Response to reviewer 1

Reviewers comment's in blue, our responses in black

In this paper the authors evaluate the added benefit of using a seamless integration (SEAM) of the outputs from ECMWF extended-range ensemble prediction system ENS-ER) and the ECMWF system 4 seasonal forecast system (SYS4) for hydrological applications. The added benefit from this approach is evaluated by comparing the continuous rank probability scores for the outputs from the hydrological model LISFLOOD forced by SYS4, SEAM, and a climatological ensemble (CLIM) over the hindcast period.

The authors find that hydrological hindcasts made using SEAM show better skill, over those made using SYS4, for much of Europe with lead times up to seven weeks. In some areas like the parts of Alps and northern Finland the reverse was true; however these results are uncertain due to the general poor performance of LISFLOOD in these regions. They argue that the increased skill can be attributed to the better initial conditions of the hydrological and meteorological conditions (models are initialised biweekly as opposed to once per month for SYS4) as well as the use of a better atmospheric model in SEAM (the atmospheric model used in SYS4 is locked at the initial version released with system 4 while the one used in ENS-ER is updated regularly). They conclude that the use of SEAM for hydrological forecasting at the seasonal scale has an added value for decision makers given the higher frequency of updates and improved skill, especially at the sub seasonal scale, making the forecasts more actionable.

The topic of this paper is of great interest at the moment considering the increased focus on forecasting at the sub seasonal to seasonal scales in recent years. Although the concept is not new this paper is the first, that I am aware of, that makes an attempt to evaluate a system that utilises current 'off-the-shelf' operational products. The pa per is well written with a good structure and generally clearly formulated, the methods are scientifically sound, and the results are interesting. Additionally, the research presented in this paper is very relevant to the topic of this special issue. In my opinion, the manuscript has a lot of potential for publication in this HESS special issue. However the authors need to clarify some points and revise some statements so that the paper is more easily understood.

General comments:

1) I feel that it is not clear for what periods the study was performed, something which has a bearing on the quality of the results. The authors state that (P5, L132-L135)

"This study focuses on the performance of SYS4 and SEAM over the hindcast period of the operational forecast with a sequence of starting dates over the period 2015-05-14 (the first available date with 11-member hindcast for ENS-ER) to 2016-06-02 producing daily output time series of discharge over the 20-year hindcast period."

The first part of the sentence suggests that the evaluation period is between the dates 2015-05-14 and 2016-06-02 yet the second part says that the hindcast period has a length of 20 years. The next line has a similar mixed message. From the paper I get the general impression that the evaluation is done for the 20 year period so I assume that the issue is to do with how section 2.3 is worded. This should be addressed as there is some confusion in the way that the paragraph (p5, L231-L238) explains it. Further it has implications on the robustness of the results, should the evaluation period be just the 13 months between the aforementioned dates this would give a limited data sample from which to draw the wider reaching conclusions made by the authors. How can the authors know

whether the performance of the different approaches during that period was typical of their general performance?

The hindcast has 20 years of rerun forecasts, so it is not just one year of integration. The section has now been clarified and we have changed the wording to:

"This study focuses on the performance of SYS4 and SEAM over the hindcast period of the operational forecast. The hindcast of the ensemble forecast is produced twice per week (Mondays and Thursdays) by running an ensemble of 11 members with for that particular day and month, for each of the previous 20 years. The hindcast is run up to 46 days, similar to the ENS-ER. For this experiment, the hindcasts with a sequence of starting dates from 2015-05-14 (the first available date with 11-member hindcast for ENS-ER) to 2016-06-02 were used. This provided 13 monthly starting dates for SYS4 and 111 biweekly starting dates for SEAM with corresponding hindcast set covering all seasons over the previous 20-year period, each with 11 ensemble members. The output was averaged to weekly means before the skill score analysis."

Further, we will add a figure to explain how the hindcasts of the extended-range forecasts are set up.

2) The results show that SEAM has skill over SYS4 in the first 3-8 weeks (Figure2b), mostly concentrated in the first 6 weeks. This would imply that there may be a benefit of merging the two meteorological forecasts before day 46. Did the authors consider this and if not why?

We are not sure if we understand this comment. Fig2b shows that SEAM has more skill than SYS4 for the first 2-3 weeks, but that after that there are some areas where the SYS4 performs better and vice versa. The differences can have many explanations, where geography and altitude plays a part (Fig 3). For those areas where the SYS4 performs better than SEAM, it could as Kean suggests be beneficial to switch to SYS4 earlier than after 46 days. However, that would be interesting from an operational point-of view and is out of scope for this study. A system where you would switch between two systems in an optimal way would have to be carefully calibrated and the effect of switching forecasts would have to be significantly better to justify it. We would rather advocate that SYS4 is used in areas where it is clearly better than SEAM, or as a complement to SEAM. However, we will in future studies dive deeper into the differences in skill between the different forecasts.

Specific comments:

P2, L26: "TSYS4 is also ..." I assume that this is a typo and should read, "SYS4 is also..."

Yes, it was corrected to SYS4.

P7, L224-L225: Although this line is factually correct it appears to contradict the preceding ones. The reader is being told how the low flows during this period caused substantial economic losses due to it affecting inland navigation in the Danube and Rhine basins only then to be told that navigations are regulated during high flows and not low flows. I suggest rewording this or removing this sentence to remove the perceived contradiction or removing this line altogether as it does not add anything significant to the discussion.

The sentence was there to point to the fact that there are no regulated restrictions on the low flow; it is down to the transport companies to make the decision. We agree that it does not add any significant information and the sentence will be deleted in the revised version..

P8, L249-L250: The second part of this line is awkward to read and should be rephrased.

The sentence was rephrased to: “The onset of the second low period was correctly modeled by the SEAM system, whereas the timing of the low flow was missed by SYS4”

P11, L350-L35: I think the reference is - Pappenberger, F., Wetterhall, F., Dutra, E., Di Giuseppe, F., Bogner, K., Alfieri, L., and Cloke, H. L.: Seamless forecasting of extreme events on a global scale, pp. 3–10, Proceedings of H01, IAHS-IAPSO-IASPEI Assembly, Gothenburg, Sweden, July 2013 (IAHS Publ. 359, 2013)

Yes, that is correct; the reference has now been updated.

P16, Caption to figure 3: The last line states, “The dimension of the circles is proportional to the number of days while the color scale refers to progressive weeks.” What do the authors mean by number of days?

The size of the circles are proportional to the number of day of predictability. The circle size was missing in the plot legend that has now been revised. To make the plot more readable we had also colour-coded the circles in broad weekly changes. Clearly, there is a correlation between colour and circle sizes as the darker the colour the larger the symbol dimension. However we found that the colour breaks made the plot more readable. The graphics of the plot has been slightly revised and is as follows: