Authors' response to interactive comment by Anonymous Reviewer #2

Black text: Reviewer comment

Blue text: Authors' response

This manuscript presents a large scale study (39 493 catchments) that aims at gaining a better understanding of the main factors that drive the skill of ensemble streamflow forecasts in Sweden. Most similar studies in seasonal forecasting aim at distinguishing the contribution of initial conditions and that of meteorological forcings. In this manuscript, the authors rather want to distinguish the hydrological processes that drive the skill of seasonal forecasts across space and time. They also study the influence of aggregating the forecasts at different timescale (2 weeks, 1 months, 2 months, etc.) affects the skill, which I find very interesting. The authors show that forecasts are mostly skillful when initialized during the winter months, and for base flow dominated catchments. They also propose a classification of catchments into clusters with similar characteristics and behavior relative to seasonal forecasts. I think this is an interesting study that can bring new information to better understand where we should concentrate our efforts to improve the skill of seasonal forecasts in hydrology. I only have very minor comments that relate to methodological choices that I would like the authors to explain in greater detail.

We thank the reviewer for his/her valuable comments and suggestions that will undoubtedly help us improve our manuscript. Below we reply to each of these and explain how we will incorporate them into the manuscript.

Detailed comments:

• Line 34: I am always bothered when people change the original name of a technique. The authors here define ESP as "Ensemble Streamflow Predictions", but this is not exactly what ESP originally stands for. In Day (1985), who originally proposed the technique, ESP refers to "Extended Streamflow Prediction". This may seem like a small detail, but 1) I think it is only fair to use the exact name that Day proposed for his own technique and 2) "Ensemble" prediction is very general and could very well be obtained using a dynamical meteorological model rather than past climatological scenarios. Therefore, designating ESP as "Ensemble" streamflow prediction can be confusing to some readers (I'm thinking especially about people who are unfamiliar with ensemble forecasting in general). ESP should refer to a very specific technique, but I have also heard people using it to refer to ensemble forecasts obtained using dynamical meteorological forecasts. Also, I think that Day (1985) should be cited, as it is the original reference for ESP.

We agree with the reviewer in that terminology should be used in a restrictive sense and that original ideas and their naming should be respected and used. That being said, the term Ensemble Streamflow Prediction referring to Day's 1985 technique has been widely adopted by the community and has nowadays almost replaced the original term (Extended Streamflow Prediction), which is why we use it even here. Nonetheless, the reviewer makes a good point here, and following his/her reflection we will also include a reference to the original name of the ESP technique and publication in the revised manuscript.

• Lines 34-50 and lines 291-299: Speaking of dynamical forecasts: ESP is quite an old technique. And I agree that it is still what is used operationally for long-term hydrological forecasting by many operational agencies, and that it works well. However, long-term dynamical meteorological forecasts also exist and some studies focus on assessing their skill for hydrology, often using ESP as a reference for comparison (e.g. MeiBner et al. 2017; Baker et al. 2019, Slater et al. 2019, Bazile et al. 2017 and others). I don't have any problem with the authors using ESP instead of dynamical forecasts, but I think the use of dynamical meteorological for seasonal hydrological forecasting should also be included in the literature review. There is a good discussion about NWP later in the paper (291-299), but I think it appears much too late. I strongly suggest including examples of NWP-based hydrological seasonal forecasting systems in the introduction, and possibly moving some elements from the discussion (a portion of lines 291-299) also in the introduction. I think it is important to explain why you chose to use ESP rather than NWP based forecasts, and to do so before the discussion!

In the revised manuscript, we will include a short description of NWP-based techniques in the introduction and further clarify the reasons behind the choice of ESP in this study. These reasons include the fact that the objective of this manuscript was to assess the existing system at SMHI's operational service, which uses ESP forecasts, and that the ESP method offers the best study object to focus on the role of initial hydrologic conditions alone (best explained through catchment characteristics than the role NWP forcings).

• Page 4 lines 101-110: I'm not sure I understand why it is relevant to include regulated rivers in the study. They all end up in the same cluster (7), which unsurprisingly has a negative median skill. It would certainly be interesting to forecasts long-term inflows to reservoirs, as it could be useful for long term water management/hydropower production planning, but if I understand those lines correctly, this doesn't seem to be the case here (I understand that there are forecast points downstream from reservoirs, correct?). I think the rationale for including regulated catchment in the study needs to be better explained.

In our view, a clear explanation to this is provided in the discussion section of the manuscript, as the reviewer states in a later comment. The rationale behind this is an operational one: since the operational service we are trying to evaluate here includes regulated rivers (which are, additionally, of special interest for such a system), they should be taken into the account in the analysis as well. It should be noted though, that the degree of regulation is not explicitly considered as one of the indicators for the clustering analysis. Nevertheless, since the regulation scheme affects the hydrological response, it is plausible that regulated catchments become clustered together.

Regarding the evaluation of inflows to reservoirs, we agree with the reviewer in that this would be very relevant for long term water management and hydropower production planning. However, in this manuscript we focused on the operational forecasting setup from the perspective of public service, which provides information based on catchment outflows. Nevertheless, even if this analysis is out of the scope of the present manuscript, it is something we plan to investigate further in the future for the exact same reasons the reviewer stated here.

Overall, we understand the reviewer's comment and we will therefore include the reasoning earlier in the text so as to make the purpose more understandable to readers.

• Page 14 lines 240-253 and Figure 6: I would find it helpful if the abbreviations from Table 1 were used in this paragraph, which analyses Figure 6 (even though a sort of synthesis is presented in Table 2). I find it difficult to remember acronyms and abbreviations, so I had to go back and forth between the figure, the text and Table 1.

We agree with the reviewer. Initially, we tried to avoid repetition and including yet more information in this already dense paragraph. However, we will follow the reviewer's advice and add relevant abbreviations there.

• Table 2: How are potential and actual evapotranspiration obtained? Is it really important to include both in the table?

Both potential and actual evapotranspiration are S-HYPE model outputs. In our case, potential evapotranspiration is calculated based on mean temperature and a land use dependent rate parameter. An additional parameter adjusts the potential evaporation rate depending on the season. Regarding actual evapotranspiration, it is calculated by a linear function depending on soil moisture and it ranges between 0 and the potential evaporation value (when water content exceeds field capacity).

We included both in the table since, originally, we had the intention to include a short analysis based on the Budyko framework. However, since most catchments in Sweden are energy limited, it did not have much explanatory power.

We agree with the reviewer that it is not necessary here to present both parameters and we will therefore remove the potential evapotranspiration column in the revised manuscript.

• Page 16 line 268: Do you have any possible explanation why the cluster (5) with the highest general skill also have the largest spread? Is it possible that those two things (skill and spread) are related? What I mean is that if the skill is assessed by the CRPS and the CRPS is very sensitive to spread, then maybe the high skill is (at least in part) a consequence of this high spread? In any case, I think it would be interesting if the authors could provide a possible explanation.

Results from cluster 5 are indeed interesting. The forecasts in the cluster 5 catchments generally show the highest skill (for all lead times) among all cluster groups, yet results are widely spread. In this paper we conclude that the forecast skill is strongly linked to the various hydrological regimes (see also a more detailed investigation in Pechlivanidis et al. 2020), and hence we argue that the answer is within a deeper understanding of the hydrological signatures in cluster 5. As we state in P14 L241-242, the catchments in cluster 5 are characterized by a high baseflow contribution (BFI), a slow response to precipitation (Flash) and a generally small intra-annual variability (DPar). In Figure 6a we observe that although the mean values for RLD (rising limb density) are below the 33rd percentile of this signature (which represent 'below normal' signature values); however the boxplot for RLD driven by all 4355 catchments in cluster 5 indicates high variability, with some catchments experience ‘normal’ RLD values and yet some others even higher than the 66th percentile of this signature. Consequently this indicates that some catchments in cluster 5 despite their high baseflow contribution experience sharp increases in their hydrographs, which is an indication of low skill as seen in Figure 5 (CRPSS and RLD are strongly, but negatively, correlated). We will explain the above argument for the large spread in cluster 5 in the revised manuscript.

• Page 18 lines 316-326: You mention the idea of using more sophisticated data assimilation techniques, such as the EnKF, but I think it would also be worth mentioning the possibility of assimilating other observations than streamflow, for instance soil moisture and/or snow water equivalent. This has been done in some studies (e.g. Huan et al. 2017), but there are still not that many in direct relation to seasonal forecasting.

In the revised manuscript we will acknowledge the possibility of assimilating other observations and refer to relevant studies such as Huan et al. 2017 or Musuuza et al. 2020, which is already cited in the manuscript.

• Page 19 lines 335-337: "This exercise shows that the regulation routines in . . ." There I finally found the justification for including regulated rivers in the study. I think this should be expressed earlier in the manuscript, around lines 105-120. At the moment, the explanations provided in lines 105-120 remain too general and it is hard to understand what it is that you want to test by including regulated rivers. At lines 335-337 it becomes clear, but it is too late.

Please see the previous comment on the same issue for a description on the reasoning behind this as well as the planned modifications to the revised manuscript.

References:

Day, G. (1985). Extended Streamflow Forecasting Using NWSRFS, J. Wat. Res. Plan. Mgmt., 10.1061/(ASCE)0733-9496(1985)111:2(157), 157-170

MeiBner et al. 2017 (already cited in the manuscript)

Baker S.A., Wood A.W. and Rajagopalan B. (2019). Developing Subseasonal to Seasonal Climate Forecast Products for Hydrology and Water Management, Journal of the American Water Resources Association, 55(4), 1024-1037.

Slater L.J., Villarini G., Bradley A.A. and Vecchi G.A. (2019) A dynamical statistical framework for seasonal forecasting in an agricultural watershed, Climate Dynamics, 53(12), 7429-7445.

Bazile R., Boucher M-A, Perreault L. And Leconte R. (2017) Verification of ECMWF System 4 for seasonal hydrological forecasting in a northern climate, Hydrology and Earth System Sciences, 21, 5747–5762.

Huang C., Newman A.J., Clark M.P., Wood A.W. and Zheng X. (2017) Evaluation of snow data assimilation using the ensemble Kalman filter for seasonal streamflow prediction in the western United States, Hydrology and Earth System Sciences, 21, 635–650.

Musuuza, J. L., Gustafsson, D., Pimentel, R., Crochemore, L. and Pechlivanidis, I. (2020): Impact of Satellite and In Situ Data Assimilation on Hydrological Predictions, Remote Sens., 12(5), 811, doi:10.3390/rs12050811.

Pechlivanidis, I. G., Crochemore, L., Rosberg, J. and Bosshard, T. (2020): What Are the Key Drivers Controlling the Quality of Seasonal Streamflow Forecasts?, Water Resour. Res., 56(6), doi:10.1029/2019WR026987.