Response to reviewer 3

This paper evaluates the performance of the hydrological forecast by merging sub-seasonal and seasonal rainfall forecasts. As expected authors found that the hydrological forecasts with merged rainfall forecasts are better than with the seasonal rainfall forecasts for first few weeks. Given that this is revised submission, I had opportunity to read the responses from the authors to address the comments given by the referees from initial submission. The authors have addressed most of those comments and revised the paper accordingly. However there are still some minor issues and errors which needs to be addressed.

The authors have compared hydrological forecasts forcing by merged rainfall forecasts with the seasonal forecasts (SYS4). I wonder whether they have also compared merged rainfall forecast with the SYS4?

Response: No, we have not looked at the rainfall in particular. The idea of this paper was to focus on the discharge, as it is an integrating variable. It is not possible to disentangle the sources of bias without specifically looking at all elements of the water balance, and this will be done as the next step.

Page 3, Line 60-65: These sentences are misleading and confusing. How simple concatenation of the best forecast can be complex in the simplification, then the concatenation is technically difficult?

R: The term simple is meant to be referring to the fact that you do not have to create a new model that produces daily forecast from day 1 up to 6 months for each day of the year. We have changed the sentences to make it clearer:

"The seamless idea could be translated into a concatenation of "the best" forecast at each lead-time. The clear advantage of this off-the-shelf seamless prediction conversion is that it utilizes products that are already available and operational, thereby avoiding the complications of new developments, while at the same time generating forecast products to meet different types of users (Pappenberger et al., 2013). There is however an underlying complexity in this simplification; the difference in design between the various forecasting systems makes the concatenation not entirely straight-forward. The forecasting systems are related since they are from different generations of the same model development, however they have non-matching temporal and spatial resolutions, different hindcast span and different ensemble sizes."

P 3 Line 69: Can authors explain what the bias characteristics are?

R: The bias characteristics means simply the model error. We have changed the text to this:

"One important consequence of this is that the more frequent updates to the extended range compared to the seasonal forecasting system at ECMWF causes the model errors from the two systems to diverge over time, and only closing this gap when the seasonal system is updated to a newer model version.."

Page 3, Line 91: delete "such as"

R: This would change the meaning of the sentence, instead the whole section "such as soil and ground water interactions" was deleted.

Figure 1 is not needed. It does not show anything more than the text in caption.

R: The figure was added because it was requested to explain the hindcasts better. We think that this figure helps explains how the reforecasts are created and that it is therefore justified.

Page 5, Line 130: I suggest authors to explain in one sentence or two “mass-conserving interpolation”.

R: This was added: “The mass-conservative interpolation summarizes the partial contribution of the meteorological input fields onto the LISFLOOD grid.”

Page 5, Line 137-138: “full hydro-meteorological integrations”, but never see this term for the rest of the paper.

R: The terms that are used throughout the paper are SEAM and SYS4, which refer to the hydrometeorological integrations rather than just the meteorological forcing. The sentence was changed to:

For simplicity SYS4 and SEAM will from now on refer to the full hydro-meteorological model chain and not only the meteorological forcing for the remainder of this paper

Page 5, Line 140-145: What is the actual period of forecast evaluation? Is it 2015-05-14 to 2016-06-02 or 1995 to 2015. I remembered this was raised by referees in earlier submission, but still it is not clear in the paper.

R: The period of evaluation are the hindcasts starting at 2015-05-14, (which are 1995-05-14 - 2014-05-14) to 2016-06-02, which have the corresponding hindcast 1996-06-02 - 2015-06-02.

This is very difficult to explain in words, there it was accompanied with Fig 1. I also added a sentence *“As described above, the hindcasts are the reforecasts over the previous 20 years and is produced for each individual run of the ENS-ER”*

Page 5, line 147: Authors mentioned ENS-ER issued Monday and Thursday not Wednesday?

R: Corrected to Thursday.

Page 5, Line 161: “modeled discharge”, did not authors define this as WB runs. I suggest to use consistent terminology throughout the paper.

R: Thanks for that observations, we changed this to WB

Page 6, Line 168: replace “tofN” to “n, N”

R: Replaced with : “step *t* of *N* number…”

Page 6, Line 169: Define “RPS”

R: Corrected to “ranked probability score (RPS)”

Page 6, Line 185: By definition forecast error is difference between observation and forecast, so need to say “against observations”. It is better to give formula for mean relative error as some people define error as forecast – observation.

R: Formula for MRE was added

Figure 3: Legends for black solid lines should be median seasonal (not 10/90 percentiles) in c) and d). I can assume that 10 and 90 percentiles are computed from CRPSS of all river gauges, but never mentioned in the paper.

R: The 10 and 90th percentiles are computed from the 679 outletpoints that were used for initial calibration. This was added to the figure text. Figure 3 was also corrected regarding the median and 10 and 09th percentiles.

Page 7, Line 193: In figure 3a, the CRPSS at week six is less than 0.1 (not 0.2).

R: Yes, this was corrected

Page 7, Line 193: All river point? Do authors mean all river gauges?

R: Was changed to "All points used in the validation"

Figure 3a: I think it is not fair comparison between SYS4 and SEAM when all start dates in SEAM are considered. Firstly, the sample sizes of both forecasts are not same, then more importantly the target dates of given lead time are not same.

R: Yes, we agree with that, and we have discussed that in the text and also shown more fair comparisons.

Page 7, Line 215-219: Not sure what the authors want to say "An explanation can …. driving forecast is used."

R: This was further explained: *"An explanation can be that the ensembles from the two meteorological forecasts are not matched member by member in terms of their relative deviation from the mean, for example matching members from each distribution according to their wetness"*

Figure 4 is hard to understand. "The dimension of the circles is proportional to the number of days while the color scale refers to progressive weeks". Number of days of what?

R: The colours and the circle size show the same thing, the colours are added to make the figure easier to understand. The figure text was changed to :

"The number of weeks (days) before the CRPSS goes below zero using only the first forecast of the month for a) SEAM against CLIM; b) SYS4 against CLIM c) SEAM against SYS4; and d) difference between SEAM against CLIM and SYS4 against CLIM. The dimension of the circles is proportional to the number of days while the color scale refers to the number of weeks. The size and colour of the circles are therefore showing the same information and are both added for clarity."

Page 8, Line 228: effect?

R: Changed to "effect"

Page 8, Line 236-240: Please rephrase this paragraph. What do authors mean by recent "development of the precipitation forecast?"

R: Precipitation forecast was changed to "precipitation model scheme"

Figure 5 caption: replace "functionality" with "function"

R: Changed

Figure 5: Legend for SYS4 is wrong (see previous comments)

R: The legend was corrected

Figure 7: In the previous figures, authors have shown 10 and 90 percentile, why they chose 25 and 75 percentiles. How are the percentiles computed for reliability diagrams?

R: The percentiles are computed as the probabilities of exceeding (subceeding) the 75 (25) percentiles. The figure will be changed to be more consistent with the rest of the paper.

Page 8, Line 252: “Both forecast systems are over-confident”. How? For forecast probability less than 0.5, it is lower than the observed frequency, while for higher than 0.5, it is higher than the observed frequency.

R: For low probabilities, the forecast is not predicting the event often enough, and for the higher end of the probabilities the event is predicted too often, which means that the forecast is over-confident.

However, Figure 7 will be reworked to better reflect the reliabilities.