Hydrol. Earth Syst. Sci. Discuss., 11, C6779–C6792, 2015
www.hydrol-earth-syst-sci-discuss.net/11/C6779/2015/

Hydrology and Earth System Sciences
Open Access
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: *Raleigh et al. present an interesting attempt to systematically determine how uncertainty in forcing data influences uncertainty in snow simulations. Some of their conclusions seem quite obvious (biases are more significant than random errors, and uncertainty in measured precipitation is the most important factor), but a large effort is often required to demonstrate things that appear obvious with hindsight in*

hydrological modelling. The increased sensitivity to biases when random errors are introduced is a striking result, however, and there should be more exploration of how this arises.

Response: Thank you for your interest and your review of the manuscript. We have revisited the striking result of enhanced sensitivity to biases when random errors are introduced, and found this to actually be related to a deficiency in the introduction of random errors into the system (eq. 4). Specifically, we discovered that the random number generator (randn.m in Matlab) used to create the "noise" (i.e. random errors) did not always have a mean of 0 (though it was a value close to 0). This is because it is a discrete array with samples drawn from a population of mean 0; hence the sample mean is not guaranteed to be 0. Because of a non-zero mean in the noise, the "random error" term also introduced additional systematic errors that were not accounted for in the bias terms.

Manuscript Revisions: We have corrected this coding issue, reran NB+RE and have found that this minimized the problem you have found (see Figure 1 in this document, below). We find that the sensitivity to biases (after introducing random errors) is less pronounced in this case for most outputs/sites considered. The most obvious outlier is for ablation rates at IC, where there is heightened sensitivity to biases after random errors are introduced. In this case, the total-sensitivity indices are amplified because of more interactions in the system (e.g., first-order sensitivities were small compared to the total-order indices). We surmise that the relatively short ablation season at IC (on order of 10-20 days) is a critical reason why there is enhanced sensitivity across all error types; errors in a variety of factors can yield large impacts on ablation rates during the brief melt period.

Comment: *I am slightly concerned about how the error distributions have been assigned. It is variances in model outputs that are examined but ranges in model inputs that are specified. The variance of a uniform distribution is larger than a normal distribution over the same range, so these scenarios are not really comparable.*

Response: These concerns are reasonable, and it is true that the variance of a uniform distribution is larger than the variance of a normal distribution over the same range. That is part of the purpose of this particular experiment, namely, to examine how the assumed probability distribution of errors influences model sensitivity. It is by design that the ranges are made the same for the uniform and normal distributions for a given forcing; this allows us to test in a controlled fashion whether/how more frequently occurring extreme errors (in the uniform distribution) change model sensitivity. If we had not matched the ranges in the two distributions, then there would be two confounding reasons why the distributions were different (probability distribution shape and range), and we wished to isolate the differences due to shape only.

You are correct that we could have alternatively constructed the experiment such that we specified variance (or standard deviation) instead of the range. However, we constructed the experiment with range instead of variance because it was more straightforward and provided a more direct approach to encompass all magnitudes of errors found in our literature review for different forcing observation/estimation approaches. We also note that most sensitivity analyses use uniform distributions (e.g., Nossent et al., 2011; Peeters et al., 2014), which are specified by the range and not the variance. In considering normal distributions, we are extending these methods to other types of distributions.

Regardless of whether we base our distributions on range or variance, we note that there is "uncertainty in uncertainty." In other words, we are not always certain about what the spread of uncertainty should be. Our understanding of the spread of uncertainty is poor due to the relatively low sample size of papers that report error statistics for different forcings. It can be shown experimentally (with a Monte Carlo sampling experiment) that for low sample sizes (n<150), we have higher confidence in the range of a given normal distribution than in the variance. Figure 2 (this document, below) shows the relative uncertainty in range and variance derived from such a Monte Carlo experiment. Given that there are few papers that systematically assess forcing errors

in mountainous areas, we argue that it is not necessarily a bad idea to work in terms of range because we have comparatively higher confidence in the range than in the variance. For example, if we only have 10 papers that specify the mean bias of shortwave radiation, the confidence interval (CI) for our variance estimate (of the probability distribution) is about 80% greater (in a relative sense) than the CI for the range of the distribution (see Figure 2 below).

Manuscript Revisions: We clarify in section 3 why we prescribed the probability distributions in this manner.

Comment: *It is not clear, in any case, how the ranges given in Table 3 determine the means and variances that would characterize the normal and lognormal distributions. Can this be clarified?*

Response: Thank you for pointing this out. Yes, this can be clarified.

Manuscript Revisions: We now clarify in detail in section 3.3.5 how the normal and lognormal distributions are constructed based on specified characteristics.

Comment: *Uncertainty in measurements of snowfall is certainly a major concern, but the upper bound chosen for precipitation biases in the model forcing (+300%) is enormous – much bigger than the stated error of less than 20% of peak SWE for most SNOTEL sites. The reason given for choosing this large uncertainty is to represent areas with drifting snow, but I would argue that the neglect of drifting snow is a missing process in the model, not an uncertainty in its forcing.*

Response: We partially agree with you on this point. We think that the model scale and the process scale are important considerations in how we categorize the uncertainty due to drifting snow. For the case when the 1-d model is applied at a model element resolution that is greater than the process scale, we would classify the wind drift uncertainty as structural uncertainty in terms of sub-grid variability accounting. However, when the model element resolution is less than the process scale of drifting snow, it is impossible to account for snowdrift processes within the structural uncertainty because the model is applied independently of neighboring locations (i.e., no lateral snow mass transfer, by definition of a 1-d model). In this case, we argue that the drift uncertainty is somewhat ambiguous for the 1-d UEB model but still must be accounted for in either the parametric or the forcing uncertainty. We argue that drift uncertainty is analogous to the precipitation undercatch uncertainty (both are cases of wind-affecting precipitation), and therefore we treat the drift uncertainty as a source of forcing uncertainty for our 1-d model.

For models of higher dimensionality (e.g., 3-d), then we agree with your point. A 3-d snow model should account for lateral mass transfer via snow drifting. In this case, it is clear that large uncertainties in snow accumulation arise due to omission of the snowdrift processes in the model, and this is a case where the uncertainty is attributed to structural (and parametric) uncertainty.

While we make this point, we share your interest in how the study would have been different if we had "standard SNOTEL precipitation errors" as the upper limit of uncertainty in precipitation. To that end, we introduced a new scenario (NBgauge) that repeated Scenario NB with all factors the same except we changed the ranges of precipitation bias of [-10% to +10%]. Figures 3 and 4 (this document, below) show the new sensitivity indices for these ranges with lower precipitation, and compare these new indices to scenario NB. When we consider this lower range in precipitation uncertainty, we find that precipitation bias is never a major factor for these four outputs at the four sites, and other dominant factors emerge. At IC, longwave bias emerges as the most important factor for ablation rates and snow disappearance while humidity bias matters most for peak SWE and sublimation. At the other sites, biases in shortwave and longwave radiation and air temperature are most important for peak SWE, ablation rates, and snow disappearance. Humidity bias is an important factor for sublimation at IC, CDP, and RME, while wind bias is important to sublimation at SASP.

Manuscript Revisions: We now introduce the new error scenario "NBgauge" in the

analysis and have updated the manuscript text to introduce this scenario and report/discuss the results.

Specific comments:

Comment: *page 13749, line 25 SWE is measured at Col de Porte using a cosmic ray detector, not a snow pillow.*

Response: Thank you for catching this mistake.

Manuscript Revisions: We have corrected the sentence.

Comment: *13750, 5 How was reasonable representation of observed SWE judged?*

Manuscript Revisions: We now state in section 2, "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."

Comment: *13751, 18 It could be made clear at this stage that normal distributions are used for additive errors and lognormal distribution for multiplicative errors.*

Manuscript Revisions: Done.

Comment: *13752, 3 In contrast, data assimilation techniques often address random errors that are assumed to be unbiased.*

Manuscript Revisions: Thank you for pointing this out – we now note this in the sentence.

Comment: *13755, 1 Overview of what?*

Manuscript Revisions: We now change the title of this subsection to "Overview: model conceptualization and sensitivity".

Comment: *13756, 12 "due to bias in forcing"*

Manuscript Revisions: Done.

Comment: *13757, 15 "For all cases, final STi values were generally close ..." sounds a little contradictory; were they all close, or generally close?*

Manuscript Revisions: We now rephrase this sentence to be more quantitative: "For all cases, final STi values were generally 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: *13758, 20 Non-physical values would be less common if multiplicative perturbations were applied to all forcing variables that cannot be negative, not just precipitation.*

Response: This is a valid point, but we are attempting to follow typical error reporting conventions and to provide easy interpretation of errors. For example, it is often the case that radiation errors are reported in the literature in an additive context (e.g., +35 W m-2) and not in a multiplicative context (e.g., +10%). In the case of radiation, a multiplicative error (e.g., +/-10%) is not straightforward to interpret because the magnitude of the error will change with seasonality (e.g., 10% error in winter shortwave radiation is much less than 10% error in summer shortwave). Additionally, some errors only make sense in an additive context (e.g., temperature errors). Our treatment of errors reflects common practices in the literature to make it more easily understood by the community.

Manuscript Revisions: We clarify in section 3.3.5 why we prescribed multiplicative vs. additive errors.

Comment: *13761, 8 Can differences in which variables are of secondary importance be linked to differences in climate at the sites?*

Response: The links with climate in these secondary variables are not always clear to us. At the warm maritime CDP site in scenario NB, it makes sense that Tair bias is important to peak SWE, as it helps control the partitioning of rain and snow. In contrast, it is not clear why Qsi and Qli biases are of secondary importance for sublimation at IC

and CDP but not at RME and SASP (where Tair bias is the second most importance factor).

While there may be interesting climate linkages, we note that we are hesitant to generalize relationships between site geo-characteristics/climates 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. 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.

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.

Comment: *13762, 3 It is not so surprising that Qli biases are more important than Qsi biases because of the high albedo of snow.*

Response: We agree with you here. We also note that given how the literature often emphasizes the importance of net shortwave over all other terms for snowmelt (e.g., Bales et al., 2006), this could be considered a surprising result.

Manuscript Revisions: We have rephrased this sentence (section 4.2) to provide a more physically based explanation of what is happening here: “However, the albedo of snow minimizes the amount of energy transmitted to the snowpack from Qsi, thereby rendering Qsi errors less important that Qli errors. Additionally, the non-linear nature of the model may enhance the role of Qli through interactions with other factors.”

Comment: *13765, 11 Please consider doi:10.1029/2010EO450004*

Response: Thank you for making us aware of this article and for helping us to see the problem with using parentheses to indicate the opposite meaning.

Manuscript Revisions: We have reworded the sentence to avoid this issue and have ensured that there are no other instances of this convention in the manuscript.

Comment: *13766, 2 Note that “constraining P uncertainty in snow-affected catchments” is the aim of WMO-SPICE http://www.rap.ucar.edu/projects/SPICE/*

Manuscript Revisions: We now state: “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/).”

Comment: *13766, 10 Probabilistic forcing is a common and long-standing approach in data assimilation*

Manuscript Revisions: We now note: “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.”

Comment: *The forcing error scenarios are described in Figure 1, Table 3 and section 3.2. Is the figure really necessary?*

Response: We considered removing Figure 1, but other reviewers thought this figure was helpful in summarizing the scenarios, and so we have left it in the manuscript. We have also expanded Figure 1 to include the new NBgauge scenario, which was added to address your concerns about the level of precipitation uncertainty.

Manuscript Revisions: None taken here.

REFERENCES

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

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

Peeters, L. J. M., G. M. Podger, T. Smith, T. Pickett, R. H. Bark, and S. M. Cuddy, 2014: Robust global sensitivity analysis of a river management model to assess nonlinear and interaction effects. Hydrol. Earth Syst. Sci., 18, 3777–3785, doi:10.5194/hess-18-3777-2014.

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.

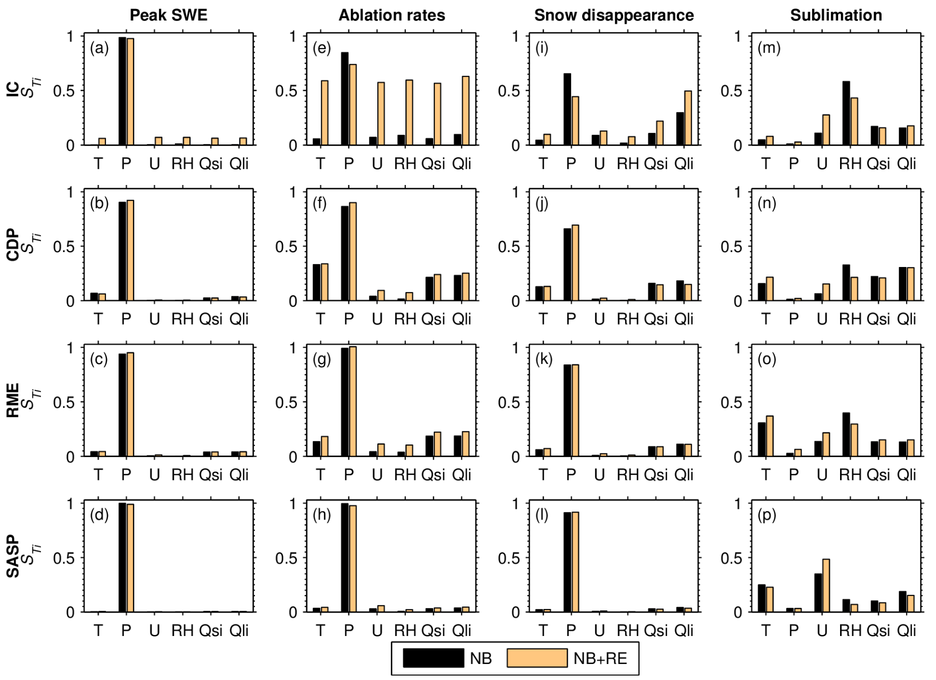

Fig. 1. Comparison of total sensitivity indices for the six bias factors in the NB and NB+RE scenarios after fixing the code to force random errors to have mean=0.

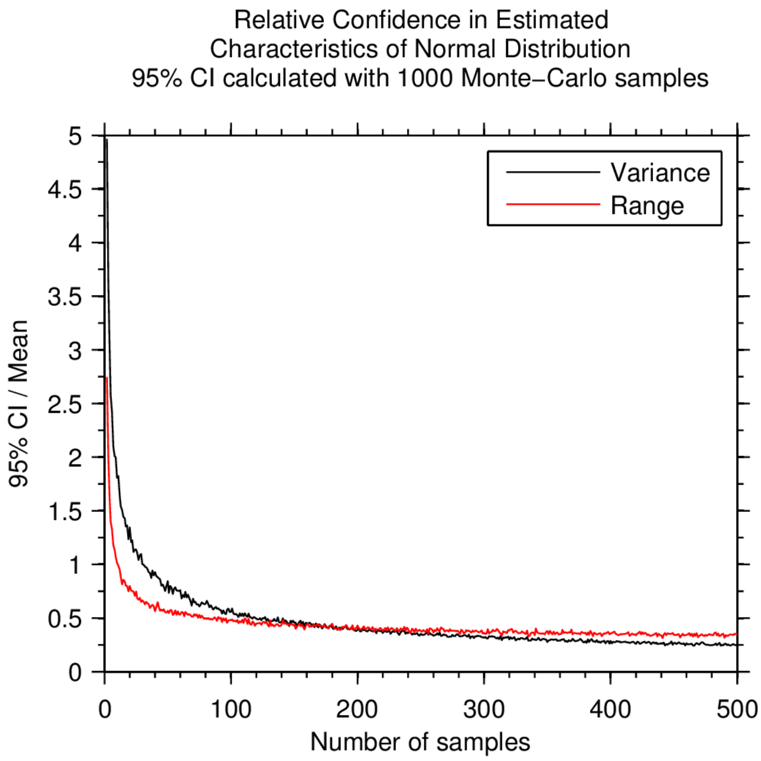

Fig. 2. Confidence in the variance and range of a normal distribution determined with Monte-Carlo sampling (n=1000) with a random dataset (106 samples, mean=0, variance=1) as a function of sample size.

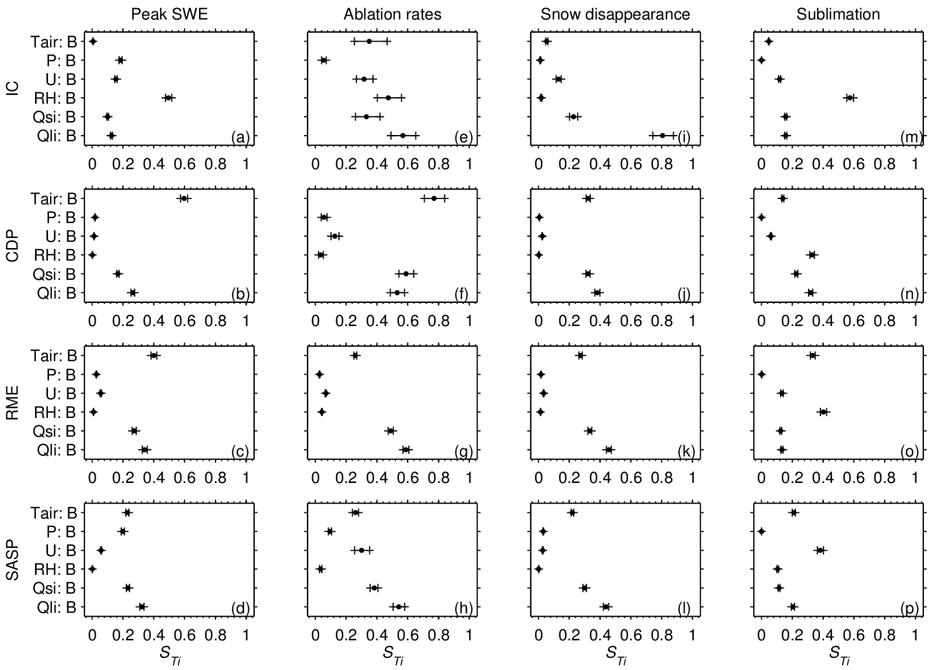

Fig. 3. Total sensitivity indices for the new NBgauge scenario. This is identical to NB except the precipitation uncertainty is constrained to the [-10% +10%] range.

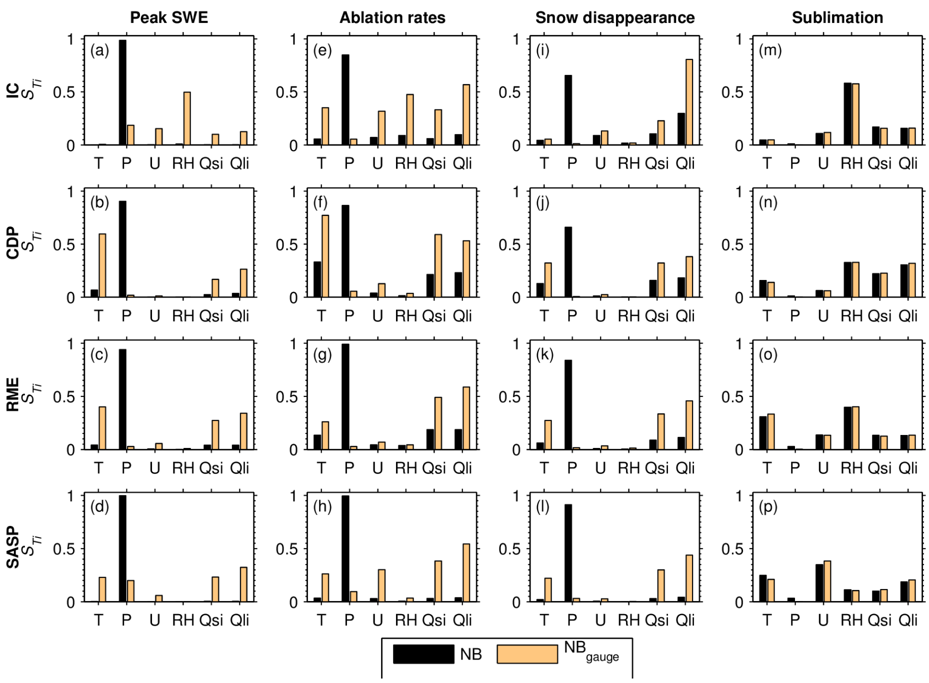

Fig. 4. Comparison of total sensitivity indices for the NB and NBgauge scenarios.