https://doi.org/10.5194/egusphere-2025-2506
Preprint. Discussion started: 24 July 2025

Improved seasonal hydrological forecasting for Great Britain

Mark D. Rhodes-Smith[1], Victoria A. Bell[1], Nicky Stringer[2], Helen Baron[1], Helen Davies[1] and Jeff Knight[2]

[1]UK Centre for Ecology and Hydrology, Maclean building, Benson Lane, Crowmarsh Gifford, Wallingford, Oxfordshire, OX10 8BB, UK
[2]Hadley Centre, Met Office, Fitzroy Road, Exeter, Devon, EX1 3PB, UK

Correspondence to: Mark D. Rhodes-Smith (marrho@ceh.ac.uk)

Abstract. Great Britain's variable maritime climate has until relatively recently limited the utility of seasonal hydrological forecasts. The latest generations of seasonal atmospheric forecasting systems have created new opportunities to improve flow forecasting across Great Britain, such as for the UK Hydrological Outlook. Here, newly-developed high-resolution rainfall forecasts derived from historical weather analogues (HWA) conditioned on large-scale circulation patterns are used to drive a monthly-resolution national-scale hydrological model. We use rainfall hindcasts from 1993-2016 to evaluate the performance of these flow forecasts and demonstrate their skill, particularly for the UK winter. We show that the high performance of the rainfall forecasts is spatially complementary to the skill provided by hydrological memory in groundwater-dominated catchments. Our analyses pinpoint the regions which would benefit most from future improvements in the rainfall forecasting or hydrological modelling systems. The introduction of these rainfall forecasts now enables hydrological forecasting at unprecedented levels of detail across Great Britain and is a model that may be similarly beneficial elsewhere in the world.

Copyright statement

1 Introduction

The UK Hydrological Outlook (https://hydoutuk.net/; Prudhomme et al., 2017; Boorman and Turner, 2019) is a monthly publication issued by the UK Centre for Ecology and Hydrology (UKCEH), the British Geological Survey (BGS), and the Met Office since 2013. It comprises estimates of the current hydrological conditions and forecasts of river flows and groundwater levels for the following one- and three-month periods. The Outlook pools forecasts from several methods, including a gridded hydrological model driven by meteorological observations and rainfall forecasts (Bell et al., 2013, 2017), ensemble streamflow predictions (e.g. Harrigan et al., 2018), and historical analogues (Svensson, 2014, 2016). Many of the modelled outputs are also made available on the Hydrological Outlooks Portal (https://ukho.ceh.ac.uk/), an interactive web service allowing users to inspect detailed forecasts from each method.

These hydrological forecasts at the sub-seasonal to seasonal timescales complement shorter-term weather and longer-term climate forecasts in providing actionable intelligence for water resources managers and emergency responders to take early actions to prepare for and mitigate extreme events (e.g. Murphy et al., 2001; Hamlet, Huppert and Lettenmaier, 2002; White, Franks and McEvoy, 2015; Anghileri et al., 2016; Portele et al., 2021; Jackson-Blake et al., 2022). Management strategies that have longer lead times, such as repositioning equipment or altering reservoir stocks, will benefit most from these seasonal hydrological forecasts. Although sub-seasonal to seasonal forecasts continue to have only moderate skill across the UK (compared to real-time weather and long-term climate forecasts; e.g. Vitart and Robertson, 2018), even with current skill they can enable low- and no-regret actions (those that are always net beneficial or have negligible or only opportunity costs) to be taken in a timely fashion.

The UK Hydrological Outlook was established in 2013 following the 2010-2012 drought (Kendon, Marsh and Parry, 2013) and the subsequent extreme rainfall that caused flooding over the rest of 2012 (Prudhomme et al., 2017). Comparable forecasting systems have also been developed for use in the USA (Demargne et al., 2014), Australia (Schepen and Wang, 2015), and sub-Saharan Africa (Sheffield et al., 2014). More recently, through the HydroSOS project, the World Meteorological Organisation has sponsored efforts to develop an equivalent global service, integrating the outputs of many national products (Jenkins et al. 2020).

Previously, Bell et al. (2013, 2017) have demonstrated the use of seasonal weather forecasts to drive the Grid-to-Grid/Water Balance Model (G2G/WBM), one of several hydrological forecasting schemes used to inform the Hydrological Outlook. The G2G/WBM scheme uses observed meteorology and the G2G hydrological model (Bell et al., 2009) to generate initial conditions (subsurface water stores). These initial conditions were then combined with rainfall forecasts produced by the Met Office's Global Seasonal forecasting system version 5 (GloSea5; Scaife et al., 2014; MacLachlan et al., 2015; Williams et al., 2018) in a simple monthly-resolution water balance model to derive estimates of river flows. However, this prior work was significantly limited by the lack of spatial resolution in the weather forecasts, with only a national mean rainfall provided. Weather forecasts were therefore downscaled to the resolution of the hydrological model by multiplying the national rainfall *anomaly* by the 1km standard average monthly rainfall. Although this method has been shown to perform reasonably well in Great Britain and is certainly superior to assuming uniform rainfall (e.g. Kay et al., 2023), it assumes a spatially uniform rainfall anomaly and thus fails to capture the significant variations that are often present across the country. As a result, Bell et al. (2017) limited their assessment of hydrological forecast skill to regional-mean flow anomalies and operational forecasts issued in the Hydrological Outlook were similarly aggregated, reducing their utility.

In this paper, we discuss the use of recently-developed ensemble rainfall forecasts derived from Historic Weather Analogues (HWA) to drive the Grid-to-Grid/Water Balance Model (G2G/WBM). These weather forecasts can be constructed at spatial resolutions equivalent to the hydrological model, potentially enabling the release of more useful high-resolution river flow forecasts. Section 2 describes our hydrological forecasting scheme in detail.

The ensembles of HWA-derived rainfall forecasts used in this work are taken from observed sequences selected from corresponding periods of the historical record and are described in detail in Sect. 2.2 and by Stringer et al. (2020). This approach is similar to the widely used ensemble streamflow prediction (ESP) method, in which a hydrological model is driven by each sequence of observed rainfall from previous years, providing a forecast consisting of a probability distribution constructed by weighting each ensemble member uniformly (e.g. Day, 1985; Harrigan et al., 2018; Troin et al., 2021). However, we benefit from two significant improvements over the standard ESP procedure:

(1) Rather than only using observations directly from the historical record, we use a synthesised sample of pseudo-observations. Each synthetic weather sequence consists of a unique combination of three sub-samples drawn from historical observations, increasing the sample from 63 possible sequences to 63^3=250047 synthetic sequences. This allows physically plausible extreme rainfall events that have not previously been observed to be included (e.g. van den Dool 2007; Svensson et al., 2016; Stringer et al., 2020).

(2) Rather than using an ensemble consisting of all (pseudo-)observations, we select a subset by matching the associated large-scale atmospheric circulation patterns that drive UK precipitation. This can be viewed as adopting a non-uniform weighting (in this work, a binary weight) on each candidate ensemble member, rather than assuming all are equally probable (e.g. Yao and Georgakakos, 2001; Svensson, 2016; Sabzipour, Arsenault and Brissette, 2021).

This work presents the first application of these HWA rainfall forecasts to gridded and national-scale seasonal hydrological forecasting in Great Britain. However, at catchment scales, Donegan et al. (2021) used HWA-derived rainfall forecasts (which they called ‘conditioned ESP’ to contrast with ‘historical ESP’) to drive hydrological forecasts for selected catchments in the Republic of Ireland over the winter. They concluded that this approach outperforms ‘historical ESP’ in almost all catchments at one- and two-month lead times. It furthermore remained skilful at three-month lead times where ‘historical ESP’ had practically no skill. On this basis they recommend that conditioned ESP is made operational in Ireland. Similar work is now in progress to apply HWA forecasts to catchment models in Great Britain (Chan et al. in prep).

The present work is structured as follows. In Sect. 2 we describe the G2G/WBM hydrological forecasting scheme, including the approaches used to estimate the hydrological initial conditions and to derive rainfall forecasts. In Sect. 3 we assess the performance of the HWA rainfall forecasts, extending the assessment of Stringer et al. (2020) beyond the winter season. In Sect. 4 we assess the performance of the G2G/WBM hydrological forecasting scheme and in Sect. 5 explore the relative contributions of rainfall forecasts and hydrological initial conditions to the forecast skill. We conclude in Sect. 6.

2 Forecasting methodology

The G2G/WBM method for forecasting river flows comprises three stages, summarised in Fig. 1:

(1) Hydrological initial conditions (HICs) are estimated using a run of the G2G hydrological model over preceding months driven by observed rainfall and actual evaporation, up to the forecast time origin.

(2) Rainfall forecasts are constructed from historical rainfall data using HWA forecasts.

(3) These two inputs are used to drive the Water Balance Model to forecast river flows.

Here, we describe each of these steps in greater detail. All data are generated, and models run, on the Ordnance Survey GB National Grid (EPSG:27700) at 1km spatial resolution.

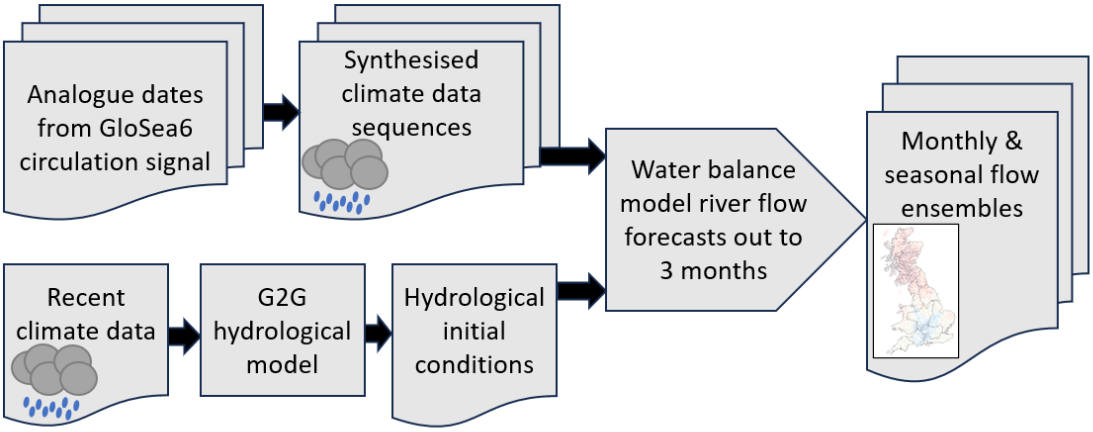

Analogue dates from GloSea6 circulation signal
Synthesised climate data sequences
Water balance model river flow forecasts out to 3 months
Monthly & seasonal flow ensembles
Recent climate data
G2G hydrological model
Hydrological initial conditions

Figure 1: Schematic of the G2G/WBM hydrological forecasting scheme, here driven by rainfall forecasts derived from Historic Weather Analogues.

2.1 Hydrological Initial Conditions

The initial conditions for the Water Balance Model consist of simulated estimates of the subsurface water stores at 9am GMT on the first day of each month (i.e. the start of the first ‘water day’). Each month’s initial stores are estimated using a G2G run beginning from the initial states at 9am on the first day of the previous month and incrementing in 15-minute timesteps over that month. The model runs on 365-day years (and 366-day leap years), correctly handling variable month lengths and leap days.

Meteorological inputs for the G2G run are provided by the Met Office and comprise the latest 1km resolution daily total rainfall from HadUK-Grid[1] and potential evaporation at 40km resolution produced using MORECS v2.0 (Hough and Jones, 1997), the latter uniformly downscaled to the 1km model resolution. The model connects together a probability-distributed model (PDM; Moore, 2007) for rainfall-runoff generation, a modified kinematic-wave routing scheme (Bell et al., 2009) for lateral flow on the surface and in a subsurface layer with flow directions derived from the UKCEH Integrated Hydrological Digital Terrain Model (IHDTM), and a return flow path. The G2G model is described in detail by Bell et al. (2009).

Hydrological initial conditions are available from 31 December 1962, based on model runs initialised from January 1961 and allowing 2 years for the model to reach equilibrium.

[1] https://www.metoffice.gov.uk/research/climate/maps-and-data/data/haduk-grid/haduk-grid

https://doi.org/10.5194/egusphere-2025-2506
Preprint. Discussion started: 24 July 2025

2.2 Rainfall Forecasts

The Water Balance Model requires monthly total rainfall forecasts as an input. There are a variety of methods that could be used to generate such forecasts, which fall on a spectrum ranging from the use of raw outputs from process-based atmospheric models to the use of historical sequences with no knowledge of contemporary atmospheric conditions (above described as 'historical ESP'). The original G2G/WBM forecasting scheme sat at one extreme of this spectrum, using raw weather forecasts from GloSea5 (Bell et al., 2017). By contrast, the new HWA forecasts sit in the centre ground. Much like the use of raw model outputs (and unlike historical ESP), they are sensitive to contemporary atmospheric conditions. However, like traditional historical ESP (and unlike purely process-based models) they also benefit from our understanding of historical variations in weather patterns at high spatial resolutions.

A key weakness of the original scheme was the necessity of downscaling national mean rainfall forecasts to the 1km spatial resolution of the hydrological models. This was done by assuming a spatially-uniform rainfall anomaly, as described above. Since observed rainfall anomalies are rarely spatially uniform, the resulting hydrological forecasts were only published as regional averages rather than at the native 1km resolution. The introduction of analogue-based forecasts has substantially mitigated this concern. Analogue rainfall forecasts have the resolution of the observations from which they were constructed. In this work we use the 1km Had-UK Grid dataset (v1.3.0.ceda; Hollis et al., 2019; Met Office, 2024a) which matches the native resolution of the hydrological model with no need for any downscaling.

The HWA approach taken in this work is an extension of that presented in Stringer et al. (2020) and Donegan et al. (2021). In outline, a c.40-member lagged ensemble of GloSea6 (Met Office, 2024b) MSLP forecasts covering the North Atlantic and Europe is constructed. A set of synthetic MSLP observations, each consisting of the observed weather averaged from three analogue years, are matched to these forecasts, with the best ten matches selected as the analogues. However, for seasonal forecasts initialised with start dates between November and February (inclusive), selection is made following the amplification of the NAO signal in these patterns. This is to correct for the 'signal to noise paradox' (Scaife and Smith, 2018; Stringer et al., 2020), in which the ensemble mean variance is smaller than is consistent with its relatively high correlation with observations. The final rainfall forecasts are then constructed from these best-matching combinations of analogue years. To correct for the effects of climate change, a linear detrending factor calculated from the HadUK-Grid observations is applied. Full details of this method will be given in Stringer et al. (in prep.).

The quality of HWA rainfall forecasts follows from the skill of the atmospheric model in predicting the proxy variable (here, the North Atlantic MSLP field) on which the forecasts are conditioned, and the strength of the correlation between that proxy and local (1km) rainfall. Thus, the use of the proxy allows us to avoid model biases associated with the calculation of rainfall, but it introduces into the forecast any biases in the prediction of the proxy, and the scatter of the correlation. This means that in the summer, where the performance of GloSea at predicting MSLP patterns is less good than in winter (Lockwood et al., 2023), the quality of rainfall forecasts and thus hydrological forecasts will suffer to some extent.

There are a few caveats to note. The use of observations means that synthesised weather forecasts can, in principle, be constructed with arbitrary temporal resolution. However, if we were to construct daily resolution rainfall sequences from forecasts conditioned on the monthly- or seasonal-mean atmospheric circulation patterns, the uncertainty in each day's forecast rainfall would be significant. Additionally, since each forecast ensemble member consists of three sub-periods, there will be discontinuities at the boundaries. We therefore do not consider sub-monthly rainfall forecasts and instead use only monthly totals.

The resampling process increases the ensemble size (from c.40 raw GloSea6 ensemble members to c.400 analogue resamples) to a size that would be too computationally intensive to produce by running GloSea6 many more times. This means that the resulting forecast probability distributions are smoother, and the extremes better sampled, but these ensemble members are not truly independent samples of the underlying distribution. This since each set of resamples are derived from the same raw ensemble member. The effect of the increased ensemble size in reducing the uncertainty in the predictand will thus be less than naively expected, with the degree of this reduction a function of these artificial covariances between ensemble members.

2.3 River flow forecasts using the Water Balance Model

Forecasts of river flows are produced by a simple water balance model originally presented by Bell et al. (2013). This model operates on a 1km spatial resolution and 1-month temporal resolution and forecasts subsurface water stores and river flows. First, stores are estimated by assuming that at each timestep the new store is the combination of the initial store, inputs from precipitation, and outputs from evaporation and outflows. The store at the end of month $m + 1$ is thus given by

$$S_{m+1} = S_m + P_{m+1} - E_{m+1} - Q_{m+1} \tag{1}$$

where $S_m$is the total subsurface water store at the end of month m, and P_{m+1}, E_{m+1}, and Q_{m+1} the forecasted precipitation, actual evaporation and outflow over month $m + 1$.

For the UK Hydrological Outlook implementation of the Water Balance model, we derive the initial store $S_m$from the G2G run described in Sect. 2.2 and the forecast precipitation as described in Sect. 2.1. Evaporation is estimated using the monthly mean actual evaporation as estimated from a long-baseline run of G2G (see Eq. 7 of Bell et al. 2009). Outflows are estimated as a function of water storage and subject to simplifying assumptions, as discussed by Bell et al. (2017).

The output of the water balance model consists of 1km grids of a forecast estimate of the next month's store $S_{m+1}$and runoff Q_{m+1}. Runoff from each upstream cell is accumulated spatially to produce the final river flow forecast estimate. For three-month forecasting, the first month uses the G2G hydrological initial conditions described in Sect. 2.1, with each subsequent month adopting the forecast stores from the water balance model for the preceding month.

2.4 Forecast products

Forecast river flows for each 1km grid-cell are converted into flow anomalies by dividing by the corresponding long-term mean river flow from an observation-driven baseline run of the WBM from 1965-2019 (inclusive). For three-month forecasts, after calculating anomalies for each month separately, the three months are averaged to produce a three-month ahead river flow forecast anomaly.

Finally, river flows are categorised into seven classes based on the historical distribution of flow anomalies. The classes follow those used in the Environment Agency Water Situation Report for ease of comparison (Table 1). The threshold river flow anomaly is calculated for each cell for each month, and for the three-month forecast for each combination of three sequential months.

Table 1: Classes used to categorise forecasts relative to their historical distributions. Each class has a name and characteristic colour.

Class	Descriptor	Colour	Percentile
7	Exceptionally high	🟦	P>0.95
6	Notably high	🟦	0.87<P<0.95
5	Above normal	🟦	0.72<P<0.87
4	Normal	🟨	0.28<P<0.72
3	Below normal	🟧	0.13<P<0.28
2	Notably low	🟥	0.05<P<0.13
1	Exceptionally low	🟥	P<0.05

3 Rainfall forecast performance assessment

In this Section we use a sample of hindcasts (re-forecasts of past cases) generated using the forecasting scheme described in Sect. 2.2 to assess the spatial and seasonal variations in the performance of the HWA rainfall forecasts. These hindcasts emulate the 3-month ahead forecasts that would have been produced at the start of each season (DJF: Winter, MAM: Spring, JJA: Summer, SON: Autumn) and for 1-month ahead at the start of each month. The 1-month forecasts span the period February 1993 to January 2017 (inclusive) while the 3-month forecasts span Spring 1993 to Winter 2016/17. The properties of this sample are summarised in Table 2.

Stringer et al. (2020) have already demonstrated, using only the winter hindcast sample, that the ensemble means of HWA rainfall forecasts correlate well with observed rainfall over much of the UK (see Fig. 9b of Stringer et al., 2020). The best performing areas of Great Britain are seen in northern Scotland and along the west coasts of England and Wales. Importantly, and as we will discuss in Sect. 5, these are the regions where accurate rainfall forecasts are most important due to the short hydrological memory.

Table 2: Availability of hindcasts for the HWA rainfall forecasts.

Forecast period	First hindcast	Last hindcast	Hindcast count[a]	Ensemble size[b]

https://doi.org/10.5194/egusphere-2025-2506
Preprint. Discussion started: 24 July 2025

1 month	Feb. 1993	Jan. 2017	24	140
3 month	MAM 1993	DJF 2016/17	23	510

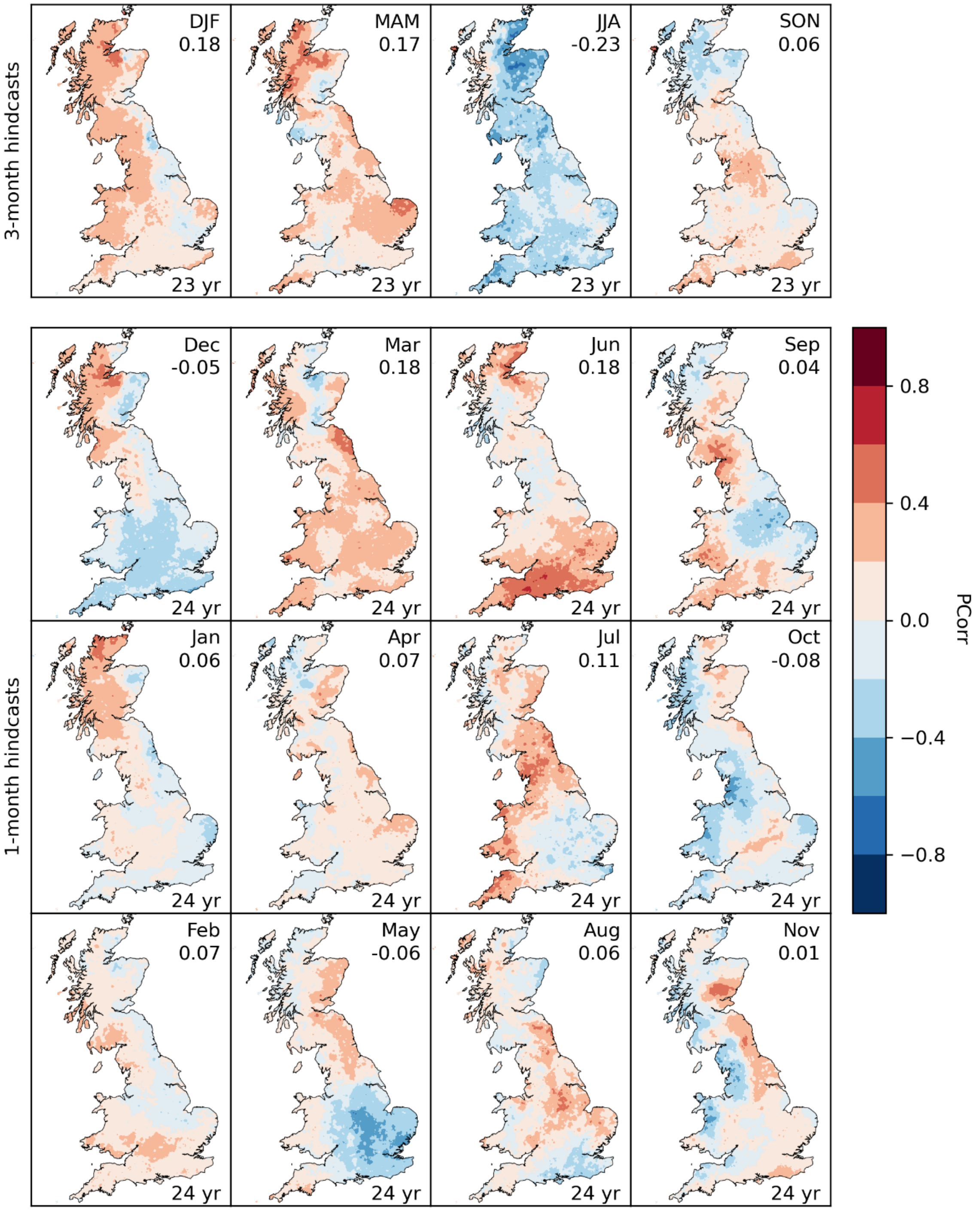
[a] The number of years of hindcasts available.

[b] The number of rainfall ensemble members within each hindcast.

We extend the analysis of Stringer et al. (2020) in our Fig. 2, which shows the Pearson correlation (PCorr) between the HWA rainfall forecast ensemble mean and observed rainfall from the HadUK-Grid dataset at 1km resolution across the full sample of hindcasts (3-months ahead for all four seasons and 1-month ahead for each month). The top-left panel (Winter) is equivalent to Fig. 9b of Stringer et al. (2020), but at 1km rather than 0.25° (c. 28km) resolution, while the other panels are new in this work.

As can be clearly seen in Fig. 2, correlations between ensemble mean and observed rainfall vary both spatially and temporally. Across the winter months positive correlations follow the west coast, consistent with the good correlation between observed rainfall and NAO index in these regions. The correlations for the three-month hindcasts (top row) vary significantly over the year, with summer exhibiting the worst performance. This follows from the lack of skill in GloSea6 in predicting summer MSLP (e.g. Fig. 5 of Lockwood et al., 2023), which is consistent with the difficulties experienced by most seasonal prediction systems in capturing summertime teleconnections (Knight and Scaife, 2024).

Overall, these correlations are not particularly strong (even given the small hindcast sample) with only a few areas rising above the threshold for statistical significance (PCorr>0.33 for p>0.95). One reason for this is that they are performed at the 1km scale, which is inevitably much noisier than at larger scales. Correlations at catchment or regional scales are somewhat larger. However, even for these modest correlations, if the skill is in the right places – those regions where hydrological memory is short and thus river flows are more sensitive to recent rainfall – the rainfall forecasts can still add value to our hydrological forecasts. This is discussed in detail in Sect. 5.



225 **Figure 2: Pearson correlation between HWA forecast ensemble mean and observed total rainfall for 3-month (top row) and 1-month (bottom three rows) hindcasts. A positive correlation is indicated in red while a negative correlation is in blue. In each panel the number in the top-right corner is the spatial mean while the number of hindcasts used to calculate the correlation is in the bottom-right corner.**

4 River flow forecast performance assessment

In this Section we consider the performance of river flow forecasts constructed using the G2G/WBM scheme outlined in Sect. 2. We use the same sample of hindcasts described in Sect. 3, consisting of 3-month forecasts for each of the four seasons and 1-month forecasts for every month.

We adopt simulated river flows using the G2G hydrological model driven by observed rainfall and potential evaporation as the 'ground truth' river flows against which we will validate predictions. This has the advantage over actual physical measurements of providing flow estimates at ungauged locations, which comprise the vast majority of our model outputs. Bell et al. (2009) compared the performance of the G2G model to observed flows at selected catchments and showed that, considering the relatively simple parameterisation, the model performs reasonably well. In addition, it should be noted that we are modelling only natural flows (unaffected by human influences such as abstractions and discharges), and so using observed flows (which are often affected in this way) would be an unfair comparison. Finally, our operational forecast products are also expressed as anomalies relative to a modelled baseline, making this approach particularly appropriate. If predictions for gauged locations are ultimately required, these anomalies can be applied to the observed flow distribution to provide a first-order bias correction.

The results of our forecasting system are an ensemble of deterministic predictions, which approximate the underlying probability distribution. Assuming that the ensemble is a fair sample of that underlying probability distribution (both of sufficient size and unbiased) we can apply the performance analysis metrics appropriate to probabilistic forecasts to these ensembles. We will therefore divide this Section into three parts. In the first we will assess the performance of the ensemble mean, in a manner similar to Stringer et al. (2020) and our Sect. 3 for rainfall forecasts, and to Bell et al. (2017) for river flows in the original G2G/WBM configuration. We will then assess the performance of the full forecast distribution, crediting a forecast for being both confident (a narrow distribution) and correct (centred on the observation). We will finally assess the performance of the distribution tails in detecting above or below normal flows, the events in which end-users of the Hydrological Outlook are most interested. We defer discussion of the relative contributions of the rainfall forecasts and hydrological initial conditions to Sect. 5.

4.1 Performance of the ensemble mean

The ensemble mean performance is assessed, as for rainfall forecasts in Sect. 3, by the Pearson correlation (Fig. 3) between that mean and the 'ground truth' (simulated) river flow, both expressed in m^3s^{-1}. When flows are expressed as percentiles of the climatological distribution (as is done operationally) the correlations follow very similar spatial patterns.

Here, we only show the correlation for pixels with non-negligible annual mean runoff (>0.05 m^3s^{-1}) which excludes upper reaches with high volatility. The performance is good overall, with weaker performance only shown in the summer,

https://doi.org/10.5194/egusphere-2025-2506
Preprint. Discussion started: 24 July 2025

consistent with results for rainfall in Sect. 3. Even in summer (JJA) a statistically significant correlation is present across the chalk aquifers in South-East England, where hydrological memory provides additional skill not present in the rainfall 260 forecasts (Fig. 2). In addition (and unlike for rainfall forecasts), the positive correlations are statistically significant over much of the country in many months, with the GB-average of the correlations exceeding the threshold for statistical significance in all months.

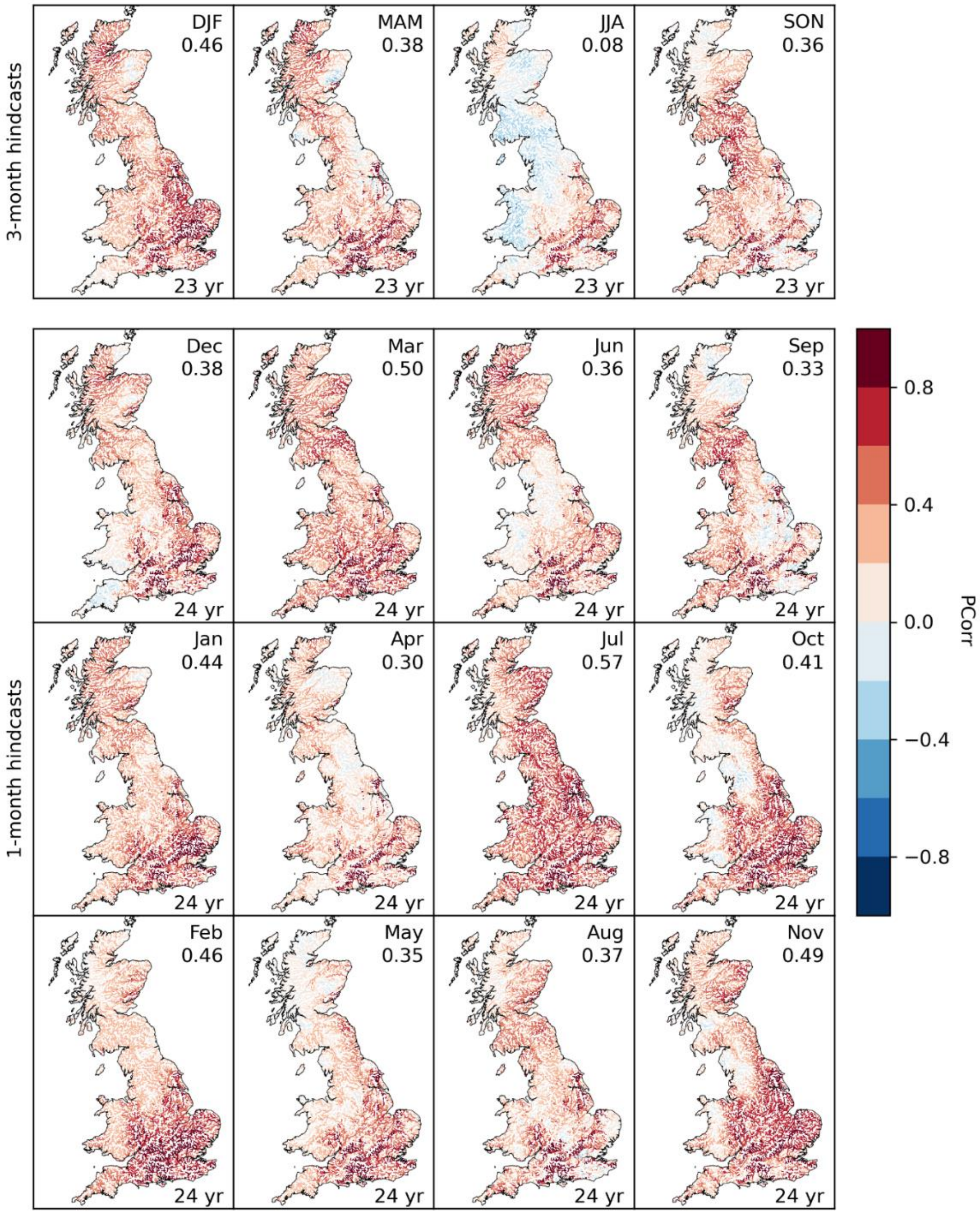


Figure 3: As Fig. 2, but for the correlation between G2G/WBM forecast ensemble mean and 'ground truth' river flows for the 265 **hindcast sample.**

In general, forecast ensemble means perform less well in northern England, where hydrological memory is not particularly long, nor rainfall forecasts especially skilful. However, there is also substantial variation over the year in which areas exhibit no correlation. Moreover, for any particular river cell skill at different lead times can vary. For instance, forecast river flows in Cornwall are not correlated with simulated observed flows in the summer at 3-months ahead but are correlated in all three summer months at 1-month ahead. Such spatial and temporal variations in skill underline the importance of understanding where, when and how far ahead a forecasting scheme is skilful when interpreting results.

4.2 Performance of the forecast distribution

Any probabilistic forecast expresses the likelihood (or confidence) of differing predictions. Forecast quality is partly characterised by the ability to predict signals (i.e. through mean skill – see above) and partly by reliability, which requires the forecast to predict an outcome with the same frequency that the outcome is observed. A useful performance statistic that compares a forecast distribution to an observation will therefore need to penalise forecasts for being over- or under-confident. A suitable choice for the latter is the Continuous Ranked Probability Score (CRPS; e.g. Brown, 1974; Matheson and Winkler, 1976; Hersbach, 2000; Wilks, 2011), which is defined as

$$CRPS = \int_{-\infty}^{\infty} \left(CDF(x) - H(x,o)\right)^2 dx \tag{2}$$

where $CDF(x)$ is the cumulative distribution function (the probability of the predictand having value less than or equal to x), $H(x,o)$ is the Heaviside step function, i.e.

$$H(x,o) = \begin{cases} 0 & x < o \\ 1 & x \geq o \end{cases} \tag{3}$$

where o is the observed value of the predictand. The CRPS can thus be viewed as analogous to the squared error, comparing the cumulative distribution functions of the forecast and the observation. In practice we do not have the probability distribution function but only an ensemble drawn from the distribution. The finite size of the ensemble can bias the CRPS, so we use the 'fair' CRPS, which corrects the score towards that which would be obtained from an infinite ensemble (Ferro, 2013).

In the limit of a deterministic system, a perfect forecast (a delta function at the correct value) will have a CRPS of zero. However, when a system has intrinsic uncertainty the best forecast probability distribution will match that intrinsic uncertainty, and thus the optimal CRPS will be greater than zero. Since the intrinsic uncertainty is unknown, we instead compare the CRPS of one forecasting system to another to determine the skill.

We thus calculate the Continuous Ranked Probability *Skill* Score (CRPSS; e.g. Wilks, 2011), defined as

$$CRPSS = 1 - \frac{\langle CRPS_{\text{forecast}} \rangle}{\langle CRPS_{\text{reference}} \rangle} \tag{4}$$

where angled brackets denote an average over all hindcasts and subscripts indicate the examined forecasting scheme and the reference forecasting scheme. We adopt the climatological distribution of river flows as our reference forecasting scheme.

https://doi.org/10.5194/egusphere-2025-2506
Preprint. Discussion started: 24 July 2025

This is a 55-member ensemble consisting of the WBM estimates of river flows obtained using the observed 1km HadUK-Grid rainfall and MORECS PE over the equivalent month or season of the years 1965-2020. The CRPSS can therefore be thought of as indicating the *value added* by atmospheric and hydrological modelling over assuming river flows will follow their historical patterns.

It should be noted that the CRPSS is not a direct measure of the performance of the forecasts issued to the Hydrological Outlook. This is because we categorise flows into the 7 classes given in Table 1. The equivalent statistics are then the Ranked Probability Score (RPS; Epstein, 1969; Murphy, 1971) and Ranked Probability Skill Score (RPSS; Wilks, 2011), which is also corrected for ensemble size (e.g. Buizza and Palmer, 1998; Müller et al., 2005; Weigel, 2007). Using the 7-class HOUK classification scheme shifts the domain over which the predictand is calculated from river flows (m^3s^{-1}) to quantiles of the historic distribution of river flows, which are then categorised into unequally-sized classes (varying between 5% of the distribution at the extremes to 44% for the central class). We thus calculate both the CRPSS (performance at predicting river flows) and the RPSS (performance at predicting the observed class) separately. Spatial and temporal patterns are, however, broadly similar in both cases.

Figure 4a shows the CRPSS for the G2G/WBM river flow forecasting scheme, as evaluated for the hindcast sample previously described. Seasonal forecasts perform well over much of Great Britain, especially in the South East, except in summer. Figure 4b shows the RPSS, which has similar patterns. We note that in regions with long hydrological memory (particularly the chalk aquifers of south-east England), the classified forecasts (RPSS; Fig. 4b) are particularly skilful. Indeed, they are more skilful than the actual river flows (CRPSS; Fig. 4a). This is likely because in these regions residuals between forecasts and 'ground truth' river flows are smaller than the classes, meaning that the forecast consistently selects the observed class.

https://doi.org/10.5194/egusphere-2025-2506
Preprint. Discussion started: 24 July 2025

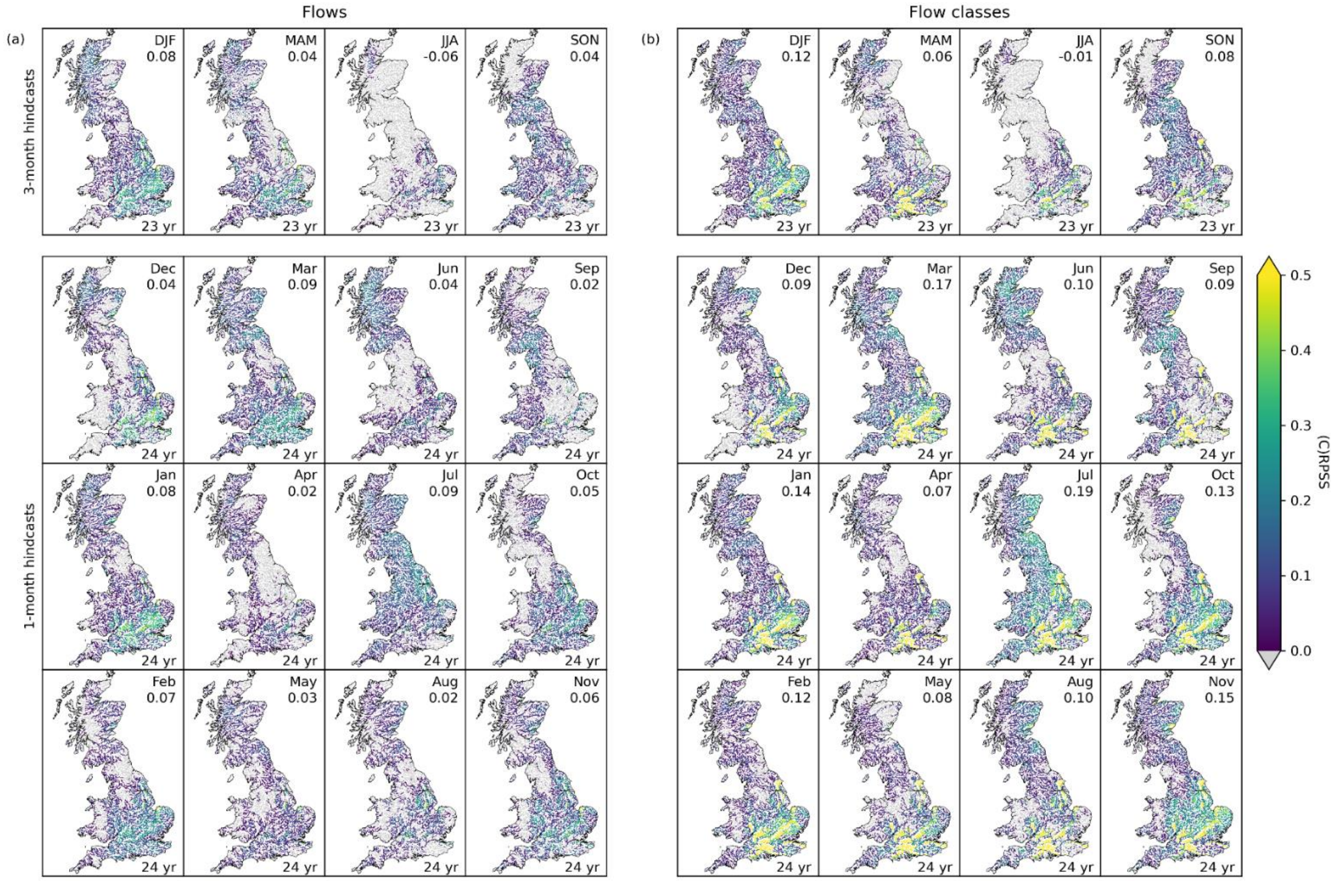


Figure 4: The skill of the G2G/WBM river flow forecasting scheme over assuming the climatological distribution. Left: CRPSS for predicting volumetric river flows (m3/s). Right: RPSS for predicting classified flows (Table 1). Scores greater than 0 (coloured) indicate the G2G/WBM scheme outperforms the climatological distribution, while those less than zero (grey) indicate the converse. The spatial mean is shown in the top-right of each panel, with the number of hindcasts in the bottom-right.

4.3 Performance of the forecast distribution tails

While the CRPSS served to assess the overall performance of the forecasting scheme, it does not emphasise the most important feature of any hydrological forecast, the prediction of extrema. It is thus useful to also use a measure of the forecasting scheme's ability to discriminate between events and non-events.

The Relative Operating Characteristic (ROC; e.g. Swets, 1973; Mason, 1982; Mason and Graham, 1999) is such a statistic. The ROC assessment consists of assuming that the forecasting system triggers an alert whenever the forecast probability of some event exceeds a pre-defined threshold. The proportion of observed events for which an alert was triggered (the 'hit rate') and the proportion of observed non-events for which an alert was triggered (the 'false alarm rate') can then be placed onto a graph. By varying the threshold at which an alert is triggered, a curve is constructed to indicate the performance of the alert system at these thresholds. Different forecasting schemes can then be compared by integrating the area under this curve

(AUC), which will vary from 0 to 1. A forecasting scheme in which an alert is triggered entirely randomly, but with the probability that the event occurs in the sample, will have an AUC=0.5. For convenience, however, we map the skilful domain of the score to the match our other performance scores, such that

$$S_{\mathrm{ROC}} = 2AUC - 1 \tag{5}$$

where $S_{\mathrm{ROC}} = 1$ indicates a perfect forecast (detects all events with no false positives at all probability thresholds) and $S_{\mathrm{ROC}} > 0$ indicates the forecasting scheme is skilful over the random forecast described above.

In this work we use S_{ROC} to evaluate the success in predicting less-common rainfall/flow events. Specifically, we define 'high' rainfall/flow events as those falling in the 'above normal' classification or above (classes 5, 6 & 7), and a 'low' event as those falling in the 'below normal' classification or below (classes 1, 2 & 3), as defined in Table 1. This is a very conservative definition of flow 'events' and will include values close to the 28th or 72nd percentiles which are less likely to have significant impacts. However, it is adopted due to the small number of hindcasts available. If only very extreme categories were considered, the expected number of events in the hindcast would be negligible, and the S_{ROC} would have a large uncertainty.

Figure 5 shows the S_{ROC} for the performance of the G2G/WBM river flow forecasting scheme at detecting above and below normal flows. Although skill varies spatially and over the year, following similar patterns to those discussed above, there are a few particularly interesting features to note. At seasonal timescales, both above- and below- normal flows are best forecast in Winter and Autumn, while at monthly timescales below-normal flows are particularly well-forecast for March, and above-normal for July. We note that although one might naively expect that it is easier to detect a 'wet' summer or a 'dry' winter, since we define our events from the climatological distribution of flows over each forecast month or season this effect has already been removed. We are thus only showing the performance of the forecasting scheme at predicting a wetter-than-normal (or drier-than-normal) month/season, and so the plots in Fig. 5 can be compared fairly.

https://doi.org/10.5194/egusphere-2025-2506
Preprint. Discussion started: 24 July 2025

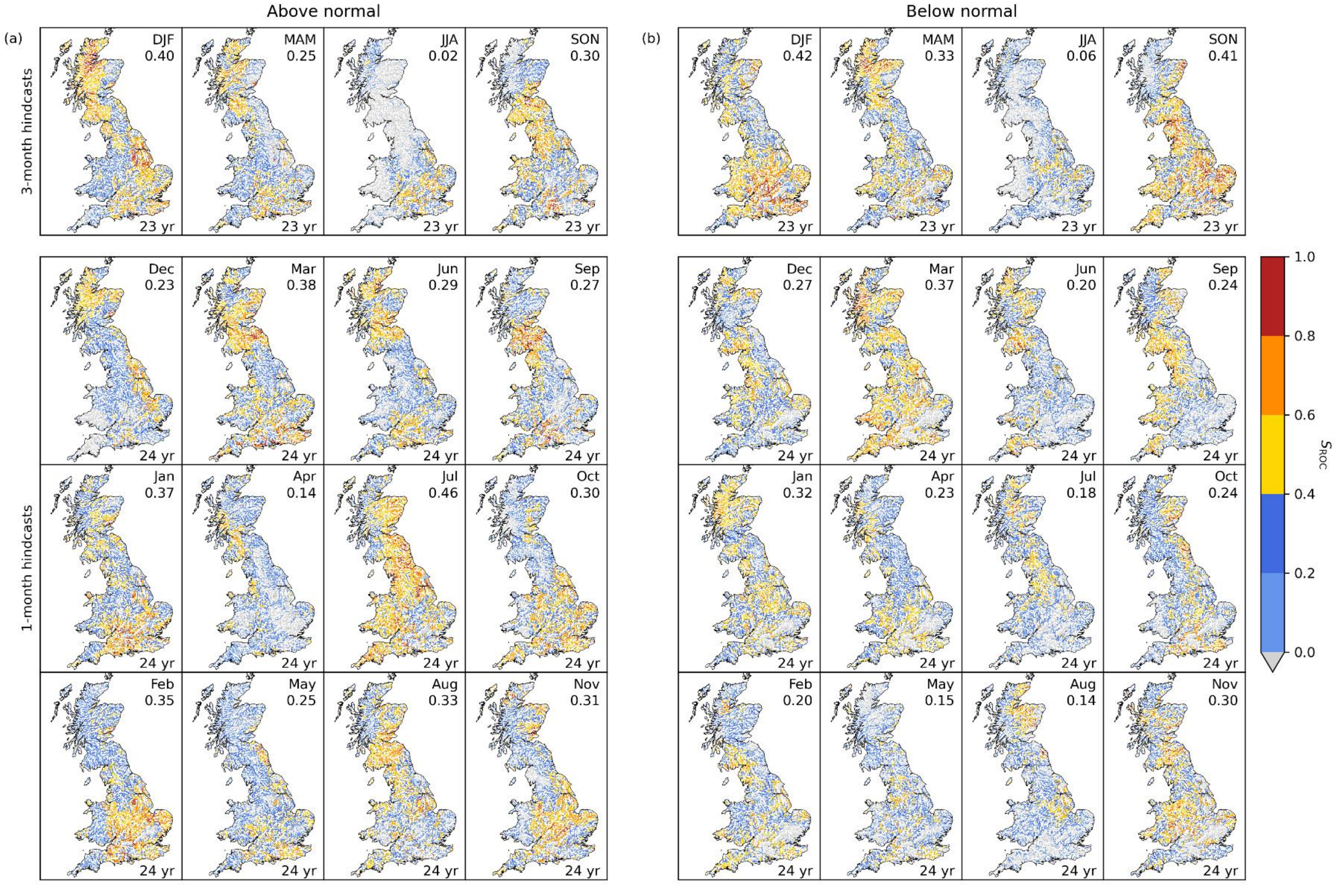


Figure 5: S_{ROC} for the G2G/WBM river flow forecasting scheme. The left array of panels (a) shows the performance of the model at forecasting above normal flows (>72nd percentile) and the right array (b) shows the performance for below normal flows (<28th percentile). Colours indicate the rescaled area under the ROC curve, with scores below 0 (grey) being inferior to a purely random process. Spatially-averaged values are shown in the top-right corner of each panel and the number of hindcasts used to calculate S_{ROC} is shown in the bottom-right.

In many parts of the country the forecasting scheme does not have the same ability to detect high flows as it does low (and vice versa). This may follow from differing relative contributions from hydrological initial conditions and recent rainfall to flows in different parts of the flow regime, and the associated variations in skill in predicting the inputs. Further work will be required to disentangle these effects.

4.4 Comparison with previous GloSea5-derived seasonal forecasts

Bell et al. (2017) assessed the skill of the G2G/WBM hydrological forecasting scheme using the spatially-uniform Glosea5 rainfall forecasts used in the Hydrological Outlook prior to December 2023. A particular strength of their analysis is that the relative contributions of rainfall forecasts and hydrological initial conditions to forecast skill, and the skill over persistence is considered. However, the statistical analysis was limited by the sample of hindcasts, which were only available for 13 years for selected 1-month (March, June, September and December) and 3-month (MAM, JJA, SON, DJF) lead times, and consisted of relatively small ensembles (either 12 members for March, September, MAM and SON or 24 members for June, December, JJA and DJF). As discussed above, the lack of spatial resolution meant that forecasts were only made at regional scales, and skill scores were calculated for temporally-aggregated combinations of hindcasts (such as a combined Spring-Summer sample) to increase the sample size. Bell et al. (2017) evaluated the Pearson correlation and the S_{ROC} score for their hindcasts, allowing us to make some comparisons between their results and ours (our Sect. 4.1 and 4.3).

Bell et al. (2017) combined Autumn-Winter and Spring-Summer hindcasts to obtain a sufficiently large sample to determine the correlation coefficient. They show that the combined Spring-Summer performance is poor across much of the country, with only a small number of regions in south-east England at both 1-month and 3-month and in southern Scotland at 1-month lead times showing a correlation. Since their forecasts were also based on rainfall outputs from an earlier GloSea version, it is probable that this originates in a similar lack of summer meteorological skill to that shown in Fig. 2 and Lockwood et al. (2023) for GloSea6. By contrast, their Autumn-Winter sample showed good correlations in almost all regions at both lead-times, with similar spatial patterns. Our Fig. 3 shows that the winter performance is better than autumn, but the correlations remain strong over almost the entire country in both seasons.

Bell et al. (2017) also performed an S_{ROC} analysis based on the same classifications of above normal and below normal flows as used in this work. However, their skill scores were not only aggregated over spatial regions compared to our 1km resolution results but were further averaged spatially (their Fig. 2) or temporally (their Fig. 3) and were not published as separate skill scores for above- or below-normal flows. While this aggregation may have been justified by their small ensembles and limited hindcast sample, it will have obscured many of the systematic variations in skill that are uncovered above.

Table 3: GB-mean S_{ROC} averaged between above- and below-normal flows for the previous GloSea5-derived forecasts (Bell et al., 2017) and in this work.

Month	Bell et al. (2017)	This work	Season	Bell et al. (2017)	This work
December	-0.07	0.25	**Winter (DJF)**	0.37	0.41
March	0.08	0.38	**Spring (MAM)**	0.10	0.29
July	-0.06	0.25	**Summer (JJA)**	-0.13	0.04
September	0.35	0.26	**Autumn (SON)**	0.41	0.36

An approximate comparison can nevertheless be made between the GB-mean S_{ROC} skill scores of our Fig. 5 and those given by Bell et al. (2017). The scores are shown in Table 3. Our 1-month GB-mean S_{ROC} for December is now positive, compared to Bell et al. (2017)'s negative score (indicating there is skill where there previously was not). The Spring GB-mean skill score has more than doubled. These examples likely indicate that the introduction of HWA forecasts offers an improvement over the previous method. It should, however, be noted that this comparison is inexact, since the GB-mean scores in Bell et al. (2017) are averages over the scores for the 17 regional-mean forecasts and are not weighted by the size of those regions as is done in this work.

5. Sources of forecast skill

The varied hydrogeology of Great Britain drives spatially-varying sensitivity to recent rainfall between different subsurface aquifers. For instance, in the chalk aquifers in the South East of England, river flows are only weakly sensitive to recent rainfall, dominated instead by subsurface storage. This means they have long hydrological memory, with forecasting methods based on persistence generally performing well and flows only slowly varying. River flow forecasting performance is predominantly derived from the quality of the hydrological initial conditions used.

By contrast, in more responsive catchments such as those in the north of Scotland, subsurface water storage constitutes a much smaller contribution to the flow. Forecast skill in these areas is more dependent on the quality of the input rainfall forecasts. A strength of conditioning GB rainfall forecasts on the atmospheric circulation is that the winter correlation between NAO and rainfall is strongest in the regions where rainfall is most important (see Sect. 3).

To investigate the relative importance of these contributions, we have run the WBM hindcasts using three different combinations of inputs:

(a) HWA rainfall forecasts with climatological-mean Hydrological Initial Conditions (HICs). This combination isolates the contribution of the *rainfall forecasts* to the river flow forecast skill.

(b) All historical rainfall sequences (1965-2020) with G2G-derived HICs, equivalent to historical ESP, quantifying the contribution of the *hydrological initial conditions* (existing surface and subsurface water stores at the start of the forecast) to the river flow forecast skill.

(c) HWA rainfall forecasts with G2G-derived HICs, the forecasting scheme presented in the preceding sections which combines *both rainfall and HIC forecast skill*.

https://doi.org/10.5194/egusphere-2025-2506
Preprint. Discussion started: 24 July 2025

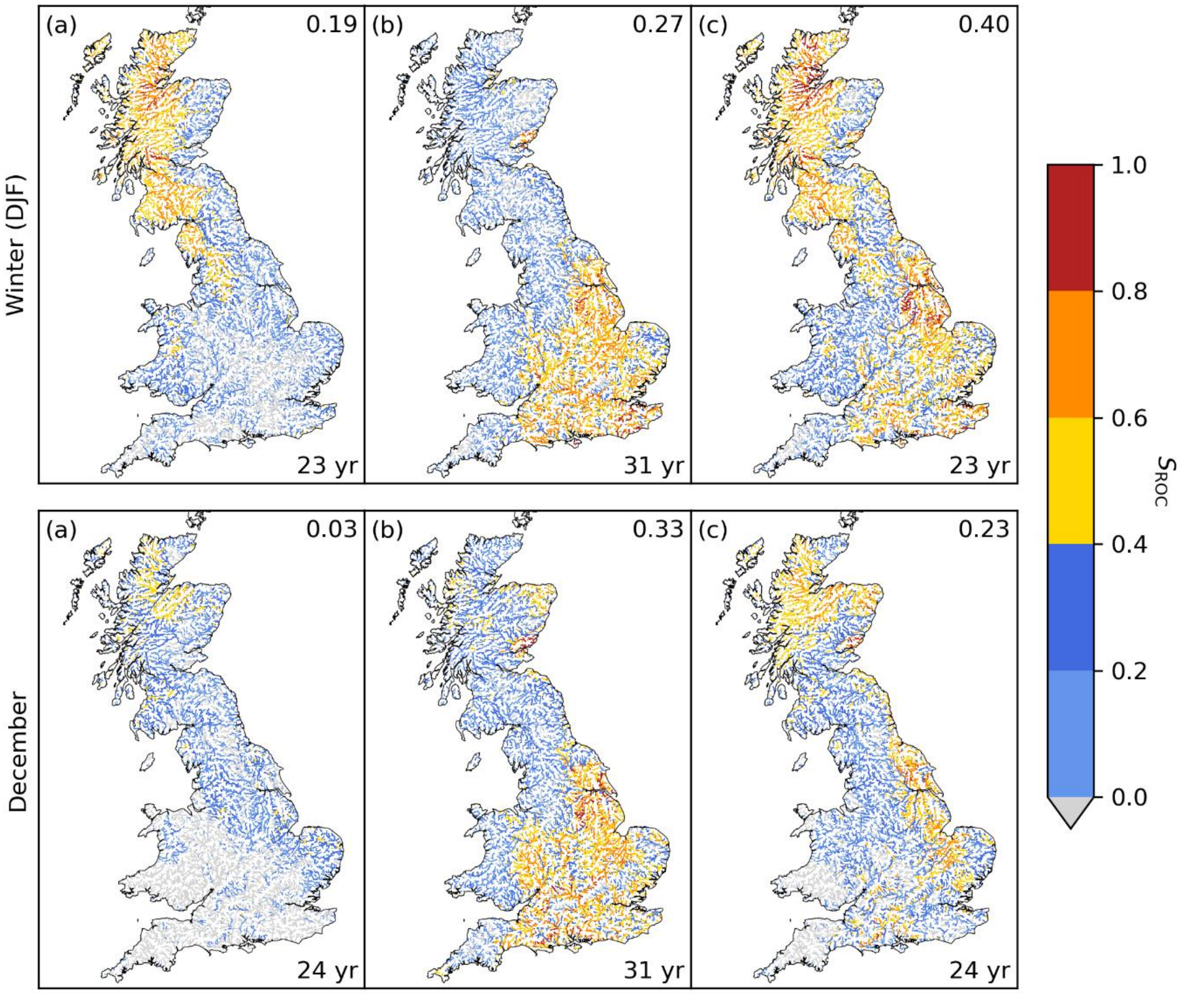


Figure 6: S_{ROC} for above normal river flows in Winter (top) and December (bottom) using different combinations of inputs to the WBM to (a): highlight HWA rainfall forecast skill (b): highlight skill derived from G2G HICs, and (c): the combined skill. Scores greater than 0 (coloured) indicate the forecasting scheme outperforms guesswork, while scores less than zero (grey) indicate the converse. The spatial means are shown in the top-right corner of each panel, and the number of hindcasts used to calculate the skill score is in the bottom-right.

Figure 6 shows the S_{ROC} score for forecasting above-normal flows across the winter (seasonal mean flows) and for the month of December (monthly mean flows) for each combination of WBM inputs. The winter maps show that the regions of high skill for each forecast driver are highly complementary, with the combination of HWA forecasts (exhibiting excellent skill in the North-West of Great Britain) and the HICs (in the South and East) providing good skill over the whole of Great Britain. Svensson et al. (2015; Fig. 2 & 3) showed that flows in selected catchments in the North and West correlate closely with the NAO index while those in the South-East correlate with the preceding month's flows. Since the NAO index correlates with rainfall (indeed, this is the correlation exploited by Stringer et al., 2020) while flow persistence follows from hydrological memory, these patterns mirror those found in this work. However, in our Fig. 6 they are revealed for the first time at national scale and for ungauged sites, revealing the pattern extends across the entire 1km river network.

However, the spatial complexity of this picture is shown by the lower row of Fig. 6, which shows the same skill scores for above-normal monthly-mean flows in December. In western Scotland neither rainfall forecasts nor HICs provide particularly high skill but once both are combined the skill metric significantly improves. In eastern parts of England the HICs continue to provide most of the skill. However, in southern Wales and central and southern England the lack of skill in the high-flow tail of the HWA rainfall forecasts degrades the good performance we might otherwise expect from the HICs. Further improvements to the rainfall forecasts should therefore be targeted at those regions where they will make the most difference. For example, to increase the skill of the December flow forecasts (Fig. 6, bottom row) the most valuable improvements would be to increase the skill in the rainfall forecasts in southern Wales and south west England.

A complementary approach to exploring the contribution of meteorological forecasting skill is to contrast the skill scores in river pixels representing catchments with high or low hydrological memory. We can infer these from the baseflow index (BFI; Gustard, Bullock & Dixon, 1992), which estimates the fraction of river flows that originate in subsurface stores. At the 1km grid of our model these are derived from the estimates of baseflow contributions associated with the hydrological properties of soils (BFIHOST; Boorman, Hollis & Lilly, 1995).

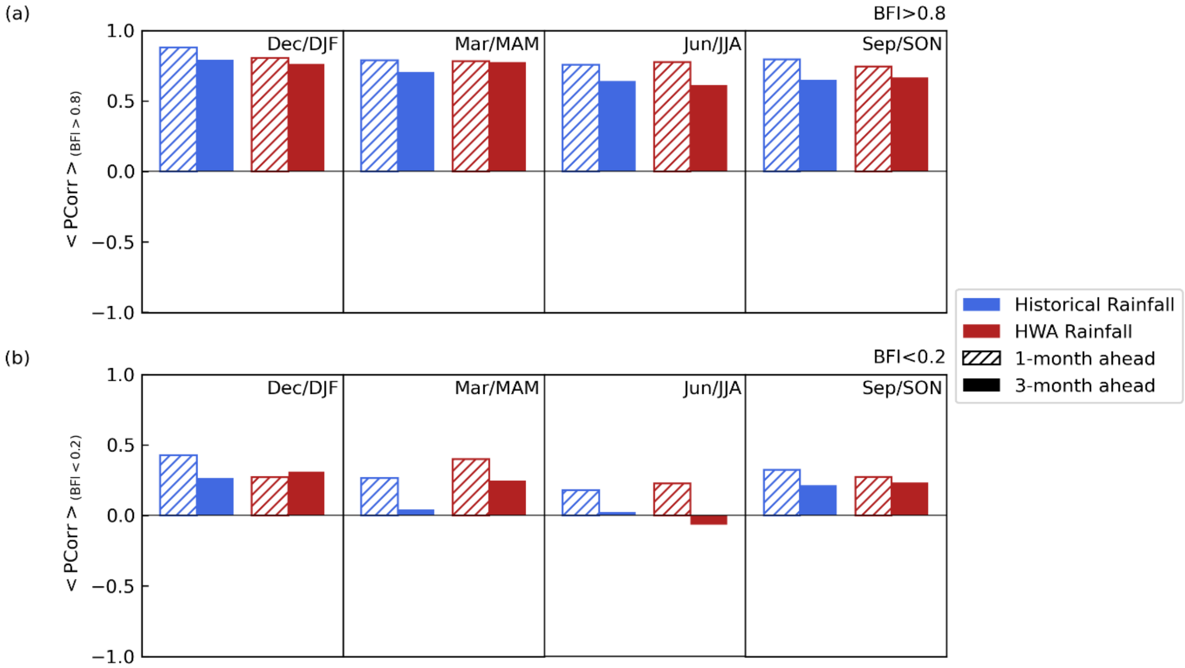


Figure 7: Pearson's correlation between river flow forecast ensemble mean and 'ground truth' river flows for monthly (hatched) and seasonal (solid) forecasts generated using the WBM driven by historical ESP (blue) and HWA (red) rainfall forecasts. Shown are the correlations averaged over river cells with (a) high-BFI and (b) low-BFI.

Figure 7 shows the correlation between forecast ensemble mean and simulated 'ground truth' flows for each of the four seasons, averaged over river cells with high BFI (top row) and low BFI (bottom row). The figure indicates the exceptional performance of the forecast ensemble mean in regions where flows are dominated by subsurface stores (Fig. 7a), with only modest reductions in skill between 1-month and 3-month ahead forecasts. This suggests that the length of hydrological memory persists beyond the seasonal timescale in regions with high BFI. In addition, we observe that there is very little variation in the strength of the correlation (and thus forecast skill) over the year because the hydrogeology does not change.

By contrast, the low-BFI sample (Fig. 7b) highlights only the catchments for which river flows are particularly sensitive to recent rainfall. The temporal variation in flow forecast skill in these catchments follows that of the rainfall forecasts, with the skill in seasonal flow forecasts improving in the three seasons in which the rainfall forecasts are skilful, excepting summer where the forecasts are poor (see Fig. 2). In most cases, the skill declines from 1-month ahead to 3-months ahead as the predictability of the rainfall declines with increasing lead time. The exception to this trend is the December/Winter skill, where predictability for both low and high BFI catchments is enhanced by the analogue forecasts being conditioned on the large-scale circulation patterns that correlate best at a 3-month-ahead timescale.

The analysis in this section has demonstrated the importance of both rainfall forecasts and hydrological initial conditions for reliable flow forecasts across Great Britain. However, the relative contribution of each source of forecast skill varies between regions where flows are dominated by subsurface stores and those sensitive to recent rainfall. Further efforts to improve the quality of UK hydrological forecasts by improving the rainfall forecasts should thus be directed at these regions where they can best complement the skill provided by hydrological initial conditions. The weather analogue forecasting approach offers the potential to do this by conditioning the ensemble on meteorological variables that correlate best with precipitation in these areas.

6. Conclusions

In this paper we have presented and assessed improvements to the G2G/WBM river flow forecasting scheme used in the UK Hydrological Outlook. The rainfall forecasts used to drive the hydrological modelling are now based on historical weather analogues (HWA) conditioned on large-scale circulation patterns, which replace an earlier scheme in which raw rainfall forecasts were spatially downscaled. We examine the skill of the new scheme, evaluating the performance of the probabilistic forecasts by comparing the forecast ensemble means (Sect. 4.1), distributions (Sect. 4.2) and high- and low-flow tails (Sect. 4.3) to simulated observed flows at 1km resolution. We determine that the forecasting scheme performs well across much of Great Britain in most months and seasons, with a particular exception in the summer deriving from the lower level of meteorological forecast skill in that season.

In line with previous work by Svensson et al. (2015), we found systematic variation in the sensitivity of river flows to recent rainfall in different parts of the country, with parts of the north and west being most sensitive and the south-east being least. The work here has demonstrated this phenomenon at 1km resolution over the whole of Great Britain, extending the small sample of gauged catchments originally sampled by Svensson et al. (2015). We thus emphasise the importance of targeting improvements to the modelling chain at the regions where they will have the most impact – for instance just as the use of historic weather analogues in this work has enabled the exploitation of correlations between the winter NAO index and rainfall in the north west of the UK, where this rainfall has a strong influence on river flows.

It is important to acknowledge that the analysis in this work has focused on only a relatively small number of available hindcasts, meaning that we have not been able to assess the performance of the forecasting scheme at detecting high-impact extremes (such as floods and droughts with >20-year return periods). We have also not explicitly assessed the ongoing impacts of climate change on the performance of these analogue forecasts, but the resampling and detrending steps used in the rainfall forecasting methods should mitigate this.

Further work over the coming years will continue to improve both the weather forecasting and hydrological modelling components of the forecasting scheme. Initially, the atmospheric model ensemble size will be increased as more supercomputer resources become available. This should provide greater prediction skill and better constraints on the probabilistic forecast distribution, especially in the high-impact tails. We also plan to investigate whether we are now able to exploit the finer temporal resolution offered by analogue rainfall forecasts to replace the simplistic monthly-resolution water balance model currently used with a process-based representation of the water cycle (such as the Grid-to-Grid hydrological model). Research is also underway to explore whether introducing a post-processing bias correction scheme would be appropriate to mitigate any remaining systematic biases, and to develop a system for blending the forecasts from the multiple models used in the Outlook together to better exploit each model's strengths.

The development of reliable 1km rainfall and flow forecasts has recently enabled the release of monthly high-resolution seasonal forecasts for Great Britain through the Hydrological Outlook Portal (https://ukho.ceh.ac.uk/). As seasonal hydrological forecasts continue to improve, they will become increasingly useful to a wide community of both professional and lay users.

Data availability

Grids of the forecast performance scores at 1km resolution will be made available on reasonable requests to the corresponding author.

Author Contribution

Mark Rhodes-Smith: Methodology, Software, Formal Analysis, Investigation, Writing, Visualisations

Victoria Bell: Methodology, Writing, Supervision

Nicky Stringer: Methodology, Investigation, Writing

Helen Baron: Software

Helen Davies: Methodology, Data curation

Jeff Knight: Conceptualisation, Methodology, Writing

Competing interests

515 The authors declare that they have no conflict of interest.

Acknowledgements

The UK Hydrological Outlook is supported by the Environment Agency, Natural Resources Wales, Scottish Environment Protection Agency and, for outputs that cover Northern Ireland, the Department for Infrastructure – Rivers. The authors thank Dr Rosanna Lane for helpful discussions and the suggestion of one of the skill metrics used.

520 Financial support

This work was supported by the Natural Environment Research Council under the UK Status, Change And Projections of the Environment (NE/R016429/1) and National Capability for UK Challenges 'Understanding the UK Environment' (NE/Y006208/1) programmes.

References

525 Anghileri, D., Voisin, N., Castelletti, A., Pianosi, F., Nijssen, B., and Lettenmaier, D. P.: Value of long-term streamflow forecasts to reservoir operations for water supply in snow-dominated river catchments, Water Resour. Res., 52(6), 4209–4225, doi:10.1002/2015WR017864, 2016.

Bell, V. A., Kay, A. L., Jones, R. G., Moore, R. J., and Reynard, N. S.: Use of soil data in a grid-based hydrological model to estimate spatial variation in changing flood risk across the UK, J. Hydrol., 377, 335–350, doi:10.1016/j.jhydrol.2009.08.031, 530 2009.

Bell, V. A., Davies, H. N., Kay, A. L., Marsh, T. J., Brookshaw, A., and Jenkins, A.: 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, 2013.

Bell, V. A., Davies, H. N., Kay, A. L., Brookshaw, A., and Scaife, A.: A national-scale seasonal hydrological forecast 535 system: development and evaluation over Britain, Hydrol. Earth Syst. Sc., 21(9), 4681–4691 doi:10.5194/hess-21-4681-2017, 2017.

Boorman, D. B., Hollis, J. M., and Lilly, A.: Hydrology of soil types: a hydrologically-based classification of the soils of United Kingdom, IH Report no.126, Institute of Hydrology, Wallingford, 146pp., https://nora.nerc.ac.uk/id/eprint/7369/, 1995.

Boorman, D. B. and Turner, S.: Assessing the skill of the UK Hydrological Outlook, Hydrolog. Sci. J., 64(15), 1932–1942, doi:10.1080/02626667.2019.1679375, 2019.

Brown, T. A.: Admissible scoring systems for continuous distributions, Manuscript P-5235, The Rand Corporation, Santa Monica, CA, 22 pp., https://www.rand.org/content/dam/rand/pubs/papers/2008/P5235.pdf, 1974.

Buizza, R. and Palmer, T. N.: Impact of Ensemble Size on Ensemble Prediction, Mon. Weather Rev., 126(9), 2503-2518, doi:10.1175/1520-0493(1998)126<2503:IOESOE>2.0.CO;2, 1998.

Day, G. N.: Extended Streamflow Forecasting Using NWSRFS, J. Water Res. Pl. Man., 111(2), 157–170, doi:10.1061/(ASCE)0733-9496(1985)111:2(157), 1985.

Demargne, J., Wu, L., Regonda, S. K., Brown, J. D., Lee, H., He, M., Seo, D., Hartman, R., Herr, H. D., Fresch, M., Schaake, J., and Zhu, Y: The Science of NOAA's Operational Hydrologic Ensemble Forecast Service, B. Am. Meteorol. Soc., 95(1), 79–98, doi:10.1175/BAMS-D-12-00081.1, 2014.

Donegan, S., Murphy, C., Harrigan, S., Broderick, C., Quinn, D. F., Golian, S., Knight, J., Matthews, T., Prudhomme, C., Scaife, A., Stringer, N., and Wilby, R.: Conditioning ensemble streamflow prediction with the North Atlantic Oscillation improves skill at longer lead times, Hydrol. Earth Syst. Sc., 25(7), 4159–4183, doi:10.5194/hess-25-4159-2021, 2021.

Epstein, E. S.: A Scoring System for Probability Forecasts of Ranked Categories, J. Appl. Meteorol. Clim., 8(6), 985–987, doi:10.1175/1520-0450(1969)008<0985:ASSFPF>2.0.CO;2, 1969.

Ferro, C. A. T.: Fair scores for ensemble forecasts, Q. J. R. Meteorol. Soc., 140, 1917-1923, doi:10.1002/qj.2270, 2014.

Gustard, A., Bullock, A., and Dixon, J. M.: Low flow estimation in the United Kingdom, IH Report No.108, Institute of Hydrology, Wallingford, 88pp., https://nora.nerc.ac.uk/id/eprint/6050/, 1992.

Fowler, H. J. and Kilsby, C. G.: Precipitation and the North Atlantic Oscillation: A study of climatic variability in Northern England, Int. J. Climatol., 22, 843-866, doi:10.1002/joc.765, 2002.

Hamlet, A. F., Huppert, D., and Lettenmaier, D. P.: Economic Value of Long-Lead Streamflow Forecasts for Columbia River Hydropower, J. Water Res. Pl. Man., 128(2), 91–101, doi:10.1061/(ASCE)0733-9496(2002)128:2(91), 2002.

Harrigan, S., Prudhomme, C., Parry, S., Smith, K., and Tanguy, M.: Benchmarking ensemble streamflow prediction skill in the UK, Hydrol. Earth Syst. Sc., 22(3), 2023–2039, doi:10.5194/hess-22-2023-2018, 2018.

Hersbach, H., 2000: Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems, Weather Forecast., 15, 559–570, doi:10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2, 2000.

Hollis, D., McCarthy, M., Kendon, M., Legg, T., and Simpson, I.: HadUK-Grid - A new UK dataset of gridded climate observations, Geosci. Data J., 6(2), 151-159, doi:10.1002/gdj3.78, 2019.

Hough, M. N. and Jones, R. J. A.: The United Kingdom Meteorological Office rainfall and evaporation calculation system: MORECS version 2.0-an overview, Hydrol. Earth Syst. Sc., 1(2), 227–239, doi:10.5194/hess-1-227-1997, 1997.

Jackson-Blake, L. A., Clayer, F., de Eyto, E., French, A. S., Frías, M. D., Mercado-Bettín, D., Moore, T., Puértolas, L., Poole, R., Rinke, K., Shikhani, M., van der Linden, L., and Marcé, R.: Opportunities for seasonal forecasting to support water management outside the tropics, Hydrol. Earth Syst. Sc., 26(5), 1389–1406, doi:10.5194/hess-26-1389-2022, 2022.

Jenkins, A., Dixon, H., Barlow, V., Smith, K., Cullmann, J., Berod, D., Kim, H., Schwab, M., and Silva Vara, L. R.: HydroSOS - The Hydrological Status and Outlook System towards providing information for better water management, WMO Bulletin, 69(1), 14-19, https://library.wmo.int/idurl/4/57750, 2020.

Kay, A. L., Rudd, A. C. and Coulson, J: Spatial downscaling of precipitation for hydrological modelling: Assessing a simple method and its application under climate change in Britain, Hydrol. Process., 37(2), 14823, doi:10.1002/hyp.14823, 2023.

Kendon, M., Marsh, T. and Parry, S. (2013) 'The 2010–2012 drought in England and Wales, Weather, 68(4), 88–95, doi:10.1002/wea.2101, 2013.

Knight, J. R. and Scaife, A. A.: Influences on North-Atlantic summer climate from the El Niño-Southern Oscillation, Quarterly Journal of the Royal Meteorological Society, 150(764), 4498–4510, doi:10.1002/qj.4826, 2024.

Lockwood, J. F., Stringer, N., Hodge, K. R., Bett, P. E., Knight, J., Smith, D., Scaife, A. A., Patterson, M., Dunstone, N., and Thornton, H. E.: Seasonal prediction of UK mean and extreme winds, Q. J. Roy. Meteor. Soc., 149(757), 3477-3489, doi:10.1002/qj.4568, 2023.

MacLachlan, C., Arribas, A., Peterson, K. A., Maidens, A., Fereday, D., Scaife, A. A., Gordon, M., Vellinga, M., Williams, A., Comer, R. E., Camp, J., Xavier, P. and Madec, G.: 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, 2015.

Mason, I.: A model for assessment of weather forecasts, Aust. Meteorol. Mag., 30, 291–303, http://www.bom.gov.au/jshess/docs/1982/mason.pdf, 1982.

Mason, S. J. and Graham, N. E.: Conditional Probabilities, Relative Operating Characteristics, and Relative Operating Levels, Weather Forecast., 14(5), 713–725, doi:10.1175/1520-0434(1999)014<0713:CPROCA>2.0.CO;2, 1999.

Matheson, J. E. and Winkler, R. L., Scoring Rules for Continuous Probability Distributions, Management Science, 22(10), 1087–1096, http://www.jstor.org/stable/2629907, 1976.

Met Office; Hollis, D., Carlisle, E., Kendon, M., Packman, S., and Doherty, A.: HadUK-Grid Gridded Climate Observations on a 1km grid over the UK, v1.3.0.ceda (1836-2023), NERC EDS Centre for Environmental Data Analysis [data set], doi:10.5285/b963ead70580451aa7455782224479d5, 2024a.

Met Office: https://www.metoffice.gov.uk/research/climate/seasonal-to-decadal/gpc-outlooks/user-guide/global-seasonal-forecasting-system-glosea6, last access 16 May 2024b.

Moore, R. J.: The PDM rainfall-runoff model, Hydrol. Earth Syst. Sc., 11(1), 483–499, doi:10.5194/hess-11-483-2007, 2007.

Müller, W. A., Appenzeller, C., Doblas-Reyes, F. J., and Liniger, M. A.: A Debiased Ranked Probability Skill Score to Evaluate Probabilistic Ensemble Forecasts with Small Ensemble Sizes, J. Climate, doi:10.1175/JCLI3361.1, 2005.

Murphy, A. H.: A Note on the Ranked Probability Score, J. Appl. Meteorol. Clim., 10(1), 155–156, doi:10.1175/1520-0450(1971)010<0155:ANOTRP>2.0.CO;2, 1971.

Murphy, S. J., Washington, R., Downing, T. E., Martin, R. V., Ziervogel, G., Preston, A., Todd, M., Butterfield, R., and Briden, J.: Seasonal Forecasting for Climate Hazards: Prospects and Responses, Nat. Hazards, 23(2), 171–196, doi:10.1023/A:1011160904414, 2001.

Portele, T. C., Lorenz, C., Dibrani, B., Laux, P., Bliefernicht, J., and Kunstmann, H.: Seasonal forecasts offer economic benefit for hydrological decision making in semi-arid regions, Scientific Reports, 11(1), 10581, doi:10.1038/s41598-021-89564-y, 2021.

Prudhomme, C., Hannaford, J., Harrigan, S., Boorman, D., Knight, J., Bell, V. A., Jackson, C., Svensson, C., Parry, S., Bachiller-Jareno, N., Davies, H. N., Davis, R., Mackay, J., McKenzie, A., Rudd, A., Smith, K., Bloomfield, J., Ward, R., and Jenkins, A.: Hydrological Outlook UK: an operational streamflow and groundwater level forecasting system at monthly to seasonal time scales, Hydrolog. Sci. J., 62(16), 2753–2768, doi:10.1080/02626667.2017.1395032, 2017.

Sabzipour, B., Arsenault, R., and Brissette, F.: Evaluation of the potential of using subsets of historical climatological data for ensemble streamflow prediction (ESP) forecasting, J. Hydrol., 595, 125656, doi:10.1016/j.jhydrol.2020.125656, 2021.

Scaife, A.A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R. T., Dunstone, N., Eade, R., Fereday, D., Folland, C. K., Gordon, M., Hermanson, L., Knight, J. R., Lea, D. J., MacLachlan, C., Maidens, A., Martin, M., Peterson, A. K., Smith, D., Vellinga, M., Wallace, E., Waters, J., and Williams, A.: Skillful long-range prediction of European and North American winters, Geophys. Res. Lett., 41(7), 2514–2519, doi:10.1002/2014GL059637, 2014.

Scaife, A.A. and Smith, D.: A signal-to-noise paradox in climate science, npj Climate and Atmospheric Science, 1, 28, doi:10.1038/s41612-018-0038-4, 2018.

Schepen, A. and Wang, Q.J.: Model averaging methods to merge operational statistical and dynamic seasonal streamflow forecasts in Australia, Water Resour. Res., 51(3), 1797–1812, doi:10.1002/2014WR016163, 2015.

Sheffield, J., Wood, E. F., Chaney, N., Guan, K., Sadri, S., Yuan, X., Olang, L., Amani, A., Ali, A., Demuth, S., and Ogallo, L.: A Drought Monitoring and Forecasting System for Sub-Sahara African Water Resources and Food Security, B. Am. Meteorol. Soc., 95(6), pp. 861–882, doi:10.1175/BAMS-D-12-00124.1, 2014.

Stringer, N., Knight, J., and Thornton, H.: 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, 2020.

Svensson, C.: Seasonal UK river flow forecasts based on persistence and historical analogy, in: Geophysical Research Abstracts, EGU General Assembly, Vienna, Austria, EGU2014-3868, 2014.

Svensson, C., Brookshaw, A., Scaife, A. A., Bell, V. A., Mackay, J. D., Jackson, C. R., Hannaford, J., Davies, H. N., Arribas, A., and Stanley, S.: Long-range forecasts of UK winter hydrology, Environ. Res. Lett. 10(6), 064006, doi:10.1088/1748-9326/10/6/064006, 2015.

Svensson, C.: Seasonal river flow forecasts for the United Kingdom using persistence and historical analogues, Hydrolog. Sci. J., 61(1), 19–35, doi:10.1080/02626667.2014.992788, 2016.

Swets, J. A.: The Relative Operating Characteristic in Psychology, Science, 182(4116), 990–1000, doi:10.1126/science.182.4116.990, 1973.

Troin, M., Arsenault, R., Wood, A. W., Brissette, F., and Martel, J.-L.: Generating ensemble streamflow forecasts: A review of methods and approaches over the past 40 years, Water Resour. Res., 57, e2020WR028392, doi:10.1029/2020WR028392, 2021.

van den Dool, H.: Empirical methods in short-term climate prediction, Oxford University Press, Oxford, UK, doi:10.1088/1748-9326/10/6/064006, 2007.

Vitart, F. and Robertson, A. W.: The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events, npj Climate and Atmospheric Science, 1(1), 1–7, doi:10.1038/s41612-018-0013-0, 2018.

Wilby, R. L., O'Hare, G., and Barnsley, N.: The North Atlantic Oscillation and British Isles climate variability, 1865-1996, Weather, 52(9), 266-296, doi:10.1002/j.1477-8696.1997.tb06323.x.s, 1997.

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, 3rd Edition, Academic Press, Oxford, 2011.

Weigel, A. P., Liniger, M. A. and Appenzeller, C.: The Discrete Brier and Ranked Probability Skill Scores, Mon. Weather Rev., 135(1), 118–124, doi:10.1175/MWR3280.1, 2007.

White, C. J., Franks, S. W. and McEvoy, D.: Using subseasonal-to-seasonal (S2S) extreme rainfall forecasts for extended-range flood prediction in Australia, Proceedings of IAHS, 370, 229–234, doi:10.5194/piahs-370-229-2015, 2015.

Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., Davis, P., Graham, T., Hewitt, H. T., Hill, R., Hyder, P., Ineson, S., Johns, T. C., Keen, A. B., Lee, R. W., Megann, A., Milton, S. F., Rae, J. G. L., Roberts, M. J., Scaife, A. A., Schiemann, R., Storkey, D., Thorpe, L., Watterson, I. G., Walters, D. N., West, A., Wood, R. A., Woollings, T., and Xavier, P. K.: The Met Office Global Coupled Model 3.0 and 3.1 (GC3.0 and GC3.1) Configurations, J. Adv. Model. Earth Sy., 10(2), 357–380, doi:10.1002/2017MS001115, 2018.

Yao, H. and Georgakakos, A.: Assessment of Folsom Lake response to historical and potential future climate scenarios: 2. Reservoir management, J. Hydrol., 249(1), 176–196, doi:10.1016/S0022-1694(01)00418-8, 2001.