Responses to Reviewers

The following responses to Reviewer #3 and Reviewer #4 include their comments and a point by point response by the authors, which is highlighted by bold font.

Reviewer #3

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.

Include a couple sentences about the density variation found in the fieldwork.

Response: The snow depth to SWE conversion is required because a SWE depth (m) is the input for the physical equations in the model code that govern the assimilation process SnowAssim (Liston and Heimstra, 2008, their section 3, paragraph 1). Assimilating snow depth would have required altering the model code and physical equations, and this was outside of the scope of our research questions. In Smyth et al. (2019) a different snow model, Snobal, is used with a particle filter to assimilate gridded snow depth data. This is a fundamentally different approach to data assimilation and snow dataset acquisition.

The authors have decided to alter the following sentence in line 214-215 of the current manuscript to further clarify that the model requires SWE depth for assimilation:

"Note that CSO measurements are submitted as snow depth (m), but the SnowAssim model code and physical equations require observational inputs to be SWE depth (m), so a conversion from depth to SWE was necessary."

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.

Response: The authors have attempted to address multiple important points brought up by the reviewer. In the calibration section (Section 4), we added several sentences that aim to clarify our decision making process during calibration and highlight the deficiencies in the precipitation inputs. We also added text that paraphrases our response to a previous reviewer that requested more information about the calibration results. This was added so the readers have easy and quick access to the substance of this conversation without needing to look into the comment discussion section. We tried to more clearly communicate that we chose to let the assimilation of CSO measurements address the precipitation deficiencies. Please see the text below that has been added to the calibration section of the current manuscript (Section 4, lines 372-386):

"CFSv2 precipitation totals at the UTS station were nearly 1.6 times the measured precipitation at the UTS station during the calibration period. The improvements that could be gained by adjusting a subset of the model parameters (wind, temperature, and precipitation lapse rates due to differences in elevation) during calibration were not likely to overcome this extreme precipitation deficiency, explaining why the final calibrated NSE and KGE values were low. There are two ways to address this precipitation deficiency using SnowAssim. One is to adjust the precipitation inputs during calibration, and the other is to allow the assimilation to adjust the precipitation inputs. Both ways are functionally equivalent because they apply a simple, scalar-based correction surface to the precipitation fluxes. In our calibration process we chose to use SnowAssim to address the precipitation deficiencies in the reanalysis product, following the approach of other recent studies in mountainous regions of Alaska, and following the original purpose of the SnowAssim model (Cosgrove et al., 2021, their Calibration of SnowModel section; Liston and Heimstra, 2008; Young et al., 2020, their section 3.4). This calibration decision supports the primary goal of the current study, which is to test whether or not participant-submitted snow depth measurements can improve physically-based modeling efforts through data assimilation.

These calibration results and the precipitation deficiencies motivated us to design an experiment to supplement the main findings of this research … "

In our response and additional text above, we have attempted to more explicitly communicate why we chose to let the assimilation address the precipitation deficiencies. Since the goal of the current study is to test whether or not participant-submitted snow depth measurements can improve physically-based modeling through data assimilation, we think our choice to show the results (Sections 6.1 to 6.5) without a precipitation correction is reasonable. We ran additional calibration statistics for the precipitation correction factor experiment results (Section 6.6) and we have a full ensemble of runs and results that correspond to this experiment. However, we chose not to show both sets of results side-by-side throughout the results section for the sake of simplicity.

We think the choice to keep the calibration process simpler is defensible because it requires less *in-situ* weather station data. This is a real-world data challenge that affects water resources managers and scientists working in remote, mountainous terrain worldwide. We think the results that include a snow model, a simple data assimilation scheme, coarse reanalysis product forcing datasets, and modest citizen scientists' efforts is more accessible and more likely to be applied by readers than a study that requires high quality, *in-situ* weather station data.

Cosgrove, C.L., Wells, J., Nolin, A.W., Putera, J. and Prugh, L.R., 2021. Seasonal influence of snow conditions on Dall's sheep productivity in Wrangell-St Elias National Park and Preserve. PloS one, 16(2), p.e0244787.

Young, J.C., Pettit, E., Arendt, A., Hood, E., Liston, G.E. and Beamer, J., 2020. A changing hydrological regime: Trends in magnitude and timing of glacier ice melt and glacier runoff in a high latitude coastal watershed. Water Resources Research, p.e2020WR027404.

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?

Response: We agree that there is some complexity in our methods and that this section would benefit from some additional clarification. We have added text to explain that the temporal subsetting of CSO measurements is the same throughout the results section, and includes only observations that occur after April 15th. We also explain that we do not restrict the spatial locations of the assimilated measurements, rather they just need to be located somewhere within the model domain. The temporal restrictions are simply that all results shown in the results section used CSO measurements that were taken on or after April 15th of both water years. We added

several sentences to Section 5: Experimental Design and slightly reordered the presentation of these sentences. See new Section 5 below (lines 395-415 of the current manuscript):

"We carried out a series of simulations in order to (1) quantify the improvement in model performance due to the assimilation of CSO measurements and to (2) understand the effects of the number of CSO data points selected for assimilation. First, we set up geographic and temporal requirements for the assimilated data. The only geographic requirement was that the CSO measurements must be located within the larger 5,736 km2 model domain. We subset the CSO measurements temporally to the peak SWE time period or later. According to the UTS station, peak SWE in the study area generally occurs mid- to late-April and consequently the earliest assimilation date was set to April 15th. The CSO measurements were aggregated by week by assuming all measurements in a given week occurred on the same day for the purposes of assimilation. This weekly aggregation allows the correction surfaces generated by SnowAssim time to adjust the precipitation fluxes and snowmelt factors between observations, thereby altering the model outputs during assimilation. Additionally, CSO participation in the Thompson Pass region during the early accumulation season was infrequent in WY2018 and non-existent in WY2017. Since peak SWE is important for mountain hydrology and ecology, with many snow studies using it as an indicator metric, the time restrictions are acceptable for the research questions addressed in this study (Bohr and Aguado, 2001; Trujillo et al., 2012; Kapnick and Hall, 2012; Mote et al., 2018; Wrzesien et al., 2017).

With these geographic and temporal filters defined for assimilation, we decided to vary the number of CSO data points selected for assimilation. Model simulations without CSO measurements provide a baseline for comparison, referred to as the NoAssim case. Ensemble model simulations were carried out with various numbers of CSO measurements assimilated, referred to as the CSO simulation case. An ensemble of 60 trials per year were carried out with n = 1, n = 2, n = 4, n = 8, n = 16, and n = 32, where n equals the number of CSO measurements assimilated per WY. In each instance (n value), 10 realizations of the numerical experiment were carried out. With the ensemble model simulations defined in terms of the spatial and temporal restrictions, the number of CSO measurements was the only feature modified during assimilation."

Lastly, the authors note that there is a new section in the results (Section 6.3: Spatial and Temporal Characteristics of the Assimilated Data, lines 538-566 of the current manuscript) that includes more details about 1) the spatial characteristics of the assimilated measurements that were selected for the time-series validation analysis and 2) the temporal characteristics of the assimilated measurements that were selected for the remote sensing validation analysis.

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)

Response: With the clarification in the answer to the previous question and the new Section 5 in the manuscript, the reviewer's question regarding geographic area and longer time period may be

clarified. However, the authors note that the wording of the referenced sentence is unintentionally vague and should be reworded for further clarity. A more precise sentence has been added to the current manuscript (lines 481-483) and can be found below:

"However, WY2017 has a smaller range in KS values as the number of assimilated measurements increases. Additionally, the number of simulations that outperform the NoAssim case in WY2017 gradually increases as the number of CSO measurements increases from 1 to 32."

The exact cause of this increase in performance in WY2017 with more CSO measurements is not known or investigated by the research questions in the current study, although it is admittedly one of the most interesting areas for future research. One may speculate that incorporating 32 measurements, that span multiple days and multiple geographic locations, would be better than incorporating a single measurement on a single day and in a single location. But the results show that assimilating a single measurement sometimes is just as advantageous as assimilating 32 measurements. There are a multitude of terrain controls on the spatial distribution of snow in mountain environments, including aspect, prevailing wind direction, elevation, curvature (concavity or convexity), etc. These all play a role in geographic location and require a different study design and a more in-depth investigation into the spatial characteristics of the study area than fits within the scope of the current study.

- 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?

Response: Yes, aggregated by week means we assume all observations in a given week occurred at the same time for the purposes of assimilation. We agree that sentences here were potentially vague and needed to be clarified. We chose this type of weekly aggregation because of the way SnowAssim adjusts precipitation fluxes / snowmelt factors between observation dates. Correction factors for precipitation / snowmelt are applied to the time period between observations (Liston and Heimstra, 2008, their section 3). Our initial model runs suggested that daily increments (consecutive SWE observations) did not allow enough time between observations for the correction factors to influence the SWE outputs. We aggregated all CSO measurements to the same day of the week to allow the correction factors time to adjust precipitation and snowmelt, thus altering the SWE depth outputs during assimilation. We changed the text of the current manuscript (lines 400-403) to add clarification.

"The CSO measurements were aggregated by week by assuming all measurements in a given week occurred on the same day for the purposes of assimilation. This weekly aggregation allows the correction surfaces generated by SnowAssim time to adjust the precipitation fluxes and snowmelt factors between observations, thereby altering the model outputs during assimilation."

<u>Minor Comments</u>
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.

Response: The authors agree that the reviewer adds an important extension to this sentence and we have changed the current manuscript (lines 83-85) accordingly:

"Regardless of the method of assimilation, the goal is the same: to produce a more accurate modeled state variable (snow depth or SWE) in space and time and to reduce uncertainty in the state variable by using in-situ observations to modify the process model output."

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?

Response: This is a good suggestion and we have changed the text in the current manuscript (lines 111-113) to the following sentence:

"The CSO project adds to a growing body of research accomplished by citizen scientists in the natural sciences, and demonstrates how CSO measurements can be assimilated into the process model workflow using a simple data assimilation technique to sometimes improve model results."

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.

Response: The authors agree that the following sentence should be placed earlier in the introduction, and we added it to the current manuscript at lines 88-90.

"The potential of mobilizing a new type of *in-situ* snow dataset collected by snow professionals and snow recreationists is significant because these participants often travel to remote mountainous environments worldwide where *in-situ* snow observations are sparse."

240: "State" missing an "s"

Response: This change has been made in the current version of the manuscript.

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

Response: These figures have been combined in the current version of the manuscript.

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.

Response: We've added the SNOTEL SWE vector to Appendix C in the current draft of the manuscript.

Reviewer # 4

This is a fairly polished manuscript that has already been through one round of reviews. Following comments from Reviewer #1 on the time and location of assimilation in relation to validation, the authors provide some useful information in their response. This information is therefore now available in the discussion, but it would be better to incorporate it in the paper.

Response: After reviewing our previous response to Reveiwer #1, we agree with Reviewer #4 that including additional information about the time and location of the assimilation data points in relation to the validation datasets adds important context for the readers. We've adapted our original response to Reviewer #1 into several paragraphs and a new figure in a new Section 6.3. These new paragraphs and figure are below (lines 538-536 in the current manuscript):

"The geographic locations of the CSO measurements used in the temporal and spatial results are an important factor that can shed some light on our understanding of the assimilation process. First, the time-series analysis validation metrics were quantified for all days in the water year at the UTS location. The CSO measurements that were assimilated in 2017 range in distance from 4.1 km to 30.5 km away from the UTS location, while the Best CSO simulation measurements (n=2) were located 5.5 and 6.9 km away. In 2018 the assimilated measurements range in distance from 2.1 km to 17.4 km away from the UTS location, and the Best CSO simulation measurements (n=2) were located 9.1 and 17.5 km away. Figure 10 includes a map of the assimilated measurements and a histogram of the distance between the CSO measurements and the UTS station from both water years, subset by the assimilation time period (on or after April 15th of each year). This distance analysis demonstrates that the CSO measurements used in the time-series assimilation do not coincide with the SNOTEL grid cell location. The histogram shows that improvements made at the SNOTEL location during assimilation were due to snow depth measurements taken by CSO participants kilometers away.

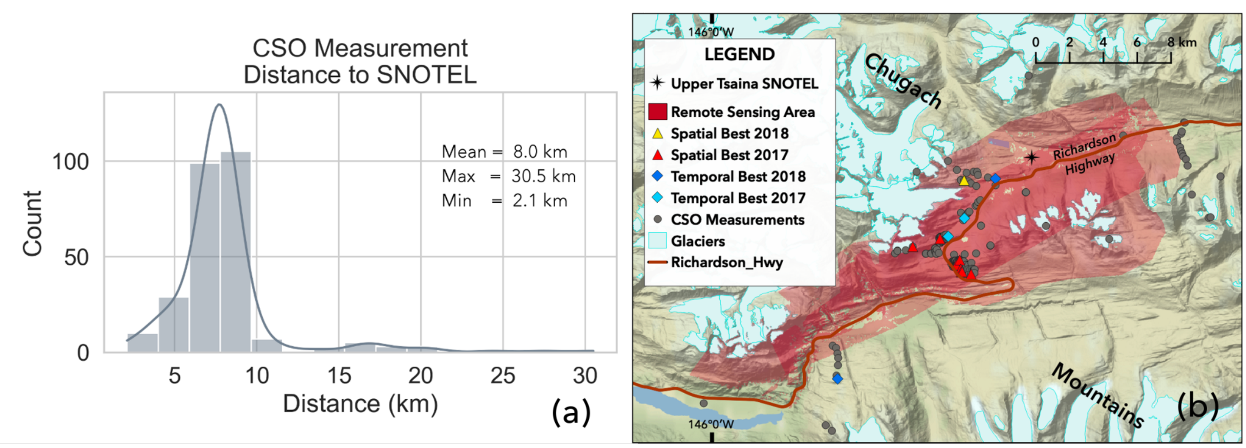

Figure 10: Assimilated measurements.

(a) A histogram showing the distance between the CSO measurements available for assimilation and the Upper Tsaina SNOTEL station, subset by the assimilation time period, on or after April 15th (n=266). A kernel density estimator is used to smooth the distribution. (b) A map of the CSO measurement locations that includes the best spatial and temporal CSO simulations for both water years. The map is zoomed in on the area of the highest density of CSO measurements.

Secondly, the remote sensing datasets were collected on April 29th in 2017 and April 7th and 8th in 2018. These validation datasets are essentially a spatial snapshot of snow depth from a single day in both water years. In water year 2017, there were a total of 9 CSO measurements submitted on April 29th, the same day as the remote sensing dataset collection. For the presented results in Section 6.2, none of these 9 CSO measurements from April 29th were used. For water year 2018, the remote sensing dataset was collected on April 8th and the measurements were not assimilated temporally until at least April 15th (see the experimental design outlined in Section 5). Figure 10b displays the locations of the CSO measurements assimilated in the Best CSO simulation from both water years (WY2017 n=1; WY2018 n=8). This analysis of the assimilated data demonstrates that the CSO measurements used in the spatial assimilation do not coincide with the dates of the remote sensing acquisition, revealing that improvements were made during assimilation by measurements that were taken at a different time."

Related to this, I agree with the reviewer's suggestion of putting Figure 1 and 3 next to each other. They have too much overlap to be combined in one figure, but being able to compare the assimilation and validation points in a Figure 1a and 1b pair would be useful. Readers will surely be curious about which CSO measurements were assimilated, and at least the reason for not giving this information should be given in the paper.

Response: This is a request from several reviewers and we are pleased to accommodate this request. We combined the two figures in the current draft of the manuscript.

Reviewer #1 also asked for a comparison of estimated and measured snow densities. In addition to the insertion of rmse and bias statistics for SWE conversion from snow depth, it would be useful to state the mean SWE here for context.

Response: The fieldwork measurement mean SWE is 51 cm and this statistic has been added to the final sentence in Section 6.4 (lines 584-585):

"The fieldwork measured mean SWE is 51 cm, the RMSE in SWE is 10.5 cm, and the Bias in SWE is 0.6 cm when using the Hill method for all fieldwork sites."

Reviewer #2 asked for a more complete description of the DA approach. A crucial piece of information that is missing in the authors' response to this is that SnowAssim applies corrections retroactively, so simulations also change before the observation time. Readers could learn that by reading Liston and Hiemstra (2008) or figure it out by puzzling over Figure 5 (it is more obvious from Figure 6), but they could be spared that effort.

Response: The authors agree that a clarification should be added to the manuscript regarding the retroactive corrections by SnowAssim. We adjusted several sentences in the SnowAssim method review in Section 3.2.4 (lines 208-212).

“SnowAssim requires the model to be run twice and pauses at the end of the first model run. During this pause, differences between the observed SWE depths and modeled SWE depths in time and location are calculated and interpolated to the entire model domain in the form of a correction surface. The final correction surface is spatially distributed (for each day of observations) using the Barnes interpolation scheme. These correction surfaces are then applied to the precipitation inputs and snowmelt factors during the second model run.”