Met Office
Hadley Centre

Hadley Centre Technical Note 109: Comparison between climate change projections from the UKCP land scenarios and CMIP6 models

J.M. Murphy, B.B.B. Booth, G.R. Harris, C.F. McSweeney, T.E. Palmer, D.M.H. Sexton, K. Yamazaki

Document history

Version	Purpose	Date
1.0	Version for HCCP deliverable.	21/03/2023
2.0	Revised version for issue as HCTN.	18/05/2023
3.0	Final version of HCTN.	21/06/2023

Prepared by

James Murphy, Science Fellow, Climate Change Projections.

Reviewed by

Prof. Lizzie Kendon, Strategic Head, Understanding Climate Change, (V1.0, 28/03/2023).

Prof. Jason Lowe, Principal Fellow, Climate Science, (V2.0, 05/06/2023).

Authorised for issue by

Katrina Macneill, Knowledge Integration team, 28/06/2023.

Comparison between climate change projections from the UKCP land scenarios and CMIP6 models

J.M. Murphy, B.B.B. Booth, G.R. Harris, C.F. McSweeney, T.E. Palmer, D.M.H. Sexton, K. Yamazaki

Met Office Hadley Centre, Exeter, U.K.

June 2023

Contents

Abstract

The CMIP6 ensemble continues a trend of steady improvement through successive phases of Coupled Model Intercomparison Project (CMIP) multi-model experiments. We compare CMIP6 simulations against those of UKCP-Global, a component of the 2018 release of the UK Climate Projections (UKCP18). UKCP-Global consists of a screened subset of CMIP5 models combined with a perturbed parameter ensemble (PPE) based on the Met Office HadGEM3-GC3.05 model. Overall, we assess the CMIP6 and UKCP-Global ensembles to be similar in quality, based on multivariate climatological averages, three major modes of coupled ocean-atmosphere variability and historical changes in global average surface temperature. Each individual simulation exhibits a range of variable-specific errors, none standing out as clearly superior to the others. Some errors (e.g. a systematic warm bias in marine stratocumulus regions) are common to both ensembles. During 2010-2020 the observed global warming (of ~1.0°C since 1900-1930) lies near the middle of the CMIP6 and UKCP-Global ranges. However, most PPE members and some CMIP6 models are cooler than observations during 1960-1990, due partly to strong aerosol forcing.

Projected changes for Scotland and England are compared for the fossil-fuel intensive RCP8.5 and SSP5-8.5 emissions scenarios and the strong mitigation scenarios RCP2.6 and SSP1-2.6. The CMIP6 ensemble confirms the main features of the distributions of surface temperature and precipitation changes projected by UKCP-Global and UKCP-Probabilistic (the probabilistic projections component of UKCP18), although some notable differences are also found. By the 2070s, all three sets of projections show central estimates of warming that are larger in summer than in winter, in England than in Scotland, and under intensive compared with mitigated greenhouse gas forcing. However, few CMIP6 simulations lie within the lowest 25% of winter warming outcomes projected by UKCP-Probabilistic for Scotland.

All precipitation distributions show signals for increase in winter that are confident under strong forcing (likelihood ~90%) and reasonably confident (likelihood ~75%) under mitigated forcing. In summer, drying is projected to be more likely than not for both countries and scenarios, with the exception that CMIP6 indicates no preferred sign for Scotland under low emissions. For England, the intensity of summer rainfall reductions at the dry end of the distributions is somewhat less pronounced in CMIP6 than in the UKCP datasets, under the high emissions scenarios.

Neither CMIP6 nor UKCP-Global produces a preferred sign for future changes in the winter North Atlantic Oscillation (NAO). However, studies of daily weather regimes report shifts towards cyclonic or zonal types at the expense of anticyclonic patterns, in both datasets. UKCP-Global (but not CMIP6) favours a future shift to the positive phase of the summer NAO, arising mainly from the PPE simulations. Both ensembles show that a positive (negative) shift in the phase of summer NAO is associated with reductions (increases) in future July-August rainfall. However, the modelled relationships tend to be

weaker than observed, implying that the influence of summer NAO changes on UK rainfall and soil moisture may be underestimated.

UKCP18 also includes sets of regional and local projections at high spatial resolution. Currently these are based exclusively on driving PPE simulations and show narrower uncertainty ranges than CMIP6, UKCP-Global and UKCP-Probabilistic, especially in summer. This applies when the regional projections are presented per unit global warming, as well as in a scenario-based format as changes for specific periods that include uncertainties in global temperature response. These results underline the importance of accounting for structural modelling uncertainties in regional impact assessments, with addition of CMIP5-driven simulations planned for the regional and local UKCP datasets. In addition, more research is needed to understand the influences and sampling characteristics of a wider set of drivers of regional extremes in UKCP, including changes in regional storm dynamics as well as thermodynamic influences linked to average levels of future warming.

Experience in using observations to screen or weight CMIP6, UKCP-Global and UKCP-Probabilistic results demonstrates the importance of constraints based on both historical climate change trends and aspects of recent climatology. In this “first look” assessment we simply included UK projections from all available CMIP6 models, but there is future potential to refine the information by applying suitable performance criteria. For example, three configurations of one CMIP6 model, that simulate much stronger warming than other models but exhibit some significant climatological biases, could potentially be excluded by such an exercise. In general, we conclude that screening or weighting CMIP6 models should be based on performance over several variables, and that models should not be screened out simply on the basis that they run ‘hot’.

1. Introduction

The UK climate projections (UKCP) provide national scenarios of climate variability and change for the 21st century (see https://www.metoffice.gov.uk/research/approach/collaboration/ukcp/index). The latest major release (UKCP18) included updated projections for the land and marine environments in and around the UK (Lowe et al., 2018) as well as datasets of observations from the National Climate Information Centre, updated annually (Kendon et al., 2022).

In UKCP18 the land projections included a probabilistic product (UKCP-Probabilistic), and sets of global, regional and local projections (UKCP-Global, UKCP-Regional and UKCP-Local). The original releases are described in Murphy et al. (2018) and Kendon et al (2021), with subsequent revisions and additions documented in Murphy et al. (2020), Sexton et al. (2020), Harris et al. (2022) and Kendon et al. (2023). The probabilistic projections are a statistical synthesis of 348 climate model simulations that provide information on uncertainties in future changes. The global, regional and local projections all consist of raw model data forming flexible datasets for impacts analysis. These are provided at horizontal resolutions of 60km (UKCP-Global), 12km (UKCP-Regional) and 2.2km (UKCP-Local). All products cover the UK, while UKCP-Regional and UKCP-Global also include information for Europe and worldwide regions, respectively.

Since UKCP18 was published the international modelling community has produced a new generation of climate change projections, via the sixth phase of the Coupled Model Intercomparison Project (CMIP6, Eyring at al., 2016). Lee et al. (2021) assessed CMIP6 projections in the Sixth Assessment Report (AR6) of the Inter-Governmental Panel on Climate Change (IPCC). They will be widely used in studies of global and regional climate change during the next few years (e.g. Brunner et al., 2020a; Palmer et al., 2021).

In this report we compare the UKCP-Global and UKCP-Probabilistic projections against CMIP6 results. Section 2 describes the data, followed by assessment of the historical performance of coupled ocean-atmosphere model simulations from CMIP6 and UKCP-Global in section 3. A comparison of the CMIP6 and UKCP projections follows in section 4. We focus on a limited selection of key diagnostics presented in a similar format to their counterparts in the UKCP18 science report (Murphy et al., 2018), tracing how CMIP6 models perform against the criteria upon which UKCP18 results were originally assessed.

Our approach is to assemble all CMIP6 models from which data is available for the diagnostic in question. We do not attempt to screen the multi-model ensemble to eliminate poorly performing models (e.g. McSweeney et al., 2015), or to reduce the influence of “near-neighbour” models that share a majority of component modules and hence simulate similar outcomes for historical and future climate (e.g. Sanderson et al., 2015). Neither do we attempt to attach relative weights to alternative models, based on either performance or similarity criteria (e.g. Knutti et al., 2017).

In this sense, our comparison provides a first look at CMIP6 models in the context of the UKCP scenarios. There is potential to update this simple approach in the future, by applying screening and/or weighting methods to derive refined estimates of UK climate changes from CMIP6. Since such methods were used extensively in the production of the UKCP18 Global and Probabilistic projections (see section 2), the present comparison is not like-for-like in this respect. We provide examples of the potential impact of screening CMIP6 models in section 4, to demonstrate how such an approach can influence projected changes provided for impacts analysis.

2. Data and scenarios

a. UKCP-Probabilistic

The probabilistic projections in UKCP18 were based on 348 perturbed parameter ensemble (PPE) simulations derived from a single climate model (HadCM3), augmented by twelve earth system model projections from the CMIP5 archive (Taylor et al., 2012). These add uncertainties arising from alternative choices in model structure to the parametric uncertainties represented by the PPE simulations. A Bayesian framework (Sexton et al., 2012; Harris et al., 2013; Sexton and Harris, 2015) is used to combine this information with a set of observational constraints, using emulation techniques to construct probability distribution functions (pdfs) that express broad ranges of possible outcomes for climate variability and change during the 21st century. The updated methodology for UKCP18 is described in more detail by Murphy et al. (2018).

The probabilistic projections are presented in an "emissions-driven" format, representing the combined effects of uncertainties in physical and carbon cycle feedbacks in response to a prescribed pathway of future greenhouse gas emissions. They are available for five such pathways: The SRES A1B scenario (Nakicenovic and Swart, 2000) plus the RCP2.6, 4.5, 6.0 and 8.5 scenarios (Moss et al., 2010). In this report, we consider results from RCP2.6 and 8.5 (see section 2d).

In section 4b we compare 5th-95th percentile ranges from UKCP-Probabilistic against the full ranges available from CMIP6 and other UKCP components. This presentation of the pdfs is chosen because we expect extreme outcomes in the tails of the distributions to be more sensitive to limitations in the statistical assumptions required in the methodology (UKCP18 Technical Note, 2019). Nevertheless, the occurrence of simulated outcomes in other products outside the 5-95% ranges of UKCP-Probabilistic does not necessarily indicate inconsistency with the latter, since we expect a 10% probability of seeing such outcomes.

b. UKCP-Global

The 28 simulations included in UKCP-Global (Murphy et al., 2018) were sourced from two climate model ensembles.

Fifteen members were contributed by a PPE based on HadGEM3-GC3.05, a coupled model closely related to the Met Office model contributed to CMIP6 (HadGEM3-GC3.1, Williams et al., 2018). One PPE member used standard (unperturbed) parameter values, the others included perturbations to 47

atmosphere model parameters. These were selected from ~3000 candidate parameter sets, reduced to 25 by applying multivariate screening criteria to retrospective 5-day weather forecasts and five-year atmosphere-only simulations using prescribed sea surface temperatures (Sexton et al., 2021). These were then reduced to 15 members for use in UKCP, following further evaluation of coupled ocean-atmosphere simulations (Yamazaki et al., 2021). The coupled simulations (GC3.05-PPE) included flux adjustments to restrict the development of regional biases in sea surface temperature (SST) and surface salinity.

Thirteen members were provided by a subset of CMIP5 models, selected from 31 candidates by McSweeney et al. (2018). The selection criteria included several global and regional performance metrics drawn mainly from published literature. Application of these identified 13 models (CMIP5-13) that provided good multivariate skill in their historical simulations (in the context of other CMIP5 models). The selection also prioritised structural diversity in the selection of alternative model components, achieved by measuring the degree of similarity in historical simulation biases (Sanderson et al., 2015). The emissions scenarios and forcing strategies for UKCP-Global and CMIP6 are described in section 2(d) below.

c. CMIP6

For each diagnostic in sections 3 and 4, we use results from all CMIP6 models for which data are available (Appendix). Ensemble sizes vary from 29 to 64 and are stated in each figure. Data is derived from coupled ocean-atmosphere models of the physical climate system and from earth system models. The latter simulate additional biogeochemical processes including dynamic vegetation modules (Song et al., 2021), the Earth's carbon cycle and aspects of non-aerosol atmospheric chemistry (e.g. Mulcahy et al., 2022). However, all simulations considered here were run in "concentration-driven" mode, in which atmospheric concentrations of CO_2 and other greenhouse gases are prescribed from standard time-dependent profiles rather than being predicted interactively. No flux adjustments were used in the CMIP5-13 or CMIP6 simulations.

No pre-screening based on model quality criteria is applied to the CMIP6 data. We use one simulation per model. This avoids implicitly upweighting models for which several simulations are contributed to the CMIP6 data archive as an ensemble distinguished by different starting conditions. Some modelling centres are now producing large initial-condition ensembles to provide insights into the impacts of the internal (unforced) component of climate variability in a changing climate (Deser at al., 2020; Lehner et al., 2020). However, we focus here on CMIP6 as a multi-model ensemble that represents the combined effects of uncertainties due to internal variability and model formulation in a "one-model-one-vote" format.

d. Emissions scenarios

In UKCP18, results from UKCP-Global were initially made available for RCP8.5, a fossil-fuel intensive scenario widely used in CMIP5 (Collins et al., 2013). It drives strong responses useful for identifying characteristic climate change signals and understanding their origins. More recently, UKCP-Global data

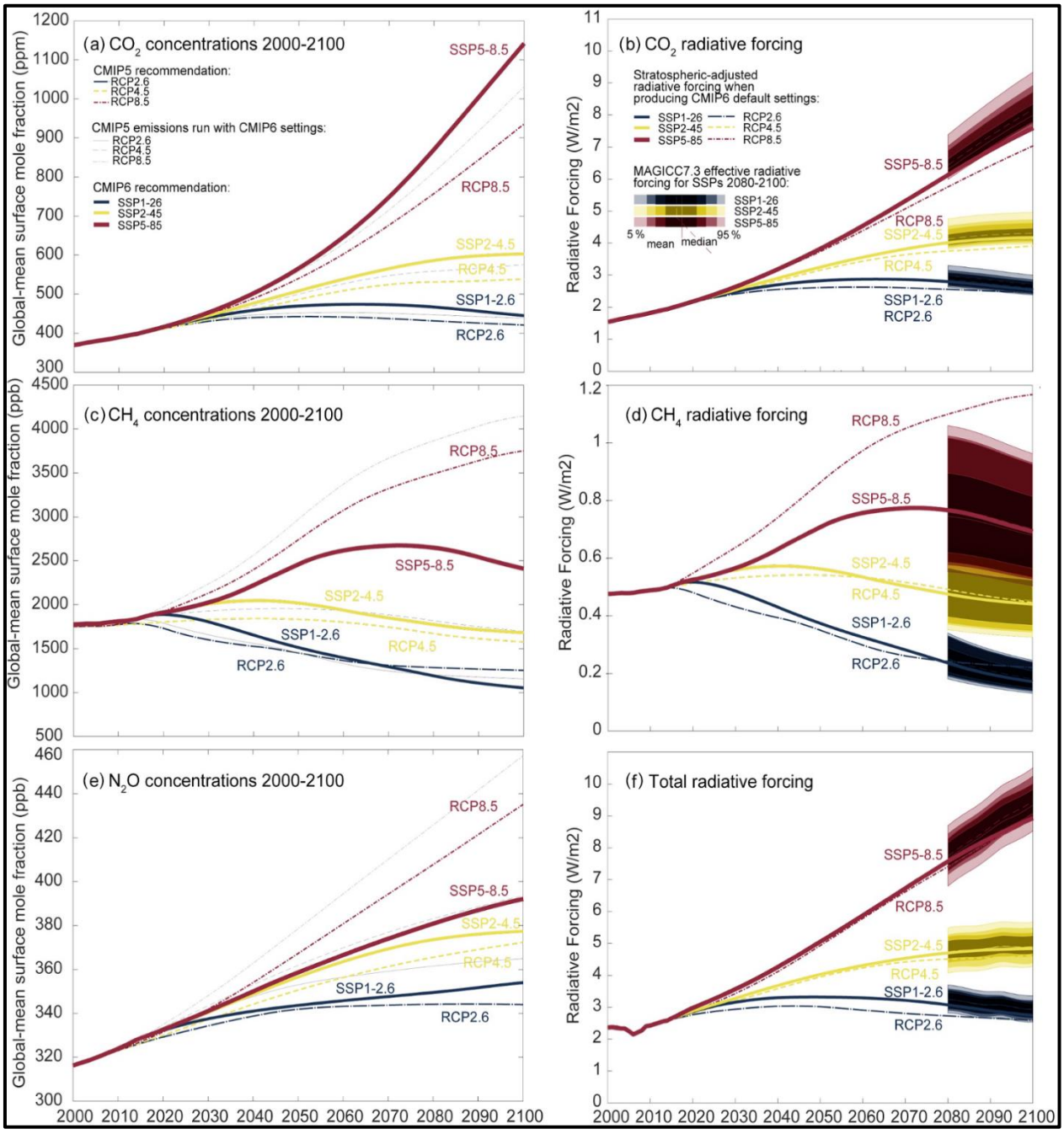

Figure 1. *Comparison of CO_2, CH_4 and N_2O concentrations (a, c, e) and effective radiative forcing (ERF, b, d, f) for the RCP and SSP scenarios used in concentration-driven CMIP5 and CMIP6 simulations respectively. The SSP5-8.5 scenario features higher CO_2 concentrations than RCP8.5, largely due to use of updated carbon cycle settings in the SSP5-8.5 calculations. RCP8.5 emissions with the SSP5-8.5 carbon cycle settings (shown as a thin dashed line in panel a) would produce intermediate CO_2 concentrations. However, methane and nitrous oxide concentrations are lower in SSP5-8.5 than in RCP8.5. Panel (f) shows the total ERF for each scenario, resulting from the combined effects from all greenhouse gas and aerosol constituents. The coloured plumes show uncertainty ranges for ERF in the SSPs, during 2080-2100. Reproduced from Tebaldi et al. (2021) (their Figure A7), in which more details are available. This work is distributed under the Creative Commons Attribution 4.0 License.*

for RCP2.6 was added. This assumes strong climate change mitigation measures and reaches a CO_2 concentration of approximately 400 parts per million (ppm) by 2100, compared with ~900 ppm in RCP8.5 (Figure 1, reproduced from Tebaldi et al., 2021). The UKCP-Global results for RCP2.6 include the 15 members of GC3.05-PPE, plus ten CMIP5-13 members for which RCP2.6 results are available (Fung et al. (2020) report results from nine of these).

The CMIP5-13 and GC3.05-PPE ensembles both used coupled ocean-atmosphere model configurations in which greenhouse gas concentrations are prescribed (Yamazaki et al., 2021). The CMIP5-13 simulations all use standard concentration pathways for the relevant RCP scenario (Figure 1). For trace gases other than CO_2, standard concentration time series were also used in GC3.05-PPE. However, for CO_2 a range of pathways were prescribed in different GC3.05-PPE members, consistent with uncertainties in carbon cycle feedbacks diagnosed from UKCP-Probabilistic (Figure 2, taken from Murphy et al., 2018). For a given RCP scenario, ranges of future change projected in UKCP-Global are driven purely by uncertainties in physical climate responses in the case of CMIP5-13, whereas those from GC3.05-PPE arise from uncertainties in both physical responses and the component of radiative forcing associated with carbon cycle processes (Yamazaki et al., 2021).

New emissions scenarios were used in CMIP6. In these, alternative pathways of future radiative forcing and socioeconomic activity were combined to develop a set of Shared Socioeconomic Pathways (SSPs, O'Neill et al., 2013). We focus on CMIP6 results from two SSPs, chosen to provide continuity with the RCPs (Tebaldi et al., 2021). These are SSP1-2.6 and SSP5-8.5, the numbers representing the approximate radiative forcing[1] in Wm^{-2} reached by the end of the 21st century (Figure 1).

The total radiative forcing in SSP1-2.6 is slightly higher than in RCP2.6, the largest differences (of ~0.3Wm^{-2}) arising during the 2060-2080 period. This arises partly from higher CO_2 forcing, and partly from other agents.

While the total radiative forcing is similar between RCP8.5 and SSP5-8.5, Figure 1 shows contrasts between the contributions from different greenhouse gases: CO_2 concentrations are higher in RCP8.5 than in SSP5-8.5, whereas the converse applies to methane and nitrous oxide concentrations. The CO_2 concentrations are derived by driving a simple climate model with the relevant RCP or SSP emissions (Riahi et al., 2017), hence the results depend on the carbon cycle feedback prescribed in the simple model. The standard RCP concentration pathways were derived assuming a relatively weak carbon cycle response (e.g. Figure 2, in which the standard pathway for RCP8.5 (red diamond) lies at the low end of the corresponding probability distribution). The SSP pathways were generated using updated carbon

[1] These numbers express the radiative forcing following rapid adjustments of the stratosphere. The forcing is slightly different when expressed as effective radiative forcing (ERF), which allows for rapid adjustments of the troposphere and land surface as well as the stratosphere (Smith et al., 2020a). The lines in Fig. 1f show the time series of the stratospherically-adjusted forcing for SSPs, whereas the coloured ranges for 2080-2100 show probability distributions for ERF estimated by Tebaldi et al. (2021).

cycle settings, which partly explain their higher CO_2 concentrations compared to RCP counterparts: The thin lines in Fig. 1a show that combining the SSP carbon cycle setting with RCP emissions produces CO_2 concentrations closer to the SSP pathways. In response to RCP8.5 emissions, the UKCP-Probabilistic distribution encompasses the SSP5-8.5 pathway (whose CO_2 concentration in 2099 lies slightly above the most likely value shown in Fig. 2), as well as including the standard RCP8.5 pathway at its lower end and outcomes above the standard SSP5-8.5 pathway.

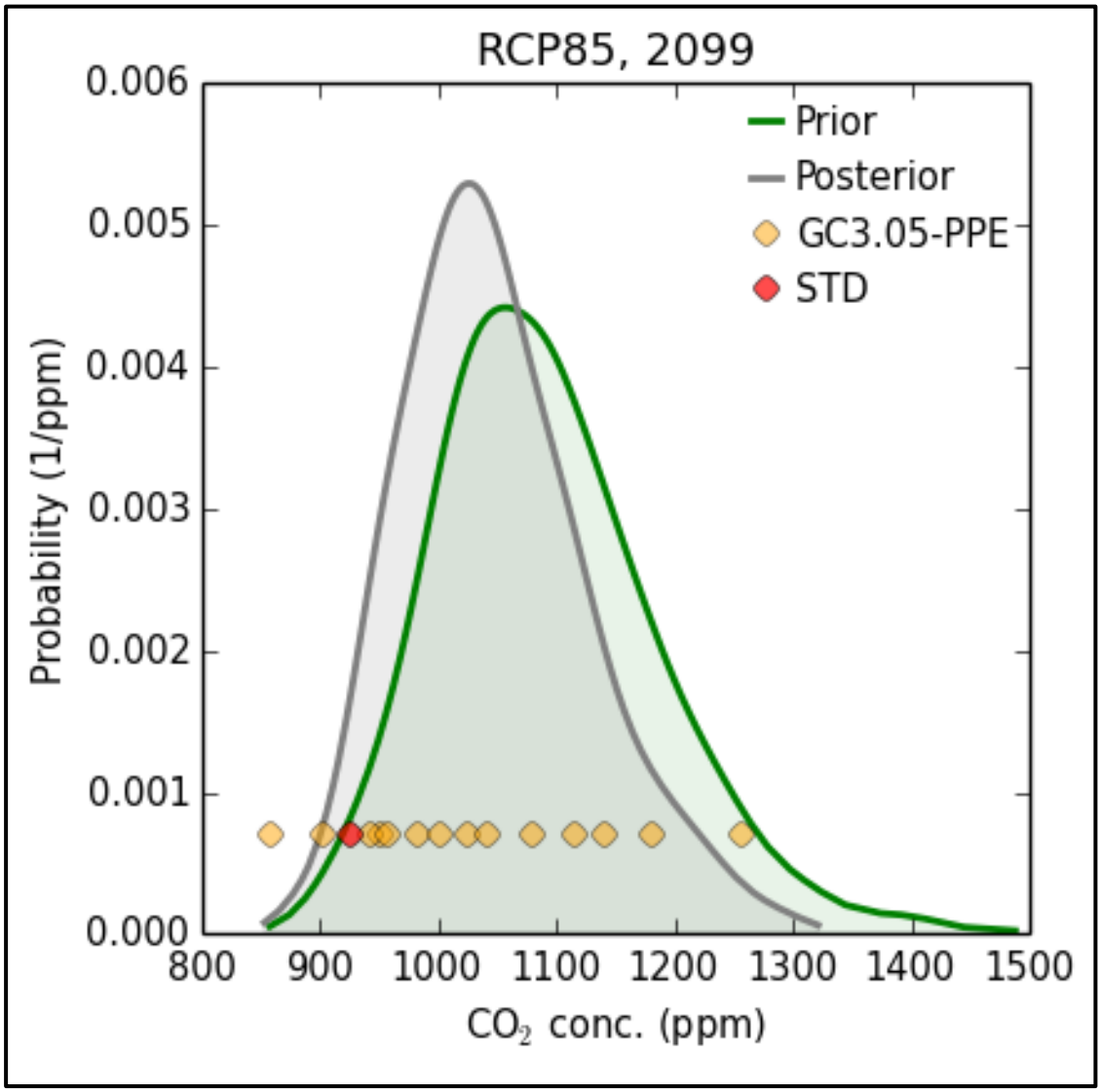

***Figure 2**. Posterior probability distribution (grey curve) from the UKCP probabilistic projections for atmospheric CO_2 concentration (ppm) in 2099, under carbon emissions prescribed by RCP8.5. Comparing this with the corresponding prior distribution (green curve) shows the impact of applying observational constraints. Orange circles show concentrations from the CO_2 pathways used to drive 14 of the 15 GC3.05-PPE members. The red circle denotes the standard pathway used in concentration-driven RCP8.5 simulations. This pathway is used in the CMIP5-13 projections, as well as the GC3.05-PPE member with unperturbed parameter settings (STD). Reproduced from Murphy et al. (2018) (their Figure 3.7), in which more details are available.*

Fyfe et al. (2021) used the CanESM2 model to demonstrate that time-dependent changes in global mean surface air temperature (GMST) differ somewhat between corresponding SSP- and RCP-driven simulations. For 2081-2100 they found warmer responses in the SSPs (relative to pre-industrial climate),

despite the similarity in total radiative forcing (Fig. 1). Differences amounted to 0.16°C for SSP1-2.6 cf RCP2.6 and 0.55°C for SSP5-8.5 cf RCP8.5. This was attributed to higher CO_2 concentrations in the SSP datasets, partly offset by the inclusion of a background stratospheric aerosol loading due to small- and medium-sized volcanic eruptions that was missing from the RCP simulations. Differences in the responses of GMST and Arctic sea-ice extent were also found during the historical period.

e. Summary

Below we compare the UKCP projections and CMIP6 results as alternative datasets available for use in impacts studies and risk assessments. Several factors are potential drivers of differences in their projected outcomes:

- Use of different climate model ensembles (sections 2a-2c), liable to drive different ranges of projected outcomes through alternative representations of physical drivers of climate change response (Palmer et al., 2021).
- Differences in postprocessing strategies to account for model performance (sections 2a, b), or the absence of such a strategy in the case of CMIP6 models (section 2c), can also drive differences in ranges of response.
- Variations in applied radiative forcing (section 2d), arising from (a) differences between the RCP- and SSP-based concentration pathways (when comparing CMIP5-13 against CMIP6); (b) use of an emissions-driven concept in UKCP-Probabilistic and GC3.05-PPE, versus a concentration-driven concept in CMIP5-13 and GC3.05-PPE.

Since our aim is to compare the datasets "as they come" for user applications, we do not attempt adjustments to account for differences in external forcing. However, in section 4 we present examples of future UK changes expressed per unit change in GMST, which removes approximately the globally-averaged influence of contrasts in external forcing and sensitivity of response between the alternative datasets.

3. Comparison of CMIP6 and UKCP projections: Historical evaluation

a. Assessment of CMIP6 models in IPCC AR6

Bock et al. (2020) compare the CMIP3, CMIP5 and CMIP6 ensembles. Compared to the average skill across the three CMIP phases, most CMIP3 models score worse in simulating the climatological seasonal cycle, while most CMIP6 models score better. Improvements across successive generations are found in latitude-longitude fields of radiation, hydrological cycle, surface temperature and dynamical variables. CMIP6 also shows improvements in simulated vertical distributions of temperature, water vapour, and zonal wind speed.

Based on this study and additional evidence, an assessment of CMIP6 performance was conducted in the AR6 Physical Science Basis report. Overall, Arias et al. (2021) assess that CMIP6 models: "Include new and better representation of physical, chemical and biological processes, as well as higher resolution, compared to climate models considered in previous IPCC Assessment Reports. This has

improved the simulation of the recent mean state of most large-scale indicators of climate change and many other aspects across the climate system. Some differences from observations remain, for example in regional precipitation patterns." Some CMIP6 models demonstrate improvements in the simulation of clouds, including more sophisticated representations of aerosol-cloud interactions and more realistic simulation of supercooled liquid droplets with a corresponding increase in cloud optical depths. These improvements lead to a stronger shortwave cloud radiative effect in present-day climate, notably over the Southern Ocean.

Figure TS.2 in Arias et al. (2021) summarises developments through successive CMIP experiments, demonstrating that average skill (based on pattern correlations of spatial climatological fields) is higher for CMIP6, but there is considerable overlap between the envelopes of performance shown by the CMIP3, 5 and 6 ensembles. Similarly, the average spatial resolution (horizontally and vertically in the atmosphere and ocean components) is finer in CMIP6, but individual models in each multi-model ensemble possess a range of resolution choices that is larger than the average inter-ensemble differences.

Overall, the above results confirm that CMIP6 can be viewed as the continuation of a steady trend in model improvement across successive CMIP phases. In sections 4b-4d we update our previous assessment of climatology, variability and historical changes in UKCP-Global (Murphy et al., 2018; Yamazaki et al., 2021), by adding CMIP6 models to provide a new performance benchmark.

b. Climatology in CMIP6 and UKCP-Global simulations

For surface air temperature, Figure 3 compares ensemble median biases in the GC3.05-PPE and CMIP5-13 components of UKCP-Global against the median bias across 47 CMIP6 coupled models. Results are presented for the 20-year average of 1981-2000, for December to February (DJF) and June to August (JJA). The patterns of bias for CMIP6 and CMIP5-13 are broadly similar, indicating that in most regions the systematic components of model error have the same sign. For example, both ensembles show positive median biases in the marine stratocumulus regions off the western coasts of North and South America and southern Africa. Bock et al. (2020) conclude that biases in the stratocumulus regions arise partly from limited horizontal resolution, since these biases are smaller in higher resolution models (e.g. Roberts et al., 2019). Both CMIP5-13 and CMIP6 also show a median warm bias over the Southern Ocean and Kazakhstan (Fig. 3), while in DJF many parts of the northern hemisphere continental landmasses show a cool bias.

In GC3.05-PPE the median cool bias in northern hemisphere winter is more pronounced and more widespread than in the multi-model ensembles. The Southern Ocean warm bias is also larger, despite the use of flux adjustments (absent from the multi-model ensembles). However, in many low-latitude marine regions the median bias is smaller in GC3.05-PPE. This is (at least partly) because here the application of flux adjustments is more successful in limiting the development of SST biases (Yamazaki et al., 2021).

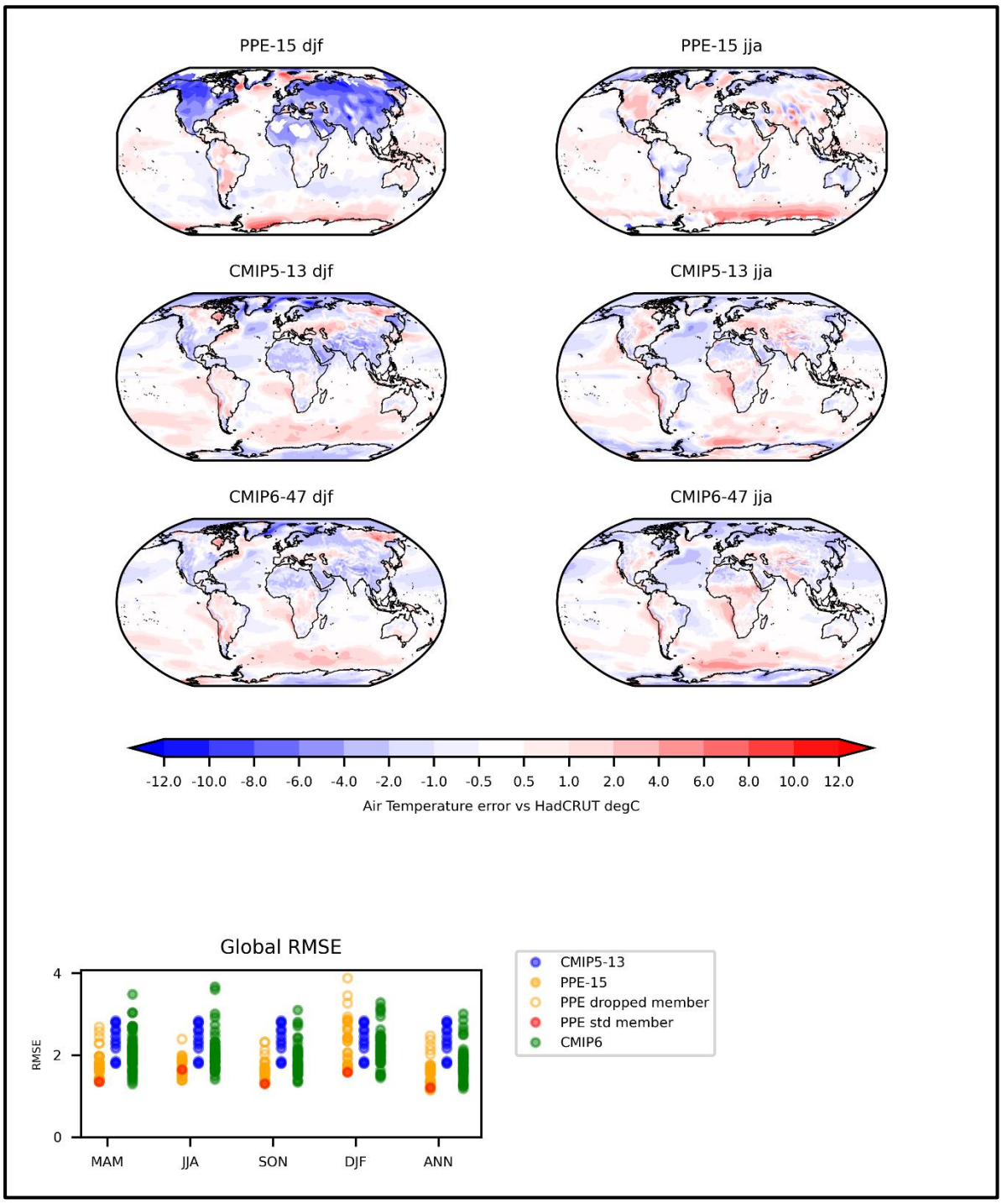

***Figure 3**. Maps show ensemble median biases in 20-year averages of surface air temperature (°C) in December-February (DJF, left) and June-August (JJA, right), relative to observations for 1981-2000 from HadCRUT5 (Morice et al., 2021). The top rows show results from the GC3.05-PPE and CMIP5-13 ensembles that form the UKCP-Global dataset, and the bottom row shows the median bias from 47 CMIP6 models. The panel below shows root-mean-square errors (RMSE, °C) calculated from the global bias fields of every member of each ensemble, for 1981-2000 averages of each meteorological season (including March-May (MAM) and September-November (SON)), and of the annual mean. STD is shown in red, while open orange circles show five potential PPE members that were excluded from UKCP-Global following assessment of their simulations of historical climate (Yamazaki et al., 2021). Some of these lie within the range of GC3.05-PPE biases, so are not visible in the plot.*

Figure 3 also shows root-mean-square errors (RMSE) for the global climatological fields from each simulation, for each season and the annual mean. In all cases, the CMIP6 ensemble shows a wider range

of scores, including members scoring both better and worse than all CMIP5-13 members. For surface air temperature, future application of screening methods to identify a skilful subset of CMIP6 models could potentially yield a set that improves on CMIP5-13 performance, by retaining some of the best CMIP6 models while excluding those showing the most serious biases (e.g. Palmer et al., 2023). The GC3.05-PPE scores also lie within the CMIP6 range, the member with standard parameter settings (STD, shown in red) being invariably competitive with the best CMIP6 models. Other PPE members produce global scores equal to (or slightly better than) the best CMIP5-13 members in all seasons, despite the enhanced cold bias found in northern hemisphere winter.

For context, Figure 3 also shows results from five additional PPE simulations that were dropped from the UKCP-Global dataset (Murphy et al., 2018) but included in the 20-member PPE assessed by Yamazaki et al. (2021). The screening for UKCP was based on criteria including the strength of the Atlantic meridional overturning circulation (AMOC), trends in northern hemisphere surface temperature during 1900-2012 and climatological averages of surface temperature and precipitation over Europe (Murphy et al., 2018). The additional simulations gave larger annual RMSE compared to the members retained in UKCP-Global. However, seasonal and annual RMSE for the additional members lie within the ranges spanned by the two multi-model ensembles except in DJF, for which three of the additional members simulated particularly large northern hemisphere cool biases. Yamazaki et al. (2021) reported (based on an evaluation of multiple variables) that the 20-member version of the PPE was generally suitable for use in global climate research applications.

For precipitation (Fig. 4), the two sets of multi-model patterns again agree on the signs of median bias in most regions, suggesting similarity in the systematic component of model error. Both CMIP5-13 and CMIP6 show an equatorial dry bias in the Pacific basin, with off-equatorial wet biases to the north and south in the tropical west Pacific. In DJF an equatorial dry bias is also found in the Atlantic with a wet bias to the south, while in JJA this pattern shifts slightly northward, the wet bias lying close to the equator. Also common to both ensembles are median dry biases in the south Asian summer monsoon and in parts of the extratropical North Atlantic Ocean, with wet biases over the Indonesian warm pool. In the Indian ocean sector, wet biases occur off the coast of Madagascar in DJF (in the region associated with the southern lobe of the Indian Ocean dipole (Behera and Yamagata, 2001)) and in parts of the Arabian Sea in JJA. Terrestrial dry biases are apparent in parts of northern Eurasia (in DJF) and in south-eastern Europe (in JJA).

The bias patterns in the PPE median are generally in phase with those of CMIP5-13 and CMIP6. In JJA, the wet biases in the tropical west Pacific in JJA are more pronounced in the PPE. On the other hand, dry biases in the north Atlantic are less widespread than in the multi-model ensembles, particularly in DJF. The RMSE values reveal a general shift towards reduced biases in individual CMIP6 models, compared to CMIP5-13. This is apparent in all seasons except JJA, and (in particular) in the

scores for the annual mean. The scores for GC3.05-PPE members invariably lie within the CMIP6 range whereas this is not the case for CMIP5-13 models.

In general, considerable overlap exists between the scores for the three ensembles in Figs. 3 and 4 (consistent with the AR6 results summarised in section 3a). An exception is annual mean precipitation, for which only a few CMIP5-13 members achieve scores better than the worst of the CMIP6 or PPE results.

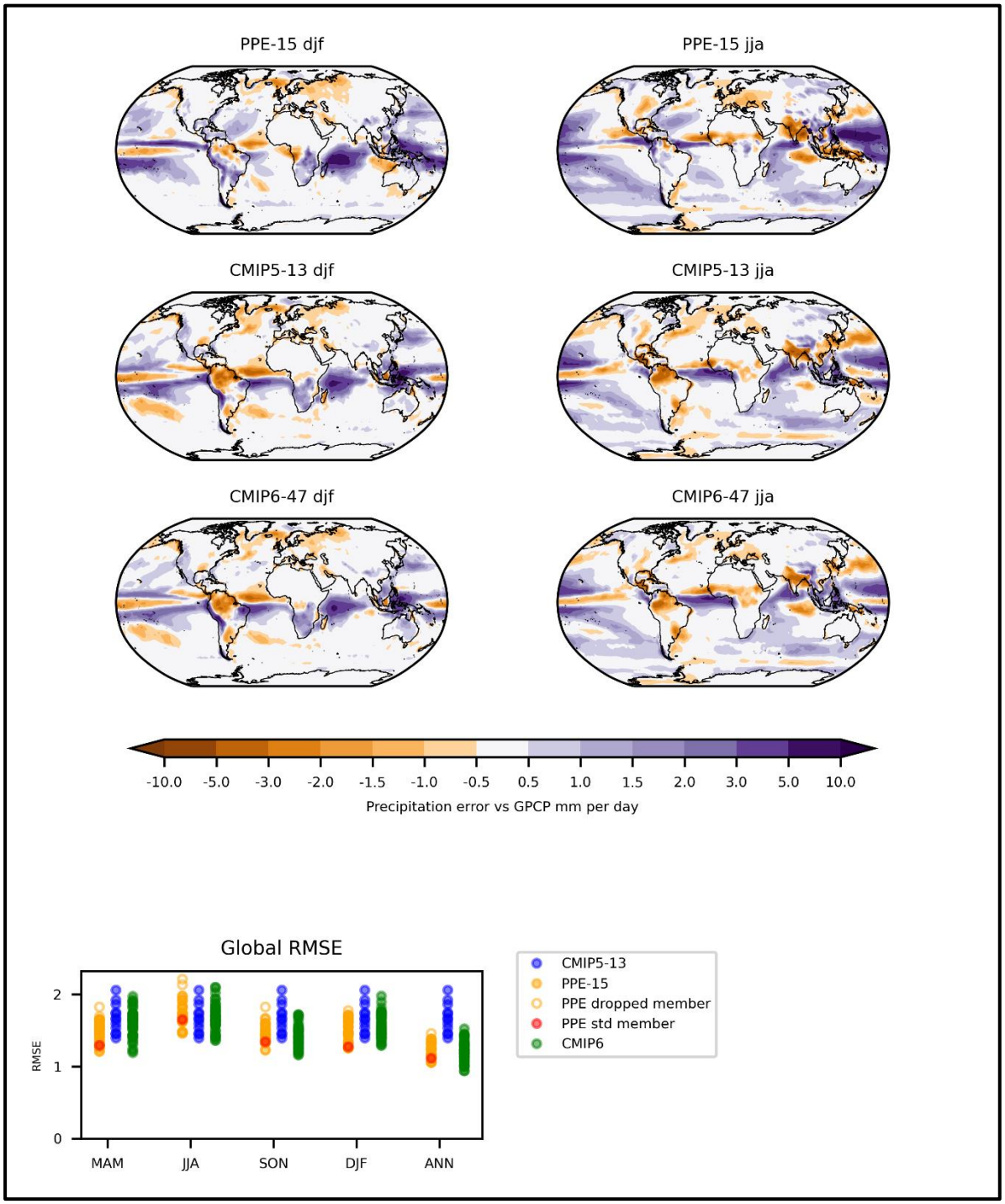

***Figure 4**. As Figure 3, for 20-year average biases in precipitation (mm/day) relative to 1981-2000 observations from the GPCP dataset (Adler et al., 2003). Maps show ensemble median biases, the bottom panel shows RMSE values for individual ensemble members.*

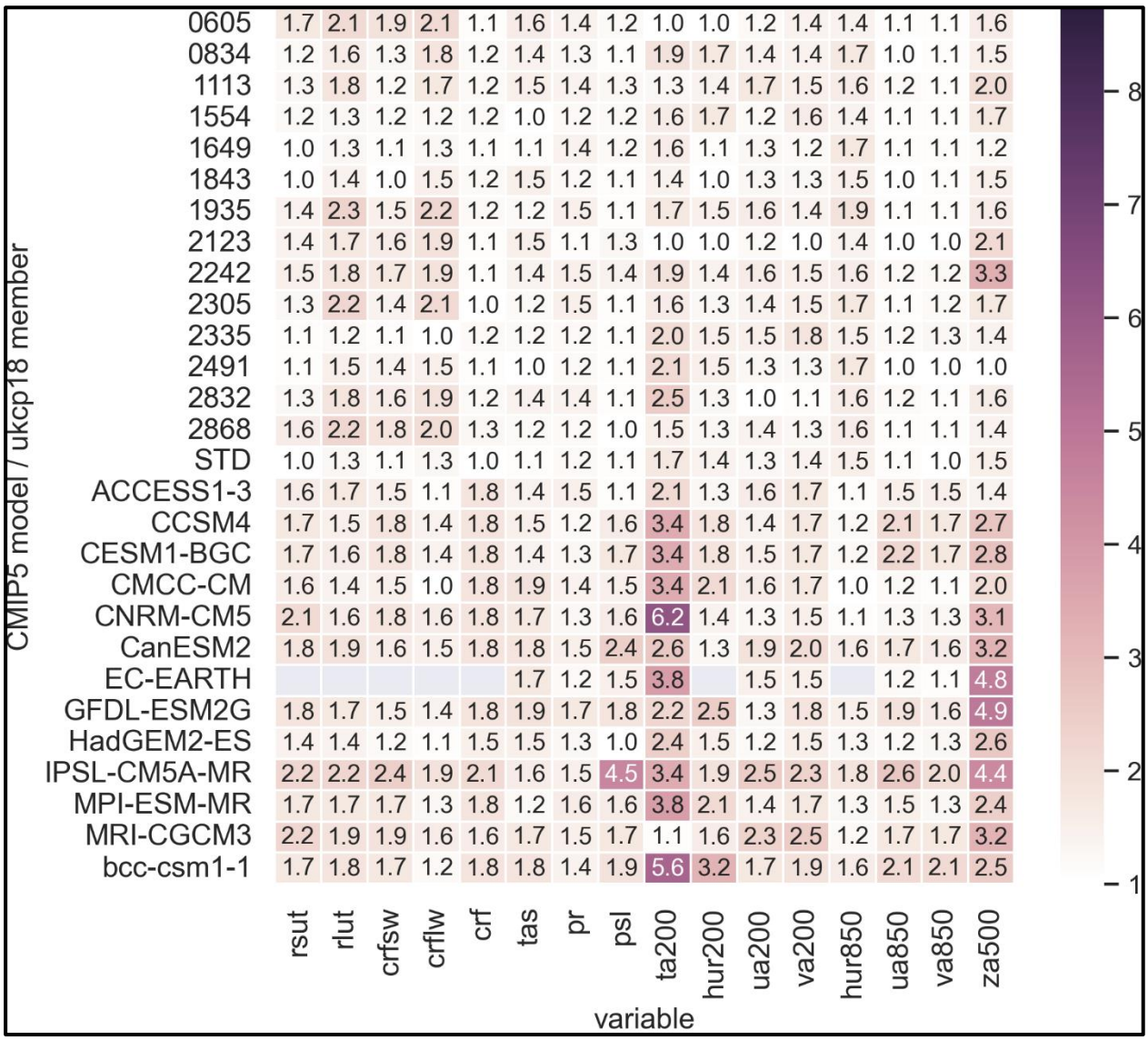

***Figure 5**. Normalised r.m.s. errors in global, annual spatial fields simulated by the 28 members of UKCP-Global for 1981-2000. Members of CMIP5-13 are named (see Table 3.1 of Murphy et al., 2018) while members of GC3.05-PPE are numbered, with the unperturbed member identified as STD. The scores for each variable are normalised by the score for the best-performing simulation across three ensembles: GC3.05-PPE, CMIP5-13 and the 64-member CMIP6 ensemble shown in Figures 6 and 7. Verifying observational datasets: HadSLP2 (Allen and Ansell, 2006) for sea-level pressure (slp); CERES (Loeb et al., 2009) for outgoing short- and long-wave radiative flux at the top of the atmosphere (rsut, rlut); ERA-Interim reanalyses (Dee et al., 2011) for other variables. In this "heatmap" plot, the worst-performing simulations are shown in the darkest shades of purple. The lightest shade denotes the best simulations possessing normalised errors of 1.0. Several entries are blank for EC-Earth, due to missing data. Precipitation is denoted by pr, surface air temperature by tas, cloud radiative effect and its short- and long-wave components by crf, crfsw and crflw; geopotential height at 500 hPa by za500; eastward and northward wind and relative humidity at 850 and 200 hPa by ua850, ua200, va850, va200, hur850, hur250; temperature at 200 hPa by ta200. Observed climatologies cover 1981-2000 apart from CERES, for which 2001-2005 is used because data for the full period do not exist.*

We also assess a broader set of standard climate variables, for UKCP-Global simulations in Figure 5 and 64 CMIP6 simulations in Figures 6 and 7. RMSE values are provided for annual, global spatial fields of climatological averages for 1981-2000. The variables are: surface air temperature, precipitation, sea-level pressure, outgoing shortwave and longwave radiation at the top of the atmosphere, total,

shortwave and longwave cloud radiative effect, 500hPa geopotential height and atmospheric temperature, relative humidity and zonal and meridional wind components at the 850hPa and 200hPa pressure levels. Each RMSE value is normalised by that of the best-performing simulation for the relevant variable, which therefore receives a normalised score of 1.0. High values indicate relatively poor performance, denoted by darker shading. In general, the UKCP-Global scores in Figure 5 differ slightly from those in Murphy et al. (2018), because the best scores are calculated across the larger set of simulations resulting from the addition of CMIP6 results.

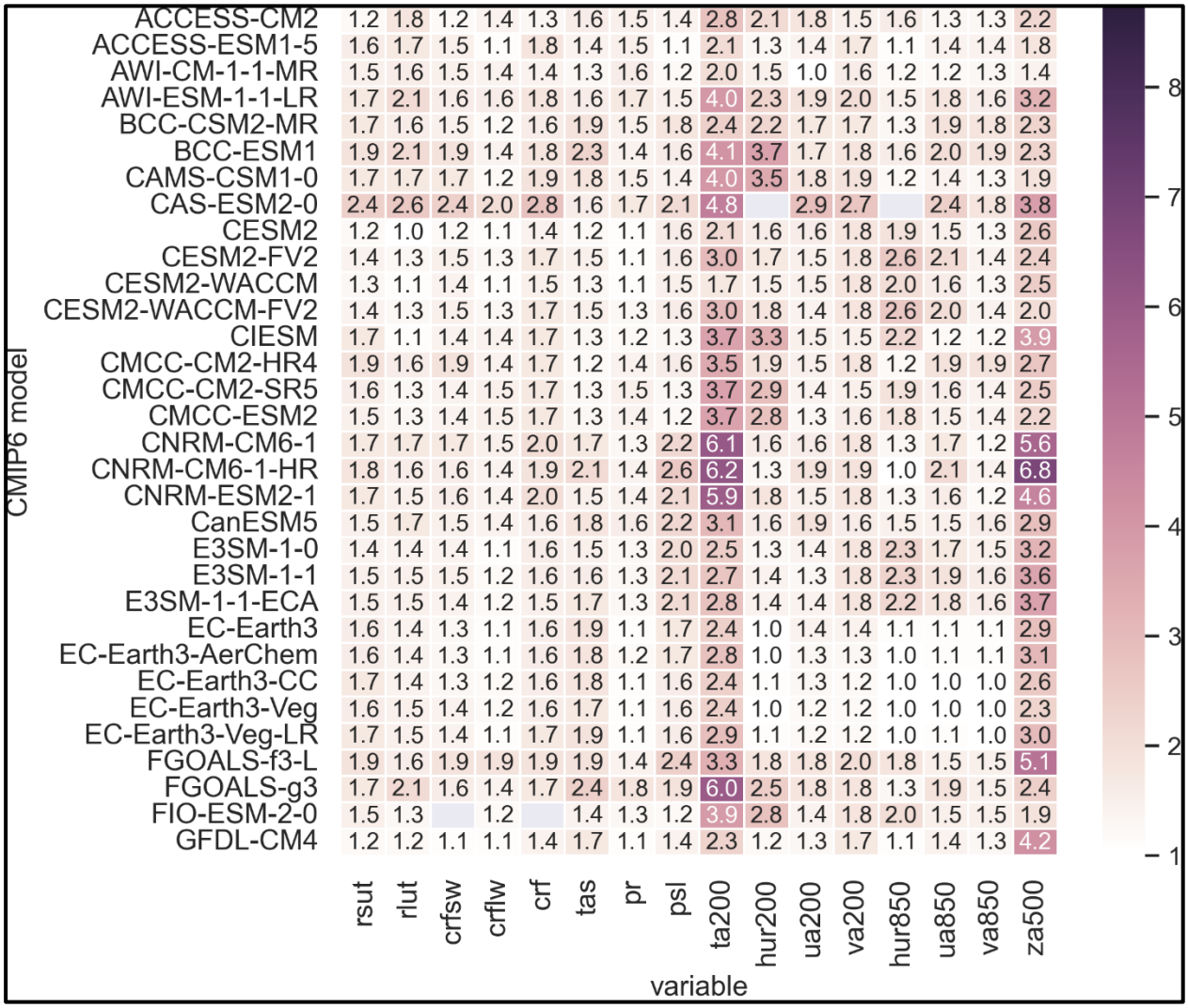

***Figure 6**. Normalised r.m.s. errors in global, annual spatial fields simulated by 32 CMIP6 models for 1981-2000. The scores for each variable are normalised by the score for the best-performing simulation across three ensembles: GC3.05-PPE, CMIP5-13 (shown in Fig. 5) and the 64-member CMIP6 ensemble shown here and in Figure 7. See Figure 5 for further details.*

There are 45 instances of the (best) 1.0 score across the 16 variables, since 2-3 models typically achieve results indistinguishable to one decimal place. Of these, 25 instances belong to GC3.05-PPE, three to CMIP5-13 and 17 to CMIP6. No simulation achieves a score of 1.0 for more than four variables, thus none stand out as clearly superior to other members of the three ensembles.

If (for the sake of illustration) we take a normalised RMSE exceeding five as an indicator of poor relative performance, we find zero instances in GC3.05-PPE, two in CMIP5-13 and 21 in CMIP6. Many of the worst scores occur for the 200hPa temperature (ta200) and 500hPa geopotential height (za500) fields. For ta200, these arise from pervasive cold biases relative to observations, while for za500 most of the poor scores are driven by widespread low biases (also related to cool tropospheric temperatures). However, the worst score of 22.7 (for KIOST-ESM) arises from a pattern of large positive biases over land accompanied by substantial negative biases over the oceans (not shown).

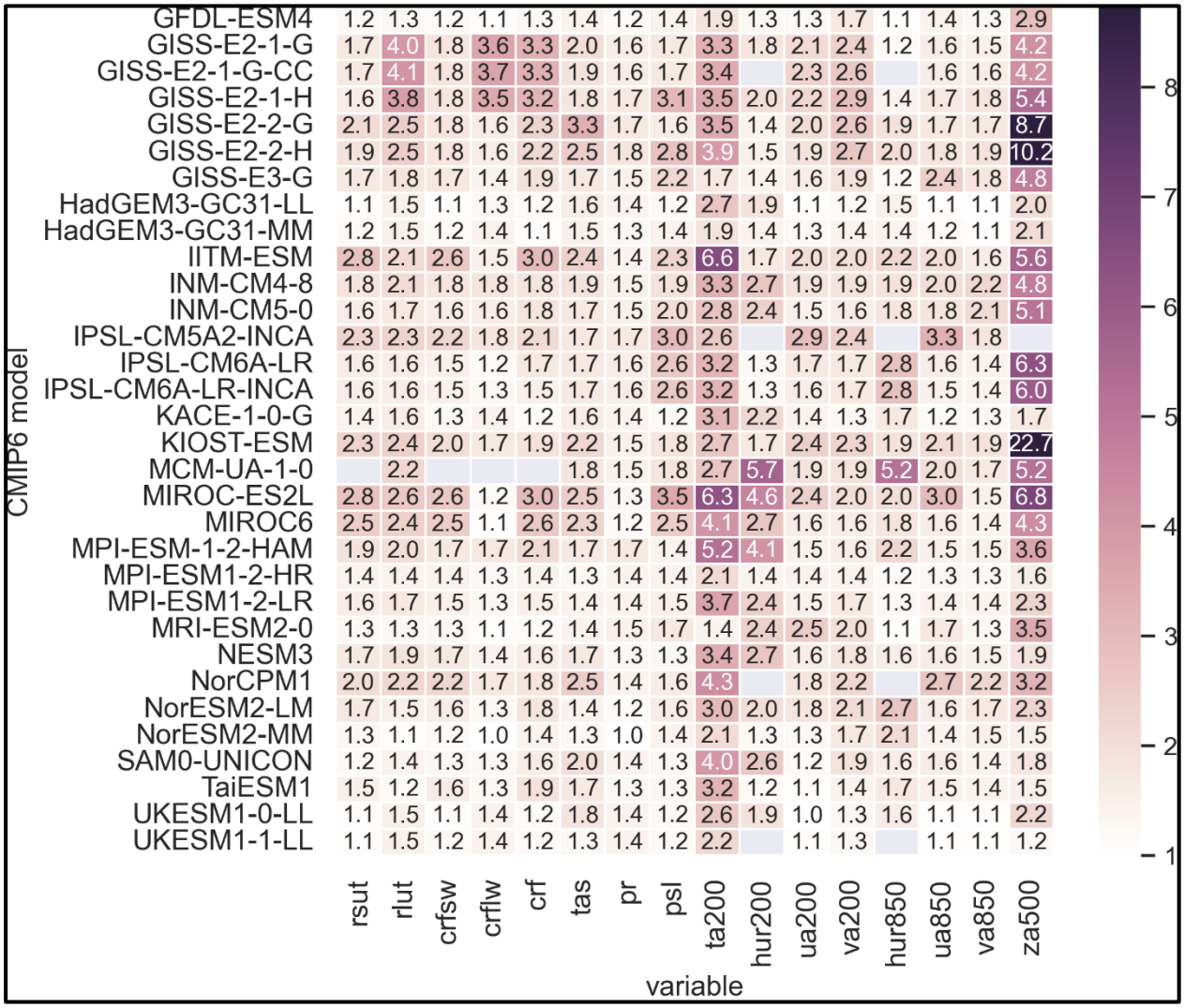

***Figure 7**. Normalised r.m.s. errors in global, annual spatial fields simulated by 32 CMIP6 models for 1981-2000. The scores for each variable are normalised by the score for the best-performing simulation across three ensembles: GC3.05-PPE, CMIP5-13 (shown in Fig. 5) and the 64-member CMIP6 ensemble shown here and in Figure 6. See Figure 5 for further details.*

Amongst GC3.05-PPE members, the worst relative score for an individual variable is 3.3, while for CMIP5-13 members there are six instances of scores exceeding 4.0. If we were to take (say) 3.5 as a maximum tolerable bias for a future subset of screened CMIP6 simulations, we would identify 28 models whose worst scores lie below this threshold leaving 36 candidate models. This indicates that CMIP6 could form an important component of a future update to UKCP-Global, by contributing

simulations that show consistently good performance in simulating global climatology, to a degree competitive with GC3.05-PPE members and somewhat better than found in CMIP5-13.

c. Large-scale modes of variability

We evaluate El Niño-Southern Oscillation events (Figure 8), Atlantic Multidecadal Variability (AMV, Figure 9) and the winter North Atlantic Oscillation (NAO, Figure 10) comparing UKCP-Global with results from 58 CMIP6 models in Figs. 8 and 9, and 64 in Fig. 10. Our assessment is limited to a simple "top-down" comparison of the intensity of variability in the relevant index values, preceded by a brief survey of current understanding of their properties and influences.

ENSO influences climate anomalies in many world regions (Trenberth and Caron, 2000), especially in the tropics through its effects on the Walker circulation. Teleconnections from ENSO to seasonal European climate are weaker, but specific influences are revealed by detailed analysis of observations. For example, Fereday et al. (2008) found negative seasonal correlations between El Niño and a circulation pattern resembling the positive phase of winter NAO, except for November–December which shows the opposite relationship. In addition, Toniazzo and Scaife (2006) found that only weak or moderate El Niños show the canonical negative NAO response in winter, while the strongest events drive high pressure to the west of Europe that results from a tropospheric wave train originating from the tropical Atlantic.

The NAO and AMV indices are important drivers of variability in European climate (e.g. Hurrell et al., 2003; Sutton and Dong, 2012; O'Reilly et al., 2017; Rousi et al., 2020). The positive (negative) phases of NAO drive higher (lower) winter precipitation and temperatures in northern Europe. For AMV, notable influences on the UK include (in the positive phase) dry precipitation anomalies in spring and wet anomalies in summer. AMV also exerts broader influences on decadal and longer time scales that include rainfall in northeast Brazil and the Sahel, Atlantic hurricanes and North American (as well as European) summer climate (Knight et al., 2006).

AMV is associated with low-frequency variability in the ocean circulation in the North Atlantic (Delworth and Mann, 2000), but understanding the relative influences of internally-generated and externally-forced variability remains an active topic of research. Booth et al. (2012), Bellucci et al. (2017) and Bellomo et al. (2018) identify an external influence from anthropogenic aerosol forcing in climate model simulations of the 20th century, while evidence of volcanic influence is also reported in studies of observations and models (e.g. Swingeduow et al., 2017; Mann et al., 2021).

Internal NAO-driven variability contributes significantly to uncertainty in projected European anomalies during the next few decades, but in historical data the dominance of internal variability makes it difficult to identify any potential forced signal (Deser et al., 2017). Blackport and Fyfe (2022) do find positive trends in observed NAO and North Atlantic jet stream data from 1951-2020. These lie outside (for the jet data) and at the top end of (for NAO data) ranges found in a large ensemble of coupled model simulations. Blackport and Fyfe conclude that the discrepancies could arise from either

the presence in observations of an anthropogenically-forced signal missing from the models, or from the failure of models to simulate sufficient internal variability in North Atlantic-European circulation on multidecadal time scales.

On the decadal time scale, Ineson and Scaife (2011) find a model response to minima in solar forcing resembling the negative phase of NAO. O'Reilly et al. (2019) suggest that observed AMV can (to a substantial degree) be explained as an ocean response to accumulated atmospheric forcing from NAO. Evidence from initialised decadal predictions supports this conclusion (Smith et al., 2020b). O'Reilly et al. (2019) find a weaker forcing from the NAO in model simulations, compared to observations. However, Lai et al. (2022) report that NAO is the dominant driver of AMV in HadGEM3-GC3.1-MM (the version using a horizontal atmospheric resolution of ~60km, shared by GC3.05-PPE members), in a 500-year pre-industrial simulation containing fixed external forcing.

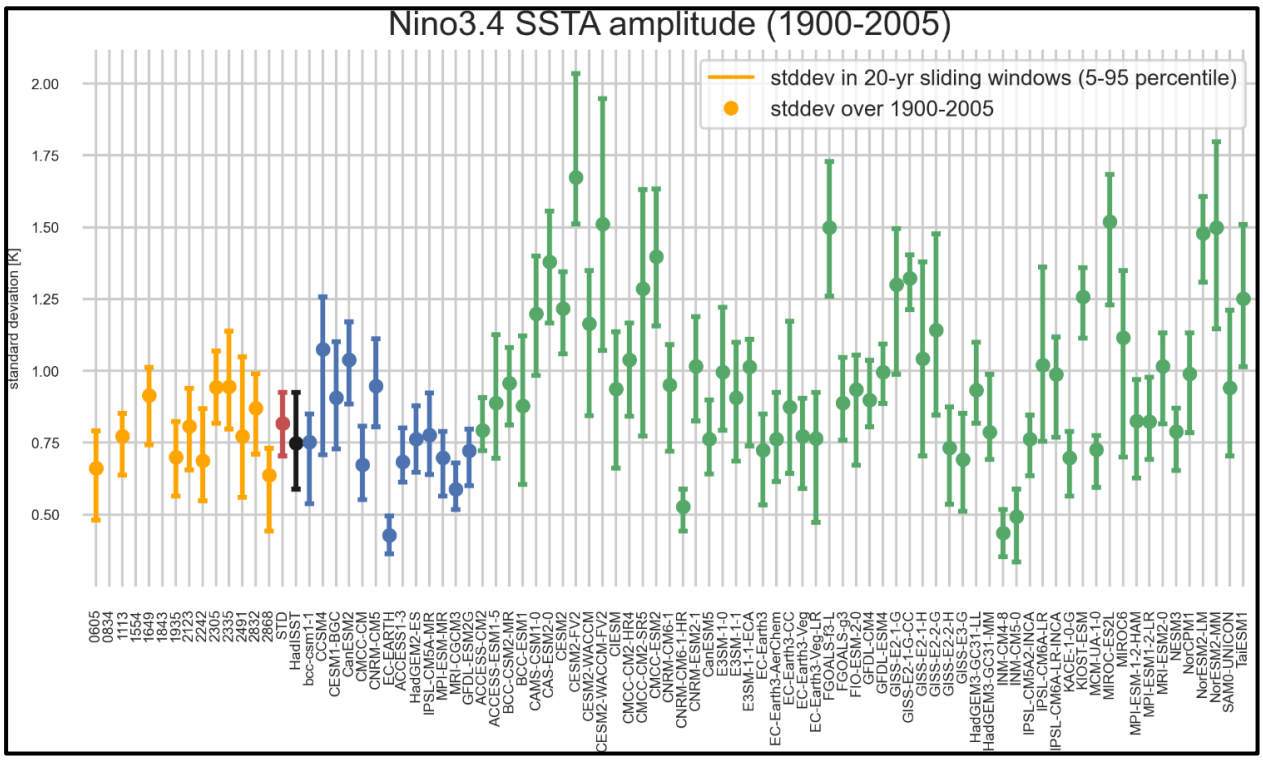

***Figure 8**. Intensity of El Niño events, measured as the standard deviation of monthly anomalies of the NINO3.4 index (°C, defined in main text). NINO3.4 values are provided for each member of the GC3.05-PPE (orange, with STD in red) and CMIP5-13 (blue) ensembles (excluding EC-EARTH), and from an ensemble of 58 CMIP6 models (green). The observed value (black) is obtained from the HadISST dataset (Rayner et al., 2003). In each case, the standard deviation is calculated by pooling linearly detrended time series for each month of the year during 1900-2005. Whiskers show 5th-95th percentile sampling ranges for each standard deviation, calculated by applying a 20-year sliding window to the pooled monthly anomalies.*

We quantify ENSO intensity using monthly SST anomalies in the NINO3.4 region of the equatorial Pacific (Trenberth, 1997). AMV is defined as annual means of North Atlantic SST (averaged over 0-60°N and 0-80°W) minus global mean SST (Trenberth and Shea, 2006). Winter NAO is defined as the December-February mean difference in sea-level pressure (slp) between Iceland and Gibraltar. This

"Atlantic pressure gradient" definition of McSweeney and Yamazaki (2020) differs from the standard index definition (Hurrell, 1995) by avoiding normalisation of annual slp differences by their standard deviation. This is done to include the absolute magnitude of temporal slp variability in the model assessment metric.

In each case, time series of values from 1900-2005 are linearly detrended to remove (approximately) forced signals on the centennial time scale, and then used to calculate the standard deviation of index variability. These are shown as dots in Figs. 8-10. In the case of NINO3.4, the standard deviation is calculated from pooled time series of monthly anomalies, whereas one value per year is available for the NAO and AMV calculations. The circles represent our best estimate of the long-term average intensity, while the whiskers show associated 5^{th}-95^{th} percentile uncertainty ranges. These ranges reflect the influence of low frequency modulation in the intensity of monthly or annual anomalies, estimated by applying a sliding window to the data of 20 years for NINO3.4 and NAO and 50 years for AMV.

For NINO3.4 (Fig. 8), the uncertainty ranges for UKCP-Global simulations show significant overlap with the observed range (black whiskers), the sole exception being EC-EARTH for which the simulated intensity is too low. However, the best-estimate intensities lie above the upper limit of the observed range for two PPE members and three CMIP5-13 models. CMIP6 models show a wide range of ENSO amplitudes: Three models show clear negative biases and eleven show clear positive biases (based on lack of overlap with the observed range), while 34 models simulate a best-estimate value above the observed range. This NINO3.4 metric could offer useful discriminatory power as a screening criterion in selection of a future subset of CMIP6 models for use in impacts studies, alongside other model assessment variables covering a range of influences on global and European climate (section 4d).

For AMV (Fig. 9), every CMIP6 model exceeds the observed standard deviation. In CMIP5-13, five members lie close to the observed value, the remainder simulating larger intensities than observed. GC3.05-PPE simulates a range of best-estimate values that encompasses the observations, avoiding the large positive biases seen in many of the multi-model simulations but including four members with weak AMV intensities. The largest uncertainty ranges are also found in members of the CMIP6 ensemble, some of which exceed the observed range considerably. This indicates large variability on time scales up to 50 years.

On time scales of 50 years or longer, Mavilia et al. (2018) find that observations and some climate models show spectra revealing larger variability than would be expected from a red noise process. They also conclude that longer time series are required to estimate spectral power reliably for periods exceeding 30 years, and that (partly for this reason) the diagnosed characteristics of AMV can be non-stationary when calculated from longer (multi-century) model simulations. In Figure 9, the intensity calculated from the full time series lies above or near the top end of the uncertainty ranges derived from the 50-year sub-samples, in observations and some simulations. Consistent with the results of

Mavilia et al. (2018), this suggests the presence of a significant component of AMV variability on the centennial time scale.

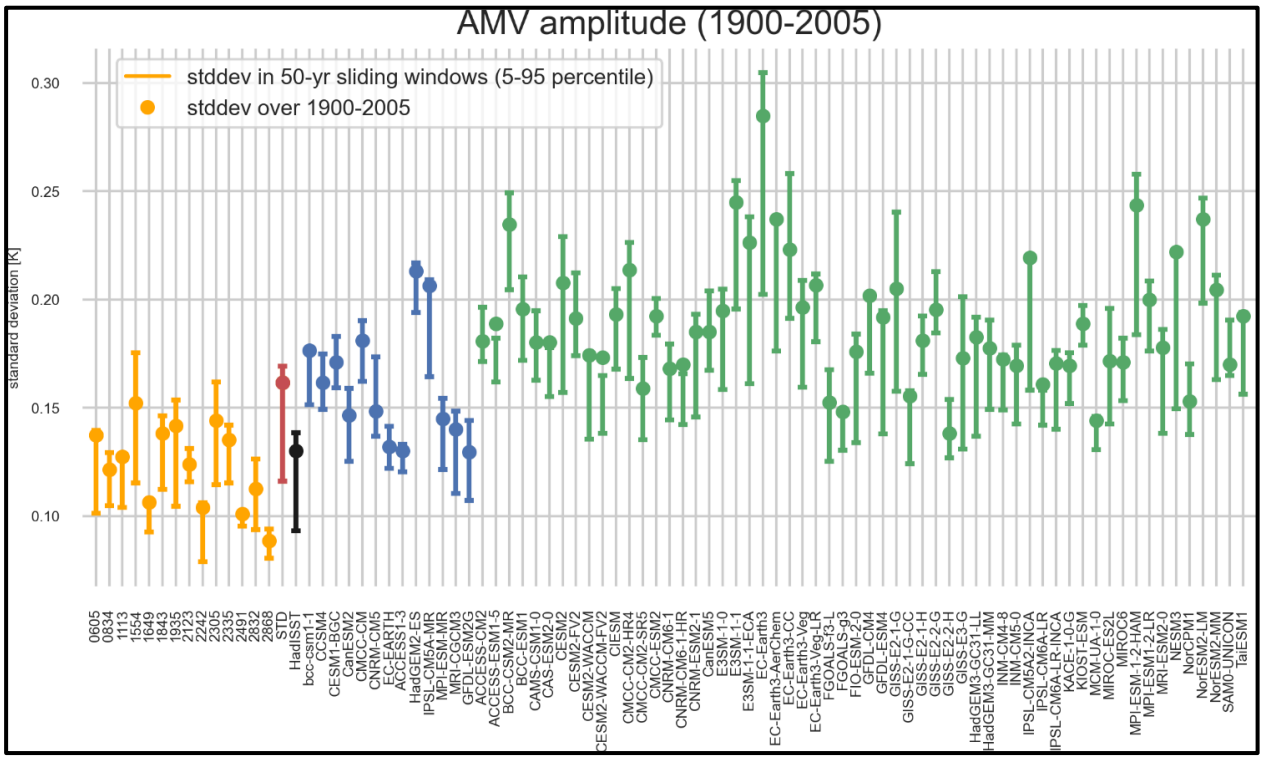

***Figure 9**. Standard deviation of the annual mean AMV index (defined in main text), calculated from time series for 1900-2005 following linear detrending. AMV values are provided for each member of the GC3.05-PPE (orange, with STD in red) and CMIP5-13 (blue) ensembles (excluding EC-EARTH), and from an ensemble of 58 CMIP6 models (green). The observed value (black) is obtained from HadISST. Whiskers show 5th-95th percentile sampling ranges for each standard deviation, calculated by applying a 50-year sliding window to the AMV time series.*

In some simulations, long-term climate drift may contribute to AMV on centennial and longer time scales. For example, Ridley et al. (2022) find that secular adjustment of the deep ocean to a small imbalance in the planetary radiation budget causes a step change of ~0.5°C in global surface temperature in HadGEM3-GC3.1-LL after ~500 years of a simulation with constant external forcing. This is caused by the onset of deep convection in the Weddell and Ross Seas. Therefore, some coupled models may need to be spun up for 1000 years or more to achieve an initial state sufficiently stable to avoid such adjustments during their climate change scenario experiments. These caveats imply that the results from Figure 9, while useful as an indication of model performance, should be treated with caution in a more formal screening or weighting exercise.

All UKCP-Global simulations exceed the observed amplitude of interannual variability in winter NAO (Fig. 10), though the associated uncertainty ranges overlap with the observed range in all but three cases. Most CMIP6 models also show positive biases, though a few ensemble members reproduce the observed value closely while GISS-E2-2-H simulates a level weaker than observed. Many simulations reveal broader uncertainty ranges than observations. Our uncertainty ranges relate to 20-year time windows. However, Woollings et al. (2015) find clear physical differences between NAO variability on

interannual-decadal (<30 year) and multidecadal (>30 year) timescales: On the shorter timescale the NAO is dominated by variations in the latitude of the North Atlantic jet and storm track, whereas on the longer timescale it represents changes in their strengths. In the HiGEM model, Woollings et al. (2015) identify SSTs in the sub-polar gyre (associated with AMOC variability) as an important influence on these multidecadal variations. Given the importance of winter NAO as a driver of variability in European climate, establishing a suitable tolerance level for simulation biases may be an important consideration in the selection of CMIP6 models for impacts analysis (section 4d).

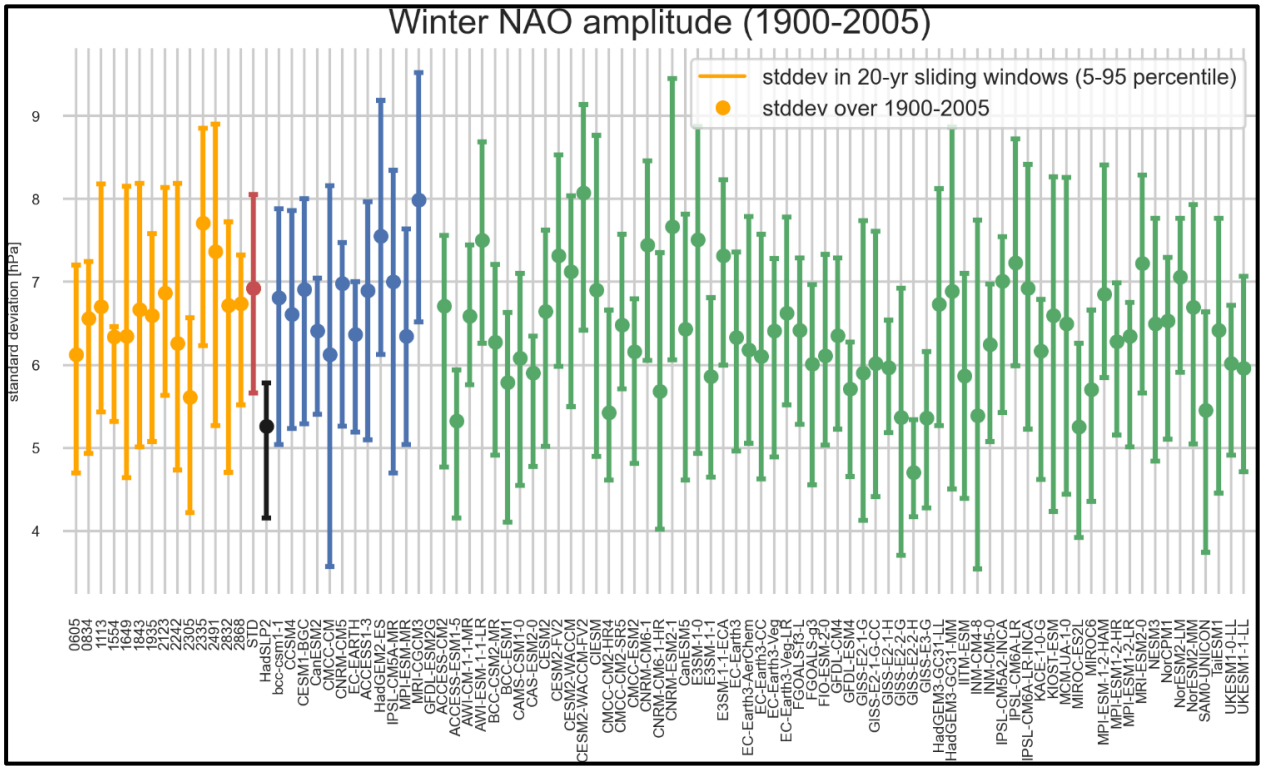

***Figure 10**. Standard deviation of the winter NAO index (hPa, defined in main text), calculated from time series for 1900-2005 following linear detrending. NAO values are provided for each member of the GC3.05-PPE (orange, with STD in red) and CMIP5-13 (blue) ensembles (excluding EC-EARTH), and from an ensemble of 64 CMIP6 models (green). The observed value (black) is obtained from HadISST. Whiskers show 5th-95th percentile sampling ranges for each standard deviation, calculated by applying a 20-year sliding window to the NAO time series.*

d. Historical changes in global mean surface temperature

Observed changes in GMST since the industrial revolution are influenced by factors that also play a key role in determining the magnitude of future changes. These include anthropogenic forcing from greenhouse gases and aerosols, and climate system responses in the form of ocean heat uptake and climate feedbacks (e.g. Smith et al., 2021, Charles et al., 2020). The magnitude of anthropogenic aerosol forcing remains highly uncertain (e.g. Smith et al., 2021), and is influenced by natural emissions that are also uncertain (Carslaw et al., 2013). Volcanic eruptions and internal variability affect the diagnosis of feedbacks from the historical record, particularly through their effects on global patterns of SST anomalies (Gregory et al., 2020). Episodes featuring relatively warm SST anomalies in the tropical western Pacific, with cooler anomalies to the east, are associated with periods of relatively weak

feedback. Due to this “pattern effect”, feedback estimates derived from 1981-2010 indicate lower sensitivity to greenhouse gas forcing than estimates derived from earlier decades in the historical record, or from simulations of the response to large increases in greenhouse gas forcing (Andrews et al., 2022).

While simulated historical and future changes in GMST are correlated, the relationship therefore involves significant uncertainties (e.g. Tokarska et al., 2020), and is likely to depend on which parts of the historical record are used to derive the observational constraint (e.g. Nijsse et al., 2020). For example, the prescribed biomass burning dataset used to force CMIP6 models contained a spurious increase in interannual variability during 1997-2014. This led to additional warming in the northern hemisphere extratropics (and hence GMST) in the CESM2 model, due to an increase in net downward short-wave radiation seen also in other CMIP6 models (Fasullo et al., 2022). Nevertheless, comparison between observed and simulated changes in GMST is an important test of model credibility, underlined by the decision in AR6 to present an observationally-constrained range for future GMST changes (Lee et al., 2021). This was obtained by synthesising three approaches based on different methods and using historical surface temperature changes from different periods (Liang et al., 2020; Tokarska et al., 2020; Ribes et al., 2021).

Here, we defer consideration of formal constraints for future work (see section 1), showing unconstrained results from 38 CMIP6 models alongside UKCP-Global simulations in Figure 11 (top panel). The model GMST data are simple global averages of annual surface air temperature, expressed as anomalies relative to 1900-1930. Two datasets of verifying observations are provided (Cowtan and Way, 2014; Morice et al., 2021), both of which blend terrestrial surface air temperature data with SSTs to construct GMST values. In CMIP5 model data, blended estimates of GMST for recent decades are slightly cooler than those based on air temperature alone, the average difference amounting to ~0.1°C by 2010 (Richardson et al., 2016). Masking to exclude regions of missing data in observational datasets was also found to reduce CMIP5 estimates of warming since ~1990. The datasets used in Fig. 11 apply statistical infilling techniques to adjust for this masking bias. The Cowtan and Way (2014) data (their version 2.0) was created by infilling the HadCRUT4 dataset of Morice et al. (2012) to create globally complete fields. The Morice et al. (2021) values use the newer HadCRUT5 dataset: Compared to HadCRUT4, it extends coverage in data-sparse regions and provides an improved analysis in data-rich regions, but some areas of missing data remain. The (relatively small) differences between the two observed time series illustrate the impact on the observed anomalies of differing data coverage and processing strategies.

By 2010-2020, both datasets show an overall warming close to 1.0°C. For this decade, AR6 concludes that global surface temperature increased by 1.09°C (relative to the 1850-1900 baseline used in their assessment), with a 90% uncertainty range of 0.95 to 1.20°C (Arias et al., 2021). The recent observed

warming in Fig. 11 lies well within the envelope of simulated changes, for both the CMIP6 ensemble and UKCP-Global.

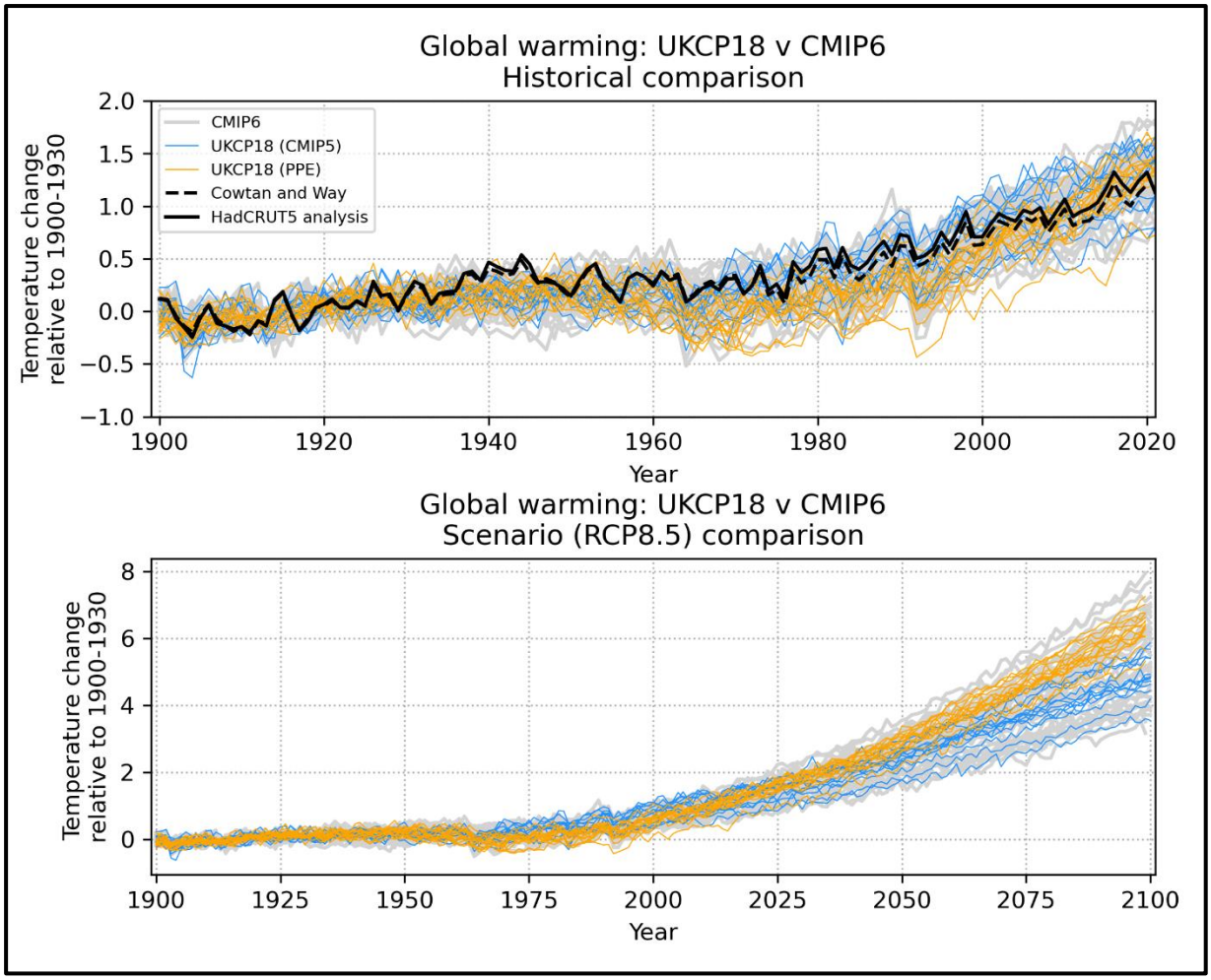

***Figure 11**. Time series of annually and globally averaged surface air temperature anomalies relative to a 1900-1930 baseline. Lower panel shows results from 1900-2100, from members of the GC3.05-PPE (orange) and CMIP5-13 (blue) ensembles contributing to the set of 28 UKCP-Global projections, and from an ensemble of 38 CMIP6 models (grey). The historical simulations account for external forcing due to anthropogenic and natural agents, switching to the RCP8.5 scenario after 2005 (UKCP-Global projections) and to the SSP5-8.5 scenario after 2014 (CMIP6 projections). The upper panel shows the same data to 2020, using a smaller y-axis temperature range and adding verifying observations. These are taken from HadCRUT5 (solid black line, Morice et al. 2021) and Cowtan and Way (2014) (dashed black line). Both datasets combine SST data with air temperatures over land, using statistical techniques to estimate results in regions of missing data. See text for further details.*

The ranges of historical warming in UKCP-Global and CMIP6 models are similar throughout, apart from one GC3.05-PPE member that shows a cooler outcome than other simulations between 1990 and 2010. This is one of two members that exhibit a significant long-term weakening of AMOC, along with a cooling in GMST, when run for 200 years using anthropogenic and natural external forcings held fixed at early-industrial values for 1900 (Sexton et al., 2020). This climate drift is likely to have contributed to the particularly cool outcome simulated for recent decades by this PPE member. More generally, most PPE members lie below the observations during 1950-1990. This is probably due (at least partly) to strong

anthropogenic aerosol forcing in the northern hemisphere, with strong cooling following major volcanic eruptions also a likely influence (Yamazaki et al., 2021).

During 1950-1990, CMIP5-13 and CMIP6 ranges encompass observations. Bock et al. (2020) find a mean cool bias of -0.07°C during this period, from an analysis in which CMIP6 data was masked to match data coverage in HadCRUT4. Zhang et al. (2021) analyse six CMIP6 earth system models (three of which are included in Fig. 11), finding excessive cooling in the northern hemisphere during this period. This "pothole cooling" is attributed mainly to strong forcing from aerosol-cloud interactions. Zhang et al. (op. cit.) also assess corresponding coupled -ocean atmosphere simulations containing less sophisticated representations of aerosol and/or tropospheric chemistry processes, four of which are included in Fig. 11: Most of these also show a pothole cooling during 1960-1990, albeit less pronounced than the corresponding earth system model configurations.

After 1990 GC3.05-PPE members tend to warm more rapidly than observations, resulting in levels of response broadly consistent with observations by 2010-2020 (excepting the single member discussed above). The multi-model average warming in CMIP5-13 and CMIP6 is also close to observations by 2010-2020, while the range of anomalies across all three ensembles covers ~0.7-1.8°C. The observed increase in GMST was slower during 1998-2012 compared with the previous and subsequent periods. The muted warming during 1998-2012 (often termed the "hiatus", e.g. Trenberth, 2015) was assessed by AR6 to be temporary, with very high confidence (Arias et al., 2021).

e. Summary

The evaluation diagnostics of sections 3b-d demonstrate an overall level of performance from the UKCP-Global simulations that is competitive with the CMIP6 ensemble. Both sets of simulations exhibit a range of biases across their individual members, with evidence of a systematic component to the biases for some variables. The results also demonstrate potential to identify a subset of the most skilful CMIP6 models that could contribute to an improved UKCP-Global dataset in the future.

4. Comparison of CMIP6 and UKCP projections: Future changes

a. Future changes in global mean surface temperature

Figure 11 (bottom panel) shows projected future changes to 2100 for annual GMST (relative to 1900-1930), comparing UKCP-Global results for RCP8.5 with 38 CMIP6 projections for SSP5-8.5, a scenario with similar total radiative forcing by 2100 (section 2d). Neither set of projections is formally constrained by historical GMST changes (see section 3d), though their status differs through application of climatology-based screening to select UKCP-Global but not CMIP6 members (section 2b cf 2c). Both sets of projections provide a broad range of changes that grows with time as the applied forcing and modelled responses develop through the 21st century. By 2100, the combined set of changes (across all 66 simulations) ranges from ~3.0-8.0°C, with a central estimate of ~5.0°C.

From 1990 onwards, GC3.05-PPE members warm faster than most CMIP5-13 counterparts, hence there is limited overlap between their envelopes of change by the end of the 21st century. Therefore, combination of the perturbed parameter and multi-model ensembles was essential to achieve a diverse range of GMST outcomes in UKCP-Global, with the PPE and CMIP5-13 supplying the upper and lower portions respectively. The stronger warming in GC3.05-PPE is driven by weaker levels of total climate feedback (and hence higher values of ECS[2]) compared to most CMIP5-13 members (Yamazaki et al., 2021). This restricted range of PPE feedbacks arises partly from a performance constraint exerted by the simulation of longwave cloud radiative effect, biases in which led to exclusion of potential PPE members simulating lower values of ECS (Rostron et al., 2020). Yamazaki et al. (2021) identify use of a range of CO_2 concentration pathways as a second driver of high warming in GC3.05-PPE, since those sampled by most PPE members lie above the standard pathway used in all CMIP5-13 members (Fig. 2) and hence supply a larger radiative forcing.

Another potential driver of the large projected GMST changes in GC3.05-PPE is the efficiency of global ocean heat uptake. GC3.05 uses the NEMO (Nucleus for European Modelling of the Ocean) model, in common with the HadGEM3-GC31-LL configuration of HadGEM3-GC3.1. However, the ACCESS-CM2 model (Bi et al., 2020), which uses the same atmosphere component as HadGEM3-GC31-LL but a different ocean model, simulates stronger ocean heat uptake and therefore a lower characteristic rate of transient warming (Liu et al., 2023).

The CMIP6 range of GMST projections is slightly broader than that of UKCP-Global. This underlines the importance of considering a diverse range of warming outcomes in applications using UKCP data, particularly where these are framed to consider time-dependent impacts conditioned on a particular choice of emissions scenario (e.g. Arnell et al., 2021).

In contrast to Figure 11, projected GMST changes from UKCP-Probabilistic are observationally constrained using historical surface temperature changes, by assigning Bayesian weights to alternative outcomes (Sexton et al., 2012; Harris et al., 2013; Murphy et al., 2018). For 2080-2099 relative to 1981-2000, applying the constraints reduces the 5-95% range for GMST changes under RCP8.5 emissions from ~2.3-6.6°C to ~2.7-5.6°C, the upper limit lying well below that of those of the unconstrained ranges in Fig. 11. While the constrained GMST projections in AR6 are derived exclusively from historical GMST changes (section 3d), the UKCP-Probabilistic results include also constraints based on recent climatology. These play a role alongside historical temperature changes in reducing the projected range (Hegerl et al., 2021). Challenges relevant to use of screened or constrained projections are discussed further in section 4d.

[2] Equilibrium climate sensitivity (ECS) is a standard benchmark of the climate system response to external forcing, defined as the steady-state response of GMST to a sustained doubling of atmospheric CO_2 concentration (IPCC, 2021: Annex VII: Glossary). Assuming the climate system is stable, the total climate feedback parameter (λ) must be negative. It determines ECS through the relationship $ECS = -\Delta F/\lambda$, where ΔF is the radiative forcing arising from a doubling of CO_2 (e.g. Sherwood et al., 2020). Hence, weaker (less negative) values of λ result in higher ECS.

b. Future changes for the UK

Figure 12 shows time series of winter precipitation anomalies (relative to 1981-2000) for England from 41 CMIP6 models, to illustrate basic properties of the CMIP6 dataset. Grey lines are individual seasons from each model, while coloured lines show multi-decadal anomalies estimated using a low-pass Butterworth filter to attenuate variability on time scales of 1-30 years. The spread in the coloured

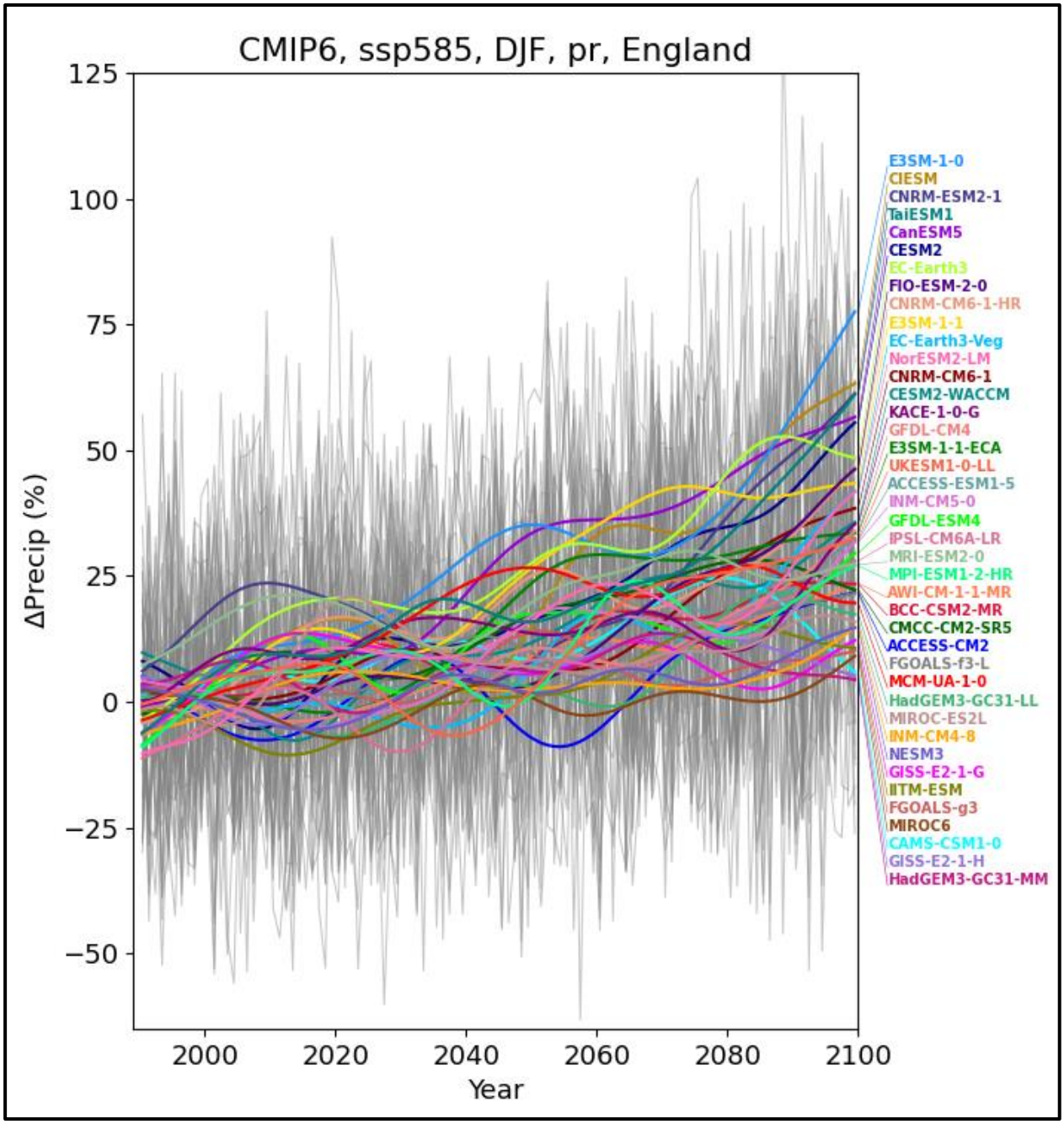

***Figure 12**. Time series of winter precipitation anomalies (%, relative to a 1981-2000 baseline) for England, from 41 CMIP6 models for the SSP5-8.5 scenario. Grey lines show annual anomalies, coloured lines show low-frequency anomalies. The latter are obtained by applying a Butterworth filter to remove variability on 1-30 year time scales. The spread in these reveals the impacts of modelling uncertainties in long-term climate change signals, plus the effects of multidecadal climate variability. The wider spread in the envelope of the grey lines demonstrates the additional effects of variability on 1-30 year time scales.*

lines indicates the impact of different long-term climate responses to the concentration-driven SSP5-8.5 forcing. By 2100, the low-pass filtered responses range from little change to an increase of ~75%. Undulations in each coloured line suggest the presence of significant multi-decadal variability (see section 3c). In a multi-scenario analysis of CMIP6 models, Lehner et al. (2020) find that internal variability dominates the total spread in decadal averages of regional precipitation anomalies during the 2020s, while model uncertainty grows to become the largest contributor by the middle of the 21st century. Scenario uncertainties exert a minor influence at these time scales, but typically explain ~15% of total uncertainty by the end of the century. In a single-scenario setting such as Fig. 12, we may therefore expect internal variability to contribute substantially to the range of low-frequency responses during the next few decades, while the influence of long-term climate change feedbacks is likely to become the main driver of spread at longer lead times. Figure 12 shows a growth of spread beyond the 2040s that appears to corroborate this, however a quantitative analysis of components of response is not attempted here.

The CMIP6 results show several extreme winters during the recent historical period, in the same 50-75% range explored by the wettest observed events during the 20th century (e.g. Fig. 5.6a of Murphy et al., 2018). The largest simulated event during this period reaches 92%, a level not exceeded till after 2070 beyond which several events beyond 100% emerge. Intermittency in the occurrence of record precipitation extremes is also found in hourly events, based on analysis of UKCP-Local projections Kendon et al. (2023). In these UKCP-Local results, accounting for climate change (via the RCP8.5 scenario) leads to an increase of 20% in the occurrence of regional hourly records in winter. For UK average precipitation, the increases in occurrence are 40% for hourly accumulations, and 70% for the seasonal average.

While occurrence of the very wettest winters is noisy in Fig. 12, there is a notable increase in events exceeding 50%, associated with a secular shift of the seasonal distribution towards higher precipitation. Nevertheless, very dry winters (precipitation 50% or more below historical climatology) continue to occur through the 21st century. Qualitatively, these changes in seasonal distributions are consistent with those from UKCP-Probabilistic (Sexton and Harris, 2105; Murphy et al., 2018).

Figure 13 shows corresponding results for surface air temperature in summer. Results for individual summers show that anomalies of 2°C become commonplace by the middle decades of the 21st century, and unusually cool by 2100. This is consistent with results reported by McCarthy et al. (2019), using UKCP-Probabilistic to assess the changing risk of exceeding the joint UK record experienced in 2018 (of ~1.7°C, calculated relative to 1982-2010 in their case).

The range of multidecadal changes becomes very wide by 2100 (~2-12°C). The top end is inflated by three configurations of the E3SM1 model. The most extreme changes should not be ruled out of

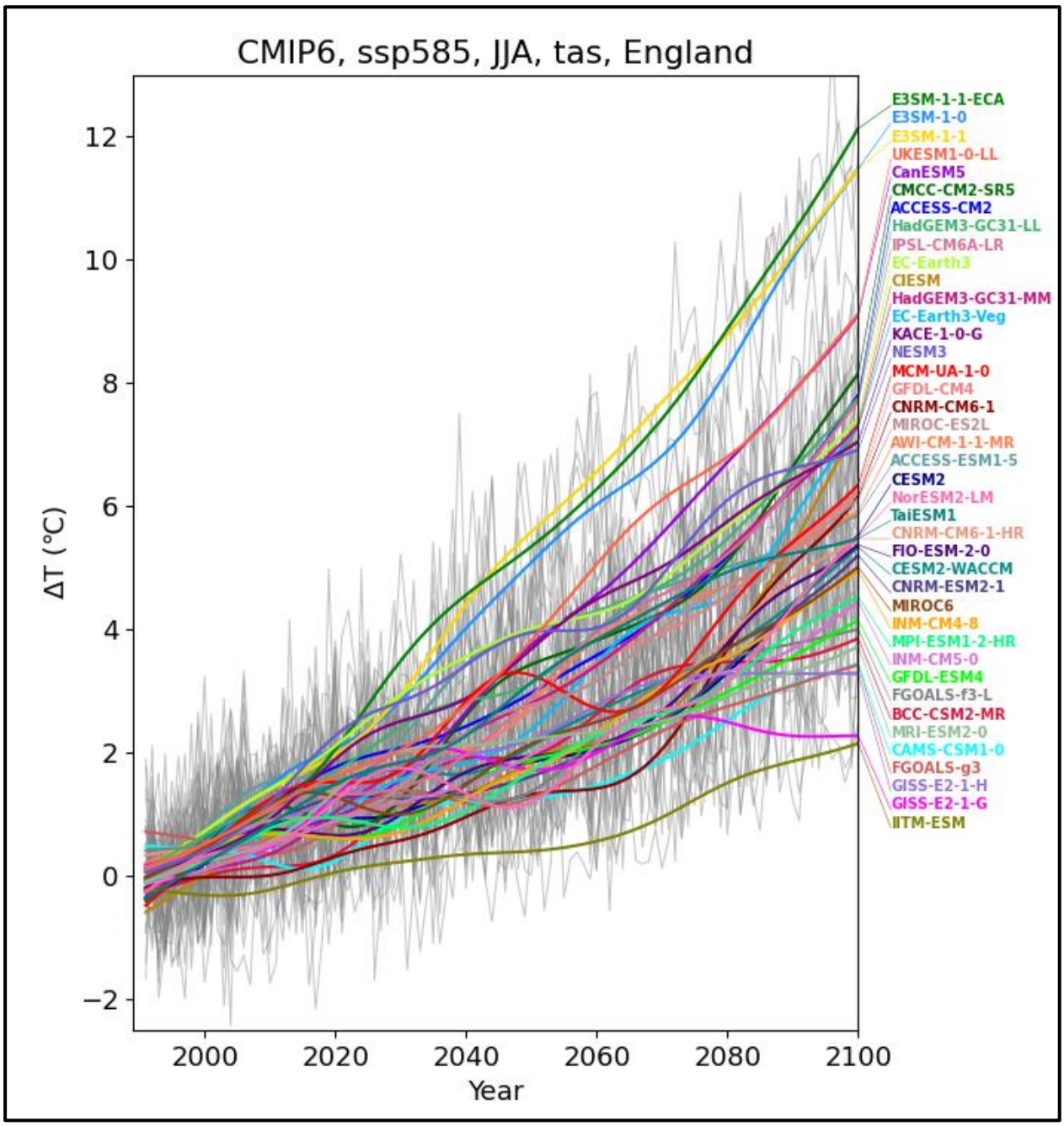

***Figure 13**. Time series of anomalies in summer surface air temperature (°C, relative to a 1981-2000 baseline) for England, from 41 CMIP6 models for the SSP5-8.5 scenario. As in Figure 12, grey lines show annual anomalies, coloured lines show low-frequency anomalies following application of a Butterworth filter to remove variability on 1-30 year time scales.*

consideration, simply on the basis that they are outliers compared to other models. However, E3SM simulates a weak AMOC and a strong cool bias in North Atlantic SSTs, accompanied by excessive sea -ice in the Labrador Sea (Golaz et al., 2019). This may contribute to an extremely strong warming projected at high latitudes in the North Atlantic/European sector (not shown), in comparison with other CMIP6

models. E3SM1 also simulates the highest transient climate response (TCR[3]) and the lowest ocean heat uptake efficiency amongst 31 CMIP6 models analysed by Liu et al. (2023). They find that CMIP6 models with fresh biases in surface salinity (amongst which E3SM exhibits the largest) may be less plausible because excessive vertical stratification in ocean density restricts heat uptake, leading to transient GMST warming that is too rapid. This is likely to play a role in enhancing the UK warming in E3SM1.

Similarly, the strong surface and atmospheric feedbacks implied by a high ECS of 5.3°C may also contribute to the large UK changes in E3SM1. However, several other CMIP6 models (including HadGEM3-GC3.1) also possess ECS values exceeding 5°C, a level assessed as very unlikely but plausible by AR6 (Arias et al. (2021), see also discussion in section 4d). A future screening exercise could potentially lead to exclusion of the E3SM1 simulations, based partly on the biases described above. However, this is left for future consideration.

Projected 20-year mean changes for Scotland and England are shown in Figure 14, for 2061-2080 relative to 1981-2000. For CMIP6 models, the surface temperature and precipitation changes confirm that the broad ranges of response shown in Figs. 12 and 13 apply to Scotland as well as England. CMIP6 models with high GMST warming tend to produce larger regional warming in the UK, through the expected influence of global changes on regional responses (e.g. Frieler et al., 2012). In addition, Pan et al. (2023) find that CMIP6 models using the NEMO ocean tend to simulate a larger polar amplification of projected warming in the northern hemisphere in winter, compared to those using a different ocean component. This is driven by larger increases in heat transport through the Barents Sea opening and stronger increases in convective mixing in the Arctic in NEMO models, which include the Met Office submissions to CMIP6. The warmest CMIP6 responses in Fig. 14 all feature high GMST responses and/or use of NEMO. HadGEM3-GC3.05 also used NEMO, which may contribute (alongside large GMST changes) to the large UK temperature changes found in GC3.05-PPE.

There is invariably substantial overlap between the CMIP6 ranges and those of UKCP-Global and UKCP-Probabilistic, indicating a reasonable level of consistency between the CMIP6 and UKCP projections. For winter precipitation in England, for example, each dataset suggests that an increase is almost certain, with potential values extending to ~40% at the upper ends of the ranges. The advent of CMIP6 information therefore reaffirms that projections derived from contemporary climate models involve substantial uncertainties, that should be accounted for robustly in user applications (Fung et al., 2018).

[3] TCR is defined as the GMST response to a hypothetical 1% per year (compounded) increase in atmospheric carbon dioxide concentration from the pre-industrial level to the time of doubling at year 70 (IPCC, 2021: Annex VII: Glossary). It provides a characteristic measure of the time-dependent response in scenarios featuring steady increases in greenhouse gas forcing, accounting for both global ocean heat uptake and the surface and atmospheric feedbacks that influence ECS.

Nevertheless, some noteworthy differences exist between the three sets of projections:

- The three E3SM1 variants discussed above contribute the outlying surface temperature responses in Fig. 14, for both countries and seasons.
- For winter surface temperature in Scotland, relatively few CMIP6 simulations and no UKCP-Global outcomes lie within the lowest quartile of the UKCP-Probabilistic distribution. This results partly from broader sampling of low-end GMST responses in UKCP-Probabilistic (Murphy et al., 2018), and partly from the representation of patterns of warming per unit rise in GMST (Fig. 15, discussed below).
- For winter precipitation in Scotland, all but one of the CMIP6 and UKCP-Global members suggests a maximum increase of ~30%, while UKCP-Probabilistic suggests a chance of about 20% of a larger increase.
- For summer rainfall, the CMIP6, UKCP-Probabilistic and UKCP-Global datasets all suggest that a reduction in the 20-year mean is more likely than in increase (for both regions). However, for Scotland the CMIP6 and UKCP-Probabilistic data suggests a higher chance of an increase than UKCP-Global. For England, UKCP-Probabilistic and UKCP-Global results suggest a maximum plausible drying exceeding 50%, whilst for the CMIP6 ensemble the largest drying is slightly less pronounced (~40%). No single climate model ensemble can be assumed *a priori* to cover all possible outcomes. Therefore (pending further investigation) the driest responses in the UKCP datasets should continue to be regarded as plausible outcomes for consideration in impacts analyses.

Figure 14 also includes projections from UKCP-Regional and UKCP-Local. The black lines show the nesting strategy used to produce these limited-area simulations, in which SSTs and lateral boundary conditions from twelve GC3.05-PPE members drove 12km regional climate model simulations for Europe (Murphy et al., 2018; Tucker et al., 2021), which in turn drove twelve 2.2km simulations for the UK (Kendon et al., 2021). Results from both datasets are being used to study spatially-detailed hazards and impacts in future UK climate (e.g. Arnell et al., 2021; Garry et al., 2021; Hanlon et al., 2021; Slater et al., 2022). UKCP-Local also offers new capability to obtain credible projections of phenomena such as hourly precipitation (Chan et al., 2023; Kendon et al., 2023), hail, lightning (Kahraman et al., 2022), extreme windstorms (Manning et al., 2022) and heat stress in urban environments (Keat et al., 2021).

The other projections in Fig. 14 provide an uncertainty context for UKCP-Regional and UKCP-Local. This is because UKCP-Probabilistic and UKCP-Global include results from multi-model ensembles whereas UKCP-Regional and UKCP-Local are currently derived exclusively from perturbed variants of GC3.05. This limits representation of temperature changes in the regional and local ensembles, particularly in summer - UKCP-Probabilistic, UKCP-Global and CMIP6 suggest a change of ~50% or more for a cooler outcome than any of the limited-area simulations. For precipitation, average summer

responses in UKCP-Regional and UKCP-Local all show reductions (with one exception for Scotland), levels of drying for England amounting to ~20-55%. These capture the dry ends of the distributions from CMIP6, UKCP-Probabilistic and UKCP-Global, but not their less dry or positive outcomes.

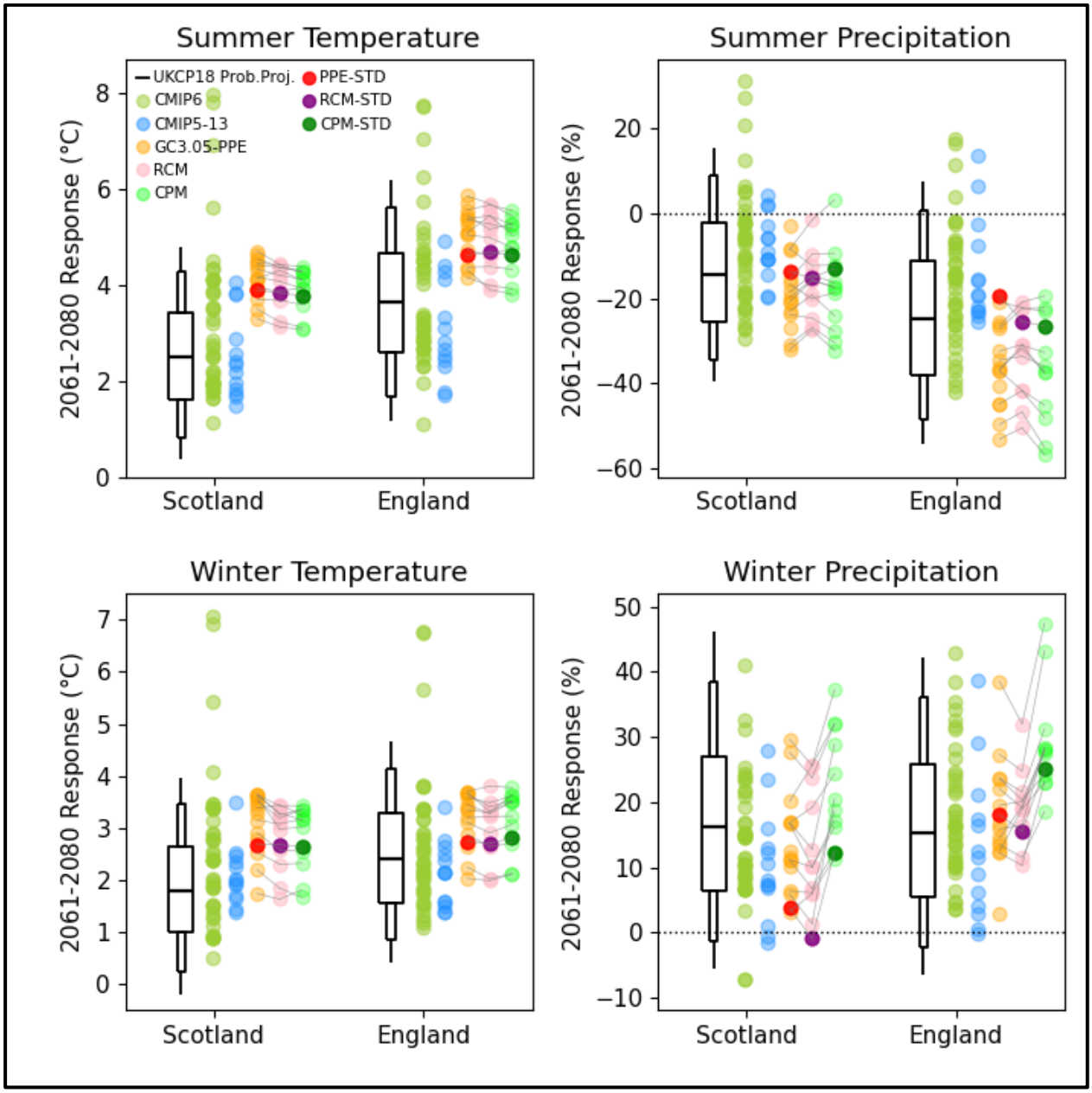

***Figure 14**. Projected changes for 2061-2080 relative to 1981-2000 for Scotland and England in winter (lower panels) and summer (upper panels), for surface air temperature (°C, left panels) and precipitation (%, right panels). Boxes and whiskers show the 5, 10, 25, 50, 75, 90 and 95% probability levels from UKCP-Probabilistic. Orange dots (STD in red) denote members of GC3.05-PPE and blue dots members of CMIP5-13, together comprising the UKCP-Global projections. Green dots show results from the 41 CMIP6 models of Figs. 12 and 13. Also shown for context are the regional climate model projections of UKCP-Regional (pink) and the convection-permitting projections of UKCP-Local (light green). The black lines show which global and regional simulations drive which regional and local simulations, respectively. The members driven by STD are shown in darker shading. These CMIP6 and UKCP projections were forced by the SSP5-8.5 and RCP8.5 emissions scenarios respectively.*

A forthcoming update (planned for late 2023 or early 2024) will add four UKCP-Regional and UKCP-Local projections driven by CMIP5 models (three of which are members of CMIP5-13), introducing some multi-model information. In cases where skilful large-scale predictor variables can be identified (e.g. Chan et al., 2018), it may be possible to emulate a broader set of plausible impacts by building downscaling relationships between UKCP-Regional or UKCP-Local and their driving global simulations (Addison et al., 2022).

In winter, UKCP-Local projects an envelope of increases in precipitation shifted higher than its driving inputs. For England, one UKCP-Local member projects an increase exceeding 45%, higher than any CMIP6 simulation and beyond the 95th percentile of UKCP-Probabilistic. The larger increases in the 2.2km model are associated with an increase in convective showers, triggered over the sea and advected inland (Kendon et al., 2020). In the driving 12km simulations, maritime increases in convective precipitation do not extend over the land, since the showers are parameterised locally and the model lacks a dynamical representation of their movement. The same limitation applies to all the simulations included in UKCP-Global and UKCP-Probabilistic.

Figure 15 shows regional results equivalent to Figure 14, but with the linear influence of global average warming removed by regressing 21st century winter and summer changes against GMST. This allows the influence of projected patterns of response to be isolated, subject to the caveat that potential non-linear influences of GMST on the regional patterns are neglected (Tebaldi and Arblaster, 2014). This presentation is also a useful guide for UK changes expected under prescribed levels of global warming (GWLs), such as the set of derived projections for 2°C and 4°C GWLs provided by Gohar et al. (2018) in UKCP18. Such a characterisation features prominently in AR6 (Arias et al. (2021) and https://interactive-atlas.ipcc.ch/) and the third UK climate change risk assessment (CCRA3, Betts and Brown 2021). Use of GWLs is also a common form of presentation in impacts studies (e.g. Hanlon et al., 2021; Arnell et al., 2021b).

The distributions of normalised response from CMIP6 and UKCP-Probabilistic show a good level of consistency for Scotland, the exception being that CMIP6 does not support the upper quartile of winter precipitation responses found in the pdfs.

For England, there is good consistency in winter for both variables. In summer, there is reasonable overlap of ranges of normalised rainfall response, with the CMIP6 range shifted to slightly less dry responses compared with UKCP-Probabilistic. For summer surface temperature, CMIP6 populates the lower three quartiles of the UKCP pdf, but no CMIP6 simulations lie within the upper quartile. However, the non-normalised CMIP6 projections for English summer do sample the upper quartile of the pdfs (Fig. 14), due to the regional impact of high GMST responses in some CMIP6 members.

In UKCP-Global, normalised responses from CMIP5-13 generally cover much of the CMIP6 range, but with gaps in sampling due to the smaller ensemble size. Summer rainfall in Scotland is an exception,

since several CMIP6 models produce wetter normalised changes than any UKCP-Global members. For winter precipitation and summer temperature in England, one CMIP5-13 member lies clearly above the CMIP6 range in each case. However, these outcomes are plausible (albeit unlikely) according to UKCP-Probabilistic, lying near the 95th percentile of the relevant pdfs.

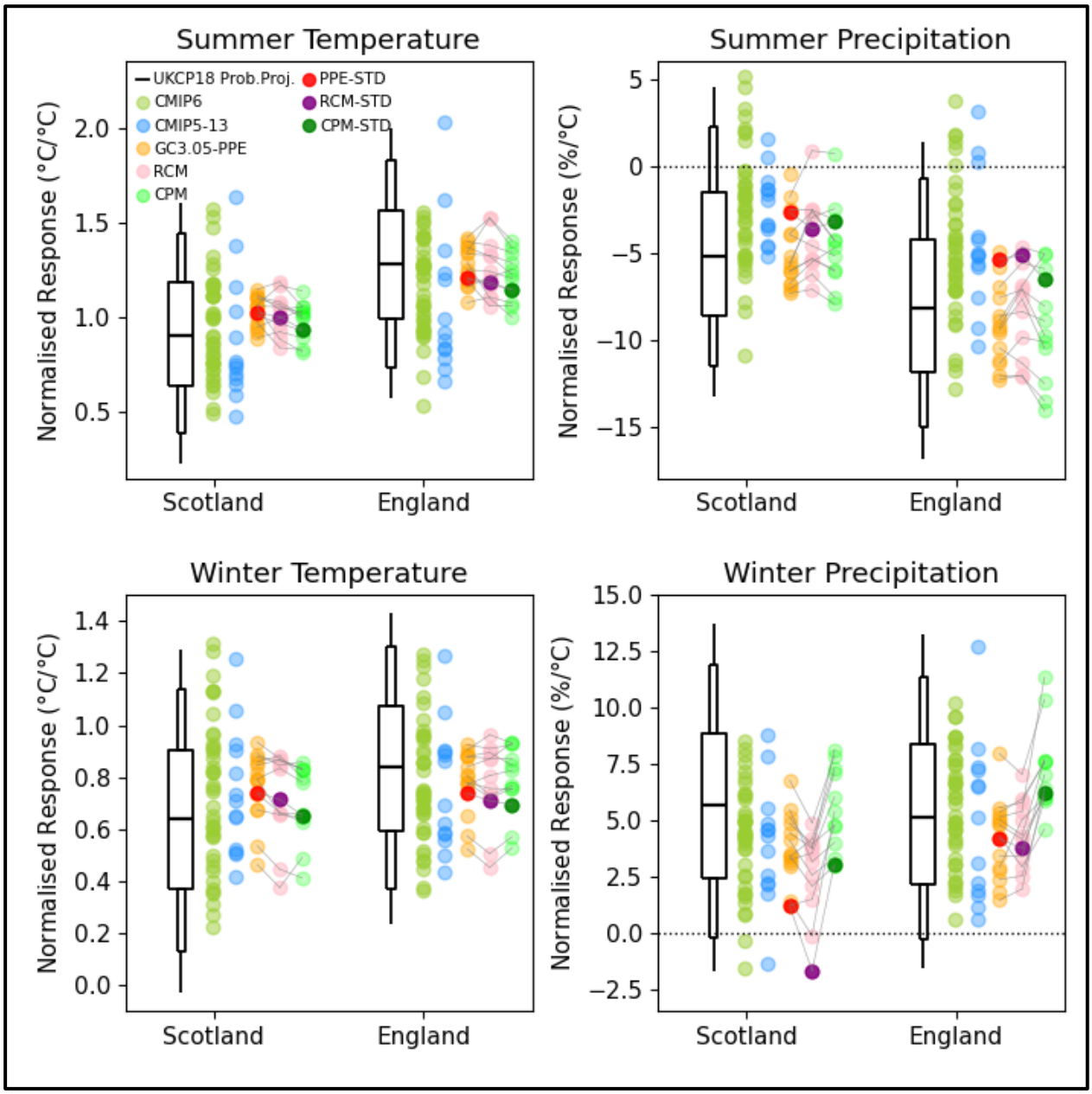

Figure 15. *Projections of normalised changes during the 21st century (relative to 1981-2000) for Scotland and England in winter (lower panels) and summer (upper panels), for surface air temperature (°C, left panels) and precipitation (%, right panels). Normalised changes are calculated by linearly regressing seasonal changes against corresponding annual changes in GMST. Projections are from UKCP-Probabilistic (boxes and whiskers), UKCP-Global (orange and blue dots), UKCP-Regional (pink dots), UKCP-Local (light green dots) and 41 CMIP6 models (darker green dots). See Fig. 14 for further details. These CMIP6 and UKCP projections were forced by the SSP5-8.5 and RCP8.5 scenarios respectively.*

The GC3.05-PPE outcomes are more tightly constrained, particularly for surface temperature in summer. These constraints are transmitted to the UKCP-Regional and UKCP-Local projections, whose normalised temperature responses follow closely those of the driving global simulations. In studies of

regional impacts conditioned on GWLs (e.g. Hanlon et al., 2021), this implies that uncertainty ranges for indicators influenced by average levels of regional warming (such as hot weather hazards) are likely to provide an incomplete picture of future risks. However, central estimates of change for such metrics should be more reliable, since the median normalised response in GC3.05-PPE lies close to those of CMIP5-13, CMIP6, and UKCP-Probabilistic.

For precipitation, ranges of normalised responses in UKCP-Regional follow their driving GC3.05-PPE simulations closely, except for Scotland in winter where the average normalised increase is somewhat smaller. In UKCP-Local, normalised changes in winter are shifted significantly higher compared with the GC3.05-PPE results, consistent with the enhanced contribution of convective precipitation (Kendon et al., 2020, discussed above). In summer, rainfall changes per unit global warming are generally similar to the GC3.05-PPE ranges, although for England two UKCP-Local members project larger normalised drying (approaching 15% per °C) than any of the driving global projections.

In Figure 16 we return to presentation of full (non-normalised) changes, considering simulations driven by the RCP2.6 and SSP1-2.6 scenarios. In this case, results from ten CMIP5-13 members and 35 CMIP6 models are available. All fifteen GC3.05-PPE members are included, but no UKCP-Regional or UKCP-Local results are available for RCP2.6. We consider changes for 2061-2080 relative to 1981-2000 (as in Figure 14). Fung et al. (2020) also show UKCP-Probabilistic and UKCP-Global results for 2081-2100 relative to the same baseline. In these strong mitigation scenarios, GMST changes for 2061-2080 range from ~0.5-2.5°C in UKCP-Probabilistic (Murphy et al., 2018), UKCP-Global (Fung et al., 2020) and CMIP6 models (Tebaldi et al., 2021). These bounds are considerably cooler than those for RCP8.5 and SSP5-8.5 (Fig. 11). Lower increases in GMST play a leading-order role in limiting regional warming (e.g. Frieler et al., 2012), hence the individual England and Scotland changes in Fig. 16 never exceed ~3.2°C. In contrast, there are several outcomes exceeding 5°C for the high emissions scenarios (Fig. 14), especially for England in summer. Note, however, that SSP1-2.6 results are not available for the three E3SM1 configurations projecting extreme warming in Fig. 14.

Under SSP1-2.6 the CMIP6 models project wide ranges of change for all UK variables (Fig. 16), that encompass those found in UKCP-Probabilistic and UKCP-Global. Modelling uncertainties in regional climate response are likely to play the main role in driving the projected ranges of change (e.g. Lehner et al., 2020). Differences in applied radiative forcing are also likely to contribute, for two reasons: (a) The standard RCP2.6 concentration pathway used in CMIP5-13 models gives a radiative forcing ~10% smaller than the SSP1-2.6 pathway used in the CMIP6 ensemble (section 2d) - this may contribute to the higher average levels of warming found in CMIP6 compared with CMIP5-13 (Fig. 16); (b) GC3.05-PPE and UKCP-Probabilistic sample a range of CO_2 pathways in their RCP2.6 distributions. As in the RCP8.5 case (Fig. 2), the standard RCP2.6 concentration pathway lies near the low end of the sampled range of CO_2 outcomes. Therefore, the average radiative forcing applied in CMIP5-13 is smaller than in GC3.05-PPE

and UKCP-Probabilistic, while their sampling of forcing uncertainties contributes alongside response uncertainties to their ranges of projected outcomes.

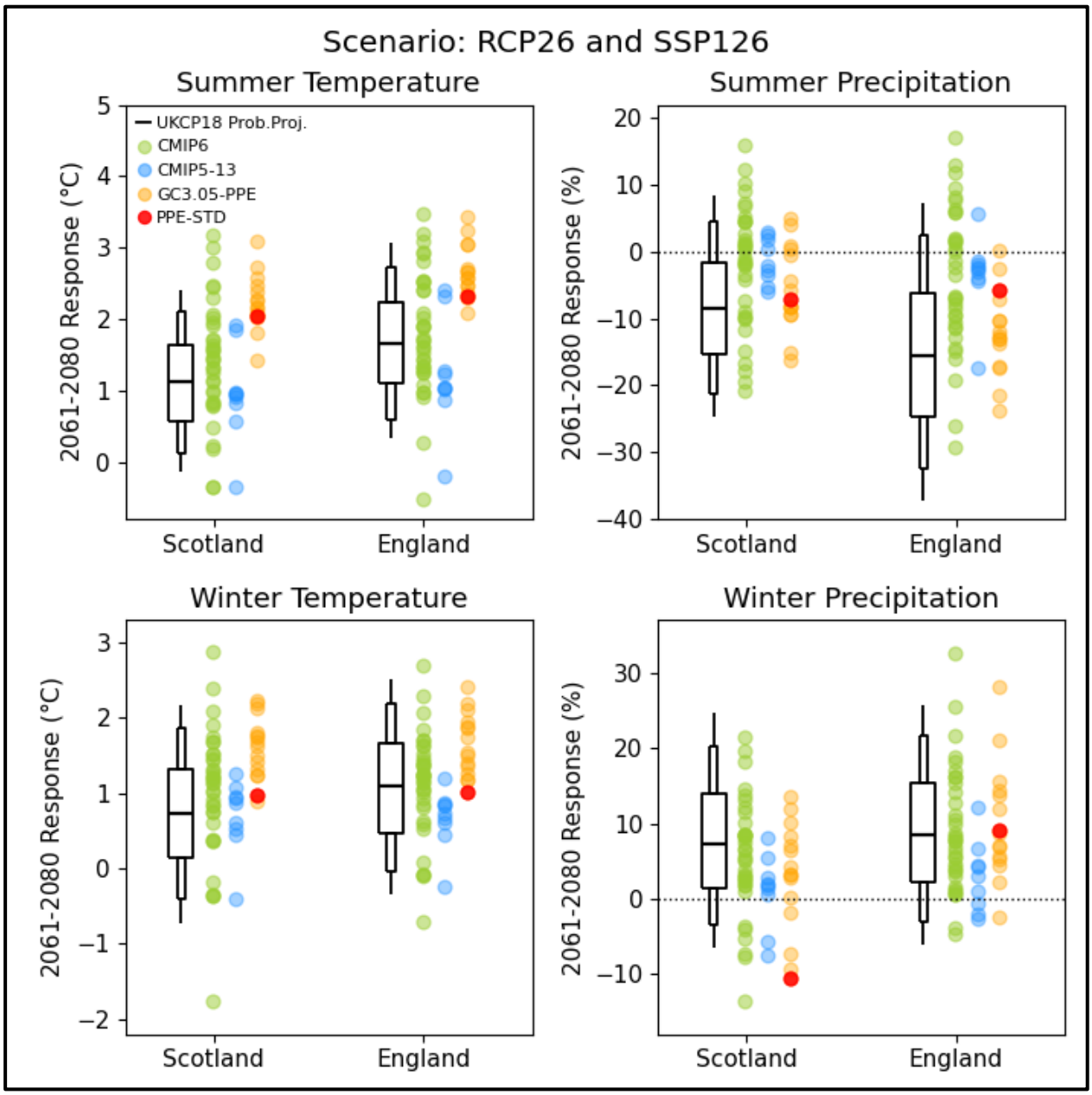

***Figure 16**. As Figure 14, but for CMIP6 and UKCP projections forced by the SSP1-2.6 and RCP2.6 emissions scenarios respectively. Projected changes are for 2061-2080 relative to 1981-2000. Details as in Fig. 14, except that only 10 ensemble members are available for CMIP5-13 (blue dots) and 35 for the CMIP6 multi-model ensemble (green dots). Results for RCP2.6 are not available from UKCP-Regional or UKCP-Local.*

Several simulations show changes outside the 5-95% range of the pdfs, or beyond the extremes of the set of UKCP-Global realisations. For summer rainfall, the CMIP6 results are distributed around zero response, suggesting (in contrast to the UKCP products) only a weak preference for drying over England, and no clear preference for Scotland. In both cases, several CMIP6 models simulate increases of 10% or more, wetter than any UKCP-Global members and higher than the 95th percentile of UKCP-Probabilistic. For England, the 95th percentile is associated with a summer rainfall increase of 7.3%. Six CMIP6 models

lie above this level, suggesting a larger chance than UKCP-Probabilistic of a higher outcome. However, the 99th percentile of the pdfs (not shown in Figs. 14-16) reaches +15.2%, higher than all but one of the CMIP6 results.

The CIESM model (Lin et al., 2020) projects a winter cooling of ~2°C for Scotland and ~0.8°C for England, both outliers compared to other simulations (Fig. 16). This model also simulates a small cooling for Scotland in summer and a small warming for England, the coolest and second coolest response respectively amongst CMIP6 members. This cool UK response occurs despite a high ECS of ~5.7°C in CIESM, although its TCR[3] of ~2.2°C lies well below those of several other models possessing ECS>5.0°C, including CanESM5, E3SM-1-0, HadGEM3-GC31-LL and UKESM1-0-LL (Liu et al., 2023).

In CMIP6, the seasonal surface temperature increases invariably show examples warmer than the 95th percentile of UKCP-Probabilistic. This occurs most frequently for Scotland temperature in summer, five CMIP6 models exceeding the 95th percentile warming of 2.4°C. This suggests a chance of ~15% of exceeding this level, cf 5% according to the pdfs. Whilst the warmest CMIP6 responses lie well above 2.4°C only two exceed 2.9°C, the warming associated with the 99th percentile of the pdfs (not shown).

For winter precipitation, most CMIP6 members simulate an increase in both countries, though a few produce a decrease in common with UKCP-Global. The largest reductions in CMIP6 occur in Scotland. These arise from UKESM1-0-LL and HadGEM3-GC31-LL, whilst HadGEM3-GC31-MM simulates little change in this period. During 2061-2080, UKESM1-0-LL also simulates reduced Scottish precipitation under SSP5-8.5 emissions, while HadGEM3-GC31-LL projects little change. All other CMIP6 models project wetter responses (Fig. 14). The UK models are closely related to STD (red dots in Figs. 14 and 16), which also simulates little change for Scottish precipitation under RCP8.5 and a reduction under RCP2.6. The LL configurations use coarser horizontal resolution in their atmosphere and ocean components, but in most respects all four models possess similar representations of physical climate system processes. However, a configuration of HadGEM3-GC3.1 using higher resolution (~50km in the atmosphere and ~1/12th degree in the ocean) projects a large increase in winter UK precipitation. This is traced to a northward shift in the Gulf Stream (absent from lower resolution configurations), that drives a strengthening of cyclone activity across the North Atlantic (Moreno-Chamarro et al., 2021).

Further investigation is needed to understand differences between the extremes of the precipitation and temperature ranges projected by the different products in Fig. 16, including the potential impact of screening CMIP6 models. This is underlined by the relevance of these strong mitigation scenarios to the national target to achieve net zero carbon emissions by 2050, set out in the 2008 Climate Change Act (see https://www.legislation.gov.uk/ukpga/2008/27/contents).

c. Future changes in circulation

We provide a brief discussion of projected circulation changes, focusing on winter and summer North Atlantic Oscillation (NAO and SNAO) as the leading modes of seasonal variability in slp (Hurrell et

al., 2003; Folland et al., 2009). We use the Atlantic pressure gradient definition of winter NAO (McSweeney and Yamazaki (2020), described in section 3c). Following Folland et al. (2009), the SNAO is defined as the leading empirical orthogonal function (EOF) of slp in July-August ("high summer"), over the region 90°W-30°E, 40-70°N.

Figure 17 shows projected winter NAO changes for 2061-2080 relative to 1981-2000, comparing UKCP-Global results (RCP8.5) against those from 36 CMIP6 models (SSP5-8.5). The CMIP6 results show no clear preferred sign for NAO response, projecting a range of changes between ~±6hPa. However, only five models simulate a reduction larger than -2hPa, whereas ten models project an increase exceeding +2hPa. A similar range emerges from UKCP-Global. The ensemble-mean change in CMIP5-13 is small (McSweeney and Yamazaki, 2020), with a subset of members contributing negative NAO responses to UKCP-Global. In contrast, GC3.05-PPE members favour a positive response, contributing most of the largest changes. Overall, combination of the PPE and MME datasets achieves a fuller representation of plausible outcomes in UKCP-Global, that is supported by the CMIP6 evidence.

Projected decline in Arctic sea-ice drives an equatorward shift in the winter mid-latitude westerlies in the northern hemisphere. This contributes to a "tug-of-war" with the effects of global ocean warming since the latter drives a poleward shift by increasing the meridional temperature gradient in the upper troposphere. However, current climate models underestimate the influence on the mean zonal flow of feedbacks from transient eddy activity. If this systematic bias was corrected, the poleward shift driven by sea-ice decline might "win" the tug-of-war (Smith et al., 2022; Screen et al., 2022). This would lead to weakening of the zonal flow in the latitude band 55-65°N, hence favouring a shift to the negative phase of winter NAO (Deser et al., 2016).

Whilst Fig. 17 provides an overview of projected changes at the seasonal time scale, analysis of sub-seasonal variability is required to understand the implications for future winter weather. For UKCP-Global, Murphy et al. (2018) showed a projected increase in the occurrence of winter storms over the UK, with a more muted increase in CMIP5-13. Pope et al. (2022) analysed future changes in daily weather patterns in UKCP-Global, using the 30-pattern classification of Neal et al. (2016). Pope et al. found increased occurrences of patterns associated with cyclonic conditions and strong westerly winds in the PPE, at the expense of anticyclonic weather patterns associated with blocked conditions. Similar results were found in CMIP5-13, but typically with smaller ensemble-mean signals of change and greater spread in the ensemble ranges for specific weather types.

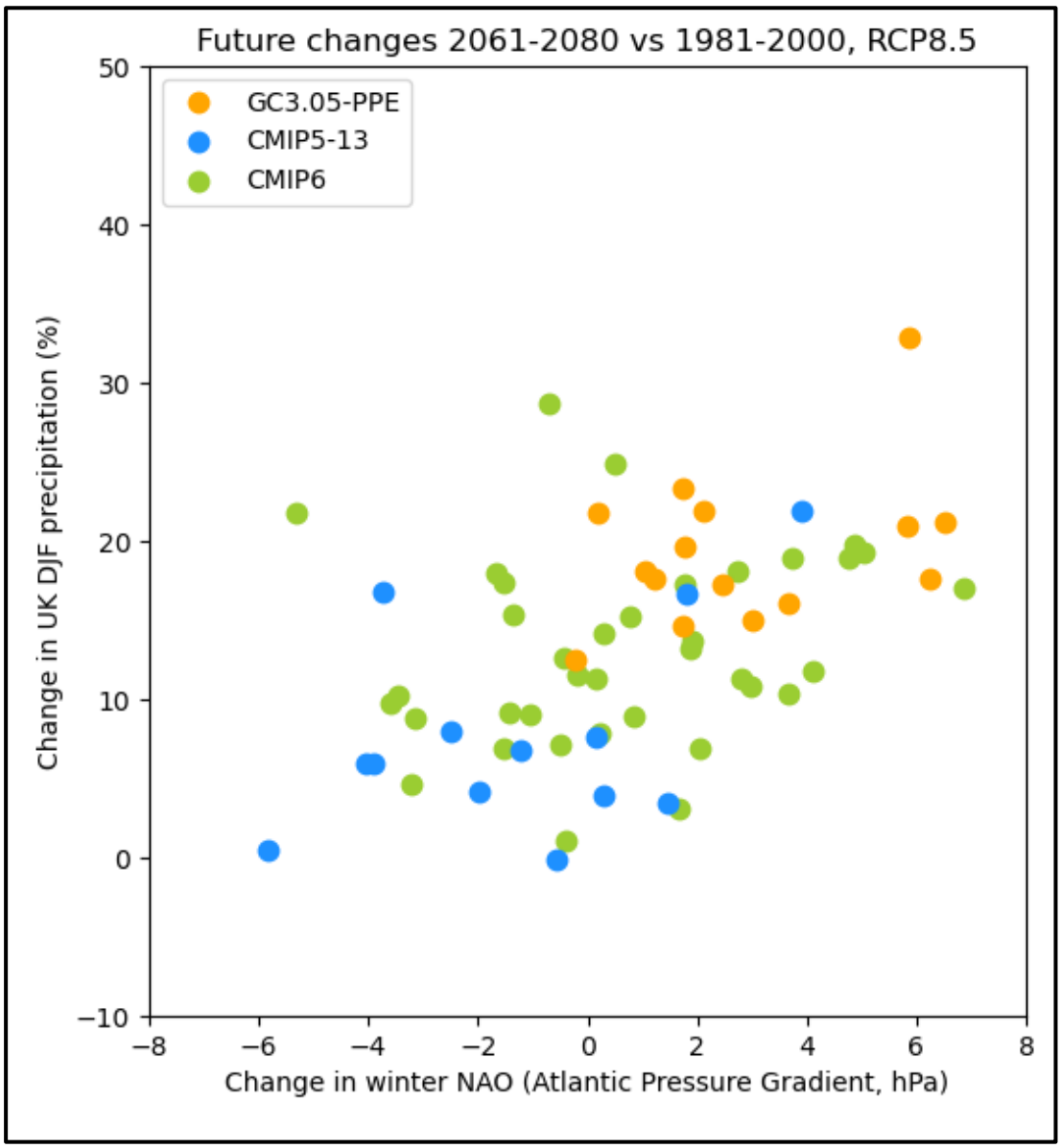

***Figure 17**. Projected changes in the winter NAO index for 2061-2080 relative to 1981-2000, with corresponding UK precipitation changes (%). Results show the GC3.05-PPE (orange) and CMIP5-13 (blue) members that comprise UKCP-Global, and simulations from 36 CMIP6 models (green). The NAO index is expressed as the average Atlantic pressure gradient (hPa) between Gibraltar and Iceland (McSweeney and Yamazaki, 2020). The CMIP6 and UKCP projections were forced by the SSP5-8.5 and RCP8.5 scenarios respectively.*

In a set of CMIP6 models, Fabiano et al (2021) use an EOF-based method to identify four daily weather clusters in the Euro-Atlantic sector: two of these resemble the positive and negative phases of the NAO, the others Scandinavian blocking and Atlantic ridge regimes. They find a general improvement in skill in CMIP6, compared with a CMIP5 ensemble. In the future projections, they find increases in the frequency and persistence of the positive NAO pattern, alongside decreases in the frequencies of the Scandinavian blocking and Atlantic ridge patterns. In SSP5-8.5, Fabiano et al. find little change in the ensemble-mean frequency of their negative NAO pattern. However, modest increases do occur for SSP scenarios with lower forcing, implying that both the positive and negative phases of NAO become more frequent at the expense of the other two regimes. In SSP5-8.5, Dorrington et al. (2022) find a future

trend towards less persistent Euro-Atlantic blocking regimes in a set of 20 CMIP6 models, coupled with an increase in zonal flow conditions associated with the low-level jet in the central North Atlantic.

A caveat is that the papers discussed above use different statistical methodologies to identify daily weather regimes. A common analysis (beyond the scope of this report) would be required to understand more precisely the similarities and differences between the changes projected in UKCP-Global and CMIP6, and their process-drivers.

Figure 18 shows projected SNAO anomalies and their relationship to July-August (JA) rainfall, for 2061-2080 relative to 1981-2000. Here, we consider rainfall averaged over the whole UK (cf the national averages of Fig. 14). Following Bladé et al. (2012), slp anomalies from observations and each model simulation are projected onto the observed EOF to establish time series of an SNAO index. This index represents the principal component (PC) time series for observations, and a "pseudo" PC for each model simulation[4]. The top row of Fig. 18 confirms that this approach identifies (essentially by construction) an index of temporal variability in the models representing similar patterns of regional slp to that associated with the observed SNAO. This pattern features an slp dipole with (in its positive phase) high pressure over the UK and low pressure over Greenland. The positive phase is also associated with dry anomalies in JA rainfall over NW Europe, the largest anomalies being centred over the UK and southern Norway. Ensemble-mean patterns from GC3.05-PPE, CMIP5-13 and CMIP6 all reproduce this pattern reasonably well (Fig. 18, second row), although the intensities of the UK and Norwegian anomalies are underestimated. In the observations, a wet response over southern Europe is also associated with positive SNAO. Again, the ensemble-mean patterns capture this, albeit with reduced amplitude. In the CMIP3 multi-model ensemble, Bladé et al. (2012) found a similar error to be related to poor simulation of the upper-level circulation response to a positive SNAO.

In CMIP6, projected changes in SNAO show no preferred sign (Fig 18, bottom panel). Simulations showing increases in JA rainfall possess neutral or negative SNAO index changes, while the driest simulations (reductions in rainfall of 25% or more) show neutral or positive index changes. This relationship is expected, given the historical SNAO-rainfall relationship. The weaker amplitude of the UK rainfall response in the CMIP6 ensemble -mean suggests that the ensemble may underestimate the impact of a potential future shift towards the positive or negative phases of the SNAO in the real world. However, the scatter in the relationship also reveals that additional drivers must be at play in the CMIP6 projections of summer rainfall change. Potential candidates include earlier snowmelt, positive feedbacks between soil moisture and cloud cover and increased moisture content in maritime air carried over the

[4] This procedure for deriving simulated SNAO index values differs from that used to create a similar diagram in the UKCP18 Science Report (Murphy et al. (2018), their Figure 5.5). There, model-specific EOF patterns most resembling the observed EOF were derived separately for each simulation. In the present report, projecting the model data onto the *observed* EOF improves the correspondence between the simulated and observed regional correlations of slp and precipitation (top two rows of Fig. 18). A caveat is that the SNAO index derived from the pseudo-PC will explain slightly less variance of simulated psl, cf the Murphy et al. (2018) approach.

UK by the prevailing westerlies (Rowell and Jones, 2006), and also decline in AMOC strength (Jackson et al., 2015).

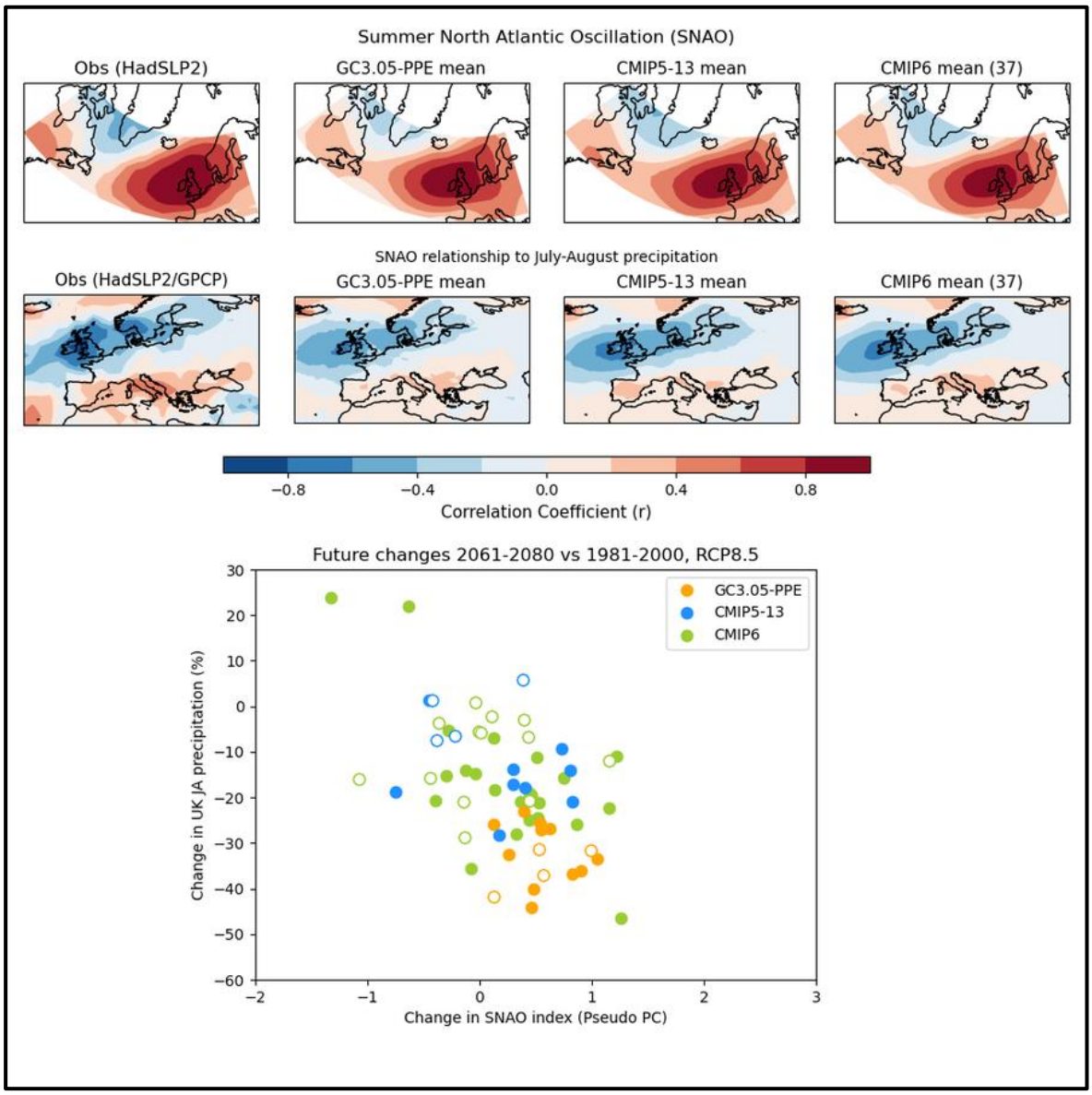

***Figure 18**. Summer North Atlantic Oscillation (SNAO) and its teleconnection to mean July-August (JA) precipitation in observations and the UKCP-Global and CMIP6 ensembles. SNAO is defined as the first EOF of sea-level pressure (slp) anomalies over the region 90°W-30°E, 40-70°N in the HadSLP2 dataset of observations. Observed and simulated slp anomalies are projected onto the observed EOF to establish time series of SNAO indices. This index represents the principal component (PC) time series for observations, and a "pseudo" PC that captures the effects of the same circulation pattern in each model simulation (see text). The top row shows temporal correlations of local slp with these SNAO index timeseries for observations and the GC3.05-PPE, CMIP5-13 and CMIP6 ensembles, during 1950-2010. The model maps are ensemble-mean correlations averaged across 15, 13 and 36 members respectively. The second row shows correlations of JA precipitation with the SNAO index during 1979-2010 for observations (GPCP, Adler et al., 2003) and during 1950-2010 for the three model datasets, displaying ensemble-mean correlations in the latter cases. Lower panel shows projected changes in the SNAO index for 2061-2080 relative to 1981-2000 with corresponding JA precipitation changes (%) for the UK, for members of the GC3.05-PPE (orange), CMIP5-13 (blue) and CMIP6 (green*) ensembles. *The filled circles denote models that produce a realistic SNAO pattern as a leading mode of simulated variability, open circles denoting those that do not (details in text). The CMIP6 and UKCP projections were forced by the SSP5-8.5 and RCP8.5 scenarios respectively.*

The UKCP-Global projections cover slightly narrower ranges of change for both variables, showing a greater preference for positive SNAO changes with a larger fraction of simulations projecting drying of 30% or more. This tendency is clearest in GC3.05-PPE, suggesting that a shift towards positive SNAO is a major driver of the strong projected drying in these simulations. Analysis of daily weather patterns (Pope et al., 2022) supports this conclusion, revealing a shift in the PPE towards anticyclonic circulation types in summer (in contrast to the winter results discussed earlier).

The fixed-pattern approach of Fig. 18 allows us to diagnose the impacts of the observed SNAO circulation pattern in the model simulations. However, it does not tell us how closely the leading modes of slp variability generated by the models[4] correspond to the observed SNAO. This was evaluated using the first and second EOFs of simulated slp variability. Models assessed to generate a realistic SNAO (filled circles in Fig. 18) are those for which either EOF includes a distinct region of high pressure influencing the UK (determined subjectively) and explains at least 20% of slp variance over the analysis region. The open circles denote model simulations failing these criteria. On this basis, four GC3.05-PPE members, three CMIP5-13 members and eight CMIP6 members do not simulate a realistic SNAO. Such an assessment could potentially form part of a future strategy to select a subset of simulations likely to provide credible projections of summer changes for NW Europe.

d. Screening or constraining ensemble projections

While the comparisons in this report take the simple approach of including all available CMIP6 models (section 1), the application of observational constraints to raw climate model output is recognised as a key step in providing reliable climate projection information for impacts applications (e.g. Hegerl et al., 2021). Such strategies were used in UKCP-Probabilistic and UKCP-Global, based on weighting (for UKCP-Probabilistic) or screening (for UKCP-Global) sets of historical variables derived from the relevant simulations (section 2a, b). Below we provide illustrative examples showing how such postprocessing could alter CMIP6 ranges of projected change.

These are based on a screening approach developed by Palmer et al. (2023) (hereafter P23) for European projections, in which an initial set of 29 CMIP6 models was reduced to a performance-filtered subset of 16 models. The performance filtering used observables selected to capture several driving phenomena and regional characteristics important for present and future European climate. These include: AMOC, North Atlantic SSTs, 850hPa winds, the North Atlantic/European storm track, blocking frequency, terrestrial surface air temperature and precipitation. Performance was assessed using a mix of quantitative error metrics and subjective judgement. For each observable poor performers were flagged as either “inadequate” or “unsatisfactory”, dependent on the level of bias encountered. Models were excluded if they incurred either one or more “inadequate” flags.

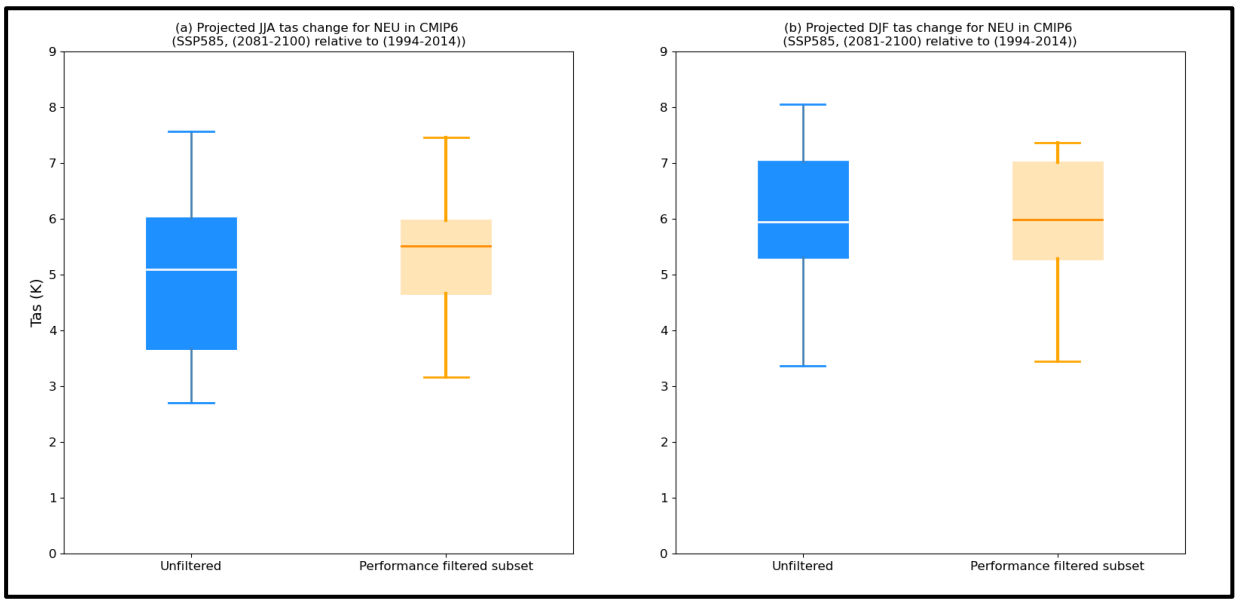

***Figure 19**. Projected changes in surface air temperature for Northern Europe in summer (left) and winter (right) for 2081-2100 relative to 1994-2014, from CMIP6 models driven by the SSP5-8.5 scenario. Northern Europe is defined as a region including the UK and Scandinavian countries – see Figure S1 of Palmer et al. (2023), Supplementary Information at https://esd.copernicus.org/articles/14/457/2023/esd-14-457-2023-supplement.pdf. Blue boxes and whiskers show the 5th, 25th, 50th (white line) 75th and 95th percentiles from 29 CMIP6 models, labelled "unfiltered" because each available model was included without applying performance criteria. Orange boxes and whiskers show the same percentiles from a filtered subset of 16 models, selected by assessing how well individual models capture a range of large-scale processes important for the representation of present-day European climate (Palmer et al., 2023).*

We show winter and summer examples for northern Europe, consisting of surface air temperature (Fig. 19) and precipitation (Fig. 20) changes for 2081-2100 relative to 1981-2000 in response to SSP5-8.5 forcing. For winter temperature, the performance filtering has negligible impact on the median and low-end (5th percentile) warming but reduces the upper-end (95th percentile) response slightly. The impact of screening is somewhat larger in summer, resulting in an increase of ~0.5°C in median warming and a smaller interquartile range caused by an upward shift in the 25th percentile response. The upper limit of summer warming remains just below 8°C, while the lower limit increases slightly. P23 also find a higher median warming in summer for their pan-European region. This occurs because the performance filtering excludes all the models possessing low ECS (below 2.0°C) and most with ECS values below 3.0°C, whilst more of the high-ECS models are retained.

For winter precipitation, performance filtering shifts the median response wetter and narrows the associated uncertainty range. In summer, the 5th-95th percentile range changes little, however the filtered ensemble shows stronger median drying than the unfiltered ensemble. In central Europe, using the P23 filtered subset narrows the range of summer rainfall changes because three ensemble members simulating small changes are removed, leaving a consistent drying signal in the retained simulations.

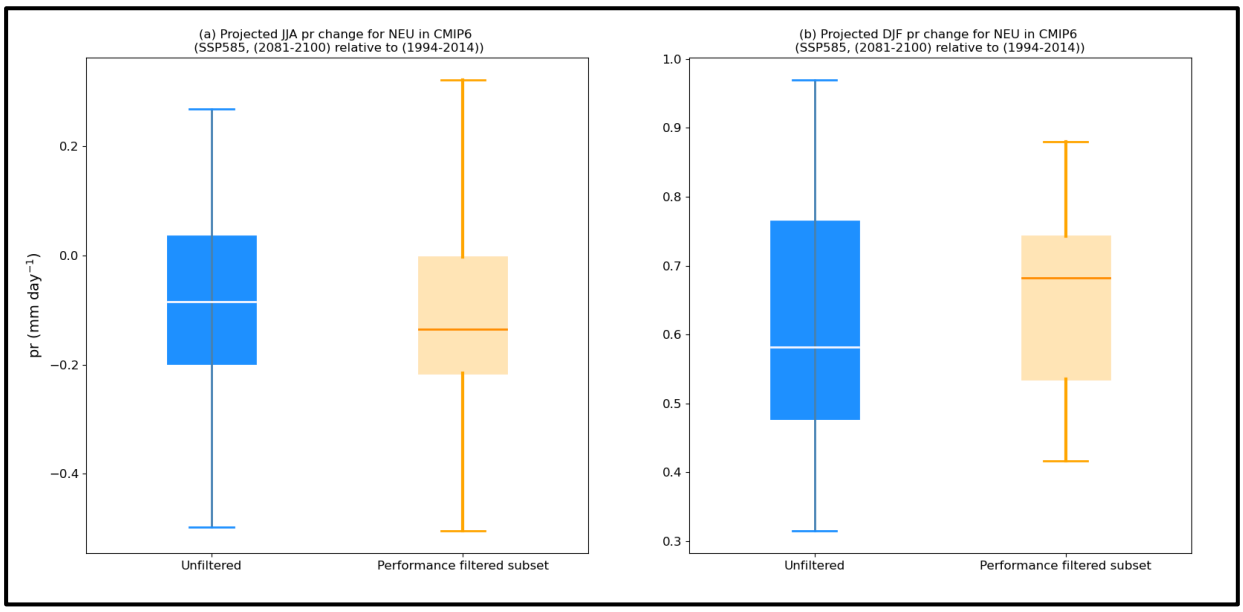

***Figure 20**. Projected changes in precipitation (mm/day) for Northern Europe in summer (left) and winter (right) for 2081-2100 relative to 1994-2014, from CMIP6 models driven by the SSP5-8.5 scenario. Blue boxes and whiskers show ranges of change from 29 unfiltered CMIP6 models. Orange boxes and whiskers show corresponding ranges from 16 models surviving filtering to remove those providing the least realistic simulations of present-day European climate (Palmer et al., 2023). Further details in Figure 19.*

The model assessment criteria used in Figs. 19 and 20 were derived exclusively from observables expressing aspects of recent climatology. P23 considered an alternative method (adapted from Brunner et al., 2020a), in which model weights are derived using the historical trend in GMST during 1981-2014. In contrast to the performance-filtering method, the global trend approach shifts the distribution of projected European warming to lower values (P23, their Fig. 6a). A potential caveat is that the period since 1980 favours lower ECS estimates compared with earlier parts of the historical record (Andrews et al., 2022), due to the pattern effect discussed in section 3d.

The UKCP-Probabilistic method uses a weighting approach that combines information based on historical trends and recent climatology (see section 4a). Brunner et al (2020a) use a different probabilistic framework (ClimWIP) to demonstrate that combining both types of constraint improves GMST projections in leave-one-out perfect model tests. In the UKCP-Probabilistic method, the historical trend information is derived from large-scale patterns of surface temperature change, global upper ocean heat content and atmospheric CO_2 concentration (Harris et al., 2013; Booth et al., 2017; Murphy et al., 2018), while the climatological information is derived from global spatial patterns of a variety of standard model assessment variables (similar in scope to the information in Figs. 5-7), reduced in dimensionality using eigenvector analysis (Sexton et al., 2012). The UKCP-Probabilistic approach is based on the premise that the weight should represent a broadly-based measure of the overall physical credibility of climate model output. This supports potential usage of its results in applications covering a

range of climate variables across different world regions (e.g. Harris et al. (2010, 2013), Brunner et al., 2020b).

In comparison with the P23 screening approach, UKCP-Probabilistic uses a larger quantity of observables but lacks the tailored regional focus of the latter. The selection of constraining observables is a key choice in the design of screened or weighted regional projections. There is a variety of potential choices in the current literature (e.g. Brunner et al., 2020b; Hegerl et al., 2021), for both evaluation metrics and the methodological frameworks used to apply them.

For example, Hausfather et al. (2022) discuss the issue that CMIP6 contains several "hot models" with ECS values of 5°C or higher. This contrasts with CMIP3 and CMIP5 (Bock et al., 2020), while AR6 identifies 2-5°C as the very likely range for ECS (Arias et al., 2021). Hausfather et al. suggest that models lying outside the likely AR6-assessed ranges for ECS (or, alternatively, TCR) should be excluded from scenario-based impacts studies in which the temporal trajectory of future global warming is important, partly on the grounds that they do a poor job of replicating historical temperatures over time. However, Bloch-Johnson et al. (2022) dispute such binary recommendations, pointing out that: (a) historical warming exerts only a moderate constraint on ECS (significant factors being the large uncertainties associated with historical aerosol forcing and time-dependence of climate feedbacks, see section 3d)); (b) the upper tail of probabilistic assessments of ECS is sensitive to reasonable variations in methodological assumptions (Sherwood et al., 2020).

In screening and constraint methods, another key challenge is to estimate (either qualitatively or quantitatively) a tolerable level of model error. The famous aphorism "all models are wrong, but some are useful" (Box and Draper, 1987), along with the presence of common systematic biases in multi-model ensembles (section 3), frames this task as one of discrimination between "useful" and "wrong" models. This challenge applies both to univariate selection methods (e.g. Hausfather et al., *op. cit*.) and multivariate approaches, requiring sources of uncertainty in the model evaluation process to be specified. Dependent on the application in question, these can include observational errors, simulated internal variability, emulation errors (where these are used to estimate climate model results) and structural model biases. In cases where an observable lies outside the available ensemble of simulated historical values but is linked mechanistically to projected changes in impact-relevant variables (e.g. Screen et al., 2022), the caveat that the true future outcome could lie outside the range of projected changes should be clearly communicated.

In the UKCP projections, methods to estimate suitable error tolerances are included in: (i) the calculation of weights in UKCP-Probabilistic (Sexton et al., 2012); (ii) the screening of potential PPE members for GC3.05-PPE in UKCP-Global (Sexton et al., 2021); (iii) screening of CMIP5 simulations for UKCP-Global (McSweeney et al., 2018). These specifications involve expert judgements capable of

exerting a significant influence on the weighted projections or selection of ensemble members (e.g. Sexton and Murphy, 2012; Rostron et al., 2020).

The importance of such choices emerges particularly clearly in cases where different evaluation criteria lead to exclusion or down-weighting of different subsets of ensemble members. For example, P23 highlight a tension between their historical trend and climatology constraints, which favour different ranges of ECS. This underscores the importance of choosing multivariate observables and model error tolerances that achieve a suitable balance between quality and sampling criteria.

In section 4b we noted the growth of interest in GWLs. These provide a way of presenting climate projections in the light of the policy-relevant warming targets in the Paris agreement (see https://unfccc.int/process-and-meetings/the-paris-agreement). Hausfather et al. (2022) and P23 suggest that presenting projections as GWLs might circumvent the need to select or weight CMIP6 models according to historical warming trends (see discussion above), on the basis that high ECS may be of lesser concern if emphasis is placed on regional patterns of response for a given GWL.

From an applications perspective, the role of ECS in GWL-based projections depends on whether information on timing is also required. In CCRA3 authors were asked to describe climate risks and opportunities "for time slices of the 2050s (2040-2070) and 2080s (2070-2100) for scenarios that project global warming to stabilise at 2°C ± 0.5°C by 2100, or project global warming to reach 4°C ± 0.5°C in 2081- 2100" (Betts and Brown, 2021). Thus, emphasis was placed on documenting the time that specific GWLs were reached and the temporal pathways towards them, both aspects likely to be influenced by ECS.

From a scientific perspective, an interesting question is whether GWL-based projections (in cases where the temporal trajectory is not important) could be based on different sets of observational constraints to projections presented as time-dependent scenarios. While ECS may not be of direct importance when the time dimension is removed, key processes that drive occurrence of high ECS in climate models (especially cloud and surface albedo feedbacks) are spatially heterogeneous in nature (Zelinka et al., 2020). Therefore, it is not clear that constraints on climate feedbacks would necessarily be of lesser importance in applications focused on regional patterns of climate change for prescribed warming levels (e.g. Fig. 16).

5. Concluding Remarks

The CMIP6 multi-model ensemble provides a rich new dataset of global climate projections, likely to be widely used in studies of regional impacts during the next few years (e.g. Sobolowski et al., 2023). Compared with earlier CMIP phases, the CMIP6 ensemble shows improvements in simulation of recent climatology across standard model assessment variables covering temperature, cloud, radiative, hydrological and circulation variables (Bock et al., 2020; Arias et al., 2021). However, no single model

stands out as clearly superior to other CMIP6 members and the range of performance scores overlaps considerably with that amongst CMIP5 models.

a. Evaluation of UKCP18 and CMIP6 historical simulations

We revisited the set of global coupled ocean-atmosphere simulations included in the UKCP-Global component of UKCP18, using the CMIP6 ensemble as a new comparative benchmark. Despite the general improvements in CMIP6 models compared to CMIP5, the historical performance of UKCP-Global simulations is competitive with CMIP6 models. This applies to seasonal climatologies of surface air temperature and precipitation, in which systematic regional biases in CMIP6 are similar to those in the perturbed parameter and multi-model ensembles contributing members to UKCP-Global (see section 2b). The same applies to annual climatologies for a broader set of climate variables, particularly in the case of the perturbed parameter ensemble (GC3.05-PPE). The PPE is derived from a model closely related to HadGEM3-GC3.1, the configuration contributed by the Met Office to the CMIP6 ensemble.

Three major modes of coupled-ocean-atmosphere variability (ENSO, AMV and winter NAO), all of which exert important influences on European climate, were also evaluated. All UKCP-Global and most CMIP6 simulations exceed the observed amplitude of interannual variations in NAO. For AMV, all CMIP6 members simulate greater interannual variability than observed, while UKCP-Global members exhibit a range of biases of either sign. Observed and simulated estimates are subject to significant sampling uncertainties across different periods, so in many cases the simulation biases are not statistically significant. However, some models do show ranges of variability that fail to overlap with observations, especially for some CMIP6 simulations of AMV. A caveat is that AMV exhibits significant variability on the centennial time scale (Mavilia et al., 2018), limiting the extent to which firm conclusions can be drawn from the 1900-2005 time series considered here. For ENSO amplitude (measured using monthly anomalies in the NINO3.4 index), uncertainty ranges from all but one UKCP-Global members overlap substantially with observations. This is also true for many CMIP6 members, though a minority simulate ENSO variability that is clearly too low or too large.

By the most recent decade (2010-2020), the observed change in global mean surface temperature (GMST) relative to 1900-1930 amounts to ~1.0°C. This lies near the middle of the distributions simulated by the CMIP6 and UKCP-Global ensembles, whose ranges of change are similar. During 1960-1990, most GC3.05-PPE members and some CMIP6 members simulate GMST anomalies cooler than observations, due at least partly to strong negative forcing from anthropogenic aerosol emissions. However, most GC3.05-PPE simulations warm more rapidly than observations beyond 1990, removing the earlier cool bias by the 2010s.

In this report CMIP6 data was assembled for evaluation on a simple “one-model-one-vote” basis. In contrast, members of UKCP-Global were selected from wider sets of CMIP5 models and GC3.05 perturbed parameter variants, through application of screening criteria to identify a subset of high-quality simulations offering a diversity of projected future changes. Our CMIP6 evaluation suggests that

a screened subset of CMIP6 models could form a potential component of a future update to UKCP-Global. This depends, however, on future decisions concerning the nature and timing of developments to the UKCP land projections, including its probabilistic, regional and local components (see sections 1 and 2). Further developments in global modelling can also be expected. For example, some modelling groups are testing increases in horizontal resolution in the atmosphere and ocean, finding improvements in various aspects of simulated climatology compared to lower-resolution counterparts included in the CMIP6 scenario experiments (Roberts et al., 2019; Bock et al., 2020).

b. Projected future changes

For GMST, CMIP6 models and UKCP-Global both provide a wide range of changes under fossil-fuel intensive emissions scenarios (SSP5-8.5 and RCP8.5 respectively). The CMIP6 spread is slightly broader, with a combined range (across both ensembles) covering ~3.0-8.0°C by 2100, relative to 1981-2000. In the present analysis, neither ensemble is formally constrained by observations of historical surface temperature changes. However, the CMIP6 range is reduced when such constraints are applied (Lee et al., 2021). This also occurs in UKCP-Probabilistic, the probabilistic projections component of UKCP18.

We also compared projected changes in surface temperature and precipitation for England and Scotland, from CMIP6, UKCP-Global and UKCP-Probabilistic. Qualitatively, the CMIP6 results support projected trends in extreme seasonal events from UKCP18. For example, occurrence of winter precipitation anomalies exceeding 50% increases during the 21st century under SSP5-8.5 emissions, as the distribution of events in individual winters shifts to wetter outcomes. Seasonal anomalies of 2°C (that exceed the current UK record) become normal events by ~2050, consistent with UKCP18 advice (McCarthy et al., 2019).

Considering multidecadal average changes, the CMIP6 ensemble develops a very broad range of warming signals in response to SSP5-8.5. We illustrated these using changes for 2061-2080 relative to 1981-2000. Three configurations of the E3SM model contribute large outlier responses that could potentially be ruled out in a formal screening exercise, though this is left to future work. Nevertheless, the wide CMIP6 ranges show substantial overlap with those of UKCP-Global and UKCP-Probabilistic for both precipitation and surface temperature, confirming the continued importance of accounting for the relevant uncertainties in impacts studies.

Some differences occur between CMIP6 and the UKCP18 products. For example, few simulations from CMIP6 or UKCP-Global lie within the lowest quartile of UKCP-Probabilistic outcomes for surface temperature changes in Scotland. For summer rainfall all three datasets suggest that reductions are more likely than increases. However, CMIP6 and UKCP-Probabilistic suggest a larger chance than UKCP-Global of an increase over Scotland, while in CMIP6 the maximum drying over England is slightly less intense than found in the UKCP18 results.

Surface temperature and precipitation changes were also compared for scenarios including strong mitigation measures (RCP2.6 for UKCP18 cf SSP1-2.6 for CMIP6). As expected, ranges of UK warming are

shifted lower compared to the RCP8.5/SSP5-8.5 results. For UK-Probabilistic, the median warming ranges from ~0.8°C (for Scotland in winter) to ~1.6°C (England in summer) for RCP2.6, cf ~1.9°C and ~3.8°C for RCP8.5. Under SSP1-2.6 the maximum warming amongst CMIP6 models is ~3.2°C (for England in summer), though results for the E3SM model (see above) are not available for this scenario. More generally, CMIP6 models project ranges of changes for both variables, seasons and countries that encompass the UKCP18 results. This underlines the importance of considering alternative adaptation pathways under mitigation scenarios, given that regional changes (while smaller than under more fossil-fuel intensive scenarios) are nevertheless diverse in both the UKCP and CMIP6 datasets.

Under both SSP scenarios, most CMIP6 models project increases in winter Scottish precipitation during 2061-2080. Two configurations of the Met Office model (UKESM1-0-LL and HadGEM3-GC31-LL) lie at the low end of the range, with UKESM1-0-LL simulating reduced precipitation in both scenarios. Configurations with higher horizontal resolution (HadGEM3-GC31-MM and the corresponding member of GC3.05-PPE using standard parameter settings) simulate modest increases or little change. However, a configuration of HadGEM3-GC3.1 using even higher resolution (~50km in the atmosphere and ~1/12th degree in the ocean) projects a substantial increase in Scottish (and UK-wide) winter precipitation, linked to a northward shift in the Gulf Stream absent from the lower resolution configurations (Moreno-Chamarro et al., 2021). This demonstrates potential for further evolution in the understanding of projected climate change signals as the international community moves to the next phase of CMIP, subject to further experiments to assess the effects of enhanced horizontal resolution in different models.

Projected changes in winter NAO show no preferred sign in the CMIP6 ensemble, revealing a range of positive and negative changes consistent with UKCP-Global. More detailed analysis of daily weather regimes reveals a future shift towards reduced occurrence of anticyclonic patterns accompanied by increased frequency of cyclonic or zonal flow conditions, in both sets of projections (Fabiano et al., 2021; Dorrington et al., 2022; Pope et al., 2022).

The CMIP6 ensemble also shows no preferred sign in its projected changes in July-August SNAO. Simulations showing negative (positive) changes in the SNAO index are more likely to show increases (reductions) in July-August rainfall. The models tend to underestimate the strength of the observed historical relationship between SNAO and UK rainfall. Therefore, the hydrological influence of future SNAO changes is also likely to be underestimated, though additional factors will affect the projected rainfall changes. The UKCP-Global projections cover slightly narrower ranges of change for both variables, showing a greater preference for positive SNAO changes with a larger fraction of simulations projecting drying of 30% or more. This arises from the influence of GC3.05-PPE members, which favour a shift to the positive phase accompanied by strong future drying.

c. Context for detailed studies of regional impacts

The global modelling results provide context for the downscaled UK projections available from the UKCP-Regional and UKCP-Local components of UKCP18. These are based on limited-area regional simulations run at 12km and 2.2km respectively (see section 1) that are used extensively to investigate future impacts, hazards and extreme events at high spatial resolution (see section 4b). UKCP-Local brings new capability for these studies, supporting understanding of potential future changes in hourly precipitation, hail, lightning, extreme windstorms and urban heat impacts. In winter, the 2.2km simulations project larger increases in seasonal-mean precipitation than the GC3.05-PPE driving simulations, because their explicit representation of convective-scale dynamics captures advection of maritime showers over land, a process missing from current global models (Kendon et al., 2020).

The CMIP6 results join the other components of UKCP18 in providing a broader uncertainty context for UKCP-Regional and UKCP-Local. The downscaled products are currently derived only from GC3.05-PPE members, and consequently sample a limited range of (high-end) surface temperature changes in the UK, particularly in summer when the multi-year mean rainfall changes in England are also limited to outcomes featuring strong drying. The CMIP6, UKCP-Global and UKCP-Probabilistic results emphasise the importance of adding representation of structural modelling uncertainty, by including projections derived from alternative international climate models. Accordingly, an update to UKCP-Regional and UKCP-Local is being prepared, that will add 12km and 2.2km regional projections driven by four CMIP5 models.

Presentation of regional changes conditioned on specific global warming levels (GWLs) is a format growing in popularity and used extensively in the latest UK climate change risk assessment (Betts and Brown, 2021). When regional changes are expressed per unit global warming, CMIP6 results generally show reasonable consistency with ranges of change from UKCP-Global and UKCP-Probabilistic. However, no CMIP6 outcomes lie in the upper quartile of the UKCP-Probabilistic distributions for winter precipitation in Scotland, or summer temperature in England.

The GC3.05-PPE (and hence UKCP-Regional and UKCP-Local) members show relatively narrow ranges of normalised warming. This implies that GWL-based assessments of indicators bearing a strong relationship to average warming are likely to be overconfident, if based solely on UKCP-Regional or UKCP-Local in their current form. Central estimates of response are likely to be more reliable, since median national warming per unit increase in GMST is similar for the downscaled products and CMIP6, UKCP-Global or UKCP-Probabilistic.

Many applications of UKCP-Regional and UKCP-Local relate to assessments of future extreme events, which are subject to a range of potential influences (e.g. Kendon et al., 2010). Some of these drivers (such as increases in atmospheric specific humidity or changes in the seasonal cycle of soil moisture) may possess relatively strong thermodynamic links to regional temperature changes. Others (such as changes in regional storm dynamics, e.g. Chan et al., 2018) could exert influences on local

extremes that are relatively independent of regional average warming. Therefore, more research is needed to understand the influences and sampling characteristics of a broader set of large-scale drivers, to evaluate the utility of the existing regional and local UKCP data and understand the potential benefits of future updates.

d. Creating reliable projections

Our analysis of UK projections from CMIP6 models provides a "first look" inclusive of all available models. The CMIP6 multi-model ensemble provides a rich source of data for potential use in studying impacts and risks in the UK and other regions. This would require sub-selection or weighting of CMIP6 simulations based on qualitative and/or quantitative performance indicators. For such exercises, two distinct types of information include metrics expressing: (i) recent climatology (either as long-term averages of basic variables, measures of climate variability, or specialised indicators capturing specific phenomena); (ii) changes in climate observed since the industrial revolution, in variables such as surface temperature or ocean heat content.

In UKCP-Global, information of type (i) was used to screen potential CMIP5 models, while information of both types was used to select GC3.05-PPE variants. In UKCP-Probabilistic, types (i) and (ii) were combined to assign weights to alternative realisations of future climate anomalies, each type contributing significantly to the constraining impact of the observations (Hegerl et al., 2021). Regarding CMIP6 models, IPCC provided constrained projections of GMST by combining three methods based on historical surface temperature changes (Lee et al., 2021). Brunner et al. (2020a) found reductions in projected GMST changes when applying a set of constraints combining types (i) and (ii), because some CMIP6 models with high future warming received lower performance weights.

For regional projections, Hausfather et al. (2022) suggest that models with possessing ECS or TCR values outside likely ranges assessed by IPCC (for example, models with ECS>5°C) should be excluded from scenario-based impacts studies in which the temporal trajectory of future global warming is important, irrespective of other performance metrics. However, such an approach is questionable (Bloch-Johnson et al., 2022), since historical warming exerts only a moderate constraint on ECS due to the influences of uncertain factors including historical aerosol forcing and the influence of time-varying patterns of warming (Andrews et al., 2019; Sherwood et al., 2020; Smith et al., 2021).

Palmer et al. (2023) use a set of type (i) criteria to recommend a screened subset of CMIP6 models for impacts studies in Europe (section 4d), finding that this approach favours models with high ECS. When applying type (ii) criteria, Palmer et al. find that lower ECS models are favoured and the distribution of projected European warming shifts to lower values. This highlights the importance of specifying appropriate error tolerances in weighting or screening model simulations, to retain an appropriate balance between ruling out unrealistic simulations while retaining an ensemble large enough to support realistic uncertainty estimates for user applications.

The CMIP6 results are a potential source of information for a future update to UKCP, alongside new results from improved models developed by the Met Office and other modelling centres for the next phase of CMIP. Another potential source is ensemble projections created by applying alternative initial states to a single climate model (e.g. Lehner et al., 2020). In such an update, priorities would include resolving, where feasible, limitations of the current UKCP datasets. For example, the impact of advection of maritime showers in enhancing winter precipitation changes over the terrestrial UK is currently captured only in UKCP-Local. This effect could potentially be included in a future release of UKCP-Probabilistic, by combining UKCP-Global and UKCP-Local information in an updated methodology. The incomplete representation of uncertainties in UKCP-Regional and UKCP-Local could also be addressed, by designing new projections in which the selection of global driving models samples more evenly the distributions of UKCP-Probabilistic and UKCP-Global outcomes derived by combining single- and multi-model information.

However, plans for a potential major update to the UKCP scenarios are contingent upon ongoing research and assessment of priorities and needs for new UK climate information. Prior to any major update, additions to the existing UKCP18 results will be produced as opportunities and user needs arise. One driver for these will be the next statutory Climate Change Risk Assessment (https://www.legislation.gov.uk/ukpga/2008/27/contents).

References

Addison H, Kendon E, Ravuri S, Aitchison L, Watson PAG (2022). Machine learning emulation of a local-scale UK climate model. Tackling Climate Change with Machine Learning workshop at NeurIPS 2022. Available from https://www.climatechange.ai/papers/neurips2022/21.

Allan RJ, Ansell TJ (2006). A new globally complete monthly historical gridded mean sea level pressure data set (HadSLP2): 1850–2003. J. Clim. 19: 5816-5842.

Andrews T, Bodas-Salcedo A, Gregory JM, Dong Y, Armour KC, Paynter D, et al. (2022). On the effect of historical SST patterns on radiative feedback. J. Geophys. Res. 127: e2022JD036675. https://doi.org/10.1029/2022JD036675.

Arias PA, Bellouin N, Coppola E, Jones RG, Krinner G, Marotzke J, Naik V, Palmer MD, Plattner G-K, Rogelj J, Rojas M, Sillmann J, Storelvmo T, Thorne PW, Trewin B, Achuta Rao K, Adhikary B, Allan RP, Armour K, Bala G, Barimalala R, Berger S, Canadell JG, Cassou C, Cherchi A, Collins W, Collins WD, Connors SL, Corti S, Cruz F, Dentener FJ, Dereczynski C, Di Luca A, Diongue Niang A, Doblas-Reyes FJ, Dosio A, Douville H, Engelbrecht F, Eyring V, Fischer E, Forster P, Fox-Kemper B, Fuglestvedt JS, Fyfe JC, Gillett NP, Goldfarb L, Gorodetskaya I, Gutierrez JM, Hamdi R, Hawkins E, Hewitt HT, Hope P, Islam AS, Jones C, Kaufman DS, Kopp RE, Kosaka Y, Kossin J, Krakovska S, Lee J-Y, Li J, Mauritsen T, Maycock TK, Meinshausen M, Min S-K, Monteiro PMS, Ngo-Duc T, Otto F, Pinto I, Pirani A, Raghavan K, Ranasinghe R, Ruane AC, Ruiz L, Sallée J-B, Samset BH, Sathyendranath S, Seneviratne SI, Sörensson AA, Szopa S, Takayabu I, Tréguier A-M, van den Hurk B, Vautard R, von Schuckmann K, Zaehle S, Zhang X, Zickfeld K (2021). Technical Summary. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte V, Zhai P, Pirani A, Connors SL, Péan C, Berger S, Caud N, Chen Y, Goldfarb L, Gomis MI, Huang M, Leitzell K, Lonnoy E, Matthews JBR, . Maycock TK, Waterfield T, Yelekçi O, Yu R, Zhou B (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 33−144. doi:10.1017/9781009157896.002.

Arnell NW, Kay AL, Freeman A, Rudd AC, Lowe JA (2021). Changing climate risk in the UK: A multi-sectoral analysis using policy-relevant indicators. Clim. Risk Man. 31: 100265. https://doi.org/10.1016/j.crm.2020.100265.

Behera SK, Yamagata T (2001). Subtropical SST dipole events in the southern Indian Ocean. Geophys. Res. Lett. 28: 327-330.

Bellomo K, Murphy LN, Cane MA, Clement AC, Polvani LM (2018). Historical forcings as main drivers of the Atlantic multidecadal variability in the CESM large ensemble. Clim. Dyn. 50: 3687-3698.

Bellucci A, Mariotti A, Gualdi S (2017). The role of forcings in the twentieth-century North Atlantic multidecadal variability: The 1940–75 North Atlantic cooling case study. J. Clim. 30: 7317-7337.

Betts RA, Brown B (2021). Introduction. In: The Third UK Climate Change Risk Assessment Technical Report [Betts RA, Haward AB, Pearson KV (eds.)]. Prepared for the Climate Change Committee, London. Available from https://www.ukclimaterisk.org/independent-assessment-ccra3/technical-report/.

Bi D, Dix M, Marsland S, O'Farrell S, Sullivan A, Bodman R, Law R, Harman I, Srbinovsky J, Rashid HA, Dobrohotoff P, Mackallah C, Yan H, Hirst A, Savita A, Dias FB, Woodhouse M, Fiedler R, Heerdegen A (2020). Configuration and spin-up of ACCESS-CM2, the new generation Australian Community Climate and Earth System Simulator Coupled Model. J. S. Hem. Earth Syst. Sci. 70: 225-251. https://doi.org/10.1071/ES19040.

Blackport R, Fyfe JC (2022). Climate models fail to capture strengthening wintertime North Atlantic jet and impacts on Europe. Sci. Adv. 8: eabn3112

Bladé I, Fortuny D, van Oldenborgh GJ, Liebmann B (2012). The summer North Atlantic Oscillation in CMIP3 models and related uncertainties in projected summer drying in Europe. J. Geophys. Res. 117: D16104.

Bloch-Johnson J, Rugenstein M, Gregory J, Cael BB, Andrews T (2022). Climate impact assessments should not discount 'hot' models. Nature 608: 667. doi: https://doi.org/10.1038/d41586-022-02241-6.

Bock L, Lauer A, Schlund M, Barreiro M, Bellouin N, Jones C, et al. (2020). Quantifying progress across different CMIP phases with the ESMValTool. J. Geophys. Res. 125: e2019JD032321. https://doi.org/10.1029/2019JD032321.

Booth B, Dunstone N, Halloran P, Andrews T, Bellouin N (2012). Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability. Nature 484:228-232.

Booth BBB, Harris GR, Murphy JM, House JI, Jones CD, Sexton D, Sitch S (2017). Narrowing the range of future climate projections using historical observations of atmospheric CO_2. J. Clim. 30: 3039-3053.

Box GEP, Draper NR (1987). Empirical model-building and response surfaces. John Wiley & Sons.

Brunner L, Pendergrass AG, Lehner F, Merrifield AL, Lorenz R, Knutti R (2020a). Reduced global warming from CMIP6 projections when weighting models by performance and independence. Earth Syst. Dynam. 11: 995-1012.

Brunner L, McSweeney C, Ballinger AP, Befort DJ, Benassi, M, Booth B, Coppola E, de Vries H, Harris G, Hegerl GC, Knutti R, Lenderink G, Lowe J, Nogheretto R, O'Reilly C, Qasmi S, Ribes A, Stocchi P, Undorf S (2020b). Comparing methods to constrain future European climate projections using a consistent framework. J. Clim. 33: 8671-8692. https://doi.org/10.1175/JCLI-D-19-0953.1

Carslaw K, Lee L, Reddington C et al. (2013). Large contribution of natural aerosols to uncertainty in indirect forcing. Nature 503: 67-71. https://doi.org/10.1038/nature12674.

Chan SC, Kendon EJ, Roberts N, Blenkinsop S, Fowler HJ (2018). Large-scale predictors for extreme hourly precipitation events in convection-permitting climate simulations. J. Clim. 31: 2115-2131. https://doi.org/10.1175/JCLI-D-17-0404.1.

Chan S, Kendon E, Fowler H, Youngman BD, Dale M, Short C (2023). New extreme rainfall projections for improved climate resilience of urban drainage systems. Climate Services, submitted.

Charles E, Meyssignac B, Ribes A (2020). Observational constraint on greenhouse gas and aerosol contributions to global ocean heat content changes. J. Clim. 33: 10579-10591.

Collins M, Knutti R, Arblaster J, Dufresne J-L, Fichefet T, Friedlingstein P, Gao X, Gutowski W, Johns T, Krinner G, Shongwe M, Tebaldi C, Weaver A, Wehner M (2013). Long-term climate change: projections, commitments and irreversibility. In*: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Stocker T, Qin D, Plattner G-K, Tignor M, Allen S, Boschung J, Nauels A, Xia Y, Bex V, Midgley P, eds.). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1029-1136, doi:10.1017/CBO9781107415324.023.

Cowtan K, Way RG (2014). Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. Q. J. R. Meteorol. Soc. 140: 1935-1944.

Dee DP, et al. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137: 553-597.

Delworth T, Mann, M (2000). Observed and simulated multidecadal variability in the Northern Hemisphere. Clim. Dyn. 16: 661-676. https://doi.org/10.1007/s003820000075.

Deser C, Sun L, Tomas RA, Screen J (2016). Does ocean coupling matter for the northern extratropical response to projected Arctic sea ice loss? Geophys. Res. Lett. 43: 2149-2157.

Deser C, Hurrell JW, Phillips AS (2017). The role of the North Atlantic Oscillation in European climate projections. Clim. Dyn. 49: 3141-3157.

Deser C, Lehner F, Rodgers KB, Ault TR, Delworth TL, DiNezio PN, Fiore AM, Frankignoul C, Fyfe JC, Horton DE, Kay JE, Knutti R, Lovenduski NS, Marotzke J, McKinnon KA, Minobe S, Randerson JT, Screen JA, Simpson IR, Ting M (2020). Insights from Earth system model initial-condition large ensembles and future prospects. Nat. Clim. Change 10: 277-286. https://doi.org/10.1038/s41558-020-0731-2.

Dorrington J, Strommen K, Fabiano F, Molteni F (2022). CMIP6 models trend toward less persistent European blocking regimes in a warming climate. Geophys. Res. Lett. 49: e2022GL100811.

Eyring V, Bony S, Meehl GA, Senior CA, Stevens B, Stouffer RJ, Taylor KE (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organisation. Geosci. Model Dev. 9: 1937-1958.

Fabiano F, Meccia VL, Davini P, Ghinassi P, Corti S (2021). A regime view of future atmospheric circulation changes in northern mid-latitudes. Wea. Clim. Dynam. 2: 163-180.

Fasullo JT, Lamarque J-F, Hannay C, Rosenbloom N, Tilmes S, DeRepentigny P, et al. (2022). Spurious late historical-era warming in CESM2 driven by prescribed biomass burning emissions. Geophys. Res. Lett. 49: e2021GL097420. https://doi. org/10.1029/2021GL097420.

Fereday DR, Knight JR, Scaife AA, Folland CK, Philipp A (2008). Cluster analysis of North Atlantic-European circulation types and links with tropical Pacific sea surface temperatures. J. Clim. 21: 3687-3703.

Folland CK., Knight J, Linderholm HW, Fereday D, Ineson S, Hurrell JW (2009). The summer North Atlantic Oscillation: Past, present, and future. J. Clim. 22: 1082-1103.

Frieler K, Meinshausen M, Mengel M (2012). A scaling approach to probabilistic assessment of regional climate change. J. Clim. 25: 3117-3144.

Fung F, Lowe J, Mitchell JFB, Murphy J, Bernie D, Gohar L, Harris G, Howard T, Kendon E, Maisey P, Palmer M, Sexton D (2018). UKCP18 Guidance: How to use the UKCP18 Land Projections. Met Office Hadley Centre, Exeter. Available from https://www.metoffice.gov.uk/research/approach/collaboration/ukcp/data/guidance.

Fung F, Bett P, Murphy J, Gohar L, Lowe J (2020). UKCP Factsheet: UKCP Global (60km) – Low emissions scenario (RCP 2.6). Met Office. Available from https://www.metoffice.gov.uk/research/approach/collaboration/ukcp/ukcp18-project-news/index.

Fyfe JC, Kharin VV, Santer BD, Cole JNS, Gillett NP (2021). Significant impact of forcing uncertainty in a large ensemble of climate model simulations. Proc. Nat. Acad. Sci. 18: e2016549118. https://doi.org/10.1073/pnas.2016549118.

Garry FK, Bernie DJ, Davie JCS, Pope ECD (2021). Future climate risk to UK agriculture from compound events. Clim. Risk Man. 100282. https://doi.org/10.1016/j.crm.2021.100282.

Gohar L, Bernie D, Good P, Lowe JA (2018). UKCP18 derived projections of future climate over the UK. Available from https://www.metoffice.gov.uk/research/collaboration/ukcp/guidance-science-reports.

Golaz J-C, Caldwell PM, Van Roekel LP, Petersen MR, Tang Q, Wolfe JD, et al. (2019). The DOE E3SM coupled model version 1: Overview and evaluation at standard resolution. J. Adv. Model. Earth Syst. 11: 2089-2129. https://doi.org/10.1029/2018MS001603.

Gregory JM, Andrews T, Ceppi P, Mauritsen T, Webb MJ (2020). How accurately can the climate sensitivity to CO_2 be estimated from historical climate change? Clim. Dyn. 54: 129-157. https://doi.org/10.1007/s00382-019-04991-y.

Hanlon HM, Bernie D, Carigi G, Lowe JA (2021). Future changes to high impact weather in the UK. Clim. Change 166: 50. https://doi.org/10.1007/s10584-021-03100-5.

Harris GR, Collins M, Sexton DMH, Murphy JM, Booth BBB (2010). Probabilistic projections for 21st century European climate. Nat. Hazards Earth Syst. Sci. 10: 2009-2020.

Harris GR, Sexton DMH, Booth BBB, Collins M, Murphy JM (2013). Probabilistic projections of transient climate change. Clim. Dyn. 40: 2937-2972.

Harris GR, Murphy JM, Pirret JSR, Sexton DMH (2022), Update to UKCP18 probabilistic projections, Met Office. Available from https://www.metoffice.gov.uk/research/approach/collaboration/ukcp/guidancescience-reports.

Hausfather Z, Marvel K, Schmidt GA, Nielsen-Gammon JW, Zelinka M (2022). Climate simulations: recognize the 'hot model' problem. Nature 605: 26-29. doi: https://doi.org/10.1038/d41586-022-01192-2.

Hegerl GC, Ballinger AP, Booth BBB, Borchert LF, Brunner L, Donat MG, Doblas-Reyes FJ, Harris GR, Lowe J, Mahmood R, Mignot J, Murphy JM, Swingedouw D, Weisheimer A (2021). Toward consistent observational constraints in climate predictions and projections. Front. Clim. 3: 678109.

Hurrell JW (1995). Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. Science 269: 676-679. doi: 10.1126/science.269.5224.676.

Hurrell JW, Kushnir Y, Ottersen G, Visbeck M. (eds) (2003). The North Atlantic Oscillation: Climatic significance and environmental impact (American Geophysical Union).

Ineson I, Scaife AA, Knight JR, Manners JC, Dunstone NJ, Gray LJ, Haigh JD (2011). Solar forcing of winter climate variability in the northern hemisphere. Nat. Geosci. 4: 753-757.

IPCC, 2021: Annex II: Models [Gutiérrez JM, Treguier A-M (eds.)]. In: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [Masson-Delmotte V, Zhai P, Pirani A, Connors SL, Péan C, Berger S, Caud N, Chen Y, Goldfarb L, Gomis MI, Huang M, Leitzell K, Lonnoy E, Matthews JBR, Maycock TK, Waterfield T, Yelekçi O, Yu R, Zhou B (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 2087–2138, doi:10.1017/9781009157896.016.

IPCC, 2021: Annex VII: Glossary [Matthews JBR, Möller V, van Diemen R, Fuglestvedt JS, Masson-Delmotte V, Méndez C, Semenov S, Reisinger A (eds.)]. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [Masson-Delmotte V, Zhai P, Pirani A, Connors SL, Péan C, Berger S, Caud N, Chen Y, Goldfarb L, Gomis MI, Huang M, Leitzell K, Lonnoy E, Matthews JBR, Maycock TK, Waterfield T, Yelekçi O, Yu R, Zhou B(eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 2215–2256, doi:10.1017/9781009157896.022.

Jackson LC, Kahana R, Graham T, Ringer MA, Woollings T, Mecking JV, Wood RA (2015). Global and European climate impacts of a slowdown of the AMOC in a high resolution GCM. Clim. Dyn. 45: 3299-3316.

Kahraman A, Kendon EJ, Fowler HJ, Wilkinson JM (2022). Contrasting future lightning stories across Europe. Environ. Res. Lett. 17: 114023. https://doi.org/10.1088/1748-9326/ac9b78.

Keat WJ, Kendon EJ, Bohnenstengel SI (2021). Climate change over UK cities: The urban influence on extreme temperatures in the UK Climate Projections. Clim. Dyn. 57: 3583-3597. https://link.springer.com/article/10.1007%2Fs00382-021-05883-w.

Kendon EJ, Rowell DP, Jones RG (2010). Mechanisms and reliability of future projected changes in daily precipitation. Clim Dyn 35: 489–509. https://doi.org/10.1007/s00382-009-0639-z.

Kendon EJ, Roberts NM, Fosser G, Martin GM, Lock AP, Murphy JM, Senior CA, Tucker S (2020). Greater future UK winter precipitation increase in new convection-permitting scenarios. J. Clim. 33: 7303-7318.

Kendon E, Short C, Pope J, Chan S, Wilkinson J, Tucker S, Bett P, Harris G (2021). Update to UKCP Local (2.2km) projections. Available from https://www.metoffice.gov.uk/research/collaboration/ukcp/guidance-science-reports.

Kendon E, Short C, Cotterill D, Pirret J, Chan S, Pope J (2023). UK Climate Projections: UKCP Local (2.2km) transient projections. Available from https://www.metoffice.gov.uk/research/approach/collaboration/ukcp/guidancescience-reports.

Kendon M, McCarthy MP, Jevrejeva S, Matthews A, Sparks T, Garforth J, Kennedy J (2022). State of the UK climate 2021. Int. J. Climatol. 42: 1-80.

Kendon EJ, Fischer E, Short CJ (2023). Variability conceals emerging trend in 100yr projections of local hourly rainfall extremes. Nature Comms.14: 1133. https://doi.org/10.1038/s41467-023-36499-9.

Knight JR, Folland CK, Scaife AA (2006). Climate impacts of the Atlantic Multidecadal Oscillation. Geophys. Res. Lett. 33: L17706. doi:10.1029/2006GL026242.

Knutti R, Sedláček J, Sanderson BM, Lorenz R, Fischer EM, Eyring V (2017). A climate model projection weighting scheme accounting for performance and interdependence. Geophys. Res. Lett. 44: 1909-1918.

Lai WKM, Robson JI, Wilcox LJ, Dunstone N (2022). Mechanisms of internal Atlantic Multidecadal Variability in HadGEM3-GC3.1 at two different resolutions. J. Clim. 35: 1365-1383.

Lee JY, Marotzke J, Bala G, Cao L, Corti S, Dunne JP, Engelbrecht F, Fischer E, Fyfe JC, Jones C, Maycock A, Mutemi J, Ndiaye O, Panickal S, Zhou T (2021) Future global climate: Scenario-based projections and near-term information. In*: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [Masson-Delmotte V, Zhai P, Pirani A, Connors SL, Péan C, Berger S, Caud N, Chen Y, Goldfarb L, Gomis MI, Huang M, Leitzell K, Lonnoy E, Matthews JBR, Maycock TK, Waterfield T, Yelekçi O, Yu R, Zhou B (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 553–672, doi:10.1017/9781009157896.006.

Lehner F, Deser C, Maher N, Marotzke J, Fischer EM, Brunner L, Knutti R, Hawkins E (2020). Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. Earth Syst. Dynam. 11: 491-508. https://doi.org/10.5194/esd-11-491-2020.

Liang Y, Gillett NP, Monahan AH (2020). Climate model projections of 21st century global warming constrained using the observed warming trend. Geophys. Res. Lett. 47: e2019GL086757. https://doi.org/10.1029/2019GL086757.

Lin Y, Huang X, Liang Y, Qin Y, Xu S, Huang W, et al. (2020). Community Integrated Earth System Model (CIESM): Description and evaluation. J. Adv. Model. Earth Syst. 12: e2019MS002036. https://doi.org/10.1029/2019MS002036.

Liu M, Soden BJ, Vecchi BA, Wang C (2023). The spread of ocean heat uptake efficiency traced to ocean salinity. Geophys. Res. Lett. 50: e2022GL100171. https://doi.org/10.1029/2022GL100171.

Lowe JA, Bernie D, Bett P, Bricheno L, Brown S, Calvert D, Clark R, Eagle K, Edwards T, Fosser G, Fung F, Gohar L, Good P, Gregory J, Harris G, Howard T, Kaye N, Kendon E, Krijnen J, Maisey P, McDonald R, McInnes R, McSweeney C, Mitchell JFB, Murphy J, Palmer M, Roberts C, Rostron J, Sexton D, Thornton H, Tinker J, Tucker S, Yamazaki K, Belcher S (2018). UKCP18 Science Overview Report. Available from https://www.metoffice.gov.uk/research/collaboration/ukcp/guidance-science-reports.

Loeb NG, Wielicki BA, Doelling DR, et al. (2009). Toward optimal closure of the Earth's top-of-atmosphere radiation budget. J. Clim. 22: 748-766. doi:10.1175/2008JCLI2637.1.

Mann ME, Steinman BA, Brouillette DJ, Miller SK (2021). Multidecadal climate oscillations during the past millennium driven by volcanic forcing. Science 371: 1014-1019.

Manning C, Kendon EJ, Fowler HJ, Roberts NM, Berthou S, Suri D, Roberts MJ (2022). Extreme windstorms and sting jets in convection-permitting climate simulations over Europe. Clim. Dyn. 58: 2387-2404.

Mavilia I, Bellucci A, Athanasiadis PJ, Gualdi S, Msadek R, Ruprich-Robert Y (2018). On the spectral characteristics of the Atlantic multidecadal variability in an ensemble of multi-century simulations. Clim. Dyn. 51: 3507-3520. https://doi.org/10.1007/s00382-018-4093-7.

McCarthy M, Christidis N, Dunstone N, Fereday D, Kay G, Klein-Tank A, Lowe J, Petch J, Scaife A, Stott P (2019). Drivers of the UK summer heatwave of 2018. Weather 74: 390-396.

McSweeney CF, Jones RG, Lee RW, Rowell DP (2015). Selecting CMIP5 GCMs for downscaling over multiple regions. Clim. Dyn. 44: 3237-3260.

McSweeney CF, Murphy J, Sexton D, Rostron J, Yamazaki K, Harris G (2018). Selection of CMIP5 members to augment a perturbed–parameter ensemble of global realisations of 21st century climate for the UKCP18 scenarios. Hadley Centre Technical Note No. 102, Met Office, Exeter, UK.

McSweeney C, Yamazaki K (2020). UKCP European circulation indices: Winter Atlantic pressure gradient (North Atlantic Oscillation – NAO). UKCP Factsheet. Met Office. Available from https://www.metoffice.gov.uk/research/approach/collaboration/ukcp/about/project-news.

Moreno-Chamarro E, Caron L-P, Ortega P, Tomas SL, Roberts MJ (2021). Can we trust CMIP5/6 future projections of European winter precipitation? Environ. Res. Lett. 16: 054063. https://doi.org/10.1088/1748-9326/abf28a.

Morice CP, Kennedy JJ, Rayner NA, Jones PD (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 dataset. J. Geophys. Res. 117: D08101.

Morice CP, Kennedy JJ, Rayner NA, Winn JP, Hogan E, Killick RE, et al. (2021). An updated assessment of near-surface temperature change from 1850: the HadCRUT5 data set. J. Geophys. Res. 126: e2019JD032361. https://doi.org/10.1029/2019JD032361.

Moss RH, Edmonds JA, Hibbard KA, Manning MR, Rose SK, van Vuuren DP, Carter TR, Emori S, Kainuma M, Kram T, Meehl GA, Mitchell JFB, Nakicenovic N, Riahi K, Smith SJ, Stouffer RJ, Thomson AM, Weyant JP, Wilbanks TJ (2010). The next generation of scenarios for climate change research and assessment. Nature 463:747-756. doi:10.1038/nature08823.

Mulcahy JP, Jones CG, Rumbold ST, Kuhlbrodt T, Dittus AJ, Blockley EW, Yool A, Walton J, Hardacre C, Andrews T, Bodas-Salcedo A, Stringer M, de Mora L, Harris P, Hill R, Kelley D, Robertson E, Tang Y (2022). UKESM1.1: Development and evaluation of an updated configuration of the UK Earth System Model. Geosci. Model Dev. Discuss. https://doi.org/10.5194/gmd-2022-113.

Murphy JM, Harris GR, Sexton DMH, Kendon EJ, Bett PE, Clark RT, Eagle KE, Fosser G, Fung F, Lowe JA, McDonald RE, McInnes RN, McSweeney CF, Mitchell JFB, Rostron JW, Thornton HE, Tucker S, Yamazaki K (2018). UKCP18 Land Projections: Science Report. Available from https://www.metoffice.gov.uk/research/collaboration/ukcp/guidance-science-reports.

Murphy JM, Brown SJ, Harris GR (2020). UKCP Additional Land Products: Probabilistic projections of climate extremes. Available from https://www.metoffice.gov.uk/research/approach/collaboration/ukcp/science/science-reports.

Nakicenovic N, Swart R (2000). Special Report on Emissions Scenarios. Cambridge University Press.

Neal R, Fereday D, Crocket R, Comer R (2016). A flexible approach to defining weather patterns and their application in weather forecasting over Europe. Meteorol. Appl. 23:389-400.

Nijsse FJMM, Cox PM, Williamson MS (2020). Emergent constraints on transient climate response (TCR) and equilibrium climate sensitivity (ECS) from historical warming in CMIP5 and CMIP6 models. Earth Syst. Dynam. 11: 737-750.

O'Neill BC. Kriegel E, Riahi K, Ebi KL, Hallegatte S, Carter TR, Mathur R, van Vuuren DP (2013). A new scenario framework for climate change research: The concept of Shared Socioeconomic Pathways. Clim. Change 122: 387-400. https://doi.org/10.1007/s10584-013-0905-2.

O'Reilly CH, Woollings T, Zanna L (2017). The dynamical influence of the Atlantic Multidecadal Oscillation on continental climate. J. Clim. 30: 7213-7230.

O'Reilly CH, Zanna L, Woollings T (2019). Assessing external and internal sources of Atlantic multidecadal variability using models, proxy data, and early instrumental indices. J. Clim 32: 7727-7745.

Palmer TE, Booth BBB, McSweeney CF (2021). How does the CMIP6 ensemble change the picture for European climate projections? Environ. Res. Lett. 16: 094042.

Palmer TE, McSweeney CF, Booth BBB, Priestley MDK, Davini P, Brunner L, Borchert L, Menary MB (2023). Performance based sub-selection of CMIP6 models for impact assessments in Europe. Earth Syst. Dynam. 14: 457-483. https://doi.org/10.5194/esd-14-457-2023.

Pan R, Shu Q, Wang Q, Wang S, Song Z, He Y, Qiao F (2023). Future Arctic climate change in CMIP6 strikingly intensified by NEMO-family climate models. Geophys. Res. Lett. 50: e2022GL102077. https://doi.org/10.1029/2022GL102077.

Pope JO, Brown K, Fung F, Hanlon HM, Neal R, Palin EJ, Reid A (2022). Investigation of future climate change over the British Isles using weather patterns. Clim. Dyn. 58: 2405-2419.

Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, Kent EC, Kaplan A (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. J. Geophys. Res. 108: 4407. doi:10.1029/2002JD002670.

Riahi K, van Vuuren DP, Kriegler E, Edmonds J, O'Neill BC, Fujimori S, Bauer N, Calvin K, Dellink R, Fricko O, Lutz W, Popp A, Cuaresma JCKCS, Leimbach M, Jiang L, Kram T, Rao S, Emmerling J, Ebi K, Hasegawa T, Havlik P, Humpenöder F, Da Silva LA, Smith S, Stehfest E, Bosetti V, Eom J, Gernaat D, Masui T, Rogelj J, Strefler J, Drouet L, Krey V, Luderer G, Harmsen M, Takahashi K, Baumstark L, Doelman JC, Kainuma M, Klimont, Z, Marangoni G, Lotze-Campen H, Obersteiner M, Tabeau A, Tavoni M (2017). The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions Implications: An overview. Global Environ. Change 42: 153-168. https://doi.org/10.1016/j.gloenvcha.2016.05.009.

Ribes A, Qasmi S, Gillett NP (2021). Making climate projections conditional on historical observations. Sci. Adv. 7: eabc0671.

Richardson M, Cowtan K, Hawkins E, Stolpe MB (2016). Reconciled climate response estimates from climate models and the energy budget of Earth. Nat. Clim. Change 6: 931-936.

Ridley JK, Blockley EW, Jones GS (2022). A change in climate state during a pre-industrial simulation of the CMIP6 model HadGEM3 driven by deep ocean drift. Geophys. Res. Lett. 49: e2021GL097171. https://doi.org/10.1029/2021GL097171.

Roberts MJ, Baker A, Blockley EW, Calvert D, Coward A, Hewitt HT, Jackson LC, Kuhlbrodt T, Mathiot P, Roberts CD, Schiemann R, Seddon J, Vannière B, Vidale PL (2019). Description of the resolution hierarchy of the global coupled HadGEM3-GC3.1 model as used in CMIP6 HighResMIP experiments. Geosci. Model Dev. 12: 4999-5028. https://doi.org/10.5194/gmd-12-4999-2019.

Rostron JW, Sexton DMH, McSweeney CF, Yamazaki K, Andrews T, Furtado K, Ringer MA, Tsushima Y (2020). The impact of performance filtering on climate feedbacks in a perturbed parameter ensemble. Clim. Dyn. 55: 521–551.

Rousi E, Rust HW, Ulbrich U, Anagnostopoulou C (2020). Implications of winter NAO flavors on present and future European climate. Climate 8: 13.

Rowell DP, Jones RG (2006). Causes and uncertainty of future summer drying over Europe. Clim. Dyn. 27: 281-299.

Sanderson BM, Knutti R, Caldwell P (2015). A representative democracy to reduce interdependency in a multimodel ensemble. J. Clim. 28: 5171-5194.

Screen JA, Eade R, Smith DM, Thomson S, Yu, H (2022). Net equatorward shift of the jet streams when the contribution from sea-ice loss is constrained by observed eddy feedback. Geophys. Res. Lett. 49: e2022GL100523. https://doi.org/10.1029/2022GL100523.

Sexton DMH, Murphy JM, Collins M, Webb MJ (2012). Multivariate probabilistic projections using imperfect climate models part I: Outline of methodology. Clim. Dyn. 38: 2513-2542.

Sexton D, Harris G (2015). The importance of including variability in climate change projections used for adaptation. Nat. Clim. Change 5: 931-936.

Sexton DMH, Murphy JM (2012). Multivariate prediction using imperfect climate models part II: robustness of methodological choices and consequences for climate sensitivity. Clim. Dyn. 38: 2543-2558. doi:10.1007/s00382-011-1209-8.

Sexton D, Yamazaki K, Murphy J, Rostron J (2020). Assessment of drifts and internal variability in UKCP projections. Available from https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/research/ukcp/ukcp-climate-drifts-report.pdf.

Sexton DMH, McSweeney CF, Rostron JW, Yamazaki K, Booth BBB, Murphy JM, Regayre L, Johnson J, Karmalkar A (2021). A perturbed parameter ensemble of HadGEM3-GC3.05 coupled model projections: Part 1: Selecting the parameter combinations. Clim. Dyn. 56: 3395-3436. https://doi.org/10.1007/s00382-021-05709-9.

Sherwood SC, Webb MJ, Annan JD, Armour KC, Forster PM, Hargreaves JC, et al. (2020). An assessment of Earth's climate sensitivity using multiple lines of evidence. Rev. Geophys. 58: e2019RG000678. https://doi.org/10.1029/2019RG000678.

Slater L, Huntingford C, Pywell RF, Redhead J, Kendon EJ (2022). Resilience of UK crop yields to changing climate extremes. Earth Syst. Dynam. 13: 1377-1396. https://doi.org/10.5194/esd-2021-92.

Smith CJ, Kramer RJ, Myhre G, Alterskjær K, Collins W, Sima A, Boucher O, Dufresne J-L, Nabat P, Michou M, Yukimoto S, Cole J, Paynter D, Shiogama H, O'Connor FM, Robertson E, Wiltshire A, Andrews T, Hannay C, Miller R, Nazarenko L, Kirkevåg A, Olivié D, Fiedler S, Lewinschal A, Mackallah C, Dix M, Pincus R, Forster PM (2020a). Effective radiative forcing and adjustments in CMIP6 models. Atmos. Chem. Phys. 20: 9591-9618. https://doi.org/10.5194/acp-20-9591-2020.

Smith DM, Scaife AA, Eade R, Athanasiadis P, Bellucci A, Bethke I, Bilbao R, Borchert LF, Caron L-P, Counillon F, Danabasoglu G, Delworth T, Doblas-Reyes FJ, Dunstone NJ, Estella-Perez V, Flavoni D, Hermanson L, Keenlyside N, Kharin V, Kimoto M, Merryfield WJ, Mignot J, Mochizuki T, Modali K, Monerie P-A, Müller WA, Nicolí D, Ortega P, Pankatz K, Pohlmann H, Robson J, Ruggieri P, Sospedra-Alfonso R, Swingedouw D, Wang Y, Wild D, Yeager S, Yang X, Zhang L (2020b). North Atlantic climate far more predictable than models imply. Nature 583: 796-800.

Smith DM, Eade R, Andrews MB et al. (2022). Robust but weak winter atmospheric circulation response to future Arctic sea-ice loss. Nat. Commun. 13: 727. https://doi.org/10.1038/s41467-022-28283-y.

Smith CJ, Harris GR, Palmer MD, Bellouin N, Collins W, Myhre G, Schulz M, Golaz J-C, Ringer M, Storelvmo T, Forster PM (2021). Energy budget constraints on the time history of aerosol forcing and climate sensitivity. J. Geophys. Res. 126: e2020JD033622. https://doi.org/10.1029/2020JD033622.

Sobolowski S, Somot S, Fernandez J, Evin G , Maraun D, Kotlarski S, Jury M, Benestad RE, Teichmann C, Christensen OB, Bülow K, Buonomo E, Katragkou E, Steger C, Sørland S, Nikulin G, McSweeney C, Dobler A, Palmer T, Wilcke R, Boé J, Brunner L, Ribes A, Qasmi S, Nabat P, Sevault F, Oudar T, Brands S (2023). EURO-CORDEX CMIP6 GCM selection and ensemble design: Best practices and recommendations. Zenodo. https://doi.org/10.5281/zenodo.7673400.

Song X, Wang D-Y, Li F, Zeng X-D (2021). Evaluating the performance of CMIP6 Earth system models in simulating global vegetation structure and distribution. Adv. Clim. Change Res. 12: 584-595.

Sutton R, Dong B (2012). Atlantic Ocean influence on a shift in European climate in the 1990s. Nat. Geosci. 5: 788-792.

Swingedouw D, Mignot J, Ortega P, Khodri M, Menegoz M, Cassou C, Hanquiez V (2017). Impact of explosive volcanic eruptions on the main climate variability modes. Global and Planetary Change 150: 24-45.

Taylor KE, Stouffer RJ, Meehl GA (2012). An overview of CMIP5 and the experiment design. Bull. Am. Met. Soc. 93: 485-498.

Tebaldi C, Arblaster J (2014). Pattern scaling: Its strengths and limitations, and an update on the latest model simulations. Clim. Change 122: 459-471. https://doi.org/10.1007/s10584-013-1032-9.

Tebaldi C, Debeire K, Eyring V, Fischer E, Fyfe J, Friedlingstein P, Knutti R, Lowe J, O'Neill B, Sanderson B, van Vuuren D, Riahi K, Meinshausen M, Nicholls Z, Tokarska KB, Hurtt G, Kriegler E, Lamarque J-F, Meehl G, Moss R, Bauer SE, Boucher O, Brovkin V, Byun Y-H, Dix M, Gualdi S, Guo H, John JG, Kharin S, Kim YH, Koshiro T, Ma L, Olivié D, Panickal S, Qiao F, Rong X, Rosenbloom N, Schupfner M, Séférian R, Sellar A, Semmler T, Shi X, Song Z, Steger C, Stouffer R, Swart N, Tachiiri K, Tang Q, Tatebe H, Voldoire A, Volodin E, Wyser K, Xin X, Yang S, Yu Y, Ziehn T (2021). Climate model projections from the Scenario Model Intercomparison Project (ScenarioMIP) of CMIP6. Earth Syst. Dynam. 12: 253-293. https://doi.org/10.5194/esd-12-253-2021.

Tokarska KB, Stolpe MB, Sippel S, Fischer EM, Smith CJ, Lehner F, Knutti R (2020). Past warming trend constrains future warming in CMIP6 models. Sci. Adv. 6: eaaz9549.

Toniazzo T, Scaife AA (2006). The influence of ENSO on winter North Atlantic climate (2006). Geophys. Res. Lett. 33: L24704.

Trenberth KE (1997). The definition of El Niño. Bull. Amer. Met. Soc. 78: 2771-2777.

Trenberth KE, Caron JM (2000). The Southern Oscillation revisited: Sea level pressures, surface temperatures, and precipitation. J. Clim. 13: 4358-4365.

Trenberth KE, Shea DY (2006). Atlantic hurricanes and natural variability in 2005. Geophys. Res. Lett. 33: L12704.

Trenberth KE (2015). Has there been a hiatus? Science 349: 691-692.

Tucker SO, Kendon E, Bellouin N, Buonomo E, Johnson B, Murphy J (2021). Evaluation of a new 12km regional perturbed parameter ensemble over Europe. Clim. Dyn. 58: 879-903.

UKCP18 Technical Note: Clipping and Baseline advice on Land Strand 1 data in UKCP18 (2019). Available from https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/research/ukcp/ukcp18-technical-noteclipping-and-baseline-guidance-on-land-strand-1-data-in-ukcp18.pdf.

Williams KD, Copsey D, Blockley EW, Bodas-Salcedo A, Calvert D, Comer R, Davis P, Graham T, Hewitt HT, Hill R, Hyder P, Ineson S, Johns TC, Keen AB, Lee RW, Megann A, Milton SG, Rae JGL, Roberts MJ, Scaife AA, Schiemann R, Storkey D, Thorpe L, Watterson IG, Walters DN, West A, Wood RA, Woollings T, Xavier PK (2018). The Met Office Global Coupled model 3.0 and 3.1 (GC3.0 & GC3.1) configurations. J. Adv. Model Earth Sys. 10: 357-380.

Woollings T, Franzke C, Hodson DLR, Dong B, Barnes EA, Raible CC, Pinto JG (2015). Contrasting interannual and multidecadal NAO variability. Clim. Dyn. 45:539-556.

Yamazaki K, Sexton DMH, Rostron JW, McSweeney CF, Murphy JM, Harris GR (2021). A perturbed parameter ensemble of HadGEM3-GC3.05 coupled model projections: Part 2: Global performance and future changes. Clim. Dyn. 56: 3437-3471. https://doi.org/10.1007/s00382-020-05608-5.

Zelinka MD, Myers TA, McCoy DT, Po-Chedley S, Caldwell PM, Ceppi P, Klein SA, Taylor KE (2020). Causes of higher climate sensitivity in CMIP6 models. Geophys. Res. Lett. 47: e2019GL085782.

Zhang J, Furtado K, Turnock ST, Mulcahy JP, Wilcox LJ, Booth BB, Sexton D, Wu T, Zhang F, Liu Q (2021). The role of anthropogenic aerosols in the anomalous cooling from 1960 to 1990 in the CMIP6 Earth system models. Atmos. Chem. Phys. 21: 18609-18627. https://doi.org/10.5194/acp-21-18609-2021.

Acknowledgements

This work was supported by the Met Office Hadley Centre Climate Programme funded by BEIS. We thank:

- Prof. Chris Jones for advice on the use of models with high climate sensitivity in regional impact studies.
- Dr. Claudia Tebaldi for the use of Figure 1, reproduced from Figure A7 of Tebaldi et al. (2021) and distributed under the Creative Commons Attribution 4.0 License.
- Profs. Lizzie Kendon and Jason Lowe for their reviews.

Appendix: CMIP6 models used in each figure

The table below maps usage of CMIP6 models in Figures 3-20. The set of models varies, dependent on data availability and timings of the analyses collated in this report. Model configuration details are available in IPCC (2021). In all cases, we use the ensemble member denoted by "r1i1" in the variant label used to identify each CMIP6 simulation. This identifies the realisation number and initialisation method used in cases where several simulations started from different initial states are available. See https://pcmdi.llnl.gov/CMIP6/Guide/dataUsers.html.

CMIP6 models	Figures								
	3-4	5-7	8-9	10	11	12-15	16	17-18	19-20
ACCESS-CM2	X	X	X	X	X	X	X	X	
ACCESS-ESM 1-5	X	X	X	X	X	X	X	X	X
AWI-CM-1-1-MR	X	X		X	X	X	X	X	
AWI-ESM-1-1-LR		X		X					
BCC-CSM2-MR	X	X	X	X	X	X	X	X	X
BCC-ESM1	X	X	X	X					
CAMS-CSM1-0	X	X	X	X	X	X	X	X	X
CAS-ESM2-0		X	X	X					
CESM2	X	X	X	X	X	X		X	X
CESM2-FV2	X	X	X	X					
CESM2-WACCM	X	X	X	X		X			X
CESM2-WACCM-FV2	X	X	X	X					
CIESM	X	X	X	X	X	X	X	X	
CMCC-CM2-HR4		X	X	X					
CMCC-CM2-SR5	X	X	X	X	X	X	X	X	
CMCC-ESM2		X	X	X					
CNRM-CM6-1	X	X	X	X	X	X	X	X	X
CNRM-CM6-1-HR	X	X	X	X		X	X	X	X
CNRM-ESM2-1	X	X	X	X	X	X	X	X	X
CanESM5	X	X	X	X	X	X	X	X	X
E3SM-1-0	X	X	X	X	X	X		X	
E3SM-1-1	X	X	X	X	X	X			
E3SM-1-1-ECA	X	X	X	X	X	X		X	
EC-Earth3	X	X	X	X	X	X	X	X	X
EC-Earth3-AerChem	X	X	X	X					
EC-Earth3-CC		X	X	X					
EC-Earth3-Veg	X	X	X	X	X	X	X	X	X
EC-Earth3-Veg-LR		X	X	X					
FGOALS-f3-L	X	X	X	X	X	X	X		X
FGOALS-g3	X	X	X	X	X	X	X	X	X
FIO-ESM-2-0	X	X	X	X	X	X	X	X	
GFDL-CM4	X	X	X	X	X	X		X	X
GFDL-ESM4	X	X	X	X	X	X	X	X	X
GISS-E2-1-G	X	X	X	X	X	X	X	X	X
GISS-E2-1-G-CC	X	X	X	X					
GISS-E2-1-H	X	X	X	X	X	X	X	X	
GISS-E2-2-G		X	X	X					
GISS-E2-2-H		X	X	X					
GISS-E3-G		X	X	X					

Table of CMIP6 models used in Figures 3-20, continued from previous page.

CMIP6 models (continued)	Figures								
	3-4	5-7	8-9	10	11	12-15	16	17-18	19-20
HadGEM3-GC31-MM	X	X	X	X	X	X	X	X	X
HadGEM3-GC31-LL	X	X	X	X	X	X	X	X	X
IITM-ESM	X	X		X	X	X	X	X	
INM-CM4-8	X	X	X	X	X	X	X	X	X
INM-CM5-0	X	X	X	X	X	X	X	X	X
IPSL-CM5A2-INCA		X	X	X					
IPSL-CM6A-LR	X	X	X	X	X	X	X	X	
IPSL-CM6A-LR-INCA		X	X	X					
KACE-1-0-G	X	X	X	X	X	X	X	X	X
KIOST-ESM		X	X	X					
MCM-UA-1-0		X	X	X		X	X		
MIROC-ES2L	X	X	X	X	X	X	X	X	X
MIROC6	X	X	X	X	X	X	X	X	X
MPI-ESM-1-2-HAM		X	X	X					
MPI-ESM1-2-HR	X	X		X	X	X	X	X	X
MPI-ESM1-2-LR		X	X	X					
MRI-ESM-2.0	X	X	X	X	X	X	X	X	X
NESM3	X	X	X	X	X	X	X	X	X
NorCPM1	X	X	X	X					
NorESM2-LM	X	X	X	X	X	X	X		X
NorESM2-MM		X	X	X					
SAM0-UNICON	X	X	X	X					
TaiESM1	X	X	X	X	X	X	X	X	X
UKESM1-0-LL	X	X		X	X	X	X	X	X
UKESM1-1-LL		X		X					

Met Office
FitzRoy Road
Exeter
Devon
EX1 3PB
United Kingdom