Projections of northern hemisphere extratropical climate underestimate internal variability and associated uncertainty

Article

Published Version

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

ARTICLE

Check for updates

Projections of northern hemisphere extratropical climate underestimate internal variability and associated uncertainty

Christopher H. O'Reilly [1✉], Daniel J. Befort[2], Antje Weisheimer [3,4], Tim Woollings[2], Andrew Ballinger [5] & Gabriele Hegerl [5]

Internal climate variability will play a major role in determining change on regional scales under global warming. In the extratropics, large-scale atmospheric circulation is responsible for much of observed regional climate variability, from seasonal to multidecadal timescales. However, the extratropical circulation variability on multidecadal timescales is systematically weaker in coupled climate models. Here we show that projections of future extratropical climate from coupled model simulations significantly underestimate the projected uncertainty range originating from large-scale atmospheric circulation variability. Using observational datasets and large ensembles of coupled climate models, we produce synthetic ensemble projections constrained to have variability consistent with the large-scale atmospheric circulation in observations. Compared to the raw model projections, the synthetic observationally-constrained projections exhibit an increased uncertainty in projected 21st century temperature and precipitation changes across much of the Northern extratropics. This increased uncertainty is also associated with an increase of the projected occurrence of future extreme seasons.

---

[1] Department of Meteorology, University of Reading, Reading, UK. [2] Atmospheric, Oceanic & Planetary Physics, University of Oxford, Oxford, UK. [3] NCAS-Climate, Atmospheric, Oceanic & Planetary Physics, University of Oxford, Oxford, UK. [4] European Centre for Medium-range Weather Forecasts, Reading, UK. [5] School of Geosciences, University of Edinburgh, Edinburgh, UK. ✉email: c.h.oreilly@reading.ac.uk

nternal variability has a strong influence on decadal-to-multidecadal climate variability and trends, particularly on regional scales in the extratropics[1–3]. The dominance of internal variability explains why regional temperatures have exhibited markedly different trends on decadal timescales, despite persistent global warming due to increasing greenhouse gas concentrations since the pre-industrial period (e.g. Eurasian winter cooling[4]). The dominant source of internal variability for continental climate in the extratropics is large-scale atmospheric circulation. For example, the extratropical warming over land during the Northern Hemisphere winter over the later part of the twentieth century was enhanced substantially by anomalies in the large-scale atmospheric circulation and their associated impact on surface-air temperature[5–7]. Internal variability in the large-scale atmospheric circulation is expected to make a similarly large contribution to the climate we will experience in the future[8]. Over the coming decades, trends in extratropical temperature and precipitation are expected to be dominated by internal variability, particularly over North America and Eurasia[2,9]. Therefore, to provide useful projections of extratropical climate over the twenty-first century, it is crucial that models accurately represent the contribution from the internal variability associated with large-scale atmospheric circulation[10].

Recent studies, however, have highlighted some disparities between the observed large-scale circulation variability and that seen in current climate models. In particular, over the North Atlantic sector during winter, there have been significant multi-decadal fluctuations in the leading mode of large-scale atmospheric circulation variability, the North Atlantic Oscillation[7,11,12], and related behaviour in the strength of the North Atlantic jetstream[13,14]. Several studies have argued that the weak multidecadal atmospheric circulation variability reflects the apparently relatively weak response of the atmospheric circulation to variability in North Atlantic sea surface temperatures (SSTs)[15–20]. It has also been suggested that the response of stratospheric polar vortex to Atlantic SSTs and the subsequent influence on the extratropical large-scale circulation in the troposphere is poorly represented in climate models, which may contribute to the weak atmospheric circulation variability[21]. There is also substantial multidecadal variability in the large-scale circulation during the summer season in the North Atlantic sector, which has exhibited a clear influence on the variability of European summer climate[22–24] and—similarly to the winter season—seems to be too weakly represented in coupled climate models[25].

While there has been much attention given to the mechanisms of multidecadal variability in the atmospheric large-scale circulation, relatively little attention has been given to the implications for future projections[26]. The apparent disparities between the large-scale circulation variability in observations and models are discussed in the recent IPCC Sixth Assessment Report[27] (Chapter 3); however, the implications for future projections are not clearly considered. In this study, we use a novel method to produce climate projections using observationally constrained estimates of large-scale circulation variability.

Results

Multidecadal circulation variability in observations and models. We begin our investigation by examining the multidecadal variability of sea-level pressure (SLP) in observational data sets. SLP is a useful proxy for the atmospheric circulation because outside of the tropics the large-scale flow is effectively in geostrophic balance, such that SLP provides a direct measure of the near-surface winds. Large-scale SLP anomalies are also typically associated with wind anomalies higher in the troposphere[17,28], so

analysis of the SLP fields implicitly reflects variability throughout the troposphere. For example, the only atmospheric observations that are used to constrain the 20th Century Reanalysis are surface pressure observations. Despite this, the 20th Century Reanalysis closely matches the variability of the upper-tropospheric extratropical circulation seen in more comprehensive reanalysis products that assimilate upper-atmosphere observations[29]. An advantage of analysing SLP here is that there is a long and extensive observational record[30], with data from ship observations over the ocean and long station records over land[31].

Figure 1a, b shows the multidecadal variability of the winter and summer SLP in observations (shown here for the HadSLP2 data set), over the historical period 1901–2005 (see 'Methods'). Multidecadal variability is defined here as the standard deviation of the 20-year running means, normalised by the interannual standard deviation. The results are not sensitive to the length of the moving average timescale, and similar behaviour is found for decadal (i.e. 10-year) and 30-year running means (Supplementary Figs. 1 and 3). There are large areas with substantial multidecadal variability in winter SLP over the North Atlantic sector, significantly more than would be expected from a simple white-noise process as shown by the black contours in Fig. 1a, b (this pattern is consistent across multiple observational data sets; Supplementary Figs. 1–3). There are also areas with substantial multidecadal variability for the summer season. The regions with high levels of summer multidecadal variability centred over Northern Europe and North America are particularly consistent across multiple observational data sets (Supplementary Figs. 1–3).

The analysis of multidecadal SLP variability was repeated for a large ensemble of 246 historical simulations from 54 different coupled climate models in the Coupled Model Intercomparison Project (CMIP) 5 and 6 archive[32,33]. The CMIP ensemble displays relatively weak multidecadal variability everywhere (Fig. 1e, f), with the multidecadal variability from the observations falling in the top 5% of the CMIP distribution and higher over large areas around the North Atlantic sector in both winter and summer (similar results are found in the separate CMIP5 and CMIP6 ensembles, Supplementary Fig. 4). Internal variability can be difficult to separate from the forced responses in the multi-model CMIP ensemble as different models have different forced responses that require large single-model ensembles to determine[34]. Therefore, in the analysis that follows we use data from the Max Planck Institute for Meteorology Grand Ensemble[35] (MPI-GE), which is a large ensemble of coupled climate model simulations consisting of 99 members (see 'Methods'). The MPI-GE demonstrates similarly weak multidecadal SLP variability as the CMIP ensemble in both winter and summer (Fig. 1g, h). Taken at face value, the CMIP ensemble and MPI-GE both seem to be inconsistent with the observed multidecadal SLP variability, exhibiting a similar deficiency with low variability over the North Atlantic compared with observations (Fig. 1a, b and Supplementary Fig. 1–3).

To examine the nature of the multidecadal SLP variability seen over the North Atlantic sector in observations, we performed an empirical orthogonal function (EOF) analysis using the regions shown in Fig. 1 (see 'Methods'). Decomposing the interannual SLP data into the leading EOFs reveals that, in winter, the leading mode of variability (EOF1)—often referred to as the North Atlantic Oscillation[36]—exhibits substantial multidecadal variability in all of the observational data sets (Supplementary Fig. 5), as highlighted in previous studies[7,13,37,38]. However, there is also substantial multidecadal variability in next two modes, EOF2 and EOF3. The contribution of the leading EOFs to the multidecadal SLP variability seen in the observations is assessed by replacing the principal component (PC) timeseries of the leading EOFs with random white-noise timeseries with the same standard deviation
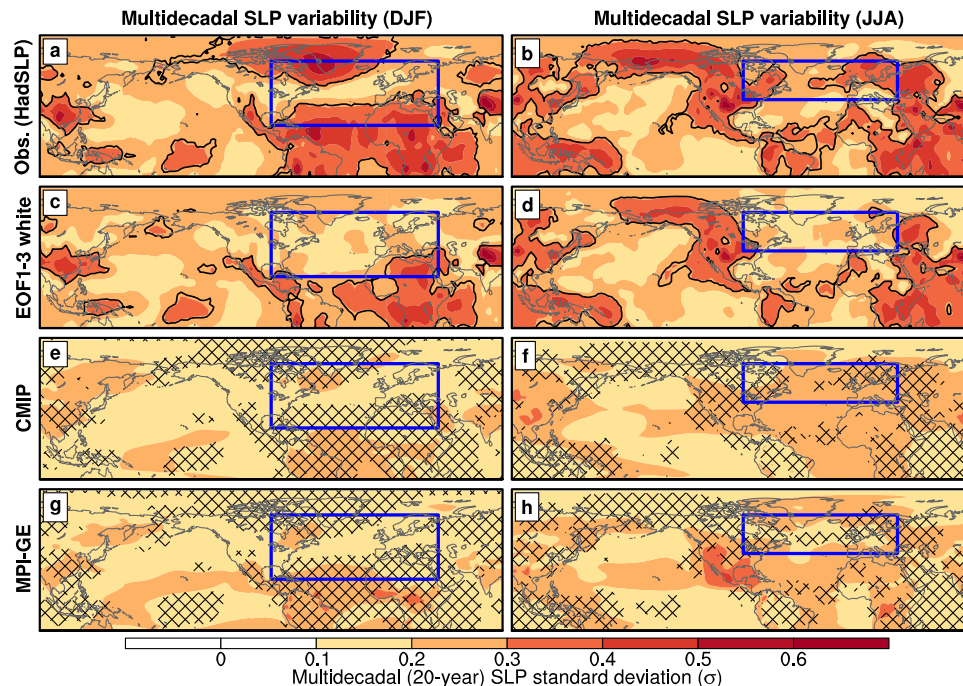
Fig. 1 Multidecadal large-scale circulation variability in observations and models. a The multidecadal variability of winter (DJF) sea-level pressure in the HadSLP2 observational data set over the period 1901–2005, where the SLP variability has been normalised by the interannual variability at each grid point. The multidecadal component is calculated using 20-year running mean. **c** The average multidecadal variability of winter sea-level pressure in a synthetic data set in which the first three EOFs of interannual SLP variability over the North Atlantic sector in the HadSLP2 data set are replaced with random white-noise timeseries (the average of 10,000 random realisations is shown). **e** The ensemble mean of the multidecadal large-scale circulation variability in CMIP model realisations (from both CMIP5 and CMIP6) over the period 1901–2005 (the multidecadal SLP variability was calculated in each model realisation individually, prior to taking the ensemble mean). **g** As in **e** but for the 99-member MPI-GE simulations over the period 1901–2005. **b, d, f, h** are as in **a, c, e, g**, respectively, but for the boreal summer season (JJA). The black contours in **a**–**d** show where the multidecadal variability is significantly different from that expected from a white-noise process (at the 95% level, estimated using a Monte Carlo test with 10,000 realisations). The black hatching in **e**–**h** show where the multidecadal variance in the HadSLP2 observations for the respective seasons exceeds the 95th percentile of each of the ensemble distributions. The blue boxes show the regions used to calculate the SLP EOF calculations.

and repeating this process 10,000 times. The observed multidecadal wintertime SLP variability in the North Atlantic sector can be largely attributed to the first three EOFs (Fig. 1c), while higher-order EOFs make no clear contribution (Supplementary Fig. 6). During the summer season, the two leading EOFs of SLP variability in the North Atlantic sector both exhibit substantial multidecadal variability (Supplementary Figs. 5 and 7), with EOF3 making relatively small contribution to the observed multidecadal variability. Similar to the winter season, the overall observed multidecadal summer SLP variability in the North Atlantic sector can be largely attributed to the first three EOFs (Fig. 1d and Supplementary Fig. 6). In contrast to the observational data sets, in the CMIP ensemble and MPI-GE the variability of the leading EOF modes are not significantly different from random white-noise processes (Supplementary Fig. 5), consistent with maps of multidecadal SLP variability (Fig. 1e–h).

Generating synthetic observationally constrained projections. The discrepancies between the observed multidecadal SLP variability and that seen in climate models are substantial—but what are the implications for the climate projections made using these models? To investigate this, we have generated synthetic temperature and precipitation projections that are consistent with the observed large-scale circulation variability (the key steps are outlined in the following; for full details, see 'Methods'). These projections are based on the MPI-GE climate model simulations, performed from 1901 to 2100 for various forcing scenarios[35].

First, the signature of SLP variability was subtracted from temperature and precipitation fields in the raw 99-member MPI-GE ensemble (MPI-GE-raw hereafter) using linear regression. Only the first three EOFs of SLP were used, as these were found to make the dominant contributions to the multidecadal SLP variability (i.e. Fig. 1). A random member is selected from MPI-GE-raw and the temperature/precipitation variability associated with the first three EOFs of SLP are removed. This variability is then replaced with the same patterns of temperature/precipitation anomalies but multiplied by a random surrogate PC timeseries that is constructed to have the same spectral characteristics as the observations (Supplementary Fig. 7). The surrogate PC timeseries, therefore, tend to have more power on multidecadal timescales (Supplementary Fig. 5), though the overall standard deviation is unchanged on average, across the ensemble. This process was repeated 10,000 times to produce a synthetic observationally constrained 10,000-member ensemble (MPI-GE-obs hereafter). To include the influence of observational uncertainty, the surrogate PC timeseries were calculated from four different observational data sets, with each contributing equally to produce the 10,000 members in MPI-GE-obs. In the analysis that follows, we compare the raw 99-member ensemble, MPI-GE-raw, with the synthetic 10,000-member observationally constrained ensemble, MPI-GE-obs.

A key assumption in producing MPI-GE-obs is that the future large-scale circulation variability will have the same characteristics as the large-scale circulation variability that we have observed in the past. This is highly uncertain, of course, but our

aim here is to estimate what we might expect the variability of the large-scale atmospheric circulation to contribute in projections of future climate. We also prescribe there to be no forced large-scale circulation changes in MPI-GE-obs (at least in ones that project onto the leading EOF patterns). To some extent, this is justified by the modest forced changes in the large-scale circulation that we see in MPI-GE-raw but, usefully, this approach allows us to interpret any differences in the median projections as being driven by forced changes in large-scale circulation in MPI-GE-raw. The forced changes are shown to be very modest in the results that follow; however, there is the potential that the forced future circulation changes are underestimated by the model[39,40].

Impact on projected climate in the Northern extratropics. To analyse the influence the observationally constrained large-scale circulation variability on future climate projections, we examine the changes in surface-air temperature and precipitation for the mid-century period (2041–2060) with respect to a present-day baseline period (1995–2014) in MPI-GE-raw and MPI-GE-obs. The median projected changes under the Representative Concentration Pathway (RCP) 4.5 scenario for the boreal winter season consist of widespread warming across the extratropical regions and is almost identical in MPI-GE-raw and MPI-GE-obs, demonstrating that changes in the large-scale atmospheric circulation are not responsible for the distribution of average temperature changes (Fig. 2a, b). There are, however, substantial differences in interquartile range (i.e. 25–75%) of the projected changes, with MPI-GE-obs exhibiting a larger range over most of Europe and North America (Fig. 2c, d). The difference in the interquartile range between MPI-GE-obs and MPI-GE-raw is plotted in Fig. 3a and shows that the interquartile range increases by over 50% across Northern Europe, North America and Mediterranean regions, almost doubling in some regions. Therefore, for many extratropical regions, the MPI-GE-raw ensemble underestimates the contribution of large-scale atmospheric circulation to the uncertainty in climate projections for the mid-twenty-first century.

Distributions of the projected regional change of temperature and precipitation for the winter and summer seasons are summarised in Fig. 4. For winter precipitation, the MPI-GE-obs shows substantial increases in the interquartile range compared with MPI-GE-raw, most notably over Northern Europe and Mediterranean regions (Figs. 3b and 4b). For the Mediterranean region, there is more than a doubling of the likely range, with more substantial drying becoming much more likely in MPI-GE-obs. The broadening of the distributions is also clear in the tails of

the distributions, where mid-century changes in the winter temperature and precipitation that would be deemed highly unlikely are now well within the range of likely outcomes in the presence of observationally constrained large-scale circulation variability (i.e. MPI-GE-obs).

The differences between the projections for changes in summer climate are relatively muted compared to the winter season. The reason for this is that the SLP EOFs exhibit a stronger relationship with temperature and precipitation anomalies in the summer season compared to the winter season (Supplementary Figs. 8–10). Nonetheless, there are still significant increases in the interquartile range of the projected summertime precipitation changes over Northern Europe (Figs. 3d and 4d). There are also significant changes in the distribution of projected temperatures over the East North America region (Figs. 3c and 4c).

From the analysis of the MPI-GE-obs projections, it is clear that, with large-scale circulation variability that is consistent with that seen in the observations, there is substantially more uncertainty in climate projections for many regions in the extratropics. We have shown this for the mid-century period 2041–2060 but analysis of other 20-year periods demonstrate similar increases in the spread of the distributions throughout the twenty-first century (Supplementary Figs. 11 and 12). Here we have presented results for 20-year future climate periods because these are commonly considered in climate assessment reports[41]; however, the uncertainty of the regional climate projections are similarly found to increase substantially in the Northern extratropics for 10-, 30- and 40-year future periods (Supplementary Figs. 13–18). Therefore, over any meaningful climatological averaging period in the future we expect there to be a substantial contribution from internal variability that is underestimated in regional climate model projections.

Impact on future extreme seasons. As well as influencing the distribution of future regional climate change, it is also possible that the characteristics of the large-scale circulation variability considered here can influence extreme events on seasonal timescales. By design, the distributions of seasonal mean anomalies measured across the MPI-GE-raw and MPI-GE-obs ensembles have equal variance (see 'Methods'). However, the characteristics of when particular extreme seasons occur in individual model realisations can be quite different. Here we define an extreme season as the highest or lowest seasonal mean value over the baseline climate period, 1995–2014, representing a 1/20 year event based on a present-day climate period, which one could also
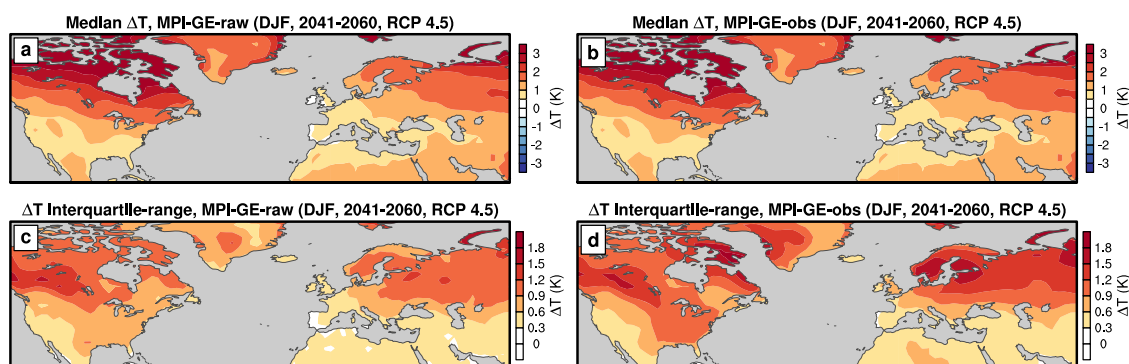
Fig. 2 Projection of mid-century climate change for boreal winter. a Median projection of the winter (DJF) surface-air temperature change in the period 2041–2060, with respect to the 1995–2014 baseline period, in the 99-member MPI-GE-raw ensemble (under the RCP 4.5 emissions scenario). **b** As is **a** but for the median of the synthetic 10,000-member MPI-GE-obs ensemble. **c** Interquartile range (i.e. 25–75% range) of the 2041–2060 winter surface-air temperature change in the 99-member MPI-GE-raw ensemble. **d** As is **c** but for the interquartile range of the synthetic 10,000-member MPI-GE-obs ensemble.
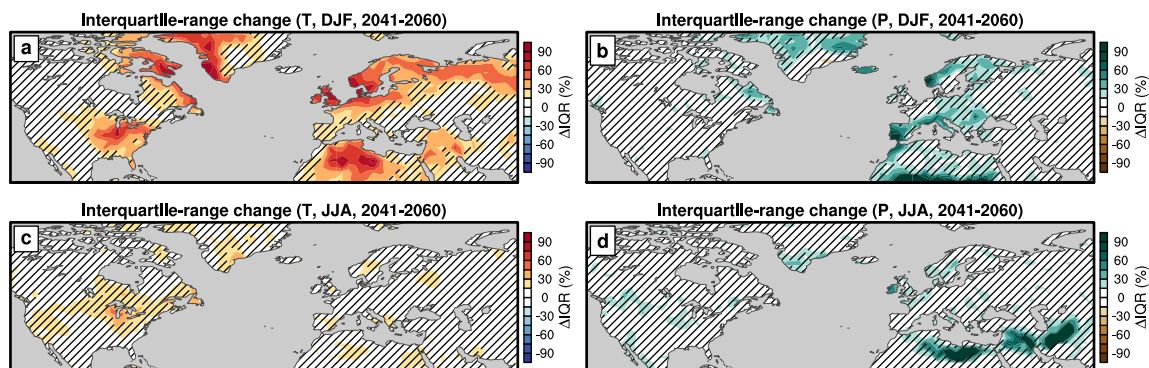
Fig. 3 Change in projection uncertainty for mid-century climate change. a The change in the 2041–2060 winter (DJF) surface-air temperature interquartile range in the MPI-GE-obs ensemble, with respect to the interquartile range in the MPI-GE-raw ensemble. **b** as in **a** but for the winter precipitation. **c** as in **a** but for the summer surface-air temperature. **d** as in **a** but for the summer precipitation. The black hatching indicates regions where the difference in the interquartile range between MPI-GE-obs and MPI-GE-raw is not statistically significant (at the 95% level, based on a Monte Carlo resampling, see 'Methods' for further details).
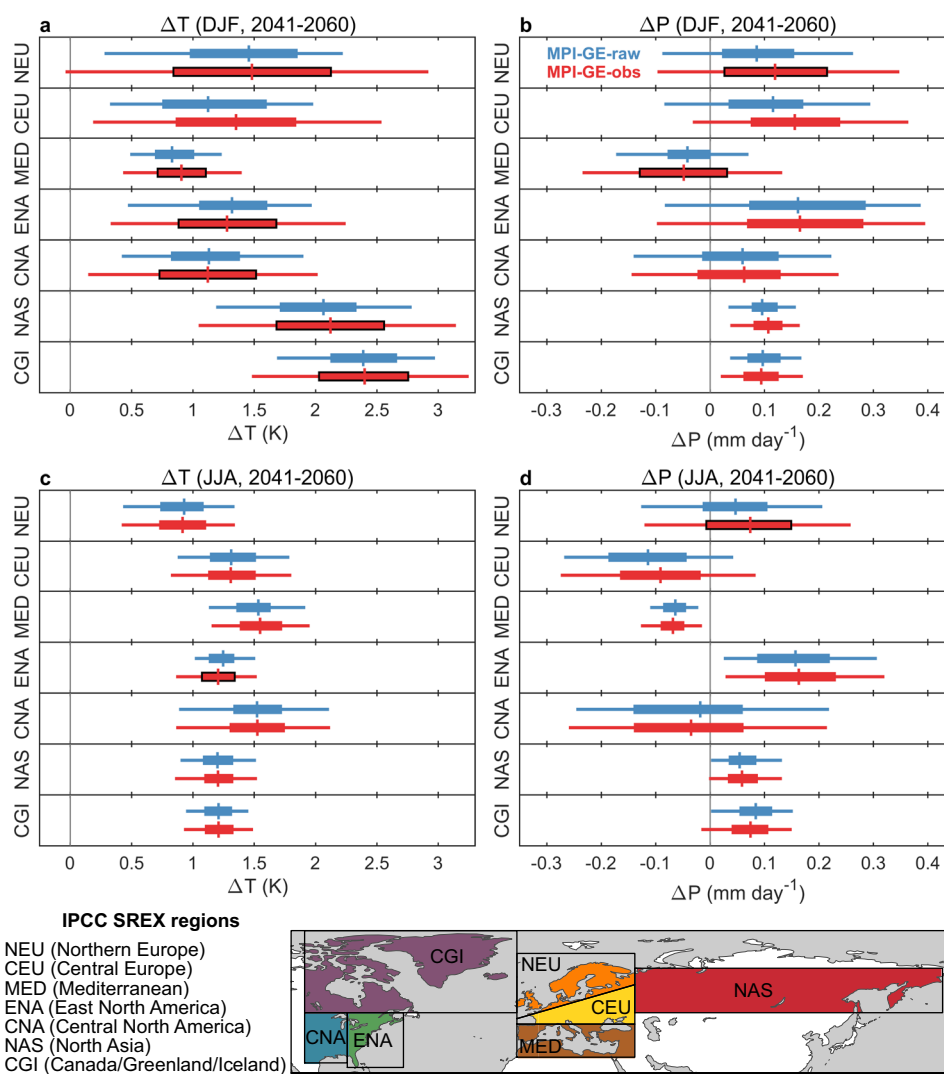

Fig. 4 Projected regional changes and the role of large-scale atmospheric circulation variability. Distributions of projected regional changes for the 2041–2060 mean from a 1995–2014 baseline period for: **a** Winter surface-air temperature, **b** winter precipitation, **c** summer surface-air temperature and **d** summer precipitation. The thin horizontal lines show the 5–95% range, the rectangular boxes show the interquartile range (i.e. 25–75%) and the vertical lines show the median change. Distributions are shown for the MPI-GE-raw ensemble (in blue) and the MPI-GE-obs ensemble (in red). The black boxes show where the interquartile range of the 2041–2060 projection in the MPI-GE-obs ensemble is significantly different from the respective interquartile range in the MPI-GE-raw ensemble (at the 95% level, based on a Monte Carlo resampling, see 'Methods' for further details).
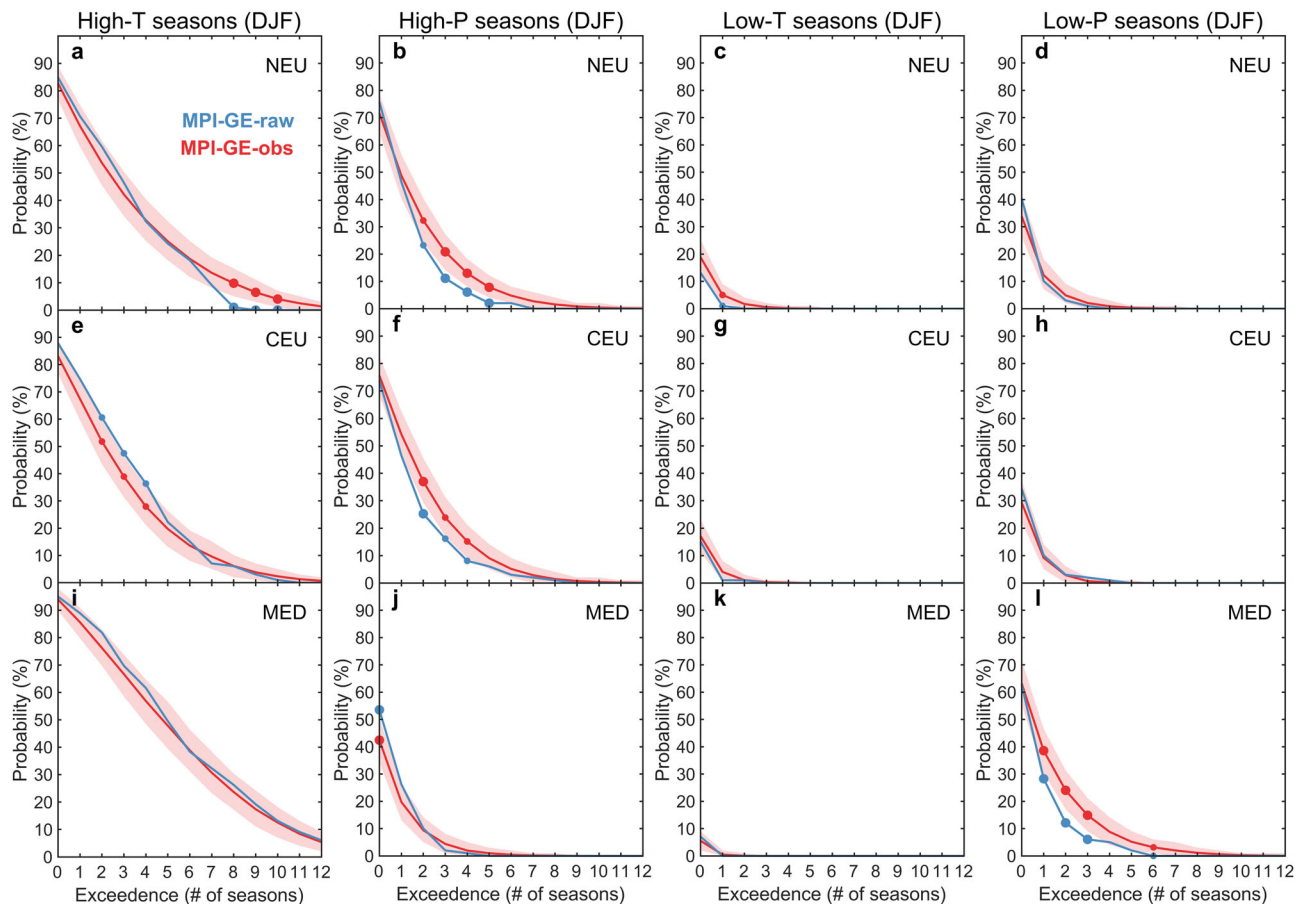
Fig. 5 Changes in the projected occurrence of extreme seasons for European winter. The probability of exceeding a number of extreme seasons over the period 2041–2060, where extreme seasons are defined as the highest/lowest seasonal mean value over the baseline period 1995–2014. The panels show probability of exceeding a number of extreme seasons in the Northern Europe, Central Europe and Mediterranean regions for: **a**, **e**, **i** high temperatures; **c**, **g**, **k** low temperatures; **b**, **f**, **j** high precipitation; **d**, **h**, **l** low precipitation. Blue lines show the probability of exceeding a given number of extreme seasons in the 99-member MPI-GE-raw ensemble, while the red lines show the same for the MPI-GE-obs ensemble. The shaded regions around the MPI-GE-obs ensemble shows the 5–95% range of the MPI-GE-obs when only 99-members are resampled at random (10,000 times). Large dots show where the probabilities in the MPI-GE-raw and MPI-GE-obs ensembles are significantly different at the 95% level and the small dots show significance at the 90% level (based on a Monte Carlo resampling, see 'Methods' for further details).

estimate in the observational record. The number of extreme seasons in a future climate period is then calculated in each ensemble member.

The occurrence rate of extreme seasons for European winters are shown in Fig. 5 for the mid-century period, 2041–2060. The occurrence of a greater number of extreme seasons within the 20-year window is larger in MPI-GE-obs than in MPI-GE-raw in a number of instances, particularly in the tails of the distribution, whereas the occurrence of relatively few events tends to be higher in MPI-GE-raw. The change in occurrence rate is most pronounced over Northern Europe, with a >10% probability of exceeding 8 seasons with extreme high temperatures in the MPI-GE-obs data set, whereas the probability of an equally extreme realisation occurring in MPI-GE-raw is about 1% (Fig. 5a). Similarly, increases in the numbers of extremely wet winters over the mid-century period are found to be significantly more likely in MPI-GE-obs (Fig. 5b, f). Another notable feature is that the occurrence of having a number of extremely dry Mediterranean winters over the mid-century period in the future is significantly higher in MPI-GE-obs (Fig. 5l). In the summer and the other extratropical regions, there are less clear differences in the occurrence of extreme seasons (see Supplementary Information). The higher probability of a large number of extreme winter seasons occurring in a future period is related to the relatively

large variability on multidecadal timescales in the MPI-GE-obs, which is absent in MPI-GE-raw. An explanation for this is that the influence of low-frequency variability in the large-scale circulation can set a relatively high background anomaly over a 20-year period, meaning that the year-to-year variability superimposed onto this can produce clusters of extreme seasons. In MPI-GE-raw, however, there is relatively little low-frequency variability so the occurrence of future extreme seasons in a given year is largely independent of the surrounding years.

Discussion and wider implications

The analysis presented here demonstrates that factoring the influence of the observed variability of the large-scale atmospheric circulation into future climate projections substantially increases the uncertainty arising from internal variability. The current generation of coupled climate models, which are used to produce future climate projections, are therefore likely to underestimate the contribution of internal variability in the extratropics. There are some significant differences in the projections of the MPI-GE-obs and MPI-GE-raw ensembles in the summer season around the North Atlantic sector but the influence of the observed large-scale atmospheric circulation on future projections is largest

during the winter season, influencing most regions in the Northern extratropics.

It is important to note, however, that the synthetic ensemble method used here likely misses some feedback mechanisms that will contribute to extratropical climate variability. One example is the increase in the uncertainty in Mediterranean precipitation in the winter season (i.e. Fig. 4b). Studies show that a wintertime precipitation deficit in the Mediterranean makes European heat waves more likely during the following summer[42,43]; however, feedbacks relating to this are not captured in the synthetic MPI-GE-obs ensemble, in which winter and summer variability are effectively decoupled. Another example of a missing feedback is that North Atlantic multidecadal SST variability has been shown to be driven in part by low-frequency variability of the wintertime large-scale atmospheric circulation[20,44–48]. In observational analysis, the multidecadal SST variability in the North Atlantic has been implicated for low-frequency climate variability during the summer season[23,49,50], including the occurrence of heat waves[25,51]; these possible feedbacks are also not captured in the synthetic MPI-GE-obs ensemble. Each of these feedbacks would be expected to further increase the uncertainty in summer climate projections over the North Atlantic sector.

The observed large-scale circulation variability that is used to produce the MPI-GE-obs projections is subject to substantial sampling uncertainty, particularly at the lower frequencies, due to the relatively short observational period. While there is clearly substantial variability on timescales of decades and longer, the precise magnitude of this variability is uncertain. However, there is evidence from early instrumental and proxy reconstructions that the large-scale circulation over the North Atlantic sector exhibits distinct variability on multidecadal timescales[52–54], in periods independent from the observational period considered here. Therefore, while the precise degree to which internal variability is underestimated is fairly uncertain, it seems clear that future climate projections using the current coupled climate models significantly underestimate the internal variability of the large-scale atmospheric circulation.

For future twenty-first century periods, the underestimation of the uncertainty due to large-scale atmospheric circulation is comparable with the structural uncertainty in the forced response[55,56]. An example of where this underestimation could be important is the recent literature considering the differing impacts of 1.5 and 2 °C of global warming[57]; the underestimation of internal variability in the extratropics implies that regional differences between 1.5 and 2 °C warming are likely to be somewhat overconfident. Furthermore, the increased uncertainty also raises questions about the treatment of internal variability in regional model projections[58]. The EURO-CORDEX ensemble[59], for example, use a relatively small subset of global coupled climate model simulations that, as has shown here, themselves underestimate the contribution of internal variability and this will be compounded in projections made using regional model ensembles. The increased projection uncertainty may also be important to factor into future risk assessment and decision making exercises.

Methods

Data sets. We analyse data from the MPI-GE, which is a 99-member ensemble comprising of a historical forcing simulation (over the period 1850–2005) and the same members then follow the RCP 4.5 scenario forcing (from 2006 to 2100)[35]. Here we use monthly averaged SLP, surface-air temperature (TAS) and precipitation output from MPI-GE.

We also analyse data from CMIP5[32] and CMIP6[33] ensembles. We use monthly averaged data from the historical forcing simulations over the common historical period (1901–2005), with 82 ensemble members from CMIP5 and 164 ensemble members from CMIP6.

We use four gridded observational SLP data sets:

- *HadSLP2*: A gridded observational data set provided on a 5° × 5° grid from 1850 to 2019[30]. HadSLP2 uses an optimal interpolation procedure using marine and land observations to reconstruct a gridded SLP field.
- *TREN*: The Trenberth SLP data set[60] (TREN) is provided on a 5° × 5° grid from 1899 to 2019. The TREN data set is a gridded analysis produced from a variety of historical SLP analysis maps. It is only provided on a grid north of 20°N and has missing values where the data was deemed unreliable. Seasonal mean anomalies were only used where no more than one constituent month was missing.
- *20CR*: The 20th Century Reanalysis v2 (20CR) is a reanalysis data set that is provided on a 2° × 2° grid from 1871 to 2012[29]. 20CR was produced by assimilating observations of surface pressure and SLP.
- *CERA-20C*: The ECMWF coupled climate reanalyses of the twentieth century, provided at 125 km horizontal resolution from 1901 to 2010[29]. CERA-20C was produced by assimilating observations of surface pressure, SLP and surface winds, along with ocean observations, into the coupled ECMWF model.

The observations and model data were compared (e.g. in Fig. 1) over the common historical period 1901–2005, which is covered by all observational data sets and historical forcing simulations.

All data sets were regridded to a common 2.5° × 2.5° resolution prior to further analysis.

EOF decomposition of observations and models. The large-scale circulation variability was decomposed using EOFs over the North Atlantic sector. The precise region during the winter (20°N–70°N, 90°W–40°E) was chosen as this has been commonly used to define the winter (DJF) North Atlantic Oscillation (shown by blue boxes in Fig. 1). For the summer season (JJA), a slightly smaller region was used (40°N–70°N, 90°W–30°E), corresponding the a region often used to define the summer North Atlantic Oscillation[22,61] (shown by blue boxes in Fig. 1), which omits data at lower latitudes due to some recorded discrepancies in a band across North Africa and Asia[62]. The EOF patterns are by definition the three modes that explain the most variance of the area-weighted SLP the North Atlantic sector. The EOF patterns were calculated from the HadSLP2 data set over the common historical period (1901–2005). To calculate the corresponding PC timeseries in the other observational data sets, the SLP anomalies from each of the other data sets were projected onto the EOF patterns calculated from the HadSLP2 data set. The PC timeseries of the EOF variability in the model data sets (i.e. CMIP5, CMIP6 and MPI-GE) was calculated individually for each ensemble member. First, SLP anomalies were defined by removing the mean over the common historical period (1901–2005) and the SLP anomalies were projected onto the EOF patterns calculated from the HadSLP2 observational data set. The resulting PC timeseries from the model (covering the period 1901–2100) were normalised over the common historical period (1901–2005) to be comparable with the observational data sets (e.g. Supplementary Fig. 5). The projection approach described here was used to ensure that the PC timeseries correspond to the same patterns in all the observational and model data sets (though tests calculating the EOF patterns from the different observational and model data sets leads to results that are qualitatively unchanged).

Generating synthetic ensembles with observationally constrained large-scale atmospheric circulation variability. To generate the synthetic observationally constrained ensembles (MPI-GE-obs), we first decompose the raw ensemble variables, X_{raw} (corresponding to the MPI-GE-raw ensemble in the main text), using the first three PC timeseries, as follows:

$$X_{raw}(x,y,t) = \sum_{n=1}^{3} a_n^{raw}(t) Y_n(x,y) + X_{residual}(x,y,t) \quad (1)$$

where $a_n^{raw}(t)$ are the PC timeseries calculated from the SLP data (see previous subsection of 'Methods') and $Y_n(x,y)$ are maps of the linear regression coefficients calculated between the (normalised) PC timeseries and X_{raw} over the historical period (1901–2005). The first three EOFs account for 74% of the variance in winter and 66% of the variance in summer (in the HadSLP2 data set), over their respective regions. $X_{residual}(x,y,t)$ denotes the variability that is not explained by the first three EOFs. The decomposition is performed for all 99 members individually.

The synthetic observationally constrained ensemble variables, X_{obs}, are produced as follows:

$$X_{obs}(x,y,t) = \sum_{n=1}^{3} a_n^{obs}(t) Y_n(x,y) + X_{residual}(x,y,t) \quad (2)$$

where $Y_n(x,y)$ and $X_{residual}(x,y,t)$ are selected from 1 of the 99 MPI-GE-raw members at random. The timeseries, $a_n^{obs}(t)$, are randomly generated surrogate PC timeseries calculated from the corresponding observational PC timeseries. This process was repeated 10,000 times to generate the MPI-GE-obs data set.

The key difference between the MPI-GE-raw and MPI-GE-obs ensembles, therefore, are the surrogate PC timeseries, $a_n^{obs}(t)$. To generate the surrogate each timeseries, we use the method of Theiler et al.[63]. First, the corresponding PC timeseries from one of the observational SLP data sets is selected and the discrete Fourier transform is computed. A random phase is added to each of the

components, and the inverse Fourier transform is taken to return a random timeseries with similar spectral characteristics to the observed PC timeseries. It is important to note that the surrogate method acts to constrain the timeseries across all timescales. The process is repeated to produce 10,000 sets of $a_n^{obs}(t)$, consisting of 2500 sets of surrogate PC timeseries from each of the observational SLPs data sets (i.e. HadSLP2, TREN, 20CR and CERA-20C). Power spectra from the observational PC timeseries, surrogate PC timeseries (i.e. $a_n^{obs}(t)$) and the raw model PC timeseries (i.e. $a_n^{raw}(t)$) are shown in Supplementary Fig. 7. Synthetic ensemble data, $X_{obs}(x, y, t)$, is calculated for precipitation and surface-air temperature using the surrogate PC timeseries (i.e. $a_n^{obs}(t)$) to produce the MPI-GE-obs ensemble.

In addition to MPI-GE-obs, we also tested a synthetic ensemble that was produced in a similar way but with the first three EOFs were replaced using EOF patterns of temperature/precipitation anomalies calculated from observations (MPI-GE-obsTELE—see Supplementary Methods and Supplementary Fig. 19). The conclusions drawn from this ensemble are not qualitatively different, so here we only present results from MPI-GE-obs in the main text.

One key assumption in the generation of MPI-GE-obs is that the temperature and precipitation anomalies associated with the leading EOFs are unchanged between the historical period and the future. This is not entirely obvious, as it has been documented that interannual variability exhibits some robust changes in future climate model simulations[64], which is also evident in some areas in the MPI-GE-raw ensemble (Supplementary Fig. 20). We tested the sensitivity to calculating the temperature/precipitation anomalies associated with the leading EOFs over the historical period by using a more complex approach in which the temperature/precipitation anomalies were calculated over a future mid-century period (2031–2070) in each respective ensemble member in MPI-GE-raw. These values for Y_n were used in the calculations of X_{obs} in Eq. (2) above for all years in the future period. The results of the test (shown in Supplementary Fig. 21) are very similar to those shown in Fig. 3, indicating that the results are insensitive to changes in the patterns of temperature and precipitation anomalies associated with the leading SLP EOFs.

Significance testing. To test the statistical significance of the difference between the 99-member MPI-GE-raw ensemble and the synthetic 10,000-member MPI-GE-obs ensemble, we used a Monte Carlo sampling technique. From the full MPI-GE-obs ensemble, 99 members were selected without replacement; the required statistic (e.g. the interquartile range) was then calculated from this subsample and recorded. This subsampling was repeated 10,000 times to estimate the uncertainty associated with only having 99 ensemble members. The statistic calculated from the MPI-GE-raw ensemble was compared with the distribution of the 10,000 random samples to ascertain the probability that MPI-GE-raw could have been selected from MPI-GE-obs.

Data availability
The coupled model data and observational data used in the study are all publicly available data sets. The MPI Grand Ensemble is available from a dedicated project page on ESGF (https://esgf-data.dkrz.de/projects/mpi-ge/). The CMIP5 and CMIP6 data are available on ESGF (https://esgf-node.llnl.gov/projects/esgf-llnl/). The HadSLP2 data set is available from the UK Met Office (https://www.metoffice.gov.uk/hadobs/hadslp2/). The CERA-20C data set is available from the Copernicus Climate Data Store (https://cds.climate.copernicus.eu/). The 20th Century Reanalysis is available from the NOAA website (https://psl.noaa.gov/data/20thC_Rean/). The Trenberth gridded SLP data set is available from the NCAR Research Data Archive (https://rda.ucar.edu/datasets/ds010.1/).

Code availability
The custom code used to create the MPI-GE-obs projections and the figure data has been uploaded to an open-access Zenodo repository[65] (https://doi.org/10.5281/zenodo.52115300).

References
1. Hawkins, E. & Sutton, R. The potential to narrow uncertainty in regional climate predictions. *Bull. Am. Meteorol. Soc.* **90**, 1095–1108 (2009).
2. Deser, C., Phillips, A., Bourdette, V. & Teng, H. Uncertainty in climate change projections: the role of internal variability. *Clim. Dyn.* **38**, 527–546 (2012).
3. Marotzke, J. & Forster, P. M. Forcing, feedback and internal variability in global temperature trends. *Nature* **517**, 565–570 (2015).
4. McCusker, K. E., Fyfe, J. C. & Sigmond, M. Twenty-five winters of unexpected eurasian cooling unlikely due to arctic sea-ice loss. *Nat. Geosci.* **9**, 838–842 (2016).
5. Deser, C., Terray, L. & Phillips, A. S. Forced and internal components of winter air temperature trends over north america during the past 50 years: mechanisms and implications. *J. Clim.* **29**, 2237–2258 (2016).
6. Wallace, J. M., Fu, Q., Smoliak, B. V., Lin, P. & Johanson, C. M. Simulated versus observed patterns of warming over the extratropical Northern Hemisphere continents during the cold season. *Proc. Natl Acad. Sci. USA* **109**, 14337–14342 (2012).
7. Iles, C. & Hegerl, G. Role of the North Atlantic Oscillation in decadal temperature trends. *Environ. Res. Lett.* **12**, 114010 (2017).
8. Maher, N., Lehner, F. & Marotzke, J. Quantifying the role of internal variability in the temperature we expect to observe in the coming decades. *Environ. Res. Lett.* **15**, 054014 (2020).
9. Deser, C., Hurrell, J. W. & Phillips, A. S. The role of the North Atlantic Oscillation in European climate projections. *Clim. Dyn.* **49**, 3141–3157 (2017).
10. Shepherd, T. G. Atmospheric circulation as a source of uncertainty in climate change projections. *Nat. Geosci.* **7**, 703–708 (2014).
11. Kravtsov, S. Pronounced differences between observed and CMIP5-simulated multidecadal climate variability in the twentieth century. *Geophys. Res. Lett.* **44**, 5749–5757 (2017).
12. Wang, X., Li, J., Sun, C. & Liu, T. NAO and its relationship with the Northern Hemisphere mean surface temperature in CMIP5 simulations. *J. Geophys. Res. Atmos.* **122**, 4202–4227 (2017).
13. Woollings, T., Czuchnicki, C. & Franzke, C. Twentieth century North Atlantic jet variability. *Q. J. R. Meteorol. Soc.* **140**, 783–791 (2014).
14. Bracegirdle, T. J., Lu, H., Eade, R. & Woollings, T. Do CMIP5 models reproduce observed low-frequency North Atlantic jet variability? *Geophys. Res. Lett.* **45**, 7204–7212 (2018).
15. Ba, J. et al. A multi-model comparison of atlantic multidecadal variability. *Clim. Dyn.* **43**, 2333–2348 (2014).
16. Peings, Y., Simpkins, G. & Magnusdottir, G. Multidecadal fluctuations of the North Atlantic ocean and feedback on the winter climate in CMIP5 control simulations. *J. Geophys. Res. Atmos.* **121**, 2571–2592 (2016).
17. Simpson, I. R., Deser, C., McKinnon, K. A. & Barnes, E. A. Modeled and observed multidecadal variability in the North Atlantic jet stream and its connection to sea surface temperatures. *J. Clim.* **31**, 8313–8338 (2018).
18. Kim, W. M., Yeager, S., Chang, P. & Danabasoglu, G. Low-frequency North Atlantic climate variability in the community earth system model large ensemble. *J. Clim.* **31**, 787–813 (2018).
19. Simpson, I. R., Yeager, S. G., McKinnon, K. A. & Deser, C. Decadal predictability of late winter precipitation in western europe through an ocean–jet stream connection. *Nat. Geosci.* **12**, 613–619 (2019).
20. O'Reilly, C. H., Zanna, L. & Woollings, T. Assessing external and internal sources of atlantic multidecadal variability using models, proxy data, and early instrumental indices. *J. Clim.* **32**, 7727–7745 (2019).
21. Omrani, N.-E., Keenlyside, N. S., Bader, J. & Manzini, E. Stratosphere key for wintertime atmospheric response to warm atlantic decadal conditions. *Clim. Dyn.* **42**, 649–663 (2014).
22. Bladé, I., Liebmann, B., Fortuny, D. & van Oldenborgh, G. J. Observed and simulated impacts of the summer NAO in Europe: implications for projected drying in the Mediterranean region. *Clim. Dyn.* **39**, 709–727 (2012).
23. O'Reilly, C. H., Woollings, T. & Zanna, L. The dynamical influence of the atlantic multidecadal oscillation on continental climate. *J. Clim.* **30**, 7213–7230 (2017).
24. Ghosh, R., Müller, W. A., Baehr, J. & Bader, J. Impact of observed North Atlantic multidecadal variations to European summer climate: a linear baroclinic response to surface heating. *Clim. Dyn.* **48**, 3547–3563 (2017).
25. Qasmi, S., Cassou, C. & Boé, J. Teleconnection between atlantic multidecadal variability and European temperature: diversity and evaluation of the coupled model intercomparison project phase 5 models. *Geophys. Res. Lett.* **44**, 11–140 (2017).
26. McKinnon, K. A. & Deser, C. Internal variability and regional climate trends in an observational large ensemble. *J. Clim.* **31**, 6783–6802 (2018).
27. Masson-Delmotte, V. et al. *IPCC, 2021: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, 2021).
28. Madonna, E., Li, C., Grams, C. M. & Woollings, T. The link between eddy-driven jet variability and weather regimes in the North Atlantic-European sector. *Q. J. R. Meteorol. Soc.* **143**, 2960–2972 (2017).
29. Compo, G. P. et al. The twentieth century reanalysis project. *Q. J. R. Meteorol. Soc.* **137**, 1–28 (2011).
30. Allan, R. & Ansell, T. A new globally complete monthly historical gridded mean sea level pressure dataset (HadSLP2): 1850–2004. *J. Clim.* **19**, 5816–5842 (2006).
31. Worley, S. J., Woodruff, S. D., Reynolds, R. W., Lubker, S. J. & Lott, N. Icoads release 2.1 data and products. *Int. J. Climatol.* **25**, 823–842 (2005).
32. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93**, 485–498 (2012).
33. Eyring, V. et al. Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* **9**, 1937–1958 (2016).

34. Maher, N., Power, S. B. & Marotzke, J. More accurate quantification of model-to-model agreement in externally forced climatic responses over the coming century. *Nat. Commun.* **12**, 788 (2021).

35. Maher, N. et al. The Max Planck Institute grand ensemble: enabling the exploration of climate system variability. *J. Adv. Model. Earth Syst.* **11**, 2050–2069 (2019).

36. Hurrell, J. W., Kushnir, Y., Ottersen, G. & Visbeck, M. An overview of the North Atlantic Oscillation. *Geophys. Monogr. Am. Geophys. Union* **134**, 1–36 (2003).

37. Hurrell, J. W. Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation. *Science* **269**, 676–679 (1995).

38. Stephenson, D. B., Pavan, V. & Bojariu, R. Is the North Atlantic Oscillation a random walk? *Int. J. Climatol.* **20**, 1–18 (2000).

39. Scaife, A. A. & Smith, D. A signal-to-noise paradox in climate science. *npj Clim. Atmos.c Sci.* **1**, 1–8 (2018).

40. Klavans, J. M., Cane, M. A., Clement, A. C. & Murphy, L. N. NAO predictability from external forcing in the late 20th century. *npj Clim. Atmos. Sci.* **4**, 1–8 (2021).

41. Stocker, T. F. et al. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of IPCC the Intergovernmental Panel on Climate Change (Cambridge University Press, 2014).

42. Vautard, R. et al. Summertime European heat and drought waves induced by wintertime Mediterranean rainfall deficit. *Geophys. Res. Lett.* https://doi.org/10.1029/2006GL028001 (2007).

43. Zampieri, M. et al. Hot European summers and the role of soil moisture in the propagation of Mediterranean drought. *J. Clim.* **22**, 4747–4758 (2009).

44. Li, J., Sun, C. & Jin, F.-F. NAO implicated as a predictor of Northern Hemisphere mean temperature multidecadal variability. *Geophys. Res. Lett.* **40**, 5497–5502 (2013).

45. McCarthy, G. D., Haigh, I. D., Hirschi, J. J.-M., Grist, J. P. & Smeed, D. A. Ocean impact on decadal atlantic climate variability revealed by sea-level observations. *Nature* **521**, 508–510 (2015).

46. Sun, C., Li, J. & Jin, F.-F. A delayed oscillator model for the quasi-periodic multidecadal variability of the NAO. *Clim. Dyn.* **45**, 2083–2099 (2015).

47. O'Reilly, C. H. & Zanna, L. The signature of oceanic processes in decadal extratropical sst anomalies. *Geophys. Res. Lett.* **45**, 7719–7730 (2018).

48. Delworth, T. L. et al. The central role of ocean dynamics in connecting the North Atlantic Oscillation to the extratropical component of the atlantic multidecadal oscillation. *J. Clim.* **30**, 3789–3805 (2017).

49. Sutton, R. T. & Hodson, D. L. Atlantic ocean forcing of North American and European summer climate. *Science* **309**, 115–118 (2005).

50. Sutton, R. T. & Dong, B. Atlantic ocean influence on a shift in European climate in the 1990s. *Nat. Geosci.* **5**, 788–792 (2012).

51. Ruprich-Robert, Y. et al. Impacts of the atlantic multidecadal variability on North American summer climate and heat waves. *J. Clim.* **31**, 3679–3700 (2018).

52. Goodkin, N. F., Hughen, K. A., Doney, S. C. & Curry, W. B. Increased multidecadal variability of the North Atlantic Oscillation since 1781. *Nat. Geosci.* **1**, 844–848 (2008).

53. Trouet, V. et al. Persistent positive North Atlantic Oscillation mode dominated the medieval climate anomaly. *Science* **324**, 78–80 (2009).

54. Ortega, P. et al. A model-tested North Atlantic Oscillation reconstruction for the past millennium. *Nature* **523**, 71–74 (2015).

55. Lehner, F. et al. Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. *Earth Syst. Dyn.* **11**, 491–508 (2020).

56. Brunner, L. et al. Comparing methods to constrain future European climate projections using a consistent framework. *J. Clim.* **33**, 8671–8692 (2020).

57. Masson-Delmotte, V. et al. *Global Warming of 1.5 °C. An IPCC Special Report on the Impacts of Global Warming of 1.5 °C above Pre-industrial Levels and Related Global Greenhouse Gas Emission Pathways, in the Context of Strengthening the Global Response to the Threat of Climate Change, Sustainable Development, and Efforts to Eradicate Poverty* (IPCC, 2018).

58. Christensen, O. B. & Kjellström, E. Partitioning uncertainty components of mean climate and climate change in a large ensemble of European regional climate model projections. *Clim. Dyn.* **54**, 4293–4308 (2020).

59. Jacob, D. et al. EURO-CORDEX: new high-resolution climate change projections for European impact research. *Reg. Environ. Change* **14**, 563–578 (2014).

60. Trenberth, K. E. & Paolino Jr, D. A. The Northern Hemisphere sea-level pressure data set: trends, errors and discontinuities. *Monthly Weather Rev.* **108**, 855–872 (1980).

61. Folland, C. K. et al. The summer North Atlantic Oscillation: past, present, and future. *J. Clim.* **22**, 1082–1103 (2009).

62. Greatbatch, R. J. & Rong, P.-p Discrepancies between different Northern Hemisphere summer atmospheric data products. *J. Clim.* **19**, 1261–1273 (2006).

63. Theiler, J., Eubank, S., Longtin, A., Galdrikian, B. & Farmer, J. D. Testing for nonlinearity in time series: the method of surrogate data. *Phys. D* **58**, 77–94 (1992).

64. Holmes, C. R., Woollings, T., Hawkins, E. & De Vries, H. Robust future changes in temperature variability under greenhouse gas forcing and the relationship with thermal advection. *J. Clim.* **29**, 2221–2236 (2016).

65. O'Reilly, C. Code and figure data from "Projections of Northern Hemisphere extratropical climate underestimate internal variability and hence uncertainty". https://doi.org/10.5281/zenodo.52115300 (2021).

Acknowledgements

Author contributions

C.H.O'R. designed the study, developed the methodology, carried out the analysis, and led the writing of the paper. D.J.B., A.W., T.W., A.B. and G.H. contributed to the discussion of the results and writing the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43247-021-00268-7.

Correspondence and requests for materials should be addressed to Christopher H. O'Reilly.

Peer review information *Communications Earth & Environment* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary handling editor: Heike Langenberg.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.