Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2017-449-RC3, 2017

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

Interactive comment on "Benchmarking Ensemble Streamflow Prediction skill in the UK" *by* Shaun Harrigan et al.

Anonymous Referee #3

Received and published: 30 August 2017

This paper investigates the performance of the ESP forecast method in the United Kingdom. The authors investigate when, where and why the ESP is skillful, based on a set of 314 catchments and 50 years of hindcasts generated with the GR6J model and data from the UK National River Flow Archive. The forecasts are evaluated with a deterministic and a probabilistic criterion, and compared to modelled streamflow climatology. The authors conclude that the skill decreases exponentially with lead time. Higher skill are observed in forecasts initialized in summer months for lead times up to one month, and in winter and autumn months for seasonal and annual lead times. Higher skill is observed in slow responding catchments with high soil moisture and groundwater reservoirs and less skillful in highly responsive catchments.

General comment

Printer-friendly version

Discussion paper

I think that this paper is very well-written and of great quality. The objectives and methods are clearly defined, and therefore easy to read and to follow the scope of the paper. The length of the article and the number of figures were appropriate and the content was always relevant. In addition, this paper fits nicely in the Subseasonal-to-seasonal special issue. This study provides a useful diagnostic of ESP over the UK. I particularly enjoyed how the authors made the link between the spatial and temporal skill patterns and catchment characteristics and seasonal features. I listed some comments and questions below, most of them dealing with methodological aspects, and none of them being major.

Major comments and general questions

In both Twedt et al. (1977) and Day (1985), the abbreviation ESP actually stands for "Extended Streamflow Prediction". It is true that "Ensemble Streamflow Prediction" is widely used, but I think that the original term better conveys the purpose of the method and should be used instead.

P5 L24-25 : "Each of the 51 generated hindcast time-series were then temporally aggregated to provide a forecast of streamflow volume with seamless lead times of 1-day to 12-months, resulting in 365 lead times LT per forecast (leap days were removed)." Do I understand correctly that the streamflow volume for 30 days is obtained by aggregating daily forecasts from day 1 to day 30, and that the streamflow volume for the year aggregates all daily forecasts from day 1 to day 365? If not, could you please clarify? If so, I was confused by the word "lead time" and the analysis involves more factors than just the lead time. Rather than an analysis on lead times, it is an analysis on both aggregation periods and lead times that can be argued to be between 0 days and the last day of the aggregation period. I don't believe this to be real issue, but maybe the authors could be more careful in the way they used the term "lead time". To be more specific, it is the occurrence of "lead times" in Figures 3, 4 and 5 and Section 3.1.1 that triggered this comment.

Printer-friendly version

Discussion paper

P5 L28 : Regarding the implementation of the L3OCV method, I was wondering why the authors excluded the subsequent two years but not the preceding two. My guess would be that, operationally, the preceding two years are always available, in any case, while the succeeding two are still missing on the day of the forecast, and adding them will add missing and non-independent information to the calibration-validation procedure. Could the authors say a bit more on that?

P6 L25-27 : "It was found in testing that ESP skill was artificially advantaged (disadvantaged) if cross-validation was not carried out in historic climate forcings (benchmark forecasts), in some cases by $\pm$ 15 %." Could you please clarify this sentence?

I was wondering about the authors' choice to use the MSE as deterministic score in this case. If the purpose of the two scores is simply to distinguish between deterministic and probabilistic performances, I would recommend using the Mean Absolute Error (the CRPS value of a deterministic forecast is MAE, Hersbach, 2000) so that, when comparing both scores (e.g. Figure 3), the difference in value is solely due to considering the probabilistic side of the forecast.

Still on the evaluation criteria, given that ESP is a probabilistic ensemble that translates the uncertainty from climatology, I would have liked the authors to focus more on the CRPS than on the MSE, e.g. in Figures 6, 7 and, possibly, 8). Was there a reason to focus on MSE instead?

P7 L17-21 : Is the scale defined for MSESS values or CRPSS values? In the interpretation of Figure 6, it also seems that the threshold value for "Very Low" has shifted to (0, 0.1).

Figure 4 and Table 2: To which extent does the performance of GR4J for each month of the year explains the results obtained for short to medium lead times and presented in Figure 4?

Figure 7: Here, I would have liked to see the maps for November which is cited earlier

in the analysis.

Minor comments

P2 L27 : Please change “out to at a least 7-month lead time” to “out to at least a 7-month lead time”

P3 L28 : “132 catchments that are part the new version” to “132 catchments that are part of the new version”

P6 L2: Please change “initilisation” to “initialisation”

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-449, 2017.

Printer-friendly version

Discussion paper