Hydrol. Earth Syst. Sci. Discuss., 11, C6766–C6778, 2015
www.hydrol-earth-syst-sci-discuss.net/11/C6766/2015/

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework" *by* M. S. Raleigh et al.

M. S. Raleigh et al.

raleigh@ucar.edu

Received and published: 15 April 2015

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

Comment: *Overview This manuscript explores the relative effects of bias and error distributions on the Utah Energy Balance model's sensitivity across Peak SWE, Ablation Rates, Snow Disappearance, and Sublimation predictions. The work exploits detailed forcing observations at 4 seasonally snow covered sites: (1) the tundra Imnavait Creek in the Brooks Range in Alaska, (2) the Col de Porte site in the Chartreuse*

Range in France, (3) Reynolds Mountain in Idaho, and the Swamp Angle Study Plot in the San Juan mountains of Colorado. The core contention of the work is that forcing bias and errors could dominate structural and parametric uncertainties for snow-affected regions with strong observation limitations. Overall I found this hypothesis somewhat self-evident, although the overall study does highlight the importance of observation errors and uncertainties. I believe this manuscript requires revision to reach its full potential.

Response: We thank you for your careful review of the manuscript.

Major Comments

Comment: *1. Limited Analysis: The core results in Figures 5-11 are discussed with extreme brevity and little analysis. The authors have made the chose to provide a more detailed exposition in their Discussion but at present the Results do not even orient the reader very well across individual plots. Figures 5-8 are summarized in text that mixes results across figures and severely limited in its value. The question that emerges when reading this is that either the authors could compress their results into fewer and better designed figures or they could tease more model related insights in their analysis text.*

Response: This is a reasonable comment. Given the number of dimensions that we are examining (4 sites, 6-12 error parameters, 4 model outputs, and now 5 scenarios), we do not think it is feasible to compress the results into fewer or more efficient figures. Hence, we have opted to provide more context and explanation of the results in the text.

Manuscript Revisions: We now provide expanded description of the core results in sections 4.2-4.5, but reserve discussion of the results in section 5.

Comment: *2. Discussion Disconnected from Results: The most interesting portions of the discussion relate to the contention of the relative importance of structural uncertainty to forcing errors. Unfortunately, this text references other published work*

strongly and does a very poor job connecting to directly to the Results/Figures of this paper. Transitioning from Section 4 to Section 5 almost feels like your reading a different paper. Overall the structure and writing of the work varies significantly from the well written Introduction, the detailed Methods, and more detailed Discussion versus the extremely cursory Results.

Response: We can understand how this is problematic and agree that the exposition of these sections can be improved.

Manuscript Revisions: We have rewritten and reorganized some parts of the the discussion to provide better correspondence with the results and better connection to other published works. As an example of the latter, we have acquired the model results of Essery et al. (2013), which is referenced heavily in the early part of the discussion and now create a new figure (see Figure 1 in this document, below) that directly compares our results (due to forcing uncertainty) and Essery's results (due to structural uncertainty):

Comment: *3. It is unclear how generalizable the results are beyond this study: Many of the results are not very insightful and seem to convey a very place-based specificity for deviating cases. The reporting of sensitivities in the Results are not well articulated in terms of their dependency on site location, the nuances of the Utah Energy Balance model, and scenarios. In its present form, I am not convinced that manuscript provides insights and it may be conflating several factors that could influence the differences in sensitivity (model choice, site selection, scenarios). Explanation of the stronger results, such as distribution choice minimally impacts computed sensitivities, is limited and not compelling. The core of the Discussion section is the best overall text of the paper. It would have been far better to lead with your core hypotheses in the Results section and test them explicitly through the analysis of your results. The Discussion would then emphasize key caveats, insights, and implications.*

Response: While recognizing the importance of generalizing the results, we are hes-

itant to generalize relationships between site geo-characteristics/climate and sensitivities indices because of the relatively low number of sites represented (n=4 sites, 1 year each) and the confounding number of differences between our sites (e.g., snow climate, latitude, elevation, wind exposure/sheltering, etc.). We would require a much larger population of snow measurement sites in order to more robustly test relationships between sensitivity indices and site characteristics such as elevation and latitude. A successful example of relating climate characteristics to sensitivity can be found in van Werkhoven et al. (2008), which had 12 sites and 39 years each, making it possible to explore inter-site and inter-annual variations in climate and linkages to model sensitivity.

Manuscript Revisions: We now emphasize in Section 2 that we selected the four sites to check for climate dependencies, but are unable to generalize the results due to the low sample size. We note in the discussion however, that there are common results that emerge across all sites, such as the dominance of precipitation bias on SWE, ablation rates and snow disappearance (NB scenario) and longwave bias on all four outputs (NBlab scenario). This suggests that there may be common features in model sensitivity to forcing errors across distinct climates.

Minor Comments

Comment: *1. It would have been interesting to explore 2nd order and 1st order differences from the total indices in the results.*

Response: While we agree this would be interesting, we argue that this could make the study less focused and therefore elect to focus only on the total sensitivity indices. The total sensitivity indices provide a summative measure of both first-order and interaction effects and therefore convey the overall importance in a straightforward manner. Calculation of the second-order terms would require nearly double the number of simulations (compare n(2k+2) vs. n(k+2) in the current analysis) (Saltelli, 2002), and hence we have not pursued this extended analysis due to the additional computational ex-

penses required.

While we do not present them in the manuscript, we can calculate the first-order indices with the existing model simulations. The comparison of the first- and total-order indices provides insights into how much of the variance is due to direct effects vs. interactions, and broad justification for only reporting one type of sensitivity indices. Figures 2 and 3 (this document, below) show the first- (Si) and total-order (STi) indices for the NB and NB+RE scenarios. From these figures, it is evident that in many cases, the sensitivity is dominated by first-order effects, as suggested by the close alignment of Si and STi values. There are cases however when the interactions have greater importance (e.g., factors of secondary importance for the ablation rates). The general correspondence between the first- and total-order indices suggests to us that most of the story is captured with just a single index; hence, we focus on just total-order sensitivity indices for simplicity/clarity.

Manuscript Revisions: We have made no changes to the analysis, but we now comment in section 3.3.2, "First-order and higher-order sensitivities can be resolved; here, only the total-order sensitivities are examined (see below) for clarity and because the first-order sensitivity indices were typically comparable to the total-order sensitivity indices."

Comment: *2. A better explanation of the scales assumed in the measures used to report sensitivities and caveats as to what cannot capture would be helpful.*

Response: We assume you are referring to numerical scales in this comment, and can comment on this in the text.

Manuscript Revisions: In section 3.3.3, we explain that interpretation of the total sensitivity indices is straightforward because they represent the fraction of output variance due to a specific factor, and state that these indices scale from 0 to 1. We now include a caveat that the Sobol' total sensitivity indices cannot account for the case of correlated errors (section 3.3.2), which may occur in the real-world.

Comment: *3. Very little treatment is provided for the convergence rates of the total order indices and their associated bootstrap intervals as a function of your sampling.*

Response: The reviewer is correct that we did not provide much information on convergence rates. Figure 4 (this document, below) shows the time history and convergence of the total sensitivity indices (as a function of sample size) for Scenario NB (other scenarios exhibited similar levels of convergence). Examining the figure, it is evident that the same conclusions for the study (at least qualitatively) could have been drawn with fewer simulations. A dynamic system of calculating sensitivity indices as model completes simulations would optimize the analysis by stopping the process once convergence has been reached, but such a system was not implemented here.

We can quantitatively assess the level of convergence by examining the ratio of the 95% confidence interval (from the bootstrapping procedure) to the mean STi values. Figure 5 (this document, below) shows this ratio (as a percentage) for the error parameter with the highest STi for each model output, scenario, and site. If we assume convergence has been reached when the ratio is less than 10% (based on Herman et al., 2013), then we can see that the majority cases in our study reached convergence, and only three out of 64 cases had a ratio greater than 15

Manuscript Revisions: We now provide some description of the convergence rates and on the bootstrap confidence intervals (sections 4.2 and 5), but do not provide any additional figures in the manuscript.

Comment: *4. How stable and/or separable are the factor prioritization rankings? What results have higher confidence?*

Response: The 95% confidence intervals (from the bootstrapping procedure) are already shown in Figures 5-8 in the original manuscript, and these provide a measure of our confidence in the rankings. The difference between the bootstrap mean and the final mean Sti values also provides a measure of stability.

Manuscript Revisions: We note in section 3.3.4: "For all cases, final STi values 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."

Comment: *5. It would improve the manuscript to better understand the justification of the ranges tested in the Sobol analysis. Would a slight change in your a priori ranges change factor rankings?*

Response: The original manuscript outlines the justification for the ranges, but we can provide more information in the methods section. While we did not test for "slight changes" in the a priori ranges, we know that more substantial changes in these ranges can change the hierarchy of factors. Our original results already suggest that a change in the error ranges will change the rankings of factors (compare NB to NBlab, where the only difference is field vs. laboratory levels of uncertainty). We also now include the new scenario (identical to NB but with lower precipitation error ranges to reflect gauge undercatch), and find again that the factor ranges do change with the a priori ranges in the forcing uncertainty.

Manuscript Revisions: We now expand on our justification of the error ranges (section 3.2.3). Additional treatment of this topic is included in the discussion.

REFERENCES

Essery, R., S. Morin, Y. Lejeune, and C. B Ménard, 2013: 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.

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

Saltelli, A., 2002: Making best use of model evaluations to compute sensitivity indices.

Comput. Phys. Commun., 145, 280–297, doi:10.1016/S0010-4655(02)00280-1.

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

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 11, 13745, 2014.

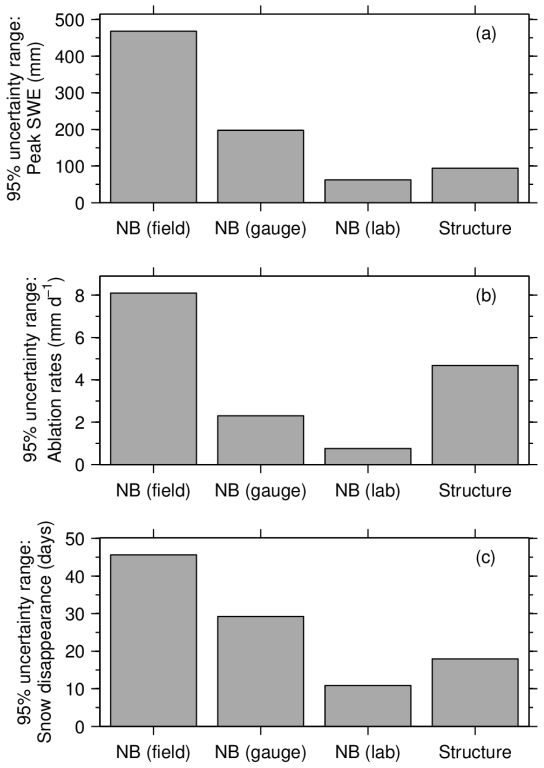

Fig. 1. 95% intervals in (a) peak SWE, (b) ablation rates, and (c) snow disappearances date at CDP in WY2006 for three forcing uncertainty scenarios and the Essery et al. (2013) structural uncertainty.

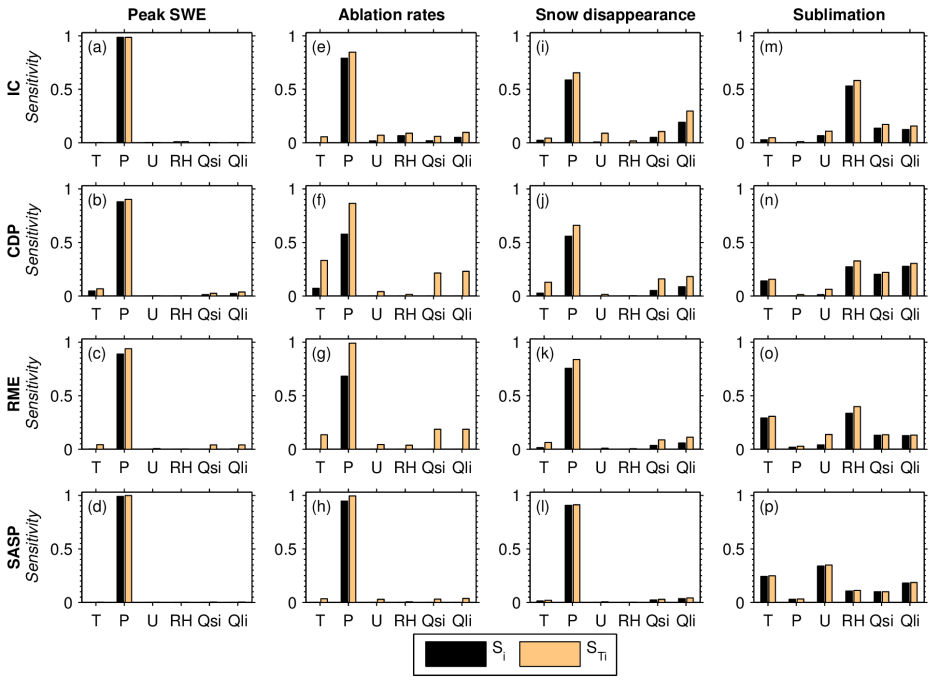

Fig. 2. First order (Si) and total-order (STi) sensitivity indices for bias factors in the NB scenario.

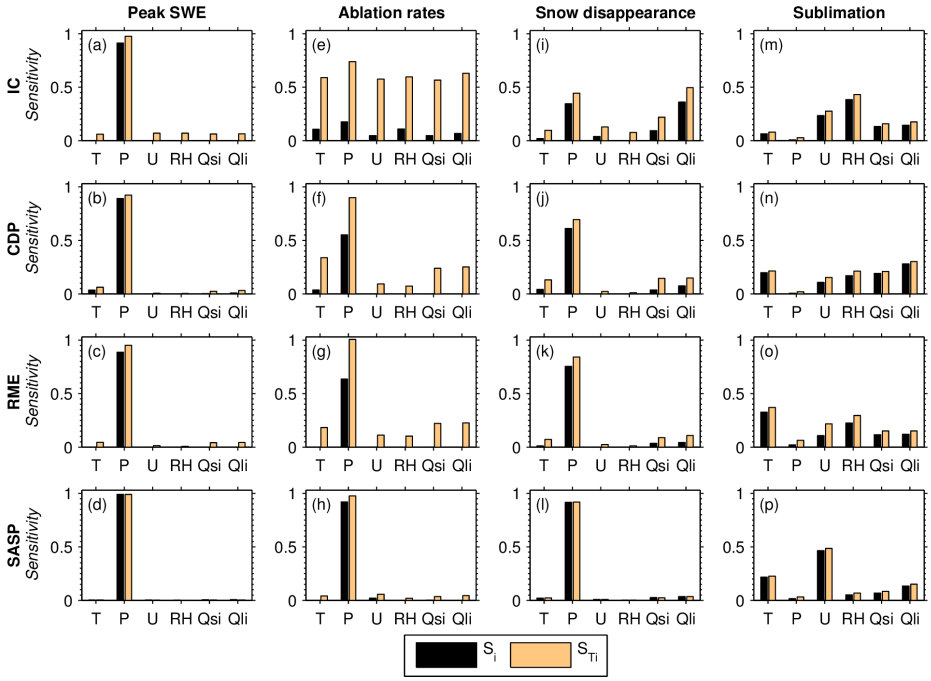

Fig. 3. First order (Si) and total-order (STi) sensitivity indices for bias factors in the NB+RE scenario.

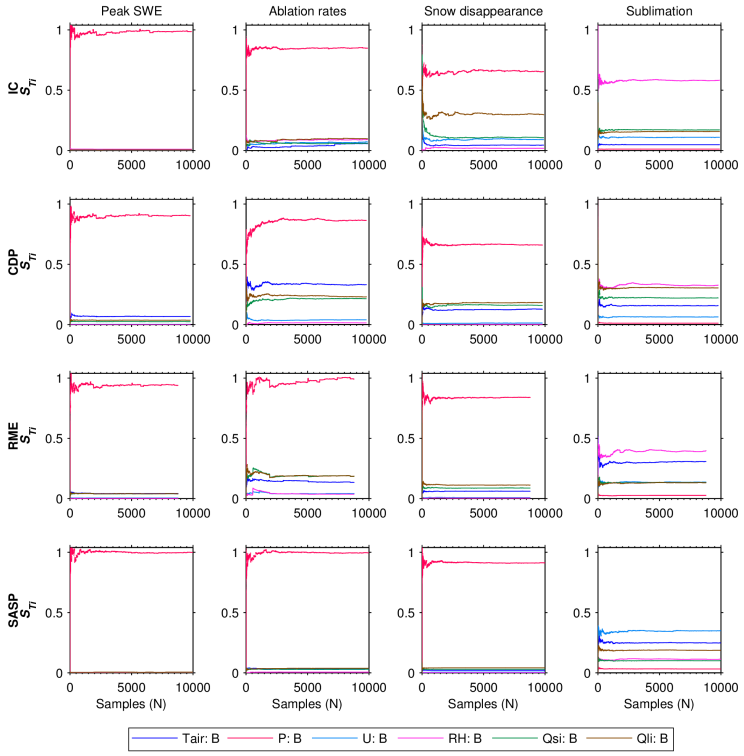

Fig. 4. Convergence of total-order sensitivity indices in scenario NB for the four model outputs at the four sites, as a function of sample size.

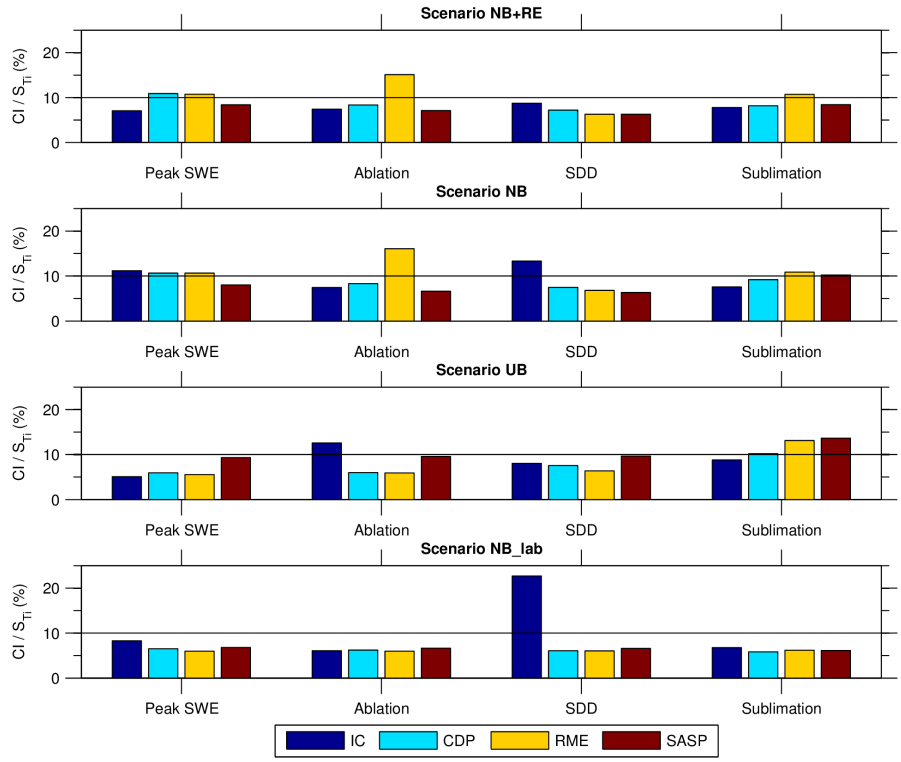

Fig. 5. Ratio of 95% confidence intervals to the bootstrapped mean total-order sensitivity indices (%) for the most important factor for each scenario, site, and model output.