Reply to reviewer #2 comments on "Improving the characterization of initial condition for ensemble streamflow prediction using data assimilation"

We would like to thank the reviewer for providing feedback on our manuscript. These comments provided motivation to further clarify the discussion of elevation representativeness of SNOTEL as well as clarify the data assimilation framework. The reviewer suggested a scatter-plot between model and SNOTEL elevations and we have added this to the manuscript. In addition, we believe there were some misunderstandings for the reviewer on the portions of the data assimilation experiment, which has prompted us to give further attention to the methodology descriptions to clarify ambiguities in our earlier presentation.

General Comments:

Ensemble Streamflow Prediction main hypothesis for snowmelt controlled basins is that the snowpack initial conditions drive the snowmelt and summer flow. The authors try to quantify the potential improvement in accuracy and uncertainty quantification in seasonal streamflow prediction if the errors in the initial conditions, in particular snowpack, were minimized and quantified via data assimilation. The results show that the current uncertainty in the data assimilation are too large in order to see a consistent improvement either in accuracy or in the quantification of the uncertainties of the seasonal flowforecasts. The paper complements previous analyses on how improving SWE via data assimilation could improve flow forecasts. The paper would benefit from adding more details on the DA (see specific comments) , and an extended literature review and related discussion in order to further define the scientific contribution with respect to those analyses: SNOTEL vs MODIS, seasonal scale vs short range prediction, different data assimilation approaches - what worked and what did not. Also, only a couple of them assessed the improvement in the probabilistic forecast sense. This is where this paper's contribution is the most significant in my opinion. It means that the paper would also benefit from improving the organization on the prediction verification section, i.e. not only assess the accuracy and the range of uncertainty but further assess the information in the uncertainty (reliability).

Suggested references:

Andreadis K. M., and D. P. Lettenmaier, 2006: Assimilating remotely-sensed snow observations into a macroscale hydrology model. Adv. Water Resour.,29, 872–886.

Wood, A.W. and D.P. Lettenmaier, 2006: A testbed for new seasonal hydrologic forecasting approaches in the western U.S., Bulletin of the American Meteorological Society, 87(12), 1699-1712, doi:10.1175/BAMS-87-12-1699.

Mcguire M., Wood A.W., Hamlet A.F., Lettenmaier D.P., 2006: Use of satellite data for streamflow and reservoir storage forecasts in the Snake River Basin, ID, J. Water Res. Planning and Mgt 132, 97-110.

Clark, and A. G. Slater, 2006: Probabilistic quantitative precipitation estimation in complex terrain. J. Hydrometeor., 7, 3–22.

Slater A. G., and M. P. Clark, 2006: Snow data assimilation via an ensemble Kalman filter. J Hydrometeor.,7, 478–493.

Clark, A. G. Slater, A. P. Barrett, L. E. Hay, G. J. McCabe, B. Rajagopalan, and G. H. Leavesley, 2006: Assimilation of snow covered area information into hydrologic and land-surface models. Adv. Water Resour.,29, 1209–1221.

Tang, Q., and D.P. Lettenmaier, 2010. Use of satellite snow-cover data for streamflow prediction in the Feather River Basin, California. International Journal of Remote Sensing, 31(14), 3745-3762. doi:10.1080/01431161.2010.483493.

Specific Comments:

Comment:

Describe the DA in more details: Is the DA assimilating SWE only or also other snow information like the snowpack content (depending on the model structure). Is the DA for one elevation band and one basin taking into account all the nearby SNOTEL stations in this elevation band used or which one are used and on which criteria. Is the period for training the DA approach different than the period of evaluation? I.e are those results the maximum or expected improvements from the DA?

Reply:

The DA is only assimilating SWE as suggested on P. 7211 L. 23 of the original manuscript. This statement has been made more explicit that SWE value used in the assimilation. The DA finds the closest SNOTEL station (a combination of spatial and vertical proximity) and assimilates that SWE, with an assumed error. With this data assimilation technique there is no training period. The DA is run beginning in September of 2002 so that it will capture the entire accumulation period. This means that the results presented would be expected if implemented operationally.

Comment:

Describe the goals of the performance metrics; the volume for accuracy, the ensemble range for the uncertainty. The RPSS is presented as a performance metrics but is not used in the results section. Also the rank histogram (should also refer to Talagrand and Vautard 1997) is used at the end and is more meaningful than the ensemble range in order to quantify that reliability in the ensemble spread. Please clarify what the QQ plot measure – is it equivalent to a reliability diagram? In the paper, the uncertainties are defined as the range of the ensemble. The larger ensemble does not imply a better representation of the uncertainties as long as this is not compared with the observed variability (the rank histogram used at the end).

Reply:

The RPSS is not shown in a chart but is used as a performance metric on P. 7219 L. 25 in the text. We added Talagrand, Vautard and Strauss (1997) to the reference section. We agree that the Rank

Histogram is more valuable than the ensemble range and this was our motivation for showing it towards the end. In addition, the Q-Q plot is similar to the rank histogram as was noted in the text. The visualization of the data in the Q-Q plot gives essentially the same information as the rank histogram but allows for comparison with the uniform distribution and allows for the calculation of the reliability scores (changed from Q-Q score). In addition, we also agree that a larger ensemble does not imply a better representation of uncertainties. We suggest that a better representation of uncertainty is provided by ESP-DA because it is much more reliable according to the rank histograms and Q-Q plot.

Comment:

the data assimilation approach used does not specify if both SWE and the snowpack cold content are assimilated. The authors remind the reader that the DA is not flawless, but they then assume that uncertainties in the DA approach are smaller than the uncertainties in the initial conditions. This point alone would clarify some points of the conclusions

Reply:

Only the SWE is used in the assimilation as noted above. In the manuscript, we give a detailed explanation that elevation differences between SNOTEL and the upper/lower elevation bands create problems in the timing of snowpack melt. It is understood that this will have negative effects early and late in the snow ablation period but is likely still valid for the forecasts beginning in the months that were discussed (March, April and May). For spatial heterogeneities that are not associated with elevation difference, additive error is applied to the SNOTEL observations (note the update to the SNOTEL description in section 2.4). In addition, we assume that the uncertainties are larger in the DA than the spin-up because the PF estimates the probability distribution of the initial states, which is neglected in conventional ESP.

Comment:

In addition, even though the results show otherwise, the conclusions remain optimistic that an improved DA should improve seasonal flow prediction, but still without assessing why it did not work.

Reply:

We respectfully disagree that the assimilation did not work. The results in this paper suggest that the technique was effective in improving the reliability of seasonal volumetric forecasts beginning in March, April and May. We remain optimistic about the general method of ESP-DA assuming that in time the snow data assimilation community will develop more effective methods than simple SNOTEL DA or gain a better understanding of how to minimize errors in a SNOTEL DA framework. We believe that it was discussed in detail the shortcomings of SNOTEL data assimilation while still discerning useful information about the general technique.

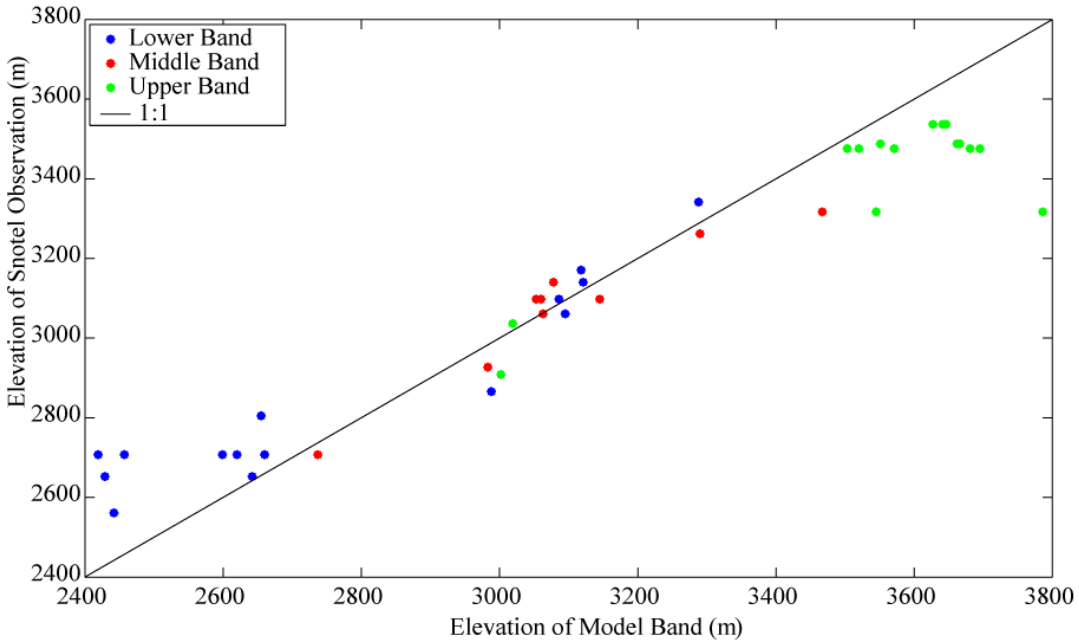
Comment:

The presented explanation is the inadequate SNOTEL spatial distribution and representation of the

elevation band according to Figure 2. Figure 2 would be more representative if instead of showing the average absolute different it was showing the scatter plot of the SNOTEL elevations and the corresponding elevation bands for the different basins so that the diversity does not get averaged. What was the improvement in a basin where the elevation bands were the best represented? and in the worse represented?

Reply:

Thanks for the comment. This new figure (pasted below) shows some important details about the SNOTEL representativeness that we think are quite important but did not address in the original manuscript. The new figure shows that the representativeness of SNOTEL in the elevation bands between 2800 and 3400 meters. Though this includes some of the upper, middle and lower elevation bands from the study basins, in general this is the range in which the middle elevation band falls and therefore in the text we have referred to this elevation range as the middle elevation band. In general, melt from this range is contributing to the March, April and May streamflows, thus providing more convincing evidence of the technique for those months. This figure will be added to the text as well as pertinent discussion.



Comment:

clarify the description on how the DA ensemble members are merged with the ESP ensemble members. (for example, just specify that from each DA ensemble members/ state – and say how many you used- and ESP is performed, driving to a seasonal flow forecast prediction of X times Y members. The figure is fine.

Reply:

This is described in the last paragraph on page 7215 of the original manuscript. Lines 24-26 explain that the DA is performed up to the given forecast start date at which time 50 initial condition ensemble members are sampled from the PDF obtained via DA, and then ESP is performed from each initial condition.

Comment:

Intuitively, it would be expected that the DA_ESP ensemble would be more reliable for shorter time steps (lead one month) when usually the ensemble is overconfident (too narrow)– but apparently it is not obvious that the ensemble reliability increases afterwards, in fact it doesn't. This is not discussed in the paper presently, although this is illustrated in Figures 5-6.

Reply:

We respectfully disagree that the reliability is not increased at lead times more than one month. The final two figures of the paper show that for 3 month forecasts (only the ones starting in March, April and May) the reliability of the forecasts are improved. Figure 6 shows that the June 3 month forecasts are quite poor due to poor SNOTEL representativeness of the upper elevation band. Figures 5 and 6 are less revealing about the reliability of seasonal forecast than Figures 8 and 9 and therefore we point the readers to these figures for drawing more finalized conclusions.

Comment:

The technique applied here is a direct insertion – i.e. the simulated SWE is 100% replaced if I understand it right. In order to better explain what is going on, I would recommend showing the current calibration performance of the model. The DA might as well be correcting for a systematic bias and improve some of the accuracy metrics – but might not be appropriate for the calibration parameters used. One of the suggested references addresses it.

Reply:

The PF is an ensemble based DA technique which uses the SWE observation to constrain the uncertainty propagated through the model. This is not a direct insertion technique. Also, the DA is not correcting for a systematic bias, it estimates the posterior state distribution at each time-step with information from the model and observation.

Comment:

specify Seasonal streamflow prediction in the title

Reply:

Although the application presented here focused on seasonal forecasting, the procedure can be used for shorter lead time forecasts as well. In order to maintain the generality of the procedure, we prefer not to change the title.

Comment:

p7209 – add the studies (McGuire et al. ***, Clark and Hay 2004, etc)

We added McGuire et al. 2006, Talagrand, Vautard and Strauss (1997), Tang and Lettenmaier 2010 as we believe these are pertinent studies.