Reply to Bart Forman's comments on "Improving the characterization of initial condition for ensemble streamflow prediction using data assimilation"

We would like to thank Dr. Forman for reviewing our manuscript. Several of these comments have motivated us to provide more description of the methods used in this study and pertinent discussion. In addition, constructive comments on the notation in the equations was greatly appreciated. We believe that the revised manuscript makes the study more reproducible and equations more understandable and straightforward.

Reviewer Summary:

The authors employ a particle filter to generate an ensemble of initial condition estimates of snow fields for the purpose of ensemble streamflow prediction. Ground-based SNOTEL measurements are assimilated by the particle filter into the SAC-SMA model. The results presented show improvement to streamflow predictions using the proposed framework, but that significant limitations exist particularly related to the representativeness of the SNOTEL measurements used during the assimilation. Overall, this manuscript should be of interest to both the snow data assimilation community as well as the ensemble streamflow forecasting community, but that the manuscript could benefit from a more careful description of the methods used as well as the interpretation of some of the results.

Comment:

P. 7211, L.15: What motivated the selection of these 15 particular station locations? Was it based on data availability? Do these 15 stations run the gambit of ecotones/climatologies pertinent to snow data assimilation? Please justify with a sentence or two.

Reply:

Thanks for bringing up this issue as it is not discussed but very pertinent to the discussion. These basins are chosen based on data availability and reliability. The available data for other headwater basins within the upper Colorado appeared to be corrupted or compromised in some way (e.g. large portions of missing data or lots of negative precipitation or streamflow values). Since this study only focuses on the upper Colorado river basin, this study is not representative of the majority of cold weather/mountainous climates. We would suppose that similar results will be found in regions of similar SNOTEL coverage but to be certain would require similar studies in other climates. This has been noted in the methods section and the discussion and conclusions section.

Comment:

P. 7213, L.15 and L.17: Should the subscript inside the model operator be "t-l"? Also, please try to be more consistent with your notation. For example, "i" appears in the subscript in some places and in the superscript in others. Please make your usage of the notation more consistent as to make it easier for the reader.

Reply:

Thanks for the observation. This was an oversight and the indexing has been updated in the revised manuscript.

Comment:

P. 7213, L.15: Does the model error have to be additive? What about multiplicative errors, e.g.? How might your formulation change using multiplicative errors?

Reply:

This error is not necessarily additive but this is the most common methodology for implementation in DA studies. Please note that this error is heteroscedastic (larger variance for larger values of SWE) because the errors will likely grow as SWE increases. However, the precipitation error is multiplicative in DA and we assume it follows a log-normal distribution. This has been addressed in the revised text.

Comment:

P. 7213, L.15 and L.22: How did you define the model and measurement errors? Are they subjectively-chosen scalars? Are they a function of SWE such that the error increases with increasing snow amount? Also, what about representativeness errors (e.g. point-scale versus model-scale error)? Please include a few sentences as to how you selected these error quantities and the rationale for doing so.

Reply:

All SWE errors are normally distributed with a standard deviation of 25% of the SWE value. Errors in Bayesian analysis are always at least partially subjective and in this study 25% was found to provide the greatest skill in hindcasting streamflow, in comparison to other levels of error. Within a given elevation band, this is assumed to be sufficient for managing the different scales the model and SNOTEL, but at different elevations the large difference in the timing of melt makes it nearly impossible to account for in this format. Thus the discussion about elevation representativeness was presented. The manuscript now explains the SWE prediction error and observation error in the data assimilation and SNOTEL section respectively.

Comment:

P. 7214, L. 9: What is “R” in Equation 4? Also, what does the subscript “k+1” represent? Please define accordingly.

Reply:

Thanks for the comment as this had not been properly described in the text. This is the variance of the SWE observation within the likelihood function. Also, this was originally supposed to be R_k but somehow a “+1” was added to the subscript. This was an oversight and has since been corrected.

Comment:

P. 7216, L. 1: Is “more accurate” the same as “less uncertainty”?

Reply:

In this sense accuracy means the correctness of the uncertainty estimation. This means that a model with more uncertainty may be more accurate if it correctly represents the true uncertainty of the streamflow volume prediction. A lowered uncertainty is preferred but accurate representation of uncertainty is a priority over lowering uncertainty.

Comment:

P. 7216, L. 14: Similar to Comment #2 above, please be more consistent with your notation. "i" was once the replicate number and now it is the probability category. I understand these quantities are somewhat related, but there is no harm in using different notation for the two as to make the paper more reader-friendly.

Reply:

Thanks for the comment. These are quite related but in light of the suggestion, we adjusted the notation to use a "j" instead of an "i" to avoid confusion.

Comment:

P. 7220, L. 15: Why not apply larger measurement errors in the upper elevations? You clearly state why the performance of your method varies as a function of elevation as related to the SNOTEL locations, but you've done nothing to try and alleviate this shortcoming. Perhaps a more judicious use of error model(s) would be beneficial? In addition, please refer to Comment #4 shown above and the need to include some description of the error models you selected and your rationale for doing so.

Reply:

Thanks for the suggestion but we had attempted to improve the results through increasing the error in the upper elevation band but this is more difficult than it first seems. This issue is the strong difference in melt timing. While the middle elevation band may begin melting in March, the upper elevation band may continue to accumulate through April and May. Since there are very different melt dynamics in the upper and middle elevation bands, the difference in SWE between the two bands grows quite fast in the melt season. This leads to a very poor representation of the upper elevation band shortly after the onset of ablation in the middle elevation band. It is quite difficult to assess the temporally changing errors that would need to be applied to the observations at this point in the season. Due to the complexity of the differing melt dynamics, we decided to not focus on fixing this problem in this study as we believe there is enough content for an entirely separate study of how to handle differing melt dynamics in data assimilation. This issue is addressed more simply on page 7218 of the original manuscript to avoid distracting the reader from the current study.

Comment:

P. 7221, L.3: I don't believe the term "accurate" is best used here. Just because you have a relatively uniform rank histogram is no guarantee that your ensemble is accurate (see discussions by Talagrand as well as Hamill). Perhaps the term "consistent" is a better choice?

Reply:

In Hammill 2000, it is stated "The rank histogram, when correctly used and interpreted, measures the reliability of the ensemble". A more uniform rank histogram suggests a more reliable probabilistic prediction according to Hammill 2000. For this reason, we will opt to use the term reliable over "accurate" or "consistent".

Comment:

P. 7221, L. 18: Perhaps a synthetic study would be a worthwhile pursuit in a follow-on study?

Reply:

Correct implementation of a synthetic study would be quite beneficial to this new technique but definition of such a study is difficult within a synthetic framework because of the ESP methodology. In a synthetic state estimation experiment, the researcher knows the true values for the forcing data, state values and model outputs (e.g. streamflow). In the ESP framework, the forcing cannot be known. This would lead to the question of how representative the historically resampled forcing is of the true forcing. Since the framework of the synthetic experiment will likely come into question, we found it more valuable to show results for a real experiment to demonstrate that real results can be improved through the addition of ensemble data assimilation to ESP.

Comment:

P. 7221, L. 19-20: Again, I think a more effective error model might be beneficial here. The results in Figure 2 could be an excellent starting point for developing a more accurate/useful error model. It would be interesting to see a discussion on error model sensitivity and how accounting for the known deficiencies in the SNOTEL station locations could translate into improved particle filter performance.

Reply:

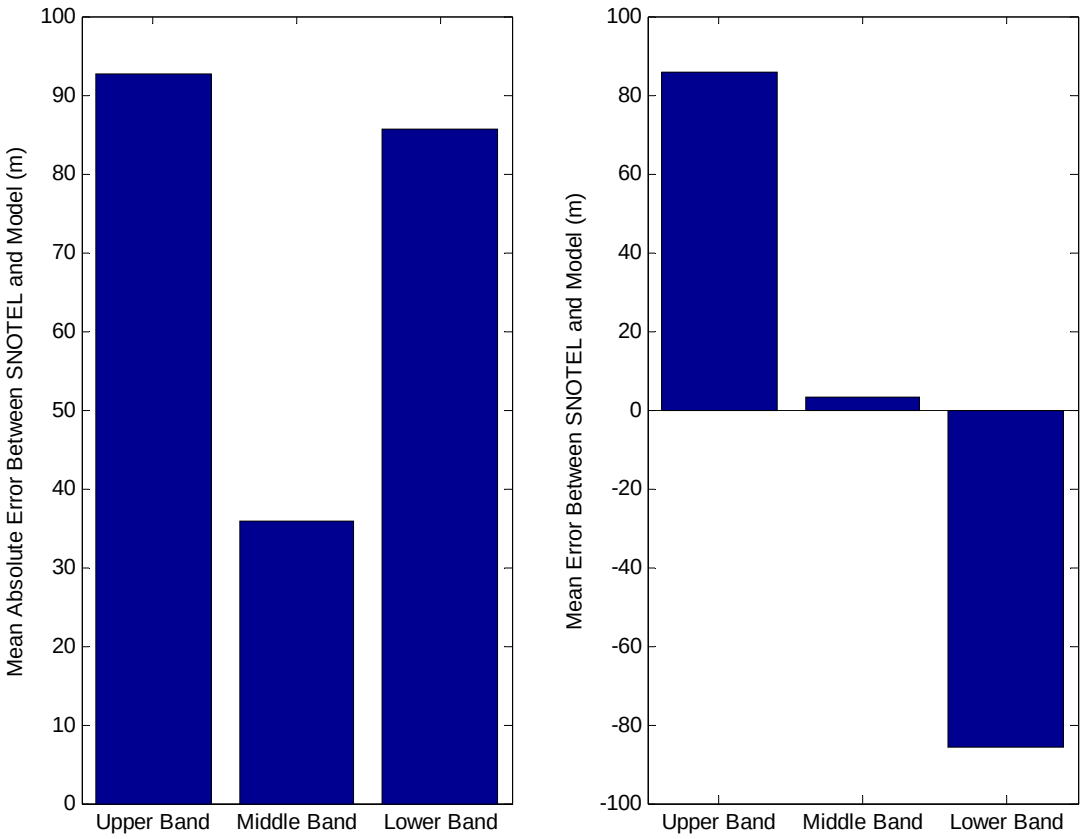
This was discussed in the response to the eighth comment. As was stated earlier, this is an interesting topic to explore but not the focus of this study, and would require a separate study to make conclusions about how to handle errors resulting from elevation differences.

Comment:

P. 7228, Figure 3: Is the "absolute" part necessary? If the elevation difference is negative, it would benefit the reader to clearly see if the station locations are too low or two high for a given elevation band. Therefore, I suggest you make the very simple change of presenting average elevation difference rather than average "absolute" elevation difference.

Reply:

Thanks for the suggestion. We feel that both mean absolute difference and mean difference are valuable information so we have replaced figure 2 with the following plot. In addition, the values in the plot have been updated because the values in this figure were slightly off due to difference in the SNOTEL site selection in the final ESP-DA experiments. We also added a new figure as requested by reviewer #2 and added a short description about our definition of the middle elevation band.



Comment:

P. 7230: The particle filter is effectively removing mass even though the same forcing is being applied. Is this solely a function of the location of the SNOTEL stations within a certain elevation band? Or are there more factors at play? Land DA systems are notorious for not conserving mass and it would behoove the reader to have you elaborate as to whether or not your system conserves mass. You hint at the benefits of the particle filter as related to its ability to conserve mass (P. 7213, L. 5-10), but I'm not convinced based on your results in Figure 4. Please comment on the issue of mass conservation with a particular emphasis on convincing the reader that mass is, in fact, being conserved and that the particle filter is really doing everything you claim it can do.

Reply:

The PF appears from these figures to be removing mass but the definition of mass conservation may be somewhat ambiguous in this sense. The PF does not perform any adjustments of the states after model simulations. The only operations are weighting and resampling of particles. The only cause for difference in mass come from the uncertainty propagated through the forcing data. We claim that mass is conserved because the PF cannot directly remove mass from the system, which is the case in the EnKF, variational and direct insertion strategies. If one defines conservation of mass as the conservation of observed precipitation, which is assumed to be imperfect through the PF, then the mass balance is violated. Since precipitation that is input into the model (though perturbed from the observed precipitation) is incapable of being removed from the system through the data assimilation process, we consider mass to be conserved. A short explanation was added to our claim of mass conservation.

Comment:

P. 7234: The x-axis is not a "Rank", per se. Perhaps "Normalized Rank" would be more appropriate?

Reply:

We agree that Normalized Rank is a more appropriate label for the x-axis and this was changed in the revised version of the paper.

Comment:

P. 7234: Is the ESP-DA rank histogram underdispersed? Overdispersed? Please elaborate with a sentence or two.

Reply:

The ESP-DA is slightly overconfident (underdispersed) but in relation to ESP is much more reliable. This overconfidence is more apparent in the Q-Q plot and is noted on page 7221 lines 4-6 of the original manuscript.

Minor Changes/Questions/Concerns:

1. P. 7214, L. 6: There is a noun missing in this sentence, but I'm not certain what it is. Perhaps you omitted the word "terms"?

Reply:

This was actually supposed to be a "y" but was somehow omitted in the conversion from a doc file to pdf. This has been addressed.

2. P. 7214, L.15: This is an incomplete sentence as it is missing a verb. Should “equal” be changed to “are set equal”?

Reply:

This should in fact be “are set equal” and is reflected in the revised version.

3. P. 7215, L. 20: This is a good point made in this sentence. Be certain to re-iterate that point elsewhere (e.g. Conclusions) to hammer that point home for the reader.

Reply:

The results and conclusion sections have been edited to more clearly re-iterate this point.

4. P. 7215, L. 26: I believe there are 10 ensemble replicates in the referenced figure rather than the 8 as specified?

Reply:

A close inspection shows that there are only 8 initial conditions sampled from the hypothetical distribution. This is noted by the 8 dots on the initial condition line.

5. P. 7220, L. 2: “Beginning” rather than “begging.”

Reply:

Thanks for the recognizing this error. It has been fixed.