Hydrology and Earth System Sciences
Review of Manuscript hess-2020-321

Summary

The authors present a method for assimilating snow depth observations collected by citizen scientists into a process-based snow model, which improves snow depth, mass, and disappearance date estimates relative to model runs without assimilation. The model outputs are improved regardless of the number of assimilated observations. The paper is well-written. I recommend the paper for publication after addressing the following comments and suggestions.

Major Comments

1) Why is the snow depth to SWE conversion necessary before assimilation? It is unclear if assimilating SWE is a literal requirement of the SnowAssim submodel, or if you are making a choice to assimilate SWE rather than assimilating snow depth directly. Previous studies (e.g., Smyth et al., 2019) have shown that assimilating snow depth improves modeled depth, and can also improve modeled density and SWE. I am not asserting that assimilating depth is better than assimilating SWE, but I think the paper needs a justification for including the depth to SWE conversion (and its associated error, even if minimal) as part of the workflow at all.

2) I was going to comment on the negative NSE values in table 1 (model calibration results), and then saw that a previous reviewer already did so. I think the authors' response makes sense, and the supplemental experiment with different precipitation adjustment factors is logical. However, I have two suggestions:
- The sentence on lines 361-362 should be more clear. Without reading your response to the previous reviewer, I would not understand that your model doesn't have a "snowfall correction factor" parameter that could be calibrated, as some other models do. Otherwise, a reader would probably not be satisfied that you simply say that NSE is "lower than expected." A negative NSE implies that you would be better off throwing away your model and using the mean of the observed data, which is a poor starting point for a supposedly calibrated model. Again, I realize that the model IS calibrated at this point, and the precipitation data is just terrible, but that point doesn't come across clearly enough in the paper.
- While the experiment with the precipitation adjustment factors is sufficient from my point of view, it sure would be a lot cleaner if you simply added a "precipitation correction factor" to your model and calibrated it along with everything else. Is that feasible? As you say, one of the benefits of your assimilation framework is that it "fixes" this bad-precipitation problem for you. But that also means you are choosing to calibrate some things, and choosing to let the assimilation fix the rest – which again, does not come through clearly in the text until later in the results.

3) I am confused by the methodology and conclusions relating to the experiments where the authors vary the number of CSO observations that are assimilated. I understand that the number of assimilated observations is varied between 1 and 32. As fewer observations are assimilated, the model receives less information – but are we talking about restricting the amount of information across space, time, or both? When you say that "Any number of CSO measurements

assimilated show improvements in model performance" (429) do you mean that any number AND any timing of assimilated observations leads to improvements? Or again, on line 462, where you say that "WY2017 has a smaller range in KS values as the number of assimilated measurements increases, more CSO simulations outperforming the NoAssim case" – is this because "more CSO simulations" cover a larger geographic area, or because they cover a longer time period? (or both)
- I have a related (more minor) question on line 391: what does "aggregated by week" mean? You assume all observations in a given week occurred at the same time, for the purposes of assimilation?

Minor Comments

82-84: A related goal (at least for statistical assimilation methods like the Kalman Filter, Particle Filter etc.) is to reduce the *uncertainty* of the given state variable.

109: This is the first mention of SnowAssim, and it is not defined/explained yet. Maybe simply omit the reference to the specific name here?

121-123: I was wondering when the motivation for using CSO depths (as opposed to lidar, etc.) would be mentioned in the introduction. It feels like this sentence belongs somewhere in the paragraph starting on line 86.

240: "State" missing an "s"

Figures 1 and 3: I agree with a previous reviewer that these figures should be combined (side-by-side).

Appendix C: Consider adding SNOTEL SWE to the plot, as another data point. For example, hopefully the measured cumulative precipitation is not less than measured peak SWE, indicating undercatch, etc.