Italicized text: Reviewer's comment

AR: Authors' response

Comments to reviewer #3 (Kevin He)

The authors present a set of synthetic experiments in assessing the potential added value of assimilating streamflow, SWE, SCA (via EnKF) into the CEQUEAU model in short- to medium-range streamflow forecasting at the Nechako watershed located in BC, Canada. Results indicate that streamflow DA and SWE DA lead to improvements in short-term forecast and medium-term forecast (during snow melt period), respectively. Assimilation of streamflow and SWE simultaneously yields even better results at both scales. However, assimilating SCA does not show any benefit. Overall, the methodology and results are sound and meaningful, yet not innovative. The paper is very well written and organized. I think it will be of interest to the readership of HESS.

My major comment is that, from the perspective of water resources management, the bias of the mean (or median) ESP forecast is typically an important factor considered in water-related decision making (e.g., water supply allocation, reservoir release/hydropower generation schedule, among others). In light of this, when assessing EPS forecast skill, the bias is normally analyzed.

AR: We have been working closely with Rio Tinto and they, like many other water resources managers, also consider the bias to be an important metric (maybe even the most important), especially during the melt period. While this has been computed for all scenarios, as with many other metrics, we felt this actually added little to the discussion that did not justify the doubling of the number of figures for forecasts. This is because we have generated bias-free synthetic observations and meteorological input. This approach will likely always result in a nonzero bias due to the non-linearity of the hydrological model (e.g. two distributions of the same amount of rain can lead to different quantities of cumulated streamflow due to evapotranspiration, etc.) and the finite period used in the study, but it should ideally be very small.

In our case, average streamflow bias is less than 1 % for the open loop compared with the true state. Zooming on each year at a time, simulated and true cumulated streamflow difference oscillates around 5 % on average over the 10 years considered. This is mainly why the assimilation of SWE can lead to some improvement. If bias was always perfect, adding or removing snow would not lead to improvements. Although this bias value is improved in various ways with data assimilation, the window for improving bias remains small however. For the real world Nechako basin, average bias is estimated at around 20 %.

In current works that has not yet been published, biased precipitations are purposely added to test the robustness of the approach. Real data assimilation has also been evaluated. In both of those cases, bias becomes a central part of the results as it is significantly far from null to begin with (there is something to potentially improve upon). As it is currently however, we feel that adding figures of bias would contribute very little to the discussion due to the near-ideal framework used in the synthetic experiment.

In the case of this study, the score MSSS is applied in the sensitivity analysis part (Figures 4, 6 and 7) but not the forecast part (Figures 8-12). The relevant results should be added (either in tabular or graphic form) and discussed.

AR: The use of MSSS during the sensitivity analysis part and CRPSS during the forecast part was done to facilitate comparisons between other similar studies. The MSSS and CRPSS are relatively similar metrics used in different contexts. The mean square error (non-normalized version of the MSSS) is often used as a metric duration calibration or comparison between two curves, while the CRPS is mostly used to evaluate ESPs since it is adapted to ensembles. In the limit where an ensemble reaches 1 member, the CRPS simply becomes the mean absolute error. Although it is not exactly the same as the mean square error, the two carry much of the same information for non-ensemble curves. For ensembles, one should use the mean or median to compute the MSSS, which could yield very different results than the CRPSS since the latter is sensitive to the precision of the ensemble and not only its accuracy. However, in our case, the two respond in a very similar fashion as can been seen in fig. 1 (below) in this response. Similar graphs can be generated for all comparisons made using the CRPSS, but we believe this would add very little to the discussion.

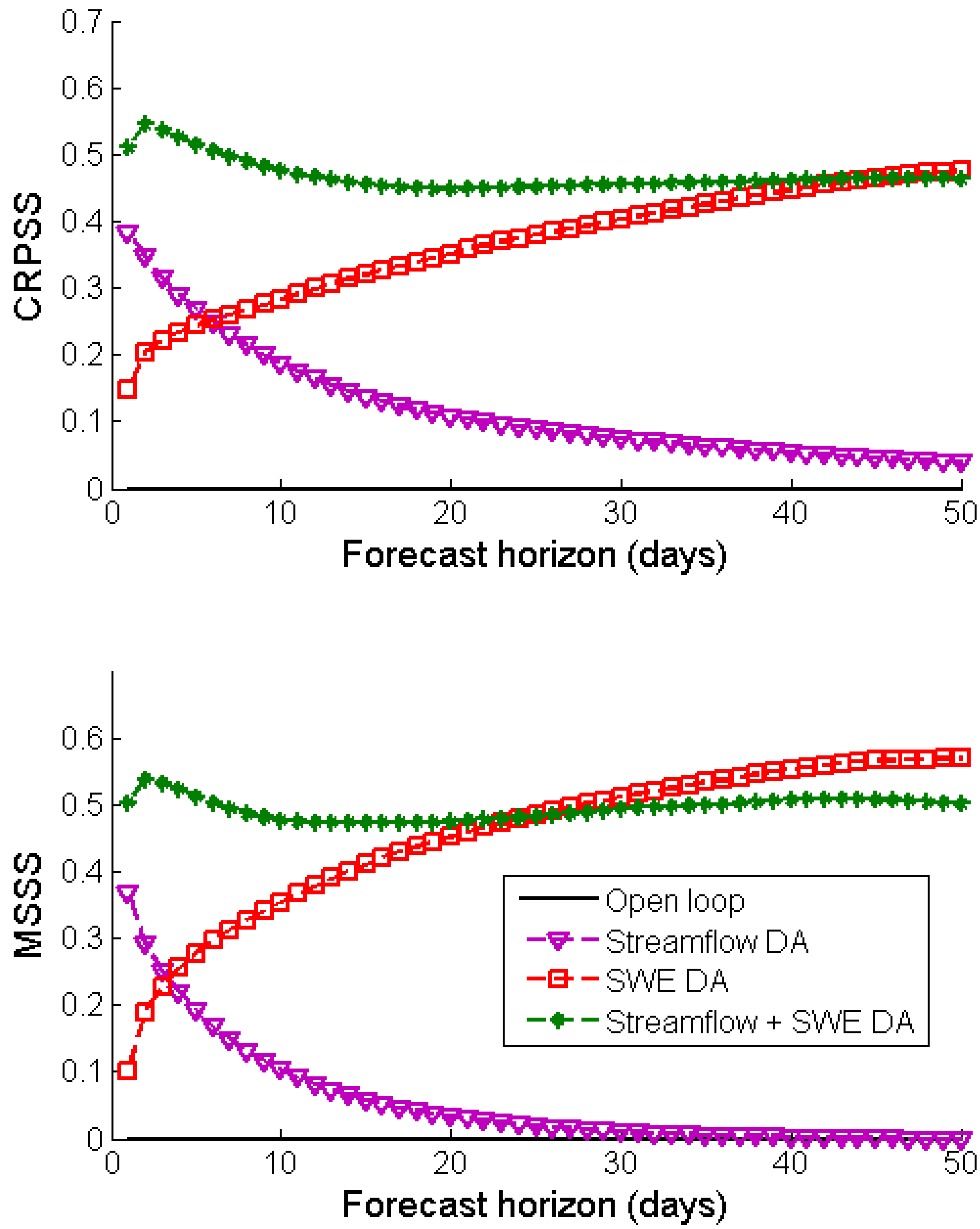

Figure 1. CRPSS and MSSS relative to the true state for a sample of forecasts.

My minor comments include 1) the authors need to be clear about how often the forecasts are issued (every day, once a week, or once per month in the study period from 8/15/1990 to 8/14/2000). If it is once a month, the authors need to discuss the sample size issue (10 years) when discussing the skill scores;

AR: This has been clarified in a new section (3.1.3) of the revised manuscript dedicated to ensemble streamflow predictions.

2) Line 7 of Page 2, "Franz" should be "Franz et al.";

AR: This has been corrected in the revised manuscript.

3) Lines 26-27 of Page 3, august should be August; delete "to";

AR: This has been corrected in the revised manuscript.

4) Line 8 of Page 4, (Fig. 2), Army Corps of Engineers;

AR: This has been corrected in the revised manuscript.

5) Lines 14-23 of Page 6, other than use (1), (2),. . ., it would less confusing when using (#1), (#2), . . ., or (Step 1), (Step 2), etc. ;

AR: This has been modified in the revised manuscript by using the term "step" before each number.

6) Line 5 of Page 7, delete "to";

AR: This is a typo and has been replaced with "the".

7) Line 6 of Page 8, modify "than";

AR: This has been corrected in the revised manuscript by removing that part of the sentence. This was done in rearrangement of the sections to split the information relating to the generation of synthetic observations and observation ensembles.

8) Lines 20-22 of Page 8, rework on the sentence;

AR: The sentence has been reworked in the revised manuscript to:

“This approach avoids introducing a systematic bias when assimilating SCA values at 0 or 100 %. When values are at 0 % (or 100 %), perturbations can only introduce higher (or lower) values in order to remain within the physical limits of the observations. This approach also gives the observations a greater uncertainty during the transition periods when SCA, which loosely follows the greater uncertainty attributed to MODIS observations over the same period (Hall and Riggs, 2007).”

9) Line 12 of Page 10, change MSS to MSSS;

AR: This has been corrected in the revised manuscript.

10) Line 20 of Page 10, in order to.

AR: This has been corrected in the revised manuscript.