

Referee #1

Streamflow forecasts play an important part in water resources management. In this paper, a method of seasonal hydrological forecasting is developed for the Great Britain. There are five comments for further improvements of the paper.

We thank the referee for their interest in this work, agreeing to review our paper, and for their helpful suggestions. Our responses are below in blue.

First of all, the period under investigation is from 1993 to 2016. As it is notably 10 years from now, the authors are suggested to update the forecasting model and take recent years into consideration.

- The 1993-2016 hindcast period used in this work is a standard period set by seasonal forecasting centres for use in e.g. Copernicus Climate Change Service. It is not expected to be updated for another couple of years. As our work relies upon post-processing the MSLP hindcasts generated automatically by the Met Office's operational system (MacLachlan et al., 2015; Kettleborough et al., 2026) we must use this period; it would be too computationally expensive to generate new hindcasts for this project. We add to the revised text to explain the ultimate source of these hindcasts (Lines 200-208).

Secondly, global climate models (GCMs) present the state-of-the-art weather/climate forecasting. The historical weather analogues (HWA) used in the analysis is relatively simple. Is it possible to incorporate GCM forecasts into the analysis?

- A GCM is already part of the forecasting chain, GloSea is based on HadGEM3 (a family of GCM models); we will update the text to indicate this explicitly (Lines 62-70). While we could use GloSea precipitation forecasts directly, as was done by Bell et al. (2017), these lack the spatial resolution required (now further emphasised in lines 66-70) and, like all other GCM-based seasonal forecasts, are known to struggle over the UK (see references in Sect. 4.2). This is the main driver behind post-processing the GCM-based forecasts with the HWA method.

Thirdly, as is shown in Figure 2, the Pearson's correlation between HWA forecasts and observations is below 0.20 across the Great Britain. In comparison with GCM forecasts, this correlation can be too low to be informative. Therefore, a comprehensive comparison of HWA forecasts with GCM forecasts is in demand.

- Spatial resolution is the key difference here – our ~0.2 correlation is for precipitation at *1 km resolution*. While a GCM may achieve higher correlations, these are at coarser resolution and would still require downscaling to 1 km to drive the hydrological forecasts (see revisions to lines 66-70).
- A comparison between GCM and HWA forecasts was done by Stringer et al. (2020), who demonstrated that at *the 50 km resolution of the GCM* no reduction in the correlation was found when adopting HWA methods (their Sect. 5 & Fig. 9). The reduction in correlation found in this work is thus caused by the higher resolution used, rather than by the HWA method itself (see our lines 328-333).
- A more systematic comparison with GCM-based forecasts, including statistical analysis of other models and methods is beyond the intended scope of this work. We

agree that such a work would be desirable in the future and note the interest in multi-model ensembles for UK seasonal meteorological & hydrological forecasting as future work (revision to lines 550-555).

Fourthly, in terms of GCM forecasts, the authors are suggested to refer to Slater et al. (2019) that presented a comprehensive of the NMME forecasts across the continental USA.

- We thank the referee for suggesting this work. Multi-model ensembles are also of interest to the future development of the Hydrological Outlook, a preliminary study has been presented at this year's EGU (Bulut et al., 2026). We have included the suggested works in our list of references for seasonal hydrological forecasting in the USA (revisions to lines 45-50).

Fifthly, how does the HWA forecasts perform under historical droughts?

- Most droughts in the UK occur in the summer. The HWA forecasts presented in this work show negative average correlations in this season (see Figure 2, row 1, panel 3), this is attributable to the poor performance of GloSea at predicting the MSLP fields in this season (Lockwood et al. 2023). Significant droughts in the hindcast period occurred in the UK in summer 2003, 2004-6, 2012-13 (UKCEH drought inventory). The HWA rainfall forecasts from those summers do not outperform the average skill over the full hindcast period (assessed using the CRPS scores). Improving the summer skill is important, as we identify in the text the key to skilful summer forecasting is ensuring the underlying GloSea forecasts capture any predictable summer signal (lines 337-342; see also Dunstone et al., 2023).
- For dry winters, some analysis of the performance of HWA forecasts has been done by Smith et al. (2019), Stringer et al. (2020; Fig. 10 & 11) and Chen et al. (2026; Sect. 3.2, Fig. 9). For instance, Smith et al. (2019) considered the 2010-11 dry winter and showed that the HWA captures an ensemble shift towards higher likelihood of low rainfall that is not present in an ESP forecast.

References:

Kirtman, Ben P., et al. "The North American multimodel ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction." *Bulletin of the American Meteorological Society* 95.4 (2014): 585-601.

Slater, Louise J., Gabriele Villarini, and Allen A. Bradley. "Evaluation of the skill of North-American Multi-Model Ensemble (NMME) Global Climate Models in predicting average and extreme precipitation and temperature over the continental USA." *Climate dynamics* 53.12 (2019): 7381-7396.

References

Bulut et al. (2026): Advancing the UK Hydrological Outlook using skill-based forecast blending, EGU General Assembly, [doi:10.5194/egusphere-egu26-5012](https://doi.org/10.5194/egusphere-egu26-5012).

Chan et al. (2026): UK Hydrological Outlook using Historic Weather Analogues, *Hydrol. Earth Sys. Sci.*, 30, 905-927, [doi:10.5194/hess-30-905-2026](https://doi.org/10.5194/hess-30-905-2026).

Dunstone et al. (2023): Skilful predictions of the Summer North Atlantic Oscillation, *Comm. Earth and Env.*, 4, 409, [doi:10.1038/s43247-023-01063-2](https://doi.org/10.1038/s43247-023-01063-2).

Kettleborough et al. (2026): Global Seasonal Forecasting System 6 (GloSea6): A Large Ensemble Seasonal Forecasting System, *Mon. Weather Rev.*, 154, 39-57, [doi:10.1175/MWR-D-25-0007.1](https://doi.org/10.1175/MWR-D-25-0007.1).

Lockwood et al. (2023): Seasonal prediction of UK mean and extreme winds, *Q. J. Roy. Meteor. Soc.*, 149(757), 3477-3489, [doi:10.1002/qj.4568](https://doi.org/10.1002/qj.4568).

MacLachlan et al. (2015): Global Seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system, *Q. J. Roy. Meteor. Soc.*, 141(689), 1072–1084, [doi:10.1002/qj.2396](https://doi.org/10.1002/qj.2396).

Smith et al. (2019): Improving Winter Ensemble Streamflow Prediction in the UK using North Atlantic Oscillation Analogues, *Geophysical Research Abstracts Vol. 21*, EGU2019-5770.

Stringer et al. (2020): Improving Meteorological Seasonal Forecasts for Hydrological Modeling in European Winter', *J. Appl. Meteorol. Clim.*, 59(2), 317–332, [doi:10.1175/JAMC-D-19-0094.1](https://doi.org/10.1175/JAMC-D-19-0094.1).

Referee #2

Dear authors,

I was not included in the original review. I acknowledge the relevance of your study, but I think you end where it starts to get interesting, namely where you could explore optimal ways to combine the original approach and this extension. I put my comments directly in the version with tracked changes.

We thank the referee for their interest in the study and for agreeing to review our work. We agree that our results point to the potential for skill-based blending of forecasts from different models. There is an ongoing project to investigate applying this method to the hydrological forecasts (we have revised the description of this work on lines 551-555). Preliminary results from this study have been presented at this year's EGU (Bulut et al., 2026; <https://doi.org/10.5194/egusphere-egu26-5012>). Our paper is a foundation for this work.

The below line numbers in bold at the start of each line relate to the position of the comments in the reviewed text. Line numbers for changes relate to the revised text.

Fig 1: Revised figure including the step-by-step HWA process added to the supplement (Fig. S1).

120: We agree that different assumptions could be made to temporally disaggregate daily total rainfall and PE to the resolution of the model. However, we don't think the difference between uniform and other simple assumptions is significant for predicting monthly (and seasonal) mean flows; driving data with finer temporal resolution will have a stronger influence more important for shorter-term forecasting (e.g. Wallbank et al., 2022). This is now stated in the revised text (lines 125-127). We note that some more sophisticated methods, such as using sunlight hours to disaggregate PE, require additional data and would add timeline risk and potentially cost to the operational system.

140: Our synthetic seasons and months consist of combinations of three analogue years. Three was chosen to strike a balance, we seek to:

- Generate enough synthetic seasons/months to select from. Here we have of order 250000 potential analogues, dramatically reducing the risk of duplication within a 100-500 member ensemble.
- Adequately span the range of possible forecasts. Stringer et al. (2020) investigated this, showing that this set of synthetic seasons captures the range of forecast NAO values (see their Section 4.1 and Fig. 5).
- Minimising the computational cost of selecting analogues to match the GloSea MSLP field – using more than three analogues in each synthetic season/month increases the number of potential combinations.
- Minimise the number of analogues within each synthetic season/month to reduce the number of discontinuities as the forecast time-series transitions from one analogue year to another.

These points have been added to the supplemental material, end of paragraph 3.

155: Bell et al. (2013) developed the monthly-resolution water balance model to exploit the monthly-resolution rainfall forecasts produced by GloSea5. HWA forecasts allow us to produce rainfall forecasts at a higher temporal resolution and potentially to drive a higher-resolution hydrological model. We allude to this in the future work section (lines 548-551). Sub-monthly HWA forecasts are the subject of ongoing work by some of the authors, preliminary results were

reported by Rhodes-Smith et al. (2024; unpublished project report); we intend to prepare this work for publication later this year.

170a: Routing behaviours are implicitly included as runoff is summed over each catchment within the water balance model, but for a monthly or seasonal forecast it is not necessary to explicitly model the temporal evolution of river flows within the UK (where travel times along rivers are short, of order only a few days).

170b: We agree that water management decisions are usually important to accurately represent river flows, and in heavily abstracted catchments our modelled natural flows will not be as accurate – this is now stated (lines 171-174). Rameshwaran et al. (2022) have shown that knowledge of actual abstractions and discharges improves flow modelling in such catchments, but this information is not available for our operational forecasts. This deficit will be somewhat mitigated by the forecast being for flow categories, rather than actual flows.

215: We have added references to additional works that include verification of modelled flows (lines 227-230). The most recent (unpublished) performance assessment using c.300 NRFA stations shows a median NSE of 0.86 for monthly and seasonal mean flows.

220: A systematically biased forecast will not perform as well as an unbiased forecast in the selected performance metrics; while the bias in the ensemble mean could also be presented as an additional metric we decided that would be excessive in this already lengthy paper. It is important to note that the operational forecasts are bias-corrected, this was described in Section 2.1.4, and we have extended that section to make the effect of this procedure explicit (lines 189-192). The final river flow forecasts are categorised into 7 classes based on the modelled historical flow distribution. This applies a simple correction scheme (similar to quantile mapping, although we use the boundaries of each class rather than the quantiles) to the forecasts to account for potential biases in the model. Bias-correction is the subject of ongoing work within the UK Hydrological Outlook multi-model research (now mentioned line 553).

Fig 2: The referee notes that the seasonal forecast for June-July-August performs significantly worse than the monthly forecasts for June, July and August. These are not directly comparable – the seasonal forecasts are based on observations in the month preceding the start of the season, while the monthly forecasts are initialised in the month preceding the forecast. This means the seasonal forecast's July and August are predicted at two and three months ahead, while the July and August monthly forecasts are at one-month ahead. We thus cannot assume that the performance of the seasonal forecasts should look like the mean of the three monthly forecasts. In this instance the June-July-August seasonal HWA forecast suffers from the poor performance of the underlying GloSea forecasts, attributed to inadequately capturing summertime teleconnections (Lockwood et al., 2023). At one-month lead times this effect is reduced thanks to the longer timescales over which these teleconnections operate.

342: Yes, we accumulate all upstream flows (runoff) as described in Sect. 2.1.3 (lines 175-178), so this section describes skill in estimating total river flows, rather than each pixel's runoff. As suggested, we have calculated whether the correlation of the ensemble mean with observed flow varies with catchment size for river flows. The smallest catchments (<15km²) have been excluded from the skill maps and forecast products, these are likely to be the most volatile and difficult to forecast accurately. Above 15km² we find no systematic variation.

425: Some interpretation of the differences in seasonal skill scores is now included as requested (lines 443-445).

Fig 6: Thank you for this suggestion, we have added labels to the top of the figure.

461: We agree that the spatially and seasonally variable skill encourages the use of a multi-model approach. While this work presents the skill of one model, the Hydrological Outlook is investigating blending the forecasts from the multiple hydrological models (see extended description of this ongoing work on lines 550-555). Preliminary results were presented at this year's EGU (Bulut et al., 2026).

477: "Exceptional" changed to "high" as suggested (line 495).

499: We now emphasise the often-marginal skill of the HWA forecasts at this resolution compared to that achieved at coarser resolutions. We now comment in more detail on the degree of improvement in seasonal river flow forecasts by comparison to prior work. Changes to lines 517-531.

Data availability: As requested, we have added extra information on the available data.

- The GloSea6 MSLP fields are available through the Copernicus Climate Change Service's Climate Data Store (doi:[10.24381/cds.68dd14c3](https://doi.org/10.24381/cds.68dd14c3)).
- The analogue years derived from these and used to construct the HWA rainfall forecasts can be provided by the authors on request.
- The hydrological models used are proprietary and thus cannot be made available by the authors.
- We cannot justify submission of the intermediate river flow and rainfall hindcast datasets to a repository; their value to other researchers is limited and the datasets very large.

References:

Bell et al. (2013): Developing a large-scale water-balance approach to seasonal forecasting: application to the 2012 drought in Britain, *Hydrol. Process.*, 27(20), 3003–3012, doi:[10.1002/hyp.9863](https://doi.org/10.1002/hyp.9863).

Bulut et al. (2026): Advancing the UK Hydrological Outlook using skill-based forecast blending, EGU General Assembly, doi:[10.5194/egusphere-egu26-5012](https://doi.org/10.5194/egusphere-egu26-5012).

Lockwood et al. (2023): Seasonal prediction of UK mean and extreme winds, *Q. J. Roy. Meteor. Soc.*, 149(757), 3477–3489, doi:[10.1002/qj.4568](https://doi.org/10.1002/qj.4568).

Rameshwaran et al. (2022): Use of abstraction and discharge data to improve the performance of a national-scale hydrological model, *Water Resour. Res.*, 5, e2021WR029787, doi:[10.1029/2021WR029787](https://doi.org/10.1029/2021WR029787).

Rhodes-Smith et al. (2024): Quantitative river flow forecasts for the Flood Outlook: high flow forecasting over the UK winter, UK Centre for Ecology & Hydrology, 18pp. (Unpublished), <https://nora.nerc.ac.uk/id/eprint/538560>.

Stringer et al. (2020): Improving Meteorological Seasonal Forecasts for Hydrological Modeling in European Winter', *J. Appl. Meteorol. Clim.*, 59(2), 317–332, doi:[10.1175/JAMC-D-19-0094.1](https://doi.org/10.1175/JAMC-D-19-0094.1).

Wallbank et al. (2022): Assessing precipitation from a dual-polarisation X-band radar campaign using the Grid-to-Grid hydrological model, *J. Hydrol.*, 613, 128311, doi:[10.1016/j.jhydrol.2022.128311](https://doi.org/10.1016/j.jhydrol.2022.128311).