB_lab**
IC	9898 (79 184)	10 000 (140 000)	8608 (68 864)	10 000 (80 000)	10 000 (80 000)
CDP	9792 (78 336)	9869 (138 166)	8925 (71 400)	9999 (79 992)	10 000 (80 000)
RME	8799 (70 392)	9233 (129 262)	9102 (72 816)	10 000 (80 000)	10 000 (80 000)
SASP	9984 (79 872)	9984 (139 776)	3399 (27 192)	10 000 (80 000)	10 000 (80 000)

~~7.~~8. Figures

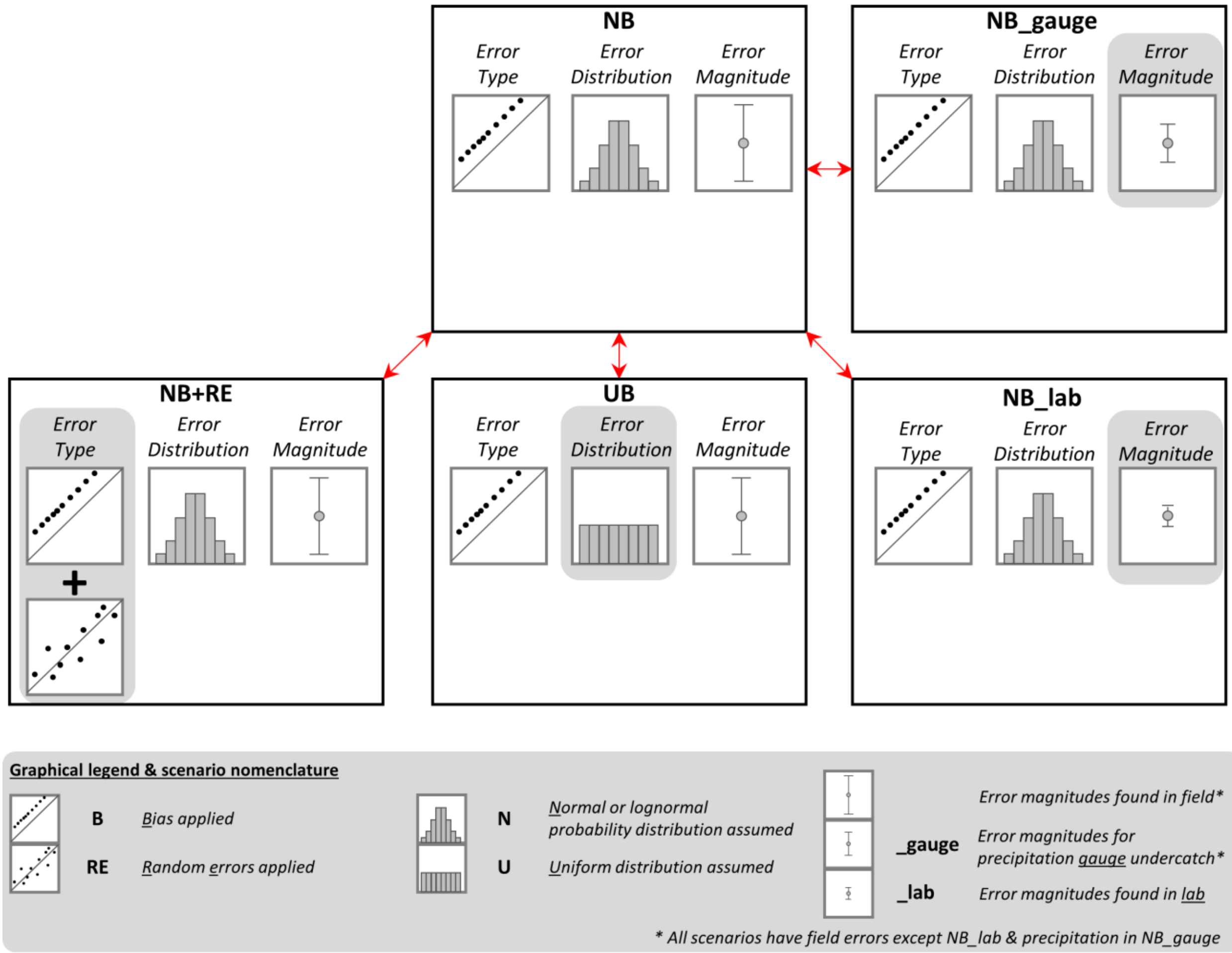

Figure 1 Scenarios of interest and the type, distribution, and magnitude of errors considered in each. NB considers normally (or lognormally) distributed biases with error magnitudes found in the field. NB+RE is the same as NB but also considers random errors. UB is the same as NB but considers uniformly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

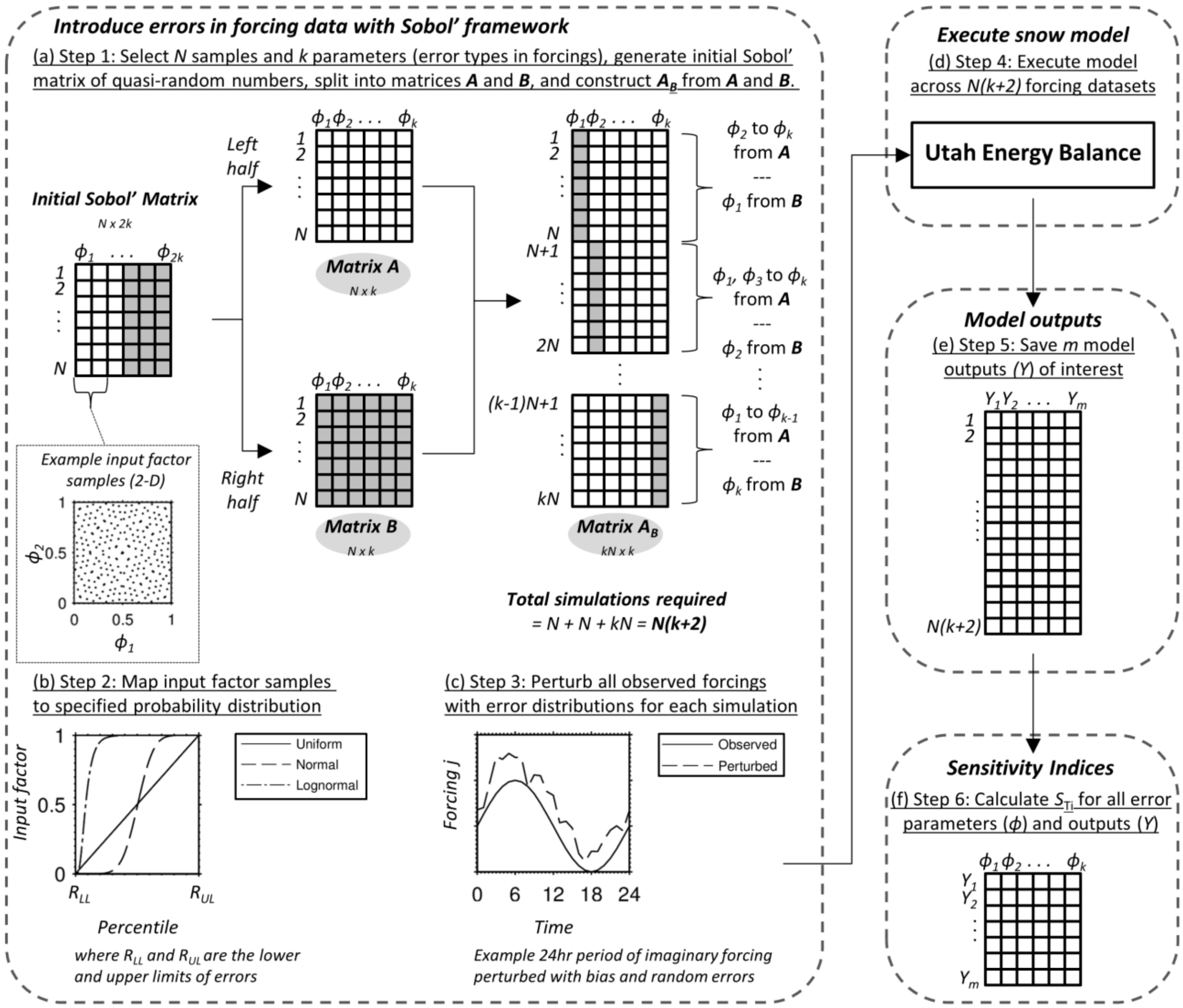

Figure 2 Conceptual diagram showing methodology for imposing errors on the forcings with error parameters (ϕ) within the Sobol' sensitivity analysis framework, and workflow for model execution and calculation of sensitivity indices on model outputs (Y).

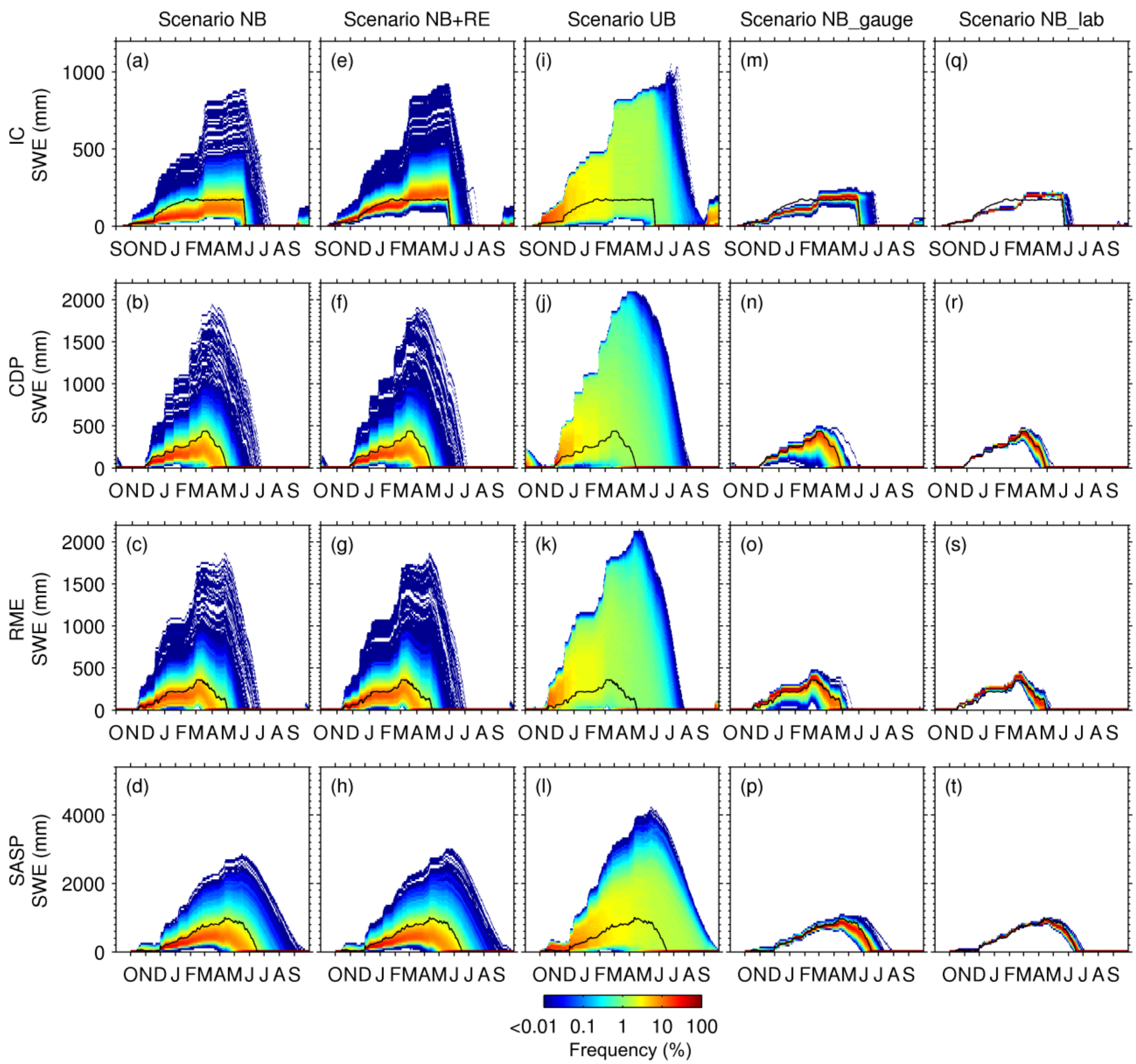

Figure 3 Observed (black line) and modeled SWE (color density plot) at the four sites across the five uncertainty scenarios (see Figure 1 and Table 3). The number of model simulations in the density plots varies with the site and scenario (see Table 4). The density plots were constructed using 100 bins 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), normally distributed bias with precipitation gauge uncertainty NB_gauge), and normally distributed bias at laboratory error magnitudes (NB_lab).

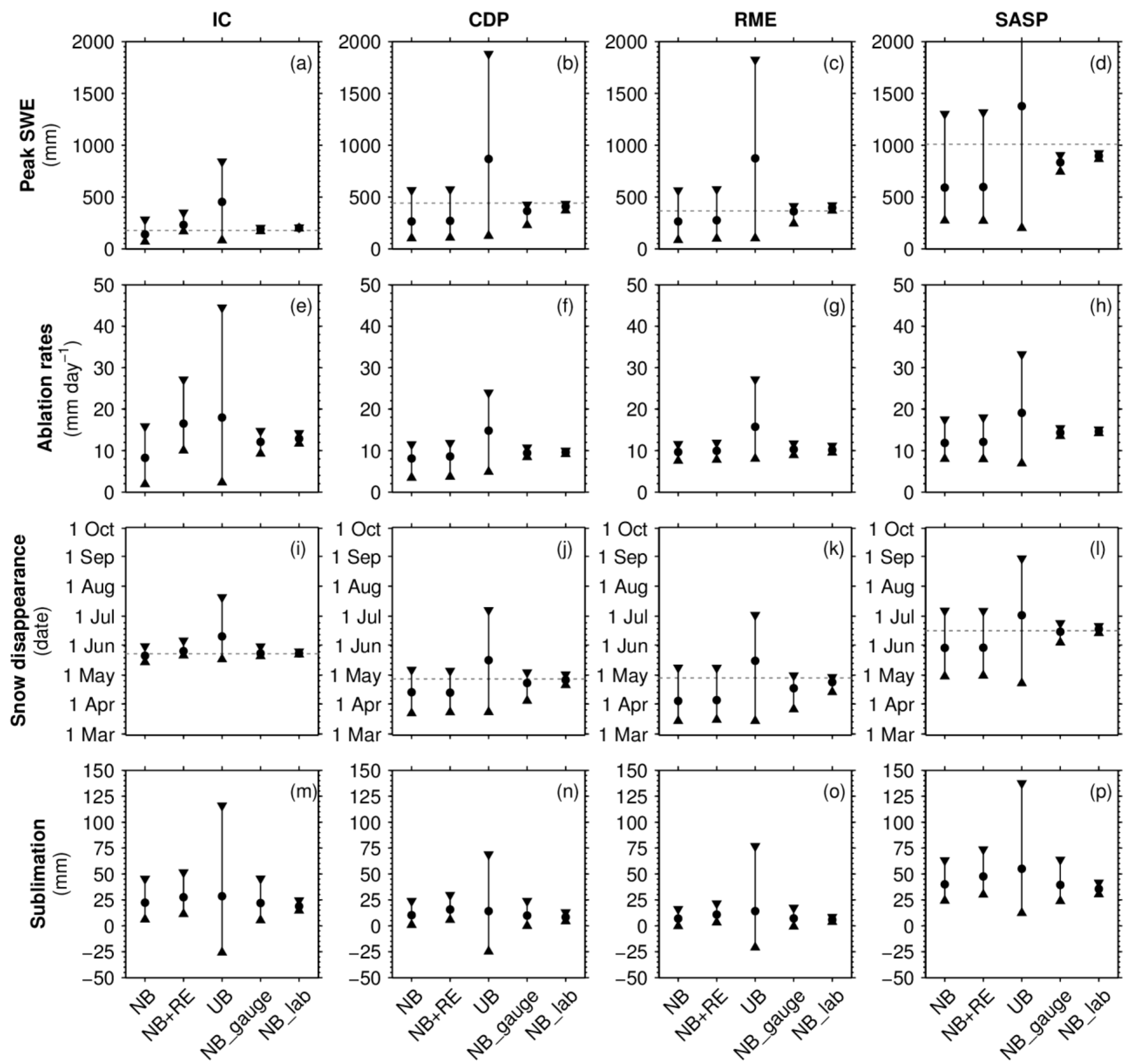

Figure 4 Distributions of model outputs (rows) at the four study sites (columns) arranged by scenario. For each scenario, the circle is the mean and the whiskers show the range encompassing 95% of the simulations (see Table 4 for number of simulations for each site and scenario). The dashed lines in (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), normally distributed bias with precipitation gauge uncertainty NB_gauge), and normally distributed bias at laboratory error magnitudes (NB_lab).

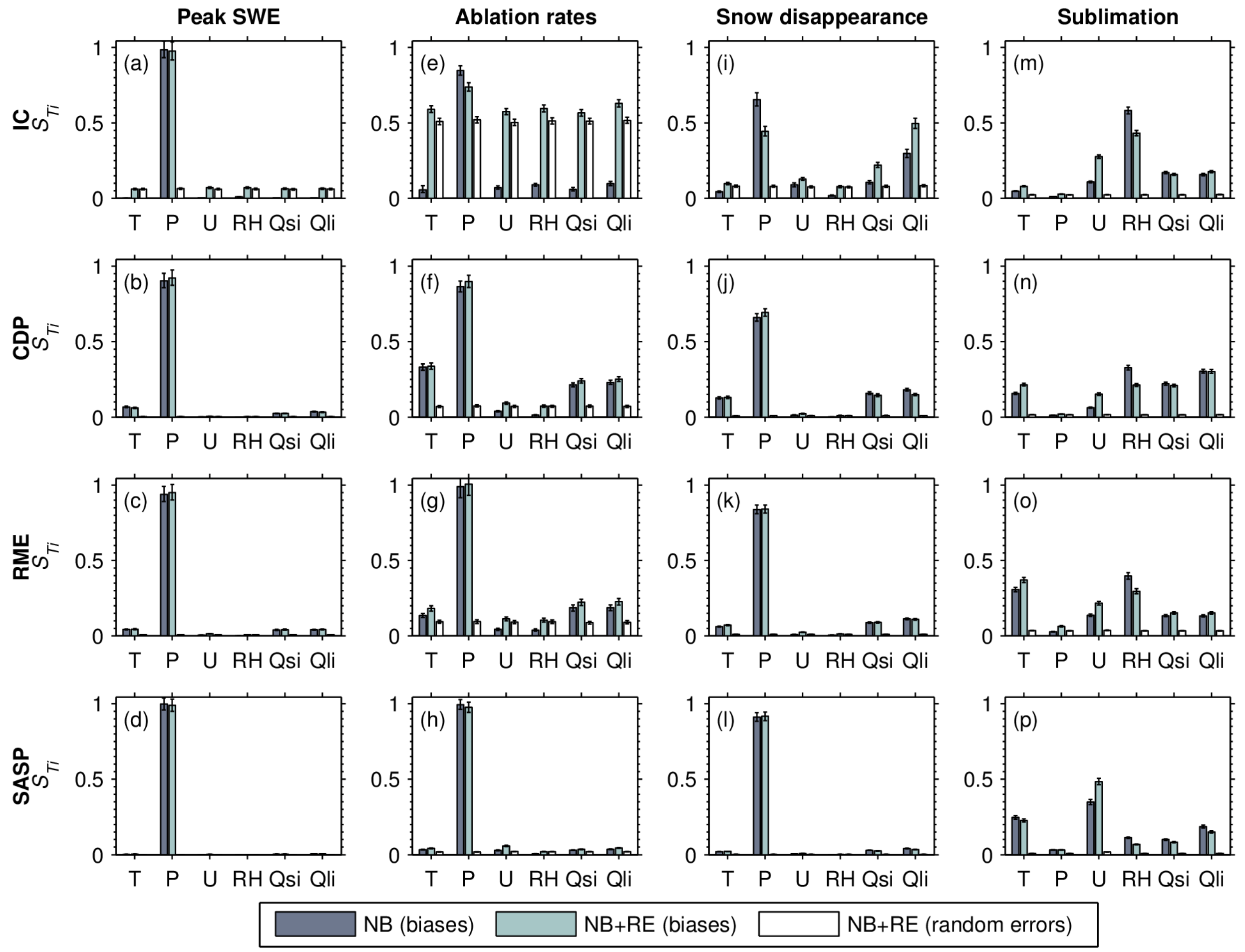

Figure 5 Model sensitivity as a function of forcing error type. Shown are the total-order sensitivity indices (S_{Ti}) of four model response variables (columns) at the four sites (rows) from scenarios NB and NB+RE. In NB+RE, bias and random error parameters are shown separately. NB+RE considers normally distributed bias and random errors, while NB considers normally distributed bias only. The bar indicates the mean (bootstrapped) sensitivity indices and associated 95% confidence intervals.

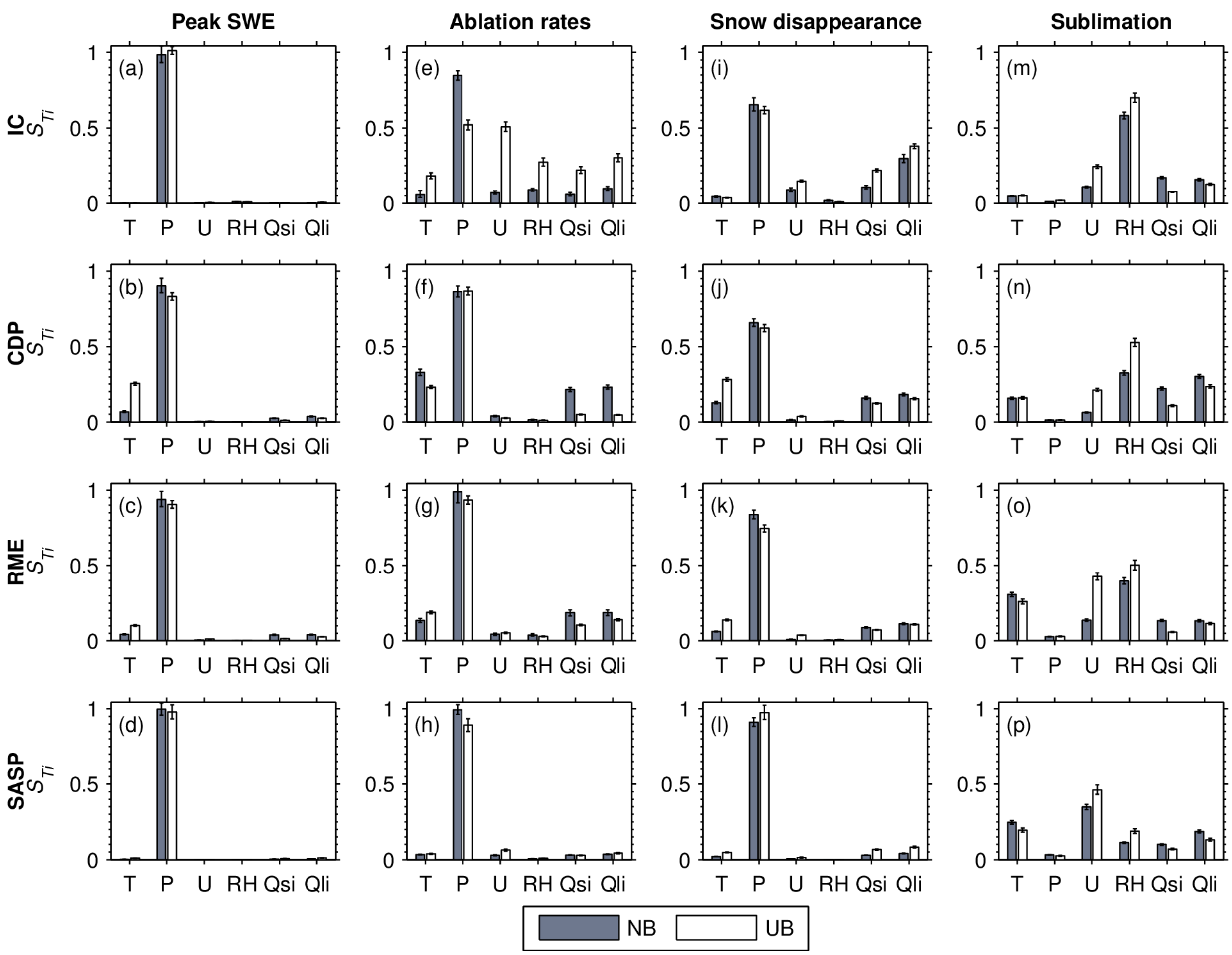

Figure 6 Same as Fig. 5, but comparing S_{Ti} values from scenarios NB and UB to test model sensitivity as a function of error distribution. UB considers uniformly distributed bias, while NB considers normally distributed bias.

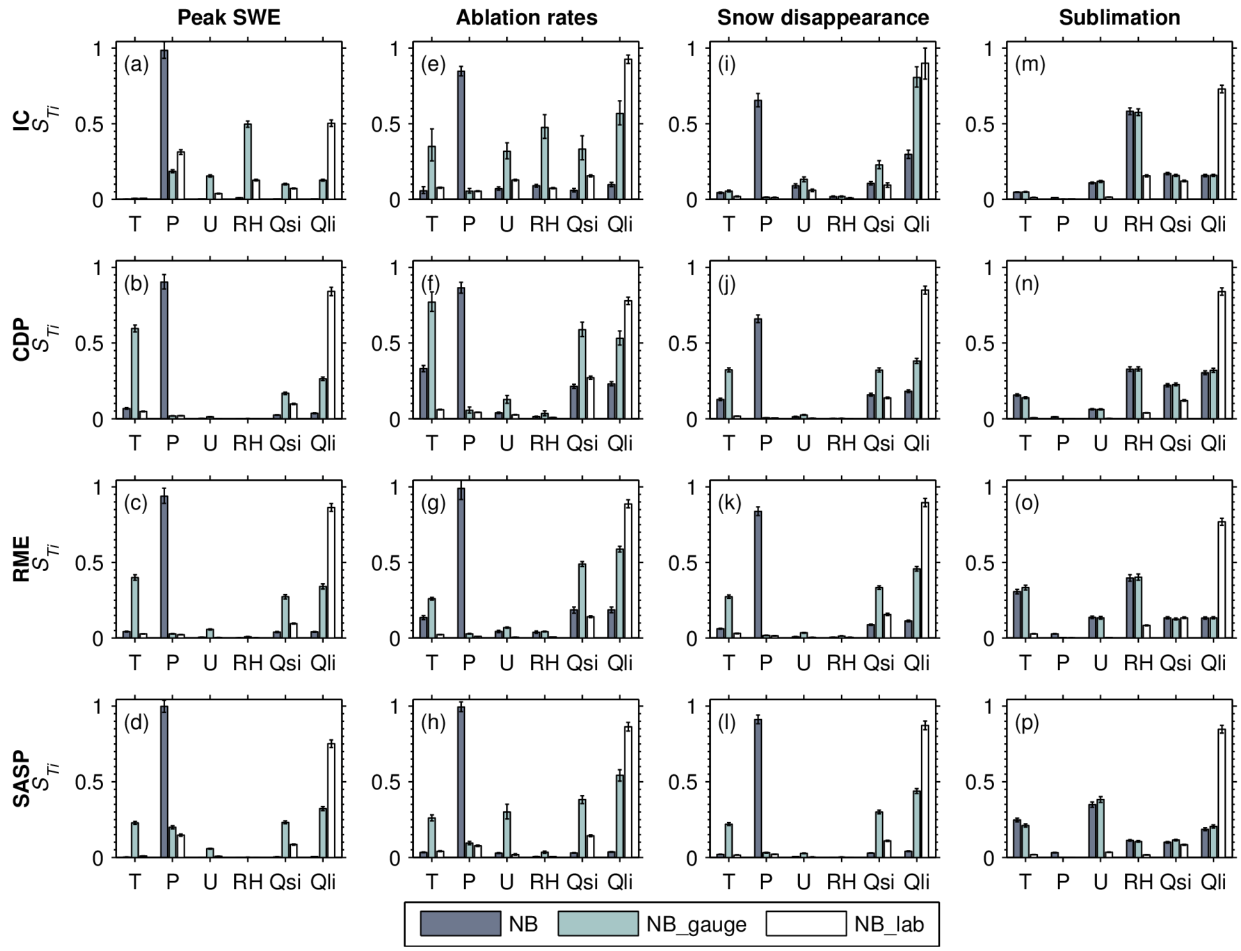

Figure 7 Same as Fig. 5, but comparing S_{Ti} values from scenarios NB, NB_gauge, and NB_lab to test model sensitivity as a function of error magnitudes. NB considers normally distributed bias at error magnitudes found in the field. NB_gauge has lower precipitation uncertainty (gauge undercatch) than NB but is otherwise identical. NB_lab considers normally distributed bias at error magnitudes found in the laboratory.

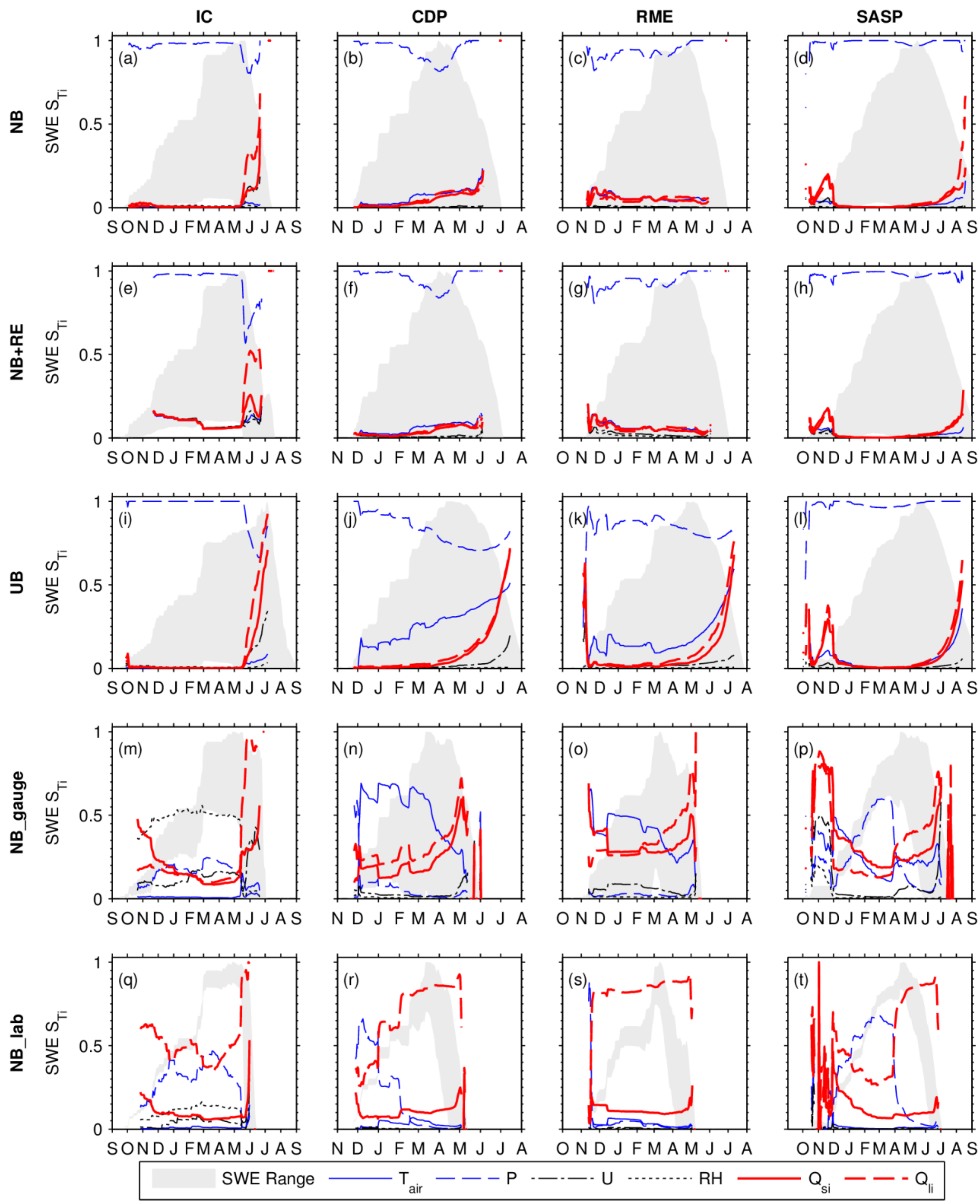

Figure 8 Variation of daily SWE sensitivity to forcing bias based on site (columns) and error scenario (rows). The normalized range (where 1 = maximum SWE) in modeled SWE is shown (gray area) for context. Sensitivity indices in the early and late part of the snow season were screened out, as a high number of simulations with SWE=0 yielded invalid sensitivity indices.

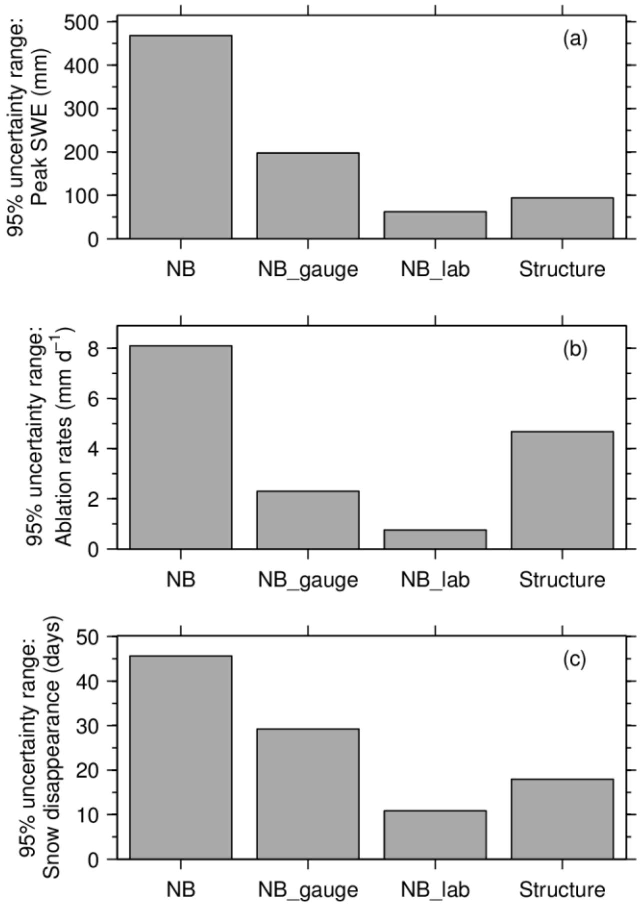

Figure 9 Uncertainty ranges (95% intervals) in (a) peak SWE, (b) ablation rates, and (c) snow disappearances date at CDP in WY2006 for three forcing uncertainty scenarios and the Essery et al. (2013) structural uncertainty.