Review of "Seasonal streamflow forecasts for Europe – II. Sources of skill" by Wouter Greuell et al.

This paper looks at the sources of skill in the WUSHP model-based seasonal hydrological forecasting system, producing hydrological forecasts for up to seven months of lead time over Europe. To this end, the authors first analysed the skill of the meteorological forcings of WUSHP (ECMWF's System 4 seasonal meteorological hindcasts). They also produced several hindcast datasets, based on which they carried out a complete investigation of the sources of skill (i.e. initial conditions of soil moisture and snow and meteorological forcings) in the seasonal runoff, discharge and evapotranspiration hindcasts.

The results presented in this paper are based on a thorough and original analysis and provide a great contribution to the field's existing literature. Furthermore, this paper is overall coherently written and I recommend it to be accepted after minor revisions. Below, a few comments which should hopefully guide the authors in revising this paper for publication.

Main comments:

- Since the last version of this paper, the authors do not seem to have updated their references list with the latest literature within the field, more specifically the papers published in the same HESS special issue on "Sub-seasonal to seasonal hydrological forecasting". This is for instance apparent on P26 L841-844, where the authors mention that to their knowledge, only two other studies looked at the sources of skill in seasonal hydrological forecasts produced by a similar system, over Europe. There are several papers in the same special issue which look at the skill of (System 4-driven) seasonal hydrological forecasts over the globe, Europe, or parts of Europe. These papers all benchmark the skill of the state-of-the-art seasonal hydrological forecasts to forecasts such as the ESP. I would highly recommend for the authors to update their references with this latest literature.

- This paper is very rich in results and figures. I would recommend for the authors to present the results in a more concise manner when possible and to remove non-essential figures from the main body. This would help highlight the key results of this paper (in my opinion and according to the paper's title, the analysis of the origins of skill in the runoff hindcasts) and keep the readers' focus throughout the paper. I have made below a few more specific comments about this.

- The legend is too small for most of the sub-figures.

Title: in my opinion, the title doesn't fully reflect the content of the paper. Firstly, this paper looks at runoff and discharge and the word "streamflow" is not used much in the paper. Furthermore, this paper presents much more than just a streamflow analysis. I would therefore recommend changing the title to something more representative of the content of the paper, for example: "Seasonal hydro-meteorological forecasts for Europe: sources of skill".

Abstract: the abstract is overall too long and dives into the results in great detail before explaining the methods used. I would recommend laying out the methods used more clearly before mentioning the results. The results' paragraph should be much shorter and highlight the key results of the paper, leaving the details for the main body. You use the word "streamflow" here and on a few occasions in the paper (e.g. P30 L973) while the rest of the time the words "runoff" or "discharge" are used. Please consider removing the word "streamflow" from the paper or clarifying what it refers to.

P4 L95-96: please specify here which variables this analysis is based on.

P4 L109-110: you need to explain here what the ESP is for readers not yet familiar with it.

P4 L109-116: it would be good to explain more clearly here that the hindcasts you produced for this paper are inspired from the conventional ESP and reverse-ESP, but that they differ in their set up and why that is (with regards to the overall aim of the paper).

P4 L118: could you please reference here one or multiple papers that have looked at the effect of evapotranspiration on runoff, more specifically on seasonal timescales.

P5 L165-166: "a length of seven months" is ambiguous. Using the word lead time would be clearer.

P6 L195-196: do the observations you mention here refer to the pseudo-observations?

P7 L235-236: this explanation of the target month number is slightly confusing and not necessarily needed. You could simply say that the target month refers to the month for which the forecast is made. I also find jumping between the terms "lead month" and "lead time" throughout the paper a bit confusing.

P7 L245-251: please remind the reader that in this paper you use alternative methods to the standard ESP and reverse-ESP (widely used in the literature) because you want to keep these hindcasts as close as possible to the FullSH for the aim of this paper.

P8 L252-271: I like the experiments' new names (after the first revision of the paper). Does snow conditions refer to the snow cover?

P8 L274: do you mean "random" instead of "uneven"?

P9 L308: please clarify what you mean by "most important input variables". Most important for what?

3.1: did you look at the effect of trends in the System 4 precipitation hindcasts?

Figure 1: in the caption, it says that "yellow cells have insignificant skill". There are however multiple shades of yellow on the map. You could instead say that the lightest yellow colour shows insignificant skill. P10 L328-329: for which lead months?

P12 L372-373: the lead month 0 line for the un-detrended hindcasts looks very close to the line for the detrended hindcasts, except for April, June and November.

P12 L373-374: it however doesn't drop as quickly as for precipitation.

P12 L395-396: I agree with this observation for JJA, while for the other months it is not so evident.

P13 L415: why was lead month 5 selected to illustrate this point? Do the other lead months show (dis)similar results?

P13 L421-423: for the sake of conciseness, I would maybe remove these results.

P13 L428-429: could this be due to some trend or predictor that takes shape in autumn and affects temperatures in Europe in winter? I wouldn't dismiss it as spurious.

P14 L472-474: the difference is however not very significant, except in late summer-autumn.

Figure 4: I would suggest moving the bottom two sub-figures to the supplementary material as it won't affect the main storyline of the results.

Figure 5: green and red should never be used together on a plot. A more colourblind friendly palette should be used instead (same for Figure 13). Since you mention the FullSH in the results that correspond to this figure it would make sense to add a line for the FullSH here as well.

P17 L537-547: these are great results!

P17 L549-550: could that be because the snowpack is at its maximum around February in Europe?

P17 L565-569: I wouldn't mention these results here unless you explain what the differences could be due to.

Figure 6: is this figure essential in the main body? You only discuss the additivity of skill from the initial soil moisture and snow conditions, which Figure 5 already shows as an average over Europe. In my opinion, the text is sufficient here.

P18 L583-585: this sentence is slightly confusing.

P19-20 L619-620: could this be rather due to the groundwater initialisation?

Figure 8: this figure is not necessary and a few sentences are sufficient to raise this point.

P20-21 L637-653: couldn't this be due to the fact that we can expect most of the snow in Europe (except for high mountain ranges) to have melted already by May? Knowing the snow cover at the start of May is therefore of not much added value for forecasting future runoff compared to knowing what it is earlier in the year, in April for example (when more snow is still present and available for runoff).

3.3: this part is too long and steals the spotlight from the runoff section, which should be the highlight of the paper according to the title. I would therefore suggest to summarise the text corresponding to Figures 10 and 11 in just a few lines and remove these figures from the main body.

P22 L695-698: I would describe what is observed on sub-figure 9b before comparing those results to the ones obtained for runoff.

P22 L700-703: please mention that these results correspond to the FullSH.

P26 L819-822: many studies (published in the same special issue) found higher skill in the ESP (compared to a state-of-the-art seasonal hydrological forecasting system) beyond the first or second month of lead time over Europe or for specific basins/regions of Europe and are worth mentioning here.

P29 L935: I am not sure to understand what is meant by "inter-member variability".

P29 L942-944: one further difference is that the initial conditions in the reverse-ESP are the full range (or ensemble) of historical initial conditions, instead of a single value (i.e. climatological average).

P30 L953-954: this is a repetition of what was said in the previous paragraph.

P31 L1007-1013: this is a very good point!

P33 L1080-1081: I would remove this example from the conclusions.

Technical corrections:

- P14 L445: lead month 1 instead of 0?

- P16 L524: “run-off” is used instead of “runoff” (comes up again after).

- P22 L686: the “o” is missing in “MeteoSH”.

- P27 L879: “taken from” instead of “taking from”.