Response to Anonymous Referee #2

Referee comments are labelled consecutively (e.g., R#2-1 is comment 1) and given in blue text.

This paper has presented an investigation of ensemble streamflow prediction (ESP)for 46 catchments in Ireland. The GR4J model is employed to formulate the rainfall-runoff relationship and perform streamflow forecasting. The forecast skill is evaluated and then related to a range of catchment attributes, e.g., base flow index, flashiness index, and runoff ratio. The results show that skillful forecasts are generated using ESP and that the skill can be attributed to catchment attributes and North Atlantic Oscillation (NAO). Overall, the paper is well-written with the methods and results clearly presented. There are a few comments for further improvements of the paper.

We thank the referee for their constructive review of our manuscript. Please find our point-by-point response below.

R#2-1. First, reliability is an important feature of ensemble forecasts. Specifically, reliability indicates the agreement between forecast probability and mean observed frequency. For streamflow forecasting, attention is usually paid to high- and low-flow events. Therefore, it would be meaningful to show whether ensemble forecasts generated by ESP yield reliable probabilistic forecasts of high- and low-flow events at different lead times.

We agree that reliability is an important consideration when evaluating ensemble hydrological forecasts. To address this, we will include an additional verification metric, the probability integral transform (PIT) diagram (Gneiting et al., 2007) as a means of assessing the reliability of high- and low-flow forecasts. The PIT diagram is the cumulative distribution of the PIT values, which measure the position of the observations relative to the forecast distribution. For a perfectly reliable forecast, the observations will fall uniformly within the forecast distribution and the PIT diagram will correspond to the 1-to-1 diagonal. Forecasts that systematically under- (over-) predict will have a PIT diagram below (above) the diagonal, whereas under- (over-) dispersive forecasts will have a transposed S-shaped (S-shaped) PIT diagram (Arnal et al., 2018). For comparison on large datasets, the area between the PIT diagram and the 1-to-1 diagonal can be calculated to provide a numerical measure of reliability. This can further be converted to a skill score for ease of interpretation (Arnal et al., 2018; Crochemore et al., 2017). We believe incorporating the PIT diagram into our manuscript will complement our use of the ROC score and help improve our overall analysis.

R#2-2. Second, there recently is an interesting paper on the influence of NAO on flooding and drought over Europe (Changes in North Atlantic atmospheric circulation in a warmer climate favor winter

flooding and summer drought over Europe, E Rousi, F Selten, S Rahmstorf, D Coumou, Journal of Climate, 2020). This paper can offer some climatological insights when relating forecast skill to NAO.

Thank you for bringing this paper to our attention. We will review its contents and revise our manuscript accordingly to include any additional insight it may provide.

References

Arnal, L., Cloke, H. L., Stephens, E., Wetterhall, F., Prudhomme, C., Neumann, J., Krzeminski, B., and Pappenberger, F.: Skilful seasonal forecasts of streamflow over Europe?, Hydrol. Earth Syst. Sci., 22, 2057–2072, https://doi.org/10.5194/hess-22-2057-2018, 2018.

Crochemore, L., Ramos, M.-H., Pappenberger, F., and Perrin, C.: Seasonal streamflow forecasting by conditioning climatology with precipitation indices, Hydrol. Earth Syst. Sci., 21, 1573–1591, https://doi.org/10.5194/hess-21-1573-2017, 2017.

Gneiting, T., Balabdaoui, F., and Raftery, A. E.: Probabilistic forecasts, calibration and sharpness, J. R. Statist. Soc. B, 69, 243–268, https://doi.org/10.1111/j.1467-9868.2007.00587.x, 2007.