HESSD

Interactive
comment

Interactive comment on "Seasonal streamflow forecasts for Europe – II. Explanation of the skill" *by* Wouter Greuell et al.

Wouter Greuell et al.

w.greuell@hetnet.nl

Received and published: 14 April 2017

We are very happy with the fact that the editor found three anonymous referees to give their highly informed opinions on our paper. We thank all three for their respective efforts to produce such extensive and constructive reviews.

We adopt most of the suggested textual improvements and specified our action to every remark in the annotated report hess-2016-604-RC2-author-reply.pdf1.

RC2 raises concerns regarding 6 technical issues (RC2's numbering): 1. Choice of metric used to present most results. We use R mostly, RC2 prefers the use of RPSS. [….] [. . . discuss merit of ROCSS rather than RPSS. . . in context of early warning of extremes. . .] We believe that RPSS underestimates skill as it addresses all percentiles, whereas it is well known that skill for the extreme percentiles is generally higher and

more relevant to the user than skill for the central percentiles. Similarly, we consider ROCSS for AN and/or BN terciles more relevant than for NN (and these AN/BN are nearly always higher as well. Though the spatio temporal patterns of the dynamics of skill are similar between the three metrics R, RPSS and ROCSS, we agree with RC2 that their statistical significance levels are not necessarily similar (generally the fraction of EU with sign. scores decrease from R > ROC-AN/BN > RPSS » ROC-NN, see figure SM2 paper 1). So first we will better explain how significance is computed. Next we will much better illustrate, describe and explain in the main text of paper 1 the spatio temporal differences found, between the various skill metric, adding to its supplementary material also maps for the other skill scores. In paper 2 we believe this is less relevant as this is about (dynamics of) the sources of skill. However, we will check how different fig 4 and 5 of paper 2 become when using RPSS as skill metric, and add these (in the main paper or SM) depending on the outcome of this check.

2. RC2 is the only reviewer that recommends the analysis of reliability in addition to the other metrics already assessed. We will make a basic assessment on the reliability of WUSHP forecasting system, and present its results in paper 1. Depending on its outcome we might come back to some aspects of reliability in paper 2.

3. We do not fully understand this comments as it seems to mix up some different things. E.g., the quantile mapping referred to in paper 1 is applied to the forcing data, and has nothing to do with VIC model calibration. The latter implies calibrating some model parameters using observed stream flows. We are aware some authors (e.g. Crochmore) also apply a 'leave one year out method' to bias correction (i.e. determine BC factors from other 29 years and apply it to year x). Though, there may be some theoretical merits to do it this way, we believe it practical effects are very small, and as a result many studies derive BC factors from the same set of years as to which they are subsequently applied. We are not sure how cross validation applies to other part our methods. E.g. the 'leave year x out' is not relevant to our ESP implementation since we take the forcing from the hindcasts , not from the reanalysis, so they are never identical

between ESP and reference.

4. We will better explain how any trend in the observations (whether it is due to climate change signal is in principle not relevant, so we will stress this aspect less), that is reproduced in the hindcast, will (mathematically, not physically) lead to an increase in skill.

5. and 6. Yes, we do propose and use an alternative form of both rev-ESP and ESP experiments. With our form of the rev-ESP we assess the value of the actual meteorological forecast quality; not the uncertainty in the forecasts due to uncertainty in the meteo forcing, which is the goal of many of the published rev-ESP experiments. This while assuming we have no knowledge of Initial Hydrological States. So we set IHCs to values unrelated to actual IHCs, by using climatological means. We agree some correlations between IHC might get lost, and thus decrease skill. But sampling IHCs from pseudo observations (i.e. historical simulations), as in the original will also decrease skill as dry IHCs might be used for meteorologically wet years and vice versa, which is less a problem when using means, so the average difference between actual and used IHC will be smaller. We will better discuss the implications of the alternative rev-ESP and ESP in the discussion sections, and like suggested by RC2 in his comment 6, to prevent confusion we will give it a different name. [see also our reply to RC1 comment #2]

Other issues raised by RC2 are mostly minor and responded to in the annotated pdf.

Please also note the supplement to this comment:
http://www.hydrol-earth-syst-sci-discuss.net/hess-2016-604/hess-2016-604-AC2-supplement.pdf

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-604, 2016.