Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2017-449-RC2, 2017

Hydrology and Earth System Sciences
Discussions

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

G. Thirel (Referee)

guillaume.thirel@irstea.fr

Received and published: 30 August 2017

This manuscript presents an evaluation of ESP over the UK. The ensemble forecasts are based on the lumped conceptual GR4J model and past P and PET observations that were resampled as used as input to GR4J. These forecasts are compared to proxy observations (GR4J streamflows using P and PET observations) and a benchmark (resampling of these GR4J streamflows).

This paper is generally well written, very clear, and it makes a significant contribution to the HESS journal. However, I of course have some remarks that would deserve some attention from the authors, some of them not being minor. I am convinced that the authors will be able to handle that efficiently and allow the paper to be published.

Major comments:

The way ESP is thought of in this manuscript is a bit old fashioned in my opinion. It is true that first ESPs were using IHCs and past data, but this is not really the standard nowadays. Indeed, the standard is more what is called in the article NWS ESP. These forecasts are now a well-established method and are the reference, especially up to a month of lead time. I would advise the authors using a more modern terminology in the abstract and article or at least being more specific. Moreover, the justification of the choice of this method should be given.

IHCs influence is high for short lead times and low for large lead times. Following the authors' sentence (P. 8, L. 2-4), that would mean that for short lead times, MSESS and CRPSS should be closer than for long lead times. However, we don't see that on Fig. 4, all lead times seem to have a similar difference between both SSs.

Section 4.1.2: this analysis is interesting. However, there is a second possible entry, in addition to the initialisation month, to take into account in my opinion: the lead time month. Indeed, some periods of the period are easier to predict (typically in between seasons are more prone to changing weather, which is difficult to predict sometimes); that may reflect on the scores, and could explain the differences that are highlighted here. Moreover, some scores can be impacted, for instances, by the streamflow characteristics. It is known that Nash-Sutcliffe (not used here) is higher for rivers with strong seasonality, or that CRPS is impacted by the streamflow magnitude (Trinh et al., 2013). I'm wondering to which extent the seasonal analysis (but also the spatial analysis actually!) can be impacted by such issues.

P. 9, L. 21-22: X1 is the production store capacity, and X3 the routing store capacity. It seems difficult to actually link them directly and specifically to soil and groundwater. However, their sum can be considered of the maximum amount of water in the basin (excluding the water in the river and snowpack) and as such it could be of interest including it in Fig. 8.

Section 4.3 aims at finding factors for skill in the model. Did the authors check if the initial states of the model show a correlation with skill? For example, the initial amount of water in the basin, S + R in Fig. 1 of Perrin et al., 2003 (production store + routing store fillings) and the initial snow pack (if a snow model is used) can give good insights (see Singla et al., 2012).

Minor comments:

Abstract: there is a mix between present tense and past tense. Line 14: missing S at ensembleS. Also, lines 21-22 there is a mix between lower, lowest, higher and highest. It is not known from the abstract what the rho symbol represents.

P. 3, L. 21: Section 5 should be Sect. 5 to be consistent with the other occurrences.

P. 3, L. 28: please check all fonts sizes

P. 6, L. 2: initialisation is misspelled

P. 6, L. 3: at p. 5, L. 21, m is the ensemble, not the ensemble size. Also, LT means lead time, it is therefore better not to use LT for designing the number of lead times

P. 6, L. 4: no need for volumes, I think that streamflow is enough

P. 6, L. 15: remove the comma after Wilks

Section 4.1.1, P. 7, L. 26 and later on: do we really need such a precision for all the scores?

P. 9, L. 6: replace “is” with “in” (I think). In this section, percentages sometimes have a space between the figure and the percent sign, sometimes not.

P. 9, L. 13: is “E” actually “SE”?

P. 12, L. 4-6: yes, that definitely has an impact in some basins!

Ghannam et al. reference has some misspelling in the authors’ list

Table 1 caption: I would add “R package (Coron et al., 2016, 2017)” after “airGR” and “(Perrin et al., 2003)” at the end of the caption

Table 2 caption: please remind the GR4J calibration period for the parameters that are given here.

Figure 3: I think that “short”, “extended”, “monthly”, “seasonal” and “annual” should be indicating more precisely what they refer to. Maybe use some arrows for this.

References:

Singla, S., Céron, J.P., Martin, E., Regimbeau, F., Déqué, M., Habets, F., Vidal, J.-P. Predictability of soil moisture and river flows over France for the spring season (2012) Hydrology and Earth System Sciences, 16 (1), pp. 201-216.

Trinh, B.N., Thielen-del Pozo, J., Thirel, G. The reduction continuous rank probability score for evaluating discharge forecasts from hydrological ensemble prediction systems (2013) Atmospheric Science Letters, 14 (2), pp. 61-65.

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