Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2016-72-RC2, 2016

Hydrology and Earth System Sciences
Discussions
Open Access

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

Anonymous Referee #2

Received and published: 14 March 2016

This paper presents a two-pronged approach for conditioning ESP forecasts on ENSO conditions. In the first step, a sub-sample of ESP forecasts are selected from an ensemble (e.g. of size 50) by conditioning on a climate index. This reduces the number of ensemble members. In the second step, the ensemble is augmented to the original size by sampling precipitation and temperature from the historical record, conditioned on the climate index, and thereafter producing additional ESP forecasts.

I think the paper presents a pragmatic approach to incorporating climate information into ESP forecasts and for enlarging the ensemble size. These types of technique are of wide interest in the hydrologic ensemble forecasting community.

The writing is generally of publication quality but several figures need improvement.

Printer-friendly version

Discussion paper

I have some issues with the clarity, execution and explanation of the science. If the authors can thoroughly address the issues, some of which are not simple, my opinion is the paper should eventually be published in HESS.

General comments 1) A number of parameters are tuned on the basis of subjective analysis for the whole period of interest. Because this is a forecasting paper, the parameter values ought to be determined from an objective analysis that can then be cross-validated using a leave-out scheme. If the results are not cross-validated then the results are potentially inconclusive. Given that the results are marginal, and perform best for the period tuned to (4–6 months lead time), I suggest this is quite important. Ideally, the following elements would be cross-validated: a. The climate index selection b. The number of optimal ESP sub-samples selected c. The "weight", w. If cross-validation isn't used, justification is required.

2) The results use the Brier score (for 80% exceedance probability forecasts) and CRPS as probabilistic measures. I think the paper would be much stronger if accuracy skill and reliability results were separated. Whether skill is attributable to accuracy or reliability or both may vary significantly with lead time. Also, it is stated repeatedly throughout the paper that a small effective number of ensemble members is associated with "degradation of the statistical properties" of the ensemble forecast. What exactly does this mean? I suggest be specific and explain exactly which properties are affected and how they are affected. This is particularly important in the results (P15 L6) and discussion (P15 L18–19).

3) The resampling approach performs poorly for short lead times. Particularly, as shown by Figure 9, the forecasts at short lead times are up to 16% worse. The resampler produces much too narrow forecasts for the first couple of months. This is a problem with the ad-hoc nature of the approach, the spread in the ensembles at any given lead time could be either too narrow or too wide or somewhere in between. What happens if the resampling begins several months prior to the forecast date (i.e. lag 2 or lag 3 MEI)? It's a hard sell to say that forecasts get worse as lead time shortens. At

what point should the forecasts be ignored? I encourage a resolution.

Specific comments 4) Abstract, last sentence: This needs to explicitly say when and where improvements of up to 10% are found and probably should also say that the results for short lead times are worsened.

5) P4 L1 suggests selecting climate indices based on correlations with MAT/MAP. But P8 L4–5 reports MEI was selected on the basis of correlation with streamflow. Please make more consistent.

6) MEI is a two-month index. Were two-month values of the other indices considered?

7) Equation (1): The summation appears to be the squared Euclidean distance (no square root). Also, how are indices in different units handled (is it implicitly through scaling/weighting)?

8) Figure 2: It might be better to show percentile intervals rather than statistics based on normal distributions (unless of course the data is very normal).

9) P13 L10–13: The BSS is marginally negative for some cases for Libby Dam, so the statement saying BSS is positive for all cases needs correcting. Also, re the comment about Figure 8, the text says the BSS is a function of "number of the original ESP members", but I think it means the number of sub-sampled years (hence less than 50 on the x-axis is Figure 8).

10) The introduction states that section 5 summarises and concludes the paper, but section 5 is headed "Discussion". Suggest renaming.

11) P15 L10 should say in two of the test basins *at lead times greater than X*

12) P15 L13–15: Operational applications should be flexible enough to adapt to different methods if there's a proven benefit. So this argument doesn't carry a lot of weight.

13) P16 L13–14: I'm confused by this. PDO was apparently investigated already in this study and disregarded.

Technical corrections (typing errors, etc.) 14) Figure 2 and Table 1. Abbreviations do not match for Hungry Horse and Libby Dam.

15) P12 L19 and elsewhere: Text refers to June flow instead of May–June flow.

16) There are some instances of weigh and weighing instead of weight and weighting. Will be easy to find and correct.

17) Improve the figure quality. Many are blurry.

18) Is Figure 5 one figure or four? There are four captions.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-72, 2016.