

Replies to Referee #2

“An intercomparison of approaches for improving predictability in operational seasonal streamflow forecasting”

Pablo A. Mendoza, Andrew W. Wood, Elizabeth Clark, Eric Rothwell, Martyn P. Clark, Bart Nijssen, Levi D. Brekke, and Jeffrey R. Arnold

We thank this reviewer for his time in commenting on our paper. We provide responses to each individual point below. For clarity, comments are given in italics, and our responses are given in plain text.

The manuscript compares different methods for seasonal water supply forecasts in several catchments in the Pacific Northwest region of the US. A large variety of different models was applied: purely statistical methods, methods based on watershed modelling as well as hybrid approaches using initial hydrologic conditions and / or climate information as input. Additionally, different post-processing and merging methods are tested. The snow-dominated test catchments cover a wide range of hydrometeorological conditions and different atmospheric teleconnection signals.

The literature review of the most commonly used methods in seasonal streamflow forecasting is exhaustive and the results are nicely presented and compared. The real value of this study is the comparison of a large variety of methods based on a common hindcast/verification framework using rigorous three years out cross validation. Using such a common framework an objective comparison of the performance of the different methods is possible. The paper is well written and should be foreseen for publication in HESS after minor revisions.

We are very pleased that this reviewer appreciates the contributions of this study.

You should probably use SI units instead of KCFS (Thousands of Cubic Feet Per Second) and MAF (Million acre Feet) (Standard in HESS)

We appreciate this sentiment. However, we consider that the paper would have much more value if the results are presented in units that are familiar to water managers in the US, since this is basically a water supply forecasting study. Therefore, we kindly ask the Editor and Publisher for permission to preserve the current flow units throughout the manuscript.

P 5 line 155: missing first three year period in the brackets could be confusing why it is missing, add period: “... (e.g., 1981-1983, 1984-1986, 1987-1989, 1990-1992, etc.),...”

We have added the first three-year period (1981-1983), following the reviewers’ suggestion, and taken out the part of the series after the first two.

P 5 line 171: “...predictant data are normalized before ...” what do you mean by normalizing in this context? I assume z-scores are calculated or do you apply a normalization method such as box-cox? Please specify!

The reviewer is correct: z-scores are computed using $z = (x - \mu)/\sigma$, where x represents the original variable, and μ and σ represent the mean and standard deviation of the population, respectively. We clarify this procedure in the revised manuscript.

Explain why you have used log-transformation of the predictant data and no other transformation method (e.g. BoxCox, ...).

We regret that we did not try other transformations as we were focused on relative outcomes, though this would have been a reasonable thing to do. In truth, we did place a great deal of importance on the transformation when the work was done, though since then our interactions with CSIRO has opened our eyes to the variation in the effectiveness of different transformations (including, for instance, the log-sinh). We do not have the bandwidth to go back and explore this issue, but for now we will highlight it for the readers based on the text of the comment. Hence, we have added the following sentences:

“In practice, forecasters use a variety of transforms such as linear, square root, cube root, log and log-sinh (Wang et al. 2012). We did not explore alternative transforms, using the log consistently throughout, but recognize that the choice of transform can affect the quality of the forecast.”

P 6 line 201: If the predictand was normalized (subtracting its expected value and dividing the difference by its standard deviation) before as stated above, the predictand has to be multiplied with the standard deviation and the mean has to be added before exponentiation. Is this correct? In this case the explanation of this procedure should be added.

The reviewer is correct, and a proper explanation has been included in the revised manuscript.

P 7 line 210: Replace MRL with MLR

We have corrected the text, following the reviewer's suggestion.

P 7 line 213: "... predicting seasonal predictor average and seasonal streamflow volume..." Is the MLR applied to log-transformed streamflow? Do you normalize the climate indices? Please specify!

The reviewer is correct: MLR is applied to log-transformed streamflow, and then both predictand (flow in log space) and predictors (e.g., climate indices) are normalized. The general procedure used in this paper for statistical approaches is clarified, following this and a previous comment from this reviewer:

“In the statistical approaches, seasonal flows are log-transformed, and predictor and predictand data are normalized before training statistical method parameters or weights (i.e., z-scores are computed using $z = (x - \mu)/\sigma$, where x represents the original variable, and μ and σ represent the mean and standard deviation of x , respectively). The statistical models were applied in log-standard-normal space for forecast generation, then predictands are transformed from z-scores to log space (i.e., apply $x = z\sigma + \mu$, with $x = \log(Q)$), and finally transformed back to streamflow space”.

P 8 line 246: Re-transformation of predicted streamflow should be added as additional step

The step suggested by the reviewer is actually done, and therefore it has been added as part of the method description:

“Ensemble forecasts are transformed from z-scores to log space, and finally exponentiated for conversion to flow space”.

P 9 line 284: Please explain shortly how the weighted resampling using the weights $1/\text{RMSE}$ works.

RWE performs a weighted resampling from the two forecast sources (i.e., the best climate-only and best watershed-only forecasts) according to their skill during the training period. I.e., two weights $1/\text{RMSE}$ are obtained, where RMSE the root mean squared error of the ensemble median. These weights are normalized to make them sum 1, and finally obtain the fraction of the new 500-member ensemble coming from each forecast source. For example, if the resulting normalized weights are 0.4 and 0.6 for the best climate-only and best watershed-only forecasts, respectively, the RWE ensemble will contain 200 and 300 members from each prediction. This explanation has been added to the text in the revised manuscript.

P 24 Table 1: In the table and in the main section the abbreviation RE (runoff efficiency) is used, in the table caption runoff ratio RR is used, please harmonize

We have replaced runoff efficiency (RE) by runoff ratio (RR). Thanks for noting this.

P 34 and p 36: Please add explanation of red line (observation?) to figure caption

Indeed, the red line represents the observed flow volumes. We have incorporated this information to the caption of Figures 9 and 11.

References

Wang, Q. J., D. L. Shrestha, D. E. Robertson, and P. Pokhrel, 2012: A log-sinh transformation for data normalization and variance stabilization. *Water Resour. Res.*, **48**, 1–7, doi:10.1029/2011WR010973.