

Improved seasonal hydrological forecasting for Great Britain

Mark D. Rhodes-Smith¹, Victoria A. Bell¹, Nicky Stringer², Helen Baron¹, Helen Davies¹ and Jeff Knight²

¹UK Centre for Ecology and Hydrology, Maclean Building, Benson Lane, Crowmarsh Gifford, Wallingford, Oxfordshire, OX10 8BB, UK

²Hadley 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, we combine newly-developed high-resolution rainfall forecasts derived from historical weather analogues (HWA) conditioned on large-scale circulation patterns with a monthly-resolution national-scale hydrological model to produce seasonal and monthly forecasts. This forecasting scheme is evaluated using hindcasts from 1993-2017. We show that the skill of the rainfall forecasts is comparable to that achieved at coarser resolutions by other models. However, increasing the spatial resolution has degraded some of the skill found by previous authors producing HWA forecasts at lower resolutions. The river flow forecasts perform better than rainfall, with much of the improvement attributable to the skill provided by hydrological initial conditions. Both rainfall and river flow forecasts are most skilful in winter and perform adequately in spring and autumn. We show that in responsive catchments, primarily found in the northwest, rainfall forecasts contribute most of the skill in the river flow predictions but are much less important in regions with long hydrological memory. We thus indicate where flow forecasts would benefit most from improved rainfall forecasts or from better hydrological modelling. The introduction of these high-resolution rainfall forecasts now enables hydrological forecasting at unprecedented levels of detail across Great Britain and demonstrates a forecasting approach that may be similarly beneficial elsewhere in the world.

Copyright statement

The works published in this journal are distributed under the Creative Commons Attribution 4.0 License. This licence does not affect the Crown copyright work, which is re-usable under the Open Government Licence (OGL). The Creative Commons Attribution 4.0 License and the OGL are interoperable and do not conflict with, reduce or limit each other.

© Crown copyright 2025

1 Introduction

Seasonal and sub-seasonal hydrological forecasts, here defined as those spanning the next three months and one month respectively, should 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). The performance of seasonal forecasts for the UK (in common with those for much of the extra-tropics) remains marginal (e.g. Johansson, 2007; Arribas et al., 2011; Kim et al., 2012; Vitart and Robertson, 2018; Quaglia et al., 2022; Nikrafter et al., 2024), but nevertheless these forecasts already underpin decisions to take low- and no-regret actions (those with minimal or only opportunity costs). These can include advertising to encourage reduced consumption when a seasonal forecast indicates a risk of water scarcity, or increasing the readiness of flood defences, relocating equipment and alerting staff when the forecasts indicate high river flows are probable. Improved forecast skill would support the use of management strategies with higher costs or longer lead times, such as altering reservoir stocks or activating regional water transfer schemes (e.g. Peñuela et al., 2020).

The UK national capability for seasonal hydrological forecasting is delivered through the UK Hydrological Outlook (<https://hydoutuk.net/>; Prudhomme et al., 2017; Boorman and Turner, 2019). It 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). Since then, the Hydrological Outlook has been issued monthly by a collaboration between the UK Centre for Ecology and Hydrology (UKCEH), the British Geological Survey (BGS), and the Met Office. Similar systems are also available in the USA (Demargne et al., 2014; Kirtman et al., 2014; Slater et al., 2016), Australia (Schepen and Wang, 2015), continental Europe (Alfieri et al., 2013), 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).

The UK Hydrological Outlook comprises estimates of the current hydrological conditions and forecasts of river flows and groundwater levels for the following one- and three-month periods. River flow forecasts are derived from several methods, including a gridded hydrological model driven by meteorological observations and rainfall forecasts (Bell et al., 2013, 2017), traditional ensemble streamflow predictions (ESP; e.g. Day, 1985; 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.

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). Rainfall forecasts are produced by the Met Office’s Global Seasonal forecasting system version 5 (GloSea5), an ensemble prediction system based on a HadGEM3 climate model (Scaife et al., 2014; MacLachlan et al., 2015; Williams et al., 2018). The initial conditions and rainfall forecasts are then combined in a simple monthly-resolution water balance model to derive estimates of river flows.

A limitation in the system demonstrated by Bell et al. (2013) was the lack of spatial resolution in the weather forecasts. Although the climate model had a resolution of 50km, only the UK-mean rainfall was provided operationally. Hydrological forecasting requires significantly finer resolution. The provided weather forecasts were therefore downscaled to the 1km resolution of the hydrological model by multiplying the national rainfall *anomaly* by the 1km standard average monthly rainfall. Although this method has been shown by Kay et al. (2023) to outperform assuming uniform rainfall across the country, it still 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.

More recently, Stringer et al. (2020) presented a new approach to produce high resolution seasonal rainfall forecasts, exploiting improvements in the prediction of the large-scale circulation patterns that drive UK rainfall (e.g. Scaife et al., 2014; Dunstone et al., 2016; Thornton et al., 2023). In this ‘Historic Weather Analogues’ (HWA) method, a forecast ensemble is constructed by combining observed rainfall patterns in years with similar circulation patterns to those predicted by GloSea. This allows the rainfall forecasts to be constructed at the native 1km resolution of the G2G/WBM hydrological models, potentially offering improved performance at these resolutions. Stringer et al. (2020) assessed the correlation of the ensemble mean of winter (December, January, February) total rainfall forecasts over northern Europe and showed that the HWA forecasts correlate better with observations than the GloSea5 forecasts from which they are derived.

In this work, we evaluate the use of HWA rainfall forecasts in the G2G/WBM hydrological forecasting scheme across the entire year. This scheme has been used operationally within the UK Hydrological Outlook since December 2023. An illustration of the forecasts over this period is provided in the Supplementary Material. This is the first application of these HWA rainfall forecasts to gridded and national-scale seasonal hydrological forecasting in Great Britain. At catchment scales, Donegan et al. (2021) have compared the performance of traditional ‘historical ESP’ with HWA-derived rainfall forecasts (referred to as ‘conditioned 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. They

therefore recommended that ‘conditioned ESP’ (HWA) is made operational in Ireland. Similar work is now in progress to
95 apply HWA forecasts to catchment models in Great Britain (Chan et al., 2025).

This paper is structured as follows. In Sect. 2 we describe our hydrological forecasting scheme and the set of hindcasts and
performance scores used to assess its performance. In Sect. 3 we examine the performance of the HWA rainfall forecasts,
extending the earlier assessment of Stringer et al. (2020) beyond the winter season. We then assess the performance of river
100 flow forecasts driven by the rainfall forecasts in Sect. 4, concluding with a comparison to the earlier forecasting scheme by
Bell et al. (2017). In Sect. 5 we explore the relative contributions of rainfall forecasts and hydrological initial conditions to
the skill of river flow forecasts across the UK. We conclude in Sect. 6.

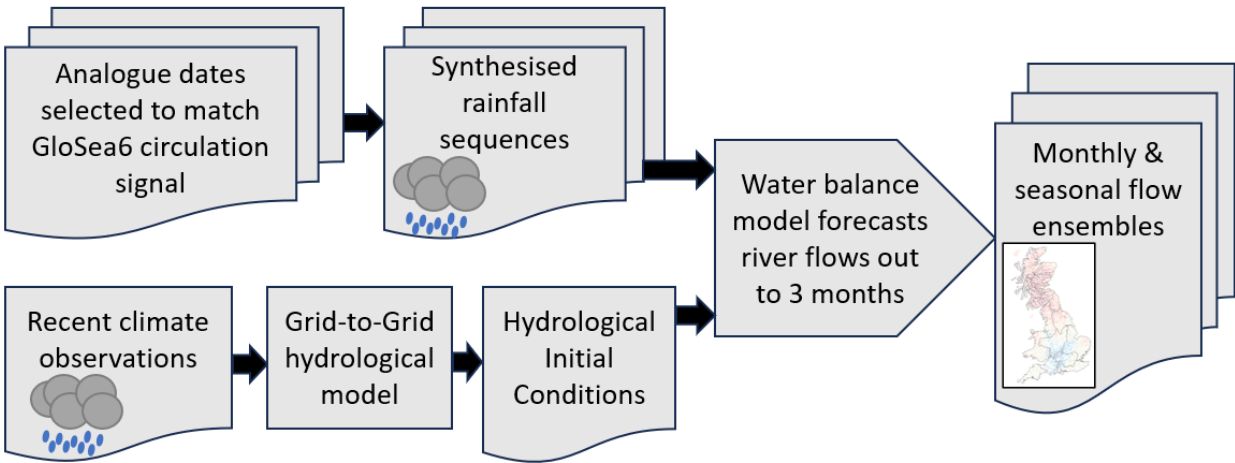
2 Methods

2.1 Forecasting River Flows

105 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.

110 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. We refer below to ‘seasonal’ and ‘monthly’ forecasts as those spanning the next three months and one month respectively.



115 **Figure 1: Schematic of the G2G/WBM hydrological forecasting scheme, here driven by rainfall forecasts derived from Historic Weather Analogues. An extended version of this figure with greater detail on the rainfall forecasting method is in the Supplementary Material.**

2.1.1 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
120 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-Grid¹, and potential evaporation at 40km resolution produced using the Met Office Rainfall and
125 Evaporation Calculation System v2.0 (MORECS; Hough and Jones, 1997), the latter uniformly downscaled to the 1km model resolution and both uniformly disaggregated from daily totals to the 15-minute resolution of the model. Although simplistic, adopting this uniform temporal downscaling will be sufficient for monthly and seasonal forecasting; more sophisticated approaches are important for nearer-term forecasts. 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
130 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.

2.1.2 Historic Weather Analogues rainfall forecasts

135 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 HWA forecasting approach sits in the centre ground. Much like the use of raw model outputs (and unlike historical ESP), such forecasts are sensitive to contemporary atmospheric conditions. However, like
140 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.

The HWA approach taken in this work is an extension of that presented in Stringer et al. (2020) and Donegan et al. (2021). A detailed description of this method is given in the Supplementary Material; here we outline the key steps. We first construct
145 a c.40-member lagged ensemble of GloSea6 (Kettleborough et al., 2026) Mean Sea Level Pressure (MSLP) forecasts covering the North Atlantic and Europe from runs initialised over the preceding days. A set of synthetic MSLP observations, constructed by combining observations from three ‘analogue’ years, are also produced. For each GloSea6 ensemble member,

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

we then rank synthetic observations by their root mean squared difference to the member’s forecast MSLP field and select the best ten to become members of our HWA forecast ensemble. The final HWA rainfall forecasts are then made from the
150 HadUK-Grid (Met Office, 2024) observed rainfall in these analogue years, with a linear correction for the effects of climate change applied.

However, for seasonal forecasts from November-December-January to February-March-April (inclusive), we first amplify the North Atlantic Oscillation (NAO) signal in the GloSea6 MSLP ensembles. This is to correct for the ‘signal to noise
155 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 approach taken to construct the set of synthetic seasons is modified to emphasise this signal by only choosing analogue years with a similar (within 4hPa) monthly-mean NAO indices of that in the adjusted GloSea6 ensemble member (Stringer et al., 2020). These synthetic seasons are then ranked by the root mean squared difference in their residual MSLP fields, the best ten selected, and rainfall forecasts constructed in the same manner
160 as in other seasons.

2.1.3 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
165 precipitation, and outputs from evaporation and outflows. The store at the end of month $m + 1$ is thus given by

$$S_{m+1} = S_m + P_{m+1} - E_{m+1} - Q_{m+1} \quad (2)$$

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 $m + 1$.

170 For the UK Hydrological Outlook implementation of the WBM, we derive the initial store S_m from the G2G run described in Sect. 2.1.1 and the forecast precipitation as described in Sect. 2.1.2. 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). We note that the model estimates natural flows with anthropogenic influences currently neglected; in some catchments in Great Britain these can be
175 significant (Rameshwaran et al., 2022).

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 seasonal forecasting, the first month uses the G2G hydrological initial conditions described in Sect. 2.1.1, with each subsequent
180 month adopting the forecast stores from the water balance model for the preceding month.

2.1.4 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 seasonal forecasts, after calculating anomalies for each month separately, the three months are averaged to produce a seasonal 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 (Fig. 2). The threshold river flow anomaly is calculated for each cell for each month, and for seasonal forecasts for each combination of three sequential months. This classification scheme expresses the forecast as a deviation from the ‘normal’ modelled conditions for the time of year, with ‘exceptionally high/low’ defined as exceeding 20-year return periods. Since the classes are defined as percentiles of the modelled climatology, this approach incorporates a first-order bias correction. At gauged locations bias-corrected flow forecasts can be derived using the known distribution of observed flows over the period of record, but this is not currently done operationally.

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		$P < 0.05$

Figure 2: Classes used to categorise river flow forecasts relative to their historical distributions. Each class has a name and characteristic colour.

2.2 Performance Assessment methods

2.2.1 Hindcast sample

To assess the performance of both the HWA rainfall forecasts and the G2G/WBM river flow forecasts, we use a sample of hindcasts (re-forecasts of past cases). These hindcasts emulate the seasonal forecasts that would have been produced at the start of each season (DJF: Winter, MAM: Spring, JJA: Summer, SON: Autumn) and monthly forecasts that would have been produced at the start of each month. They are generated automatically within the operational GloSea6 forecasting system,

205 which runs an array of hindcasts starting from historic initial atmospheric conditions on the 1st, 9th, 17th, and 25th of each month (MacLachlan et al., 2015; Kettleborough et al., 2026). From these we select the hindcasts that are generated at a similar lead time to the operational system (see Supplementary Material, Sect. S1). For the seasonal hindcasts, these are the 17 GloSea6 hindcast members initialised on each of the 1st, 9th and 17th of the month preceding the forecast. For the monthly hindcasts, we use 7 GloSea6 members from each of the 9th and 17th. Monthly hindcasts are available from February 1993 to 210 January 2017 and seasonal hindcasts from Spring 1993 to Winter 2016/17. This is the current international standard period used to verify seasonal forecasting systems participating in the Copernicus Climate Change Service², which includes the Met Office’s GloSea6. The hindcast samples’ properties are summarised in Table 1.

Using these GloSea6 hindcasts, we generate seasonal (monthly) total rainfall ensembles as described in Sect. 2.1.2. From the 215 rainfall ensembles and the associated hydrological initial conditions (Sect. 2.1.1) we derive seasonal (monthly) mean river flow ensembles as described in Sect 2.1.3.

Table 1: Availability of hindcasts for the HWA rainfall forecasts.

Forecast period	First hindcast	Last hindcast	Hindcast count ^a	Ensemble size ^b
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.				

2.2.2 Ground truths

220 The hindcasts described in Sect. 2.2.1 are to be compared with observations of the seasonal (monthly) total rainfall and mean river flows in the corresponding period. Ideally, one would compare with observations taken at the resolution of the dataset; in practice observations are not available everywhere we would wish. In this study we choose to prioritise coverage rather than restrict ourselves to only use observations taken from weather observing sites and gauging stations, and thus for both rainfall and river flows we have used ‘ground truth’ datasets that include an element of modelling to achieve the spatial 225 completeness we require.

For the rainfall ‘ground truth’, we use the 1km HadUK-Grid dataset, which relies on interpolation to infill between observing sites, accounting for variations in topography and other local effects (Hollis et al., 2019; Met Office, 2024). For river flows, we estimate ‘ground truth’ flows at 1km resolution across Great Britain using the G2G model, HadUK-Grid and MORECS 230 data as described in Sect. 2.1.1. Bell et al. (2009) verified G2G daily flows against National River Flow Archive

² <https://climate.copernicus.eu/seasonal-forecasts>

observations and found that it achieves a median Nash-Sutcliffe efficiency (Nash and Sutcliffe, 1970) of ≈ 0.7 . Subsequent works have found similar performance across the flow regime, with no systematic variations across different catchment types (Bell et al., 2012; Bell et al., 2016; Rudd et al., 2017; Formetta et al., 2018; Kay et al. 2021; Rameshwaran et al., 2022).

2.2.3 Skill scores

In this work we use three different skill scores to measure the performance of different aspects of our probabilistic forecasts: the Pearson correlation, the (Continuous) Ranked Probability Skill Score and the Relative Operating Characteristic. Our principal goal in this work is to understand the quality of the river flow forecasts, so to these we apply all three metrics. The Pearson correlation alone suffices to indicate where and when HWA forecasts contribute skill.

The Pearson correlation (PCorr) measures the linear correlation between a single forecast and the observed value of a predictand. It is thus commonly used for deterministic forecasts but is also suitable for measuring the performance of the ensemble mean of a probabilistic forecast, which is often interpreted as representing the ability of the forecast to recover the ‘predictable component’ (the signal) of the predictand (e.g. Weisheimer et al., 2024).

The distribution of ensemble members is a measure of the modelled uncertainty in the forecast, and thus our confidence in its predictions. A performance statistic that compares a forecast distribution to an observation which penalises over- or under-confident forecasts is thus also useful. A suitable choice is the Continuous Ranked Probability Score (CRPS; e.g. Brown, 1974; Matheson and Winkler, 1976; Hersbach, 2000; Wilks, 2011), which is defined as

$$CRPS = \int_{-\infty}^{\infty} (CDF(x) - H(x, o))^2 dx \quad (3)$$

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.

$$H(x, o) = \begin{cases} 0 & x < o \\ 1 & x \geq o \end{cases} \quad (4)$$

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 that 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
 265 Probability *Skill* Score (CRPSS; e.g. Wilks, 2011), defined as

$$CRPSS = 1 - \frac{\langle CRPS_{\text{forecast}} \rangle}{\langle CRPS_{\text{reference}} \rangle} \quad (5)$$

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. This is a 55-member ensemble consisting of the WBM estimates of river flows obtained using the observed 1km HadUK-
 270 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
 275 Outlook. This is because we categorise flows into the 7 classes given in Fig. 2. 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
 280 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).

Finally, we choose a performance statistic that assesses the ability of the forecasting system to predict extreme events, the Relative Operating Characteristic (ROC; e.g. Swets, 1973; Mason, 1982; Mason and Graham, 1999). The ROC assessment
 285 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
 290 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 match our other performance scores, such that

$$S_{\text{ROC}} = 2AUC - 1 \quad (6)$$

where $S_{\text{ROC}} = 1$ indicates a perfect forecast (detects all events with no false positives at all probability thresholds) and
 295 $S_{\text{ROC}} > 0$ 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 river flow events. Specifically, we define ‘high’ flow events as those falling in the ‘above normal’ classification or above (classes 5, 6 & 7), and ‘low’ events as those falling in the ‘below normal’ classification or below (classes 1, 2 & 3), as defined in Fig. 2. This is a very conservative definition of ‘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.

3 Performance Assessment – Rainfall Forecasts

3.1 Results

Figure 3 shows spatial maps of the Pearson correlation between the ensemble means of seasonal- (monthly-) total rainfall hindcasts and HadUK-Grid observations at 1km resolution across Great Britain. The spatially-averaged correlation is shown in the top right of each panel as a simple indication of the relative performance in different seasons (months). However, it is the spatial variations that we consider to be of greatest importance, since these will determine over which river catchments the rainfall forecasts can offer skill. Moreover, as we will examine in Sect. 5., rivers vary in their sensitivity to recent rainfall, and thus the performance of the rainfall forecasts over regions of high sensitivity matters more than over regions of low sensitivity.

The clearest patterns in skill are apparent in the seasonal forecasts, where better performance is seen in winter (spatially averaged PCorr=0.18, 49% of pixels with PCorr>0.2) and spring (0.17; 43%) than in summer (-0.23; 0%) or autumn (0.06; 15%). In winter, there is a spatial pattern of higher correlations along the west coast and in northern Scotland compared to other parts of the country, while in spring the skill is less spatially coherent, with higher skill persisting in northern Scotland, but now also being seen in parts of northern England and in East Anglia. However, in only a few areas can these correlations be determined to be statistically significant (PCorr>0.33 at 95% confidence; Fisher and Yates, 1963). Correlations in monthly forecast skill (Fig. 3) are much more spatially and temporally variable than for seasonal forecasts, and do not systematically follow the same patterns.

3.2 Discussion

Forecasting seasonal precipitation in the extra-tropics remains extremely challenging, particularly at the high spatial resolutions demanded by hydrological forecasts. Recent studies have shown that the GloSea model used in this work is currently among the best performing models at predicting European seasonal precipitation (e.g. Quaglia et al., 2022; Nikraftar et al., 2024). However, even for the best models, correlations over the UK remain weak (typically PCorr≈0.3; Baker et al., 2017, Quaglia et al., 2022). Other models which perform well in the tropics, such as the European Centre for

Medium Range Weather Forecasting's SEAS5 model, show little correlation over almost all of the UK in winter and summer (PCorr<0.2; Johnson et al., 2019). Adopting the HWA approach in winter achieves PCorr≈0.4, a small improvement on the raw skill from GloSea (Stringer et al., 2020). However, all the quoted results are at much coarser resolutions (0.5-1°≈25-50km) than is required by hydrological models. At higher resolutions some of the skill will be lost as less noise is averaged out. This is demonstrated by comparing the top-left panel of our Fig. 3 with the corresponding Fig. 9 of Stringer et al. (2020), which show exactly the same hindcasts at 1km (spatially-averaged PCorr=0.18) and at 25km (≈0.4) resolution. Reaching comparable skill at high resolutions will be necessary to fully exploit operational seasonal forecasts. Although beyond the scope of this work, additional post-processing is often used to improve seasonal rainfall forecasts elsewhere in the world and may prove useful here also (e.g. Kharin et al., 2017; Zhao et al., 2017; Shepen et al. 2018; Wang et al., 2019).

The correlations we present in Fig. 3 indicate that the performance of our HWA rainfall forecasts varies spatially as well as temporally. The best performance is visible in the winter and spring maps, where the spatial patterns correspond to the influence of the NAO upon UK rainfall. The strongest NAO influence is in December and January in northern and western parts, before it declines in later months (West et al., 2021). The widespread negative correlations in summer indicate that in this season the HWA forecasts are misleading. This results from the inability of GloSea to accurately predict summer MSLP (Lockwood et al. 2023) which, consistent most seasonal prediction systems, struggles to capture summertime teleconnections (Knight and Scaife, 2024). However, in summer the influence of the NAO on UK rainfall becomes strong again (although with opposite sign to the winter; West et al., 2021), suggesting that the HWA method might remain appropriate if only it were conditioned on more skilful predictions.

The low spatially-averaged skill of these forecasts suggest that overall the 1km HWA forecasts presented here offer only small improvements over climatology (which would have PCorr=0). This explains the results of Chen et al. (2025), who found that while HWA forecasts often drove slightly more skilful catchment predictions than ESP, the latter remained difficult to beat. However, in spite of the low spatially-averaged skill, some parts of the country show substantially better performance than others, such as the north and west of Great Britain. As these are the regions where rivers are most sensitive to recent rainfall (Svensson et al., 2015), hydrological forecasts will derive more value from rainfall forecasting skill here than in less sensitive regions.

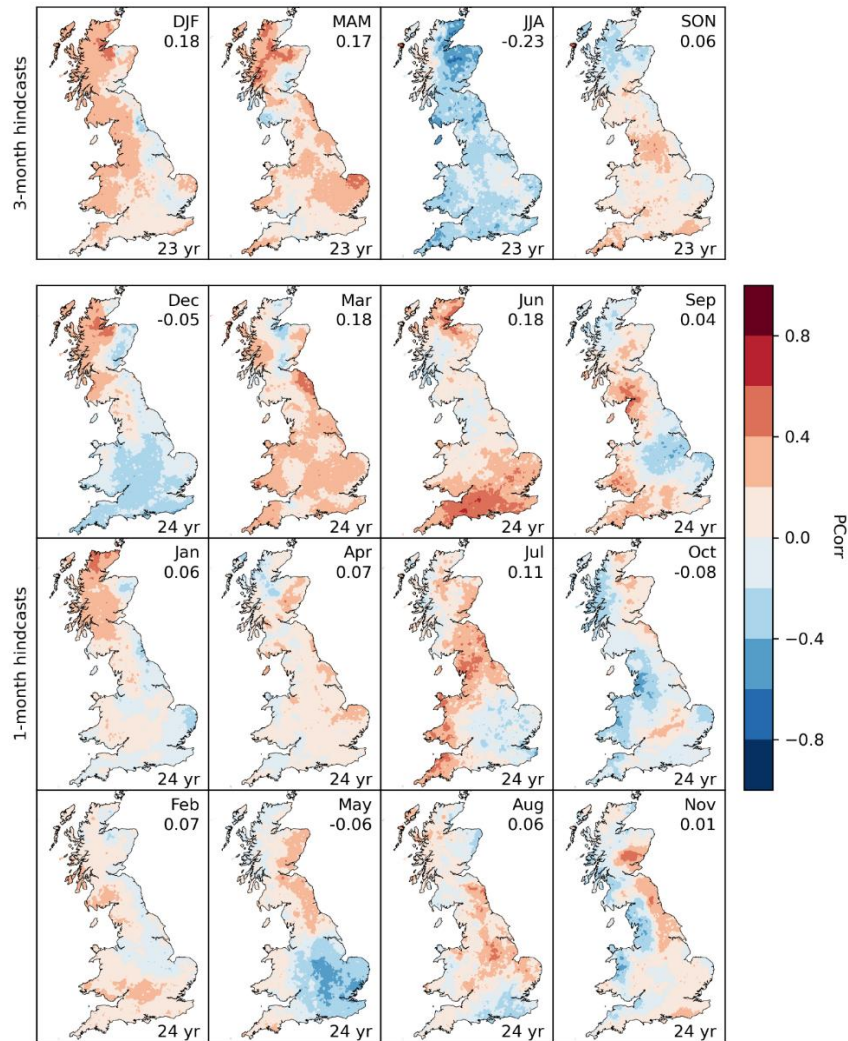


Figure 3: Pearson correlation between HWA forecast ensemble mean and observed total rainfall for 3-month (top row) and 1-month (bottom three rows) hindcasts. A positive correlation is indicated in red while a negative correlation is in blue. In each panel the number in the top-right corner is the spatial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

4 Performance Assessment – River Flow Forecasts

4.1 Results

Figure 4 shows the Pearson correlation between hindcast ensemble means and ‘ground truth’ (simulated) river flows, 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, and in the other figures in this section, we only show pixels with non-negligible annual mean runoff ($>0.05 \text{ m}^3\text{s}^{-1}$) which excludes highly-volatile upper reaches. Seasonal forecasting performance

remains best in winter (spatially-averaged PCorr=0.48; 90% of river pixels with PCorr>0.2), has similar skill in both spring (0.38; 80%) and autumn (0.36; 75%), but shows a substantial drop in summer (0.08; 31%) consistent with our earlier findings for rainfall (Sect. 3). Even in summer, the strongest correlations (≈ 0.4 -0.7) are found across the chalk aquifers in South-East England. Monthly forecasts show spatial patterns that are similar to the seasonal forecasts, but there is significant variability in their spatially-averaged correlations over the year.

Figure 5a shows the CRPSS for the G2G/WBM river flow forecasting scheme, while Figure 5b shows the RPSS after flows are classified according to the scheme in Fig. 2. In addition to the spatial and seasonal variability observed for the ensemble means, the skill indicated for the classified forecasts systematically improves upon that for the actual river flows, with the most notable improvements across the chalk aquifers. This improvement is expected, as our classification scheme has the effect of both bias-correcting and binning forecasts. In the chalk, the uncertainty in the forecasts can often be comparable to, or smaller than, the classes such that the forecast consistently selects the observed class with high confidence.

Figure 6 shows the S_{ROC} for the performance of the G2G/WBM river flow forecasting scheme at detecting above and below normal flows. 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 naïvely expect that it is easier to detect a ‘wet’ summer or a ‘dry’ winter, our events are defined relative to the climatological distribution of flows for the forecast period. We are thus showing the performance of the forecasting scheme at predicting a wetter-than-normal (or drier-than-normal) month/season, and the plots in Fig. 6 can be compared to one another fairly.

4.2 Discussion

Overall, the river flow forecasts perform better than the rainfall forecasts on which they were based. This is because flows in many catchments have long hydrological memory, meaning that they are less sensitive to recent rainfall. The effects of this memory are captured in our hydrological initial conditions, which contribute a significant proportion of the forecast skill for many rivers (Bell et al., 2017). A particularly clear example of this is in the southern chalk where the deep aquifer is the dominant influence on river flows. By contrast, hydrological forecasts perform less well in northern England, a region where hydrological memory is short, and where rainfall forecasts do not provide sufficient skill to underpin reliable flow forecasts.

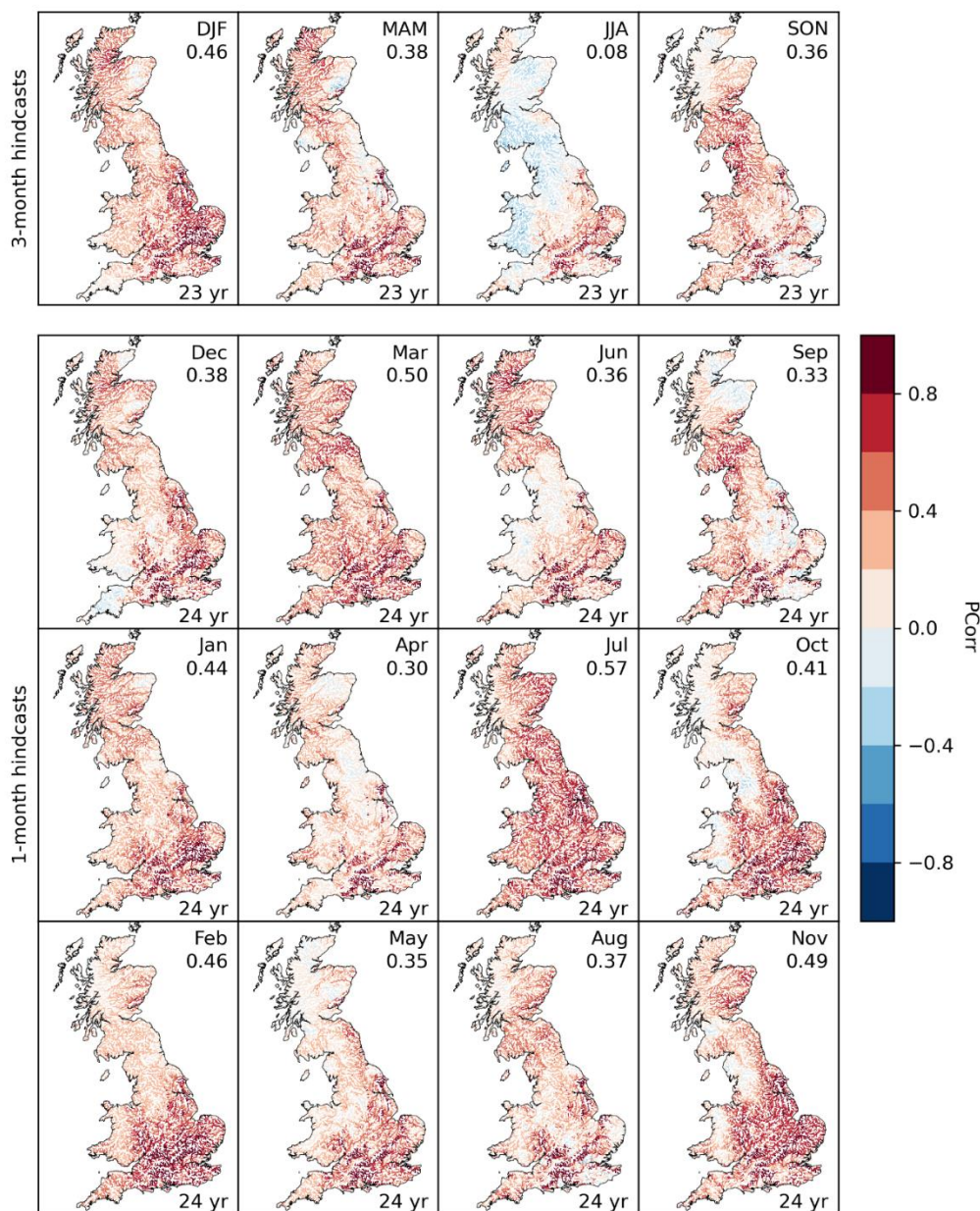
The seasonal variations in the spatially averaged skill scores (top-right of each panel of Fig. 4-6) are consistent with the patterns for the rainfall forecasts in Fig. 3, with better relative performance in the same seasons. The skill maps show high spatial variability arising from both varying meteorological forecast skill and the differing responses of different catchments to rainfall across Great Britain. 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.

400 We also find that the forecast skill is not uniform across the flow regime. Fig. 6 demonstrates that the forecasting scheme does not always have the same ability to detect high flows as it does low (and vice versa). An example of this can be seen by comparing the plots for July, where the scheme has significantly higher skill at detecting above-normal flows (spatially-averaged $S_{ROC}=0.46$) than below (0.18). High and low flows arise from non-linear interactions between hydrological storage and rainfall. These non-linear interactions mean that any forecasting system is likely to exhibit such variations in
405 performance over the flow regime.

4.3 Comparison with previous GloSea5-derived seasonal forecasts

Bell et al. (2017) assessed the skill of the previous version of the G2G/WBM hydrological forecasting scheme, which differed from that presented here in that it was driven by spatially-uniform GloSea5 rainfall forecasts. This version was used in the Hydrological Outlook prior to December 2023. Bell et al. (2017) evaluated the Pearson correlation and the S_{ROC} score
410 for regional-mean river flow forecasts, which we can compare to our results (shown in Fig. 4 and 6). However, they only had a limited selection of 13 years of monthly (March, June, September, December) and seasonal (spring, summer, autumn, winter) hindcasts. Bell et al. (2017) combined Autumn-Winter and Spring-Summer hindcasts to obtain a sufficiently large sample to determine the correlation coefficient. This obscures some of the systematic variations with season found in our work.

415 In Autumn-Winter, Bell et al. (2017) found good correlations ($PCorr\approx0.4-0.7$) in almost all regions for both monthly and seasonal forecasts, with similar spatial patterns to ours. Our Fig. 4 shows that the winter performance is better than autumn, but the correlations remain strong over almost the entire country in both seasons. Their work also noted poor ($PCorr\approx0.0-0.3$) Spring/Summer performance across much of the country. Since Bell et al. (2017)'s forecasts were also based on rainfall
420 outputs from GloSea, it is probable the same lack of summer meteorological skill is present (our Fig. 3 and Lockwood et al., 2023).



425 **Figure 4: Pearson correlation between G2G/WBM forecast ensemble mean and 'ground truth' river flows 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.**

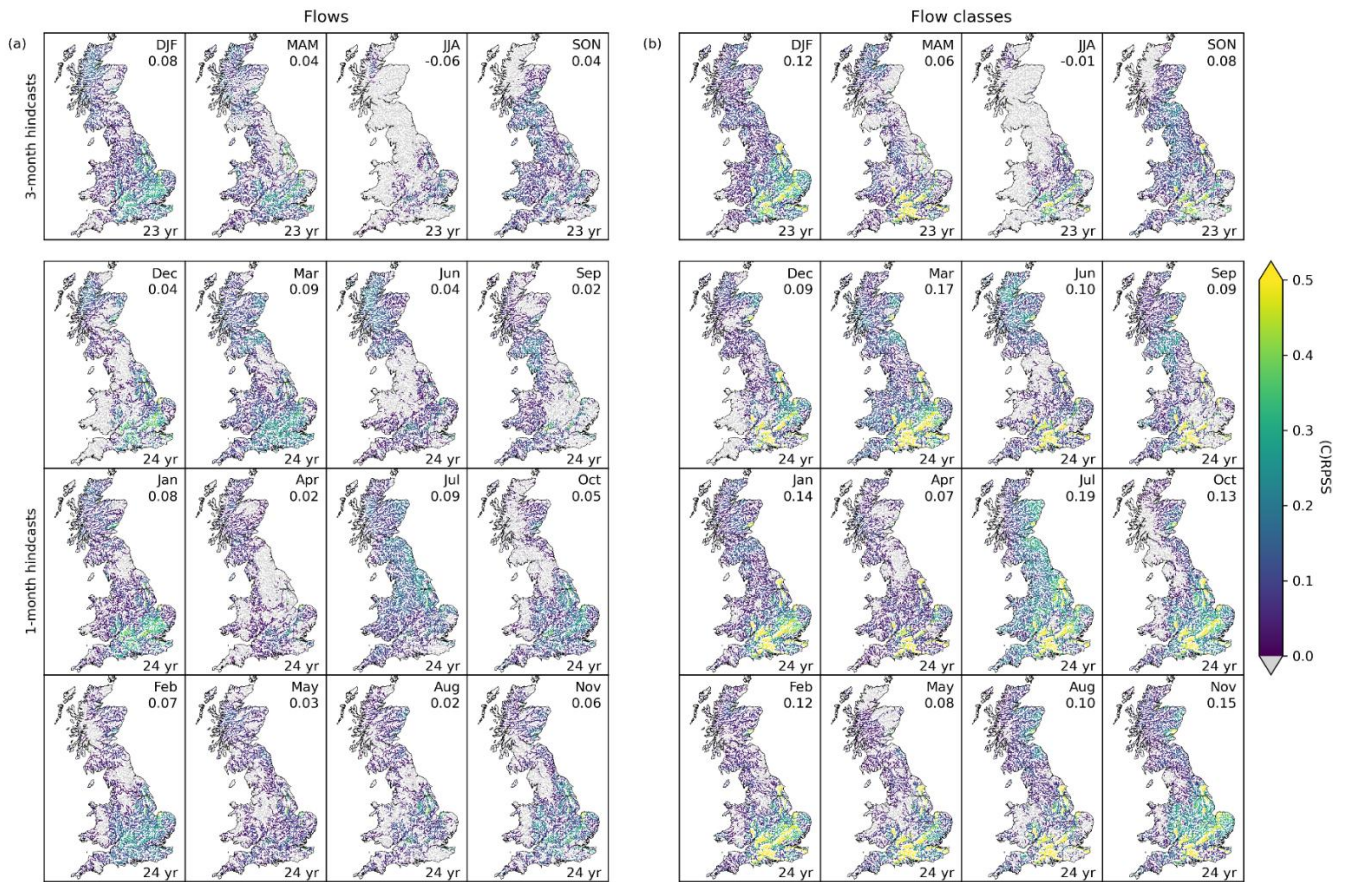


Figure 5: The skill of the G2G/WBM river flow forecasting scheme over assuming the climatological distribution. Left: CRPSS for predicting volumetric river flows (m³/s). Right: RPSS for predicting classified flows (Fig. 2). Scores greater than 0 (coloured) indicate the G2G/WBM scheme outperforms the 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.

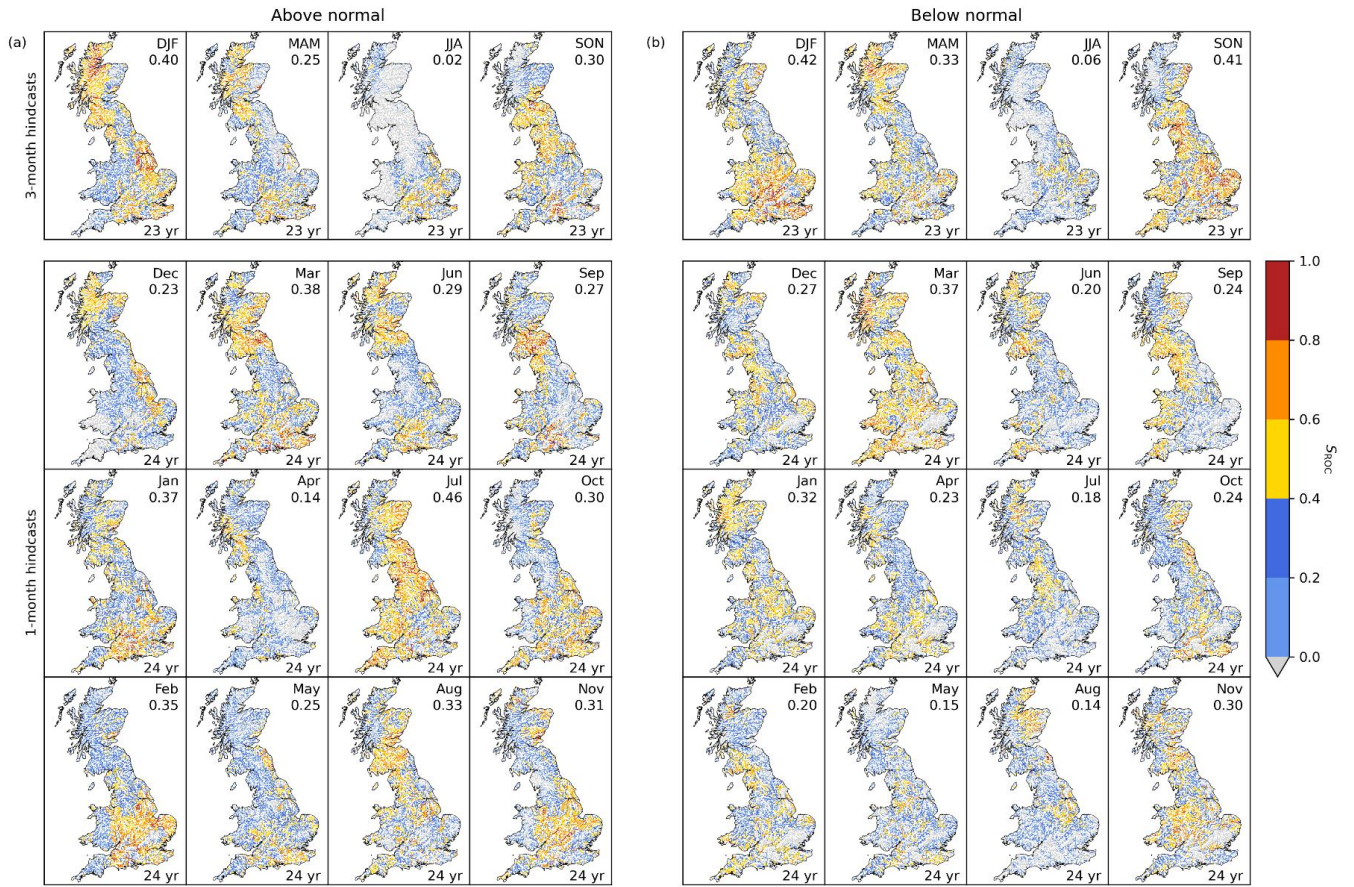


Figure 6: S_{ROC} for the G2G/WBM river flow forecasting scheme. The left array of panels (a) shows the performance of the model at forecasting above normal flows (>72 nd percentile) and the right array (b) shows the performance for below normal flows (<28 th percentile). Colours indicate the rescaled 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.

Bell et al. (2017) also performed an S_{ROC} analysis based on the same classifications of above normal and below normal flows as in this work. Their results were aggregated spatially and temporally, and only the average S_{ROC} score for above- and below-normal flows presented and so their results are not directly comparable to ours. Nevertheless, we can make an approximate comparison between the GB-mean S_{ROC} skill scores of our Fig. 6 and those given by Bell et al. (2017); these scores are shown in Table 2. Our 1-month GB-mean S_{ROC} for December is now positive, compared to Bell et al. (2017)'s negative score (indicating there is now some skill). The Spring GB-mean skill score has more than doubled. These examples suggest that the introduction of HWA forecasts offers some improvement over the previous method, particularly in Winter and Spring. In these seasons, the correlations between large-scale circulation patterns and UK rainfall are strongest (Stringer

et al. 2020, West et al., 2021) so the HWA forecasts (and particularly the amplification of the NAO signal) are expected to offer the greatest improvement at this time of year.

450

Table 2: 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

5. Sources of forecast skill

455 The varied hydrogeology of Great Britain drives spatially-varying sensitivity to recent rainfall and hydrological memory. Skilful river flow forecasts thus require rainfall forecasts that perform well in the regions where the rainfall is most influential. Fortuitously, the north and west of Great Britain is where both the highest sensitivity to rainfall and best performing forecasts are found (Svensson et al., 2015). Here, we are now able to explore this phenomenon for ungauged catchments across Great Britain. To do so, we have run the WBM hindcasts using three different combinations of inputs:

- 460
- (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.
 - 465 (c) HWA rainfall forecasts with G2G-derived HICs, the forecasting scheme presented in the preceding sections which combines *both rainfall and HIC forecast skill*.

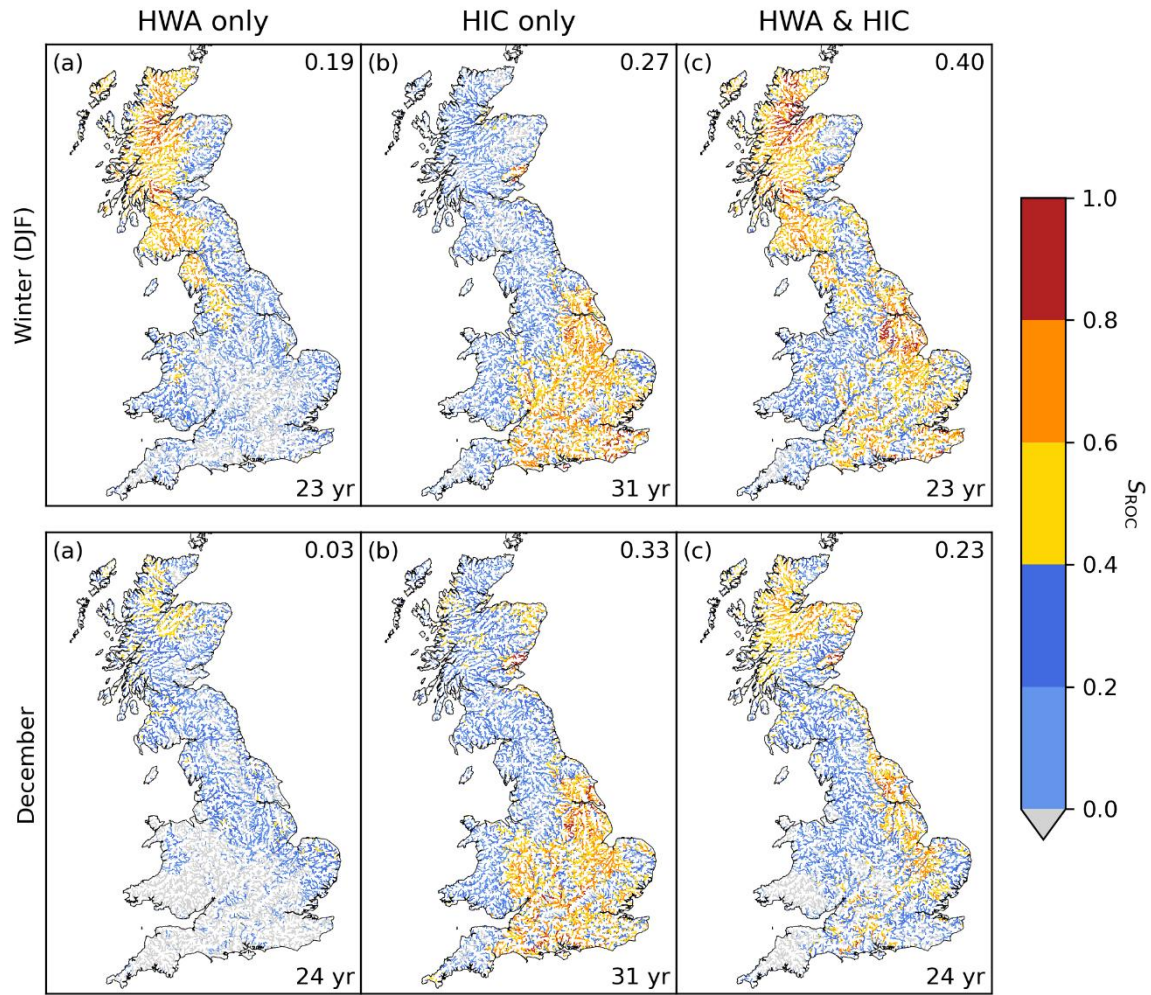


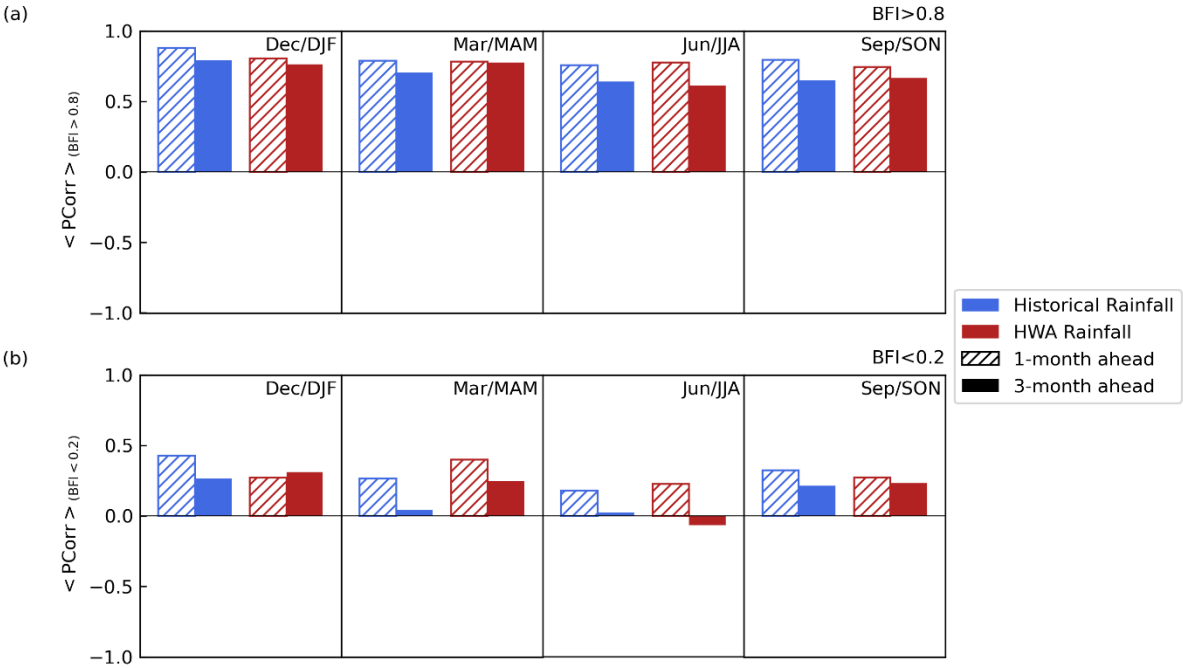
Figure 7: S_{ROC} for above normal river flows in Winter (top) and December (bottom) using different combinations of inputs to the WBM to highlight the contributions to forecast skill from (a): HWA rainfall forecasts (b): G2G HICs, and (c): the combination of both. Scores greater than 0 (coloured) indicate 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 7 shows the S_{ROC} score for forecasting above-normal flows across the winter (seasonal-mean flows) and for 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 winter rainfall while flow persistence follows from hydrological memory, these patterns mirror those found in this work. In our Fig.

480 7 they are revealed for the first time at national scale and for ungauged sites, revealing the pattern extends across the entire 1km river network.

The lower row of Fig. 7 shows the same skill scores for above-normal monthly-mean flows in December. Here, there is greater spatial complexity to the contributions of HICs and rainfall forecasts. For instance, in western Scotland neither rainfall forecasts nor HICs provide particularly high skill; 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 HWA rainfall forecasts degrades the performance we would otherwise expect to be provided by HICs.

490 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 the length of hydrological memory from the baseflow index (BFI; Gustard et al., 1992), which estimates the fraction of river flows that originate in subsurface stores. Here we use the estimates of BFI derived from the hydrological properties of soils derived by Boorman et al. (1995).



495 **Figure 8: 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 8 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 high

500 performance of the forecast ensemble mean in regions where flows are dominated by subsurface stores (Fig. 8a), with only modest reductions in skill between monthly and seasonal forecasts. This suggests that hydrological memory is longer than three months in regions with high BFI. In addition, we observe that there is very little variation in the strength of the correlations (and thus forecast skill) over the year; this is because hydrogeology does not change seasonally.

505 By contrast, the low-BFI sample (Fig. 8b) includes 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 seasonal flow forecasts improved by HWA rainfall in winter, spring and autumn (see Fig. 3). In most cases, the correlation for a monthly forecast is stronger than that for the seasonal forecast that begins in that month, as the predictability of the rainfall declines with increasing lead time. The exception to this trend is the December/Winter skill, where rainfall
510 predictability for both low and high BFI catchments is enhanced by the large-scale circulation patterns that correlate well over the later months of winter (West et al., 2021).

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
515 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 HWA forecasting approach offers the potential to do this by conditioning the ensemble on meteorological variables that correlate best with precipitation in these areas.

520 6. Conclusions

In this paper we have presented and assessed the performance of a new configuration of 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 high-resolution historical weather analogues (HWA) conditioned on the large-scale circulation patterns predictions produced by GloSea6. This approach replaces an earlier scheme in which raw rainfall forecasts were spatially
525 downscaled to the resolution of the hydrological model. We show that the HWA rainfall forecasts offer at higher spatial resolutions, skill that is comparable to that achieved over the UK by other models, but that some skill is lost compared to lower-resolution HWA forecasts. This results in forecasts that are, on average, marginally skilful in some seasons and regions. The best performance seen in winter in the north-west of Great Britain, attributable to the strong influence of the North Atlantic Oscillation. By contrast, in summer the rainfall forecasts are unreliable due to a lack of skill in the underlying
530 GloSea6 forecasts.

We also examine the skill of the resulting river flow forecasts. We find that the G2G/WBM forecasting scheme performs well across much of Great Britain for most months and seasons, notwithstanding the marginal skill in the HWA rainfall forecasts. We compare our results to those found by Bell et al. (2017) for an earlier configuration of the forecasting system and conclude that HWA-based forecasts better predict the likelihood of low- and high-flow events in three of the four seasons. A small reduction in the skill score is found in autumn.

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 the north and west being most sensitive and the south-east being least. The work here has demonstrated this phenomenon at high resolution for ungauged catchments over the whole of Great Britain, extending the small sample of gauged catchments originally used by Svensson et al. (2015). These findings allow future improvements to rainfall forecasts and hydrological models to be targeted at the regions where they will have the most impact.

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). The hindcasts cover the period 1993-2017 and thus do not include more recent hydro-climatic extreme events and continuing changes in our climate. The argument could be made that the current operational performance may be different to that evaluated over the hindcast period. In the Supplementary Material (Section S2) we briefly compare the performance of the operational scheme to that of the hindcasts and demonstrate that there is minimal evidence of a difference in performance.

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). Across the UK Hydrological Outlook, there is a desire to make best use of the multiple available models by giving varying weight to different forecasting methods in proportion to their (seasonally- and spatially-varying) skill. Currently this is done subjectively by expert forecasters, but research is also underway into bias-correcting and statistically blending forecasts to better exploit each model's strengths. There are similar ambitions to explore multi-model ensembles to support seasonal weather forecasting in the UK.

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

The GloSea6 MSLP hindcasts upon which this work is based are available via the Copernicus Climate Change Service's Climate Data Store (doi:[10.24381/cds.68dd14c3](https://doi.org/10.24381/cds.68dd14c3)). The analogue years used to construct the HWA forecasts and the final grids of the forecast performance scores at 1km resolution will be made available on reasonable requests to the corresponding author.

Author Contribution

MR-S: Methodology, Software, Formal Analysis, Investigation, Writing, Visualisations; **VB:** Methodology, Writing, Supervision; **NS:** Methodology, Investigation, Writing; **HB:** Software; **HD:** Methodology, Data curation; **JK:** Conceptualisation, Methodology, Writing

Competing interests

The authors declare that they have no conflicts 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.

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

- Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., and Pappenberger, F: GloFAS – global ensemble streamflow forecasting and flood early warning, *Hydrol. Earth Sys. Sci.*, 17(3), 1161–1175, [doi:10.5194/hess-17-1161-2013](https://doi.org/10.5194/hess-17-1161-2013), 2013.
- 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](https://doi.org/10.1002/2015WR017864), 2016.
- Arribas, A., Glover, M., Maidens, A., Peterson, K., Gordon, M., MacLachlan, C., Graham, R., Fereday, D., Camp, J., Scaife, A. A., Xavier, P., McLean, P., Colman, A., and Cusack, S.: The GloSea4 Ensemble Prediction System for Seasonal Forecasting, *Mon. Weather Rev.*, 139(6), 1891–1910, [doi:10.1175/2010MWR3615.1](https://doi.org/10.1175/2010MWR3615.1), 2011.
- Baker, L. H., Shaffrey, L. C., Scaife, A. A.: Improved seasonal prediction of UK regional precipitation using atmospheric circulation, *Int. J. Clim.*, 38, e437–e453, [doi:10.1002/joc.5382](https://doi.org/10.1002/joc.5382), 2017.
- 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](https://doi.org/10.1016/j.jhydrol.2009.08.031), 2009.
- Bell, V. A., Kay, A. L., Cole, S. J., Jones, R. G., Moore, R. J., and Reynard, N. S.: How might climate change affect river flows across the Thames Basin? An area-wide analysis using the UKCP09 Regional Climate Model ensemble, *J. Hydrol.*, 442–443, 89–104, [doi:10.1016/j.jhydrol.2012.04.001](https://doi.org/10.1016/j.jhydrol.2012.04.001), 2012.
- 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](https://doi.org/10.1002/hyp.9863), 2013.
- Bell, V. A., Kay, A. L., Davies, H. N., and Jones, R. G.: An assessment of the possible impacts of climate change on snow and peak river flows across Britain, *Climatic Change*, 136, 539–553, [doi:10.1007/s10584-016-1637-x](https://doi.org/10.1007/s10584-016-1637-x), 2016.
- Bell, V. A., Davies, H. N., Kay, A. L., Brookshaw, A., and Scaife, A.: A national-scale seasonal hydrological forecast system: development and evaluation over Britain, *Hydrol. Earth Syst. Sc.*, 21(9), 4681–4691 [doi:10.5194/hess-21-4681-2017](https://doi.org/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](https://doi.org/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.

- 620 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](https://doi.org/10.1175/1520-0493(1998)126<2503:IOESOE>2.0.CO;2), 1998.
- Chan, W., Facer-Childs, K., Tanguy, M., Magee, E., Bulut, B., Stringer, N., Knight, J., and Hannaford, J.: UK Hydrological Outlook using Historic Weather Analogues, *EGUsphere* [preprint], <https://doi.org/10.5194/egusphere-2025-2369>, 2025.
- 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\(157\)](https://doi.org/10.1061/(ASCE)0733-9496(1985)111:2(157)), 1985.
- 625 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](https://doi.org/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](https://doi.org/10.5194/hess-25-4159-2021), 2021.
- 630 Dunstone, N., Smith, D., Scaife, A., Hermanson, L., Eade, R., Robinson, N., Andrews, M., and Knight, J.: Skilful predictions of the winter North Atlantic Oscillation one year ahead, *Nature Geoscience*, 9, 809-814, [doi:10.1038/ngeo2824](https://doi.org/10.1038/ngeo2824), 2016.
- Eade, R., Smith, D., Scaife, A., Wallace, E., Dunstone, N., Hermanson, L., and Robinson, N.: Do seasonal-to-decadal climate predictions underestimate the predictability of the real world?, *Geophys. Res. Lett.*, 41, 5620–5628, [doi:10.1002/2014GL061146](https://doi.org/10.1002/2014GL061146), 2014.
- 635 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](https://doi.org/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](https://doi.org/10.1002/qj.2270), 2014.
- 640 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.
- Fisher, R. A., and Yates, F.: Statistical tables: for biological, agricultural and medical research (6th ed.), Oliver & Boyd: Edinburgh, 1963.
- Formetta, G., Prosdocimi, I., Stewart, E., and Bell, V. A.: Estimating the index flood with continuous hydrological models: an application in Great Britain, *Hydrol. Res.*, 49, 123–133, 2018.
- 645 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](https://doi.org/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\(91\)](https://doi.org/10.1061/(ASCE)0733-9496(2002)128:2(91)), 2002.
- 650 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](https://doi.org/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](https://doi.org/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
655 observations, *Geosci. Data J.*, 6(2), 151-159, [doi:10.1002/gdj3.78](https://doi.org/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](https://doi.org/10.5194/hess-1-227-1997), 1997.

Jackson-Blake, L. A., Clayner, F., de Eyto, E., French, A. S., Frías, M. D., Mercado-Bettfán, D., Moore, T., Puértolas, L.,
660 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](https://doi.org/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.

Johansson, Å.: Prediction Skill of the NAO and PNA from Daily to Seasonal Time Scales, *J. Clim.*, 20(10), 1957-1975,
665 [doi:10.1175/JCLI4072.1](https://doi.org/10.1175/JCLI4072.1), 2007.

Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., Tietsche, S., Decremmer, D.,
Weisheimer, A., Balsamo, G., Keeley, S. P. E., Mogensen, K., Zuo, H., and Monge-Sanz, B. M.: SEAS5: the new ECMWF
seasonal forecast system, *Geosci. Model Dev.*, 12, 1087–1117, [doi:10.5194/gmd-12-1087-2019](https://doi.org/10.5194/gmd-12-1087-2019), 2019.

Kay, A. L., Rudd, A. C. and Coulson, J: Spatial downscaling of precipitation for hydrological modelling: Assessing a simple
670 method and its application under climate change in Britain, *Hydrol. Process.*, 37(2), 14823, [doi:10.1002/hyp.14823](https://doi.org/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](https://doi.org/10.1002/wea.2101), 2013.

Kay, A. L., Davies, H. N., Lane, R. A., Rudd, A. C., and Bell, V. A.: Grid-based simulation of river flows in Northern
Ireland: Model performance and future flow changes, *J. Hydrol: Reg. Studies*, 38, 100967, [doi:10.1016/j.ejrh.2021.100967](https://doi.org/10.1016/j.ejrh.2021.100967),
675 2021.

Kettleborough, J. A., Davis, P. J., Comer, R. E., Martin, M. J., Barbosa Aguiar, A., Collier, T., Scaife, A. A., Clark, A.,
Mancell, J., Smout-Day, K. A., and Hardiman, S. C.: 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), 2026.

Kharin, V. V., Merryfield, W. J., Boer, G. J., and Lee, W.-S.: A Postprocessing Method for Seasonal Forecasts Using
680 Temporally and Spatially Smoothed Statistics, *Mon. Weather Rev.*, 145, 3545-3561, [doi:10.1175/MWR-D-16-0337.1](https://doi.org/10.1175/MWR-D-16-0337.1), 2017.

Kim, H., Webster, P. J., Curry, J. A.: Seasonal Prediction skill of ECMWF System 4 and NCEP CFSv2 retrospective
forecast for the Northern Hemisphere Winter, *Climate Dynamics*, 39, 2957-2973, [doi:10.1007/s00382-012-1364-6](https://doi.org/10.1007/s00382-012-1364-6), 2012.

Kirtman, B. P., Min, D., Infanti, J. M., Kinter, J. L. III, Paolino, D. A., Zhang, Q., van den Dool, H., Saha, S., Mendez, M.
P., Becker, E., Peng, P., Tripp, P., Huang, J., DeWitt, D. G., Tippett, M. K., Barnston, A. G., Li, S., Rosati, A., Schubert, S.
685 D., Rienecker, M., Suarez, M., Li, Z. E., Marshak, J., Lim, Y.-K., Tribbia, J., Pegion, K., Merryfield, W. J., Denis, B., and
Wood, E. F.: 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), 585–601, [doi:10.1175/BAMS-D-12-00050.1](https://doi.org/10.1175/BAMS-D-12-00050.1), 2014.
- Knight, J. R. and Scaife, A. A.: Influences on North-Atlantic summer climate from the El Niño-Southern
690 Oscillation, *Quarterly Journal of the Royal Meteorological Society*, 150(764), 4498–4510, [doi:10.1002/qj.4826](https://doi.org/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](https://doi.org/10.1002/qj.4568), 2023.
- MacLachlan, C., Arribas, A., Peterson, K. A., Maidens, A., Fereday, D., Scaife, A. A., Gordon, M., Vellinga, M., Williams,
695 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](https://doi.org/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
700 Levels, *Weather Forecast.*, 14(5), 713–725, [doi:10.1175/1520-0434\(1999\)014<0713:CPROCA>2.0.CO;2](https://doi.org/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],
705 [doi:10.5285/b963ead70580451aa7455782224479d5](https://doi.org/10.5285/b963ead70580451aa7455782224479d5), 2024.
- Moore, R. J.: The PDM rainfall-runoff model, *Hydrol. Earth Syst. Sc.*, 11(1), 483–499, [doi:10.5194/hess-11-483-2007](https://doi.org/10.5194/hess-11-483-2007), 2007.
- Müller, W. A., Appenzeller, C., Doblas-Reyes, F. J., and Liniger, M. A.: A Debaised Ranked Probability Skill Score to
Evaluate Probabilistic Ensemble Forecasts with Small Ensemble Sizes, *J. Climate*, [doi:10.1175/JCLI3361.1](https://doi.org/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](https://doi.org/10.1175/1520-0450(1971)010<0155:ANOTRP>2.0.CO;2), 1971.
710
- 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](https://doi.org/10.1023/A:1011160904414), 2001.
- Nash, J. E., and Sutcliffe, J., V.: River flow forecasting through conceptual models part I — A discussion of principles, *J.*
715 *Hydrology*, 10(3), 282–290, [doi:10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6), 1970.
- Nikraftar, Z., Mbuva, R., Sadegh, M., and Landman, W. A., Impact-Based Skill Evaluation of Seasonal Precipitation
Forecasts, *Earth’s Future*, 12(11), e2024EF004936, [doi:10.1029/2024EF004936](https://doi.org/10.1029/2024EF004936), 2024.
- Peñuela, A., Hutton, C., and Pianosi, F.: Assessing the value of seasonal hydrological forecasts for improving water resource
management: insights from a pilot application in the UK, *Hydrol. Earth Sys. Sc.*, 24(12), 6059–6073, [doi:10.5194/hess-24-6059-2020](https://doi.org/10.5194/hess-24-6059-2020), 2020.
720

- 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](https://doi.org/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.,
725 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](https://doi.org/10.1080/02626667.2017.1395032), 2017.
- Quaglia, C. F., Terzago, S., and von Hardenberg, J.: Temperature and precipitation seasonal forecasts over the Mediterranean region: added value compared to simple forecasting methods, *Climate Dynamics*, 58, 2167–2191,
730 [doi:10.1007/s00382-021-05895-6](https://doi.org/10.1007/s00382-021-05895-6), 2022.
- Rameshwaran, P., Bell, V. A., Brown, M. J., Davies, H. N., Kay, A. L., Rudd, A. C., and Sefton, C.: 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), 2022.
- Rudd, A., Bell, V. A., and Kay, A.: National-scale analysis of simulated hydrological droughts (1891–2015), *J. Hydrol.*, 550,
735 368–385, [doi:10.1016/j.jhydrol.2017.05.018](https://doi.org/10.1016/j.jhydrol.2017.05.018), 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](https://doi.org/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.,
740 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](https://doi.org/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](https://doi.org/10.1038/s41612-018-0038-4), 2018.
- Schepen, A. and Wang, Q.J.: Model averaging methods to merge operational statistical and dynamic seasonal streamflow
745 forecasts in Australia, *Water Resour. Res.*, 51(3), 1797–1812, [doi:10.1002/2014WR016163](https://doi.org/10.1002/2014WR016163), 2015.
- Shepen, A., Zhao, T., Wang, Q. J., and Robertson, D. E.: A Bayesian modelling method for post-processing daily subseasonal to seasonal rainfall forecasts from global climate models and evaluation for 12 Australian catchments, *Hydrol. Earth Sys. Sci.*, 22, 1615–1628, [doi:10.5194/hess-22-1615-2018](https://doi.org/10.5194/hess-22-1615-2018), 2018.
- Sheffield, J., Wood, E. F., Chaney, N., Guan, K., Sadri, S., Yuan, X., Olang, L., Amani, A., Ali, A., Demuth, S., and Ogallo,
750 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](https://doi.org/10.1175/BAMS-D-12-00124.1), 2014.
- Slater, L. J., Villarini, G., and Bradley, A. A.: 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, *Clim. Dyn.*, 53, 7381–7396, [doi:10.1007/s00382-016-3286-1](https://doi.org/10.1007/s00382-016-3286-1), 2019.

- 755 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](https://doi.org/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.,
- 760 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](https://doi.org/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](https://doi.org/10.1080/02626667.2014.992788), 2016.
- Swets, J. A.: The Relative Operating Characteristic in Psychology, *Science*, 182(4116), 990–1000,
- 765 [doi:10.1126/science.182.4116.990](https://doi.org/10.1126/science.182.4116.990), 1973.
- Thornton, H. E., Smith, D. M., Scaife, A. A., and Dunstone, N. J.: Seasonal Predictability of the East Atlantic Pattern in Late Autumn and Early Winter, *Geophysical Research Letters*, 50(1), e2022GL100712, 2023.
- Troin, M., Arsenaault, 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](https://doi.org/10.1029/2020WR028392),
- 770 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](https://doi.org/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](https://doi.org/10.1038/s41612-018-0013-0), 2018.
- 775 Wang, Q. J., Shao, Y., Song, Y., Schepen, A., Robertson, D. E., Ryu, D., Pappenburger, F.: An evaluation of ECMWF SEAS5 seasonal climate forecasts for Australia using a new forecast calibration algorithm, *Env. Modelling & Software*, 122, 104550, [doi:10.1016/j.envsoft.2019.104550](https://doi.org/10.1016/j.envsoft.2019.104550), 2019.
- 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](https://doi.org/10.1002/j.1477-8696.1997.tb06323.x.s), 1997.
- 780 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](https://doi.org/10.1175/MWR3280.1), 2007.
- Weisheimer, A., Baker, L. H., Bröcker, J., Garfinkel, C. I., Hardiman, S. C., Hodson, D. L. R., Palmer, T. N., Robson, J. I., Scaife, A. A., Screen, J. A., Shepherd, T. G., Smith, D. M., and Sutton, R. T.: The Signal-to-Noise Paradox in Climate
- 785 Forecasts: Revisiting Our Understanding and Identifying Future Priorities, *Bulletin of the American Meteorological Society*, 105(3), E651-E659, [doi:10.1175/BAMS-D-24-0019.1](https://doi.org/10.1175/BAMS-D-24-0019.1), 2024.
- West, H., Quinn, N., and Horswell, M.: Spatio-Temporal Variability in North Atlantic Oscillation Monthly Rainfall Signatures in Great Britain, *Atmosphere*, 21(6), 763, [doi:10.3390/atmos12060763](https://doi.org/10.3390/atmos12060763), 2021.

White, C. J., Franks, S. W. and McEvoy, D.: Using subseasonal-to-seasonal (S2S) extreme rainfall forecasts for extended-
 790 range flood prediction in Australia, *Proceedings of IAHS*, 370, 229–234, [doi:10.5194/piahs-370-229-2015](https://doi.org/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](https://doi.org/10.1002/2017MS001115), 2018.

795 Zhao, T., Bennet, J. C., Wang, Q. J., Schepen, A., Wood, A. W., Robertson, D. E., and Ramos, M.-H.: How suitable is quantile mapping for postprocessing GCM precipitation forecasts? *J. Clim. Sci.*, 30(9), 3185-3196, [doi:10.1175/JCLI-D-16-0652.1](https://doi.org/10.1175/JCLI-D-16-0652.1), 2017.