Reply to reviewer #1 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. Overall these comments motivated us to provide a better description of the assumed uncertainties with respect to SNOTEL and improve our literature review to differentiate our work from previous studies. This makes the study more reproducible and our motivation more clear.

Reviewer Summary and General Comments:

The paper addresses the question if ensemble streamflow predictions can be improved by assimilation of observed snow water equivalent (SWE) data into a hydrological model. Specifically the paper is concerned with the problem of initial conditions of a forecast model run. Here a case study of the Upper Colorado river basin and the existing method of ensemble streamflow predictions (ESP) is presented. It is aimed to improve ESP by an assessment of the uncertainty in estimating the initial conditions of the hydrological model in use. Thereby the authors make use of SNOTEL observations of snow water equivalent (SWE). They show that there is some improvement compared to the traditional ESP technique and conclude that data assimilation (DA) of SWE datahas potential for streamflow predictions. They also point out to the problem of SNOTEL station representativity, which shows to effect the resulting predictions in various ways. However, this apparent and interesting problem is not quantified by the authors. The question of how additional data sources can be used to improve hydrological forecasts is practically and scientifically interesting and important. However, the approach taken by the authors is not new and the paper may be regarded as another case study. Even though the authors show that on average improvements over the existing method can be made, it is hard to judge these improvements as no other comparisons e.g. to other studies are given. There are already other studies which showed that DA can improve probabilistic predictions. However, the authors touch a generally important question, which is about the representativity of data sources fed into a model. This is however not investigated in detail. Thus I recommend to improve the manuscript along these lines and would be available for reviewing a resubmitted version. Further remarks and comments on the paper follow below.

General Comments

The authors show that the SNOTEL observations may not really be representative for the basin as the observation density in high and low elevations is rather poor. However, with assuming that the SWE states by the SNOTEL observations is correct, they may introduce a bias into the model predictions. Fig. 4 shows that the DA technique has usually lower values of SWE and as stated on P7219L10 -L13 ESP-DA shows smaller ensemble prediction ranges than the traditional ESP, which is counter-intuitive but related to this bias introduced by the low representativity of SNOTEL. I feel that the uncertainty introduced by the different observation densities of SNOTEL observations should be quantified to get more confidence in the resulting predictions. This is important for (i) acceptability of the method itself and (ii) the question which properties of data sources for DA are relevant (such as representativity).

Further, it may be worthwhile to constrain the DA technique also with observations of the target variable, i.e. streamflow.

Reply to main comments:

Two major concerns were expressed in these comments

1.) The framing of the current study within the current literature of hydrologic data assimilation and ensemble streamflow forecasting
2.) The potential for introducing bias into the modeling framework stemming from low representativeness of SNOTEL data

Reply to major concern #1:

With respect to the first concern, the reviewer states "the approach taken by the authors is not new and the paper may be regarded as another case study". While separately ensemble based data assimilation and ensemble streamflow prediction (ESP) have been widely studied in the hydrologic literature, to our knowledge, the use of ensemble data assimilation to improve initial condition uncertainty estimation for ESP has not been presented in the published literature. A number of studies have similarities and are listed below. Of these references, Mcguire et al., (2006) and Tang and Lettenmaier (2010) use a direct insertion method to assimilate satellite snow cover into the VIC model. While this is a data assimilation technique, it differs significantly from our approach. Direct insertion is a method to sequentially correct states but unlike ensemble based filters (EnKF and PF) does not quantify state uncertainty. On the other hand, Li et al., (2009) and Wood and Lettenmaier (2008) examine a method called reverse ESP that examines initial condition uncertainty by spinning up the model with resampled forcing data, not data assimilation. The hypothesis in our study is that by accounting for the initial condition uncertainty through Bayesian data assimilation, the uncertainty in the overall ESP framework can be more accurately estimated and therefore lead to more effective risk management in reservoir operation (discussed in the introduction and section 3.2). We would like to acknowledge that McGuire et al., (2006) and Tang and Lettenmaier (2010) are pertinent studies to discuss in this paper and it was an oversight to not include these in our literature review.

Li, H., Luo, L., Wood, E. F., and Schaake, J.: The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting, Journal of Geophysical Research, 114, D04114, 2009.

McGuire, M., Wood, A. W., Hamlet, A. F., and Lettenmaier, D. P.: Use of satellite data for streamflow and reservoir storage forecasts in the Snake River Basin, Journal of Water Resources Planning and Management, 132, 97, 2006.

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

Wood, A. W., and Lettenmaier, D. P.: An ensemble approach for attribution of hydrologic prediction uncertainty, Geophysical Research Letters, 35, L14401, 2008.

In addition, it was suggested that the paper must compare these results to other studies. In light of the clarification above, we believe that there is no consistent study to compare against. The only other studies that use data assimilation to initialize seasonal forecasts are those discussed above. Since these studies only examine deterministic measures of forecast accuracy, and our study is focused on examining the reliability of probabilistic prediction, these studies do not have comparable results.

Reply to major concern #2:

We appreciate the discussion about the representativeness of in-situ measurements because this has been a difficult issue to deal with in many aspects of hydrologic data assimilation. This can be an incredibly difficult problem to quantify due to the lack of a reliable dataset. Though this will always be a concern, we find it necessary to make some clarifications in light of the reviewers comments. First, we must respond to the statement "with assuming that the SWE states by the SNOTEL observations is correct, they may introduce a bias into the model predictions." This is true, but in our study we do not assume that the SNOTEL observations are correct. As with other sources of information (precipitation, temperature) there is an assumption that it carries some error. We assume that the SNOTEL observations have an error with a standard deviation of 25% of the SWE at the time of assimilation. This was a variable we tuned while performing this study to achieve improvements in streamflow simulation. A heteroscedastic error, as was used here, is capable of managing spatial errors within certain elevation bands, assuming that the error is sufficiently quantified. This is because within a small range in elevation in a basin, the dominant processes causing heterogeneities of SWE will be precipitation distribution, vegetation distribution, and wind redistribution. Elevation differences are a much more difficult issue to deal with and this led to our discussion of the elevation representativeness of the SNOTEL sites as opposed to spatial representativeness (Note the explanation on page7218 lines 11-20). Second, we respond to the comment that "ESP-DA shows smaller ensemble prediction ranges than the traditional ESP, which is counter-intuitive but related to the bias introduced by the low representativity of SNOTEL." While it is acknowledged in the paper that there is a bias in relation to the standard model spin-up, we cannot completely quantify the bias in relation to the true states of the basin. Because we cannot know the true average snowpack properties, we do not know if there is a bias. In this paper, we use a water balance model which allows for us to make the assumption that a more accurate initial condition will lead to a more accurate seasonal forecast due to the conservation of mass. Since the ESP-DA provides a more accurate characterization of the uncertainty in volumetric flows during months affected by the middle elevation band, we find that this is evidence that ESP-DA is providing more accurate representation of the uncertainty in the water held in the snowpack. This was discussed at the end of Page 7219 and the beginning of 7220 in the original manuscript.

Comment:

I feel that the uncertainty introduced by the different observations densities of SNOTEL observations should be quantified to get more confidence in the results.

Reply:

It is stated that this is necessary to show the “acceptability” of the method and to show what data sources are relevant for DA. Since the issue of representativeness is such a great problem for data assimilation, we believe that a more convincing account of the uncertainty in SNOTEL data would warrant an entirely different publication. With the reasoning presented above and the evidence in the paper to support the hypothesis that data assimilation has potential to improve ESP, we believe that this paper is fit for publication. However, a discussion about the reasoning for SNOTEL uncertainty was expanded in the revised paper in the methods section.

Comment:

P7219L13 and Figure 4: Is there a relation of the prediction uncertainty (deviation, IQR) and the value of SWE initial conditions? I think that the paper could be improved by a thorough characterization of the initial conditions (e.g. volume stored as snow vs. seasonal streamflow volumes or the spatial variability in SWE, etc.).

Reply:

There is inherently a relationship between prediction uncertainty and SWE initial conditions. This is known because of the conservation of mass. Though we know there is a relationship between initial condition and streamflow volume, it is difficult to show directly in a figure because watershed processes are quite nonlinear and the timing of melt is complex through the different elevation bands. This problem was a motivating factor in performing this study. The addition of water (i.e. precipitation) and the removal of water (evapotranspiration) necessitates the use of a water balance model to compare the effects of initial condition and streamflow uncertainty. The ESP-DA framework performs this directly but also includes forcing uncertainty. The seasonal streamflow volume box-plots, cumulative volume runoff time-series graphs, rank histogram and Q-Q plot in conjunction with the SWE TWS box-plots provide insight into the effects of the initial condition uncertainty on streamflow.

Comment:

Figures 4-7: Generally I find it tedious to derive general conclusion from these multipanel figures. First, there are 15 sub-basins introduced. But it is not clear to the reader if the results shown are aggregated for all 15 basins and then how large are the differences of prediction accuracy between these basins. Further, there seems to be some effect on prediction accuracy of (i) season, (ii) individual years (is dataset large enough?) and (iii) lead time. But there is no assessment of these effects in terms of verification measures. Instead the verification presented with Figures 8-9 aggregated almost all data (15 basins * 3 month * 3 years). Further, while Fig.4-7 also shows results for June (has poorest DA results) this month is left out in the verification figures.

Reply:

Thanks for the comment as we have realized the nature of these four figures was not entirely clear. Figure 4 is the sum of the TWS across all basins as stated in P7218L1 of the original manuscript, but were also specified in the figure caption in the revised version. Figure 5 and 6 are the cumulative runoff (summed across all 15 basins) for each day in the seasonal forecast. This is specified in the revised text and figure captions**.** Last Figure 7 is the total seasonal flow across all the basins, which was described in the figure caption but not in the original text. This is now included in the revised version. Summing up the runoff in this way gives the reader an idea of how implementation of ESP-DA would affect reservoir management. Since the runoff from many basins would be available to large reservoirs, this type of a forecast would be more valuable than the individual basins. The results will differ from basin to basin and year to year but we believe that our presentation is prudent given the demands on operational seasonal forecasting. Also, we removed June from Figures 8 and 9 because it is known that this month had poor data assimilation. Since we are attempting to draw conclusions about the use of ESP-DA and not just the SNOTEL data assimilation, it was necessary to only use those months in which it is plausible that accurate assimilation was performed. This is reflected in our conclusion that the main limitation of this technique is the lack of a SWE observation that is representative through all elevation bands.

Comment:

P7219L25: it is not clear what reference forecast has been used to compute the RPSS. Section 3.3 suggests that a climatological reference has been used. From my point of view the ESP should be used as a reference forecast.

Reply:

Thanks for the comment. The reference forecast is the climatology as described in equation 8. While it is a viable option to use the ESP as a reference forecast, we believe comparing the RPSS of each with respect to climatology is an equally useful comparison. Both methods show whether there is an improvement in probabilistic forecasting in comparing ESP to ESP-DA.

Comment:

Results and Discussion: The results obtained are not discussed in the light of existing literature. Thus it is hard for the reader to judge the results and improvements obtained by the approach. Further the Discussion and Conclusion section contains repetitions and the conclusions reached are not very significant.

Reply:

The first half of this comment is similar to the first major concern which was responded to above. We explained that this study is significantly different from previous studies. We respectfully disagree with the statement that the conclusions reached are not very significant. The importance in our conclusions is illustrated most clearly by the rank histogram and the Q-Q plot. Both show that the regular ESP had a tendency to overpredict the seasonal streamflow volume. This is quite problematic because reservoir operators would have expected larger flows than were actually received. ESP-DA much more accurately predicted the uncertainty in seasonal flow and reservoir operators would have therefore benefitted from the new forecast during the months in which snow data assimilation was effective. This reasoning is briefly explained on lines P7220 L29 through P7221 L2 of the original version of the paper.

Response to minor comments:

Comment:

P7209L23: Andreadis and Lettenmaier (2006) paper not in references list

Reply:

Thanks for the observation. In addition to Andreadis and Lettenmaier 2006, Matgen et al., 2010 was cited in the text but not in the reference section. These were both added to the reference section in the revised paper.

Comment:

P7215 - 3: why do you use 500 ensemble members, is this recommended?

Reply:

500 ensemble members was found to be suitable through experimentation. This is not necessarily recommended but for this specific application was found to be suitable.

Comment:

References: 9 out of 34 refs are written or co-authored by the authors

Reply:

We believe that the number of references written or co-authored by the authors is warranted given the relevance of our previous studies in both data assimilation and ESP to our current research.

Comment:

Fig. 1 Some elevation information might be useful, Eventually use a hill-shading.

Reply

While it may be helpful to give more information about the elevation through hill-shading in figure 1, this would make much of the other information (i.e. location of the basins and SNOTEL station) difficult to see. For the purpose of this study, we believe that an examination of Table 1 and Figures 1-2 (and an additional figure as suggested by reviewer #2) provide the necessary information to show the SNOTEL representativeness. (Note the updates in Table 1 and Figure 2 and the new figure/discussion)

Comment:

Fig. 2 text size and bar widths are not proportional, add elevation levels to xlab annotations

Reply:

Thanks for the comments about Figure 2. The text and bars need to be adjusted to look more appealing to the reader and this is fixed in the revised version with some additional information provided in the plot.

Comment:

Fig. 4 display is not very informative

Reply:

Figure 4 is used strictly as a comparison of the spin-up and DA. It is important to show that a distribution of initial conditions is being used in ESP-DA as opposed to the single initial condition used in ESP and how they are related. For this reason we believe that this figure is necessary to illustrate to the reader how the initial conditions in these two approaches are treated.