Interactive comment on "ENSO-Conditioned Weather Resampling Method for Seasonal Ensemble Streamflow Prediction" *by* J. V. L. Beckers et al.

Anonymous Referee #1

Received and published: 7 March 2016

Summary:

In this paper, the authors propose a technique that combines a post-processing step – i.e., sub-sampler of raw ensemble streamflow prediction (ESP) outputs based on climate index similarity – with a pre-processing step that generates synthetic precipitation and temperature time series via resampling, based on climate index similarity, to force hydrologic model simulations and re-populate the previously sub-sampled ensemble forecast. The method is applied in three catchments located in the Pacific Northwest, using the SAC-SMA and Snow-17 models, for seasonal (May-June) streamflow forecasting. The authors conclude that their framework is an improvement in skill (RMSE, Brier Score and Continuous Ranked Probability Score) over both standard ESP and

climate-based subsampling.

The paper is in general well written and well organized, the proposed technique is scientifically sound and the results are quite interesting. Further, the connection with the existing literature on this topic is nicely conducted. In my opinion, the manuscript has a lot of potential for publication in HESS, but the authors need to clarify some methodological choices, revise some statements, and include omitted results to show if the method is actually robust.

Major comments:

1. Why didn't the authors include the results for improvement in skill (as in Figure 9) for Libby and Hungry Horse? I think that showing the results at these locations is critical to demonstrate that the proposed technique is an advance over raw ESP and climate-based subsampling (see comment #14 for more details on this).

2. P4, L29: It is inferred from this paragraph that the reference date is set to the day when the forecast is initialized. Further, it is also mentioned that "the year of the reference date even has the highest probability of being re-selected". However, later in the paper the authors mention that "the year of reforecast was excluded from the subsampling and resampling schemes" (P8, L24). These statements are confusing, so the authors should clarify what was actually done. In my opinion, the year of the reference date (or initialization time) should NOT be included in the subsampling/resampling procedures, since that year is the one forcing the forecast.

3. P8, L2: The authors state that "several climate mode indices and combinations of indices for ensemble member selection and conditioning of the subsampler were evaluated". However, from the same paragraph it is implied that MEI was selected because it provided the highest correlation with historical streamflow. Did the authors actually test several combinations of climate indices? Moreover, it has been shown that PDO strongly affects interannual variability of runoff in this region (e.g., McCabe, G.J., Wolock 2014; Sagarika et al. 2015). Did the authors perform any experiments

including both MEI and PDO in the subsampling process? I think this manuscript would greatly benefit if - at least for the subsampler method - additional experiments showing the use of PDO were included. My guess is that the poor results obtained at Libby may be related to this issue.

Minor comments:

4. P1, L23: The authors should note that the hydrologic model does not necessarily have to be conceptual in ESP frameworks.

5. Throughout the manuscript: the authors refer to "reforecasts" or "forecasts in retrospect" when reporting results, but it might be better to use the word "hindcasting" (Beven and Young 2013).

6. P2, second paragraph: the text may be enriched by adding a few more references (Hamlet and Lettenmaier 1999; Tootle et al. 2007; Abudu et al. 2010; Sagarika et al. 2015).

7. P2, L18: Several studies recommend developing custom climate indices for the basin(s) of interest using reanalysis datasets (e.g., Grantz et al. 2005; Regonda et al. 2006; Block et al. 2009; Opitz-Stapleton et al. 2007; Bracken et al. 2010; Mendoza et al. 2014), instead of using standard climate indices for predicting seasonal runoff volumes. This point could be made in the introduction.

8. P2, L21: The reference is missing here.

9. P5, L17: A better title for section 3 would be "Example Application".

10. P7, Table 1: It would be more informative to add mean basin elevation (or elevation range), mean annual runoff and mean annual precipitation (mm/yr), and runoff ratio. I think that powerhouse capacity is not relevant here.

11. I strongly encourage the authors to improve the quality (resolution) of Figures 1, 4, 5, 7 and 8. This is critical to enhance the readability of the paper.

C3

12. Figures 7 and 8: The authors could merge the results displayed here into a single figure, using different colors for different methods (for instance, red for subsampler, and black for combined subsampler-resampler), and keeping the title of x-axis label as "Number of historical years in ensemble". This would allow a direct comparison between the proposed method and the benchmark technique (i.e. only sub-sampling). I also think that the authors should add two additional panels (similar to the one described) with results of CRPSS – which is in my opinion a much more interesting score to assess the skill of ensemble systems – and RMSE. Further, it should be mentioned in the caption that results are averaged over lead times of 1-12 months.

13. Figures 7-9: The captions indicate that results are for May-June flows, but the text refer to June flows. What is actually being presented? If results are for May-June flows, are these aggregated (i.e. how many values are used for computing the scores, Nyears or 2 x Nyears)? Is the 80% flow computed from all monthly streamflow values, or only from May and June historical flows?

14. Figure 9: As pointed in comment #1, the authors are encouraged to add and discuss results for Libby and Hungry Horse in this figure. This could be done by or adding two panels (b and c, for instance), or extra lines with different colors for each basin. The improvement in skill could also be compared to that obtained from using only sub-sampling (the benchmark method) to understand the added value of re-populating the ensemble.

15. P13, L10-16: The authors might want to re-word or delete a couple of sentences. For instance, they point for Figure 8 that "in contrast to Fig. 7, the BSS for all test basins are now positive over the full range", which is NOT true for the Libby reservoir (there are still negative BSS values). Moreover, the authors mention that "a mix of 10 historical years from the subsampler ESP and 40 additional resampled traces produces the best result for these sub-basins", which is innacurate again when looking at Libby (higher BSS is obtained using five historical years).

C4

Suggested minor edits:

16. P1 L23: "forcing" -> "forcings".

17. P2, L27: "case study" -> "case study basin".

18. P2, L26: "weigh" -> "weight".

19. P3, L19-21: "Sect." -> "Section".

20. P5, L13: "needs" -> "need".

21. P7, L9: "of e.g." -> "with"; "into the states" -> "into model states".

22. P8, L1: "parameter tuning" -> "parameter calibration".

23. P12, L18: "the most variation" -> "the largest variation".

References:

Abudu, S., J. P. King, and T. C. Pagano, 2010: Application of Partial Least-Squares Regression in Seasonal Streamflow Forecasting. J. Hydrol. Eng., 612–623.

Beven, K., and P. Young, 2013: A guide to good practice in modeling semantics for authors and referees. Water Resour. Res., 49, 5092–5098, doi:10.1002/wrcr.20393.

Block, P. J., F. A. Souza Filho, L. Sun, and H.-H. Kwon, 2009: A Streamflow Forecasting Framework using Multiple Climate and Hydrological Models. JAWRA J. Am. Water Resour. Assoc., 45, 828–843, doi:10.1111/j.1752-1688.2009.00327.x.

Bracken, C., B. Rajagopalan, and J. Prairie, 2010: A multisite seasonal ensemble streamflow forecasting technique. Water Resour. Res., 46, W03532, doi:10.1029/2009WR007965.

Grantz, K., B. Rajagopalan, M. Clark, and E. Zagona, 2005: A technique for incorporating large-scale climate information in basin-scale ensemble streamflow forecasts. Water Resour. Res., 41, W10410, doi:10.1029/2004WR003467.

Hamlet, A. F., and D. L. Lettenmaier, 1999: Columbia River Streamflow Forecasting Based on ENSO and PDO Climate Signals. J. Water Resour. Plan. Manag., 125, 333–341.

McCabe, G.J., Wolock, D. M., 2014: Spatial and temporal patterns in conterminous United States streamflow characteristics. Geophys. Res. Lett., 41, 6889–6897, doi:10.1002/2014GL061980.Received.

Mendoza, P. A., B. Rajagopalan, M. P. Clark, G. Cortés, and J. McPhee, 2014: A robust multimodel framework for ensemble seasonal hydroclimatic forecasts. Water Resour. Res., 50, 6030–6052, doi:10.1002/2014WR015426.

Opitz-Stapleton, S., S. Gangopadhyay, and B. Rajagopalan, 2007: Generating streamflow forecasts for the Yakima River Basin using large-scale climate predictors. J. Hydrol., 341, 131–143, doi:10.1016/j.jhydrol.2007.03.024.

Regonda, S. K., B. Rajagopalan, M. Clark, and E. Zagona, 2006: A multimodel ensemble forecast framework: Application to spring seasonal flows in the Gunnison River Basin. Water Resour. Res., 42, W09404, doi:10.1029/2005WR004653.

Sagarika, S., A. Kalra, and S. Ahmad, 2015: Interconnections between oceanic–atmospheric indices and variability in the U.S. streamflow. J. Hydrol., 525, 724–736, doi:10.1016/j.jhydrol.2015.04.020. http://www.sciencedirect.com/science/article/pii/S0022169415002693.

Tootle, G. A., A. K. Singh, T. C. Piechota, and I. Farnham, 2007: Long Lead-Time Forecasting of U.S. Streamflow Using Partial Least Squares Regression. J. Hydrol. Eng., 442–451.