Check for updates

AGU ADVANCING EARTH AND SPACE SCIENCE

Geophysical Research Letters

RESEARCH LETTER
10.1029/2022GL100650

Skillful Decadal Flood Prediction

S. Moulds[1], L. J. Slater[1], N. J. Dunstone[2], and D. M. Smith[2]

[1]School of Geography and the Environment, University of Oxford, Oxford, UK, [2]Met Office Hadley Centre, Exeter, UK

Key Points:

- Multiyear predictions of mean winter floods 2–5 years ahead are skillful across much of the UK
- Skill is improved by "NAO-matching" to overcome spuriously weak modeled signals
- The higher the sensitivity of streamflow to the North Atlantic Oscillation at a given gauge, the greater the benefit of NAO-matching for decadal flood prediction

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

S. Moulds,
simon.moulds@ouce.ox.ac.uk

Citation:

Moulds, S., Slater, L. J., Dunstone, N. J., & Smith, D. M. (2023). Skillful decadal flood prediction. *Geophysical Research Letters*, *50*, e2022GL100650. https://doi.org/10.1029/2022GL100650

Received 29 JUL 2022
Accepted 4 DEC 2022

Author Contributions:

Conceptualization: S. Moulds, L. J. Slater
Data curation: S. Moulds
Formal analysis: S. Moulds, L. J. Slater
Funding acquisition: L. J. Slater
Investigation: S. Moulds, L. J. Slater
Methodology: S. Moulds, L. J. Slater
Project Administration: L. J. Slater
Resources: N. J. Dunstone
Software: S. Moulds, L. J. Slater, D. M. Smith
Validation: S. Moulds, L. J. Slater
Visualization: S. Moulds, L. J. Slater
Writing – original draft: S. Moulds
Writing – review & editing: S. Moulds, L. J. Slater, N. J. Dunstone, D. M. Smith

© 2022 The Authors.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Abstract

Accurate long-term flood predictions are increasingly needed for flood risk management in a changing climate, but are hindered by the underestimation of climate variability by climate models. Here, we drive a statistical flood model with a large ensemble of dynamical CMIP5-6 predictions of precipitation and temperature. Predictions of UK winter flooding (95th streamflow percentile) have low skill when using the raw 676-member ensemble averaged over lead times of 2–5 years from the initialization date. Sub-selecting 20 ensemble members that adequately represent the multiyear temporal variability in the North Atlantic Oscillation (NAO) significantly improves the flood predictions. Applying this method we show positive skill in 46% of stations compared to 26% using the raw ensemble, primarily in regions most strongly influenced by the NAO. Our findings reveal the potential of decadal predictions to inform flood risk management at long lead times.

Plain Language Summary

Reliable predictions of flooding can help society to manage the associated risk to lives and property. Seasonal predictions of flooding over the coming months already form the basis of many operational services around the world. In contrast, decadal predictions with lead times of up to 10 years are more challenging, due to the difficulty of simulating dynamic changes in atmospheric circulation at these timescales. Here, we show that a large ensemble of climate models can predict average winter flood conditions over the UK in the next decade. The climate models underestimate the magnitude of atmospheric variability in the north Atlantic and identifying a subset of skillful climate model simulations improves the ability to predict floods. Our results suggest that decadal climate predictions may be useful in the context of flood risk management. However, the use of multiyear averages for flood prediction is still poorly studied, and therefore further work should help determine how such predictions can be used in an operational setting.

1. Introduction

Timely and accurate predictions of flooding are crucial to flood risk management and efforts to adapt to our rapidly changing climate. Research to improve the accuracy of flood predictions has focused on short to medium lead times of up to 12 months in advance (Emerton et al., 2016). In contrast, decadal climate predictions, which forecast the climate system up to 10 years ahead (Done et al., 2021), remain underexplored for flood prediction (Neri et al., 2019). However, predictions of hydrological extremes at annual to decadal lead times are important for near-term climate planning (e.g., adequate flood protection systems), and may therefore offer significant benefits to society (Meehl et al., 2009). We address this research gap by exploring the predictability of UK winter flooding at decadal timescales.

Decadal climate predictions specify changes in radiative forcing and initialize slowly varying sources of climate variability (Dunstone et al., 2020; Kushnir et al., 2019). Prediction skill is typically clearest at multiyear forecast periods (e.g., years 2–9 from the initialization date) (Dunstone et al., 2020), and depends on the ability of climate models to simulate the evolution of the climate system in response to internal and external sources of variability. Hence, the objective of decadal flood predictions is to forecast the occurrence and magnitude of multiyear flood-rich and flood-poor periods, which vary over multidecadal timescales (Blöschl et al., 2019; Wilby & Quinn, 2013). Although the current generation of decadal predictions shows high skill for surface temperature (Smith et al., 2019), their ability to simulate dynamical changes in atmospheric circulation and precipitation is less clear (Knutti & Sedláček, 2012; Shepherd, 2014), presenting a barrier to reliable flood predictions (Neri et al., 2019).

The North Atlantic Oscillation (NAO) is the principal mode of variability in atmospheric circulation in the north Atlantic and the main driver of flood predictability over the British Isles (Biggs & Atkinson, 2011; Hannaford & Marsh, 2008). The current generation of climate models skilfully predict the phase of observed NAO variability but severely underestimate its amplitude at decadal timescales (Smith et al., 2020—henceforth S20). This

19448007, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022GL100650 by Test, Wiley Online Library on [06/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

occurs because the signal-to-noise ratio of decadal forecasts is too low (Dunstone et al., 2016; Eade et al., 2014; Scaife & Smith, 2018; Smith et al., 2019). With a sufficiently large climate ensemble the noise can be reduced to produce a skillful representation of decadal NAO variability (Smith et al., 2019; S20). However, the weak NAO signal in models implies that the magnitude of teleconnections between the NAO and northern European weather is underestimated relative to other factors driving climate variability. S20 proposed that this weakness could be overcome by selecting members for which the simulated NAO index, and hence related teleconnections, are as close as possible to the forecast magnitude after scaling to account for the signal underestimation. Applying this "NAO-matching" technique, S20 showed significant improvements in multiyear forecasts of mean north Atlantic climate. Here, we test whether this improvement in NAO predictability enhances flood prediction skill.

Most operational flood forecasts are produced by forcing a conceptual or physics-based hydrological model with dynamical daily or sub-daily climate inputs (Emerton et al., 2016). Statistical-dynamical, or hybrid, methods present an attractive alternative approach to produce streamflow forecasts, as they avoid the need to run an offline land model and benefit from advances in statistical modeling (AghaKouchak et al., 2022; Slater et al., 2022). Streamflow quantiles are predicted using statistical or machine-learning models driven by dynamical weather or climate predictions. For example, Slater and Villarini (2018) used monthly mean temperature and precipitation from the North American Multi-Model Ensemble to make skillfull streamflow forecasts for the Midwestern United States at seasonal lead times. Hybrid methods have several advantages compared to dynamical methods. They are computationally efficient, and can integrate a wide variety of nonstationary predictor variables, including climate forecasts, large-scale climate indices, and teleconnections (Slater et al., 2022).

We develop a statistical-dynamical framework to assess the predictability of boreal winter (December-March; DJFM) high streamflows in the United Kingdom. We focus on DJFM as it represents the major UK flood season (Huntingford et al., 2014; Robson, 2002). Moreover, previous work has shown that positive NAO (NAO+) phases are strongly correlated with higher rainfall in northwest UK, making DJFM an appropriate season for assessing the potential of NAO-matching to improve decadal flood predictions (Simpson & Jones, 2014; West et al., 2019). We use a distributional regression approach where the distribution parameters depend on time-varying predictor variables derived from retrospective decadal forecasts (i.e., hindcasts) from the Coupled Model Intercomparison Project (CMIP), initialized each year between 1960 and 2005. We first quantify the baseline hindcast flood skill over multiyear timescales using the mean forecast of precipitation and temperature from a large climate multi-model ensemble. We then assess the extent to which NAO-matching can improve decadal flood forecasts by better representing the teleconnection between the NAO and northern European precipitation and temperature.

2. Materials and Methods

2.1. Flood Prediction

We used a statistical-dynamical framework (Slater et al., 2022) to predict multiyear average winter (DJFM) high flows in the UK using precipitation and temperature. Decadal climate predictions are currently only publicly available as monthly aggregates. At this level of temporal aggregation the prediction of daily streamflow quantiles is not possible using either a conceptual or physics-based hydrological model. We focused on the 95th percentile of daily winter streamflow (Q95—the daily streamflow equaled or exceeded 5% of the time during each winter season), which we took as broadly representative of flood conditions across the UK. We evaluated prediction skill for year 2–5 multiyear averages (Figure 1a). This period is amongst those recommended for the verification of decadal predictions (Goddard et al., 2013), and provides a reasonable number of data points for fitting the statistical models. We modeled DJFM Q95 streamflow using a Gamma distribution, which has been shown to work well in this context as it models continuous positive variables and is flexible enough to describe data that is symmetrical through to highly positively skewed (Villarini et al., 2009; Villarini & Strong, 2014). The distribution has location and shape parameters (μ and σ, respectively) that depend linearly on the predictor variables through a logarithmic link function. The location parameter μ is allowed to vary over time as a linear function of predictor variables. We kept σ constant, as previous work has shown that using variable σ does not improve model fit for streamflow (Villarini & Strong, 2014). We used the Generalized Additive Models for Location, Scale and Shape implementation using the *gamlss R* package (Rigby & Stasinopoulos, 2005). We verified the independence and normality of the residuals by computing their mean, variance, kurtosis, and skewness (Figure S1 in Supporting Information S1).

19448007, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022GL100650 by Test, Wiley Online Library on [06/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

10.1029/2022GL100650

To obtain the decadal flood hindcasts we fitted two separate statistical models to each Q95 time series (Table S1 in Supporting Information S1). The simplest model (P in Table S1 in Supporting Information S1) has mean DJFM northern European precipitation (x_p) as the only predictor, while the second model (PT) includes precipitation and mean DJFM temperature (x_t). We used precipitation and temperature averaged over northern Europe (Figure 1) instead of the UK because our focus is on large-scale drivers of streamflow variability. We found the Pearson's correlation coefficient between the observed 8-year DJFM mean NAO and northern European precipitation is stronger than the correlation coefficient with UK precipitation (0.85 vs 0.66). This indicates that the teleconnection between the NAO and north Atlantic climate is stronger over the larger region, suggesting that the larger region has less unpredictable noise. One alternative to using spatially averaged values would be to take precipitation/temperature values from the nearest climate grid cell to a given river catchment. However, selecting climate inputs in this way could potentially introduce greater noise to the statistical models which would otherwise be smoothed by averaging over a larger spatial domain (Guemas et al., 2015).

To fit the models we used a cross-validation approach where each year in the time series was excluded once and the model trained on the remaining years. As the multiyear averages overlap (e.g., the year 2–5 forecast initialized in 1980 shares 3 years with the forecasts initialized in 1979 and 1981; Figure 1b), it was necessary to additionally leave out 3 years either side of the year in question to prevent training data contamination. In each case the fitted model was used to predict the Q95 streamflow for the excluded data point until a prediction was available for each year in the study period. We used this cross-validation approach because the hindcast period (1960–2005) is too short to separate into strictly sequential training and testing periods. We evaluated predictive skill using the Continuous Ranked Probability Score (CRPS) and associated skill score (Continuous ranked probability skill score (CRPSS)). The CRPSS evaluates probabilistic forecasts with respect to a reference forecast over the whole probability distribution, penalizing forecasts with large biases or low sharpness. As the reference forecast we used a probabilistic climatology that consisted of the observations from all years included in the training data set. We compute the CRPS of the forecast and reference for each test year and then take the mean across all test years (see Methods in Supporting Information S1). At each gauging station we selected the best performing model out of P and PT as the one with the lowest CRPS value. Lastly, we calculated the coefficient of determination (R^2) between the observed multiyear DJFM Q95 streamflow and the observed NAO to quantify the influence of NAO on streamflow variability at each gauging station.

2.2. Streamflow Observations

We used daily streamflow data from the UK Benchmark Network (UKBN2) (Harrigan et al., 2018). This network is a subset of the UK national hydrometric network, which comprises "near-natural" catchments with good quality streamflow records of long duration (Harrigan et al., 2018). In total UKBN2 has 146 stations, which are classified according to their suitability for assessments of low, medium and high flows. We selected stations that were (a) suitable for the analysis of high flows (Harrigan et al., 2018); (b) had streamflow records between 1975 and 2005; and (c) were missing no more than 20% of daily values in any given DJFM season. This resulted in a total of 98 stations. We calculated the DJFM Q95 streamflow for each year on record.

2.3. Observed and Modeled Climate Variables

Observations of monthly mean sea-level pressure were retrieved from HadSLP2 (Allan & Ansell, 2006). Observed near-surface temperature was taken as the average of HadCRUT4 (Morice et al., 2012), NASA-GISS (Hansen et al., 2010), and NCDC (Karl et al., 2015). Observed precipitation was retrieved from GPCC (Schneider et al., 2014) (Table S2 in Supporting Information S1). These datasets correspond with the observational datasets used by S20. We used a large multi-model ensemble of decadal hindcasts from the CMIP phases 5 (Taylor et al., 2012) and 6 (Boer et al., 2016), consisting of 169 members from 13 climate models (Table S3 in Supporting Information S1). The models differ both in structure and initialization. We used hindcasts initialized each year between 1960 and 2005, from which we extracted monthly sea-level pressure, surface temperature, and precipitation fields. We chose this time period because the number of members in the multimodel ensemble is consistent throughout. Observations and models were resampled to a common 5° by 5° latitude/longitude grid prior to the analysis.

From the observed and modeled climate data we computed the mean winter (DJFM) NAO index and northern European precipitation and temperature (Figure 1e; see Methods in Supporting Information S1).

19448007, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022GL100650 by Test, Wiley Online Library on [06/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19448007, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022GL100650 by Test, Wiley Online Library on [06/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.4. Lagged Ensemble

Although consecutive year 2–5 forecasts share three identical years (overlap of a 4-year moving window; Figure 1b), there are large interannual variations in the hindcast time series data. These occur because of the insufficient signal-to-noise ratio in the current generation of climate models, combined with the fact that model simulations initialized at different times contain different samples of the noise. We reduced the interannual variability by combining the required forecast with the preceding three forecasts (e.g., the year 2–5 hindcast starting in 1970 includes year 2–5 forecasts initialized in 1970, 1969, 1968, and 1967—Figure 1b). In effect, this quadrupled the ensemble size to 676 members, albeit at the cost of some additional temporal smoothing (S20).

2.5. NAO-Matching

The use of a large multimodel ensemble provides skillful NAO predictions in the sense that the ensemble mean is highly correlated with the observations. However, because the predictable signal is underestimated in the CMIP5-6 outputs, the magnitude of the ensemble-mean NAO predictions is too small. This means that the teleconnections associated with NAO variability are similarly misrepresented by the raw ensemble mean model outputs (S20). The NAO-matching technique developed by S20 overcomes this problem by selecting from the raw ensemble (676 members) a subset of 20 members at each time point in which the NAO and related teleconnections have more realistic magnitudes (see Methods in Supporting Information S1). The need for NAO-matching fundamentally arises because predictable atmospheric circulation changes are too weak in models. Consequently, regions impacted by the NAO, including UK precipitation, have low skill in the ensemble mean (Figure 2e in S20 and Figure S3e in Supporting Information S1). Ensemble selection relies on there being skill in the ensemble mean, and therefore could not be applied to precipitation and temperature directly. We applied the NAO-matching technique and tested whether climate inputs from the NAO-matched ensemble improved the skill of decadal flood forecasts.

3. Results

We present results for the year 2–5 multiyear mean DJFM Q95 streamflow predictions. These are skillful (positive CRPSS) at 25 out of 98 stations in the UK using the raw ensemble of 676 members (Figure 1c; see Figure S2 in Supporting Information S1 e.g., time series predictions). The median CRPSS across all stations is −0.105, taking the best performing model out of *P* and *PT*. We contrast this result with the case where the statistical models are fitted against perfect climate predictions (i.e., predictors derived from observed north Atlantic climate variables) for the same time period (1960–2005; Figure 1d). Here, 53 out of 98 stations have positive skill taking the best performing model. Across all stations the median CRPSS is 0.006, whereas if only stations with positive skill are considered the median is 0.146 (Figure 1d).

The inclusion of temperature in model *PT* increases model skill relative to model *P* in 28 stations using the raw ensemble mean (Figure 1a), and 18 stations using observed predictors (Figure 1b). In some places (e.g., Scotland, north England) this could be related to snowmelt-dominated flood generation (Berghuijs et al., 2019), as the decadal forecasts pick up a monotonic trend of increasing temperature from around 1970 (Figure S3 in Supporting Information S1). However, it is also possible that streamflow is responding to a concurrent trend such as urbanization (Han et al., 2022). Although the UKBN2 streamflow data set controls for non-climatic drivers of streamflow to a reasonable extent, few UK catchments are unaffected by anthropogenic activity (Harrigan et al., 2018).

Applying NAO-matching increases the predictive skill of models *P* and *PT* in the majority of catchments (Figure 2b) relative to predictions which use precipitation and temperature from the raw decadal hindcast ensemble (Figure 1c; Figure 3). On average, the predictive skill of models *P* and *PT* is significantly higher ($p < 0.05$; one-sided Wilcoxon signed rank test; Figure 2c) using the NAO-matched ensemble (Figure 2a) compared to the raw ensemble (Figure 1c) across all gauging stations. As mentioned previously, using observed predictors we find skill in 53 out of 98 stations (Figure 1b). Using predictors from the NAO-matched ensemble we find skill in 74% of these stations, compared to only 32% using the raw ensemble. The increase in skill arises from improvements in predictions of both temperature and precipitation (Figure S3 in Supporting Information S1; S20). We also considered the ability of statistical models to reproduce the Q95 DJFM streamflow anomaly during the extreme NAO+period (1985–1989). We find that models forced with predictor variables taken from the NAO-matched

10.1029/2022GL100650

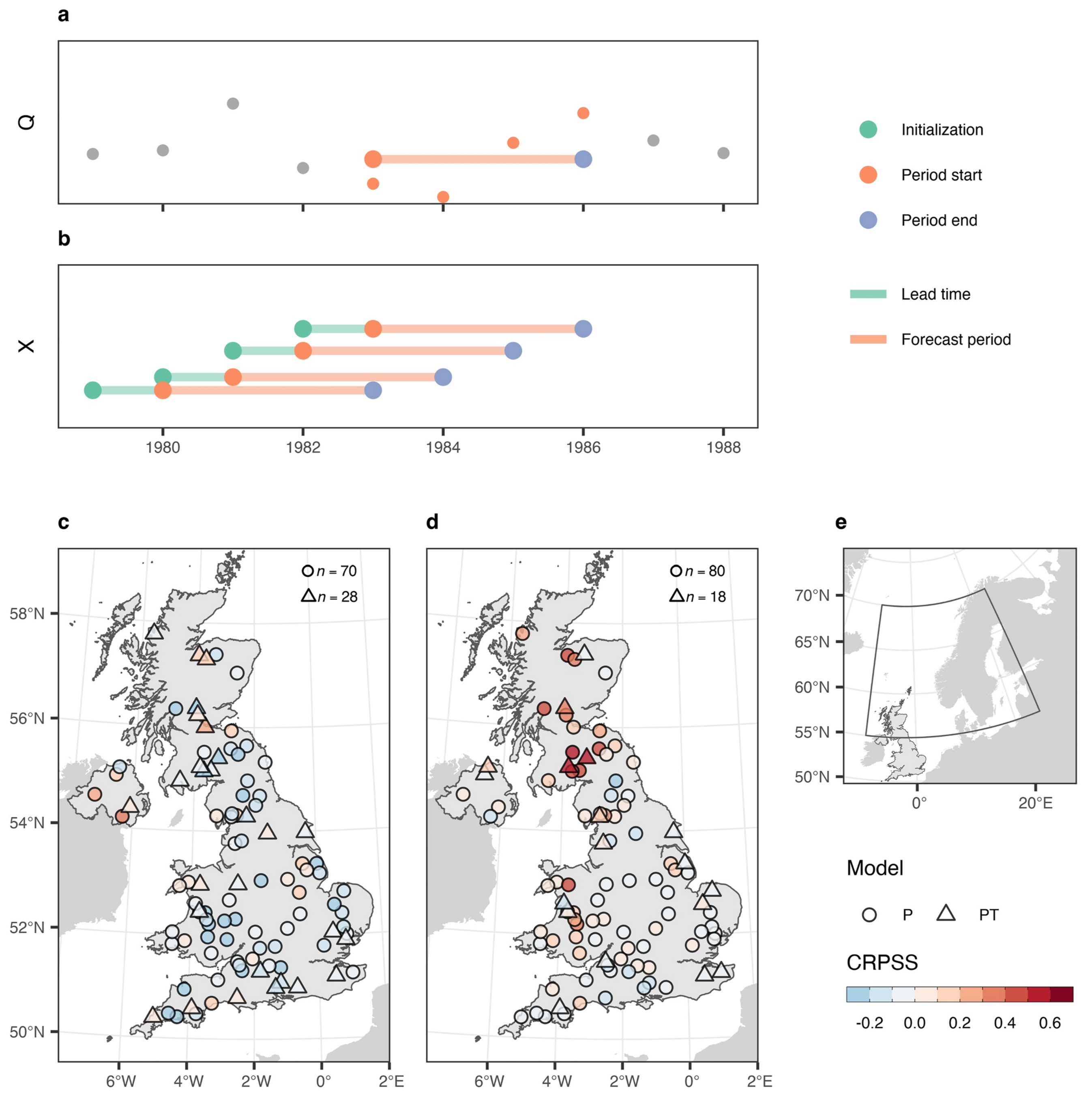

Figure 1. Overview of multiyear prediction scheme where the predictand is the Q95 DJFM streamflow averaged across each boreal winter season in the multiyear period (a). Small color circles indicate daily Q95 streamflow in each winter season; those which contribute to the displayed multiyear period are shown in orange. The horizontal orange line indicates the multiyear forecast, with start and end years as large orange and blue circles, respectively. The forecasting systems are initialized once per year, with start times between 1 November and 1 January. The raw ensemble is quadrupled in size from 169 to 676 members by combining the required forecast with the previous three (b). During NAO-matching we consider each of 676 members for inclusion in the NAO-matched ensemble. We compare the predictability of winter high flow (DJFM Q95) using precipitation and temperature averaged across the raw ensemble (676 members) (c) with the best possible skill obtained using precipitation and temperature from observations (d) for the period 1960–2005. At each station the model shown is the one with the lowest Continuous Ranked Probability Score, indicated with a circle (*P*) or triangle (*PT*). Climate input variables are averaged over the North Atlantic region (e).

19448007, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022GL100650 by Test, Wiley Online Library on [06/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

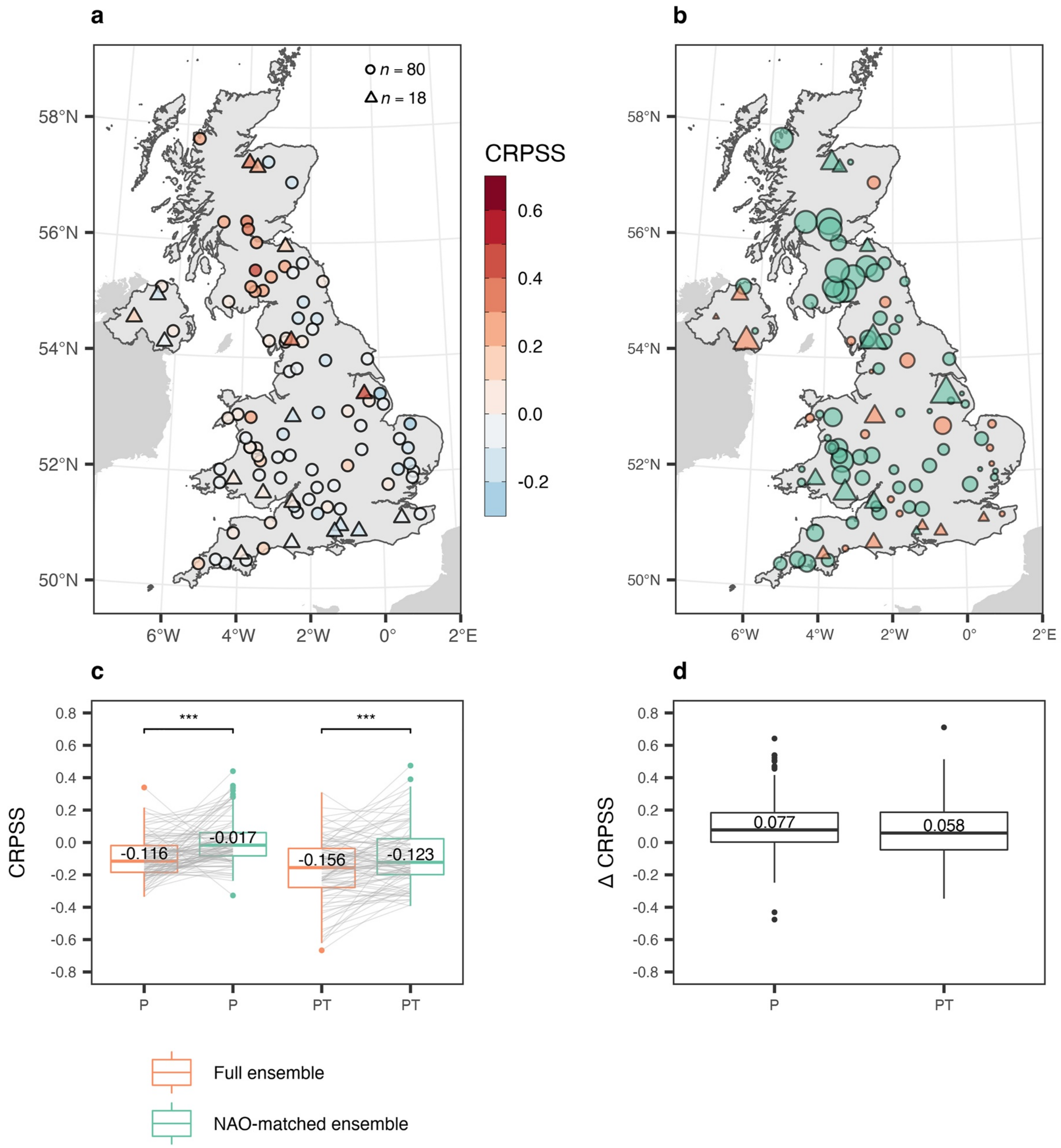

Figure 2. Effect of NAO-matching on predictive skill. Model skill for the best performing model (out of *P* and *PT*) using predictors from the NAO-matched ensemble 20 members; (a). The best performing model (*P* or *PT*) using predictor variables from the NAO-matched ensemble increases model skill at 74 out of 98 stations compared to the equivalent model using predictors from the raw ensemble (b). We used a one-sided Wilcoxon signed rank test to test whether the skill scores for models using predictor variables from the NAO-matched ensemble were significantly higher ($p < 0.01$, indicated by ***) than those using predictor variables from the raw ensemble (c). The median increase in Continuous ranked probability skill score as a result of NAO-matching was positive for both models *P* and *PT* (d). The sample size is 98 NRFA gauging stations in each map and boxplot.

ensemble better reproduce the observed magnitude of the anomaly compared to those which use predictor variables from the raw ensemble (Figure S4 in Supporting Information S1).

The improvement in CRPSS is largest amongst stations where decadal DJFM streamflow variability is influenced by the NAO (Figure 4). We find that the higher the sensitivity of streamflow to the NAO at a given gauge (i.e., the more variance in observed streamflow i.e., explained by the observed NAO), the greater the benefit of the NAO-matching technique for decadal flood prediction. Stations where the NAO is a poor explanatory variable tend to have either a small positive or negative change in skill as a result of applying the NAO-matching technique (Figure 4).

19448007, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022GL100650 by Test, Wiley Online Library on [06/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

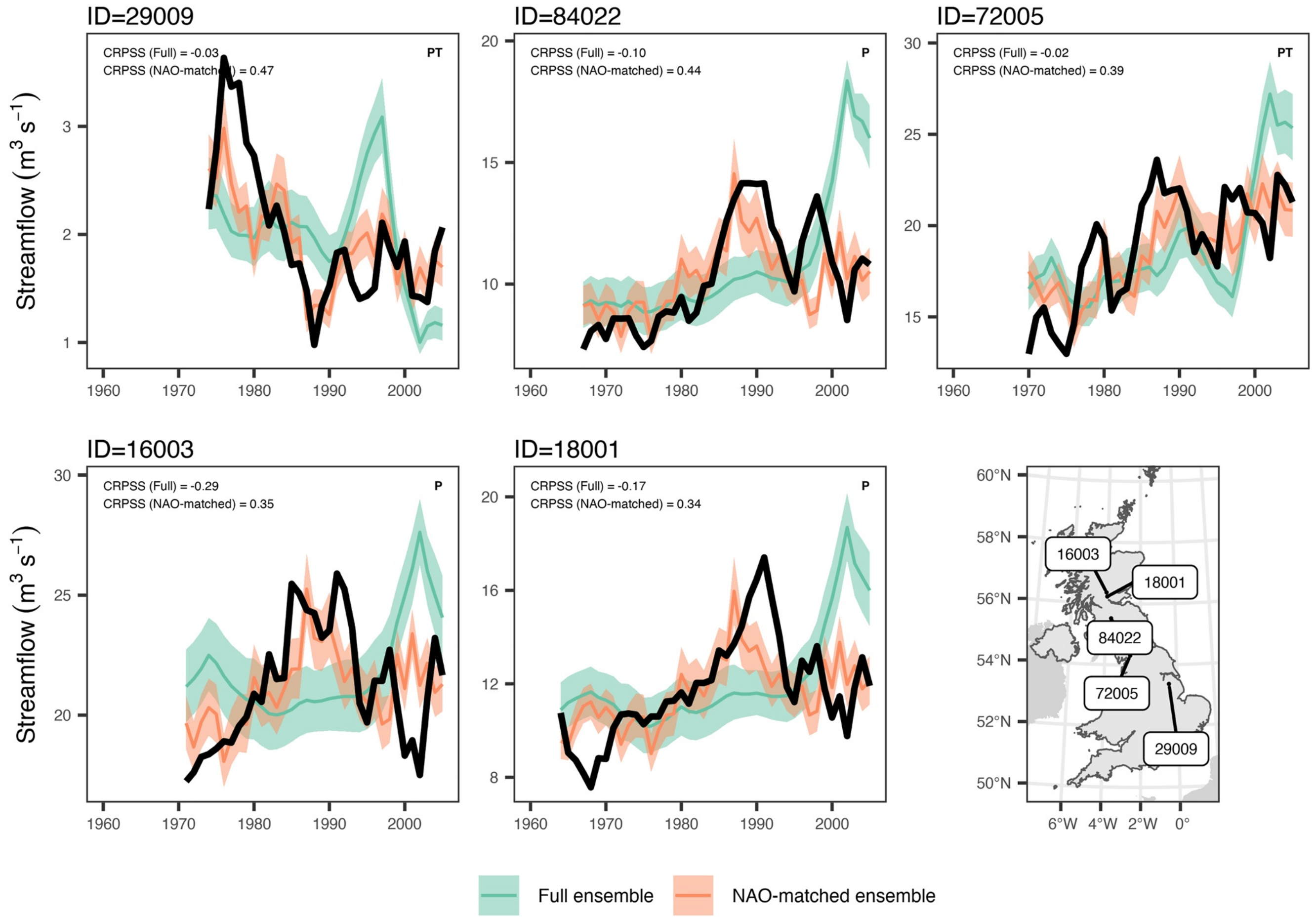

Figure 3. Examples of forecasts of DJFM Q95 streamflow in catchments where year 2–5 predictability improves as a result of using predictor variables from the NAO-matched ensemble (20 members, Figure 2a) compared to predictor variables from the raw ensemble (676 members; Figure 1c). In each case we show the timeseries for the streamflow model (*P* or *PT*, shown in the top right corner of each panel) showing the largest increase in Continuous ranked probability skill score. The black line in the time series plots is the observed multiyear streamflow. Solid color lines are the median forecasts (i.e., 50th percentile), with confidence bands representing 25th–75th percentiles of the raw Coupled Model Intercomparison Project ensemble and NAO-matched ensemble.

Decadal climate variability in northern Europe is strongly linked to the NAO, and NAO-matching improves the skill of precipitation and temperature predictions for this region. The NAO mainly influences UK winter streamflow in the north and northwest UK, and has less influence in east and south UK where antecedent conditions are a better predictor of winter streamflow (Svensson et al., 2015; West et al., 2019). Other climate modes may therefore have greater value for enhancing flood prediction in different regions of the UK and globally, but further work would be required to assess the extent to which they improve the predictability of climate variables.

4. Conclusions

We developed a framework for UK flood prediction at decadal time scales using a large ensemble of CMIP decadal predictions (676 members). We found that multiyear predictions of DJFM high flows exhibit some skill using a statistical-dynamical framework driven by the raw predictions from the large ensemble. Flood predictability varies across the UK, with the highest skill observed in northwest UK, where the NAO is known to influence winter streamflow. Our results suggest that skillful decadal flood prediction is dependent on understanding the dominant large-scale drivers of climate variability in specific catchments. There is low skill in parts of the UK where the NAO does not drive streamflow variability. This is the case both for observed and modeled covariates, suggesting that in these locations north Atlantic precipitation and temperature do not explain streamflow variability at the 2–5 multiyear timescales we considered here. We find a strong positive association between the increase in the decadal prediction skill obtained as a result of applying NAO-matching, and the strength of the association

19448007, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022GL100650 by Test, Wiley Online Library on [06/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

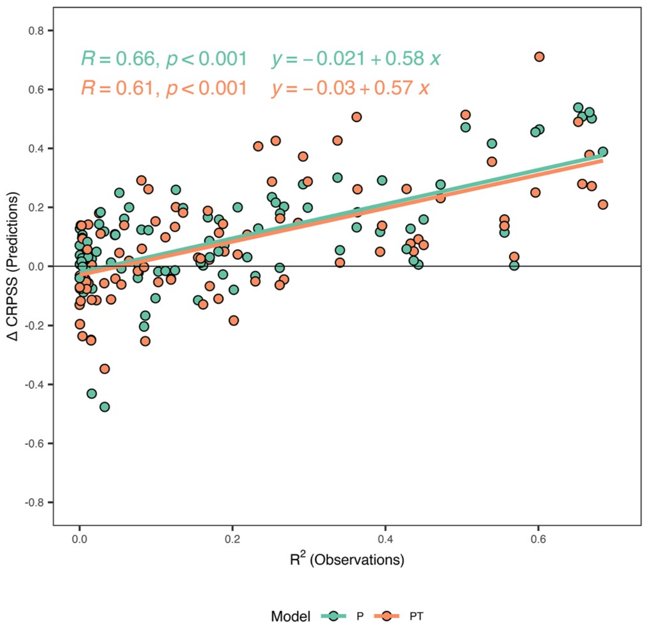

Figure 4. The higher the sensitivity of streamflow to the North Atlantic Oscillation (NAO) at a given gauge, the greater the benefit of NAO-matching for decadal flood prediction. Change in skill of statistical flood models (*P* and *PT*) obtained as a result of NAO-matching (*y*) against the coefficient of determination (R^2) between the observed multiyear DJFM Q95 streamflow and the observed NAO (*x*). Each circle indicates one site, with every site shown for each model.

between observed NAO and the year 2–5 DJFM Q95 variability. In other words, the benefits of the NAO-matching approach for flood predictability are greatest in locations where the NAO exerts control on the magnitude and frequency of flooding. It would be beneficial to trial the same approach in other regions of the world where alternative modes of variability are known to control the occurrence of flood-rich and flood-poor periods.

Our results show that decadal flood predictions can be skillful when averaged over multiyear periods. Earlier work, in contrast, found little skill in predicting individual years (rather than multiyear periods) using CMIP5 decadal climate predictions (Neri et al., 2019). Most of the previous research on decadal climate prediction has concentrated on the predictability of large-scale climate variability. The decadal timescale is a key planning horizon for many flood risk managers (Done et al., 2021). However, they typically require more fine-grained information to inform decisions (Dunstone et al., 2022). Our work demonstrates that a dynamical-statistical approach can bridge these spatial scales to provide skillful predictions of high streamflow at the catchment scale. It therefore represents an important step toward useable decadal climate services.

Although multiyear predictions of high streamflow exhibit skill, operationalizing such information would require a shift in the way that stakeholders understand and act upon flood predictions. A key priority is to identify ways to communicate this skill and its associated uncertainty such that multiyear decadal predictions deliver social and economic value (Meehl et al., 2009). Multiyear predictions could be used to provide information on flood risk at the annual scale (Ward & Conway, 2020). For example, given skillfull predictions of year 2–5 DJFM high flows, it should be possible to estimate the increased risk of flooding in each year of the forecast period. As decadal predictions start to be made on annual basis for operational purposes (Hermanson et al., 2022; Smith et al., 2013), more research is needed to understand how their utility can be maximized for the benefit of society.

Data Availability Statement

The hindcasts used in this study are available from the CMIP data archives: https://esgf-node.llnl.gov/projects/cmip5/ and https://esgf-node.llnl.gov/projects/cmip6/. All observed datasets are available online (HadSLP2r: https://psl.noaa.gov/gcos_wgsp/Gridded/data.hadslp2.html HADCRUT4: https://www.metoffice.gov.uk/hadobs/hadcrut4/ GISS: https://data.giss.nasa.gov/pub/gistemp/ NCDC: ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/blended GPCC: https://downloads.psl.noaa.gov/Datasets/gpcc/full_v2018/). The raw climate data was processed using ESMValTool (https://www.esmvaltool.org/), which is free and open-source. The processed input data, consisting of observed and simulated climate indices, and the scripts that are needed to reproduce the results of this study, have been uploaded to a Zenodo data repository (https://doi.org/10.5281/zenodo.6940449) under an MIT license.

Acknowledgments
This work is funded by UK Research and Innovation Grant MR/V022008/1 (SM and LJS). NJD and DMS were supported by the Met Office Hadley Centre Climate Programme funded by BEIS and Defra. The authors would like to acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility in carrying out this work (http://dx.doi.org/10.5281/zenodo.22558).

References

AghaKouchak, A., Pan, B., Mazdiyasni, O., Sadegh, M., Jiwa, S., Zhang, W., et al. (2022). Status and prospects for drought forecasting: Opportunities in artificial intelligence and hybrid physical–statistical forecasting. *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences*, *380*(2238), 20210288. https://doi.org/10.1098/rsta.2021.0288

Allan, R., & Ansell, T. (2006). A new globally complete monthly historical gridded mean sea level pressure data set (HadSLP2): 1850–2004. *Journal of Climate*, *19*(22), 5816–5842. https://doi.org/10.1175/JCLI3937.1

Berghuijs, W. R., Harrigan, S., Molnar, P., Slater, L. J., & Kirchner, J. W. (2019). The relative importance of different flood-generating mechanisms across Europe. *Water Resources Research*, *55*(6), 4582–4593. https://doi.org/10.1029/2019WR024841

Biggs, E. M., & Atkinson, P. M. (2011). A characterisation of climate variability and trends in hydrological extremes in the severn uplands. *International Journal of Climatology*, *31*(11), 1634–1652. https://doi.org/10.1002/joc.2176

Blöschl, G., Hall, J., Viglione, A., Perdigão, R. A. P., Parajka, J., Merz, B., et al. (2019). Changing climate both increases and decreases European river floods. *Nature*, *573*(7772), 108–111. https://doi.org/10.1038/s41586-019-1495-6

19448007, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022GL100650 by Test, Wiley Online Library on [06/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19448007, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022GL100650 by Test, Wiley Online Library on [06/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., et al. (2016). The decadal climate prediction Project (DCPP) contribution to CMIP6. *Geoscientific Model Development*, *9*(10), 3751–3777. https://doi.org/10.5194/gmd-9-3751-2016

Done, J. M., Morss, R. E., Lazrus, H., Towler, E., Tye, M. R., Ge, M., et al. (2021). Toward useable predictive climate information at decadal timescales. *One Earth*, *4*(9), 1297–1309. https://doi.org/10.1016/j.oneear.2021.08.013

Dunstone, N., Lockwood, J., Solaraju-Murali, B., Reinhardt, K., Tsartsali, E. E., Athanasiadis, P. J., et al. (2022). Towards useful decadal climate services. *Bulletin of the American Meteorological Society*, *103*(7), E1705–E1719. https://doi.org/10.1175/BAMS-D-21-0190.1

Dunstone, N., Smith, D., Scaife, A., Hermanson, L., Eade, R., Robinson, N., et al. (2016). Skilful predictions of the winter North Atlantic Oscillation one year ahead. *Nature Geoscience*, *9*(11), 809–814. https://doi.org/10.1038/ngeo2824

Dunstone, N., Smith, D., Yeager, S., Danabasoglu, G., Monerie, P.-A., Hermanson, L., et al. (2020). Skilful interannual climate prediction from two large initialised model ensembles. *Environmental Research Letters*, *15*(9), 094083. https://doi.org/10.1088/1748-9326/ab9f7d

Eade, R., Smith, D., Scaife, A., Wallace, E., Dunstone, N., Hermanson, L., & Robinson, N. (2014). Do seasonal-to-decadal climate predictions underestimate the predictability of the real world?: Seasonal to decadal predictability. *Geophysical Research Letters*, *41*(15), 5620–5628. https://doi.org/10.1002/2014GL061146

Emerton, R. E., Stephens, E. M., Pappenberger, F., Pagano, T. C., Weerts, A. H., Wood, A. W., et al. (2016). Continental and global scale flood forecasting systems. *WIREs Water*, *3*(3), 391–418. https://doi.org/10.1002/wat2.1137

Goddard, L., Kumar, A., Solomon, A., Smith, D., Boer, G., Gonzalez, P., et al. (2013). A verification framework for interannual-to-decadal predictions experiments. *Climate Dynamics*, *40*(1–2), 245–272. https://doi.org/10.1007/s00382-012-1481-2

Guemas, V., García-Serrano, J., Mariotti, A., Doblas-Reyes, F., & Caron, L.-P. (2015). Prospects for decadal climate prediction in the Mediterranean region. *Quarterly Journal of the Royal Meteorological Society*, *141*(687), 580–597. https://doi.org/10.1002/qj.2379

Han, S., Slater, L., Wilby, R. L., & Faulkner, D. (2022). Contribution of urbanisation to non-stationary river flow in the UK. *Journal of Hydrology*, *613*, 128417. https://doi.org/10.1016/j.jhydrol.2022.128417

Hannaford, J., & Marsh, T. J. (2008). High-flow and flood trends in a network of undisturbed catchments in the UK. *International Journal of Climatology*, *28*(10), 1325–1338. https://doi.org/10.1002/joc.1643

Hansen, J., Ruedy, R., Sato, M., & Lo, K. (2010). Global surface temperature change. *Reviews of Geophysics*, *48*(4), RG4004. https://doi.org/10.1029/2010RG000345

Harrigan, S., Hannaford, J., Muchan, K., & Marsh, T. J. (2018). Designation and trend analysis of the updated UK Benchmark Network of river flow stations: The UKBN2 data set. *Hydrology Research*, *49*(2), 552–567. https://doi.org/10.2166/nh.2017.058

Hermanson, L., Smith, D., Seabrook, M., Bilbao, R., Doblas-Reyes, F., Tourigny, E., et al. (2022). WMO global annual to decadal climate update: A prediction for 2021–25. *Bulletin of the American Meteorological Society*, *103*(4), E1117–E1129. https://doi.org/10.1175/BAMS-D-20-0311.1

Huntingford, C., Marsh, T., Scaife, A. A., Kendon, E. J., Hannaford, J., Kay, A. L., et al. (2014). Potential influences on the United Kingdom's floods of winter 2013/14. *Nature Climate Change*, *4*(9), 769–777. https://doi.org/10.1038/nclimate2314

Karl, T. R., Arguez, A., Huang, B., Lawrimore, J. H., McMahon, J. R., Menne, M. J., et al. (2015). Possible artifacts of data biases in the recent global surface warming hiatus. *Science*, *348*(6242), 5–1472. https://doi.org/10.1126/science.aaa5632

Knutti, R., & Sedláček, J. (2012). Robustness and uncertainties in the new CMIP5 climate model projections. *Nature Climate Change*, *3*(4), 369–373. https://doi.org/10.1038/nclimate1716

Kushnir, Y., Scaife, A. A., Arritt, R., Balsamo, G., Boer, G., Doblas-Reyes, F., et al. (2019). Towards operational predictions of the near-term climate. *Nature Climate Change*, *9*(2), 94–101. https://doi.org/10.1038/s41558-018-0359-7

Meehl, G. A., Goddard, L., Murphy, J., Stouffer, R. J., Boer, G., Danabasoglu, G., et al. (2009). Decadal prediction: Can it Be skillful? *Bulletin of the American Meteorological Society*, *90*(10), 1467–1486. https://doi.org/10.1175/2009BAMS2778.1

Morice, C. P., Kennedy, J. J., Rayner, N. A., & Jones, P. D. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *Journal of Geophysical Research*, *117*(D8), D08101. https://doi.org/10.1029/2011JD017187

Neri, A., Villarini, G., Salvi, K. A., Slater, L. J., & Napolitano, F. (2019). On the decadal predictability of the frequency of flood events across the U.S. Midwest. *International Journal of Climatology*, *39*(3), 1796–1804. https://doi.org/10.1002/joc.5915

Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *54*(3), 507–554. https://doi.org/10.1111/j.1467-9876.2005.00510.x

Robson, A. J. (2002). Evidence for trends in UK flooding. *Philosophical Transactions of the Royal Society of London,Series A: Mathematical, Physical and Engineering Sciences*, *360*(1796), 1327–1343. https://doi.org/10.1098/rsta.2002.1003

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

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Ziese, M., & Rudolf, B. (2014). GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theoretical and Applied Climatology*, *115*(1–2), 15–40. https://doi.org/10.1007/s00704-013-0860-x

Shepherd, T. G. (2014). Atmospheric circulation as a source of uncertainty in climate change projections. *Nature Geoscience*, *7*(10), 703–708. https://doi.org/10.1038/ngeo2253

Simpson, I. R., & Jones, P. D. (2014). Analysis of UK precipitation extremes derived from Met Office gridded data: Analysis of UK precipitation extremes. *International Journal of Climatology*, *34*(7), 2438–2449. https://doi.org/10.1002/joc.3850

Slater, L., Arnal, L., Boucher, M.-A., Chang, A. Y.-Y., Moulds, S., Murphy, C., et al. (2022). Hybrid forecasting: Using statistics and machine learning to integrate predictions from dynamical models. Hydrology and Earth System Science Data. [preprint]. https://doi.org/10.5194/hess-2022-334

Slater, L. J., & Villarini, G. (2018). Enhancing the predictability of seasonal streamflow with a statistical-dynamical approach. *Geophysical Research Letters*, *45*(13), 6504–6513. https://doi.org/10.1029/2018GL077945

Smith, D. M., Eade, R., Scaife, A. A., Caron, L.-P., Danabasoglu, G., DelSole, T. M., et al. (2019). Robust skill of decadal climate predictions. *Npj Climate and Atmospheric Science*, *2*(1), 13. https://doi.org/10.1038/s41612-019-0071-y

Smith, D. M., Scaife, A. A., Boer, G. J., Caian, M., Doblas-Reyes, F. J., Guemas, V., et al. (2013). Real-time multi-model decadal climate predictions. *Climate Dynamics*, *41*(11–12), 2875–2888. https://doi.org/10.1007/s00382-012-1600-0

Smith, D. M., Scaife, A. A., Eade, R., Athanasiadis, P., Bellucci, A., Bethke, I., et al. (2020). North Atlantic climate far more predictable than models imply. *Nature*, *583*(7818), 796–800. https://doi.org/10.1038/s41586-020-2525-0

Svensson, C., Brookshaw, A., Scaife, A. A., Bell, V. A., Mackay, J. D., Jackson, C. R., et al. (2015). Long-range forecasts of UK winter hydrology. *Environmental Research Letters*, *10*(6), 064006. https://doi.org/10.1088/1748-9326/10/6/064006

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An Overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

Villarini, G., Serinaldi, F., Smith, J. A., & Krajewski, W. F. (2009). On the stationarity of annual flood peaks in the continental United States during the 20th century. *Water Resources Research*, *45*(8), W08417. https://doi.org/10.1029/2008WR007645

Villarini, G., & Strong, A. (2014). Roles of climate and agricultural practices in discharge changes in an agricultural watershed in Iowa. *Agriculture, Ecosystems & Environment*, *188*, 204–211. https://doi.org/10.1016/j.agee.2014.02.036

Ward, N., & Conway, D. (2020). Applications of interannual-to-decadal climate prediction: An exploratory discussion on rainfall in the Sahel region of Africa. *Climate services*, *18*, 100170. https://doi.org/10.1016/j.cliser.2020.100170

West, H., Quinn, N., & Horswell, M. (2019). Regional rainfall response to the North Atlantic Oscillation (NAO) across Great Britain. *Hydrology Research*, *50*(6), 1549–1563. https://doi.org/10.2166/nh.2019.015

Wilby, R. L., & Quinn, N. W. (2013). Reconstructing multi-decadal variations in fluvial flood risk using atmospheric circulation patterns. *Journal of Hydrology*, *487*, 109–121. https://doi.org/10.1016/j.jhydrol.2013.02.038

References From The Supporting Information

Chikamoto, Y., Kimoto, M., Ishii, M., Mochizuki, T., Sakamoto, T. T., Tatebe, H., et al. (2013). An overview of decadal climate predictability in a multi-model ensemble by climate model MIROC. *Climate Dynamics*, *40*(5–6), 1201–1222. https://doi.org/10.1007/s00382-012-1351-y

Counillon, F., Keenlyside, N., Bethke, I., Wang, Y., Billeau, S., Shen, M. L., & Bentsen, M. (2016). Flow-dependent assimilation of sea surface temperature in isopycnal coordinates with the Norwegian Climate Prediction Model. *Tellus A: Dynamic Meteorology and Oceanography*, *68*(1), 32437. https://doi.org/10.3402/tellusa.v68.32437

Haarsma, R., Acosta, M., Bakhshi, R., Bretonnière, P.-A., Caron, L.-P., Castrillo, M., et al. (2020). HighResMIP versions of EC-Earth: EC-Earth3P and EC-Earth3P-HR–description, model computational performance and basic validation. *Geoscientific Model Development*, *13*(8), 3507–3527. https://doi.org/10.5194/gmd-13-3507-2020

Jordan, A., Krüger, F., & Lerch, S. (2019). Evaluating probabilistic forecasts with scoringRules. *Journal of Statistical Software*, *90*(12). https://doi.org/10.18637/jss.v090.i12

Karl, T. R., Arguez, A., Huang, B., Lawrimore, J. H., McMahon, J. R., Menne, M. J., et al. (2015). Possible artifacts of data biases in the recent global surface warming hiatus. *Science*, *348*(6242), 5–1472. https://doi.org/10.1126/science.aaa5632

Kharin, V. V., Boer, G. J., Merryfield, W. J., Scinocca, J. F., & Lee, W.-S. (2012). Statistical adjustment of decadal predictions in a changing climate. *Geophysical Research Letters*, *39*(19), L19705. https://doi.org/10.1029/2012GL052647

Mochizuki, T., Chikamoto, Y., Kimoto, M., Ishii, M., Tatebe, H., Komuro, Y., et al. (2012). Decadal prediction using a recent series of MIROC global climate models. *Journal of the Meteorological Society of Japan. Series II*, *90A*(0), 373–383. https://doi.org/10.2151/jmsj.2012-A22

Müller, W. A., Baehr, J., Haak, H., Jungclaus, J. H., Kröger, J., Matei, D., et al. (2012). Forecast skill of multi-year seasonal means in the decadal prediction system of the Max Planck Institute for Meteorology. *Geophysical Research Letters*, *39*(22), L22707. https://doi.org/10.1029/2012GL053326

Pohlmann, H., Müller, W. A., Bittner, M., Hettrich, S., Modali, K., Pankatz, K., & Marotzke, J. (2019). Realistic Quasi-Biennial oscillation variability in historical and decadal hindcast simulations using CMIP6 forcing. *Geophysical Research Letters*, *46*(23), 14118–14125. https://doi.org/10.1029/2019GL084878

Sospedra-Alfonso, R., & Boer, G. J. (2020). Assessing the impact of initialization on decadal prediction skill. *Geophysical Research Letters*, *47*(4), e2019GL086361. https://doi.org/10.1029/2019GL086361

Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., et al. (2019). The Canadian Earth System Model version 5 (CanESM5.0.3). *Geoscientific Model Development*, *12*(11), 4823–4873. https://doi.org/10.5194/gmd-12-4823-2019

Wang, Y., Counillon, F., Bethke, I., Keenlyside, N., Bocquet, M., & Shen, M. (2017). Optimising assimilation of hydrographic profiles into isopycnal ocean models with ensemble data assimilation. *Ocean Modelling*, *114*, 33–44. https://doi.org/10.1016/j.ocemod.2017.04.007

Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., et al. (2018). The met office global coupled model 3.0 and 3.1 (GC3.0 and GC3.1) configurations. *Journal of Advances in Modeling Earth Systems*, *10*(2), 357–380. https://doi.org/10.1002/2017MS001115

Yang, X., Rosati, A., Zhang, S., Delworth, T. L., Gudgel, R. G., Zhang, R., et al. (2013). A predictable AMO-like pattern in the GFDL fully coupled ensemble initialization and decadal forecasting system. *Journal of Climate*, *26*(2), 650–661. https://doi.org/10.1175/JCLI-D-12-00231.1

Yeager, S. G., Danabasoglu, G., Rosenbloom, N. A., Strand, W., Bates, S. C., Meehl, G. A., et al. (2018). Predicting near-term changes in the Earth system: A large ensemble of initialized decadal prediction simulations using the community Earth System Model. *Bulletin of the American Meteorological Society*, *99*(9), 1867–1886. https://doi.org/10.1175/BAMS-D-17-0098.1

19448007, 2023, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022GL100650 by Test, Wiley Online Library on [06/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License