Response to reviewer 2

B. Klein

The manuscript shows the development and the skill of a seamless hydrological forecasting system from sub-seasonal to seasonal scales. Meteorological forecasts from ENS extended (day 1 – 46) and SYS4 (47 to month 7) are merged by randomly selecting ensemble members of SYS4 after ENS extended ends. The skill analysis shows that most of the skill improvement by using SEAM is due to the more frequent model initializations and the more recent NWP model version of ENS extended. The paper is well written, the methodology and results are nicely presented and compared. The real value of this study is the application of products off the shelf (available operational products). Hence the results can be directly incorporated in real-time operational streamflow forecasting practice. The paper should be foreseen for publication in HESS after minor revisions.

Thank you very much Bastian for the comments. Below are detailed point-to-point answers to your remarks, but we have also taken on board your comments of expanding the discussion and conclusion part on the predictability and limitations of the two systems and will expand the discussion part further.

Comments:

p 2, l 24: Typo, replace TSYS4 with SYS4

The typo has been corrected.

p 2, l 30: please add the forecast length published in the seasonal outlook of EFAS

"...with a lead-time of 7 month." was added to clarify.

p 2, l 119: please add possible drawbacks of selecting a random member of SYS4 (one point was raised p 6 l 192- p7 l 195). Another possible drawback could be that ensemble members are combined originating from complete different climatological conditions day 1 – day 46.

We are aware of this problem and we tried to address it on p6, but will expand on this and discuss the drawbacks further. However, the regimes over Europe can shift quite rapidly and it is not certain that matching the ensembles would increase the skill of the seamless.

p 3, l 89: Are the 5kmx5km grid cells of Lisflood further subdivided in elevation zones?

Yes, they are divided into three sub elevation zones to account for differences in snow accumulation and snowmelt. See more details in the answer to P7 below.

p 4, l 124: Are bias/drift correction methods applied to correct the meteorological forecasts?

No bias correction is applied to the meteorological forecasts

p 5, l 135: the description of the hindcast period used in this study is a little bit confusing due to the mixture of forecast dates (2015-05-14 – 2016-06-02) used to produce the hindcast dataset and the forecast dates of the retrospective forecasts. Please clarify!

One possibility would be probably to add the range of forecast dates. Something like: "...the hindcast data set of SEAM covers the period 1995-05-14 to 2016-06-02..." "...the SYS4 re-forecasts used in this study are initialized each month over the period 1995-05-01 to 2016-06-01..."

Thank you for the suggestion. Also Kean commented on the difficulty of understand the setup of the experiment. We have taken care in explaining the hindcast and experiment setup more in detail. We will also add a figure to explain the setup of the hindcast system.

p 5, l 160: replace SEAS with SYS4

Corrected.

p 6, l 161: Incomplete sentence, I assume: ".... as in SEAM to account for the difference in ensemble size...."

The incomplete sentence was deleted

p 7, l 206: Another option of the poor performance of Lisflood in these regions could be the snow modelling component. In steep orography a 5km x 5km grid is relatively coarse to model snow adequately, are grid cells of Lisflood further subdivided in elevation zones? Please add a comment/discussion of the snow modelling performance of Lisflood.

The snow modelling in LISFLOOD is a degree-day method with elevation zones to further differentiate the snow processes in steep orography. This could explain differences in the model performances if the results were compared with observed runoff. However, the model results are compared with a climatology run using observed precip and temperature, and it is more likely that the poor NWP representation of temperature and precipitation are the culprits.However, the snow modelling component could also play a role in this, and we will add a description and discussion on this to the paper.

"The snow accumulation and snowmelt are further divided into three elevation zones within a grip in LISFLOOD to better account for orographic effects in mountainous regions. However, this increase in sub grid resolution is not likely to be high enough to capture the snow variability during the snow accumulation and snowmelt in mountainous regions. Further, precipitation forecasts have documented biases in steep orography (Haiden et al., 2014).

p 8, l 231: add Figure to the figure number "...Cologne (Figure 4)...".

Corrected

p 8, l 233: I assume 3% of its climatological value is derived from the simulated climatology and not from the observed climatology? Please specify!

Yes, it is correct, we are throughout the paper comparing against modelled climatology. We have taken care to make this very clear wherever this is mentioned in the paper. At the above mentioned passage we have changed the sentence to: "went below the 3% percentile of *the modelled* climatological value". Italics denote the addition

p 8, l 240: It should be mentioned that the second low flow event was hit by the SYS4 forecast initialized 2003-09-01. This signal towards a low flow event is missing in the SEAM forecasts published after 2003-09-01. In SEAM a signal towards an extreme low flow event first appears about 3 days before the begin of the event (forecast date 2015-09-14). I would add the real forecast dates to

Figure 4 and not the forecast dates the hindcast data set is produced. This could be a little bit confusing for a reader not familiar to the hindcast procedure of ENS extended.

Yes, and the example is chosen to illustrate a situation where the SYS4 performed well. We also point to the fact that SYS4 does perform well in this particular case in the discussion. However, the higher frequency of the SEAM would give it an advantage when you are closer to the event, since you would get more detailed information about the timing. The following was added to stress the point: “SYS4 does indicate the second low flow with a longer lead time than SEAM. However, SYS4 misses the timing of the event.

Figure 4 was also improved to show more clearly the forecast dates vs the verification dates.
p 8 Conclusion: I miss a discussion of potential improvements of the presented seamless forecasting system. Are there any ideas how to reduce the higher spread of the CRPSS of SEAM compared to SYS 4 in figure 2 c, d? Probably an improvement of the methodology of the concatenation of the forecasts from the two systems? Please add this aspect to the conclusions.

This is a good point, and still be investigated, however outside the scope of this paper. We will add the following to the Conclusions.

“Future work with the seamless forecasting system is to further explore the limits of predictability to assess the strengths and limitations of the current setup. The assumption that the forecasts can be randomly concatenated would also need to be tested against a system where the forecasts are matched according to their respective climatology.”

Another aspect I miss is the conclusion from Figure 2 b):
The improved boundary condition of the first 46 days originating from the more recent model version with a higher resolution doesn’t improve the predictability (forecast skill) after day 46.

This is also a good point, and more is to say on the predictability of the two forecast systems. This will however as mentioned above be dealt with in another study, so we are reluctant to speculate too much at this time.

Figure 3: Are all forecast dates used in this analysis? Please add to the caption to be consistent with the caption of Figure 2.

No, in Figure 3 only the first forecast of the month is used to avoid too much the effect of the initial conditions of the hydrological model. This will be clarified in the revised manuscript.