Comparison of EuroCORDEX output with UKCP18 regional ensemble

Project CR20-3: Enabling the use and producing improved understanding of EuroCORDEX data over the UK

Clair Barnes*
Department of Statistical Science, UCL

Richard Chandler
Department of Statistical Science, UCL

Chris Brierley
Geography Department, UCL

July 29, 2022

*Please direct all communications to clair.barnes.16@ucl.ac.uk

Contents

Executive Summary

This report summarises the results of analyses designed to compare several collections ('ensembles') of future climate projections for the UK under Representation Concentration Pathway (RCP) 8.5, derived from multiple climate models. The report also evaluates the skill of the models upon which these projections are based. It fulfils Milestones M3.5 and M4.3 of Project CR20-3. The ensembles considered are as follows:

UKCP18 global (60km): this contains 12 members, each produced using variants of the HadGEM3-GC3.05 global climate model (GCM) and with outputs at a spatial resolution of 60km. This ensemble forms part of the 2018 UK Climate Projections (UKCP18).

UKCP18 regional (12km): this also contains 12 members, obtained by downscaling the UKCP18 global ensemble to 12km resolution using corresponding variations of the HadREM3-GA7-05 regional climate model (RCM). This ensemble also forms part of UKCP18.

CMIP5-13: this contains 13 members, each from one of the GCMs in the CMIP5 archive, at spatial resolutions of between 0.75° and 2.8°. The ensemble was included in UKCP18 to supplement the global ensemble described above.

EuroCORDEX regional (12km): this contains 65 members, each combining one of 10 Global Circulation or Earth System Models (ESMs) with one of 13 RCMs. The ensemble was produced as part of the EuroCORDEX programme and was not included in UKCP18. The spatial resolution is 0.11°, similar to the UKCP18 regional ensemble.

EuroCORDEX evaluation: these are the 'evaluation runs' from the EuroCORDEX programme, obtained by using the ERA-Interim reanalysis to drive each of the EuroCORDEX RCMs over the period 1989–2008. They allow us to evaluate the performance of the RCMs in the absence of errors or biases inherited from the driving GCMs. The availability of these evaluation runs dictates the time period used for skill assessment elsewhere in the report.

CMIP5-EC: contains the ten CMIP5 GCM / ESM variants used to drive the EuroCORDEX regional ensemble members, with model output at spatial resolutions of 1.1°-1.9° latitude and 1.1°-2.5° longitude. These variants are derived from six separate models, of which two (EC-EARTH and MPI-M-ESM) each contribute three realisations.

Aims of the analyses

The aims of the analyses described in the report are:

1. To understand the extent to which the UKCP18 ensembles are representative of the broader international body of evidence on current and future climate in the UK. This is achieved by comparing the skill and future projections from the UKCP18 ensembles with those from the other ensembles considered, for a collection of user-relevant climate indices connected with

temperature, precipitation and wind speed.[1]

2. To understand the value added by using regional climate simulations instead of coarse-resolution information from global models. This is achieved by comparing the skill and future projections from the UKCP18 and EuroCORDEX regional ensembles with those from the global ensembles used to drive them (UKCP18 global and CMIP5-EC) as well as with the supplementary CMIP5-13 ensemble, for the same set of climate indices considered under Aim 1 above.

3. To assess the physical plausibility of the ensembles, in terms of the models' ability to reproduce key large-scale modes of variation that influence the climate of the UK: the North Atlantic Oscillation (NAO); Atlantic Meridional Overturning Circulation (AMOC); strength and location of the jetstream; the frequencies of occurrence of circulation-related weather types over the UK; and the climate sensitivity i.e. equilibrium change in global mean surface temperature that one would see in response to a doubling of carbon dioxide concentrations. The regions over which these modes are defined are larger than those covered by the regional ensembles: this assessment therefore focuses on the models in the global ensembles.

4. To provide interpretable summaries of the regional ensembles, that can help users:

    - to identify and understand the main features of the data provided;
    - to identify subsets of the data that are appropriate for use in their own analyses — for example, so as to ensure that a small sample of ensemble members is representative of the entire ensemble.

To examine the skill of the various ensemble members in Aims 1–3, their outputs are compared to the historical estimates of the corresponding climate indices from the following sources:

HadUK-Grid: this data set contains interpolated daily estimates of precipitation and temperature on a grid identical to that used for the UKCP18 regional ensemble but restricted to the UK land surface. Other variables are available as monthly means. To address Aims 1 and 2, comparisons are made between the various ensembles and HadUK-Grid for the period from 1st January 1989 to 31st December 2008. The time period is chosen to match the availability of EuroCORDEX evaluation runs (see above).

ERA-Interim: this reanalysis data set, with a spatial resolution of 0.7°, is used to evaluate indices of large-scale structure under Aim 3.

For the indices considered in Aims 1 and 2, all model outputs have been regridded and cropped to the 12km UK land surface domain used by the Met Office to present UKCP18 and HadUK-Grid data. In all cases, the regridding process replicates that used to produce the land surface portion of the UKCP18 regional datasets.

Aims 1 and 2 are addressed by comparing the relevant ensembles to the UK-Grid dataset, in terms of:

[1] Analyses for a broader range of indices are available from the online project plot explorer at https://github-pages.ucl.ac.uk/UKCORDEX-plot-explorer/.

- The distributions of UK-averaged values of the indices considered. This analysis provides an overall comparison of the ensembles in terms of bias and future change.
- The spatial pattern of the indices over the UK land area. For each ensemble member, this is assessed in terms of the correlation between the observed (HadUK-Grid) and modelled spatial pattern as well as the standard deviations of the two patterns.
- Maps of the observed (HadUK-Grid) values of the indices for the UK land area, together with maps of the mean biases computed from the members of each ensemble.

For the regional ensembles, Aim 4 is addressed using multivariate analyses of variance and deviance. These analyses partition the variation within the ensemble into contributions from the RCMs, the driving GCMs and 'unstructured' residual variation. Results are displayed as maps, with summaries indicating the overall proportions of variation attributable to each source. Moreover, dominant modes of spatial variation are identified for both the RCMs and GCMs within a given ensemble, using methodology developed during the project: we refer to this methodology as 'Ensemble Principal Pattern' (EPP) analysis. As well as helping to visualise the ensemble structure, EPPs can be used to diagnose features that are deserving of further investigation — this can be particularly useful for perturbed physics ensembles, for example to identify perturbations that have led to particularly unusual patterns. An example is shown in Figure 7, where a single ensemble member is associated with a tendency to simulate temperatures that are too cool over Scotland but not elsewhere.

Main findings

The main findings from the analyses are as follows:

Projected changes in UK climate: although not a specific aim of the analyses reported herein, it is worth summarising the projected changes in UK climate for which all ensembles agree in qualitative terms. These changes are evaluated relative to a reference period from 1980 to 2010[2], and are as follows:

- A fairly uniform increase in future temperature across the whole of the UK. The CMIP5 and EuroCORDEX ensembles generally display a similar degree of warming to one another, of around 1°C in both winter and summer in the period 2020–2050 and around 2°C by 2050–2080. The two UKCP18 ensembles suggest a greater degree of warming. In all ensembles, the warming by 2050–2080 is projected to be around double that seen by 2020–2050 in both winter and summer.
- More (resp. less) rapid increases in the temperature on the coolest (resp. warmest) winter days than for the mean winter temperature, indicating a narrower distribution of temperatures about the mean. Conversely, in summer the hottest days are projected to warm more rapidly on average than the coolest. The increased warming of the hottest summer days is most pronounced in the south of England.

[2]This period differs from some previous UKCP18 analyses which used 20-year references. It has been used for consistency with World Meteorological Organisation recommendations.

- An increase in total winter precipitation across the UK, coupled with increases in mean wet-day winter precipitation rates. The magnitude and spatial pattern of changes varies between ensembles, however.
- Fewer wet days in summer, leading to decreases in total summer precipitation. These changes are fairly small in the period to 2050, but become more substantial (from 10% to 25% less precipitation compared to the 1980–2010 period) by 2080.
- Little change in the proportion of rain simulated on the wettest days in winter; in summer however, there is a consistent suggestion that the wettest days will contribute increasing proportions of the total precipitation.
- Little change over time in average wind speed, or in the frequency of 'windy days' (defined as days with wind speed greater than 10.8ms^{-1}).

Comparison of UKCP18 with other ensembles:

- The biases and ranges of both UKCP18 ensembles for the 1989–2008 period are broadly comparable with those of the other ensembles for mean temperatures in both winter and summer, with the exception of some particularly cool runs in the EuroCORDEX regional ensemble in winter. In winter, all model runs tend to overestimate the spatial variability in mean temperatures: the UKCP18 regional runs are among those with the highest correlation with the observed field, however.
- The UKCP18 ensembles have a tendency to underestimate the temperature of the coolest summer days during the 1989–2008 period, while the UKCP18 regional ensemble tends to overestimate the temperature of the hottest summer days in urban areas. These tendencies are not seen in the other ensembles.
- From 1989 to 2008, almost all ensembles tend to underestimate the mean precipitation in the wettest areas of the UK in both summer and winter due to underestimation of wet-day precipitation; and to overestimate it in the driest areas due to the simulation of too many wet days. The UKCP18 regional ensemble is the exception, being slightly too wet almost everywhere. The individual UKCP18 ensemble members have a much smaller range of biases in winter precipitation indices than the EuroCORDEX regional ensemble.
- All of the ensembles fail to simulate high enough wind speeds at higher elevations in both summer and winter: this bias is largest in the two UKCP18 ensembles. A further unusual feature of the UKCP18 ensembles is that there is very little variation between the members with respect to wind speeds in urban areas.
- Both UKCP18 ensembles tend to project larger changes in the future than other ensembles, in some cases with little or no overlap (Figure 80 on page 89 gives a typical selection of results, relating to changes in percentiles of daily summer temperatures for two different periods). The issue seems to be associated primarily with the properties of the GCM used to drive the UKCP18 projections (see 'Physical plausibility and large-scale

modes of variation' below), although for some indices there are indications that the RCM may also play a role.

- Considering future temperature changes over the UK, the spread of both UKCP18 ensembles is substantially larger than that for either the EuroCORDEX or CMIP5 ensembles in winter. The UKCP18 distributions are also skewed towards larger increases in both winter and summer temperatures.
- The UKCP18 ensembles project less uniform warming over the UK than the other ensembles in summer, with nighttime temperatures in urban areas tending to increase more than in the surrounding areas, and with daytime temperatures tending to increase more in southern and central England.
- Although all ensembles project an increase in winter precipitation overall across the UK, the changes are less spatially uniform in the UKCP18 ensembles which project smaller relative increases at higher elevations and, in particular, a reduction in winter precipitation in northern Scotland in the near future. This reduction is not seen in the other ensembles, and seems to be primarily associated with a decrease in the frequency of wet days in the UKCP18 projections, although the wet-day precipitation intensity tends to increase on average. The other ensembles project little change in wet day frequency.
- In summer, the CMIP5 and EuroCORDEX regional ensembles all project a fairly uniform increase of around 2% in mean wet day precipitation intensity by 2050. The UKCP18 ensembles project larger increases in rain shadow areas and in eastern England and Scotland, as well as over London; however, they also project a reduction of up to 10% in mean wet-day intensity elsewhere, with the largest reductions in the areas that currently receive the most summer rainfall. Combined with the tendency of all ensembles to project fewer wet days in summer, the net effect is that the UKCP18 ensemble members all indicate future reductions in total summer precipitation by 2050 whereas 22 of the 65 EuroCORDEX ensemble members suggest increases. By 2080 however, all regional ensembles project reductions in summer precipitation — with the exception of EuroCORDEX runs driven by the CNRM-CM5 GCM.
- The UKCP18 and EuroCORDEX regional ensembles also differ with respect to their dominant modes of variation in future relative winter precipitation changes. The dominant mode within the UKCP18 ensemble (see Figure 89) relates to the strength of orographic enhancement and rain shadow effects: this corresponds to the *secondary* mode of variation within the EuroCORDEX regional ensemble GCMs (Figure 88). The dominant mode in the latter ensemble corresponds to a uniform relative increase, which does not appear as either of the first two modes of variation in UKCP18. This difference in structures is not seen in summer however, where the ensembles' dominant modes of variation are similar. This result suggests that the two ensembles' different sampling of uncertainty manifests itself particularly with respect to the processes controlling winter precipitation.

Comparison of regional and global ensembles: the two sets of ensembles project broadly similar patterns of future change, in qualitative terms. We therefore note here only salient features relating to the evaluation period from 1989 to 2008.

- During winter the global ensembles tend to be too warm at high elevations and around the coast, but too cool over much of central England. By contrast, the regional ensembles tend to be uniformly too cool. This can be attributed to an underlying cold bias in many of the GCMs, offset by local warm biases due to unresolved orography and blurring between land and sea surface temperatures due to the coarse resolution: the regional ensembles inherit the cool bias but not the blurring. This effect is also present in summer, but to a lesser extent.
- All of the global ensembles underestimate the average diurnal temperature range in both winter and summer, as do many of the regional runs although the underestimation is less severe for the latter.
- Summer night-time temperatures are too warm on average in the global ensembles, although to a lesser extent in the UKCP18 global ensemble; this warm bias is reduced or eliminated in the regional ensembles.
- The global ensembles do a poor job of capturing the spatial pattern of mean winter precipitation over the UK: the regional ensembles improve on this. All ensembles tend to simulate too many wet days with too little wet-day precipitation on average in both winter and summer, although the biases are again lower in the regional ensembles.
- The two CMIP5 global ensembles generally generate too little of their rainfall on the wettest days in both summer and winter: in the UKCP18 global ensemble this effect is restricted to the wettest days in western parts of the UK that typically receive the most precipitation overall. All of the regional ensembles, however, tend to simulate too high a proportion of their rainfall on these wetter days.
- Overall, there is little difference between the biases of the global and regional ensembles in terms of wind speed indices. The spatial pattern of surface wind climatologies across the UK is generally poor in all ensembles: correlations with the corresponding patterns in the HadUK-Grid data set are close to zero for some regional ensemble members.

Physical plausibility and large-scale modes of variation: for some indices, the assessment is based on reported results from the literature; for the remainder it is based on a comparison of indices calculated from the global model outputs with the ERA-Interim reanalysis, for the 1989–2008 period.

North Atlantic Oscillation. All of the global models correctly simulate a stronger positive NAO index in winter than in summer, although the CMIP5 models tend to overestimate the index in winter and underestimate in summer. The seasonality is represented more accurately by the UKCP18 global ensemble members. Several members of CMIP5 also

show substantially less interannual variability in the winter NAO index than observations, unlike in UKCP18-Global.

Atlantic Meridional Overturning Circulation. The models tend to slightly underestimate the AMOC strength, with this tendency being more pronounced for the models in CMIP5-EC than in the other global ensembles.

Jetstream. The models do a fair job in capturing variability in the strength of the jetstream, with a similar range of values in the UKCP18-Global and CMIP5-EC ensembles. The jetstream position is less well reproduced however: on average, its position in most of the CMIP models is slightly too far south, while in the UKCP18 global ensemble it is too far north in the summer.

Circulation types. The proportions of days classified into each of 30 different circulation-related weather types are mostly well captured by all of the models, with some variation. Our results agree closely with those of a recent study by other authors (Pope et al., 2021), in which the model outputs were compared to a different reanalysis dataset and over different time periods to those considered here.

Climate sensitivity. Most of the CMIP5 models have an effective climate sensitivity (ECS) within the IPCC "likely" range of 2.5°C to 4.5°C for the Earth itself: two models (IPSL-CM5A-MR and HadGEM2-ES) fall beyond the upper limit, yet still within the "very likely" range which extends up to 5°C. No ECS is published for the HadGEM3-GC3.05 model used in the UKCP18 global ensemble; however, the CMIP6 version of HadGEM3 has an ECS exceeding the upper limit of the IPCC "very likely" range when run at the same resolution. Although this is a different variant of the model, it is reasonable to expect that HadGEM3-GC3.05 also has an ECS above the IPCC likely upper limit. This is probably the main reason that the UKCP18 ensembles tend to project larger changes than the other ensembles considered in this report.

Main features of the regional ensembles: in all situations considered, the RCMs account for the majority of the joint variation in the EuroCORDEX regional ensemble when considering all grid cells simultaneously and accounting for the correlations between them. However, this result appears to be dominated by the between-cell correlations: a more nuanced picture emerges when considering variation across the ensemble at each grid cell individually. The points below summarise the main points in this picture.

- Considering winter temperatures at each grid cell from 1989 to 2008 across the Euro-CORDEX ensemble members, about two thirds of the variance in mean temperature can be attributed to the driving GCM, with 25-30% attributed to the downscaling RCM. The GCM-induced variation is almost all associated with a tendency for each model to be uniformly warmer or cooler over the entire UK. By contrast, the RCM-induced variation is associated with a pattern that is fairly constant except to the east of upland areas where its strength is reduced.

- The structure of EuroCORDEX ensemble variation in summer mean temperatures is qualitatively similar to that in winter, although inter-RCM variation is slightly greater in summer. Moreover, the greatest variations in summer temperatures within the ensemble occur in urban areas, with inter-RCM differences being primarily responsible. The dominant mode of spatial variation attributable to the GCMs is broadly uniform as for winter temperatures, with the warmest runs driven by HadGEM2-ES; the dominant mode of variation for the RCMs picks out the urban centres (see Figure 19b).
- Within the EuroCORDEX regional ensemble, the RCMs contribute a greater proportion of ensemble variation in 'extreme' temperatures (i.e. the warmest summer days and coolest winter days) than they do to variation in seasonal means; and a smaller proportion to variation in the coolest summer days and warmest winter days. These changes are mostly compensated by corresponding changes in the proportions of GCM-attributable variation.
- Biases in mean daily maximum temperature among the EuroCORDEX regional ensemble members are primarily attributable to the driving GCMs in winter, while the RCMs and GCMs are almost equally important sources of bias in mean daily minimum temperature. In summer, the relative importance of the RCMs is increased in both cases: the RCMs and GCMs are almost equally important when considering mean daily maximum temperature, but the RCMs contribute almost two thirds of the total variance in the mean daily minimum.
- The UKCP18 regional ensemble variation in both winter and summer temperatures, which is associated with different parametrisations within the perturbed physics of the models, is almost entirely associated with tendencies for runs to be uniformly warmer or cooler over the entire UK.
- For the EuroCORDEX regional ensemble, the RCMs dominate the within-ensemble variation for all precipitation indices over the 1989–2008 period, accounting for 75% to almost 100% of the variance in winter and slightly less in summer where the GCMs are more influential in low-lying areas. The spatial structure of ensemble variation is more complex for precipitation than for temperature: for example, of approximately 80% of the total variability in mean wet-day winter precipitation that can be ascribed to the choice of RCM, 55% is associated with a pattern representing an increase in precipitation across the UK and with increased enhancement on the western faces of upland areas. In summer however, the dominant pattern indicates precipitation enhancement over the upland areas themselves rather than their west-facing slopes. For extreme precipitation, a high proportion of variation (38% of the total in winter, 41% in summer) is not systematically attributable either to the GCMs or to the RCMs: this reflects the contribution of unpredictable individual storm events on the wettest days in both seasons.
- For indices of wind speed, the members of the EuroCORDEX regional ensemble are more dispersed over higher elevations where the highest wind speeds occur, and also over

urban areas. The within-ensemble variation is dominated by the RCMs which explain around 85% of the variation overall, but up to 100% of the variation in the areas with highest ensemble standard deviations. The dominant mode of RCM spatial variation is very similar to the observed climatology, suggesting a fairly uniform relative bias for each RCM.

- For the EuroCORDEX regional ensemble, future changes in both summer and winter temperatures are driven primarily by the GCMs, especially in the more distant future. For most temperature-related indices the dominant mode of variation within the ensemble is a spatially uniform increase attributable to the driving GCMs: this accounts for 70% of the total ensemble variance of mean temperature changes in winter and 50% in summer. 'Extreme' temperatures (the coolest and warmest summer and winter days) are an exception to this rule however: here, a greater part of the ensemble variation is unstructured and cannot be attributed to the GCMs or the RCMs.
- Similarly, inter-GCM differences are more important than inter-RCM differences when considering future changes in most precipitation indices, although to a lesser extent than for temperature: the principal mode of GCM-related variation represents a fairly uniform relative change in precipitation over the UK. Inter-RCM variation increases over time and becomes locally dominant in some areas by 2050–2080. Comparing projected changes in relative precipitation to those in temperature, a greater proportion of the ensemble variation is unstructured.

Conclusions and recommendations

- Projections of future change based on UKCP18 are likely to be larger than those based on other ensembles, due mainly to the properties of the underlying GCM which is likely to overestimate the Earth's climate sensitivity according to current IPCC assessments. This means that **planning and adaptation decisions based solely on UKCP18 projections may be conservative**.
- The global ensembles suffer from blurring of local climate features associated with land-sea boundaries and with variations in topography, due to the coarse spatial scale of the underlying model outputs: the regional ensembles resolve this without introducing noticeable additional biases. **The regional ensembles are therefore recommended in preference to the global ones, for use in the assessment of future climate change impacts and adaptation strategies**.
- Overall, apart from the spatial blurring in the global model runs, the various ensembles do a reasonably good job of representing most of the climate indices considered for the evaluation period. However, the ensemble spread increases when considering future projections for some indices: in the absence of a full uncertainty analysis (to be delivered later in the project) **a prudent planning and adaptation strategy should consider information from across the**

range of the ensembles. This is particularly relevant for users who, due to time or other resource constraints, are unable to work with more than a small subset of ensemble members.

- To identify a small subset of members that capture a representative range of information from an ensemble, **users are advised to identify and focus on the dominant sources of variation, with respect to the specific climate indices that are most relevant to their application(s)**. For example, in a situation where the GCMs are responsible for 90% of the variation in the indices of interest, the chosen sample of ensemble members should contain runs driven by contrasting GCMs but the choice of RCM is relatively unimportant.
- Within an ensemble, both the dominant sources of variation and the models that emerge as 'contrasting' will usually depend on the quantities (indices) being studied. Therefore, **the selection of appropriate representative samples from an ensemble will be application-specific**. The project plot explorer at https://github-pages.ucl.ac.uk/UKCORDEX-plot-explorer/ is designed to enable rapid visualisation of the key structures of ensembles for a wide range of indices, and hence to inform the selection of appropriate subsets for use in applications.

1 Introduction

The UK Climate Projections (UKCP) provide several different products giving information about the future climate of the UK, including projections at global (60km), regional (12km) and local (2.2km) scales. While the regional projections are proving beneficial in understanding future climate risks (for example, Kennedy-Asser et al., 2021; Arnell et al., 2021), these were derived by perturbing the physical parameters of a single global climate model (GCM) and a single regional climate model (RCM), and so may not sample the full range of possible futures that is consistent with current scientific understanding. Project CR20-3 aims to complement the UKCP regional projections by combining them with information from the EuroCORDEX ensemble at the same 0.11° resolution. This combined set will allow a better understanding of structural uncertainty in UK climate projections (Murphy et al., 2019; Arnell et al., 2021) and substantially enhances the information base available to support future climate change risk assessments.

A major contribution of the project is to analyse and understand the EuroCORDEX outputs (both the GCM and the RCM components) over the UK region, in particular comparing them with the existing information from UKCP18. There are two aspects to this, the first being to examine the skill of the EuroCORDEX and UKCP18 models in capturing the current climate, and the second to characterise the variation and uncertainty in future projections.

Sections 4–6 of this report focus on the first aspect, evaluating the EuroCORDEX and UKCP18 regional components against gridded climate observations from the HadUK-Grid dataset. The evaluation is designed to provide information to users who want to understand the quality of the projections in terms of how well they reproduce various aspects of climatology that are represented by the indices identified in our earlier project report (Barnes et al., 2021). Detailed results are presented for the core variables of temperature, precipitation and surface wind speed, interpolated to 12km resolution over the UK land surface. Plots of the same analysis for the remaining UKCP18 core variables listed in Fung et al. (2018) are available online at `https://github-pages.ucl.ac.uk/UKCORDEX-plot-explorer/`, along with plots of selected derived indices of these variables. For these variables, the evaluation methodology, and descriptions of the plots used to present the results, are described in Section 3.

The regional model output is also compared to that of the driving GCMs and the UKCP18 global runs, in order to assess the performance of the regional models against that of the lower-resolution models; and to evaluation runs driven by ERA-Interim reanalysis, to examine the RCMs' performance when not inheriting errors from the driving GCMs. This evaluation is carried out over a twenty-year period starting on January 1st 1989 and finishing on December 31st 2008, chosen because this is the longest period for which reanalysis-driven evaluation runs are available (note that this differs from the 1980-2000 period used by Tucker et al. (2021)). From 1989-2005, the runs were driven by CMIP5 historical forcings; from 2006-2008, by the RCP8.5 scenario (Taylor et al., 2012).

In Section 7, the ensembles are further evaluated in terms of their ability to capture the physics of the wider climate system, as manifested in indices of large-scale activity that are related to the

climate of the UK. This is done partly by collating existing results from the literature, and partly by direct comparisons.

Sections 8–10 focus on the second aspect of the comparison, examining the changes in each model's climatology between a historical reference period (from December 1st 1980 to November 30th 2010) and future time periods. The analysis is focused primarily on the near future period (December 1st 2019 to November 30th 2049) and the future period (December 1st 2049 to November 30th 2079), although results are presented for other time periods where this is necessary to identify or highlight trends. Table 1 gives details of the time periods considered. This reference period is not the same as that used in the original UKCP18 analysis, which considered twenty-year periods (Murphy et al., 2019): instead, this report focuses on the thirty-year time-slices recommended by the World Meteorological Organisation (WMO, 2017), reflecting the potential interest in these results to a wider international audience.

Table 1: Definitions of time slices used in this report

Period	Start date	End date	Description
1989-2008	01-Jan-1989	31-Dec-2008	Evaluation period
1980-2010	01-Dec-1980	30-Nov-2010	Baseline period
1990-2020	01-Dec-1989	30-Nov-2019	
2000-2030	01-Dec-1999	30-Nov-2029	
2010-2040	01-Dec-2009	30-Nov-2039	
2020-2050	01-Dec-2019	30-Nov-2049	Near future
2030-2060	01-Dec-2029	30-Nov-2059	
2040-2070	01-Dec-2039	30-Nov-2069	
2050-2080	01-Dec-2049	30-Nov-2079	Future

Finally, Section 11 provides some conclusions and recommendations for users of future climate projections over the UK.

2 Datasets used in the report

The main focus of the project and of this report is a comparison of the EuroCORDEX and UKCP18 regional ensembles, but several other model ensembles are included in the analysis: brief descriptions of these ensembles are provided here.

CMIP5-13 ensemble: this consists of simulations from thirteen of the coupled ocean-atmosphere models run as part of the Coupled Model Intercomparison Project Phase 5 (CMIP5) experiment. This ensemble was constructed to supplement the 60km projections in the UKCP18 'Strand 2' global projections (Murphy et al., 2019), with the constituent models selected on the basis of performance and diversity criteria to in order to increase the range of global and regional changes sampled. The models are run at varying resolutions, ranging from 0.75° to 2.8° in latitude and from 0.75° to 2.8° in longitude (Taylor et al., 2012). The members of the CMIP5-13 ensemble are listed in Table 2.

CMIP5-EC ensemble: this contains the ten CMIP5 variants that were used to drive the EuroCORDEX 12km ensemble members. These ten driving realisations consist of six separate

models, two of which — EC-EARTH and MPI-M-ESM — each contribute three realisations. The models are run at varying resolutions, ranging from 1.1° to 1.9° in latitude and from 1.1° to 2.5° in longitude (Taylor et al., 2012). The models used to drive EuroCORDEX runs are listed in Table 2.

Table 2: CMIP5 global climate models appearing in either the CMIP5-EC or CMIP5-13 ensembles.

			Ensemble		Resolution	
Modelling group	**Model name**	**Run**	**CMIP5-13**	**CMIP5-EC**	**Lat**	**Lon**
CSIRO-BOM	ACCESS1-3	r1i1p1	■		1.25	1.88
BCC	BCC-CSM1-1	r1i1p1	■		2.8	2.81
NCAR	CCSM4	r1i1p1	■		0.94	1.25
NSF-DOE-NCAR	CESM1-BGC	r1i1p1	■		0.94	1.25
CMCC	CMCC-CM	r1i1p1	■		0.75	0.75
CNRM-CERFACS	CNRM-CM5	r1i1p1	■	■	1.4	1.41
CCMA	CanESM2	r1i1p1	■		2.8	1.81
ICHEC	EC-EARTH	r12i1p1		■	1.12	1.13
		r1i1p1	■	■	1.12	1.13
		r3i1p1		■	1.12	1.13
NOAA GFDL	GFDL-ESM2G	r1i1p1	■		2.02	2
MOHC	HadGEM2-ES	r1i1p1	■	■	1.25	1.88
IPSL	IPSL-CM5A-MR	r1i1p1	■	■	1.27	2.5
MPI-M	MPI-ESM-LR	r1i1p1		■	1.87	1.88
		r2i1p1		■	1.87	1.88
		r3i1p1		■	1.87	1.88
	MPI-ESM-MR	r1i1p1	■		1.87	1.88
MRI	MRI-CGCM3	r1i1p1	■		1.12	1.13
NCC	NorESM1-M	r1i1p1		■	1.89	2.5

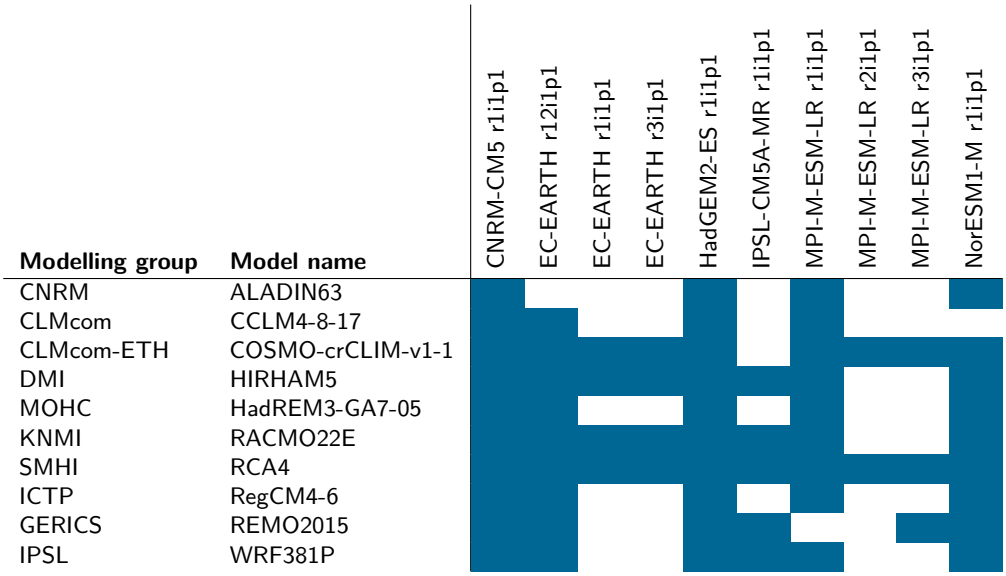
EuroCORDEX regional (12km) ensemble: this consists of 65 climate simulations produced by multiple dynamical downscaling models forced by multiple global climate models from the CMIP5 experiment. The simulations are run at a resolution of 0.11°, with the exact spatial extent varying according to the downscaling RCM: all cover an area from approximately 27-72° N, and from 22° W-45° E (Jacob et al., 2014).

At time of writing, the EuroCORDEX project has produced outputs at the required 0.11° resolution from six different GCMs (two of which provided three independent realisations each) and thirteen RCMs. However, runs have only been produced for a subset of the possible GCM-RCM pairs. Table 3 shows the combinations of GCMs and RCMs for which runs are currently available, and which are evaluated in this report. Each marked GCM-RCM pair contributes a single run to the 65-member EuroCORDEX ensemble.

Three available EuroCORDEX runs were omitted from the ensemble: a single run produced by ALADIN53, and two runs produced by REMO2009, which were driven by realisations of a single GCM. All of these runs lack some of the core variables, and realisations produced by later versions of both of these RCMs driven by the same GCMs also appear in the ensemble, so these earlier versions were omitted from the analysis.

ERA-EuroCORDEX evaluation runs: for each of the ten RCMs listed in Table 3 a single evaluation run, forced by ERA-Interim reanalysis rather than by a GCM, was produced. The evaluation periods for which these runs were produced differ between models, with only the

Table 3: The 65 GCM-RCM pairs included in the EuroCORDEX ensemble.



Modelling group	Model name	CNRM-CM5 r1i1p1	EC-EARTH r12i1p1	EC-EARTH r1i1p1	EC-EARTH r3i1p1	HadGEM2-ES r1i1p1	IPSL-CM5A-MR r1i1p1	MPI-M-ESM-LR r1i1p1	MPI-M-ESM-LR r2i1p1	MPI-M-ESM-LR r3i1p1	NorESM1-M r1i1p1
CNRM	ALADIN63	■				■		■			■
CLMcom	CCLM4-8-17	■	■			■		■			
CLMcom-ETH	COSMO-crCLIM-v1-1	■	■	■	■	■	■	■	■	■	■
DMI	HIRHAM5	■	■	■	■	■	■	■			■
MOHC	HadREM3-GA7-05	■				■		■			■
KNMI	RACMO22E	■	■	■	■	■	■	■			■
SMHI	RCA4	■	■	■	■	■	■	■	■	■	■
ICTP	RegCM4-6	■				■		■			■
GERICS	REMO2015	■	■			■	■			■	■
IPSL	WRF381P	■	■			■	■	■			■

period from December 1989 to November 2008 covered by all of the models.

UKCP18 global (60km) ensemble: this is a twelve-member perturbed physics ensemble (PPE), produced by using HadGEM3-GC3.05 as the driving model, with ensemble member 01 being run with the standard parameters, and each remaining ensemble member having slightly different perturbations to the parametrisation of the model physics (Murphy et al., 2019). The models have a spatial resolution of 0.55° latitude × 0.83° longitude, or approximately 60km over the UK.

UKCP18 regional (12km) ensemble: this ensemble downscales the twelve runs of the UKCP18 global ensemble using the HadREM3-GA7-05 regional model with the same parameter perturbations as those used to generate the 60km 'parent' run (Murphy et al., 2019). Ensemble member 01 was produced without perturbations; this is the same RCM that appears in the EuroCORDEX ensemble (Tucker et al., 2021).

HadUK-Grid observations: in Sections 4-6, surface weather quantities simulated by the model runs are evaluated against the HadUK-Grid dataset, a collection of gridded climate variables over the UK land surface derived from the network of UK land surface observations. Observations from irregularly spaced station data were interpolated to a 1km grid using inverse-distance weighting, then averaged to the 12km grid used in this report. For temperatures, observations from around 600 stations were used, and the interpolation was carried out over the residuals from a regression model relating the normalised temperatures to latitude, longitude, altitude, coastal influence and, for minimum temperatures only, density of urban land use; for precipitation, observations from between 2,500 and 3,500 stations were interpolated directly; and for surface wind speeds, a regression model relating the observations from 70 stations to latitude, longitude, altitude and coastal influence was fitted, and the residuals interpolated. Precipitation and daily maximum and minimum temperatures are available at daily resolution,

with daily mean temperatures reconstructed by taking the mean of the daily maximum and minimum; average surface wind speeds are available monthly (Perry and Hollis, 2004; Perry et al., 2009).

ERA-Interim reanalyses: in Section 7, indices of synoptic-scale weather phenomena are evaluated against indices calculated from ERA-Interim reanalyses at 0.7° resolution (Dee et al., 2011). ERA-Interim reanalyses are used in preference to the more recent ERA5 dataset to maintain consistency with the EuroCORDEX evaluation runs, which are driven by ERA-Interim.

2.1 Comparison of data at different spatial resolutions

The ensembles of models described above cover different spatial domains, and within each ensemble, different models produce data on different model grids, using different coordinate systems. To facilitate direct comparison of runs produced at different resolutions and on different axes, model output must be interpolated from the native grid onto a common grid. In this report, all data is presented on a 12km-square grid using OSGB coordinates; this is the same grid used to provide the 12km-resolution HadUK-Grid data, as well as the UKCP18 regional output over the UK.

Indices are first computed on each model's native grid, then interpolated to the 12km grid using a conservative area-weighting scheme. When regridding the EuroCORDEX and UKCP18 12km outputs, a land-sea mask is applied to both the source and target grids, so that there is no blurring between the land and sea boundaries. When regridding the lower-resolution CMIP5 datasets, this approach is not used: partly because land-sea masks are not available for all of the models; and partly because removing cells flagged as belonging to the sea surface before regridding the data would result in an absence of data in large areas of the UK. Instead, the low-resolution data is regridded directly onto the land surface 12km grid, and the effect of any blurring of land and sea surface variables is highlighted when comparing the ensembles later in the report. The same approach is applied to the UKCP18 60km model output.

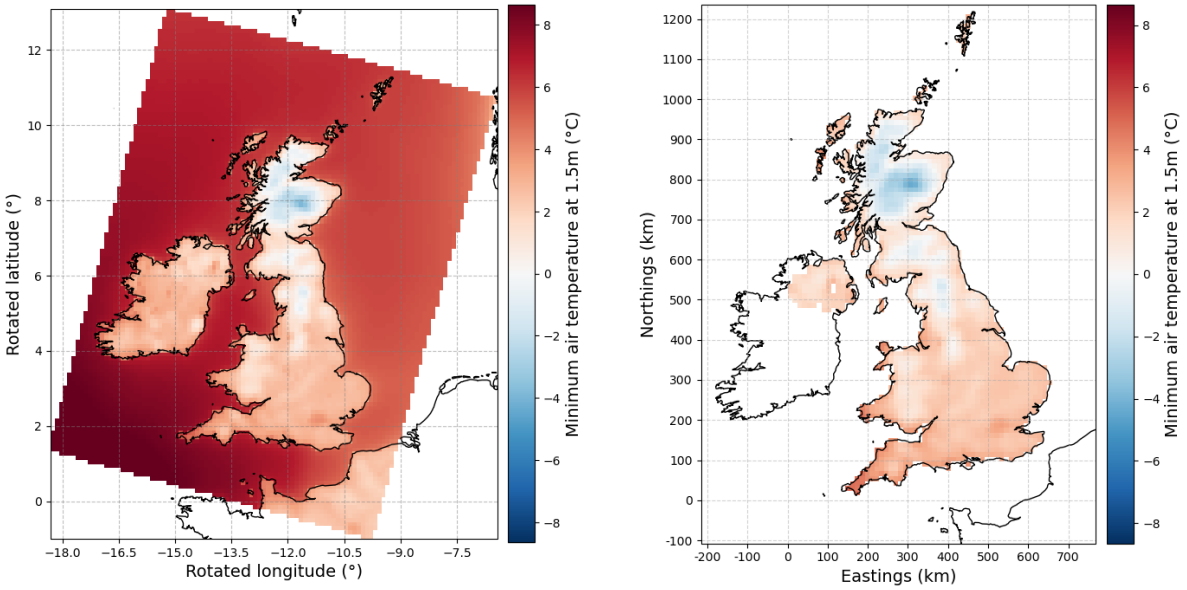
3 Evaluation methodology

The results in this report are based on maps of seasonal climatologies derived from daily data: these are computed by averaging (or, where specified, computing some other function of) the daily model output over the winter (December, January, February) or summer (June, July, August) months for the time period of interest. The one exception to this is the surface wind climatology computed from HadUK-Grid, which is computed from monthly averages, because daily data are not available. In Sections 4-6, biases are calculated by subtracting the HadUK-Grid climatology from the model climatology during the evaluation period from 1989-2008; in Sections 8-10, changes in each run's climatology are computed by subtracting the climatology for the reference period of 1980-2010 from the period of interest. Several graphical representations of this information are used to summarise different aspects of changes within the various ensembles: descriptions of the data used to produce

Figure 1: Maps of model output showing the area covered by the OSGB domain: (a) on the native rotated-pole grid and (b) after masking the land surface and interpolating to the 12km OSGB grid used in this report. The maps here show the average of minimum daily temperatures simulated by a single EuroCORDEX run during the winter months from 1980-2010.

(a) Area of EuroCORDEX native grid covered by OSGB domain

(b) UK land surface on OSGB grid



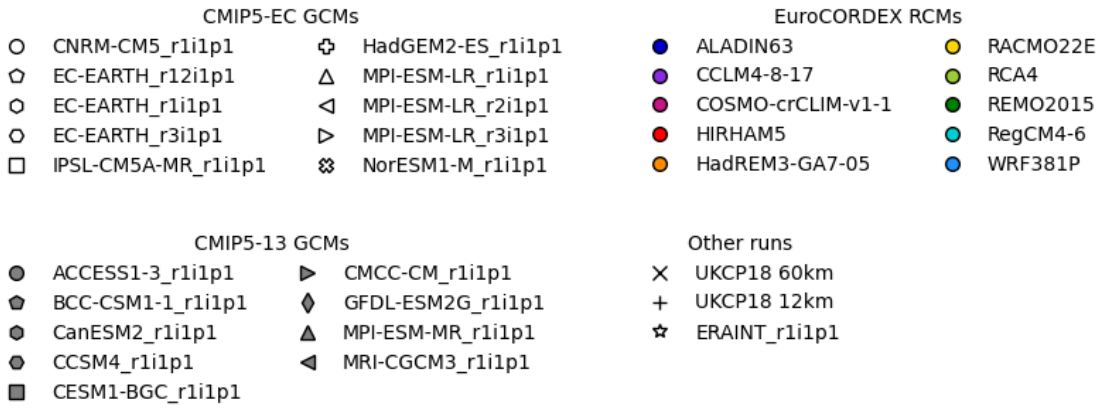
each plot, and where necessary some information about the interpretation of the plots, are provided here. For clarity, descriptions are given only with reference to plots of biases during the evaluation period, but apply equally to the plots of changes in later sections.

The same legend (shown in Figure 2) is used throughout this report. Members of the CMIP5-EC ensemble are represented by symbols shaded in white, with members of the CMIP5-13 ensemble filled in grey; where a model appears in both ensembles, a white symbol is used. Some symbols are used in both ensembles, but this does not necessarily connote any relationship between the corresponding runs. Members of the EuroCORDEX ensemble are represented by coloured symbols, with the shape indicating the member of the CMIP5-EC ensemble used to force the run, and the colour indicating the RCM. UKCP18 ensemble members are represented by dots in boxplots; by crosses (60km runs) or pluses (12km runs) in Taylor diagrams; and by ensemble member number in EPP score plots. In all cases the unperturbed ensemble member is highlighted in orange.

3.1 Maps of ensemble mean biases (changes)

Systematic differences between the ensembles — both in spatial patterns common to all ensemble members and in the magnitude of biases — are identified using maps of the mean biases within each ensemble, after regridding to the 12km OSGB grid. During the evaluation period, the mean bias in each ensemble is presented as a map alongside the HadUK-Grid climatology as in Figure 3, while for future periods, maps of the mean change in each ensemble are provided, as in Figure 63).

Figure 2: Model legend used in all plots



3.2 Boxplots of average values of indices simulated by each run

The distribution of the average values of an index (or, when considering future periods, of the average changes in an index) across the UK are summarised in boxplots like that shown in Figure 4a. The boxes indicate the central 50% of the distribution of UK-averaged values of the index within each ensemble, with both the mean and median values marked; the whiskers of the boxplot extend to values lying 1.5 times the interquartile range beyond the upper and lower quartiles, with any values lying beyond the whiskers assumed to be outliers. The shaded region behind each boxplot shows a kernel density estimate of the empirical distribution of the values, which provides more detail about the shape of the distribution: for example, by highlighting bimodality within an ensemble. Each point represents the mean value of the index across the UK land surface for a single run, with points corresponding to the output of a single GCM jittered horizontally for ease of viewing. For plots of biases, the UK-wide average of the HadUK-Grid climatology is shown by a dashed red line; for plots of changes, the dashed red line indicates no change.

3.3 Plots of spread of biases in each run

Sections 4.2 and 4.5 examine biases in the percentiles of the distributions of daily temperatures simulated by each run, in order to evaluate how well the models capture the natural variability of the climate. Biases in the mean temperature typically affect the whole temperature distribution, so boxplots of biases in each index tend to look rather similar, and it can be difficult to identify patterns in the behaviour of individual runs or models. Instead, it is convenient to visualise the information from all indices relating to a particular temperature distribution simultaneously, as in the spread plots in Figure 10. In these plots, the distance between the first and 99th percentiles of the temperature distribution of each run is represented by a vertical line, with the mean temperature marked by a symbol denoting the run, with each ensemble appearing in a separate panel. The corresponding values from the HadUK-Grid dataset are marked with horizontal red lines across all panels. The width of the interval after subtracting the bias in the mean temperature is marked with red dashes,

so that the width of the interval between the two percentiles can be compared directly for all runs.

The same approach is used when evaluating the biases in the diurnal temperature range simulated by each run. Patterns of changes in the percentiles of the temperature distribution, and of changes in the diurnal temperature range, are less closely related to changes in the mean temperature, and so this approach is not used when considering changes in the distribution.

3.4 Correlation of spatial patterns in modelled and observed climatology (Taylor diagrams)

Taylor diagrams like that in Figure 4b are used to assess which runs produce the most realistic spatial variation of climate during the evaluation period. For each run, two statistics are computed: the Pearson correlation coefficient between the simulated climatology and the 'observed' HadUK-Grid climatology, which measures the similarity between the patterns in the maps of simulated and observed climate; and the standard deviation of the simulated climatology, which measures the amplitude of the pattern. These two statistics are plotted on radial axes, with the correlation coefficient related to the azimuthal angle, and the standard deviation proportional to the radial distance from the origin (Taylor, 2001). Where only positive correlations are observed, the plot will cover a quarter circle; if negative correlations are observed, the plot will be semicircular, as in Figure 55.

The standard deviation of the HadUK-Grid climatology is marked with a heavy dashed arc: points lying close to this arc represent runs with climatologies that have a similar amplitude to the HadUK-Grid climatology, while points lying closer to the origin represent runs that underestimate the spatial variability, and points lying outside the arc represent runs that overestimate the spatial variability. Points close to the x-axis represent runs with high correlation coefficients, indicating that those runs produce similar spatial patterns to those in HadUK-Grid; points that lie closer to the y-axis represent runs with correlation coefficients close to zero, indicating very different spatial patterns to those in HadUK-Grid.

Ideally, points should lie as close as possible to the black star on the x-axis that represents the HadUK-Grid climatology against which the runs are being evaluated: runs represented by points closer to the black star more closely replicate the HadUK-Grid climatology in terms of both the location of spatial features and the range of values simulated across the UK. The radial axes are scaled in such a way that the distance from each point to the black star is proportional to the centred root mean squared error in the simulated field (Taylor, 2001); however, note that this centred root mean squared error excludes UK-wide biases in the runs, which must be assessed separately using boxplots like those described in Section 3.2.

3.5 Sources of variation in the ensemble

The analyses described above are designed to provide information about variation within an ensemble, together with some characteristics of the individual ensemble members. Moreover, the legend in Figure 2 enables some key features to be identified from those analyses, in terms of how the GCMs

and RCMs contribute the overall variation. For example, in Figure 4a the points in the EuroCORDEX ensemble boxplot are grouped by symbol rather than colour: this suggests that in this particular case, the GCMs tend to produce results that are distinct from each other while the RCMs do not contribute much additional variation. By contrast, in Figure 49b the corresponding points are more obviously grouped by colour suggesting that the RCMs are more important in this case. In this section we develop methodology for quantifying more formally the relative contributions of the GCMs and RCMs to the ensemble structure. One purpose is to guide users who wish to select a representative sample of runs: if, for example, their application involves indices for which the RCMs contribute little to the overall ensemble variation, then the user's precise choice of RCMs will be relatively unimportant.

As a first step in a formal analysis, analysis of variance (ANOVA) can be used to partition the ensemble variation into contributions that can be attributed to each potential source. ANOVA approaches are being used increasingly to study the structure of climate model ensembles, for example to understand the extent to which differences between future projections are attributable to the use of different GCMs or different emissions scenarios (Yip et al., 2011) or, as here, to quantify the respective effects of GCMs and RCMs in an ensemble of RCM simulations. This has been done by Christensen and Kjellström (2020) for a subset of the EuroCORDEX ensemble, separately for each grid cell within the full EuroCORDEX domain described in Section 2. We here develop some extensions of the methodology used by Christensen and Kjellström (2020), allowing us both to analyse the complete ensemble and to partition the joint variation across an entire spatial domain (rather than considering grid cells individually).

To introduce the methodology, it is helpful first to consider a situation in which the ensemble is "balanced" in the sense that it contains a run for every RCM:GCM combination. In this case, let R and G denote the numbers of RCMs and GCMs respectively; and let $\boldsymbol{Y}_{rg}$ be an $S \times 1$ vector containing quantities of interest derived from the run of RCM r driven by GCM g, at each of S spatial locations (for example, $\boldsymbol{Y}_{rg}$ might contain the biases or future changes in some climate index, at each node of the OSGB grid). Moreover, let $\overline{\boldsymbol{Y}}_{r\cdot} = G^{-1} \sum_{g=1}^{G} \boldsymbol{Y}_{rg}$ and $\overline{\boldsymbol{Y}}_{\cdot g} = R^{-1} \sum_{r=1}^{R} \boldsymbol{Y}_{rg}$ denote the means of all available runs for the rth RCM and gth GCM respectively; and let $\overline{\boldsymbol{Y}}_{\cdot\cdot} = (RG)^{-1} \sum_{r=1}^{R} \sum_{g=1}^{G} \boldsymbol{Y}_{rg}$ denote the overall ensemble mean field. Then we can write

$$\boldsymbol{Y}_{rg} = \overline{\boldsymbol{Y}}_{\cdot\cdot} + \left(\overline{\boldsymbol{Y}}_{\cdot g} - \overline{\boldsymbol{Y}}_{\cdot\cdot}\right) + \left(\overline{\boldsymbol{Y}}_{r\cdot} - \overline{\boldsymbol{Y}}_{\cdot\cdot}\right) + \left(\overline{\boldsymbol{Y}}_{rg} - \overline{\boldsymbol{Y}}_{\cdot g} - \overline{\boldsymbol{Y}}_{r\cdot} + \overline{\boldsymbol{Y}}_{\cdot\cdot}\right) \ . \tag{1}$$

Consider now the sum of squares and cross-products (SSCP)

$$\sum_{r=1}^{R} \sum_{g=1}^{G} \left(\boldsymbol{Y}_{rg} - \overline{\boldsymbol{Y}}_{\cdot\cdot}\right) \left(\boldsymbol{Y}_{rg} - \overline{\boldsymbol{Y}}_{\cdot\cdot}\right)' = \boldsymbol{T}, \text{ say,} \tag{2}$$

which is a matrix of dimension $S \times S$ that plays the same role as the total sum of squares in a standard univariate ANOVA (Krzanowski, 1988, Section 13.3). To aid interpretation, it may be helpful to note that if the individual runs $\{\boldsymbol{Y}_{rg} : r = 1, \ldots, R; g = 1, \ldots, G\}$ were stored in a $(RG \times S)$ data

matrix with a row for each run and a column for each spatial location, then the sample covariance matrix between all pairs of locations would be $\boldsymbol{T}/(RG-1)$: the diagonal elements of this covariance matrix are the sample variances (across the ensemble) of the values at the individual locations, and the off-diagonal elements are the covariances between the values at each pair of locations.

From (1) and (2), a little algebra shows that the SSCP can be decomposed as

$$\boldsymbol{T} = \boldsymbol{T}_G + \boldsymbol{T}_R + \boldsymbol{T}_E \text{ , say,} \tag{3}$$

where $\boldsymbol{T}_G = R\sum_{g=1}^{G}\left(\overline{\boldsymbol{Y}}_{\cdot g} - \overline{\boldsymbol{Y}}_{\cdot\cdot}\right)\left(\overline{\boldsymbol{Y}}_{\cdot g} - \overline{\boldsymbol{Y}}_{\cdot\cdot}\right)'$ and $\boldsymbol{T}_R = G\sum_{r=1}^{R}\left(\overline{\boldsymbol{Y}}_{r\cdot} - \overline{\boldsymbol{Y}}_{\cdot\cdot}\right)\left(\overline{\boldsymbol{Y}}_{r\cdot} - \overline{\boldsymbol{Y}}_{\cdot\cdot}\right)'$ are contributions from the GCM-specific means $\{\overline{\boldsymbol{Y}}_{\cdot g}\}$ and RCM-specific means $\{\overline{\boldsymbol{Y}}_{r\cdot}\}$ respectively; and $\boldsymbol{T}_E = \sum_{r=1}^{R}\sum_{g=1}^{G}\left(\boldsymbol{Y}_{rg} - \overline{\boldsymbol{Y}}_{\cdot g} - \overline{\boldsymbol{Y}}_{r\cdot} + \overline{\boldsymbol{Y}}_{\cdot\cdot}\right)\left(\boldsymbol{Y}_{rg} - \overline{\boldsymbol{Y}}_{\cdot g} - \overline{\boldsymbol{Y}}_{r\cdot} + \overline{\boldsymbol{Y}}_{\cdot\cdot}\right)'$ represents unstructured 'residual' variation. The relative magnitudes of these three terms can therefore be used to summarise the relative contributions of the GCMs, RCMs and the unstructured component to the overall ensemble variation.

Unfortunately, if the number of locations S exceeds 1, there is no unique way to compare the magnitudes of $\boldsymbol{T}_G$, $\boldsymbol{T}_R$ and $\boldsymbol{T}_E$ (if $S = 1$, then each of these matrices is a single number and can be expressed as a proportion of the total variation $\boldsymbol{T}$). Two simple options are as follows:

- Produce maps showing the diagonal elements of $\boldsymbol{T}$, and the proportional contribution of $\boldsymbol{T}_G$, $\boldsymbol{T}_R$ and $\boldsymbol{T}_E$ to each of these diagonal elements. These maps (e.g. Figure 5) are directly equivalent to the "total variability partition" maps in, for example, Figure 5 of Christensen and Kjellström (2020).
- Calculate the traces of the matrices (i.e. the sums of the diagonal elements): the trace operator is additive, so that $\text{trace}(\boldsymbol{T}) = \text{trace}(\boldsymbol{T}_G) + \text{trace}(\boldsymbol{T}_R) + \text{trace}(\boldsymbol{T}_E)$ and the total trace can be unambiguously partitioned into single-number summaries representing the contributions from each source of variation.

In Section 3.5.3 we will consider an alternative to these approaches, that accounts for the full structure of the matrices $\boldsymbol{T}_G$, $\boldsymbol{T}_r$ and $\boldsymbol{T}_E$ instead of just using their diagonal elements.

3.5.1 A model-based formulation

The decomposition (3) is purely arithmetical, requiring no assumptions except that the ensemble is balanced. To go beyond this however, it is helpful to note the connection with a statistical model in which the individual ensemble runs are represented as

$$\boldsymbol{Y}_{rg} = \boldsymbol{\mu} + \boldsymbol{\alpha}_g + \boldsymbol{\beta}_r + \boldsymbol{\varepsilon}_{rg} \qquad (g = 1, \ldots, G; r = 1, \ldots, R) \text{ ,} \tag{4}$$

where the collections of coefficient vectors $\{\boldsymbol{\alpha}_g\}$ and $\{\boldsymbol{\beta}_r\}$ are each constrained to sum to zero: $\sum_{g=1}^{G}\boldsymbol{\alpha}_g = \sum_{r=1}^{R}\boldsymbol{\beta}_r = \boldsymbol{0}$. In this case, $\boldsymbol{\mu}$ represents an overall mean; $\boldsymbol{\alpha}_g$ and $\boldsymbol{\beta}_r$ represent the respective extents to which the gth GCM and rth RCM deviate systematically from that overall mean; and $\boldsymbol{\varepsilon}_{rg}$ is a random 'error' representing unstructured variation, treated as though it is drawn

independently for each run from a multivariate normal (Gaussian) distribution with mean vector $\mathbf{0}$ and some covariance matrix, $\boldsymbol{\Sigma}$ say (we write $\boldsymbol{\varepsilon}_{rg} \sim MVN(\mathbf{0}, \boldsymbol{\Sigma})$). All of these quantities are vectors of length S.

Expression (4) can be considered as the most complex of a collection of models:

Model 0: $\boldsymbol{Y}_{rg} = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_{rg}$. Here and throughout, $\boldsymbol{\varepsilon}_{rg} \sim MVN(\mathbf{0}, \boldsymbol{\Sigma})$. This model implies an absence of any systematic variations associated with the GCMs or RCMs in the ensemble.

Model 1a: $\boldsymbol{Y}_{rg} = \boldsymbol{\mu} + \boldsymbol{\alpha}_g + \boldsymbol{\varepsilon}_{rg}$. Under this model, there are systematic variations associated with the GCMs, but not with the RCMs.

Model 1b: $\boldsymbol{Y}_{rg} = \boldsymbol{\mu} + \boldsymbol{\beta}_r + \boldsymbol{\varepsilon}_{rg}$. Here there are systematic variations associated with the RCMs, but not with the GCMs.

Model 2: this is the full model (4), allowing for systematic variations associated with both RCMs and GCMs.

In each of these models, the least-squares estimate of $\boldsymbol{\mu}$ is $\overline{\boldsymbol{Y}}_{\cdot\cdot}$ (this follows from standard derivations as given in, for example, Davison 2003, Section 9.2.2). Moreover, in models 1a and 2 the estimate of the GCM effect $\boldsymbol{\alpha}_g$ is $\overline{\boldsymbol{Y}}_{\cdot g} - \overline{\boldsymbol{Y}}_{\cdot\cdot}$; and in models 1b and 2 the estimate of the RCM effect $\boldsymbol{\beta}_r$ is $\overline{\boldsymbol{Y}}_{r\cdot} - \overline{\boldsymbol{Y}}_{\cdot\cdot}$. These estimates can be mapped individually for each RCM and GCM, to visualise their estimated effects within the ensemble.

The estimates for a model can also be substituted into the model equation, omitting the error term $\boldsymbol{\varepsilon}_{rg}$, to obtain *fitted values* which are denoted generically by $\{\hat{\boldsymbol{Y}}_{rg}\}$. The differences $\{\boldsymbol{Y}_{rg} - \hat{\boldsymbol{Y}}_{rg}\}$ are *residual vectors*, representing variation within the ensemble that cannot be related to (or 'explained by') the covariates in the model. For example, the $\{\boldsymbol{Y}_{rg} - \overline{\boldsymbol{Y}}_{\cdot\cdot}\}$ are the residual vectors for Model 0 (there are no covariates in this model, so these residual vectors encompass all of the variation in the ensemble); the $\{\boldsymbol{Y}_{rg} - \overline{\boldsymbol{Y}}_{\cdot g}\}$ are those from Model 1a, representing variation that cannot be explained by differences between the GCMs; and so on. Denoting the (r, g) residual vectors for the respective models by $\boldsymbol{e}_{rg}^{(0)}$, $\boldsymbol{e}_{rg}^{(1a)}$, $\boldsymbol{e}_{rg}^{(1b)}$ and $\boldsymbol{e}_{rg}^{(2)}$, equation (2) can be rewritten as $\boldsymbol{T} = \sum_{r=1}^{R}\sum_{g=1}^{G} \boldsymbol{e}_{rg}^{(0)} \left[\boldsymbol{e}_{rg}^{(0)}\right]'$. The expression for $\boldsymbol{T}_E$ following equation (3) is also equal to a residual sum of squares and cross-products (RSSCP): $\boldsymbol{T}_E = \sum_{r=1}^{R}\sum_{g=1}^{G} \boldsymbol{e}_{rg}^{(2)} \left[\boldsymbol{e}_{rg}^{(2)}\right]'$. It is, moreover, straightforward to show that $\boldsymbol{T}_G = \boldsymbol{T} - \sum_{r=1}^{R}\sum_{g=1}^{G} \boldsymbol{e}_{rg}^{(1a)} \left[\boldsymbol{e}_{rg}^{(1a)}\right]'$ and that $\boldsymbol{T}_R = \boldsymbol{T} - \sum_{r=1}^{R}\sum_{g=1}^{G} \boldsymbol{e}_{rg}^{(1b)} \left[\boldsymbol{e}_{rg}^{(1b)}\right]'$.

Now consider fitting Model 0 followed by Model 1a and then Model 2. This sequence of model fits produces a corresponding sequence of RSSCPs: $\boldsymbol{T}$, $\boldsymbol{T} - \boldsymbol{T}_G$, $\boldsymbol{T}_E = \boldsymbol{T} - \boldsymbol{T}_G - \boldsymbol{T}_R$. Thus, according to this fitted model sequence, $\boldsymbol{T}_G$ is the 'reduction' (in quotes because it is a matrix-valued quantity) in RSSCP obtained by using Model 1a in place of Model 0: this reduction represents a portion of the total variation $\boldsymbol{T}$ that is attributable to the GCMs. Moving from Model 1a to Model 2, $\boldsymbol{T}_R$ represents the additional reduction in RSSCP that is attributable to the RCMs after accounting for the GCM effects. The same partitioning of the overall RSSCP is obtained by fitting model 0 followed by model 1b and then model 2. Under this model-based approach therefore, the algebraic decomposition (3) can be interpreted as characterising the extent to which the total ensemble

variation can be 'explained' by the combined effects of the GCMs and RCMs.

3.5.2 Handling unbalanced ensembles

The development above is all for a balanced ensemble: the balance condition (i.e. that the ensemble contains the same number of runs for every RCM:GCM combination) underpins the unique decomposition (3) of the total explained variation into contributions from the GCMs and RCMs, as well as the simple expressions for the parameter estimates in model (4). Unfortunately however, the EuroCORDEX ensemble is not balanced (see Table 3). Informally, this makes it harder to disentangle the effects of the individual RCMs and GCMs: if, say, a particular RCM has only been run in conjunction with a subset of the GCMs, then systematic differences between the runs of this RCM and the remainder of the ensemble could be due either to the characteristics of the RCM itself, or to the particular subset of GCMs that have been used to drive it. For this reason, some published analyses of climate model ensembles have chosen to work only with the largest available balanced subset of runs (e.g. Yip et al. 2011); others have tried to estimate the values for runs that are missing (e.g. Christensen and Kjellström 2020 who replaced the outputs from a single missing combination with their expected values; and Evin et al. 2019 who used Bayesian methods to try and account for the associated uncertainty), while others still have opted to escape from the constraints of a balanced design within a more general random-effects formulation (e.g. Northrop and Chandler 2014, again in a Bayesian framework). Bayesian approaches are computationally intensive however, and hence not well suited for the rapid exploration of large numbers of indices. Moreover, the present analysis aims primarily to characterise the structure of the EuroCORDEX ensemble: adding more than a very small number of imputed values for runs that are not present would be inconsistent with this aim.

Fortunately, under a single emissions scenario the EuroCORDEX regional ensemble has a relatively simple structure because there are only two possible sources of systematic variation (the RCMs and GCMs). It is therefore feasible to conduct two separate analyses: in the first, the maximum possible portion of variation is attributed to the GCMs, with the RCM information used only to try and account for variation that cannot otherwise be explained. In the second analysis, the roles of GCMs and RCMs are reversed. The difference between the two analyses indicates the extent to which the partitioning is affected by the lack of balance. In the results reported below therefore, we carry out both analyses and plot the resulting ranges of attributable variation as, for example, in Figure 11.

The analyses are most easily described by considering the RSSCPs obtained from different sequences of statistical models. Using the same residual notation as in Section 3.5.1, the RSSCPs from Models 0 and 1a are $\boldsymbol{T} = \sum_{r=1}^{R}\sum_{g=1}^{G}\chi_{rg}\boldsymbol{e}_{rg}^{(0)}\left[\boldsymbol{e}_{rg}^{(0)}\right]'$ and $\sum_{r=1}^{R}\sum_{g=1}^{G}\chi_{rg}\boldsymbol{e}_{rg}^{(1a)}\left[\boldsymbol{e}_{rg}^{(1a)}\right]'$ respectively, where χ_{rg} is an indicator variable taking the value 1 if the (r, g) coupling appears in the ensemble and zero otherwise. The RSSCP from Model 1a is denoted as $\boldsymbol{T} - \boldsymbol{T}_G^{(a)}$, so that $\boldsymbol{T}_G^{(a)}$ represents the reduction in RSSCP explained by the GCMs under this model. Similarly, the RSSCP from Model 2 is $\boldsymbol{T}_E = \sum_{r=1}^{R}\sum_{g=1}^{G}\chi_{rg}\boldsymbol{e}_{rg}^{(2)}\left[\boldsymbol{e}_{rg}^{(2)}\right]'$, and the difference $\boldsymbol{T} - \boldsymbol{T}_G^{(a)} - \boldsymbol{T}_E$ $(= \boldsymbol{T}_R^{(a)}$, say) is the additional reduction explained by the RCMs. This sequence of models therefore partitions the total variation as $\boldsymbol{T} = \boldsymbol{T}_G^{(a)} + \boldsymbol{T}_R^{(a)} + \boldsymbol{T}_E$. A similar partitioning $\boldsymbol{T} = \boldsymbol{T}_R^{(b)} + \boldsymbol{T}_G^{(b)} + \boldsymbol{T}_E$ is obtained from the sequence

Model 0 $\rightarrow$ Model 1b $\rightarrow$ Model 2.

In these two partitionings obtained from alternative model sequences, the quantities $\boldsymbol{T}$ and $\boldsymbol{T}_E$ (representing total and unstructured variation) are the same. The variation attributed to the GCMs and RCMs may differ however, because in general $\boldsymbol{T}_G^{(a)} \neq \boldsymbol{T}_G^{(b)}$ and $\boldsymbol{T}_R^{(a)} \neq \boldsymbol{T}_R^{(b)}$. In the first sequence ($0 \rightarrow 1\text{a} \rightarrow 2$), the GCM effects are considered first so that $\boldsymbol{T}_G^{(a)}$ represents their maximum possible contribution to the total variation. Since $\boldsymbol{T}_R^{(a)} = \boldsymbol{T} - \boldsymbol{T}_G^{(a)} - \boldsymbol{T}_E$, and $\boldsymbol{T}$ and $\boldsymbol{T}_E$ are the same for both sequences, it follows that $\boldsymbol{T}_R^{(a)}$ represents the minimum possible contribution from the RCMs: this is RCM-related variation that cannot otherwise be explained using the GCMs. Conversely, for the second sequence $0 \rightarrow 1\text{b} \rightarrow 2$ the RCM effects are considered first: $\boldsymbol{T}_R(b)$ and $\boldsymbol{T}_G^{(b)}$ therefore represent respectively the maximum possible contribution from the RCMs and the minimum possible contribution from the GCMs. The proportion of the total trace associated with the GCMs therefore lies between trace $\left(\boldsymbol{T}_G^{(b)}\right)$ /trace($\boldsymbol{T}$) and trace $\left(\boldsymbol{T}_G^{(a)}\right)$ /trace($\boldsymbol{T}$); and the proportion associated with the choice of RCM lies trace $\left(\boldsymbol{T}_R^{(a)}\right)$ /trace($\boldsymbol{T}$) and trace $\left(\boldsymbol{T}_R^{(b)}\right)$ /trace($\boldsymbol{T}$). Similar ranges apply for the scaled deviance contributions.

Although this process is a natural generalisation of that for a balanced ensemble, the computational aspects — and the fitting of Model 2 in particular — are less straightforward. Details are provided in Appendix A.

A further implication of the non-unique decomposition of variation is that there is no unique analogue of the total variability partition maps of Christensen and Kjellström (2020) for an unbalanced ensemble. There are various possible approaches to this. One is to map the proportional contributions of either or all of $\boldsymbol{T}_G^{(a)}$, $\boldsymbol{T}_G^{(b)}$, $\boldsymbol{T}_R^{(a)}$ and $\boldsymbol{T}_R^{(b)}$ to the diagonal elements of $\boldsymbol{T}$. An alternative exploits the connection between $\boldsymbol{T}_G$ and $\boldsymbol{T}_R$ and the Model 2 coefficient estimates $\{\hat{\boldsymbol{\alpha}}_g\}$ and $\{\hat{\boldsymbol{\beta}}_r\}$: for a balanced ensemble, $\boldsymbol{T}_G = R\sum_{g=1}^{G} \hat{\boldsymbol{\alpha}}_g \hat{\boldsymbol{\alpha}}_g'$ and $\boldsymbol{T}_R = G\sum_{g=1}^{G} \hat{\boldsymbol{\alpha}}_g \hat{\boldsymbol{\alpha}}_g'$, suggesting that the analogous quantities could be computed directly from the parameter estimates in an unbalanced ensemble. These analogous quantities are $\boldsymbol{T}_G = \sum_{g=1}^{G} n_{\cdot g} \hat{\boldsymbol{\alpha}}_g \hat{\boldsymbol{\alpha}}_g'$ and $\boldsymbol{T}_R = \sum_{g=1}^{G} n_{r\cdot} \hat{\boldsymbol{\alpha}}_g \hat{\boldsymbol{\alpha}}_g'$, where $n_{\cdot g}$ and $n_{r\cdot}$ are respectively the numbers of runs of GCM g and RCM r in the ensemble. For mapping purposes, this approach has the advantages that the contributions from the GCMs, RCMs and residuals can be shown in a single set of maps instead of having to consider both the '(a)' and '(b)' variants; and that these maps (such as Figure 5) are based on the optimal coefficient estimates that make full use of the available data. Moreover, the approach is directly connected with the analysis of spatial structure considered in the next section, in the sense that the spatial analysis provides a further exact decomposition of $\boldsymbol{T}_G$ and $\boldsymbol{T}_R$ if they are defined in this way. A slight disadvantage of the approach is that the resulting mapped percentages of total variation do not sum exactly to 100%, because the RCM and GCM contributions are not independent in an unbalanced ensemble. However, experimentation suggests that for the EuroCORDEX ensemble used here, the sum of the contributions is rarely far from 100% (e.g. in the decompositions of ensemble biases for the period 1989–2005, the sum is between 95% and 105% in all cases); and also that the overall appearance of such maps is insensitive to the precise details of their construction.

3.5.3 Decomposition of joint variation: MANOVA and analysis of deviance

To summarise the results of the ANOVA described above, we have proposed using maps and summary measures based on the diagonal elements of the matrices $\boldsymbol{T}_G$, $\boldsymbol{T}_R$ and $\boldsymbol{T}_E$: this is equivalent to the cell-by-cell analysis of Christensen and Kjellström (2020). Such an analysis, however, has the disadvantage of failing to account for potential correlations between neighbouring locations: for example, it cannot reveal anything about the spatial extent of 'typical' differences between simulations. To account for this in the present work therefore, a multivariate ANOVA (MANOVA) approach is adopted: this can be considered as a first step towards the functional ANOVA methodology of Sain et al. (2011), although it avoids the computational complexities of that methodology and hence is perhaps better suited to routine application for exploratory visualisation.

The key to the MANOVA approach is the model-based perspective of Section 3.5.1, which suggests an alternative partitioning of the ensemble variation by considering the close connection between variation and likelihood. Noting that the least-squares parameter estimates from each of Models 0, 1a, 1b and 2 are also the maximum likelihood estimates (MLEs) (Krzanowski, 1988, Section 15.2), the fitted values for each model can be used to compare its maximised log-likelihood with the highest log-likelihood attainable from the ensemble (i.e. the log-likelihood from a statistical model that fits the ensemble outputs perfectly). The *scaled deviance* for a model is defined as twice the difference between the achieved and highest attainable log-likelihoods: in linear regression, this quantity is proportional to the residual sum of squares (RSS) (Davison, 2003, Section 10.2), and it can be interpreted in the same way as the RSS in other models. It can be shown (see Appendix B) that the scaled deviance for Model 0 can be partitioned into contributions from the GCMs, RCMs and residuals in exactly the same way as the total trace, and that the proportions of scaled deviance attributable to the GCMs and RCMs are trace $\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{T}_G\right]$ /trace $\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{T}\right]$ and trace $\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{T}_R\right]$ /trace $\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{T}\right]$ respectively. This deviance-based partitioning accounts for the spatial correlation structure of the ensemble members via the $\boldsymbol{\Sigma}$ matrix: any differences compared with the partitioning based on trace$(\boldsymbol{T}_G)$/trace$(\boldsymbol{T})$, trace$(\boldsymbol{T}_R)$/trace$(\boldsymbol{T})$ and trace$(\boldsymbol{T}_E)$/trace$(\boldsymbol{T})$ must be associated with this correlation structure, therefore. The calculations require some care however, to avoid excessive computational storage requirements when working with matrices of dimension $S \times S$; and also to deal with the fact that $\boldsymbol{\Sigma}$ must be estimated in practice and that its estimate will usually be singular if the number of locations S is large. These issues are tackled in Appendix B.1.1.

3.6 Spatial structure of ensemble variation

A final analysis of ensemble structure seeks to characterise the nature of the contributions from each GCM and each RCM. To some extent, this can be achieved by inspecting maps of the estimated coefficient vectors $\{\boldsymbol{\alpha}_g\}$ and $\{\boldsymbol{\beta}_g\}$ in Model (4): however, it is not always easy to draw clear conclusions from visual comparisons of large numbers of maps. As an alternative, we aim to find a small number of dominant 'modes of variation' within the estimated effects, that can be interpreted analogously to principal components (Krzanowski, 1988) or empirical orthogonal functions (von Storch and Zwiers, 1999) in conventional analyses of space-time climate fields.

In fact, for a balanced ensemble these dominant modes of variation can be extracted directly from the Singular Value Decompositions (SVDs) used for the deviance partitioning in Appendix B.1.1. For an unbalanced ensemble, additional SVDs are needed although these can be computed quickly. We discuss the general procedure in relation to the GCM effects $\{\boldsymbol{\alpha}_g\}$: the RCM effects $\{\boldsymbol{\beta}_g\}$ can be treated similarly. Further details of the underpinning mathematics can be found in Krzanowski (1988, Section 4.1).

Consider stacking the scaled estimates $\{\sqrt{n_{\cdot g}}\hat{\boldsymbol{\alpha}}_g\}$ into a $G \times S$ matrix $\boldsymbol{\alpha}$, with $\sqrt{n_{\cdot g}}\hat{\boldsymbol{\alpha}}_g'$ as its gth row. The SVD of this matrix is

$$\boldsymbol{\alpha} = \boldsymbol{U}_G \boldsymbol{\Lambda}_G \boldsymbol{V}_G' \ ,$$

where $\boldsymbol{U}_G$ and $\boldsymbol{V}_G$ are matrices with dimensions $G \times G$ and $S \times G$ respectively, and $\boldsymbol{\Lambda}_G$ is a $G \times G$ diagonal matrix with non-negative elements that are sorted in decreasing order. For the balanced case, Appendix B.1.1 — in which the matrix $\boldsymbol{M}_G$ is equal to the current $\boldsymbol{\alpha}$ — provides more details of these matrices.

Denote the diagonal elements of $\boldsymbol{\Lambda}$ by $\lambda_1 > \lambda_2 > \ldots > \lambda_G$. Now consider the submatrices formed from just the first $d < G$ columns of both $\boldsymbol{U}_G$ and $\boldsymbol{V}_G$ (denoted $\boldsymbol{U}_G^{[1:d]}$ and $\boldsymbol{V}_G^{[1:d]}$ say), and define the $G \times G$ matrix

$$\boldsymbol{\alpha}^* = \boldsymbol{U}_G^{[1:d]} \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_d \end{pmatrix} \left[\boldsymbol{V}_G^{[1:d]}\right]' \ .$$

This can be rewritten as

$$\boldsymbol{\alpha}^* = \lambda_1 \boldsymbol{u}_1 \boldsymbol{v}_1' + \lambda_2 \boldsymbol{u}_2 \boldsymbol{v}_2' + \ldots + \lambda_d \boldsymbol{u}_d \boldsymbol{v}_d' \ ,$$

where $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ are vectors, of length G and S, containing the ith columns of $\boldsymbol{U}_G^{[1:d]}$ and $\boldsymbol{V}_G^{[1:d]}$ respectively. Denoting the gth element of $\boldsymbol{u}_i$ by u_{ig}, the gth row of $\boldsymbol{\alpha}^*$ is thus

$$\lambda_1 u_{1g} \boldsymbol{v}_1' + \lambda_2 u_{2g} \boldsymbol{v}_2' + \ldots + \lambda_d u_{dg} \boldsymbol{v}_d' \ ,$$

so that each row of $\boldsymbol{\alpha}^*$ is a weighted sum of the d vectors $\boldsymbol{v}_1', \boldsymbol{v}_2', \ldots, \boldsymbol{v}_d'$. These vectors are each of length S and, in the current context, represent distinct spatial patterns: in fact, if the matrix $\boldsymbol{\alpha}$ were regarded as a data matrix in its own right then the vectors would be the first d principal components and the weights $\{\lambda_i u_{ig} : i = 1, \ldots, d; g = 1, \ldots, G\}$ would be the corresponding principal component scores. The matrix $\boldsymbol{\alpha}^*$ is therefore a low-rank approximation of $\boldsymbol{\alpha}$, obtained by projecting the original set of G spatial patterns in $\boldsymbol{\alpha}$ onto the first d columns of $\boldsymbol{V}_G$. Moreover, the sum of squared approximation errors $\sum_{g=1}^{G}\sum_{s=1}^{S}\left(\alpha_{gs} - \alpha_{gs}^*\right)^2 = \text{trace}\,(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)\,(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)'$ is smaller than under any other rank d approximation;[3] and the ith spatial pattern $\boldsymbol{v}_i$ accounts for a proportion $\lambda_i^2 / \sum_{g=1}^{G} \lambda_g^2$ of the total variation $\text{trace}\,(\boldsymbol{\alpha}\boldsymbol{\alpha}') = \text{trace}\,(\boldsymbol{\alpha}'\boldsymbol{\alpha})$ — which is proportional to

[3] Here, α_{gs} and α_{gs}^* denote the (g, s) entries of $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}^*$ respectively.

trace $(\boldsymbol{T}_G)$ in the balanced ensemble discussion above.

As presented above, the $\{\lambda_i\}$ are multiplied by the $\{\boldsymbol{u}_i\}$ by analogy with the convention in principal component analysis. In the current context however, it is arguably more interpretable to multiply the $\{\lambda_i\}$ by the spatial patterns $\{\boldsymbol{v}_i\}$ instead: the resulting patterns $\{\lambda_i \boldsymbol{v}_i\}$ have the same units of measurement as the original index and hence can be interpreted directly as contributions to the overall variation of that index, while the scores $\{\boldsymbol{u}_i\}$ are dimensionless and hence can be interpreted as 'weights' attached to each pattern in a particular GCM's output. We refer to the patterns $\{\lambda_i \boldsymbol{v}_i\}$ as *ensemble principal patterns* (EPPs).

Subsequently in this report, maps of the first $d = 2$ EPPs $\lambda_1 \boldsymbol{v}_1$ and $\lambda_2 \boldsymbol{v}_2$ are presented for all indices considered, together with the associated proportions of GCM/RCM-specific variation: these pairs of maps are supplemented with scatterplots of the scores $\{(u_{1g}, u_{2g}) : g = 1, \ldots, G\}$. Figure 6 provides an example. The GCMs / RCMs with the largest scores (positive or negative) on either axis of such a scatterplot are those for which the associated pattern contributes most to the estimated effects: the effects for GCMs / RCMs with positive (resp. negative) scores tend to be positive (resp. negative) in areas where the pattern itself is positive, and vice versa. Throughout the report, patterns are oriented such that positive scores tend to be associated overall with high values of the underlying index. A potential use of such plots is to help users select a small number of GCMs or RCMs than span the range of variation in an ensemble.

The connection with trace $(\boldsymbol{T}_G)$ above shows that this spatial analysis is a refinement of the earlier trace-based partitioning of variation. In principle, the same approach could be used to refine the deviance-based partitioning. However, such an exercise would be based on the SVD of matrices such as $\hat{\boldsymbol{\alpha}}\hat{\boldsymbol{\Sigma}}^{-1/2}$, each element of which is a linear combination of all the elements in $\boldsymbol{\alpha}$ and hence is not linked to an individual spatial location. It is hard to see how the results of such an exercise could usefully be interpreted, so we do not attempt it.

4 Evaluation of temperature climatologies

This section evaluates the distribution of temperatures simulated by each model during the evaluation period from January 1st 1989 to December 31st 2008, considering the spread of biases within each ensemble. We begin with a discussion of each model's seasonal mean temperature, obtained by averaging the daily mean temperature variable `tas`. The spread of daily temperatures is characterised by maps of the 1st and 99th percentiles of the daily mean temperatures in a given season, denoted `tas01` and `tas99`, which can be interpreted as approximately representing, respectively, the typical coldest and warmest days in one season. Because HadUK-Grid does not provide daily time series of the mean temperature (model variable `tas`), daily mean temperatures were constructed by averaging the daily minimum and maximum temperatures (model variables `tasmin` and `tasmax`) (Perry et al., 2009). Finally, the range of diurnal temperatures simulated by each model is assessed, through evaluation of the seasonal averages of the maximum and minimum daily temperatures `tasmin` and `tasmax`. The temperature variables considered are summarised in Table 4.

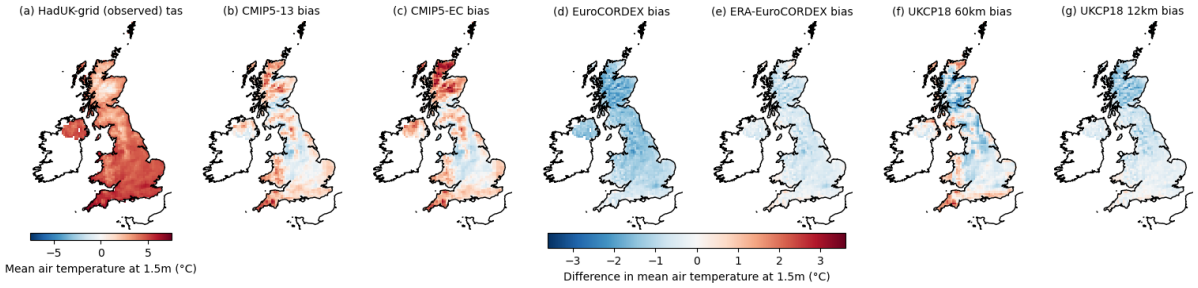
Table 4: Temperature indices used to characterise the seasonal distribution and average diurnal range of daily temperatures

Index	Description
tas	Average of daily mean temperature tas
tas01	1st percentile of daily mean temperature tas
tas99	99th percentile of daily mean temperature tas
tasmin	Average of daily minimum temperature tasmin
tasmax	Average of daily maximum temperature tasmax

4.1 Average winter temperatures

Figure 3 shows a map of the HadUK-Grid mean daily temperature in winter (panel a), together with the mean deviation from that temperature for each of the ensembles described in Section 2 (panels b-g). The two CMIP5 ensembles share a similar pattern of biases, being too warm at high elevations and around much of the coast, but up to 1°C too cold over much of central England. This pattern can be attributed to an underlying cold bias in many of the GCMs, offset by local warm biases due to unresolved orography and blurring between land and sea surface temperatures due to the coarse resolution. The RCMs are able to resolve the features causing the warm biases in the driving GCMs, with the EuroCORDEX ensemble as a whole having a fairly constant bias of between -1 and -2°C across the UK land surface, with London having a slightly larger cold bias than the surrounding area. For the evaluation runs driven by ERA-Interim reanalysis, the magnitude of this bias is somewhat reduced, suggesting that the error is largely inherited from the driving GCMs: however, the London offset remains, suggesting that this is related to the RCMs' representation of urban temperatures. The UKCP18 ensemble has a small but fairly constant bias of around -0.5°C over the UK at both 60km and 12km resolutions, with the 60km ensemble also displaying warm biases around the coast similar to those seen in the CMIP5 GCMs.

Figure 3: Maps of (a) HadUK-Grid daily mean temperature (in °C) during the winter months from 1989 to 2008, and (b-g) of the mean climatological biases in each of the ensembles of models.



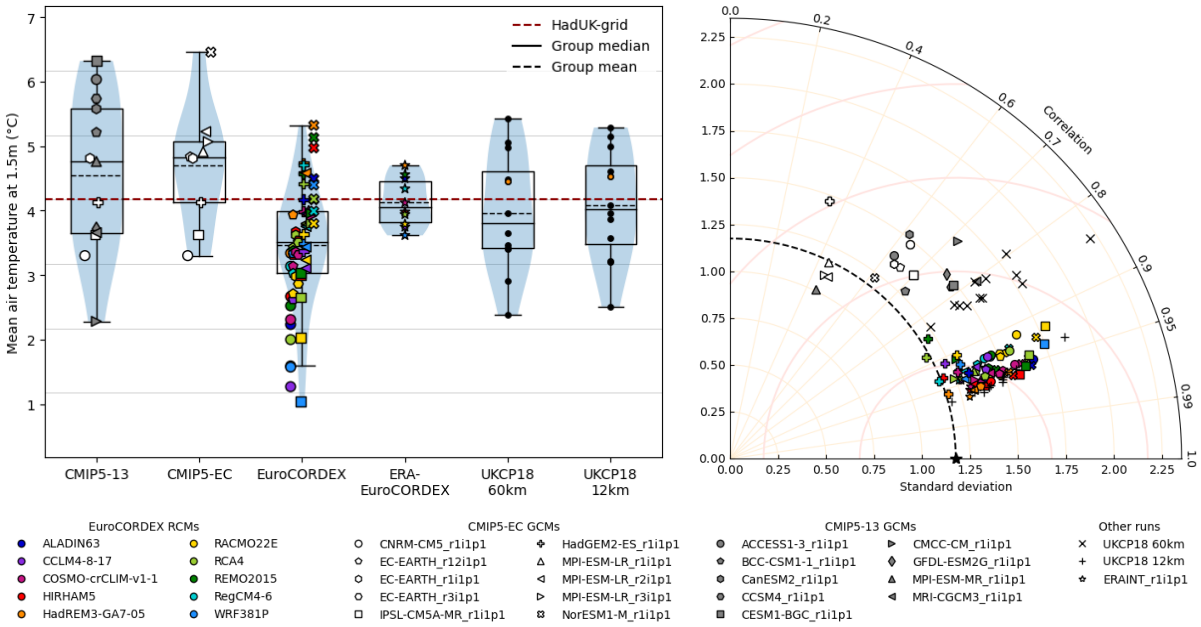
The boxplots in Figure 4a show the distributions of UK-averaged temperatures simulated by each run. The two GCM ensembles cover a similar range of biases, although the CMIP5-EC ensemble does not include the coldest member of the CMIP5-13 ensemble. The reanalysis-driven EuroCORDEX ensemble has a much narrower range of biases than the main EuroCORDEX ensemble, where runs represented by the same symbol (denoting the same driving GCM) are more closely grouped than runs represented by the same colour (denoting the same downscaling RCM), suggesting that differences between the model biases in the main EuroCORDEX ensemble are driven largely by the GCMs.

The UKCP18 runs have similar UK-averaged biases in mean winter temperature at 60km and 12km resolution, and cover much of the range of the EuroCORDEX ensemble, with the exception of the coldest runs.

Figure 4: Bias boxplots & Taylor diagrams for daily mean temperatures during the winter months from 1989 to 2008.

(a) Boxplots of average daily mean temperature over the UK land surface for each run

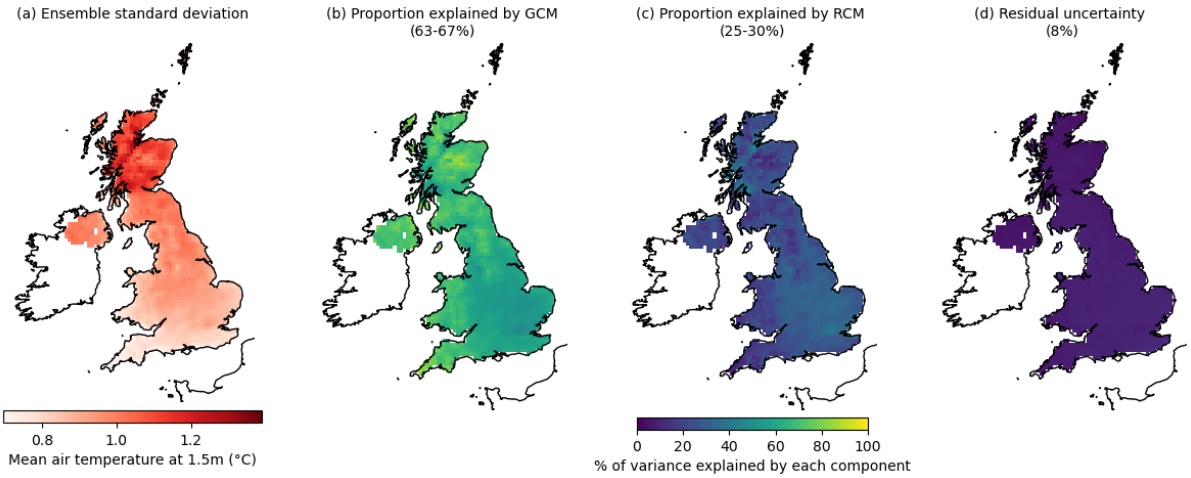
(b) Taylor diagram comparing model climatologies to HadUK-Grid climatology



Complementing Figure 4a, the Taylor plots in Figure 4b summarise the success of the models in capturing the climatological pattern of winter temperatures. All runs are denoted by the same symbols as those used in the boxplots and summarised in Figure 2. Correlation between the downscaled GCMs and the HadUK-Grid observations is quite variable and, due to the unresolved features mentioned above, sometimes poor, ranging from 0.35 to 0.8. Correlation between the model fields and the observed field increases with resolution, with the UKCP18 60km runs having correlation scores between 0.8 and 0.85, and the 12km runs between 0.9 and 0.96. Almost all of the points lie to the right of the dashed arc indicating the standard deviation of the climatological field, indicating that the runs overestimate the spatial variability of daily mean temperatures somewhat; the RCMs tend to do so to a greater degree than their driving GCMs, although the UKCP18 60km runs do so more than their 12km-resolution counterparts. The regional runs from both the EuroCORDEX and UKCP18 ensembles are fairly closely grouped, with the UKCP18 runs among those with the highest correlation with the observed field, while those driven by HadGEM2-ES lie close to the dashed arc, indicating that they tend to better simulate the amplitude of the spatial variability in average winter temperatures.

Turning next to the structure of the ensemble biases: between 63 and 67% of the variance of the runs in the EuroCORDEX ensemble about the ensemble mean can be attributed to the driving GCM, with 25-30% attributed to the downscaling RCM. Maps of the spatial distribution of the standard

Figure 5: Maps of (a) the standard deviation of winter mean temperature climatologies in the EuroCORDEX ensemble (including evaluation runs driven by ERA-Interim reanalysis), and (b-d) of the percentage of the total ensemble variance in each grid cell explained by the choice of GCM and RCM. The ranges of proportions of the total variance explained by each component, computed using the methodology described in Section 3.5.2, are given in parentheses.

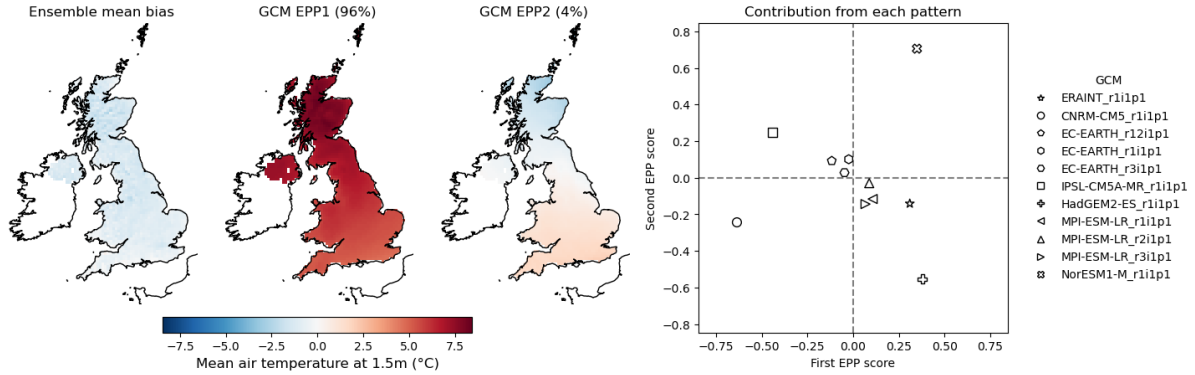


deviation of the EuroCORDEX model biases, and of the proportion of this variability explained by the GCM and RCM, are shown in Figure 5. In panel a, showing the ensemble standard deviation, lighter colours indicate a lower standard deviation within the ensemble, and so indicate greater agreement between the ensemble members; the ensemble is more variable in Scotland — where the RCM contributes a larger proportion of the variance — than in the south of England, although there is visibly slightly higher variability in winter temperatures in London. The proportion of variance explained by the GCMs is fairly constant across much of the UK, but greater at higher elevations; differences between the RCMs contribute slightly more to the variability in temperatures in London. A multivariate ANOVA, as described in Section 3.5, found that the RCMs accounted for 84-87% of the total deviance once interactions between locations were taken into account, and the GCMs for just 10-13%: this may be because differences between the RCMs account for a greater proportion of the spatial variability between grid cells than the GCMs.

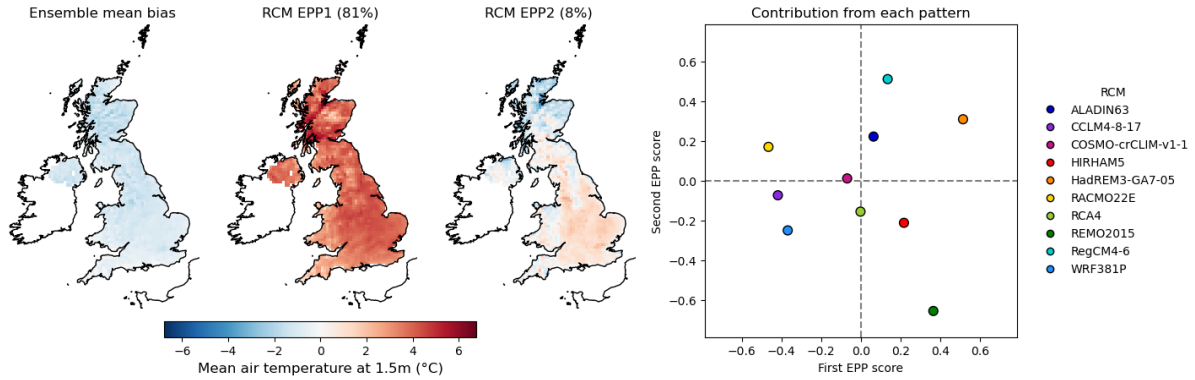
GCM and RCM effects were estimated for the unbalanced EuroCORDEX ensemble — including the reanalysis-driven evaluation runs — using the methodology described in Section 3.5.2, and Ensemble Principal Patterns (EPPs) and scores calculated for the resulting maps as described in Section 3.6, in order to identify the dominant patterns of variability between the GCMs and between the RCMs. Of the 63-67% of the total variability due to the GCMs, 96% is determined by the pattern labelled 'GCM EPP1' in Figure 6a, which is spatially fairly uniform over the UK, with a slight north-south gradient indicating larger differences from the ensemble mean in Scotland than in the south of England. Runs driven by a GCM with a positive (resp. negative) score corresponding to this pattern tend to be warmer (resp. cooler) than the ensemble average: the scores assigned to this component are closely correlated with the biases in individual model runs in Figure 4a. The second EPP, which contributes just 4% of the variability due to the driving GCM, is a north-south gradient opposed to that seen in the first component: runs driven by a GCM with positive or negative scores for both EPPs will have

Figure 6: Ensemble principal patterns in winter temperature climatologies from 1989-2008, showing dominant patterns of contributions from each GCM and RCM to variation about the mean bias in the EuroCORDEX ensemble. The proportion of the between-GCM or between-RCM variation attributed to each pattern is given in parentheses.

(a) Contribution of each driving GCM to EuroCORDEX ensemble spread



(b) Contribution of each downscaling RCM to EuroCORDEX ensemble spread



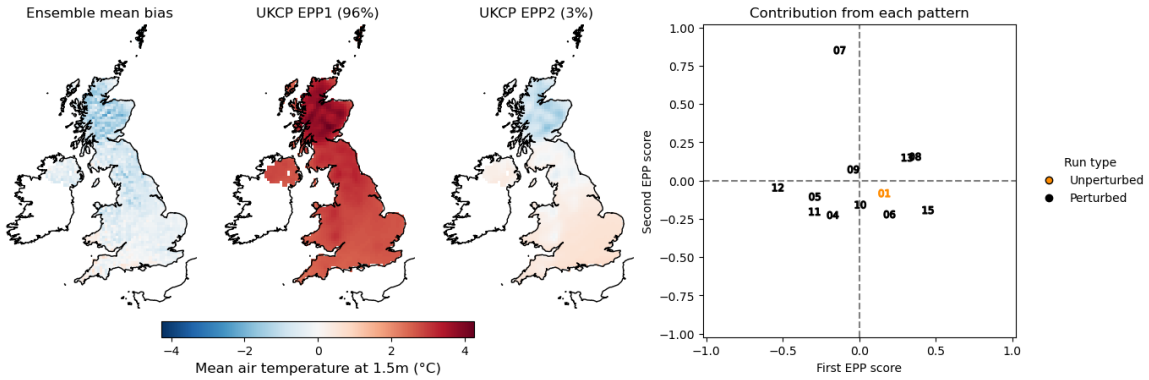
a less pronounced north-south temperature gradient than a GCM with scores with different signs.

Of the 25-30% of the total variability due to the RCM, 81% was associated with the spatial pattern labelled 'RCM EPP1' in Figure 6b: a bias that is fairly constant across the extent of the UK, but smaller in areas to the east of mountainous regions, suggesting that this is related to the strength of topographic influences within the models. Runs downscaled by an RCM with a positive (resp. negative) score for this first EPP tend to be warmer (resp. cooler) than the average. The second EPP, labelled 'RCM EPP2', primarily reflects a contrast between areas that lie to the east of mountainous regions and those that do not; however, this component is largely unimportant.

EPP analysis was also carried out for the members of the UKCP18 regional ensemble (Figure 7). The main patterns of variation are very similar to those occurring between EuroCORDEX runs driven by different GCMS: 96% of the variability within the 12km ensemble is associated with the pattern labelled 'EPP1' in Figure 7, which, like the first GCM EPP in Figure 6a, represents a fairly uniform temperature offset across the UK, with a larger offset in Scotland. Similarly, 3% of the total variability is attributed to a reduction in the offset in Scotland relative to the rest of the UK, although only ensemble member 07 receives a particularly high score in this component. It would

be of interest to examine the precise perturbation of the model physics for this ensemble member: the result suggests an alternative use of EPP analysis, as a diagnostic tool to highlight features of an ensemble that are deserving of further investigation.

Figure 7: EPP analysis of mean winter temperatures simulated by the UKCP18 regional models, showing the dominant patterns of variation. The proportion of the total variation attributed to each pattern is given in parentheses.

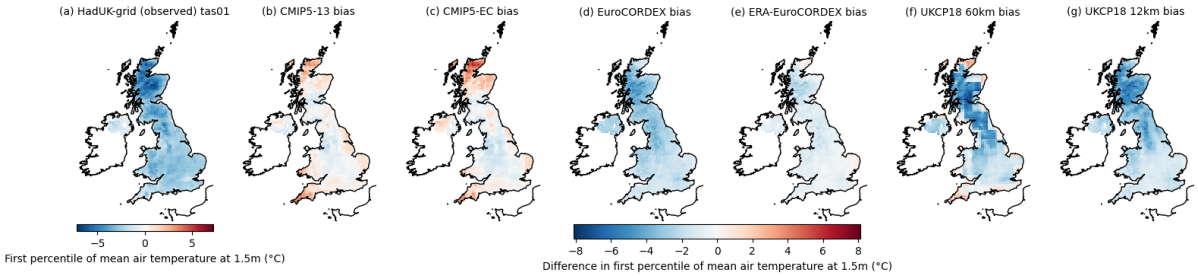


4.2 Spread of winter temperatures (`tas01` and `tas99`)

Having evaluated the models' seasonal climatologies, we now move on to assessing the spread of daily temperatures simulated by each run. This is done by considering the 1st and 99th percentiles of the winter daily mean temperatures, `tas01` and `tas99`.

Figure 8 shows maps of `tas01` for HadUK-Grid, along with each ensemble's mean bias. As in Figure 3, the two CMIP5 ensembles and the 60km UKCP18 ensemble (panels b, c and f) remain too warm around the coast; however, the warm biases at higher elevations due to unresolved topography are somewhat reduced. In the regional ensembles (panels d, e and g) the UK-wide cold biases in `tas` that were seen in Figure 3 are still apparent.

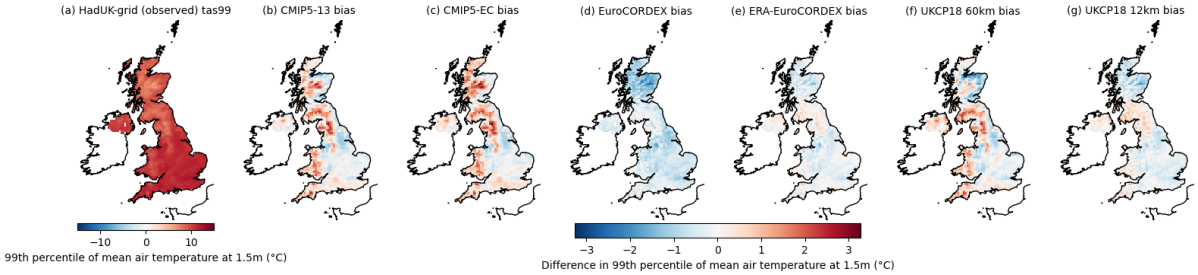
Figure 8: Maps of `tas01`, the first percentile of daily temperatures (in °C) during the winter months from 1989 to 2008: (a) in HadUK-Grid and (b-g) biases in each ensemble.



The same maps are shown for `tas99` in Figure 9. Here, there is almost no coastal effect in the GCM and UKCP18 60km ensemble biases (panels b, c and f), because the land and sea surface temperatures are typically fairly similar on the warmest days in winter; however, warm biases at higher elevations are more pronounced. The GCM-driven EuroCORDEX runs are slightly too cold across the UK, while the patterns of biases in the reanalysis-driven and UKCP18 12km ensembles

are very similar, with both ensembles being on average slightly too cold in central and southeastern England and northern Scotland, and slightly too warm elsewhere.

Figure 9: Maps of tas99, the 99th percentile of daily temperatures (in °C) during the winter months from 1989 to 2008: (a) in HadUK-Grid and (b-g) biases in each ensemble.



The spread of UK-averaged temperatures in each run is shown in Figure 10. These plots represent the range of daily temperatures simulated by each model as a vertical line from tas99 to tas01, with the mean temperature tas marked by a symbol denoting the driving GCM and, where applicable, a colour denoting the downscaling RCM. Red dashes mark tas01 and tas99 for each run after the bias in tas has been removed, so that the width of the interval between tas01 and tas99 can be compared directly for all runs. The corresponding quantities in the HadUK-Grid dataset are marked by dashed red lines: tas99 at 10.3°C, tas at 4.2°C and tas01 at -2.6°C, representing a range of 12.8°C between tas01 and tas99 in winter.

The GCMs in the CMIP5-13 and CMIP5-EC ensembles typically simulate narrower ranges of temperatures than the RCMs of the EuroCORDEX ensemble, which may in part be due to warm coastal biases affecting the UK average of tas and tas01 in the GCMs: these shorten the lower tail of the distribution but not the upper. Of the CMIP5-EC runs used to drive the EuroCORDEX ensemble, most simulate a similar range of temperatures to those in HadUK-Grid, although the two EC-EARTH runs fail to replicate the coldest temperatures, and NorESM1-M is systematically around two degrees warmer than the observed temperatures. As already noted in Section 4.1, UK-wide mean biases in the temperatures simulated by the RCMs in the EuroCORDEX ensemble tend to be slightly lower than the biases in their driving GCMs: most of the EuroCORDEX and ERA-EuroCORDEX runs simulate tas99 fairly well, with the top end of the line close to or slightly above the upper red line, while many of the runs underestimate tas01 by more than 1°C. Likewise, at both 60km and 12km resolutions all but one of the runs in the UKCP18 ensemble simulate tas99 close to the observed value, while nine of the twelve runs underestimate tas01 by more than 1°C.

The proportions of variance within the EuroCORDEX ensemble attributable to the choice of GCM and the choice of RCM are shown in Figure 11. As noted above, on average and warm days, the temperature is largely driven by the choice of GCM, while on cold days, both the choice of RCM and residual variability play a greater part. Plots of the EPPs of the between-GCM and between-RCM variability within the EuroCORDEX ensemble are not presented here; the spatial patterns of the dominant modes of variation were very similar to those seen in Figure 6, as were the scores assigned to the majority of the models.

Figure 10: Plots showing the typical range of daily temperatures simulated in winter by each model, averaged over the UK. The top of each line is at `tas99`, the 99th percentile of daily mean temperatures; the bottom is at `tas01`, the first percentile of daily mean temperatures; and the symbol denotes `tas`, the mean of the daily mean temperatures. Individual red dashes mark `tas01` and `tas99` for each run after the bias in `tas` has been removed. The corresponding dashed red horizontal lines denote `tas99`, `tas` and `tas01` in the HadUK-Grid dataset; the short red markers show the distance between `tas99` and `tas01` with the bias removed. The grey lines mark intervals of 1°C.

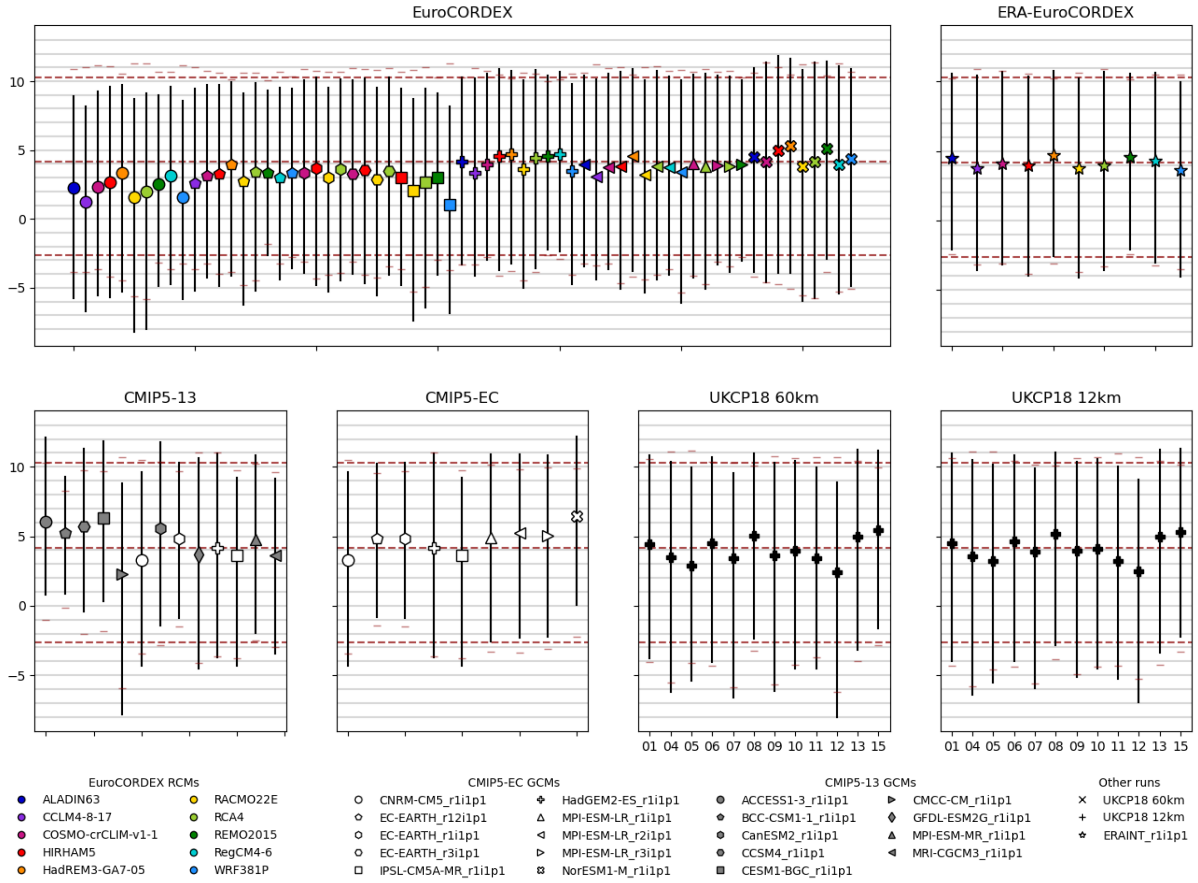
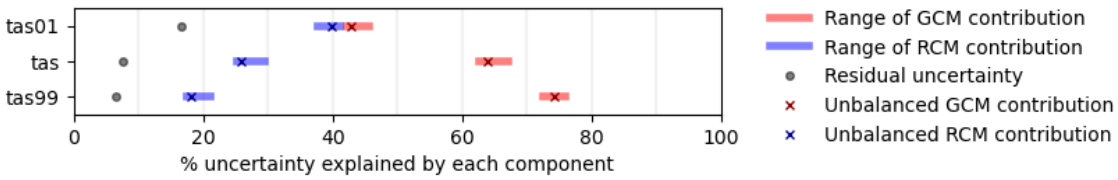


Figure 11: Range of percentages of variance within the EuroCORDEX ensemble explained by the driving GCM, by the downscaling RCM, and left unexplained, computed using the method described in Section 3.5.2.



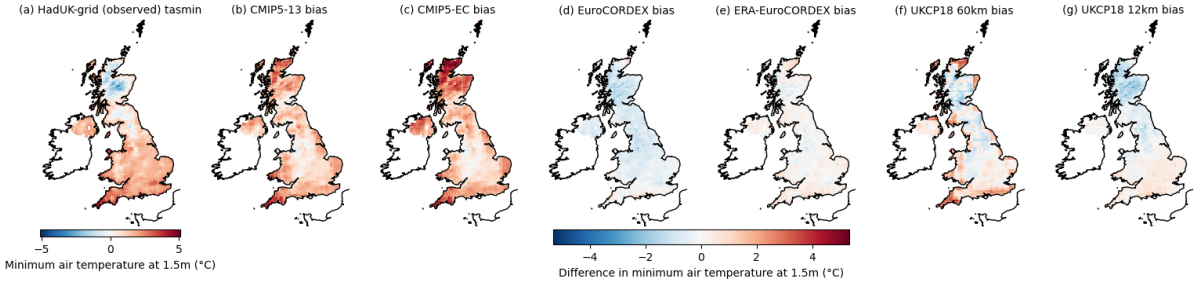
4.3 Diurnal range of winter temperatures (`tasmin` & `tasmax`)

We now consider the diurnal range of winter temperatures simulated by each run, by evaluating maps of the average daily minimum and maximum temperatures, denoted `tasmin` and `tasmax` respectively.

Maps of the HadUK-Grid observed `tasmin`, and the average bias in each ensemble, are shown in

Figure 12. Both of the CMIP ensembles tend to overestimate daily minimum temperatures across the UK, particularly in Scotland and Northern Ireland, and around the coast (panels b and c). The EuroCORDEX ensemble has a collective cold bias across much of the UK, which appears to be largely inherited from the driving GCMs, being reduced to almost zero in the reanalysis-driven ERA-EuroCORDEX ensemble (panels d and e). The UKCP18 60km ensemble, like the CMIP5 ensembles, retains warm biases around the coastline due to blurring of the land surface with the warmer sea surface, but is otherwise slightly too cold across Scotland (panel f): this bias is also apparent in the 12km ensemble (panel g).

Figure 12: Maps of `tasmin`, the average of the daily minimum temperatures during the winter months from 1989 to 2008: (a) in HadUK-Grid and (b-g) biases in each ensemble.



The corresponding maps for `tasmax` are shown in Figure 13. Here, the GCM ensembles (panels b and c) are up to 2°C too cool across much of the UK, but too warm at the highest elevations. This cool bias is inherited by the EuroCORDEX ensemble, which is between 1 and 3°C too cold across the whole of the UK; the reanalysis-driven EuroCORDEX ensemble retains a smaller cold bias of up to 2°C, suggesting that this bias is partly inherent to the RCMs. The average maximum daily temperatures within the two UKCP18 ensembles are also 1-2°C too cold across most of the UK.

Figure 13: Maps of `tasmax`, the average of the daily maximum temperatures during the winter months from 1989 to 2008: (a) in HadUK-Grid and (b-g) biases in each ensemble.

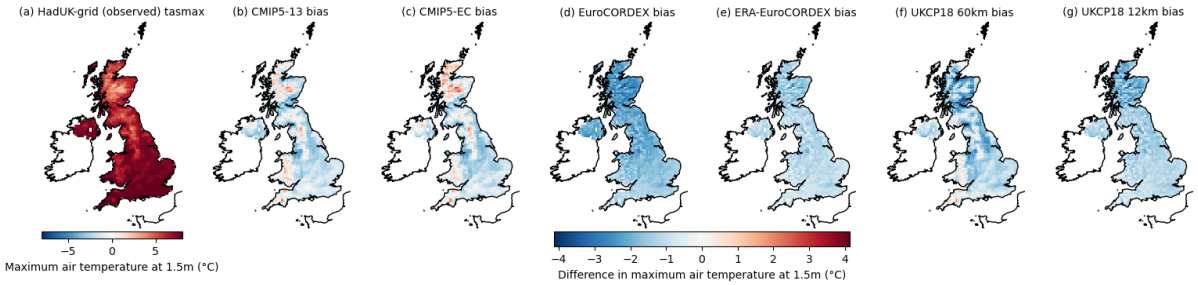
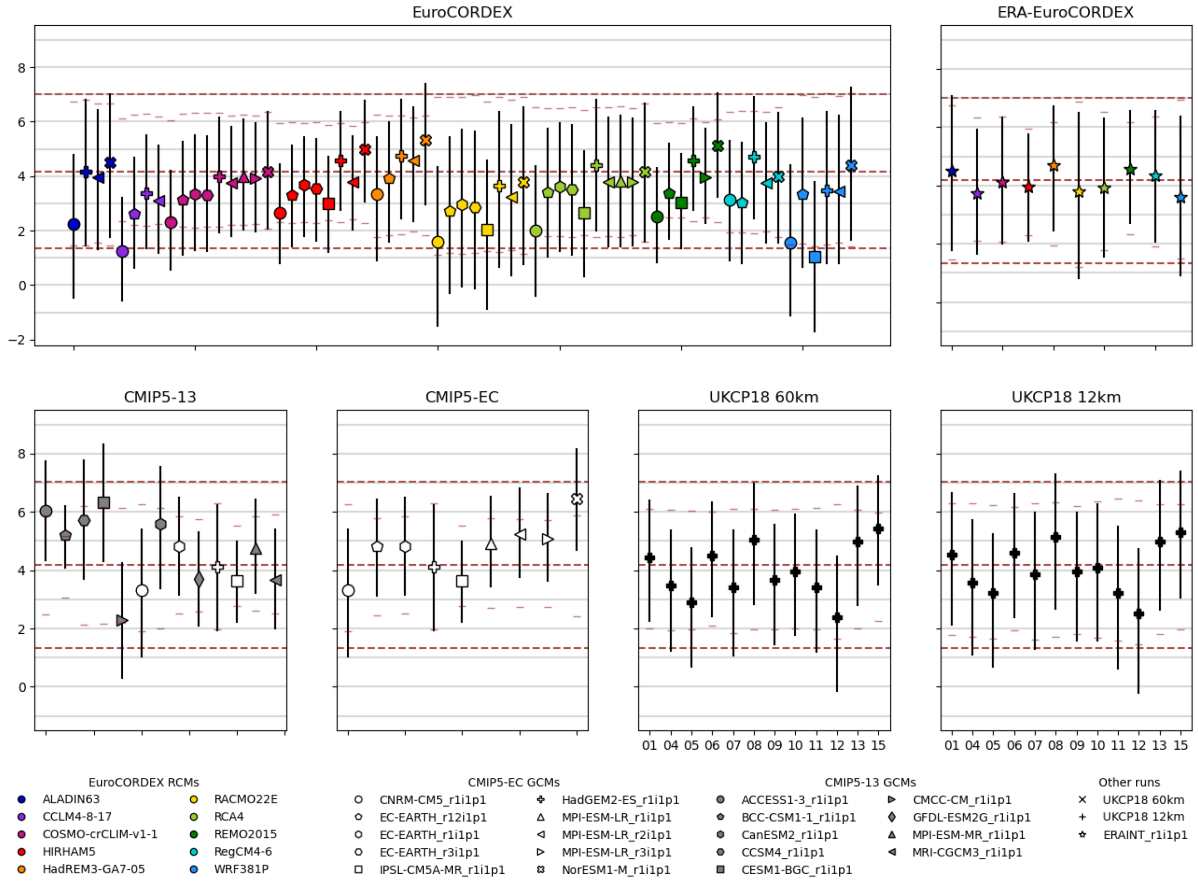


Figure 14 shows the average diurnal temperature range simulated by each run, represented as a vertical line between `tasmin` and `tasmax`, with the daily mean temperature `tas` marked as a point; the colour and shape of the marker denote, where applicable, the driving GCM and downscaling RCM. The width of the interval between `tasmin` and `tasmax` for each run, with biases in `tas` removed for easier comparison of the interval between runs, is represented by red dashes. The red lines mark the corresponding quantities in the HadUK-Grid dataset: `tasmin` at 1.3°C, `tasmax` at 7.0°C, and `tas` at 4.2°C, representing a difference of 5.7°C between `tasmin` and `tasmax`.

Figure 14: Plots showing the typical ranges of daily temperatures simulated in winter by each model, averaged over the UK. The top of each line is at tasmax, the average of the daily maxima; the bottom is at tasmin, the mean of the daily minima; and the symbol denotes tas, the mean of the daily mean temperatures. The corresponding dashed red lines denote tasmax, tas and tasmin in the HadUK-Grid dataset; the grey lines mark intervals of 1°C.



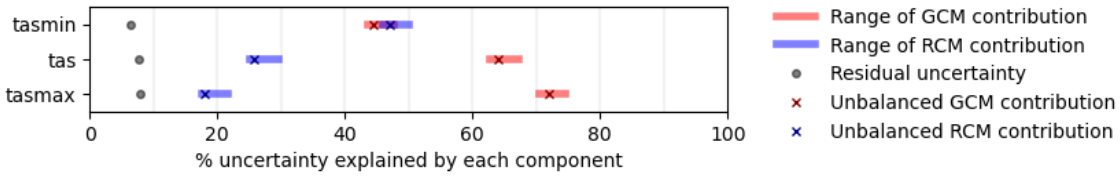
Almost all of the runs underestimate the diurnal temperature range, with only a handful of the EuroCORDEX and ERA-EuroCORDEX runs coming close to the observed values of both tasmin and tasmax. While a large part of the overall temperature bias is inherited from the driving GCMs, the width of the diurnal temperature range is dependent on the RCM, with ALADIN63, RACMO22E and WRF381P (denoted by symbols coloured dark blue, yellow and sky blue, respectively) all simulating a diurnal range close to tas ± 2.8°C, and all other models simulating diurnal ranges as short as tas ± 1.8°C, regardless of whether the runs were driven by reanalysis or by a GCM.

In the UKCP18 ensemble, the 60km runs all have a very similar diurnal range of tas ± 2°C, while the 12km regional runs estimate a slightly wider diurnal range of temperatures, reaching on average around tas $- 2.5$°C at night, and a slightly shorter range of tas $+ 2.2$°C during the day.

The contributions of the choice of GCM and the choice of RCM to the spread of biases within the EuroCORDEX ensemble are summarised in Figure 15. Systematic differences between the runs in daytime temperatures tasmax are largely determined by differences between the driving GCMs, while nighttime temperatures tasmin receive an equal contribution from the RCM. Plots of the EPPs of

the between-GCM and between-RCM variability within the EuroCORDEX ensemble are again not presented here, but the spatial patterns of the dominant modes of variation were found to be very similar to those seen in Figure 6, as were the scores assigned to the majority of the models.

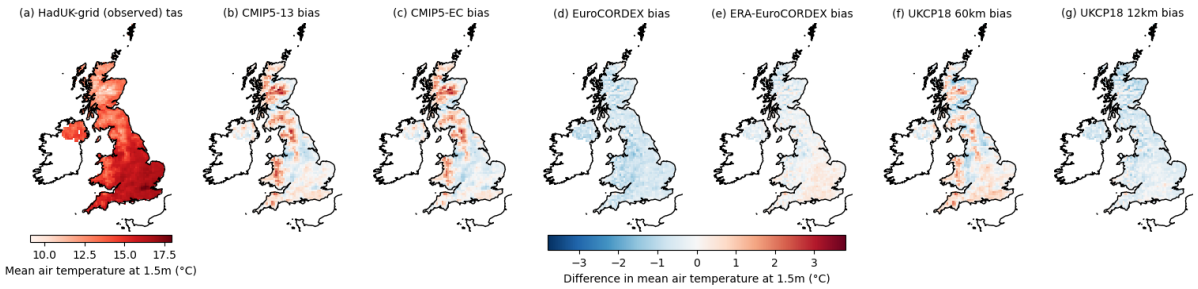
Figure 15: Range of percentages of variance within the EuroCORDEX ensemble explained by the driving GCM, by the downscaling RCM, and left unexplained, computed using the method described in Section 3.5.2.



4.4 Average summer temperatures

A map of the average observed daily summer temperature is shown in panel a of Figure 16, alongside maps of each ensemble's mean bias with respect to that observed average temperature. The two CMIP5 ensembles overestimate the temperature at higher elevations due to their low resolution, as does the UKCP18 60km ensemble, albeit to a lesser extent. Both the EuroCORDEX and UKCP18 12km ensembles have a cold bias of around 1°C across the UK, which is much reduced in the reanalysis-driven ensemble.

Figure 16: Maps of (a) HadUK-Grid mean daily temperature (in °C) during the summer months from 1989 to 2008, and (b-g) of the mean climatological biases in summer temperatures in each of the ensembles of models.

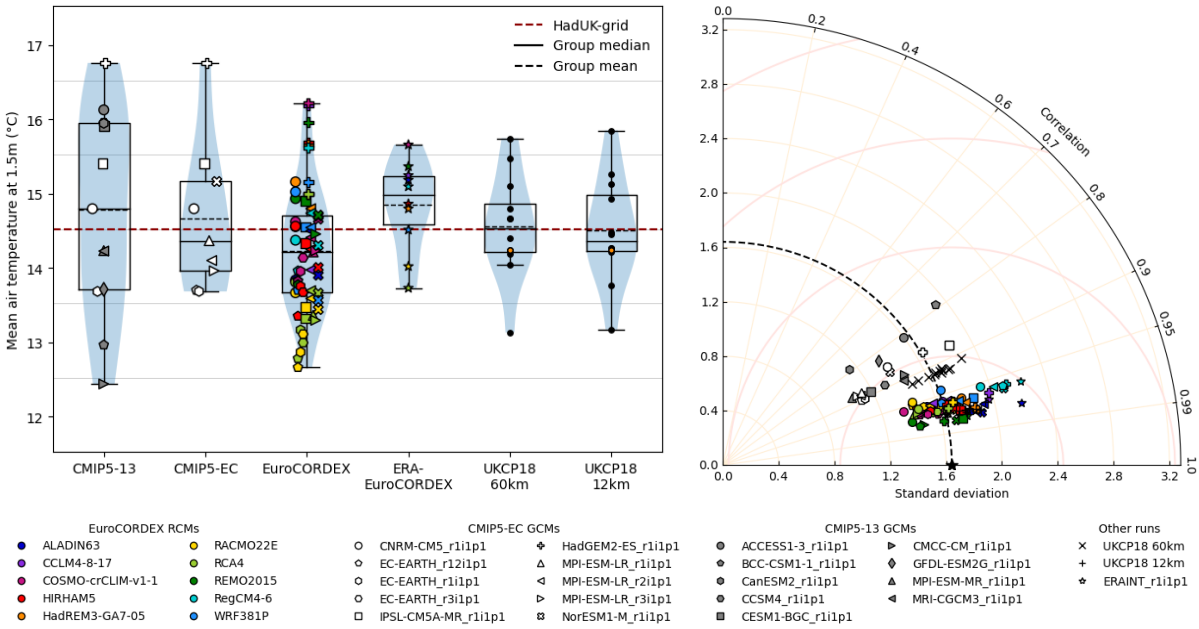


The distributions of UK-averaged summer temperatures simulated by the runs in each ensemble are shown in the boxplots in Figure 17a. The spread of the CMIP5-EC ensemble is somewhat smaller than that of the CMIP5-13 ensemble and heavily skewed by the inclusion of one GCM — HadGEM2-ES — with a particularly large bias of over 2°C. Within the EuroCORDEX ensemble, the average summer temperatures are broadly grouped according to the driving GCMs, with the warmest runs driven by HadGEM2-ES; however, there is also a degree of clustering by RCM, with the coolest summers simulated by runs downscaled using RACMO22E and RCA4 (coloured yellow and lime green), suggesting that the RCMs also contribute systematic differences of their own. The two UKCP18 ensembles again have similar distributions of UK-averaged summer temperatures, and span almost the full range covered by the EuroCORDEX ensemble.

Figure 17: Bias boxplots & Taylor diagrams for mean daily temperatures during the summer months from 1989 to 2008.

(a) Boxplots of average daily mean temperature over the UK land surface for each run

(b) Taylor diagram comparing model climatologies to HadUK-Grid climatology



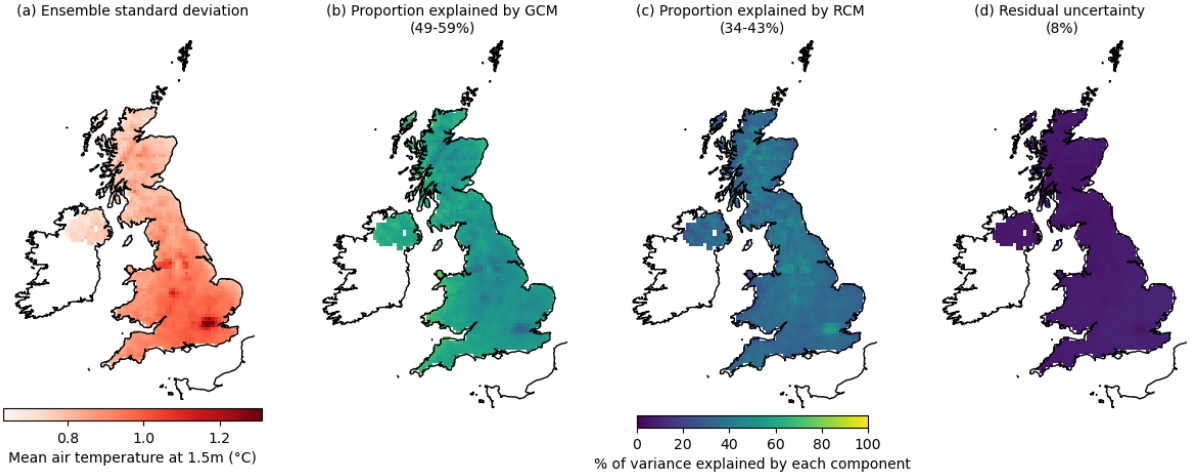
All of the models are more successful at capturing the spatial distribution of summer temperatures — summarised in the Taylor plots in Figure 17b — than winter temperatures (shown in Figure 4b). All of the ensembles span the dashed arc denoting the reference standard deviation, whereas the amplitude of the spatial patterns of average winter temperatures was systematically overestimated. Correlations between the model temperatures and the HadUK-Grid temperatures are also high in all ensembles, with all EuroCORDEX and UKCP18 12km runs having correlation scores between 0.95 and 0.99, the UKCP18 60km runs having scores close to 0.9, and the GCMs achieving scores between 0.8 and 0.9.

We now consider the structure of the uncertainty within the EuroCORDEX ensemble. An ANOVA was carried out over the unbalanced EuroCORDEX ensemble, including the reanalysis-driven evaluation runs, using the methodology described in Section 3.5.2: between 49 and 59% of the total variance of the EuroCORDEX runs about the ensemble mean bias was attributed to differences between the GCMs, with between 34 and 43% attributed to differences between the RCMs. Multivariate ANOVA found that, as with the winter temperatures, the RCMs accounted for 84-88% of the total deviance once interactions between locations were taken into account, and the GCMs for just 10-14%: again, this may be because differences between the RCMs account for a greater proportion of the spatial variability between grid cells than the GCMs.

Maps of the standard deviation of the EuroCORDEX runs at each grid cell, and of the proportion of this variability explained by the GCM and RCM in each grid cell, are shown in Figure 18. The greatest variation in summer temperatures occurs in urban regions, with London, Birmingham, Liverpool,

Manchester, Sheffield and Leeds clearly visible as darker patches in panel a. These areas of greater variability are associated more closely with the choice of RCM than the choice of GCM, appearing as darker spots in Figure 18b, and brighter spots in Figure 18c; this is most likely related to the RCMs' differing representation of urban heat island effects in summer.

Figure 18: Maps of (a) the standard deviation of summer temperature climatologies in the EuroCORDEX ensemble (including evaluation runs driven by ERA-Interim reanalysis), and (b-d) of the percentage of the total ensemble variance in each grid cell explained by the choice of GCM and RCM. The ranges of proportions of the total variance explained by each component are given in parentheses.



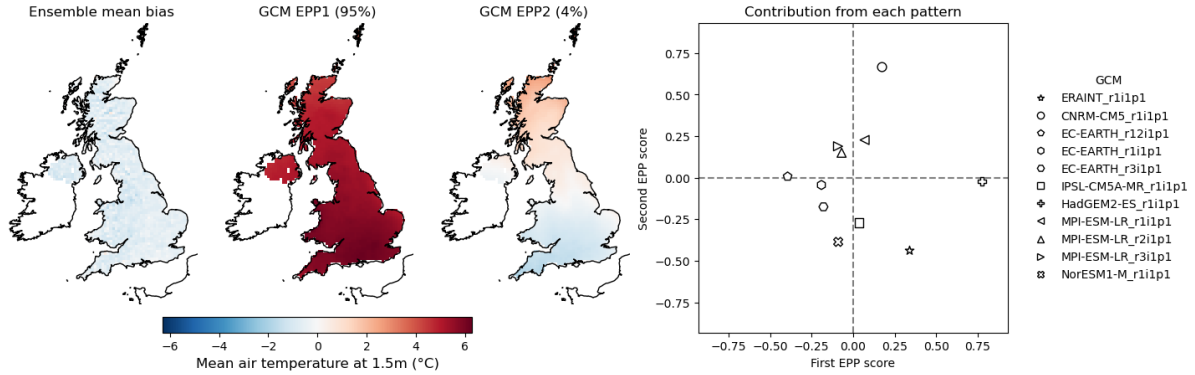
EPP analysis of the GCM and RCM effects fitted using the methodology described in Section 3.5.2 was carried out to learn more about the dominant spatial patterns of variation in the biases: the results are shown in Figure 19. These patterns are broadly similar to those seen in the winter temperatures (Figure 6): 95% of the between-GCM variability (and therefore between 47 and 56% of the total variability in the EuroCORDEX ensemble) is attributed to the first EPP ('GCM EPP1' in Figure 6a), a relatively uniform temperature offset across the UK with a slight north-south gradient, with models that receive positive (negative) scores in the first EPPs tending to be warmer (cooler) than the ensemble average, and warmer (cooler) in the south of England than in Scotland. The second component, which accounts for only 4% of the between-GCM variability, adjusts the north-south gradient, so that runs driven by a GCM with scores of the same sign in both components will have a less pronounced north-south temperature gradient than runs driven by a GCM with one positive and one negative score.

Of the 34-43% of the total variability associated with differences between runs downscaled using different RCMs, 81% is attributed to the pattern labelled 'RCM EPP1' in Figure 6b: again, this represents a fairly uniform offset from the ensemble mean across the UK, with larger offsets in urban areas such as London and Birmingham. The second component, which contributes 10% of the between-RCM variation (3-4% of the total), is a pattern of higher-than-average temperatures at high elevations, in Scotland, Northern Ireland and the south-west of England, and lower-than-average temperatures elsewhere: one model (RegCM4-6) receives a particularly large negative score in this component, suggesting that runs downscaled using this model tend to be warmer than the

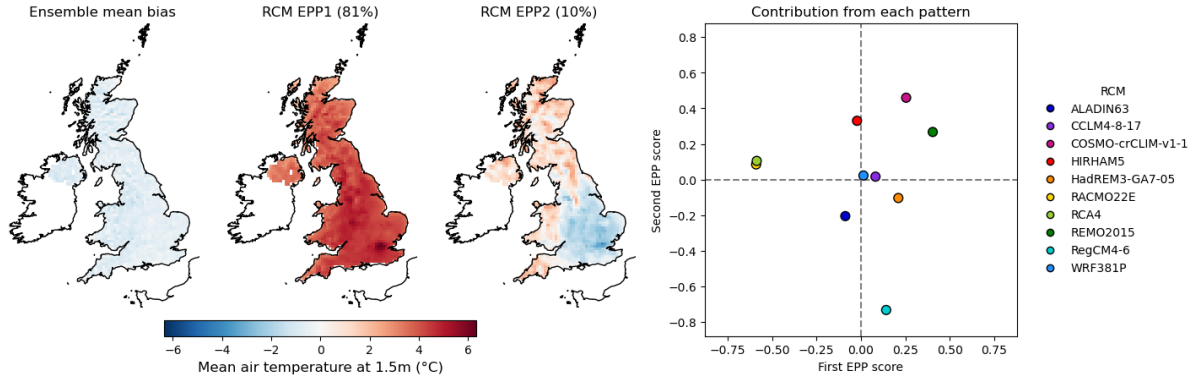
ensemble average across much of England, and cooler than the ensemble average elsewhere.

Figure 19: Ensemble principal patterns in summer temperature climatologies from 1989-2008, showing dominant patterns of contributions from each GCM and RCM to variation about the mean bias in the EuroCORDEX ensemble. The proportion of the between-GCM or between-RCM variation attributed to each pattern is given in parentheses.

(a) Contribution of each driving GCM to EuroCORDEX ensemble spread



(b) Contribution of each downscaling RCM to EuroCORDEX ensemble spread



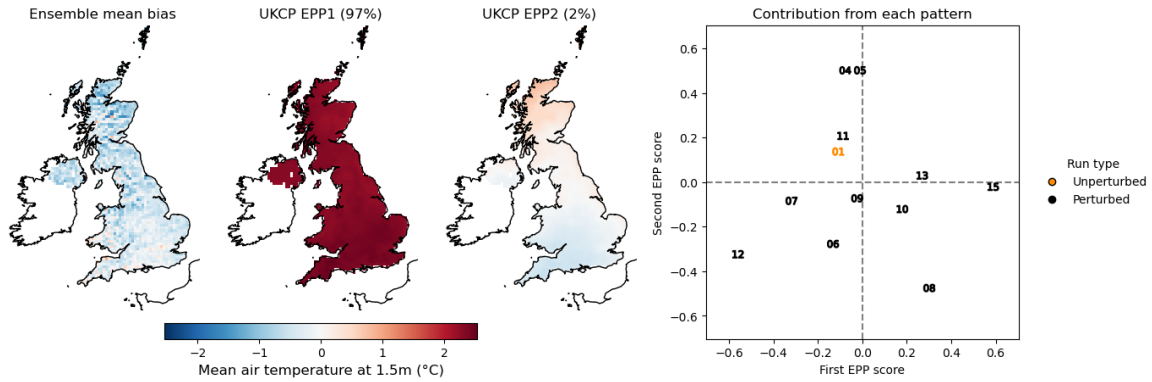
EPP analysis was also carried out for the members of the UKCP18 regional ensemble (Figure 20). As with the winter temperatures, the main patterns of variation between the runs are very similar to those occurring between EuroCORDEX runs driven by different GCMS: 97% of the variability within the 12km ensemble is associated with the pattern labelled 'EPP1' in Figure 20, which, like the first GCM EPP in Figure 6a, represents a fairly uniform temperature offset across the UK, while 2% is attributed to differences in the north-south gradient in temperatures simulated by the runs.

4.5 Spread of summer temperatures (`tas01` and `tas99`)

We now consider the variables `tas01` and `tas99`, which describe the 1st and 99th percentiles, respectively, of the daily mean temperatures during the summer months of June, July and August. These variables can be interpreted as characterising typical temperatures on the coldest and warmest days, respectively, in a single summer; together, they indicate how well the models are able to capture the full range of observed daily temperatures during the summer months.

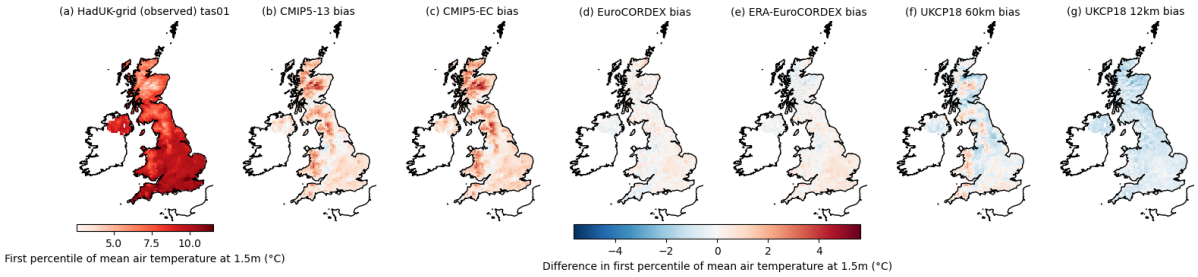
Maps of `tas01` in the HadUK-Grid dataset, and of the mean bias in each ensemble with respect

Figure 20: EPP analysis of mean summer temperatures simulated by the UKCP18 regional models, showing the dominant patterns of variation. The proportion of the total variation attributed to each pattern is given in parentheses.



to HadUK-Grid, are shown in Figure 21. Both of the CMIP5 ensembles overestimate `tas01` across the UK, but particularly at higher elevations; the biases in the GCM-driven and reanalysis-driven EuroCORDEX ensembles are close to zero across much of the UK, although both are slightly too warm in the south-east. The UKCP18 60km ensemble is up to 2°C too cold across much of the UK on the coolest summer days, except in areas of higher elevation, where the temperature is overestimated due to the unresolved topography. The UKCP18 12km ensemble is able to resolve these features, but retains the cold bias, being around 1°C too cool across the UK.

Figure 21: Maps of `tas01`, the first percentile of daily temperatures (in °C) during the summer months from 1989 to 2008: (a) in HadUK-Grid and (b-g) biases in each ensemble.



The corresponding maps of HadUK-Grid `tas99` and the mean biases in `tas99` in each of the ensembles are shown in Figure 22. When simulating the warmest summer days, the CMIP5 ensembles have quite different underlying mean biases, suggesting substantial differences between the temperatures produced by the individual models — although both ensembles overestimate the temperature in areas of unresolved high elevation, and underestimate the warmest temperatures in London by around 2°C. On average, the EuroCORDEX ensemble underestimates the hottest summer temperatures fairly uniformly across the UK, while the reanalysis-driven ERA-EuroCORDEX ensemble is slightly too warm across central and southern England and Wales. The UKCP18 60km ensemble displays a similar mean bias to the CMIP5-13 ensemble, while the 12km ensemble is generally too cool in Scotland and northern England, with visible localised warm biases of up to 1°C around London, Oxford, Birmingham and Manchester, suggesting that the members of this ensemble tend to overestimate

the effect of urban heating on the warmest days.

Figure 22: Maps of tas99, the 99th percentile of daily temperatures (in °C) during the summer months from 1989 to 2008: (a) in HadUK-Grid and (b-g) biases in each ensemble.

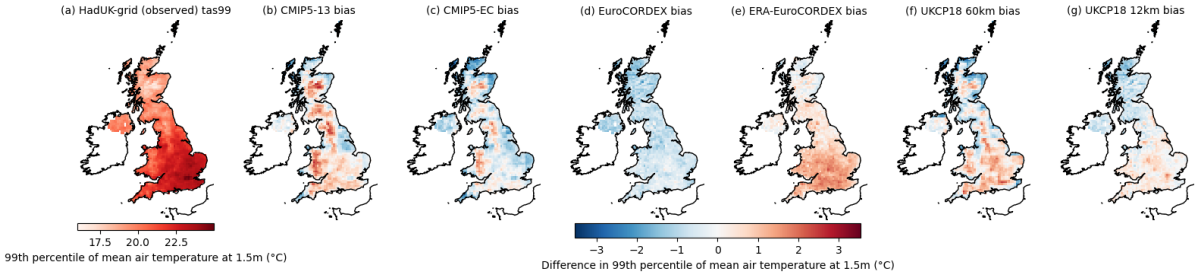


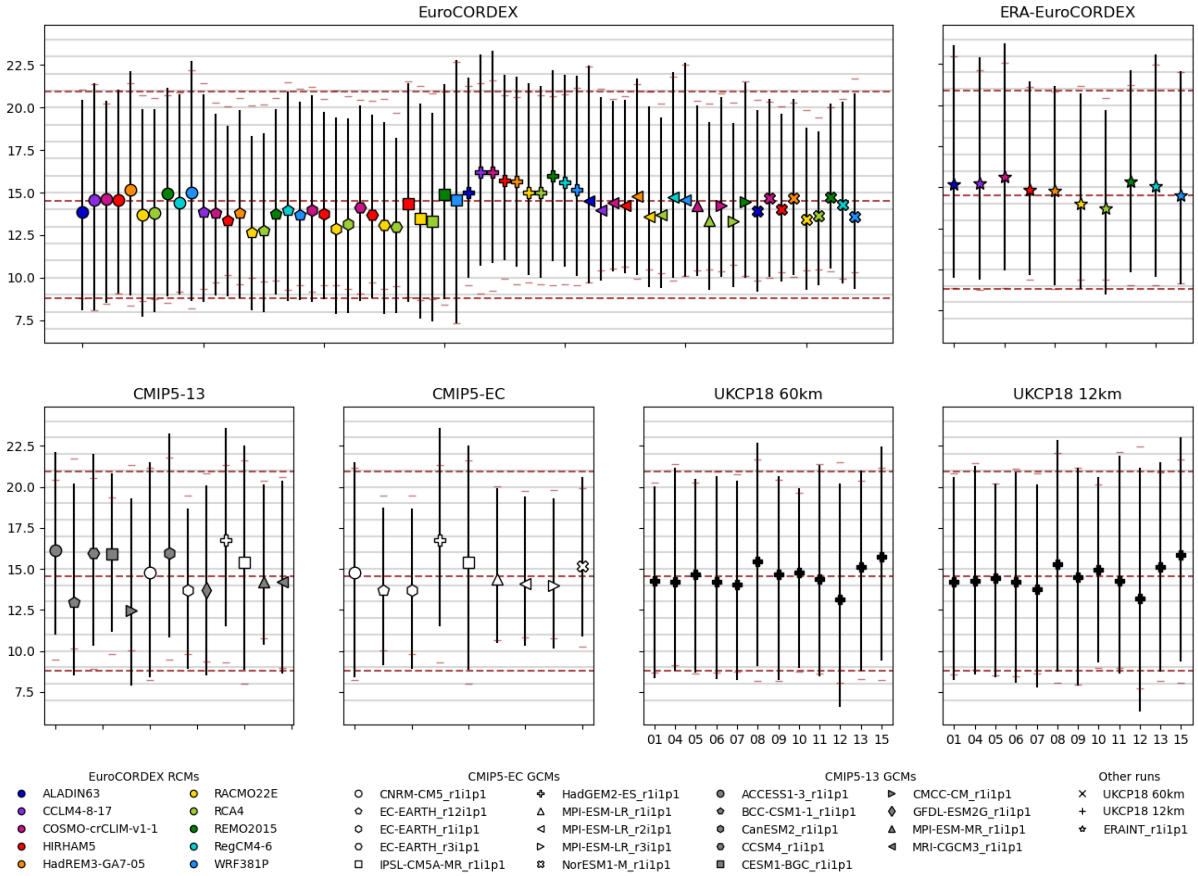
Figure 23 shows the spread of summer temperatures simulated by each run, along with the average daily temperatures. Six of the nine GCM-EC runs fail to capture the full spread of observed summer temperatures; of the runs that do simulate approximately the correct spread of temperatures, IPSL-CM5A-MR (marked by a square) overestimates tas99 by more than 1°C, while the whole distribution of temperatures simulated by HadGEM2-ES is over 2°C warmer than the observed distribution. The CMIP5-13 ensemble contains only three models that underestimate the full range of temperatures, with four of the remaining models displaying systematic biases that shift the entire distribution towards warmer temperatures than those observed by more than 1°C, and two models displaying systematic cold biases of more than 1°C.

The spread of summer temperatures attained by the runs in the EuroCORDEX ensemble depends largely on the GCM: the majority of runs driven by the underdispersive GCMs (EC-EARTH, MPI-ESM-LR, and NorESM1-M) remain too narrow, while those driven by CNRM-CM5, HadGEM2-ES and IPSL-CM5A-MR tend to cover a similar range of temperatures to the HadUK-Grid dataset, albeit with biases in the location of the distribution. While some RCMs — notably RACMO22E and RCA4 — contribute systematic biases to the location of the whole temperature distribution, only WRF381P substantially changes the spread of temperatures simulated, producing somewhat wider ranges of temperatures then are attained by other RCMs downscaling the same GCM. The reanalysis-driven runs generally cover a wider range of temperatures than their GCM-driven counterparts: all of the RCMs simulate tas01 around 6°C below tas, close to the observed range, but tas99 is more than 1°C too high in five of the ten models.

The members of the UKCP18 PPE have very similar temperature distributions at 60km and 12km resolution, suggesting that downscaling has relatively little influence in this model. Most of the runs have biases of less than 0.5°C across the UK and capture the spread of summer temperatures well, although ensemble members 08 and 15 overestimate the hottest temperatures by around 2°C and ensemble member 12 underestimates the coldest by nearly 3°C, leading to much wider temperature distributions in these runs.

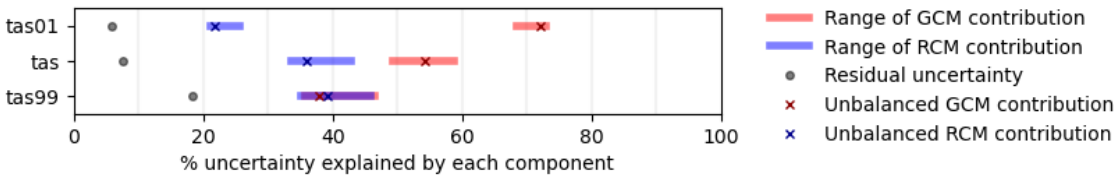
An unbalanced analysis of variance (Figure 24) confirms that the GCM plays a large part in determining the temperatures simulated on cool and average days (tas01 and tas); when simulating the warmest summer days (tas99), the choice of RCM contributes as much to the variability of the

Figure 23: Plots showing the 1st to 99th percentiles of summer daily temperatures (in °C) simulated by each model, averaged over the UK. See caption to Figure 10 for explanation of plots.



ensemble as the choice of GCM, and the residual uncertainty is around 10% higher.

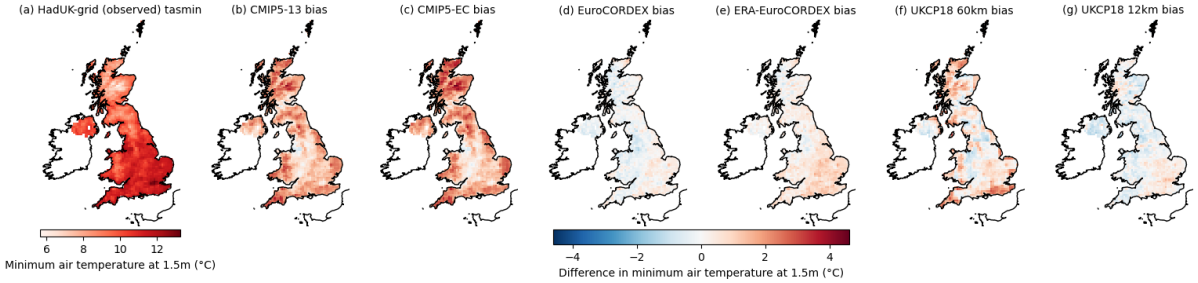
Figure 24: Range of percentages of variance in summer temperature indices within the EuroCORDEX ensemble explained by the driving GCM, by the downscaling RCM, and left unexplained, computed using the method described in Section 3.5.2.



4.6 Diurnal range of summer temperatures (`tasmin` & `tasmax`)

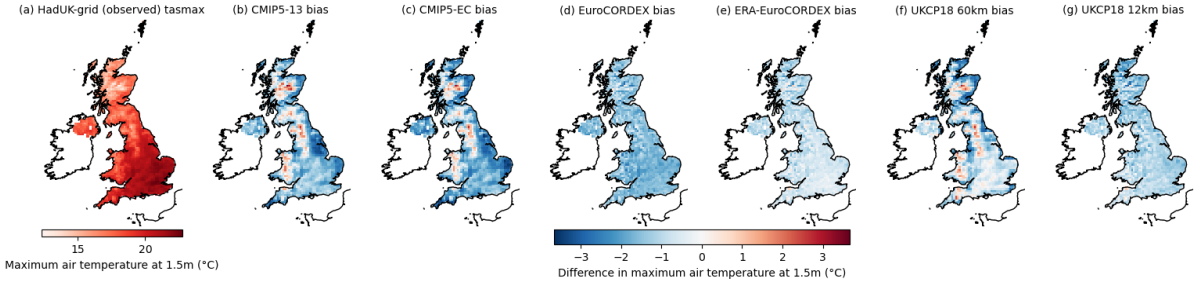
We now turn to consideration of how well the models simulate daily minimum and maximum temperatures, characterised by the variables `tasmin` and `tasmax`: these represent, respectively, the average minimum nighttime temperature and the average maximum daytime temperature. Maps of the observed HadUK-Grid `tasmin`, and the average bias in `tasmin` within each ensemble, are shown in Figure 25.

Figure 25: Maps of (a) HadUK-Grid minimum daily temperature (in °C) during the summer months from 1989 to 2008, and (b-g) of the mean climatological biases in each of the ensembles of models.



Both of the CMIP5 ensembles are too warm at night over almost the whole of the UK, with the largest biases once again occurring at higher elevations and around the coast. The UKCP18 60km ensemble shares these biases due to unresolved features, but to a lesser extent, and is too cool across low-lying parts of central and northern England. Of the 12km-resolution ensembles, the EuroCORDEX and UKCP18 12km ensembles have close to zero bias across much of the UK, while the reanalysis-driven EuroCORDEX is slightly too warm overall. The corresponding plots for tasmax in Figure 26 show that, with the exception of localised warm biases at the highest elevations in the CMIP5 and 60km ensembles, maximum daytime temperatures are systematically underestimated across the UK by all of the ensembles

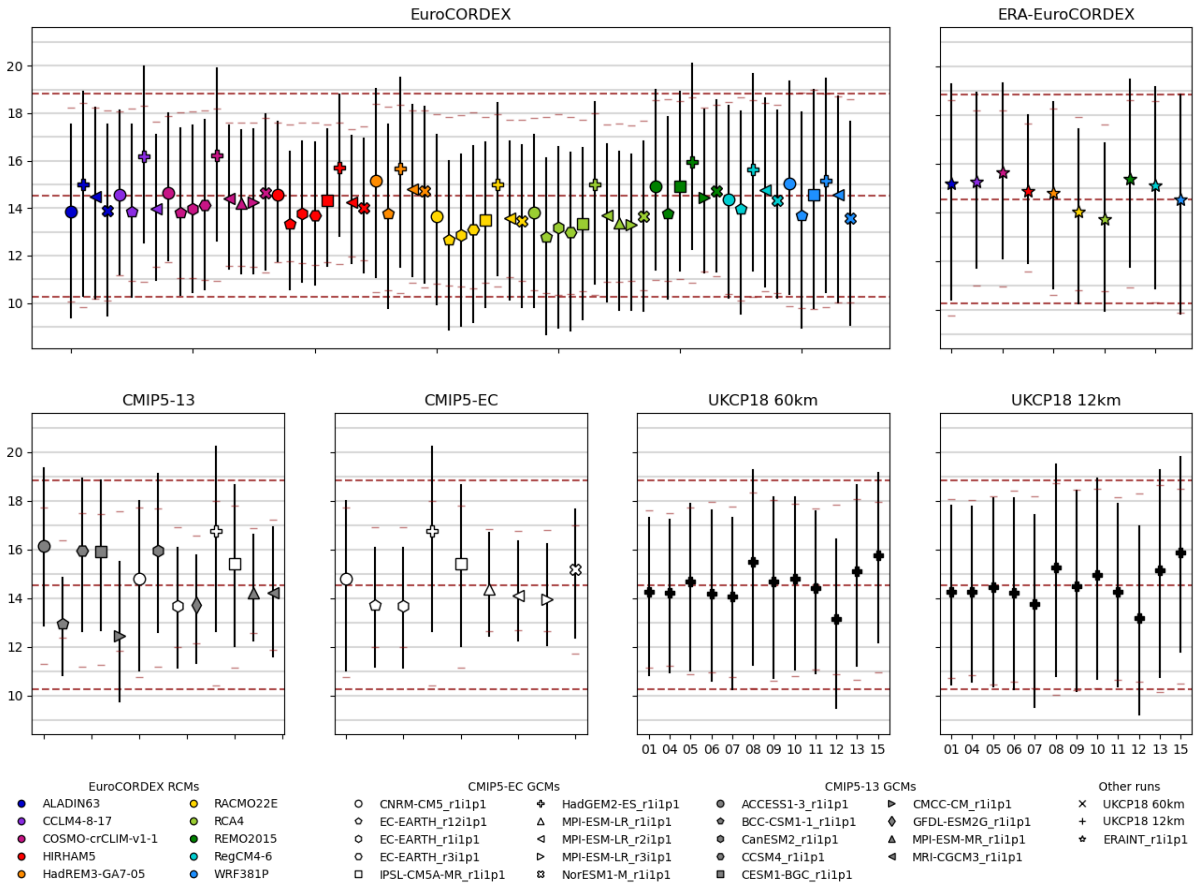
Figure 26: Maps of (a) HadUK-Grid maximum daily temperature (in °C) during the summer months from 1989 to 2008, and (b-g) of the mean climatological biases in each of the ensembles of models.



The ranges of average diurnal temperatures over the UK for each model are shown in Figure 27. In the HadUK-Grid observations, the average of tasmin is 10.3°C; of tas, 14.5°C; and of tasmax, 18.8°C: the average diurnal temperature range is therefore around 8.5°C. All of the GCMs underestimate this diurnal range: those runs that simulate values of tasmax greater than the observed value or tasmin lower than the observed value do so only because of large systematic biases that shift the whole temperature distribution. This is also true of the majority of the reanalysis-driven ERA-EuroCORDEX runs, although the diurnal ranges of temperatures in those runs are wider, and in four of the ten models both tasmin and tasmax are close to the observed values. Members of the EuroCORDEX ensemble inherit biases from their driving GCMs — for example, all runs driven by EC-EARTH are too cool, and all runs driven by HadGEM2-ES are too warm — but as with winter temperatures, the width of the intervals between tas and tasmin and, particularly, between tas and tasmax is more similar for all runs using the same RCM than for runs using the same GCM. In

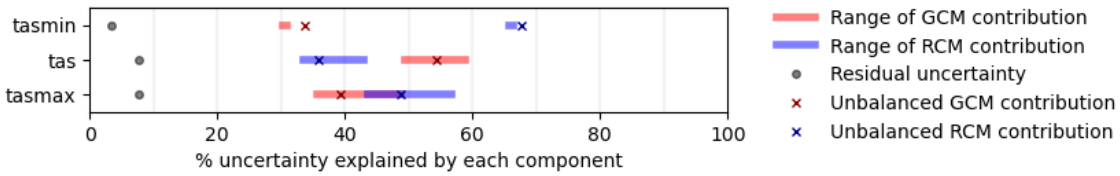
the UKCP18 ensemble, while each run's mean temperature bias is largely inherited from the 60km model, the 12km members have slightly wider diurnal temperature ranges that more closely approach the observed range, although most members underestimate the daytime temperatures tasmax, and runs 07, 08, 12 and 15 suffer from systematic biases of more than 1°C that affect the whole of the diurnal temperature range.

Figure 27: Plots showing the typical range of daily temperatures simulated in summer by each model, averaged over the UK. See caption to Figure 10 for explanation of plots.



The analysis of variance described in Section 3.5.2 was again carried out over the full EuroCORDEX ensemble, including the evaluation runs. The proportion of the variability between the runs that can be explained by the systematic differences attributed to the GCMs and RCMs is summarised in Figure 28. As already discussed in Section 4.4, differences between runs driven by different GCMs contribute slightly more (49-59%) than differences between runs downscaled by different RCMs (35-43%) to the variability in the mean temperature tas. The RCM contributes almost two-thirds of the total variance in tasmin, while the ranges of proportions of variance in tasmax contributed by the two model components — which, together, account for 92% of the total variation — are wide enough to overlap, suggesting that neither can be identified definitively as a dominant source of variation.

Figure 28: Range of percentages of variance within the EuroCORDEX ensemble explained by the driving GCM, by the downscaling RCM, and left unexplained, computed using the method described in Section 3.5.2.



5 Evaluation of precipitation

This section evaluates the distribution of precipitation simulated by each model in the winter and summer months during the evaluation period from January 1st 1989 to December 31st 2008. For each season, we begin with a discussion of each ensemble's mean daily precipitation rate, obtained by averaging the daily precipitation rate `pr`, and how this is affected by biases in both the mean wet-day precipitation rate (also known as the standardised daily intensity index, `sdii`) and the proportion of wet days, `fwd`. The proportion of precipitation occurring on very wet days, `r99ptot`, is also discussed: this is the proportion of the total seasonal precipitation falling on days on which the precipitation rate exceeds the 99th percentile of daily precipitation rates for that season. These quantities are summarised in Table 5.

Table 5: Indices used to characterise the seasonal distribution of daily precipitation

Index	Description
`pr`	Average daily precipitation rate `pr` (mm/day)
`sdii`	Average daily precipitation rate on wet days (mm/day)
`fwd`	Percentage of days on which `pr` exceeded 1mm/day
`r99`	99th percentile of daily precipitation rate `pr`
`r99ptot`	Percentage of total precipitation falling on days on which `pr` exceeded `r99`

Daily precipitation rates are reported in preference to seasonal accumulated rainfall due to the different calendar lengths used by the various models. All biases are reported as a percentage error relative to the HadUK-Grid observations, for easier comparison and interpretation, unless otherwise noted.

5.1 Winter precipitation

Figure 29 shows a map of the HadUK-Grid winter mean daily precipitation rate (panel a), together with the average of the relative deviations from that rate in each of the ensembles (panels b-g). The mean daily precipitation rate is highest in west-facing areas of high elevation, with the heaviest rainfall concentrated in western Scotland. All of the ensembles tend to underestimate the total precipitation in the wettest areas and to overestimate it in the driest areas, with the exception of the UKCP18 12km ensemble, which is slightly too wet everywhere except for the northwest of Scotland. Corresponding maps of the observed wet-day precipitation rate `sdii` and the proportion of wet days `fwd` are shown in Figures 30 and 31: comparison of these with Figure 29 show that the dry biases

in pr in the wettest areas generally correspond to an underestimation of the wet-day precipitation rate, while the wet biases in the drier areas are related to simulation of too many wet days. This latter bias is minimised in the reanalysis-driven ERA-EuroCORDEX ensemble, suggesting that it is to some extent driven by the global models.

Figure 29: Maps of (a) daily mean precipitation rate pr during the winter months from 1989 to 2008 in HadUK-Grid (mm/day); (b-g) average relative biases in mean precipitation rate in each ensemble.

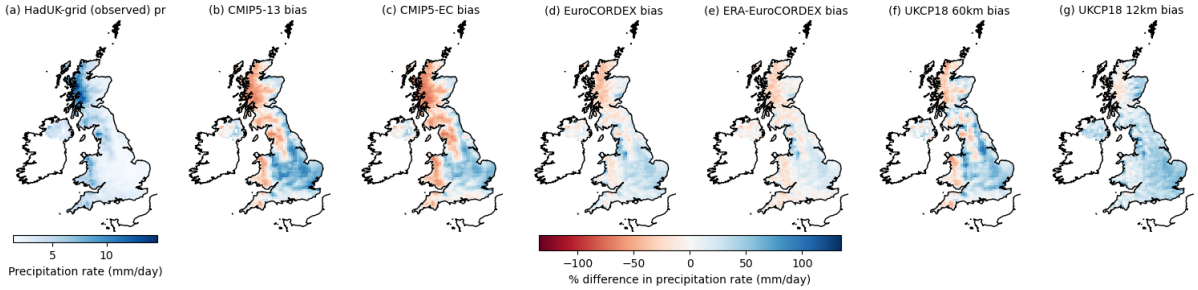


Figure 30: Maps of standardised daily intensity index sdii during the winter months from 1989 to 2008: (a) HadUK-Grid winter mean wet-day precipitation in mm/day; (b-g) average relative biases in mean wet-day precipitation rate in each ensemble.

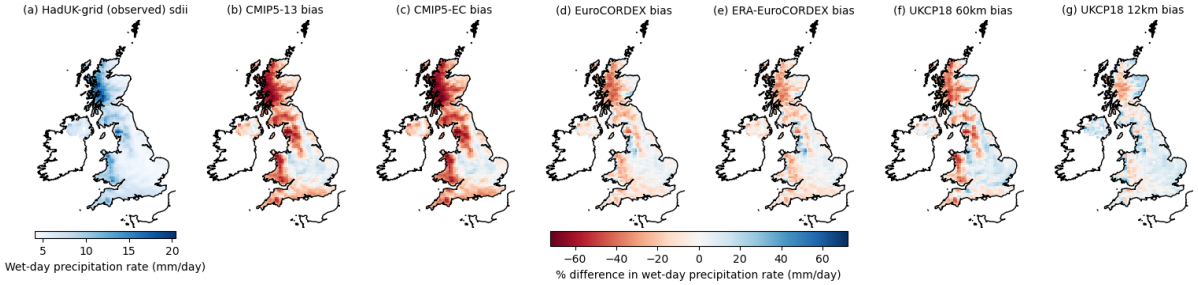
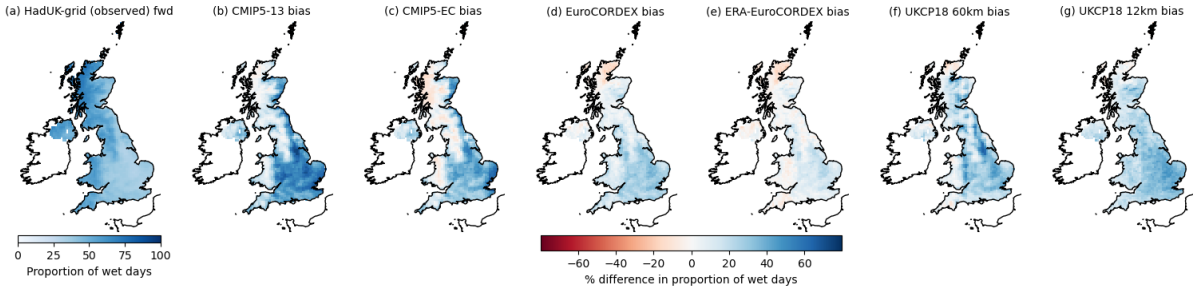


Figure 31: Maps of proportion of wet days fwd during the winter months from 1989 to 2008: (a) HadUK-Grid proportion of days with pr $> 1mm/day$; (b-g) average relative biases in proportion of wet days in each ensemble.



The distributions of relative biases in UK-averaged pr, sdii and fwd within each ensemble are shown in Figure 32. Within the CMIP5-13 ensemble there are two clusters of models, one containing four runs that overestimate the mean precipitation rate pr by up to 20%, and one containing seven runs that underestimate by 20-25%; the CMIP5-EC ensemble is dominated by models from the latter group. However, both ensembles display a similar, fairly narrow range of biases in sdii (panel b): the difference in the overall average rate pr is driven largely by differences in the proportion of wet

days simulated (panel c), where the two groups of GCMs are more clearly separated.

The pattern of relative biases in fwd in the CMIP5-EC ensemble is reflected, to an extent, in the EuroCORDEX ensemble: for example, for runs produced by any given RCM, those driven by IPSL-CM5A-MR and NorESM1-M tend to have the largest relative biases in fwd. However, the RCM clearly also plays a substantial part in driving these biases, with the EuroCORDEX runs more closely grouped by colour (denoting the RCM) than by shape (denoting the GCM), and with the reanalysis-driven ERA-EuroCORDEX runs having almost as wide a spread of biases as the CMIP5-EC ensemble. Biases in the rate of precipitation on wet days, sdii, also depend largely on the choice of RCM, being closely grouped by colour. The biases exhibited by the CMIP5-EC ensemble members are not reflected in the biases of runs driven by the CMIP5 output, although the driving GCM does affect the ordering of the biases to some extent, with runs driven by NorESM1-M (represented by crosses) being wetter (relative to HadUK-Grid) than runs using the same RCM but driven by CNRM-CM5 (represented by circles). While the relative biases in fwd and sdii are positively correlated, most of the EuroCORDEX runs overestimate the number of wet days and underestimate the rate of precipitation on wet days, and as a result, the spread of relative biases in overall precipitation pr is wide, ranging from -45 to +50%.

The UKCP18 runs have a much smaller range of relative biases in all three precipitation indices than the EuroCORDEX ensemble, with all of the regional ensemble members having average wet-day precipitation rates within 10% of the observed rate, and all but two of the regional ensemble members simulating 20-30% more wet days than were observed across the UK: the overall precipitation rate is 15-30% too high in all but the two driest runs, largely as a result of the bias in fwd. The rate of precipitation on wet days is somewhat lower across the UK in the 60km runs, which are around 20% drier than the 12km runs overall as a result.

Figure 32: Boxplots showing UK-averaged relative biases in winter precipitation indices in each ensemble.

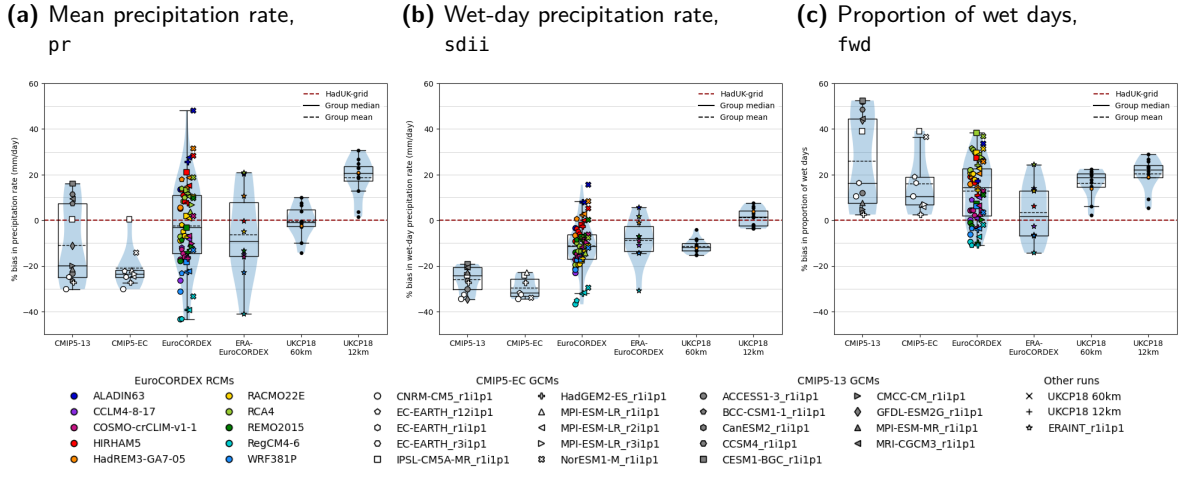
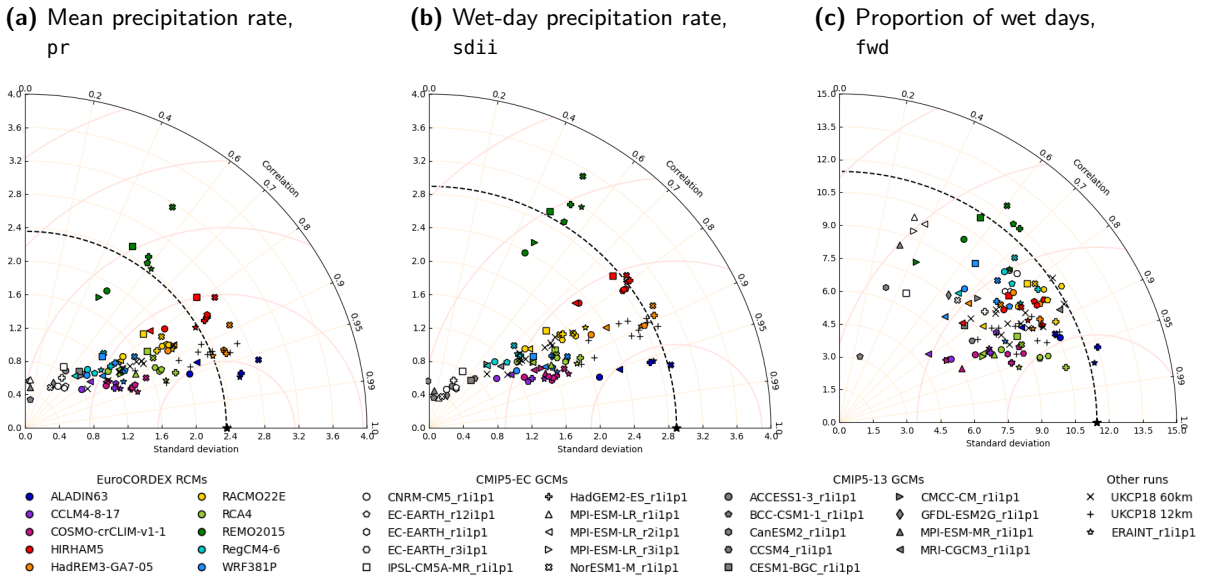


Figure 33 summarises how well each of the models captures the spatial patterns of observed precipitation. The patterns of scores in panels a and b are very similar: almost all of the runs underestimate the amplitude of the spatial patterns, lying to the left of the dashed reference arc that indicates the

standard deviation of the observed daily precipitation rates; generally, correlation with the observed patterns increases as the standard deviation approaches the correct level, with the GCMs achieving both the lowest correlation scores and the lowest standard deviations. The members of the EuroCORDEX ensembles are closely grouped by colour (i.e. by RCM); the ordering of the models from lowest to highest standard deviation approximately reflects the ordering of the biases in the ERA-RCM ensemble in Figure 32a, suggesting that the UK-averaged biases may be due more to underestimation of the spatial variability than to a simple offset of the type seen in the temperature distributions. One exception to this pattern is the group of runs downscaled by REMO2015 (shown in dark green); this RCM produces patterns of precipitation that are poorly correlated with those observed, but which capture the variability of the precipitation better than might be expected, given that it also underestimates the UK-wide precipitation rate. The UKCP18 60km ensemble members, indicated by $\times$, fall among the EuroCORDEX ensemble members with the lowest standard deviations, while the 12km runs indicated by $+$ symbols are grouped with the HadGEM3-GA7-05 runs (coloured orange), and are among the runs with the highest standard deviations and the highest correlations.

The pattern in the spatial distribution of `fwd`, shown in panel c, is less clear. Again, most of the runs underestimate the spatial variability in the observed pattern, but there is no clear relationship between spatial variability and correlation. Although the EuroCORDEX runs are loosely grouped by colour, reflecting the role of the downscaling RCM in simulating wet days, there is no clear pattern of skill improving with increased resolution: several of the low-resolution GCM runs (indicated by grey or white shapes) appear in the midst of the EuroCORDEX ensemble, along with the UKCP18 runs at both resolutions.

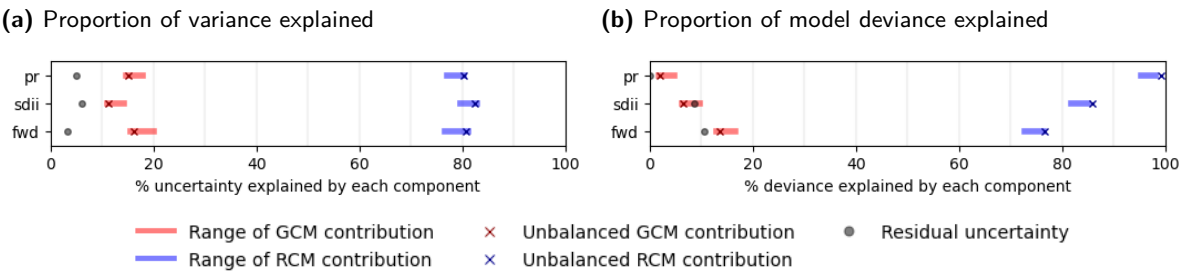
Figure 33: Taylor diagrams comparing model climatologies of winter precipitation indices to HadUK-Grid climatology



In order to gain a better understanding of the structural uncertainties within the EuroCORDEX ensemble, analysis of variance was carried out over the full dataset, using the methodology described

in Section 3.5.2, along with an analysis of deviance, as described in Section 3.5.3. The proportions of uncertainty and deviance in each of the precipitation indices explained by the GCM and RCM effects are shown in Figure 86. Systematic differences between the RCM effects contribute between 75 and 80% of the variation between the runs in all three indices, and almost 100% of the model deviance in pr, while differences between the GCM effects account for between 10 and 20% of the total variation.

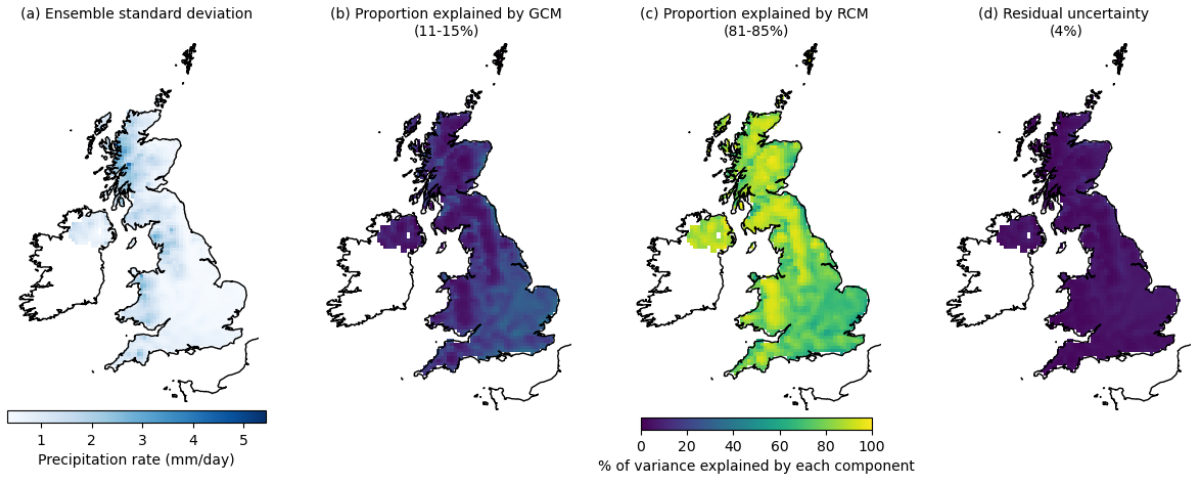
Figure 34: Range of percentages of variance and model deviance within the EuroCORDEX ensemble explained by the driving GCM, by the downscaling RCM, and left unexplained, computed using the method described in Sections 3.5.2 and 3.5.3.



Maps of the standard deviation of the daily precipitation rate pr simulated by the members of the EuroCORDEX ensemble (including the runs driven by reanalysis) and of the contributions of the GCM and RCM components to the variability within the ensemble are shown in Figure 35. Precipitation rates within the ensemble are more variable in the wettest areas (panel a), and the proportion of local variance explained by the RCMs in these areas and in Northern Ireland is close to 100%. In drier areas, a slightly lower proportion (60-80%) of the variance is attributed to the RCMs, with the GCMs contributing up to 40%. This is to be expected given that the RCMs calculate cloud properties and rainfall, but the GCMs determine the occurence of weather conditions leading to cloudless skies.

The contributions of the GCM and RCM effects shown in Figure 35 can be further decomposed into their EPPs. Of the approximately 80% of the total variability in the biases in the mean daily precipitation rate pr that can be ascribed to the choice of RCM, 62% is captured by the pattern labelled 'RCM EPP1' in Figure 36a. This pattern represents a fairly uniform relative bias across the UK, with slightly larger biases at higher elevations: any runs downscaled using an RCM with a positive (resp. negative) first EPP score in 36a will tend to be wetter (resp. drier) relative to the HadUK-Grid climatology than the ensemble average in these areas. The second pattern, labelled 'RCM EPP2', describes 20% of the variability between the fitted RCM effects: runs using RCMs with positive scores in this component will tend to simulate less precipitation at higher elevations and in the rain shadows of those higher elevations, and more precipitation to the west of those areas. One RCM — REMO2015, which also stood out as having particularly poor correlation with the observed precipitation field in Figure 33a — receives a particularly large score in this component: this suggests that this model may simulate too much rain at lower elevations, relative to the HadUK-Grid climatology.

Figure 35: Maps of (a) the standard deviation of mean winter precipitation rates in the EuroCORDEX ensemble (including evaluation runs driven by ERA-Interim reanalysis), and (b-d) of the percentage of the total ensemble variance in each grid cell explained by the choice of GCM and RCM. The ranges of proportions of the total variance explained by each component, computed using the methodology described in Section 3.5.2, are given in parentheses.



The GCM fitted effects contribute around 20% to the total variability within the EuroCORDEX ensemble; 69% of this variability (14% of the total) is associated with the pattern labelled 'GCM EPP1' in Figure 36b, with runs driven by GCMs having a positive score in this component having larger relative biases than the ensemble average across much of England, where winter precipitation rates are typically lower (Figure 29a). The second component represents lower rainfall than the ensemble average in the east of the UK, and slightly higher in the wettest areas to the west.

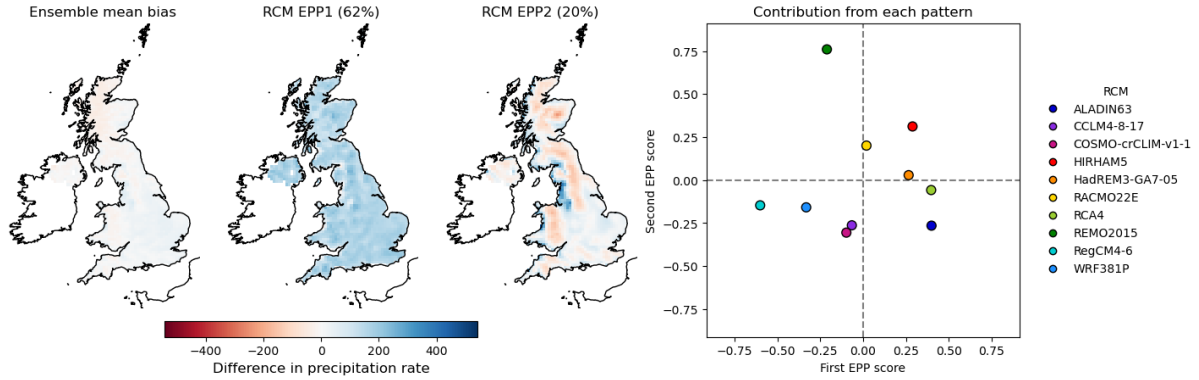
An EPP analysis was also carried out to identify the dominant patterns of variation between the UKCP18 regional runs (Figure 37). 46% of the variation between the individual runs can be attributed to the pattern labelled 'EPP1', with runs with positive scores in this component simulating more precipitation in rain shadows than the ensemble average, and close to the ensemble average in the wettest areas along the west coast. The second component, which accounts for 25% of the total variability, represents a contrast between higher-than-average precipitation in the wettest areas of the UK and lower-than-average precipitation in rain shadows and in northeastern Scotland.

5.2 Summer precipitation

Figure 38 shows a map of the HadUK-Grid mean daily precipitation rate `pr` during the summer months, together with the average relative bias in each of the ensembles with respect to the HadUK-Grid climatology. The lower-resolution GCM runs (panels b and c) are too dry in the wettest areas and around the south coast due to unresolved orography and blurring of the land-sea boundary, and otherwise slightly too wet; the EuroCORDEX ensemble (panel d) simulates too much summer precipitation over much of the UK, while the reanalysis-driven ERA-EuroCORDEX ensemble mean is slightly too dry in the very wettest areas and in East Anglia, but also slightly too wet elsewhere (panel e). This pattern of biases is shared by the UKCP18 12km ensemble (panel g), while the 60km

Figure 36: EPP analysis of relative biases in winter precipitation rates from 1989-2008, showing dominant patterns of contributions from each GCM and RCM to variation about the mean bias in the Euro-CORDEX ensemble. The proportion of the between-RCM or between-GCM variation attributed to each pattern is given in parentheses.

(a) Contribution of each downscaling RCM to EuroCORDEX ensemble spread



(b) Contribution of each driving GCM to EuroCORDEX ensemble spread

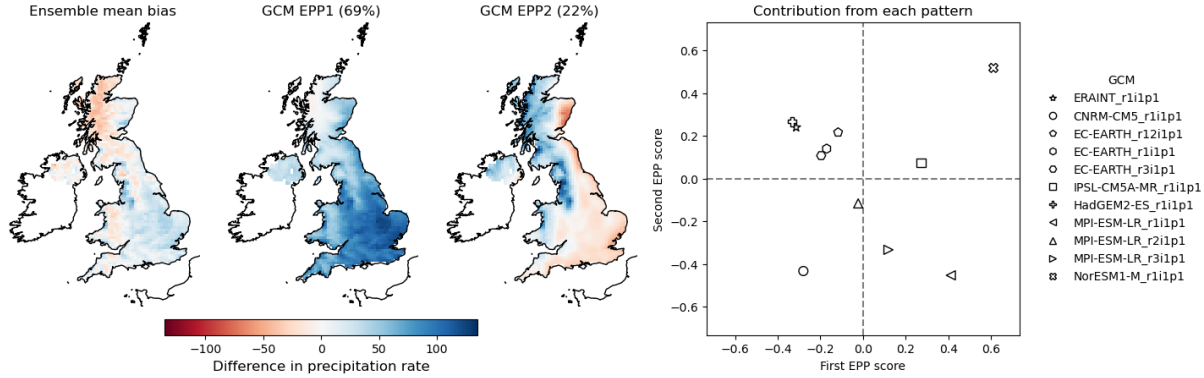
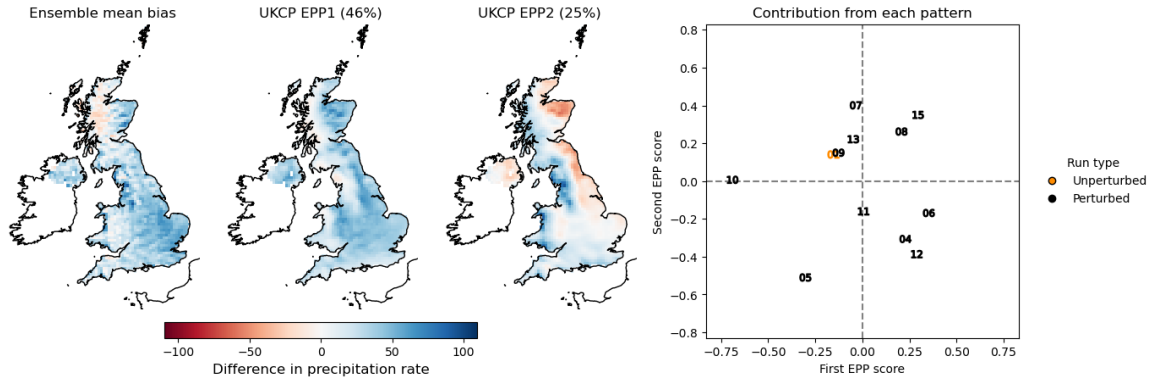


Figure 37: EPP analysis of winter precipitation rates simulated by the UKCP18 regional models, showing the dominant patterns of variation. The proportion of the total variation attributed to each pattern is given in parentheses.



ensemble is too dry across much of England (panel f).

Maps of the relative biases in mean wet-day precipitation rate `sdii` and of the proportion of wet days `fwd`, which contribute to these biases in `pr`, are presented in Figures 39 and 40. All of the

ensembles simulate too little precipitation on wet summer days across most of the UK (Figure 39), while the three regional ensembles simulate too many wet days on average (Figure 39); the CMIP5 ensembles simulate too many wet days across most of the UK but too few in the wettest parts of Cornwall, west Wales and the Lake District, while the UKCP18 60km ensemble simulates too many wet days in rain shadows, and too few elsewhere.

Figure 38: Maps of (a) daily mean precipitation rate `pr` during the summer months from 1989 to 2008 in HadUK-Grid(mm/day); (b-g) average relative biases in mean precipitation rate in each ensemble.

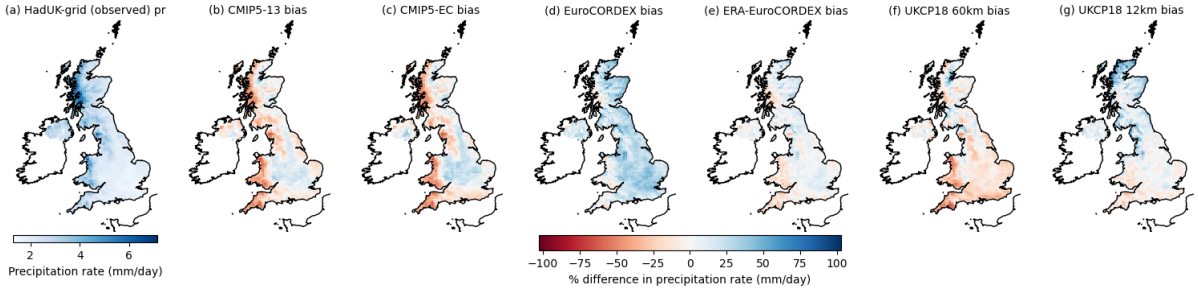


Figure 39: Maps of (a) mean wet-day precipitation rate `sdii` during the summer months from 1989 to 2008 in HadUK-Grid (mm/day); (b-g) average relative bias in mean wet-day precipitation rate in each ensemble.

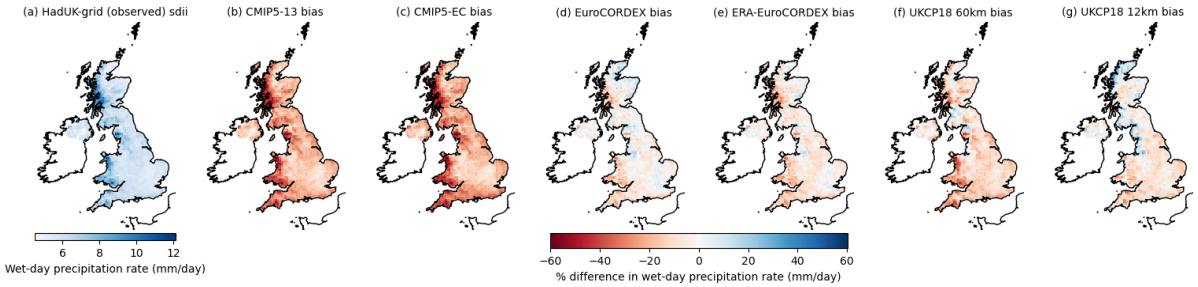
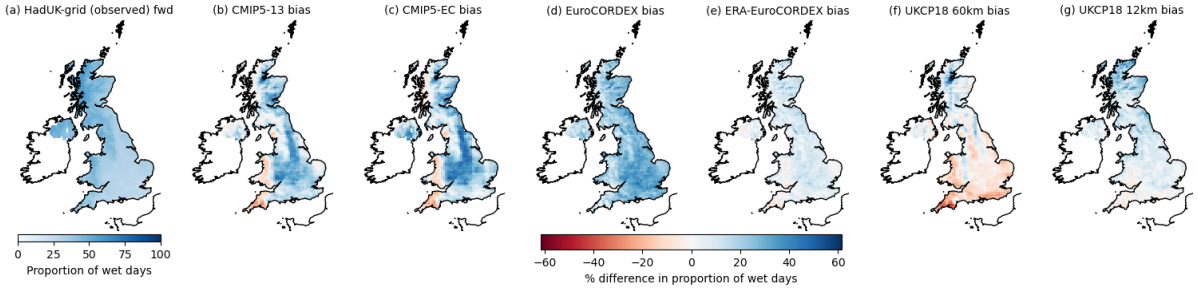


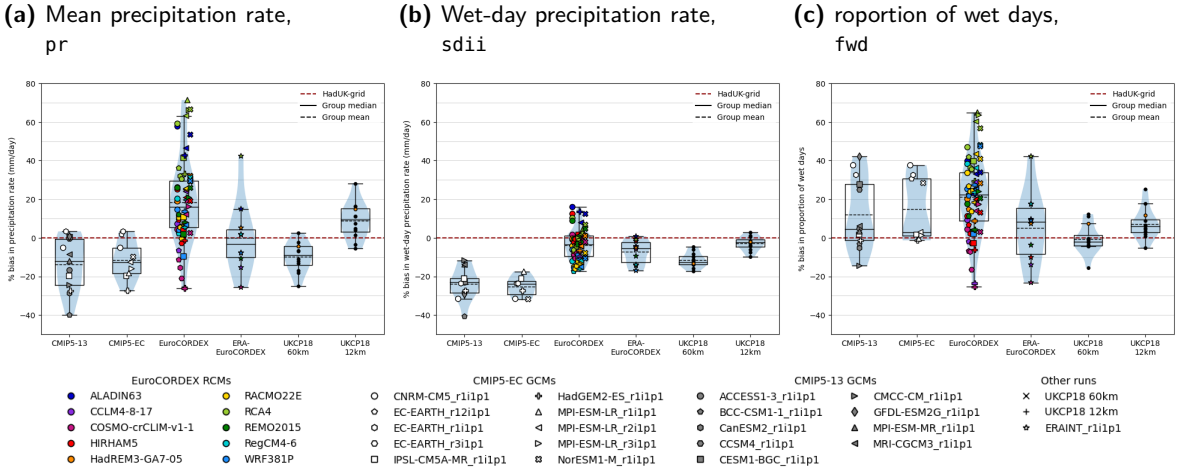
Figure 40: Maps of (a) proportion of wet days `fwd` during the summer months from 1989 to 2008 in HadUK-Grid (%); (b-g) average relative bias in proportion of wet days in each ensemble.



The distributions of UK-averaged relative biases in summer `pr`, `sdii` and `fwd` in each ensemble are shown in Figure 41. The GCMs all have similar relative biases in `sdii` (panel b), underestimating wet-day precipitation rates by 15-35% on average across the UK, but fall into two clusters in panel c, one simulating close to the correct proportion of wet days, and one group simulating 25-45% more wet days than were observed: the members of the two clusters do not correspond exactly to those seen in Figure 32c. Relative biases in the CMIP5 runs are not closely related to relative

biases in the EuroCORDEX runs, which are loosely grouped according to the downscaling RCM: the range of relative biases in `sdii` is narrow compared to the range of biases in `fwd`, and the pattern of relative biases in `pr` closely resembles the pattern in `fwd`, suggesting that the proportion of wet days simulated contributes more to overall precipitation in summer than the rate of precipitation on wet days. The spread of relative biases in the two UKCP18 ensembles is much smaller than the spread of biases in the EuroCORDEX ensembles: all of the global ensemble members simulate too little rain across the UK on wet days but approximately the correct number of wet days, while most of the regional ensemble members simulate close to the observed average rate of wet-day precipitation across the UK and a slightly higher proportion of wet days.

Figure 41: Boxplots showing the UK-averaged biases in summer precipitation indices within each ensemble.



The Taylor diagrams in Figure 42 summarise how well each of the models captures the spatial patterns of observed summer precipitation. Patterns in the wet-day precipitation rate (panel b) are generally less well represented in the models than patterns in the proportion of wet days (panel c) or overall precipitation (panel a), with all runs receiving lower correlation scores in `sdii` than in `pr` or `fwd`. In all three panels the EuroCORDEX runs are loosely grouped by RCM, with runs downscaled by two models (REMO2015 and WRF381P) receiving lower correlation scores than the low-resolution GCM output, and often of less than 0.5. The UKCP18 runs at both resolutions fall within the EuroCORDEX ensemble, generally close to the runs downscaled by HadREM3-GA7-05 (marked with orange symbols), with the 12km runs (indicated by $+$) having slightly higher correlation scores than the 60km runs (marked as $\times$), but tending to overestimate the amplitude of spatial variability.

An analysis of variance over the full EuroCORDEX ensemble including the evaluation runs confirms that, for all three precipitation indices, systematic differences between the fitted RCM effects contributes most of the variation in the EuroCORDEX ensemble, although slightly less than in the case of winter temperatures (between 66 and 70% of the total variability in `pr` depending on the order of model fitting, compared to 81-85% for the winter precipitation rate). Multivariate ANOVA found that the RCMs accounts for 71-75% of the total model deviance when interactions between locations are taken into account, and the GCMs for 11-15% of the total deviance.

Figure 42: Taylor diagrams comparing model climatologies of summer precipitation indices to HadUK- Grid climatology

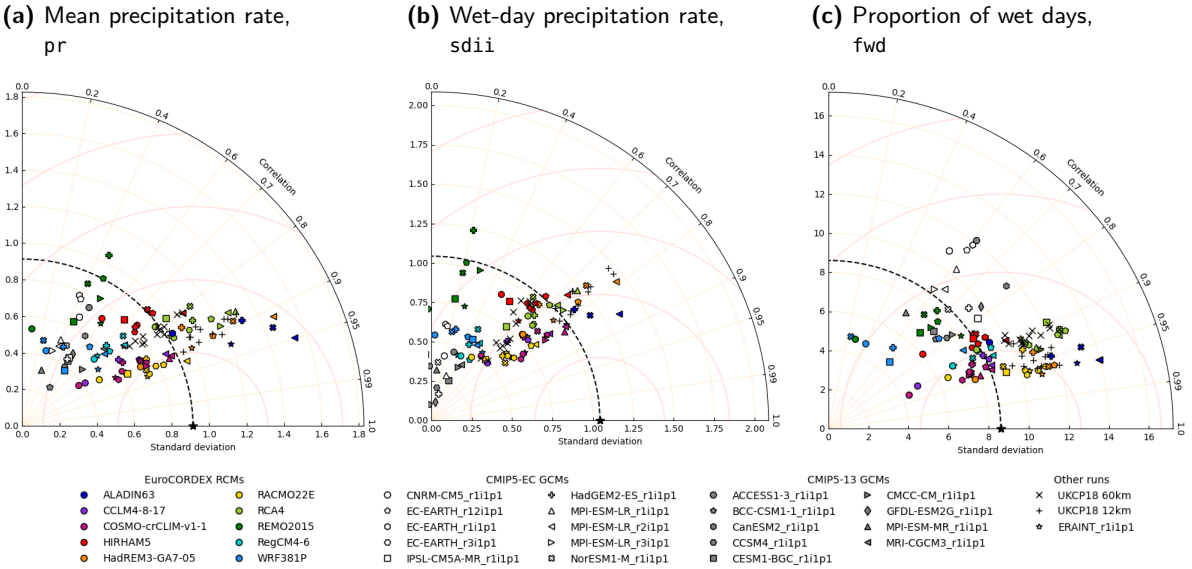
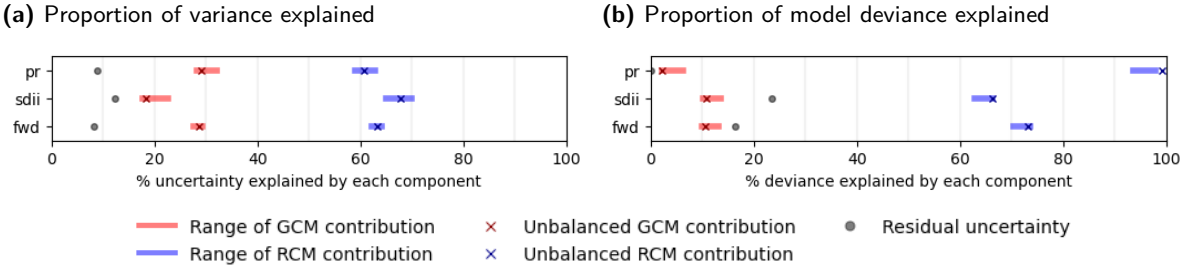


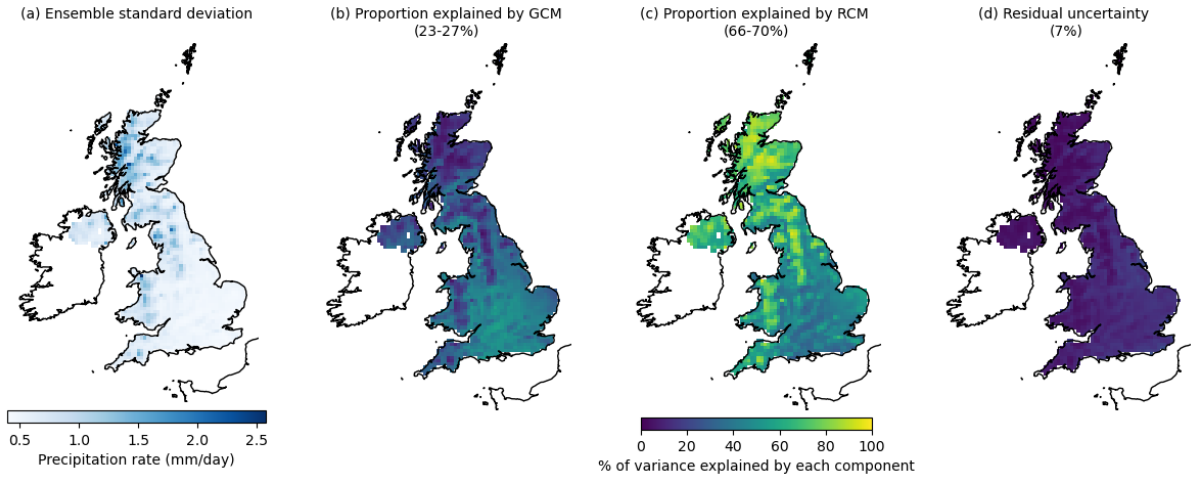
Figure 43: Range of percentages of variance and model deviance within the EuroCORDEX ensemble explained by the driving GCM, by the downscaling RCM, and left unexplained, computed using the methods described in Sections 3.5.2 and 3.5.3.



Maps of the ensemble standard deviation in relative biases in summer pr, and of the proportion of this variance explained by the GCM and by the RCM in each grid cell, are shown in Figure 44: these are very similar to the corresponding maps for winter precipitation, shown in Figure 35, with the RCMs accounting for almost 100% of local variability in the wettest areas and at higher elevations, where the greatest variability occurs; in low-lying areas, where the prevailing weather regime may play a greater part in determining the occurrence and intensity of rainfall, the choice of GCM contributes up to 60% of the local variation.

The principal spatial patterns of the contributions of the RCM and GCM effects to variability of the relative biases within the EuroCORDEX ensemble can be identified using EPP analysis. Figure 36a shows the principal patterns of variability in relative biases in pr associated with the RCMs, which account for around 70% of the total variance. Positive scores in the first component, which accounts for 67% of the variation attributed to the RCM effects, correspond to a higher relative precipitation bias than the ensemble average across the UK, with larger relative biases at higher elevations; these scores are strongly correlated with the UK-averaged model biases in Figure 41a.

Figure 44: Maps of (a) the standard deviation of relative biases in the summer precipitation climatologies in the EuroCORDEX ensemble (including evaluation runs driven by ERA-Interim reanalysis), and (b-d) of the percentage of the total ensemble variance in each grid cell explained by the choice of GCM and RCM. The ranges of proportions of the total variance explained by each component, computed using the methodology described in Section 3.5.2, are given in parentheses.



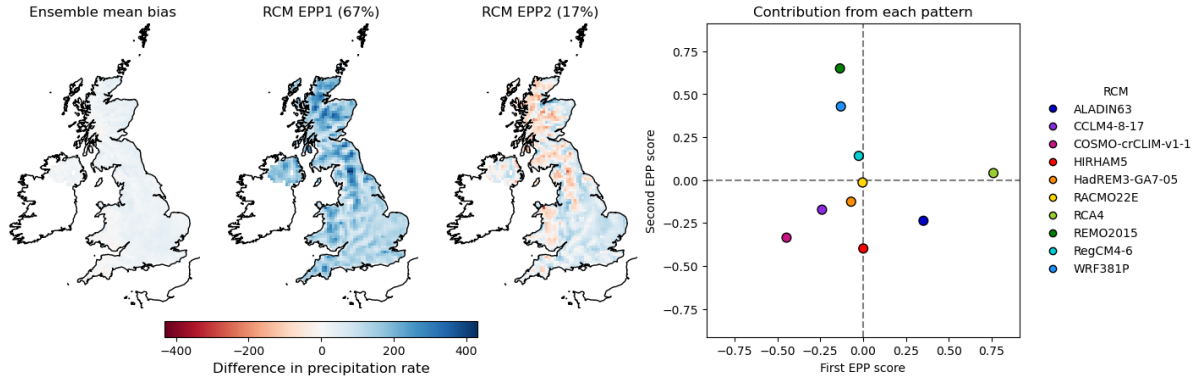
A positive score in the second component, which accounts for 17% of the between-RCM variation, indicates lower-than-average precipitation at high elevations and higher-than-average precipitation at lower elevations, reducing the difference in the relative biases in pr at high and low elevations, while a negative score will increase the difference further. REMO2015 and WRF381P, the RCMs producing runs with the poorest correlation with the observed spatial patterns in Figure 42, have particularly high scores in the second component.

The corresponding plots of the GCM components — which account for a much smaller proportion of the variation in these precipitation indices — are shown in Figure 45b. 84% of the variation attributed to the GCM effects is associated with the first EPP, which represents higher-than-average relative biases in precipitation across the UK, particularly in central and southern England. Almost identical maps and scores were obtained for the fitted GCM effects in fwd (not shown here), again suggesting that the choice of GCM affects the summer precipitation rate largely through the proportion of wet days simulated, rather than through the amount of precipitation simulated on wet days.

EPPs were also calculated for the UKCP 12km runs, in order to identify the dominant spatial patterns of differences between their climatologies (Figure 46). 72% of the variation within the UKCP18 regional ensemble can be attributed to the first EPP, with runs with positive (negative) scores in this component tending to be wetter (drier) than the ensemble average across the UK, but particularly at low elevations and in urban areas: both London and Birmingham are clearly visible as darker patches, suggesting that the UKCP18 runs differ in their treatment of urban precipitation. The second component, which accounts for 11% of the variation, represents higher-than-average precipitation around the coast of Wales, Scotland and northern England, but lower-than-average precipitation elsewhere.

Figure 45: EPP analysis of relative biases in summer precipitation rates from 1989-2008, showing dominant patterns of contributions from each GCM and RCM to variation about the mean bias in the Euro-CORDEX ensemble. The proportion of the between-RCM or between-GCM variation attributed to each pattern is given in parentheses.

(a) Contribution of each downscaling RCM to EuroCORDEX ensemble spread



(b) Contribution of each driving GCM to EuroCORDEX ensemble spread

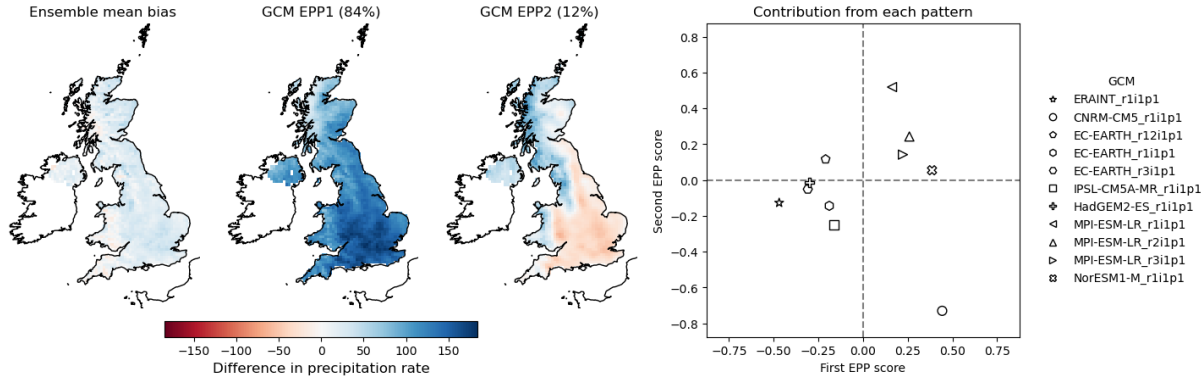
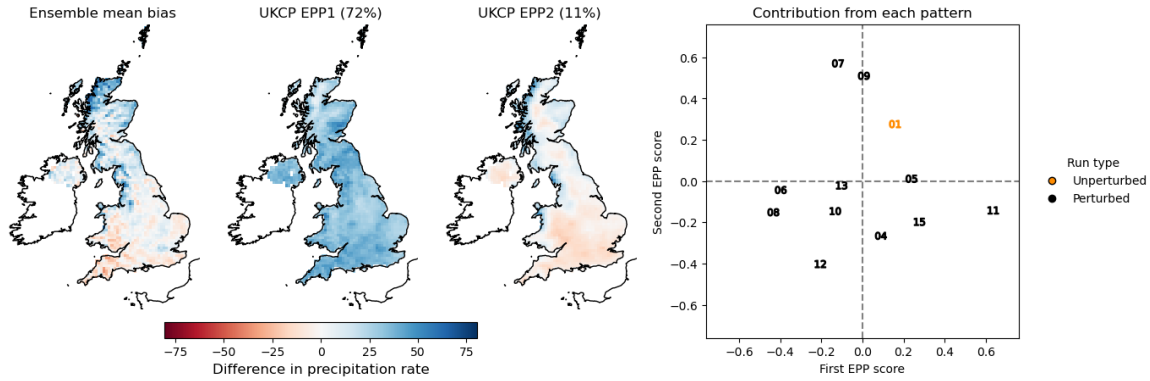


Figure 46: EPP analysis of summer precipitation simulated by the UKCP18 regional models, showing the dominant patterns of variation. The proportion of the total variation attributed to each pattern is given in parentheses.



5.3 Extreme precipitation

The proportion of precipitation falling during intense rainfall events is captured through `r99ptot`, the proportion of the total precipitation falling on days on which the precipitation rate exceeding

the 99th percentile of daily wet-day precipitation rates. Maps of r99ptot in HadUK-Grid, and of the absolute mean biases in r99ptot in each ensemble, during the winter and summer months are shown in Figures 47 and 48. In both 47a and 48a, localised areas of particularly high contributions from the wettest days are visible: these correspond to individual storm events during the evaluation period, which are not simulated by the model runs, and consequently appear as corresponding highly localised dry biases in the ensemble means. Disregarding the areas affected by these individual events in the HadUK-Grid dataset, the two CMIP5 ensembles generally generate too little of their rainfall on the wettest days in both summer and winter, while the EuroCORDEX and UKCP18 ensembles tend to simulate too high a proportion of their rainfall on these wetter days. The exception to this is the UKCP18 60km ensemble, which simulates too little of its rainfall on the wettest days in the regions to the west of the UK that typically receive the most precipitation overall.

Figure 47: Maps of r99ptot during the winter months from 1989 to 2008: (a) HadUK-Grid; (b-g) absolute biases in r99ptot within each of the ensembles of models.

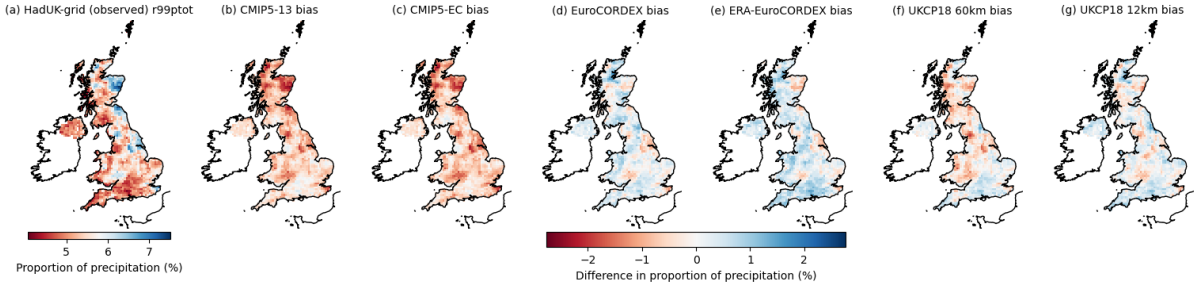
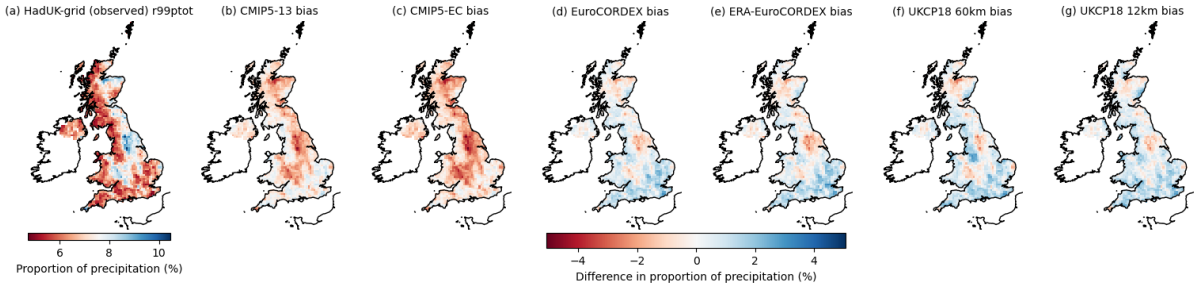


Figure 48: Maps of r99ptot during the summer months from 1989 to 2008: (a) HadUK-Grid; (b-g) mean biases in each ensemble.



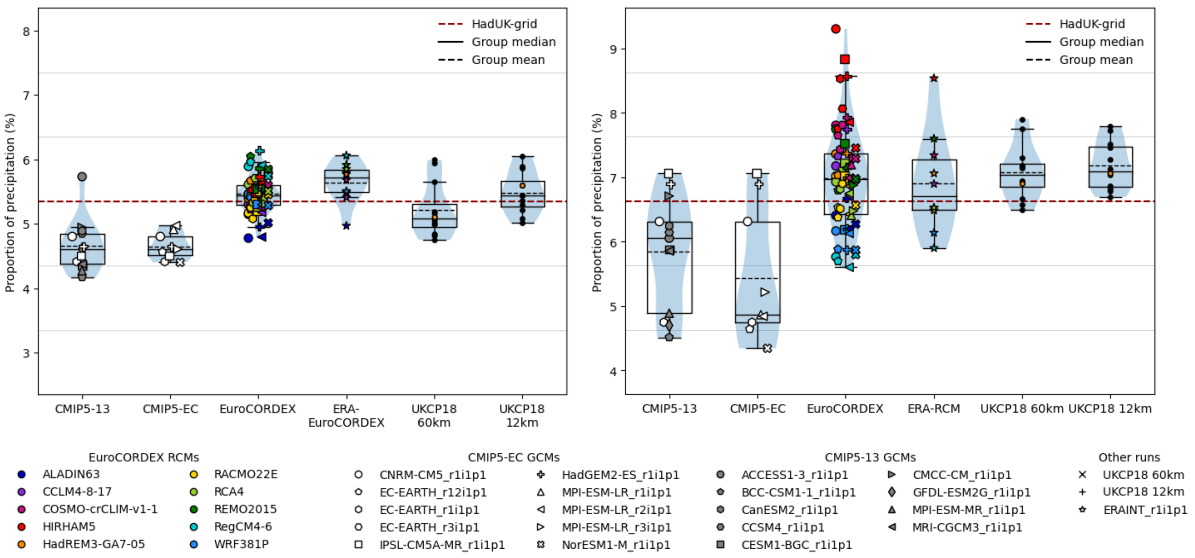
Boxplots of the biases in each model, averaged over the UK, are shown in Figure 49. The spread of biases in r99ptot in the CMIP5 and EuroCORDEX ensembles is narrower in winter than in summer, reflecting the fact that winter precipitation — and particularly intense winter precipitation — tends to arise from frontal processes, which are generally less variable than the convective processes that typically produce intense summer precipitation, and which are likely to be represented more consistently by most models. Only the UKCP18 ensemble produces a similar spread of biases in both winter and summer r99ptot, with each member producing similar biases at both 60km and 12km resolutions. The biases in both winter and summer r99ptot in the EuroCORDEX ensemble are grouped predominantly by colour, and are similar to the biases seen in the reanalysis-driven ERA-EuroCORDEX ensemble, indicating that these are driven largely by the choice of RCM. In both

winter and summer, biases in r99ptot are negatively correlated with the biases in fwd seen in Figures 32c and 41c, with models that simulate a higher proportion of wet days tending to simulate less intense rainfall on the wettest days.

Figure 49: Boxplots of the bias in r99ptot in each model in the winter and summer months, averaged over the UK and grouped by ensemble.

(a) Biases in winter r99ptot

(b) Biases in summer r99ptot



An ANOVA carried out over the full EuroCORDEX ensemble, including the reanalysis-driven ERA-EuroCORDEX runs, confirms that differences between the RCMs contribute more systematic variation (48% of the total in winter, 46-49% in summer) to the model biases in r99ptot than differences between the GCMs (14-15% of the total in winter and 11-13% in summer), although a high proportion (38% of the total in winter, 41% in summer) remains unexplained by either model component: this high residual uncertainty reflects the contribution of unpredictable individual storm events to precipitation on the wettest days in both seasons. Maps of the standard deviation of r99ptot in the EuroCORDEX ensemble, and of the contribution of each model component to this variability, are shown in Figures 50 and 51. In winter, the choice of RCM is able to explain much of the variation between the EuroCORDEX runs in the wettest areas, with up to 90% explained in some areas; however, in areas where precipitation is not driven primarily by frontal processes, the residual variability is much higher. This is not the case in summer, where the contribution from each model component is fairly constant over the UK, and the variation between the model runs is greatest in central and southern England, where precipitation from parameterised thunderstorms is more important.

6 Evaluation of wind speed climatologies

In this section, the distribution of wind speeds simulated by each model during the evaluation period is evaluated. Due to the difficulty of obtaining observations of gust speeds over the evaluation period, the analysis here is limited to mean wind speeds, which are available at monthly frequency

Figure 50: Maps of (a) the standard deviation of winter `r99ptot` in the EuroCORDEX ensemble (including evaluation runs driven by ERA-Interim reanalysis), and (b-d) of the percentage of the total ensemble variance in each grid cell explained by the choice of GCM and RCM.

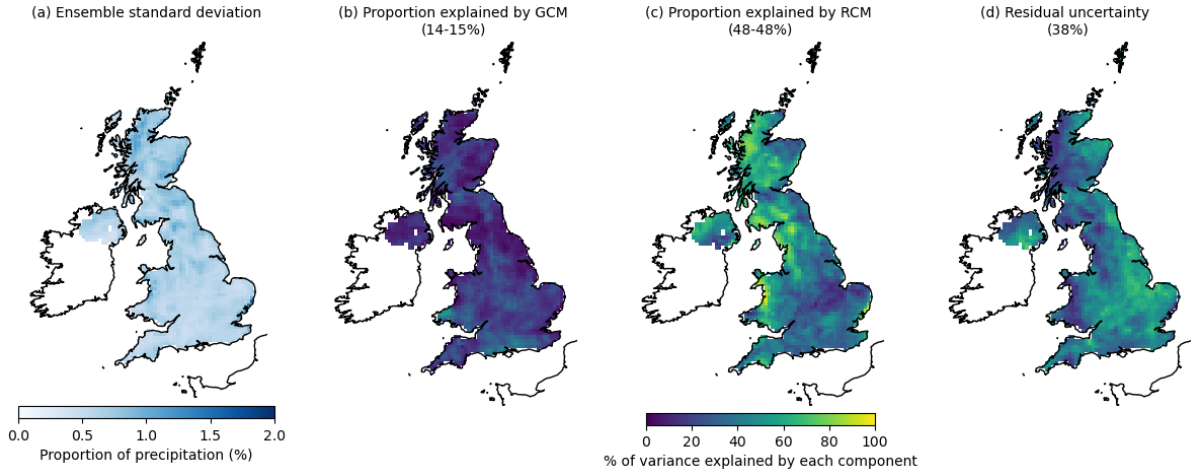
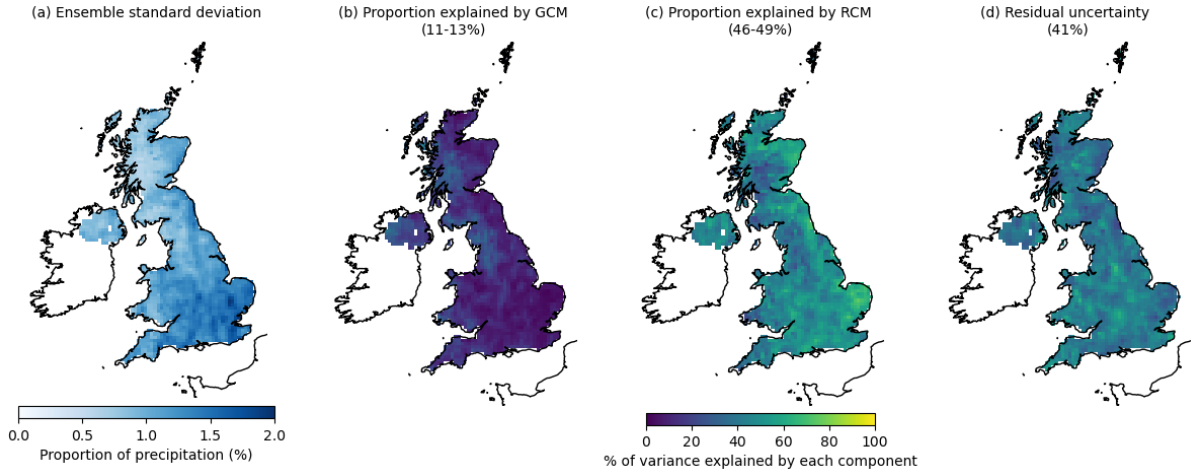


Figure 51: Maps of (a) the standard deviation of summer `r99ptot` in the EuroCORDEX ensemble (including evaluation runs driven by ERA-Interim reanalysis), and (b-d) of the percentage of the total ensemble variance in each grid cell explained by the choice of GCM and RCM.



as part of the HadUK-Grid dataset.

The HadUK-Grid observed wind speeds in winter and summer are shown, respectively, in Figures 52 and 53, along with the mean biases in each ensemble. The spatial patterns of observed wind speeds (panel a in both figures) are very similar in both seasons, with higher winds observed at higher elevations, although wind speeds in winter are typically around double those observed in summer. The ensemble mean biases are also very similar in both seasons — again, up to a scaling factor — with all of the ensembles overestimating wind speeds slightly at low elevations, and failing to simulate high enough wind speeds in at higher elevations: the magnitude of this bias is largest in the two UKCP18 ensembles. The fact that the patterns of ensemble mean biases so closely resemble the patterns of observed wind speed in both spatial distribution and magnitude suggests that all of the ensembles, regardless of resolution, simulate insufficiently variable wind speeds across the UK in

both summer and winter.

Figure 52: Maps of (a) HadUK-Grid mean daily wind speed during the winter months from 1989 to 2008, and (b-g) of the climatological biases in each ensemble.

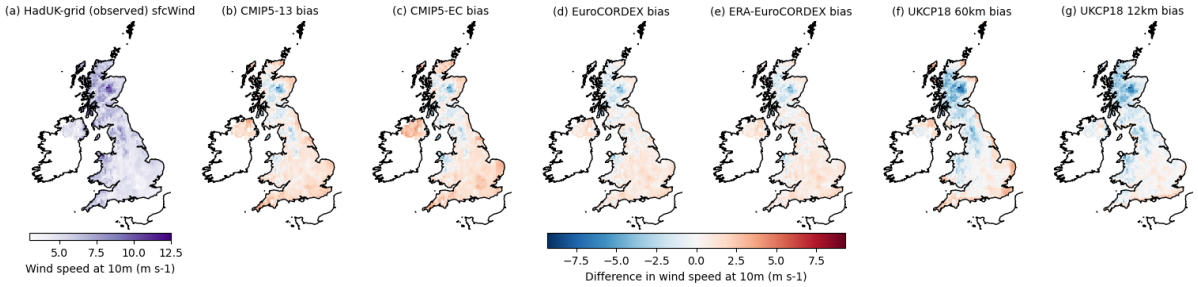
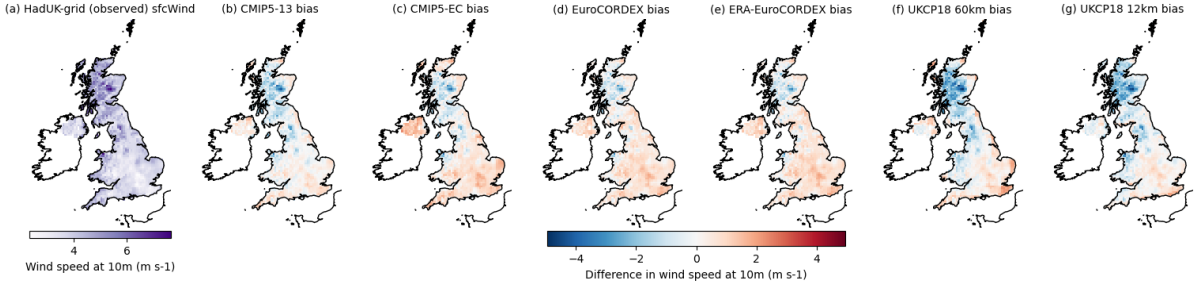


Figure 53: Maps of (a) HadUK-Grid mean daily wind speed during the summer months from 1989 to 2008, and (b-g) of the climatological biases in each ensemble.



The distribution of individual run biases in the average wind speed across the UK is shown in Figure 54. With the exception of one GCM (CanESM2), the CMIP5 and EuroCORDEX models simulate similar ranges of wind speeds in both winter and summer; the EuroCORDEX runs are closely grouped by colour, indicating that the biases in particular runs are driven largely by the RCM, rather than the GCM. The UKCP18 ensembles tend to underestimate wind speeds in winter to a greater degree than in summer, and also simulate a narrower range of UK-averaged wind speeds in summer than in winter.

The success of the models at capturing the spatial patterns of summer wind speeds shown in Figures 53a is summarised in the Taylor diagrams in Figure 55. As with the biases already discussed, the models' performance is very similar in both winter and summer, so the Taylor diagram for winter wind speeds is not shown here. Almost all of the models substantially underestimate the amplitude of the spatial patterns in the observed wind speeds, supporting the earlier suggestion that the models tend to simulate too-smooth fields of wind speeds, although the UKCP18 60km runs simulate patterns with higher standard deviations. Correlation between the models' surface wind climatologies and the climatologies derived from HadUK-Grid is generally poor and highly variable, with the performance of the EuroCORDEX runs depending very strongly on the downscaling RCM: the best correlation scores achieved were close to 0.75, with most model runs scoring less than 0.5, and several RCMs having extremely low or even negative correlation with the observed wind speeds.

Figure 56 shows the results of an ANOVA of the biases in summer wind speeds for the full Eu-

Figure 54: Boxplots of average daily wind speeds (in m/s) over the UK land surface for each run

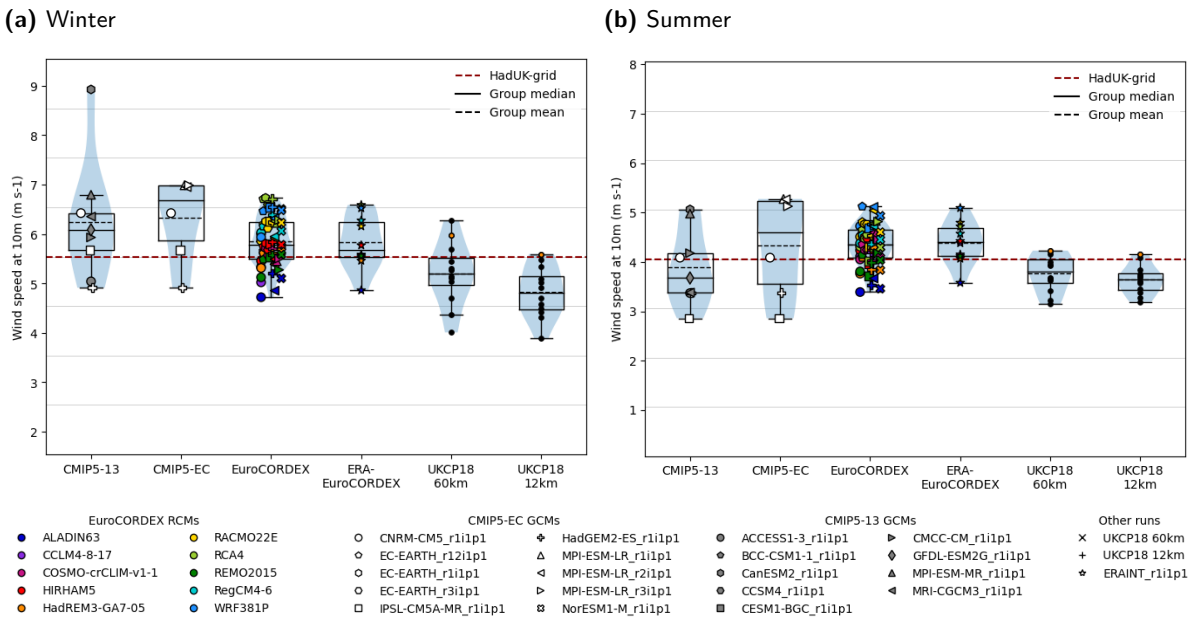
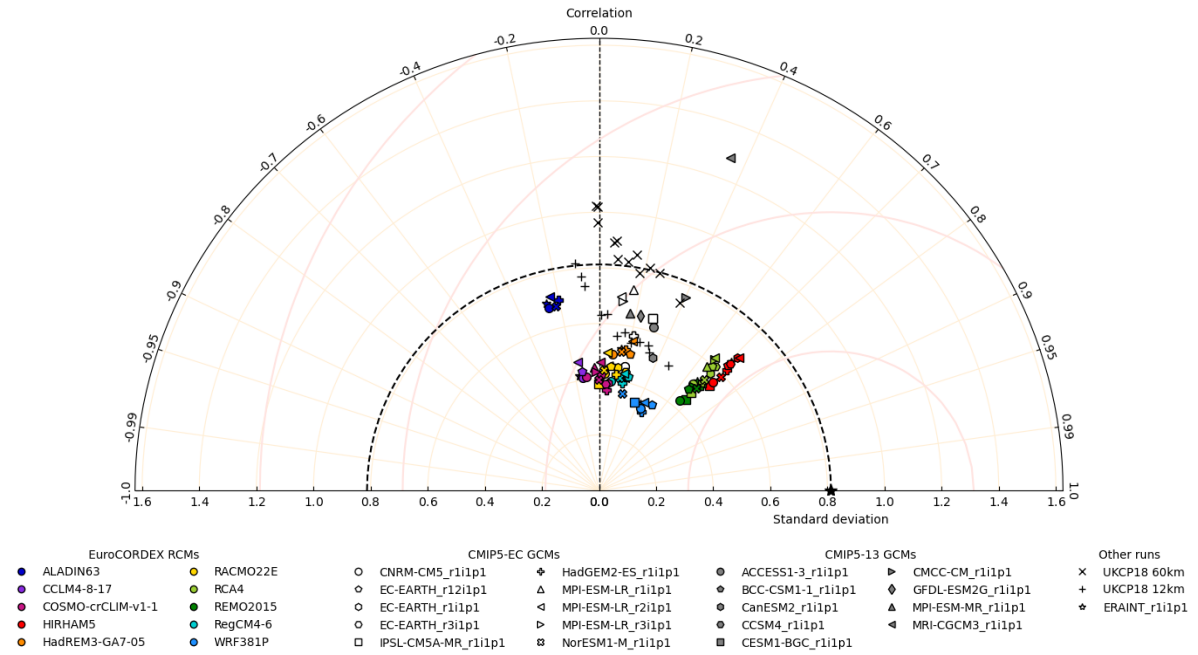


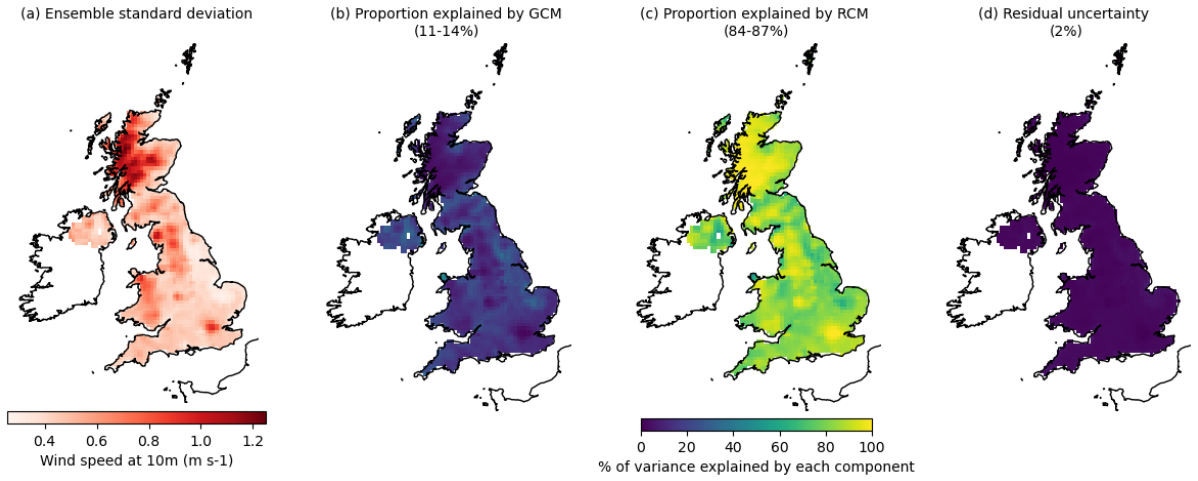
Figure 55: Taylor diagram comparing model climatologies of `sfcWind` during the summer months from 1989-2008 to HadUK-Grid climatology



roCORDEX ensemble, including reanalysis-driven runs; the plots for winter wind speeds are very similar, so are not shown here. The within-ensemble standard deviation at each grid cell is shown in Figure 56a: the largest differences between the EuroCORDEX model runs are not only, as might be expected, at the higher elevations where the highest wind speeds occur, but also over cities: London and Birmingham are both clearly visible as darker spots, indicating a higher level of variation between the modelled windspeeds than in the surrounding area. Between 84 and 87% of the total

variability across the UK is explained by differences between the RCMs, but locally in urban areas and at higher altitudes the proportion approaches 100%. This is probably due to the effect of surface roughness parametrisations within the RCMs, which are particularly influential in urban areas.

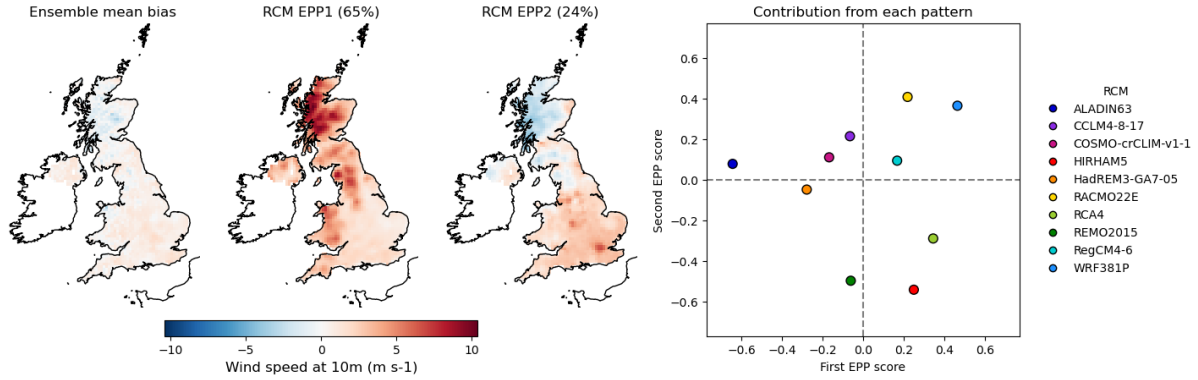
Figure 56: Maps of (a) the standard deviation of mean summer wind speeds at 10m in the EuroCORDEX ensemble (including evaluation runs driven by ERA-Interim reanalysis), and (b-d) of the percentage of the total ensemble variance in each grid cell explained by the choice of GCM and RCM. The ranges of proportions of the total variance explained by each component, computed using the methodology described in Section 3.5.2, are given in parentheses.



The dominant spatial patterns arising from the differences between the fitted RCM effects are shown for summer wind speeds in Figure 57; again, the patterns for winter wind speeds are very similar, so are not shown. The first component, labelled 'RCM EPP1', captures 65% of the variability between the RCMs: this pattern reflects a positive bias across the UK with respect to the ensemble mean bias, increasing with elevation, meaning that runs downscaled using RCMs with a positive (negative) score in this component tend to be windier (less windy) than the ensemble average, particularly in mountainous regions. The second component, which accounts for 24% of the between-RCM variance, has two distinct features: positive scores in this component indicate a tendency to simulate lower wind speeds in much of highland Scotland, the Pennines and western Northern Ireland than in the rest of the UK, and so to produce smoother fields of wind speeds across the UK; but also to simulate higher winds in cities than in the surrounding areas, with London, Birmingham, Manchester, Leeds and Sheffield all visible as darker patches. These two spatial patterns jointly describe 89% of the variation between the RCMs, and so describe between 76 and 78% of the total variation between the EuroCORDEX runs. The first EPP score is highly correlated with the individual model biases seen in Figure 54b; the largest scores in the second component (assigned to REMO2015 and HIRHAM5) are negative, suggesting that these models tend to overestimate wind speeds in Scotland and to underestimate urban wind speeds to a greater extent than other models in the ensemble.

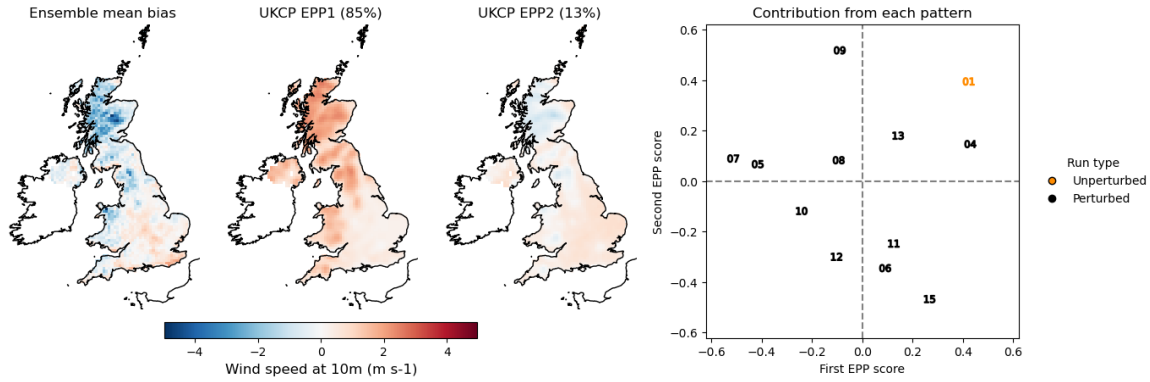
EPP analysis was also carried out over the UKCP18 ensemble to identify the dominant spatial patterns of variation between the regional runs (Figure 58). The first EPP, which accounts for 85% of the variation between the runs, is similar to the first EPP of the EuroCORDEX RCM effects, with runs with positive scores in this component being windier than the ensemble average,

Figure 57: Ensemble principal patterns in summer wind speed climatologies from 1989-2008, showing dominant patterns of contributions from each RCM to variation about the mean bias in the EuroCORDEX ensemble. The proportion of the between-RCM variation attributed to each pattern is given in parentheses.



particularly at higher elevations. Positive scores in the second component, which accounts for 13% of the variation, likewise indicate less variable wind speeds across the UK; however, unlike in the EuroCORDEX ensemble, this component is not associated with higher wind speeds in urban areas. In fact, London appears as a lighter patch in both of these components, suggesting that there is very little variation in the wind speeds simulated by the twelve ensemble members in urban areas.

Figure 58: EPP analysis of summer wind speeds simulated by the UKCP18 regional models, showing the dominant patterns of variation



7 Evaluation of large-scale features of the climate

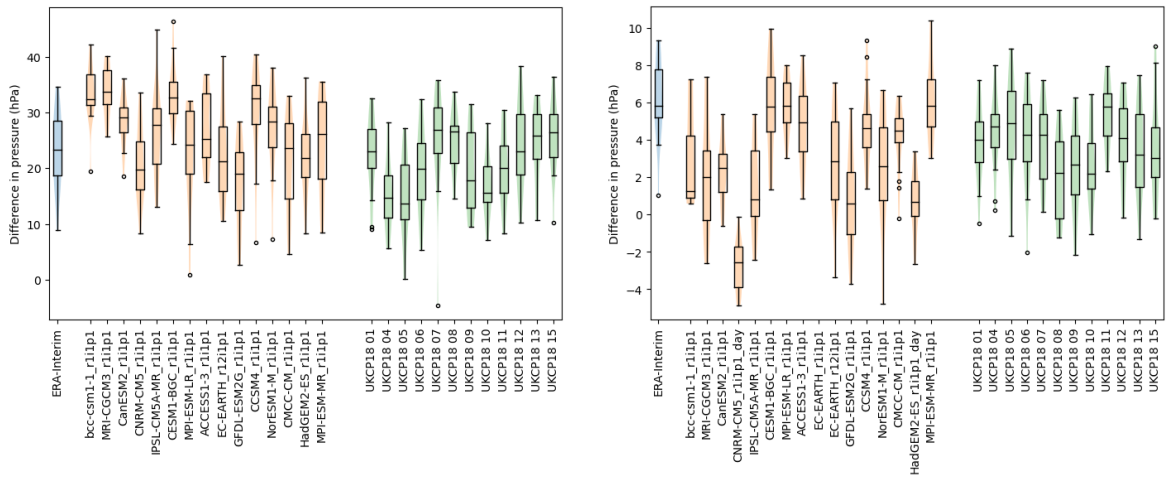
The above evaluation has focused on comparison with well-observed surface variables over the UK. In this section the ability of the ensembles to capture the physics of the wider climate system is evaluated. Generally these features occur over larger areas than the UK or Europe. Therefore, this strand of the analysis mainly involves assessment of the underlying GCMs, rather than the higher-resolution regional output produced by by the RCMs. We first focus on several climate indices that play an important role in determining the meteorology of the UK, and then move onto assessing the various kinds of weather that it experiences.

7.1 Climate Indices

The North Atlantic Oscillation index measures interannual variations in the seasonal location and strength of the stormtrack that brings low pressure systems to the UK (Scaife et al., 2014). It consists of the pressure difference between the Azores and Iceland (Hurrell, 1996). It is generally much stronger in the winter than summer (as can be seen in the ERA-Interim reanalysis in the leftmost columns of Figure 59), with higher values associated with a stronger influence from prevailing conditions from the North Atlantic. This seasonal cycle is simulated by all GCMs, in both the CMIP5 and UKCP18 60km ensembles. CMIP5 generally overestimates the winter pressure gradient (panel a) and underestimates the summer gradient (panel b), and also exhibits a wider range of mean states than UKCP18. There are several members of CMIP5 that show substantially less interannual variability in the winter NAO than ERA-Interim (Figure 59a), while all UKCP18 runs simulate close to the expected level of variability.

Figure 59: Representation of the North Atlantic Oscillation, as measured by the pressure differences between the Azores and Iceland. Boxplots indicate the distribution of daily values over the period 1989 to 2008. Note the difference in vertical scales between the winter and summer plots.

(a) Winter **(b)** Summer



A second useful index of large-scale climate variability is the frequency with which anticyclonic weather patterns 'block' the movement of low pressure systems over the UK (Masato et al., 2013): there are several patterns that can induce this effect (Section 7.2). In winter, the frequency of these conditions in the CMIP5 models ranges from 14% (IPSL-CM5A-MR) to 26% (HadGEM2-ES) of days on average (Table 6). There appears to be little connection between a model's blocking frequency and either the winter storm track strength or its variability (as measured by the NAO index, Figure 59).

All of the regional climate models consist of solely atmospheric components, meaning that any ocean climate indices are determined solely by the driving CMIP5 model. An important ocean feature for the climate of the UK is the Atlantic Meridional Overturning Circulation (AMOC), which assists in bringing heat polewards. Observations from the RAPID array suggest that the strength of the AMOC in recent years lies between 15-20 Sv (Smeed et al., 2018), although this reflects a period of

reduced overturning compared to the period from 2004-2008, and so represents a period of weaker overturning than that simulated by the GCMs. The models analysed here either fall within that range or are slightly below it (Table 6, 'AMOC (Sv)'). They also represent the weaker side of the spread exhibited in the full CMIP5 ensemble (Reintges et al., 2017). The AMOC strength is not available for the UKCP18 global ensemble.

Table 6: Table of large-scale climate indices. **ECS** means the Effective Climate Sensitivity (Zelinka et al., 2020). **TCR** means the Transient Climate Response and is taken from Flato et al. (2014). The Atlantic Meridional Overturning Circulation (**AMOC**) was calculated by Wang et al. (2014). The blocking frequency is computed at 0° in winter (Masato et al., 2013) and represents the proportion of days with blocking high pressure systems. Climate indices are available only for each GCM, rather than for each individual simulation, and are not available for the UKCP18 global PPE members.

GCM	ECS (°C)	TCR(°C)	AMOC (Sv)	Blocking Frequency
CNRM-CM5	3.25	2.1	14	0.17
EC-Earth	3.3	2	18	0.21
HadGEM2-ES	4.6	2.5	20	0.26
IPSL-CM5A-MR	4.11	2	11	0.14
MPI-ESM-LR	3.63	2	20	0.19
NorESM1-M	2.87	1.4	12	0.13

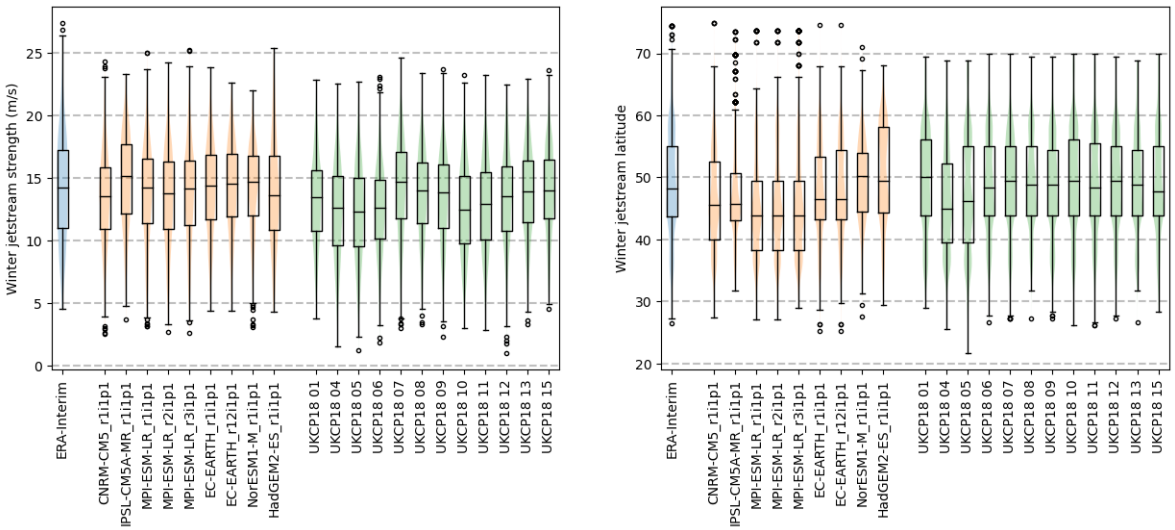
The jetstream is a band of strong winds occurring at the top of the troposphere, crossing the north Atlantic towards Europe (Woollings et al., 2010). It is stronger in winter (Figure 60a) than in summer (Figure 60b), a feature that is replicated well in all models. In both seasons, both the CMIP5 and the UKCP18 models are more likely to underestimate the strength of the jetstream than to overestimate it — although not by much — and all of the models seem to do a fair job in capturing variability in its strength (Figure 60). The average position of the jet stream is at about 50° N, with most of the CMIP5 models lying a little too far south, while the UKCP18 models are a little too far north in the summer (Figure 60b). There is substantial variability in the location of the jetstream throughout the evaluation period, which appears to be captured adequately in most models.

7.2 Weather types

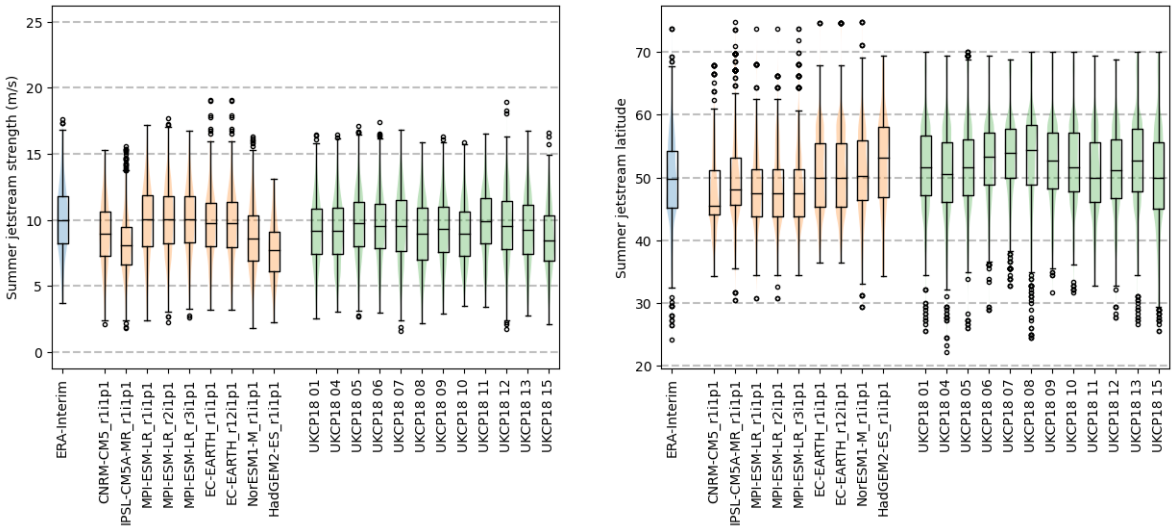
In this section we evaluate whether the GCMs used to drive the EuroCORDEX runs are simulating realistic patterns of weather regimes, using a methodology originally developed by the Met Office for use in operational weather prediction (Neal et al., 2016), and later used by Pope et al. (2021) to assess projected changes in the frequency of thirty weather patterns representing a range of circulation types that are known to affect the weather over the UK and northern Europe (Table 7). Each day of model output is classified as a single weather type by calculating the differences between anomalies in the daily mean sea level pressure (MSLP) field and each of the thirty daily weather patterns, and selecting the pattern for which the area-weighted sum of squared differences across all grid cells is smallest. Lower-numbered patterns tend to occur most frequently overall, to occur more frequently in the summer months, and to have smaller MSLP anomalies; higher-numbered patterns occur less frequently overall, to occur more frequently in the winter months, and to have

Figure 60: Strength and location of the jetstream, determined using the methodology of Woollings et al. (2010). Boxplots indicate the distribution of daily values over the period 1989 to 2008.

(a) Winter



(b) Summer



larger MSLP anomalies, and so can represent more severe weather regimes. For further details of the methodology and of the characteristics associated with the weather types, see Neal et al. (2016) and Pope et al. (2021).

Pope et al. (2021) carried out a detailed analysis of the proportion of days assigned to each weather type in the CMIP5-13 and UKCP18 global ensembles, and of changes in the proportion of days assigned to each weather type; we have followed a similar approach here, to allow direct comparison with the results in that paper. Key differences between that investigation and the results presented here are that, in order to main consistency with the rest of the analysis, the 'true' weather types here are computed from ERA-Interim rather than ERA-5; the evaluation period is from 1980-2010, rather than from 1979-2010; and changes are reported between the reference period of 1980-2010

Table 7: Table of weather types (Pope et al., 2021)

Pattern	Description
1	Unbiased north-westerly
2	Cyclonic south-westerly, returning Polar maritime airmass
3	Anticyclonic south-westerly, ridge over Northern France
4	Unbiased westerly
5	Unbiased southerly, high over Scandinavia
6	Anticyclonic, Azores high extension
7	Cyclonic south-westerly, low west-northwest of Ireland
8	Cyclonic westerly, low near Shetland
9	Anticyclonic north or north-easterly, high near Iceland
10	Anticyclonic west or south-westerly, slight Azores ridge
11	Cyclonic, low centred over southern UK
12	Anticyclonic southerly, high over Poland
13	Anticyclonic north-westerly, high south west of Ireland
14	Cyclonic north or north-westerly, low near southern Sweden
15	Unbiased south-westerly, very windy over north-west Britain
16	Anticyclonic south or south-easterly, high east of Denmark
17	Anticyclonic east or south-easterly high over Denmark
18	Anticyclonic south-westerly, high over northern France
19	Unbiased northerly, low east of Denmark
20	Cyclonic westerly, intense low near Iceland
21	Cyclonic south-westerly, deep low south of Iceland
22	Cyclonic southerly, low west of Ireland
23	Unbiased westerly, windy in north of the UK
24	Cyclonic northerly, low in North Sea
25	Anticyclonic northerly, high centre in Irish Sea
26	Cyclonic north-westerly, low near Norway, windy
27	Anticyclonic easterly, high in Norwegian Sea
28	Cyclonic south-easterly, low SW of UK
29	Cyclonic south or south-westerly, deep low west of Ireland
30	Cyclonic west or south-westerly, deep low south-east of Iceland

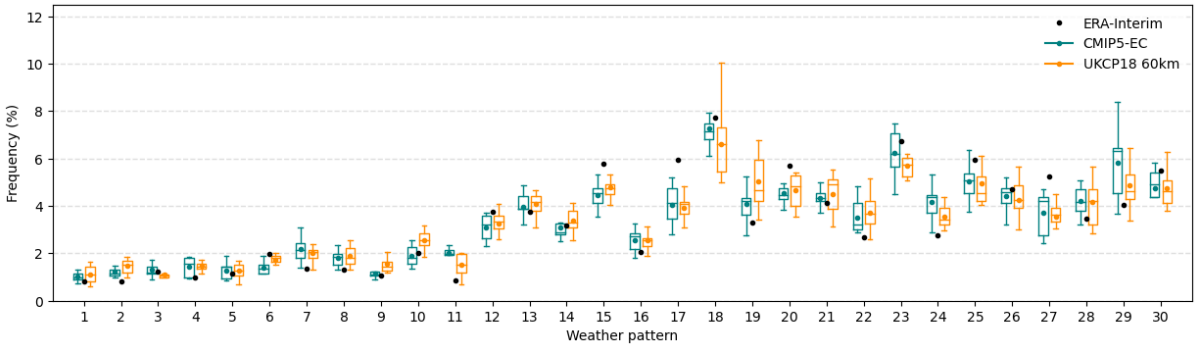
and the future period of 2050-2080, rather than a historical period from 1900-2010 and a future period of 2070-2099. Despite these differences, the results of the analysis of changes in the UKCP18 global ensemble are very similar to those presented in that paper, so the discussion here is kept short, and focuses more on differences between the UKCP18 and CMIP5-EC ensembles, which may be expected to result in differences between the UKCP18 and EuroCORDEX regional ensembles.

Figure 61 shows the proportion of days assigned to each of the thirty weather types by the CMIP5-EC and UKCP18 ensembles, along with the proportion assigned to each weather type by ERA-Interim. The pattern of weather types assigned is very similar to that seen in Figure 3 of Pope et al. (2021), with both ensembles generally simulating similar proportions of each weather type and approximately the expected proportion of most weather types. In winter, too few occurrences of types 15, 17, 20 and 27 are simulated in both ensembles; pattern 27 in particular is associated with cooler conditions and can potentially bring severe weather, especially in the east of the UK. In summer, both ensembles overestimate the prevalence of type 1, representing generally settled and dry conditions over the UK, and the two ensembles simulate quite different proportions of type 9, a pattern associated with the negative phase of the NAO, with CMIP5-EC producing slightly too few and UKCP18 slightly too

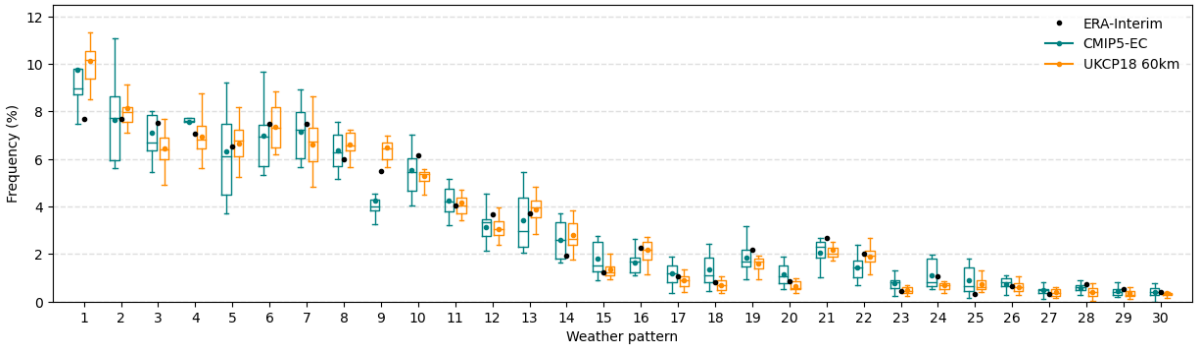
many occurrences.

Figure 61: Proportion of days assigned to each weather type by the CMIP5-EC and UKCP18 ensembles and by ERA-Interim. The boxes display the upper and lower quartiles of each ensemble, with the median marked by a horizontal line and the mean by a dot. The whiskers extend to the furthest point lying within 1.5 times the inter-quartile range of the upper or lower quartile; points lying beyond the whiskers are not shown.

(a) Winter months



(b) Summer months



Boxplots of the relative changes in the proportion of each weather type simulated by the ensembles are shown in Figure 62. The orange points denoting the mean of the UKCP18 ensemble are equivalent to the red bars in Figure 4 of Pope et al. (2021), and exhibit broadly similar patterns of changes.

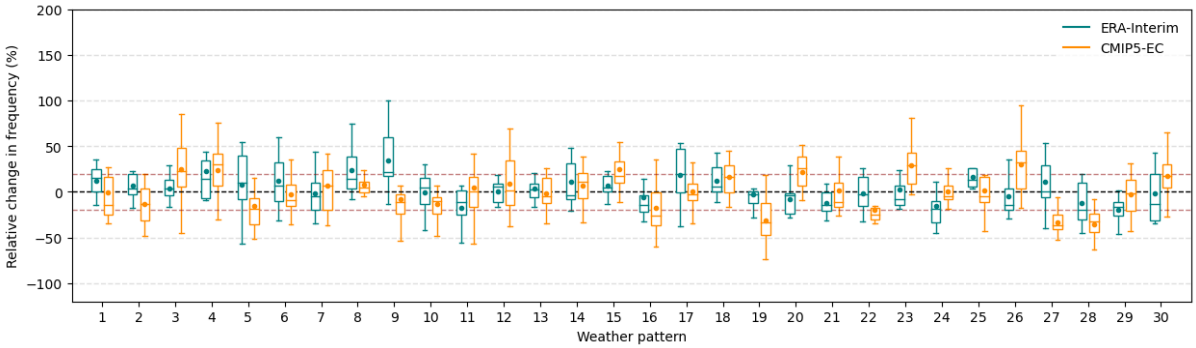
The changes in weather type proportions are generally smaller in winter than in summer, with occurrences of most weather types changing by less than 20% in both ensembles (panel a). The distributions of changes in each weather type simulated by the two ensembles generally overlap, with the CMIP5 ensemble typically producing a somewhat wider range of changes than the UKCP18 ensemble. The UKCP18 ensemble mean change is typically larger than that of the CMIP5-EC ensemble, although the CMIP5-EC ensemble simulates changes of more than 20% in weather types 8 and 9; both of these weather types are fairly rare in winter and neither is strongly associated with extreme weather. Overall, the UKCP ensemble projects a reduction in weather types associated with anticyclonic, settled regimes (Table 7), indicating a shift towards warmer, wetter winters; however, most members of the CMIP5-EC ensemble project an increase — of around 10% on average — in occurrences of these weather types.

Changes in the frequency of each weather type during the summer months are much more variable

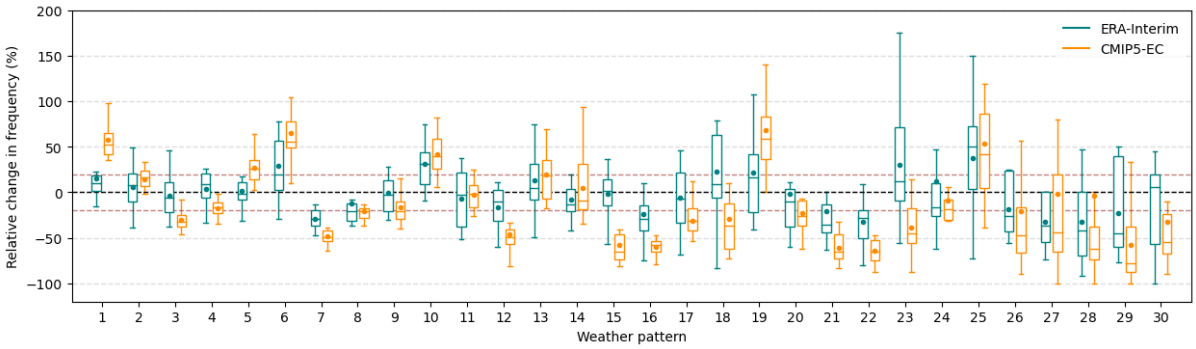
(panel b), with the greatest variability among the higher-numbered weather patterns that tend to occur only rarely in summer (Figure 61b). As noted by Pope et al. (2021), the UKCP18 ensemble simulates a decrease in the frequency of the higher-numbered patterns associated with wet and stormy weather, and this behaviour is also seen in the CMIP5-13 ensemble, although the changes are typically smaller. Similarly, the CMIP5 ensemble consistently simulates smaller changes than the UKCP18 ensemble in the lower-numbered patterns more typically seen during the summer. During the summer months, both ensembles exhibit an increase in anticyclonic weather types that typically bring dry, settled weather, suggesting a shift towards warmer, drier summers.

Figure 62: Changes in the proportion of days assigned to each weather type by the CMIP5-EC and UKCP18 ensembles between the reference period from 1980–2010 and the future period from 2050–2080. The boxes display the upper and lower quartiles of each ensemble, with the median marked by a horizontal line and the mean by a dot. The whiskers extend to the furthest point lying within 1.5 times the inter-quartile range of the upper or lower quartile; points lying beyond the whiskers are not shown. Red dashed lines indicate changes of more than 20% from the reference period.

(a) Winter months



(b) Summer months



7.3 Rates of warming

An implicit assumption of the evaluation undertaken so far in this report is that a model's ability to simulate the present conditions constrains the fidelity of its climate change response. An alternate approach to categorise a global climate model's warming response is through its behaviour in some idealised experiment. One such experiment involves the atmospheric carbon dioxide concentration increasing by 1% per year cumulatively (e.g. Hansen et al., 1981), which has since become adopted as a key benchmarking experiment in CMIP5 and CMIP6 (Eyring et al., 2016). The global mean

surface air temperature after 70 years in this experiment is referred to as the transient climate response (TCR, Cubasch et al., 2001). The TCR values for each of the CMIP5-EC GCMs for which this value is available are given in Table 6, and are generally reflective of the spread of the wider CMIP5 ensemble (Flato et al., 2014).

A more derived, but more intuitive, index is the climate sensitivity (Sherwood et al., 2020). This is the equilibrium change in global mean surface temperature that one would see in response to a doubling of carbon dioxide concentrations. The method of using an abrupt change in forcing to diagnose a model's climate sensitivity has become the de facto standard (Gregory et al., 2004), and is termed the 'effective climate sensitivity' (ECS, Forster et al., 2021). The IPCC assesses that our best estimate of the Earth's climate sensitivity is 3°C, with a likely range of 2.5°C to 4°C (Forster et al., 2021). The majority of the CMIP5 GCMs have an ECS within this range (Table 6); both IPSL-CM5A-MR and HadGEM2-ES fall beyond the upper limit, yet still within the very-likely range which extends upto 5°C. The CMIP6 version of HadGEM3 has an ECS of 5.55°C when run at the same 60km resolution that was used for the UCKP18 global ensemble (Zelinka et al., 2020). Although the UKCP18 global ensemble uses different variants of HadGEM3 whose climate sensitivity has not been determined, it is reasonable to expect them to have an ECS outside the IPCC likely range, and the majority to even fall outside the IPCC very-likely range (Flato et al., 2014).

Table 8: Estimates of the year at which global models first reach specified levels of warming above pre-industrial climate, under the RCP8.5 emissions scenario (Gutiérrez et al., 2021). The NorESM1-M model did not reach 4°C warming during the simulations considered.

GCM	1.5°C	2°C	3°C	4°C
CNRM-CM5	2030	2045	2067	2077
EC-Earth_r12	2018	2034	2060	2082
EC-Earth_r3	2020	2034	2054	2070
HadGEM2-ES	2023	2035	2054	2071
IPSL-CM5A-MR	2009	2025	2047	2065
MPI-ESM-LR_r1	2017	2037	2061	2081
NorESM1-M	2032	2048	2072	–

The regional EuroCORDEX models are driven by global models, each of which portrays a different trajectory of warming under a given RCP. These scenarios cross key warming levels at different times. These warming levels have been determined using a 'time-shift' approach, which involves identifying segments in a pre-existing simulation that match the desired global mean temperature (Herger et al., 2015). Using a climate averaging period of 20 years, the time associated with global warming levels ranging from 1.5°C to 4°C above pre-industrial climate have been calculated by the IPCC (Gutiérrez et al., 2021). The dates are shown in Table 8. The IPSL-CM5A-LR simulation is the first to exceed all thresholds, whilst the NorESM1-M is the last to cross them, and in fact has not crossed the 4°C threshold by the end of the simulation; however, these rates of global warming are not directly reflected in the temperature changes simulated by the regional models, as discussed in Section 8.1.

8 Projected changes in surface temperature

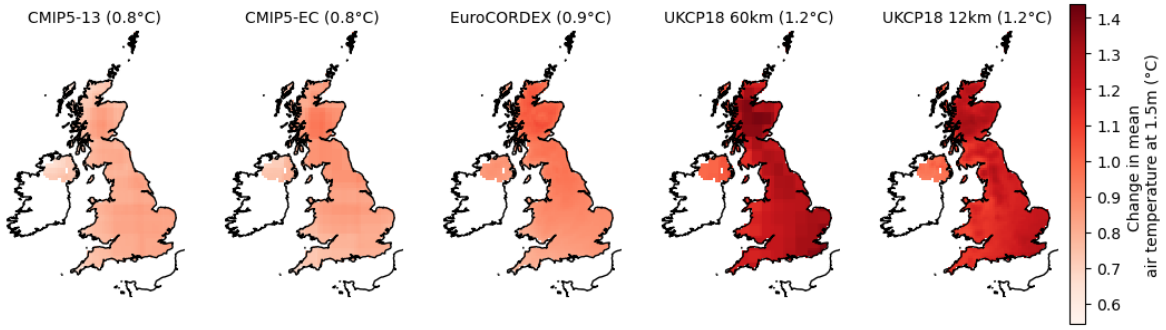
The previous sections considered the structure of biases in the model outputs with respect to the observed climatology. We now turn our attention to projected changes in climatology with respect to a reference period of 1980 to 2010, and in particular, to investigating the extent to which the changes projected by the members of the various ensembles differ. Changes are considered for both the near future period (2020–2050) and the future period (2050–2080). The temperature indices considered are summarised in Table 4 in Section 4.

8.1 Changes in average temperatures

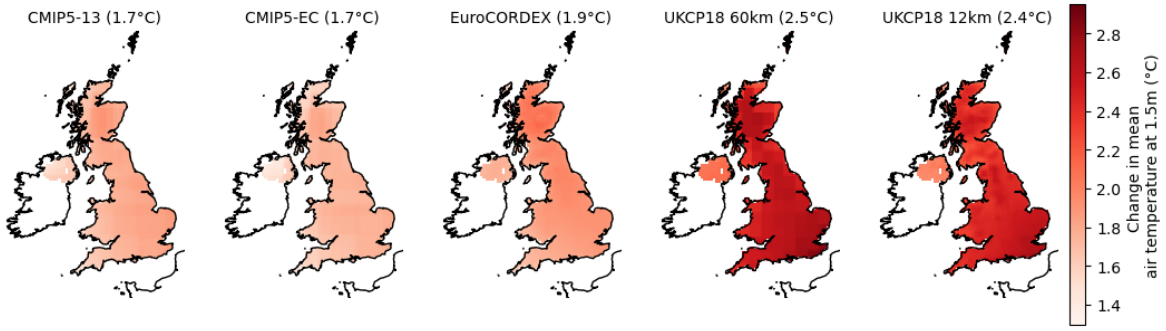
Figure 63 shows maps of the mean changes in winter temperatures for each of the ensembles; the corresponding plots of ensemble mean changes in summer temperatures are shown in Figure 64.

Figure 63: Maps of mean change in winter temperatures (in °C) in each ensemble. The average change over the UK land surface for each ensemble is shown in parentheses.

(a) Change in winter temperatures, 2020–2050



(b) Change in winter temperatures, 2050–2080

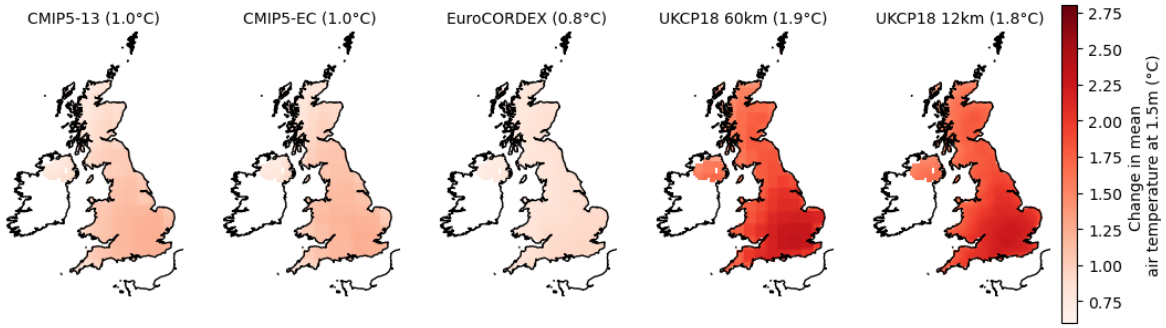


In both of these figures the maps are generally smooth, indicating a fairly constant increase in temperature across the whole of the UK in both periods, although the UKCP18 12km ensemble warms slightly more over higher elevations in winter and in southern England in summer. The CMIP5 and EuroCORDEX ensembles generally display a similar degree of warming to one another, of around 1°C in both winter and summer in the period 2020–2050 and around 2°C in both winter and summer by 2050–2080. The two UKCP18 ensemble means, which are very similar to one another, display a somewhat greater degree of warming: in winter by 0.3°C more in the near future

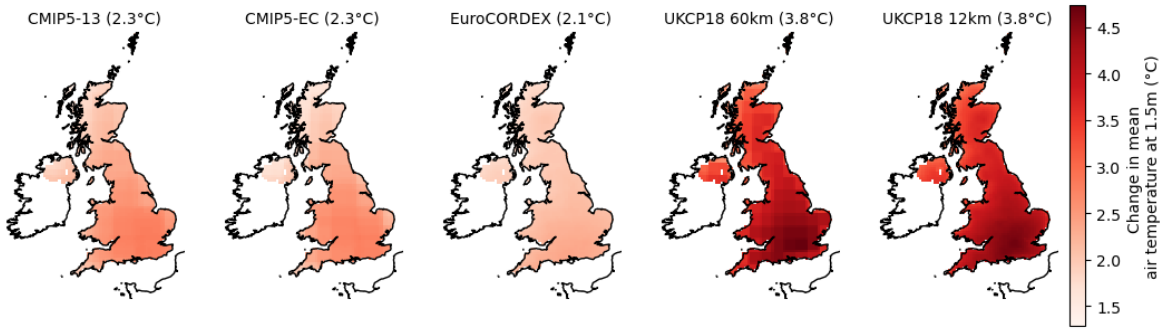
and 0.6°C more in the longer term; and in summer, by almost 1°C more in the near future and by 1.7°C more in the longer term. In all of the ensembles, the degree of warming by 2050–2080 is projected to be around double that seen by 2020–2050 in both winter and summer.

Figure 64: Maps of mean change in summer temperatures (in °C) in each ensemble. The average change over the whole of the UK for each ensemble is shown in parentheses.

(a) Change in summer temperatures, 2020–2050



(b) Change in summer temperatures, 2050–2080



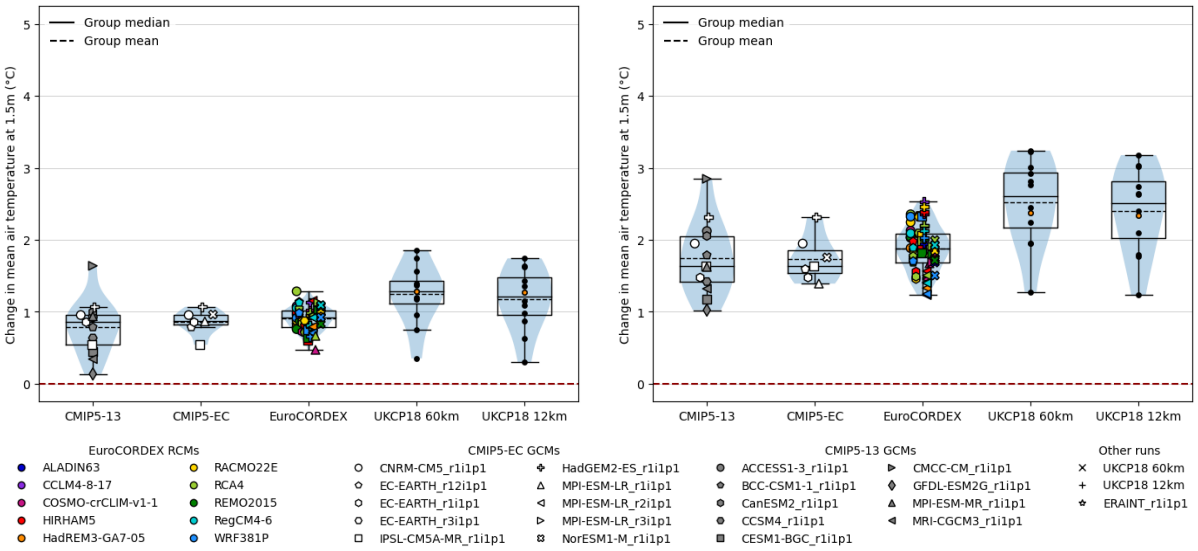
The distribution of the average changes in winter temperature projected by individual runs is shown in the boxplots in Figure 65. In both periods, the spread of changes in the CMIP5-EC and EuroCORDEX ensembles is narrower than the spread of changes in the other ensembles, with a wider spread of changes projected in all ensembles by 2050–2080 than by 2020–2050. Within the EuroCORDEX ensemble, runs denoted by the same symbol (and therefore driven by the same GCM) are closely grouped together, indicating a strong contribution from the driving models. The two UKCP18 ensembles simulate a much wider range of temperature changes and generally warm more than the EuroCORDEX ensemble, with nine of the ensemble members projecting changes above the upper quartile of changes in the EuroCORDEX ensemble in 2020–2050 (ten members in 2050–2080), and the ensemble mean and median falling close to the maximum change in the EuroCORDEX ensemble in both periods.

The spread of changes in summer temperatures in the CMIP5 and EuroCORDEX ensembles (Figure 66) is wider than the spread of changes in winter temperatures, while the spread of changes in the UKCP18 ensembles remains similar to that seen in Figure 65. HadGEM2-ES stands out among the CMIP5 runs as projecting particularly large changes in summer temperatures, warming by almost 1°C more than the ensemble average in the near future: this behaviour is reflected in the EuroCORDEX

Figure 65: Boxplots of changes in UK-averaged winter temperature climatologies in individual model runs between the reference period of 1980–2010 and future periods. Grey lines indicate intervals of 1°C change from the reference period.

(a) Change in winter temperatures, 2020–2050

(b) Change in winter temperatures, 2050–2080



runs driven by this GCM, which project larger temperature increases than runs driven by other models. In the UKCP18 ensembles, the distribution of changes in summer temperatures is centred very close to the change projected by HadGEM2-ES, with all but one of the UKCP18 runs warming by more than any EuroCORDEX run not driven by HadGEM2-ES in both periods: although the UKCP18 PPE is driven by HadGEM3-GC3.05, a model with substantial differences to HadGEM2-ES (Williams et al., 2018; Murphy et al., 2019), it appears to retain a similar degree of climate sensitivity.

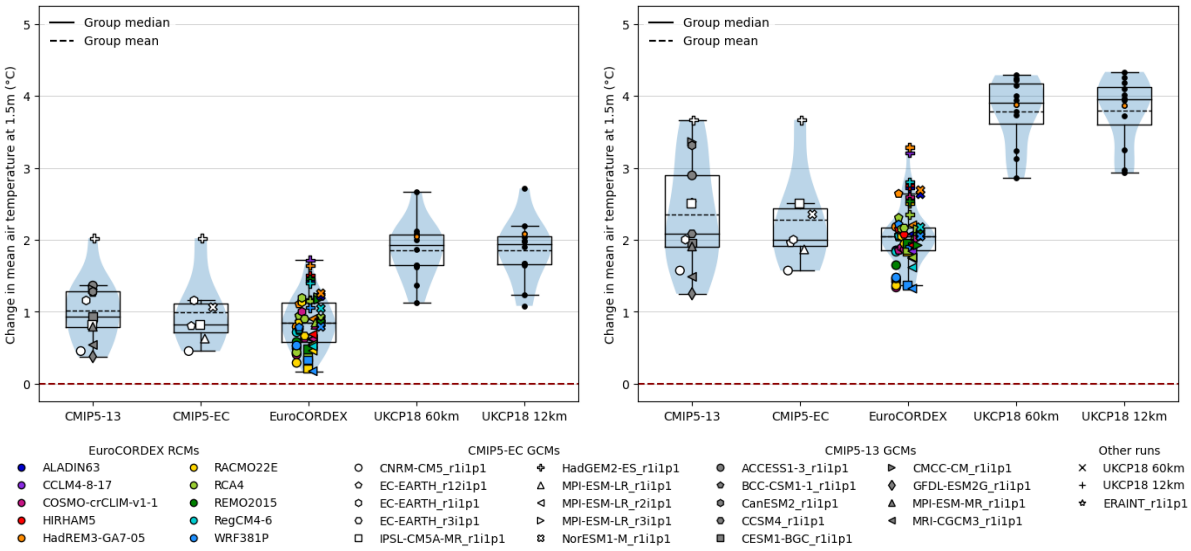
Taylor plots were produced to compare the spatial patterns of temperatures in the future periods with those in the reference period, but are not shown here: in both summer and winter, the correlation coefficient between the reference and future periods was over 0.99 for almost all of the model runs, indicating that the climatological spatial patterns of projected temperatures in the future periods are largely unchanged from the reference period. The standard deviations of the future winter temperature climatologies were between 0.9 and 1.1 times those of the reference climatologies in both periods, suggesting very little change in the amplitude of the spatial patterns, while in summer, the standard deviations were up to 20% higher than those in the reference period in the regional model runs and up to 40% higher in the global model runs.

An analysis of variance was carried out using the methodology described in Section 3.5.2 to investigate systematic variation in the projected changes in seasonal mean temperatures within the unbalanced EuroCORDEX ensemble. The proportions of variation assigned to each of the model components in each period are shown in Figure 67. The changes in both summer and winter temperatures are driven primarily by the GCM effects, with the proportion of variance in winter temperatures explained by the GCMs increasing with increasing length of the projections; in summer temperatures,

Figure 66: Boxplots of changes in UK-averaged summer temperature climatologies in individual model runs between the reference period of 1980–2010 and future periods. Grey lines indicate intervals of 1°C change from the reference period.

(a) Change in summer temperatures, 2020–2050

(b) Change in summer temperatures, 2050–2080

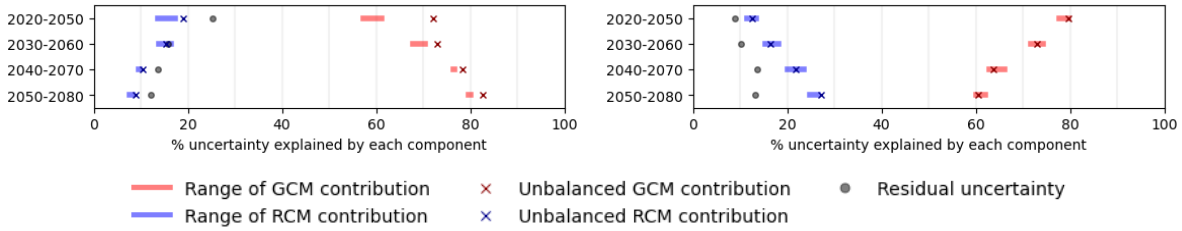


the contribution from the RCM effects increases slightly with time, although the contribution to the change in mean summer temperatures by 2050–2080 is still only around 25%. Maps of the regions

Figure 67: Plots of the proportion of the variance of changes in temperature within the EuroCORDEX ensemble that are explained by effects due to the driving GCM and the downscaling RCM, or unexplained. Coloured bars show the range of variance explained when the maximum possible variation is ascribed to either the GCM or the RCM, and the crosses show the proportions of variance explained by GCM and RCM effects fitted simultaneously using the unbalanced ANOVA described in Section 3.5.2. Residual uncertainty is shown in grey.

(a) ANOVA of change in winter temperatures

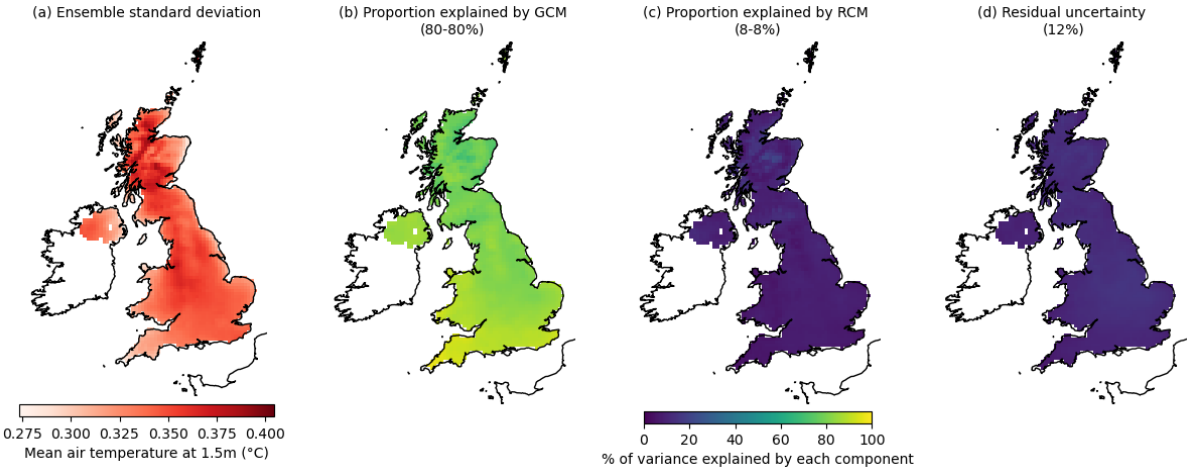
(b) ANOVA of change in summer temperatures



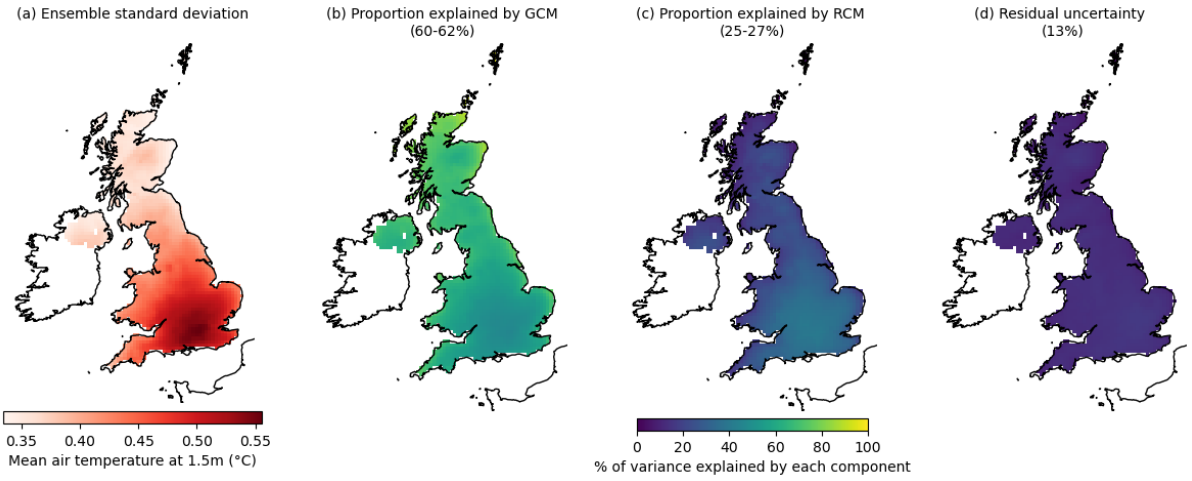
of greatest variability in temperature changes by 2050–2080 within the EuroCORDEX ensemble, and of the proportions of this variability explained by differences between the GCMs and between the RCMs, are shown in Figure 68. In winter, variation between the models is fairly uniform across the UK, with slightly larger differences between the runs occurring along river valleys in Scotland; the proportion of this variation explained by each model component is fairly constant across the UK. In summer, the largest differences occur across England from the Midlands southwards, and particularly in urban areas. Up to 40% of the variability in these regions is driven by differences between the RCMs, rather than the GCMs.

Figure 68: Maps of (a) the standard deviation of changes in temperatures in the EuroCORDEX ensemble, and (b-d) of the percentage of the total ensemble variance in each grid cell explained by the choice of GCM and RCM.

(a) Spatial distribution of variation in winter temperatures, 2050–2080



(b) Spatial distribution of variation in summer temperatures, 2050–2080

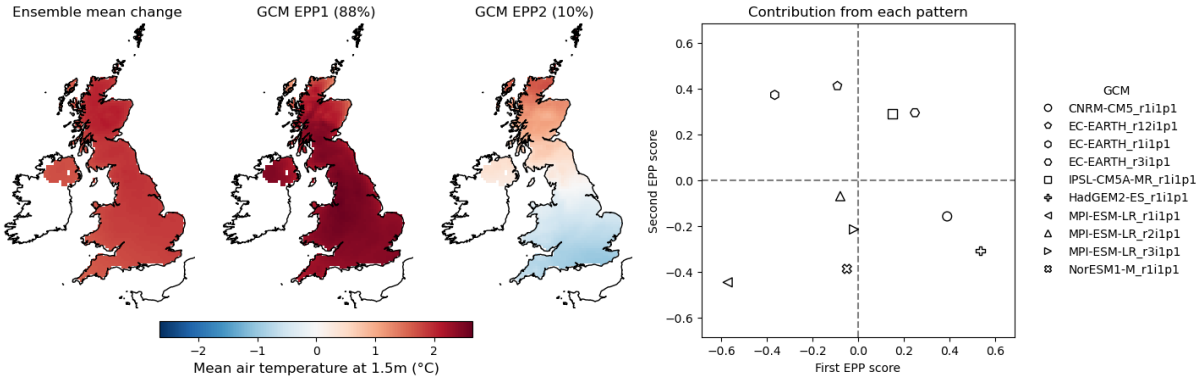


The dominant patterns of variation in the change in mean temperatures in 2050–2080 that can be attributed to systematic differences between the GCM effects are shown in Figure 69: the plots obtained for the period 2020–2050 were very similar, and so are not shown. In both winter and summer almost all of the variation (88% in winter, 96% in summer) between runs driven by different GCMs can be represented by a uniform temperature offset from the ensemble mean across the whole of the UK, with the scores in this first EPP strongly correlated with the changes projected by individual runs in Figures 65 and 66: this means that around 70% of the variation in changes in winter temperatures in the EuroCORDEX ensemble, and around 56% of the variation in changes in summer temperatures, can be attributed to a simple linear offset between runs driven by different GCMs.

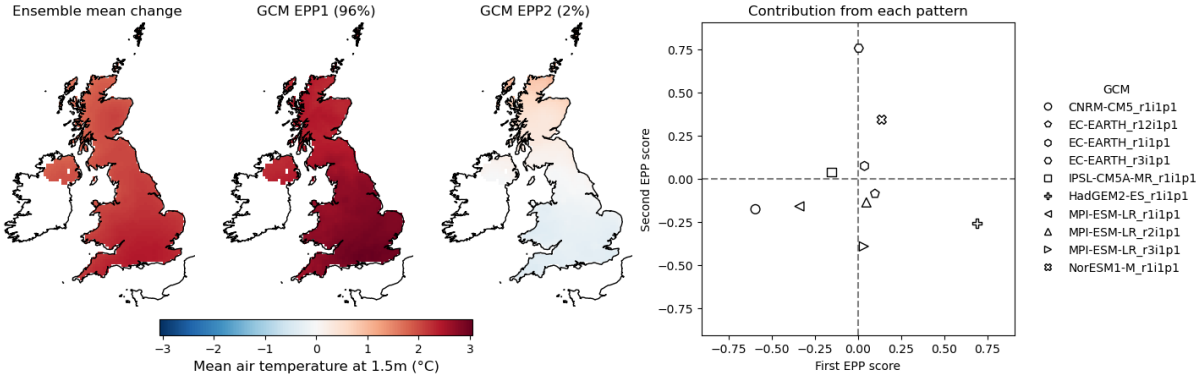
Figure 70 shows the contributions from the RCMs to within-ensemble variability in the projected changes in summer temperatures; the plots for winter temperatures, where only 8% of the total

Figure 69: Ensemble principal patterns of changes in winter temperatures between the reference period and 2050–2080, showing dominant patterns of contributions from each GCM to variation about the EuroCORDEX ensemble mean change. The proportion of the between-GCM variation attributed to each pattern is given in parentheses.

(a) Contribution of each GCM to spread of changes in winter temperatures



(b) Contribution of each GCM to spread of changes in summer temperatures



variance was contributed by the RCM, are not presented here. However, in summer, systematic differences between the RCMs contributed a larger and increasing proportion of the variance between runs, and up to 40% across central and southern England. 96% of the variability between runs downscaled by different RCMs (hence around a quarter of the total variation within the ensemble) can be attributed to the pattern labelled 'RCM EPP1', which indicates a positive offset from the ensemble mean change across the UK, with a slightly larger offset in central and southern England. HadREM3-GA7-05 has a score of 0.7 in this component, suggesting that runs using this model — also the model from which the UKCP18 PPE was generated — may project particularly high levels of warming, especially over southern England.

The results of an EPP analysis of the changes in summer temperatures projected by the UKCP18 12km runs in 2050–2080 are shown in Figure 71; the pattern of changes in 2020–2050 was almost identical, so is not shown here. In both summer and winter more than 95% of the variation within the ensemble can be attributed to a fairly uniform offset from the ensemble mean across the whole of the UK.

Figure 70: EPP analysis of RCM contributions to the variance of changes in summer temperatures (2050–2080) within the EuroCORDEX ensemble, showing the dominant spatial patterns of variation. The proportion of the between-GCM variation attributed to each pattern is given in parentheses.

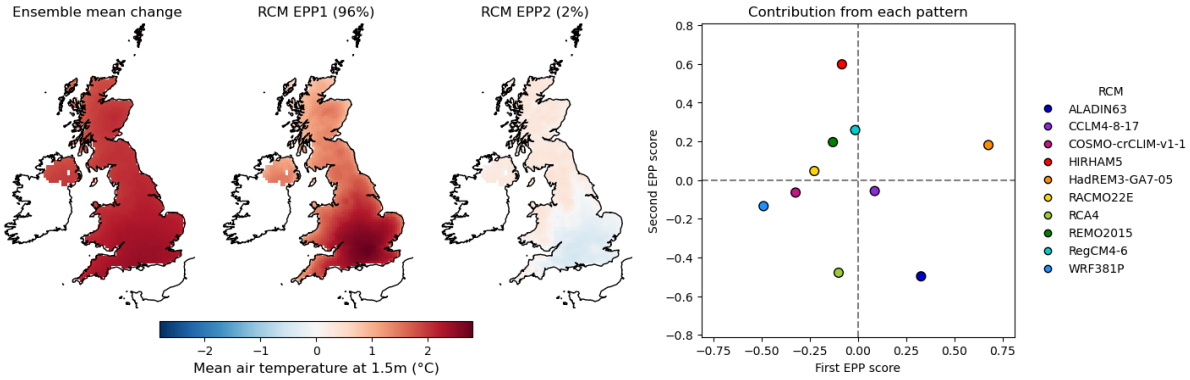
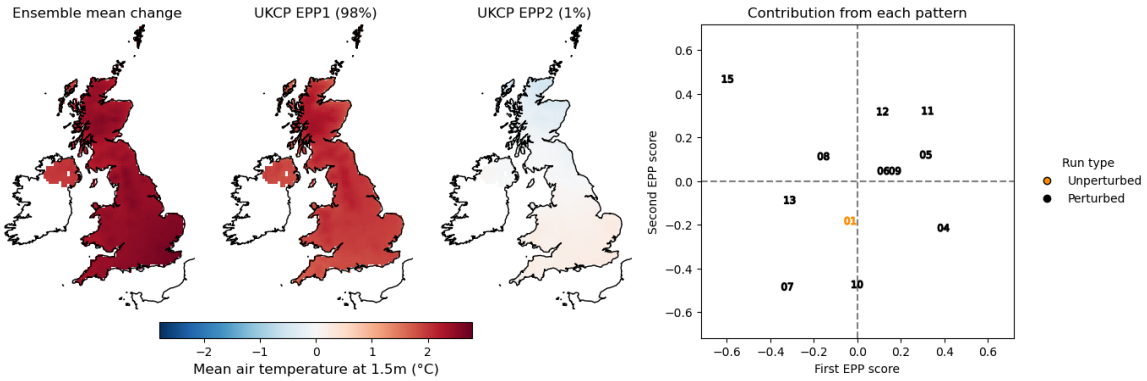
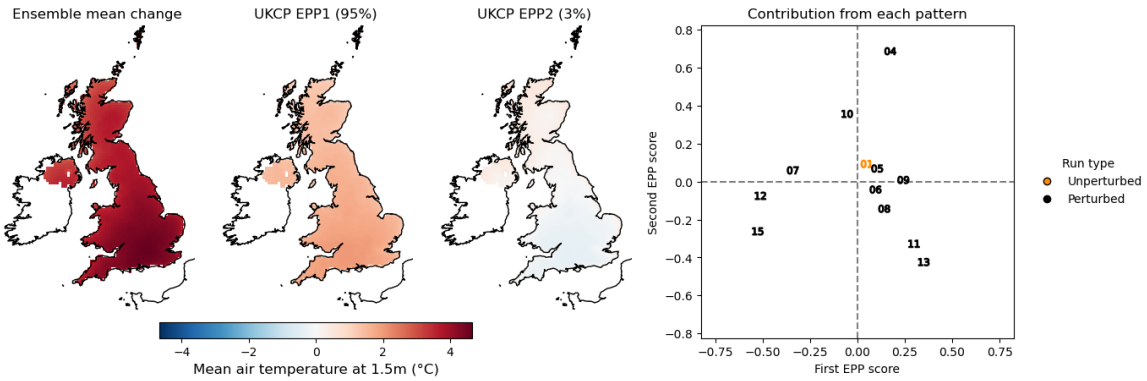


Figure 71: EPP analysis of changes in daily mean temperatures (2050–2080) simulated by the UKCP18 regional models, showing the dominant patterns of variation. The proportion of the total variation attributed to each pattern is given in parentheses.

(a) EPPs of changes in winter temperatures



(b) EPPs of changes in summer temperatures



8.2 Changes in the diurnal temperature range

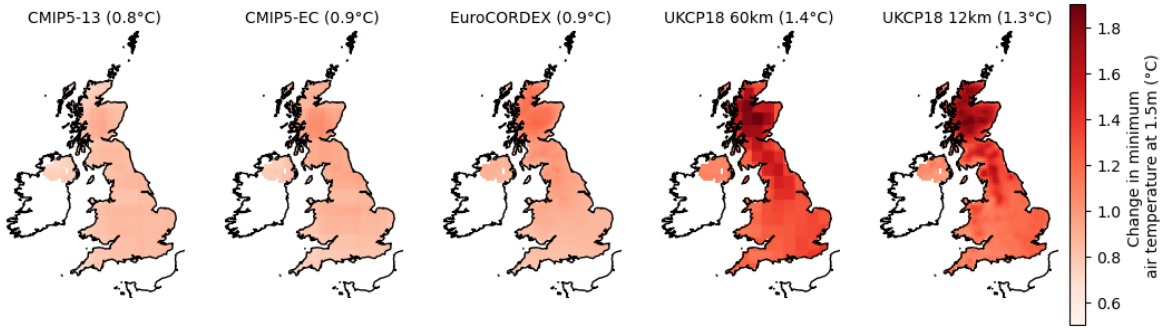
Having considered the projected changes in the seasonal mean temperature tas, we now turn to consideration of changes in daily minimum and maximum temperatures, characterised by the variables

tasmin and tasmax : these represent, respectively, the average minimum nighttime temperature and the average maximum daytime temperature.

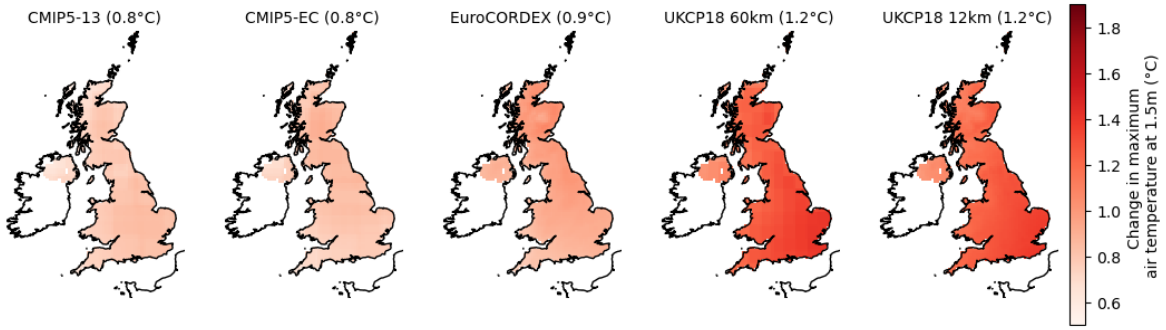
Maps of the mean change in winter tasmin and tasmax in each of the ensembles between the reference period and 2020–2050 are shown in Figure 72; plots for 2050–2080 are not shown, but are qualitatively very similar. The changes in the average daily maximum and minimum temperatures in winter are commensurate with those seen in the daily mean temperature (Figure 63), although in the UKCP18 ensemble, minimum winter temperatures at high elevations increase by more than elsewhere in the UK, indicating in a reduction in the diurnal temperature range in these areas.

Figure 72: Maps of mean change in winter nighttime and daytime temperatures in each ensemble (2020–2050). The average change over the UK land surface for each ensemble is shown in parentheses.

(a) Change in winter daily minimum temperature tasmin, 2020–2050



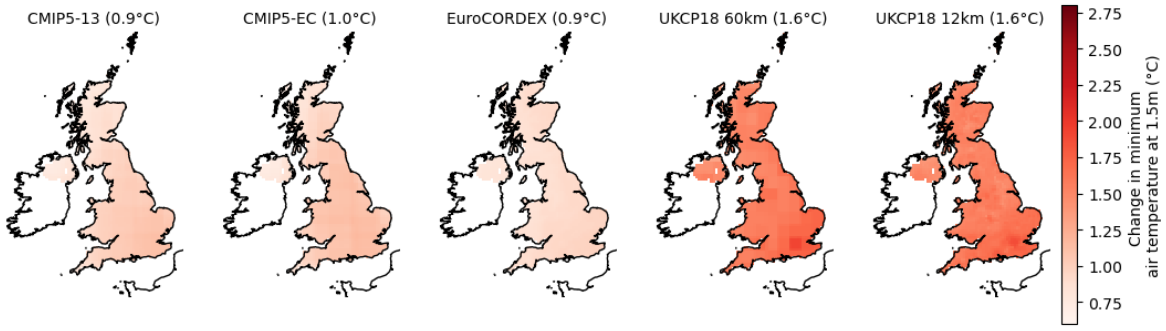
(b) Change in winter daily minimum temperature tasmax, 2020–2050



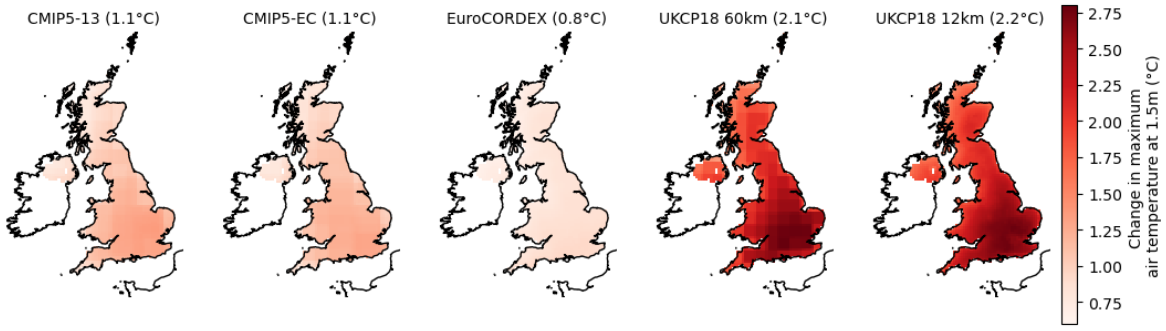
Corresponding plots showing the mean changes in summer tasmin and tasmax are shown in Figure 73. While the CMIP5 and EuroCORDEX ensembles all project similar levels of warming in minimum and maximum daily temperatures, the same is not true of the UKCP18 ensembles, both of which project on average smaller increases in the daily minima than in the daily maxima, suggesting a corresponding increase of around 0.5°C in the average diurnal temperature range. The UKCP18 ensembles also project less uniform warming over the UK in summer, with nighttime temperatures in urban areas projected to increase more than in the surrounding areas, and with daytime temperatures projected to increase more in southern and central England than elsewhere: this drives the greater increase in mean temperatures in this region that was shown in Figure 64. The same pattern is seen in the changes projected in the period 2050–2080 (maps of which are not shown), where tasmax was projected to increase by, on average, 1.2°C more than tasmin over the UK.

Figure 73: Maps of mean change in summer nighttime and daytime temperatures in each ensemble (2020–2050). The average change over the UK land surface for each ensemble is shown in parentheses.

(a) Change in summer daily minimum temperature tasmin, 2020–2050



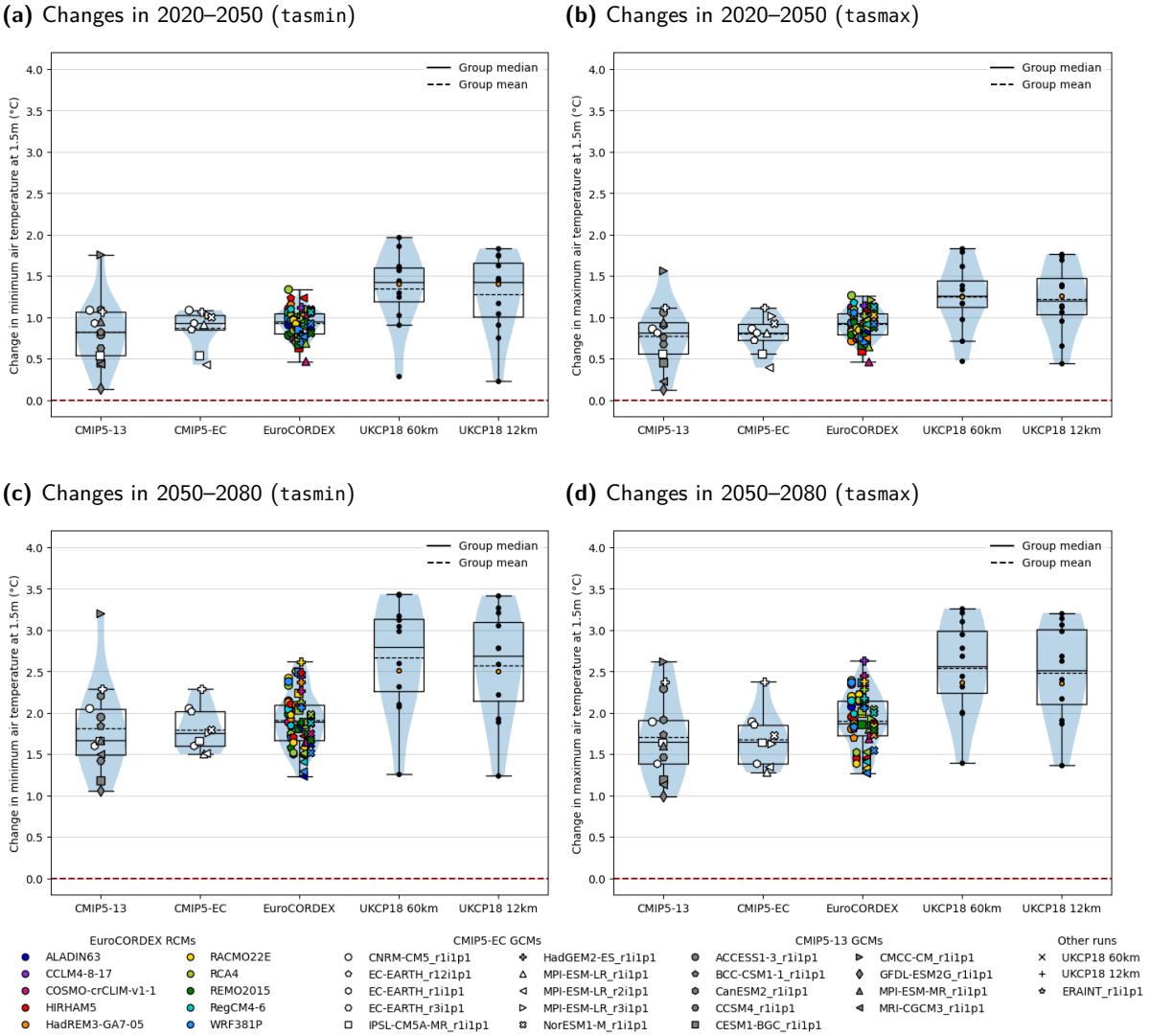
(b) Change in summer daily minimum temperature tasmax, 2020–2050



Boxplots showing the UK-averaged changes in tasmin and tasmax in winter are shown in Figure 74. In both periods, most of the runs project similar changes in tasmin and tasmax. The CMIP5 and EuroCORDEX ensembles project similar ranges of changes in minimum and maximum temperatures, while the UKCP18 ensembles project larger increases on average, and also a much wider range of possible changes; one run — produced by ensemble member 15 — consistently projects one of the smallest increases in temperatures of any of the runs considered.

The corresponding changes in summer temperatures are shown in Figure 75. Most of the CMIP5 and EuroCORDEX runs project similar changes in tasmin and tasmax, although a wider range of changes is projected in tasmax than in tasmin in both of the future periods considered, largely due to HadGEM2-ES simulating a particularly large change in tasmax and CNRM-CM5 a particularly small change in tasmax. The EuroCORDEX runs driven by HadGEM2-ES simulate among the largest changes in both tasmin and tasmax in both periods, while runs using RACMO22E, RCA4 and RegCM4-6 (denoted by yellow, light green and sky blue symbols) simulate smaller changes in tasmax than in tasmin in both periods, indicating a reduction in the diurnal temperature range. All of the UKCP18 ensemble members simulate larger changes in tasmax than in tasmin, with increases in diurnal temperature range of between 0.3 and 1.2°C in the near future, and between 0.7 and 1.8°C in the period from 2050–2080. These results are broadly in agreement with those reported by Lo et al. (2020), who examined changes in the intensity of urban heat islands during summer in both the UKCP18 and EuroCORDEX regional ensembles.

Figure 74: Boxplots of average change in daily minimum and maximum winter temperatures over the UK from the 1980–2010 reference (in °C).

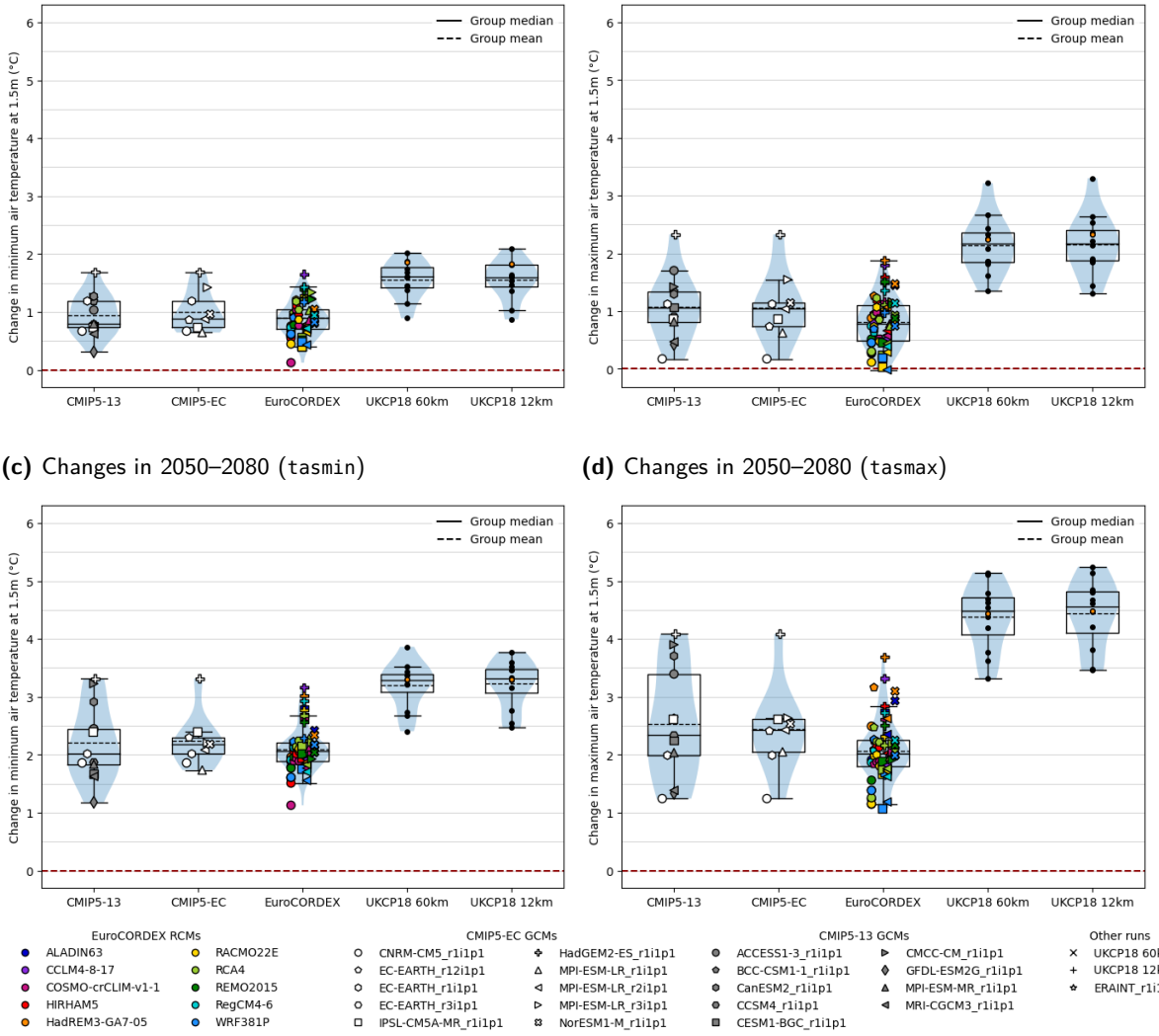


An analysis of variance was again carried out using the methodology described in Section 3.5.2 to investigate systematic variation in the projected changes in seasonal daily maximum and minimum temperatures within the unbalanced EuroCORDEX ensemble; the results of this analysis are shown in Figure 76. In all four panels, differences between the fitted GCM effects account for the largest proportion of the variation between individual runs, with the proportion increasing with each time step when considering winter temperatures (panels a and b) and decreasing with each time step when considering summer temperatures (panels c and d). The RCMs also contribute between 20 and 30% of the variation in winter minima and between 20 and 45% of the variation in summer maximum temperatures, with systematic differences between summer maximum temperatures in runs downscaled by different RCMs contributing almost as much to the spread of changes in 2050–2080 as differences between runs driven by different GCMs. Maps of the proportions of the variance explained by the GCM and RCM effects were found to be very similar to those shown in Figure 68;

Figure 75: Boxplots of average change in daily minimum and maximum summer temperatures over the UK from the 1980–2010 reference (in °C).

(a) Changes in 2020–2050 (tasmin)

(b) Changes in 2020–2050 (tasmax)



(c) Changes in 2050–2080 (tasmin)

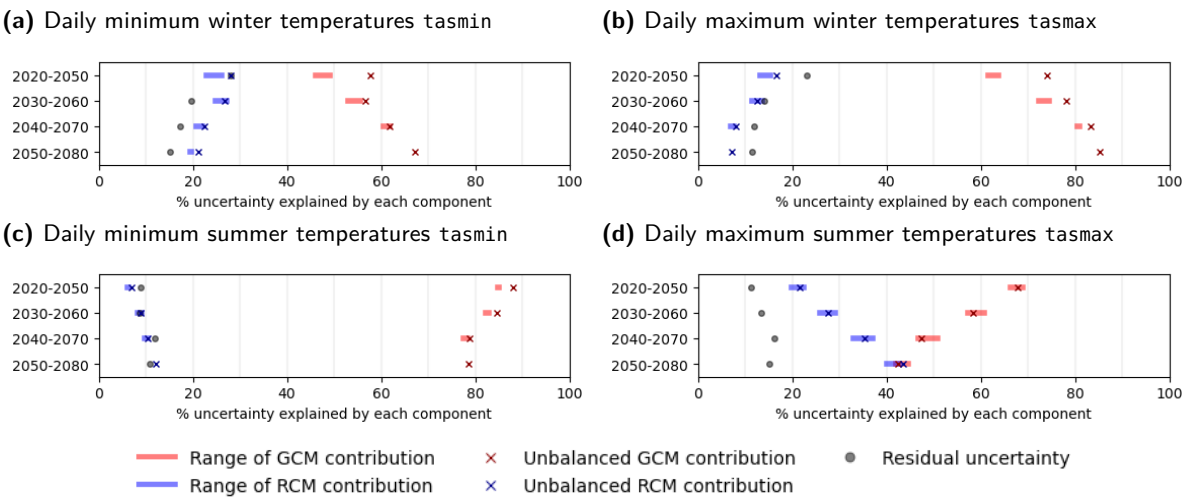
(d) Changes in 2050–2080 (tasmax)

similarly, as in Figures 69, 70 and 71, EPP analysis found that over 90% of the variance attributed to each effect could be described by a relatively uniform offset from the ensemble mean. Plots of these analyses are therefore not presented here.

8.3 Changes in the distribution of daily temperatures

Having considered the projected changes in the seasonal mean temperature tas, and how these are related to changes in the daily minimum and maximum temperatures tasmin and tasmax, we now consider changes in the variables tas01 and tas99, which describe, respectively, the 1st and 99th percentiles of the daily mean temperatures during the winter and summer months. These variables characterise typical temperatures on the coolest and warmest days in a season, and indicate the range of daily temperatures simulated by each model.

Figure 76: Plots of the proportion of the variance of changes in daily maximum and minimum temperature within the EuroCORDEX ensemble that are explained by effects due to the driving GCM and the downscaling RCM, or unexplained. Coloured bars show the range of variance explained when the maximum possible variation is ascribed to either the GCM or the RCM, and the crosses show the proportions of variance explained by GCM and RCM effects fitted simultaneously using the unbalanced ANOVA described in Section 3.5.2. Residual uncertainty is shown in grey.



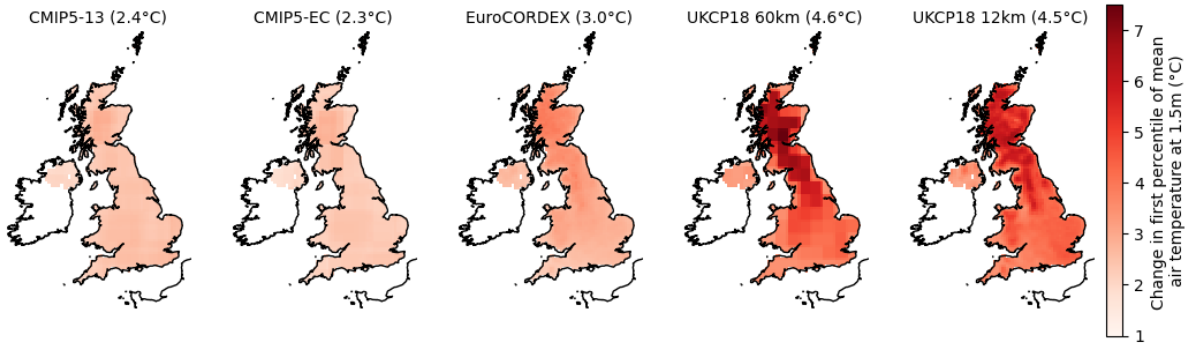
Maps of the ensemble mean changes in `tas01` and `tas99` between the reference period and 2050–2080 during the winter months are shown in Figure 77. Once again, most of the maps show uniform warming across the UK, although the lowest temperatures in the UKCP18 runs increase to a greater extent at higher elevations (panel a). The maps of changes in the near future period are almost identical up to a scaling factor and are therefore not shown. In all of the ensembles, the projected increase in the first percentile of daily winter temperatures is somewhat larger than the projected increase in the 99th percentile, indicating a narrower distribution of daily temperatures about the mean. The UKCP18 ensembles project the largest increases in both `tas01` and `tas99`, and also the largest reduction in the range of daily winter temperatures simulated.

Corresponding maps of the ensemble mean changes in `tas01` and `tas99` during the summer months between the reference period and 2050–2080 are shown in Figure 78; again, maps of changes in the near future period are almost identical to these up to a scaling factor and are therefore not shown. The ensemble mean changes in `tas01` are fairly constant across the UK, while all of the ensembles project a larger increase in `tas99` in the south of England than in Scotland; this pattern is similar to that seen in the changes in daily mean temperatures `tas` in Figure 64. In all of the ensembles, the temperature on the hottest summer days `tas99` is projected to increase by more than on the coolest summer days: the spread of summer temperatures is therefore projected to increase over time. The UKCP18 ensembles again project the largest increases in both `tas01` and `tas99`, and also in the spread of temperatures.

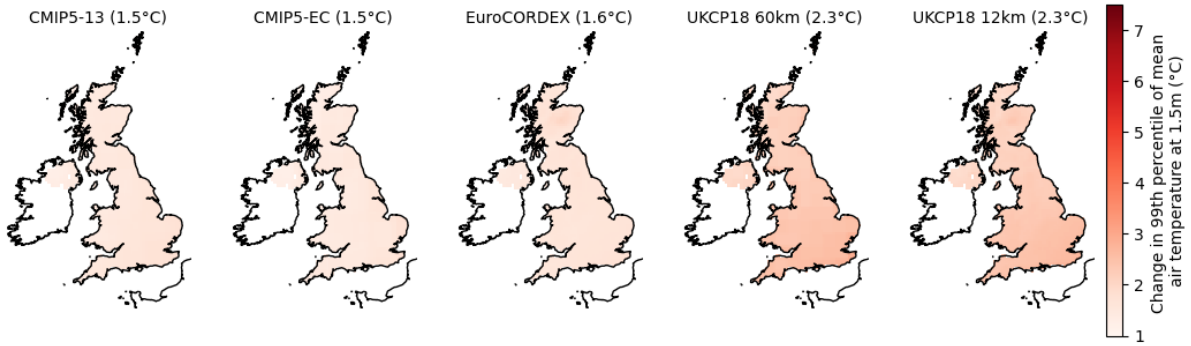
Boxplots of the average change in UK-averaged winter `tas01` and `tas99` in each run are shown in Figure 79. The coldest winter temperatures `tas01` are not only projected to increase by more than the warmest temperatures `tas99`, but the differences between individual projections are much

Figure 77: Maps of mean change in 1st and 99th percentiles of daily winter mean temperatures in each ensemble (2050–2080). The average change over the UK for each ensemble in shown in parentheses.

(a) Changes in 1st percentile of distribution of daily winter temperatures, `tas01`



(b) Changes in 99th percentile of distribution of daily winter temperatures, `tas99`

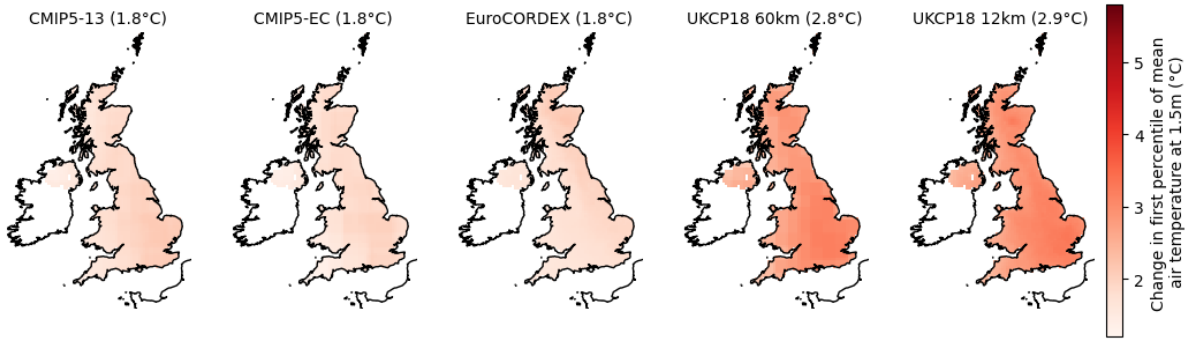


larger in all ensembles. The UKCP18 runs project a wider range of changes in `tas01` than any other ensemble, although all but one or two of the runs project increases in both `tas01` and `tas99` that are larger than the upper quartile of the changes projected by the EuroCORDEX ensemble members.

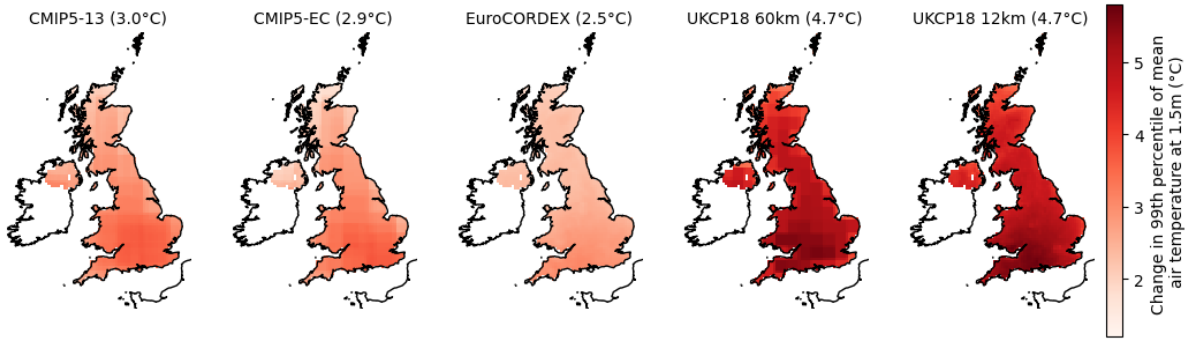
The distribution of changes in summer `tas01` and `tas99` projected by the individual models in each ensemble are shown in the boxplots in Figure 80. In both periods, the spread of changes in `tas01` within each of the ensembles is around 1.5°C, with the UKCP18 ensemble again projecting larger increases than the CMIP5 and EuroCORDEX ensembles; within the EuroCORDEX ensemble, runs are again grouped according to the driving GCM — denoted by different symbols — and again the runs driven by HadGEM2-ES produce the largest increases in temperature. The 99th percentiles of the projected temperatures are more variable, with several of the EuroCORDEX runs projecting a decrease in the hottest summer temperatures in the near future: these runs are driven by IPSL-CM5A-MR (squares), MPI-ESM-LR ensemble member r2i1p1 (left-pointing triangles), and in one case CNRM-CM5 (circles). In both periods, the largest increases in `tas99` in the EuroCORDEX and CMIP5 ensembles are again projected by HadGEM2-ES, and are broadly in line with the increases projected by the UKCP18 ensembles: the difference between the increases in `tas99` projected by HadGEM2-ES and UKCP18 and by the majority of the other models is larger in 2050–2080 than in 2020–2050, with most of the UKCP18 ensemble members projecting changes somewhat above the

Figure 78: Maps of mean change in 1st and 99th percentiles of daily summer mean temperatures in each ensemble (2050–2080). The average change over the UK for each ensemble in shown in parentheses.

(a) Changes in 1st percentile of distribution of daily summer temperatures



(b) Changes in 99th percentile of distribution of daily summer temperatures



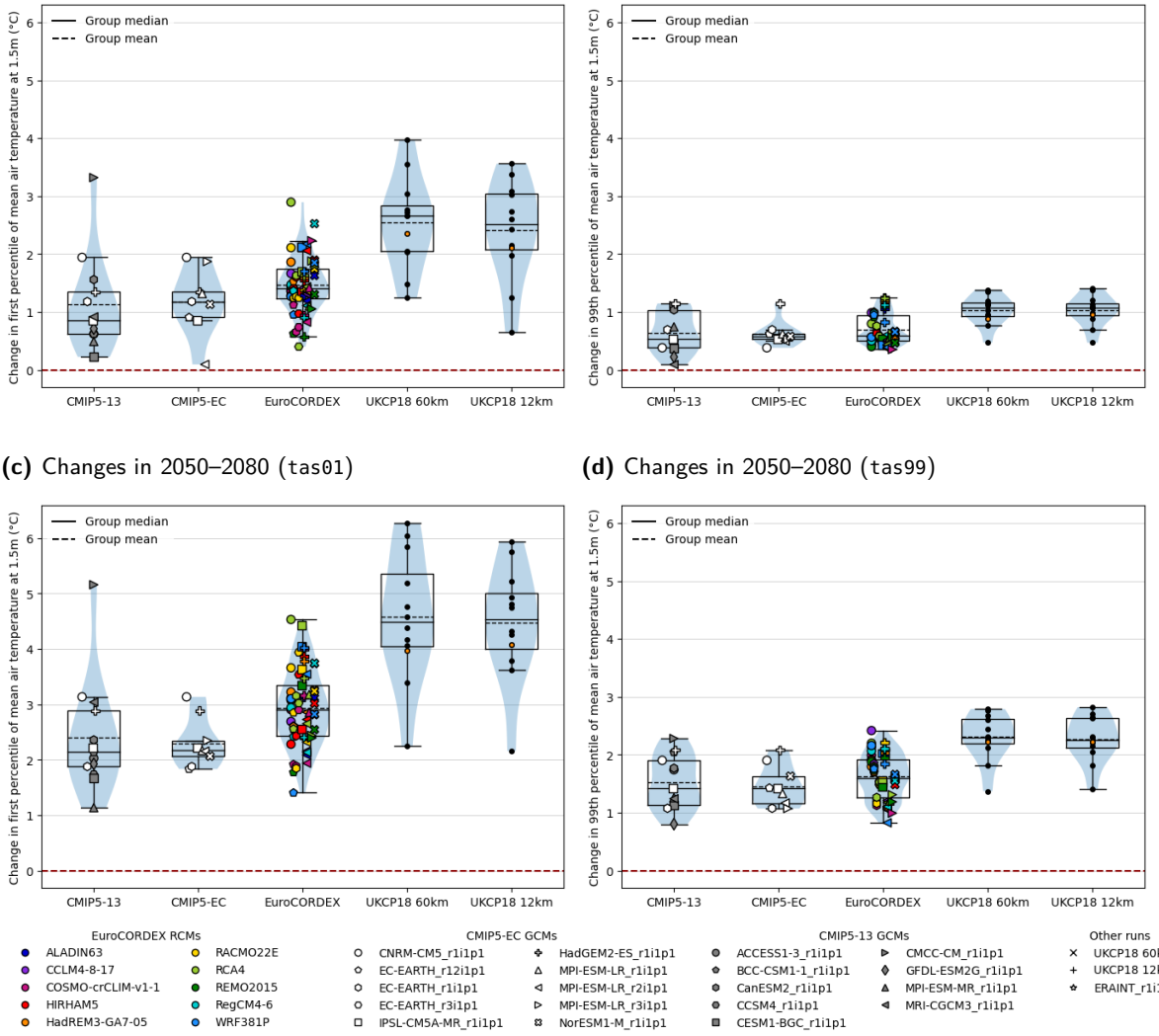
upper quantile of changes in the EuroCORDEX ensemble by 2050–2080.

An analysis of variance was carried out using the methodology described in Section 3.5.2 to investigate the sources of systematic variation in the projected changes in the 1st and 99th percentiles of temperatures within the unbalanced EuroCORDEX ensemble: plots showing the proportion of variance assigned to each model component are shown in Figure 81. Most of the variability in the 99th percentile of winter temperatures (panel b) and in the 1st percentile of summer temperatures (panel c) can be attributed to systematic differences between the fitted GCM effects. The majority of the variability in the 1st percentile of winter temperatures — which are generally projected to increase by more than the 99th percentile of winter temperatures, with a wider range of increases projected — cannot be attributed to systematic differences between either the GCMs or the RCMs (panel a). Residual uncertainty plays a similarly large role in variation in the 99th percentiles of summer temperatures (panel 81b), although the GCMs also contribute a substantial part of the residual uncertainty, particularly in the nearer term. Maps of the proportion of the total variance explained by the GCMs were very similar to those in Figure 68, and so are not shown here; similarly, in those cases where the GCM effects were found to explain a large proportion of the variance within the ensemble, at least 85% of the variance was attributed to a fairly uniform offset in temperatures across the UK, so maps of the EPP analysis are not shown here.

Figure 79: Boxplots of average change in 1st and 99th percentiles of daily winter temperatures (`tas01` and `tas99`) from the 1980–2010 reference (in °C).

(a) Changes in 2020–2050 (`tas01`)

(b) Changes in 2020–2050 (`tas99`)



(c) Changes in 2050–2080 (`tas01`)

(d) Changes in 2050–2080 (`tas99`)

9 Projected changes in precipitation

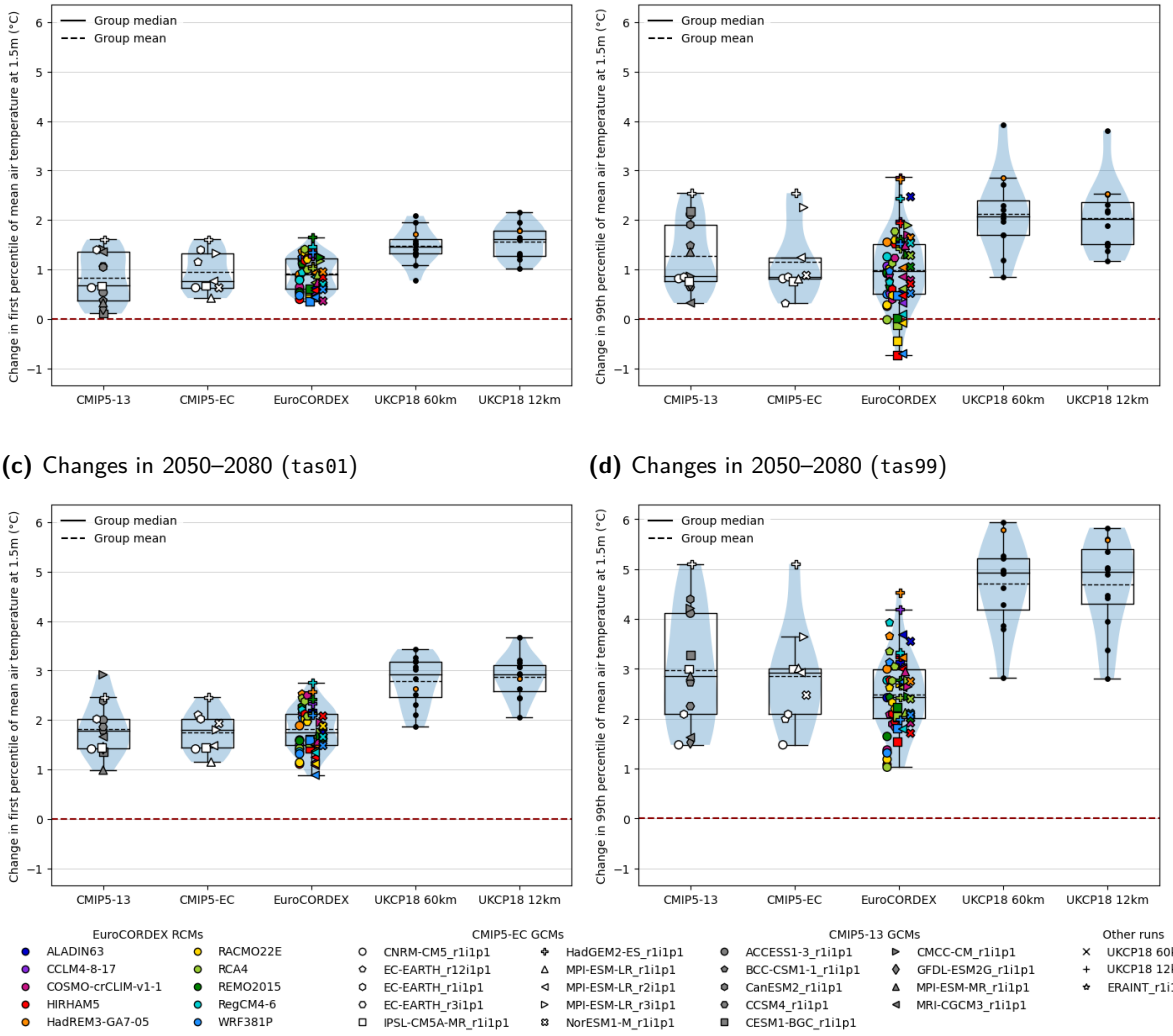
Having considered projected changes in temperature, we now turn to considering projected changes in precipitation over the UK in the summer and winter months, both in the near future period (2020–2050) and in the future (2050–2080). As in Section 5, the analysis will first consider changes in the average daily precipitation rate `pr`, which is related to the total seasonal accumulated rainfall, and how those changes are driven by changes in either the mean wet-day precipitation rate (also known as the standardised daily intensity index, `sdii`) and the proportion of wet days, `fwd`. The change in the proportion of precipitation occurring on very wet days, `r99ptot`, is also considered. These variables are summarised in Table 5.

All changes in precipitation indices are reported as a percentage change with reference to the reference period (1980–2010) unless otherwise stated.

Figure 80: Boxplots of average change in 1st and 99th percentiles of daily summer temperatures (`tas01` and `tas99`) from the 1980–2010 reference (in °C).

(a) Changes in 2020–2050 (`tas01`)

(b) Changes in 2020–2050 (`tas99`)



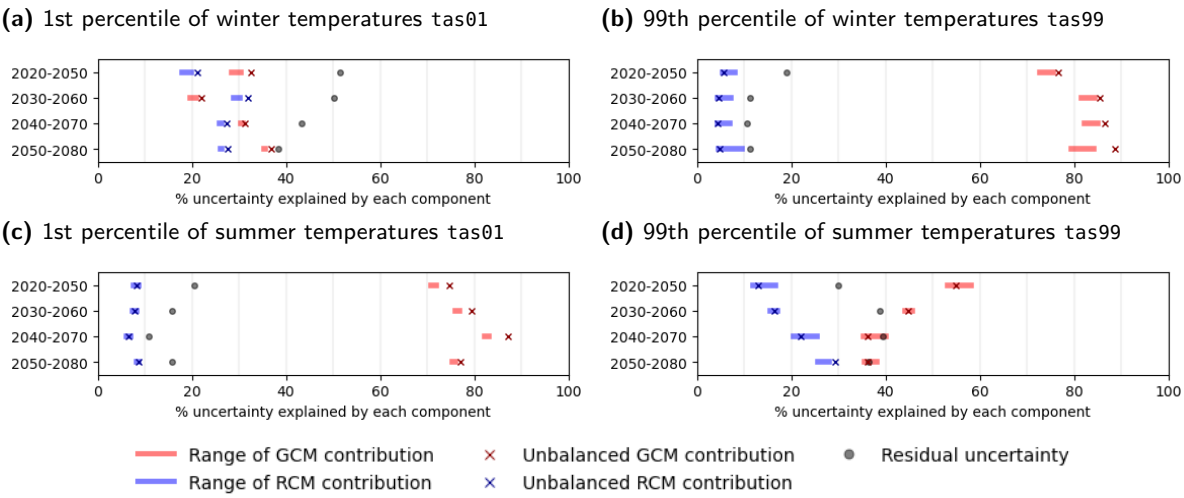
(c) Changes in 2050–2080 (`tas01`)

(d) Changes in 2050–2080 (`tas99`)

9.1 Projected changes in winter precipitation

Maps of the ensemble mean changes of the projected percentage changes in precipitation rates in the near future are shown in Figure 82a. In the CMIP5 ensembles there is a fairly uniform percentage increase in precipitation across the UK, while the EuroCORDEX ensemble mean has a somewhat smaller increase over higher elevations. While precipitation rates averaged over the whole of the UK also increase in the UKCP18 ensembles, they do not do so uniformly across the UK, with smaller relative increases projected at higher elevations, and with northern Scotland projected to be drier in the near future. The patterns of relative changes in `sdii` in the CMIP5 and EuroCORDEX ensemble means are closely related to those in `pr`, with very little change projected in the proportion of wet days occurring across the UK (panels 82b and 82c). In the UKCP18 ensemble, more precipitation is simulated on wet days, but fewer wet days are simulated, particularly in rain shadows and in the east of England.

Figure 81: Plots of the proportion of the variance of changes in temperature within the EuroCORDEX ensemble that are explained by effects due to the driving GCM and the downscaling RCM, or unexplained.



Looking further ahead to the period from 2050–2080, the spatial patterns of projected percentage changes in precipitation rates are broadly similar up to a scaling factor (Figure 83, panels a and b), although the projected drying over northern Scotland in the UKCP18 ensemble is less severe. The CMIP5 and EuroCORDEX ensemble means again exhibit very small changes in fwd across the UK, while the UKCP18 ensemble means project a small percentage increase in the proportion of wet days along the western coast, and a small decrease in rain shadow areas, along the east coast of England, and particularly along the east coast of Scotland.

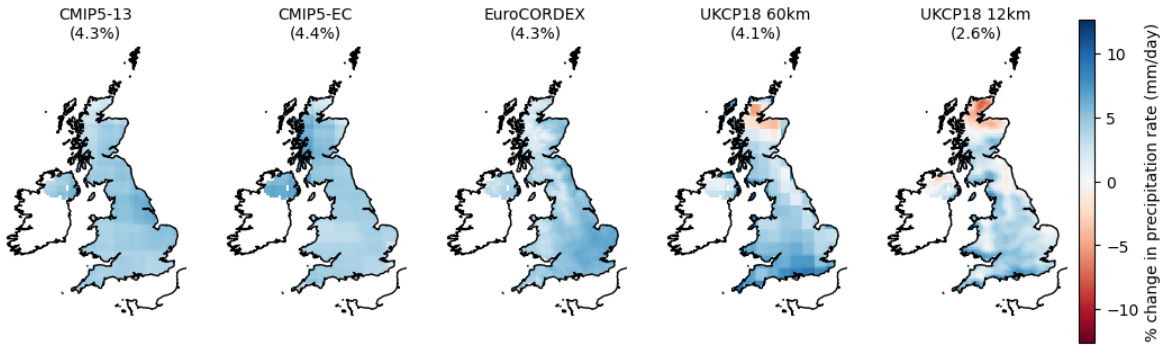
Boxplots showing the distribution of the UK-averaged percentage changes in the near future are shown in Figure 84. Around 75% of the runs in each ensemble project an increase in winter precipitation, with most projecting up to 10% more precipitation (panel a). The CMIP5 and EuroCORDEX runs project very little change in the proportion of wet days (panel b) but almost all runs simulate a small percentage increase in the wet-day precipitation rate, which drives the change in overall precipitation. The UKCP18 runs also simulate more precipitation on wet days, but with precipitation occurring on fewer days.

Looking further ahead (Figure 85a), both the EuroCORDEX and CMIP5-13 ensembles project a wide range of changes, with most of the runs projecting between 5 and 15% more precipitation in winter, although a handful of runs simulate less precipitation during this period than during the reference period. The UKCP18 ensembles project somewhat larger percentage increases in overall precipitation by this period, with all of the runs simulating larger changes than the lower quartile of the EuroCORDEX ensemble and the UKCP18 mean approximately equal to the upper quartile of the EuroCORDEX ensemble. There is no strong trend in the proportion of wet days simulated by any of the ensembles, with the relative change in overall precipitation driven primarily by the relative change in wet-day precipitation (panel c).

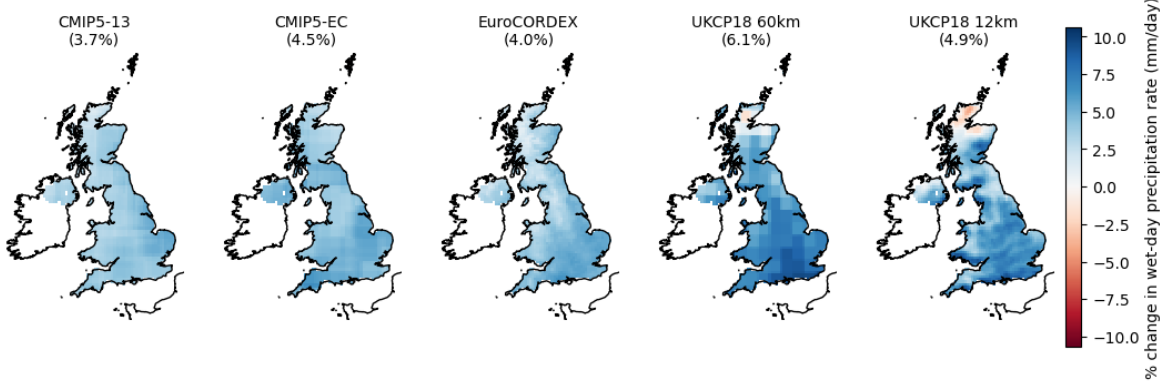
Taylor diagrams similar to those in Figure 33a were produced by plotting the correlation of the

Figure 82: Maps of average relative changes in mean winter precipitation rate `pr`, wet-day precipitation rate `sdii` and proportion of wet days `fwd` in each of the ensembles between the reference period (1980–2010) and the near future (2020–2050). The average change over the whole of the UK for each ensemble is shown in parentheses.

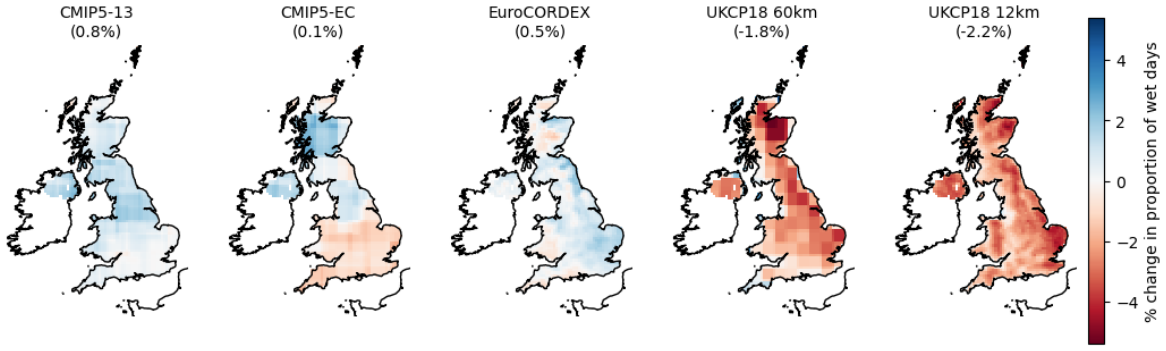
(a) Change in daily mean winter precipitation rates `pr`, 2020–2050



(b) Change in winter wet-day precipitation rates `sdii`, 2020–2050



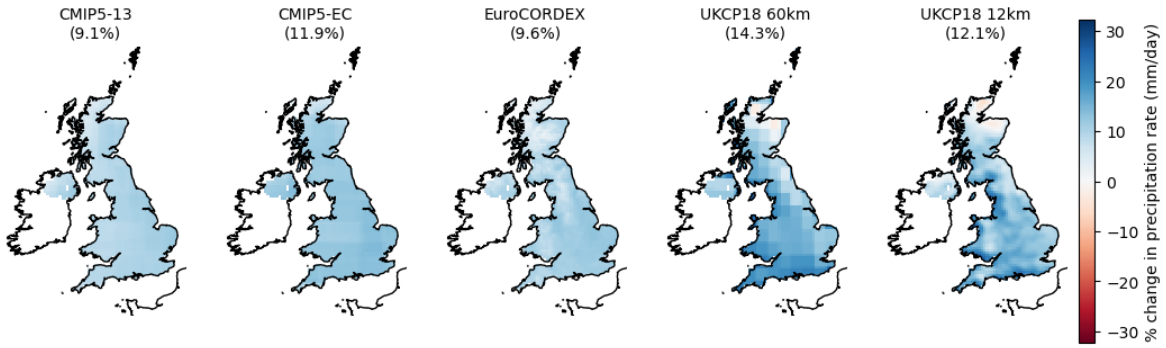
(c) Change in proportion of wet winter days `fwd`, 2020–2050



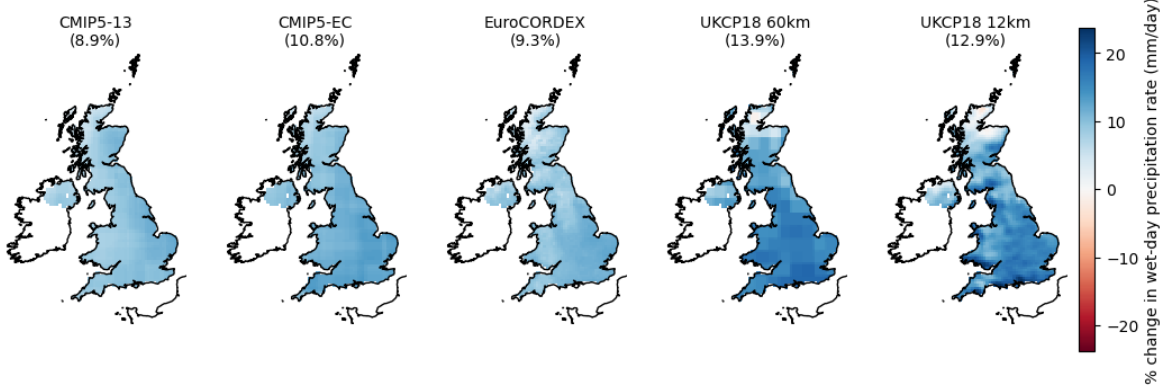
spatial fields of `pr`, `sdii` and `fwd` in the future projections with those in the reference period, and the standard deviation of the future projections normalised by dividing by the standard deviation of the reference simulation. The plots are not included in this report, but in both periods and in all three indices, correlation between the future and reference patterns was greater than 0.95 in all model runs, and the standard deviations of the spatial fields in the future period between 80% and 120% of those in the reference period, suggesting that the climatological spatial patterns of precipitation do not change significantly in future projections.

Figure 83: Maps of average changes in mean winter precipitation rate `pr`, wet-day precipitation rate `sdii` and proportion of wet days `fwd` in each of the ensembles between the reference period (1980–2010) and the future period (2050–2080). The average change over the whole of the UK for each ensemble is shown in parentheses.

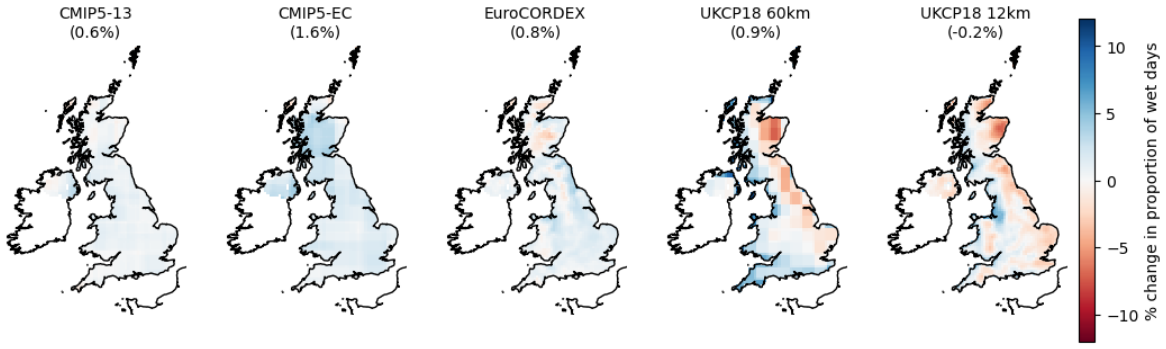
(a) Change in daily mean precipitation rates `pr`, 2050–2080



(b) Change in daily mean wet-day precipitation rates `sdii`, 2050–2080



(c) Change in proportion of wet days `fwd`, 2050–2080



An unbalanced ANOVA was carried out over the relative changes in `pr`, `sdii` and `fwd` projected by the members of the EuroCORDEX ensemble, to investigate sources of systematic variation between the model runs; the results of this analysis are shown in Figure 86. Between 50 and 60% of the variation between the model runs in both `pr` and `fwd` is attributed to the choice of GCM in all time periods, and around 40% of the variation in `sdii`, although a similar proportion of the variation in `sdii` cannot be explained by differences between the GCM or RCM effects. In all three indices the contribution from the fitted RCM effects, while smaller, increases in successive time steps as residual

Figure 84: Boxplots of projected changes in winter precipitation indices during the period 2020–2050, averaged over the UK and expressed as a percentage change from the UK average of the index during the reference period.

(a) Change in `pr`, 2020–2050 **(b)** Change in `fwd`, 2020–2050 **(c)** Change in `sdii`, 2020–2050

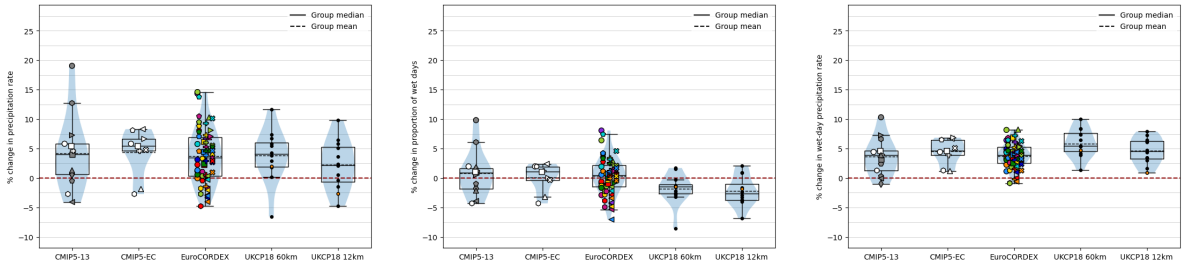
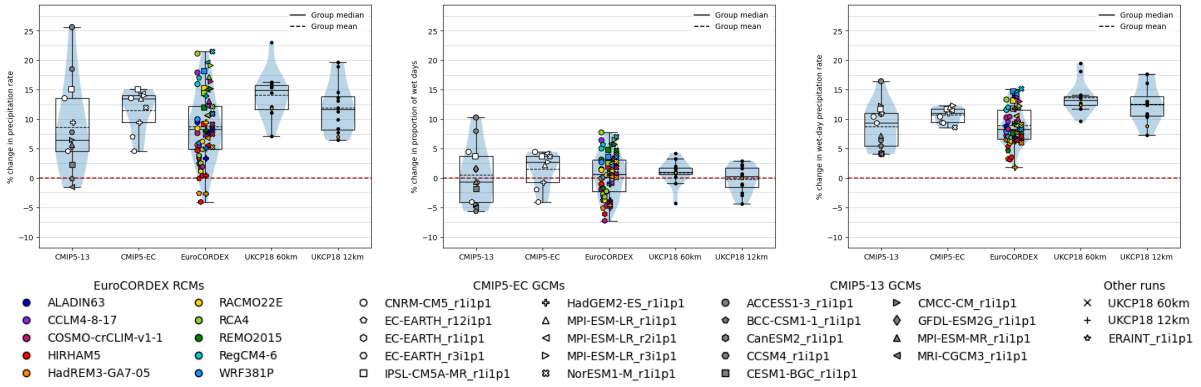


Figure 85: Boxplots of projected changes in winter precipitation indices during the period 2050–2080, averaged over the UK and expressed as a percentage change from the UK average of the index during the reference period.

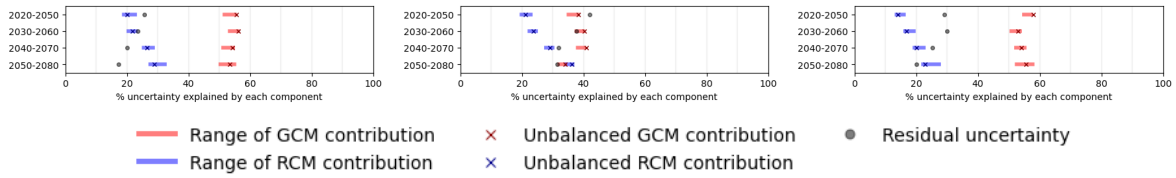
(a) Change in `pr`, 2050–2080 **(b)** Change in `fwd`, 2050–2080 **(c)** Change in `sdii`, 2050–2080



uncertainty decreases.

Figure 86: Plots of the proportion of the variance of changes in precipitation indices within the EuroCORDEX ensemble that are explained by effects due to the driving GCM and the downscaling RCM, or unexplained.

(a) `pr` **(b)** `sdii` **(c)** `fwd`

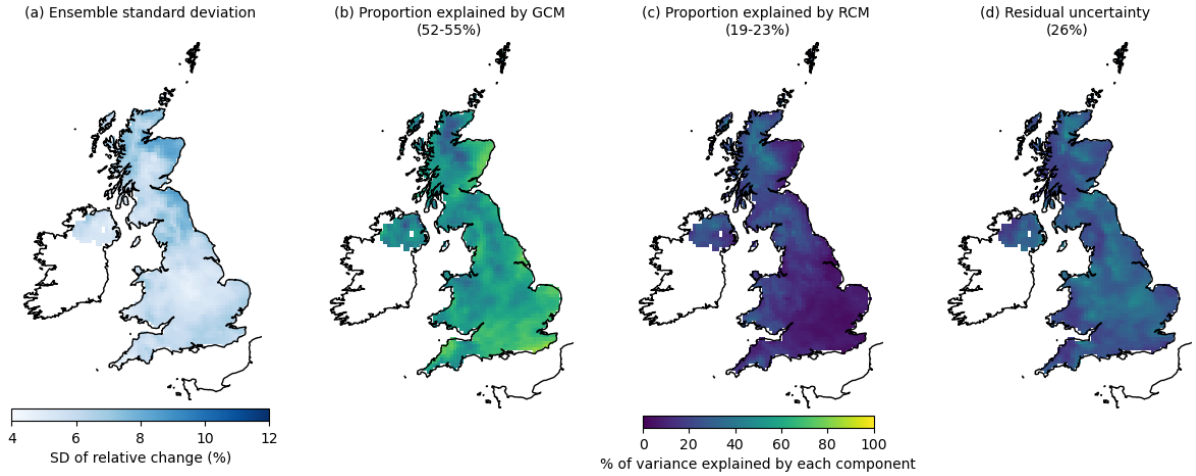


Maps of the areas of greatest variation in percentage changes in overall precipitation rate within the EuroCORDEX ensemble, and of the proportion of that variation explained by each model component, are shown in Figure 87. The highest variation within the ensemble occurs in Scotland and around the edges of mountainous regions. In the near term, systematic variation arising from differences between runs driven by different GCMs account for between 50 and 55% of the total variance across

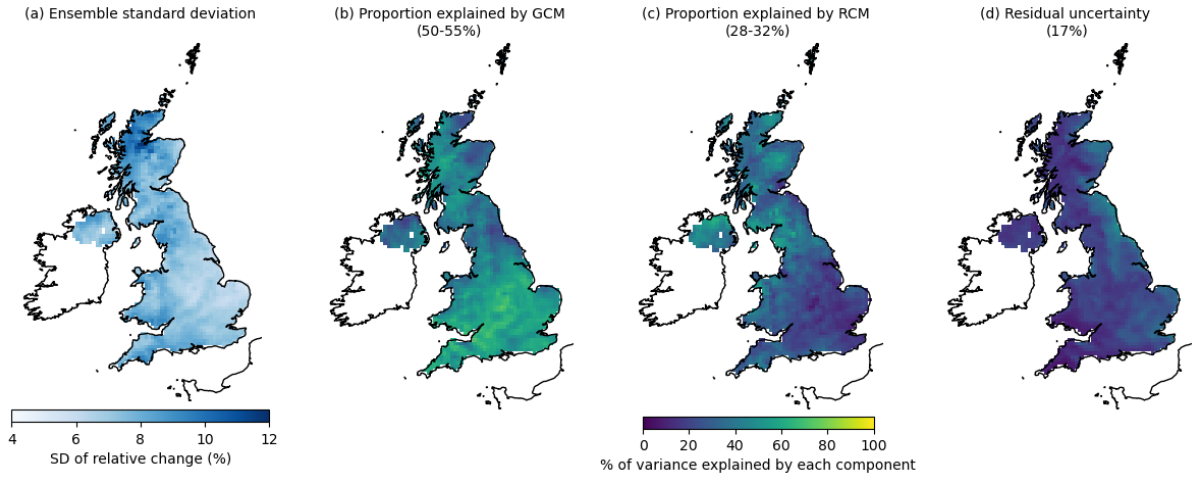
the UK, and up to 70% in coastal and low-lying regions. Differences between runs downscaled using different RCMs account for up to 30% of the local variation across the UK in the near term, with the largest contributions in Scotland and northern England, while residual uncertainty can contribute up to 60% of the variation in rain shadows. A similar spatial pattern appears in the period from 2050–2080, although the proportion of local variation attributed to the RCM effects increases, with up to 70% of the variation in Northern Ireland and over high elevations explained by the RCMs.

Figure 87: Maps of the standard deviation of relative changes in the winter precipitation rate pr in the EuroCORDEX ensemble in each grid cell, and of the percentage of the total ensemble variance in each grid cell explained by the choice of GCM and RCM.

(a) Variation in winter precipitation rate pr, 2020–2050



(b) Variation in winter precipitation rate pr, 2050–2080

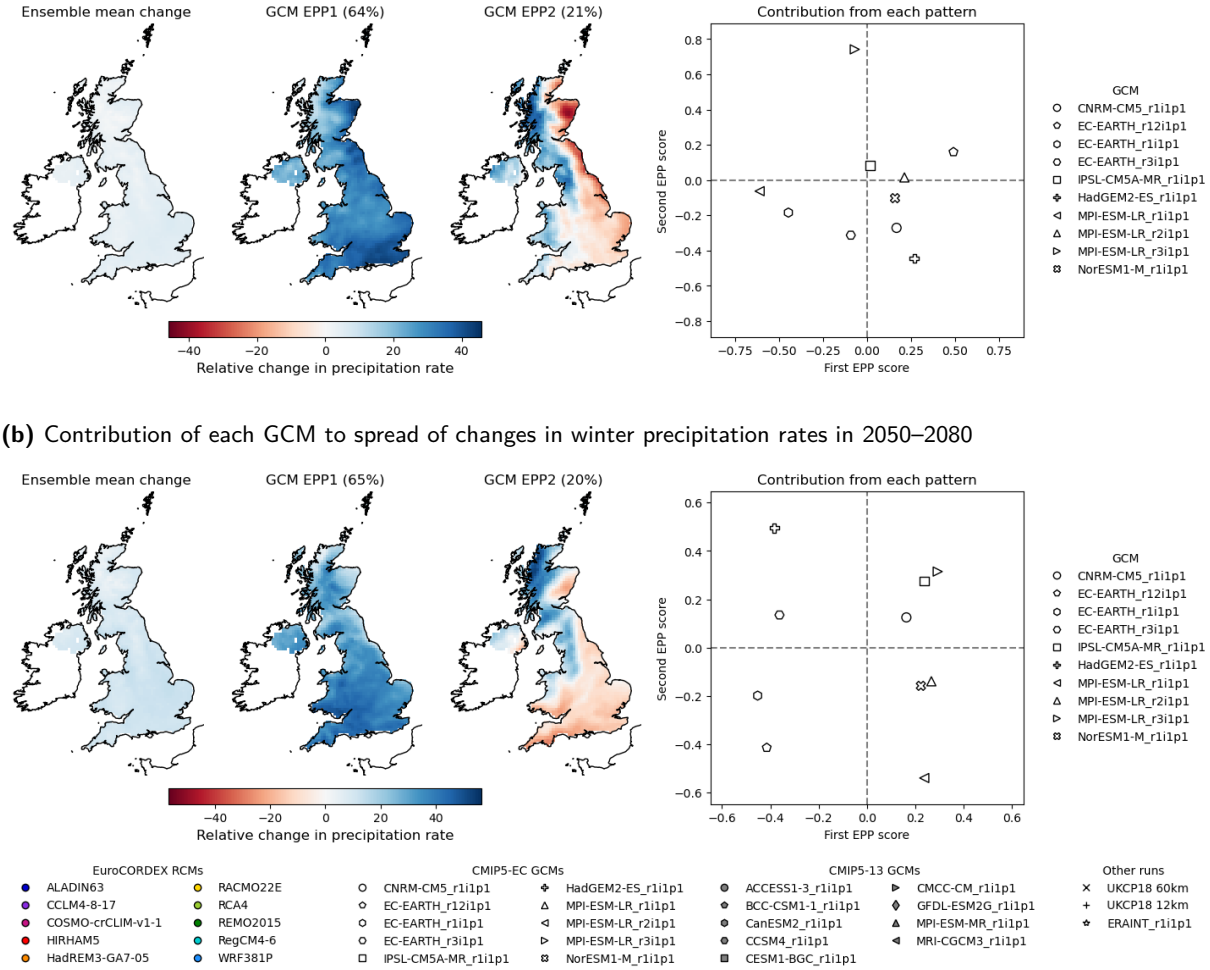


EPP analysis was used to investigate the dominant modes of spatial variation arising from differences between the fitted GCM effects. Maps of the principal modes of variation in the relative change in precipitation are shown in Figure 88, along with the scores assigned to each of the GCMs for the first two EPPs. The spatial patterns are similar in both periods, with the first EPP (EPP1), which accounts for 64-65% of the systematic variation between the GCM effects, representing a fairly uniform increase in mean daily precipitation rates across the UK; runs driven by GCMs with positive

scores in this component project larger changes over the UK than the EuroCORDEX ensemble average, while runs driven by GCMs with negative scores project smaller changes than the ensemble average. The second component, which accounts for 20-21% of the systematic variation, reflects a contrast between increased precipitation to the west of the highest elevations, and decreased precipitation to the east.

Figure 88: Ensemble principal patterns of changes in winter precipitation rates pr between the reference period and future periods, showing dominant patterns of contributions from each GCM to variation about the EuroCORDEX ensemble mean change. The proportion of the between-GCM variation attributed to each pattern is given in parentheses.

(a) Contribution of each GCM to spread of changes in winter precipitation rates in 2020–2050



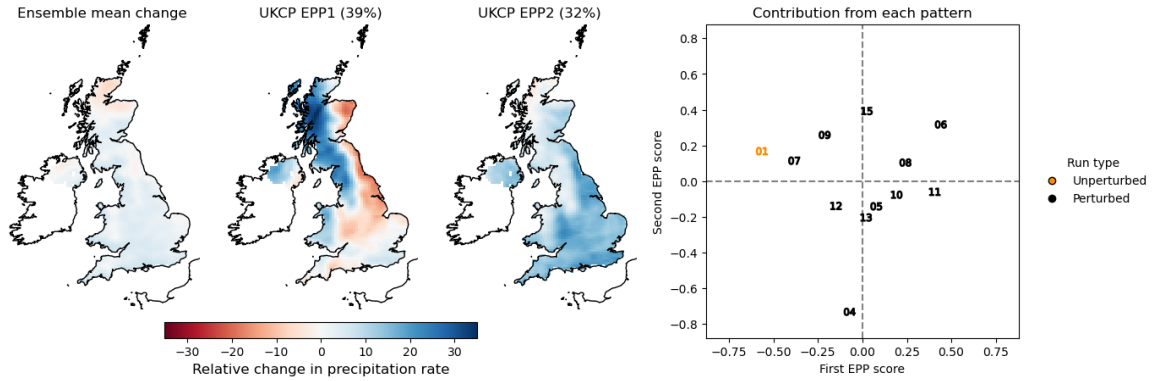
(b) Contribution of each GCM to spread of changes in winter precipitation rates in 2050–2080

EPP analysis of the dominant modes of spatial variation was also carried out over the UKCP18 runs. The patterns of variation in between the runs are not so easily categorised as the patterns of variation between runs driven by different GCMs: in both periods, the first EPP accounts for slightly less than half of the total variance, and the second for around one third of the total variance in the near term, and a quarter looking further ahead. Runs with positive scores in the first component will be wetter to the west of higher elevations and drier to the east; runs with positive scores in the second will have higher-than-average relative changes in precipitation in drier areas, particularly

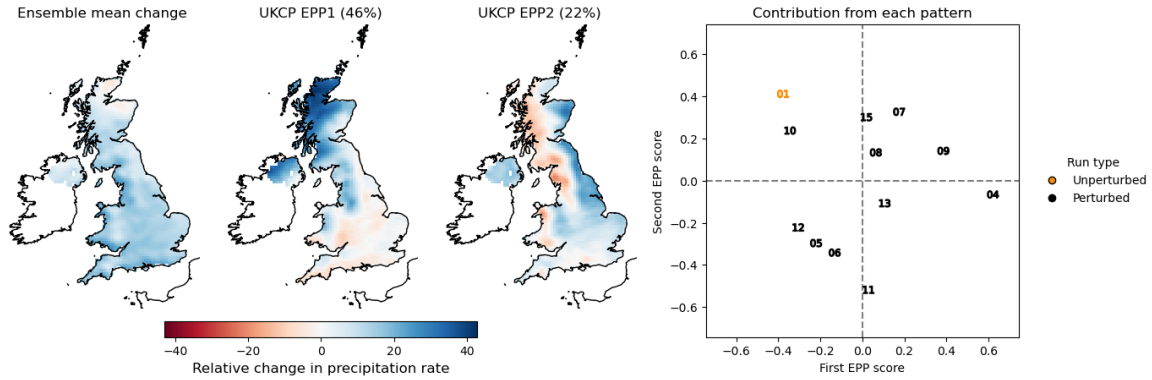
in rain shadows. Both of these patterns — which, together, account for around 70% of the total variation in the ensemble — represent differences in the simulation of precipitation in western areas where winter precipitation is typically higher, and in the east of the country where precipitation is lower due to rain shadow effects. This is a different pattern of variation to that shown in the GCMs, where a large proportion of the differences between the fitted effects can be attributed to a uniform scaling of overall winter precipitation across the UK.

Figure 89: EPP analysis of changes in winter precipitation rates pr simulated by the UKCP18 regional models between the reference period and future periods, showing the dominant patterns of variation. The proportion of the total variation attributed to each pattern is given in parentheses.

(a) EPPs of changes in 2020–2050



(b) EPPs of changes in 2050–2080



9.2 Projected changes in summer precipitation

We turn now to analysing the changes in summer precipitation over the UK. Maps of the relative changes in summer precipitation rates in 2020–2050 with respect to the reference period of 1980–2010 are shown in Figure 90, with the corresponding maps of relative changes in the period 2050–2080 in Figure 91. Almost all of the ensemble means exhibit a reduction in summer precipitation, which in the CMIP5 and EuroCORDEX ensembles is relatively small and uniform across the UK. The UKCP18 ensembles project much larger reductions in summer precipitation, increasing from almost no change in northern Scotland to around 23% less precipitation along the south coast of England (panel a). These changes are very similar both in spatial pattern and in magnitude to those

seen in the relative changes in fwd (panel b), suggesting that the change in overall precipitation is driven largely by the change in the proportion of wet days simulated. The CMIP5 and EuroCORDEX ensembles all project a small increase in the rate of precipitation on wet days across the UK (panel c), to some extent offsetting the effect of the reduced number of wet days on overall precipitation rates. The UKCP18 ensemble means show an increase of up to 6% in wet-day precipitation in rain shadow areas and in eastern England and Scotland, as well as over London, but a reduction of up to 10% elsewhere, with the largest reductions in the areas that currently receive the most summer rainfall (Figure 38a), compounding the effect of the reduction in the proportion of wet days.

The corresponding maps of relative changes in summer precipitation the period 2050–2080 are shown in Figure 91; the pattern of changes is very similar to that in Figure 90, but the changes are of greater magnitude, with the CMIP5 and EuroCORDEX ensembles projecting 10% less precipitation over the UK than in 1980–2010, and with the UKCP18 ensembles projecting around 25% less on average, and up to 40% less in southern England.

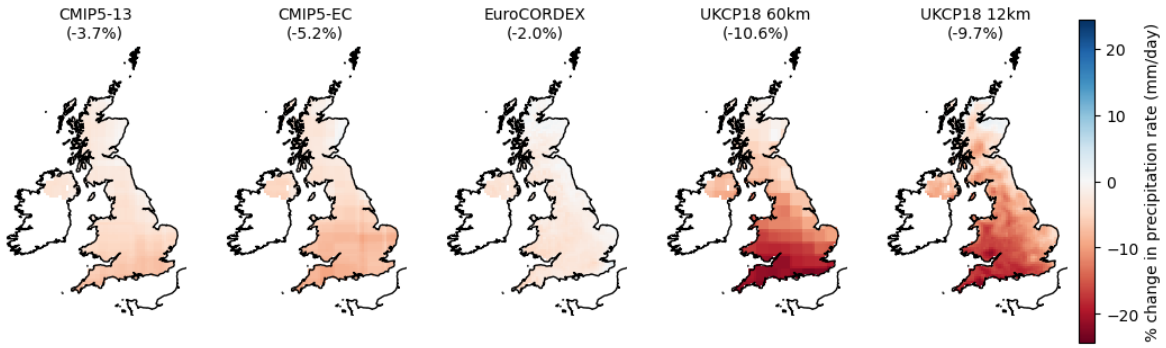
Boxplots of the distribution of UK-averaged changes in summer precipitation within each ensemble are shown in Figure 92. The distribution of near-term projected changes in the two CMIP5 ensembles are fairly skewed, with one model — CNRM-CM5 — projecting an increase of 10% in summer precipitation across the UK, and all but one of the remaining models projecting reductions of up to 12.5% (panel a). While the majority of the EuroCORDEX projections simulate less summer precipitation, the spread of changes within the 65-member ensemble is wide enough to include 22 models with projected increases during this period. The UKCP18 ensembles, in contrast, all project reductions in summer precipitation, with six of the twelve models projecting reductions of between 5 and 10%, and three projecting even larger reductions.

The distribution of changes in wet-day precipitation rates sdii is much narrower (panel b), with only a handful of runs projecting changes of more than 5% from the reference rate in either direction; the majority of the CMIP5 and EuroCORDEX runs project a small increase in wet-day precipitation rates, while the UKCP18 60km runs are evenly split between increasing and decreasing sdii, and most of the UKCP18 12km runs project a small reduction. In contrast, the distributions of changes in the proportion of wet days are very similar to those of the changes in overall precipitation rates (panel c), lending further support to the hypothesis that changes in summer precipitation are driven largely by changes in the number of wet days simulated, rather than by the amount of precipitation falling on wet days.

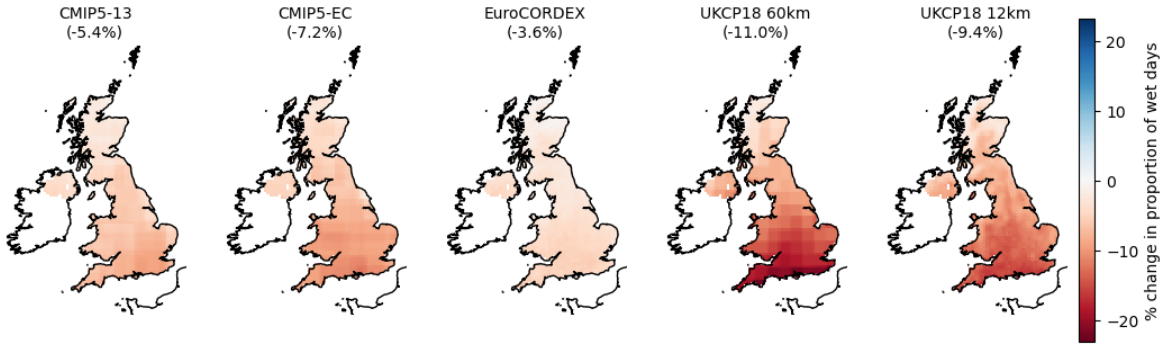
By 2050–2080 almost all of the runs project a reduction in summer rainfall, although CNRM-CM5 still simulates a 7.5% increase in precipitation compared to the reference period; six of the eight EuroCORDEX runs that simulate an increase in summer precipitation across the UK are driven by this GCM (panel d). The largest reduction in summer precipitation in the CMIP5 ensembles is produced by HadGEM2-ES, the model that also simulated the largest increase in summer temperatures in Section 8; however, this tendency is not inherited directly by the EuroCORDEX runs driven by that GCM, which produce a wide spread of changes in precipitation, including both the largest decrease and the second largest increase in the EuroCORDEX ensemble. All of the UKCP18 runs — which

Figure 90: Maps of changes in mean summer precipitation rate `pr`, wet-day precipitation rate `sdii` and proportion of wet days `fwd` in each of the ensembles between the reference period (1980–2010) and the near future (2020–2050). The average change over the whole of the UK for each ensemble is shown in parentheses.

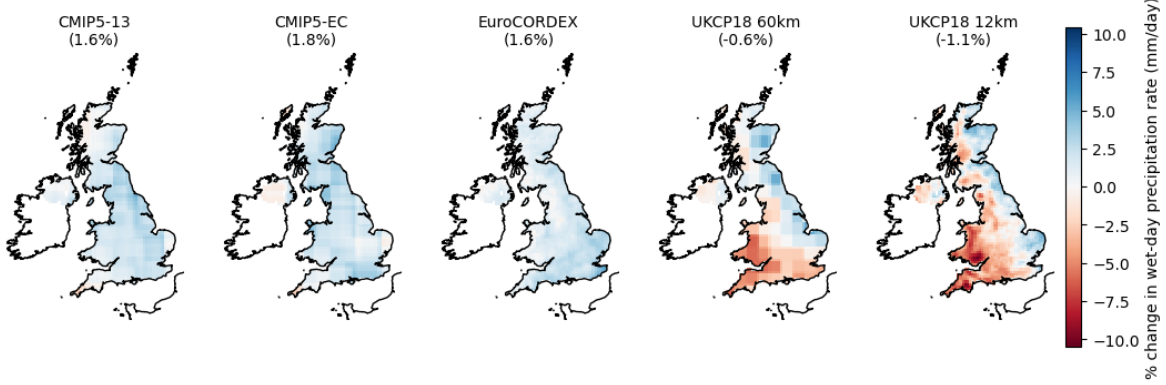
(a) Relative change in daily mean precipitation rates `pr`, 2020–2050



(b) Relative change in proportion of wet days `fwd`, 2020–2050



(c) Relative change in daily wet-day precipitation rates `sdii`, 2020–2050

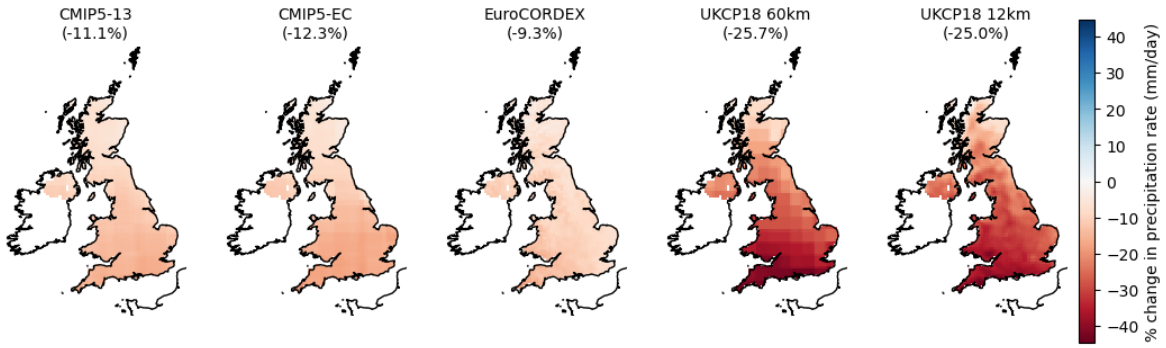


are driven by HadGEM3-GC0.5, a model related to HadGEM2-ES — project reductions of more than 10%, larger than the mean of the EuroCORDEX changes, with an average reduction of around 22.5%, larger than any run other than HadGEM2-ES. Again, the changes in overall precipitation are highly correlated with the changes in the proportion of wet days, with reductions in `fwd` in the CMIP5 and EuroCORDEX ensembles generally slightly mitigated by small increases in `sdii`, and reductions in `fwd` in the UKCP18 ensembles typically compounded by small decreases in `sdii` (panels e and f).

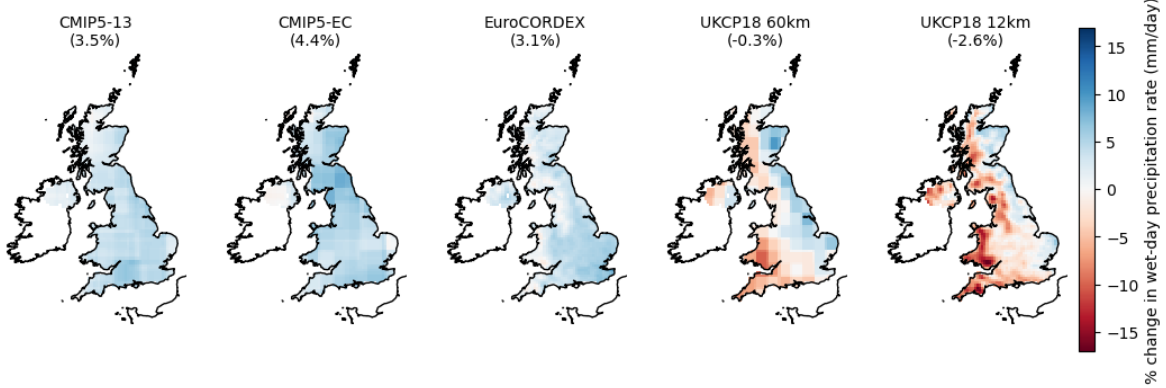
Taylor diagrams comparing the spatial patterns of summer precipitation indices in the reference

Figure 91: Maps of changes in mean summer precipitation rate `pr`, wet-day precipitation rate `sdii` and proportion of wet days `fwd` in each of the ensembles between the reference period (1980–2010) and the future period (2050–2080). The average change over the whole of the UK for each ensemble is shown in parentheses.

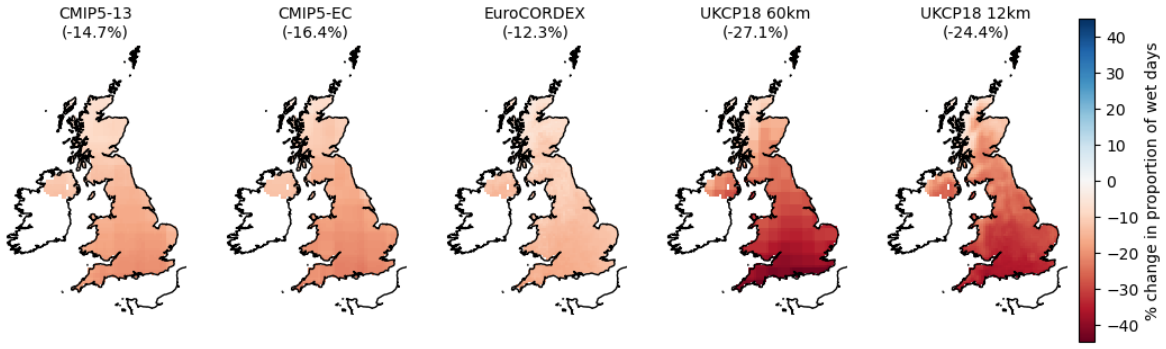
(a) Relative change in daily mean precipitation rates `pr`, 2050–2080



(b) Relative change in daily mean wet-day precipitation rates `sdii`, 2050–2080

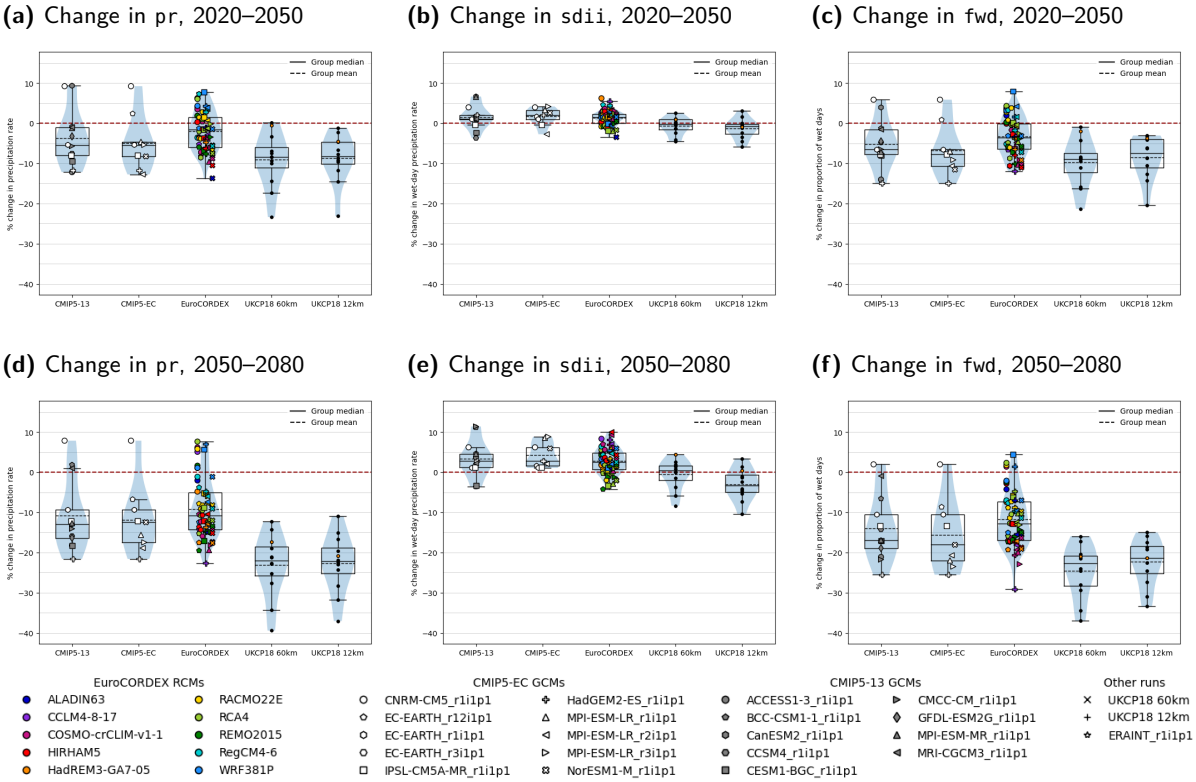


(c) Relative change in proportion of wet days `fwd`, 2050–2080



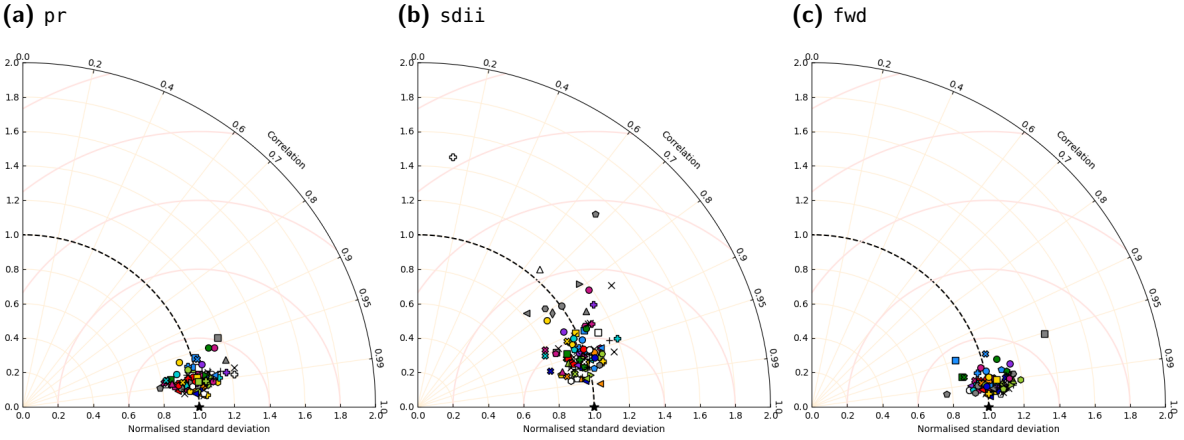
(1980–2010) and near future (2020–2050) periods are shown in Figure 93. In the plots for `pr` and `sdii` (panels a and c), almost all of the runs achieve correlation coefficients above 0.95, with most lying along the line indicating correlation of 0.99: this indicates that the spatial patterns in overall precipitation and in the proportion of wet days are fairly similar in the reference and near future periods. For many runs correlation between the patterns in `sdii` in the future and reference periods is also high; however, HadGEM2-ES again stands out as unusual in having a very low correlation coefficient (panel b), indicating that the spatial patterns in `sdii` in the reference and future periods

Figure 92: Boxplots of projected changes in summer precipitation indices during the period 2020–2050, averaged over the UK and expressed as a percentage change from the UK average of the index during the reference period.



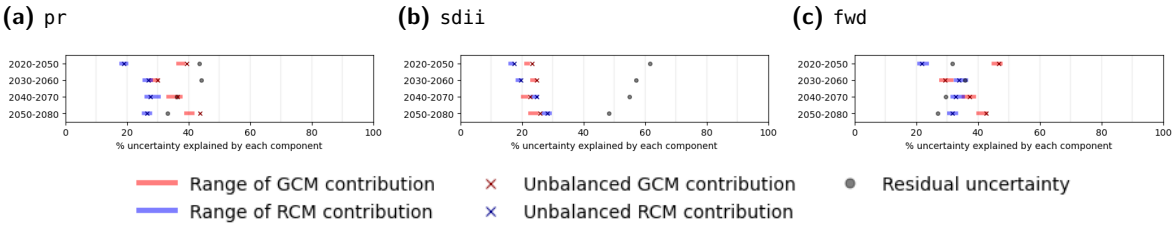
are very different.

Figure 93: Taylor diagrams comparing the spatial patterns of summer precipitation indices in the reference (1980–2010) and near future (2020–2050) periods.



ANOVA of the EuroCORDEX ensemble shows that, while the GCM effects account for slightly more of the systematic variation between the projected changes in pr than the RCM effects, a larger proportion of the variation — 30-45% — is not explained by systematic effects of any other component, with even higher residual uncertainty in sdii (Figure 94).

Figure 94: Plots of the proportion of the variance of changes in temperature within the EuroCORDEX ensemble that are explained by effects due to the driving GCM and the downscaling RCM, or remain unexplained. Coloured bars show the range of variance explained when the maximum possible variation is ascribed to either the GCM or the RCM, and the crosses show the proportions of variance explained by GCM and RCM effects fitted simultaneously using the unbalanced ANOVA described in Section 3.5.2. Residual uncertainty is shown in grey.



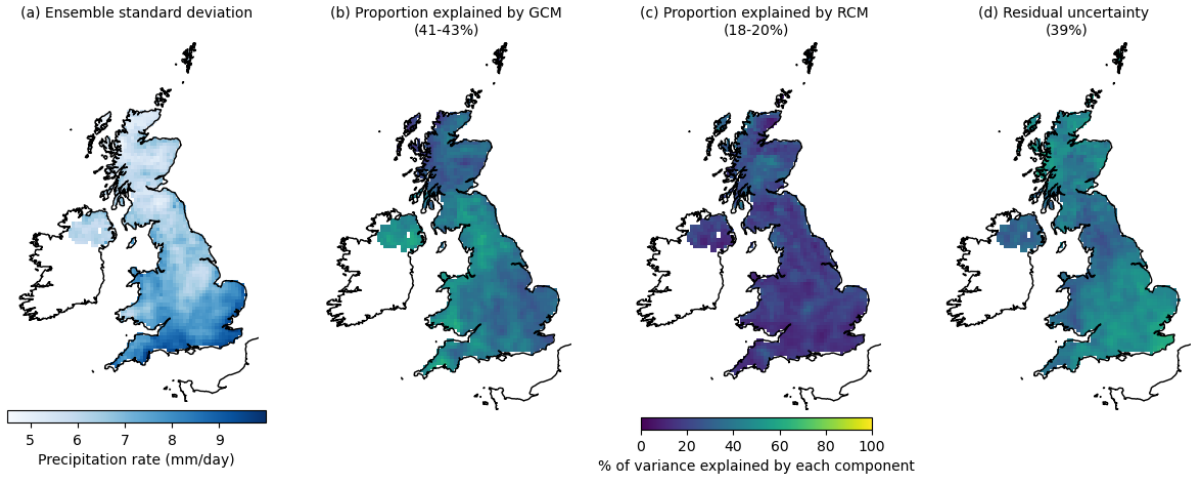
Maps of the standard deviation of the projected relative changes in pr in the EuroCORDEX ensemble, and of the proportion of that variation explained by the fitted GCM and RCM effects as described in Section 3.5.2, are shown in Figure 95. Variation between the projected changes is highest in the south of England in both periods, and decreases steadily towards Scotland. In the near future, there are no strong spatial patterns in the proportions of this variance explained by each component, with 30-50% of the local variation across the UK attributed to the GCMs, and a similar amount to residual uncertainty. By 2050–2080, the GCM contributes up to 70% of the variation in Northern Ireland and across much of England and southern Scotland, with the RCMs contributing up to 50% of the local uncertainty at higher elevations, where orographic effects may be stronger. Residual uncertainty is highest in eastern Scotland and across much of central and southeastern England, where summer precipitation is currently lowest (Figure 38a).

EPP analysis was used to identify the spatial patterns accounting for the majority of variation between the fitted GCM effects; plots of the dominant patterns of spatial variability are shown in Figure 96. In both periods, the dominant mode of variability — accounting for 70% of the between-GCM variance, and so for around a quarter of the total variation between runs in the near term, and for 85% of the between-GCM variance and 35% of the total variation in 2050–2080 — reflect an increase in precipitation across the UK, decreasing in magnitude from south to north. In both periods, CNRM-CM5 receives the highest score in this first EPP, reflecting much wetter relative changes than the ensemble average, as noted in the discussion of Figure 41. The second EPPs reflect a contrast between a percentage increase in precipitation in Scotland and to the west of the Pennines and Cambrians, and a percentage reduction elsewhere: these account for 16% of the variation between GCMs (6% of the total variation between runs) in 2020–2050 and just 8% of the variation between GCMs (3% of the total variation between runs) in 2050–2080.

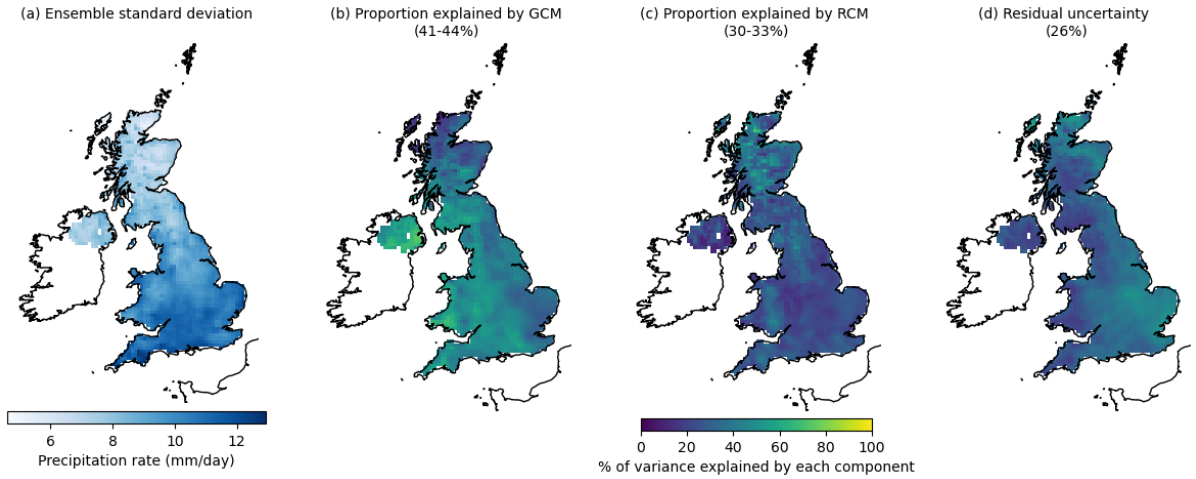
The results of an EPP analysis of the dominant patterns of spatial variability in the UKCP18 ensemble are presented in Figure 97. In both periods, most of the variance (68% in the near future and 80% by 2050–2080) is attributed to the first EPP, with positive scores in this component associated with a percentage increase in precipitation across the UK, with slightly larger relative changes over England and Wales. Positive scores in the second component indicate lower-than-average percentage

Figure 95: Maps of the standard deviation of relative changes in the summer precipitation rate `pr` in the EuroCORDEX ensemble in each grid cell, and of the percentage of the total ensemble variance in each grid cell explained by the choice of GCM and RCM.

(a) Variation in changes in `pr` between reference and 2020–2050



(b) Changes changes in `pr` between reference and 2050–2080



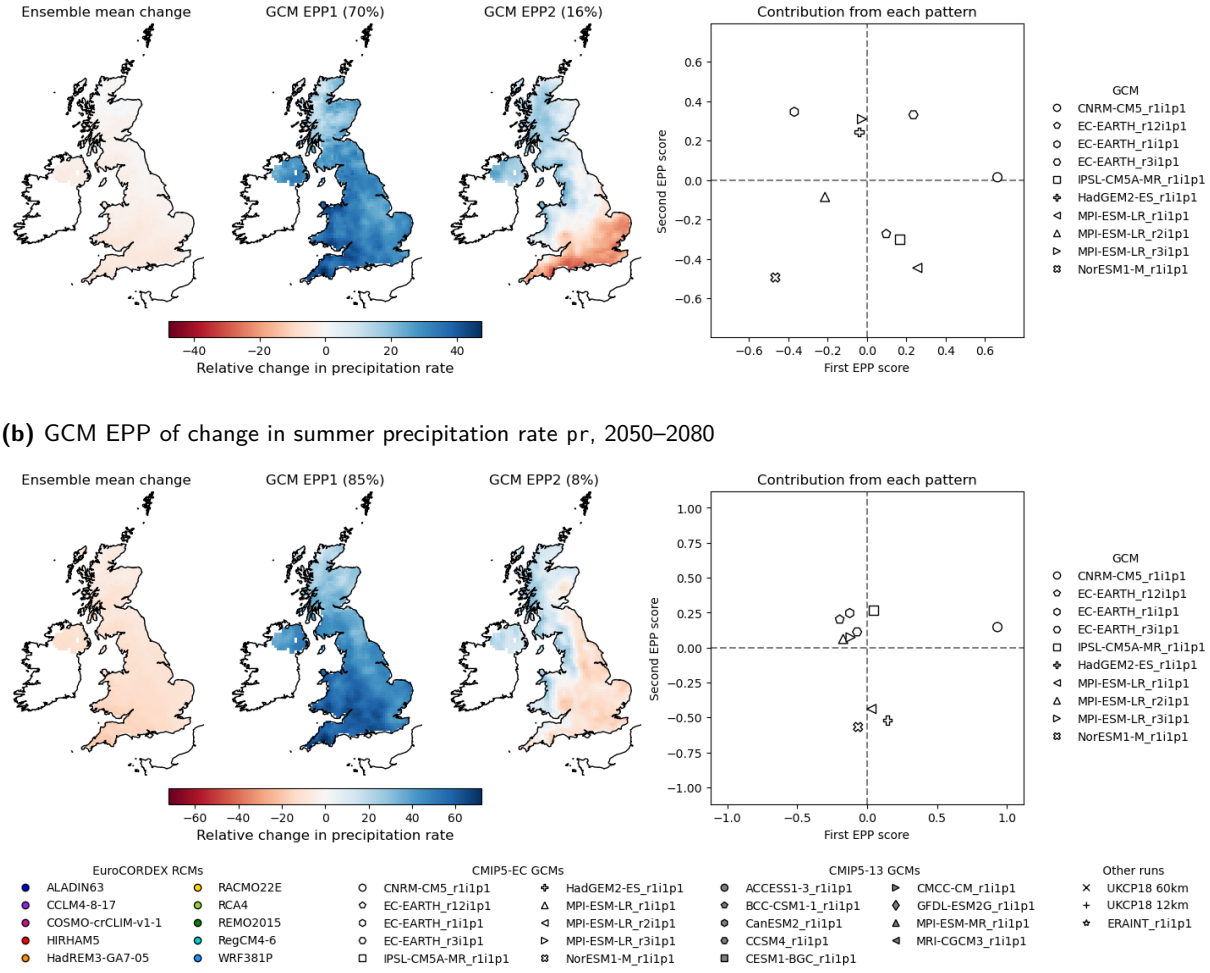
changes across southern England and higher-than-average changes elsewhere; the dominant patterns of relative changes in summer precipitation are very similar to the patterns in Figure 96 explaining difference between the GCMs.

9.3 Projected changes in extreme winter precipitation

Rainfall rates on the wettest days of a season are characterised by the 99th percentile of the distribution of daily precipitation rates, denoted `r99`, with the proportion of precipitation falling on days with precipitation exceeding this threshold denoted by `r99ptot`. As in Section 5.3, changes are expressed as absolute differences from the reference in percentage points, rather than as relative differences with respect to the reference: thus, if `r99ptot` was 5% in the reference period and 6% in the future, this is reported as an increase of 1%, rather than 20%.

Figure 96: Ensemble principal patterns of changes in winter precipitation rates pr between the reference period and future periods, showing dominant patterns of contributions from each GCM to variation about the EuroCORDEX ensemble mean change. The proportion of the between-GCM variation attributed to each pattern is given in parentheses.

(a) GCM EPP of change in summer precipitation rate pr, 2020–2050



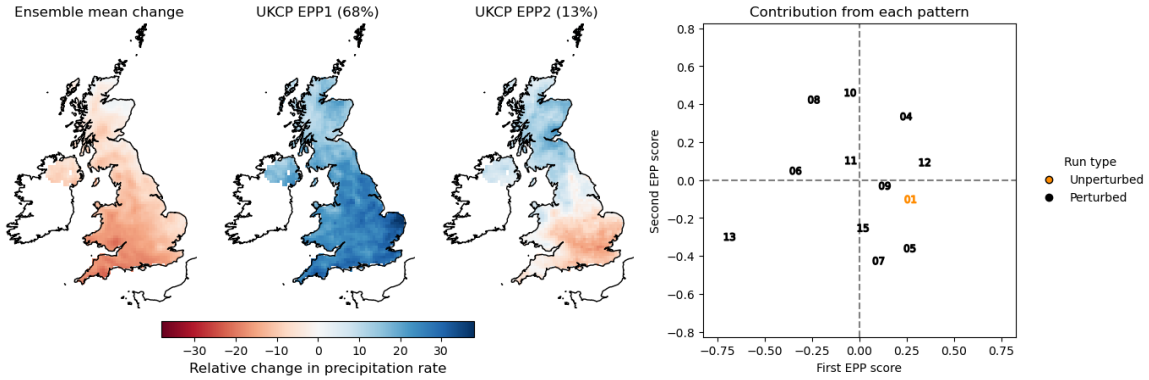
(b) GCM EPP of change in summer precipitation rate pr, 2050–2080

As noted in Section 5.3, r99ptot reflects the proportion of precipitation occurring during extreme precipitation events such as storms, the simulation of which is extremely variable in both space and time: as a result, maps of the changes in r99ptot are extremely noisy, and so maps of the ensemble mean changes are not presented here. Instead, the boxplots in Figure 98 show the distribution of changes in r99ptot during the winter months in five successive 30-year time periods, for the CMIP5, EuroCORDEX and UKCP18 12km ensembles. There is little evidence of any consistent trend in the proportion of rain simulated on the wettest days in winter in the CMIP5 or EuroCORDEX ensembles. In the UKCP18 ensemble, there is no evidence of any strong trend in the first four time periods considered; the increase in r99ptot by 2050–2080 is somewhat larger, although the change is driven largely by just two of the twelve ensemble members.

Boxplots of the changes in r99ptot during the summer months are shown in Figure 99. Here, there is clear evidence of a trend in r99ptot in all of the ensembles, with increasing proportions of summer

Figure 97: EPP analysis of changes in summer precipitation rates pr simulated by the UKCP18 regional models between the reference and future periods, showing the dominant patterns of variation. The proportion of the total variation attributed to each pattern is given in parentheses.

(a) EPPs of change in summer precipitation rate pr, 2020–2050



(b) EPPs of change in summer precipitation rate pr, 2050–2080

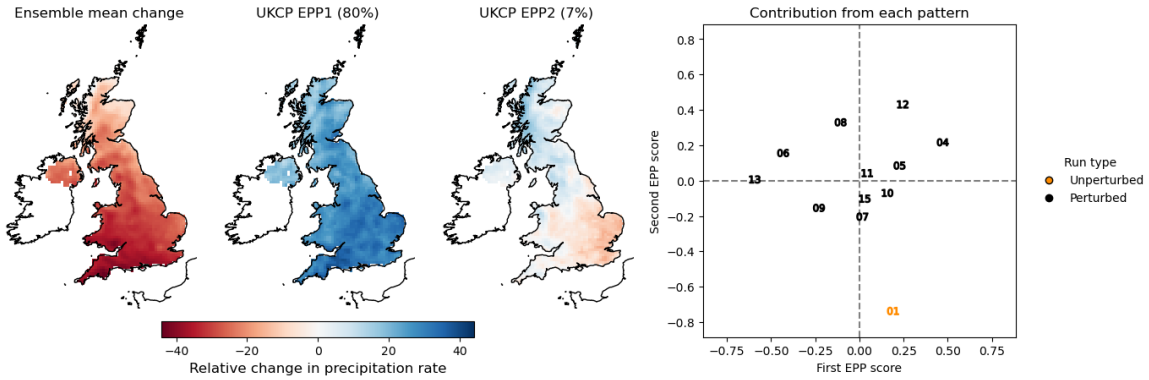
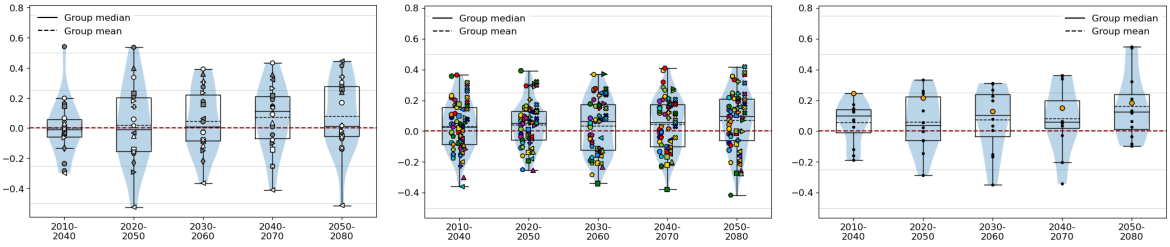


Figure 98: Boxplots of changes in r99ptot over the UK during the winter months in five successive thirty-year time periods

(a) Changes in CMIP5 runs **(b)** Changes in EuroCORDEX runs **(c)** Changes in UKCP18 runs

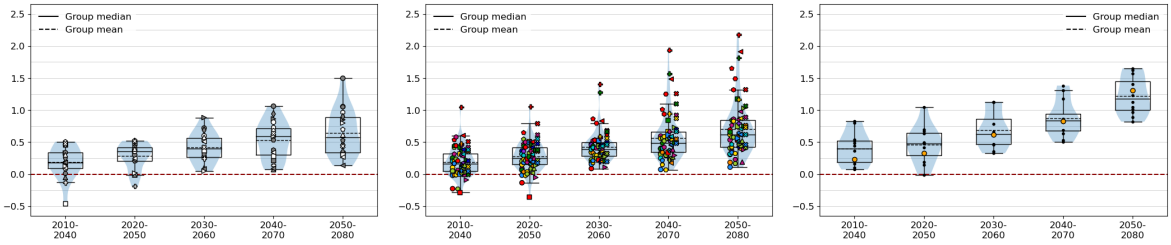


precipitation simulated on the wettest days by almost all of the runs. The spread of the CMIP5 and EuroCORDEX ensembles also increases over time, with some models — for example, EuroCORDEX runs driven by CNRM-CM5, represented by circles — consistently simulating similar proportions of precipitation on the wettest days in all time slices, while in other runs — for example, several of the runs downscaled using HIRHAM5, represented by red symbols — r99ptot increases by more than 1.5% by 2050–2080. While none of the UKCP18 ensemble members simulate such a large individual increase, the average increase in r99ptot is larger than that projected by the EuroCORDEX ensemble,

with the ensemble mean of r99ptot increasing by around 1.25% by 2050–2080.

Figure 99: Boxplots of changes in r99ptot over the UK during the summer months in five successive thirty-year time periods

(a) Changes in CMIP5 runs **(b)** Changes in EuroCORDEX runs **(c)** Changes in UKCP18 runs



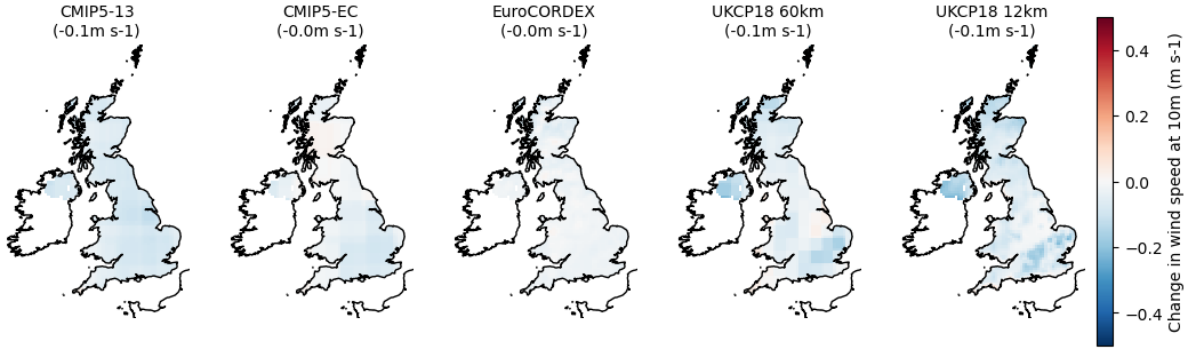
10 Projected changes in surface wind speeds

Maps of the ensemble averages of the projected changes in wind speeds during the winter months are shown in Figure 100, with the corresponding maps of the ensemble averages of the projected changes in wind speeds during the summer months in Figure 101. The CMIP5 and EuroCORDEX ensembles project a reduction in winter wind speeds on average, while the UKCP18 ensembles initially project a reduction, but project an increase by 2050–2080; in summer, all of the ensembles project a very small reduction. However, the projected changes are small: less than 5% in winter, and up to 10% in summer.

Boxplots of the average change over the UK in each of the runs show that the range of changes projected is, likewise, small, although almost all of the runs show a very small reduction in average summer wind speeds by 2050–2080 (Figure 102). Further analysis of the proportions of windy days — defined as days on which the average wind speed is greater than 10.8m/s — simulated by each run was carried out; none of the runs simulate any increasing or decreasing trends in the number of particularly windy days, so changes in surface winds are not considered any further here.

Figure 100: Maps of mean change in wind speeds during the winter months in each ensemble. The average change over the UK land surface for each ensemble is shown in parentheses.

(a) Changes in winter wind speeds — 2020–2050



(b) Changes in winter wind speeds — 2050–2080

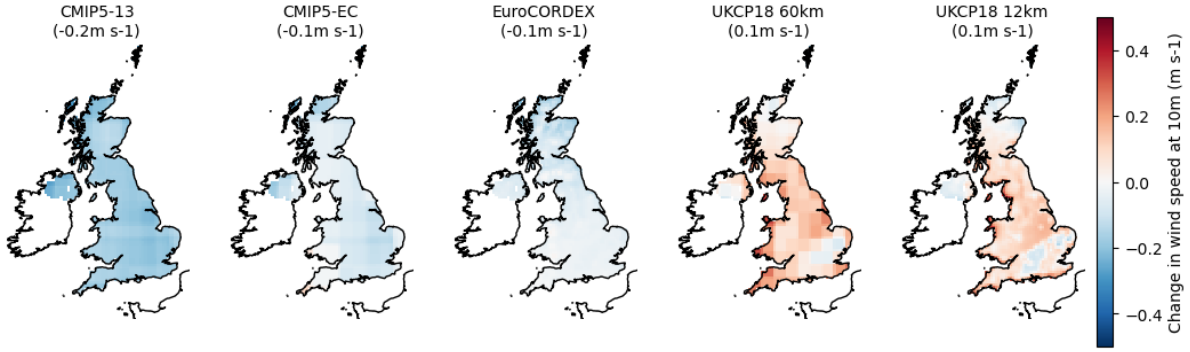
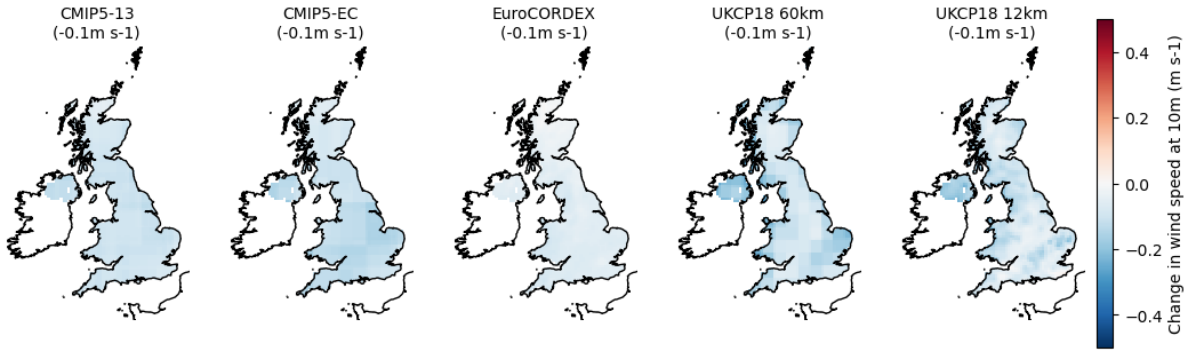


Figure 101: Maps of mean change in wind speeds during the summer months in each ensemble. The average change over the UK land surface for each ensemble is shown in parentheses.

(a) Changes in summer wind speeds — 2020–2050



(b) Changes in summer wind speeds — 2050–2080

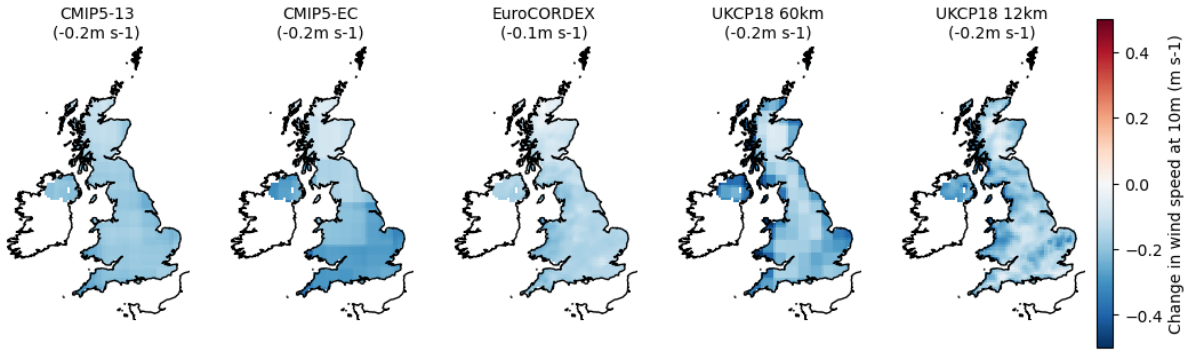
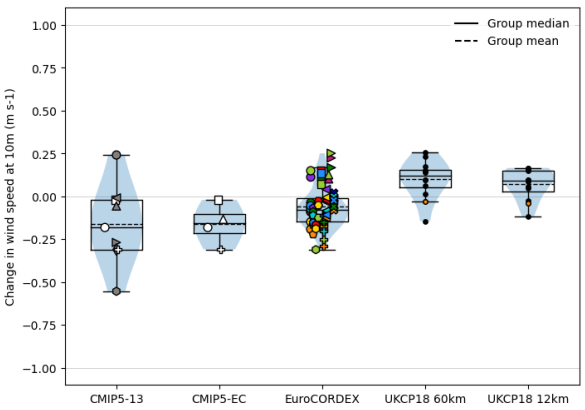
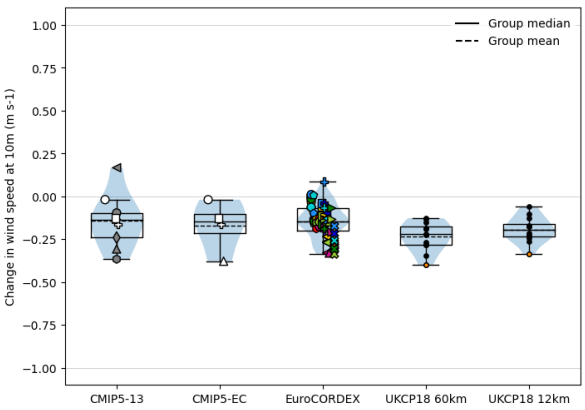


Figure 102: Boxplots of changes in UK-averaged winter wind speed climatologies in individual model runs between the reference period of 1980—2010 and 2050–2080.

(a) Winter



(b) Summer



11 Conclusions and recommendations

The results in this report carry several implications for users of UK climate change projections, whether from the UKCP18 or other ensembles. A key finding is that projections of future change based on UKCP18 are likely to be larger than those based on other ensembles, due mainly to the properties of the underlying GCM which is likely to overestimate the Earth's climate sensitivity according to current IPCC assessments. This means that planning and adaptation decisions based solely on UKCP18 projections may be conservative.

A further finding is that the information from the global ensembles suffers from blurring of local climate features associated with land-sea boundaries and with variations in topography, due to the spatial scale of the underlying outputs. According to the analyses in Sections 4–6, the regional ensembles resolve this without introducing noticeable additional biases. The regional ensembles are therefore recommended in preference to the global ones, for use in the assessment of future climate change impacts and adaptation strategies.

Overall, apart from the spatial blurring in the global model runs, the various ensembles do a reasonable job of representing most of the climate indices considered for the evaluation period. However, the ensemble spread increases when considering future projections for some indices: in the absence of a full uncertainty analysis therefore, a prudent planning and adaptation strategy should consider information from across the range of an ensemble so as to capture the range of possible futures that is consistent with the available modelling scenarios. The most obvious way to achieve this is to use all ensemble members: in many applications however, this will not be feasible due to time and resource limitations.

In situations where only a small number of ensemble members can be used, it is recommended that they are chosen so as to be as representative as possible, with respect to the specific climate indices that are most relevant in the context of the application. To achieve this, users are advised to identify and focus on the dominant sources of variation of these indices within the ensemble. For example, in a situation where the GCMs are responsible for 90% of the variation in the indices of interest, the chosen sample of ensemble members should contain runs driven by contrasting GCMs but the choice of RCM is relatively unimportant.

Our analyses show that the dominant sources of variation in an ensemble are index-dependent. To inform the selection of appropriate ensemble members for use in applications therefore, the online plot explorer at https://github-pages.ucl.ac.uk/UKCORDEX-plot-explorer/ can be used to visualise rapidly the key structure of the regional ensembles for a wide range of indices. For applications examining the impacts of, or adaptation to, future changes in UK climate, a suggested sequence of analyses is:

- Having identified the climate indices, seasons and future period of interest, examine the boxplots showing the corresponding overall UK-averaged distribution(s) of future changes for the time period of interest. This will indicate the extent to which any individual ensemble captures the full range of variation; for the EuroCORDEX ensemble, the boxplots can also give

a preliminary indication as to whether one potential source of variation (e.g. the GCMs) is highly dominant, since this will manifest itself via clustering of the plotted values by colour or plotting symbol as discussed in Section 3.5.

- Examine maps showing the ANOVA decomposition of the ensemble for the indices of interest, such as those in Figure 68. The plot titles quantify the overall partitioning of variation across the UK, and hence formalise the preliminary indications from the boxplots. Users working on applications in a specific part of the UK are advised to use the maps to determine whether the overall partitioning is representative of their area of interest — for example, for some indices the GCMs are dominant overall but the RCMs play a more important role in urban areas.
- For indices where the GCMs contribute an appreciable component of the total ensemble variation, examine maps of the first two GCM EPPs and their scores such as those in Figure 69a. If the first EPP accounts for most of the GCM-attributable variation, then this suggests choosing ensemble members driven by GCMs with scores extending along the range of the first EPP axis; if the second EPP also accounts for a substantial amount of variation, then the selected GCMs should also have scores extending along the range of the second axis. A similar approach can be used to identify an appropriate subset of RCMs, in situations where these contribute appreciably to the ensemble variation.
- For indices where the residual component of variation is dominant in the region of interest, users may either select a random sample of ensemble runs, or examine the runs individually and select a sample that spans the range of values.
- In situations where the guidelines above do not allow users to identify a clear subset of ensemble members for use in their application, Taylor diagrams and bias maps for the selected indices over the evaluation period (e.g. as presented in Sections 4–6) may be used to identify the runs that best reproduce relevant aspects of historical climate by comparison with observations: there is an argument for prioritising these runs for inclusion in any sample. It is worth noting, however, that a GCM:RCM combination that reproduces historical climate well is not necessarily guaranteed to provide the most reliable estimates of future change (the widening spread of the ensembles' future projections for some indices, compared to a relatively narrow range of historical biases, is a symptom of this). Selecting runs solely on the basis of past performance is not necessarily the most appropriate strategy therefore: this is why we suggest first identifying the combinations that span the range of future change in an ensemble.

Acknowledgements

The authors acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Table 2 of this paper) for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global

Organization for Earth System Science Portals. We also acknowledge the World Climate Research Programme's Working Group on Regional Climate, and the Working Group on Coupled Modelling, former coordinating body of CORDEX and responsible panel for CMIP5. We also thank the climate modelling groups (listed in Table 3 of this paper) for producing and making available their model output. We also acknowledge the Earth SystemGrid Federation infrastructure an international effort led by the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison, the European Network for Earth System Modelling and other partners in the Global Organisation for Earth System Science Portals (GO-ESSP).

We are grateful to Bruno Hebling Vieira of the University of São Paulo, for his suggestion to use deviance-based partitioning of variation in a multivariate setting, that formed the basis for the methodology described in Section 3.5.3.

References

Arnell, N., Kay, A., Freeman, A., Rudd, A., and Lowe, J. (2021). Changing climate risk in the UK: a multi-sectoral analysis using policy-relevant indicators. *Climate Risk Management*, 31:100265.

Barnes, C. R., Chandler, R. E., and Brierley, C. M. (2021). Indices to be used in comparison of UKCP18 and EuroCORDEX ensembles. Technical report, SPF Climate Resilience project CR20-3. Available on request from the authors.

Christensen, O. and Kjellström, E. (2020). Partitioning uncertainty components of mean climate and climate change in a large ensemble of European regional climate model projections. *Climate Dynamics*, 54:4293–4308. DOI:10.1007/s00382-020-05229-y.

Cubasch, U., Meehl, G., Boer, G., Stouffer, R., Dix, M., Noda, A., Senior, C., Raper, S., and Yap, K. (2001). Projections of future climate change. In *Climate Change 2001: The scientific basis. Contribution of WG1 to the Third Assessment Report of the IPCC (TAR)*, pages 525–582. Cambridge University Press.

Davison, A. C. (2003). *Statistical Models*. Cambridge University Press, Cambridge.

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., et al. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656):553–597.

Evin, G., Hingray, B., Blanchet, J., Eckert, N., Morin, S., and Verfaillie, D. (2019). Partitioning uncertainty components of an incomplete ensemble of climate projections using data augmentation. *J. Climate*, 32:2423–2440. DOI:10.1175/JCLI-D-18-0606.1.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9(5):1937–1958.

Faraway, J. J. (2014). *Linear Models with R (second edition)*. Chapman and Hall / CRC, Boca Raton.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., et al. (2014). Evaluation of climate models. 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*, pages 741–866. Cambridge University Press.

Forster, P., Storelvmo, T., Armour, K., Collins, W., Dufresne, J.-L., Frame, D., Lunt, D., Mauritsen, T., Palmer, M., Watanabe, M., Wild, M., and Zhang, H. (2021). The Earth's energy budget, climate feedbacks, and climate sensitivity. 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*, page in press. Cambridge University Press.

Fung, F., Stephens, A., and Wilson, A. (2018). UKCP18 Guidance: Data availability, access and formats. *Met Office Hadley Centre: Exeter, UK*.

Gentle, J. E. (2007). *Matrix Algebra: Theory, Computations, and Applications in Statistics*. Springer, New York.

Gregory, J., Ingram, W., Palmer, M., Jones, G., Stott, P., Thorpe, R., Lowe, J., Johns, T., and Williams, K. (2004). A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical research letters*, 31(3).

Gutiérrez, J., Jones, R., Narisma, G., Alves, L., Amjad, M., Gorodetskaya, I., Grose, M., Klutse, N., Krakovska, S., Li, J., et al. (2021). Atlas. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*.

Hansen, J., Johnson, D., Lacis, A., Lebedeff, S., Lee, P., Rind, D., and Russell, G. (1981). Climate impact of increasing atmospheric carbon dioxide. *Science*, 213(4511):957–966.

Herger, N., Sanderson, B. M., and Knutti, R. (2015). Improved pattern scaling approaches for the use in climate impact studies. *Geophysical Research Letters*, 42(9):3486–3494.

Hurrell, J. W. (1996). Influence of variations in extratropical wintertime teleconnections on northern hemisphere temperature. *Geophysical Research Letters*, 23(6):665–668.

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O. B., Bouwer, L. M., Braun, A., Colette, A., Déqué, M., Georgievski, G., et al. (2014). EURO-CORDEX: new high-resolution climate change projections for European impact research. *Regional environmental change*, 14(2):563–578.

Kennedy-Asser, A. T., Andrews, O., Mitchell, D. M., and Warren, R. F. (2021). Evaluating heat extremes in the UK Climate Projections (UKCP18). *Environmental Research Letters*, 16(1):014039.

Khatri, C. G. (1968). Some results for the singular normal multivariate regression models. *Sankhyā (A)*, 30(3):267–280.

Krzanowski, W. (1988). *Principles of Multivariate Analysis*. Oxford University Press.

Lo, Y. E., Mitchell, D. M., Bohnenstengel, S. I., Collins, M., Hawkins, E., Hegerl, G. C., Joshi, M., and Stott, P. A. (2020). UK climate projections: summer daytime and nighttime urban heat island changes in England's major cities. *Journal of Climate*, 33(20):9015–9030.

Masato, G., Hoskins, B. J., and Woollings, T. (2013). Winter and summer northern hemisphere blocking in CMIP5 models. *Journal of Climate*, 26(18):7044–7059.

Murphy, J., Harris, G., Sexton, D., Kendon, E., Bett, P., Clark, R., and Yamazaki, K. (2019). UKCP18 land projections: Science report. *Met Office. Retrieved from https://www. metoffice. gov. uk/pub/data/weather/uk/ukcp18/science-reports/UKCP18-Land-report. pdf*.

Neal, R., Fereday, D., Crocker, R., and Comer, R. E. (2016). A flexible approach to defining weather patterns and their application in weather forecasting over Europe. *Meteorological Applications*, 23(3):389–400.

Northrop, P. J. and Chandler, R. E. (2014). Quantifying sources of uncertainty in projections of future climate. *J. Climate*, 27:8793–8808. DOI:10.1175/JCLI-D-14-00265.1.

Perry, M. and Hollis, D. (2004). The generation of monthly gridded datasets for a range of climatic variables over the United Kingdom. *Met Office, Exeter, UK*.

Perry, M., Hollis, D., and Elms, M. (2009). The generation of daily gridded datasets of temperature and rainfall for the UK. *National Climate Information Centre, Met Office, Exeter*, 7.

Pope, J. O., Brown, K., Fung, F., Hanlon, H. M., Neal, R., Palin, E. J., and Reid, A. (2021). Investigation of future climate change over the British Isles using weather patterns. *Climate Dynamics*, pages 1–15.

Rao, C. R. (1973). *Linear Statistical Inference and its Applications (second edition)*. Wiley, New York.

Reintges, A., Martin, T., Latif, M., and Keenlyside, N. S. (2017). Uncertainty in twenty-first century projections of the Atlantic Meridional Overturning Circulation in CMIP3 and CMIP5 models. *Climate Dynamics*, 49(5):1495–1511.

Sain, S. R., Nychka, D., and Mearns, L. (2011). Functional ANOVA and regional climate experiments: a statistical analysis of dynamic downscaling. *Environmetrics*, 22:700–711. DOI:10.1002/env.1068.

Scaife, A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R., Dunstone, N., Eade, R., Fereday, D., Folland, C., Gordon, M., et al. (2014). Skillful long-range prediction of European and North American winters. *Geophysical Research Letters*, 41(7):2514–2519.

Sherwood, S., Webb, M. J., Annan, J. D., Armour, K., Forster, P. M., Hargreaves, J. C., Hegerl, G., Klein, S. A., Marvel, K. D., Rohling, E. J., et al. (2020). An assessment of Earth's climate sensitivity using multiple lines of evidence. *Reviews of Geophysics*, 58(4):e2019RG000678.

Smeed, D. A., Josey, S., Beaulieu, C., Johns, W. E., Moat, B. I., Frajka-Williams, E., Rayner, D., Meinen, C. S., Baringer, M. O., Bryden, H. L., et al. (2018). The North Atlantic Ocean is in a state of reduced overturning. *Geophysical Research Letters*, 45(3):1527–1533.

Srivastava, M. S. (2007). Multivariate theory for analyzing high-dimensional data. *J. Japan Statist. Soc.*, 37 (1):53–86.

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, 106(D7):7183–7192.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American meteorological Society*, 93(4):485–498.

Tucker, S. O., Kendon, E. J., Bellouin, N., Buonomo, E., Johnson, B., and Murphy, J. M. (2021). Evaluation of a new 12 km regional perturbed parameter ensemble over Europe. *Climate Dynamics*, pages 1–25.

von Storch, H. and Zwiers, F. W. (1999). *Statistical analysis in climate research*. Cambridge University Press, Cambridge.

Wang, C., Zhang, L., Lee, S.-K., Wu, L., and Mechoso, C. R. (2014). A global perspective on CMIP5 climate model biases. *Nature Climate Change*, 4(3):201–205.

Williams, K., Copsey, D., Blockley, E., Bodas-Salcedo, A., Calvert, D., Comer, R., Davis, P., Graham, T., Hewitt, H., Hill, R., et al. (2018). The Met Office global coupled model 3.0 and 3.1 (GC3.0 and GC3.1) configurations. *Journal of Advances in Modeling Earth Systems*, 10(2):357–380.

WMO (2017). WMO guidelines on the calculation of climate normals. World Meteorological Organization Geneva, Switzerland.

Woollings, T., Hannachi, A., and Hoskins, B. (2010). Variability of the North Atlantic eddy-driven jet stream. *Quarterly Journal of the Royal Meteorological Society*, 136(649):856–868.

Yip, S., Ferro, C. A. T., Stephenson, D. B., and Hawkins, E. (2011). A simple, coherent framework for partitioning uncertainty in climate predictions. *Journal of Climate*, 24(17):4634–4643.

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., Klein, S. A., and Taylor, K. E. (2020). Causes of higher climate sensitivity in CMIP6 models. *Geophysical Research Letters*, 47(1):e2019GL085782.

Technical appendices

These appendices provide further details of some of the methodology described in Section 3.5.

A Partitioning of variation in an unbalanced ensemble

For an unbalanced ensemble, define χ_{rg} to be an indicator variable taking the value 1 if the (r, g) coupling appears in the ensemble and zero otherwise. Also, let $n_{\cdot\cdot} = \sum_{r=1}^{R}\sum_{g=1}^{G}\chi_{rg}$ be the total number of runs in the ensemble. It is easy to show that the fitted values $\hat{\boldsymbol{Y}}_{rg}$ from a least-squares (and maximum likelihood) fit of Model 0 are once again all equal to $\overline{\boldsymbol{Y}}_{\cdot\cdot}$, the overall mean across all available runs. Thus the total sum of squares and cross-products is now

$$\boldsymbol{T} = \sum_{r=1}^{R}\sum_{g=1}^{G}\chi_{rg}\left(\boldsymbol{Y}_{rg} - \overline{\boldsymbol{Y}}_{\cdot\cdot}\right)\left(\boldsymbol{Y}_{rg} - \overline{\boldsymbol{Y}}_{\cdot\cdot}\right)' .$$

Similarly, the fitted values for Models 1a and 1b are the means over the relevant groups of runs: $\hat{\boldsymbol{Y}}_{rg} = \overline{\boldsymbol{Y}}_{\cdot g} = n_{\cdot g}^{-1}\sum_{r=1}^{R}\chi_{rg}\boldsymbol{Y}_{rg}$ for Model 1a and $\hat{\boldsymbol{Y}}_{rg} = \overline{\boldsymbol{Y}}_{r\cdot} = n_{r\cdot}^{-1}\sum_{g=1}^{G}\chi_{rg}\boldsymbol{Y}_{rg}$ for Model 1b, where $n_{\cdot g} = \sum_{r=1}^{R}\chi_{rg}$ and $n_{r\cdot} = \sum_{g=1}^{G}\chi_{rg}$ are the total numbers of runs available for GCM g and RCM r respectively. The (r, g)th residual vectors from the two models are thus $\boldsymbol{e}_{rg}^{(1a)} = \boldsymbol{Y}_{rg} - \overline{\boldsymbol{Y}}_{\cdot g}$ and $\boldsymbol{e}_{rg}^{(1b)} = \boldsymbol{Y}_{rg} - \overline{\boldsymbol{Y}}_{r\cdot}$ as before, so that the changes in the corresponding residual cross-product matrices compared with Model 0 are $\boldsymbol{T}_G^{(a)} = \boldsymbol{T} - \sum_{r=1}^{R}\sum_{g=1}^{G}\chi_{rg}\boldsymbol{e}_{rg}^{(1a)}\boldsymbol{e}_{rg}^{(1a)\prime}$ and $\boldsymbol{T}_R^{(b)} = \boldsymbol{T} - \sum_{r=1}^{R}\sum_{g=1}^{G}\chi_{rg}\boldsymbol{e}_{rg}^{(1b)}\boldsymbol{e}_{rg}^{(1b)\prime}$ respectively, as defined in Section 3.5.2.

For Model 2, in the unbalanced case there are no simple closed-form expressions, involving just the group means, for the coefficient estimates $\hat{\boldsymbol{\mu}}$, $\{\hat{\boldsymbol{\alpha}}_g\}$ and $\{\hat{\boldsymbol{\beta}}_r\}$ or the fitted values $\{\hat{\boldsymbol{Y}}_{rg}\}$. Most modern treatments of ANOVA (and, by extension, MANOVA) models handle the situation by exploiting their alternative representation as linear models with dummy covariates used to define group membership (see, for example, Faraway 2014, Section 13.2): this enables the coefficients to be estimated using the usual least-squares matrix computations. Specifically, the model can be written as

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{1} & \boldsymbol{X}_G & \boldsymbol{X}_R \end{pmatrix}\begin{pmatrix} \boldsymbol{\mu}' \\ \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix} + \boldsymbol{\varepsilon} = \boldsymbol{X}\begin{pmatrix} \boldsymbol{\mu}' \\ \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix} + \boldsymbol{\varepsilon} \text{ , say,}$$

where, in addition to the quantity $\boldsymbol{\mu}$ defined already (note that its transpose appears in the expressions above, as the first row of the respective matrices):

- $\boldsymbol{Y}$ is a $n_{\cdot\cdot} \times S$ matrix in which each row contains a single run from the ensemble;
- $\boldsymbol{1}$ denotes a vector containing $n_{\cdot\cdot}$ ones;
- $\boldsymbol{X}_G$ is an $n_{\cdot\cdot} \times (G-1)$ design matrix in which the gth column contains 1s in the rows corresponding to runs in which the gth GCM was used, -1s in the rows where the Gth GCM was used, and zeroes everywhere else;

- $\boldsymbol{X}_R$ is an $n_{\cdot\cdot} \times (R-1)$ design matrix in which the rth column contains 1s in the rows corresponding to runs in which the rth RCM was used, -1s in the rows where the Rth GCM was used, and zeroes everywhere else;
- $\boldsymbol{\alpha}$ is the $(G-1)\times S$ matrix in which the rows are the coefficient vectors $\{\boldsymbol{\alpha}_g : g = 1, \ldots, G-1\}$;
- $\boldsymbol{\beta}$ is the $(R-1)\times S$ matrix in which the rows are the coefficient vectors $\{\boldsymbol{\beta}_r : r = 1, \ldots, R-1\}$;
- $\boldsymbol{\varepsilon}$ is $n_{\cdot\cdot} \times S$ matrix in which the rows are the vectors $\{\boldsymbol{\varepsilon}_{rg}\}$.

The matrix $\boldsymbol{X} = (\mathbf{1}\ \boldsymbol{X}_G\ \boldsymbol{X}_R)$, of dimension $R+G-1$, is obtained by placing its component parts side by side. As an example: for an ensemble containing the outputs from $G = 3$ GCMs and $R = 2$ RCMs, with four runs in total corresponding to the (r, g) pairs $(1, 1)$, $(1, 2)$, $(2, 1)$ and $(2, 3)$, the combined design matrix would be

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & -1 \\ 1 & -1 & -1 & -1 \\ \underbrace{}_{\mathbf{1}} & \underbrace{}_{\boldsymbol{X}_G} & \underbrace{}_{\boldsymbol{X}_R} \end{pmatrix} .$$

In the matrix-based linear model formulation, the least-squares estimator and MLE of the coefficient matrix satisfies the equation (Rao, 1973)

$$(\boldsymbol{X}'\boldsymbol{X}) \begin{pmatrix} \hat{\boldsymbol{\mu}}' \\ \hat{\boldsymbol{\alpha}} \\ \hat{\boldsymbol{\beta}} \end{pmatrix} = \boldsymbol{X}'\boldsymbol{Y} ,$$

each side of which is an $(R+G-1) \times S$ matrix. Although S can be large, the cost of solving this system increases linearly in S so that the solution is perfectly feasible on modern computers, even for large spatial domains.[4]

Having obtained the coefficient estimates, we can compute $\hat{\boldsymbol{\alpha}}_G = -\sum_{g=1}^{G-1} \hat{\boldsymbol{\alpha}}_g$ and $\hat{\boldsymbol{\beta}}_R = -\sum_{r=1}^{R-1} \hat{\boldsymbol{\beta}}_r$, to give a complete set of effect estimates for each GCM and RCM in the ensemble: these estimates can be mapped as in the balanced case. For the partitioning of variation, the residuals can be computed as

$$\boldsymbol{e}^{(2)} = \boldsymbol{Y} - \boldsymbol{X} \begin{pmatrix} \hat{\boldsymbol{\mu}}' \\ \hat{\boldsymbol{\alpha}} \\ \hat{\boldsymbol{\beta}} \end{pmatrix} .$$

The rows of the $n_{\cdot\cdot} \times S$ matrix $\boldsymbol{e}^{(2)}$ are the vectors $\{\boldsymbol{e}^{(2)}_{rg}\}$ needed to calculate $\boldsymbol{T}_E$ in the partitioning of variation.

[4] Note that it is more accurate and more computationally efficient to solve the equations directly than it is to compute the solution as $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$ (Gentle, 2007, Chapter 6).

Having computed the contribution from $\boldsymbol{T}_E$ to the total trace, the corresponding contributions from the matrices $\boldsymbol{T}_R^{(a)}$ and $\boldsymbol{T}_G^{(b)}$ are most easily computed by subtraction (recall that the total trace and the contributions from $\boldsymbol{T}_G^{(a)}$ and $\boldsymbol{T}_R^{(b)}$ have been discussed already).

B Deviance-based partitioning of variation

B.1 Balanced ensembles

Recall that, for any of the models 0, 1a, 1b or 2 considered in Section 3.5, the fitted values are denoted generically by $\{\hat{\boldsymbol{Y}}_{rg} : r = 1, \ldots, R; g = 1, \ldots, G\}$. If the covariance matrix $\boldsymbol{\Sigma}$ is known then, for a balanced ensemble, the maximised multivariate Gaussian log-likelihood for any of these models has the form

$$\begin{aligned} \log L &= \sum_{r=1}^{R}\sum_{g=1}^{G}\left[\frac{S}{2}\text{det}\boldsymbol{\Sigma} - \frac{1}{2}\left(\boldsymbol{Y}_{rg} - \hat{\boldsymbol{Y}}_{rg}\right)'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{Y}_{rg} - \hat{\boldsymbol{Y}}_{rg}\right)\right] \\ &= \frac{SRG}{2}\text{det}\boldsymbol{\Sigma} - \frac{1}{2}\text{trace}\left[\boldsymbol{\Sigma}^{-1}\sum_{r=1}^{R}\sum_{g=1}^{G}\left(\boldsymbol{Y}_{rg} - \hat{\boldsymbol{Y}}_{rg}\right)\left(\boldsymbol{Y}_{rg} - \hat{\boldsymbol{Y}}_{rg}\right)'\right] . \end{aligned} \tag{5}$$

The rearrangement of the second term above follows from the fact that the log-likelihood is scalar-valued and hence equal to its own trace, together with the identity $\text{trace}(\boldsymbol{AB}) = \text{trace}(\boldsymbol{BA})$ when the relevant matrix products are defined (Gentle, 2007, Section 3.2).

To provide some context for any maximised log-likelihood, it can be compared with the highest attainable value i.e. the value that would be obtained from a model that fits the data perfectly such that $\hat{\boldsymbol{Y}}_{rg} = \boldsymbol{Y}_{rg}$. From (5), in a Gaussian setting the log-likelihood for such a model is just $SRG\,\text{det}\boldsymbol{\Sigma}\,/\,2$. Twice the difference between this value and the log-likelihood for any other model is called the *scaled deviance* for that model: in standard linear regression models this quantity is proportional to the residual sum of squares, and it has a similar interpretation in more general models as a 'lack of fit' measure (Davison, 2003, Section 10.2). The scaled deviance corresponding to (5) is

$$D = \text{trace}\left[\boldsymbol{\Sigma}^{-1}\sum_{r=1}^{R}\sum_{g=1}^{G}\left(\boldsymbol{Y}_{rg} - \hat{\boldsymbol{Y}}_{rg}\right)\left(\boldsymbol{Y}_{rg} - \hat{\boldsymbol{Y}}_{rg}\right)'\right] \tag{6}$$

For Model 0, the fitted values $\{\hat{\boldsymbol{Y}}_{rg}\}$ are all equal to the MLE of $\boldsymbol{\mu}$ which is $\overline{\boldsymbol{Y}}_{\cdot\cdot\cdot}$. From (6) therefore, and using (2), the scaled deviance for this model is

$$\text{trace}\left[\boldsymbol{\Sigma}^{-1}\sum_{r=1}^{R}\sum_{g=1}^{G}\left(\boldsymbol{Y}_{rg} - \overline{\boldsymbol{Y}}_{\cdot\cdot}\right)\left(\boldsymbol{Y}_{rg} - \overline{\boldsymbol{Y}}_{\cdot\cdot}\right)'\right] = \text{trace}\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{T}\right] .$$

Similarly, for Model 2 the fitted value $\hat{\boldsymbol{Y}}_{rg}$ is equal to $\overline{\boldsymbol{Y}}_{\cdot\cdot} + \left(\overline{\boldsymbol{Y}}_{\cdot g} - \overline{\boldsymbol{Y}}_{\cdot\cdot}\right) + \left(\overline{\boldsymbol{Y}}_{r\cdot} - \overline{\boldsymbol{Y}}_{\cdot\cdot}\right) = \overline{\boldsymbol{Y}}_{\cdot g} + \overline{\boldsymbol{Y}}_{r\cdot} - \overline{\boldsymbol{Y}}_{\cdot\cdot}$ and, by substitution into (6), the scaled deviance is trace $\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{T}_E\right]$. Under Models 1a and 1b, the fitted values $\hat{\boldsymbol{Y}}_{rg}$ are $\overline{\boldsymbol{Y}}_{\cdot g}$ and $\overline{\boldsymbol{Y}}_{r\cdot}$ respectively. Some algebra shows that the scaled

deviances for these models are trace $\left[\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{T}_E+\boldsymbol{T}_R\right)\right]$ and trace $\left[\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{T}_E+\boldsymbol{T}_G\right)\right]$ respectively.

Now, consider fitting model 0 followed by model 1a and then model 2. The resulting scaled deviances are trace $\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{T}\right]$, trace $\left[\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{T}_E+\boldsymbol{T}_R\right)\right]$ and trace $\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{T}_E\right]$ respectively. The reduction in moving from model 0 to model 1a (representing the amount of scaled deviance attributable to GCM effects) is trace $\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{T}\right]$ − trace $\left[\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{T}_E+\boldsymbol{T}_R\right)\right]$ = trace $\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{T}_G\right]$ from (3); and the further reduction in moving to Model 1a to Model 2 (representing the additional contribution of the RCM effects) is trace $\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{T}_R\right]$. The same partitioning of deviance into GCM and RCM effects is obtained by fitting model 0 followed by model 1b and then model 2. The proportions of deviance attributable to the GCMs and RCMs are therefore trace $\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{T}_G\right]$ /trace $\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{T}\right]$ and trace $\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{T}_R\right]$ /trace $\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{T}\right]$ respectively.

B.1.1 Implementation

The development above assumed that the residual covariance matrix $\boldsymbol{\Sigma}$ is known, which is never the case in practice (this is not a problem when $S=1$, because in case all of the quantities are scalar-valued and the $\boldsymbol{\Sigma}^{-1}$ factors cancel in the numerators and denominators to recover the standard univariate ANOVA partitioning). An obvious solution is to substitute an estimator of $\boldsymbol{\Sigma}$ into the relevant formulae: the usual estimator (Rao, 1973, Section 6c) is

$$\hat{\boldsymbol{\Sigma}}=\frac{\mathbf{T}_E}{(R-1)(G-1)}\ . \tag{7}$$

Unfortunately however, the estimator $\hat{\boldsymbol{\Sigma}}$ has rank at most $(R-1)(G-1)$ (the residual degrees of freedom), so $\hat{\boldsymbol{\Sigma}}^{-1}$ does not exist if the number of spatial locations S exceeds this. This is the case for all situations of interest in the current project. There are two potential solutions to this problem:

- Replace $\hat{\boldsymbol{\Sigma}}$ with an estimator derived from a spatial statistical model. This requires inspection of $\hat{\boldsymbol{\Sigma}}$, and appropriate spatial modelling of its structure. In the present project however, the methodology must be applied to tens of quantities: it is not feasible to carry out this inspection and spatial modelling for all of them.
- Use a reduced-rank approximation to the deviance, for example by replacing $\hat{\boldsymbol{\Sigma}}^{-1}$ with its Moore-Penrose generalised inverse ($\hat{\boldsymbol{\Sigma}}^{\dagger}$, say) wherever it appears. This was shown by Khatri (1968) to be a valid way to construct Gaussian log-likelihoods when the underlying covariance matrix $\boldsymbol{\Sigma}$ is known to be singular. Although this is slightly different from the current situation where the singularity arises instead from the high dimension of the RCM outputs relative to the ensemble size, it nonetheless suggests that the approach has some justification if the implied low-rank approximation of $\boldsymbol{\Sigma}$ is itself reasonable; and the Moore-Penrose substitution also features in more recent approaches to deal with similar high-dimensional Gaussian problems e.g. Srivastava (2007). Given that the purpose in the current exercise is merely to provide an overall measure of the contributions of GCM and RCM variation to the ensemble, this justification is probably adequate.

A further potential difficulty with the deviance-based partitioning is that — superficially at least — it requires the computation and storage of the $S \times S$ matrices $\hat{\boldsymbol{\Sigma}}$, $\boldsymbol{T}$, $\boldsymbol{T}_G$ and $\boldsymbol{T}_R$. If S is very large, the associated computational storage requirements could be problematic. Fortunately, this issue can be sidestepped by using the singular value decompositions (SVDs) of the underlying data matrices. Specifically, let:

- $\boldsymbol{M}$ be the $RG \times S$ matrix with rows containing the individual centred simulations $\{\boldsymbol{Y}_{rg} - \overline{\boldsymbol{Y}}_{\cdot\cdot}\}$, and with rank r say, which is no greater than $\min(RG-1, S)$;
- $\boldsymbol{M}_G$ be the $G \times S$ matrix with rows containing the scaled and centred GCM means $\left\{\sqrt{R}\left(\overline{\boldsymbol{Y}}_{\cdot g} - \overline{\boldsymbol{Y}}_{\cdot\cdot}\right)\right\}$, with rank $r_G \leq \min(G-1, S)$;
- $\boldsymbol{M}_R$ be the $R \times S$ matrix with rows containing the scaled and centred RCM means $\left\{\sqrt{G}\left(\overline{\boldsymbol{Y}}_{r\cdot} - \overline{\boldsymbol{Y}}_{\cdot\cdot}\right)\right\}$, with rank $r_R \leq \min(R-1, S)$;
- $\boldsymbol{M}_E$ be the $RG \times S$ matrix with rows containing the residuals $\{\overline{\boldsymbol{Y}}_{rg} - \overline{\boldsymbol{Y}}_{\cdot g} - \overline{\boldsymbol{Y}}_{r\cdot} + \overline{\boldsymbol{Y}}_{\cdot\cdot}\}$, with rank $r_E \leq \min\left[(R-1)(G-1), S\right]$.

Each of these data matrices has a compact singular value decomposition (SVD): for example, $\boldsymbol{M}$ can be written as $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}'$ where $\boldsymbol{U}$ is a $RG \times r$ matrix with $\boldsymbol{U}'\boldsymbol{U} = \boldsymbol{I}_{r\times r}$ (the $r \times r$ identity matrix); $\boldsymbol{\Lambda}$ is an $r \times r$ diagonal matrix; and $\boldsymbol{V}$ is a $S \times r$ matrix with $\boldsymbol{V}'\boldsymbol{V} = \boldsymbol{I}_{r\times r}$. The elements of the corresponding decompositions for $\boldsymbol{M}_G$, $\boldsymbol{M}_R$ and $\boldsymbol{M}_E$ are denoted with the appropriate subscripts. The matrices $\boldsymbol{T}$, $\boldsymbol{T}_G$, $\boldsymbol{T}_R$ and $\boldsymbol{T}_E$ can now be represented in terms of the SVDs, by noting that each is itself a cross-product of the corresponding data matrix: $\boldsymbol{T} = \boldsymbol{M}'\boldsymbol{M}$, $\boldsymbol{T}_G = \boldsymbol{M}_G'\boldsymbol{M}_G$ and so forth. So, for example,

$$\boldsymbol{T}_E = \boldsymbol{M}_E'\boldsymbol{M}_E = \left[\boldsymbol{U}_E\boldsymbol{\Lambda}_E\boldsymbol{V}_E'\right]'\boldsymbol{U}_E\boldsymbol{\Lambda}_E\boldsymbol{V}_E' = \boldsymbol{V}_E\boldsymbol{\Lambda}_E'\boldsymbol{U}_E'\boldsymbol{U}_E\boldsymbol{\Lambda}_E\boldsymbol{V}_E' = \boldsymbol{V}_E\boldsymbol{\Lambda}_E^2\boldsymbol{V}_E' \ ,$$

the last step following because $\boldsymbol{U}_E'\boldsymbol{U}_E = \boldsymbol{I}_{r_E\times r_E}$ and because $\boldsymbol{\Lambda}_E$ is diagonal. It follows that the generalised inverse of $\hat{\boldsymbol{\Sigma}}$, as defined at (7), is

$$(R-1)(G-1)\left[\boldsymbol{V}_E\boldsymbol{\Lambda}_E^2\boldsymbol{V}_E'\right]^{-1}$$

which, because $\boldsymbol{V}_E'\boldsymbol{V}_E = \boldsymbol{I}_{r_E\times r_E}$ and $\boldsymbol{\Lambda}_E^2$ is diagonal, is equal to $(R-1)(G-1)\boldsymbol{V}_E\boldsymbol{\Lambda}_E^{-2}\boldsymbol{V}_E'$.

In a similar vein, $\boldsymbol{T} = \boldsymbol{V}\boldsymbol{\Lambda}^2\boldsymbol{V}'$. This can be used to calculate the quantity trace $\left(\hat{\boldsymbol{\Sigma}}^{\dagger}\boldsymbol{T}\right)$ required for the deviance-based partitioning of variation: this quantity is

$$(R-1)(G-1)\text{trace}\left[\boldsymbol{V}_E\boldsymbol{\Lambda}_E^{-2}\boldsymbol{V}_E'\boldsymbol{V}\boldsymbol{\Lambda}^2\boldsymbol{V}'\right] = (R-1)(G-1)\text{trace}\left[\boldsymbol{\Lambda}_E^{-2}\boldsymbol{V}_E'\boldsymbol{V}\boldsymbol{\Lambda}^2\boldsymbol{V}'\boldsymbol{V}_E\right] \ .$$

The purpose of the final step above (another application of the identity $\text{trace}(\boldsymbol{A}\boldsymbol{B}) = \text{trace}(\boldsymbol{B}\boldsymbol{A})$) is to reduce the storage requirements from a matrix of dimension $S \times S$ to one of dimension $r_E \times r_E$, before calculating the trace. A similar approach can be used to calculate all of the other quantities required for the deviance-based partitioning of variation.

This SVD approach can be implemented straightforwardly using standard software routines such as

the `svd()` command in R, or the `numpy.linalg.svd` command in python. It can also be used to calculate quantities such as trace($\boldsymbol{T}$) without computing $\boldsymbol{T}$ explicitly: trace($\boldsymbol{T}$) = trace($\boldsymbol{V}\boldsymbol{\Lambda}^2\boldsymbol{V}'$) = trace($\boldsymbol{\Lambda}^2\boldsymbol{V}'\boldsymbol{V}$) = trace($\boldsymbol{\Lambda}^2$) = $\sum_{i=1}^{r}\lambda_i^2$ where the $\{\lambda_i\}$ are the diagonal elements of $\boldsymbol{\Lambda}$. Similar calculations provide trace($\boldsymbol{T}_G^{(a)}$), trace($\boldsymbol{T}_R^{(b)}$) and trace($\boldsymbol{T}_E$) where these are needed.

B.2 The unbalanced case

When calculating the deviance from an unbalanced ensemble, the estimator of the residual covariance matrix $\boldsymbol{\Sigma}$ must be modified to account for the fact that there are fewer than RG runs in the ensemble: instead of (7), $\boldsymbol{\Sigma}$ must now be estimated as

$$\hat{\boldsymbol{\Sigma}} = \frac{\boldsymbol{T}_E}{n_{\cdot\cdot} - R - G + 1} \ ,$$

which reduces to (7) when the ensemble is balanced so that $n_{\cdot\cdot} = RG$.

For an unbalanced ensemble, the matrices $\boldsymbol{T}$ and $\boldsymbol{T}_E$ are calculated directly from the relevant collections of residual vectors as described in Section 3.5.2: for the SVD-based approach to computation, the only difference compared with the balanced case discussed in Appendix B.1.1 is that the matrices $\boldsymbol{M}$ and $\boldsymbol{M}_E$ now contain the respective residuals only for those runs that are actually present in the ensemble (in fact, $\boldsymbol{M}_E$ is exactly the matrix $\boldsymbol{e}^{(2)}$ discussed above). The matrices needed to calculate the SVDs of $\boldsymbol{T}_G^{(a)}$, $\boldsymbol{T}_G^{(b)}$, $\boldsymbol{T}_R^{(a)}$ and $\boldsymbol{T}_R^{(b)}$ are less obvious, however.

The matrix $\boldsymbol{T}_G^{(a)}$ was defined above as $\boldsymbol{T} - \sum_{r=1}^{R}\sum_{g=1}^{G}\chi_{rg}\boldsymbol{e}_{rg}^{(1a)}\boldsymbol{e}_{rg}^{(1a)\prime}$

$$= \sum_{r=1}^{R}\sum_{g=1}^{G}\chi_{rg}\left[\left(\boldsymbol{Y}_{rg} - \overline{\boldsymbol{Y}}_{\cdot\cdot}\right)\left(\boldsymbol{Y}_{rg} - \overline{\boldsymbol{Y}}_{\cdot\cdot}\right)' - \left(\boldsymbol{Y}_{rg} - \overline{\boldsymbol{Y}}_{\cdot g}\right)\left(\boldsymbol{Y}_{rg} - \overline{\boldsymbol{Y}}_{\cdot g}\right)'\right] \ .$$

The first term here can be written as

$$\begin{aligned}
&= \sum_{r=1}^{R}\sum_{g=1}^{G}\chi_{rg}\left(\boldsymbol{Y}_{rg} - \overline{\boldsymbol{Y}}_{\cdot g} + \overline{\boldsymbol{Y}}_{\cdot g} - \overline{\boldsymbol{Y}}_{\cdot\cdot}\right)\left(\boldsymbol{Y}_{rg} - \overline{\boldsymbol{Y}}_{\cdot g} + \overline{\boldsymbol{Y}}_{\cdot g} - \overline{\boldsymbol{Y}}_{\cdot\cdot}\right)' \\
&= \sum_{r=1}^{R}\sum_{g=1}^{G}\chi_{rg}\left[\left(\boldsymbol{Y}_{rg} - \overline{\boldsymbol{Y}}_{\cdot g}\right)\left(\boldsymbol{Y}_{rg} - \overline{\boldsymbol{Y}}_{\cdot g}\right)' + \left(\overline{\boldsymbol{Y}}_{\cdot g} - \overline{\boldsymbol{Y}}_{\cdot\cdot}\right)\left(\overline{\boldsymbol{Y}}_{\cdot g} - \overline{\boldsymbol{Y}}_{\cdot\cdot}\right)'\right. \\
&\qquad\qquad \left. + \left(\boldsymbol{Y}_{rg} - \overline{\boldsymbol{Y}}_{\cdot g}\right)\left(\overline{\boldsymbol{Y}}_{\cdot g} - \overline{\boldsymbol{Y}}_{\cdot\cdot}\right)' + \left(\overline{\boldsymbol{Y}}_{\cdot g} - \overline{\boldsymbol{Y}}_{\cdot\cdot}\right)\left(\boldsymbol{Y}_{rg} - \overline{\boldsymbol{Y}}_{\cdot g}\right)'\right] \ ,
\end{aligned}$$

so that

$$\begin{aligned}\boldsymbol{T}_G^{(a)} &= \sum_{r=1}^{R}\sum_{g=1}^{G}\chi_{rg}\Big[\left(\overline{\boldsymbol{Y}}_{\cdot g}-\overline{\boldsymbol{Y}}_{\cdot\cdot}\right)\left(\overline{\boldsymbol{Y}}_{\cdot g}-\overline{\boldsymbol{Y}}_{\cdot\cdot}\right)' \\ &\qquad + \left(\boldsymbol{Y}_{rg}-\overline{\boldsymbol{Y}}_{\cdot g}\right)\left(\overline{\boldsymbol{Y}}_{\cdot g}-\overline{\boldsymbol{Y}}_{\cdot\cdot}\right)' + \left(\overline{\boldsymbol{Y}}_{\cdot g}-\overline{\boldsymbol{Y}}_{\cdot\cdot}\right)\left(\boldsymbol{Y}_{rg}-\overline{\boldsymbol{Y}}_{\cdot g}\right)'\Big] \\ &= \sum_{g=1}^{G} n_{\cdot g}\left(\overline{\boldsymbol{Y}}_{\cdot g}-\overline{\boldsymbol{Y}}_{\cdot\cdot}\right)\left(\overline{\boldsymbol{Y}}_{\cdot g}-\overline{\boldsymbol{Y}}_{\cdot\cdot}\right)' + \left[\sum_{g=1}^{G}\left(\overline{\boldsymbol{Y}}_{\cdot g}-\overline{\boldsymbol{Y}}_{\cdot\cdot}\right)\sum_{r=1}^{R}\chi_{rg}\left(\boldsymbol{Y}_{rg}-\overline{\boldsymbol{Y}}_{\cdot g}\right)'\right]' \\ &\qquad + \sum_{g=1}^{G}\left(\overline{\boldsymbol{Y}}_{\cdot g}-\overline{\boldsymbol{Y}}_{\cdot\cdot}\right)\sum_{r=1}^{R}\chi_{rg}\left(\boldsymbol{Y}_{rg}-\overline{\boldsymbol{Y}}_{\cdot g}\right)' .\end{aligned}$$

But $\sum_{r=1}^{R}\chi_{rg}\left(\boldsymbol{Y}_{rg}-\overline{\boldsymbol{Y}}_{\cdot g}\right)' = \sum_{r=1}^{R}\chi_{rg}\boldsymbol{Y}'_{rg} - n_{\cdot g}\overline{\boldsymbol{Y}}'_{\cdot g} = n_{\cdot g}\overline{\boldsymbol{Y}}'_{\cdot g} - n_{\cdot g}\overline{\boldsymbol{Y}}'_{\cdot g} = \boldsymbol{0}$. It follows that

$$\boldsymbol{T}_G^{(a)} = \sum_{g=1}^{G} n_{\cdot g}\left(\overline{\boldsymbol{Y}}_{\cdot g}-\overline{\boldsymbol{Y}}_{\cdot\cdot}\right)\left(\overline{\boldsymbol{Y}}_{\cdot g}-\overline{\boldsymbol{Y}}_{\cdot\cdot}\right)' ,$$

so that the required matrix $\boldsymbol{M}_G^{(a)}$ is once again a $G \times S$ matrix but where the scaling reflects the available number of runs of each GCM: its gth row is $\left\{\sqrt{n_{\cdot g}}\left(\overline{\boldsymbol{Y}}_{\cdot g}-\overline{\boldsymbol{Y}}_{\cdot\cdot}\right)\right\}$.

The treatment of $\boldsymbol{T}_R^{(b)}$ follows exactly the same steps: $\boldsymbol{M}_R^{(b)}$ is an $R \times S$ matrix in which row r contains the scaled and centred RCM means $\left\{\sqrt{n_{r\cdot}}\left(\overline{\boldsymbol{Y}}_{r\cdot}-\overline{\boldsymbol{Y}}_{\cdot\cdot}\right)\right\}$.

The calculations above enable the efficient calculation of trace($\boldsymbol{T}$), trace($\hat{\boldsymbol{\Sigma}}$), trace($\boldsymbol{T}_G^{(a)}$) and trace($\boldsymbol{T}_R^{(b)}$). For the partitioning of variation, trace $\left(\boldsymbol{T}_G^{(b)}\right)$ and trace $\left(\boldsymbol{T}_R^{(a)}\right)$ are also needed: these are calculated most efficiently by subtraction. Similar comments apply to the traces of the matrix products involving $\hat{\boldsymbol{\Sigma}}^{\dagger}$, that are needed for the deviance-based partitioning.