Hydrol. Earth Syst. Sci. Discuss., 11, C5846–C5851, 2014
www.hydrol-earth-syst-sci-discuss.net/11/C5846/2014/

Hydrology and Earth System Sciences Discussions
Open Access

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.

F. Pianosi (Referee)

francesca.pianosi@bristol.ac.uk

Received and published: 30 December 2014

This paper investigates how errors in meteorological observations affect the simulations of a physically based one-dimensional snow model (the Utah Energy Balance). Global sensitivity analysis (GSA) is used to quantify the relative contribution of different error characteristics (bias, magnitude, presence of random errors, error distribution) to the uncertainty in four snow variables (SWE, ablation rates, snow disappearance and sublimation). GSA results are presented for four study sites in distinct snow climate.

Detailed studies focusing on forcing uncertainty are relatively few, and they are needed particularly in snow-affected watersheds where meteorological measurements are scarce and forcing uncertainty can significantly impact model simulation. This work provides useful insights on the topic and establish a methodology that could be extended to other physically based models or error types.

I think the analysis here described is interesting and solid, the paper is clear and well-structured, and its contribution is well placed in the literature. I have some concerns about the reliability and interpretation of some of the GSA results, and a number of specific comments that the authors may consider in revising their manuscript. I think the paper should be considered for publication on HESS after such revisions.

1) Some of the results in Figure 6 and 7 are a bit surprising and need clarification.
For instance, in the cases of Fig. 5.a and 5.e, bias in P is the only influential parameter. However, when including random errors (Fig. 6.a and 6.e), all parameters become (almost equally) influential. In the text, this is explained as being due to interactions between parameters. I agree in principle but I think a more detailed analysis is needed. For instance, do bias parameters $\theta_{B,i}$ become influential through interactions with parameter $\theta_{RE,i}$ of the same meteorological variable? Or does this happen through interactions with $\theta_{RE,i}$ of different forcings (for instance, bias $\theta_{B,i}$ of T_{air} interacting with random error magnitude $\theta_{RE,i}$ of P)? I guess the physical interpretation of the result and its implications would be very different in the two cases. For instance, if the interactions occur within the same forcing error equation, it would mean that the bias in the observations is not influential per se, but it becomes influential if there are also random errors. Does this make sense from the physical point of view? Or is it a result of some inadequacy in the error structure of Eq. (4)?
Also, in all sites and for all outputs, the sensitivity indices of $\theta_{RE,i}$ are almost the same for all i. This is strange. Does it make sense that errors in all meteorological variables

have the same importance, or is there a purely numerical explanation for this?

2) I am not sure that Figure 9, 10, 11 are the most effective way to compare GSA results.
The main conclusion drawn in the text is that overall GSA results are similar across scenarios NB, NB+RE and UB. Scatter plot visually confirm this. However, they do not facilitate one-to-one comparison of sensitivity indices (bar plots with two coloured bars would be better), which in my opinion would provide more interesting information. For instance, comparing Fig. 5.o with 7.o I can see a big increase in the influence of U bias when moving from scenario NB to UB; comparing Fig. 5.e with 7.e shows that in the NB scenario only P bias is important, while in the UB scenario the bias of other meteorological variables also matter. Can you explain these behaviours? Maybe an interpretation effort of these results might lead to learning important aspects of the model behaviour.

3) Motivation of the study (in both the abstract and the introduction). I would add some comments on how the authors think that GSA results (which error characteristic matter most) could be used in practice. What are the implications of these results? How would you expect to use this piece of information? I think one way to use GSA results is to spot unexpected behaviours and thus have directions for further investigation of simulation results. However, I feel that this is somehow missing in the paper (see also my previous comment).

SPECIFIC COMMENTS

- page 13755: "The goal of sensitivity analysis is to quantify how variance in specific input factors (...) influences variance in specific outputs". This sentence is inaccurate. First, the use of output variance as a proxy of output uncertainty is a specific assumption of variance-based SA (Sobol') and it is not a general assumption of GSA. Many other GSA methods are available that do not rely on this assumption, either because they simply do not look at output distribution (e.g. the Morris method) or because they

HESSD
11, C5846–C5851, 2014

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

consider other properties of the output distribution (e.g. density-based methods, see for instance Peeters et al. 2014). Second, also within the variance-based approach, the output variance is related to generic variability of input factors (reproduced by random sampling or Sobol' sampling) and not their variance only.

- One assumption of the Sobol' method (at least in the implementation used in this work) is that input factors are uncorrelated. In this case, this means that: in the NB+NR scenario, bias and magnitude of random errors are independent; and in all scenarios, bias (and random errors) of different meteorological observations are independent. Are these reasonable assumptions?

- Page 13755: "by creating k new parameters (θ_1, θ_2,..., θ_k) that specify forcing uncertainty characteristics".
This is a bit confusing, mainly because up to this point the symbol θ was used to refer to model parameters in contrast to forcing inputs $\mathbf{F}$. The same confusion may arise in the following section, when the symbol θ and the term "parameters" may be interpreted as referring to model parameters (and Eq. (1) reinforce this misinterpretation). I would suggest to use a different symbol for the model parameters in Eq. (1) (for instance, p), and maybe insert a second equation like

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

as a companion to Eq. (1) to clarify the point (and also to link to the error model of Eq. (4)).

- Page 13759: "The number of rejected samples varied with site and scenario...".
I think the step of screening out meaningless simulations before estimating sensitivity indices is a very good practice, unfortunately not always applied in SA applications - the authors may want to stress the relevance, also referencing other works where this was done (for instance the already cited Pappenberger 2008).
Also, it would be interesting to know if this screening provided further insights about the model response surface. For instance, did you find that discarded simulations where

generated by input samples falling in a specific range or were they scattered across the input space? In the former case, can you give a physical interpretation to this result? Also, it is reported that the UB scenario at SASP had a very high number of meaningless simulations: can you give an interpretation for this? Does this relate to any specific property of the SASP site?

- Page 13762: "This was surprising given that bias magnitudes are lower for Q_{li} than for Q_{si}.
Misleading. It seems to suggest that the input with the larger variability range is expected to have the larger influence on the model output, which is not true unless the model is linear (and which motivates the use of complex SA methods to obtain input ranking).

- Page 13766: "1 520 000 simulations for examining only a single year at four sites across four error scenarios"
Misleading: the number of simulated years influences the computing time of each simulation but not the number of simulations. See also next comment on the issue of number of simulations vs computing time.

- Page 13767: "will be more feasible in the future with better computing resources and advances in sensitivity analysis methods".
The computing issue here is not completely clear. Over one million model evaluations is a big number but what is the actual computing time? Given that the model is one-dimensional I would expect every model evaluation to be rather fast, and therefore even 1 million evaluations to be a reasonable target.
Also, before Rakovec et al. (2014), there exist other well established GSA methods (for instance Morris method or FAST) requiring much less model evaluations than Sobol'. This is not a criticism of the choice of using Sobol', just a comment about the fact that computational complexity in this case is also due to the fact that you chose the GSA method that requires by far the highest number of model evaluations.

HESSD
11, C5846–C5851, 2014

REFERENCES

Peeters et al., 2014, Robust global sensitivity analysis of a river management model to assess nonlinear and interaction effects, HESS

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