Hydrol. Earth Syst. Sci. Discuss., author comment AC2
https://doi.org/10.5194/hess-2022-89-AC2, 2022

Reply on RC2

Sandra M. Hauswirth et al.

Author comment on "The suitability of a seasonal ensemble hybrid framework including data-driven approaches for hydrological forecasting" by Sandra M. Hauswirth et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2022-89-AC2, 2022

Anonymous Referee #2

Referee comment on "The suitability of a hybrid framework including data driven approaches for hydrological forecasting" by Sandra M. Hauswirth et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2022-89-RC2, 2022

The study is building heavily on results published elsewhere (Hauswirth et al., 2021). I guess that is fine and always a difficult call to decide just how much information to provide so that a paper becomes a stand-alone piece of work without unnecessary repetition. However, in places I would have liked a little more info in this paper so I would not necessarily have to read the previous paper. For example, lines 114-117; this seems important and a bit more in-depth description of the datasets (SEAS5) and the 'lagged time series approach' which I am not familiar with.

Perhaps I have overlooked something, but I am not sure how the all the forecasts made by the different models and different ensembles are aggregated into one CDF (as per Fig.2)? Also, If there is only a minor difference in the performance between the five different ML models (line 205) then what is the advantage of using all of them rather than selecting the 'best' or most credible model for a particular site and use that? Might it be useful to add a section highlighting the performance of each of the five ML models in contrast to the aggregate performance of the entire system?

Response: We want to thank the reviewer for taking the time to review our manuscript and sharing thoughts and suggestions for improvement. We will address the comments, suggestions and open questions point by point below.

We agree that our study is heavily building up on a previous study by the main author, which was a conscious decision that was made. The idea was and is to use the base modelling framework for different experiments. Therefore, we wanted to fully focus on the model development aspect in the first publication and refer to it for future work. We acknowledge that in this case, the explanation of the modelling framework was too short (as also indicated by the other reviewer). We plan to extend the description of the base modelling framework, including the lagged time series approach. Furthermore, we will extend the description of the new introduced datasets such as SEAS5 and EFAS as recommended.

Regarding the comment Fig 2 and how the different ensembles were aggregated into one CDF, we would like to refer to line 196 ('This was done by computing all the individual daily CRPSS results of all the hindcasts of every station and method, before aggregating the individual CRPSS scores to different temporal scales (weekly in Fig. 2).) and line 155 for the specific calculation procedure ('Equation 1 …was used and the CRPS computed over all ensemble members for each lead day of every hindcast before aggregating it to other temporal scales'). For creating Fig 2 the different CRPSS scores for the different station and method combinations were additionally aggregated by lead day and the average was used for plotting.

The reviewer poses a good question regarding what is the advantage of using different ML models, while they only show minor differences in their results. While performing the experiments, we were not expecting to only observe minor differences, while predictions using observed input data did show differences (Hauswirth et al., 2021). The most likely explanation is that the forecasting skill is very much dependent on the skill by which the input variables are forecasted, which apparently make the differences in skill between the ML models insignificant in comparison. We will add an explanation for this in the manuscript.

To keep the manuscript easy to follow we decided to focus on one method in more detail. However, we will add a section on this as this is an interesting aspect to compare to the previous study (Hauswirth et al., 2021), where there were differences found between the models.

Minor comments:

The acronym LSTM is not defined?
Response: We will define the acronym LSTM (Long Short Term Memory Model) to prevent confusion.

Figs.2-6 did not come off well in my black and white copy, but ok for online viewing (which in fairness is probably the most common by now).
Response: We will revisit the figures 2 - 6 and see if we can improve the readability for printed version.

Line 253: Why the 20th percentile? Or why only the 20th percentile? I could imagine that the ability to assess low-flow across a range of severities would be of interest?
Response: We chose the 20th percentile as it is a commonly chosen threshold for low flows and was also interesting to our collaboration partners.

Line 253: Do any of your models include the effect of human interventions and their potential impacts on low flow? For example, water restrictions, operation of control structures to manage low flows etc? If not, is this likely to important in a highly regulated system such as the Netherland's water ways? I think there is so mentioning of this in lines 220-225, but seems to me this is particularly important during low flow?
Response: The Netherlands is indeed a country with high water management

infrastructures and plans. We therefore are working with the National Water Authority to include expert knowledge where possible and available. In our previous study we did include a separate water management scenario where we included operational plans of the major water infrastructures. We did see that there was an improvement in our simulations, albeit very minor in terms of our modelling results (Hauswirth et al., 2021).

Line 293: I am not sure I understand how you incorporated water management into your models? Was this done in this study, or is that something that was part of the Hauswirth et al. (2021) study? I think perhaps more detail on this could be included in this manuscript as this seems interesting and important (even if you did not find a strong effect).

Response: In line with the previous comment about water management above: We did a run with water management influence using the same approach as in Hauswirth et al. (2021). This run includes operational rules of main infrastructures which are related to the Rhine discharge at Lobith (one of our main input variables) for two specific input locations and two additional observation records of locations based at smaller infrastructures. We were therefore able to use the same approach regarding the operational rules for the main infrastructures, as these are based on the Rhine discharge we obtain from the EFAS dataset. For the two other additional timeseries climatology was used as operational plans were not available.

We acknowledge that the information given in this manuscript related to water management aspects are very limited and heavily based on the previous study. We will extend the explanation so that the reader will be able to follow better.