Department
for Environment
Food & Rural Affairs

Department for
Business, Energy
& Industrial Strategy

Met Office
Hadley Centre

Environment
Agency

UKCP18 Land Projections: Science Report

November 2018 (Updated March 2019)

Update to Probabilistic Projections

Global Projections

Regional Projections

UKCP18

Derived Projections

Observations

Marine Projections
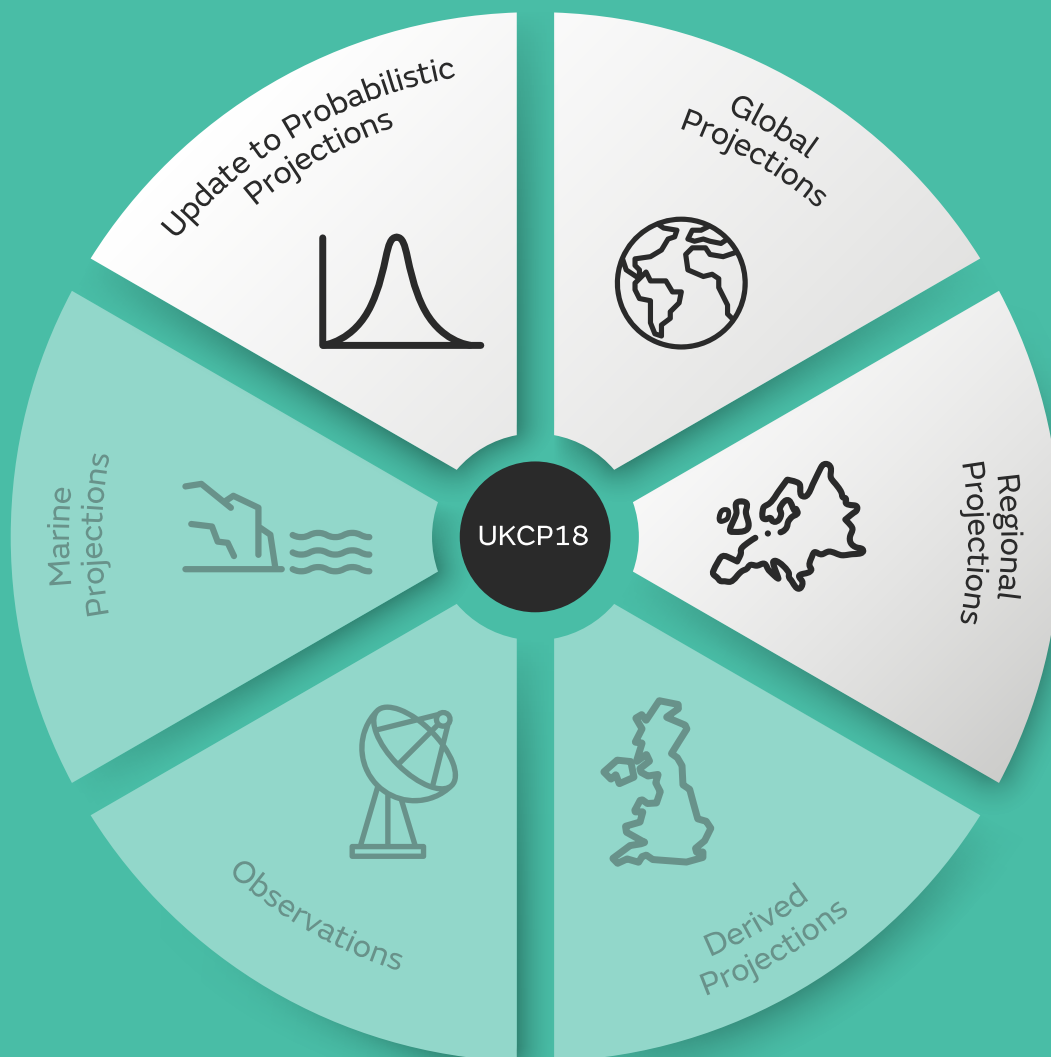
J.M. Murphy, G.R. Harris, D.M.H. Sexton, E.J. Kendon, P.E. Bett, R.T. Clark, K.E. Eagle, G. Fosser, F. Fung, J.A. Lowe, R.E. McDonald, R.N. McInnes, C.F. McSweeney, J.F.B. Mitchell, J.W. Rostron, H.E. Thornton, S. Tucker and K. Yamazaki

Version 2.0

Contents

1. Introduction

The UKCP18 land projections have been designed to replace the previous UKCP09 scenarios (Jenkins et al., 2009), taking into account subsequent feedback from users and developments in modelling capability. The new projections consist of three Strands, described in detail in sections 2-4. The Strands are built to serve types of user requirement that are related, but also somewhat distinct, and each Strand uses different climate modelling inputs. For some applications it may be appropriate to consider combining information from two or more Strands to obtain the most comprehensive advice (section 5.2). The Strands include some inter-dependencies in their designs, to help with this. Below, we discuss how developments since UKCP09 have informed the content of UKCP18, and introduce the three Strands. We describe and evaluate their underpinning projection systems in more detail in sections 2-4, and compare their future projections in section 5. This includes discussion of how the results should be interpreted and used, taking into consideration the nature of the products and the degree of consistency between their results. A summary is provided in section 6.

1.1. Recap of UKCP09

The centrepiece of the UKCP09 land component was a set of probabilistic projections (Murphy et al., 2009). These expressed a broad range of plausible outcomes for UK climate during the 21st century, for a set of key climate variables. They were provided for three future emissions pathways (the B1, A1B and A1FI scenarios of Nakicenovic and Swart (2000), labelled "low", "medium" and "high"), for a national 25km grid and two sets of aggregated regions. The probabilistic projections were presented as 30-year monthly, seasonal and annual average changes, expressed relative to a 1961-1990 baseline. They represented known uncertainties (due to internal climate variability and the modelling of key Earth system processes), consistent with available climate model simulations and the knowledge contained within them. The projections were constructed from several ensembles of variants of a single climate model (HadCM3), representing modelling uncertainties by perturbing model parameters within expert-specified ranges. These perturbed parameter ensembles (PPEs) were combined with results from the Coupled Model Intercomparison Project Phase 3 (CMIP3) ensemble of international global climate models (GCMs), and a set of observational metrics of historical model performance (Sexton et al., 2012; Harris et al., 2013). A weather generator (Jones et al., 2010) was also provided. This was driven by change factors sampled from the projections in order to assess localised impacts and risks in many applications. The UKCP09 projections were the first UK scenarios to include intrinsic, quantitative estimates of uncertainty. They formed a major component of the evidence bases for the first and second national climate change risk assessments (CCRA), published in 2012 (https://www.gov.uk/government/publications/uk-climate-change-risk-assessment-government-report) and 2017 (CCRA2, see https://www.theccc.org.uk/tackling-climate-change/preparing-for-climate-change/uk-climate-change-risk-assessment-2017/).

1.2. User and science drivers for UKCP18

In addition to the CCRAs, UKCP09 has been used in many specific impacts and adaptation studies. User feedback has accumulated steadily on strengths, limitations and gaps, to be addressed in subsequent scenarios. This insight has been further updated and refined during the development of UKCP18, through engagement with Government and non-Government users through two active groups. Key points include:

1. The assessments of uncertainty afforded by the probabilistic projections are important, and should be retained.

2. However, some users require a simpler and more flexible dataset, and prefer to work directly with climate model output. This is because the latter provides access to future projections with full spatial and temporal coherence, and a wider range of variables and time resolutions. Some impacts studies therefore used an eleven-member PPE of regional climate model (RCM) simulations (Murphy et al., 2009). For example, these included applications requiring assessment of multiple drivers of changing hazards in distributed networks, such as electricity (McColl et al., 2012) and rail (Palin et al., 2013).

3. Related to (2), UKCP18 should provide products capable of supporting development of a limited set of "storylines". These could take a number of potential forms, including high resolution narratives of how impacts related to particular types of weather event might change in future (Hazeleger et al., 2015), or characterisations of simulated regional changes that promote understanding in terms of specific remote circulation drivers (Zappa and Shepherd, 2017).

4. UKCP18 should retain and update projections that cover the whole 21st century, but provide more information on climate variability on annual to decadal time scales, and more information on extreme events.

5. The new projections should remain aligned with international activities such as the Intergovernmental Panel on Climate Change (IPCC), and account for the latest developments in climate modelling capability and scientific understanding.

6. Information on 21st century climate in other worldwide regions should be available, to support analysis of imported risks to the UK.

The major scientific driver for new projections arises from development of new generations of climate models. Since UKCP09, the CMIP5 ensemble of international climate models (Taylor et al., 2012) has replaced CMIP3 as an international focus for assessments of projected changes and impacts, notably in the IPCC Fifth Assessment Report (AR5, Collins et al., 2013). The recent CCRA2 assessment drew heavily on projections from both UKCP09 and CMIP5, along with observational results and other evidence from climate science literature (Humphrey and Murphy, 2016).

Since CMIP5, the UK and other countries have continued to develop their climate models, in preparation for the next phase (CMIP6, Eyring et al., 2016). Recent developments have improved the representation of some dynamical influences on regional climate variability and change in the European/North Atlantic sector. For example, Scaife et al. (2012) and Karpechko and Manzini (2012) showed that improving dynamical coupling between the stratosphere and troposphere (not well represented in UKCP09, or in most CMIP5 models) may influence projected changes in the winter storm track and precipitation in the North Atlantic/ European sector. Recent versions of the Met Office seasonal prediction system, based on the latest UK climate model HadGEM3, have shown significant improvements in prediction of the winter North Atlantic Oscillation (NAO). This is accompanied by skill in modelling the teleconnection patterns of potential drivers, such as the El Nino Southern Oscillation (ENSO), the quasi-biennial oscillation (QBO) of equatorial stratospheric winds, and sea-ice anomalies in the Kara Sea (Scaife et al., 2014).These developments offer potential to provide new high-resolution projections, capable of supporting advice on future extremes and impacts underpinned by better simulation of driving mechanisms. In UKCP18, we use a new coupled ocean-atmosphere model, HadGEM3-GC3.05, as the basis for developing ensembles of such projections. HadGEM3-GC3.05 (hereafter GC3.05) is a high-resolution physical climate system model (horizontal grid spacing ~60km at mid-latitudes) which also includes a sophisticated explicit representation of atmospheric aerosols. The horizontal resolution of GC3.05 is comparable to that used in regional downscaling for the UKCIP02 scenarios (Hulme and Jenkins, 2002), and provides much better representation of synoptic-scale weather systems, mountains and coastlines, compared with the ~300km-scale global simulations that formed the core of UKCP09. The GC3.05 model is described in more detail in section 3.2. It is closely related (see Appendix D) to the HadGEM3-GC3.1 model (hereafter GC3.1) that forms the basis of the UK submission to CMIP6. GC3.1 contains many changes to its physical components relative to the HadGEM2-AO model submitted to CMIP5. These include an improved dynamical core, new ocean and sea-ice models and a new parameterisation of cloud (Williams et al., 2018).

1.3. Three strands of UKCP18 land projections

In response to these motivating factors, a three-strand strategy has been implemented for UKCP18. Strand 1 provides updates to the probabilistic projections of UKCP09. The probabilistic projections express a broad range of potential outcomes derived from around 350 climate model simulations, and constrained by a set of observational measures of model performance. The results are conditional upon, and limited by, the knowledge contained within the model simulations, and are also conditional on various expert choices required in the statistical methodology used to combine the input information.

Provision of Strand 1 addresses a continuing requirement for a product focused on exploring uncertainties (user driver 1 above). It is intended to provide support for future risk assessments, in particular by providing context for applications reliant on projection systems containing more limited sampling of uncertainties (such as Strands 2 and 3, described below). In order to achieve this, Strand 1 requires ensembles of earth system model simulations that include an interactive carbon cycle component, so that uncertainties in carbon cycle feedbacks can be considered alongside those in physical climate processes and aerosol chemistry. In this way, Strand 1 can provide a more complete view of the range of future outcomes consistent with a given scenario of future emissions. On the other hand, Strand 1 does not account for uncertainties associated with earth system processes not included in the available model projections, such as potential carbon release from melting of permafrost or destabilisation of ocean methane hydrates (Collins et al., 2013).

Although a new UK community earth system model (UKESM1, see http://www.jwcrp.org.uk/research-activity/ukesm.asp) has recently been developed from GC3.1, it was not feasible to consider building a new PPE of UKESM1 variants in time for UKCP18. Therefore, PPEs derived from HadCM3 continue to form the kernel of the modelling strategy for Strand 1. These are combined with results from CMIP5 earth system models, which replace the CMIP3 climate models as the source of information on the structural component of modelling uncertainty (see section 2.1).

In the light of the user drivers of section 1.2, the main limitations of Strand 1 are that the large sample of potential outcomes needed to support probabilistic projections is not suitable for derivation of a small set of storylines, that results are only provided for a limited set of climate variables, and that the projections lack the full spatial and temporal coherence of raw climate model output. The latter feature is due to the need to apply various statistical techniques within the methodology (see section 2.2). These issues are addressed through provision of Strands 2 and 3, which are provided in response to user drivers 2 and 3 for a limited set of plausible projections capable of supporting a wide range of impacts studies and development of narratives. In order to help fulfil these drivers, new PPEs have been developed from the GC3.05 model (see section 1.2). Strand 2 consists of 15 PPE simulations run from 1900-2100 (hereafter GC3.05-PPE), plus 13 simulations from CMIP5 coupled ocean-atmosphere models (hereafter CMIP5-13). These were selected using performance and diversity criteria, in order to increase the range of global and regional changes sampled in the projections. Being global projections, the Strand 2 results are particularly suited to development of physical narratives in which future UK changes are explained in terms of large-scale drivers, such as remote changes in circulation (Zappa and Shepherd, 2017). They also provide a resource for analysis of international risks (user driver 6).

Strand 3 consists of a 12-member PPE of projections at 12km horizontal resolution. These are obtained by configuring corresponding members of GC3.05-PPE as regional climate model variants (hereafter RCM-PPE) using a European domain. The RCP-PPE members are then driven from 1980-2080 by output from the global simulations. Results from Strand 3 provide a source of impacts and storyline information for applications requiring analysis at local to regional scales, accounting in greater detail for the effects of mountains, coastlines, lakes and mesoscale atmospheric circulations. Analysis of distributed impacts at regional to national scales can potentially be carried out using either Strand 2 or Strand 3. Here, a key choice concerns the relative benefits of improved spatial detail from Strand 3, against the broader range of projections available from Strand 2. For studies with a European focus, or which seek to set UK changes in a European context, either Strand 2 or Strand 3 can be used.

In this report, we use the term 'ensemble' when describing GC3.05-PPE, CMIP5-13 and RCM-PPE, following traditional practice. However, it does not imply that the relevant subset of model projections forms a uniform sample of the full space[1] of possible model configurations. When describing the combined collection of simulations that forms Strand 2, we use the term 'set'.

---

[1] In the case of a perturbed parameter ensemble, such a space would consist of the full range of parameter space. For a multi-model ensemble it might consist of a broader space of climate models that could, in theory, be created by considering all possible combinations of alternative model sub-components.

In previous long-term climate simulations, and the new UKCP18 simulations outlined above, the dynamics of atmospheric convection is parameterised. However, Kendon et al (2014) reported the first UK climate change simulation in which the dynamics of larger convective storms is explicitly represented, using a 1.5 km resolution regional model. They found substantial increases in the intensity of hourly summer rainfall extremes that were not captured in a coarser 12 km simulation in which convection was parameterised. Ban et al. (2015) found a similar result for the Alpine region. In winter, sub-daily rainfall extremes increased substantially in both the 1.5 km and 12 km simulations of Kendon et al. (2014). Kendon et al (2017) cautioned that further kilometre-scale simulations are needed to assess the robustness of their results, and that moving to this resolution does not address all sources of bias in modelled convection. Nevertheless, this development provides a first opportunity to include convective storm dynamics in projections contributing to a set of national climate scenarios. Accordingly, an ensemble of projections using a 2.2km convective-permitting regional model is being produced.

These simulations are driven by the 12km RCM-PPE simulations, for a domain covering the whole UK. They will be documented in a separate report, forming a second phase of Strand 3. This will provide advice on the projected changes and their degree of credibility, relating in particular to potential new capabilities to provide information at higher spatial and temporal resolution than before, including changes in heavy sub-daily rainfall events.

1.4. Emissions scenarios

In Strand 1, projections are provided for each of the RCP scenarios (Moss et al., 2010) assessed in AR5 (2.6, 4.5, 6.0 and 8.5), and also the SRES A1B scenario (Nakicenovic and Swart, 2000). The latter allows a direct comparison with UKCP09. As noted above, these represent "emissions-driven" projections that describe uncertainties in the conversion of carbon emissions to CO_2 concentrations in the atmosphere, as well as uncertainty in the physical climate system response to changes in greenhouse gas concentrations. For each emissions scenario, the probabilistic projections from Strand 1 therefore account for a range of future CO_2 concentration pathways.

The Strand 2 and 3 simulations were run only for RCP8.5, due to restrictions in high performance computing (HPC) capacity. As GC3.05 does not include a carbon cycle component, the GC3.05-PPE members were run with prescribed CO_2 concentrations. However, different concentration pathways were used for each member, chosen to replicate approximately the range of outcomes projected in Strand 1. In this way, the GC3.05-PPE simulations reflect the global effects of carbon cycle uncertainties on projected changes, though carbon cycle effects on *regional patterns* of change are omitted, in contrast to Strand 1. The CMIP5 simulations added to Strand 2 are also based on physical climate system models lacking an interactive carbon cycle. In this sense they are consistent with GC3.05-PPE, however the relevant RCP8.5 experiment was a concentration-driven experiment in which all CMIP5 models used the same CO_2 pathway. The spread of outcomes in the combined set of Strand 2 projections therefore contains contributions from internal variability, uncertainties in modelling of physical processes and aerosol chemistry (from both the GC3.05-PPE and CMIP5 simulations), and from uncertainty in the globally-averaged effects of carbon cycle feedbacks (from GC3.05-PPE only).

In Strand 3, each RCM-PPE member inherits the CO_2 pathway prescribed in its driving Strand 2 global simulation. This is also the case for changes in the forcing due to anthropogenic aerosols, which are prescribed from the driving model using a technique described in section 4.2.

Derived results are also produced for the RCP2.6 scenario, and for analysis of impacts under specific global warming targets. These are provided by applying time shifting and pattern scaling methods to the RCP8.5 results from Strand 2 (Gohar et al., 2018).

1.5. Choice of baseline

In this report, the period 1981-2000 is used as a standard baseline for calculation of projected future anomalies. This is partly because the UKCP09 baseline of 1961-1990 is now out of date as a representation of recent climate, and also because a start date of 1980 was used for the new 12km RCM simulations, precluding use of an earlier baseline in Strand 3. The standard baseline of the World Meteorological Organisation is currently 1981-2010. This was not adopted for UKCP18 because the RCP emissions scenarios used in the projections (section 1.4) start from 2006, hence the simulations include an element of predictive information during 2001-2010. However, users of data from Strands 1 and 2 will have the option to choose 1961-1990 or 1981-2010 as an alternative to the UKCP18 standard baseline. Users of the 12km RCM runs will have the alternative option of 1981-2010. In section 2.3 we compare the Strand 1 projections against UKCP09 results for the A1B scenario, including an illustration of the impact of reverting to the 1961-1990 baseline.

1.6. Methodologies and evaluation

In sections 2-4 we describe the construction of the three Strands and evaluate the projection systems. We also provide examples of projected changes, in order to illustrate selected properties and applications specific to each Strand. Comparison of projections across all Strands follows in section 5.

For Strands 2 and 3 the major focus is the performance of the relevant sets of climate model simulations. For Strand 1, the evaluation involves comparing the probabilistic projections with changes given by the model simulations from which they are constructed. Appendices are used to provide some supporting information on data and methods, and comprehensive technical descriptions will appear later, in the specialist journal papers cited.

2. Strand 1: Probabilistic projections

2.1. Overview

The purpose of the probabilistic projections is to provide, for a given emissions scenario, information on known uncertainties in future climate changes. In particular, the aim is to represent uncertainties consistent with the knowledge incorporated in existing ensembles of climate model projections, plus the effects of internal climate variability. The probability distributions provide information on ranges of outcomes, and the relative likelihood of alternative outcomes within these ranges, conditioned on the evidence used to produce them. The results are Bayesian probabilities that represent the relative strength of evidence behind different potential future changes, and depend on subjective judgements required to develop the methodology used to produce them. As such, they should not be confused with (for example) probabilistic weather forecasts, which can be assessed in a frequentist sense through repeated trial and verification (e.g. Hamill, 1997).

By combining results from PPE and multi-model projections, the UKCP18 methodology (like its UKCP09 predecessor) produces broad ranges. These are wider than those derived from the multi-model information in isolation (Sexton et al., 2012; Harris et al., 2013), and therefore give a fuller picture of plausible changes that reduces the risk of overconfident decision-making. However, the probabilistic format should not be misinterpreted as an indication of high confidence in the weight of evidence behind specific outcomes. For example, developments in our physical understanding of climate change drivers and improvements in modelling capability could lead in the future to revised projections giving different probabilities for changes in UK variables. The Strand 1 projections should be seen as a source of broad guidance that forms a useful starting point for risk assessments. They can be used together with other sources of information such as the Strand 2 and 3 projections (see also discussion in section 5.2).

The methodology for Strand 1 is an updated version of the Bayesian statistical framework that underpinned UKCP09. This was outlined by Murphy et al. (2009), and described fully by Sexton et al (2012) and Harris et al. (2013). The technical description in section 2.2 focuses on summarising developments to the UKCP09 methodology. It does not repeat a detailed description of aspects that remain unchanged since UKCP09, which interested readers can review in the papers cited above. A full description of the UKCP18 method will also appear in Harris et al. (2019). A brief review of key elements of the calculations is provided in the list below:

- Definition of a "prior" parameter space representing uncertainties in a set of key earth system processes in a single climate model.

- Use of PPE integrations to sample variations in historical simulation skill and projected future response at a set of locations in parameter space.

- Calibration of climate model emulators and pattern scaling techniques to provide estimates of past and future climate at any point in parameter space.

- Estimation of the "structural" component of model error, by searching parameter space for best analogues to results of an independent multi-model ensemble.

- Production of probabilistic projections of future response by integrating over parameter space, adding in the estimated structural component of model error in past and future climate, and weighting each point according to estimated model quality (expressed as a relative likelihood).

- Use of an emissions-driven approach to capture uncertainties in both physical and carbon cycle feedbacks on the response to a given emissions scenario. Both types of feedback contribute substantially to uncertainties in projections of global mean surface temperature (GMST) (e.g. Knutti et al., 2008), which in turn influences the spread of projected regional changes. Accounting for carbon cycle uncertainties therefore ensures that the probabilistic projections take fuller account of known limitations in the current modelling of earth system processes.

Interpretation of the probabilistic projections

As discussed above, the results are conditional on the climate models used in the methodology. For example, if each international climate modelling institute could be persuaded to run a PPE of variants of its CMIP5 model, combining these could potentially produce a range of outcomes shifted and/or broadened with respect to the Strand 1 results. However, this is a hypothetical scenario, since such an experiment does not currently exist. The probabilistic projections are also conditional on other methodological choices including the expert prior distributions for uncertain model parameters, the selection of observations used to constrain the distributions and the method used to estimate the structural component of modelling uncertainty (termed "discrepancy" by Sexton et al. 2012). Sexton and Murphy (2012) tested the sensitivity of the UKCP09 results to plausible variations in these choices, finding only modest impacts on the results.

A key assumption is that the discrepancy term can be quantified by using the differences between outputs of PPE variants and other climate models as a proxy for structural simulation errors relative to the real world. This neglects the impact of systematic errors that are common to all climate models. Our rationale is that the presence of systematic errors represents an important but unavoidable caveat, and that the available models are sufficiently skilful that the conditional probabilistic projections derived from them provide useful advice about known uncertainties in future changes.

This and other methodological choices are subjective judgements, which some authors may disagree with. For example, Frigg et al. (2015) questioned the decision-relevance of the UKCP09 probabilistic projections, citing in particular concerns relating to systematic model errors. In practice, common biases occur in some but not all of the observables used to assess climate models. For example, the observed metrics of recent climatological averages used as constraints in UKCP09 were found to lie within the spread of modelled outcomes (Sexton et al., 2012). Nevertheless, current models do show common biases in their simulations of certain phenomena, such as a spurious convergence zone in the southern hemisphere of the tropical Pacific Ocean that leads to excessive precipitation in that region (Oueslati and Bellon, 2015).

In the future, climate models can be expected to improve, perhaps to the point at which a collection of climate models could be provided that is free from common biases. However, this is likely to be a slow process. For example, Knutti et al. (2013) show that whilst CMIP5 models showed some improvements relative to their CMIP3 predecessors, the progress in performance was incremental rather than substantial.

Developments for UKCP18

For UKCP18, the probabilistic projections are built from PPEs derived from the HadCM3 model, as in UKCP09 (see Introduction), while the required estimates of structural modelling uncertainty are provided by transient climate change simulations from CMIP5 earth system models. These replace the CMIP3 atmosphere-mixed layer ocean simulations used for this purpose in UKCP09. A basic prerequisite is that the PPEs at the core of the methodology remain credible, in the context of the current state-of-the-art in climate modelling. Flato et al. (2013) found evidence of a general improvement in the CMIP5 ensemble compared to CMIP3, including better simulation of surface temperature at regional scales and rainfall at large (continental to global) scales. However, the improvements were modest, and the ranges of simulation skill across members of the CMIP3 and CMIP5 ensembles show considerable overlap (Knutti et al., 2013). Sexton et al (2016) compared the UKCP09 results against CMIP5 models, considering both historical performance and projected changes. The HadCM3-based PPEs were found to be competitive with CMIP5 models in historical simulation skill, based on evaluation using a standard set of model assessment criteria[2] (see also Murphy et al., 2014). This reflected the status of HadCM3 as one of the best-performing CMIP3 models in terms of historical evaluation metrics. Ranges of projected change were also found to be broadly consistent between UKCP09 and CMIP5. An exception was that CMIP5 provided fewer projections showing substantial future reductions in summer rainfall (i.e. reductions of exceeding 10% per degree of global warming) for England and Wales, and more simulations showing modest increases, than suggested by the corresponding UKCP09 probability distributions.

These results support the chosen approach, allowing Strand 1 to provide an updated product that combines evidence from HadCM3-based PPEs and CMIP5 models in a consistent manner, retaining the Bayesian statistical framework used for UKCP09 (Goldstein and Rougier, 2004). The probabilistic projections are provided for a core set of UK climate variables (Fung et al., 2018). The new projections remove the requirement previously placed on users to evaluate the probabilistic projections and CMIP5 results as independent lines of evidence (see http://ukclimateprojections.metoffice.gov.uk/24127). However, we still recommend that the probabilistic results should, where feasible, be compared against alternative projections (notably Strands 2 and 3 of UKCP18), in order to support robust analysis of hazards relevant to impacts and decision-making.

---

[2] The assessment included quantitative evaluation of spatial fields of multi-year climatological averages of surface temperature, precipitation, sea level pressure, net radiation, net cloud radiative effect and its shortwave and longwave components, planetary albedo, and atmospheric temperature and zonal wind at 200hPa and 850hPa. Also included was qualitative evaluation of the NAO, historical trends in Arctic sea-ice extent, the North Atlantic storm track and the frequency of atmospheric blocking events.

In addition to use of CMIP5 earth system models, the updated probabilistic projections include several developments in methodology and scope:

- A 57-member PPE of variants of the earth system configuration of HadCM3 (Lambert et al., 2013; Murphy et al., 2014), produced subsequently to UKCP09, is used to represent the effects of parametric uncertainties due to ocean, aerosol and carbon cycle processes. Use of this ensemble (hereafter ESPPE) simplifies the method by allowing the number of PPE inputs to be reduced from seven to three. This is mainly because ocean, aerosol and carbon cycle uncertainties can now be quantified from one ensemble, whereas separate ensembles were used in UKCP09 (Murphy et al., 2009; Harris et al., 2013). It also allows the effects of uncertainties in ocean and carbon cycle processes on spatial patterns of climate change to be considered, alongside influences of land surface and atmospheric processes. In UKCP09, ocean and carbon cycle uncertainties were only accounted for in projections of GMST.

- Improved observational constraints are used to weight projections from different points in parameter space, by adding metrics of historical change in upper ocean heat content and CO_2 concentration (the latter to constrain carbon cycle feedbacks, following Booth et al., 2017). The use of historical surface temperature changes is also updated to consider changes up to 2017 rather than 2000, thus including the recent "warming hiatus" period (e.g. Trenberth, 2015).

- The representation of historical changes in external forcing has been improved, by using a probability distribution for anthropogenic aerosol forcing provided by AR5 (Myhre et al., 2013), and accounting for uncertainties in fossil fuel and land-use carbon emissions (Booth et al.,2017, updated).

- The methodology has been extended to present the probabilistic projections for individual years rather than for the 30-year averages of UKCP09, based on the method of Sexton and Harris (2015, updated). This allows effects of climate variability on annual to decadal time scales to be accounted for, facilitating analysis of the changing risks of seasonal extremes and allowing observed events to be placed in the context of the projections. Also, this annual presentation of data allows users flexibility to choose their own baseline period (see Fig. 2.10 and related discussion), if they wish to depart from one of the options provided on the UKCP18 website. The website options consist of the standard UKCP18 period of 1981-2000, plus 1961-1990 and 1981-2010.

- This extension allows the probabilistic projections to contribute to user requirements for more information on variability and extremes (driver 4 of section 1.2), providing context for the new modelling information in Strands 2 and 3.

2.2. Technical summary of methodology

This section provides more detail on the Strand 1 methodology, supported by further information in Appendices A-C. A full description will be given by Harris et al. (2019). Illustrative examples of key results follow in section 2.3.

 The probabilistic projections provide conditional probability density functions (pdfs) expressing a broad range of plausible outcomes for historical and future climate variability and change. The projections are driven by historical changes in natural and anthropogenic forcing agents to 2005, switching subsequently to scenarios of future emissions.
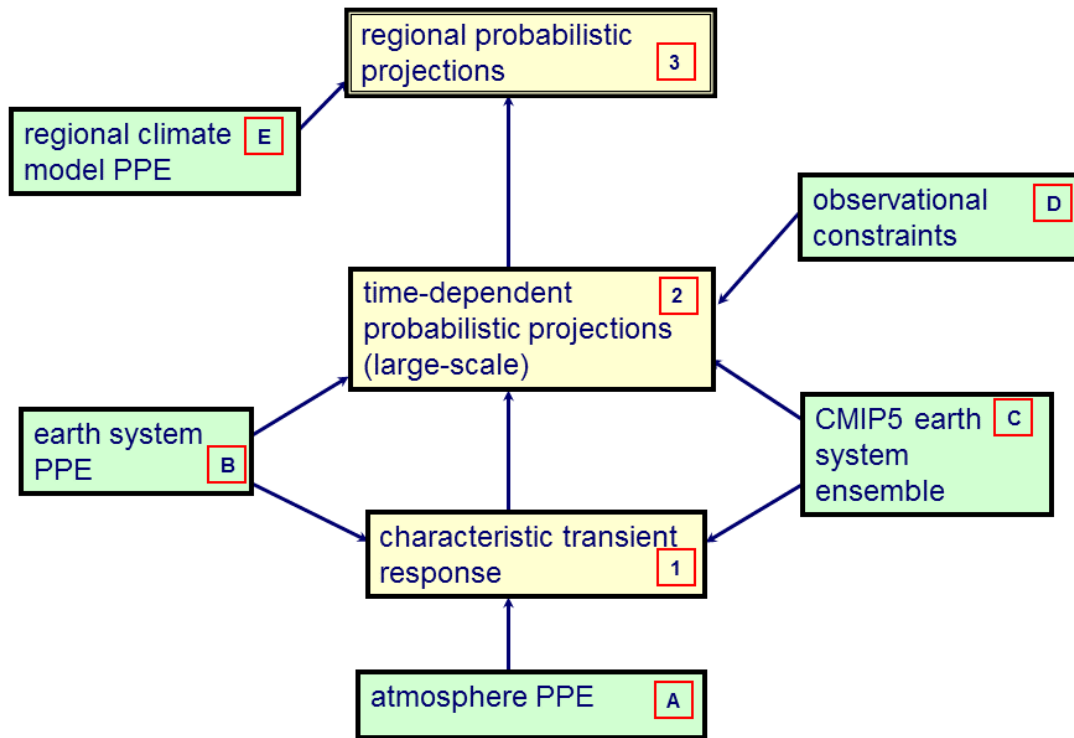

Figure 2.1. Major components of the methodology underpinning the UKCP18 probabilistic projections. Yellow boxes show the three main stages of the statistical calculations, green boxes show the main climate modelling and observational inputs to the method.

The projections are constructed using three PPEs of variants of HadCM3 (Gordon et al., 2000). In these, process uncertainties are investigated by perturbing model parameters in the land, atmosphere, sea-ice, ocean, sulphur cycle and carbon cycle components, within expert-specified ranges. As outlined above, these PPEs are combined with results from CMIP5 earth system models in order to achieve a combined sampling of parametric and structural uncertainties in physical and carbon cycle responses. A set of observational metrics of historical model performance is used to account for differences in credibility between alternative outcomes. The methodology involves three main stages (Fig. 2.1), summarised below. Figure 2.2 shows how the modelling and observational inputs feed into the main statistical components included in each stage.
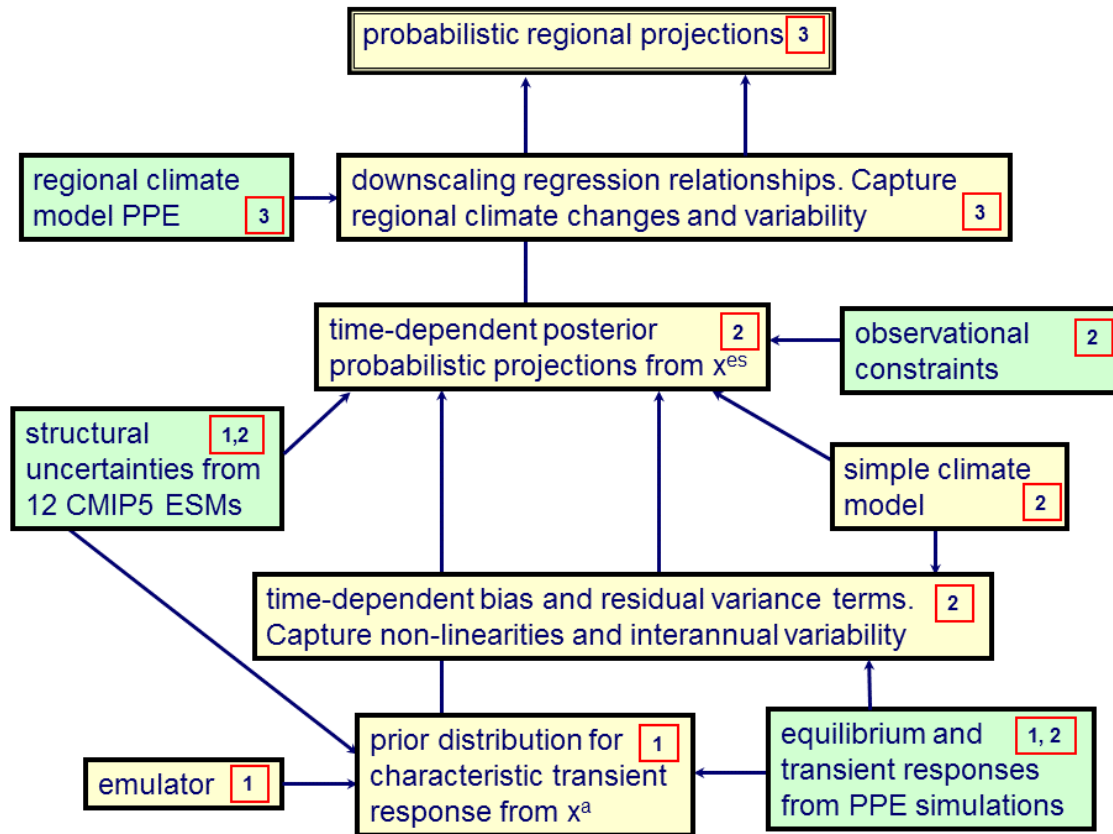
Figure 2.2. Yellow boxes show the main statistical tools and steps in the three stages of the method used to produce the probabilistic projections. x^a denotes the 31-dimensional prior distribution of uncertain parameters in the atmosphere component of HadCM3. x^{es} refers to the 50-dimensional space of uncertain earth system processes obtained by augmenting and combining x^a with prior distributions for 19 earth system variables influencing global climate changes (details in text). Prior and posterior distributions of climate variables are obtained by sampling projected changes from multiple locations in these parameter spaces. Green boxes show (as in Figure 2.1) the main climate modelling and observational inputs to the method, but labelled here according to the stage(s) to which they contribute.

Stage 1: Prior pdfs of characteristic climate response

In the first stage, pdfs (for a given climate variable, location and season) are obtained as a characteristic 20-year mean response per unit GMST rise that, when scaled by time-dependent GMST changes, maximises the explained variance of transient regional changes in coupled ocean-atmosphere simulations (Harris et al., 2006). We refer to such variables as normalised transient responses (NTR). At this stage, the pdfs are prior results which do not account for observational constraints. The calculations are updated from Sexton et al. (2012). The main elements are shown schematically in Fig. 2.2, and involve the following steps:

a. The equilibrium response to doubled CO_2 is predicted. This involves a relatively large (280 member) PPE of simulations using the coupled atmosphere-mixed-layer (hereafter SLAB) configuration of HadCM3 (Box A in Fig. 2.1). These simulations provide values of a multivariate set of historical climate means expressed as global spatial fields (SLAB_CLIM), as well as the response to doubled CO_2 (Sexton et al., 2012). The response variables (Fig 2.3, top and middle panels) are the equilibrium climate sensitivity (ECS[3]), and regional changes per unit GMST change, hereafter normalised equilibrium responses (NER).

---

[3] ECS is defined as the equilibrium (steady state) response of globally and annually averaged surface temperature to a doubling of CO_2 concentration in the atmosphere. It is a standard benchmark measuring the long term sensitivity of a climate model, or of the real climate system, to a sustained change in greenhouse gas concentrations.
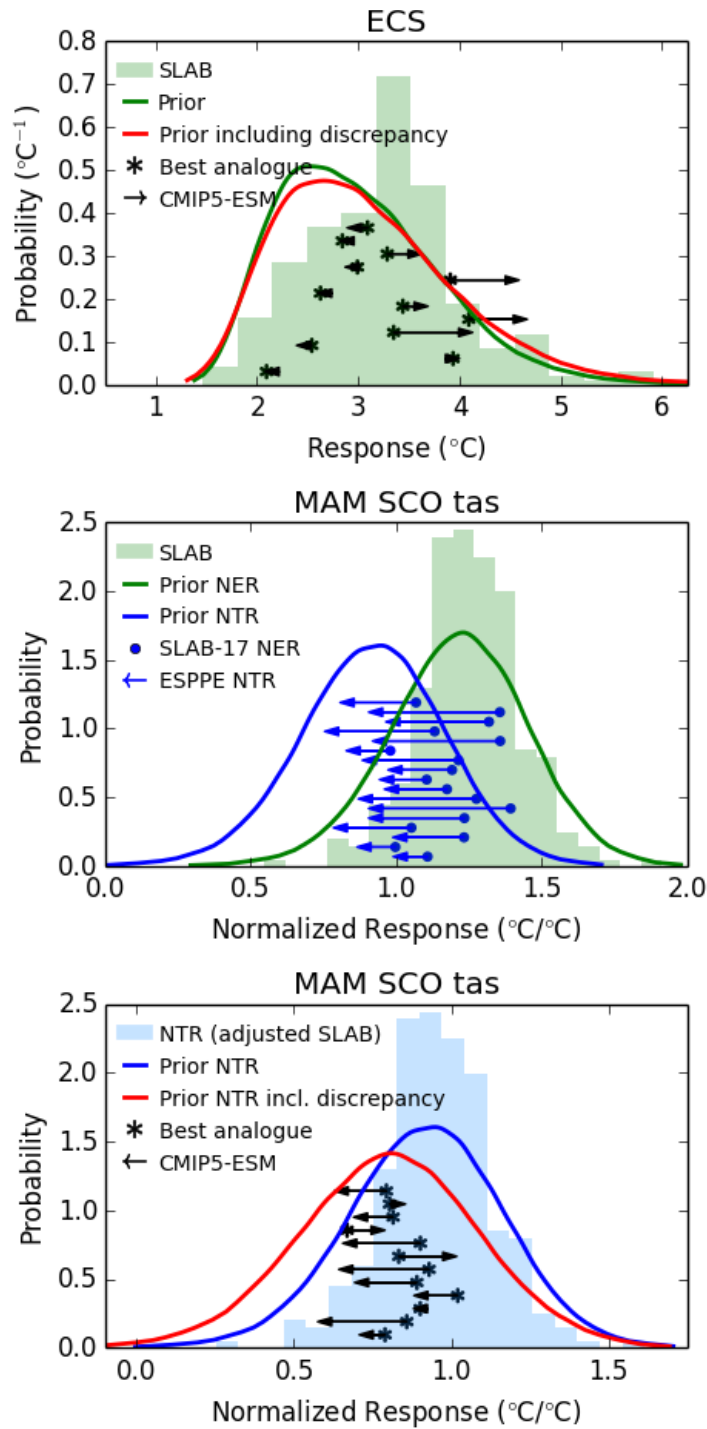
Figure 2.3. Stage 1 of the probabilistic projections method of Strand 1. Top panel shows distributions of equilibrium climate sensitivity (ECS). These consist of: simulated responses to doubled CO_2 from the 280-member SLAB PPE using the atmosphere-mixed layer ocean configuration of HadCM3 (green histogram); prior distributions extended to sample the whole atmospheric parameter space of HadCM3 using statistical emulators, with (red curve) and without (green curve) the impact of structural model uncertainties (discrepancy). The mean and variance of the discrepancy distribution is calculated from the twelve distances between the asterisks and arrow heads, where the latter shows ECS values of the twelve CMIP5 earth system models used to estimate structural uncertainty, and the asterisks their nearest analogues in the HadCM3 parameter space. Middle and lower panels show the response of surface air temperature for Scotland in March to May, expressed as a normalised response per degree of globally-averaged warming. The middle panel shows prior distributions of the normalised equilibrium response (NER) to doubled CO_2 (green), and the normalised transient response (NTR, blue), adjusted to allow for ocean circulation changes and parametric uncertainties in ocean, sulphur cycle and terrestrial ecosystem processes in the earth system configuration of HadCM3. The average offset between NER and NTR is calculated from the 17 blue lines. Circles show NER values for 17 SLAB model variants, and arrow heads represent responses averaged over variants in the 57-member earth system perturbed parameter ensemble (ESPPE) using corresponding perturbation sets in their atmosphere component. The lower panel shows NTR values for the SLAB PPE (blue histogram) following application of the offset, and illustrates the impact of adding discrepancy to the prior distribution of NTR from the middle panel (cf red and blue curves). In this panel, asterisks and arrows show the relationship between CMIP5 estimates of NTR and their best PPE analogues.

b. The 57-member ESPPE (Box B in Fig. 2.1) is used to account for the effects of coupled earth system processes omitted from SLAB.

- The ESPPE uses the earth system configuration of HadCM3, including dynamical ocean and vegetation modules with an interactive carbon cycle (Lambert et al., 2013; Murphy et al., 2014). Members are distinguished by multiple simultaneous perturbations to parameters in the atmosphere, ocean, sulphur cycle and terrestrial ecosystem components. The simulations were driven using historical and future emissions of CO_2 and aerosol precursors, and concentrations of other major greenhouse gases. Natural historical forcing, due to variations in solar irradiance and major volcanic eruptions, was also included. From previous experiments, 17 perturbed parameter sets were available for each of the four earth system components of the model (Harris et al., 2013), creating a space of 17^4 possible combinations of the 68 parameter sets. It was not feasible to run an ensemble of this size. Therefore, a Latin hypercube experimental design was used to create an initial 68 member ESPPE in which parameter sets from each model component were combined to sample the space of possible combinations as efficiently as possible (details in Lambert et al., 2013). This ensemble was subsequently reduced to 57 members, following a simple screening based on historical performance. This removed members with substantial biases in planetary net radiation, global surface temperature or simulated vegetation distributions (Lambert et al., 2013). Murphy et al. (2014) compared the performance of the surviving 57 members against CMIP5 earth system models for a few key aspects of historical climate, finding similar levels of skill in both ensembles[4]. ESPPE simulations are available for the RCP2.6 and 8.5 scenarios, and for SRES A1B.

- The perturbed atmospheric parameter sets in the ESPPE correspond to 17 members of SLAB, hereafter SLAB17. By comparing corresponding ESPPE and SLAB17 members, samples of 57 differences between their historical climatology (SLAB_CLIM cf COUPLED_CLIM) and future response variables (NTR cf NER, see Fig 2.3, middle panel) are obtained. These differences account for dynamic ocean effects and parametric uncertainties arising from ocean transport, sulphate aerosol and terrestrial carbon cycle processes. In the example of Fig 2.3, NTR is smaller than NER, partly because reductions in the North Atlantic Meridional Overturning Circulation (AMOC) restrict the regional warming in the ESPPE simulations. Values of COUPLED_CLIM are compared against their observed counterparts (CLIM) in Stage 2c below.

- The average of the 57 differences is added to SLAB_CLIM and NER to obtain estimates of COUPLED_CLIM and NTR respectively. This simple approach assumes that the offset is independent of location in the atmosphere parameter space, because the ESPPE is too small to support reliable estimation of potential parametric dependencies. However, an estimate of uncertainty in the offset is accounted for below. This is derived from 15 of the SLAB-17 members for which more than one corresponding ESPPE member is available (Lambert et al., 2013).

c. The estimated values of COUPLED_CLIM and NTR are then extended using statistical emulators trained on the climate model results. The emulators are used to estimate outcomes for parameter combinations for which no climate model simulation is available, allowing comprehensive sampling of prior distributions for 31 parameters in the atmosphere component of HadCM3 (blue curve in Fig 2.3, middle panel). The sampling accounts for uncertainties in the emulator output, and in the offsets discussed above (cf blue and green curves in Fig 2.3, middle panel).

---

[4] The comparison with CMIP5 models considered recent climatological patterns of surface air temperature, precipitation and sea level pressure, plus simulations of changes in GMST during the 20th century.

d. Results from (c) are then modified to account for structural modelling uncertainties, since these are not represented in PPEs constructed from a single climate model.

- Structural uncertainties (the aforementioned discrepancy term of Sexton et al., 2012) are estimated using an ensemble of CMIP5 earth system models (Box C in Fig. 2.1). Like the ESPPE, these transient simulations were driven by historical and future CO_2 emissions, though an ensemble of sufficient size was only produced for RCP8.5. Fifteen models were available, however two were excluded due to data gaps, and one of a pair of closely-related models submitted by the Geophysical Fluid Dynamics Laboratory was excluded, leaving the 12 models (hereafter CMIP5-ESM) listed in Table 2.1. The structural uncertainty calculations involve searching the atmospheric parameter space for emulated estimates of COUPLED_CLIM, ECS and NTR that are closest to those of each CMIP5-ESM member. Differences between each CMIP5-ESM model and their best PPE analogues (e.g. Fig 2.3, top and bottom panels) are then used to define a multivariate Gaussian distribution of discrepancy.

- Prior pdfs of ECS and NTR (Fig. 2.3, red curves) are then produced using the Bayesian method of Sexton et al (2012), based on the general framework of Goldstein and Rougier (2004). These are presented as a set of 106 outcomes obtained by using the emulator to sample the prior atmospheric parameter space of HadCM3 (x^a), and adding estimates of structural uncertainty sampled from the above discrepancy distribution (boxes labelled "1" in Fig. 2.2). Weighting according to relative likelihood is carried out in Stage 2 below.

Table 2.1. The 12 CMIP5 earth system models selected to provide estimates of structural modelling uncertainty in the probabilistic projections of Strand 1, derived from simulations driven by prescribed future emissions of CO_2 under the RCP8.5 scenario.

Modelling group	Group acronym	Model designation
Univ. Tokyo, National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	MIROC	MIROC-ESM
Beijing Climate Centre, China Meteorological Administration	BCC	BCC-CSM1.1(m)
Canadian Centre for Climate Modelling and Analysis	CCCMA	CanESM2
Beijing Normal University	BNU	BNU-ESM
Community Earth System Model Contributors	NSF-DOE-NCAR	CESM1-BGC
Centre National de Recherches Météorologiques / Centre Européen de Recherche et Formation Avancée en Calcul Scientifique	CNRM-CERFACS	CNRM-CM5
NOAA Geophysical Fluid Dynamics Laboratory	NOAA GFDL	GFDL-ESM2G
Met Office Hadley Centre	MOHC	HadGEM2-ES
Institut Pierre-Simon Laplace	IPSL	IPSL-CM5A-LR
Max-Planck-Institut für Meteorologie	MPI-M	MPI-ESM-LR
Meteorological Research Institute	MRI	MRI-ESM1
Russian Institute for Numerical Mathematics	INM	INM-CM4

Stage 2: Posterior pdfs of time-dependent variability and change at large scales

The second stage is an updated version of the "timescaling" approach described by Sexton and Harris (2015), involving the statistical elements shown in Fig. 2.2 (yellow boxes labelled "2"). The basic approach involves producing time-dependent projections of GMST at the annual time scale (Fig. 2.4a), and using these to apply a simple linear scaling to corresponding estimates of NTR from Stage 1. This produces estimates of time-dependent regional variability and change (TRVC, Fig. 2.4b). These estimates are produced for multiple locations in a version of the HadCM3 parameter space expanded to consider uncertainties in ocean transport and global carbon cycle processes in addition to the atmospheric processes explored in x^a. From these sampled estimates, probabilistic projections of GMST and TRVC are constructed (Figs. 2.4a,b). These TRVC pdfs are provided for five HadCM3 land boxes labelled "Scotland", "Ireland", "Northern England", "Southern England" and "Wales" in Fig. 3 of Harris et al. (2010). In Stage 3 these variables, hereafter UK_GCM, are subsequently downscaled to obtain corresponding year-by-year projections for the 25km grid and aggregated (country, river-basin and administrative) regions of UKCP18 (Fung et al., 2018). The main steps in Stage 2 are described below. Stages 2a-c describe the methodological components, and Stage 2d the production of constrained transient projections:

a. **Transient global projections**

- These are made using the two-box simple climate model (SCM) of Harris et al (2013), which uses energy balance principles to predict ocean and land surface temperature in response to prescribed natural and anthropogenic forcing. Vertical ocean heat transport is modelled using a one-dimensional diffusion-advection equation, and a carbon cycle component is included to allow the simple model to predict globally averaged carbon exchanges between atmosphere, land and ocean.

- In order to extend the Bayesian method of Stage 1, x^a is generalised to include a set of 19 input parameters to the SCM (x^{scm}) that control its response. Two of these are ECS_L and ECS_O (components of ECS calculated separately over land and ocean points). These are provided by Stage 1 as a function of location in x^a. Other parameters in x^{scm} control ocean heat uptake and the global carbon cycle. Prior distributions for these are obtained by calibrating the SCM to replicate the results of each ESPPE member, and then specifying distribution functions for SCM parameters consistent with the joint multivariate spread of the 57 sets of results. A space x^{es} of earth system process uncertainties is then defined by combining the priors of x^a and x^{scm}. In this, we assume independence between SCM parameters and outputs from x^a, apart from ECS_L and ECS_O. Sampling of x^{es} to produce transient projections is described in (d) below.

- Estimates of discrepancy are required for global, time-dependent outputs of the SCM. In addition to GMST, these include atmospheric CO_2 concentration and upper ocean heat content (OHC, representing the top 700 metres). Historical changes in these variables are used as observational constraints in (c) below. For each CMIP5-ESM member, the best PPE analogues in x^a, identified in Stage 1d, are augmented by finding best analogues in x^{scm}. The SCM is then run using these parameter settings (Fig. 2.4a, top left). Differences between the GMST, CO_2 and OHC outputs of each CMIP5-ESM member and their best analogues (Fig 2.4a, top right) are then used to quantify means and variances of the relevant discrepancy distributions. The mean discrepancy for GMST reaches ~-0.3ºC beyond 2050 (dark blue curve in Fig. 2.4a, top right), partly because the HadCM3-based analogues tend to overestimate slightly the carbon cycle feedbacks in CMIP5-ESM members. Following production, the annual discrepancy estimates are low-pass filtered[5]. This is done to retain structural uncertainty signals in long-term climate change, and exclude residual effects of internal variability arising from the limited size of the CMIP5-ESM ensemble.

- The timescaling procedure also requires specification of time-dependent bias and residual variance terms. These are calibrated by comparing ESPPE results against timescaling estimates produced from corresponding locations in x^{es} (e.g. red and yellow curves in Fig 2.4a, top left). The bias term (e.g. red curve in Fig. 2.4a, top right) represents the average error obtained in estimating transient responses from the time-invariant climate change metrics of Stage 1. In the case of GMST, contributions might arise, for example, from neglect in the SCM of potential time-dependence in the strength of global climate feedbacks (e.g. Gregory and Andrews, 2016), or variations in the efficacy of different forcing agents in driving surface temperature responses (Shindell, 2014). Potential dependencies of such errors on location in x^{es} are not accounted for (since insufficient ESPPE simulations were available to quantify these), but may contribute to the residual variance term. In the case of GMST, for example, the residual variance is derived from the orange curves in Fig. 2.4a (top right). In addition, an important contribution to the residual arises from uncertainty due to climate variability on 1-20 year time scales. The bias and residual variance terms are low-pass filtered to remove noise arising from the limited number of ESPPE simulations available for calibration, with the exception that the bias is left unfiltered during years affected by major volcanic eruptions (see Appendix A).

- Realisations[6] of GMST (Fig 2.4a, bottom left), CO_2 and OHC are produced by running the SCM using input parameters sampled from x^{es}. Like the ESPPE, the SCM is driven by historical solar and volcanic forcing, plus emissions of CO_2 and concentrations of other major greenhouse gases, switching from historical to scenario values after 2005. Anthropogenic changes in global-mean aerosol forcing are also included in the SCM. Uncertainties in these are accounted for by sampling from a skew normal distribution consistent with an estimate of the median and range for total aerosol forcing in 2011 (Myhre et al., 2013). Historical and future sulphur dioxide emissions are used to scale the forcing for other periods.

- Each SCM realisation is modified, by adding the timescaling bias term and sampling the discrepancy and timescaling residual distributions. The timescaling residual adds a noise component to the SCM outputs that accounts for climate variability, and is generated using a sampling method (AUTOVAR) that preserves temporal autocorrelations found during calibration to ESPPE simulations (Appendix A).

---

[5] In this report, low-pass filtering of time series is applied, to distinguish between long-term climate changes and variability on shorter time scales. This is done using a Butterworth filter, with a cut-off period of 20 years unless stated.

[6] The Strand 1 outputs consist of a large sample of projected changes produced using statistical methods. Individual members of this sample are referred to as "realisations" in this report. The individual simulations of climate change that constitute Strands 2 and 3, provided by global or regional climate models, are referred to as "projections".
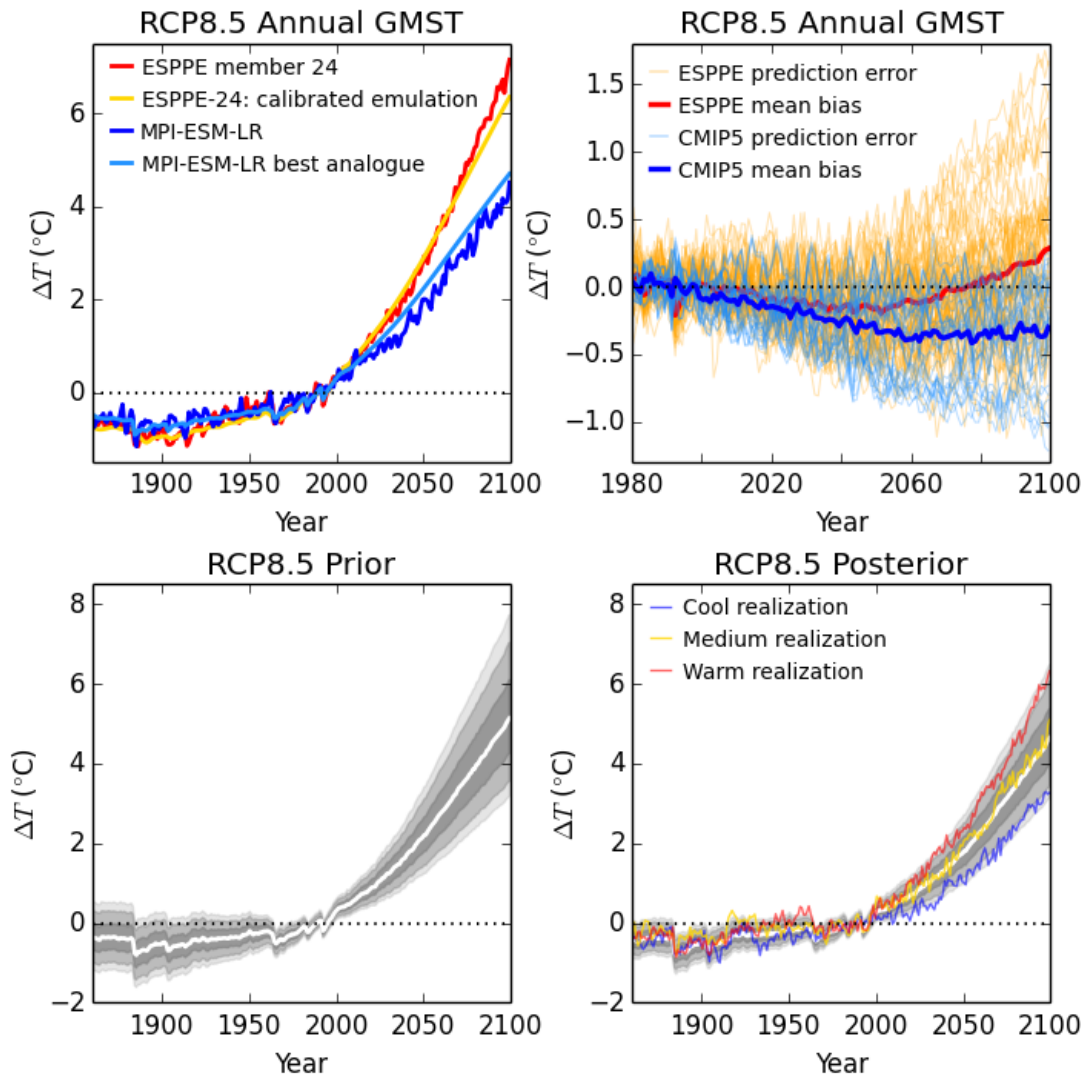
Figure 2.4a. Stage 2 of the probabilistic projections method. Top left panel shows transient projections of global mean surface temperature (GMST) from two earth system model projections, one from the ESPPE (red) and one from CMIP5 (dark blue). Also shown in yellow and light blue are reconstructions of their time series obtained using the simple climate model (SCM), run with input climate response parameters prescribed from the earth system model results. Top right panel shows time-dependent errors (orange) in SCM predictions of the response of the 57 ESPPE members, of which the differences between the yellow and red lines in the top left panel are one example. These are used to calculate time-dependent bias (red curve) and residual variance terms, that represent errors and uncertainties arising from use of time-invariant metrics of climate change from Stage 1 (such as ECS) to estimate transient changes via use of the SCM. Internal climate variability is a significant contributor to the residual term. The light blue curves show errors incurred by using the SCM to reconstruct the projections of CMIP5 earth system models. For these reconstructions, the SCM uses input parameters that represent the best analogues to the relevant CMIP5 model that can be found in the parameter space of HadCM3 earth system processes, x^{es} (see text). The twelve light blue curves are used to calculate the time-dependent mean (dark blue curve) and variance of the discrepancy distribution for SCM outputs, in this case GMST. Bottom left panel shows prior probabilistic annual projections of GMST for the RCP8.5 scenario. These are produced by using the SCM to sample x^{es} comprehensively, including the bias, residual and discrepancy terms described above. The white line shows the median, and shades of grey show the 5, 10, 25, 75, 90 and 95% probability levels. Bottom right shows corresponding transient posterior distributions, modified by application of observational constraints to weight outcomes from specific locations in x^{es}. Coloured curves show three examples of the 3000 realisations of variability and change that constitute the probabilistic projections.

b. Transient regional projections

- For a given sample member, time-dependent regional changes (TRVC) are obtained by using the prediction of GMST to scale the relevant emulated estimate of NTR (e.g. Fig. 2.4b, top left). Errors associated with this linear pattern-scaling assumption (Fig. 2.4b, top right) are accounted for by applying time-dependent bias and residual terms to the scaled regional projections (as for the global projections above). These terms are calibrated by comparing scaled projections from SLAB17 against TRVC outcomes from ESPPE members with corresponding atmosphere parameter sets. Since regional responses to volcanic eruptions cannot be assumed to scale with GMST in the same way as the response to greenhouse forcing, we modify the scaling method during historical periods affected by major eruptions (see Appendix A).

- At this stage, projected TRVC values consist of annual changes from 1860-2100 for the UK_GCM variables (e.g. Fig. 2.4b, bottom left), and also for four large-scale indices (BRAG) describing historical changes in global patterns of surface temperature (Braganza et al., 2003). These are used below to constrain the projections.
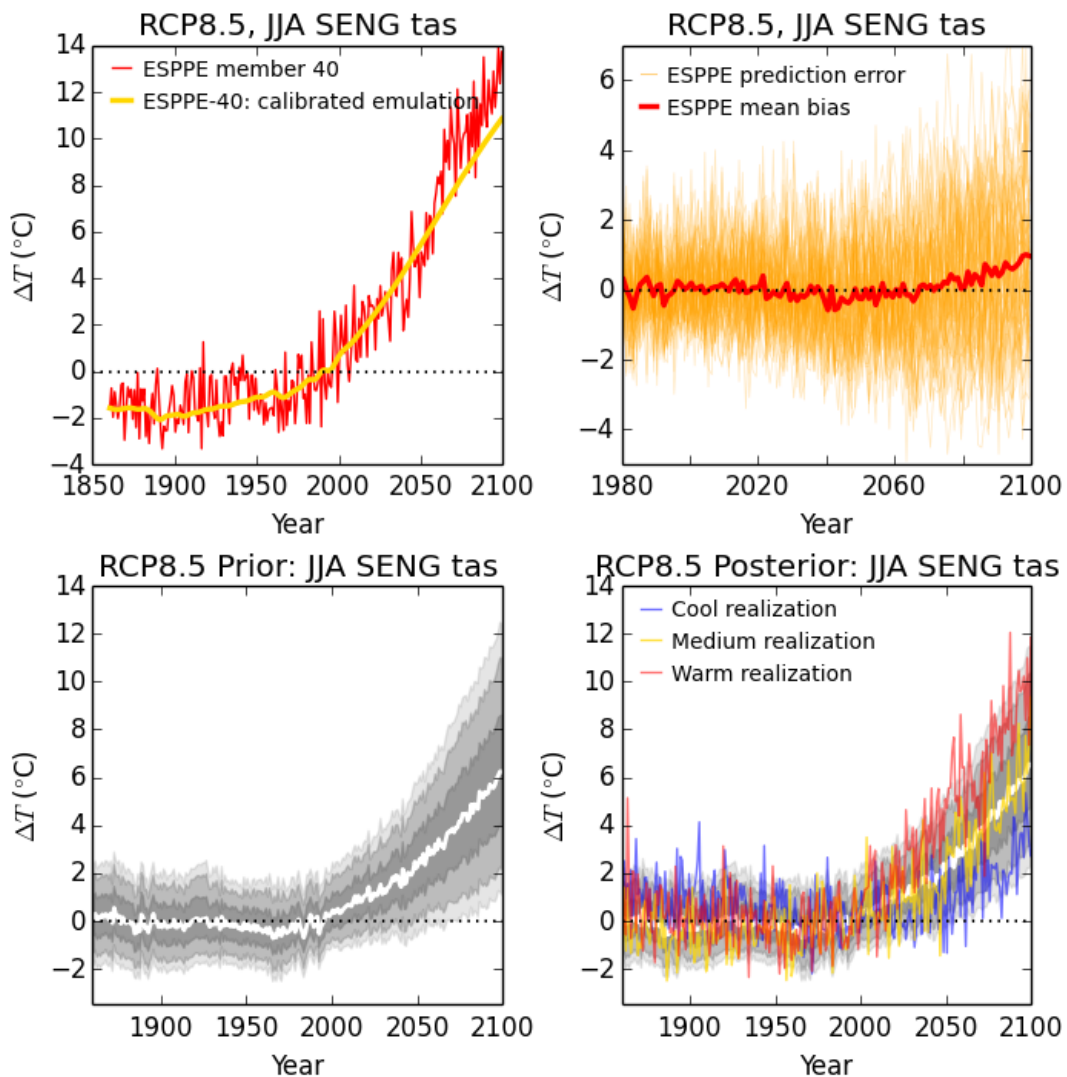

Figure 2.4b. As Fig. 2.4a, for summer surface air temperature over Southern England. In this case, reconstructed estimates (e.g. top left panel, yellow curve) are obtained by using the time series of GMST predicted by the simple model to scale NTR for the relevant ESPPE member. Orange curves in the top right panel show errors in these estimates for the 57 ESPPE members, with the average bias in red. There is no time-dependent discrepancy information in this panel, as this term is only calculated for global variables output by the SCM (GMST, CO_2 concentration and upper ocean heat content). Structural uncertainty in regional patterns of change (NTR) is accounted for in Stage 1 (Fig. 2.3), and assumed invariant in time.

c. Observational constraints (Box D in Fig. 2.1)

- The role of the specified observable quantities is to update joint prior probability distributions according to the relative likelihood that alternative sampled realisations replicate the observables. The observational constraints are applied by weighting sampled outcomes from locations in x^{es}. Following Sexton et al. (2012), weights are calculated from the multivariate distance between emulated estimates of a set of historical variables and verifying observations, normalised by a covariance matrix representing the sum of uncertainties due to emulator error, observational error and discrepancy.

- The observables consist of CLIM, plus several indicators of historical climate change, specifically GMST, BRAG, OHC and CO_2 concentration (Booth et al., 2017). Our set of constraints is informed by an assumption that climate changes are expected to be subject to influences from a wide variety of earth system processes, both from within, and remote from, any given region of interest. Therefore, we design the constraints to cover a range of climate variables, expressed mainly at global or large regional scales. Use of a broad range of variables reduces the risk that a given model variant might receive an unrealistically high or low weight, due either to an untypically large bias in one particular metric, or due to a fortuitously good match that might arise from compensating biases. Metrics of historical climate change, for example, are prone to the latter issue, since a number of distinct physical or biogeochemical processes influence the net feedbacks that influence changes in properties such as GMST or global CO_2 concentration (e.g. Gregory et al., 2009).

- The CLIM observables include seasonal climatological spatial fields for the twelve climate variables listed in Appendix B (Table B.1). This is the same set used in UKCP09, and includes surface air temperature, precipitation, sea-level pressure and a variety of additional diagnostics relating to the earth's energy and hydrological cycles. The data amounts to about 175,000 observables. It is necessary to reduce its dimensionality, in order (a) to remove dependencies between variables affected by common physical processes[7], and (b) to make the required multivariate statistical calculations computationally feasible. This is done by identifying the six leading eigenvectors of the climatological variables in the SLAB ensemble, following closely the UKCP09 approach (Sexton et al., 2012). Amplitudes of these are emulated, and adjusted as described in Stage 1b to account for coupled ocean-atmosphere processes (COUPLED_CLIM). These are then compared with amplitudes for these observables projected onto the same set of eigenvectors, and used to estimate likelihoods associated with these mean climate constraints.

---

[7] For example, simulation errors in radiative fluxes, cloud amounts and surface air temperature at neighbouring grid points are likely to arise from common sources, and are therefore unlikely to provide independent information.

- Details of the constraint variables based on historical climate change are listed in Appendix B (Table B.2). The BRAG variables consist of four global scale surface air temperature indices (Braganza et el 2003). These are identical to those used in UKCP09, except here we use 17 years of additional observed data to the end of 2017. We also use trends for two periods (Table B.2), rather the single period employed in UKCP09. Our decision to include a constraint derived from historical CO_2 changes is based on Booth et al. (2017). They demonstrated that trends in atmospheric CO_2 concentration (c^{atm}) could be used to rule out earth system model simulations with an interactive carbon cycle that produced results inconsistent with observations. As described in section 3.3g, this approach was also used to select a subset of c^{atm} pathways from the ESPPE and CMIP5-ESM ensembles to drive the GC3.05-PPE simulations in Strand 2. Following studies (e.g., Skeie et al, 2014; Bodman and Jones, 2016) which note that future projections can be sensitive to the historical trend in ocean heat content, we also use the change in global mean heat content in the upper 700m (OHC) in the likelihood estimation. The 700m depth is chosen because data with better coverage over a longer period is available, compared to other choices. For most observables, two datasets were compared to obtain estimates of observational uncertainty (Sexton et al., 2012). Four datasets were used in the case of OHC. For SST and surface air temperature constraints, single datasets provided with associated uncertainty estimates were used.

- The effects of the observational constraints on the projected changes are discussed below.

d. Production of time-dependent posterior projections (Fig. 2.2)

- For each of the 10^6 members sampling x^a from Stage 1, plausible values of ocean heat uptake and carbon cycle parameters are added by randomly selecting points in the prior space of x^{scm} (see Stage 2a). This creates a new prior sample exploring x^{es}, also of size 10^6, suitable for use in production of transient projections.

- Initially, historical realisations from 1860-2015 are produced, allowing a weight to be attached to each of the 10^6 sample members. In these realisations, historical uncertainty due to fossil fuel (FF) and land-use (LU) carbon emissions is accounted for. FF is sampled from an assumed Gaussian distribution (derived from the uncertainty estimate of Le Quéré et al., 2015). Alternative LU results are sampled from ten dynamic global vegetation models (Booth et al., 2017, updated). This ensures that the contribution to the weight of historical change in CO_2 concentration takes appropriate account of relevant emissions uncertainties.

- The 10^6 members are then resampled with replacement according to weight, to create a new 50,000 member sample that is convenient for the calculation of probability levels[8] of the time-evolving pdfs.

- A 3000-member sub-sample of timescaled realisations, of convenient size for users, is then produced (Figs. 2.4a,b show examples of pdfs derived from these, bottom right panels). In these, standard CMIP5 specifications of historical and future carbon emissions are used, rather than sampling FF and LU uncertainties as described above. For a given emissions scenario, each realisation therefore uses the same specification of past and future CO_2 emissions (as do the CMIP5-ESM and ESPPE simulations). Uncertainty in future FF and LU emissions therefore contributes to differences between the pdfs for different emissions scenarios (Fig. 2.8, discussed later), but not to the spread of outcomes for a specific emissions scenario.

---

[8] In this report, we use the term "probability level" to refer to quantiles of probability distributions from Strand 1. We use the term "percentile" where we are discussing quantiles from frequency distributions of results from Strands 2 and 3. This reflects the non-probabilistic nature of Strands 2 and 3, which are composed from sets of individual simulations that are not designed to provide estimates of the relative likelihood of different future climate outcomes.

Stage 3: Downscaled projections

The final stage (Fig. 2.2, yellow boxes labelled "3") involves downscaling the TRVC realisations to the four sets of target regions for Strand 1. The calculations use an 11-member PPE of the RCM configuration of HadCM3 (Box E in Fig. 2.1), documented in Murphy et al. (2009) and referred to henceforth as HadRM3-PPE. This was run from 1951-2100 at 25km horizontal resolution using a European domain (Fig. 3.8 of Murphy et al., 2009). The RCM simulations were driven by 11 members of the AO-PPE-A ensemble of Collins et al. (2011), a coupled ocean-atmosphere PPE containing the same atmospheric parameter settings as SLAB17. We used data from 1961-2100 to calibrate the downscaling relationships used in Strand 1. Prior to downscaling, HadRM3-PPE data is spatially interpolated to the target regions.

a. For monthly and seasonal variables (Fung et al., 2018), downscaling is achieved using linear regression relationships (e.g. Fig. 2.5, left panel) between projected changes in the RCM and driving GCM. The latter are provided by selecting a predictor grid point[9] from the UK_GCM variables described in Stage 2. This adjusts the GCM-based results to account for the finer scale information available from the RCM. For a given variable, a single regression relationship is calibrated by pooling changes for 1990-2099 relative to 1981-2000 from all eleven GCM-RCM pairs. The GCM predictor variables are low-pass filtered, to allow the regression coefficients to capture long-term climate change signals more effectively (e.g. blue curve in Fig. 2,5, right panel). The predictand RCM variables are left unfiltered, so that the regression residuals include uncertainties due to interannual variability (e.g. red curve in Fig. 2.5, right panel). These typically provides the dominant contribution to the spread of residuals. This can be seen by comparing the ranges of the black bar and grey shading during 2080-2099 in Fig. 2.5 (right panel). In addition, the residuals capture errors arising from the simple downscaling regression model, which does not account for potential non-linearities, or variations in the downscaling relationship across parameter space. The residuals are used to calibrate AUTOVAR for use in (c) below.

b. The 3000 TRVC realisations from Stage 2 are then filtered to extract low-pass climate change signals, and converted using the above downscaling relationship, under the assumptions that the relationships are independent of location in x^{es} and choice of emissions scenario (noting that the AO-PPE-A and HadRM3-PPE simulations were driven by the A1B scenario).

c. Finally, variability on 1-20 year time scales (grey shading in Fig. 2.5, right panel) is added to the downscaled realisations of long-term climate change, using AUTOVAR. For most variables the AUTOVAR sampling is as described by Sexton and Harris (2015), in which the residuals are assumed to be Gaussian. However, for specific humidity, total cloud amount and precipitation, the residual distributions are adjusted to account for errors introduced by this assumption. In the case of summer precipitation, for example, this preserves a positive skew present in the HadRM3-PPE simulations, ensuring that the statistical treatment of wet events is consistent with the underlying climate modelling results (Appendix A).

---

[9] For locations in the national 25km grid (Fung et al., 2018), candidate UK_GCM grid points include the one nearest the target location, and its neighbours. The nearest point is selected unless a neighbour can be found for which the downscaling residuals (measured by mean absolute deviation) are significantly smaller at the 10% significance level, determined by bootstrap resampling of the downscaling residuals. For the aggregated regions, predictors are chosen by selecting the UK_GCM point that is picked most often amongst the subset of 25km grid locations contained within the relevant region.
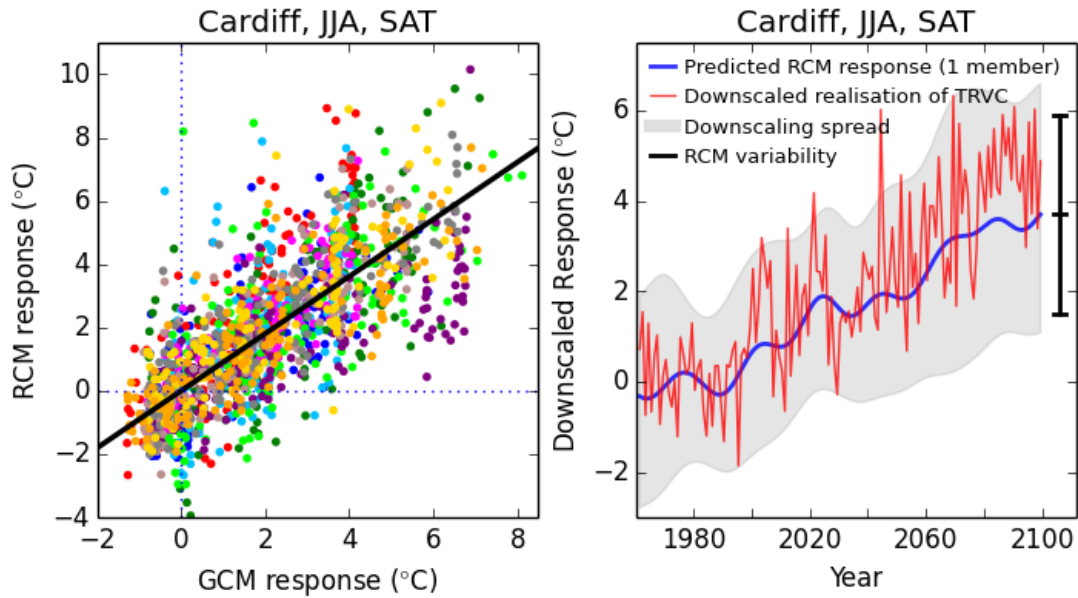
Figure 2.5. Stage 3 of the probabilistic projections method. Circles (left panel) show values of surface air temperature changes for a single 25km grid box containing Cardiff, comparing values from individual HadRM3-PPE members (distinguished by colour) against corresponding values from their driving global model simulations (which used the A1B emissions scenario, Collins et al., 2011). The regional model responses are for individual years from 1990-2099, relative to a baseline of 1981-2000. Contemporaneous global model changes are low-pass filtered to remove variability on time scales shorter than 20 years, in order to emphasise long-term climate changes. The linear regression (black line) has a coefficient of 0.90, and shows the relationship between the downscaled 25km climate change signal and that at the 300km driving model land point chosen to provide the predictor variable (see text). The scatter of the coloured dots around the regression line is due to residual uncertainties (see below). These uncertainties are sampled in the probabilistic projections, using the AUTOVAR method described in Appendix A. This is based on a Singular Value Decomposition that accounts for interannual and inter-variable covariances in the downscaling residuals, and temporal dependencies in the variance of residuals. The right panel shows the predicted low frequency signal of change (blue) obtained by applying the regression relationship to one of the driving model simulations, with the time series of variability and change in the corresponding HadRM3-PPE member in red. The grey shading shows the 5th-95th percentile range of the residual downscaling uncertainty (assumed Gaussian) around the blue curve. This uncertainty is sampled in each of the individual realisations that comprise the probabilistic projections. The black bar shows the 5th-95th percentile range of internal RCM variability on time scales of 1-20 years for 2080-2099. The corresponding residual range is slightly broader. This shows that internal variability in the RCM simulations provides a dominant contribution to the residual variance, with errors in the simple downscaling relationship providing a modest additional contribution. Such errors can arise from potential non-linearities and/or member-specific dependencies in the relationship.

Impact of observational constraints

Figure 2.6 demonstrates how posterior distributions for the transient response to RCP8.5 emissions change, as the observational constraints are applied successively. Examples are given for three variables, considering changes for 2081-2100 relative to 1981-2000: GMST (top panel), summer surface air temperature in Southern England (middle panel), summer precipitation change for Southern England (bottom panel). The starting point in each case is the red curve, which is the prior distribution emerging from Stages 2a,b above. This accounts for sampling of parametric modelling uncertainties in the earth system configuration of HadCM3 and discrepancy, estimated using CMIP5-ESM members. However, the red curves do not account for variations in the credibility of PPE variants at different locations in the parameter space of HadCM3.

In common with UKCP09, constraining the GMST response with only the historical mean climate variables leads to a substantial narrowing of the probability distribution (orange curve cf red curve in Fig. 2.6, top panel). A further narrowing is obtained when the constraint due to historical changes in surface air temperature is added (dark green curve cf orange curve). The impact is a little greater than was found in Harris et al (2013). This is due partly to the signal present in the additional 17 years of observational data used in UKCP18. The inclusion of atmospheric CO_2 trends in the likelihood weighting (an innovation in UKCP18) leads to a cooler GMST response. The 95% probability level becomes 0.34°C cooler, for example, when this constraint is added. Although subsequent inclusion of OHC does not alter the posterior substantially, this does not necessarily imply it is unimportant. This is because sequential application of constraints can be expected to lead to diminishing returns, especially when simulation errors are not independent, as is the case for the historical trends used here.

For surface air temperature in Southern England[10], successive application of constraints leads to a progressive reduction in the spread of the pdf. For example, the 90% prior range of 10.0°C is reduced to 8.4°C when the full set of constraints is applied. In common with UKCP09 (Sexton and Murphy, 2012), the relative impact of likelihood weighting is less for UK-scale variables than for GMST. In the case of summer precipitation change, the principal effect of the constraints is to reduce the estimated chance of an increase in precipitation, compared to the prior distribution. The change at the 95% probability level is reduced from 60% to 25%, for example. Application of the constraints also leads to a modest shift to greater drying, in the median response.

[10] In Fig. 2.6, one of the UK_GCM grid boxes of Harris et al. (2010) is used to define Southern England (see section 2.2, Stage 2), so the downscaling step in the methodology (section 2.2, Stage 3) is not applied.
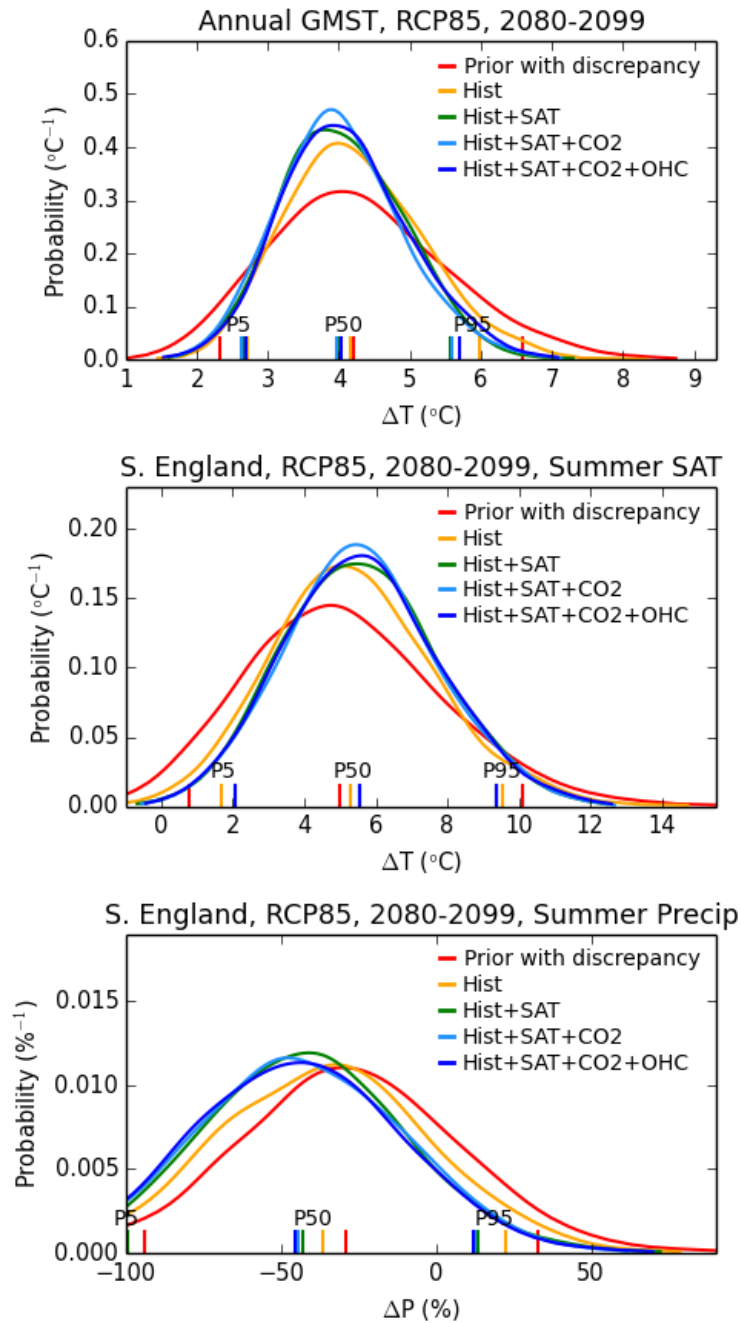
Figure 2.6 Dependence of the posterior pdfs for transient climate change on choice of observational constraints. All pdfs are for the 20-year mean response for 2081-2100 relative to 1981-2000, for the RCP8.5 scenario. The three variables are: annual GMST (upper panel), summer surface air temperature response for Southern England (middle panel), summer change in precipitation (%), also for Southern England. The red curve shows the prior (unweighted) distribution, and the others show the cumulative effects of adding constraints due to climatological spatial fields (orange), historical trends in surface air temperature (green), CO_2 concentration (light blue) and upper ocean heat content (dark blue). P5, P50 and P95 denote the 5, 50 and 95% probability levels of the various pdfs.

Checking the credibility of the probabilistic projections

For each of the twelve core UKCP18 variables (Fung et al., 2018), probabilistic projections were produced, and then checked for credibility. The main check (described in detail in Appendix C), involved comparing the probability distributions of projected change (for 2080-2099 relative to 1981-2000 under RCP8.5 emissions) against 349 changes from the global climate model simulations used in Strand 1. For relative humidity and near-surface wind speed, the tails of the probability distribution often showed a significant cumulative probability (F) of 15 or 20% for outcomes beyond the most extreme of the climate model responses, which could be traced to a failure in one of the key assumptions in the methodology. These variables were rejected. The occurrence of high values of F was substantially smaller for the other variables, so these were assessed as credible.

In Appendix C we also assess the representation of precipitation variability in the downscaling step (Stage 3 above). In UKCP09, a logarithmic transformation was applied to precipitation values prior to the statistical calculations, which was then inverted to obtain final results as percentage changes. This was done to avoid generation of negative values. However, we found that this approach led to the generation of unrealistic wet extremes in UKCP18, compared to the variability simulated in HadRM3-PPE. We therefore performed the UKCP18 calculations directly in units of percentage anomalies, which eliminated the unrealistic wet extremes at the cost of losing the automatic bounding of dry anomalies at -100%.

The final set of sampled precipitation realisations were therefore clipped at -100%. In addition, a second level of clipping was applied to all variables at the 1% and 99% probability levels of the relevant pdfs, in order to avoid provision of potentially unrealistic extremes.

Production of pdfs for different emissions scenarios

As explained in Stages 2a, b above, the generation of sampled climate change realisations requires specification of bias and residual terms associated with the timescaling procedure. For a given variable, these terms are calibrated by using global and regional metrics of climate change (assumed time-invariant) to predict the transient response of ESPPE members. This involves use of the simple climate model to predict the ESPPE GMST responses. Then, the bias and residual terms can be calibrated from the resulting errors. For the RCP2.6, 8.5 and A1B scenarios, ESPPE simulations are available for this purpose. However, this is not the case for the RCP4.5 and RCP6.0 scenarios, so direct calibration of timescaling errors is not possible.

Therefore, timescaling errors are estimated indirectly for these two scenarios, from errors obtained in the RCP8.5 case. We do this by assuming the bias term to be proportional to the response of GMST. The relationship is quantified by regressing the scaling error against the contemporaneous GMST response predicted by the simple model. A different relationship is obtained for each ESPPE member. Predicted GMST responses for RCP4.5 and RCP6.0 are then multiplied by these coefficients to estimate the time-varying bias term for these scenarios. The residual term (which accounts for interannual variability, as explained in Stage 2) is noisier than the bias term, hence it is difficult to justify expressing it as a function of GMST response. Therefore, it is simply assumed to be independent of scenario, and sampled as in the RCP8.5 case. The residual variances for the A1B and RCP8.5 scenarios are very similar, which provides some support for this choice.

A similar issue arises in the estimation of the time-dependent discrepancy terms for GMST, CO_2 and OHC (Stage 2a above). In this case, direct calibration is only possible for the RCP8.5 scenario, for which data from the CMIP5-ESM multi-model ensemble is available. We apply an analogous approach to estimate the discrepancy distributions for the other four scenarios. The mean discrepancy is assumed proportional to the predicted GMST, and regression coefficients are calculated for CMIP5-ESM members under the RCP8.5 scenario. Predictions of GMST for the other scenarios are then used to estimate the time-dependent mean values of discrepancy, while the variances of the discrepancy distributions are assumed to be identical to their RCP8.5 counterparts.

2.3. Examples of the probabilistic projections

In this section we provide a few examples demonstrating the nature and scope of the probabilistic projections. Further information is presented in section 5.1 (alongside results from Strands 2 and 3), and also in Lowe et al. (2018).

Given the extensive use of the UKCP09 probabilistic projections, it is instructive to compare these against their UKCP18 equivalents. This illustrates the impact of new information and improvements in methodology (sections 2.1 and 2.2), and allows stakeholders to assess whether studies based on UKCP09 should be updated. Figure 2.7 compares the two set of projections for GMST, and for winter surface air temperature and summer precipitation for the South East England administrative region. In order to provide a like-for-like comparison, the UKCP18 results are (in this particular case) calculated for the A1B emissions scenario, using the UKCP09 baseline of 1961-1990. They are shown as a set of overlapping 30-year average changes stepped 10 years at a time, consistent with the presentation of UKCP09.

The uncertainty ranges are considerable in both sets of projections, for all three variables. There is substantial overlap between the UKCP09 and UKCP18 distributions for all future periods, and the main characteristics of the changes are the same. Both show a monotonic[11] increase in the medians of their 30-year mean GMST and South East England winter temperature distributions, with the median warming over South East England being slightly smaller by the end of the 21st century. The median summer precipitation changes both show reductions that grow through the century, accompanied by uncertainty ranges wide enough to encompass the possibility of an increase.

---

[11] Note that at the annual time scale (e.g. Sexton and Harris (2015), Fig. 1), time series of individual realisations that comprise the probabilistic projections can show periods of 10-20 years during which GMST departs from a monotonic increase, due to the influence of internal variability. This is also true of the Strand 2 simulations (see Fig. 3.20).
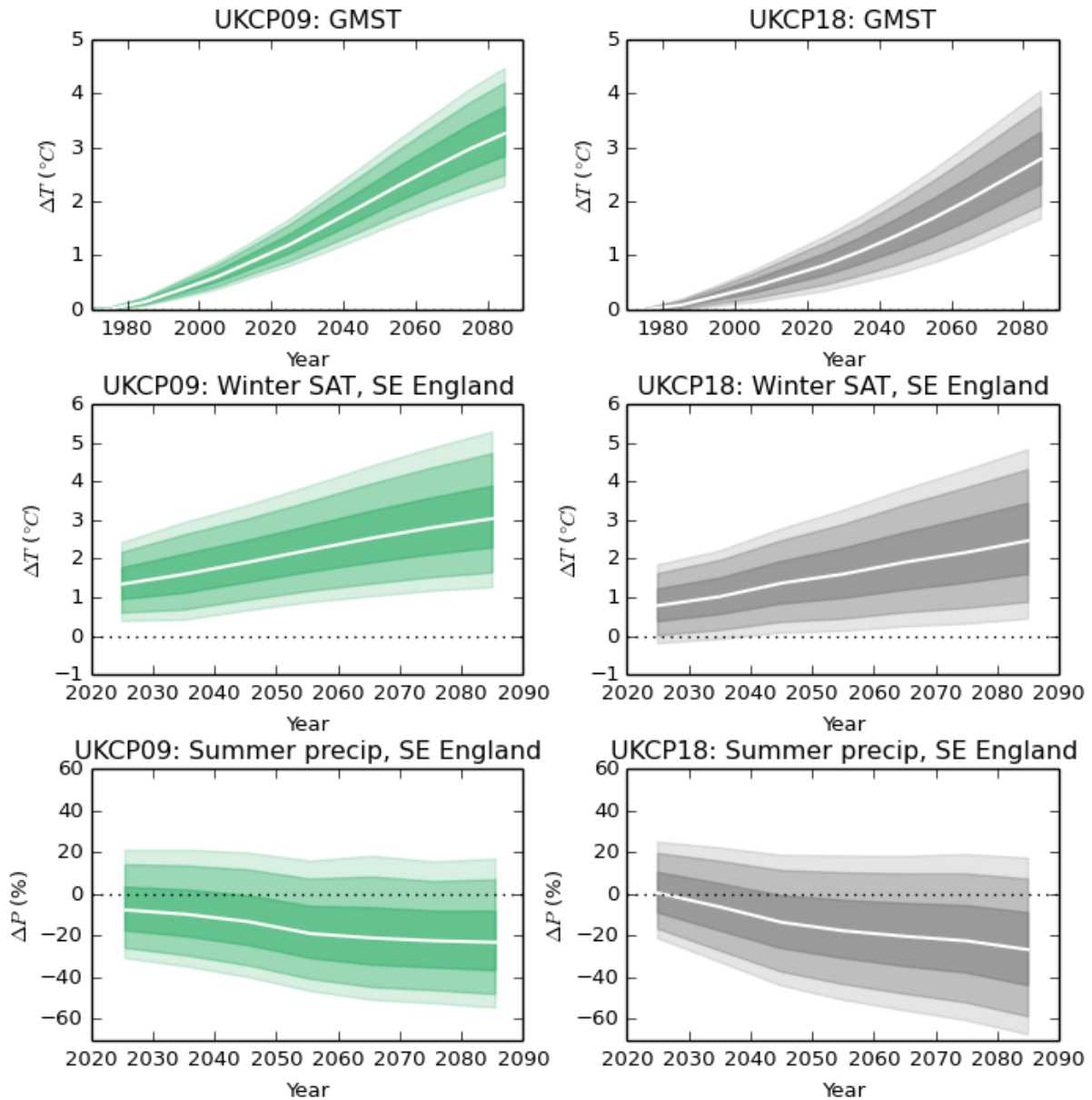
Figure 2.7. Probabilistic projections from UKCP09 (left) compared with those of UKCP18 (right), for the A1B emissions scenario. Top panels show GMST (°C), middle and lower panels show surface air temperature (°C) in winter and precipitation (%) in summer for the South-East England administrative region. The white line shows the median of the relevant probability distribution, and shading shows the 5, 10, 25, 75, 90 and 95% probability levels. Changes in 30-year averages are shown relative to the UKCP09 baseline of 1961-1990, for the seven overlapping future periods provided in UKCP09: 2011-40, 2021-50, .., 2071-00. In the case of GMST, results for earlier periods (1971-2000, 1981-2010, 1991-2020 and 2001-2030) are also shown.a

However, there are also some differences between the UKCP18 and UKCP09 results. The median GMST warming is somewhat lower in UKCP18, amounting to 2.9°C by 2071-2100, compared with 3.2°C for UKCP09. A major driver of the cooler median response in UKCP18 is the addition of a time-dependent component to the calculation of discrepancy. This provides a negative contribution of 0.2-0.4°C to GMST changes, throughout the 21st century (Fig. 2.4a, dark blue curve in top right panel). By 2071-2000, the upper end of the UKCP18 GMST distribution is ~0.5°C cooler in UKCP18. The additional reduction reflects the new constraint on CO_2 concentration (Fig. 2.6), which down-weights future outcomes with strong positive carbon cycle feedbacks (see Figure 3.7 in section 3.4).

The median South East England winter warming is also lower in UKCP18, reaching 2.7°C by 2071-2100, cf 3.0°C in UKCP09. This shift is apparent throughout the 21st century, with a median warming for 2010-2039 of 0.8°C in UKCP18 cf 1.3°C in UKCP09. Most of this shift is due to the influence of the cooler GMST distribution discussed above, the contemporaneous difference in the median GMST response amounting to 0.35°C. Also, South East England winter air temperature during 2010-2039 is projected to warm slightly more than GMST in UKCP09, but slightly less in UKCP18. As a result of these factors, UKCP18 shows a 10% probability for a small near-term cooling at the low end of its range, in contrast to UKCP09.

The median reduction in summer South East England precipitation is slightly smaller in UKCP18 during the first half of the 21st century, but reaches a similar level to UKCP09 by 2071-2100, of just over 20%. The 10-90% uncertainty range grows during the century in both products, but slightly more so at the dry end in UKCP18. By 2071-2100, the 10% probability level shows a drying of almost 60% in UKCP18, compared to ~50% in UKCP09.
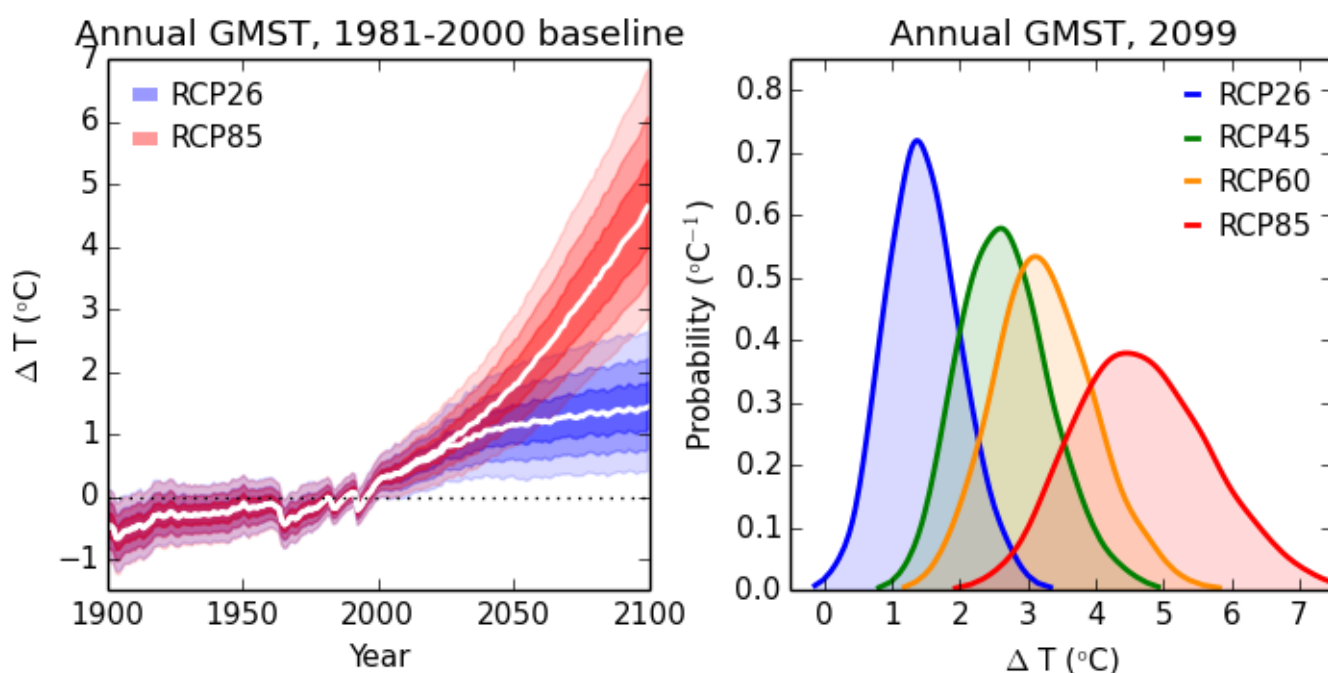

Figure 2.8. Comparison of probabilistic projections of annual GMST (°C) from Strand 1, for different emissions scenarios. Left panel (red shading) shows the 5, 10, 25, 75, 90 and 95% probability levels of the time-evolving distributions under historical changes in radiative forcing to 2005, and future responses to the RCP8.5 scenario from 2006-2100. Blue shading shows the same probability levels for the 21st century response to RCP2.6. The white lines show the medians of the relevant probability distributions. Anomalies are calculated relative to the 1981-2000 baseline. Right panel shows probability distributions of change for 2099, for the RCP2.6, 4.5, 6.0 and 8.5 scenarios.

In UKCP18, a key role of Strand 1 is to provide projections for a range of alternative emissions scenarios (see section 1.4). These include SRES A1B (for backwards-compatibility with UKCP09, see above), and RCP2.6, 4.5, 6.0 and 8.5. The new simulations in Strands 2 and 3 are available only for RCP8.5, with derived results for RCP2.6 obtained from the Strand 2 outputs using statistical methods (Gohar et al., 2018). The RCP scenarios sample a range of pathways for anthropogenic forcing (Moss et al., 2010). The numbers denote the approximate forcing in Wm^{-2} by 2100, relative to pre-industrial conditions. In RCP8.5, carbon emissions are assumed to rise substantially, approaching 30 GtC per year by 2100, whereas RCP2.6 assumes aggressive mitigation measures, leading to small, or potentially negative, anthropogenic carbon emissions by the last decades of the 21st century.

Figure 2.8 (right panel) shows probabilistic projections for GMST for 2099 (relative to 1981-2000) for each RCP scenario, as a basic benchmark of changes expected in each case. The A1B scenario (not shown in Fig. 2.8) gives very similar results to RCP6.0. Information on UK changes predicted by Strand 1, for each of the five available scenarios, is provided in Lowe et al. (2018). The different emissions profiles have a major impact on the ranges of projected change in GMST. Differences in response between RCP2.6 and RCP8.5 emerge from about 2030 onwards[12] (Fig. 2.8, left panel), and become substantial by 2050. At the end of the century, there is little overlap between the probability distributions for these two scenarios, despite the wide ranges in response in each case (particularly RCP8.5). Other pairs of scenarios show greater degrees of overlap, particularly between RCP4.5 and either 2.6 or 6.0, and 6.0 with either 4.5 or 8.5. In agreement with previous studies (e.g. Knutti et al., 2008; Hawkins and Sutton, 2009), the Strand 1 results indicate that uncertainties in emissions and climate response are comparable in their impacts on the range of plausible GMST outcomes during the coming century. For any specific emissions scenario, uncertainty in response leads to considerable uncertainty in the time at which any given threshold of GMST change might be exceeded (Joshi et al., 2011).

Table 2.2. Lower and upper limits of projected changes in global mean surface temperature for 2081-2100 relative to 1986-2005 for RCP scenarios, from AR5 and Strand 1 of UKCP18. The lower and upper limits correspond to 5% and 95% probability levels respectively. For each emissions scenario, the AR5 results are derived from CMIP5 simulations that all use the standard prescribed CO_2 concentration pathway. The UKCP18 results sample a range of alternative CO_2 pathways, obtained from emissions-driven HadCM3 and CMIP5 simulations using the Strand 1 methodology.

Emissions scenario	5-95% Probability range of change in GMST (°C) 2081-2100 relative to 1986-2005	
	IPCC AR5	UKCP18 Strand 1
RCP2.6	0.3 – 1.7	0.5 – 2.2
RCP4.5	1.1 – 2.6	1.4 – 3.3
RCP6.0	1.4 – 3.1	1.7 – 3.9
RCP8.5	2.6 – 4.8	2.6 – 5.5

[12] Although the CO_2 emissions pathway in RCP2.6 differs from that of RCP8.5 after 2005, emissions in RCP2.6 are assumed not to start reducing until after 2020, and CO_2 concentrations in the atmosphere stay close to those of RCP8.5 until the mid-2020s. Hence, there is a delay in the emergence of clear differences in GMST response between the two scenarios.

The GMST projections of Fig. 2.8 can be compared against uncertainty ranges assessed by AR5. For each RCP scenario, AR5 provided 5-95% probability ranges[13] of response for 2081-2100 relative to 1986-2005. These were derived from CMIP5 simulations using the standard RCP concentration pathway, assuming a Gaussian distribution of responses (Collins et al., 2013). These ranges are compared with corresponding UKCP18 results in Table 2.2. The lower ends of the UKCP18 5-95% ranges differ by 0.3°C or less from the AR5 results, while the upper limits are 0.5-0.8°C higher in UKCP18. The upper limit in Strand 1 is increased by the sampling of uncertainties in carbon cycle feedbacks, which are not accounted for in the AR5 results. In probabilistic projections made for RCP8.5 using the UKCP09 methodology, the upper limit in the emissions-driven projections reduced by 0.5°C when they were repeated as concentration-driven projections using the standard RCP8.5 pathway (Humphrey and Murphy, 2016).
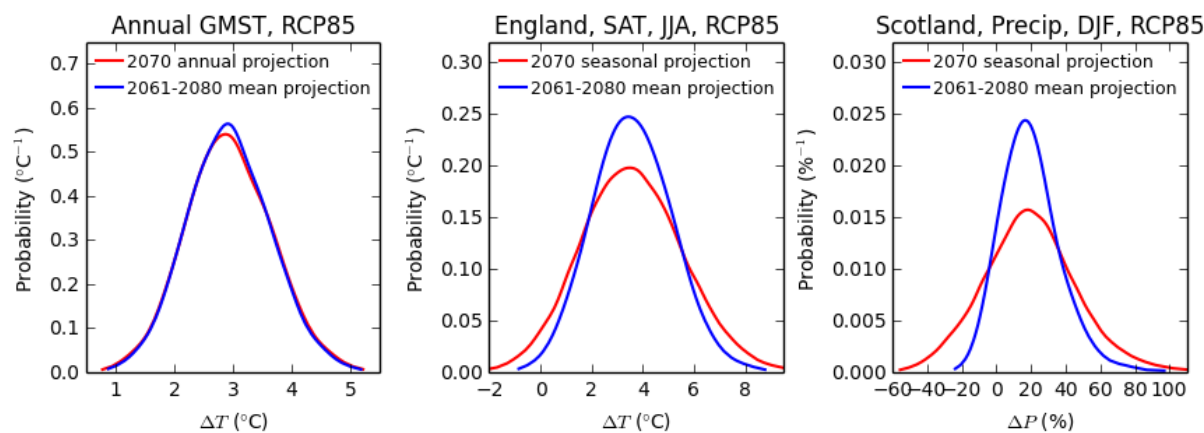

Figure 2.9. Probabilistic projections of annual GMST (°C, left), surface air temperature for England in summer (°C, middle) and precipitation for Scotland in winter (%, right), for 2061-2080 relative to 1981-2000 under the RCP8.5 emissions scenario. The curves show distributions of variability and change for the 20-year average (blue), and for all individual years within the future period (red).

As pointed out in section 2.1, the probabilistic projections are presented at the seasonal time scale, using a methodology based on Sexton and Harris (2015) to account for internal climate variability. This extends the UKCP09 information, in which probabilistic information was provided only for 30-year averages, to support analysis of changes in seasonal extremes. Users who need probabilistic information relating to long-term averages can obtain this by filtering or averaging the 3000 sampled realisations that form the output of Strand 1 (as done for Fig. 2.7, for example).

For RCP8.5, Figure 2.9 compares probability distributions of individual seasonal anomalies projected during 2061-2080 against distributions of 20-year mean changes for the same period. This illustrates the impacts of adding variability on 1-20 year time scales. For GMST (left panel), there is little difference between the two distributions, indicating that uncertainties in the global average response dominate those due to seasonal to decadal variability. For summer surface air temperature changes in England (middle panel), the range of seasonal responses is larger than for GMST. It is also dominated by the contribution from 20-year averages, although 1-20 year variability (larger at the regional scale[14]) does broaden the range more than it does for GMST. For Scotland precipitation in winter (right panel), uncertainties in the long-term average and 1-20 year variability both contribute substantially to the overall range in seasonal anomalies.

---

[13] In the specific case of global temperature projections, AR5 adjusted its quantitative language scale to interpret 5-95% probability ranges as likely rather than the usual interpretation of very likely, on the basis that the range of values of the transient climate response (TCR) found in CMIP5 models was consistent with the likely range assessed by AR5 for TCR. TCR is defined as the change in GMST at the time of doubling in a simulation in which CO_2 increases by 1% per year, and is an informative indicator of GMST change in future scenarios with steadily increasing radiative forcing. The AR5 assessment of TCR was based on multiple studies using partly independent lines of evidence, derived from a range of different climate model and observational datasets.

[14] This can be seen by comparing the annual spread of the two variables during 1981-2000 (Figs. 5.6a and 2.8 respectively), since the average forced response is by construction zero during the baseline period.

The relative importance of 1-20 year variability is larger at earlier lead times, when the forced changes, and their associated uncertainties, are smaller (Sexton and Harris, 2015). This is discussed further in section 5 (Figs. 5.6 and 5.7).
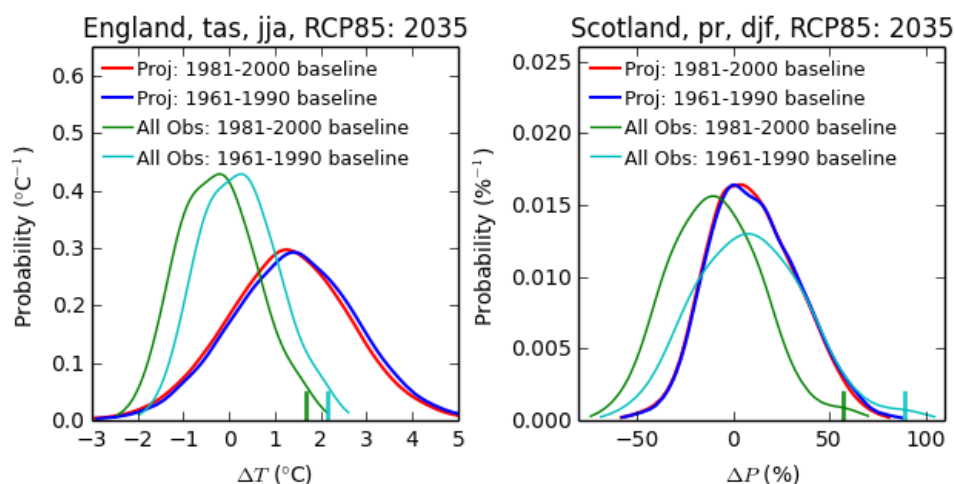

Figure 2.10. Effects of changing the baseline period. Red curves show probabilistic projections of the response to RCP8.5 emissions of surface air temperature (°C) for England in summer (left), and precipitation (%) for Scotland in winter (right), for seasonal anomalies during 2030-39. Anomalies are expressed relative to the standard UKCP18 baseline of 1981-2000 (red), and the UKCP09 baseline of 1961-1990 (blue). The green and light blue curves show distributions of seasonal observed anomalies during 1910-2017, relative to the 1981-2000 and 1961-1990 baselines respectively. The vertical bars show the warmest English summer (1976) and wettest Scottish winter (2015-16) in the historical record, relative to each of the baselines. Observations are taken from NCIC datasets (Table E.2).

Users of UKCP18 have the option of choosing alternative baselines (section 1.5). Figure 2.10 shows an example for seasonal projections during 2030-39, in which English summer surface air temperature anomalies (left) and Scottish winter precipitation anomalies (right) are expressed relative to the UKCP09 and UKCP18 baselines of 1961-1990 (blue) and 1981-2000 (red). Interestingly, there is relatively little difference between the two curves, for either variable. The distribution of summer surface air temperature anomalies is shifted ~0.1°C warmer when the earlier baseline is used, while the winter precipitation anomalies become slightly wetter.

However, distributions of observed anomalies (light blue and green curves) differ more, when the baseline is changed. For example, the anomaly associated with the hottest English summer on record (1976) increases from 1.8°C to 2.1°C when based relative to 1961-1990, while the anomaly associated with the wettest Scottish winter (2015-16) increases substantially, from just over 50% to just under 90%.

This difference probably occurs because the observed record is only one realisation of climate, whereas the probabilistic projections include 3000 realisations. The small shifts seen in the Strand 1 distributions reflect modest signals of forced climate change in their time series between 1961 and 2000, distinguished from the effects of annual to decadal variability by considering a large sample of outcomes. By contrast, the observed record contains only one sample of low frequency variability, and therefore shows larger changes between the two baseline periods. For example, total winter precipitation in Scotland increased by about 50mm between 1960 and 2000, but some of this change is likely to have been driven by an increasing positive trend in the NAO during 1961-1990 (Deser et al., 2017), in which internal climate variability may have played an important role (e.g. Selten et al., 2004).

The different effects of changing the baseline between observations and the projections are worth noting. They reveal the importance of understanding that contrasts between recent observed climatological averages for different periods reflect an uncertain combination of low frequency variability and anthropogenic climate changes (e.g. Karoly and Stott, 2006). The smaller shifts seen in the Strand 1 probability distributions are the result of considering many realisations of how baseline climate might have changed between 1961-1990 and 1981-2000, and are useful in placing recent observed trends in a broader uncertainty context.

3. Strand 2: A new set of global climate simulations from 1900 – 2100

3.1. Overview

Strand 2 includes GC3.05-PPE, a new 15-member PPE of global climate simulations from 1900-2100, using the RCP8.5 scenario beyond 2005. These simulations are provided to address several of the science and user drivers listed in section 1.2. In particular, GC3.05-PPE utilises the latest developments in UK modelling capability (driver 5) and provides a multi-variable dataset for impacts analysis with full spatial and temporal coherence (driver 2). It also supports development of storylines relating to future climate variability and extremes on a broad range of time scales, and provides new information on potential imported risks from other international regions (drivers 3, 4 and 6). These global simulations have also been used to drive new 12km regional climate model simulations for the UK and Europe (Strand 3, described in section 4).

The GC3.05-PPE simulations are augmented by a subset of 13 CMIP5 models (CMIP5-13), in order to add sampling of diversity in model structure to the diversity arising from parametric process uncertainties in GC3.05-PPE. This creates a combined set of 28 projections. The CMIP5-13 members are selected using a mixture of quantitative and qualitative assessment criteria, with the aim of removing implausible simulations to identify a surviving subset of credible projections (McSweeney et al., 2015). Similar principles are used in the design of GC3.05-PPE. CMIP5 models are also used to represent structural uncertainties in Strand 1, however the methodology there is necessarily different. This is because (a) the probabilistic presentation of Strand 1 requires use of a formal mathematical framework containing a specific discrepancy term to quantify structural uncertainties (section 2.2), and (b) results from CMIP5 earth system models driven by carbon emissions are required in Strand 1, whereas concentration-driven simulations from physical ocean-atmosphere CMIP5 models are used in Strand 2 (see section 1.4).

3.2. The new climate model underpinning GC3.05-PPE

The basis for the PPE is the GC3.05 coupled ocean-atmosphere model (see discussion in section 1.2). It incorporates the main improvements added to GC3.0 to produce GC3.1, the UK model submitted to CMIP6 (Williams et al., 2018). The atmosphere and land[15] components are configured on a regular latitude-longitude grid at N216[16] resolution, which gives a horizontal grid spacing of approximately 60km at mid-latitudes. There are 85 vertical levels (L85), 30 of which are in or above the stratosphere. This provides improved resolution of middle atmosphere dynamics compared with most CMIP5 models, including the Met Office submissions HadGEM2-AO and HadGEM2-ES. Compared to HadGEM2, the GC3.05 atmosphere includes a major revision to its dynamical core (ENDGame, Wood et al., 2014), a new cloud scheme (PC2, Wilson et al., 2008), and numerous smaller changes to other parameterisations of sub-grid scale processes (Williams et al. (2018) and references therein). Two stochastic physics schemes are also included (Sanchez et al., 2016), representing uncertainty arising from aspects of unresolved sub-grid scale variability. The Stochastic Kinetic Energy Backscatter scheme version 2 (SKEB2) represents the backscatter to the resolved flow of small-scale kinetic energy lost via numerical diffusion, and the Stochastic Perturbation of Tendencies scheme (SPT) stochastically scales the outputs of physical parameterisations to represent variability about their best-estimate outputs.

The ocean and sea-ice models (NEMO and CICE, Hewitt et al., 2011) are also replacements for their HadGEM2-AO counterparts. The ocean component uses a tripolar eddy-permitting grid with 75 levels (ORCA025L75). The horizontal resolution is ¼° at the equator, where the grid is at its coarsest. This yields an improved simulation of the Gulf Stream extension compared to HadGEM2-AO, reducing a cold bias in North Atlantic sea surface temperature (SST) and hence ameliorating an important source of error in the simulation of climate variability over Europe (Scaife et al., 2011).

Williams et al. (2018) compare GC3.0 and GC3.1 to HadGEM2-AO, finding that the new models achieve improved simulation of the atmospheric temperature structure, surface and top of atmosphere radiative fluxes, synoptic variability at mid-latitudes, ENSO and tropical cyclones. However, several long-standing systematic biases remain, including excessive global average precipitation, cool and warm sea surface temperature biases in the northern hemisphere and Southern Ocean respectively, and insufficient frequency of blocking anticyclones in the northern hemisphere. The latter is an important factor in relation to the simulation of cold extremes in UK winters. Worldwide, SST biases are generally reduced in GC3.1 compared to HadGEM2-AO, although the warm bias in the Southern Ocean is slightly larger.

[15] The land module is the Joint UK Land Environment Simulator (JULES, Best et al., 2011). In GC3.05 JULES is run using prescribed vegetation distributions, as the model does not include an interactive carbon cycle.

[16] Nx denotes a regular grid defined at 2x longitudinal points and 1.5x+1 latitudinal points, giving an approximately isotropic spacing at mid-latitudes. The coupled ocean-atmosphere GC3.05-PPE projections in Strand 2 use N216, while the preliminary simulations used to identify potential PPE members use N96, at which the mid-latitude spacing is approximately 135km.

Scaife et al (2014) found important improvements in the predictability of the winter NAO when the Met Office seasonal prediction system was upgraded to use a version of the Met Office coupled model configured at N216L85 and ORCA025L75. Senior et al (2016) used HadGEM3-GC2, another recent coupled model version, to investigate the impact on climate change simulations of increasing horizontal resolution from N96 to N216 in the atmosphere and from 1° to ¼° in the ocean. They found minimal impact on projected changes in long-term averages at continental or larger scales. However, there were some important regional impacts, including a larger increase in the frequency of the most intense winter storms at N216 over the North Atlantic and Europe.

The aerosol module in GC3.05 is GLOMAP-mode (Mann et al., 2010). This simulates the transport, growth, nucleation, coagulation, cloud-cycling and removal of sulphate, sea salt, dust, black carbon and particulate organic species, in response to primary natural and anthropogenic emissions. Direct radiative effects and those arising from aerosol-cloud interactions are accounted for, and the scheme transports both particle concentration and mass in seven multi-component modes, avoiding the simplifying assumption of a fixed particle size distribution made in most previous climate models. During the development cycle for GC3.0 (Williams et al., 2018), the total anthropogenic radiative forcing simulated for the year 2000 relative to pre-industrial conditions was found to be negative, failing a basic model acceptance criterion. This was caused by excessively strong cooling due to anthropogenic aerosols. A limited set of physical improvements was then included in the GC3.1 configuration (Walters et al., 2017), partially addressing this issue. The main change was to include a parameterisation of the effect of droplet number (and hence aerosol loading) on the spectral dispersion for the cloud droplet size distribution (Liu et al., 2008). This parameterisation, and a change to the refractive index of black carbon (Bond and Bergstrom, 2006), were included in GC3.05. Additional minor changes applied in GC3.1 (see Appendix D) were not available in time to be included in GC3.05. The GC3.05-PPE members all simulate a positive change in low-pass filtered values of total anthropogenic radiative forcing at year 2000 (relative to 1900), although several members simulate a strong aerosol cooling during the second half of the 20th century. The GC3.05-PPE simulations of historical climate change are discussed further in section 3.4.

3.3. Design of the strand 2 projections

In this section we describe the design of GC3.05-PPE and the selection of additional CMIP5-13 projections. Section 3.4 provides evaluation of the historical components of the 28 coupled ocean-atmosphere simulations, considering selected metrics of global and regional performance. The latter are focused on the UK and North Atlantic/Europe sector. Section 3.5 covers global aspects of the projected future changes given by the Strand 2 set. Projections for the UK are described in section 5, alongside those from Strands 1 and 3.

The GC3.05 model is computationally expensive. The choice of configuration reflects the aim in Strand 2 of producing a limited number of high-resolution future projections, underpinned by the best possible representation of regional climate variability. In order to maximise its utility, the design methodology for GC3.05-PPE is based on the twin principles of plausibility and diversity (Karmalkar et al., 2018).
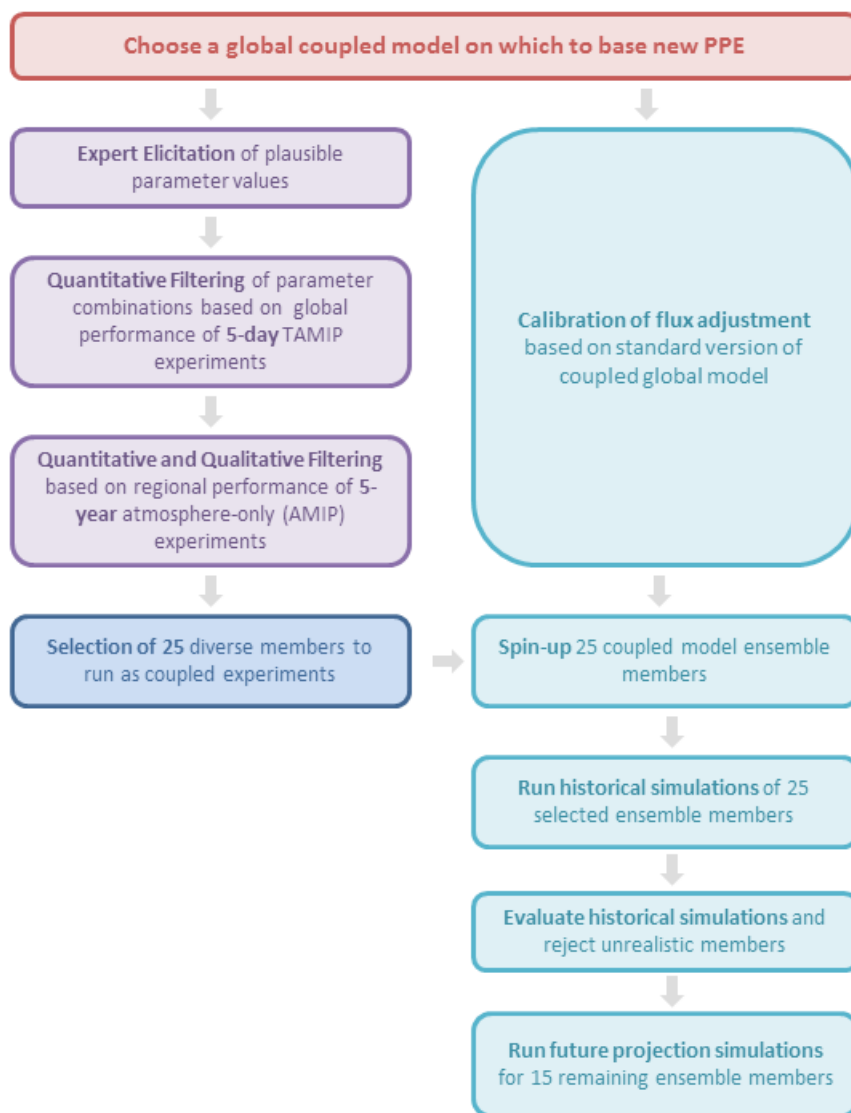
Figure 3.1. Key steps in the production of the GC3.05 perturbed parameter ensemble (GC3.05-PPE) of global climate projections in Strand 2 of UKCP18. The choice of model configuration (red box) is described in section 3.2. Light purple boxes and dark blue box (section 3.3) denote steps required to select promising parameter combinations in the atmosphere, land and aerosol components for coupling to the GC3.05 ocean model. Light blue boxes denote preparation, evaluation and production of climate change simulations from 1900-2100 (sections 3.3-3.5).

Plausibility is assessed via three successive stages of model evaluation, involving five-day weather prediction tests followed by five year climate simulations carried out with the atmosphere model, and finally multidecadal coupled ocean atmosphere simulations (Fig. 3.1). The goal is to ensure that every GC3.05-PPE member performs to a good overall standard, so that available HPC resources are used effectively.

Diversity refers to the spread of future climate responses and historical climates simulated across the ensemble. Here, the aim is to ensure that the broadest possible range of future outcomes is sampled, subject to the scientific constraint of ensuring plausibility in the selected members, and the technological constraint on ensemble size imposed by computational expense (see above). This is to ensure that process uncertainties are represented robustly in GC3.05-PPE, thus maximising its ability to provide a range of alternative future storylines. Diversity is addressed by first identifying a pool of model variants likely to provide good simulations of historical climate, and then combining this information with results from using short idealised climate change simulations to pick a subset of these (3.3d below).

Our evaluation strategy follows a seamless model assessment philosophy (Hurrell et al., 2009), in which information from simulations of a few days (started from observed initial conditions) is combined with information from multiannual climate simulations in order to identify and minimise model biases common to all time scales. Seamless assessment plays an important role in the development strategy for the Met Office Unified Model (e.g. Martin et al., 2010). The methodology below is developed from the work of Sexton et al (2018a) and Karmalkar et al (2018), who showed that this approach can be used to build a PPE. This is because strong relationships are found between errors on weather and climate time scales (Rodwell and Palmer, 2007), which allows efficient screening of the model parameter space via the short simulations, and also strengthens the evidence base for accepting or rejecting potential PPE variants.

Development of GC3.05-PPE

Following the choice of GC3.05 as its parent model, design of the PPE involves several stages of development (Fig. 3.1). These start with definition of the parameter space from which plausible model variants may be drawn, followed by filtering of potential parameter combinations in the three stages outlined above. The initial 5-day pseudo-weather forecasts and 5-year climate simulations both involve running the atmosphere model using prescribed SSTs. The filtering involves quantitative metrics of global performance augmented (in the latter case) by qualitative assessment of performance in the North Atlantic/Europe region. Twenty-five potential members are then spun-up as coupled model variants, followed by production of historical climate change simulations. Assessment of these provides a final stage of filtering, leading to continuation of 15 members out to 2100 using the RCP8.5 scenario. These three stages are described below.

a. Expert elicitation of GC3.05 parameter space

The first step is to identify a set of model parameters to perturb. Design choices include which earth system components to consider, which areas of model physics, dynamics or chemistry (usually schemes parameterising sub-grid scale processes) to perturb within a given component, and which parameters to select in each perturbed scheme. We chose the atmosphere, land surface and aerosol components for perturbation. The ocean component was not perturbed. This was due to timing constraints and the lack of an established method for extending the use of seamless assessment ideas to coupled ocean–atmosphere simulations. In earlier PPEs using the HadCM3 model, perturbing ocean parameters was found to contribute little (compared to atmosphere, land and aerosol perturbations) to the uncertainty in projections of GMST (Collins et al., 2011), but was capable of producing a contribution to spread in regional changes beyond that attributable to internal climate variability (Brierley et al., 2010).

The selection of schemes and parameters followed the procedure described in detail by Sexton et al. (2018a), and was based on guidance obtained from model development experts. It was decided to perturb parameters in schemes representing convection, boundary layer, gravity wave drag, and cloud radiative and microphysical properties in the atmosphere model, as well as parameters in the aerosol and land surface modules. The atmospheric dynamics and radiative transfer codes were not perturbed, although the simulation of dynamical transports and radiative fluxes were affected by other perturbations, for example by cloud parameters in the latter case.

Within each selected parameterisation, experts nominated a subset of parameters controlling a broad set of physical processes. This was done to avoid favouring a priori specific sub-processes within a given scheme. For example, the convection parameters cover entrainment, detrainment, radiative effects in shallow, mid-level and deep convection, and processes determining the vertical profile of cloud condensate. Figure 3.2 shows the expert distributions provided for the 47 selected parameters, and Appendix D (Table D.1) gives details of their roles and effects. The distributions describe the prior parameter space, explored using atmosphere-only simulations to identify 25 perturbed variants suitable for consideration in coupled ocean-atmosphere projections. This work is described in detail by Sexton et al. (2018b), and summarised in the following sub-sections (b), (c) and (d).
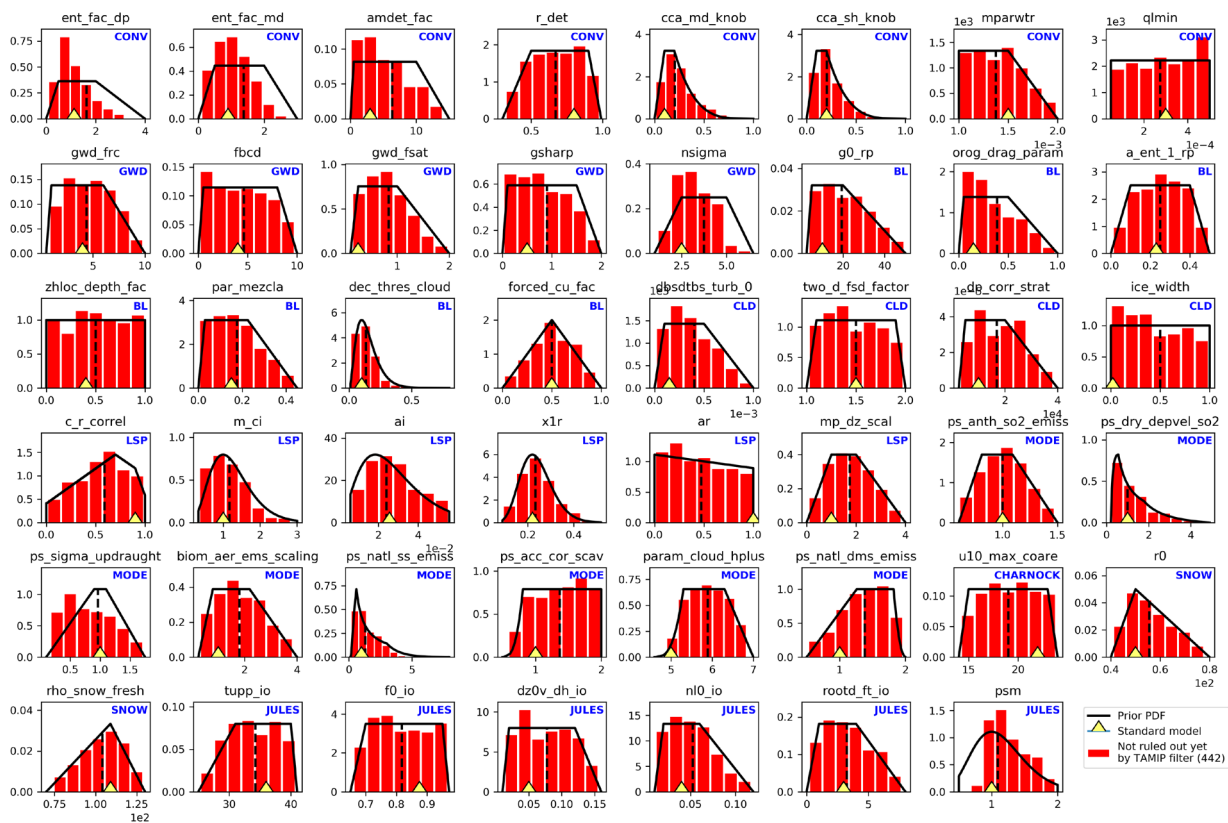

Figure 3.2. Prior probability distributions (black) provided by modelling experts for the 47 parameters perturbed in the GC3.05-PPE. Blue lettering denotes the parameterisation scheme or model component to which each parameter belongs (convection, CONV; gravity wave drag, GWD; boundary layer, BL; cloud, CLD, large-scale precipitation, LSP; aerosol, MODE; land surface, JULES). Dashed lines indicate median values and diamonds the values used in the standard, unperturbed model variant. Table D.1 in Appendix D describes the roles of the parameters and effects of perturbing them. Red histograms show the relative frequency of sampling in the 442 members surviving the TAMIP assessment, following consideration of the quantitative assessment metrics shown in Fig. 3.3 (top panel). The histograms show relative frequencies in seven bins of equal width across the range of plausible values for each parameter.

b. Filtering of parameter space using retrospective weather forecasts

The initial evaluation of combinations of parameter settings was carried out using five-day hindcasts (retrospective weather forecasts started from historical analyses of observations). A key assumption is that the predicted large-scale circulation will remain reasonably close to observations on this time scale, in which case hindcast biases should, to a substantial degree, reflect the growth of "fast" physics errors traceable to the parameterisation of sub-grid scale processes (Phillips et al., 2004). These short-term errors should be related to biases in longer term climate simulations, but can also provide complementary information on model credibility. This is because incipient error growth can be easier to relate to specific physical processes (Rodwell and Palmer, 2007), whereas emergent biases in longer simulations typically reflect the results of physical and dynamical interactions between multiple processes.

Hindcasts

We used an approach based on the Transpose-Atmospheric Model Intercomparison Project Phase II (TAMIP) experiment (Williams et al., 2013). In TAMIP, the primary state variables of atmosphere models are initialised using analyses of observations and run for five days, using SSTs and sea-ice extents also prescribed from observational analyses. The analyses are provided by the European Centre for Medium-Range Weather Forecasts (ECMWF). The full TAMIP protocol uses 64 start dates covering the period October 2008 to August 2009. We used 16 of these, reducing the number of start dates in order to support better sampling of parameter space within the available HPC resource. The dates[17] were chosen to sample the diurnal and seasonal cycles during the relevant period.

Studies of the relationship between weather forecast errors and climate simulation biases have used forecast time scales ranging from a few timesteps ahead (e.g. Rodwell and Palmer, 2007) to 1-2 weeks ahead (e.g. Wan et al., 2014). The use of tendencies during the initial forecast timesteps was not possible in our case, as this requires use of a data assimilation cycle to achieve fully balanced initial states, rather than the simpler TAMIP approach used here. However, strong relationships with climate biases have been found using day-2 or day-5 forecast errors in TAMIP experiments using CMIP5 models (Ma et al., 2014). Here we use day-5 errors following Sexton et al. (2018a). They found that error signals at this range were stronger, and therefore gave clearer relationships with climate biases, in a recent PPE study using the GA4 release of the Unified Model (Walters et al., 2014).

The TAMIP hindcasts, and also the 5-year climate simulations in (c) below, used an N96L85 configuration of the GC3.05 atmosphere. This allowed production of a 2800 member TAMIP PPE, an order of magnitude larger than would have been possible using the N216 version. Importantly, this ensemble was large enough to support use of a statistical emulator (as in Strand 1) to predict performance for untested parameter combinations.

Use of the N96 configuration was based on previous model development experience, which suggested that this would provide a reliable guide to performance at N216 resolution. Following the selection of 25 variants for coupled simulations, the five-year simulations of (c) below were repeated using the N216L85 configuration. The results confirmed the above "traceability" assumption, showing close relationships between model biases at the two resolutions accompanied (typically) by modest improvements at N216.

All parameters were perturbed in each TAMIP member, in order to investigate interactions between processes as well as the individual effects of each parameter. The TAMIP PPE was built up in three successive groups of ensemble members ("waves"). See Sexton et al. (2018b) for a detailed description of the experiment design. A key aim was to sample parameter space consistently with the expert-prior distributions, which was essentially[18] achieved. However, the choice of perturbed variants in successive waves was also influenced by accumulated experience. This included instances of model crashes due to numerical instabilities (waves 2 and 3 were designed to minimise the risk of these), and use of an emulator to identify variants (for wave 3) more likely to give acceptable performance.

---

[17] 15/10/08, 00Z; 21/10/08, 06Z; 27/10/08, 12Z; 31/10/08, 06Z; 15/01/09, 00Z; 27/01/09, 12Z; 31/01/09, 06Z; 02/02/09, 18Z; 15/04/09, 00Z; 21/04/09, 06Z; 27/04/09, 12Z; 03/05/09, 18Z; 15/07/09, 00Z; 21/07/09, 06Z; 27/07/09, 12Z; 02/08/09, 18Z.

[18] One exception was *ps_sigma_updraught*. This is a scaling parameter that defines a distribution of updraught velocities, required to determine the activation of aerosols to produce cloud droplets. In the first wave of 680 TAMIP simulations the range of values sampled for this parameter was wrong, resulting in only low values being considered. In subsequent waves this error was corrected. The net result achieved representation of the full range in the 2800 members, but with oversampling of the low end compared with the prior distribution of Fig. 3.2.

Assessment of hindcasts

TAMIP performance was assessed using a broad range of variables commonly evaluated in longer term simulations. These cover circulation, temperature, water cycle, radiative and cloud metrics (Fig. 3.3, top panel). For each ensemble member and assessment variable, the averaged day-5 difference between hindcast and observations was calculated over the 16 TAMIP start dates. Most assessment metrics consisted of single-level variables presented as a global spatial field. For these, mean-square errors (MSE) were calculated for six (overlapping) large-scale regions, consisting of: 60°S-60°N (global land and ocean domain, excluding polar regions), 30°S-30°N (tropics, land only), 30°S-30°N (tropics, ocean only), 30-90°N (northern hemisphere extra-tropics, land only), 30-90°N (northern hemisphere extra-tropics, ocean only), 30-90°S (southern hemisphere extra-tropics, ocean only). For variables defined as latitude-pressure distributions of zonal averages (temperature, specific and relative humidity, eastward and northward wind), global MSE values were calculated.

The MSE values were divided by a normalising factor consisting of the mean-square-error for the standard model variant (MSE-STD), plus terms representing observational uncertainty[19] (σ^2_{obs}) and uncertainty associated with hindcast values for a given PPE member (σ^2_{var}). In this factor, MSE-STD was used as a first-order estimate of the irreducible (structural) component of model error[20], while the sum of σ^2_{obs} and σ^2_{var} represent uncertainty in the measured MSE values as a true representation of model error. This normalisation allows us to express the MSE values relative to an approximate best attainable value. For variables where two datasets of verifying observations were available (Table E.2, Appendix E), one was used to calculate hindcast MSE values, and the other to provide a basic estimate of observational uncertainty. For variables where only one verifying dataset was available, observational uncertainty was estimated as a fraction of MSE-STD (20% in the TAMIP verification, and 10% in corresponding AMIP calculations in 3.3c below), guided by cases where two datasets could be used. The σ^2_{var} term was calculated from a 32-member ensemble of alternative TAMIP hindcasts for the standard model variant. In this ensemble, the growth of small initial differences, introduced by different realisations of the stochastic parameterisation schemes in the model, acted as a proxy for the component of weather prediction uncertainty due to initial state errors.

The normalised values (nMSE) provided a relative, dimensionless measure of skill suitable for identification of implausible members. The grey shading in Fig 3.3 shows the range of values obtained for each assessment variable. The red curve shows scores for a typical ensemble member that was not ruled out, illustrating that relative performance varies significantly across the assessment metrics. This is because the net effect of a given set of parameter perturbations tends to reduce some aspects of model bias while worsening others.

Rejection of implausible members was determined using error thresholds. The main[21] criterion was a "hard" limit (Fig 3.3, black line): any PPE member was excluded in which nMSE exceeded 4.5, for any metric. Since there is no agreed objective method of choosing model performance thresholds, this choice involved subjective judgement. The value of 4.5 was selected to eliminate PPE members clearly inferior to the unperturbed member, while retaining sufficient members to sample parameter space, and potential diversity in simulated climate changes, as broadly as possible.

---

[19] Observational uncertainty (e.g. Morice et al., 2012) can result from several sources, including measurement errors, errors of representivity (for example due to incomplete spatial coverage) and structural errors related to the chosen methodology (for example in how observations are converted into gridded datasets).

[20] This choice is based on the assumption that the standard, unperturbed model variant will provide results that are close to the best hindcasts that could be found in the entire parameter space of GC3.05 (Fig. 3.2), if resources existed to sample it thoroughly. In practice, this is likely to be the case for most but not all variables (Karmalkar et al., 2018).

[21] A second criterion was a "soft" limit (see Sexton et al., 2018b), set lower than the hard value, in which model variants were excluded if they exceeded the soft threshold for several variables. In practice, however, the hard limit was the main determinant of acceptance or rejection.

Applying the threshold left 442 surviving members. Circulation, water cycle and radiative variables all played important roles in the selection, while relatively few were ruled out based on cloud performance. While the TAMIP filtering excluded the majority of the 2800 members, the remainder sampled the full range of the prior distributions for most parameters (Fig. 3.2). This demonstrated retention of a high level of process diversity for consideration in longer simulations.
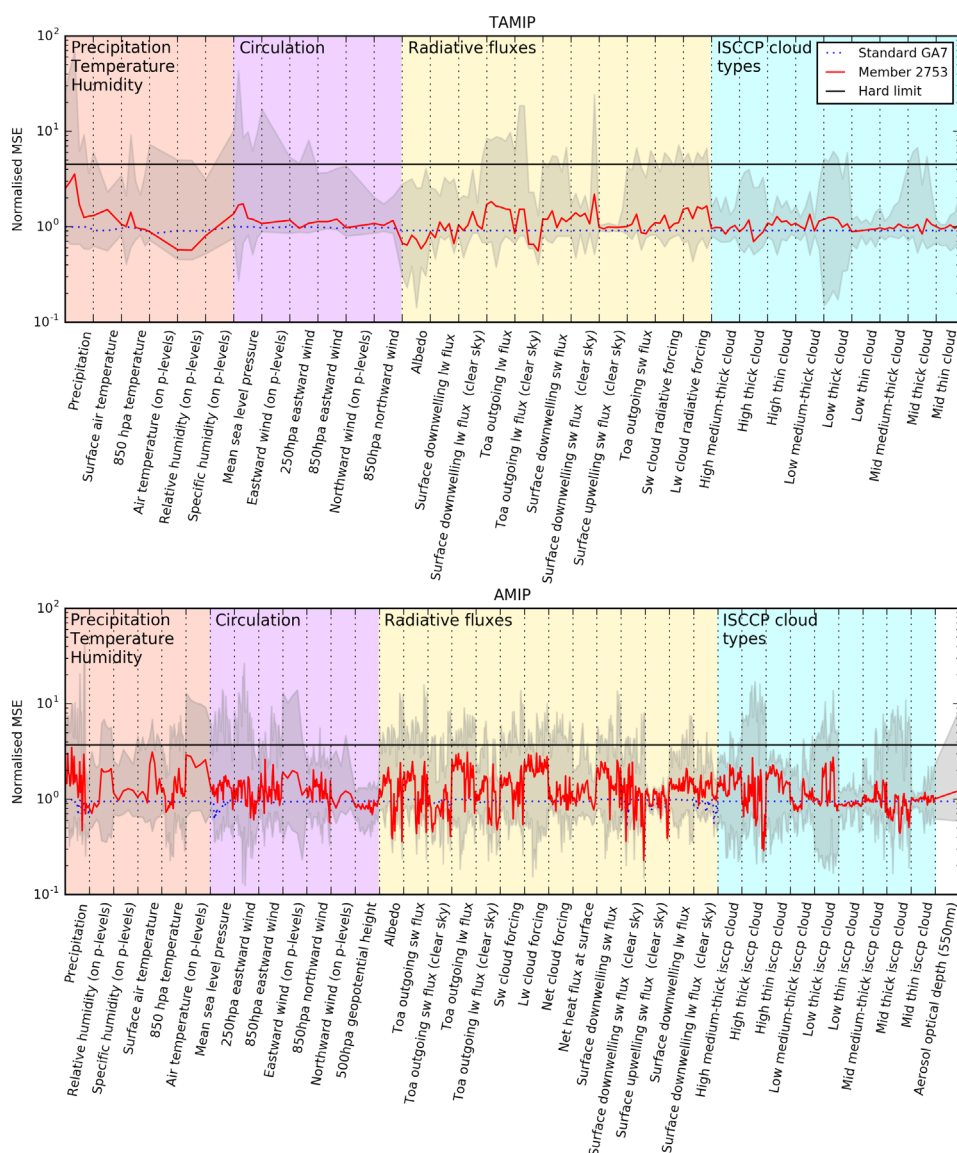

Figure 3.3. Relative performance of PPE members created from GA7 (Walters et al., 2017), the atmospheric component of GC3.05. The metric, nMSE, is mean-squared error calculated relative to (MSE-STD + σ^2_{obs} + σ^2_{var}), the sum of the MSE for the standard model variant, observational uncertainty and uncertainty associated with the MSE for a given PPE member (details in text). Values are shown for each of the variables used to filter perturbed parameter ensemble members in 5-day hindcasts (TAMIP, top), and 5-year climate simulations (AMIP, bottom), carried out using the atmosphere model driven by prescribed sea surface temperatures. For TAMIP, errors for each member are calculated from the average bias across 16 hindcast start dates covering the annual cycle. For AMIP, nMSE values are calculated separately for each season, giving four times as many entries per variable. For variables constituted as latitude-longitude spatial fields, values of nMSE are provided for each of the large-scale regions defined in the text, giving six separate ensemble distributions of values between each pair of grey lines for TAMIP, and 24 for AMIP. For variables defined as latitude-height fields of zonal averages (temperature, specific and relative humidity, eastward and northward wind on pressure levels), nMSE values for the whole globe are calculated. For aerosol optical depth (only calculated in the AMIP case), a single nMSE value is calculated from errors with respect to observations of optical depth at wavelength 550nm, at a set of specific site locations (see Holben et al., 2001). For a given variable, region and period, grey shading[22] shows the full range of scores across the 2800 TAMIP members and 557 AMIP members. Black lines denote the hard limits used to identify unacceptable members (see text). Red lines show the variation in scores by variable, for a typical ensemble member. The dotted blue line shows nMSE for the standard model variant, which is always a little smaller than 1.0 due to the σ^2_{obs} and σ^2_{var} terms.

[22] For high medium-thick cloud, the grey shaded range is incorrect in the AMIP case, because the wrong simulated cloud-type was compared against the observations. Correction of this error does not change the overall effect of the AMIP filtering, because model variants that failed on this metric also failed on other metrics.

c. Filtering of parameter space using five-year climate simulations

The five year[23] atmospheric simulations were similar in design to the Atmospheric Model Intercomparison (AMIP) experiment (Gates et al., 1999), using time-evolving SSTs and sea-ice extents prescribed from high-resolution HadISST2 analyses of observations (Titchner and Rayner, 2014). We refer to these as AMIP experiments, noting that we used a different and shorter simulation period (2005-2009) to the standard AMIP protocol. This allowed us to consider a recent period and increase the size of the PPE available for assessment. Contemporaneous[24] greenhouse gas concentrations and aerosol emissions were taken from CMIP5 datasets.

The AMIP PPE consisted of 557 members. These included the 442 variants that passed the TAMIP filtering, plus 115 variants added to augment the pool of potential coupled PPE members. The first 100 of these were predicted to perform well based on emulator results. The final 15 members were added to increase the sampling diversity of climate change forcing and response, by including more candidates with weaker climate feedbacks and aerosol forcing than the bulk of the ensemble (see sub-section (d)). The 15 GC3.05-PPE members selected for Strand 2 used 13 atmosphere variants taken from the original 442 members, plus two from the additional 115 AMIP members.

The 557 members were filtered in two steps. The first involved quantitative assessment using nMSE values (Fig 3.3, lower panel), carried out with a similar approach to the TAMIP evaluation. The same variables were assessed, based on 5-year mean error fields in the AMIP case. A measure of aerosol performance (optical depth at wavelength 550nm) was also included. For all variables bar aerosol optical depth, values of nMSE were calculated separately for each season of the year (as well as for different regions in the case of single-level spatial fields). Uncertainty in member-specific MSE values was estimated by running a 32-member ensemble of 5-year integrations using the standard model variant, distinguished by alternative realisations of internal variability created by the stochastic parameterisation outputs.

In the AMIP case, the basic hard threshold (see above) was reduced from 4.5 to 3.7[25]. In similar calculations applied to CMIP5 models, Karmalkar et al. (2018) found nMSE values ranging from ~2-10 between the best and worst performers, thus our choice provides a performance threshold typical of the spread of outcomes in CMIP5.

Model crashes ruled out 39 AMIP members, and the quantitative filtering removed a further 469 members, demonstrating the importance of assessing the credibility of multiannual climate averages in addition to performance in short-range weather forecasts. Variables in all categories played a significant role in determining the selection, with cloud errors contributing to a greater degree than at the TAMIP stage.

---

[23] Ideally, longer simulations would have been preferred, in order to achieve better sampling of the effects of internal atmospheric variability in the AMIP runs. This was prevented by limitations in HPC resources. However, we found (as reported in section 3.4b) that parallel AMIP simulations at N216 and N96 atmosphere resolution gave consistent results for model biases, suggesting that the five-year simulations were long enough to give a reasonable estimate of climatological errors in the 25 members selected for use in coupled simulations.

[24] A small error was found in the specification of time-varying CO_2 concentrations, which were shifted by 6 months compared to reality. However, the resulting error in radiative forcing was very small.

[25] The TAMIP value was set at a more tolerant level to avoid excessive restriction of diversity prior to consideration of model performance in extended climate simulations.

The 49 remaining members were then screened, to identify members showing unrealistic characteristics in selected aspects of UK and European climate. While the diagnostics used were all susceptible to quantitative analysis, a visual inspection by experts was selected as the best method of picking out member-specific biases that might undermine the credibility of simulated climate impacts. This qualitative assessment considered 5-year averages of surface air temperature, precipitation and 850hPa winds in winter (December to February, DJF) and summer (June to August, JJA), and frequency distributions of daily latitudinal positions of an eddy-driven jet index (Woollings et al., 2010). This was constructed from the maximum zonal wind speed at 850hPa.

Eight ensemble members were eliminated because their prevailing south-westerly winds over the UK were much too weak in winter. These members also failed to replicate the three preferred (northern, central and southern) locations of the North Atlantic winter jet found in observations (Woollings et al., op. cit.). Six members (four of the eight discussed above, plus two more) also simulated a substantial winter cold bias in surface air temperature over Northern Europe. In these members, biases in UK average temperature amounted to 1-2°C[26]. The two additional members were rejected on this basis, bringing the total excluded to ten.

d. Filtering to maximise diversity of PPE

As explained above, the purpose of the diversity filtering was to provide as broad a range of climate storylines as possible, subject to the constraint of plausibility. The need for this step arose from considerations of HPC capacity (a maximum of 25 coupled ensemble members was feasible), plus recognition that restricting ensemble size would help facilitate impacts studies by limiting the data processing requirements for users.

Three short, idealised climate change experiments, based largely on standard CMIP5 experiment protocols, were run for the diversity analysis. These consisted of:

- A two-year simulation to estimate the effective radiative forcing (ERF)[27] due to anthropogenic aerosol emissions between 2005-9 and 1860, measured as the change in planetary radiation balance at the top of the atmosphere (TOA) relative to the AMIP experiment. This was calculated by re-running the AMIP PPE (which prescribed aerosol emissions from both natural and anthropogenic sources) using 1860 aerosol emissions (which consisted only of the natural component). Both simulations used the same SSTs, sea-ice extents and greenhouse gas concentrations.

- A two-year simulation to estimate the ERF due to a quadrupling of CO_2 concentration relative to the AMIP experiment, holding SST, sea-ice, other greenhouse gas concentrations and aerosol emissions fixed.

---

[26] A stricter tolerance level for UK and European temperature biases was used at this stage, than during the subsequent filtering of coupled simulations in section 3.3f. This is because use of observed SSTs in the North Atlantic ocean (and elsewhere) in the AMIP simulations removed a major potential source of terrestrial surface air temperature biases.

[27] Radiative forcing (RF) measures the immediate change in radiative balance of the earth in Wm^{-2}, due to a change in an external driver of climate change such as CO_2, aerosol composition or a volcanic eruption. Effective radiative forcing (ERF) denotes the change in RF following rapid adjustments to the resulting warming or cooling of the atmosphere (for example changes in atmospheric temperatures and cloud properties), that occur on time scales of a few days to about a year. The definition excludes responses on longer time scales driven by changes to ocean temperatures and sea-ice.

- A five-year simulation replicating the AMIP simulation, but with a prescribed 4°C globally-averaged increase in SSTs, imposed as a global pattern typical of coupled ocean-atmosphere climate change simulations (e.g. Webb and Lock, 2013). This experiment provides a useful guide to the strengths of large-scale climate feedbacks expected in response to changes in greenhouse gases (Ringer et al., 2014). However, the estimates are approximate, and likely to be biased, because the nature of the forcing is different, and sea-ice extents (being prescribed) are unable to respond to the simulated warming.

Diversity was assessed using a combination of climate change and historical performance metrics. The climate change metrics consisted of two sets of nine variables from the idealised simulations described above, representing global and regional changes respectively. The global-average diagnostics were total aerosol ERF plus its radiation and cloud components, CO_2 ERF, and total feedback strength plus its clear-sky and all-sky shortwave and longwave components. The regional diagnostics were derived from surface air temperature and precipitation responses in DJF and JJA, chosen from a subset of the Giorgi and Francisco (2000) regions in which the ranges of response were significantly broader than uncertainty due to internal variability. The latter was estimated by running 32-member stochastic ensembles using the standard model variant. Nine of these regional responses were chosen, using a dimensional reduction technique (Principal Variables, Cumming and Wooff, 2007) that finds the subset of quasi-independent variables that explains most variation in the full set. The same technique was used to identify a further nine metrics measuring historical performance, drawn from the quantitative assessment variables of Fig. 3.3.

This set of metrics provided a 27-dimensional space of model characteristics, which was then used to identify the most diverse set of 25 members from the 39 candidates. We used an algorithm based on Karmalkar et al. (2018) in which the standardised Euclidean distance between successive pairs of members was maximised, starting from the standard member. Figure 3.4 shows the ranges of the resulting selection for global mean values of aerosol and CO_2 ERF and climate feedback strength. The short lengths of the simulations limits the precision of results for specific PPE members. However, uncertainty due to internal variability is considerably smaller than the spread in aerosol forcing and climate feedback found across the 25 selected members (Fig. 3.4).
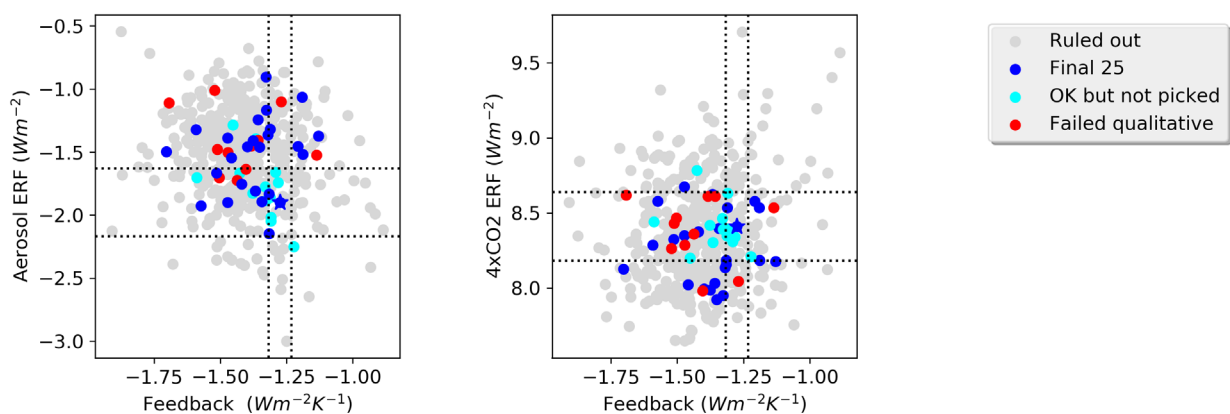

Figure 3.4. Scatter plots showing global average values of effective radiative forcing (Wm^{-2}) due to anthropogenic aerosol emissions (for 2005-9 relative to 1860, left panel) and an assumed quadrupling of CO_2 concentration (right panel), estimated from short, idealised climate change experiments (see text). These are plotted against estimates of the global climate feedback parameter (Wm^{-2}K^{-1}), for the 495 perturbed GC3.05 atmosphere model variants for which it was possible to run all four of the required AMIP simulations. Grey dots show members rejected via the quantitative analysis of Fig. 3.3, and red dots members rejected by the subsequent qualitative evaluation (see text). The blue dots show the 39 members surviving performance-based filtering. The darker blue dots indicate the 25 selected as potential coupled model variants, including the standard member (starred). These maximise diversity in a set of 27 metrics described in the text, which include the three global climate change diagnostics shown here. Dashed lines show uncertainty (best estimate plus and minus two standard deviations) due to internal climate variability. This is derived from an ensemble of 32 simulations using the standard ensemble member, distinguished by different realisations of stochastic parameterisation outputs (see text).

For these variables, the 25 members capture much of the range of the full AMIP ensemble, although members featuring a combination of relatively weak aerosol forcing and feedback strength (in the context of the GC3.05 parameter space) were not represented. These were excluded by the evaluation of climatological performance described above. For CO_2 forcing, the 25 members explore a range of approximately 8.0-8.5 Wm^{-2}, because members with higher values were removed by the performance-based filtering. Across the 25 members, the range for anthropogenic aerosol forcing is approximately -1.0 to -2.0 Wm^{-2}. The 5-95% probability range used in Strand 1 is −0.1 to -1.9 Wm^{-2}, derived from AR5. GC3.05-PPE thus contains a number of members with a strong estimated forcing compared to the bulk of the AR5 distribution. This is discussed further in section 3.4.

e. Design of GC3.05 coupled ocean-atmosphere PPE experiments

Here, we summarise the design of an initial PPE of 25 GC3.05 transient climate change simulations. These were run from 1900-2100 under historical changes in natural and anthropogenic radiative forcing till 2005, switching to the RCP8.5 scenario to 2100. Filtering of these simulations based on historical performance is described in sub-section (f) below. A full description of the experiments is provided by Yamazaki et al (2018).

Calibration

The transient simulations were preceded by calibration and spin-up phases. The purposes of these were to bring the ocean model towards an equilibrium state consistent with exchanges of heat, water and momentum simulated across the atmosphere-ocean interface, and initialise the transient simulations with realistic spatial distributions of SST and sea-ice. Given the importance of SST biases for the credibility of regional projections of climate variability (Scaife et al., 2011) and change (He and Soden, 2016), we decided to apply flux adjustments to surface heat and water fluxes, as in earlier HadCM3-based PPE experiments contributing to UKCP09 (Collins et al., 2011).

The calibration phase involved only the standard PPE variant (STD). Since the subsequent transient simulations required initial conditions typical of the beginning of the 20th century, we used constant pre-industrial values for greenhouse gas concentrations and aerosol emissions taken from standard CMIP5 data (https://cmip.llnl.gov/cmip5/forcing.html). As observational coverage of sub-surface temperatures and salinities was sparse around 1900, the simulation was started from an ocean state derived from temperature and salinity observations for 2004-8 (Ingleby and Huddleston, 2007). During this simulation, sea surface temperatures and salinities were continuously and linearly relaxed towards a seasonal cycle[28] of prescribed spatial fields representing observed conditions typical of late 19th and early 20th centuries (Rayner et al., 2003; Good et al., 2013), while sub-surface ocean conditions adjusted towards the applied trace gas, aerosol and surface forcing.

---

[28] A seasonal cycle of prescribed sea surface temperatures and salinities was applied, in order to facilitate the subsequent calculation of seasonally-varying flux adjustments that would limit the development of biases in these variables on a seasonal basis.

Following experimentation with alternative choices (Yamazaki et al., 2018), a relaxation time scale of 4 months (for a 50 metre ocean mixed layer) was chosen for both temperature and salinity. The relaxation was applied in all marine regions apart from the Arctic Ocean. This set-up was found to maintain stability in the AMOC, whereas a full global application of the relaxation led to a rapid and unrealistic weakening. The calibration integration was stopped after 115 years, once surface ocean-atmosphere fluxes and the relaxation terms had reached a quasi-equilibrium. The relaxation terms were averaged over the last 30 years to provide prescribed flux adjustment fields. These were applied to the ocean surface layer (in addition to heat and water exchanges simulated by the model interactively) in subsequent phases of the experiment. The flux adjustments counter the development of biases in long-term averages of SST and surface salinity, while allowing unconstrained internal variability to occur in the simulations. The flux adjustment fields covered the entire global ocean apart from the Arctic[29], and varied with location and season, but not from year to year. The main benefit of applying flux adjustments is that they reduce the potential for biases in SST to affect projected regional changes (He and Soden, 2016), with the caveat that the flux adjustments can also affect aspects of the ocean circulation that are sensitive to variability in surface energy or buoyancy forcing, such as the AMOC (Collins et al., 2006). Also, holding the flux adjustments constant from year to year involves the key assumption that the sources of bias that they are intended to counter are invariant under climate change.

Spin-up

The spin-up phase initially involved a 68 year simulation of STD, started using the end point of its calibration simulation. External forcing agents were specified as in the calibration phase, and STD was run with flux adjustments applied. Following this, STD plus the 24 perturbed members were run from the end point of the first spin-up phase (using the same external forcing), for periods ranging from 65-83 years[30]. For the perturbed members, offline estimates of the required flux adjustments were produced, as HPC limitations precluded running a separate calibration simulation for each model variant. For the heat flux adjustment, this was done by adding a constant offset to the spatial pattern obtained for STD. The offset consisted of the difference between global mean net TOA radiation in the AMIP simulations of the relevant member, and STD. This approach reflects evidence from earlier models that *regional patterns* of heat flux adjustment are, to leading order, determined by ocean model properties (resolution, circulation and SST errors in regions of high spatial gradients), while biases in atmosphere-ocean fluxes play an important role at broader scales (Roberts et al., 1997). For the fresh water flux adjustment, all members used the STD fields without application of offsets.

During the spin-up period, the climatological state of the upper ocean adjusted from the starting conditions. Potential drivers of these adjustments included non-linear effects of variability in SST and surface salinity (all members), ongoing trends in deeper ocean layers (all members) and the approximate nature of the applied flux adjustments (all members except STD).

Regional surface salinity values (in perturbed members as well as STD) typically remained within 1psu of the constraining observations used in the calibration step, apart from in the Arctic. Yamazaki et al (2018) will provide further discussion of the impact of simulated surface freshwater fluxes on variations between ensemble members in salinity biases, and the strength of AMOC (see Fig. 3.5 and sub-section (f) for discussion of simulated AMOC values). Regional biases in SST (typically less than ±2°C, but with significant exceptions in some ensemble members) are also discussed in sub-section (f).

---

[29] In the case of fresh water, flux adjustments were not applied at coastal river outflow points either.

[30] The lengths varied because some members ran slower than others, due to restarts caused by numerical instabilities.

By the end of the spin-up, key indicators such as AMOC, sea-ice extents in both hemispheres, SST, surface salinity and surface heat and water fluxes had reached approximate equilibrium, thus providing suitable initial conditions for the transient simulations. Inevitably, however, the deep ocean was not in equilibrium (in common with multi-model ensemble experiments, e.g. SenGupta et al., 2012). This is a consequence of the millennial time scales associated with adjustments below about 2km (Yang and Zhu, 2011). The STD spin-up was run on for a further 190 years, beyond the point of switching to the historical phase. This provided an indication of the potential for climate drift arising from incomplete equilibration (Sen Gupta et al., 2013). The results showed a warming of about 0.2°C in GMST during the first 40 years, and little change subsequently. This warming occurs largely over the Southern Ocean. For parts of the world that are reasonably well observed (which excludes the Southern Ocean, Arctic and Antarctic regions), the warming is confined to the first 20 years. This warming has only a modest influence on the simulated historical changes in GMST against observations (Fig. 3.12, discussed in section 3.4), and is small compared with projected future changes (see section 3.5). Control simulations are not yet available for the other 14 GC3.05-PPE members, so we cannot assess the drift in those.

Specification of climate forcing agents

The historical phase[31] of the transient simulations ran from 1900-2005, using CMIP5 time series of observed changes in well-mixed greenhouse gases, ozone, solar radiation, major volcanic eruptions and natural and anthropogenic aerosol precursors. In the case of sulphur dioxide, PPE members sampled uncertainty in the observed emissions, using a scaling factor (ranging from 0.5 to 1.5 – see Fig. 3.2) that constituted one of the perturbed parameters in GLOMAP-mode (Carslaw et al., 2013).

Time-dependent changes in fractional coverage of land vegetation types were prescribed[32], according to the harmonised land-use reconstructions used in CMIP5 (Hurtt et al., 2011). Following a similar approach to that taken in HadGEM2-AO (Baek et al., 2013), land use change is represented by mapping anomalies in total crop and pasture from Hurtt el al. (2011) to changes in the combined coverage due to C3 and C4 grass plant functional types (PFTs). The anomalies were applied to the reference present day land cover map by clearing or expanding the combined coverage of trees and shrubs in relevant regions. The ratios of coverage were maintained between C3 and C4 grasses, and between broadleaf trees, needleleaf trees and shrubs. Coverage of urban, soil and land ice classifications was kept unchanged.

After 2005, the applied forcing followed the RCP8.5 scenario. In the case of future sulphur dioxide emissions, the member-specific scaling factors applied during the historical phase (see above) were also applied to the RCP8.5 values, in order to avoid creating spurious trends in aerosol forcing following 2005. Volcanic forcing was prescribed from an observed estimate to 2000 (Sato et al. 1993, updated). Subsequently, it followed a profile similar to that of Jones et. al (2011), ramping down to a low level to 2020, and then recovering to the average level used in the calibration and spin-up simulations by 2040. Total solar irradiance was specified using the Lean et al (2009) data to 2008 and then a fixed 12-year cycle to 2100, obtained by continuously repeating the observations for 1996-2008.

---

[31] After the spin-up simulations, an error was found in the soil hydraulics scheme. The movement of water through the soil is modelled using the van Genuchten (1980) equations, whereas parameter values that determine the hydraulic conductivity for different soil types had been determined using a different set of equations (Brooks and Corey, 1964), leading to insufficient vertical transfer of soil moisture. This error was corrected for the transient simulations. The effects of the change were tested. It led to changes in the detailed patterns of runoff over land, but no significant impact on the ocean surface freshwater budget at basin scales.

[32] The GC3.0 configuration includes a scheme that constrains the grid-box average snow-free albedo to replicate an observed climatology. This scheme was switched off in GC3.05, in order to allow simulated surface short-wave radiation to respond to the specified changes in land-use.

While GC3.05 is driven by prescribed concentrations of CO_2, we specify a range of alternative pathways in the PPE, in order to account for uncertainties in carbon cycle feedbacks (see section 1.4). The selection of these, and the method for assigning them to PPE members, is described in section 3.3g below.

f. Filtering of GC3.05-PPE using historical performance

The final filtering stage involved assessment of the 25 GC3.05 simulations during the period 1900-2005. Two members were dropped due to frequent occurrences of numerical instabilities in the atmosphere component, which made one integration impractical to continue and the other impossible to complete in time. Remaining members were assessed against selected characteristics of 20th century climate, focusing initially on AMOC (a key driver of climate and its variability in the North Atlantic and Europe), and regional SST biases worldwide. Three simulations showed a substantial spin-down of AMOC (Fig. 3.5), leading to values at or below 8 Sv by 1950. These values were clearly inconsistent with recent observed values in the range 13-20 Sv (Smeed et al., 2018, also shown in Fig. 3.5). In these members, this was accompanied by large negative biases of 5-6 °C in North Atlantic SSTs (Fig. 3.6), substantially worse than in other PPE members and the CMIP5-13 simulations. These three PPE simulations were excluded.
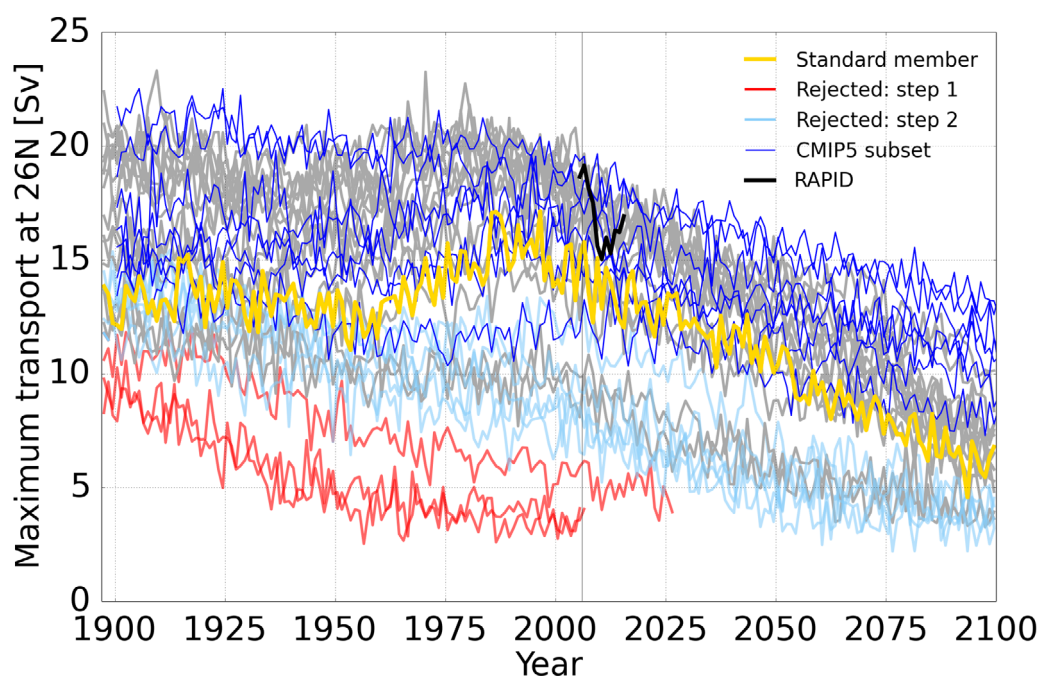

Figure 3.5. Annual simulated values from 1900-2100 of the strength in the Atlantic Meridional Overturning Circulation (AMOC), measured as net northward transport of water in the Atlantic Ocean at 26°N, in Sverdrups ($10^6 m^3 s^{-1}$). The yellow line shows values from STD, grey lines showing time series from the other 14 members of GC3.05-PPE that passed two rounds of screening of the historical stage of the coupled ocean-atmosphere simulations, prior to the switch to RCP8.5 forcing in 2005. Red lines show three of the five members excluded during the first screening step. This was due to an unrealistically weak circulation of <5Sv, suggesting a near-collapse of the AMOC inconsistent with recent observations (black line, from the National Environmental Research Council RAPID programme, Smeed et al., 2018). Light blue lines show five further PPE members excluded during the second screening step, three of which show AMOC values clearly below 10Sv during the period 1981-2000. These were excluded in conjunction with assessment of other key indicators relevant to the simulation of UK climate. Dark blue lines show results from seven of the 13 CMIP5 models included in Strand 2, for which AMOC values were available.

The remaining 20 members provided SST simulations competitive with the level of skill found in CMIP5-13 (Fig. 3.6), and were accordingly run on to 2100. However, it was decided subsequently to carry out additional assessment of the historical part of the GC3.05-PPE simulations, with a more specific focus on identifying levels of bias that might compromise their credibility for use in UK impacts assessments. This process led to the exclusion of a further five members, and is described below.
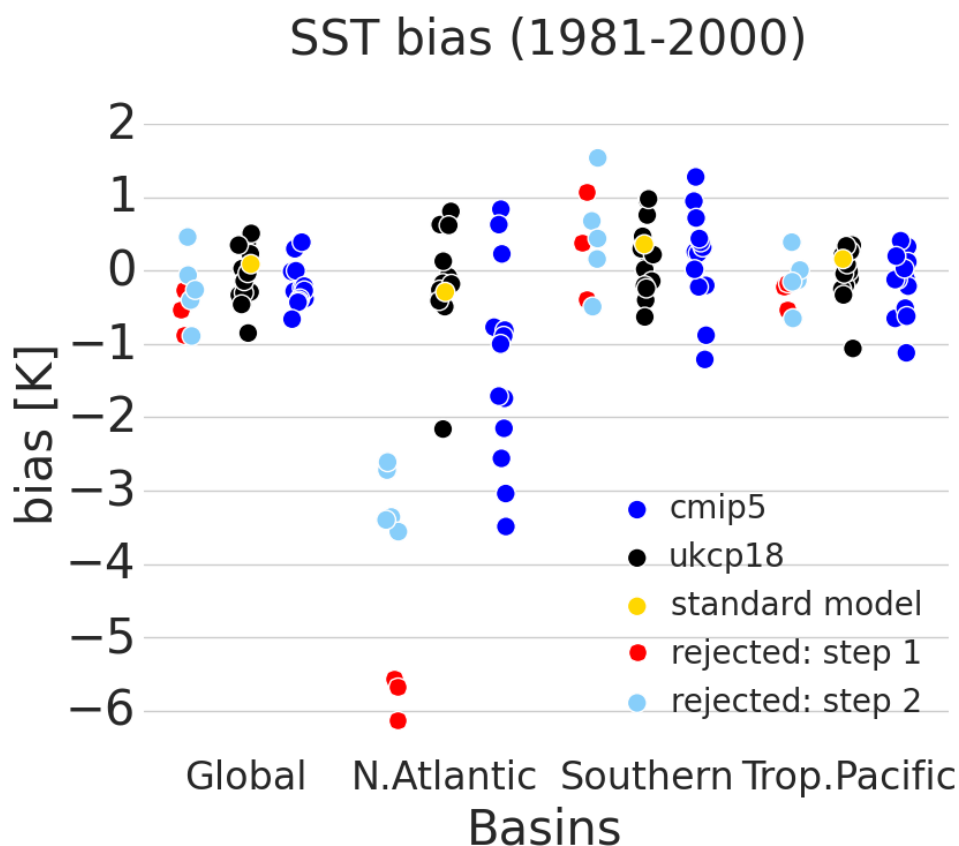
SST bias (1981-2000)

Figure 3.6. Biases in area-averaged annual sea surface temperature (SST) for 1981-2000, relative to observed values from HadISST1 (Rayner et al., 2003). "N. Atlantic" denotes 45-70°N, 75°W-30°E; "Southern" 30-90°S; "Trop. Pacific" 15°S-15°N, 105°E-75°W. Red dots show potential members of GC3.05-PPE excluded in the initial screening of the historical simulations, in which SST biases were an important aspect of the assessment criteria. Black circles show the final set of 15 GC3.05-PPE members, with light blue circles denoting the five members excluded during the second screening step (in which SST biases were not an explicit criterion). Dark blue circles show biases from the CMIP5-13 models.

The second assessment reconsidered AMOC, plus five additional variables: trends in average northern hemisphere surface air temperature during 1900-2012 and biases in winter and summer climatological averages of surface air temperature and precipitation for Europe during 1981-2000. Instances of poor performance were assessed by comparison with observations and the envelope of performance in CMIP5-13, identifying a PPE member as seriously biased if its simulation of the relevant metric was clearly inferior to the worst-performing CMIP5 members. In the case of AMOC, estimates were only available for seven members of CMIP5-13, so we used 10Sv as a threshold for an unrealistically weak circulation during the late 20th century. This was the minimum value found in a larger set of CMIP5 models by (Heuzé, 2017), based on overturning measured at 30°N.

Six of the PPE members showed a historical cooling trend of 0.3-0.6 ºC per century in northern hemisphere surface air temperature, probably driven mainly by a strong negative aerosol forcing (Fig. 3.4 and discussion in section 3.3d). However, observations showed a significant warming trend amounting to 0.7 ºC per century, while other PPE and CMIP5-13 members showed warming trends of varying magnitude. The six cooling simulations were therefore assessed as seriously biased for this metric. For AMOC, three PPE members (in addition to those rejected during the initial assessment) showed an average strength over 1981-2000 clearly below 10Sv (Fig. 3.5), and were assessed as seriously biased. For European climatology, no serious biases in PPE members were found for precipitation, or for surface air temperature in summer. In winter, three members simulated an average cold bias (over the whole European land mass) worse than -4.2 ºC, the largest found in CMIP5-13. These members, which produced regional cold biases exceeding -10 ºC over parts of northern Europe, were assessed as seriously biased.

The three criteria discussed above are inter-dependent to a degree. In particular, a hemispheric cooling and a weak AMOC both contributed directly to a cold bias over Europe. However, several different process drivers are likely to be responsible for the three types of bias. As noted above, strong aerosol forcing was probably a key driver in those members showing hemispheric cooling during the second half of the 20th century. However, aerosol forcing can also strengthen rather than weaken the AMOC (Menary et al., 2013). In the PPE, four of the six members showing the cooling trend do not simulate strengthening of the AMOC in response to aerosol forcing, suggesting competing influences on AMOC. Regarding the European cold bias, insufficient radiative heating from the longwave effects of clouds (a feature common to most of the PPE members) was an additional contributory factor. Another influence was perturbations to *nsigma*, a gravity wave drag parameter that significantly affects the strength of average mid-latitude westerly winds (see Table D.1), and hence drives variations in the magnitude of the cold bias between different PPE members.

It was decided to exclude those members recording serious biases for at least two of the three criteria discussed above, in order to avoid losing members that may suffer from only a minority of the relevant process errors. The excluded members consisted of one that failed all three criteria, two that failed the AMOC and hemispheric cooling checks, and two failing on European cold bias and hemispheric cooling. One of the six members failing on hemispheric cooling was retained, because it was assessed as acceptable on the other criteria. This left 15 members to constitute the GC3.05-PPE projections. These were added to CMIP5-13 to form a combined set of 28 projections for UKCP18. A more general evaluation of these simulations is provided in section 3.4.

g. Specification of future CO_2 pathways for GC3.05-PPE members

As explained in section 1.4, the GC3.05-PPE simulations use a range of future CO_2 pathways, consistent with uncertainties in how carbon cycle feedbacks convert emissions in the RCP8.5 scenario into atmospheric concentrations. The standard RCP8.5 concentration time series provides one. A further 72 pathways were available for selection, provided by emissions-driven earth system model outputs. These consisted of the 57 ESPPE simulations from Strand 1, and results from 15 CMIP5 models, 12 of which were also used in Strand 1 (section 2.2). Pathways were chosen by applying an observational test to the model results (Booth et al., 2017, updated), based on regressed linear trends during 1969-2014.

For each earth system model simulation, this was done by calibrating the simple climate model (SCM) of Strand 1 to replicate its transient changes in CO_2, and then running the SCM using alternative assumptions for fossil fuel and land use emissions (as in section 2.2, Stage 2d). Any model for which the spread of trends failed to encompass the observed value was rejected, leaving 31 simulated future pathways. Annual time series from these were smoothed using a Butterworth filter with a 30-year cut-off, and blended with the observed value at 2005. Adding the standard RCP8.5 concentration time series provided 32 plausible pathways.

A subset of these were then picked to drive GC3.05 ensemble members that survived the performance-based filtering of section 3.3f. The selection of CO_2 pathways was done after the first of the two filtering steps (which reduced the pool of potential GC3.05 variants from 25 to 20). The standard RCP8.5 profile was assigned to STD, and 19 additional pathways were selected. This was done by identifying 15 GC3.05 parameters that made the largest contributions to the spread in CO_2 forcing, aerosol forcing and climate feedback strength, estimated from the idealised climate change simulations of Fig. 3.4 (Rostron et al., 2018). The CO_2 pathways were assigned on the basis that member-by-member variations between CO_2 and the 15 parameters should be as independent as possible[33]. This choice was made to avoid building in strong accidental relationships between physical climate change drivers and carbon cycle feedbacks[34].

---

[33] This was achieved by sampling the assignment of CO_2 profiles to GC3.05 PPE members one million times, and identifying the sample that maximised the determinant of the 16-dimensional correlation matrix of the 15 parameters and the set of CO_2 outcomes at year 2100.

[34] Extension of the GC3.05 PPE experiments to include carbon cycle perturbations in the UKESM1 model framework (see section 1.3) would be required to provide physical justification for such relationships.

Figure 3.7 (left panel) shows the range of CO_2 pathways in the 15 members (including STD) that passed the second filtering step in section 3.3f above). The right panel shows the associated concentrations in the year 2099, compared against the posterior distribution obtained from Strand 1. The latter was derived using the same set of earth system model simulations and a similar observational constraint on historical CO_2 concentration, but accounted for the additional constraints and modelling evidence described in section 2[35]. The range of 857-1256 parts per million (ppm) essentially covers the full range of the Strand 1 probability distribution (a 1-99% range of 896-1268 ppm), with six pathways above the median (1039 ppm) and nine below. This shows reasonable consistency between the pathways and the pdf. The standard RCP8.5 concentration of 926 ppm lies near the low end of both distributions, indicating that the single pathway used in concentration-driven CMIP5 simulations represents a scenario of relatively weak carbon cycle feedback.
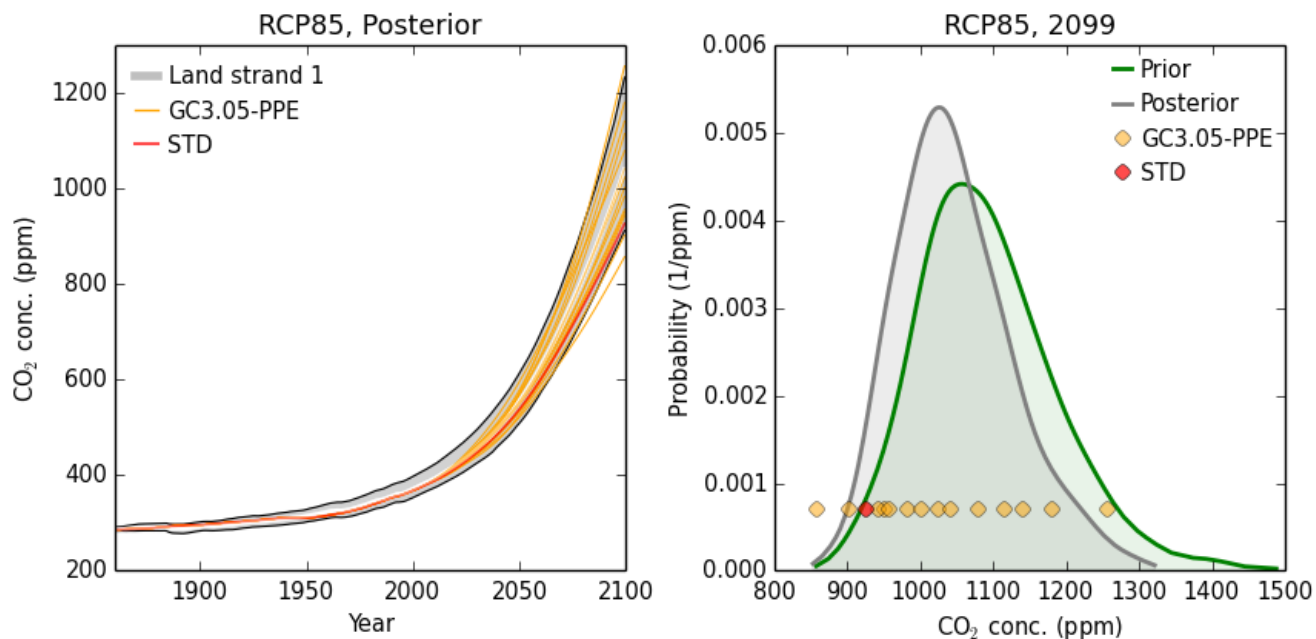

Figure 3.7. Posterior probability distribution (grey curve, right panel) from Strand 1, for atmospheric CO_2 concentration (parts per million) in 2099, under carbon emissions prescribed by the RCP8.5 scenario. Green curve shows the corresponding prior distribution, which can be compared with the grey curve to see the impact on the latter of applying the Strand 1 observational constraints (section 2.2, Stage 2). Orange circles show corresponding concentrations from the CO_2 pathways used to drive 14 of the 15 GC3.05-PPE members included in the set of 28 Strand 2 projections. The red circle denotes the standard pathway used in concentration-driven RCP8.5 simulations. This pathway is used in the CMIP5-13 projections included in Strand 2, as well as the GC3.05-PPE member with unperturbed parameter settings (STD). Left panel shows annual CO_2 values in the GC3.05-PPE members, in comparison with the time-dependent evolution of the Strand 1 probability distribution. The white line is the median of the latter, with grey shading showing the range between the 5% and 95% probability levels, which are plotted as black lines.

---

[35] It was not possible to use the probabilistic projections of CO_2 from Strand 1 to provide the pathways for the GC3.05 PPE, because the development timelines for Strand 1 and 2 ran in parallel.

---

h. Choice of CMIP5 models to provide additional projections

Of the 42 CMIP5 models used to provide concentration-driven simulations under the RCP8.5 scenario, eleven were not considered, due to insufficient availability of daily data for a core set of impacts-relevant variables (precipitation, and diurnal mean, maximum and minimum surface air temperature). The remaining 31 models were screened using a combination of global and regional performance criteria. Those passing the screening stage were clustered into groups of models sharing similar error characteristics, and Strand 2 members were then chosen by picking the best-performing model(s) in each group. McSweeney et al. (2018a) describe the selection procedure in detail, with a brief summary given below.

The global and regional performance criteria were drawn mainly from published literature, supplemented by additional diagnostics provided by McSweeney at al (op. cit.). Many of the available metrics were also used in the design of GC3.05-PPE. The global assessment considered variability in global and zonal averages of surface temperature, and worldwide patterns of SST bias. The regional assessment considered AMOC, SST errors in the North Atlantic, biases in climatological averages of 850hPa winds over the North Atlantic and Europe, and biases in surface air temperature over the UK and Europe. Errors in selected aspects of intraseasonal variability were also assessed, including track densities for North Atlantic storms (from McSweeney et al., 2015), blocking frequencies in the northern hemisphere (Christensen et al., 2013) and frequencies of weather types diagnosed from daily mean sea-level pressure fields (Perez et al., 2014).

Models were excluded if:

1. One or more of the above variables was assessed as being so poorly simulated (for example, an absence of a clear prevailing south-westerly flow over the UK) that relevant future changes in regional climate could not usefully be diagnosed for impacts assessments.

2. Two or more aspects were assessed as possessing significant biases (for example a substantial bias in SSTs in the North Atlantic, or over a large remote region such as the Southern Ocean, or a failure to replicate observed zonal maxima in storm track density).

Three models were screened out using the first of the above thresholds, failing circulation-based criteria for Europe. Six more failed on the second threshold. The remaining 22 models were clustered into groups showing a maximum degree of commonality in their simulation errors. This exercise used the results of Sanderson et al. (2015a,b), who constructed a matrix of pairwise inter-model distances using a multivariate metric. This was developed from seasonal climatology errors in global spatial fields of surface temperature, precipitation, outgoing shortwave and longwave radiation at the top of the atmosphere, and latitude-height distributions of zonally averaged atmospheric temperature and relative humidity.

Sanderson et al. found that pairs of models with similar biases in these emergent properties tend also to share components, confirming their metric as a useful way of identifying common structural assumptions. Picking models from each group ensures that a diversity of structural choices was sampled. Ten groups were found that contained at least one model that passed the screening. Two models were picked from two groups containing four eligible models, and one from the others. In groups requiring a choice, the best-performing models were chosen. One eligible model (EC-EARTH) was not included in the Sanderson et al. near-neighbour analysis. This was included in the 13 models forming the final selection for Strand 2 (see Table 3.1).

Table 3.1. The 13 CMIP5 models selected to contribute simulations of historical and future climate to Strand 2, using simulations driven by prescribed future CO_2 concentrations under the RCP8.5 scenario.

Modelling group	Group acronym	Model designation
Centro Euro-Mediterraneo per I Cambiamenti Climatici	CMCC	CMCC-CM
Beijing Climate Centre, China Meteorological Administration	BCC	BCC-CSM1.1
Canadian Centre for Climate Modelling and Analysis	CCCMA	CanESM2
Commonwealth Scientific and Industrial Research Organization (CSIRO) and Bureau of Meteorology (BOM), Australia	CSIRO-BOM	ACCESS1-3
Community Earth System Model Contributors	NSF-DOE-NCAR	CESM1-BGC
Centre National de Recherches Météorologiques / Centre Européen de Recherche et Formation Avancée en Calcul Scientifique	CNRM-CERFACS	CNRM-CM5
EC-EARTH consortium	ICHEM	EC-EARTH
NOAA Geophysical Fluid Dynamics Laboratory	NOAA GFDL	GFDL-ESM2G
Met Office Hadley Centre	MOHC	HadGEM2-ES
Institut Pierre-Simon Laplace	IPSL	IPSL-CM5A-MR
Max-Planck-Institut für Meteorologie	MPI-M	MPI-ESM-MR
Meteorological Research Institute	MRI	MRI-CGCM3
National Center for Atmospheric Research	NCAR	CCSM4

While diversity in historical simulation errors was used as an explicit criterion in the selection of CMIP5 models, diversity in projected future changes was not. This was because further reductions in ensemble size would risk reducing the influence of the CMIP5 ensemble relative to the 15 GC3.05 members, which would be difficult to justify in the absence of a clear objective basis for allowing either GC3.05-PPE or the multi-model ensemble to dominate the combined set of Strand 2 projections. Compared to the 31 CMIP5 models considered in the selection process, McSweeney et al (2018a) found that CMIP5-13 captured almost the full range of changes in GMST. For the UK region, CMIP5-13 also captured much of the full CMIP5 range of seasonal mean changes in temperature and precipitation. We note, however, that the exclusion of the poorest models does impact the range of regional projections by discounting the member with the lowest warming for the UK (INMCM4), and those with the largest temperature and precipitation increases (MIROC-ESM and MIROC-ESM-CHEM).

3.4. Evaluation of the strand 2 simulations

a. Global performance

Figure 3.8a-d shows global patterns of bias in surface air temperature and precipitation in DJF and JJA for 1981-2000, relative to analyses of observations from ERA-Interim and the Global Precipitation Climatology Project (GPCP) (Table E.2). Median biases are shown to provide estimates of the systematic component of model error in the relevant ensemble, along with those of the best- and worst-performing members. Root-mean-square error (RMSE) and global mean bias statistics are also provided, for all members.

 In DJF, median values of GC3.05-PPE show a significant cold bias over most of the northern hemisphere continental land mass (Fig. 3.8a), despite removal of members showing the worst cold biases over Europe (section 3.3f). All GC3.05-PPE members share this error, but the magnitude varies substantially across the ensemble (Fig. 3.8a). Contributing factors to the variation in bias include longwave radiative heating (insufficient in most members due to a negative bias in simulated cloud extents) and the strength of the average mid-latitude westerly winds, as well as aerosol forcing and AMOC strength discussed previously. These biases are likely to enhance future surface temperature changes in some of these regions, by increasing surface albedo feedbacks (see Fig. 3.22 and discussion in section 3.5).

The CMIP5-13 median for DJF also shows a cold bias in many northern hemisphere regions (Fig. 3.8b). In general, it is less pronounced than in GC3.05-PPE, with the exception of south-east Asia, Mexico and South-Western US. In JJA, the perturbed parameter and multi-model ensembles both show a median warm bias over central parts of North America, South-Eastern Europe and parts of Central Asia. Over the oceans, median biases in CMIP5-13 are often somewhat larger than in GC3.05-PPE. This is likely to be due, at least in part, to the use in the latter of flux adjustments, and should not be taken as an indicator that GC3.05-PPE necessarily simulates better than CMIP5-13 the processes giving rise to regional values of SST.
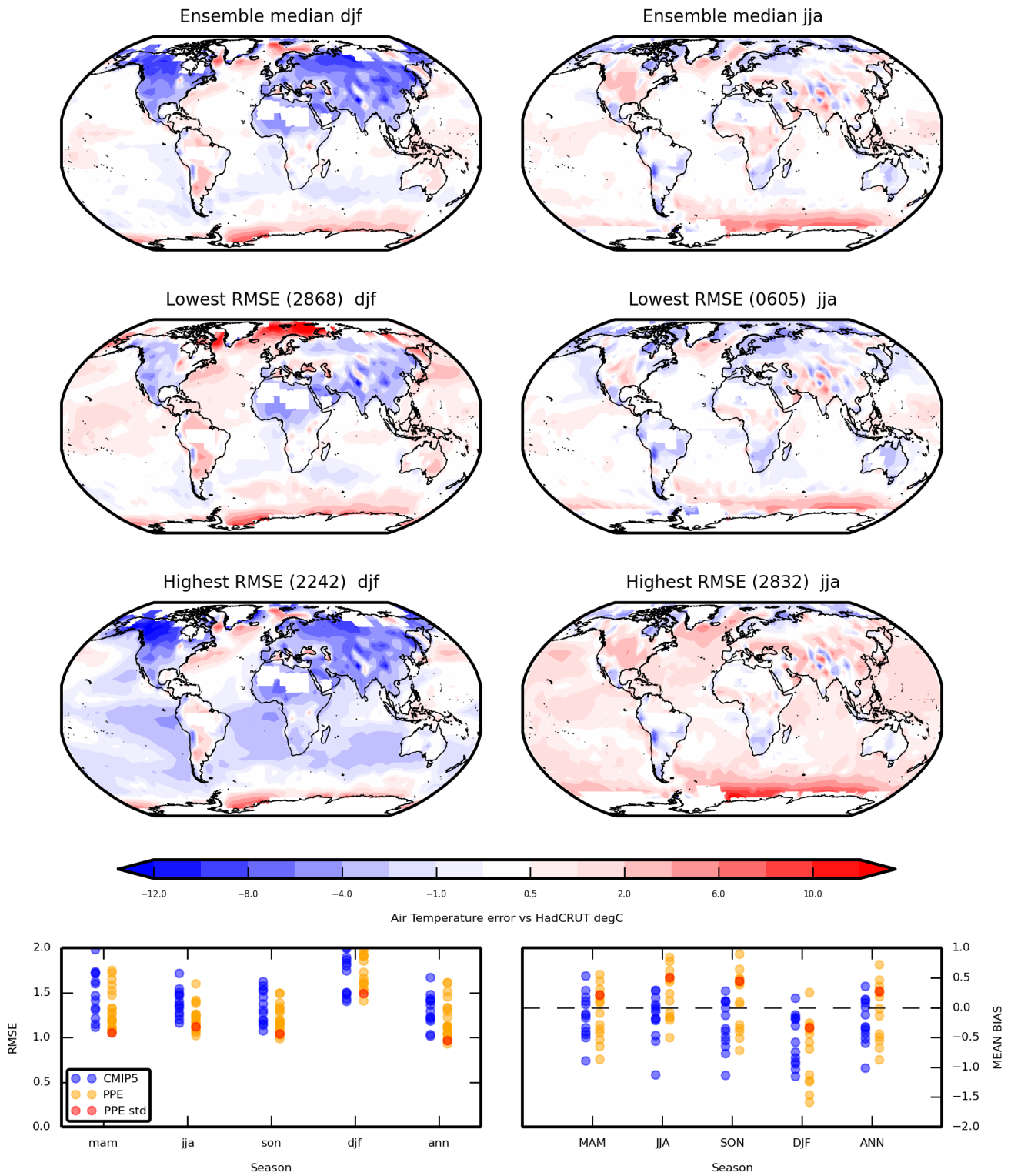
Figure 3.8a. Twenty-year mean biases in surface air temperature (°C) simulated by the 15 members of GC3.05-PPE. Errors for DJF and JJA are calculated relative to ERA-Interim reanalyses of observations, for 1981-2000. Bottom panels show values of root-mean-square error (RMSE, left) and global average bias (right), calculated from the worldwide patterns. Orange dots show GC3.05-PPE members, blue dots depicting corresponding values for CMIP5-13 models (shown in Fig. 3.8b). STD is shown in red. Top row shows ensemble median biases for DJF (left) and JJA (right), second and third rows showing biases in the ensemble members with the lowest and highest RMSE values, respectively.
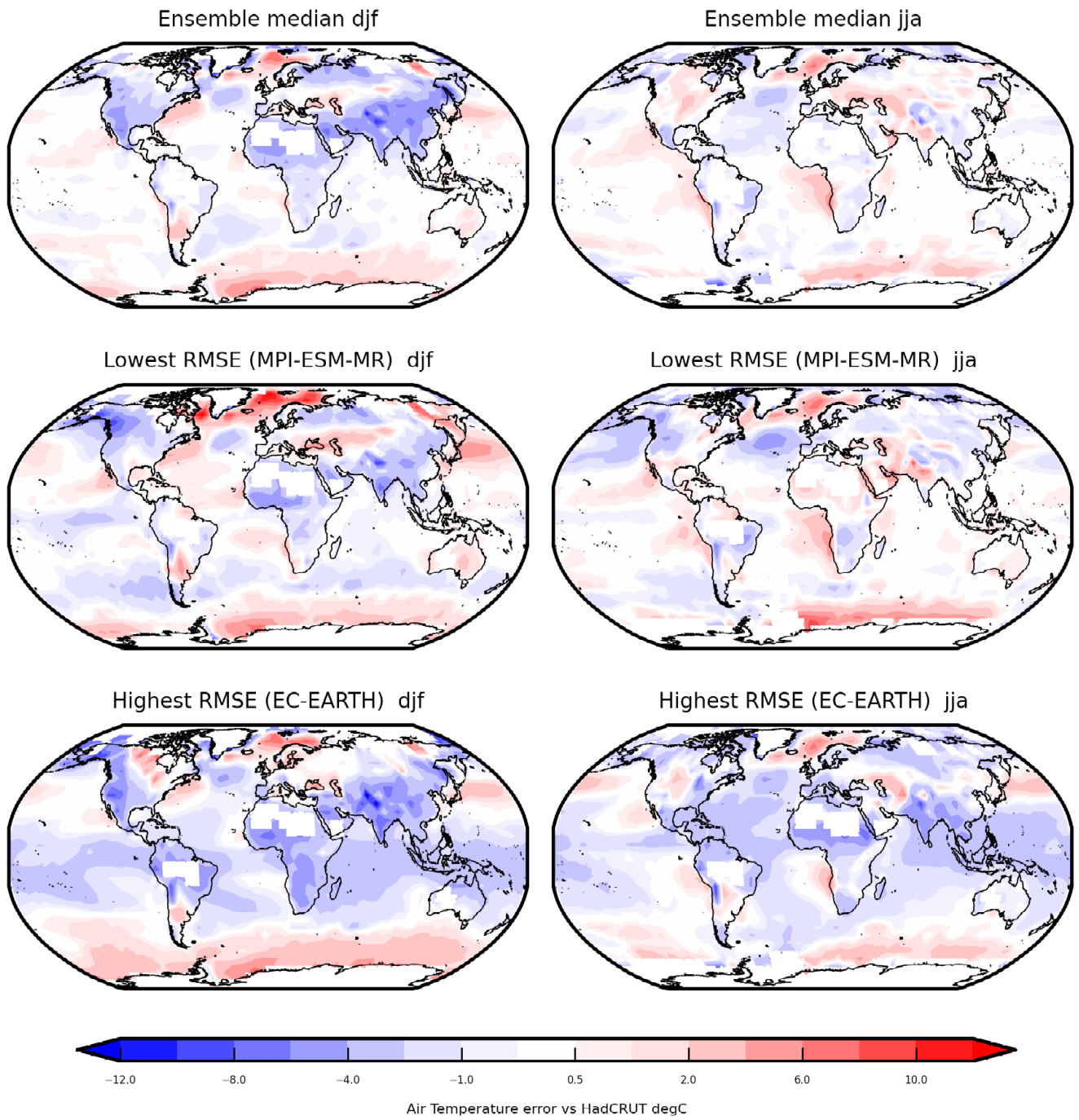
Figure 3.8b. As Figure 3.8a, for CMIP5–13 models. Relevant global average bias and RMSE statistics are shown in the bottom row of Fig.3.8a.
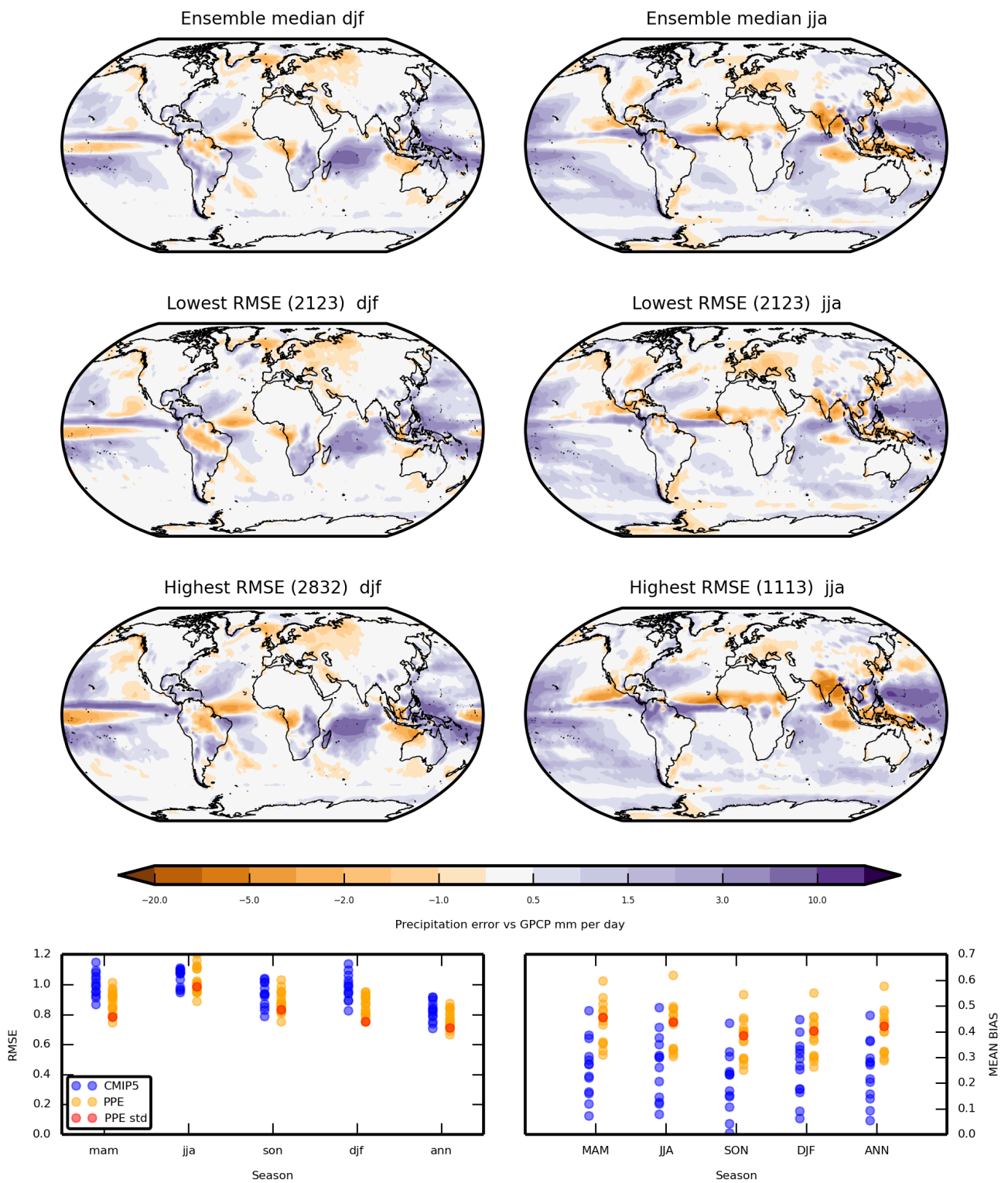
Figure 3.8c. As Figure 3.8a, for twenty-year mean biases in precipitation relative to GPCP observations (mm/day), simulated by the 15 GC3.05-PPE simulations.
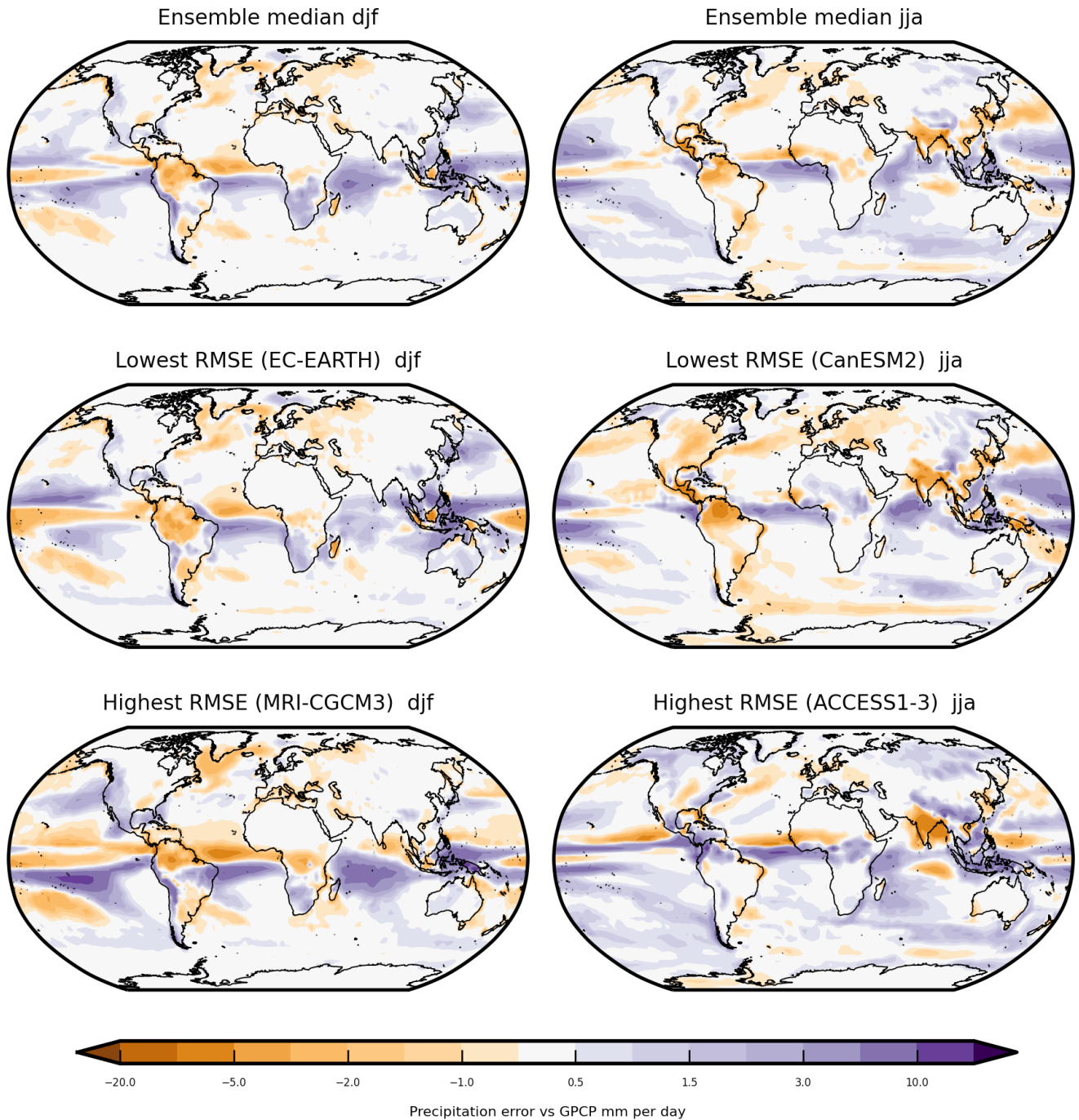
Figure 3.8d. As Figure 3.8b, for twenty-year mean biases in precipitation relative to GPCP observations (mm/day), simulated by the CMIP5-13 models.

In the Southern Ocean a majority of GC3.05-PPE members simulate a warm bias in SST (Fig. 3.6), as do most CMIP5-13 members. This bias is typical of CMIP5 models in general (Meijers, 2014), and is related to errors in the atmosphere-ocean net heat flux (Hyder et al., 2018). In most GC3.05-PPE members the average heat flux adjustment is negative in the Southern Ocean, acting to reduce the warm bias (Williams et al., 2018). The presence of residual warm biases in most GC3.05-PPE members may reflect non-linear effects of variability in SST on average surface heat fluxes. This is because SST variability was restricted in the calibration simulation used to calculate the flux adjustments (section 3.3e), but not in the transient simulations. In Antarctic circumpolar waters, median warm biases in surface air temperature are present in both ensembles. Reductions in sea-ice extent occur in all GC3.05-PPE members during 1900-2005, contributing to the bias in surface air temperature alongside increases in SST. Median biases in surface air temperature over the North Atlantic are restricted by the exclusion of simulations showing substantial cold biases (Fig. 3.6 and section 3.3f), and thus compare well with those in CMIP5-13.

The regional errors discussed above contribute to biases in global average surface air temperature, shown in Fig. 3.8a (bottom right panel) for each season and the annual mean. Ranges of bias in the global average are broadly similar in both ensembles, amounting typically to ~1.5°C. The ranges invariably encompass the observations, although all but one member of each ensemble shows a negative bias in DJF, due to the aforementioned cold biases in winter over the northern hemisphere continents.

Both ensembles share median biases associated with excessive off-equatorial precipitation in the tropical east Pacific Ocean, and south of the equator in the tropical Atlantic Ocean (Fig 3.8c,d), although the Atlantic bias is less pronounced in GC3.05-PPE. Other common biases include excessive precipitation in the western equatorial Indian Ocean, and too little in the south Asian summer monsoon. The GC3.05-PPE median shows too little precipitation over central and southern Europe in JJA, while CMIP5-13 has too little over north-east Brazil in both seasons.

Overall, RMSE values show a similar range of scores across perturbed parameter and multi-model ensemble members for surface air temperature (Fig. 3.8a). The identity of the worst performing GC3.05-PPE member is different in DJF and JJA. These show a widespread cold and warm bias, respectively. EC-EARTH has the largest RMSE values for both seasons in CMIP5-13, driven by a general cold bias in tropical and sub-tropical regions. For precipitation, several GC3.05-PPE members score better than any CMIP5-13 model in boreal winter and spring, while the range of scores is similar in boreal autumn. The GC3.05-PPE shows a somewhat wider range of scores in JJA, contributing both the best- and worst- performing members of the combined Strand 2 set. In GC3.05-PPE, the patterns of precipitation bias are similar in the best- and worst-performing members, being distinguished largely by different amplitudes of error (Fig 3.8c). Bias patterns in the best and worst CMIP5-13 members show somewhat greater variety (Fig. 3.8d), though most of the major features of the median bias pattern remain common to both.
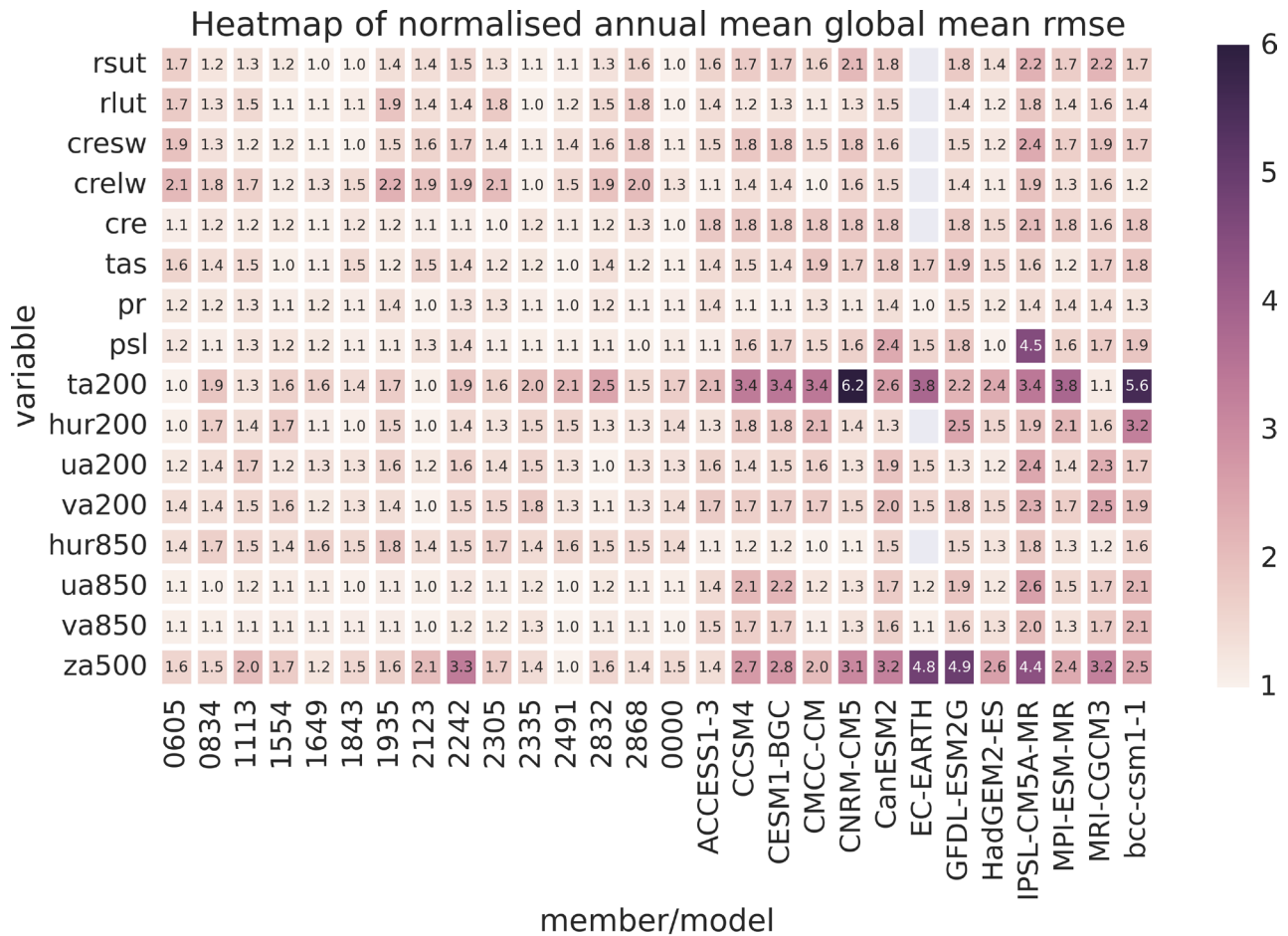
Heatmap of normalised annual mean global mean rmse

variable	0605	0834	1113	1554	1649	1843	1935	2123	2242	2305	2335	2491	2832	2868	0000	ACCESS1-3	CCSM4	CESM1-BGC	CMCC-CM	CNRM-CM5	CanESM2	EC-EARTH	GFDL-ESM2G	HadGEM2-ES	IPSL-CM5A-MR	MPI-ESM-MR	MRI-CGCM3	bcc-csm1-1
rsut	1.7	1.2	1.3	1.2	1.0	1.0	1.4	1.4	1.5	1.3	1.1	1.1	1.3	1.6	1.0	1.6	1.7	1.7	1.6	2.1	1.8		1.8	1.4	2.2	1.7	2.2	1.7
rlut	1.7	1.3	1.5	1.1	1.1	1.1	1.9	1.4	1.4	1.8	1.0	1.2	1.5	1.8	1.0	1.4	1.2	1.3	1.1	1.3	1.5		1.4	1.2	1.8	1.4	1.6	1.4
cresw	1.9	1.3	1.2	1.2	1.1	1.0	1.5	1.6	1.7	1.4	1.1	1.4	1.6	1.8	1.1	1.5	1.8	1.8	1.5	1.8	1.6		1.5	1.2	2.4	1.7	1.9	1.7
crelw	2.1	1.8	1.7	1.2	1.3	1.5	2.2	1.9	1.9	2.1	1.0	1.5	1.9	2.0	1.3	1.1	1.4	1.4	1.0	1.6	1.5		1.4	1.1	1.9	1.3	1.6	1.2
cre	1.1	1.2	1.2	1.2	1.1	1.2	1.2	1.1	1.1	1.0	1.2	1.1	1.2	1.3	1.0	1.8	1.8	1.8	1.8	1.8	1.8		1.8	1.5	2.1	1.8	1.6	1.8
tas	1.6	1.4	1.5	1.0	1.1	1.1	1.2	1.5	1.4	1.2	1.2	1.0	1.4	1.2	1.1	1.4	1.5	1.4	1.9	1.7	1.8	1.7	1.9	1.5	1.6	1.2	1.7	1.8
pr	1.2	1.2	1.3	1.1	1.2	1.1	1.4	1.0	1.3	1.3	1.1	1.0	1.2	1.1	1.1	1.4	1.1	1.1	1.3	1.1	1.4	1.0	1.5	1.2	1.4	1.4	1.4	1.3
psl	1.2	1.1	1.3	1.2	1.2	1.1	1.1	1.3	1.4	1.1	1.1	1.1	1.1	1.0	1.1	1.1	1.6	1.7	1.5	1.6	2.4	1.5	1.8	1.0	4.5	1.6	1.7	1.9
ta200	1.0	1.9	1.3	1.6	1.6	1.4	1.7	1.0	1.9	1.6	2.0	2.1	2.5	1.5	1.7	2.1	3.4	3.4	3.4	6.2	2.6	3.8	2.2	2.4	3.4	3.8	1.1	5.6
hur200	1.0	1.7	1.4	1.7	1.1	1.0	1.5	1.0	1.4	1.3	1.5	1.5	1.3	1.3	1.4	1.3	1.8	1.8	2.1	1.4	1.3		2.5	1.5	1.9	2.1	1.6	3.2
ua200	1.2	1.4	1.7	1.2	1.3	1.3	1.6	1.2	1.6	1.4	1.5	1.3	1.0	1.3	1.3	1.6	1.4	1.5	1.6	1.3	1.9	1.5	1.3	1.2	2.4	1.4	2.3	1.7
va200	1.4	1.4	1.5	1.6	1.2	1.3	1.4	1.0	1.5	1.5	1.8	1.3	1.3	1.4	1.7	1.7	1.7	1.7	1.7	1.5	2.0	1.5	1.8	1.5	2.3	1.7	2.5	1.9
hur850	1.4	1.7	1.5	1.4	1.6	1.5	1.8	1.4	1.5	1.7	1.4	1.6	1.5	1.5	1.4	1.1	1.2	1.0	1.1	1.5	1.5		1.5	1.3	1.8	1.8	1.7	1.6
ua850	1.1	1.0	1.2	1.1	1.1	1.0	1.1	1.0	1.2	1.1	1.2	1.0	1.2	1.1	1.1	1.4	2.1	2.2	1.2	1.3	1.7	1.2	1.9	1.2	2.6	1.5	1.7	2.1
va850	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.0	1.2	1.2	1.3	1.0	1.1	1.1	1.1	1.5	1.7	1.7	1.1	1.3	1.6	1.1	1.6	1.3	2.0	1.3	1.7	2.1
za500	1.6	1.5	2.0	1.7	1.2	1.5	1.6	2.1	3.3	1.7	1.4	1.0	1.6	1.4	1.5	1.4	2.7	2.8	2.0	3.1	3.2	4.8	4.9	2.6	4.4	2.4	3.2	2.5

member/model

Figure 3.9. Normalised root-mean-squared (r.m.s) errors in global, annual spatial fields of a variety of climate variables from the 28 Strand 2 simulations, averaged over 1981-2000. Members of CMIP5-13 are named (see Table 3.1) and members of GC3.05-PPE are referred to by their numerical run identifiers. The STD member is 0000. The variable acronyms are listed in Table E.1, numbers representing atmospheric pressure levels. Verifying observational datasets (see Table E.2) are CERES (for rsut and rlut), ERA-Interim (cresw, crelw, cre, tas, ta200, hur200, ua200, va200, hur850, ua850, va850, za500) and HadSLP2 (psl). The observational climatologies cover 1981-2000 apart from CERES, for which 2001-2005 is used because data for the full verification period do not exist. The scores are normalised by the r.m.s. error for the best-performing simulation for a given variable, which therefore possesses a value of 1.0. For some variables, more than one simulation has a value of 1.0, since values are recorded to one decimal place. The plot is presented as a "heatmap", in which simulations with the highest normalised errors, and therefore the worst performance, are shown as the darkest shades of purple. Some entries are missing for EC-EARTH, due to unavailability of data.

In Figure 3.9 we provide a broader statistical picture of global model performance across the combined set of 28 Strand 2 simulations. This is derived from a set of variables similar to those used in the filtering of the GC3.05 parameter space in section 3.3. The evaluation is based on RMSE values for annual, global spatial fields for climatological averages for 1981-2000. We consider surface air temperature, precipitation, sea-level pressure, outgoing shortwave and longwave radiation at the top of the atmosphere, total, shortwave and longwave cloud radiative effect, 500hPa geopotential height, and atmospheric temperature, relative humidity and zonal and meridional wind components at the 850hPa and 200hPa pressure levels. The RMSE values are normalised by that of the best-performing simulation for the relevant variable, which therefore receives a normalised score of 1.0. High values indicate relatively poor performance, denoted by darker shading in the matrix of values.

In general, different Strand 2 members provide the best levels of performance for different variables, and no single member stands out as clearly superior to the other simulations. The worst scores for individual variables (values of normalised RMSE exceeding 3.0) belong to CMIP5-13 members, with the exception of one score exceeding 3.0 in a GC3.05-PPE member for the 500hPa geopotential height field. Several CMIP5-13 models also show poor scores for this variable, as well as for upper tropospheric temperature. For circulation variables, GC3.05-PPE members perform well in comparison to CMIP5-13 simulations. This is also the case for reflected shortwave radiation and total cloud radiative effect. However, the GC3.05-PPE scores for the longwave component of cloud radiative effect are typically worse in GC3.05-PPE: Nine members exceed a value of 1.6, compared to only one CMIP5-13 member. This is due to a widespread negative bias, present in most of the GC3.05-PPE members. These underestimates are generally offset by positive biases in the shortwave component, explaining the smaller errors seen in total cloud radiative effect. The ranges of scores for surface air temperature and precipitation show substantial overlap between the two ensembles, as also shown in Figs. 3.8a,c.
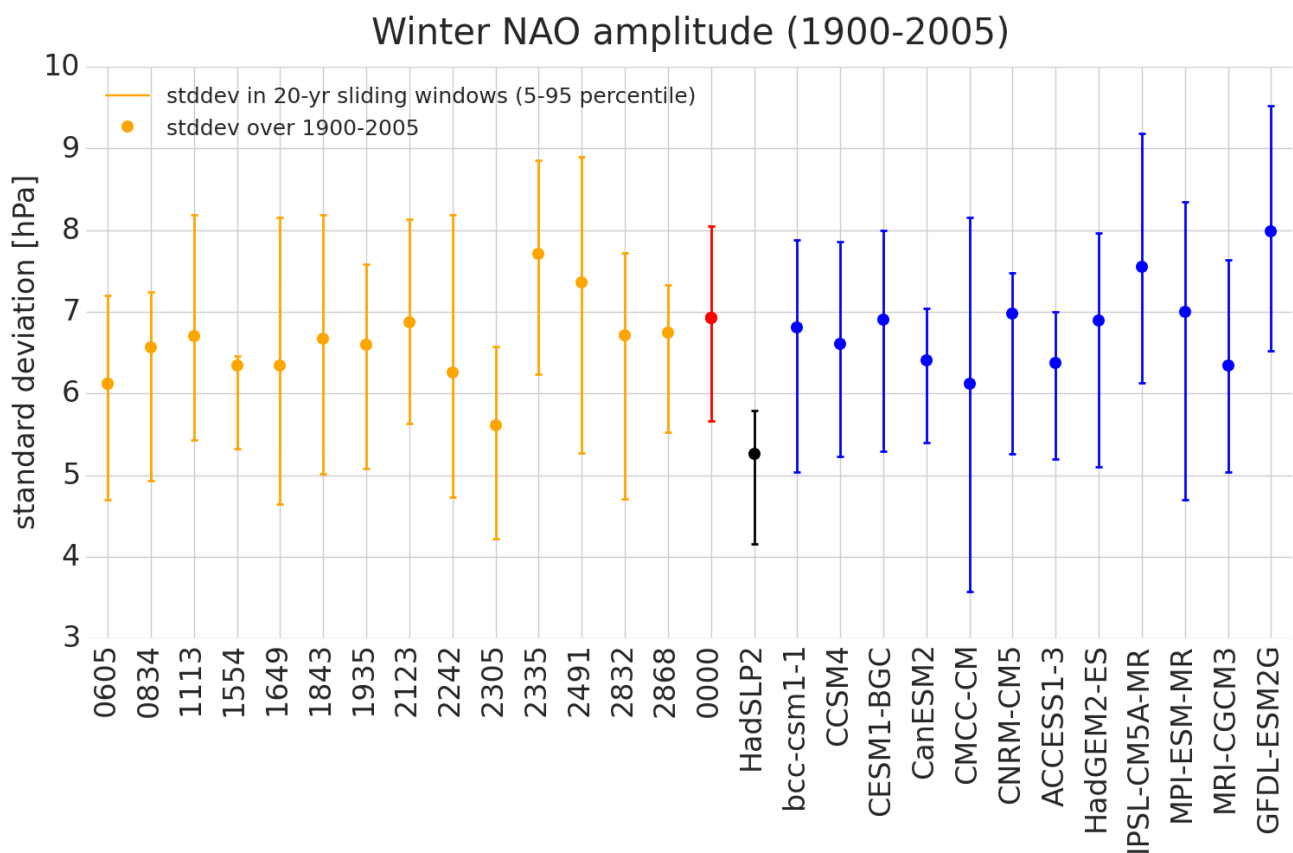
Figure 3.10. Circles show standard deviations (hPa) of linearly detrended annual values of the winter North Atlantic Oscillation index of Hurrell (1995), calculated over 1900-2005. Values are shown for the 15 GC3.05-PPE members (orange, with STD in red), twelve of the CMIP5-13 members (blue, excluding EC-EARTH as data for this model was only available from 1950), and observations (black, from the HadSLP2 dataset, Table E.2). The whiskers show 5th and 95th percentiles of the distribution of standard deviations obtained by calculating separate values for all consecutive 20 year periods between 1900 and 2005. This provides an estimate of variations in the strength of interannual NAO variability during each time series.

The NAO is key driver of interannual variability in wind, temperature and precipitation over Europe in winter. Figure 3.10 compares simulated and observed values of the NAO index (Hurrell, 1995) against observations. Circles give the standard deviation of annual values over 1900-2005, and whiskers show the 5th-95th percentile spread of standard deviations from alternative 20-year periods within each time series. This spread measures the extent to which the strength of the NAO on the interannual time scale[36] varies during the 20th century. All of the Strand 2 simulations overestimate the long-term average amplitude of NAO. The degree of overlap between the uncertainty ranges for the simulated and observed values is substantial for some Strand 2 members, partly because the ranges are generally larger in the simulations than in the observations. However, one GC3.05-PPE member and two CMIP5-13 members show no overlap, and are therefore significantly biased in their simulations of NAO.

In section 3.4b we show that intraseasonal variability of weather types related to the NAO compare quite well with observations (see Fig. 3.19 and related discussion). Therefore, the effects of biases in simulated characteristics of the NAO are likely to vary according to the time scales of interest in different impacts applications: Biases can be quite different on intraseasonal, interannual or interdecadal time scales, and will need to be evaluated on a case-by-case basis.

A similar analysis was carried out for the Atlantic Multidecadal Oscillation (AMO) index (Trenberth and Shea, 2006). The AMO is a cycle of multidecadal variability in North Atlantic SSTs that influences temperature and precipitation anomalies in many parts of the northern hemisphere. In its warm phase, it drives wet summers over northern Europe (including the UK), and hot, dry summers over southern Europe (Sutton and Dong, 2012). A majority of simulations simulate a long-term average AMO amplitude close to the observed value. However, six CMIP5-13 members simulate excessive AMO variability, and in one GC3.05-PPE member the variability is too weak.

---

[36] While the spread metric in Fig. 3.10 shows uncertainties in the *interannual* variability of the NAO, it does not measure variability in the phase of the NAO on *multidecadal* time scales.

Nino3.4 SSTA amplitude (1900-2005)
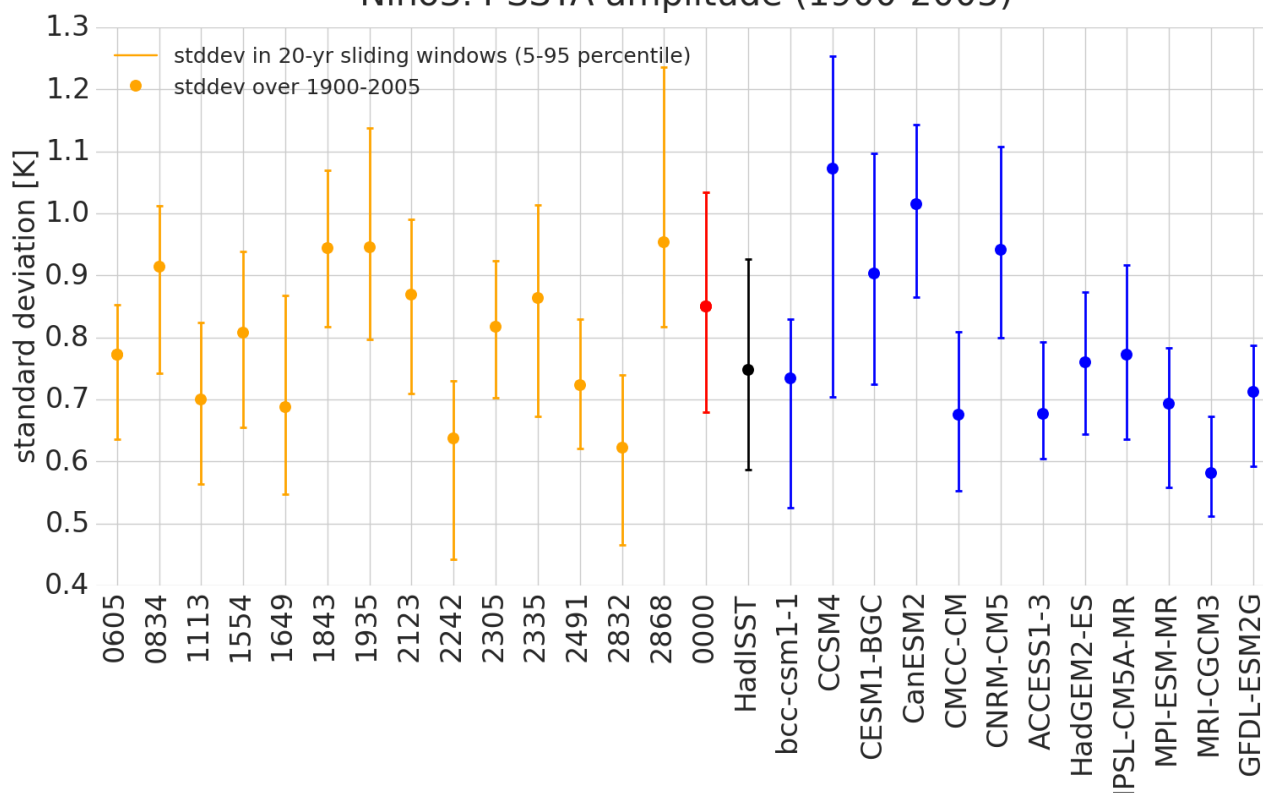
Figure 3.11. As Figure 3.10, for values of the Niño 3.4 Index (°C), a standard measure of the strength of El Niño events defined as the average SST anomaly for the region 5°N-5°S, 120-170°W (Trenberth, 1997). In this case, monthly anomalies of Niño 3.4 were linearly detrended, to provide 12x105 values from which the standard deviation (circles) of each full time series was calculated. The 5th and 95th percentile whiskers show the range of standard deviations obtained by pooling the 240 monthly anomalies in each consecutive 20 year period between 1900 and 2005.

The strength of El Niño events is assessed in Fig. 3.11, using the standard deviation of monthly SST anomalies in the Niño 3.4 region of the equatorial Pacific Ocean (Trenberth, 1997). For the majority of simulations, the average amplitude for 1900-2005 lies within the distribution of observed amplitudes found for alternative 20-year periods in the historical record. However, for seven simulations this is not the case. CCSM4, CanESM2, CNRM-CM5 and three GC3.05-PPE members simulate a long-term amplitude that is too strong, while in MRI-CGCM3 it is too weak.

Figure 3.12 compares the observed record of annual variability and change in GMST against the Strand 2 simulations, from 1900-2017. The model values are constructed to match as closely as possible the observed dataset (HadCRUT4, see Table E.2). This is done by blending SSTs and land surface air temperatures, and applying a time-varying mask to reflect the changing coverage of observations[37]. Estimates of simulated warming since about 1980 are typically reduced by ~0.2°C in such blended and masked estimates, compared to GMST values derived from a simple global average of surface air temperature values (Richardson et al., 2016).

---

[37] The model data is regridded to the HadCRUT4 grid, using SST to regrid to ocean points, and 1.5m temperature to regrid to land points. Then, SST values are set to missing where sea ice is greater than 10%; for HadCRUT4, this was done using the sea ice from HadISST1 (see Table E.2) regridded to the HadCRUT4 grid. To match the HadCRUT4 coverage month by month, model values are set to missing at grid points where there are no observed values. Then a monthly climatology for Dec 1960 to Nov 1990 is estimated at each grid point, requiring at least 20 of the 30 years to be present, otherwise the climatological value is set to missing. Monthly anomalies relative to 1961-90 are estimated by removing the relevant monthly climatology. Anomalies are set to missing in areas where there is no climatological value, which includes most of the Arctic, Southern Ocean, and Antarctica. Seasonal means are calculated when at least two of the three months are present, and then all four seasonal means must be non-missing to make an annual mean. In Figure 3.12, each anomaly time series is re-centred by removing the average for the baseline period of 1901-50.

Interannual variability in individual GC3.05-PPE and CMIP5-13 simulations (shown for the GC3.05-PPE members in Fig. 3.12) is qualitatively similar to observations. Also plotted in Fig. 3.12 is the control simulation for STD, obtained by extending its spin-up integration (see section 3.3e). This simulation (the magenta line labelled CTL_STD, blended and masked as described above[37]) provided an additional illustration of internal variability, and also of potential climate drift, which may explain the warming of ~0.2°C seen in CTL_STD during the first 20 years (see section 3.3e).

After 1920, the longer-term changes in observations show a slight warming, followed by a period of slight cooling from ~1950-1980 and then a marked warming trend in recent decades. This was punctuated by a pause between 1998 and 2014, termed the "warming hiatus" (e.g. Trenberth, 2015). Periods of temporary cooling following major volcanic eruptions (notably Agung in 1963, El Chichon in 1981 and Mt Pinatubo in 1991) are also evident. IPCC AR5 concluded that it is certain that global mean surface temperature has increased since the late 19th century, and extremely likely[38] that human activities caused more than half of the warming between 1951 and 2010 (Stocker et al., 2013).

All but one of the model simulations produce a significant warming by the end of the historical period. The exception is a GC3.05-PPE member that simulated an unrealistic cooling trend in northern hemisphere surface air temperature, but was retained because it did not fail any of the other screening criteria (section 3.3f). By the 2010s, most GC3.05-PPE members simulate a warming similar to that of STD, and close to the observed warming. During the period 1960-2000 most GC3.05-PPE members simulate long-term average values (not shown) that are within the CMIP5-13 range, but at the lower end. These values are cooler than the baseline period (1901-1950 in Fig. 3.12) and in contrast to observations. This is likely to be due mainly to the effects of relatively strong cooling due to anthropogenic aerosol forcing, compared to the range given by AR5 (section 3.3d).

---

[38] Uncertainties in AR5 were assessed using likelihood statements based on probability ranges derived from statistical analysis and expert judgement (Stocker et al., 2013). *Extremely likely, very likely* and *likely* correspond to probabilities of at least 95%, 90% and 66%, respectively.

Global temperature change
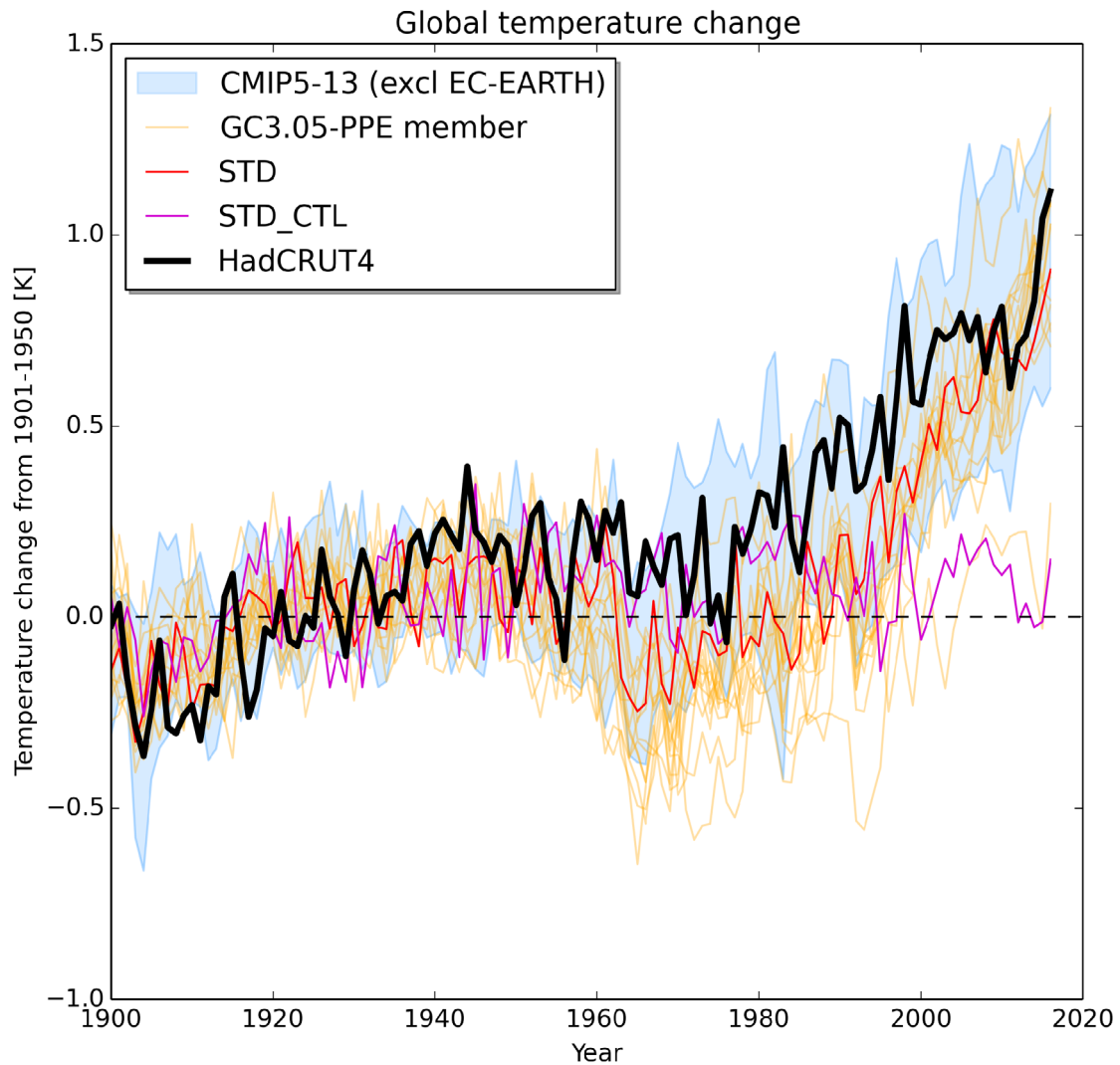
Figure 3.12. Historical annual average anomalies relative to 1901-1950 of masked and blended global mean surface temperature from observations (HadCRUT4, black), and the Strand 2 simulations. Blue shading shows the range of values from 12 of the CMIP5-13 models (EC-EARTH is omitted as data was not available prior to 1950). Values from the 15 GC3.05-PPE simulations are shown individually in red (STD) and orange (other members). The purple curve (STD_CTL) shows the control simulation of STD[31], created by extending its spin-up simulation as described in section 3.3e. This shows a warming of ~0.2°C between 1900 and 1920, suggesting that climate drift may be at least partly responsible for the warming seen in STD during this period. Control simulations for other GC3.05-PPE simulations are not available. Observed global mean values are constructed from monthly values of SST over oceans, and surface air temperature over land, with time-varying gaps in data coverage. Model values are constructed to match the observed product as closely as possible[37].

The global average aerosol forcing (in climate models and observations alike), and its impact on historical changes in GMST, is the net result of multiple species, processes and interactions. Therefore, it should not be assumed that a simulation that matches the observed GMST record relatively well will necessarily represent specific aerosol processes better than a simulation which contains a larger bias in GMST. For example, Malavelle et al. (2017) used a large recent volcanic eruption in Iceland (Holuhraun, 2014-15) to study the ability of climate models to simulate aerosol effects on cloud properties. One of these experiments used GLOMAP-mode with standard parameter settings, embedded in a development version of GA7 (Walters et al., 2017), the atmospheric component of GC3.0. This model was therefore similar to STD in its atmosphere and aerosol configuration, and gave a realistic simulation of the radiative effects of aerosol-cloud interactions following the eruption. In contrast, the cloud impacts were substantially overestimated in a parallel experiment in which GLOMAP-mode was replaced by the simpler aerosol module included in HadGEM2-AO.

Despite the cautions explained above, the extent to which climate models replicate observed changes in GMST (and associated large-scale patterns of surface temperature) is a relevant test of their credibility as tools for projecting future changes. For example, we include such changes as one of a set of formal observational constraints in the probabilistic projections of Strand 1. However, the extent to which biases in historical changes are linked to errors in future responses is limited by several sources of uncertainty. These include:

- Historical forcing, particularly related to aerosols (e.g. Aldrin et al., 2012).

- The efficacy[39] of different forcing agents in driving a surface temperature response: Shindell (2014) showed that aerosol and ozone changes are more effective than CO_2 in driving a surface temperature response, implying that future changes in the relative importance of these agents may limit the usefulness of historical changes as a guide to the future.

- Temporal dependence in the strength of climate feedbacks related to patterns of SST change, which may differ in the future from those observed since the 19th century (Gregory and Andrews, 2016).

- The role of ocean heat uptake in delaying the full response to previous changes in forcing (e.g. Appendix B).

- The role of internal variability in the observed record (Olson et al., 2013), allied to challenges in disentangling natural and anthropogenically-forced contributions to variability on multidecadal time scales (e.g. Booth et al., 2012). For example, if the warming hiatus was mainly due to internal variability, then we would expect the observations to show less warming than most of the simulations during the relevant period, and indeed the Strand 2 runs warm faster during 1998-2014.

In GC3.05-PPE, the simulated cooling between 1930-50 and 1970-90 is found to be uncorrelated with projected future warming trends beyond 2050. This is because the model parameters that control the strength of the historical cooling (which belong to the aerosol component) are not important drivers of the long-term future warming. This is not surprising, since greenhouse gas concentrations in the RCP8.5 scenario increase steadily during the 21st century (e.g. Fig. 3.7), whereas emissions of aerosol precursors reduce substantially[40]. Therefore, the historical aerosol-driven biases in Fig. 3.12 may not be informative about the credibility of projected future changes on the centennial time scale. For example, the strong aerosol cooling may have masked the response to greenhouse gas forcing during the 20th century, but such masking would reduce substantially during the 21st century, as aerosol emissions are assumed to reduce in the RCP8.5 scenario.

However, the mid-20th century cooling may be more informative regarding near-term future GMST changes, than it is about longer-term levels of warming. Assuming that the historical aerosol forcing in GC3.05-PPE is too strong in most members, it also follows that the effects of future reductions in aerosol forcing, which may be expected to enhance projected warming trends during the next few decades (Raes and Seinfeld, 2009), are likely to be overestimated.

---

[39] Efficacy is a measure of how effectively radiative forcing from a given driver changes the equilibrium global mean surface temperature, compared to an equivalent radiative forcing from CO_2 that has an efficacy of 1.

[40] For example, anthropogenic sulphur dioxide emissions in the RCP8.5 scenario reduce from 114 Teragrams (Tg) per year in 2005, to 52 $TgSO_2$/yr by 2050 and 26 $TgSO_2$/yr by 2100.

Overall, the assessment of large-scale characteristics presented above provides support for use of the Strand 2 simulations in impacts assessments. However, the presence of member-specific biases in aspects of regional climatology, or simulations of historical climate change, underlines the importance of considering bias correction strategies (Fung, 2018). In such approaches, statistical postprocessing techniques such as simple adjustments of climate averages or quantile-mapping of daily or seasonal distributions (e.g. Piani et al., 2010) can be calibrated using differences between simulations and observations of historical climate, and then applied to future projections. However, it is important to select an appropriate technique on an application-specific basis, and consider carefully its impact and limitations. Use of bias corrections cannot remove a lack of credibility that may be associated with large model errors, and is only recommended when relevant earth system phenomena are captured reasonably well by the chosen set of climate models or model variants. Such phenomena might include the location of and variability in large scale circulation features affecting the region of interest, spatiotemporal characteristics of associated regional responses in variables such as precipitation and surface temperature, and sufficient resolution of local climate change feedbacks at the spatial scale of the relevant application (Maraun et al., 2017).

b. Regional performance for UK and Europe

Figures 3.13a and b show ensemble-mean biases for European surface air temperature and precipitation, during 1981-2000. Verifying observations are taken from E-OBS datasets (Table E.2). There is a widespread cold bias for GC3.05-PPE in winter (Fig. 3.13a), the largest errors (of -5 to -10ºC) occurring mainly over Scandinavia, with values in the range 0 to -3 ºC typical over Southern, Western and Central Europe. These errors are the regional manifestation of the general cold bias found over northern hemisphere continents in DJF. Drivers of the ensemble-mean bias include the strong aerosol forcing and negative biases in long-wave cloud radiative forcing found in most GC3.05-PPE members (Fig. 3.8 and related discussion). CMIP5-13 also shows an ensemble-mean cold bias over Scandinavia. However, it is considerably smaller than that of GC3.05-PPE, and CMIP5-13 shows a modest warm bias over most of Central, Southern and Eastern Europe. In DJF, European RMSE values for most GC3.05-PPE members (not shown in Fig. 3.13a) are above the range for CMIP5-13 members. However, this is not the case in other seasons. In summer, both Strand 2 ensembles show warm biases in the range 2-3ºC over parts of South-East Europe, extending northward to the Baltic States in CMIP5-13. A cold bias over Sweden and Finland is also common to both ensembles, however regional biases are smaller than ±2ºC in most locations.

Ensemble-mean precipitation errors reveal widespread wet biases in both Strand 2 ensembles in winter (Fig. 3.13b). The largest errors exceed 100% of the observed 1981-2000 value, errors of this magnitude being more prevalent in CMIP5-13 than in GC3.05-PPE. Wet biases exceed 40% in many regions, and are therefore comparable to, or larger than, typical changes expected during the 21st century (e.g. Jacob et al., 2014).
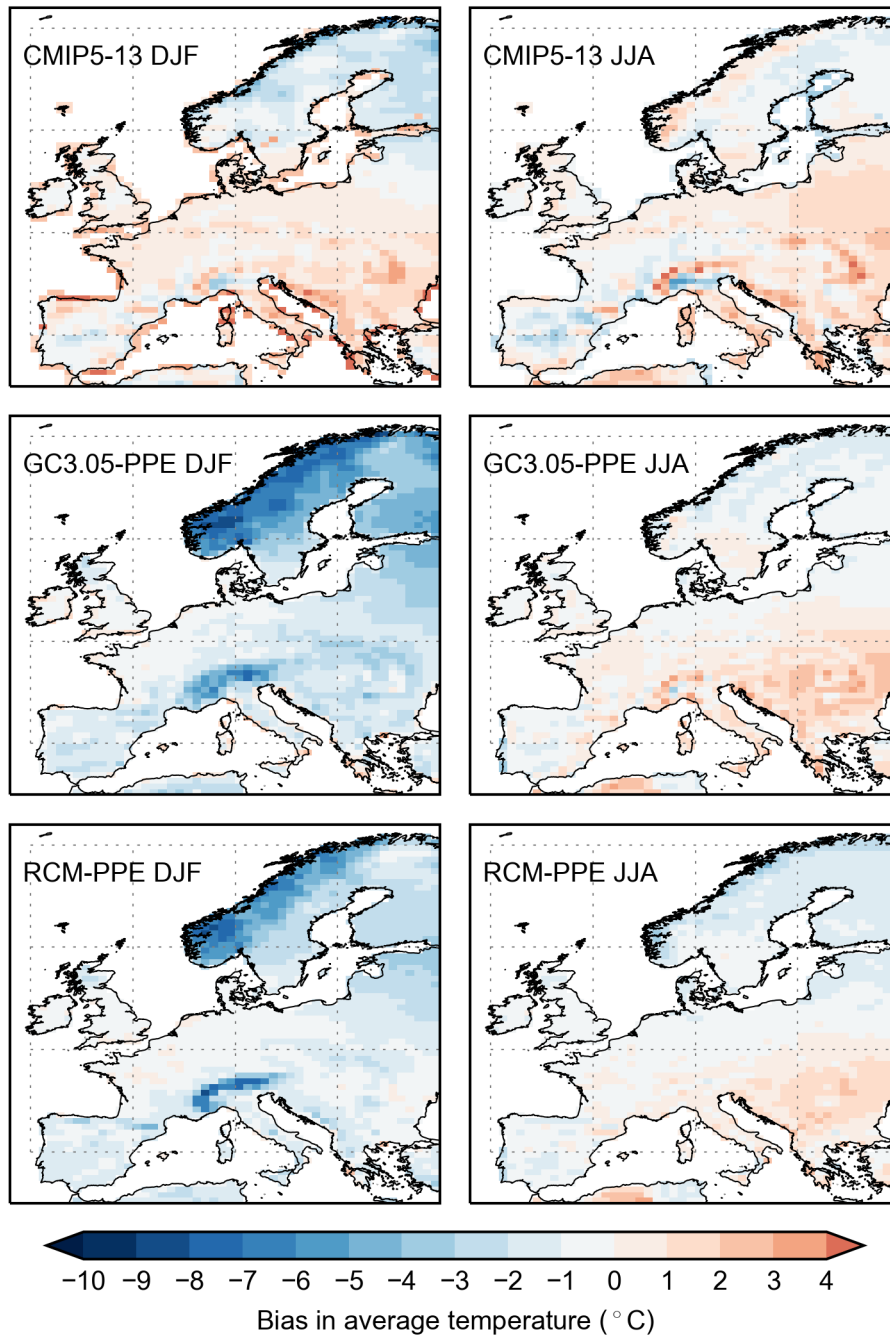
Figure 3.13a. Ensemble-mean biases in climatological averages of surface air temperature (ºC) over Europe, for 1981-2000, in winter (left) and summer (right), for the Strand 2 and 3 simulations. CMIP5-13 multi-model ensemble (top row), GC3.05-PPE (middle row), RCM-PPE (bottom row). Observed values are taken from the E-OBS dataset (Table E.2).
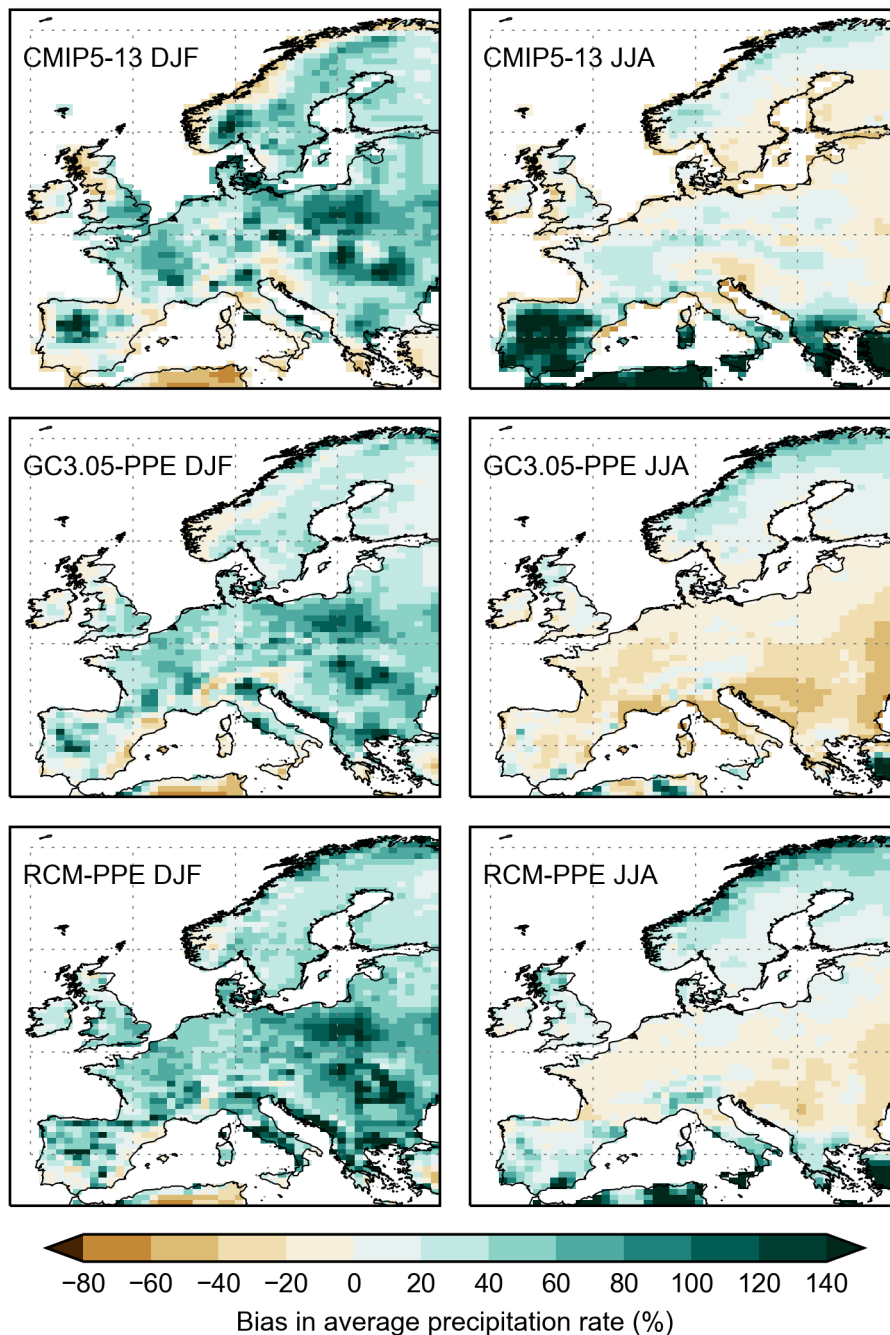
Figure 3.13b. As Fig. 3.13a, for precipitation biases, expressed as a percentage of the observed value.

Another consideration is that rain gauge observations underestimate the true level of observed precipitation due to undercatch associated with multiple factors, including the fraction of precipitation falling as snow, wind speeds and the exposure of the gauge. Biases vary with season and location, and are largest in winter, and at high elevations (e.g. Frei et al., 2003). Some authors (Kotlarski et al., 2014; Rajczak and Schär, 2017) increase observed values by 20% to estimate the effects of undercatch. We do not apply such adjustments in this report. However, we note that measurement biases provide an important caveat to our estimates of systematic biases in simulated precipitation (Figs. 3.13b, 3.14b, 3.15, and 4.1-4.6), and represent an additional source of uncertainty in the context of bias correction methods. For example, the GPCP dataset (used for global evaluation in Fig. 3.8c,d) shows larger precipitation values than E-OBS over Europe, apart from in high elevation regions. Over parts of Northern and Eastern Europe, the differences between the two verifying datasets in winter (typically ~40%) are large enough to change the sign of the diagnosed biases, which are dry relative to GPCP in Fig. 3.8c, but wet relative to E-OBS in Fig. 3.13b.

In Fig. 3.13b dry biases are limited in extent, but occur in some western coastal regions in CMIP5-13, and in islands and some coastal fringes of the Mediterranean Sea (both Strand 2 ensembles). In summer, CMIP5-13 shows large wet biases (100% or greater) over the Iberian Peninsula, Southern Italy, the Balkans, Greece and Turkey. Several CMIP5-13 models show substantial wet biases over the Iberian Peninsula, so the ensemble-mean bias is not due to the influence of outliers. Note, however, that since summer precipitation levels are typically low in these regions (e.g. below 0.5mm/day in Southern Spain and Portugal, in E-OBS), the large percentage biases do not represent especially large errors in absolute precipitation amounts.

The GPCP dataset shows 30-50% more summer rainfall than E-OBS in this region, underlining again the verification uncertainties discussed above. GC3.05-PPE produces a dry bias of varying magnitude over most of Central, Southern and Eastern Europe, reaching 60-80% in South-Eastern parts of the domain. Both ensembles simulate a wet bias over much of Scandinavia, the largest values (>40%) occurring over the Norwegian mountains.

The errors discussed above, particularly the larger cold and wet biases found in winter, motivates consideration of bias-correction techniques (Fung, 2018). This would apply particularly in applications sensitive to absolute levels of future surface temperature or precipitation. However, use of bias correction methods is subject to development of understanding of the sources of model error, particularly in regions where these are large (see Maraun et al. (2017) and discussion in section 3.4a). In practice, decisions on whether or how to perform bias corrections will be application-specific, and it may also be important to consider whether a subset of the Strand 2 projections may need to be excluded from some studies, given that biases in particular variables vary between the different simulations.

Figure 3.14a shows ensemble-mean biases in the standard deviation of seasonal average surface air temperature during 1981-2000, a basic measure of interannual variability. During this period, the CMIP5-13 error pattern reveals smaller than observed variability over most of Europe in winter, the largest errors of 1.0ºC or more occurring over Scandinavia. In this region GC3.05-PPE also simulates less variability than is seen in observations, although biases are smaller than in CMIP5-13. The negative biases over Northern Europe occur despite a systematic overestimation in both ensembles of the strength of the NAO (Fig. 3.10), suggesting that other aspects of circulation variability or thermodynamic processes may play important roles in the bias. In summer, ensemble-mean biases show greater than observed variability in most regions, in both GC3.05-PPE and CMIP5-13. The GC3.05-PPE results show larger biases (0.6 ºC or more) over much of Central Europe, while CMIP5-13 has larger biases over Northern Europe and Spain.
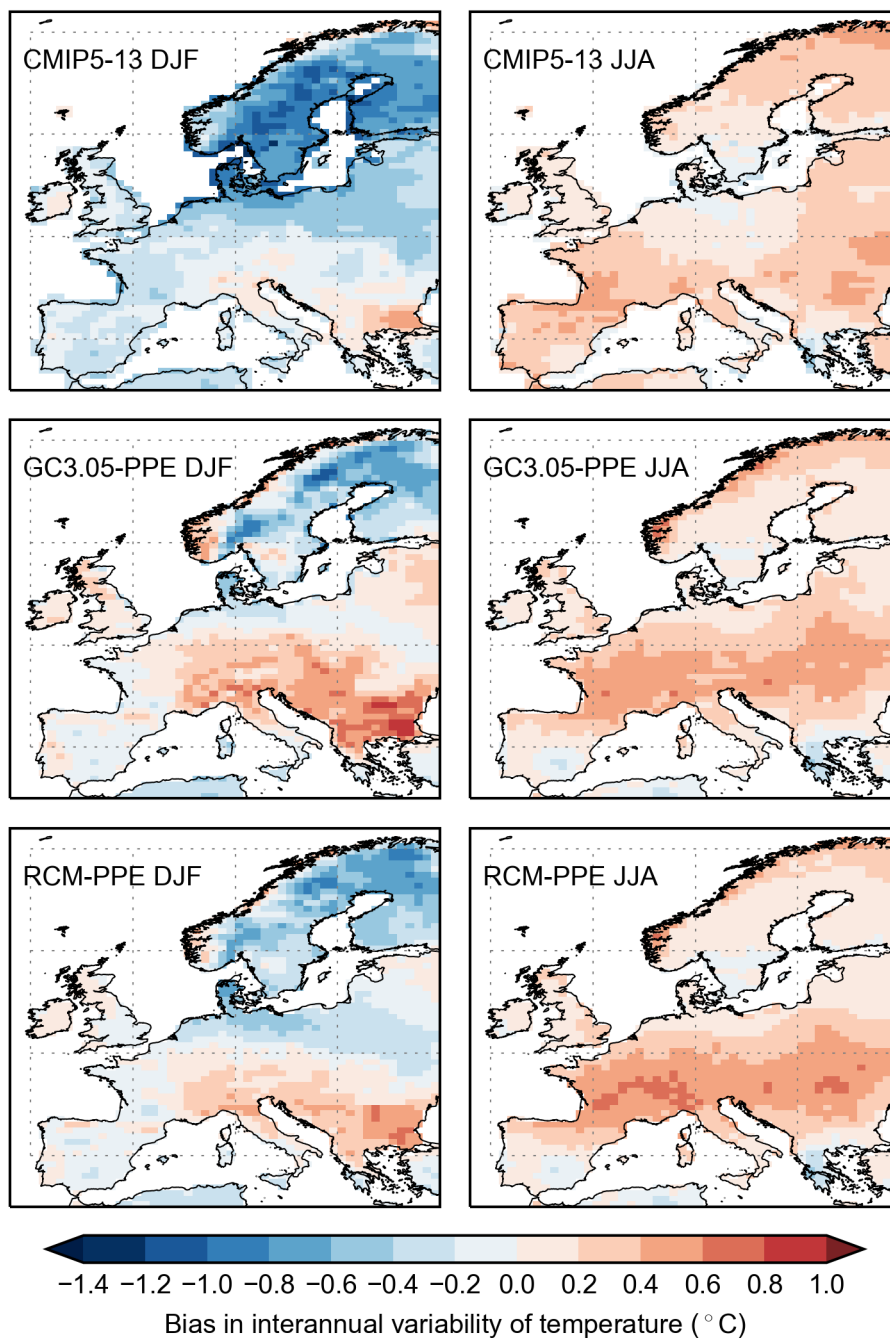
Figure 3.14a. Ensemble-mean biases in the interannual variability of surface air temperature (ºC) over Europe during 1981-2000, in winter (left) and summer (right), for the Strand 2 and 3 simulations. Here, interannual variability is defined as the standard deviation of seasonal averages. CMIP5-13 multi-model ensemble (top row), GC3.05-PPE (middle row), RCM-PPE (bottom row). Observed variability is calculated from the E-OBS dataset.

Biases in interannual variability of precipitation, expressed as a percentage of observed variability, are shown in Fig. 3.14b. The CMIP5-13 ensemble mean underestimates observed variability by 40% in some coastal regions of Western Europe (despite the strong NAO variability noted above), including parts of Ireland and Scotland. Here, biases in GC3.05-PPE are smaller, however both ensembles show excessive variability across much of Central and Eastern Europe. In summer, ensemble-mean variability in CMIP5-13 is 40% or more too small over much of North-Western Europe, whereas GC3.05-PPE shows a balance of positive and negative regional biases. Both ensembles share a positive bias over Germany, while biases over the Iberian Peninsula are positive in CMIP5-13 and negative in GC3.05-PPE. The errors in variability shown in Figs. 3.14a,b imply that bias correction techniques that account for errors in climate variability (e.g. Ho et al., 2012), as well as the climatological mean state, may be appropriate in some applications (see Fung, 2018).
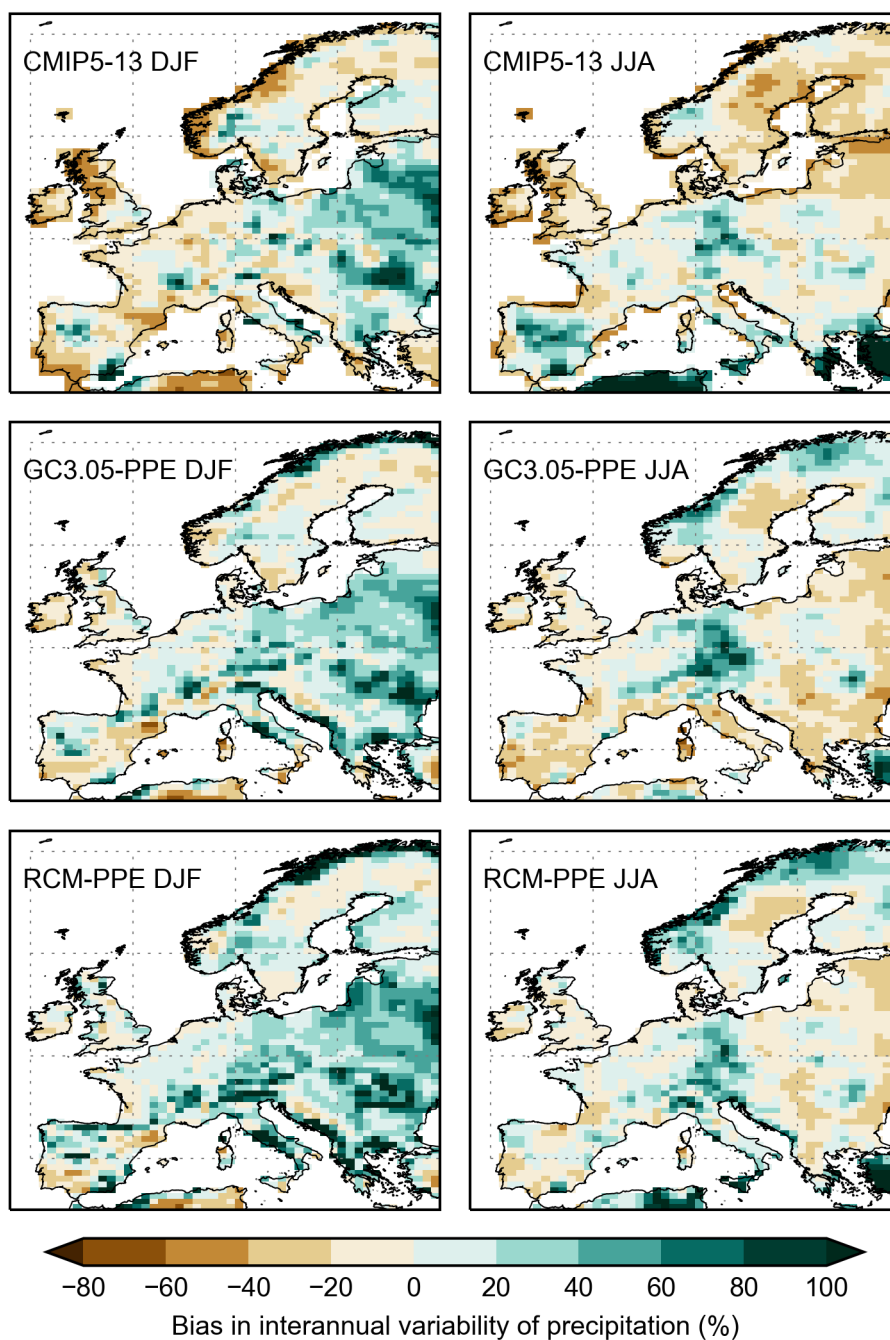
Figure 3.14b. As Fig. 3.14a, for biases in the interannual variability of precipitation, expressed as a percentage of the observed value.

Seasonal cycles of UK average surface air temperature and precipitation are shown in Figure 3.15 (left panels). While most GC3.05-PPE members show a cold bias in winter, none exceeds -2ºC, despite the larger biases seen in Northern Europe. During summer, biases in GC3.05-PPE members are relatively small, and are distributed around the two observed estimates provided (from E-OBS and the National Climate Information Centre (NCIC, Perry et al., 2009)). One CMIP5-13 model (CMCC-CM) shows a larger winter cold bias than any GC3.05-PPE member, and also shows the largest cool bias during summer and autumn. In most months of the year the largest warm biases, typically ~2 ºC, also come from CMIP5-13 members.
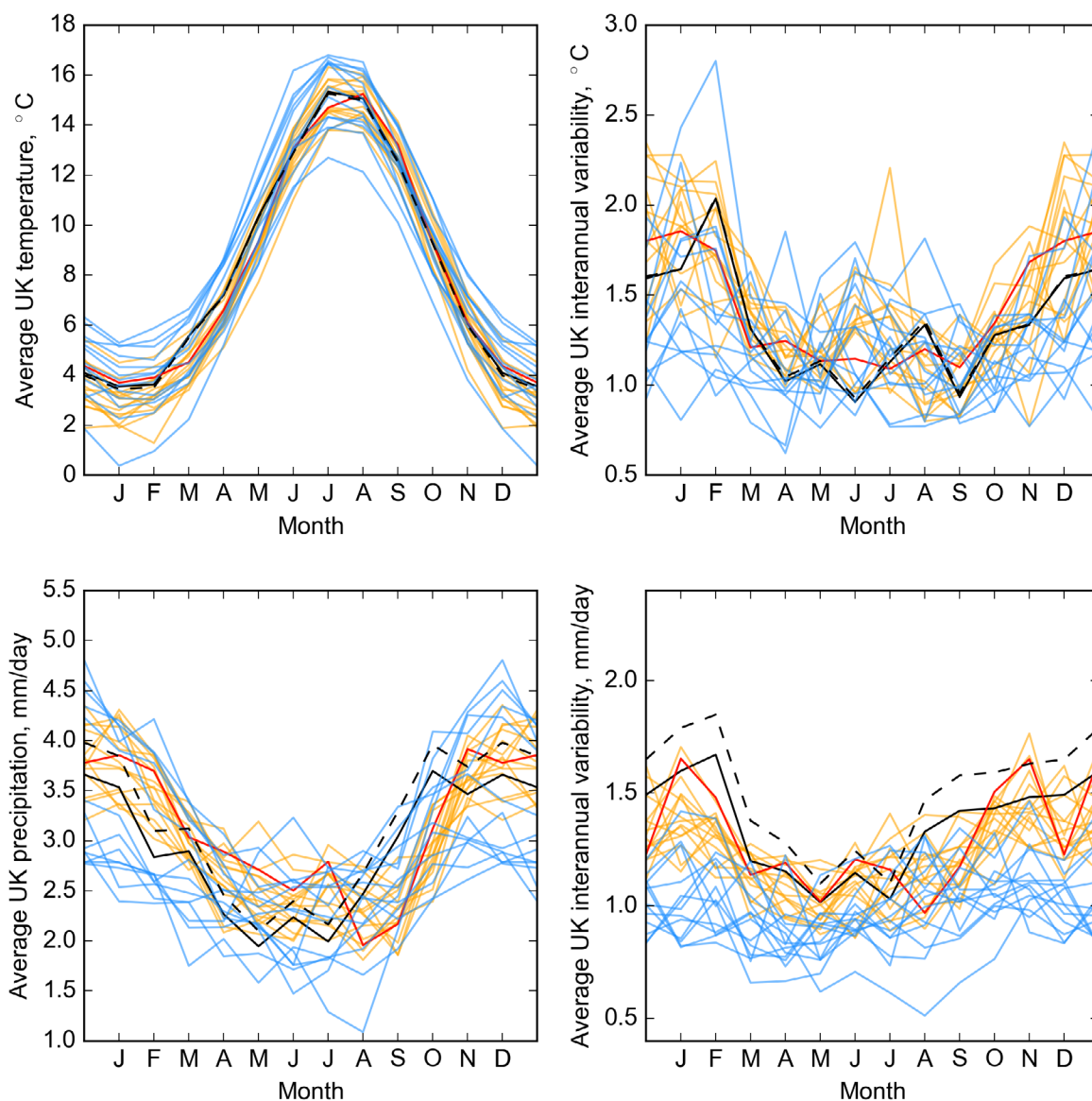
Figure 3.15. Left panels show seasonal cycles of UK average surface air temperature (ºC, top) and precipitation (mm/day, bottom) for 1981-2000. Right panels show UK averages of local values of interannual variability (defined as the standard deviation of seasonal averages during 1981-2000), for surface air temperature (top) and precipitation (bottom). Members of CMIP5-13 and GC3.05-PPE are shown in blue and orange respectively, with STD in red. Observations from the National Climate Information Centre (NCIC) and E-OBS datasets (Table E.2) are shown as dashed and solid black lines, respectively. All datasets were regridded to the GC3.05 (~60km) grid, prior to calculation of these diagnostics.

Average precipitation is slightly higher in the NCIC dataset compared with E-OBS, the largest differences (approaching 0.5 mm/day) occurring in winter. While the differences provide an estimate of observational uncertainty, the NCIC values are likely to be more reliable. This is because they are based on a much denser network of UK stations (approximately 3000) compared to the 138 used in E-OBS, which gives lower values in high-elevation regions, particularly in south-west England, Wales, the Lake District and North-East Scotland (Hofstra et al., 2009). The spread of GC3.05-PPE values encompasses NCIC in all months except September and October, in which all members show a dry bias (typical values ~ -1 mm/day). Although most members show a wet bias relative to E-OBS, this is not the case relative to NCIC. During extended winter (October-March), one subset of five CMIP5-13 models shows positive biases, while a second subset of six members shows significant dry biases. In summer, one CMIP5-13 model (BCC-CSM1.1) shows a substantial dry bias.

Seasonal cycles of interannual variability are shown in the right panels of Fig. 3.15. These are UK averages of local values on the ~60km grid of GC3.05, to which all data was regridded prior to calculation of the diagnostics. Both Strand 2 ensembles show biases distributed around the observations for temperature variability, although several CMIP5-13 models show negative biases of 0.5ºC or more during extended winter. For precipitation, both ensembles simulate too little variability, especially in comparison to NCIC which gives slightly higher values than E-OBS. However, biases are typically smaller in GC3.05-PPE members. Errors in CMIP5-13 members often exceed 0.5 mm/day, with the exception of May-July.

In addition to surface temperature and precipitation, the credibility of simulated hydrological characteristics of the land surface is an important consideration for climate impact assessments, particularly in relation to drought and agriculture. As an example, Figure 3.16 (lower left panel) shows annual cycles of soil moisture in the GC3.05-PPE simulations[41]. We consider soil moisture in the top 1m of soil, as this determines the level of water stress for vegetation with shallow roots (such as grass), whereas vegetation with deeper roots (such as trees) is rarely substantially water-stressed in the UK. Maximum soil moisture values occur in February or March, following accumulation of water during autumn and winter when precipitation exceeds evaporation. Minima occur in August or September, following depletion during spring and summer due to reduced precipitation and higher evaporation. Members of GC3.05-PPE show a range of values (typically ~0.15m for any given month), with a high degree of persistence through the annual cycle: members with relatively high or low soil moisture in winter also have relatively high or low values in summer. Biases in surface air temperature also tend to persist through the year (Fig. 3.15), however there is no clear relationship between these and the variations in soil moisture. Other potential sources of the cross-ensemble differences in soil moisture include variations in surface water fluxes, and impacts of parameter perturbations in the land surface component of GC3.05 (Table D.1).

---

[41] Corresponding results are not shown for CMIP5-13 simulations, as equivalent soil moisture diagnostics are difficult to calculate because different climate models represent the vertical distribution of soil moisture in different ways. Observational estimates are not given, because: (a) datasets made from in-situ measurements do not possess sufficiently dense spatial coverage; (b) remote sensing methods are only able to provide an estimate of surface (skin) soil moisture (Seneviratne et al. 2010), whereas depth-integrated values including the root-zone are needed to assess levels of water accessible to plants.
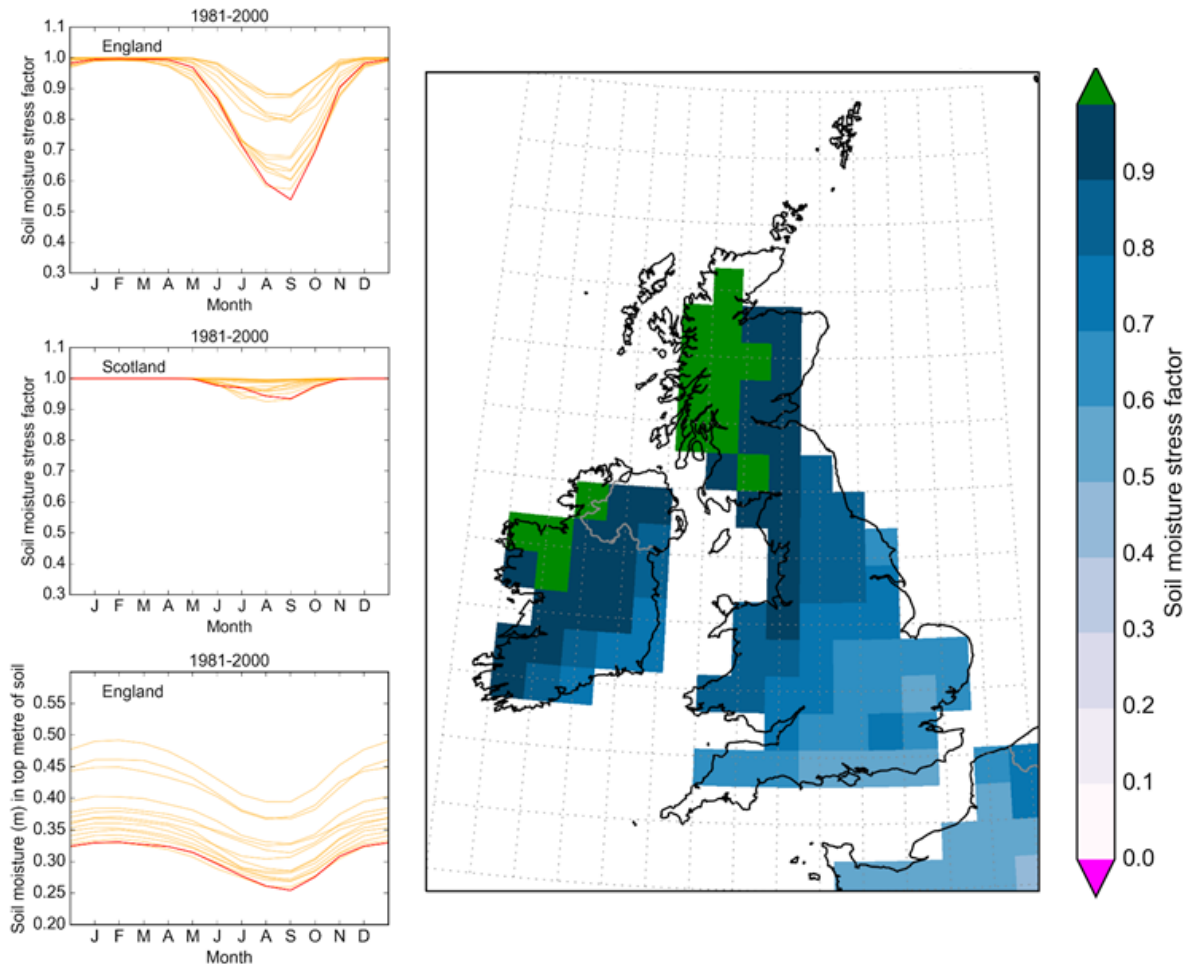
Figure 3.16. Annual cycles of spatial averages of the soil moisture stress factor in the top 1m of soil, for England (top left) and Scotland (centre left). Results are climatological averages over 1981-2000, shown for members of GC3.05-PPE with STD in red. This factor is a diagnostic (β) in the land surface component of GC3.05 (see Best et al., 2011). β=1 indicates soil moisture at or above a critical value at which no restriction is placed on transpiration from vegetated surfaces; β<1 indicates limitations in moisture availability that result in reductions in evaporation from both vegetated and bare soil surfaces; β=0 indicates root-zone soil moisture at or below a wilting point at which there is no water available to sustain plants. Bottom left panel shows member-specific annual cycles for England of average soil moisture (m) contained in the top 1m of soil. The map shows the ensemble-mean stress factor during September, at the driest point in the annual cycle.

The top and centre left panels in Fig. 3.16 show seasonal cycles of a soil moisture stress factor (β, see Best et al., 2011), for England and Scotland. In the GC3.05 land surface scheme, this measures the degree of restriction placed on evapotranspiration from vegetated surfaces, as soil moisture drops below a prescribed critical value corresponding to β=1. In Scotland, average monthly values never fall below 0.9, indicating no significant water stress. In England, average values remain at 0.8 or higher throughout the annual cycle in some ensemble members, also indicative of limited water stress. However, other members simulate higher levels of stress (β in the range 0.5-0.8) during the summer months, demonstrating a degree of moisture restriction during April-September.

By the end of summer, most parts of the UK experience some level of plant stress in the GC3.05 simulations (Fig. 3.16, right panel). The only exceptions are North-Western Scotland and parts of Northern Ireland. The lowest levels of water availability are found in South-East England, although the ensemble-mean climatological values shown in Fig. 3.16 remain well above zero, the level at which no moisture is available to plants. Lower values of β (occasionally close to zero) do occur in individual seasons during the 1981-2000 period of the GC3.05-PPE simulations. These correspond, for example, to occasional summers in the observed climate during which grass turns brown and dies in parts of the South of England.
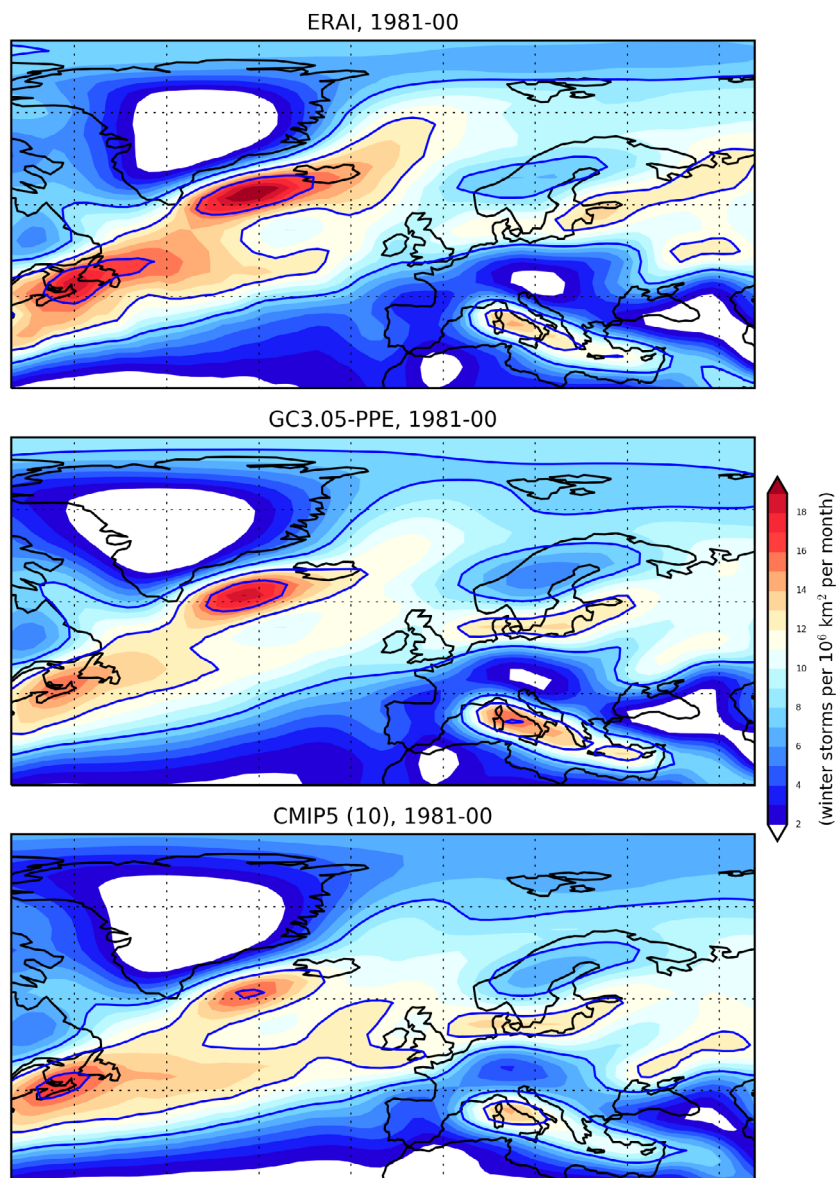
Figure 3.17. Density of winter storm tracks in the North Atlantic/Europe sector averaged over 1981-2000. Shading denotes intervals of two tracks per 10^6 km² per month, with values of 8, 12 and 16 contoured. These extra-tropical cyclone tracks are diagnosed from 6 hourly fields of relative vorticity at 850hPa using the methodology of Hodges et al. (2011), who provided the results from ERA-Interim shown in the top panel. Middle and lower panels show ensemble-mean values for GC3.05-PPE and ten of the CMIP5-13 members for which the required data was available. CMIP5 track data was provided by Zappa et al. (2013).

Extra-tropical cyclones are a key weather phenomenon that bring high winds, precipitation extremes and tidal surges to the UK. Fig. 3.17 compares the winter storm track density in observations against ensemble averages from GC3.05-PPE and ten members of CMIP5-13, noting that the required data (6 hourly values of relative vorticity at 850hPa) were not available for the other three models. The GC3.05-PPE ensemble-mean verifies quite well in general. In the core of the North Atlantic storm track, the locations of the regional maxima in track density to the south of Greenland, and over Nova Scotia and Newfoundland, are captured. The ensemble-mean intensities are underestimated somewhat, especially in the latter case, and the number of storms in the southern fork of the track are underestimated to the west of the UK, although not over the UK itself. The extension of the southern fork across the Baltic Sea and into Southern Scandinavia is represented, although the region of enhanced storm densities is located further west than in observations. The location, strength and shape characteristics of the Mediterranean storm track are also simulated quite accurately.

The CMIP5 ensemble mean also captures the main core of the North Atlantic track. The maximum south of Greenland is weaker than in observations or GC3.05-PPE, and the observed extension to the north-east of Iceland is missing. However, the maximum over Nova Scotia and Newfoundland, and the southern fork, is captured quite well, with more storms to the south of Iceland than in GC3.05-PPE. However, the number of storms tracking across the UK and into Europe is overestimated. The position and orientation of the Mediterranean track is also replicated in the CMIP5 ensemble mean, but its intensity is underestimated. Seasonal RMSE statistics (Fig. 3.18) invariably show a larger spread of errors for the ten CMIP5-13 members in comparison to GC3.05-PPE members. Several CMIP5 members score worse than any GC3.05-PPE member, in each of the four seasons. One model (BCC-CSM1.1) shows RMSE values considerably higher than the other simulations in winter, spring and autumn. This is because the storm track is shifted south, and is too zonal, with too many storms moving across the UK and western Europe. Errors in most of the other CMIP5 members are qualitatively similar, but the size of the bias in BCC-CSM1.1 is larger.
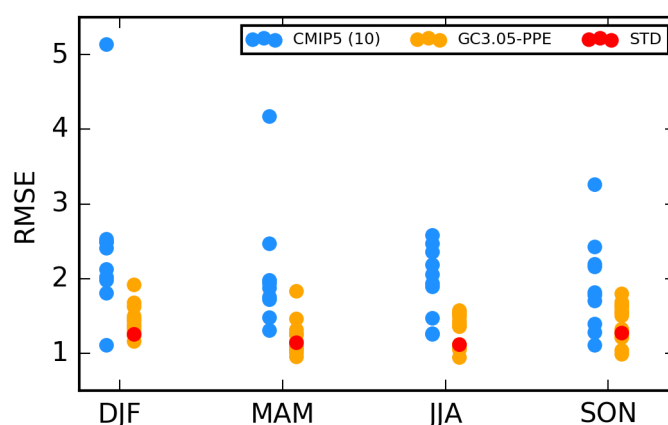
Figure 3.18. Root-mean square errors in simulations of average storm track density for 1981-2000, for winter, spring, summer and autumn. These are calculated as in Fig. 3.17, for the North Atlantic domain of 30-75ºN, 50ºW-5ºE. Blue and orange dots show ten members of CMIP5-13 and 14 GC3.05 members respectively, with STD in red. Units are tracks per $10^6 km^2$ per month.

Intra-seasonal variability can also be evaluated through assessment of weather types (WTs) that capture characteristic circulation patterns associated with different wind flows, or phenomena such as intense low pressure systems or blocking anticyclones. Neal et al. (2016) identify a set of eight WTs for the North Atlantic/Europe sector, based on a classification scheme developed using historical mean sea-level pressure (MSLP) data. Figure 3.19 (top panels) show the two leading types (WT1 and WT2), which correspond to the negative and positive phases of the NAO. WT1 is associated with a high pressure anomaly over Iceland, and a reduction in the strength of the westerly flow or reversal to easterly flow conditions. WT2 provides enhanced westerlies associated with a strengthened meridional pressure gradient. During 1900-1999, the observed daily circulation during winter is closest to WT1 on 22% of days, and WT2 on 21% of days. There is substantial interannual variability in both cases (Fig 3.19, lower panels), the number of observed days varying between zero (or close to zero for WT2) and ~60%.
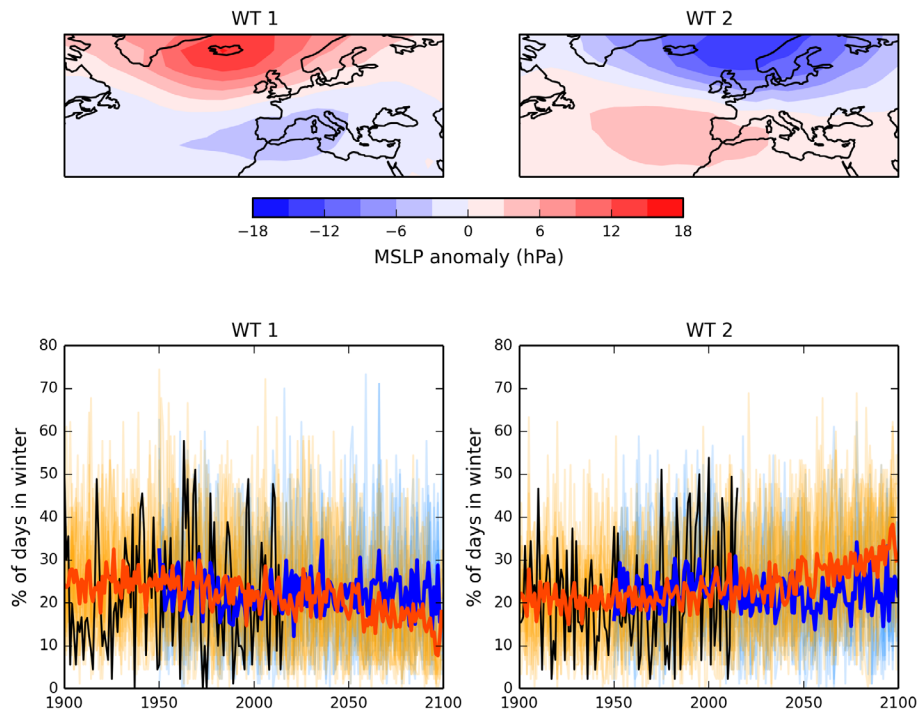
Figure 3.19. Upper panels show observed anomalies (hPa) in mean sea-level pressure (MSLP) relating to daily weather types (WT) 1 and 2 of the eight-pattern classification of Neal et al (2016). The observed patterns are generated from the EMULATE (EMSLP) dataset (Ansell et al., 2006), based on daily fields from 1900-99. The anomalies are calculated relative to a long-term annual mean climatology from 1850-2003. WT1 and 2 correspond respectively to the negative and positive phases of the NAO. Lower panels show the percentage of days during December to February assigned to WT1 (left) and WT2 (right) during 1900-2100 in members of GC3.05-PPE (orange), and during 1951-2100 in nine members of CMIP5-13 (blue) for which daily weather typing is available. The thicker, darker lines show ensemble-mean values. Black line shows observed historical values from EMSLP.

The lower panels of Fig. 3.19 also show time series of occurrence rates from the Strand 2 simulations, from 1900-2100 in the case of GC3.05-PPE members, and from 1951-2100 for CMIP5-13 (noting that weather typing data was only available for nine members of the latter, and was not available prior to 1951). During the historical period, the long-term average rates from each ensemble lie close to observations[42], and the values for individual seasons explore similar ranges to those found in the real world. The model time series show a few examples where simulated occurrences exceed the maximum observed value. However, this is not surprising, given the availability of a larger sample of seasonal values obtained by pooling multiple simulations. These results show that the intraseasonal variability in NAO-related circulation patterns corresponds quite well with observations. However, the long-term average of interannual variability in the NAO index is lower in observations than in the Strand 2 simulations, with the bias varying from member to member (see Fig 3.10 and discussion in section 3.4a). The future projections of WT1 and WT2 frequencies are discussed in section 5.1.

---

[42] Ensemble-mean values for GC3.05-PPE are 24.2% for WT1 and 21.0% for WT2.

3.5. Future global changes in the strand 2 projections

Figure 3.20 shows annual anomalies in GMST, simulated by GC3.05-PPE and CMIP5-13 members for the RCP8.5 scenario. These are compared with the probabilistic projections (see also Fig. 2.4a and discussion in section 2.3) in order to provide context, and illustrate how the Strand 1 and 2 products depend on their underlying modelling systems and methodologies.

The projections consist of full global averages of surface air temperature from 1900-2100. This contrasts with Fig. 3.12, which shows a blended SST and land air temperature product for the historical period, masked for consistency with the HadCRUT4 dataset. The observations in Fig. 3.20 are from an adjusted version (Cowtan and Way 2014, updated), in which regions of missing SST and air temperature data are filled in, using a kriging technique based on covariances derived from HadCRUT4 data. In comparison to HadCRUT4 this dataset gives slightly warmer estimates of global temperature beyond 2005, due mainly to filling of coverage gaps in the Arctic Ocean. It is more consistent with the projected values in Fig. 3.20, with the caveat that the Cowtan and Way dataset uses SSTs over open ocean regions, whereas the projection outputs use marine air temperatures. The use in Fig. 3.20 of the standard UKCP18 baseline (1981-2000) masks the cool bias found in the historical response of the GC3.05-PPE during this period (section 3.4a). This is seen more clearly in Fig. 3.12, in which a 1901-1950 baseline is used. In Fig. 3.20, the effects of this cooling appear as a warm shift in the GC3.05-PPE anomalies during 1900 to ~1960, compared to observations. During 1930-60, the CMIP5-13 ensemble mean shows a slight cool bias. Between 2000 and 2017 the observations generally lie below the ensemble-means of both GC3.05-PPE and CMIP5-13.
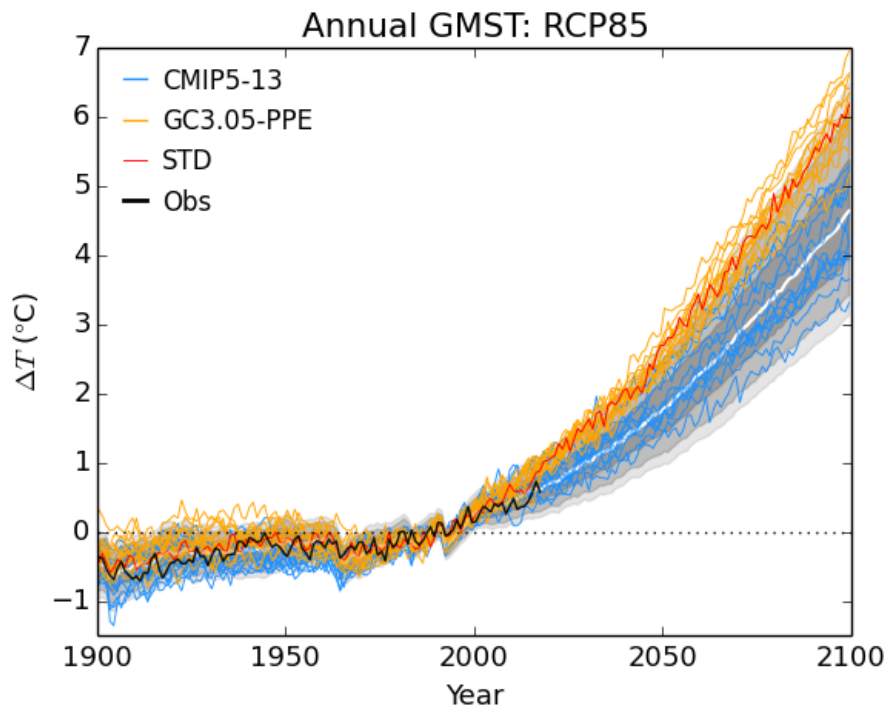

Figure 3.20. Historical and future changes in annual GMST from 1990-2100, relative to 1981-2000, from Strands 1 and 2 of UKCP18. Future changes are based on the RCP8.5 emissions scenario, applied in the projections beyond 2005. The median of the probabilistic projections is the white line, and shades of grey show the 5, 10, 25, 75, 90 and 95% probability levels. Orange lines (with STD in red) show members of GC3.05-PPE, and blue lines show CMIP5-13 projections. The black curve shows observations from Cowtan and Way (2014, updated). The model projections consist of full global averages of surface air temperature. The observations are derived from HadCRUT4, using a kriging technique to fill in missing SST values, and surface air temperature values over land and sea-ice. This improves consistency between the observed and modelled estimates, subject to the caveat that the observations use SSTs rather than marine air temperatures over the open oceans. Note: In Fig. 3.12 modelled GMST values are constructed differently, by blending and masking SST and surface air temperature in order to facilitate verification against the unmodified HadCRUT4 dataset.

This may reflect the warming hiatus, to which climate variability on decadal time scales is likely to have made a significant contribution (e.g. Dai et al., 2015). In GC3.05-PPE, another contribution may arise from warming due to the reduction in global sulphur dioxide emissions after 1980, given the strong aerosol forcing simulated in the ensemble (sections 3.3d and 3.4a). However, the observations remain fairly close to the median of the probabilistic projections during this period. Overall, the observations lie within the uncertainty ranges of the Strand 1 probability distributions, GC3.05-PPE and CMIP5-13 throughout the historical period.

The Strand 1 projections provide a range of future outcomes with steadily increasing uncertainty. This leads to a median warming exceeding 4°C by the 2090s, accompanied by a 10-90% range of approximately 3-5.5°C during this decade. The combined set of Strand 2 projections samples the upper 75% of the Strand 1 range throughout the 21st century. Beyond 2050, GC3.05-PPE consistently shows three or four outcomes above the 95% probability level of the Strand 1 projections, although these do not exceed the 99% probability level (not shown in Fig. 3.20). The ranges of change explored by GC3.05-PPE (orange lines) and CMIP5-13 (blue lines) show a clear and increasing separation through the future period, with most of the GC3.05-PPE members lying above the envelope of CMIP5-13 outcomes beyond 2050. CMIP5-13 shows a broader range of changes than GC3.05-PPE, but their combined range of outcomes is substantially wider than that of either ensemble in isolation.

However, only a few Strand 2 simulations, all from CMIP5-13, provide future outcomes below the median of the Strand 1 projections. After 2060, no Strand 2 simulation consistently populates the lowest 10% of the Strand 1 distribution. This demonstrates that Strand 2, despite possessing substantial diversity in GMST projections, does not capture all plausible outcomes for future GMST. For some international impacts assessments, it may be important to consider the benefits of outcomes consistent with lower levels of global warming than those provided in Strand 2. Strand 1 can be used to inform such studies, for example by use of scaling methods to consider regional temperature changes below the lower end of the Strand 2 envelope.

For the RCP8.5 scenario, AR5 assessed a 5-95% probability range[13] of 2.6-4.8°C for warming of GMST for 2081-2100 relative to 1986-2005 (Table 2.2). This was derived from CMIP5 simulations using the standard RCP8.5 CO_2 concentration pathway (red dot in Fig. 3.7), by assuming a Gaussian distribution of responses (Collins et al., 2013). The corresponding 5-95% probability range from Strand 1 is 2.6-5.5°C (Table 2.2), based on emissions-driven projections accounting for uncertainties in carbon cycle feedbacks. As noted in section 2.3, the upper limit in emissions-driven RCP8.5 projections derived from UKCP09 was reduced by 0.5°C when they were repeated as concentration-driven projections using the standard RCP8.5 pathway (Humphrey and Murphy, 2016). In Strand 2, the (CMIP5-13) projection with the smallest warming gives a response of 2.7°C for 2081-2100 relative to 1986-2005, while the largest warming of 5.9°C is provided by one of the GC3.05-PPE projections. In the latter simulation, the prescribed CO_2 concentration reaches 1123 ppm by 2100, compared with 931ppm in the standard concentration pathway.

The GC3.05-PPE projections of future GMST are distributed around a high level of ensemble-mean warming that lies close to the 90% probability level of Strand 1. The range of CO_2 profiles used to drive GC3.05-PPE members lies mainly above the standard RCP8.5 pathway (Fig 3.7). This is one factor that contributes to the higher level of ensemble-mean warming in GC3.05-PPE relative to CMIP5-13. The spread of outcomes in GC3.05-PPE reaches ~2°C by 2100, but is smaller, as noted above, than the spread in CMIP5-13 changes. The CMIP5-13 spread is driven entirely by uncertainty in physical climate feedbacks and ocean heat uptake, since the simulations all use the standard RCP8.5 CO_2 concentration profile (section 1.4). In

GC3.05-PPE, the range of CO_2 pathways contributes about 60%[43] to the variance of warming rates between 2050 and 2100, suggesting that the spread in global climate feedbacks sampled in the ensemble may be relatively narrow.

The sensitivity of GMST to changes in greenhouse gas forcing in a climate model is typically characterised by the equilibrium climate sensitivity (ECS) and the transient climate response (TCR). ECS is defined as the equilibrium response to a doubling of atmospheric CO_2 concentration, and is usually estimated from an idealised simulation in which CO_2 is either doubled or quadrupled, and then run for 150 years (e.g. Andrews et al., 2012). TCR is defined as the change in GMST at the time of doubling in a simulation in which CO_2 increases by 1% per year. At the time of production of this report, the idealised simulations required to calculate ECS and TCR for GC3.05-PPE members are not available. In AR5, IPCC assessed ECS to have a likely[38] range of 1.5-4.5°C (Collins et al., 2013), and also judged that there is a small probability (of up to 10%) that ECS exceeds 6°C. The levels of 21st century warming simulated for the RCP8.5 scenario suggest that most of the GC3.05-PPE members are likely to possess ECS values above 4.5°C.
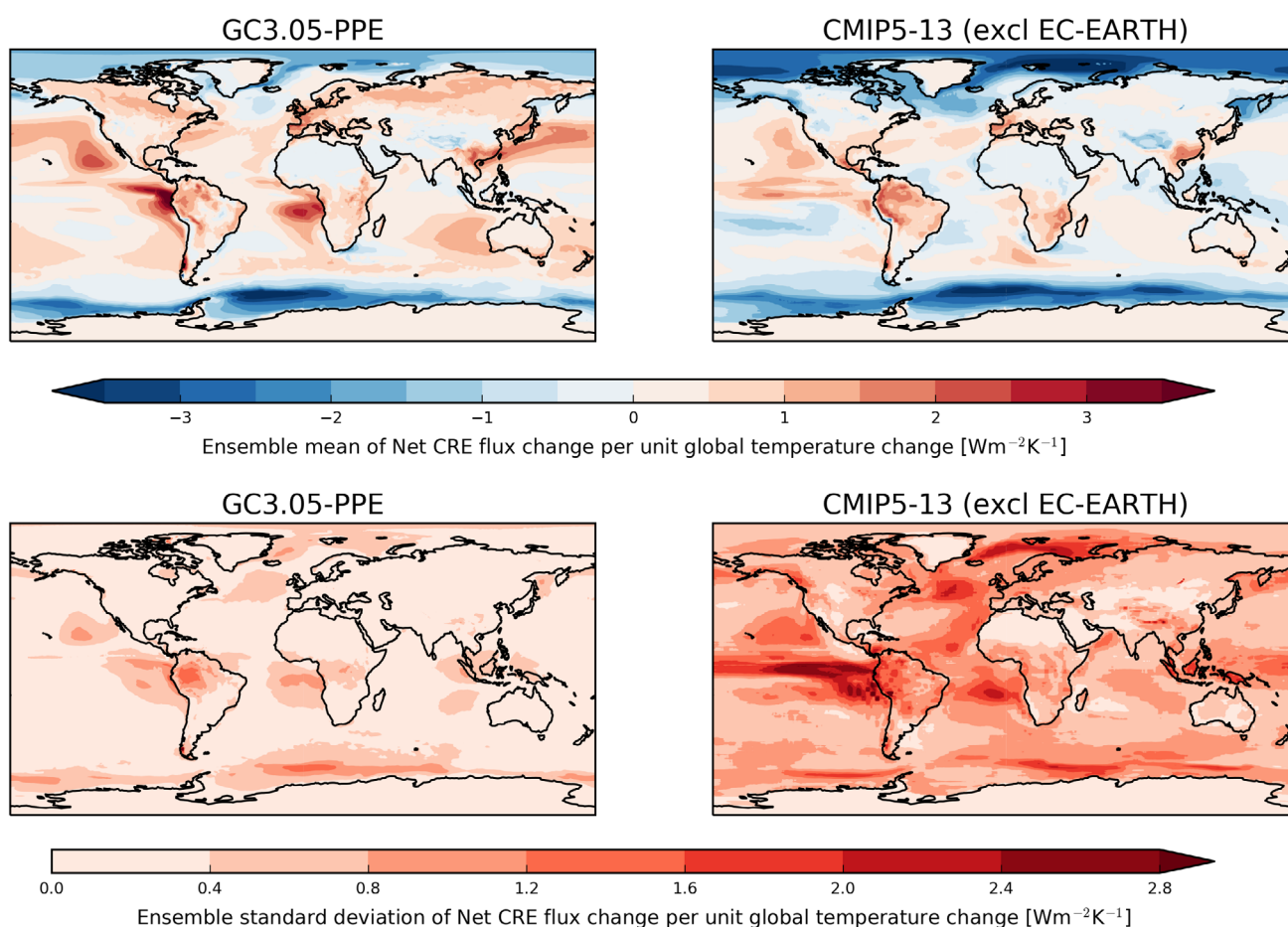
Figure 3.21. Changes in cloud radiative effect per unit change in global mean surface temperature (nCRE, in Wm^{-2}K^{-1}), in the Strand 2 simulations. Top row shows ensemble mean changes from GC3.05-PPE (left) and CMIP5-13 (right), excluding EC-EARTH. Bottom row shows the standard deviation of the changes across ensemble members. Values are calculated by linearly regressing local annual changes in CRE against GMST in each member of the relevant ensemble, from 1990-2100. This provides an estimate of total cloud feedback in response to the applied time-dependent changes in radiative forcing, including the effects of both rapid adjustments to changes in forcing from year to year, and longer-term feedbacks accounting for slower responses to accumulated changes in forcing. The former arises from climate system components with relatively little thermal inertia (atmosphere and land surface), while the latter accounts for changes in the oceans and sea-ice on decadal and longer time scales[27].

---

[43] This estimate is derived from the squared correlation across GC3.05-PPE members between the logarithm of the prescribed CO_2 concentration in 2100 (used because the relationship between concentration and radiative forcing is logarithmic) and the linear trend in simulated warming of GMST between 2050 and 2100.

Differences in ECS between models arise from differences in their simulations of contributions to global climate feedback arising from changes in water vapour, the temperature lapse rate in the atmosphere, cloud and surface albedo (e.g. Bony et al., 2006). Uncertainty in cloud feedbacks provides the largest contribution to the spread of global feedbacks, in both PPEs (e.g. Webb et al., 2006) and multi-model ensembles (e.g. Vial et al., 2013). Figure 3.21 (upper left panel) shows the ensemble-mean change in annual cloud radiative effect in GC3.05-PPE, calculated per unit rise in GMST. This metric, hereafter nCRE, provides an estimate[44] of total cloud feedback in response to the changes in radiative forcing during 1900-2100. It incorporates contributions from both rapid adjustments of the land surface and atmosphere on a time scale of ~1 year (e.g. Andrews and Ringer, 2014), and slower feedbacks accounting for the response of SSTs and sea-ice extents. The pattern shows strong positive changes in many parts of the world, in particular over many non-polar oceanic regions. This includes tropical regions subject to persistent low level clouds in the North and South Pacific and South Atlantic oceans, and also many mid-latitude regions in both hemispheres. The global average change in nCRE in GC3.05-PPE is 0.30 $Wm^{-2}K^{-1}$, compared to -0.17 $Wm^{-2}K^{-1}$ in CMIP5-13, in which the ensemble-mean pattern shows much less evidence of strong positive regional feedbacks (Fig 3.21, top right).

The lower panels of Fig. 3.21 show regional standard deviations of nCRE across members of each ensemble. In most regions the spread in GC3.05-PPE is substantially smaller than in CMIP5-13. The same applies to the standard deviation in globally-averaged nCRE, which amounts to 0.06 $Wm^{-2}K^{-1}$ in GC3.05-PPE cf 0.33 $Wm^{-2}K^{-1}$ in CMIP5-13. These results suggest that cloud feedbacks play an important role in explaining the consistently high levels of global warming found in GC3.05-PPE, compared with the lower ensemble average, and broader range, found in CMIP5-13.

Figure 3.22 shows changes in clear-sky shortwave radiative heating at the top of the atmosphere per unit rise in GMST, hereafter nSWCS, presented in the same format as Fig. 3.21. The patterns of ensemble-mean change (upper panels) show large positive values at high latitudes over the oceans, due to reductions in surface albedo arising from projected decreases in sea-ice extents. There are also relatively large values over the northern hemisphere continents at middle and high latitudes. These are probably due mainly to reductions in snow cover during winter and spring. Over much of Central and Northern Europe, Kazakhstan, Western Russia, Canada and northern parts of the US, the strength of regional nSWCS changes in GC3.05-PPE is larger than in CMIP5-13. In these regions, the winter cold bias in GC3.05-PPE (Fig. 3.8a) is likely to be an important driver of its stronger albedo feedback (Hall and Qu, 2006). In the global average, however, the ensemble-mean strength of nSWCS is slightly smaller in GC3.05-PPE than in CMIP5-13 (0.66 cf 0.82 $Wm^{-2}K^{-1}$), showing that this component of climate feedback does not play a role in explaining the higher GMST responses in GC3.05-PPE.

---

[44] The normalised changes in cloud radiative effect (Fig. 3.21) and clear-sky shortwave radiation (Fig. 3.22) provide good indications of the strength of feedbacks in response to changes in greenhouse gases due to cloud and surface albedo respectively. However, they should not be interpreted as precise estimates of these feedbacks, because the results are derived from RCP scenario simulations that are not designed to isolate the effects of feedbacks from other contributors to the changes in radiative fluxes. These include the effects of changes in greenhouse gas forcing, and effects of forcing due to aerosols and changes in land use. Nevertheless, these additional contributions are expected to be relatively small, because the greenhouse gas forcing manifests itself mainly in the clear-sky longwave component of the earth's radiation budget (e.g. Gregory and Webb, 2008), and changes in anthropogenic forcing during the 21st century are dominated by greenhouse gases.
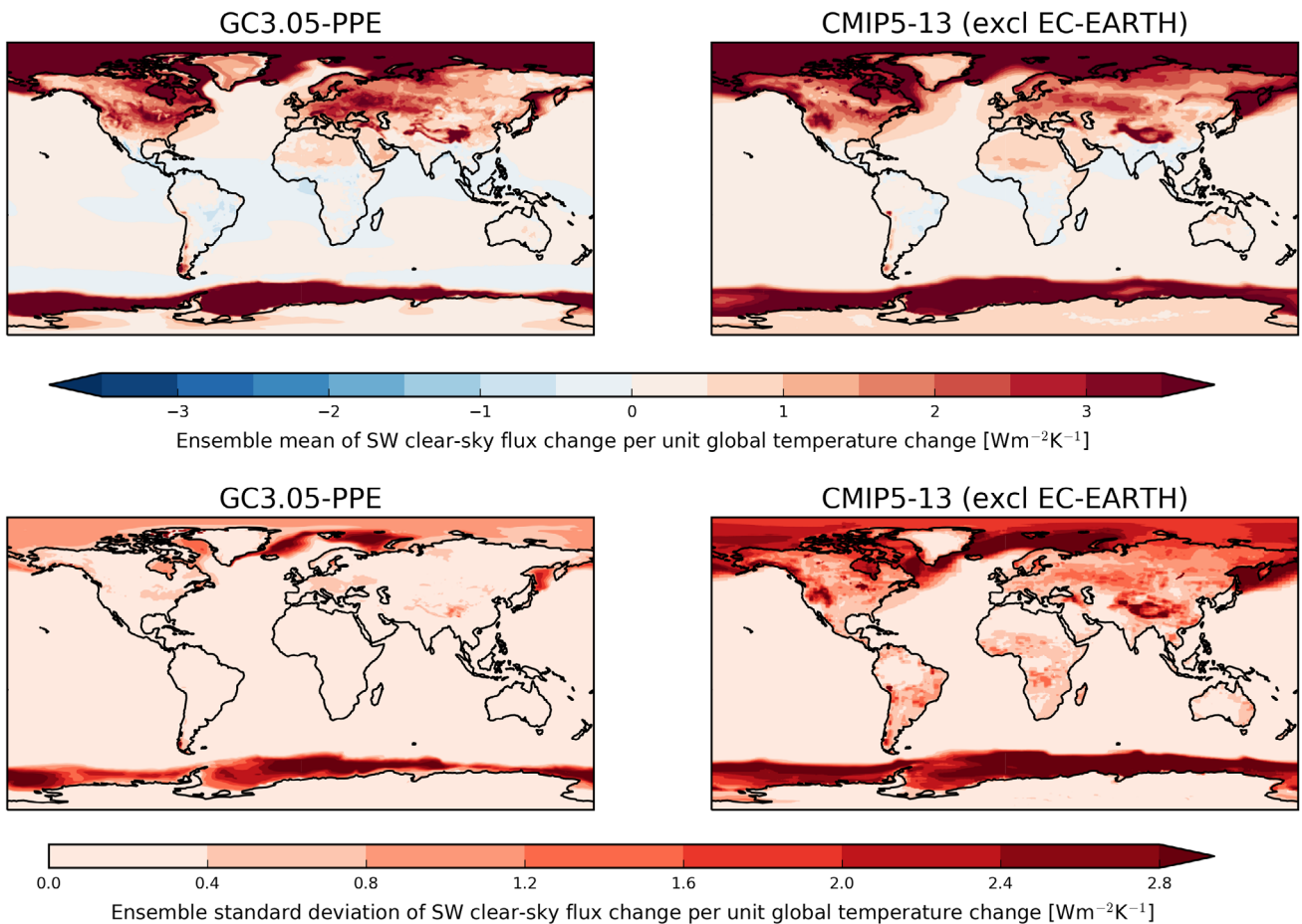
Figure 3.22. Changes in short-wave radiative heating at the top of the atmosphere under clear-sky conditions, in the Strand 2 simulations. Changes are given in normalised form (nSWCS), expressed per unit change in global mean surface temperature ($Wm^{-2}K^{-1}$). Top row shows ensemble mean changes from GC3.05-PPE (left) and CMIP5-13 (right), excluding EC-EARTH. Bottom row shows the standard deviation of the changes across ensemble members. Values of nSWCS are calculated by linearly regressing local annual changes in SWCS against GMST in each member of the relevant ensemble, from 1990-2100.

The spread of nSWCS changes (Fig. 3.22, lower panels) is larger in CMIP5-13 than in GC3.05-PPE in most terrestrial regions, as well as over regions of sea ice. In some regions, such to the south west and south east of the Kamchatka Peninsula, the spread in CMIP5-13 may be enhanced by the presence of a wider range of historical biases in sea-ice cover compared to GC3.05-PPE. In the latter, biases are restricted by use of flux adjustments (see section 3.4e). Over land, many of the local spread maxima in CMIP5-13 occur in regions of high orography, probably related to the use of different horizontal grid spacings in different models. Different implementations of land use changes may also contribute to the spread of terrestrial nSWCS values in CMIP5-13 (Prestele et al., 2017). The spread in globally averaged values is 0.29 $Wm^{-2}K^{-1}$ in CMIP5-13, compared with 0.07 $Wm^{-2}K^{-1}$ in GC3.05-PPE.
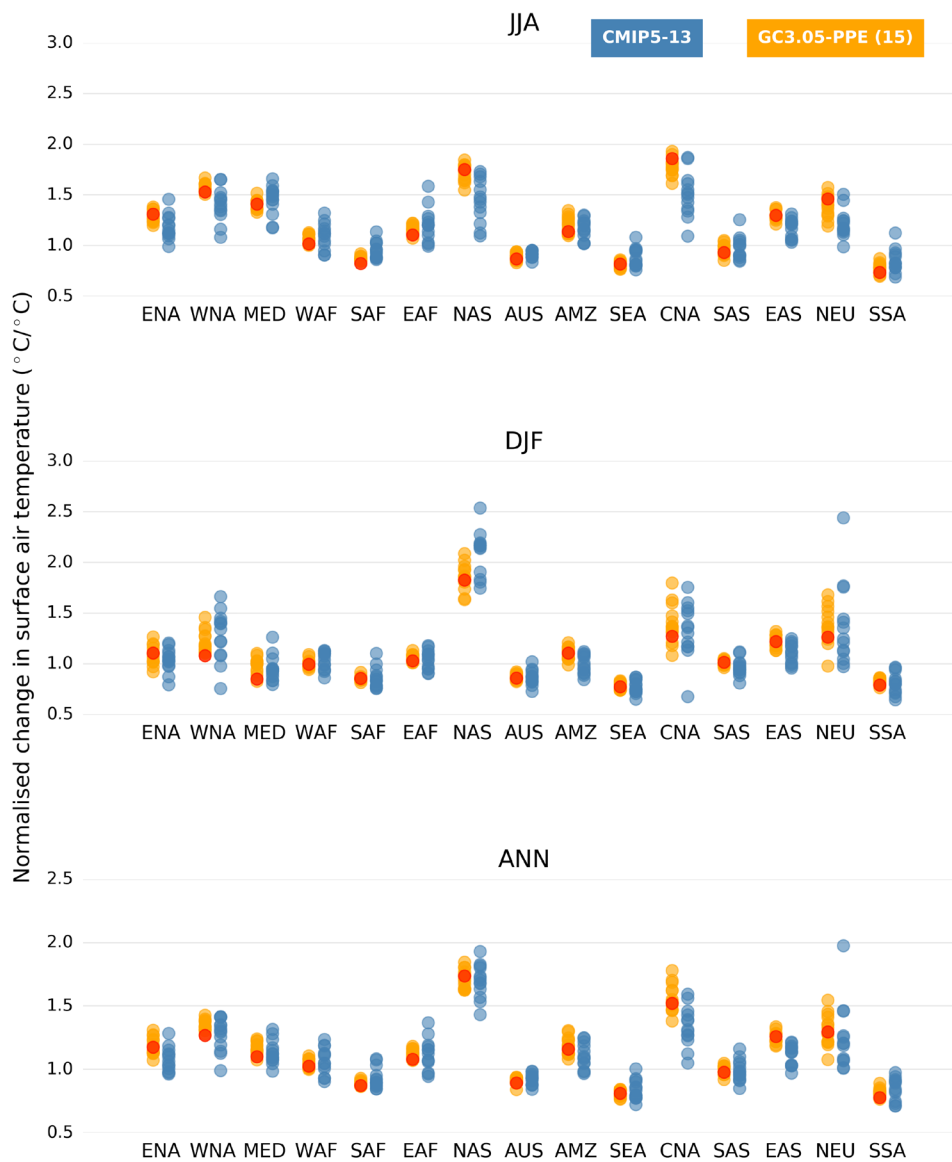
Figure 3.23a. Regional changes in surface air temperature for 2061-2080 relative to 1981-2000, normalised by corresponding changes in the global average (°C per °C). Results are shown for various worldwide regions defined by Giorgi and Francisco (2000), for JJA (top), DJF (middle) and annual mean (bottom), for the Strand 2 projections. Members of GC3.05-PPE are shown in orange (with STD in red); CMIP5-13 members in blue. ENA = Eastern North America; WNA = Western North America; MED = Mediterranean Basin; WAF = Western Africa; SAF = Southern Africa; EAF = Eastern Africa; NAS = North Asia; AUS = Australia; AMZ = Amazon Basin; SEA = South-East Asia; CNA = Central North America; SAS = South Asia; EAS = East Asia; NEU = Northern Europe; SSA = Southern South America.

In Figure 3.23a we show changes in surface air temperature for 2061-2080 with respect to 1981-2000 for each of the 28 Strand 2 projections, normalised by the corresponding change in GMST. These normalised changes are given for a set of the terrestrial regions of Giorgi and Francisco (2000), and provide a simple measure of spatial patterns of response (e.g. Fig.12.10 and discussion in Collins et al., 2013) and their associated uncertainties. The ranges of change arise from the combined effects of modelling uncertainties and internal climate variability on the 20-year time scale. The spread of normalised surface air temperature changes in GC3.05-PPE is narrower than that of CMIP5-13 in some cases, and similar in others. There are no examples where the spread in GC3.05-PPE significantly exceeds that of CMIP5-13. In some regions, combining both ensembles allows sampling of a slightly wider range of normalised responses than is provided by CMIP5-13 alone, including in North and East Asia, Central North America and Amazonia. This

could be important, for example, in studies of regional impacts associated with specific levels of global warming (e.g. Clark et al., 2010; Gohar et al., 2018). In such applications, further evaluation of how well the individual Strand 2 projections represent regional drivers of changes in temperature, precipitation and other key variables will be an important precursor to using the results for impacts assessments.
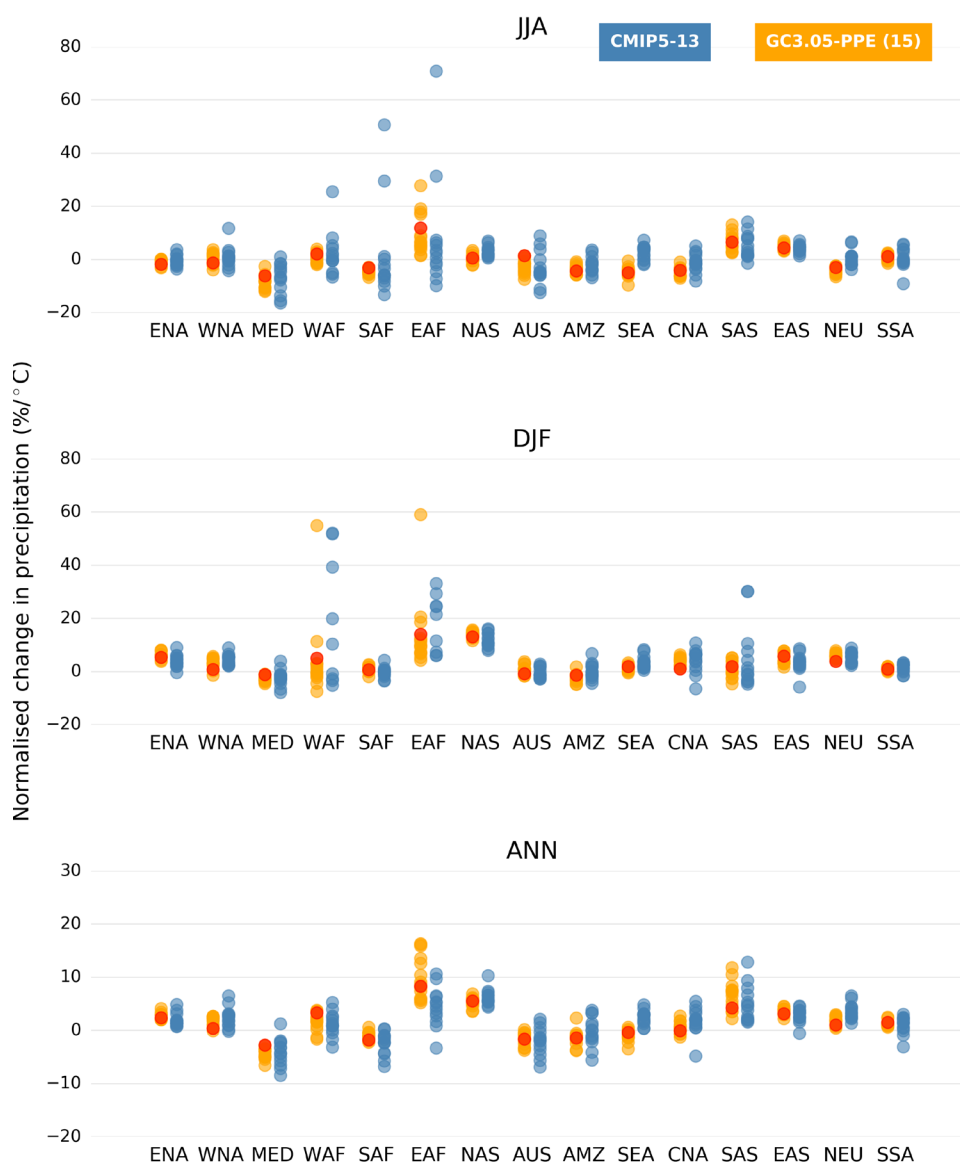

Figure 3.23b. As Fig. 3.23a for normalised changes in precipitation (% per °C of warming in global mean surface air temperature).

Corresponding results for normalised precipitation changes are shown in Fig. 3.23b. The spread of changes in GC3.05-PPE is typically smaller than in CMIP5-13. However, comparable ranges are found in some cases, including South East Asia in JJA and the annual mean for South Asia and East Africa. In the latter examples, combining both ensembles significantly enhances the overall spread of normalised changes. Comparing the GC3.05-PPE and CMIP5-13 ranges is also helpful in confirming the status of outlying members. Over East Africa, for example, one GC3.05-PPE member (in DJF) and one CMIP5-13 member (in JJA) simulates an increase in normalised precipitation much larger than any of the other 27 projections. All GC3.05-PPE members simulate drying in summer precipitation over Northern Europe, whereas some CMIP5-13 members project an increase. This is also the case for the UK (see section 5.1).

For a set of 24 countries, Murphy et al. (2014) performed a similar analysis to that of Figs. 3.23a,b, in their case comparing the ESPPE used in Strand 1 against emissions-driven CMIP5 earth system model simulations. They found several examples (notably for northern hemisphere countries in JJA) in which the spread in normalised surface air temperature from the ESPPE exceeded that in the multi-model ensemble, and more cases where combining the two ensembles led to a substantial increase in the overall range of surface air temperature or precipitation outcomes. This underlines that the diversity properties of a PPE, and how they relate to a corresponding multi-model ensemble, are conditional upon the climate model chosen for perturbation (see also Yokohata et al., 2010; Sanderson, 2011).

It is not possible to provide robust estimates of the (potentially member-specific) contribution from internal variability to the uncertainty ranges of Fig. 3.23a,b, in the absence of multiple simulations for each GC3.05-PPE or CMIP5-13 member (e.g. Deser et al., 2012). Hawkins and Sutton (2009, 2011) show that model uncertainty typically provides the larger source of spread for multidecadal lead times, while the role of internal variability becomes more important at time scales of a few decades ahead, particularly for precipitation.

For a given scenario of future emissions, uncertainties in projected regional changes are generally influenced strongly by uncertainties in both the GMST response (Mauritzen et al., 2017) and the normalised patterns of change (e.g. Murphy et al., 2014). This understanding also underpins the standard pattern-scaling approach used to emulate responses to alternative emissions scenarios, in which stationarity in the normalised pattern is assumed (e.g. Tebaldi and Arblaster, 2014). However, regional changes can also be influenced by a number of additional factors, not accounted for in this simple paradigm. These include differing responses to different forcing agents (Shiogama et al., 2013), and non-linearity in regional responses to different levels of radiative forcing (Chadwick and Good, 2013) or GMST response (Herger et al., 2015). The basic response characteristics of Fig 3.20 and 3.23a,b should not, therefore, be interpreted as a complete explanation of the time-dependent regional changes simulated by the Strand 2 members.
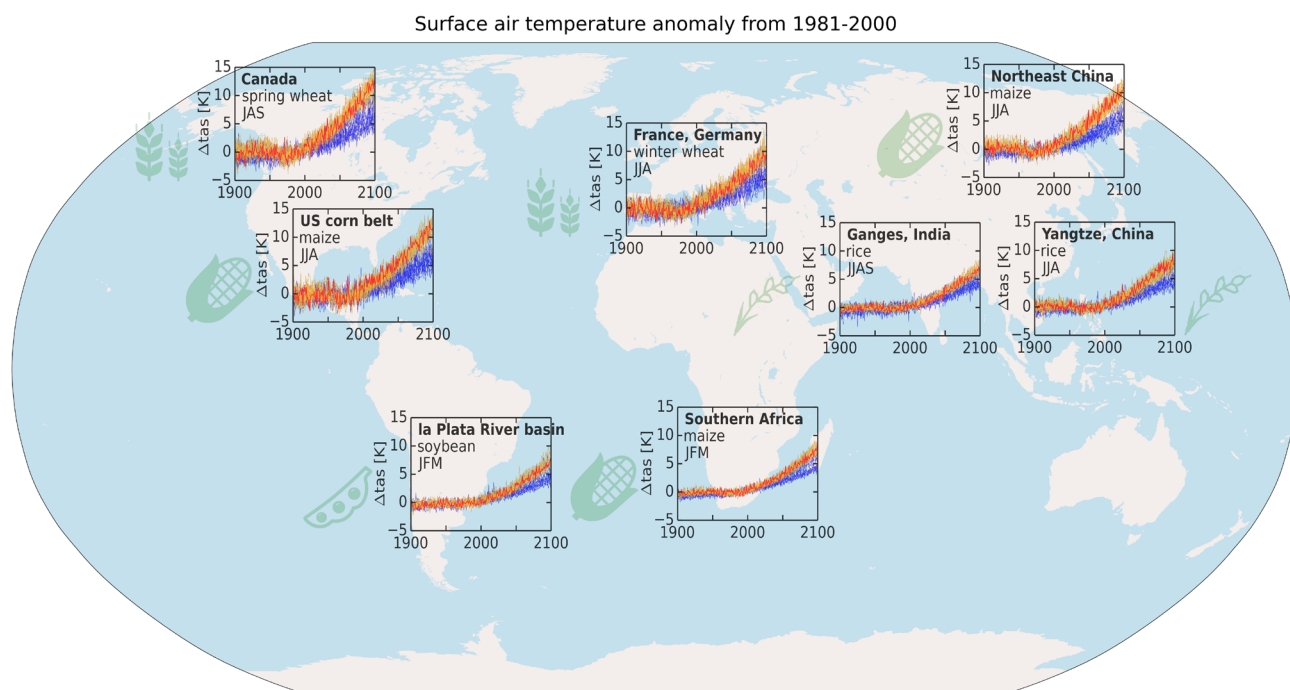

Figure 3.24a. Projected anomalies in surface air temperature (°C) relative to 1981-2000, from the Strand 2 simulations. GC3.05-PPE members shown in orange (with STD in red), and CMIP5-13 members in blue. Results are shown for the annual growing seasons of wheat, maize, rice and soybean in their major production regions, from 1900-2100.

As an example of transient regional changes, Figures 3.24a,b show time series of surface air temperature and precipitation for growing seasons in major food production regions. These relate to wheat, maize and rice (which account for 60% of global food energy intake), and soybean, which is a valuable source of protein and animal feed. The Strand 2 simulations could hypothetically be used to obtain information on the changing risks of drought events with potential to cause major reduction in yield, and hence risks to food security. For maize, for example, summers in which average temperature exceeds 23°C and precipitation is lower than 250mm are associated with major reductions in yield (Kent et al., 2017). For the US Corn Belt, the GC3.05-PPE projections suggest a clear increase in risk of breaching both thresholds in future, while CMIP5-13 shows more uncertainty in the sign of the precipitation response, with more modest increases in surface air temperature than in GC3.05-PPE. In all regions, the post-2050 surface air temperature changes in GC3.05-PPE explore outcomes beyond the upper end of the CMIP5-13 range, due in particular to the influence of large projected changes in GMST (Fig. 3.20). For precipitation there is a greater degree of overlap, although in the North American and European regions GC3.05-PPE explores stronger reductions in summer precipitation than are found in CMIP5-13. In North-East China, GC3.05-PPE explores positive and negative precipitation anomalies beyond the ranges of CMIP5-13, by the end of the 21st century. Note that in a practical application of this type, it would be important to consider carefully how to bias correct the model output for use in crop projections, given especially the importance of thresholds in the calculations (Hawkins et al., 2013).
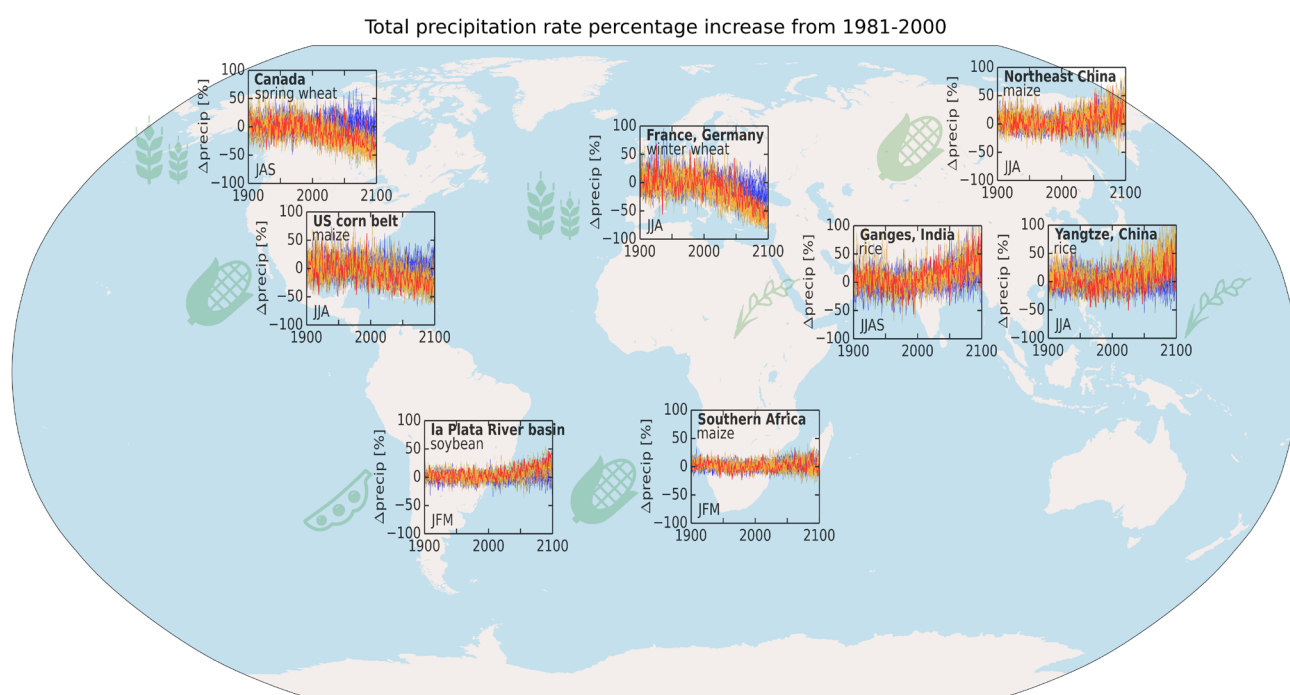

Figure 3.24b. As Fig. 3.24a, for annual changes in growing season precipitation (%)

The representation of future outcomes within the combined Strand 2 range is often somewhat uneven. This is due to its construction from two distinct climate model ensembles, with different uncertainty properties. Users should not interpret probabilistically the numbers of outcomes in different sub-ranges of the overall envelope of responses. This is because Strand 2 is designed to support a limited set of plausible storylines (section 1.3), rather than a comprehensive survey of possible changes.

Nevertheless, the dataset provides a rich sample of outcomes under the RCP8.5 scenario, to which GC3.05-PPE contributes plausible pathways not available by considering CMIP5 simulations in isolation. It therefore provides a new opportunity for investigation of international climate impacts, and inward risks to the UK. For example, the GC3.05-PPE runs may be particularly useful for stress-testing impacts sensitive to high levels of future warming (see also discussion in section 5.2). In specific regional applications remote to the UK and Europe, it will be important to consider further screening of Strand 2 members, beyond that of section 3.3f. This will be necessary to check the credibility of the 28 simulations for impacts assessments in different parts of the world.

4. Strand 3: a new perturbed parameter ensemble of regional climate model simulations from 1980 – 2080

4.1. Overview

Dynamical downscaling using RCMs has formed an important component of UK climate projections since the UKCIP02 scenarios (Hulme et al., 2002). Experience from worldwide research shows that RCMs add value to simulations from global climate models, by better resolving physiographic features such as mountains, land-sea contrast, urban effects and inland water bodies, and also mesoscale circulations and storms (Rummukainen, 2016). Regional models can also provide improvements at larger scales that might be considered well resolved by the driving GCM (Sørland et al., 2018). For the UK, RCMs simulate spatial detail in climatological distributions of surface air temperature and precipitation with considerable skill (e.g. Jones et al., 1995). They can also improve the representation of distributions of daily surface air temperature and precipitation (Massey et al., 2015), and provide credible simulations of extreme regional rainfall events on 1-10 day time scales (Fowler et al., 2005). Many regional climate model simulations (including those described below) are driven by global climate model output using a one-way nesting technique, with a domain small enough to ensure that the large-scale circulation remains fairly close to that of the driving model. The RCM therefore inherits biases from the global simulation (e.g. Noguer et al., 1998), and is subject to the same limitations in relation to the representation of large-scale driving phenomena (section 3.4).

Strand 3 provides a new ensemble of twelve RCM simulations for the period 1980-2080. This is a PPE of RCM variants (RCM-PPE), configured at 12km horizontal resolution and driven by twelve of the fifteen GC3.05-PPE simulations of section 3. It addresses requirements for a flexible dataset for impacts assessments using contemporary UK modelling capabilities (user drivers 2 and 5 of section 1.3). In particular, Strand 3 supports applications requiring detailed impacts information at local to sub-national scales, especially those likely to benefit from the addition of skilful downscaling information to global model output from GC3.05-PPE.

In such applications, RCM-PPE provides a new resource for augmenting and updating studies based on the HadRM3-PPE ensemble included in the suite of UKCP09 simulations (Murphy et al., 2009). Those 25km projections were used in many UK impacts studies, including assessments of drought (Burke et al., 2010), river flows (Prudhomme et al., 2012), water availability (Sanderson et al., 2012), flood frequency (Kay and Jones, 2012), and effects on the electricity and rail networks (McColl et al., 2012; Palin et al., 2013).

UKCP09 users were encouraged to compare the spread of projected changes against the broader sampling of uncertainties contained in the probabilistic projections (Sexton et al., 2010). The idea was that stakeholders requiring datasets with full spatial and temporal coherence could use the RCM simulations to analyse impacts, and the probability distributions to understand their limitations. This might extend to application of scaling techniques to consider potential outcomes outside the range of the RCM ensemble. Such exercises can also be carried out using UKCP18 data, through use of Strand 1 results to provide context for the Strand 3 projections (see section 5.2).

In addition, UKCP18 users have the option of deriving impacts from the spatially coherent global model projections of Strand 2. Here, a key judgement will involve the relative benefits of downscaling information, versus the opportunity to consider a more diverse set of projections that combines perturbed parameter and multi-model ensembles. In practice, the choice is likely to be application-dependent. Combining the Strand 2 and 3 information may often provide more comprehensive advice than considering either in isolation. For example, if a user wishes to consider an event like the winter floods of 2013-14 (Huntingford et al., 2014), the set of Strand 2 projections could inform changing risks of driving circulation patterns such as the NAO, and resulting UK-wide precipitation anomalies. The Strand 3 PPE could then provide input data for hydrological modelling of catchment-scale river flows and flood risks, for similar events selected from the projections.

While Strand 3 does not include a multi-model RCM ensemble alongside the PPE, users can, if they wish, consider augmenting the UKCP18 results with information from the EuroCordex coordinated downscaling experiment (Jacob et al., 2014). EuroCordex includes a multi-model ensemble of 12km RCM simulations from 1951-2100, driven by CMIP5 global models under the RCP8.5 scenario (see http://www.euro-cordex. net/) and using the same European domain as RCM-PPE. These simulations provide an option to add sampling of structural modelling choices in assessment of regional downscaling uncertainties, alongside the parametric uncertainties considered in RCM-PPE. EuroCordex simulations are also available for the RCP2.6 and 4.5 scenarios.

Strand 3 will subsequently be augmented by an ensemble of 2.2km simulations for the UK (see section 1.3). This will provide advice on changes at kilometre-scale resolution and for sub-daily rainfall extremes, providing important new information for applications such as urban planning and flash flooding.

4.2. The 12km regional climate model

The RCM is a limited area configuration of the GC3.05 atmosphere model. It uses the EuroCordex latitude-longitude grid with 0.11° resolution and a rotated pole set at 39.25°N, 198°E. This gives a quasi-uniform grid spacing of 12km over the European domain. The RCM uses a 4 min. timestep and has 63 levels (with an upper lid at ~40km), rather than the 85 levels of GC3.05 (upper lid ~80km). It is driven at its lateral boundaries in a one-way nesting approach, using time series of surface pressure, wind, temperature and moisture values archived every 3 hours from the relevant GC3.05-PPE member, and applied to the RCM across a 10-point relaxation zone. Inside this, orographic heights on the global and regional model grids are blended across a further three points. The RCM therefore adjusts to the surface and lateral boundary forcing across a 13-point external rim, which is removed in the European figures shown in section 4.4. In its internal domain, the RCM simulation evolves freely, and responds to surface orography specified at its native grid scale.

Prescribed daily fields of SST and sea-ice cover are also supplied from the driving global simulation. For some inland water bodies resolved on the RCM grid (for example some high-elevation lakes), it was not possible to provide credible characteristics by interpolation from the coarser global model fields. Therefore, the RCM land-sea mask was edited to keep only lakes for which the nearest GC3.05 sea point was expected to give a better representation of surface temperature than would be achieved by excluding the lake point. Swedish lakes were set as land points on the assumption that most of them would normally be frozen in winter, whereas nearby GC3.05 sea points may not be. However, the Finnish lakes are included, as well as the large Russian lakes (Onega and Ladoga) and the Bosporus. The latter are important for simulation of surface heat and moisture fluxes in Eastern Europe, especially during summer.

The representations of atmospheric dynamics and the parameterisations of land and atmospheric processes are identical in the RCM and global simulations, with the exception of the treatment of aerosols (see below). Parameter perturbations applied in each driving GC3.05-PPE member are mirrored in its RCM counterpart. This is done to ensure that the RCM scenarios are as consistent as possible with the global simulations at large regional scales (further discussion in section 4.4).

It was decided not to include the GLOMAP-mode aerosol scheme as an interactive component of the RCM, due both to its computational expense and a lack of previous experience of how the scheme performs in climate simulations at resolutions higher than N216. However, aerosol radiation and cloud effects simulated by the relevant GC3.05 member were approximately replicated in each RCM simulation. This was achieved by saving monthly spatial fields of shortwave and longwave optical properties (absorption, extinction, scattering and asymmetry) and cloud droplet number concentration, and then prescribing these as a time series in the RCM simulation for use in its calculation of time-varying radiative forcing. In a global model, Stevens et al. (2017) showed that this approach replicates quite well the aerosol forcing found in interactive simulations, and tests using GC3.05 supported this conclusion.

4.3. Design of regional projections

The 12 pairs of GC3.05 and RCM variants were selected in two steps, in conjunction with those used to reduce the original set of 25 GC3.05 candidates down to 20 and then 15 members, based on historical performance (section 3.3f).

Initially, 16 GC3.05 driving members were selected from the 20 that survived the first filtering step, motivated by availability of HPC resources. STD was chosen, plus the four members with the weakest and strongest estimates of global aerosol forcing and climate feedback strength (section 3.3d). From the remaining 15 members, one was excluded due to poor simulation of SSTs. An additional 11 members were then picked from the surviving 14 candidates, by successively dropping members on the basis that the retained model variants should maximise the spread of sampled parameter values in the 47-dimensional space of Fig 3.2.

The initial 16-member RCM ensemble included four members driven by GC3.05 variants dropped in the second filtering step of section 3.3f, following further assessment of European climatology, AMOC strength and historical trends in northern hemisphere surface temperature. These were subsequently excluded from RCM-PPE, leaving 12 members to form the Strand 3 contribution to UKCP18.

The RCM simulations were run in three slightly overlapping time slices (1979-2021, 2019-2061, 2059-2080), in order to expedite completion of the simulations. Initial conditions for atmospheric state variables were taken from the corresponding points of the driving GC3.05 simulations, and initial soil moisture values were taken from an ancillary dataset of soil properties derived from observations (see Table 2 in Walters et al., 2017). The first two years of each RCM time slice were used as a spin-up period, to allow fine-scale circulations and land surface properties (particularly soil moisture in the root zone) to reach approximate equilibrium. This was achieved, with no evidence of residual drift in soil moisture after this point. The two spin-up years were discarded, and continuous time series for 1980-2080 were created by combining the remainder of the data, with switchover dates of 1 December 2020 and 1 December 2060. This ensures that all meteorological seasons in the dataset are taken from one simulation only.

Changes in greenhouse gas forcing follow those in the driving GC3.05 simulations[45]. After 2005, each RCM simulation follows the member-specific CO_2 concentration pathway prescribed in its driving model projection (section 3.3g). Changes in aerosol optical and cloud properties through 1980-2080 are prescribed from the driving simulations, as described in section 4.2. Past and future changes in land use were represented by prescribing a time-varying component due to anthropogenic disturbance (Hurtt et al., 2011), which was used to modify the observed present-day distribution of land cover types. As in the global simulations (section 3.3e), this was done by modifying the combined coverage of C3 and C4 grasses relative to the combined coverage of trees and shrubs.

An additional simulation was produced using the RCM variant with standard parameter settings (RCM-STD), in order to facilitate understanding of model biases. This simulation, hereafter ERAI-RCM-STD, was run from 1981-2001[46], using lateral boundary forcing taken from ERA-Interim reanalyses rather than GC3.05 STD. Sea surface temperatures and sea-ice extents were also prescribed from analyses of observations (Reynolds et al., 2002). Aerosol properties were prescribed as described in section 4.2, but using a standard historical forcing dataset developed for CMIP6 (Stevens et al., 2017) rather than output from STD. Below, results from this simulation are compared with those of RCM-STD. This simulation allows the downscaling properties of the RCM to be assessed relatively cleanly, in a simulation driven by a reconstruction of the large-scale atmospheric circulation close to observations. Establishing the downscaling skill under such conditions is an important precondition for use of the RCM in climate change impacts studies. Through comparison with the RCM-STD simulation driven by the STD member of GC3.05-PPE, biases in the latter can also be partitioned approximately into a component inherited from the driving global model and a component arising from regional errors in the RCM.

---

[45] During production of the first two time slices, it was discovered that the CMIP5 time series of ozone forcing data had been implemented with a one-year shift compared to the correct values. This was corrected in 1990 and 2031 respectively, with the result that the periods 1980-1989 and 2020-2030 contain the time-shift error. Tests show that the offset leads to an error in total anthropogenic forcing of less than 0.5% during these periods. Hence it was judged unnecessary to repeat the relevant periods of the RCM simulations.

[46] In section 4.4, the 20-year climatology of the reanalysis-driven simulation for 1982-2001 is compared with 20-year mean values for 1981-2000 of observations, and the GC3.05-driven simulation. This period is offset by one year from the standard baseline period, because the Reynolds high resolution SST dataset did not cover 1981. However, the offset is not expected to make a significant contribution to differences between the reanalysis-driven simulation and the other datasets.

4.4. Evaluation of regional simulations for 1981–2000

We assess here the performance of selected aspects of RCM-PPE during the 1981-2000 baseline period. Further information is provided by Tucker et al. (2018).

In Figures 4.1a-d we use RCM-STD to demonstrate the impacts of higher resolution on simulations of climatological average surface air temperature and precipitation. The top row of each figure shows absolute values in ERAI-RCM-STD, RCM-STD, STD (referred to here as GCM-STD for clarity), and NCIC observations. The bottom row shows biases relative to observations in the two regional model simulations at their native 12km resolution (left and centre left). Also shown is the difference between RCM-STD at 12km resolution and a version smoothed to the 60km resolution of the driving model, to show the influence of local detail simulated at scales unresolved by the latter. The bottom right panel shows differences between the smoothed RCM-STD climatology and that of GCM-STD, in order to assess the impact of higher resolution on scales resolved in the global model.
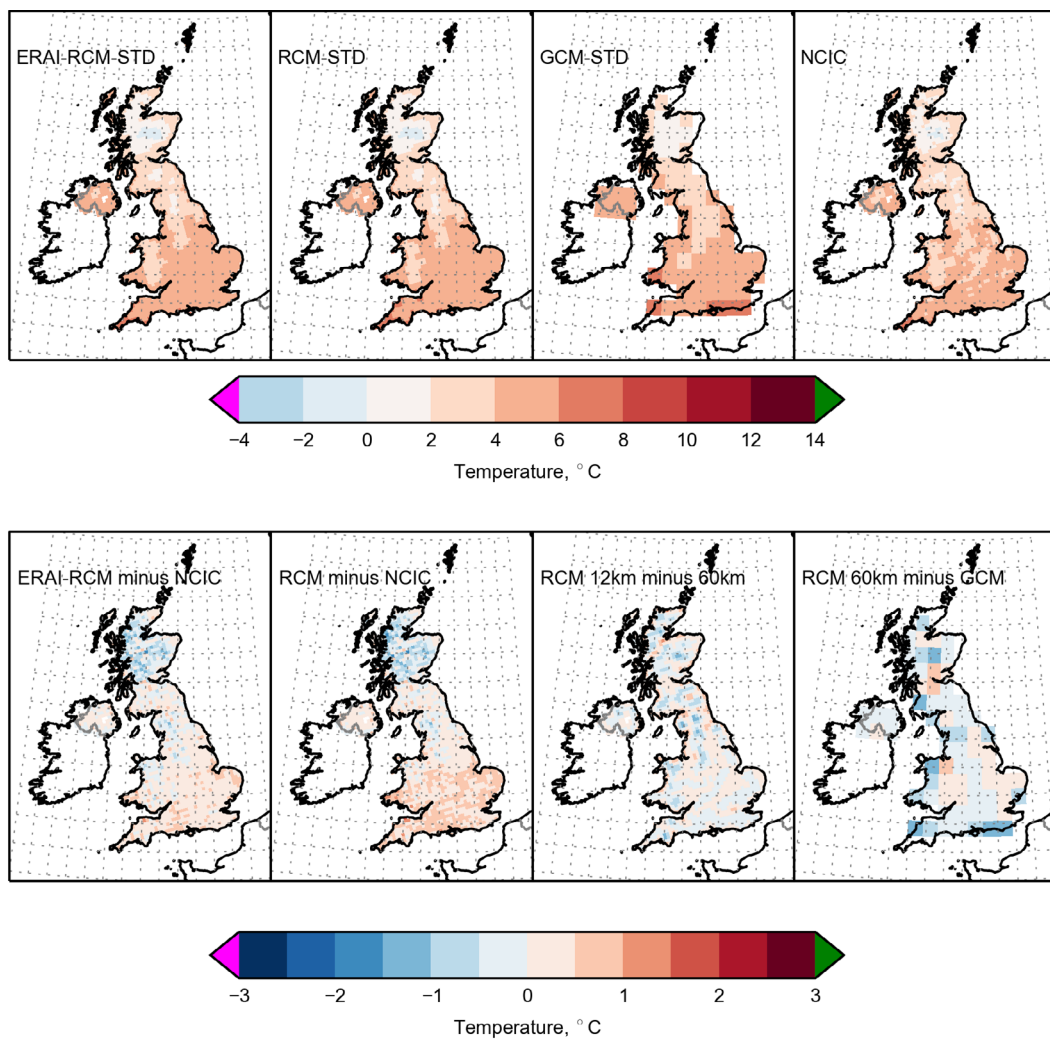
Figure 4.1a. Impact of dynamical downscaling on the climatological average of winter surface air temperature (°C) during 1981-2000, in the RCM-STD member of RCM-PPE, the Strand 3 perturbed parameter ensemble of 12km regional climate model simulations. Top left panel shows values from a simulation driven by reanalyses of observations, ERAI-RCM-STD[46]. The RCM-STD panel shows results from a 12km simulation driven by GCM-STD (centre right), the unperturbed member of the GC3.05-PPE global model ensemble with corresponding parameter settings. Verifying observations from NCIC (regridded to the RCM 12km grid) are shown in the top right panel. The bottom row shows biases in the ERAI-RCM-STD and RCM-STD simulations (left and centre left). The centre right panel shows differences between the RCM field at 12km resolution, and a version smoothed to the ~60km horizontal resolution of GCM-STD. This isolates spatial features simulated by the RCM at scales finer than the global model grid. The bottom right panel shows differences between the smoothed RCM field and GCM-STD, to show the impact of downscaling at the finest scale resolved by the driving model.

In both seasons, all the simulations capture the main spatial contrasts in the observed surface air temperature patterns (Figs. 4.1a,b), showing the coolest temperatures in the Scottish Highlands and the warmest in Southern England and East Anglia. In winter, RCM-STD is too warm in the south and too cool in Scotland (Fig. 4.1a). The southern bias is weaker in ERAI-RCM-STD, suggesting that this is partly inherited from the driving model in RCM-STD. However, the northern bias is present in both regional simulations, indicating that it probably arises from the simulation of regional physical processes. The effects of downscaling, at both 12km and 60km scales, reflect to a large extent straightforward lapse-rate effects associated with better resolution of orography.
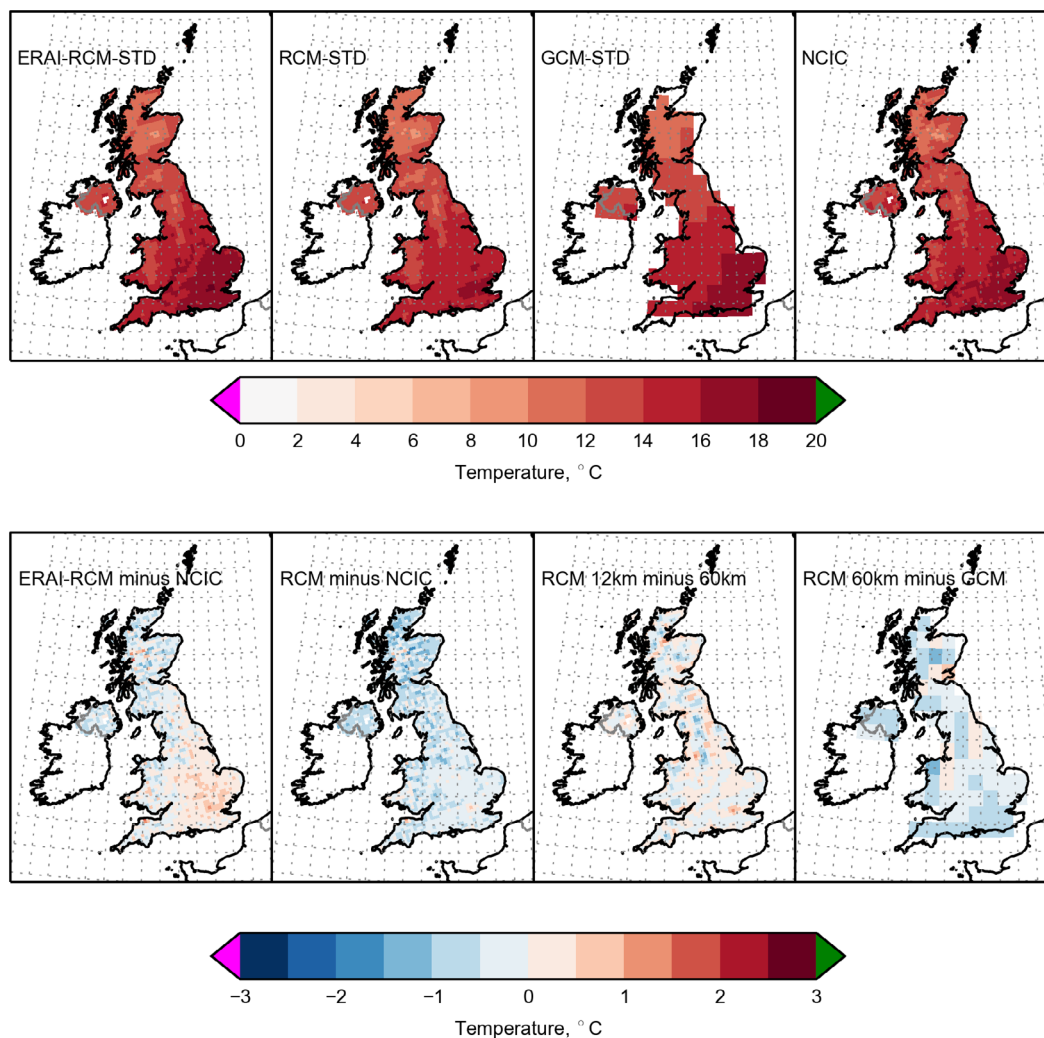
Figure 4.1b. As Fig. 4.1a, for effects of downscaling on summer air temperature during 1981-2000 in RCM-STD.

In summer (Fig. 4.1b), the impact of downscaling at the 12km scale again shows clear orographic effects. However, the difference (at 60km scale) from GCM-STD shows a general cooling in the regional model, associated with increased cloud cover. This increase is also found in other seasons. In test simulations in which the RCM is run at 60km resolution, cloud cover is similar to values found in the global model, confirming that the increase at 12km is an effect of enhanced resolution. Relative to observations, the summer cool bias found in RCM-STD is reduced, and in some regions reversed, in ERAI-RCM-STD.
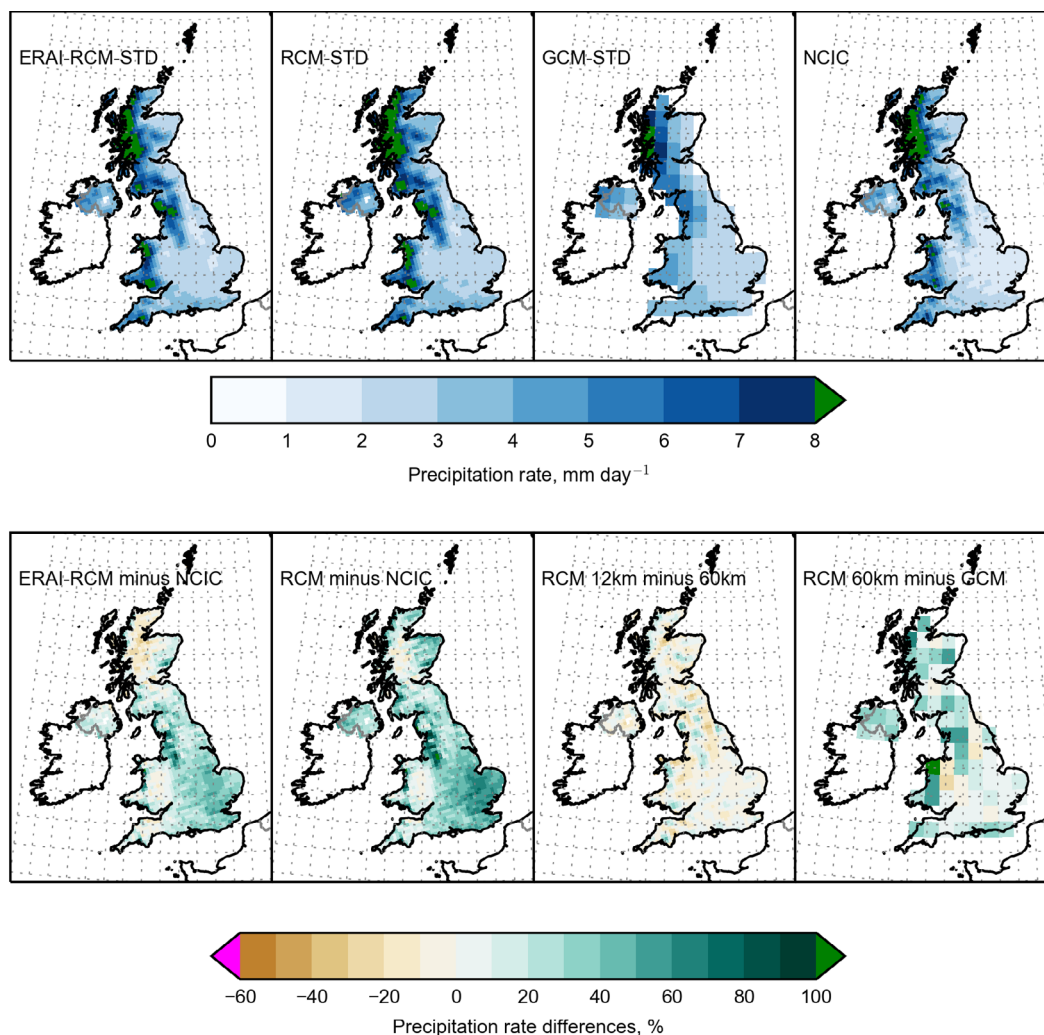
Figure 4.1c. As Fig. 4.1a, for effects of downscaling on winter precipitation during 1981-2000 in RCM-STD. Absolute precipitation values (top row) are shown in mm/day, while errors or differences (bottom row) are shown as percentage deviations from observations (left and centre left), the RCM-STD simulation smoothed to 60km scale (centre right) or the GCM-STD global simulation (right).

The climatological distributions of winter precipitation in both RCM simulations show clear benefits relative to GCM-STD, in capturing spatial details of the observed distribution. The patterns and intensities of areas of high precipitation in South West England, the Welsh mountains, the Lake District, South West Scotland and the Cairngorms are all represented with substantial skill, although there is a modest dry bias over much of North West Scotland. The climatological patterns are very similar in RCM-STD and ERAI-RCM-STD, indicating that biases in the driving model are not large enough to degrade significantly this aspect of downscaling performance. The fine-scale component of the RCM-STD pattern (Fig. 4.1c, bottom row, centre right) shows enhanced precipitation on the windward upslopes of mountains, demonstrating the effects of improved resolution of orography. In most parts of the UK there is, however, a broader-scale wet bias in RCM-STD. Some of this is inherited from the driving model, as the bias is less pronounced in ERAI-RCM-STD. In addition, precipitation is somewhat higher in RCM-STD than in GCM-STD (Fig. 4.1c, bottom right), consistent with the intensification of the hydrological cycle typically seen in regional models compared with coarser resolution driving simulations (e.g. Jones et al., 1995).
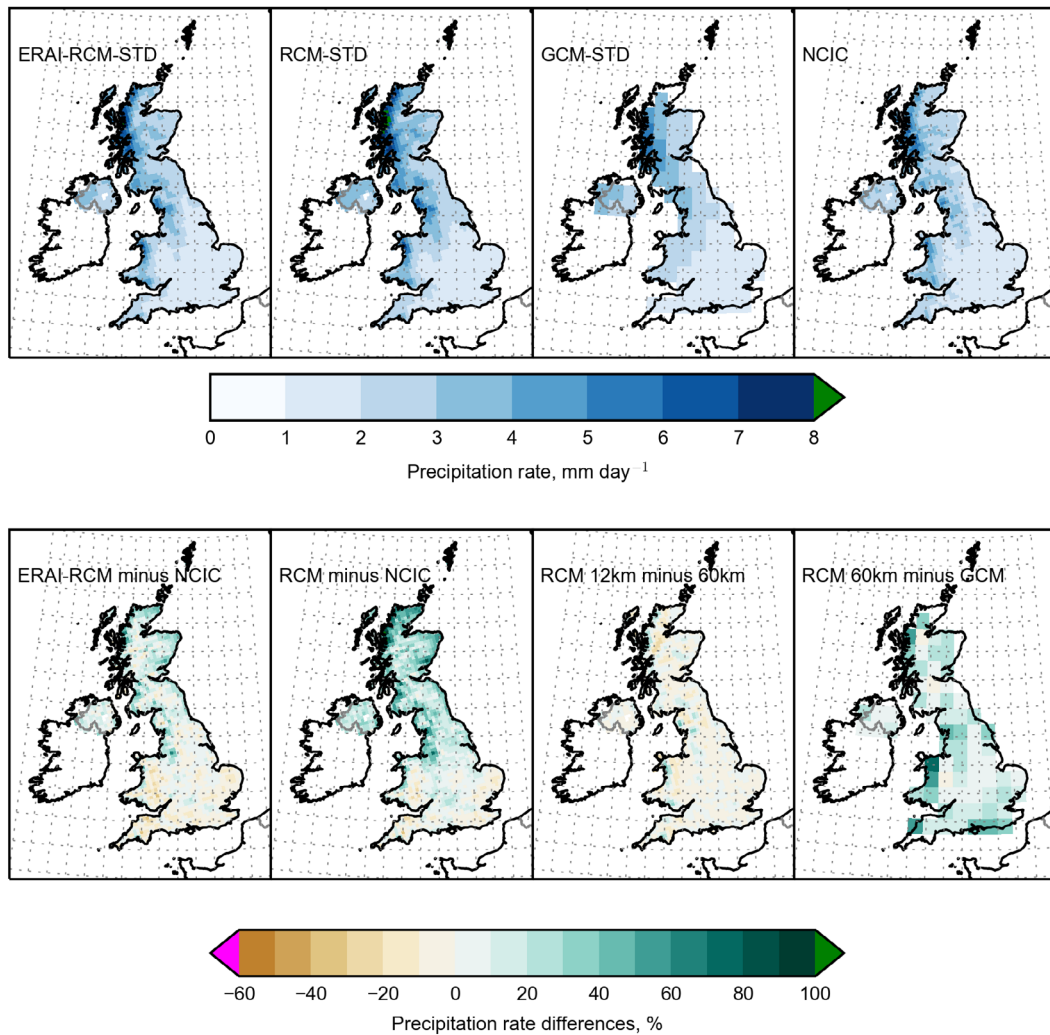
Figure 4.1d. As Fig. 4.1c, for effects of downscaling on summer precipitation during 1981-2000 in RCM-STD.

In summer, ERAI-RCM-STD shows generally good agreement with observed precipitation, with modest wet biases over parts of Northern England and Scotland (Fig. 4.1d). The RCM-STD simulation shows larger wet biases of typically 20-50% in these regions, inherited in part from GCM-STD (compare red and black curves in Fig. 4.4, bottom left panel).

We extend the assessment of precipitation in RCM-STD to consider biases in the fraction of wet days (Figure 4.2), and the average wet-day intensity (Figure 4.3). Here, wet days are defined as those in which accumulated precipitation exceeds 1mm.

The highest wet-day fractions are located in Northern Ireland, Scotland, the Lake District, Wales and South-West England, mostly over high terrain. Values are higher in winter than in summer (Fig 4.2, top left cf bottom left panels), and exceed 0.7 over North-West Scotland in winter. The wet-day fractions in RCM-STD are mostly higher than observed, especially in winter, during which the largest biases (of 30-50%) occur in East Anglia, the East Midlands and parts of South East England. In winter, the driving GCM-STD model also simulates fractions that are generally larger than observed. However, its biases are typically somewhat smaller than those of RCM-STD, consistent with the corresponding biases in average precipitation (Fig. 4.1c). In summer, wet-day fractions in GCM-STD are lower than in RCM-STD. The GCM-STD values show negative biases relative to observations over much of Southern England and Wales, and positive biases over Scotland.
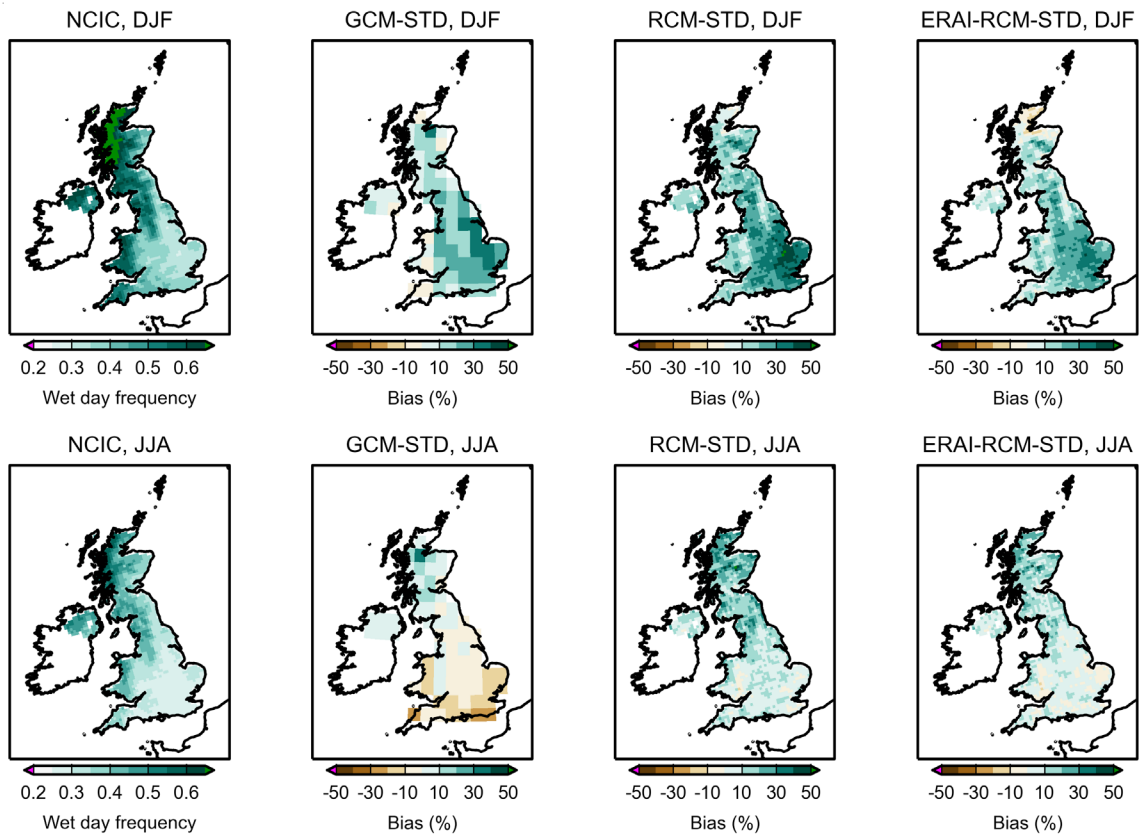
Figure 4.2. Biases in the simulation of the frequency of wet days during 1981-00 for winter (top row) and summer (bottom row). Here, a wet day is defined as a daily precipitation accumulation exceeding 1mm. Left panels show NCIC values from observations regridded onto the 12km RCM grid. Centre left and centre right panels show biases, expressed as a percentage of the observed value, in the simulations of GCM-STD and RCM-STD respectively. For GCM-STD, the biases are calculated relative to NCIC values regridded to its ~60km grid. Rightmost panels show biases from the reanalysis-driven simulation ERAI-RCM-STD. These can be compared with the RCM-STD biases to see the impacts in the latter of errors inherited from the driving model simulation (GCM-STD).

The reanalysis-driven ERAI-RCM-STD simulation gives predominantly positive biases in wet-day fraction, demonstrating that errors in the regional representation of hydrological cycle processes contribute to the biases found in RCM-STD. However, the positive biases are typically larger in RCM-STD than in ERAI-RCM-STD, indicating also a role for errors inherited from the driving model.

In RCM-STD, simulated values of mean wet-day intensity are too low in some high-elevation regions, particularly over the Welsh and Scottish mountains in winter (Fig. 4.3). There are also negative biases over much of Eastern and Southern England in summer. The patterns of bias are generally similar between RCM-STD and ERAI-RCM-STD, suggesting that the errors are to a substantial degree inherent to the regional model. Nevertheless, the negative biases over high ground are generally smaller in RCM-STD than in GCM-STD, probably due to better resolution of orographic effects in the regional model.
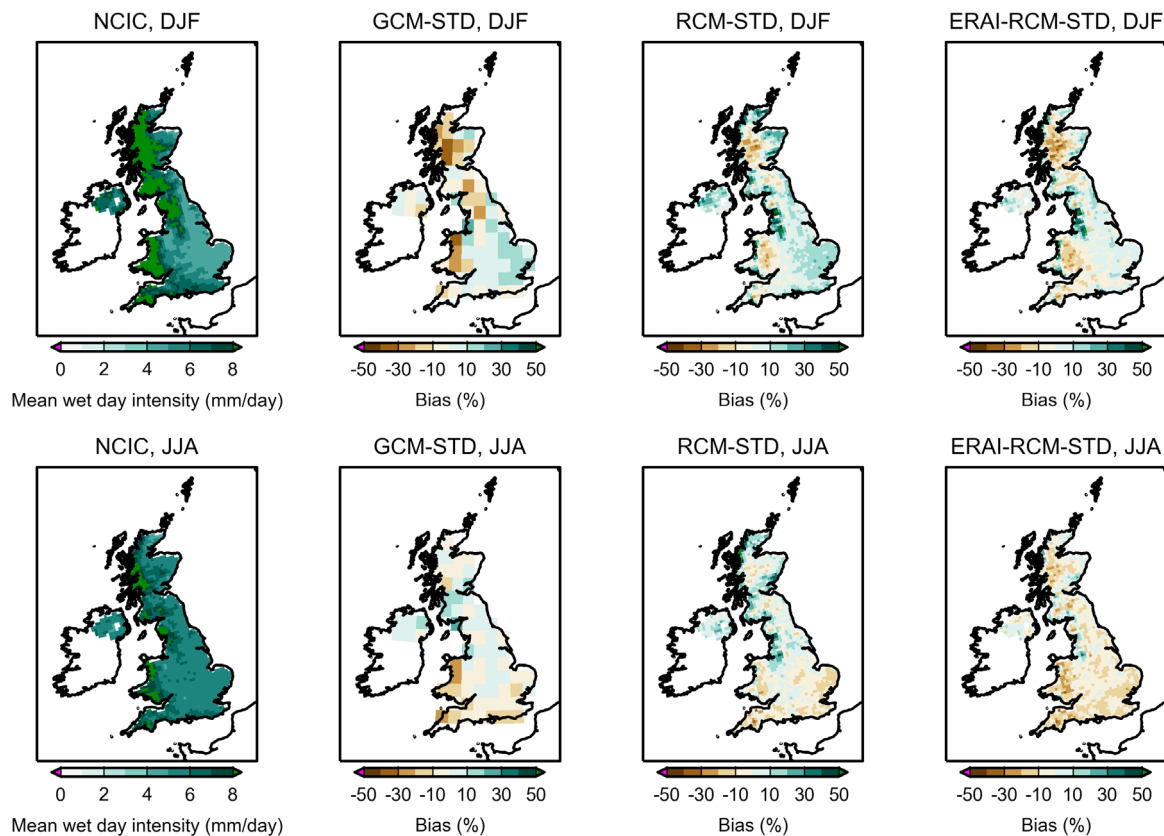
Figure 4.3. As Fig. 4.2, for biases in the simulation of the average intensity of wet-day precipitation. Wet days are identified as those with daily precipitation accumulations exceeding 1mm. The NCIC observations (left panels) are shown in mm/day, and simulation biases (other panels) are shown as percentages of the observed value.

The widespread positive bias in the fraction of wet days, shown for RCM-STD in Fig. 4.2, is present in RCM-PPE as a whole. The NCIC observations for 1981-2000 give average fractions over the UK of 0.47 in winter, and 0.37 in summer. In comparison, the ensemble-mean values from RCM-PPE are 0.58 and 0.44 respectively. Simulating too many wet days is a common error in climate models (e.g. Sun et al., 2006), and most of the multi-model RCM simulations in EuroCordex also simulate excessive wet day frequencies over the British Isles (Rajczak and Schär, 2017).

For Europe, Figures 3.13a,b show winter and summer ensemble-mean biases in RCM-PPE for surface air temperature and precipitation. The pattern of cold winter bias found in the driving global simulations is largely replicated in RCM-PPE, but the magnitude of the bias is smaller in most regions. This is likely to be due, at least in part, to the aforementioned increase in cloud cover in the regional simulations. The pattern of summer errors is again similar to GC3.05-PPE, with slightly smaller warm biases in RCM-PPE over South-Eastern Europe.

Increased precipitation in the regional model (discussed above) leads to a widespread increase in the ensemble-mean winter wet bias in RCM-PPE, compared with GC3.05-PPE (Fig 3.13b). Biases exceeding 100% occur in some coastal and high elevation regions, notably in parts of Southern Europe including Italy, Croatia, Greece, and the Carpathian mountains. Note, however, that the verifying E-OBS dataset (Table E.2) is likely (in common with all datasets derived from rain gauge measurements) to underestimate observed precipitation due to undercatch (see section 3.4b). In high elevation regions, an additional source of underestimation in E-OBS arises from an insufficient density of observing stations (Hofstra et al., 2009). In summer, the general increase in precipitation in the regional model leads to the wet bias over Scandinavia being slightly larger in RCM-PPE than in GC3.05-PPE, while the dry biases over much of Central and Southern Europe are smaller, or in some regions (e.g. Spain and Southern Italy) reversed in sign.

In winter, RCM-PPE shows a pattern of bias in interannual variability of surface air temperature similar to that of GC3.05-PPE (Fig. 3.14a). There is too little variability over Scandinavia and too much over South-Eastern Europe (the latter bias is somewhat smaller in RCM-PPE). The general similarity in the patterns reflects the strong control exerted by the driving model on the large-scale atmospheric circulation in the RCM. In summer, interannual variability is too large in RCM-PPE in most European regions (Fig. 3.14a). The biases are slightly larger in the global ensemble over the Scandinavian mountains and slightly larger in the regional model ensemble over much of Central and South-Eastern Europe (maximum values 0.6-0.8°C). The latter bias is a persistent problem in climate models, and is likely due to errors in local soil moisture feedbacks. Previous studies have argued that this overestimation of interannual variability in warm summer months could have implications for the reliability of climate change projections (e.g. Sørland et al 2018). Over the UK, where there is a strong maritime influence, the biases for 1981-2000 are lower.

As for surface air temperature, RCM-PPE shows regional patterns of bias in interannual precipitation variability that are closely related to those in the global simulations (Fig. 3.14b). The regional model ensemble generally simulates slightly higher variability than GC3.05-PPE over Eastern and Southern Europe, exacerbating the positive bias found in the driving simulations in winter, but reducing the negative biases over the Baltic States and the Balkans in summer.
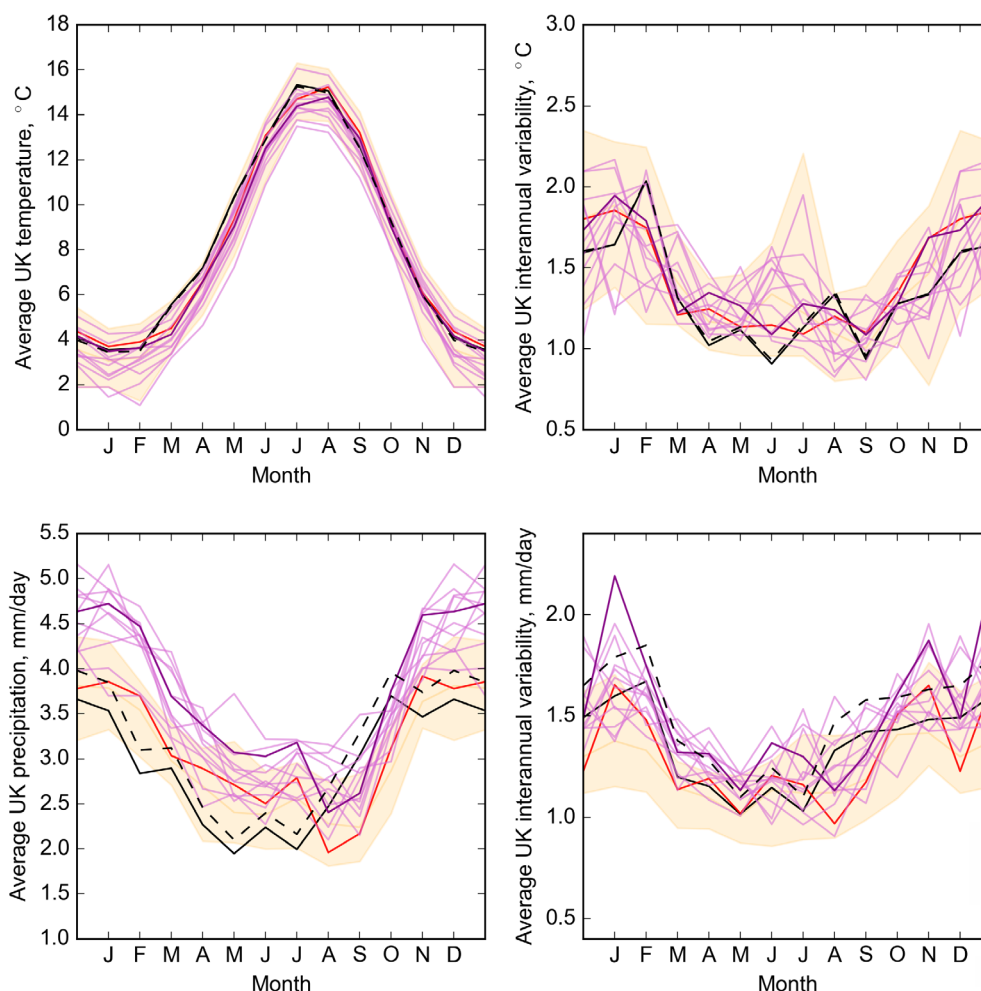

Figure 4.4. Left panels show seasonal cycles of UK average surface air temperature (ºC, top) and precipitation (mm/day, bottom) for 1981-2000. Right panels show UK averages of local values of interannual variability (defined as the standard deviation of seasonal averages during 1981-2000) for surface air temperature (top) and precipitation (bottom). Members of RCM-PPE are shown in pink, with RCM-STD in purple. The orange line shows the median, and the orange shading the range, of values from the 15-member ensemble of GC3.05-PPE simulations (see also Fig. 3.15). Twelve of these drove the RCM-PPE simulations. One of the driving global simulations, GCM-STD, is shown in red. Observations from the National Climate Information Centre (NCIC) and E-OBS (Table E.2) are shown as dashed and solid black lines, respectively. All datasets were regridded to the GC3.05 (~60km) grid, prior to calculation of these diagnostics.

Figure 4.4 shows seasonal cycles of UK-average surface air temperature and precipitation for individual RCM-PPE members, compared against observations from NCIC and E-OBS. The range of results from GC3.05-PPE is shown in orange shading for comparison (STD is plotted in red, and other individual members of GC3.05-PPE are shown in Fig. 3.15). The ranges of surface air temperature in the two ensembles are generally similar, with a slight cool shift in the envelope of values in RCM-PPE during spring and summer. The RCM-PPE range encompasses observations in all months apart from April and May. The distribution of RCM-PPE precipitation values is shifted higher than its GC3.05-PPE counterpart, throughout the annual cycle. This is due to the general increase in the regional model discussed above. Observed precipitation lies below (or at the low end of) the RCM-PPE range in most months, the exceptions being August to October. This is true even for the higher estimates in the NCIC dataset, which is based on a much denser station network than E-OBS (section 3.4b).

In RCM-PPE, the spread of monthly values for interannual variability in surface air temperature (Fig. 4.4, top right) follows closely that of GC3.05-PPE, and encompasses the observed values in all months except June (in which the simulated values are all too high). The main observed seasonal variations (highest values in December to February, lower values from late spring to early autumn) are captured in the simulations, apart from an anomalously high value in July in one member of GC3.05-PPE (Fig. 3.15) and its corresponding RCM-PPE member.

Over the UK, the ensemble average of interannual variability in precipitation is somewhat higher in RCM-PPE than in GC3.05-PPE (Fig. 3.14b). Therefore the range of values for simulated variability found across RCM-PPE members (while overlapping substantially with the GC3.05-PPE range) is shifted higher (Fig. 4.4, bottom right). This improves the correspondence with observations (particularly the NCIC dataset).

In general, good agreement is seen between RCM-PPE and GC3.05-PPE in relation to large-scale climatological averages of surface air temperature and precipitation, and their interannual variability. This is consistent with the aim of our one-way nesting approach (section 4.1), that the RCM should not diverge strongly from its driving global model simulation at large scales. It is on finer spatial scales (e.g. Fig. 4.1) and shorter time scales (discussed below) that we may expect the RCM to differ more substantially from the driving model, and where the higher resolution may offer added value.

The simulation of cold daily extremes in winter is assessed in Fig. 4.5a (top row). The metric used is the 1st percentile of daily winter values of diurnally averaged surface air temperature. This is shown for observations in the left panel, and can be interpreted approximately as a typical coldest day in an individual winter. Values range from -1 to -5°C across much of Southern England, to -9 to -11°C in some high-elevation regions of Northern Scotland. Cold biases exceeding 3°C are present in the GC3.05-PPE ensemble-mean over much of Northern England and Scotland (centre left panel), with warm biases in some coastal grid boxes, including across Southern England. Corresponding biases in RCM-PPE show a detailed spatial pattern at native 12km resolution (right panel), but are generally improved relative to the driving simulations. This can be seen by comparing the GC3.05-PPE biases against those found when the regional model results are smoothed to the same spatial resolution (centre right panel). The cold biases in the global model in high-elevation regions, and the coastal warm biases, are both reduced in the regional simulations.
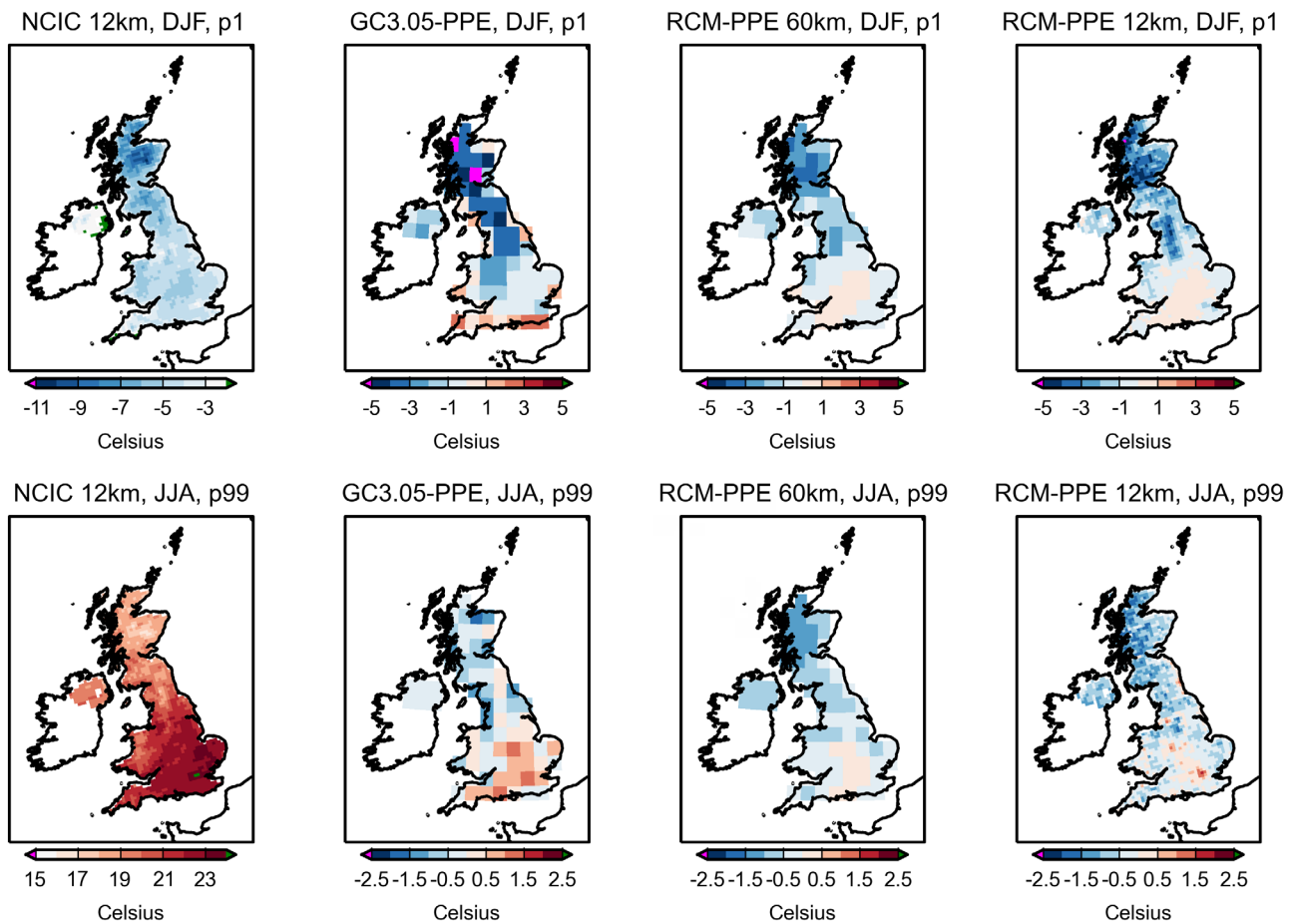
Figure 4.5a. Observed and simulated daily temperature extremes in winter (top row) and summer (bottom row). Left panels show the 1st percentile of diurnal average temperature in winter (cold days), and the 99th percentile in summer (hot days), found in the NCIC observed dataset during 1981-2000. Percentiles were calculated empirically, by ranking the simulated daily values. Centre left panels show biases in the ensemble-average simulated values for the corresponding metrics, from the 15-member GC3.05-PPE of Strand 2. Centre right panels show ensemble-average biases from the 12-member RCM-PPE of Strand 3 simulations, following smoothing to the ~60km horizontal resolution of the driving GC3.05-PPE simulations. Rightmost panels show the RCM-PPE ensemble average biases, at full 12km resolution. The NCIC data was regridded to the relevant 12km or 60km model grid, in each case.

In summer, we consider the 99th percentile of diurnal mean surface air temperature values (which corresponds approximately to a typical hottest day in an individual season), as a metric of summer heat extremes. The ensemble-mean values in RCM-PPE are generally cooler than in GCM3.05-PPE, leading to reduced warm biases over parts of the Midlands and Southern England, but larger cool biases over much of Northern England and Scotland. The distribution of RCM-PPE biases at 12km resolution (bottom right panel in Fig. 4.5a) shows positive values over Birmingham and London. In the RCM urban effects are represented in a simple way, as one of nine "tiles" used to represent sub-grid scale heterogeneity in land surface characteristics (Best et al., 2011). The forthcoming 2.2km simulations will include a more sophisticated urban scheme using two tiles to represent roof and street canyon facets, with surface parameters determined from the morphology and materials properties of relevant cities (Porson et al., 2010).
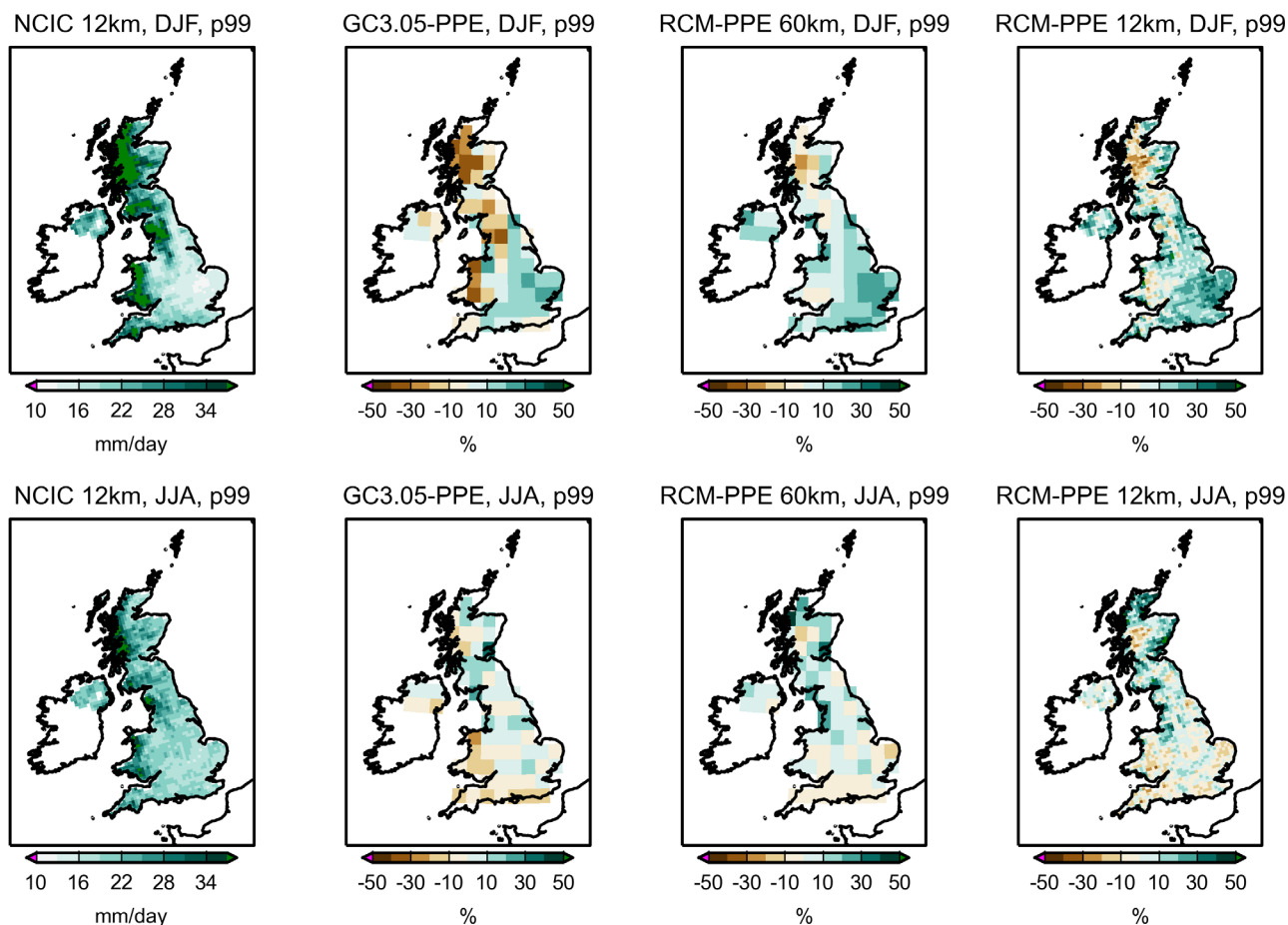
Figure 4.5b. As Fig. 4.5a, for precipitation extremes, defined here as the 99th percentile of daily accumulated values during 1981-2000. The percentile values are calculated from all daily accumulations, including dry as well as wet days.

Precipitation extremes are evaluated as the 99th percentile of daily values in Fig. 4.5b, for winter (top row) and summer (bottom row). These are all-day percentiles calculated by considering dry and wet days together. The highest observed values (in excess of 30mm, see left panels) are found in high-elevation regions on the western side of Great Britain. In winter, the RCM-PPE values are generally higher than those of GC3.05-PPE. This shows that the general increase in precipitation in the regional model affects extreme wet events as well as the climatological average. This increase leads to substantial reductions in the dry bias found in GC3.05-PPE over the Welsh mountains, the Lake District and Northern Scotland. However, RCM-PPE overestimates the intensity of observed heavy events over Eastern and Southern England, to a greater degree than GC3.05-PPE. In summer, average regional biases at the 60km scale are generally smaller than ± 20% in both ensembles. At the 12km scale, RCM-PPE produces biases exceeding ± 30% in some locations (note, for example, the winter dry bias over parts of the North-West Highlands in Scotland). However, the local biases are generally below 30%, which indicates significant skill in replicating the complex distributions seen in observations.
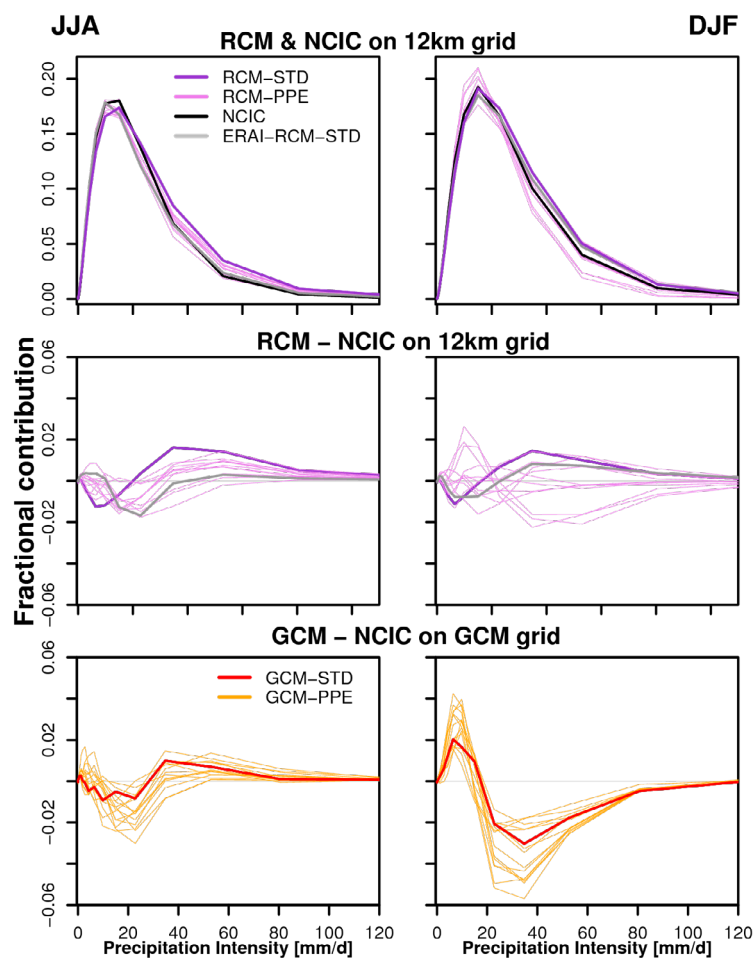
Figure 4.6. Fractional contribution of daily precipitation events within 20 intensity bins to total accumulated UK precipitation in summer (left) and winter (right), during 1981-2000. The contributions were diagnosed by assigning each day from every UK grid point to the relevant bin, and multiplying the number of counts in each bin by the average intensity. Dry days are assigned to the lowest bin. Top panels show results for each of the 12 members of RCM-PPE (pink, with RCM-STD in purple), plus the reanalysis-driven simulation ERAI-RCM-STD (grey) and NCIC observations (black). These results are calculated from data fields on the native 12km grid of the RCM. Middle row shows corresponding differences between the RCM simulations and observations. Bottom row shows differences between the 15 members of GC3.05-PPE and observations (orange, with GCM-STD in red). For the bottom row, the NCIC observations were regridded to the 60km GC3.05 grid, prior to calculating the diagnostics. The bin boundaries (mm/day) are: 0.05, 0.16, 0.18, 0.27, 0.41, 0.62, 0.95, 1.45, 2.2, 3.4, 5.1, 7.8, 11.9, 18.1, 27.5, 42.0, 63.9, 97.4, 148.4, 500.0.

Figure 4.6 evaluates the shape of the distribution of daily precipitation events, for winter (right panels) and summer (left). The curves are constructed from all days at all UK grid points during 1981-2000, showing the fractional contribution from 20 intensity bins to total UK precipitation within that period. Dry days are assigned to the lightest intensity bin. The top row shows observations and regional model simulations at 12km scale, and the middle row shows the same RCM results as differences from observations. The ERAI-RCM-STD simulation replicates the observed distributions quite well, although the relative[47] contribution from extreme events of 40mm or more is overestimated in winter, and that from events in the range 10-30mm is underestimated in summer. In winter, some members of RCM-PPE overestimate the number of 10-20mm events at the expense of heavier accumulations, while in summer most members overestimate the number of events of 40mm or more. Since these biases in heavy summer events are not present in ERAI-RCM-STD, it is likely that errors in the driving model boundary conditions are at least partly responsible for these. The regional impacts of parameter perturbations are another potential factor, particularly in explaining the spread in biases between alternative RCM-PPE members.

---

[47] As the statistic in Fig. 4.6 is a normalised metric, biases in average precipitation do not affect the distribution of values.

The bottom row in Fig. 4.6 shows differences between the GC3.05-PPE simulations and observations at the 60km scale. When smoothed to this scale, the RCM-PPE distributions (not shown) reveal error characteristics similar to those found at 12km resolution (middle row). The global model substantially underestimates the fractional contribution from events above 20mm in winter. This is captured better in RCM-PPE, albeit with a spread of member-specific biases distributed around the observed values. In summer, the global and regional model ensembles both tend to overestimate events over 30mm at the expense of lighter daily accumulations.
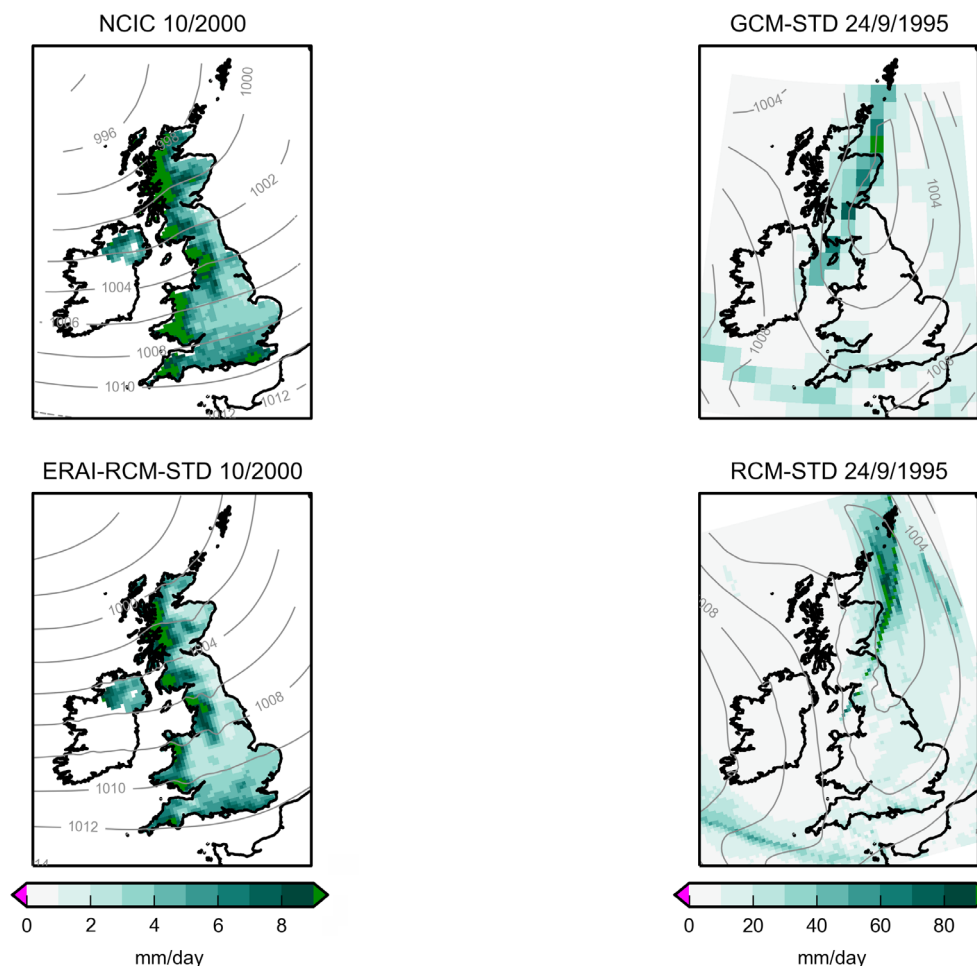

Figure 4.7. Examples of heavy precipitation events simulated by RCM-STD. Left panels shows monthly average precipitation (mm/day) and sea level pressure contours (hPa) for October 2000 from NCIC and ERA-Interim observations (top), and from the ERAI-RCM-STD simulation driven by ERA-Interim lateral boundary conditions (bottom). Right panels show daily sea level pressure and precipitation fields for 24 September 1995 from the RCM-STD simulation (bottom), and the driving GCM-STD global simulation (top). This was the day on which RCM-STD simulated the highest value of precipitation in an individual 12km grid box, during the 1981-2000 period.

In Figure 4.7 we show two specific examples of heavy precipitation events simulated by the regional model. The first relates to the floods of autumn 2000, which arose from the passage of a series of intense synoptic storms over the UK that led to insured losses exceeding £1 billion (Pall et al., 2011). The top-left panel shows the observed monthly averages of precipitation and circulation for October 2000, revealing high precipitation totals exceeding 8mm/day over most of the high ground on the western side of Great Britain, and also the South Downs. The ERAI-RCM-STD simulation, driven by the observed circulation, replicated the main features of the observed precipitation pattern well (Fig. 4.7, bottom-left panel). However, the maximum levels were underestimated, with areas exceeding 8mm/day absent from the South Downs, and confined to smaller areas than observed in South West England, Wales, the Lake District and Scotland.

The right panels in Fig. 4.7 show an example of extreme daily precipitation, from RCM-STD. The driving GCM-STD model simulated two frontal systems on the relevant day, one situated over the South Western approaches, and the other extending from North East to South West over Scotland. The regional model replicates the low pressure system inherited from the driving model (centred off the coast of North East Scotland) and simulates the aforementioned frontal systems with greater spatial detail, showing more sharply-defined bands of precipitation maxima. The front over Scotland yielded the highest local precipitation event during the 1981-2000 simulation of RCM-STD.

In summary, RCM-PPE provides significant added value to the simulation of regional variations in UK precipitation, especially in coastal areas and regions of high orography. Improvements are seen in both the climatological average and the representation of extreme heavy events. Benefits from enhanced resolution are also seen in the regional model simulations of surface air temperature, particularly in the replication of extreme winter cold days as well as in spatial patterns of long-term averages. In general, the regional ensemble simulates more precipitation compared with its driving global simulations, along with more extensive cloud cover. The latter contributes to the simulation of slightly cooler UK surface temperatures during spring and summer, compared with GC3.05-PPE. The global and regional simulations show similar spatial and seasonal characteristics in their simulations of interannual variability, for both surface air temperature and precipitation. Over the UK, however, RCM-PPE produces higher variability for precipitation, which improves the correspondence with observations.

4.5. Future changes in the strand 3 projections

In this section, we provide a few examples of projected changes in the Strand 3 simulations for 2061-2080 relative to 1981-2000, comparing with changes for the same period in GC3.05-PPE. A broader comparison between Strands 1-3 is provided in section 5.

For winter surface air temperature, Figure 4.8a shows the second lowest, central and second highest responses obtained by ranking locally the changes in the 12 member ensemble, at the native 12km resolution of the RCM. Use of the second lowest and second highest responses characterises the range of responses within the ensemble, while reducing the influence of any outliers. The central response is taken as that of the 7th ranked member. In general, different simulations will contribute the low, central or high responses in different regions, so the maps do not represent any specific model projection. The changes range from 1-3°C at the low end (dependent on location) to 3-4°C at the upper end. The UK average changes are 1.9°C, 3.1°C and 3.4°C for the low, central and high responses respectively. The gap between the central and high changes is narrower than in Strands 1 and 2 (see Fig. 5.4a and discussion in section 5.1). This is because the other Strands contain information from multi-model simulations alongside perturbed parameter ensemble results, leading to a broader range of responses (sections 3.5 and 5.1).
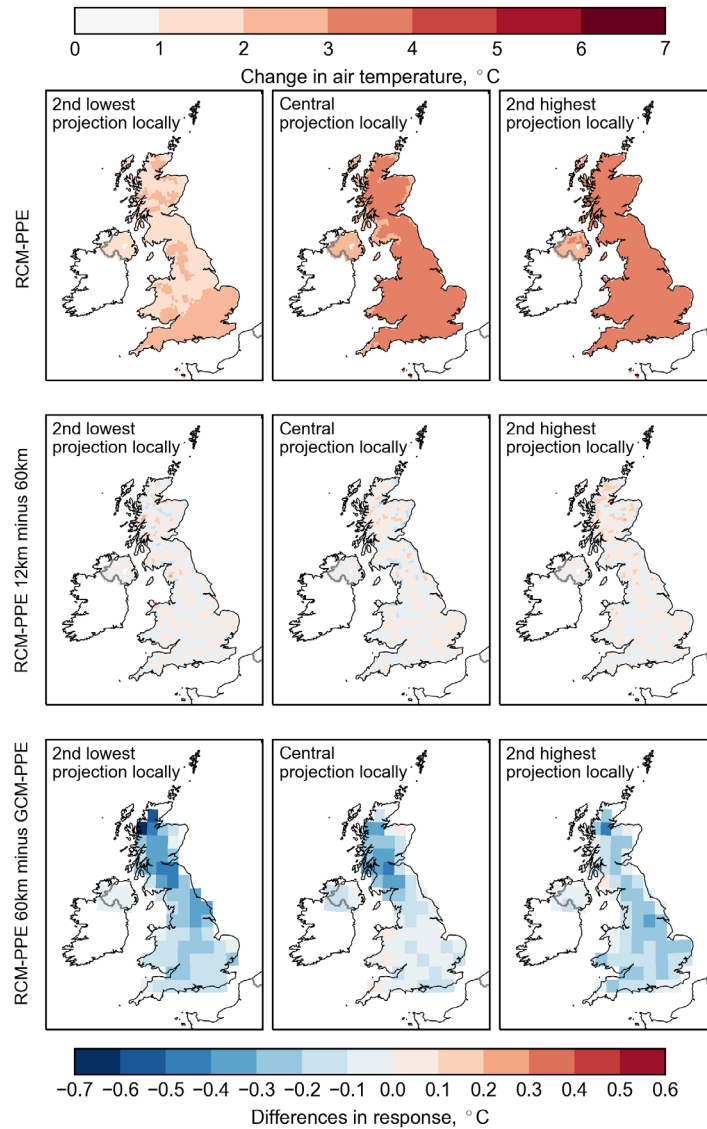
Figure 4.8a. Impact of dynamical downscaling on the range of changes in winter surface air temperature (°C) for 2061-2080 relative to 1981-2000, in the RCM-PPE projections. Top row shows the second lowest, central and second highest responses obtained by ranking the changes in the 12 member ensemble locally, at the native 12km resolution of the RCM. Use of the second lowest and second highest responses reduces the influence of any outliers on the diagnosed ranges of change. The central response is taken as that of the 7th ranked member. In general, different simulations will contribute the low, central or high responses in different regions. Middle row shows differences between the top row, and the corresponding responses obtained when the relevant RCM-PPE members are smoothed to the ~60km horizontal resolution of the driving global simulations. This isolates spatial features simulated by the RCM at scales finer than the global model grid. The bottom row shows differences between the second lowest, central and second highest smoothed RCM-PPE responses and those of the 12 driving members of GC3.05-PPE, labelled here as GCM-PPE. This shows the impact of downscaling on the range of changes obtained at the finest scale resolved by the driving model.

The bottom panels in Fig. 4.8a show differences[48] between the low, central and high responses of RCM-PPE (when smoothed to the 60km global model grid) and those of the 12 driving simulations. Comparing against these, rather than the range given by the full 15 members of GC3.05-PPE, allows us to isolate cleanly the impact of downscaling at the 60km scale. The results show a modest cool shift in the distribution of changes, the largest values (of 0.4-0.5°C) occurring over parts of Scotland. Averaged over the UK, the low, central and high changes are 0.23°C, 0.14°C and 0.17°C cooler in RCM-PPE. The middle panels show the differences in the low, central and high responses in RCM-PPE at 12km resolution, compared to those at 60km resolution. This reveals the effects of downscaling at native regional model resolution, which are small for these climatological average changes in winter surface air temperature.

---

[48] These are differences between the 2nd lowest, central and 2nd highest changes in each ensemble, rather than 2nd lowest, central and 2nd highest values of the twelve differences between each RCM-PPE member and its driving simulation.
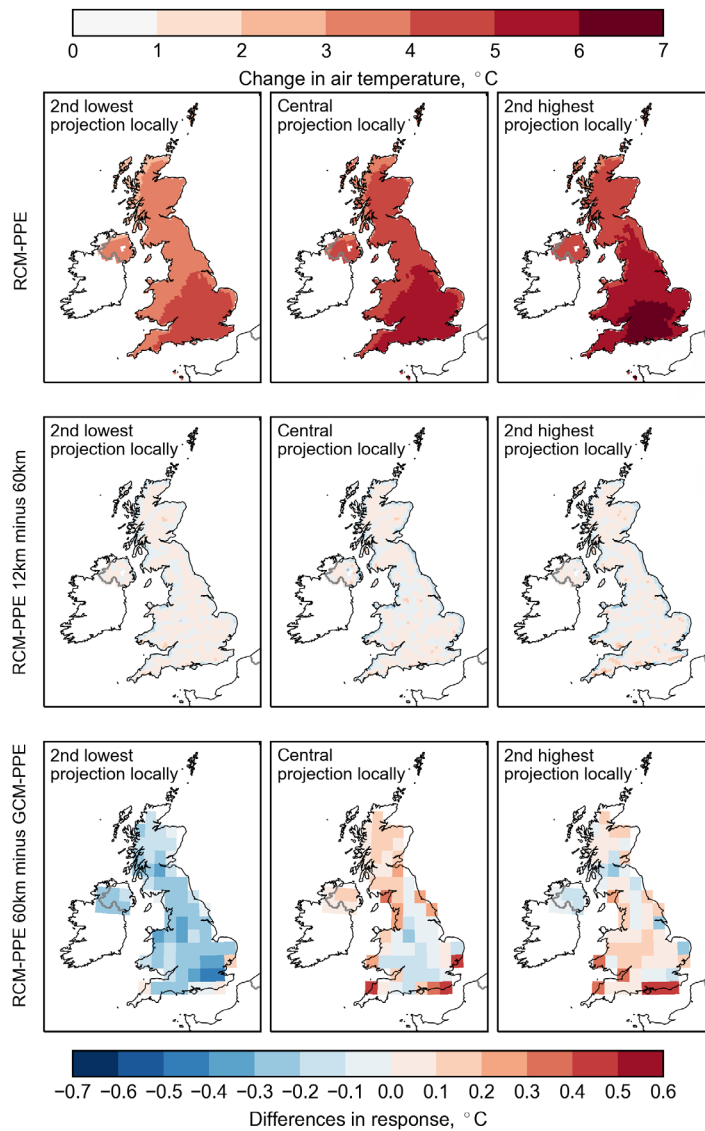

Figure 4.8b. As Fig 4.8a, for changes in summer surface air temperature for 2061-2080 relative to 1981-2000.

This is also the case for summer surface air temperature (Fig. 4.8b, middle row). However, at the global model grid scale the low-end changes are cooler by 0.1-0.4 °C in RCM-PPE, the largest differences occurring mostly in Southern England. At the upper end of the range, the impacts of downscaling vary in sign, with slightly higher warming in RCM-PPE (maximum differences 0-4.-0.5°C) at a number of coastal grid boxes in England and Wales. However, these effects are generally modest compared with the overall ranges of change in RCM-PPE (top panels). Typically, the difference in warming between the low and high outcomes amounts to ~1.5°C. The largest changes occur in Central and Southern England, in common with Strands 1 and 2 (see Fig. 5.4b).
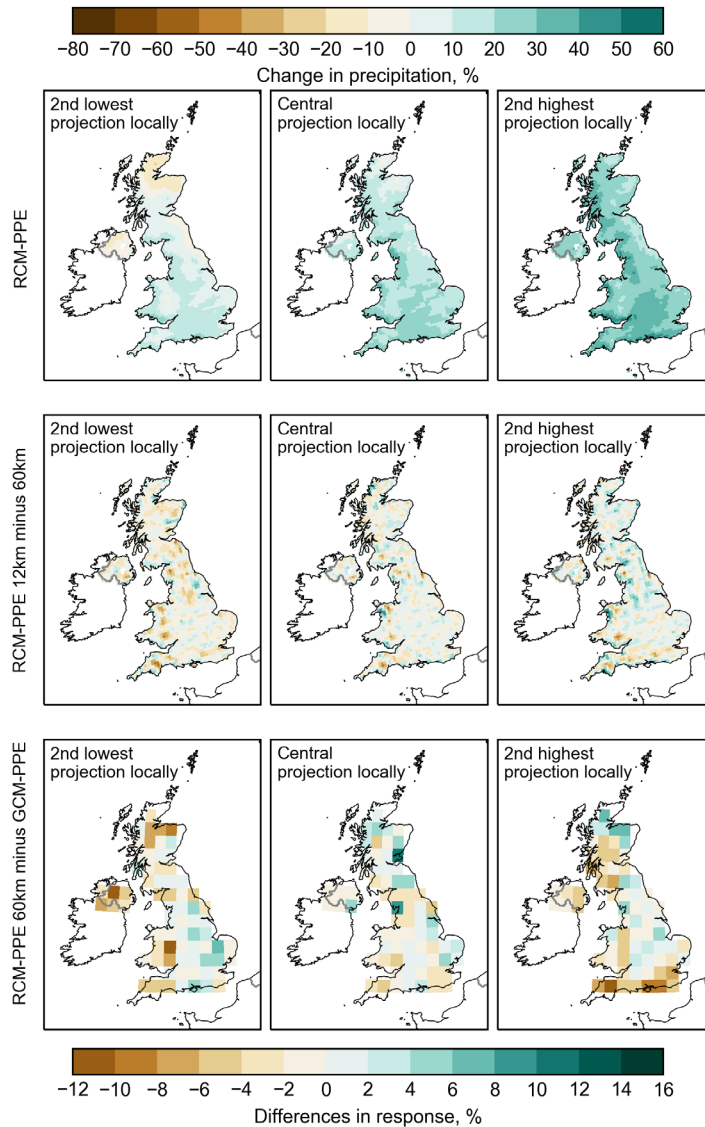
Figure 4.8c. As Fig 4.8a, for changes in winter precipitation (%) for 2061-2080 relative to 1981-2000.

For winter precipitation (Fig. 4.8c, top row), increases of uncertain magnitude are projected in most regions, although the low-end changes show the possibility of reductions exceeding 10% over parts of Northern Scotland. The influence of local (sub-60km) scales on the lower and upper limits of the RCM-PPE ranges show a variety of values, approaching ±10% in specific locations. At the global model scale, the high-end changes in RCM-PPE are smaller (typically by 4-10%) in East Anglia, Southern England and in the lee of the Welsh mountains, with higher values in Northern Scotland. Along the Buchan coast, the low end of the range is also drier, leading to a significant increase in the range of responses compared to the driving simulations. Central changes are generally similar, the UK averages amounting to 17% in both the 12 driving global projections and RCM-PPE.

The RCM-PPE simulations show little future change in the occurrence of wet winter days. The increases in average precipitation arise mainly from increases in average wet-day intensity, which typically range from 10-40%, dependent on ensemble member and location. Similar results are found in the 15-member multi-model ensemble of 12km EuroCordex simulations, which consists of a partly-filled matrix of combinations of six RCMs and five driving CMIP5 models (Rajczak and Schär, 2017).
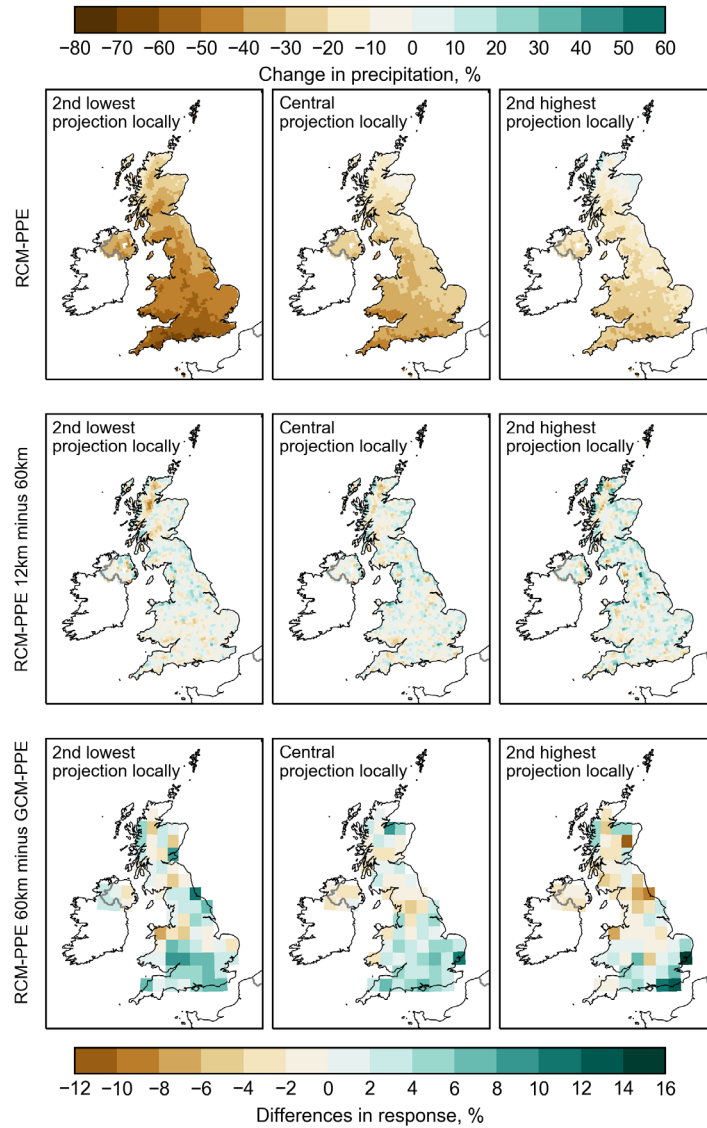
Figure 4.8d. As Fig 4.8a, for changes in summer precipitation (%) for 2061-2080 relative to 1981-2000.

Changes in summer precipitation (Fig. 4.8d) reveal a consistent drying response in RCM-PPE in most regions. The only exceptions are some coastal areas of Northern Scotland, where the high responses shows the potential for some small increases. The UK average reductions are 41% (low), 26% (central) and 18% (high). Rajczak and Schär (2017) find a smaller level of drying in the EuroCordex ensemble mentioned above. In this, spatial average changes for the British Isles[49] revealed that three simulations gave a small increase (maximum value 10%), and twelve of the fifteen simulations simulated a drying. However, the largest reduction was only 28%, similar to the median result from Strand 3.

---

[49] The Rajczak and Schär results provide an approximate but not precise comparison with those of Fig. 4.8d. Their results are for average precipitation changes over the whole British Isles (including the Republic of Ireland), and consider the period 2070-2099 relative to 1981-2010. The average changes quoted from Strand 3 are for the UK, and consider 2061-2080 relative to 1981-2000.

In RCM-PPE, changes in average wet-day intensity range from reductions of ~20% to increases of a similar magnitude. In contrast, there is a strong signal for reductions in the fraction of wet days. These typically vary between 10% and 50% across ensemble members and regions of the UK, with the largest reductions occurring in Southern England and Wales. Most of the EuroCordex simulations show an increase in average wet-day intensity and a reduction in wet-day fraction in summer. The ensemble average increase in wet-day intensity over the British Isles is 5.2% (Rajczak and Schär, 2017), compared with an ensemble average reduction of 3.3% over the UK in RCM-PPE. For wet-day fraction, only two of the fifteen EuroCordex members project changes for the British Isles that are larger than the ensemble-mean reduction of 26% (in the UK average) that is found in RCM-PPE.

In summer, the local contributions to the RCM-PPE ranges of change in average precipitation amount typically to a few percent (Fig. 4.8d), with broader uncertainties in the 60km-scale changes. In the low-end responses, the RCM-PPE results show slightly smaller decreases in precipitation than their driving model counterparts over Southern England and South Wales, however the downscaled changes still show reductions exceeding 40% in these regions. The effects of downscaling also ameliorate the drying in the central outcomes by a few percent, over much of England and Wales.
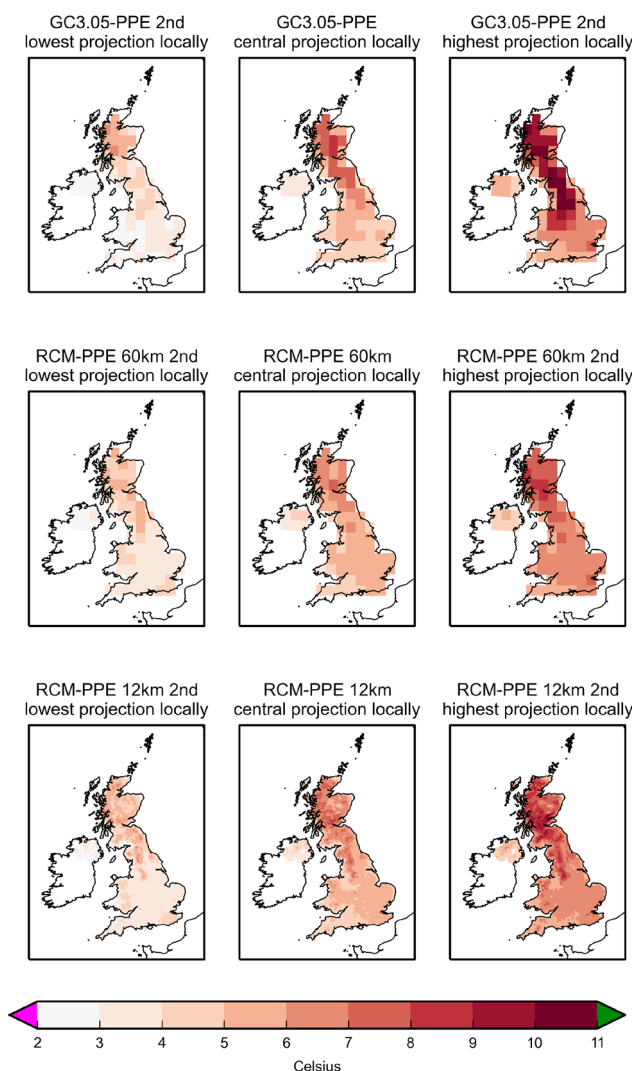

Figure 4.9a. Second lowest, central and second highest projected changes in surface air temperature (°C) on cold winter days in RCM-PPE, for 2061-2080 relative to 1981-2000. Cold winter days are defined as in Fig 4.5a. Top row shows the ranges of response given by ranking locally the responses of the 12 members of GC3.05-PPE that drove members of RCM-PPE, with the central projection defined as the 7th ranked response. In general, different simulations will contribute the low, central or high responses in different regions. Middle row shows corresponding ranges from RCM-PPE itself, following regridding of RCM results to the GC3.05 60km grid. Bottom row shows changes from RCM-PPE at its native 12km resolution, revealing local detail at scales unresolved by GC3.05.

Projected changes in surface air temperature values associated with cold winter days are presented in Figure 4.9a, for 2061-2080 relative to 1981-2000. These results employ the same metric of cold extremes evaluated against observations in Fig. 4.5a. They represent a future change in the *intensity* of a typical coldest day in winter, compared with the baseline period. The top row shows low, central and high changes from the 12 driving GC3.05-PPE projections, compared against those from RCM-PPE (at the 60km scale) in the middle row.

In GC3.05-PPE, the central warming for cold winter days is substantial over Northern England and Scotland, reaching up to 9°C in places. The high-end changes (which often exceed 10°C in these regions) are smaller in RCM-PPE. The UK averages of the high-end warming are 7.3°C (GC3.05-PPE) and 6.6°C (RCM-PPE). However, these differences, while important, are relatively small compared with the substantial uncertainty ranges across either ensemble. In RCM-PPE, for example, the UK average of the low-end change is 3.9°C. The bottom row of panels in Fig. 4.9a shows RCM-PPE changes at 12km resolution, revealing a more detailed spatial pattern that shows many of the largest increases occurring in high-elevation locales, in common with other studies (e.g. Kotlarski et al., 2015).
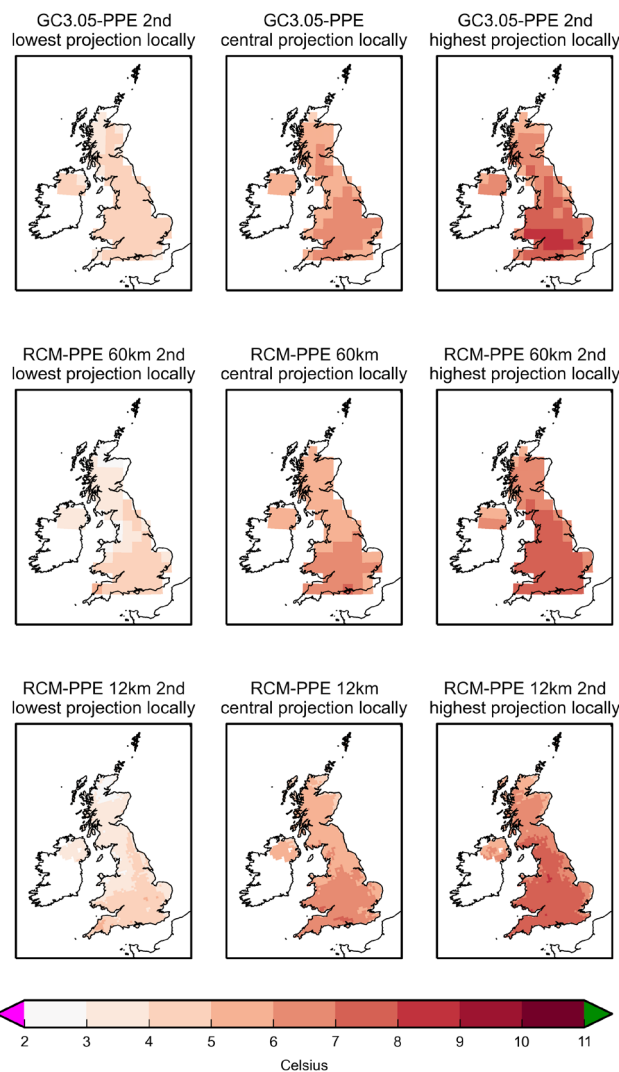

Figure 4.9b. As Fig. 4.9a, for low, central and high local projections of change in hot summer days, defined as in Fig. 4.5a.

Changes in the intensity of hot summer days (Fig. 4.9b) show increases everywhere, with broad uncertainty ranges. The UK average changes in RCM-PPE are 3.8°C, 5.9°C and 6.8°C for the low, central and high outcomes. For this variable, the additional spatial detail in the changes at 12km resolution (bottom row) reveals slightly smaller high-end changes in eastern coastal regions. Changes in GC3.05-PPE (top row) are larger at the low end of the uncertainty range of the smoothed 60km-scale changes in RCM-PPE, middle row), especially over Northern England and Scotland. The high-end changes are also slightly higher over parts of Wales and Southern England.
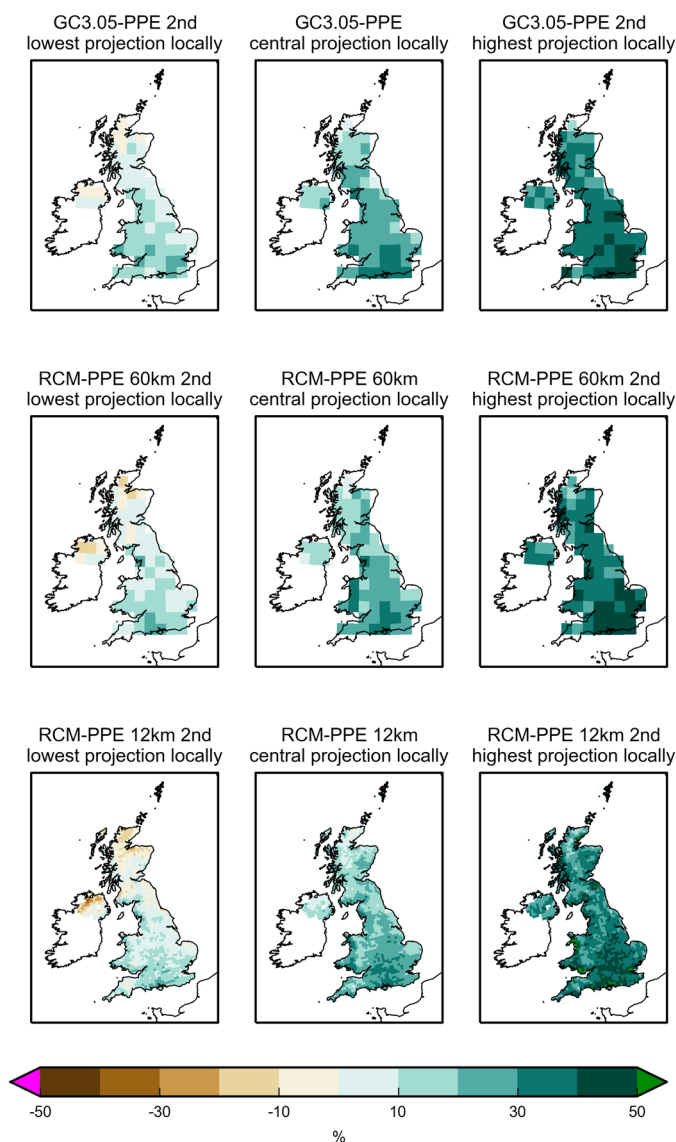

Figure 4.10a. As Fig. 4.9a, for low, central and high local projections of change in precipitation (%) on extremely wet winter days, defined as in Fig. 4.5b.
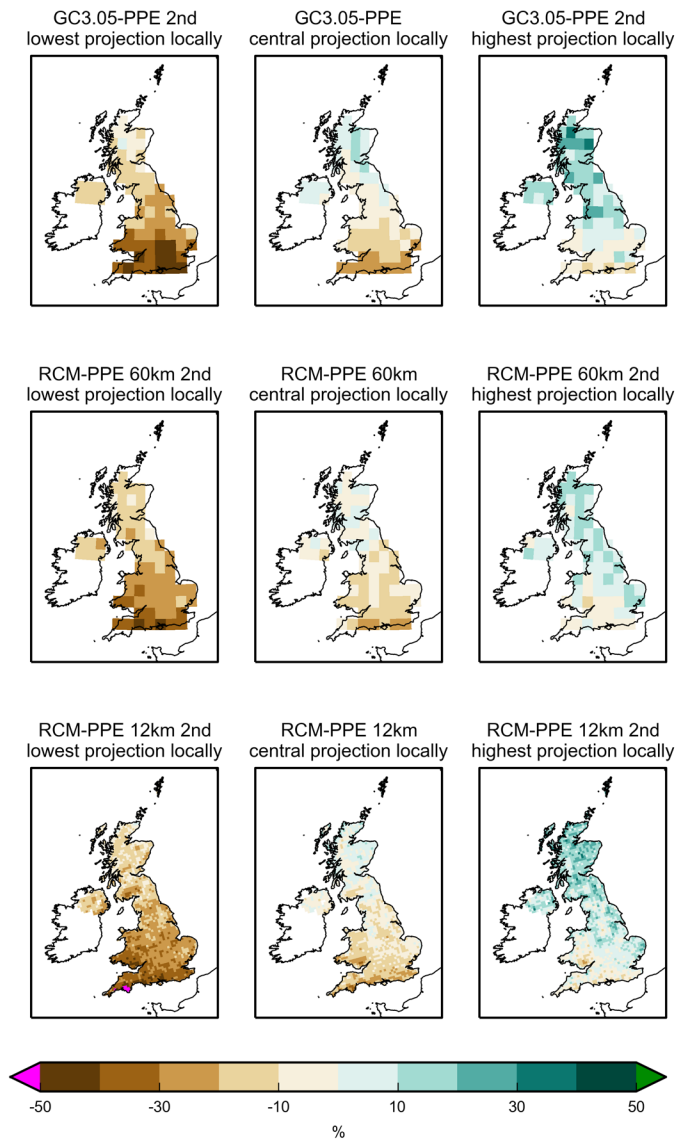
Figure 4.10b. As Fig. 4.10a, for low, central and high local projections of change in precipitation on extremely wet summer days, defined as in Fig. 4.5b.

Figures 4.10a,b show changes in the intensity of extreme wet days in winter and summer, based on the metric of Fig. 4.5b. In winter, the changes in RCM-PPE and GC3.05-PPE are quite similar. The central changes show increases exceeding 20% over most of England and Wales, with generally smaller values in Northern Ireland and most of Scotland. Low-end changes range from reductions in parts of Northern Ireland and Scotland, to increases exceeding 10% in parts of England and Wales. High-end increases exceed 30% in most parts of the UK, and exceed 50% locally in the 12km-scale RCM-PPE results. The projected increases in extreme winter precipitation are larger than corresponding changes in the climatological average (Fig. 4.10a cf 4.8c). The UK average of the central changes in RCM-PPE is 20%, for example, compared with 17% for 20-year mean precipitation. This implies a change in the shape of the distribution of daily events, shown in Fig. 4.11 below.

In summer (Fig 4.10b), the ranges of change in extreme wet day intensity span both increases and decreases. Average UK changes in RCM-PPE are -21.8% (low), -6.6% (central) and 8.0% (high). In Southern England and South Wales, decreases (in the central and low-end changes) are typically larger in GC3.05-PPE than in RCM-PPE. In Scotland, increases in the central change are more widespread in GC3.05-PPE, and the intensity of increases at the high end is generally larger.
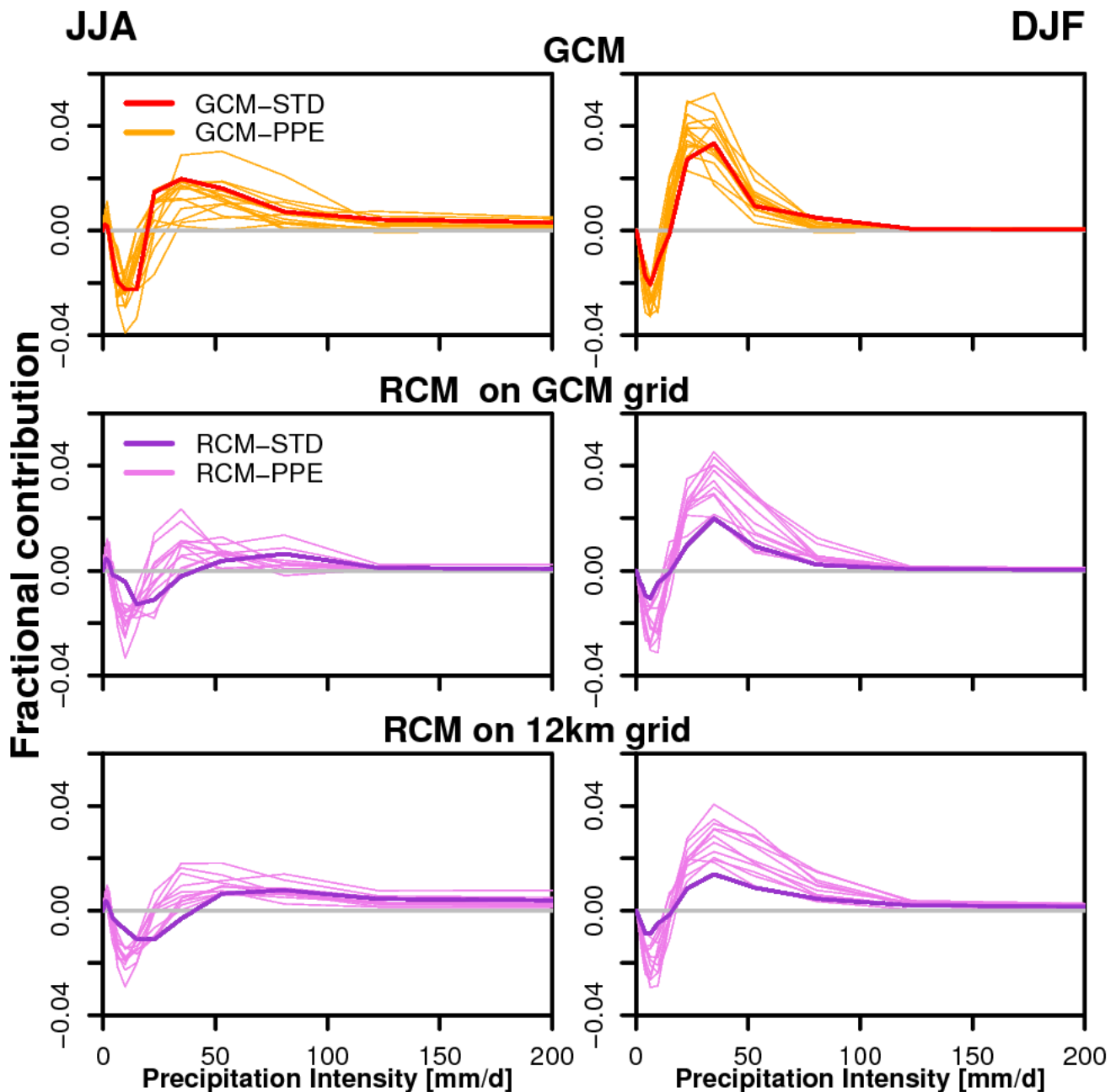
Figure 4.11 As Fig. 4.6, for future changes in the fractional contribution to total UK precipitation from each bin, for 2061-2080 relative to 1981-2000.

Projected changes in the frequency of wet days are typically smaller than 10% in winter, whereas substantial decreases are predicted by both GC3.05-PPE and RCM-PPE in summer, with central changes (not shown) amounting to more than 30% in Southern England. In Figure 4.11, we show (for the UK as a whole) how the distribution of regional daily precipitation events changes, relative to the baseline distributions of Fig. 4.6. In both seasons, the global and regional model projections show an increase in the fractional contribution of heavy events, at the expense of lighter daily accumulations[50]. The switch from reductions to increases occurs at ~20mm in winter and in the range 20-50mm in summer.

[50] The only exception is that the fractional contribution from the lightest summer events increases slightly.

The changes are quite consistent between GC3.05-PPE and RCM-PPE, and the characteristics of the regional model changes do not depend significantly on whether they are calculated on its native grid, or following smoothing to the global model grid. Since these distributions use daily data from all UK locations, they cannot be related to specific regional changes in extreme intensities (e.g. Fig. 4.10a,b). Nevertheless, the general shift to heavier events is consistent between ensemble members and the two seasons. The results are also qualitatively consistent with other regional climate model projections for Europe (e.g. Beniston et al., 2007).

Overall, the results of Figs. 4.8-4.11 demonstrate (as expected) a high degree of consistency between national-scale responses in the GC3.05-PPE and RCM-PPE projections. This applies both to changes in long-term averages of surface air temperature and precipitation, and to changes in extreme events. The enhanced spatial resolution in the regional simulations adds detail to the patterns of change at their native 12km scale. It also leads to modifications to the projected changes at larger scales resolved by the driving global projections. Taken together, these results provide support for the use of Strand 3 in stakeholder applications requiring detailed regional information (see section 4.1). This includes applications involving development of case studies or narratives, in which local consistency with larger scale circulation, temperature or precipitation changes is particularly important.

5. Projections of future variability and change for the UK

In this section we present projected changes for the UK under RCP8.5 emissions, comparing results from each Strand. The main focus is placed on surface air temperature and precipitation. A limited selection of results is shown in section 5.1, in order to highlight some key features of the projections, including differences between the ranges of future change produced by the three Strands. Additional results are provided in Lowe et al. (2018). In section 5.2, we provide some general guidelines on how to interpret and use the results, in the light of the different constructions and purposes of Strands 1-3.

5.1. Comparison of projections from strands 1-3

Figure 5.1 shows changes expressed per unit rise in GMST for England and Scotland. These are calculated by regressing winter and summer anomalies in surface air temperature and precipitation against annual GMST changes during the 21st century. The results provide a basic measure of uncertainties in spatial patterns of climate change, independent of uncertainty in the time evolution of GMST (see also the international normalised changes of Fig 3.23, and associated discussion in section 3.5). The ranges of values reflect the effects of both modelling uncertainties and internal variability. They contribute to the spread of time-dependent UK changes discussed later, and are particularly relevant to studies of impacts and risks associated with specific global warming targets (e.g. Clark et al., 2010).

The Strand 1 probability distributions provide the broadest uncertainty ranges, in all cases. For surface air temperature, the Strand 2 and 3 results all lie within the 10-90% range of the probabilistic projections. For precipitation in England, three CMIP5-13 members exceed the 90% probability level of Strand 1 in summer, and one in winter. In the winter case, the relevant CMIP5-13 member gives a much higher response (approaching 15% per °C) than other Strand 2 projections. For precipitation in Scotland, two RCM-PPE members and one CMIP5-13 member give small negative normalised responses in winter, below the 10% probability level of Strand 1. Also, most of the Strand 2 and 3 projections lie below the median response from Strand 1. In summer, all the Strand 2 and 3 simulations lie within the 10-90% probability range of Strand 1.

In winter, there is a substantial degree of overlap between the GC3.05-PPE and CMIP5-13 ensembles of Strand 2, with CMIP5-13 showing a wider spread. However, this is due mainly to extension of the range by one ensemble member, particularly in the case of precipitation. In summer, the GC3.05-PPE and CMIP5-13 ranges of normalised response show offsets not apparent in winter. Consequently, the range of the combined set of 28 projections is always larger than the range of either constituent ensemble in isolation. For surface air temperature in Scotland, GC3.05-PPE samples outcomes mainly above the median Strand 1 value, whereas CMIP5-13 outcomes are distributed more evenly above and below the Strand 1 median. Three CMIP5-13 members simulate an increasing trend in average English summer precipitation, whereas consistent summer drying signals are found in GC3.05-PPE members and HadCM3-based PPEs (Murphy et al., 2014; Sexton et al., 2016). The Strand 1 distributions in Fig. 5.1 reflect the combination of these results with the (typically) weaker drying signals, or modest increases, found in CMIP5 earth system models.
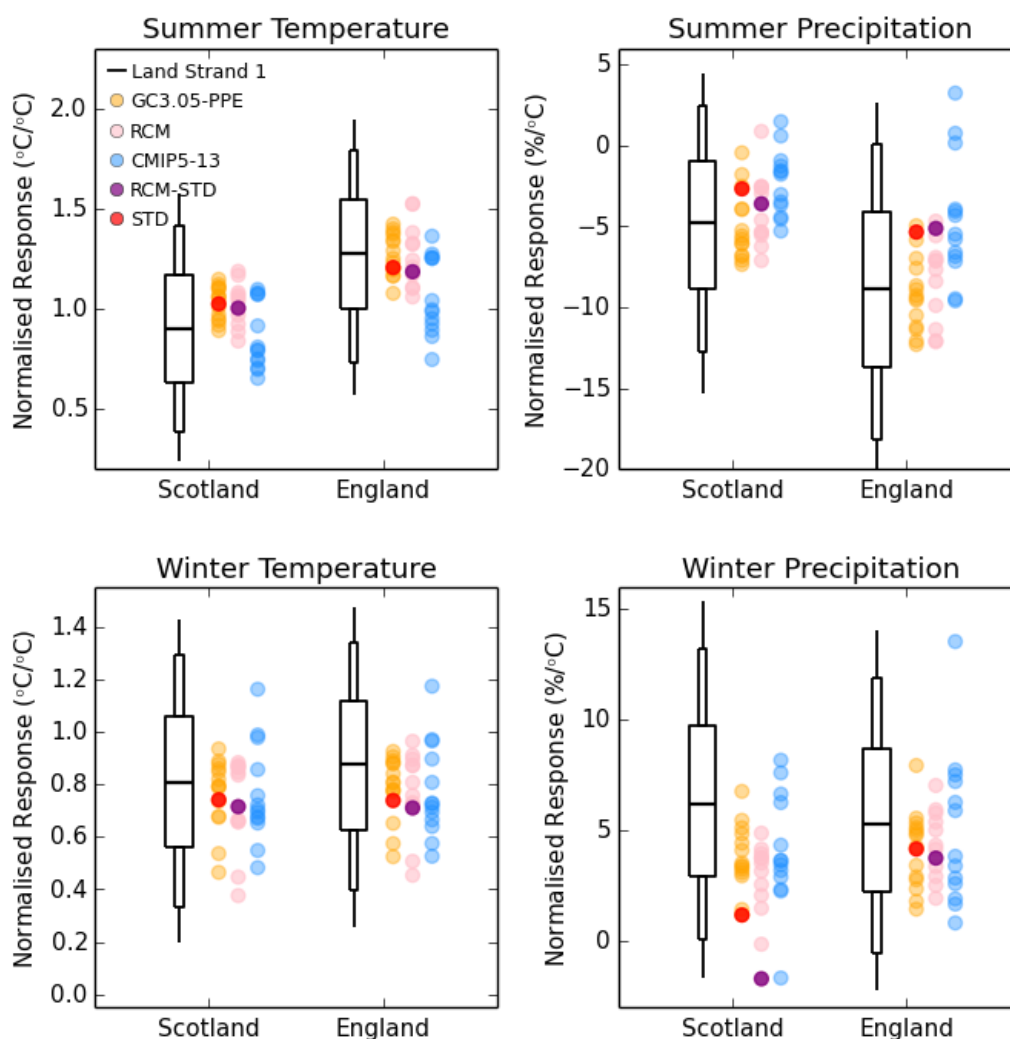
Figure 5.1. Projections of normalised changes during the 21st century under RCP8.5 emissions for surface air temperature (left panels, °C per °C) and precipitation (right panels, % per °C), expressed per unit increase in GMST. Changes for DJF (lower panels) and JJA (upper panels) are expressed relative to a 1981-2000 baseline, and are calculated by linearly regressing seasonal changes for England and Scotland against annual changes in GMST. Boxes and whiskers denote the 5, 10, 25, 50, 75, 90 and 95% probability levels of the Strand 1 probabilistic projections. Orange dots (with STD in red) denote members of GC3.05-PPE, and blue dots show those of CMIP5-13 (Strand 2). Pink dots show members of RCM-PPE, with RCM-STD in purple (Strand 3). The regressions are derived from 1980-2100 data for Strands 1 and 2, and 1980-2080 for Strand 3, as the RCM-PPE simulations were not extended to 2100.

The HadCM3- and GC3.05-based PPEs were both developed from UK climate models. Their land surface components are structurally similar, however the two models contain many structural differences, including their representations of atmospheric dynamics, large-scale cloud and boundary layer processes (Walters et al., 2017). The overall ranges of response seen in Strands 1 and 2 for GMST (Fig.3.20), and patterns of response (Fig. 5.1), are therefore a consequence of combining three quasi-independent lines of modelling evidence, namely: HadCM3-based PPEs (Strand 1), GC3.05-PPE (Strand 2) and CMIP5 multi-model ensembles (Strands 1 and 2, via emissions- and concentration-driven experiments respectively).

Ranges of normalised change from Strand 3 are generally similar to those in the driving GC3.05-PPE simulations. This is not surprising, as impacts of high resolution downscaling are likely to be greater at local scales than in the national averages of Fig. 5.1. A few differences are apparent: For example, RCM-STD produces a small negative change in Scottish winter precipitation per unit rise in GMST, whereas its driving global projection (STD) simulates a small increase. However, much of this contrast arises from the final 20 years of the STD simulation (2081-2100), as it is only during this period that a consistent increase in Scottish winter precipitation emerges. This period is not included in the RCM simulations, which stop at 2080 (section 4.3).

Projected changes for 2061-2080 relative to 1981-2000 are shown in Figure 5.2, for England and Scotland. The changes in Fig. 5.2 are influenced by projected responses in GMST, as well as by the normalised regional responses of Fig. 5.1. In the Strand 2 projections, the ranges of winter changes given by GC3.05-PPE and CMIP5-13 (Fig. 5.2, lower panels) show considerable overlap. The combined set of projections lie within the 5-95% probability range of the Strand 1 projections, for both regions and both variables. For winter surface air temperature, no Strand 2 projections lie below the 10% probability level of the Strand 1 distribution for either Scotland or England. These results reflect the broader sampling in Strand 1 of low-end responses in both the normalised response (Fig. 5.1), and in GMST changes (Fig. 3.20). For winter precipitation, the Strand 2 projections sample the range of Strand 1 outcomes quite well, with the exception that only one Strand 2 simulation lies above the 75% probability level of Strand 1 changes for Scotland.

The set of Strand 2 projections give summer surface air temperature outcomes covering the top three quartiles of the Strand 1 probability distributions for Scotland, and the 10-90% probability range for England (Fig. 5.2, upper left panel). The GC3.05-PPE projections lie almost entirely (for Scotland), or mostly (for England) above the 25% probability level, due to the combined influences of the high levels of GMST response (Fig. 3.20) and relatively strong normalised regional responses (Fig. 5.1). Several GC3.05-PPE members lie above the 90% probability level for both countries. Most of the CMIP5-13 projections give smaller levels of summer warming than any of the GC3.05-PPE members. However, as in winter, none lie below the 10% probability level for either Scotland or England.

In summer, precipitation changes from Strand 1 show broad distributions (Fig 5.2, upper right panel), with 5-95% probability ranges encompassing a small increase to a ~60% reduction for England, and an increase of ~15% to a reduction of ~40% for Scotland. For England, the set of Strand 2 projections gives a wide spread of changes that covers almost fully the 5-95% range from Strand 1, and also includes one projection lying above the wet end of this range. There is little overlap between the GC3.05-PPE changes and those of CMIP5-13, which mostly sample responses below and above the median of Strand 1 respectively. In this case, the combination of the two ensembles substantially broadens the range of projected changes available from Strand 2. For Scotland, there is more overlap between the GC3.05-PPE and CMIP5-13 ensembles, although only CMIP5-13 provides examples of projections lying in the upper quartile of the Strand 1 distribution, and all of the Strand 2 projections in the lower quartile arise from GC3.05-PPE.
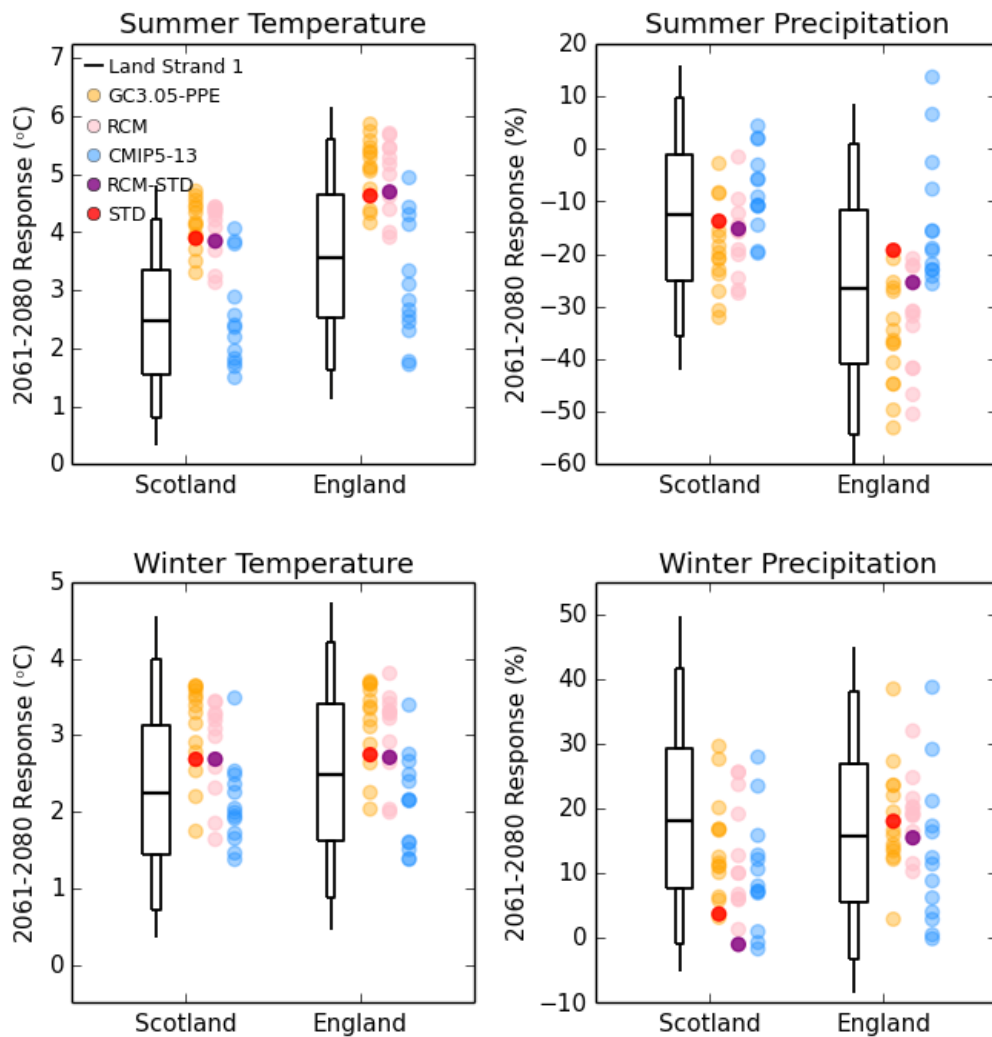
Figure 5.2. Projected changes for 2061-2080 relative to 1981-2000 for Scotland and England in DJF (lower panels) and JJA (upper panels), under RCP8.5 emissions. Results are shown for surface air temperature (left panels, °C) and precipitation (right panels, %). Boxes and whiskers denote the 5, 10, 25, 50, 75, 90 and 95% probability levels of the Strand 1 probabilistic projections. Orange dots (with STD in red) denote members of GC3.05-PPE, and blue dots show those of CMIP5-13 (Strand 2). Pink dots show members of RCM-PPE, with RCM-STD in purple (Strand 3).

The Strand 3 surface air temperature projections for Scotland and England show similar ranges of change to GC3.05-PPE, as is the case for their normalised responses (Fig. 5.2 cf Fig. 5.1). The RCM-PPE changes are slightly cooler compared with the 12 driving members of GC3.05-PPE (see Fig. 4.8a,b, discussed in section 4.5). This is reflected in the envelopes of change for both countries in summer, and for Scotland in winter, which show a small shift to cooler responses in Strand 3. Note, however, that in Fig. 5.2 we compare RCM-PPE against all 15 GC3.05-PPE members, including the three members which were not used to drive regional model projections. For precipitation, the ranges of change within the RCM-PPE and GC3.05-PPE ensembles are similar, as for surface air temperature. The only exception is winter changes for England, for which GC3.05-PPE produces two projections that are outliers with respect to the other ensemble members, and also lie respectively below and above the range given by RCM-PPE. The outlier at the low end of the range was not downscaled in Strand 3. The outlier at the high end was downscaled, and produces the highest response amongst RCM-PPE members, albeit somewhat smaller than the increase in the driving global simulation.

In comparing results from Strands 1-3, it is important to bear in mind the differing nature of the products. The Strand 1 results are derived from large samples of potential outcomes and are formally constrained by a set of observables. This allows them to be interpreted as probabilistic estimates conditional on the climate modelling inputs and expert assumptions used in their construction (see section 2.1, also discussion in 5.2 below). On the other hand, Strands 2 and 3 provide limited sets of individual projections with full spatial and temporal coherence. They are intended to be useful for impacts assessments and development of storylines (section 1.3), but not to support a probabilistic interpretation. For example, it would not be appropriate to simply count the number of Strand 2 simulations lying above or below a given level of response in Fig 5.2, and ascribe a likelihood of exceeding that response for comparison with the corresponding probability from Strand 1.

Figure 5.3a. Maps of changes in winter surface air temperature for 2061-2080 relative to 1981-2000 under RCP8.5 emissions, from individual projections selected from Strands 2 and 3. The projections are selected by ranking the UK-average changes, and selecting those giving the lowest, central and highest values within the set of 28 Strand 2 members or 12 Strand 3 members. The central members correspond to those ranked 15th and 7th lowest respectively. For Strand 2, the UK average changes are 1.4°C, 2.6°C and 3.6°C for the low, central and high members. The low UK average response is from the CESM1-BGC model, and the others from GC3.05-PPE members. For Strand 3, the UK average changes from RCM-PPE members are 1.8°C, 3.2°C and 3.6°C.

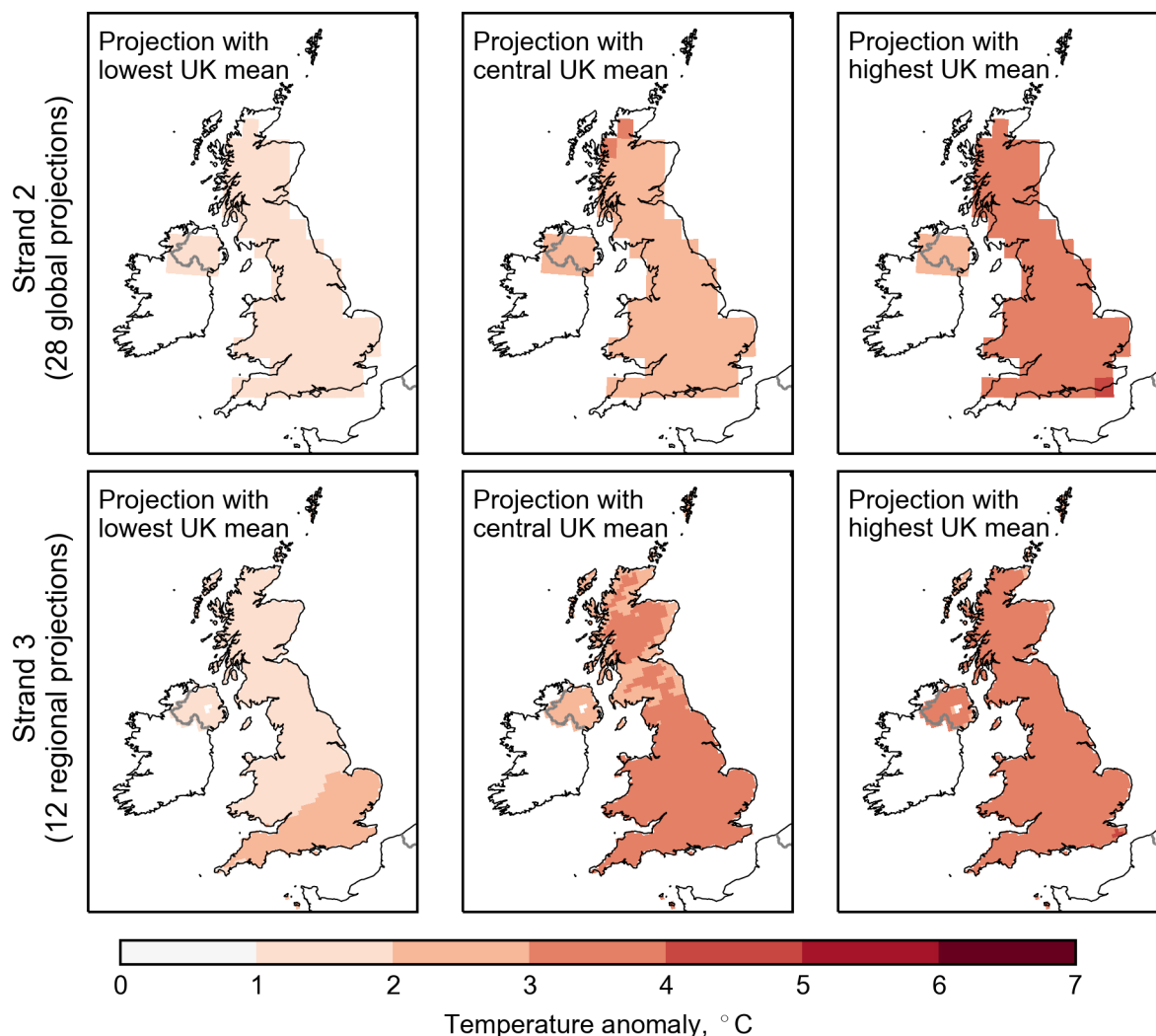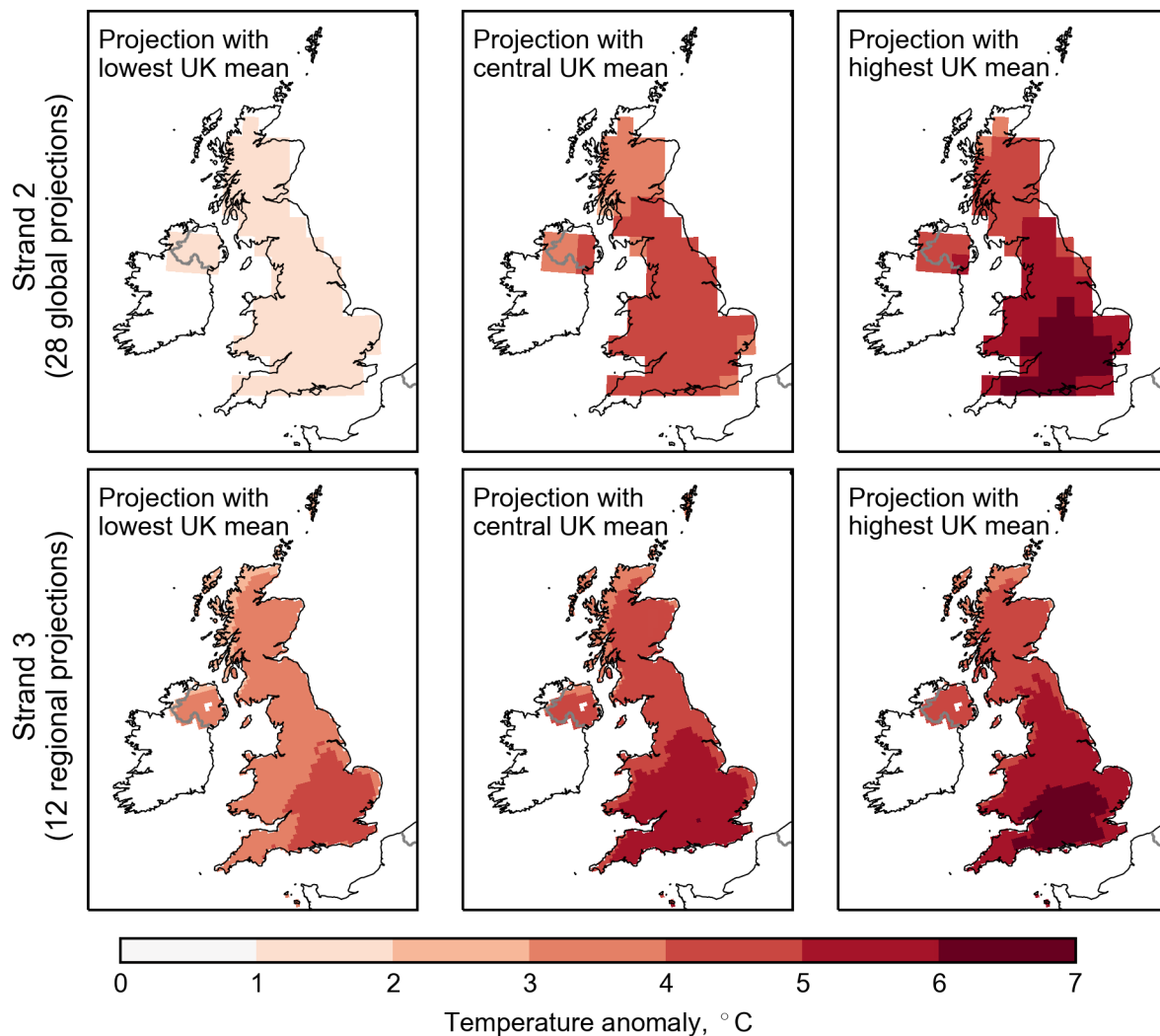Figure 5.3b. As Fig. 5.3a for changes in summer surface air temperature for 2061-2080 relative to 1981-2000 under RCP8.5 emissions, from individual projections selected from Strands 2 and 3. The central members correspond to those ranked 15th and 7th lowest respectively. For Strand 2, the UK average changes are 1.7°C, 4.2°C and 5.4°C for the low, central and high members. The low UK average response is from the GFDL-ESM2G model, the central response from the CanESM2 model, and the high response from a GC3.05-PPE member. For Strand 3, the UK average changes from RCM-PPE members are 3.6°C, 4.9°C and 5.2°C.

In addition, it should not be assumed that the set of Strand 2 or Strand 3 projections will necessarily show a higher concentration of outcomes close to the middle of their ranked set of responses. See, for example, the summer temperature changes from Strand 2 in Fig. 5.2, also Figs. 3.20 and 3.24a discussed in section 3.5, and Figs. 5.6a described below. This is another illustration of the non-probabilistic nature of these Strands. In contrast, the Strand 1 probability distributions take unimodal forms (e.g. Figs. 2.8 and 2.9) in which the highest concentration of outcomes usually lies close to the median. However, the outcome of maximum relative probability (the "mode" of the relevant distribution) does not, in general, coincide precisely with the median.

In Figure 5.3a-d we show examples of projected surface air temperature and precipitation changes for 2061-2080 relative to 1981-2000, from individual members of Strands 2 and 3. The maps show regional changes from the members of each Strand possessing the lowest, central and highest changes for the relevant UK average, for winter and summer. The central member is taken as that with the 15th and 7th lowest-ranked UK-average response, for Strands 2 and 3 respectively. Maps showing changes given by each

of the 28 Strand 2 projections are available from the UKCP18 User Interface (https://ukclimateprojections-ui.metoffice.gov.uk/). The individual projections of change in winter surface air temperature show rather uniform patterns (Fig. 5.3a). Regional changes lie almost always within 1°C of the relevant UK average response. These amount to 1.4°C, 2.6°C and 3.6°C for the low, central and high members of Strand 2, and 1.8°C, 3.2°C and 3.6°C for those of Strand 3. The maps of change in summer air temperature show larger regional variations than their winter counterparts, especially for the central and high projections, with larger warming over England and Wales than in Scotland. These spatial variations are discussed further below, in the context of Figure 5.4b.

Figure 5.3c. As Fig. 5.3a for changes in winter precipitation (%) for 2061-2080 relative to 1981-2000 under RCP8.5 emissions, from individual projections selected from Strands 2 and 3. For Strand 2, the UK average changes are 2%, 14% and 35% for the low, central and high members. The low UK average response is from the BCC-CSM1.1 model, and the central and high responses from GC3.05-PPE members. For Strand 3, the UK average changes from RCM-PPE members are 8%, 16% and 29%.

The winter precipitation maps (Figs. 5.3c) show examples consistent with different national-scale levels of increase. The UK average changes in Strand 2 range from 1.9% (low) through 14% (central) to 35% (high). The Strand 3 examples show high resolution spatial details absent from the Strand 2 projections, such as enhanced changes near the South Wales and Southern England coasts in the central projection, and in the Lake District in the high projection. Generally, the ranking of regional changes follows the ranking in the UK average response (as for the surface air temperature changes). However, the central projection from Strand 3 shows reductions over North East Scotland (exceeding 10%, and 20% in some locations) that are larger

than any reductions seen in the low projection. This demonstrates that individual projections possessing low, central or high national outcomes cannot necessarily be assumed to give corresponding changes locally.
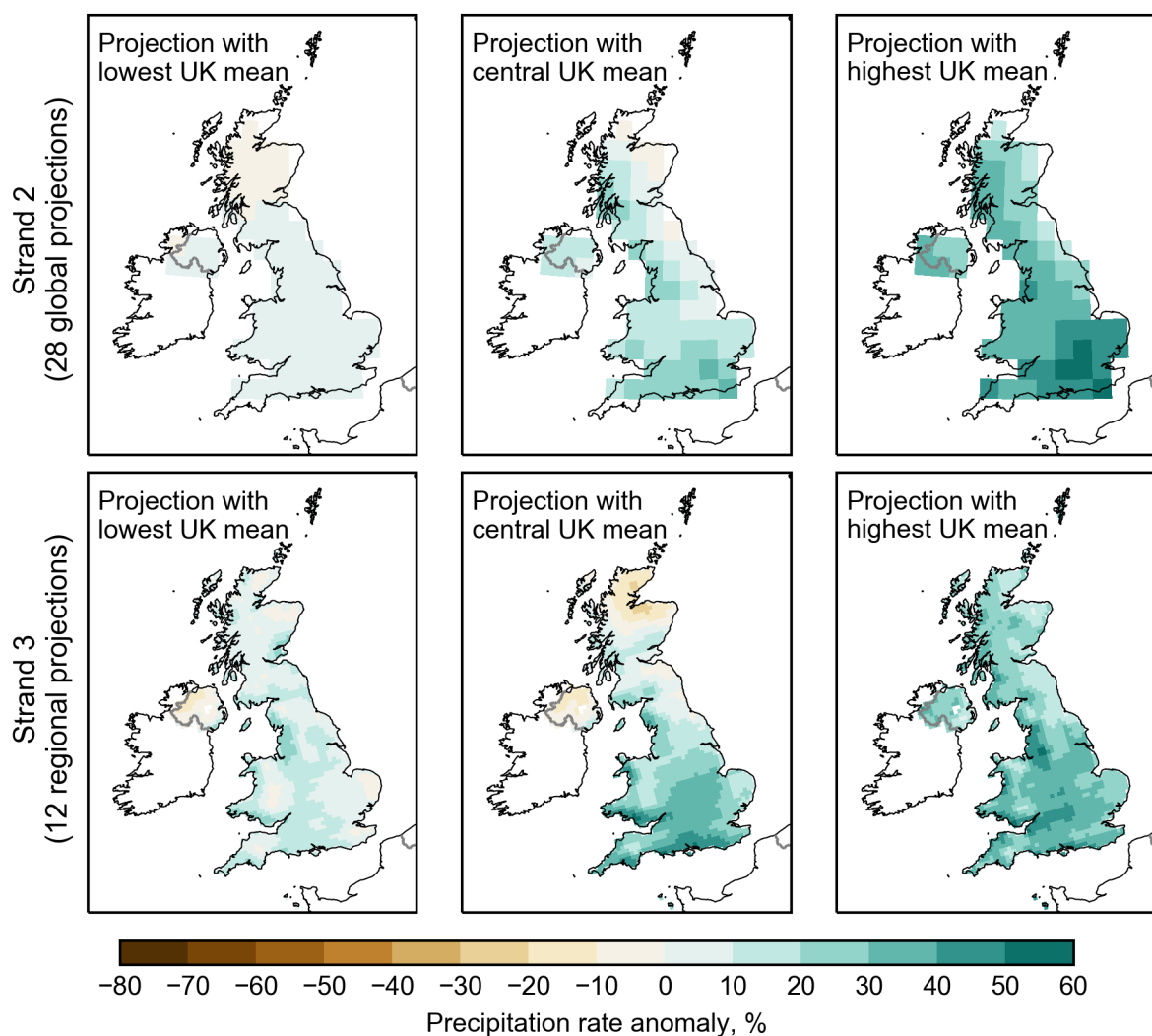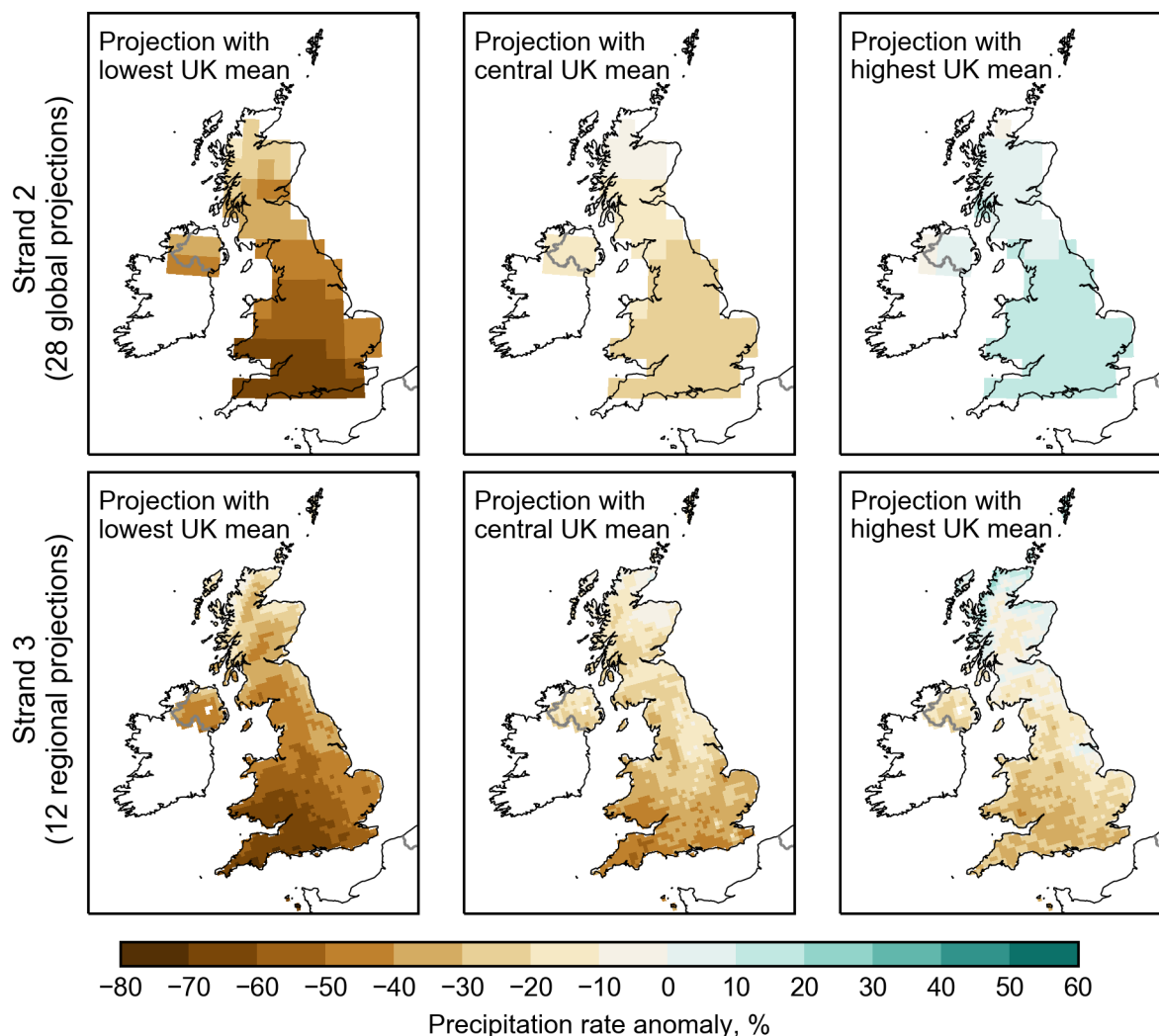

Figure 5.3d. As Fig. 5.3c for changes in summer precipitation (%) for 2061-2080 relative to 1981-2000 under RCP8.5 emissions, from individual projections selected from Strands 2 and 3. For Strand 2, the UK average changes are -47%, -21% and 10% for the low, central and high members. The low UK average response is from a GC3.05-PPE member, the central response is from the GFDL-ESM2G model and the high response from the CNRM-CM5 model. For Strand 3, the UK average changes from RCM-PPE members are -44%, -26% and -16%.

In summer (Fig. 5.3d), the Strand 2 examples range from a substantial drying (a 47.6% reduction in the UK average for the low projection) to a modest increase (10.4% increase for the high projection). The low and central Strand 3 projections show similar UK average changes to those of Strand 2, which are contributed by members of GC3.05-PPE. However, the high projection from Strand 3 gives a drying in the UK average of 16.4%. This contrasts with the 10.4% increase in the high Strand 2 member, which is provided by a CMIP5-13 model. As in winter, the Strand 3 maps show examples of detailed spatial variability arising from the downscaling capability in RCM-PPE. However, the broader scale patterns in the low and central projections are similar to their Strand 2 equivalents, with the strongest reductions in precipitation generally occurring over England and Wales.
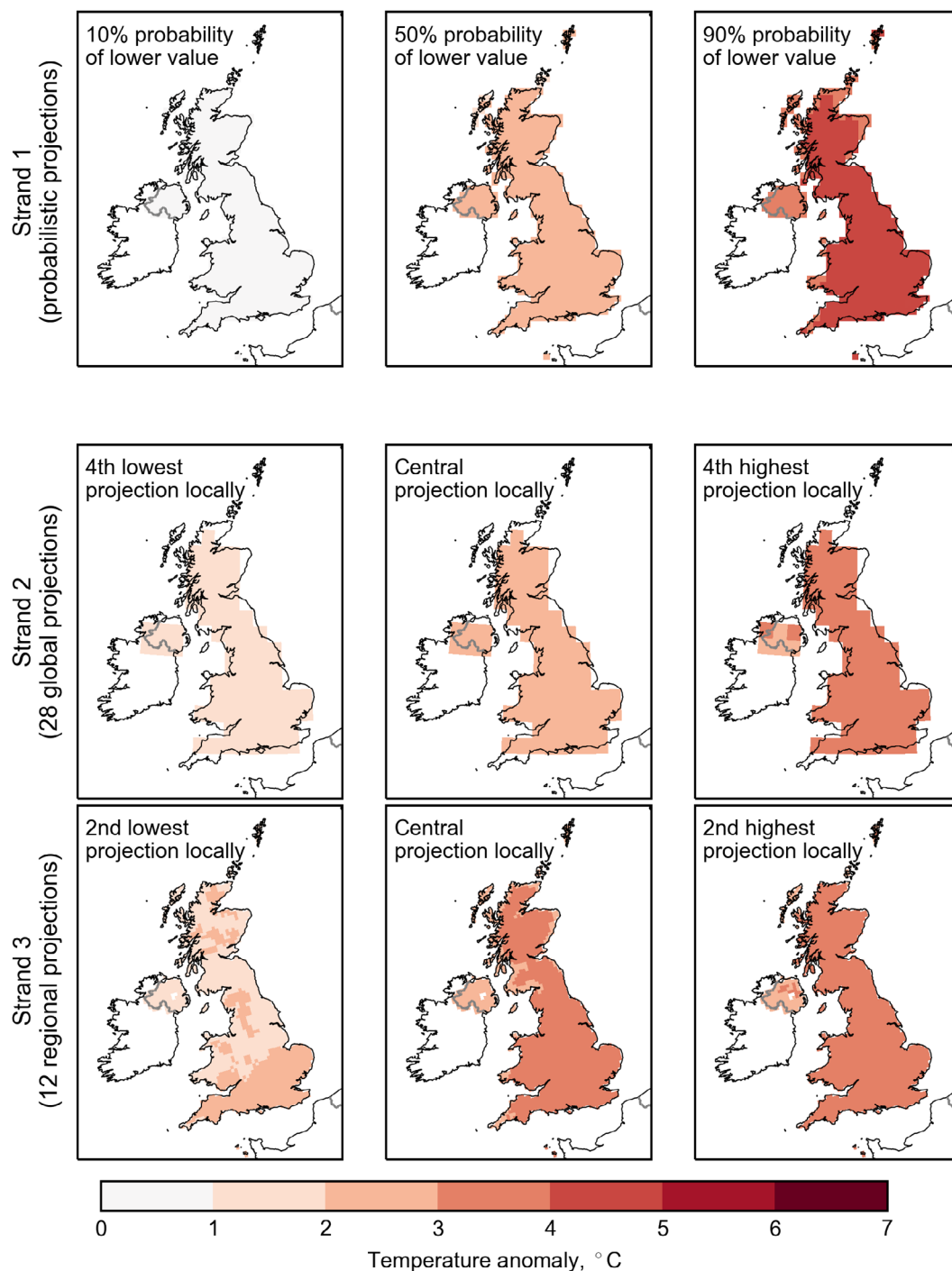
Figure 5.4a. The top row shows 10, 50 and 90% probability levels of changes in surface air temperature in winter from the probabilistic projections, for 2061-2080 relative to 1981-2000 under the RCP8.5 emissions scenario. The values represent the chance of an outcome lower than the relevant probability level, conditional upon the modelling evidence, methodology choices and assumptions used in Strand 1 (see text). The middle and bottom rows show low, central and high local changes selected from the set of 28 global climate model projections of Strand 2 (middle row, obtained by combining GC3.05-PPE and CMIP5-13), and the 12 regional climate model projections (RCM-PPE) of Strand 3 (bottom row). The Strand 2 and 3 results are obtained by considering each spatial location separately, and ranking the relevant set of local responses ($x = 1 \rightarrow N$, where N=12 or 28) into ascending order. We then identify "low", "central" and "high" projections as members close to the 10th, 50th and 90th percentiles of a frequency distribution of outcomes defined by the relevant set of N projections, assuming that member x represents percentile $(x-1)*100/N$. Thus, the 4th lowest and 4th highest responses define the range for Strand 2, and the 2nd lowest and 2nd highest for Strand 3. This provides estimates of spread which reduce sensitivity to any outliers, in a manner broadly consistent with the Strand 1 maps. Central estimates are provided by the 15th and 7th ranked members, in Strands 2 and 3 respectively. Although we choose low, central and high members of Strands 2 and 3 by considering the relevant set of outcomes as a *frequency* distribution, the outcomes should not be interpreted *probabilistically*, as explained in the text. Strand 1-3 results are all shown on their native spatial grids, of resolution 25, 60 and 12km respectively.

In Figure 5.4a-d, we show a different presentation of how projected changes vary across the UK, including Strand 1 alongside Strand 2 and 3. As in Fig. 5.3a-d, we consider changes in surface air temperature and precipitation for 2061-2080 relative to 1981-2000. However, in Fig. 5.4a-d we compare local ranges of response found in Strands 1-3. For the probabilistic projections, the maps show the 10, 50 and 90% probability levels of change. These represent the levels of warming or precipitation change for which the Strand 1 methodology predicts a 10, 50, or 90% chance that a lower change will occur. The 50% level is the median change, and the 10% and 90% levels provide lower and upper estimates of the associated uncertainty ranges. For Strands 2 and 3, we consider each spatial location separately, and rank the local responses of the relevant (28 or 12 member) set of projections from low to high, plotting the responses from members that correspond approximately to the 10th, 50th and 90th percentiles of the ranked set, when considered as a frequency distribution of outcomes (details in caption of Fig. 5.4a). These members correspond to the 4th lowest, 15th lowest and 4th highest response in Strand 2, and the 2nd lowest, 7th lowest and 2nd highest response in Strand 3.

Our choice of 10-90% probability intervals for Strand 1, and (approximate) 10th-90th percentile intervals for Strands 2 and 3, allows us to compare local ranges of change across the Strands using metrics that treat extreme outcomes in a consistent fashion. In particular, this choice reduces the sensitivity of the diagnosed ranges from Strands 2 and 3 to any outlier responses (e.g. Fig. 5.2). In the discussion of Fig. 5.4a-d below, we refer to the 10%, 50% and 90% probability levels from Strand 1 as "low", "central" and "high" responses, for convenient comparison with Strands 2 and 3. We also refer to the difference between the low-end and high-end responses as the "range". However, we emphasise again (noting the earlier discussion of Fig. 5.2) that:

- The Strand 2 and 3 results should not be interpreted probabilistically.

- Since Strand 1 is designed explicitly to provide assessments of uncertainty, the 10 and 90% probability levels from Strand 1 should be taken as the primary source of information on ranges of potential future change.

Note also that different Strand 2 or 3 members contribute the low, central or high outcomes in different regions in Fig. 5.4a-d, in contrast to Fig. 5.3a-d. Therefore, the maps in Fig. 5.4a-d should not be interpreted as examples of spatial patterns that might be simulated in an individual projection.

In winter, central changes in surface air temperature invariably lie between 2°C and 3°C for Strands 1 and 2. Strand 3 central changes are mostly in the range 3-4°C, but below 3°C over Northern Ireland and a few (mainly coastal) regions of Great Britain. Averaged over all locations, the values are 2.4°C (Strand 1), 2.7°C (Strand 2) and 3.1°C (Strand 3). In Strand 3, the effects of downscaling cause a modest cool shift in the distribution of changes in RCM-PPE, compared to the driving GC3.05-PPE simulations (Fig. 4.8a). Despite this, central responses in Strand 2 are generally lower than in Strand 3, due to the inclusion of CMIP5 models in the former. The cooler UK responses in CMIP5-13, compared with GC3.05-PPE, arise mainly from the higher levels of GMST warming in GC3.04-PPE (Fig. 3.20). Differences in normalised regional responses are not a major influence, because the envelopes of CMIP5-13 values are not shifted lower compared with GC3.05-PPE (Fig. 5.1).

The ranges of response are broadest in Strand 1. These reveal potential outcomes below 1°C at the lower end, and above 4°C at the upper end, which are not represented in Strands 2 or 3. The UK averages of the low-end changes are 0.7°C, 1.6°C and 1.9°C for Strands 1-3 respectively, and 4.2°C, 3.6°C and 3.4°C for the high-end changes. The similarity in the high outcomes for Strands 2 and 3 occurs because these arise in both cases from pairs of global and regional PPE projections.

The winter surface air temperature changes in Fig. 5.4a show only modest spatial contrasts (in some cases <1°C across the entire UK), as in Fig. 5.3a. For a given level of response, these are much smaller than the regional uncertainty ranges for the relevant Strand. Some downscaling effects are apparent, mainly through better resolution of maritime influences in limiting warming in coastal regions. These can be seen in the central response map for Strand 3 and the high response map for Strand 1. However, downscaling influences precipitation changes (discussed below) more substantially than changes in surface air temperature.
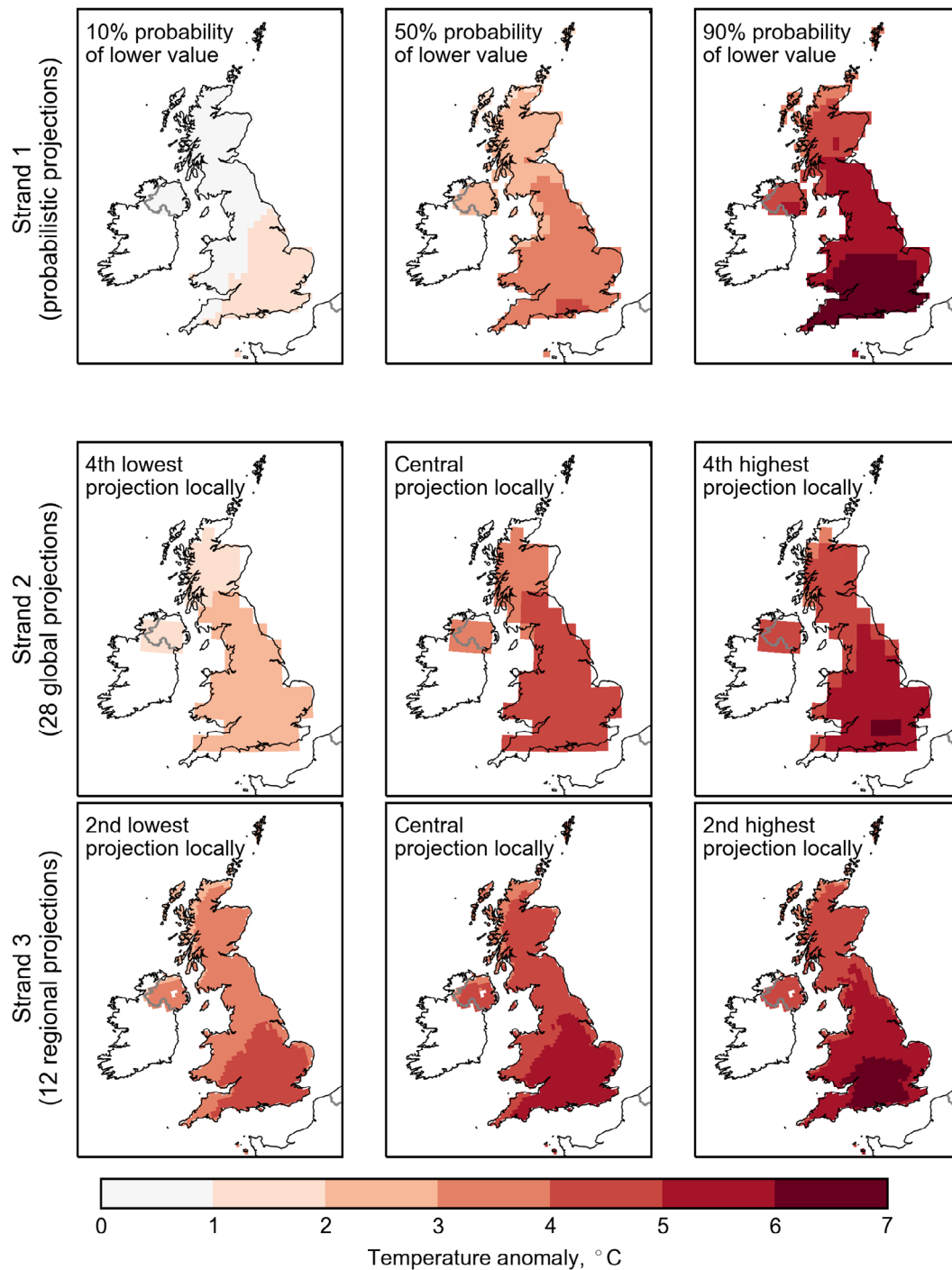

Figure 5.4b. As Figure 5.4a, for surface air temperature changes in summer.

In summer (Fig. 5.4b), Strand 1 again shows the possibility (at its 10% probability level) of warming below 1°C nationwide, whereas corresponding low-end outcomes are typically 1-3°C in Strand 2 and 3-5°C in Strand 3. However, the high responses in Strands 2 and 3 reach similar levels to Strand 1, which is not the case in winter. The UK average values are 5.4°C in Strand 1, 5.1°C in Strands 2 and 5.2°C in Strand 3. This reflects the relatively strong normalised regional responses found in GC3.05-PPE and RCM-PPE, which exceed the central Strand 1 value in all members for Scotland, and most for England (Fig. 5.1).

The UK averages of central changes are 3.1°C (Strand 1), 4.2°C (Strand 2) and 4.8°C (Strand 3). These exceed their winter counterparts, with substantial differences occurring in Strands 2 and 3. In winter, the projected changes are reduced by the effects of relatively weak values of normalised change (values are well below unity in most Strand 2 and 3 simulations, see Fig. 5.1). However, normalised changes are larger in summer than in winter in most CMIP5-13 members, as well as in the GC3.05-PPE and RCM-PPE projections.

In general, the summer surface air temperature changes in Fig. 5.4b show greater spatial contrast than in winter, as in the individual projections of Fig. 5.3b. The smallest warming occurs in North West Scotland, with progressively larger values to the south and east. In the Strand 2 simulations (McSweeney et al., 2018b), and in HadCM3-based PPE simulations contributing to Strand 1 (Joshi et al., 2008), continental landmasses warm more than the oceans. In Strand 2, for example, the average normalised response exceeds 1.25°C per unit rise in GMST over the whole of continental Europe in JJA, for both GC3.05-PPE members and CMIP5-13 members. This compares to less than 0.75°C per °C over central parts of the North Atlantic Ocean. It is likely, therefore, that the larger summer responses over Southern and Eastern England are related to the greater influence of continental Europe on this region, compared to the stronger maritime influence to the north and west. In winter, the dominance of the maritime influence extends further south and east. This is consistent with the presence of stronger prevailing south-westerly flow in the climatological near-surface wind pattern.

For winter precipitation (Fig. 5.4c), central changes by 2061-2080 show increases everywhere for Strand 2, and in all regions except parts of Northern Scotland for Strands 1 and 3. The high-end changes show increases of varying magnitude in all Strands, revealing that substantial changes cannot be ruled out. The UK average of high responses amounts to 35%, 27% and 28%, in Strands 1-3 respectively. The UK averages of the central changes are 16%, 14% and 17%. At the low end of the uncertainty ranges, the results demonstrates that modest reductions in precipitation (up to 10%) are plausible, with potential for reductions exceeding 10% in parts of Northern Scotland. In Strand 1, the UK average of the low changes is a small reduction (-1%). For Strand 2 and 3, the UK averages are positive, amounting to 2% and 5% respectively. The 12km RCM simulations in Strand 3 provide enhanced regional detail, which is particularly clear in the high response map. This shows sharper definition of enhanced changes in most westward-facing coastal regions of Great Britain, compared to the Strand 2 results. Evidence of downscaling effects is also apparent in the high response map from Strand 1, which includes information from RCM simulations at 25km resolution (section 2.2, Stage 3). Here, clear coastal enhancements are found in North Eastern Scotland and Southern England, with westward facing effects more patchy than in Strand 3.
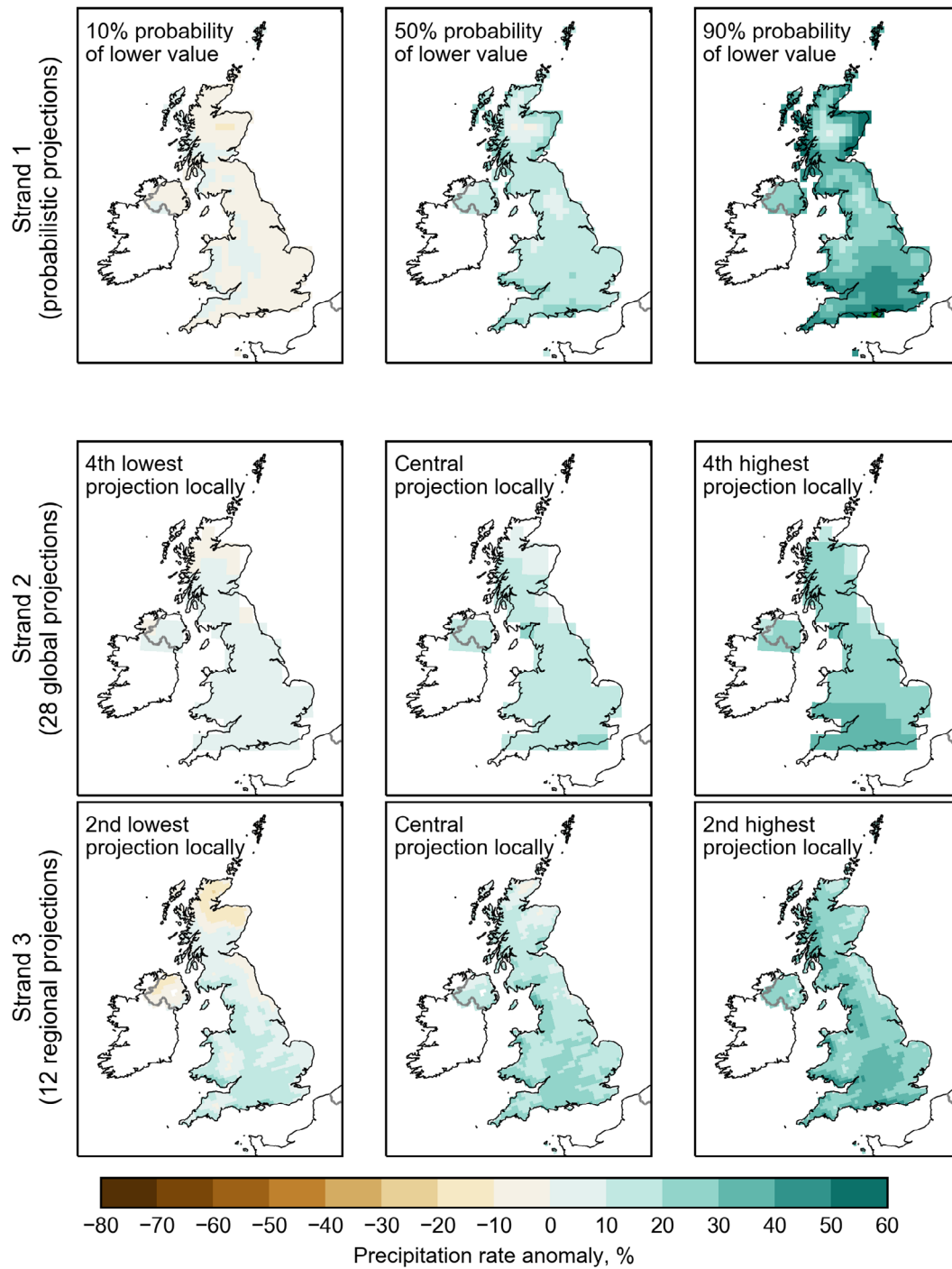
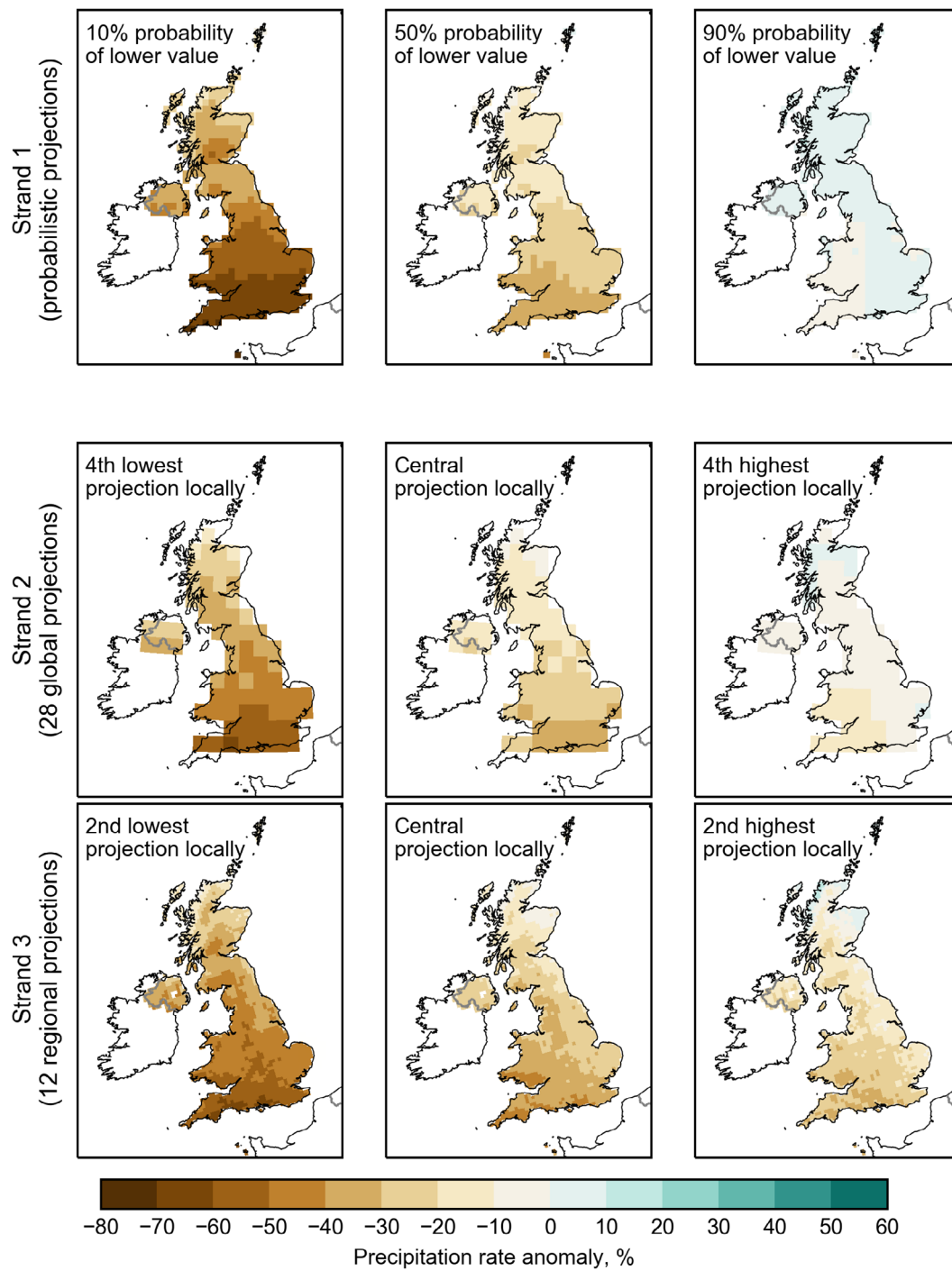**Figure 5.4c.** As Figure 5.4a, for winter precipitation changes (%).

Figure 5.4d. As Figure 5.4b, for summer precipitation changes (%).

In summer (Fig. 5.4d), the central changes show reductions in precipitation at all locations in all Strands. These range from 0-10% to over 40%, with UK average values of -23% (Strand 1), -22% (Strand 2) and -26% (Strand 3). The low response maps show that substantial reductions exceeding 50% are possible over England and Wales. Values of high-end changes from Strand 1 demonstrate that modest increases in summer precipitation are also plausible, in many parts of the UK (average value 2%). Strands 2 and 3 also show small high-end increases in parts of Scotland and eastern England, although UK average values are negative (-6% and -17% respectively). Differences between the ranges of change reflect the different constructions of the Strands. Over much of England and Wales, Strand 3 shows reductions exceeding 20% even for the high-end changes, because it is built entirely from a PPE that produces a consistent drying signal across its members. The central responses in Strand 3 are 5-10% drier than in Strands 1 and 2 in many locations (especially in much of England and Wales), but the spatial contrasts are smaller than for the high responses. The broader response ranges in Strands 1 and 2 result from the influence of CMIP5 models, particularly over England where CMIP5 provides a few projections showing an increase in summer precipitation, whereas the PPEs show only decreases. This is true for Strand 1 (not shown), as well as Strand 2 (Fig. 5.2 and related discussion). The EuroCordex RCM simulations for RCP8.5 are driven by five CMIP5 models, four of which are members of CMIP5-13. These also provide a few examples showing small increases in summer precipitation (see Racjzak and Schär (2017), discussed in section 4.5).

In Figure 5.5 we investigate the relationship between changes in atmospheric circulation and the changes in summer precipitation discussed above. For this we consider the summer North Atlantic Oscillation (SNAO). This is identified as the leading observed pattern of interannual variability in July-August sea-level pressure (SLP) in the North Atlantic/Europe sector, using empirical orthogonal function (EOF) analysis (Folland and Knight, 2009). Following Bladé et al. (2012), we show the SNAO (Fig. 5.5, top left panel) as a map of spatial correlations between local SLP variations and those of the principal component of the leading EOF. The resulting pattern has a more northerly location and a smaller spatial scale than the winter NAO pattern. The SNAO exerts a substantial influence on summer precipitation in North West Europe (including the UK) and the Mediterranean (see Fig. 5.5, bottom left panel). Specifically, the positive phase of the SNAO is associated with high SLP anomalies over the UK, and lower-than-average seasonal precipitation.

Also shown in Fig. 5.5 are ensemble-mean patterns of the SNAO, and associated precipitation anomalies, for GC3.05-PPE and CMIP5-13. Most members of both ensembles simulate an SNAO pattern showing positive and negative dipoles in similar locations to observations, with a negative correlation between the SNAO phase and July-August precipitation over the UK and NW Europe. Note, however, that this is not the case for all the members (not shown in Fig. 5.5): Several members in both ensembles position the southern dipole too far west into the Atlantic, and/or fail to represent the observed relationship with UK precipitation.
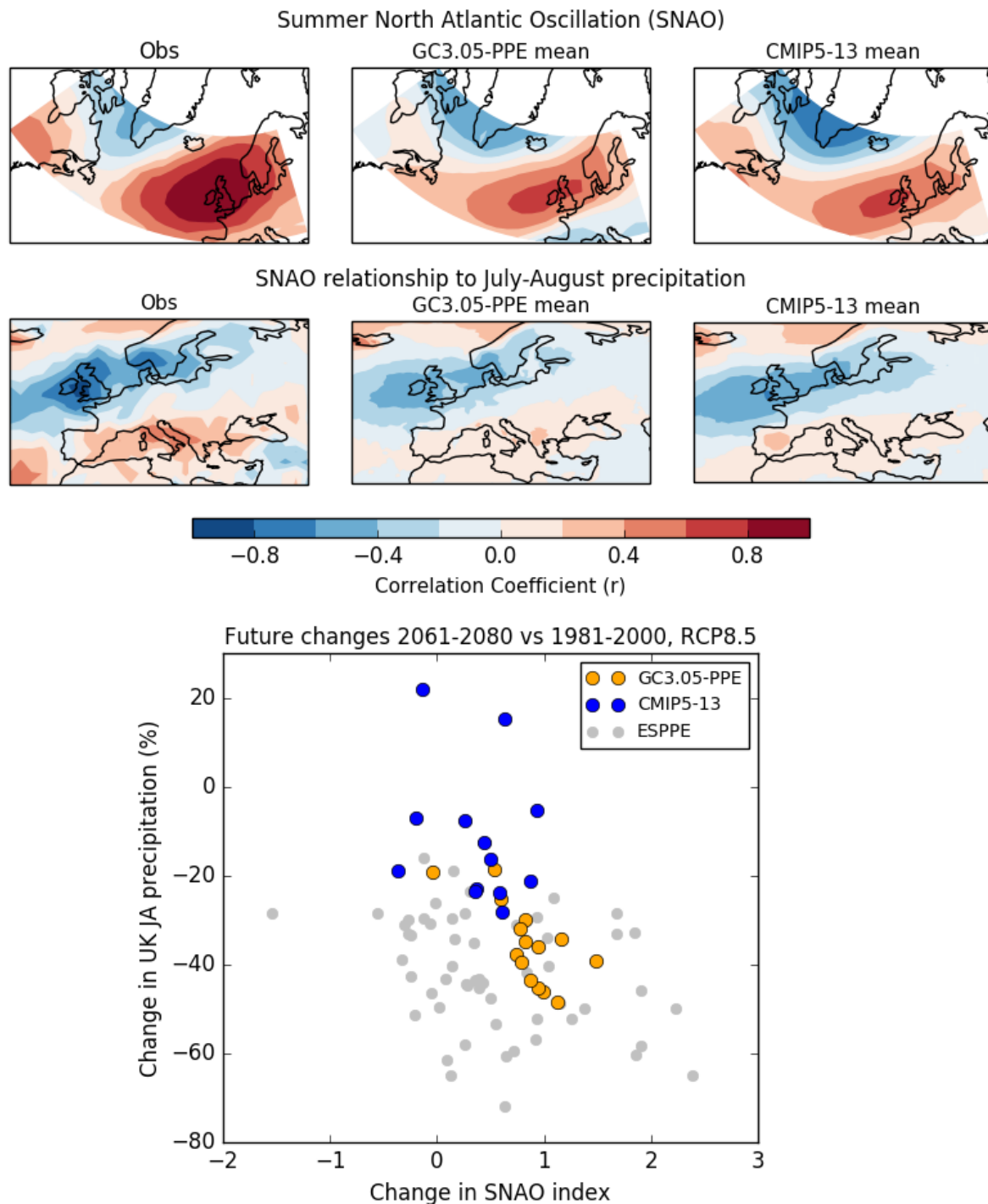
Figure 5.5. Relationship between future changes in UK precipitation and atmospheric circulation in summer, diagnosed from transient climate change simulations used in the construction of Strands 1 and 2. The top left panel shows the observed summer North Atlantic Oscillation (SNAO) for the period 1950-2010. Following Bladé et al. (2012), it is identified as the leading empirical orthogonal function (EOF) of observed anomalies of sea level pressure (SLP, from HadSLP2) for July-August in the region 40-70°N, 90°W-30°E. The EOF pattern is expressed as the spatial pattern of correlations between the time series of its principal component (the SNAO time series), and local SLP anomalies. For the GC3.05-PPE and CMIP5-13 ensembles, the SNAO is identified separately for each ensemble member, using whichever of the first and second EOFs that corresponds most closely to the observed SLP pattern. Spatial correlation patterns are then calculated from the SNAO and SLP time series for each member, and these are averaged to give the two ensemble-mean patterns shown in the centre and right panels in the top row. The middle row shows patterns of local correlations between the SNAO time series and regional precipitation anomalies during 1950-2010, corresponding to the panels above. Observed precipitation is taken from the GPCP dataset. The bottom panel shows the relationship between projected changes in UK average precipitation (%) and the SNAO for July-August, for 2061-2080 relative to 1981-2000 under RCP8.5 emissions. In addition to changes from the GC3.05-PPE (orange) and CMIP5-13 (blue) projections comprising Strand 2, the panel also provides results from the 57 ESPPE simulations (grey) used in Strand 1.

Most members of GC3.05-PPE and CMIP5-13 project a future shift towards the positive phase of the SNAO (bottom panel of Fig. 5.5). Those with larger increases in SNAO tend to display larger reductions in July-August precipitation. Bladé et al. (2012) found similar results in the CMIP3 multi-model ensemble. Several GC3.05-PPE members simulate larger increases in the SNAO than any of the CMIP5-13 members shown in Fig. 5.5. This may partially explain the greater summer drying seen in some of the GC3.05-PPE members. However, there is considerable scatter in the relationship between the SNAO and precipitation changes, and two CMIP5-13 models project an increase in July-August precipitation despite showing little change in the SNAO. Such results suggest that other factors may also be important drivers of the precipitation changes.

For example, in the HadAM3P model Rowell and Jones (2006) found that the increased moisture content of maritime air tended to offset circulation-driven reductions in summer precipitation over the UK and Southern Scandinavia. They also found that thermodynamic drivers of reduction in soil moisture (arising from either an earlier decline during spring due to enhanced snowmelt and evaporation, or from a positive feedback between precipitation and soil moisture during summer) were important over continental Europe, but less so over the UK. Summer soil moisture content in the UK reduces during the 21st century in the GC3.05-PPE simulations, even in ensemble members showing little change in the SNAO.

The bottom panel of Fig. 5.5 also shows precipitation and SNAO changes in members of the ESPPE used in Strand 1. This ensemble gives a wider range of SNAO changes than is found in the Strand 2 simulations, and also explores larger reductions in July-August precipitation (of -50% or more). This is consistent with the results of Fig. 5.4d, in which the low end of the range of Strand 1 outcomes shows greater levels of drying than its Strand 2 counterpart. Overall, the three ensembles of simulations in Fig. 5.5 suggest that projected SNAO changes play a significant role in explaining future UK summer precipitation changes. However, there is a wide spread of precipitation changes in a subset of the projections that show little change in the preferred phase of the SNAO. This motivates further study of the physical mechanisms that drive the projected summer hydrological changes and their uncertainties.

The temporal evolution of the Strand 1 and 2 projections is shown in Figure 5.6a. We use spatial averages for England as an example of a national-scale overview of the development of changes through the 21st century. The use in Fig. 5.6a of annual changes (shown for surface air temperature and precipitation in winter and summer) emphasises the combined roles of internal climate variability and forced long-term changes in driving the spread of future values (e.g. Sexton and Harris, 2015).

During the 20th century, long-term climate change signals from the Strand 1 and 2 outputs are relatively small. These can be seen, for example, in the ensemble-averages of CMIP5-13 or GC3.05-PPE (thick blue and orange curves), or the median (white) curves of the probabilistic projections (the latter are more easily visible in Fig. 5.6b, in which the Strand 1 results are repeated). Therefore, the spread of seasonal anomalies during the 20th century is likely to be due mainly to internal variability (plus a minor contribution from statistical uncertainties in the case of Strand 1, e.g. Fig. 2.5). Observations (black lines) are also shown in Fig. 5.6a. By comparing the envelope of observed anomalies with the ranges of historical results from Strands 1 and 2, we can assess how well the amplitude of observed internal variability is represented in the projection systems.

For example, if the Strand 1 results (available for the historical period from 1961) are to provide a reasonable basis for estimating seasonal risks, we would expect observed outcomes to fall mostly within the grey shaded region, as this represents the 5-95% range of the probability distributions. In the case of winter precipitation, fifteen events during 1961-2017 (26% of years) lie outside the 5-95% probability range, suggesting that seasonal variability is underestimated in Strand 1. For summer surface air temperature, no observed event lies outside the 5-95% range of Strand 1, while six cases lie outside the range for both winter surface air temperature and summer precipitation. The latter values are broadly consistent with the expectation value of 5.7 years, for a 10% occurrence rate.

Where observations fall outside the grey range, they often relate to well-known events, such as the extremely cold winter of 1962-63 (bottom left panel). Encouragingly, the Strand 2 projections produce examples of cold winters matching or exceeding 1962-63 in intensity (surface air temperature anomalies < -4°C). This provides evidence that Strand 2 can be useful in assessing the characteristics and impacts associated with future seasonal anomalies (relative to the future climatological average of the period) of a similar magnitude. Similarly, historical seasonal events can be found in Strand 2 that correspond to, or in some cases outstrip, all of the observed cool, warm, dry or wet extremes in Fig. 5.6a.
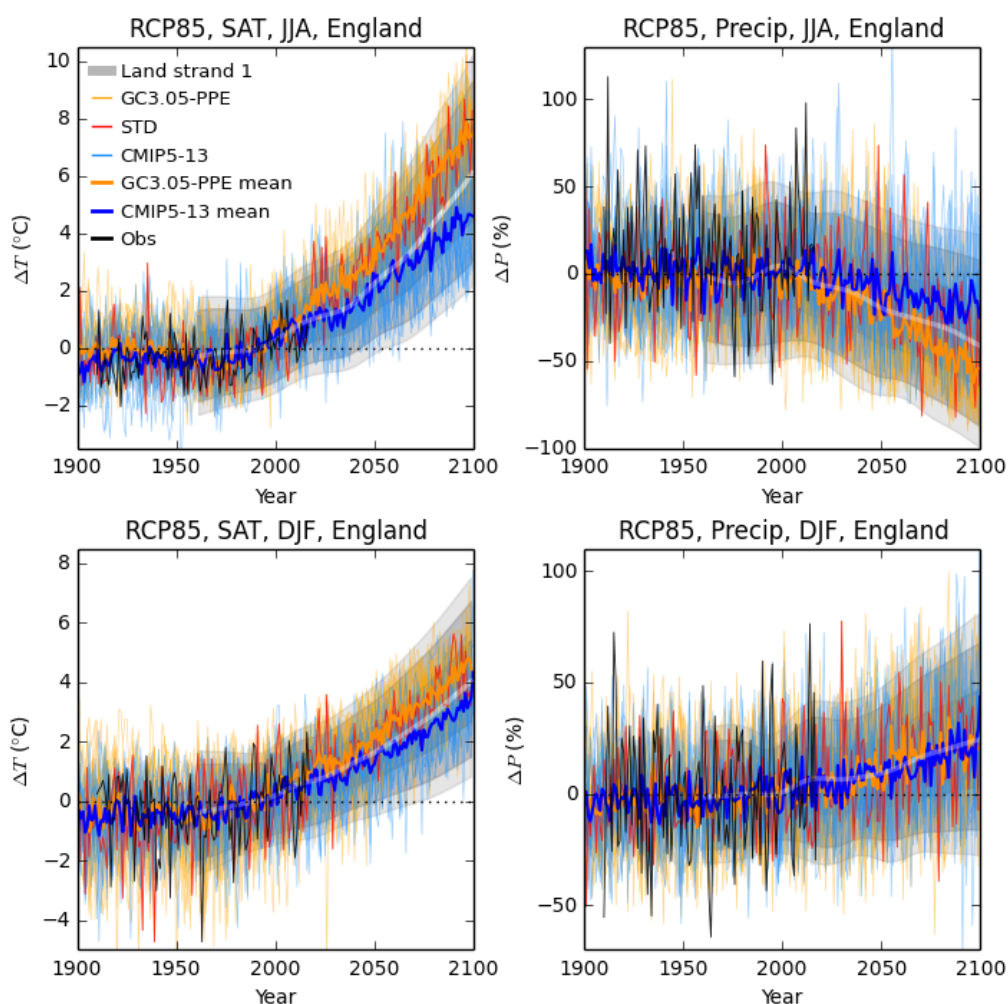
Figure 5.6a. Historical and future anomalies in seasonal surface air temperature (°C, left) and precipitation (%, right) for England. Seasonal anomalies are expressed relative to 1981-2000, and are plotted for winter and summer (lower and upper panels respectively), from Strands 1 and 2. Future changes are based on the RCP8.5 emissions scenario, applied in the projections beyond 2005. The median of the probabilistic projections is the white line, and shades of grey show the 5, 10, 25, 75, 90 and 95% probability levels. Orange and blue lines show members of GC3.05-PPE and CMIP5-13 respectively. The STD member of GC3.05-PPE is in red. The thicker lines show ensemble means, and the black curves show observations from NCIC. The probabilistic projections start from 1960 rather than 1900, as downscaling information (required to produce Strand 1 projections for UK regions, see section 2.2) is only available from that point onwards.

The Strand 1 future projections show a broad range for all four variables, which essentially covers the spread of the combined set of Strand 2 projections throughout. Examples of future seasonal extremes beyond 5-95% probability interval of Strand 1 can be found in all cases, but the frequencies of occurrence are broadly consistent with the 10% that would be expected if Strand 2 events were sampling the same distribution as Strand 1. Note that this was not guaranteed a priori to be the case, since Strand 2 is constructed using a different methodology to Strand 1, and includes (in GC3.05-PPE) a new projection system unrelated to any of the climate model ensembles used in Strand 1. Nevertheless, the inter-Strand comparison is useful on two counts. Firstly, it demonstrates that the uncertainty ranges from Strand 1 are sufficiently robust to encompass the results from new climate modelling capability in Strand 2. Secondly, the comparison confirms the status of outlying seasonal anomalies in the Strand 2 projections as extreme events, in the context of known uncertainties in future climate change.

For surface air temperature, the two Strand 2 ensembles show a significant degree of overlap in their ranges of interannual variability throughout the 21st century. However, the ensemble-mean responses also reveal a clear and increasing divergence. This is also seen in the GMST responses (Fig. 3.20), and underlines that GMST exerts an important influence on the characteristics of the UK surface air temperature changes. Beyond about 2060, the ensemble-mean response of GC3.05-PPE lies between the 75% and 90% probability levels of Strand 1 in summer. This indicates that the PPE responses in Strand 2 should form a useful source of information for studies of heat-related impacts in summer climate that are plausible, but lie near the upper end of the Strand 1 uncertainty ranges. Such studies form part of the evidence base for national Climate Change Risk Assessments (Met Office, CEH and University of Reading, 2015). The CMIP5-13 ensemble-mean generally remains close to, or slightly below, the Strand 1 median response, in both winter and summer. CMIP5-13 is therefore particularly useful for providing examples of unusually cool future seasons (for example winters below the 5% probability level of Strand 1), since GC3.05-PPE does not provide any such events after 2050.

For winter precipitation, Strands 1 and 2 both project a steady shift to wetter distributions of seasonal values. The GC3.05-PPE and CMIP5-13 ensemble means stay close to the median of Strand 1 throughout the coming century. In addition, both Strand 2 ensembles provide numerous examples of wet and dry winter seasons sampling the upper and lower deciles of the evolving Strand 1 distributions. For this variable, therefore, there is a good degree of consistency between the Strand 1 and Strand 2 results. In summer, the Strand 1 distribution shifts to a drier envelope of outcomes (as expected from Fig. 5.4d), the median drying reaching 40% by 2100. The ensemble-mean of CMIP5-13 shows a smaller drying than Strand 1, typically ~20% during the last decade of the 21st century. In contrast, the ensemble-mean of GC3.05-PPE is drier than the Strand 1 median, amounting typically to ~50% during the corresponding period. While the projections show a clear shift to higher probabilities of dry summers, they also suggest that individual wet summers are still possible. For example, the CMIP5-13 projections suggest that a positive seasonal precipitation anomaly of (say) 60% relative to the 1981-2000 average, which occurred only twice in the observed record since 1960, remains a plausible event during 2081-2100. However, Strand 1 suggests that the risk of wet summers declines during the coming century. For example, the 95% probability level reduces from a wet anomaly of ~50% at year 2000 to one of ~ 10% by 2100. This decline is more marked in UKCP18 compared with the earlier work of Sexton and Harris (2015), due to the improved statistical treatment of internal variability in UKCP18 (Appendix A).

More generally, an important application of the seasonal projections lies in assessing the changing risks of specific high-impact events experienced in the recent historical record. For example, beyond the 2020s the Strand 2 projections provide no examples of a winter (measured relative to the 1981-2000 baseline) that corresponds to the 1962-63 surface air temperature anomaly in England. In summer, the observed anomaly of about 2°C, that was experienced during the 2003 heatwave, becomes a typical event by the 2040s according to the Strand 1 median. This is consistent with the results of Christidis et al. (2014). The risk of a seasonal precipitation anomaly in excess of 50%, such as that associated with the wet winter of 2013-14 (Huntingford et al., 2014), is projected to increase. According to Strand 1, for example, the chance of such an event increases from about one in forty at 2000, to about one in five by 2100.

Examples of future extreme seasons in Strands 2 and 3 provide potentially fruitful resources for case study analysis, to support development of narratives associated with specific high-impact weather or climate events. For example, identification of simulated future winters associated with negative NAO conditions could support investigation of how the historical impacts of such events (e.g. the cold December of 2010, Maidens et al., 2013) might change in future. See also Fig. 5.11 and discussion below.
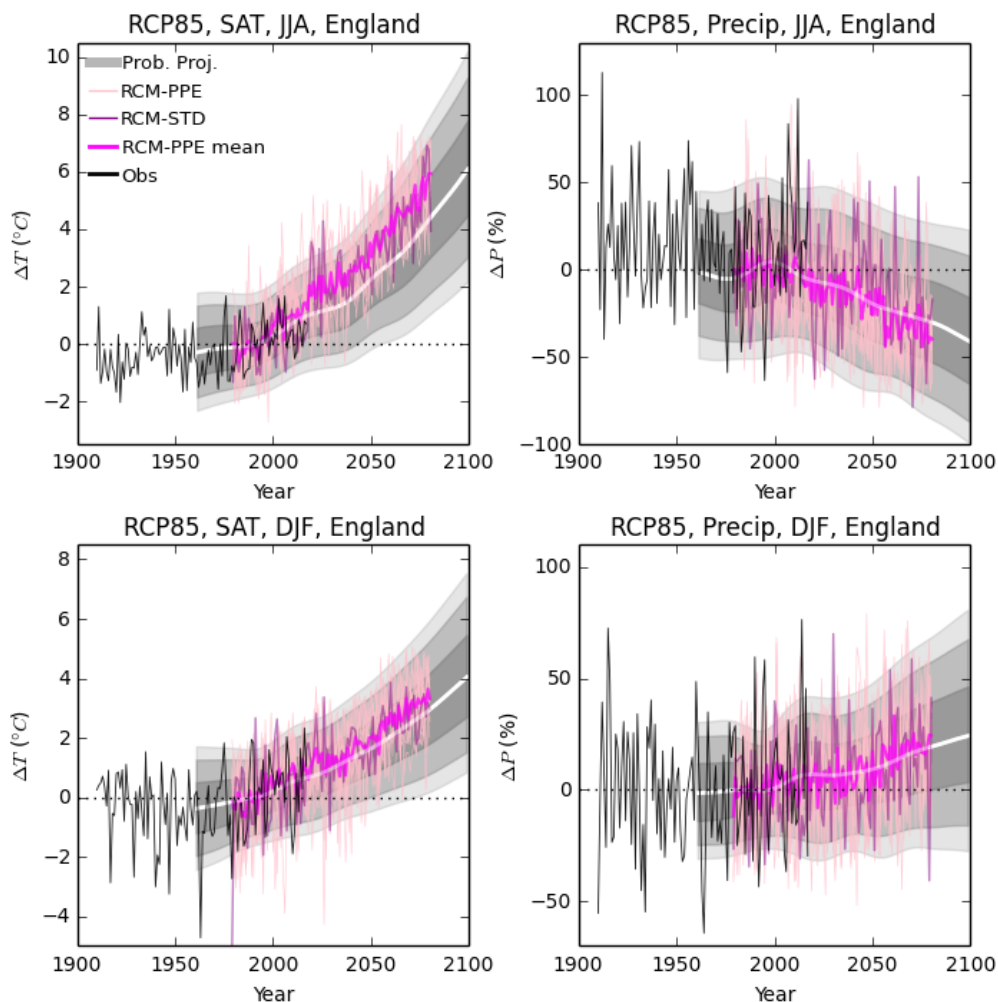

Figure 5.6b. As Fig. 5.6a, except that the pink lines show seasonal anomalies from members of the Strand 3 RCM-PPE simulations, with RCM-STD in purple. The thicker pink line shows the ensemble-mean.

Corresponding seasonal projections for England are shown for Strand 3 in Fig. 5.6b. These show ranges of response very similar to the GC3.05-PPE results of Fig. 5.6a. From the middle of the 21st century, seasonal surface air temperature anomalies are distributed around an ensemble-mean warming between the 75% and 90% probability levels of Strand 1 in summer, and between the 50% and 75% levels in winter. The ensemble-mean changes in winter precipitation track the Strand 1 median quite closely (like the Strand 2 ensembles), and RCM-PPE shares the strong long-term drying signal seen in the GC3.05-PPE simulations. This consistency *at the national scale* between the RCM-PPE and its driving simulations is an expected consequence of the design of the RCM simulations, and supports use of RCM-PPE results to add value to analysis of detailed impacts at local to regional scales (see section 4). This applies both to multidecadal average changes in surface air temperature and precipitation, and to extreme events on daily to seasonal time scales. This is likely to be beneficial, for example, in studies of future impacts on the electricity and rail networks (McColl et al., 2012; Palin et al., 2013).
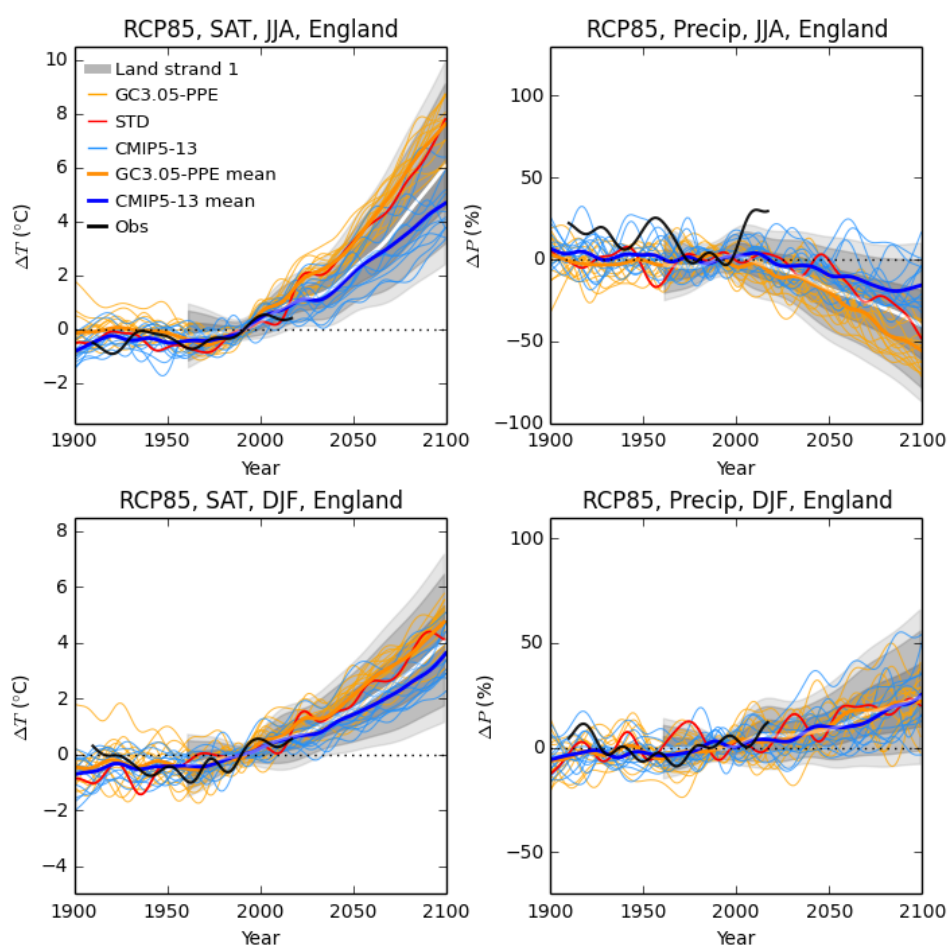
Figure 5.7. As Fig. 5.6a, showing seasonal projections filtered to retain signals of variability and long-term change on time scales of 20 years and longer.

Uncertainties in both internal variability and long-term climate change contribute to the time-dependent ranges of seasonal change shown in Figs. 5.6a,b. In Fig. 5.7, we show the results of filtering the projections of Fig. 5.6a to remove variability on 1-20 year time scales. This isolates signals on longer time scales, including the response to anthropogenic forcing and internal multidecadal variability. Uncertainties in the filtered signals increase during the 21st century. This is probably due mainly to growth in the spread of long-term climate responses contained in the ensembles of climate model simulations that underpin Strands 1 and 2 (e.g. Hawkins and Sutton, 2009, 2011). The filtered changes show a clear separation in the surface air temperature response between the two component ensembles of Strand 2, to a greater extent than the unfiltered changes of Fig. 5.6a. This underlines the advice given in discussion of Fig. 5.4 above, that Strands 2 and 3 cannot be used to obtain estimates of the relative likelihood of alternative future projections.

The impact on the future projections of variability on 1-20 year time scales can be seen by comparing the spread of outcomes in Figs. 5.6a and 5.7 (see also Fig. 2.9). For surface air temperature, the overall range of outcomes is broadened considerably by annual to decadal variability up to the middle of the 21st century. By 2100, however, the spread is dominated by uncertainty in forced climate response. This is despite the continuing presence of substantial internal variability in individual projections, which can drive year-to-year changes of several degrees. For precipitation, variability significantly broadens the spread in seasonal anomalies throughout. The low-pass filtered summer precipitation changes demonstrate the benefits of combining the CMIP5-13 and GC3.05-PPE ensembles: Beyond about 2050, CMIP5-13 provides all the projections showing 20-year outcomes with increases in rainfall, while GC3.05-PPE provides all the examples at the dry end of the Strand 1 distribution. This diversity is important for robust assessment of impacts and development of storylines (see section 1.2).

During the historical period the spread of results from the projection systems is likely to reflect mainly the effects of decadal climate variability. This spread encompasses observations (black lines in Fig. 5.7) in almost all cases. The only exceptions are the positive anomalies in summer precipitation that peak around 1960, and in the recent period since 2010. The occasional occurrence of observed events near, or slightly outside, the extremes of the projected ranges is to be expected (see discussion of Fig. 5.6 above). Nevertheless, it is important to understand the mechanisms that drive the observed events, and assess how well the projection systems reproduce these. The observed decadal wet periods are likely to have been driven, at least in part, by warm SST anomalies in the North Atlantic associated with the positive phase of the (detrended) AMO. Such events are linked with wet summers in the UK and Northern Europe (Sutton and Dong, 2012). In contrast, the AMO was primarily in a negative phase during the baseline period, when average UK summer precipitation was relatively low compared to the 20th century as a whole (Kendon et al., 2017). The ability of the projection systems to replicate such events depends on the ability of the constituent climate models to simulate AMO variability (section 3.4) and capture the observed teleconnection link to UK summer rainfall. This discussion also illustrates that observed 20-year averages for a given period (such as the UKCP18 baseline of 1981-2000), can be significantly influenced by the dominant contemporaneous phases of low frequency climate variability, as well as by the true underlying climatological state.
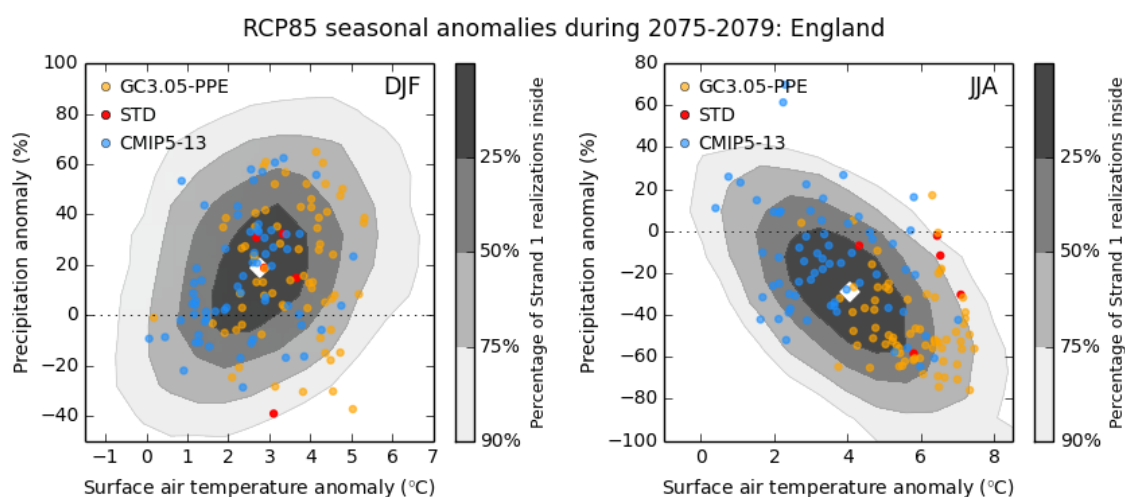
Figure 5.8a. Grey shading shows joint probability distributions from Strand 1 of projected anomalies (relative to 1981-2000) of surface air temperature (°C) and precipitation (%) for individual seasons during 2075-79, under RCP8.5 emissions. Results are shown for England, in winter (left) and summer (right). The white diamond denotes the location of maximum relative probability. The shaded regions contain 25% (darkest), 50%, 75% and 90% (lightest) of the 3000 realisations of which the probabilistic projections are comprised. The 65 blue dots show corresponding changes for individual years during 2075-2079 from each of the CMIP5-13 simulations. Orange dots show individual seasons from 14 of the GC3.05-PPE members, with results from STD in red.

Future impacts are in many cases determined by two or more climate drivers (section 1.2). For example, future occurrences of drought (e.g. Burke and Brown, 2008) are (dependent on the choice of index) sensitive to changes in both surface air temperature and precipitation, as are changes in crop yields (see Fig. 3.24 and discussion in section 3.5). In Figure 5.8a, we compare the joint distributions from Strands 1 and 2 of seasonal anomalies for surface air temperature and precipitation in England, during 2075-79. The left and right panels show winter and summer anomalies respectively, relative to 1981-2000. In both seasons, Strand 1 provides the broadest view of potential combinations of outcomes, emphasising its value in providing context for risk assessments.

In winter, the combined set of Strand 2 projections provides seasonal anomalies that cover 75%[51] of the space of Strand 1 outcomes reasonably well, with the exception of wet winters coupled with a relatively small warming (of 1°C or less). Potential wet or dry outcomes paired with higher levels of warming (>3°C) are quite well represented. The outer 25% of the Strand 1 distribution is less well sampled, although several examples of extreme dry winters coupled with high warming are present in Strand 2. The summer distributions show reasonable sampling of dry seasons coupled with warming in the range 1-4°C (from CMIP5-13 models), and good representation of very dry and warm seasons with precipitation anomalies of -50% to -80% and warming in the range 4-8°C (GC3.05-PPE members). The joint probability distribution also includes seasonal outcomes with increased summer precipitation, of which there are several examples in Strand 2. Overall, Fig. 5.8a shows that Strand 2 provides a relatively broad range of surface air temperature and precipitation changes that shows reasonable correspondence with Strand 1. However, the probabilistic projections reveal additional potential outcomes that are not represented in Strand 2. This demonstrates the difficulty of covering all combinations of impact-drivers in a limited set of climate model simulations.
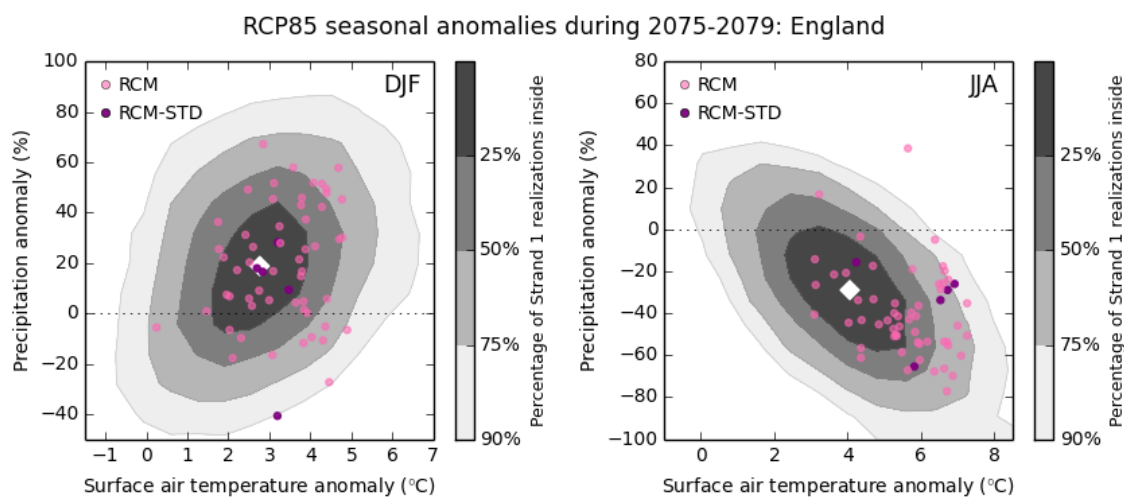
Figure 5.8b. As Fig. 5.8a, except that the 60 dots show seasonal anomalies from the twelve RCM-PPE members during 2075-79, with RCM-STD in purple and other members in pink.

---

[51] The contours in Figs 5.8a,b represent the smallest regions containing the stated percentage of outcomes that can be found in the relevant joint distribution of Strand 1.

Figure 5.8b shows how Strand 3 samples the Strand 1 space of joint outcomes. The winter projections include reasonable representation of anomalies ranging from -10 to +60% in precipitation, and 2-5°C of warming. However, cooler outcomes (either wet or dry) are not represented. In summer, dry seasons coupled with warming of 2-7°C are well covered, but outcomes with warming lower than 3°C are absent, and only two wet summers are present in the sample. Therefore, the sampling of outcomes in Strand 3 is more limited than in Strand 2. This is not surprising, given that Strand 3 is limited to 12 members, and does not include a multi-model counterpart to its PPE component.
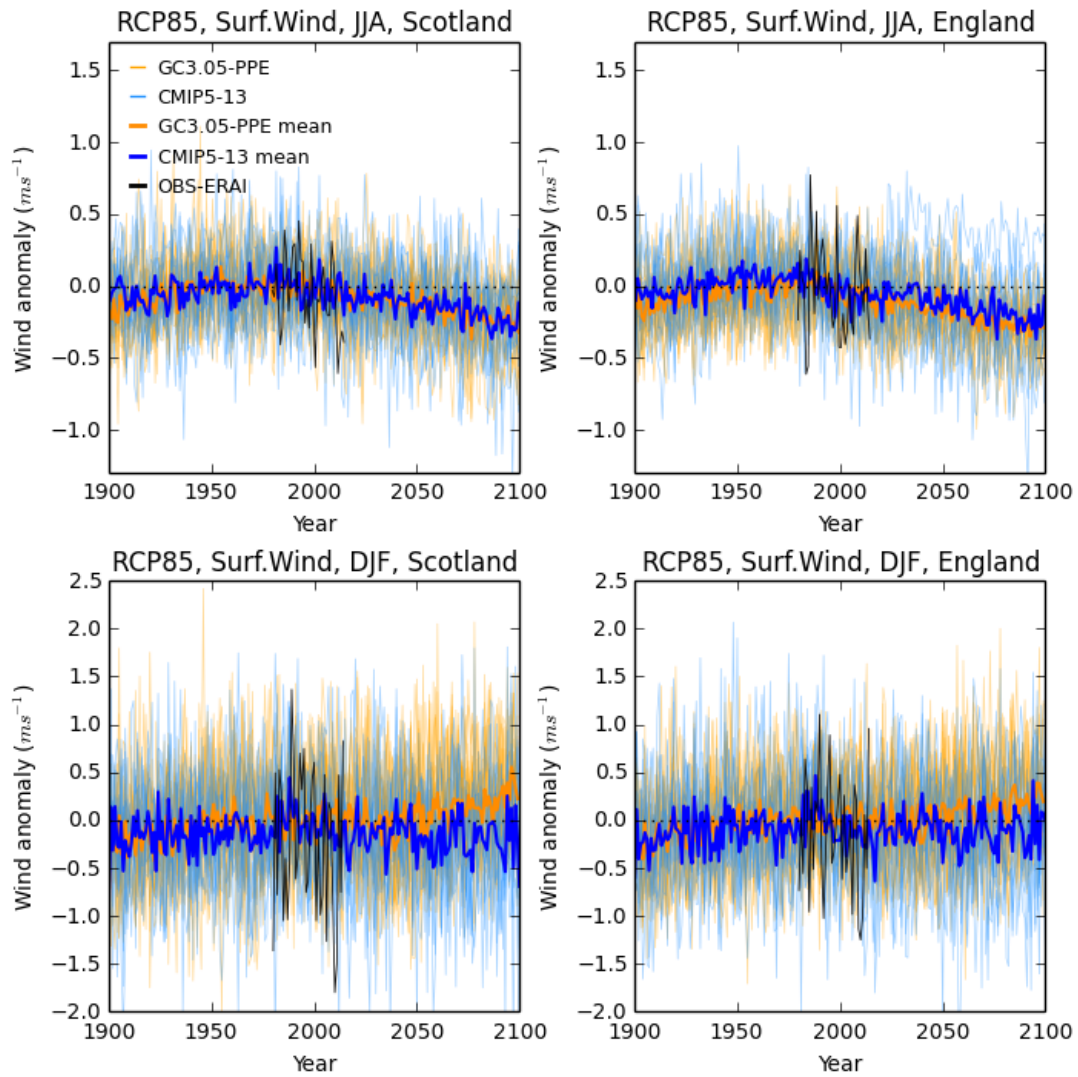
Figure 5.9. Historical and future anomalies in seasonal near-surface wind speed (ms⁻¹) for Scotland (left) and England (right), in summer (top) and winter (bottom). Seasonal anomalies are expressed relative to 1981-2000, and are plotted for Strand 2 of UKCP18. Future changes are based on the RCP8.5 emissions scenario, applied in the projections beyond 2005. Orange and blue lines show members of GC3.05-PPE and CMIP5-13 respectively. The thicker lines show ensemble means, and the black curves show observations from ERA-Interim reanalyses during 1980-2017.

Users wishing to fill such gaps could consider accessing additional climate model simulations. In the case of Strand 2, for example, the CMIP5-13 subset is recommended for general use, because its members have been shown to simulate key characteristics of European climate reasonably well. However, users could consider adding simulations from the wider CMIP5 ensemble in specific applications where including projections from other models is useful, and can be justified with further evaluation. In applications where information at local scales is essential, users may be able to strengthen their assessments of multiple hazards by considering use of multi-model RCM simulations from EuroCordex (see section 4.1). In user applications, it will therefore be important to weigh the benefits of better sampling of multivariate outcomes that is available from Strand 2, against the benefits of downscaling available from Strand 3 (see sections 4.4 and 4.5). In some cases, it may be appropriate to use additional tools such as statistical downscaling techniques or impacts models to derive projections of the relevant indicators, rather than relying on direct climate model output (see section 5.2 for further discussion).

Seasonal projections of average near-surface wind speed over England and Scotland are shown in Figure 5.9, for Strand 2[52]. In winter, there is substantial interannual variability in GC3.05-PPE and CMIP5-13, which is broadly consistent with observations (provided from ERA-Interim). In summer, variability is approximately half of the winter value, again with reasonable consistency between the simulations and observations. During the second half of the 21st century, the ensemble-mean of GC3.05-PPE shows a shift to increased winter wind speed over both Scotland and England. The signal of change is larger over England, but in both countries the changes are modest compared with interannual variability. No trend is apparent in the CMIP5-13 ensemble-mean. We note that EuroCordex simulations, driven by different CMIP5 models, show a range of changes in winter wind speed including examples of both increases and decreases (Kjellström et al., 2018). In GC3.05-PPE, the shift to higher seasonal average wind speeds is accompanied by an increase in the occurrence of winter storms over the UK and Southern Scandinavia (Fig. 5.10, left panel), with reductions to the north and south. The pattern of change implies a strengthening of the southern fork of the winter storm track (Fig. 3.17), with a weakening of the core to the north. The right panel shows a corresponding ensemble-mean pattern of change for ten of the CMIP5-13 simulations for which storm tracking data is available. This shows a similar pattern to that of GC3.05-PPE, but the band of increase in the southern fork is weaker. Zappa et al. (2013) find a similar tripolar pattern of change in the future response of 19 CMIP5 models to the RCP4.5 emissions scenario.

In winter, GC3.05-PPE also projects increases in the frequency of daily occurrences of WT2 (positive NAO) and reductions in WT1 (negative NAO - see Fig. 3.19 and discussion in section 3.4). This leads by the end of the century to annual occurrences of WT2 typically exceeding 30%, while WT1 frequencies reduce to typically 10-15%. These signals of change are not replicated in the nine CMIP5-13 members for which weather typing results are available. In these simulations, average frequencies of WT1 and WT2 remain close to historical values. These results are consistent with those of Fig. 5.9, and show the importance of including two different ensembles of projections in Strand 2. Doing so helps to avoid an overconfident interpretation, for example, of either a shift towards a predominance of positive NAO conditions (suggested by GC3.05-PPE), or of a prediction of null response (suggested by available CMIP5 members).

---

[52] Results from Strand 1 are not provided, because its methodology relies on an assumption that transient climate changes can be estimated from simulations of the equilibrium response to doubled CO_2 (section 2.2, Stage 1). This assumption was found to be invalid in the case of wind speed (Appendix C).
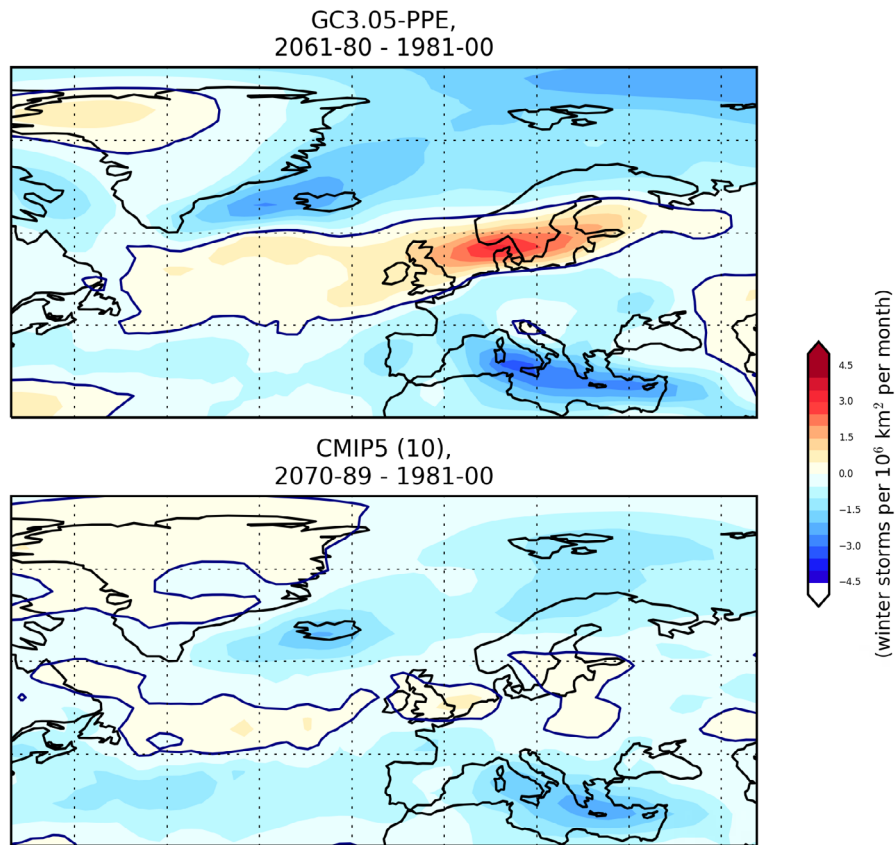
Figure 5.10. Projected changes in the density of winter storms (per 10^6km^2 per month), relative to the 1981-2000 values shown in Fig. 3.17. Increases and reductions are shown in shades of red and blue respectively, with zero change contoured. Upper panel shows the ensemble-mean change in GC3.05-PPE during 2061-2080, lower panel shows the average change for 10 members of CMIP5-13 for 2070-89. A later period is used for CMIP5, because future tracking data was only available from 2070.

In summer, the mean changes in both Strand 2 ensembles show gradual reductions in wind speed during the coming century, for both England and Scotland. In GC3.05-PPE, the reductions in average summer wind speed are accompanied by reductions in the frequency of WT1 and increases in that of WT2. This is the opposite of the winter changes, and indicates fewer episodes of strong westerly flow. However, in the nine available CMIP5 results there is no significant change in the occurrences of WT1 or WT2. Ensemble-mean changes in Strand 3 (not shown) are similar to those of GC3.05-PPE.

Average surface air temperature anomaly (K) during WT 1 days
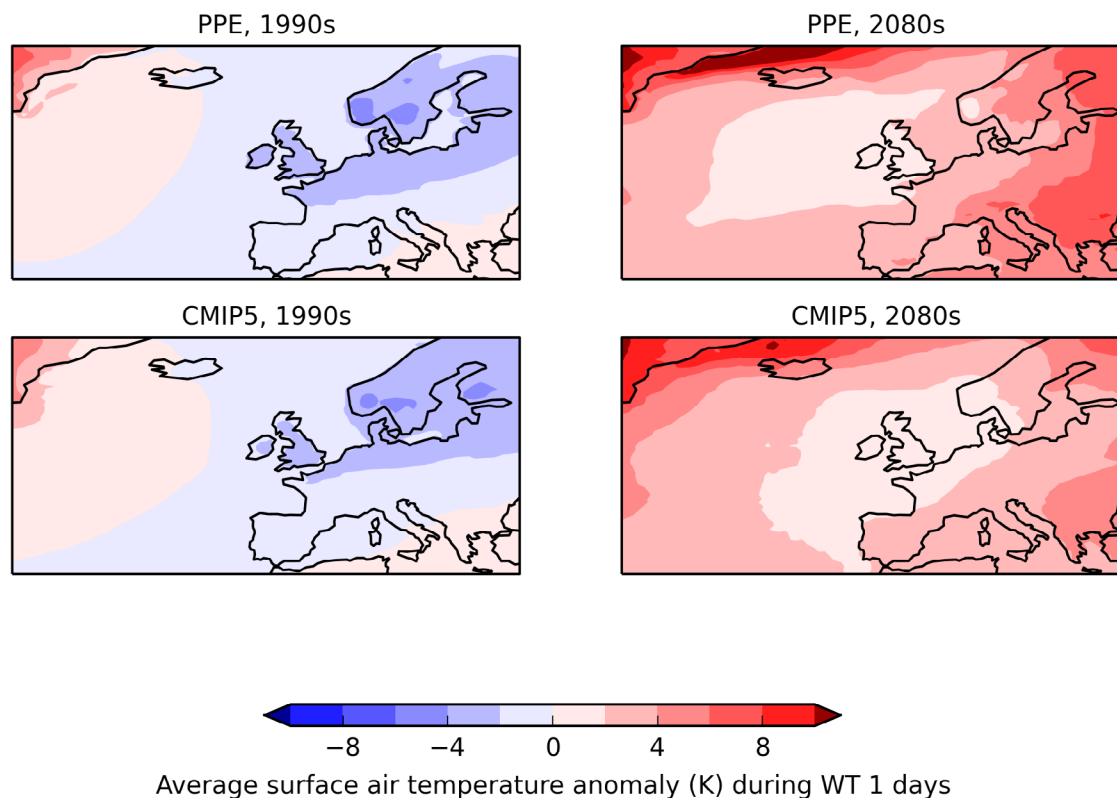
Figure 5.11. Anomalies in surface air temperature (°C) on winter days during which the simulated atmospheric circulation in the North Atlantic/ Europe sector corresponds to weather type 1 of Fig 3.19. Results are shown for the 1990s (left) and 2080s (right), expressed in both cases as anomalies relative to the 1981-2000 baseline. Top panels show composite anomalies averaged over all 15 members of GC3.05-PPE, with corresponding results averaged over nine members of CMIP5-13 shown in lower panels.

In addition to identifying potential changes in the occurrence of specific circulation patterns, the Strand 2 simulations can be used to understand how other indicators might change on typical days associated with the relevant flow regimes. This can be useful in placing the future projections in the context of recent observed events, and understanding potential future changes in impacts and vulnerability. For example, Figure 5.11 shows composite surface air temperature anomalies associated with WT1 during the 1990s and 2080s, expressed in both cases with respect to the 1981-2000 baseline. Under historical conditions, WT1 days in both the GC3.05-PPE and CMIP5 ensembles show cold anomalies of 2-4°C over the UK in winter, associated with occurrences of anomalous easterly flow. By the 2080s, such days are associated with temperatures warmer (by 0-2°C) than the baseline climatology, albeit that they are cooler than the projected average conditions for the 2080s. Under these projections, a reduction in cold-weather disruption would therefore be expected during winters featuring a high frequency of WT1 days, compared to recent experience (e.g. Prior and Kendon, 2011). A caveat is that the magnitude of the change could potentially be enhanced by the winter cold bias in the GC3.05-PPE simulations, given that it leads to a stronger surface albedo feedback than is seen in CMIP5-13 (see Figure 3.22). Nevertheless, such results illustrate the importance of understanding both changes in circulation (driven by regional dynamical responses to large-scale patterns of climate change, e.g. Zappa and Shepherd, 2017), and thermodynamic responses that alter the associated temperature and precipitation characteristics (e.g. Clark and Brown, 2013).

5.2. Interpretation and use of the projections

The three projection systems of Strands 1-3 provide the most extensive range of information delivered to date in a coordinated set of UK land scenarios. To understand how to use the Strands, either separately or in combination, it is important to recall the stakeholder drivers (section 1.2) that motivated their development.

The primary purpose of the probabilistic projections is to ensure that UKCP18 includes a product focused on expressing known uncertainties in future changes, in order to reduce risks of overconfident decision-making in user applications. For example, extensive use of the UKCP09 probabilistic projections was made in CCRA2 (Humphrey and Murphy, 2016). Strand 1 attempts to provide ranges of outcomes consistent with, and limited by, the knowledge incorporated in existing ensembles of climate model projections (see section 2.1). It is based on a much larger set (of ~350) climate model simulations than Strands 2 or 3, augmented by use of emulation techniques to estimate results for climate model variants for which no simulation is available. Its projections are available as 3000 realisations, sampled from distributions that are formally constrained by observations (section 2.2, Stage 2), in order to ensure that the uncertainty ranges of Strand 1 are not broadened excessively through inclusion of outcomes of low credibility. However, results are only available for a limited set of key climate variables (Fung et al., 2018), and the realisations do not possess the full spatial and temporal coherence available directly from climate model output, due to the extensive statistical processing required to create the results[53].

Strands 2 and 3 were developed primarily to support user requirements for more flexible datasets that provide specific examples of physically plausible climate outcomes. These:

- Allow a wide range of climate indicators to be derived (including at fine spatial scales and short timescales).

- Support case studies related to particular meteorological events and impacts studies requiring data with full spatial and temporal coherence.

- Provide a dataset useful for the development of narratives that support decision-making and build confidence in the results through improved scientific understanding (e.g. Hazeleger et al., 2015).

Serving these purposes requires data provided directly from climate model projections, rather than following the statistical postprocessing used in Strand 1. The sets of Strand 2 and 3 projections are intended to provide simulations of the best quality attainable given current modelling capabilities, both globally and in the North Atlantic and European regions. The aim is to provide, as far as possible, credible representations of phenomena (such as the Atlantic storm track) likely to be important drivers of future change, in order to improve the utility of impacts assessments and narratives. In addition, the numbers of projections are limited in size (28 and 12 in Strands 2 and 3 respectively). This choice is influenced by the heavy computational cost of developing and running ensembles of high-resolution global and regional configurations of the GC3.05 climate model, but also reflects an aim to limit the data-processing burden placed on users.

---

[53] Users requiring Strand 1 data for aggregated river-basin, administrative or country regions (Fung et al., 2018) can obtain this from the User Interface at https://ukclimateprojections-ui.metoffice.gov.uk/. Users requiring aggregated data for other regions can create this from Strand 2 or 3 results, but should not attempt to form spatial averages from the gridded 25km Strand 1 data. However, it is acceptable to form time averages of monthly or seasonal Strand 1 values for a specific location, as these have been created using a method that preserves temporal correlations present in the climate model simulations used to construct the probabilistic projections (see Appendix A).

The results of section 5.1 generally reflect the distinct purposes of the Strands, and provide pointers for how their outputs can be combined to improve assessments of impacts, risks and decision options. For example, Strand 1 (as explained above) is intended to provide the main source of information for estimates of uncertainty in UKCP18, and its 10-90% probability ranges duly reveal (in general) a wider range of regional outcomes than ranges derived from the more limited sets of projections included in Strands 2 and 3 (Figs. 5.4a-d).

However, the projections in Strands 2 and 3 provide specific illustrations of plausible outcomes that lie towards the lower or upper ends of the Strand 1 distributions. For example, GC3.05-PPE produces simulations featuring consistently strong positive cloud feedbacks that contribute to a set of projections with strong future warming, both globally and regionally (Figs. 3.20, 3.24, 5.4 and 5.6). For the UK, this results in projections of surface air temperature that sometimes exceed the 90% probability level of Strand 1. Similarly, the GC3.05 projections explore future reductions in summer precipitation close to the lower end of the Strand 1 distributions. For risk assessments, such results emphasise the importance of considering some of the more extreme outcomes shown in Strand 1. This is because a climate model derived from the latest UK capability (Williams et al., 2018), and developed more recently than those included in Strand 1, provides support for some of these.

Here, it is worth emphasising that the results from each Strand are conditional upon the climate modelling information and methodological choices used to produce it. In particular, the probability distributions of Strand 1, whilst generally successful in providing a broad view of uncertainties, are inevitably reflections of the evidence and expert assumptions upon which they are based. Therefore, they should be viewed as conditional probabilities, providing ranges of potential outcomes (characterised typically by the 10% and 90% probability levels) that are useful uncertainty estimates likely to encompass the observed outcome, *based on recent climate models and the knowledge incorporated in them.*

In UKCP09, Sexton and Murphy (2012) showed that the relative probabilities of different outcomes varied to a degree, in response to exploring plausible alternatives for expert choices such as the prior distributions for uncertain model parameters, the choice of observational constraints or the estimates of structural model uncertainties. None of these alternatives questioned the basic shapes of the distributions, and changes in the ranges of outcomes were modest. However, the results illustrate that Strand 1 users should consider the sensitivity of their application to reasonable variations in the probabilities provided, such as those outlined above.

More generally, the presence of systematic biases common to all climate models, and the incomplete nature of current understanding of earth system processes, underlines the importance of exploring the sensitivity of potential adaptation decisions or impact assessments to other sources of information beyond the scope of UKCP18. This might include, for example, idealised studies of possible "climate surprises" arising from events or processes that are either not typically simulated by, or not yet included in, current climate models. Such events might include a future collapse of the AMOC, or a substantial release of carbon from permafrost, or of methane from ocean sediments (Collins et al., 2013).

For assessments of impacts and extreme events, Strands 2 and/or 3 are expected to provide the primary source of information in many applications, as explained above. The global simulations of Strand 2 should be useful for the development of storylines, as the investigation of processes driving UK or European changes can be extended to consider large-scale patterns of climate change remote to the region. Regarding UK impacts, a key choice will be between the broader set of projections (including the CMIP5-13 contributions) available from Strand 2, versus the enhanced spatial detail[54] and improved simulation of extremes available from Strand 3. The appropriate choice is likely to be application-dependent. As noted in section 4.1, for example, UKCP09 users found the spatial detail available from its 25km ensemble of HadRM3-PPE simulations useful for studies of impacts on infrastructure networks, flood risk and water resources. The provision in UKCP18 of 12km simulations from Strand 3 (and forthcoming 2.2km simulations) is likely to offer further benefits in studies of this nature.

Users of Strands 2 and 3 will also need to consider use of bias correction techniques (Fung, 2018), for example in cases where the requirement for adaptation depends on the exceedance of absolute thresholds. Note, however, that bias correction cannot remove the effects of large model errors on projected changes, and needs to be accompanied by a good understanding of relevant earth system phenomena, and how well they are represented in climate models (Maraun et al., 2017). Here, the availability of model simulations with a range of historical error characteristics (e.g. Figs. 3.15, 3.16 and 4.4) will allow users to explore the benefits of alternative approaches, and understand how variations in historical bias are related to projected future changes in variables of interest (e.g. Sørland et al., 2018).

The examples of joint distributions of surface air temperature and precipitation changes, provided in Fig. 5.8, demonstrate that Strand 1 can be used to guide impact studies based on the other Strands. In particular, comparison with Strand 1 can reveal deficiencies in the sampling of potential responses. Strategies to address such limitations could include sourcing additional climate model simulations, as discussed in section 5.1.

Other methods can also be considered for improving sampling of potential multivariate changes. One possibility is to apply statistical methods such as scaling or time-shifting to the existing Strand 2 or Strand 3 projections (Herger et al., 2015), in order to provide additional "derived" projections in sparsely populated regions of the Strand 1 distribution. Such approaches (see Gohar et al., 2018) are worth considering in applications where the full flexibility available from time series of climate model output is required. Examples would include extreme event analysis, or assessment of changes in vulnerability arising from multiple hazards affecting different locations simultaneously (e.g. McColl et al., 2012).

However, in applications focused on a specific impact in a specific region, or in a set of locations for which the risk of simultaneous events is not a major concern, weather generators (Jones et al., 2010)[55] or other statistical downscaling tools could be used to derive additional estimates of future daily climate characteristics for particular locations, conditioned on changes in climatological average conditions sampled from Strand 1. However, such methods do not include information on climate changes at fine spatial and temporal scales.

---

[54] Note that the size of the datasets available for download (https://ukclimateprojections-ui.metoffice.gov.uk/) increases considerably with resolution.

[55] UKCP18 does not include a specific weather generator as UKCP09 did, so interested users would need to source a suitable tool independently.

Another option is to use an impacts modelling system capable of sampling a wide range of potential changes in primary climate variables such as surface air temperature and precipitation. For example, a "response surfaces" approach (Fronzek et al., 2010) can be used to predict impacts such as crop production (e.g. Ferrise et al., 2011), by combining an impacts model with a statistical emulator to convert probabilistic climate projections (such as Strand 1) into probabilistic projections of yields. Such methods have advantages of computational efficiency, and of direct calibration of the downscaling or impacts tool against historical observations. However, they typically rely on assumptions of temporal stationarity in the characteristics of internal climate variability, or in relationships between the climate variables prescribed as inputs and other variables upon which the derived impact or weather sequence may depend. In some applications, it may be advantageous to combine approaches based on direct use of climate model output with methods involving impacts models or statistical downscaling methods, given that each approach possesses specific strengths and limitations (Kay and Jones, 2012).

In CCRA2, increased use was made of the "H++" concept, originally developed in the sea level and storm surge component of UKCP09 (Lowe et al., 2009). An H++ scenario is an unlikely but possible high-end outcome that is useful for contingency planning. It may be derived from a range of sources including model projections, idealised sensitivity studies, theoretical insight or past observations, One of the CCRA2 research projects (Met Office, CEH and University of Reading, 2015) used this approach to develop advice on limits for heatwaves, cold extremes, rainfall and river flow events, droughts, windstorms and wildfires. The GC3.05-PPE and RCM-PPE results in Strands 2 and 3 are likely to be particularly useful in future H++ studies. This is because their emergent properties include levels of warming, increases in winter precipitation and reductions in summer precipitation that are plausible, but relatively extreme in the context of other evidence, including Strand 1 and CMIP5 models.

6. Summary

The UKCP18 land projections include three Strands of evidence, each motivated by a combination of user requirements and recent developments in scientific capabilities.

Strand 1

Strand 1 is an update to the probabilistic projections that formed the core component of the previous UKCP09 land scenarios (Jenkins et al., 2009). This provides continuity, by ensuring that a product focused on uncertainties remains available for use in risk assessments (https://www.theccc.org.uk/tackling-climate-change/preparing-for-climate-change/uk-climate-change-risk-assessment-2017/), and wider impacts and adaptation activities. The new probabilistic projections are provided for five emissions scenarios. The SRES A1B scenario is included to support comparison against UKCP09 results (in which it was the "medium" scenario). Also included are the Representative Concentration Pathway scenarios (RCP2.5, 4.5, 6.0 and 8.5) considered in the IPCC Fifth Assessment Report (Collins et al., 2013). These range from a scenario assuming aggressive mitigation measures (RCP2.6) to one in which carbon emissions are assumed to rise substantially (RCP8.5).

Strand 1 quantifies uncertainties consistent with the knowledge incorporated in several pre-existing ensembles of climate model projections, plus the effects of internal climate variability. The probability distributions express ranges of outcomes, and estimates of the relative likelihood of alternative outcomes within these ranges, consistent with the relative strength of evidence behind alternative future changes.

As in UKCP09 (Murphy et al., 2009), the probabilistic projections are based on global and regional perturbed parameter ensembles (PPEs) derived from the HadCM3 climate model, an alternative multi-model ensemble of international climate models and a set of formal observational constraints. These three lines of evidence are combined using a Bayesian statistical framework based on, and updated from, that used in UKCP09 (Sexton et al., 2012; Harris et al., 2013). An "emissions-driven" approach is used, in order to account for uncertainties in modelling the earth's carbon cycle alongside those associated with the physical response of the climate system to changing greenhouse gas concentrations. The probabilities are conditional on the climate modelling information used to produce them, and also upon various subjective choices required to implement the methodology, such as expert prior distributions for uncertain model parameters, the selection of observational constraints, and the design and calibration of statistical techniques needed to combine the information.

Key developments from UKCP09 include use of the recent CMIP5 generation of international climate models (instead of CMIP3) to estimate the contribution of structural modelling choices to uncertainties in projections, and expansion of the set of observational constraints to improve identification and down-weighting of low credibility outcomes. In UKCP18, the probabilistic projections are presented at the annual time scale, rather than as the 30-year average changes of UKCP09. This broadens their utility by accounting for climate variability on interannual to decadal time scales, thus supporting analysis of the changing risks of extreme seasonal anomalies (Sexton and Harris, 2015). This presentation also allows users to choose their own historical baseline period, if they wish to explore alternatives to the standard period of 1981-2000 (or the alternatives of 1961-1990 and 1981-2010) provided from the UKCP18 User Interface at https://ukclimateprojections-ui.metoffice.gov.uk/.

In comparison to UKCP09, the projected changes in global mean and UK surface temperatures (under A1B emissions) are slightly lower in UKCP18. This is partly due to the inclusion in UKCP18 of a new constraint on projected atmospheric CO_2 concentrations, which rules out outcomes with the highest future values. However, there is considerable overlap between the two sets of projections, because uncertainty ranges are broad in both cases. For summer precipitation, the median of the Strand 1 projections shows progressive reductions during the 21st century, which are largest over Southern England and South Wales, as was also the case in UKCP09.

Strand 2

Following publication of UKCP09, a considerable body of user feedback accumulated. In addition to confirming the value of uncertainty estimates from the probabilistic projections, this feedback also stressed the importance of access to flexible datasets consisting of raw climate model output. This stems from a requirement for time series of meteorological data with full spatial and temporal coherence, and capable of supporting derivation of a wide range of climate change indicators. During this period, climate model development has also continued worldwide. Recent global models developed in the UK have demonstrated improved representation of the dynamics of regional climate variability in the North Atlantic and European region (Scaife al., 2012, 2014), due at least in part to improvements in their horizontal and vertical resolution. Such improvements in capability can help to justify use of such models to provide information useful for development of "storylines": The development of physically realistic narratives of how large-scale climate responses might drive changes experienced in the UK (Zappa and Shepherd, 2017), or case studies of how specific types of high-impact event might change (Hazeleger et al., 2015).

Strand 2 was conceived in response to these user and scientific drivers. A new PPE was developed, using a coupled-ocean atmosphere model (GC3.05) configured at a higher resolution (~60km horizontally, with 85 vertical levels) than all CMIP5 models used in climate scenario simulations. It contains all the major developments in scientific formulation included in HadGEM3-GC3.1 (Williams et al., 2018), the UK model that will form part of the next generation of international climate models (CMIP6, Eyring et al., 2016).

The identification of model variants for the new PPE (GC3.05-PPE) involved three main stages of selection. These considered short-range retrospective weather forecasts, followed by five-year climate simulations (both using the atmosphere component with prescribed sea surface temperatures), and finally multidecadal coupled ocean-atmosphere simulations. During this process, a pool of 2800 candidate model variants was reduced to a final set of fifteen GC3.05-PPE members, using a range of criteria. These included global performance metrics for a set of standard climate variables (in the atmosphere-only simulations), evaluation of simulated climate in the North Atlantic/Europe sector (in both the atmosphere-only and coupled simulations), and historical climate changes simulated during the 20th century (coupled simulations). The selection also involved consideration of idealised climate change experiments, with the aim of choosing as diverse a set of GC3.05-PPE members as possible, subject to achievement of credible simulations of historical climate.

A set of thirteen CMIP5 models (CMIP5-13) were added to GC3.05-PPE, creating a combined set of 28 simulations of historical and future climate from 1900-2100, using the RCP8.5 scenario during the 21st century. CMIP5 models were selected using a set of global and regional assessment criteria drawn mainly from published literature, in order to identify models suitable for general use in climate impacts applications focused on the UK and Europe. Including CMIP5-13 adds sampling of structural diversity in model formulation, and broadens the range of future climate changes included in Strand 2.

In addition to providing a new resource for understanding future climate change in the UK and Europe, the Strand 2 projections can be used to assess potential impacts in other international regions. This includes assessment of inward risks to the UK arising from issues such as security of food or water supplies.

The Strand 2 simulations were made using models without an interactive carbon cycle, and hence used prescribed time series of CO_2 concentration. All CMIP5-13 members used the standard RCP8.5 concentration pathway. In GC3.05-PPE, a range of concentration pathways was used, in order to consider a range of outcomes consistent with those projected in Strand 1.

Although Strand 2 considers only a single emissions scenario (due to the computational expense of the GC3.05-PPE simulations), a set of derived results is also available for the RCP2.6 scenario. These were generated by applying statistical scaling and time-shifting techniques to the RCP8.5 results (Gohar et al., 2018). Some key results from the RCP8.5 simulations are summarised in the final section below.

Strand 3

Information from regional climate models (RCMs) has formed an important component of UK climate scenarios since the release of UKCIP02 (Hulme et al., 2002). Such dynamical downscaling experiments add value to the output of global climate simulations by accounting for local effects of surface forcing from mountains, coastlines and inland water bodies, while also offering better resolution of mesoscale circulations in the atmosphere.

In UKCP18, a 12-member PPE of RCM simulations (RCM-PPE) was run from 1980-2080, using the European domain of the international EuroCordex experiment (http://www.euro-cordex.net/). Each member was driven at its lower and lateral boundaries by a time series of SSTs, sea-ice extents and atmospheric variables taken from a member of GC3.05-PPE, and used an identical set of parameter perturbations to the driving simulation. The RCM was configured at 12km horizontal resolution. It provides the most detailed simulations available in UK national scenarios to date, following the 50km simulations of UKCIP02 and the 25km PPE (derived from HadCM3) that formed part of UKCP09. The new RCM-PPE is expected to be an important resource for regional impacts assessments, in cases where these are dependent on credible information at scales finer than those resolved effectively by global models. These include, for example, hydrological impacts such as river flows, flood and drought risks, and changing hazards to infrastructure networks such as electricity and rail.

Evaluation of RCM-PPE demonstrates substantial skill in simulating costal and orographic effects on UK precipitation, in both the long-term climatological average and extreme heavy daily events. Other impacts of enhanced resolution include improvements in the simulation of cold winter days, and a general increase in cloud cover and precipitation.

Projected changes under the RCP8.5 scenario

This report compares a selection of changes from Strands 1-3 for the RCP8.5 scenario. More examples are given in Lowe et al. (2018), alongside further examples of results for the other emissions scenarios included in Strand 1.

Strand 1 gives the broadest uncertainty ranges, which generally encompass the spread of projections from Strands 2 and 3. This reflects the design criteria explained above, and underlines that the probabilistic projections provide the primary source of information for use in defining low- and high-end mainstream outcomes to frame adaptation studies and risk assessments.

We also use central changes to illustrate intermediate outcomes in each Strand. For Strand 1 this is the median of the relevant probability distribution. For Strands 2 and 3 we use results from specific projections in the middle of the relevant ranked set of changes. However, for Strands 2 and 3 users should not assume that the future projections are distributed around the central change in a unimodal format resembling, for example, a Gaussian distribution. This is because these Strands are not designed to support estimates of the relative likelihood of alternative outcomes within their ranges of response.

For changes in surface air temperature over the UK, the Strand 2 projections evolve into two semi-distinct clusters during the second half of the 21st century. This is due in part to the influence of a high level of simulated global warming in GC3.05-PPE. Overall, the Strand 2 projections explore much of the spread of potential UK temperature changes defined by Strand 1, with GC3.05-PPE members providing most of the warmest outcomes, particularly after 2050. This indicates that GC3.05-PPE members will be particularly useful in investigations of plausible high-end scenarios of heat-related impacts. By 2061-2080, median changes in Strand 1 (averaged over the UK) amount to 2.4°C in winter and 3.1°C in summer, with corresponding 10-90% ranges of 0.7 to 4.2°C and 0.9 to 5.4°C, respectively.

Changes in summer precipitation also show evidence of clustering in Strand 2. The GC3.05-PPE simulations project reductions, which are strongest over England and Wales. The ensemble-average drying for England reaches 50% by the end of the 21st century, compared with a median reduction of about 40% in Strand 1. The corresponding ensemble-average of CMIP5-13 shows a smaller drying of ~20%. A few members of CMIP5-13 show little change or a small increase, consistent with the upper end of the Strand 1 range. For winter precipitation changes this separation does not occur, ensemble-mean changes in GC3.05-PPE and CMIP5-13 showing similar trends of gradual increase.

 Projected changes in Strand 3 follow those of the driving Strand 2 simulations at national scales, but add detail to the changes in surface air temperature and precipitation. This applies both to long-term averages and changes in the intensity of extreme events. For example, in many coastal regions the winter precipitation changes are enhanced in the regional model projections, indicating that they will be useful for contingency planning activities associated with potential coastal flooding events.

GC3.05-PPE projects an ensemble-mean increase in the occurrence of winter storms over the UK, with reductions to the north and south. Storm track diagnostics are also available for ten members of CMIP5-13. These show a similar ensemble-mean pattern, but with a weaker increase over the UK. GC3.05-PPE also predicts an increase in the frequency of daily circulation patterns associated with the positive (westerly) phase of the North Atlantic Oscillation (NAO). These changes are associated with ensemble-mean increases in wind speed at the national scale. However, CMIP5-13 shows no signal of change in wind speed, and the nine members for which daily weather typing information is available do not show the shift to positive NAO conditions projected in GC3.05-PPE.

In summer, reductions in wind speed are found in the ensemble-means of both GC3.05-PPE and CMIP5-13. In GC3.05-PPE, the changes are accompanied by a shift to fewer days characterised by strong westerly flow, although this is not seen in the nine CMIP5-13 members referred to above.

In general, the results show evidence of realistic levels of climate variability at seasonal time scales, in surface air temperature, wind and precipitation over the UK. This helps to justify using the results to assess the future risks of recent high-impact events such as the 2003 summer heatwave, the cold winter of 2009-10 or the wet summer of 2012. However, the Strand 1 distributions of winter precipitation anomalies underestimate the risks of extreme wet or dry seasons found in historical observations.

Many impacts are dependent on changes in multiple climate variables. In this regard, it is instructive to compare the projections of Strands 2 and 3 against corresponding joint distributions from Strand 1. This provides useful context for studies based on Strands 2 or 3, by highlighting potential combinations of outcomes that may not be represented in their projections. Considering joint temperature and precipitation changes during 2075-79 as an example, the Strand 2 projections sample much of the Strand 1 distribution, although some gaps are apparent. The sampling available from the Strand 3 projections is more limited, particularly in summer when potential wet seasons are largely absent. This is due to the consistent drying signal discussed above.

Concluding remarks

The UKCP18 land projections are built from a range of climate modelling inputs beyond that of any previous UK scenarios. This includes global and regional perturbed parameter ensembles derived from two UK models (HadCM3 and GC3.05), plus CMIP5 ensembles of earth system models (used in Strand 1) and physical atmosphere-ocean models (used in Strand 2). The specific sets of models (or model variants) used in each Strand, and the techniques used to create projection systems from them, are determined by various methodological choices.

The results that emerge from each of the three projection systems are conditional on these specific choices, and are inevitably somewhat different. In some applications, it may be appropriate to consider results from only one Strand, whereas in others it is likely to be beneficial to combine information from multiple Strands. For example, a case study of a future wet winter might use Strand 2 to select a broad-scale atmospheric circulation pattern from a GC3.05-PPE member analogous to a past event such as the winter of 2013-14, and then use the corresponding RCM-PPE simulation to analyse the regional flood risks associated with the event.

Another example might be use of Strand 1 distributions to identify limitations in the sampling of multiple drivers of impacts, as discussed above. In such cases, users could consider addressing such gaps by sourcing additional climate model simulations from other projects, or using statistical techniques or impacts models to derive additional impacts scenarios from the UKCP18 outputs.

Finally, we emphasise that the information in the UKCP18 projections reflects current modelling technologies and the knowledge incorporated in them. It is therefore liable to be updated in future, as capabilities evolve. The changes in the UKCP18 probabilistic projections, relative to their UKCP09 counterparts, are one illustration of this. Developments in capability may also lead to extensions in the type of information that can be provided. One example is the recent first use in climate research of kilometre-scale regional modelling that explicitly represents the dynamics of large convective storms (Kendon et al., 2014). An ensemble of UK projections using a 2.2km convective-permitting model is under production, and will be published in a second phase of UKCP18.This will provide new information on potential changes in sub-daily rainfall extremes, and will be helpful in applications such as urban planning and flash flooding.

References

Adler RF, Huffman GJ, Chang A, Ferraro R, Xie P, Janowiak J, Rudolf B, Schneider U, Curtis S, Bolvin D, Gruber A, Susskind J, Arkin P (2003). The version 2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present). J. Hydromet. 4:1147–1167.

Aldrin M, Holden M, Guttorp P, Skeie RB, Myhre G, Berntsen TK (2012). Bayesian estimation of climate sensitivity based on a simple climate model fitted to observations of hemispheric temperatures and global ocean heat content. Environmetrics 23:253–271.

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

Andrews T, Gregory JM, Webb MJ, Taylor KE (2012). Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. Geophys. Res. Lett. 39:L09712.

Andrews T, Ringer MA (2014). Cloud feedbacks, rapid adjustments, and the forcing–response relationship in a transient CO_2 reversibility scenario. J. Clim. 27:1799-1818.

Ansell TJ and co-authors (2006). Daily mean sea level pressure reconstructions for the European - North Atlantic region for the period 1850-2003. J. Clim 19:2717-2742.

Aumann HH, Chahine MT, Gautier C, Goldberg MD, Kalnay E, McMillin LM, Revercomb H, Rosenkranz PW, Smith WL, Staelin DH, Strow LL, Susskind J (2003). AIRS/AMSU/HSB on the aqua mission: design, science objectives, data products, and processing systems. IEEE Trans. Geosci. Remote. Sens. 41:253–264.

Baek H-J, Lee J, Lee H-S, Hyun Y-K, Cho C, Kwon W-T, Marzin C, Gan S-Y, Kim M-J, Choi D-H, Lee J, Lee J, Boo K-O, Kang H-S, Byun Y-H (2013). Climate change in the 21st Century simulated by HadGEM2-AO under representative concentration pathways. Asia-Pacific J. Atmos. Sci. 49:603-618.

Ballantyne AP, Alden CB, Miller JB, Tans PP, White JWC (2012). Increase in observed net carbon dioxide uptake by land and oceans during the past 50 years. Nature 488: 70–72.

Ban N, Schmidli J, Schär C (2015). Heavy precipitation in a changing climate: Does short-term precipitation increase faster? Geophys. Res. Lett. 42:1165-1172.

Beniston M, Stephenson DB, Christensen OB, Ferro CAT, Frei C, Goyette S, Halsnaes K, Holt T, Jylhä K, Koffi B, Palutikof J, Schöll R, Semmler T, Woth K (2007). Future extreme events in European climate: an exploration of regional climate model projections. Clim. Change 81:71-95.

Berry DI, Kent EC (2009). A new air-sea interaction gridded dataset from ICOADS with uncertainty estimates. Bull. Amer. Met. Soc. 90:645–656.

Best MJ, Pryor M, Clark DB, Rooney GG, Essery RLH, Ménard CB, Edwards JM, Hendry MA, Porson A, Gedney N, Mercado LM, Sitch S, Blyth E, Boucher O, Cox PM, Grimmond CSB, Harding RJ (2011). The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes. Geosci. Model Dev. 4:677-699.

Bindoff N, Stott P, Achuta Rao K, Allen M, Gillett N, Gutzler D, Hansingo K, Hegerl G, Hu Y, Jain S, Mokhov I, Overland J, Perlwitz J, Sebbari R, Zhang X (2013). Detection and attribution of climate change: from global to regional. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (Stocker T, Qin D, Plattner G-K, Tignor M, Allen S, Boschung J, Nauels A, Xia Y, Bex V, Midgley P eds.). Cambridge University Press, Cambridge, UK and New York, NY, USA, pp. 876-952, doi:10.1017/CBO9781107415324.022.

Bladé I, Liebmann B, Fortuny D, van Oldenborgh GJ (2012). Observed and simulated impacts of the summer NAO in Europe: Implications for projected drying in the Mediterranean region. Clim. Dyn 39:709–727.

Bond TC, Bergstrom RW (2006). Light absorption by carbonaceous particles: An investigative review. Aerosol Sci. Technol. 40, 27–67, https://doi.org/10.1080/02786820500421521, 2006

Bony S, Colman R, Kattsov VM, Allan RP, Bretherton,CS, Dufresne J-L, Hall A, Hallegatte S, Holland MM, Ingram W, Randall DA, Soden BJ, Tseloudis G, Webb MJ (2006). How well do we understand and evaluate climate change feedback processes? J. Clim. 19:3445-3482.

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

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

Braganza K, Karoly DJ, Hirst AC, Mann ME, Stott P, Stouffer RJ, Tett SFB (2003). Simple indices of global climate variability and change: Part I — variability and correlation structure. Clim. Dyn. 20:491–502.

Brierley CM, Collins M, Thorpe AJ (2010). The impact of perturbations to ocean-model parameters on climate and climate change in a coupled model. Clim. Dyn. 34:325-343.

Brohan P, Kennedy JJ, Harris I, Tett SFB, Jones PD (2006). Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850. J. Geophys. Res. 111:D12106, doi:10.1029/2005JD006548

Brooks RH, Corey AT (1964). Hydraulic properties of porous media. Hydrol. Pap. 3, Colorado State University, Fort Collins, 27 pp.

Burke EJ, Brown SJ (2008). Evaluating uncertainties in the projection of future drought. J. Hydromet. 9:292-299.

Burke EJ, Perry RHJ, Brown SJ (2010). An extreme value analysis of UK drought and projections of change in the future. J. Hydrol. 388:131-143.

Carslaw KS, Lee LA, Reddington CL, Pringle KJ, Rap A, Forster PM, Mann GW, Spracklen DV, Woodhouse MT, Regayre LA, Pierce JR (2013). Large contribution of natural aerosols to uncertainty in indirect forcing. Nature 503:67-71.

Chadwick R, Good P (2013). Understanding nonlinear tropical precipitation responses to CO_2 forcing. Geophys. Res. Lett. 40:4911–4915, doi:10.1002/grl.50932.

Christensen JH, Krishna Kumar K, Aldrian E, An S-I, Cavalcanti IFA, de Castro M, Dong W, Goswami P, Hall A, Kanyanga JK, Kitoh A, Kossin J, Lau N-C, Renwick J, Stephenson DB, Xie S-P, Zhou T (2013). Climate phenomena and their relevance for future regional climate change. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA

Christidis N, Jones GS, Stott PA (2014). Dramatically increasing chance of extremely hot summers since the 2003 European heatwave. Nat. Clim. Change 5:46-50.

Clark RT, Murphy JM, Brown SJ (2010). Do global warming targets limit heatwave risk? Geophys. Res. Lett. 37:L17703.

Clark RT, Brown SJ (2013). Influences of circulation and climate change on European summer heat extremes. J. Clim. 26:9621-9632.

Collins M, Booth BBB, Harris GR, Murphy JM, Sexton DMH, Webb MJ (2006). Towards quantifying uncertainty in transient climate change. Clim. Dyn. 27:127–147.

Collins M, Booth BBB, Bhaskaran B, Harris GR, Murphy JM, Sexton DMH, Webb MJ (2011). Climate model errors, feedbacks and forcings. A comparison of perturbed physics and multi-model ensembles. Clim. Dyn. 36:1737-1766.

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

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

Cumming JA, Wooff DA (2007). Dimension reduction via principal variables. Comp. Stat. Data Anal. 52:550-565.

Dai A, Fyfe J, Xie S-P, Dai X (2015). Decadal modulation of global surface temperature by internal climate variability. Nat. Clim. Change 5:555-559.

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

Dentener F, Kinne S, Bond T, Boucher O, Cofala J, Generoso S, Ginoux P, Gong S, Hoelzemann JJ,. Ito A, Marelli L, Penner JE, Putaud J-P, Textor C, Schulz M, van der Werf GR, Wilson J (2006). Emissions of primary aerosol and precursor gases in the years 2000 and 1750 prescribed data-sets for AeroCom. Atmos. Chem. Phys. 6:4321-4344.

Deser C, Phillips A, Bourdette V, Teng H (2012). Uncertainty in climate change projections: the role of internal variability. Clim. Dyn. 38:527–546.

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

da Silva AM, Young CC, Levitus S (1994) Atlas of surface marine data, vol 1: algorithms and procedures. Tech Rep 6, U.S. Department of Commerce, NOAA, NESDIS.

Dlugokencky E, Tans P, NOAA/ESRL (www.esrl.noaa.gov/gmd/ccgg/trends/)

Domingues CM, Church JA, White NJ, Gleckler PJ, Wijffels SE, Barker PM, Dunn JR (2008). Improved estimates of upper-ocean warming and multi-decadal sea-level rise. Nature 453:1090-1093, doi:10.1038/nature07080.

Etheridge DM, Steele LP, Langenfelds RL, France RJ, Barnola J-M, Morgan VI (1996). Natural and anthropogenic changes in atmospheric CO_2 over the last 1000 years from air in Antarctic ice and firn. J. Geophys. Res. 101:4115-4128.

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

Ferrise R, Moriondo M, Bindi M (2011). Probabilistic assessments of climate change impacts on durum wheat in the Mediterranean region. Nat. Hazards Earth Syst. Sci. 11:1293–1302.

Flato G, Marotzke J, Abiodun B, Braconnot P, Chou S, Collins W, Cox P, Driouech F, Emori S, Eyring V, Forest C, Gleckler P, Guilyardi E, Jakob C, Kattsov V, Reason C, Rummukainen M (2013). 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 (Stocker T, Qin D, Plattner G-K, Tignor M, Allen S, Boschung J, Nauels A, Xia Y, Bex V, Midgley P eds.). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 741-866, doi:10.1017/ CBO9781107415324.020.

Folland CK, Knight J (2009). The Summer North Atlantic Oscillation: Past, present, and future. J. Clim. 22:1082-1103.

Fowler HJ, Ekström M, Kilsby CG, Jones PD (2005). New estimates of future changes in extreme rainfall across the UK using regional climate model integrations. 1. Assessment of control climate. J. Hydrol. 300:212–233.

Frei C, Christensen JH, Déqué M, Jacob D, Jones RG, Vidale PL (2003). Daily precipitation statistics in regional climate models: Evaluation and intercomparison for the European Alps. J. Geophys. Res. 108(D3):4124.

Frigg R, Smith LA, Stainforth DA (2015). An assessment of the foundational assumptions in high-resolution climate projections: The case of UKCP09. Synthese 192:3979-4008.

Fronzek S, Carter TR, Räisänen J, Ruokolainen L, Luoto M (2010). Applying probabilistic projections of climate change with impact models: a case study for sub-arctic palsa mires in Fennoscandia. Clim. Change 99:515–534,doi:10.1007/s10584-009-9679-y.

Fung F (2018). UKCP18 Guidance: Bias correction.
Available from https://ukclimateprojections.metoffice.gov.uk.

Fung F, Stephens A, Wilson A (2018). UKCP18 Guidance: Data availability, access and formats. Available from https://ukclimateprojections.metoffice.gov.uk.

Gates WL, Boyle JS, Covey C, Dease CG, Doutriaux CM, Drach RS, Fiorino M, Gleckler PJ, Hnilo JJ, Marlais SM, Phillips TJ, Potter GL, Santer BD, Sperber KR, Taylor KE, Williams DN (1999). An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). Bull. Amer. Met. Soc. 80:29–55.

Giorgi F, Francisco R (2000). Uncertainties in regional climate change predictions. A regional analysis of ensemble simulations with the HadCM2 GCM. Clim. Dyn. 16:169–182.

Goldstein M, Rougier JC (2004). Probabilistic formulations for transferring inferences from mathematical models to physical systems. SIAM J. Sci. Comp. 26:467–487.

Gohar L, Bernie D, Good P, Lowe JA (2018). UKCP18 derived projections of future climate over the UK. Available from https://ukclimateprojections.metoffice.gov.uk.

Good SA, Martin MJ, Rayner NA (2013). EN4: quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates, J. Geophys. Res. Oceans 118:6704-6716, doi:10.1002/2013JC009067

Gordon C, Cooper C, Senior CA, Banks HT, Gregory JM, Johns TC, Mitchell JFB, Wood RA (2000). The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. Clim. Dyn. 16:147–168.

Gregory JM, Jones CD, Cadule P, Friedlingstein P (2009). Quantifying carbon cycle feedbacks. J. Clim. 22:5232-5250.

Gregory JM, Andrews T (2016). Variation in climate sensitivity and feedback parameters during the historical record. Geophys. Res. Lett. 43:3911-3920.

Gregory J, Webb M (2008). Tropospheric adjustment induces a cloud component in CO_2 forcing. J. Clim. 21:58-71.

Hall A, Qu X (2006). Using the current seasonal cycle to constrain snow albedo feedback in future climate change. Geophys. Res. Lett. 33: L03502.

Hamill TM (1997). Reliability diagrams for multicategory probabilistic forecasts. Weather and Forecasting 12:736-741.

Harris GR, Sexton DMH, Booth BBB, Collins M, Murphy JM, Webb MJ (2006). Frequency distributions of transient regional climate change from perturbed physics ensembles of general circulation model simulations. Clim. Dyn. 27:357–375.

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

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

Harris GR, Murphy JM, Booth BBB, Sexton DMH (2019). Constraining the climate response of emission-driven Earth System models. In preparation.

Harrison EP, Minnis P, Barkstrom BR, Ramanathan V, Cess RD, Gibson GG (1990). Seasonal variation of cloud

radiative forcing derived from the Earth Radiation Budget Experiment. J. Geophys. Res. 95:18687–18703.

Hawkins E, Sutton R (2009). The potential to narrow uncertainty in regional climate predictions. Bull. Amer. Met. Soc. 90:1095-1107.

Hawkins E, Sutton R (2011). The potential to narrow uncertainty in projections of regional precipitation change. Clim. Dyn. 37:407-418.

Hawkins E, Osborne TM, Ho CK, Challinor AJ (2013). Calibration and bias correction of climate projections for crop modelling: An idealised case study over Europe. Ag. Forest Met. 170:19-31.

Haylock MR, Hofstra N, Klein Tank AMG, Klok EJ, Jones PD, New M (2008). A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. J. Geophys. Res. 113:D20119.

Hazeleger W, van den Hurk BJJM, Min E, van Oldenborgh GJ, Petersen AC, Stainforth DA, Vasileiadou E, Smith LA (2015). Tales of future weather. Nat. Clim. Change 5:107-113.

He J, Soden BJ (2016). The impact of SST biases on projections of anthropogenic climate change: A greater role for atmosphere-only models? Geophys. Res. Lett. 43:7745–7750, doi:10.1002/2016GL069803.

Herger N, Sanderson BM, Knutti R (2015). Improved pattern scaling approaches for the use in climate impact studies. Geophys. Res. Lett. 42:3486–3494, doi:10.1002/2015GL063569.

Hewitt HT, Copsey D, Culverwell ID, Harris CM, Hill RSR, Keen AB, McLaren AJ, Hunke EC (2011). Design and implementation of the infrastructure of HadGEM3: the next-generation Met Office climate modelling system, Geosci. Model Dev. 4:223–253, doi:10.5194/gmd-4-223-2011.

Ho CK, Stephenson DB, Collins M, Ferro CAT, Brown SJ (2012). Calibration strategies: a source of additional uncertainty in climate change projections. Bull. Amer. Met. Soc. 93:21-26.

Hodges KI, Lee RW, Bengtsson L (2011). A comparison of extra-tropical cyclones in recent reanalyses ERA-Interim, NASA MERRA, NCEP CFSR, and JRA-25. J. Clim. 24:4888–4906, doi:10.1175/2011JCLI4097.1.

Hofstra N, Haylock M, New M, Jones PD (2009). Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature. J. Geophys. Res. 114:D21101.

Holben BN, Tanre D, Smirnov A, Eck TF, Slutsker I, Abuhassan N, Newcomb WW, Schafer J, Chatenet B, Lavenue F, Kaufman YJ, Vande Castle J, Setzer A, Markham B, Clark D, Frouin R, Halthore R, Karnieli A, O'Neill NT, Pietras C, Pinker RT, Voss K, Zibordi G (2001). An emerging ground-based aerosol climatology: Aerosol Optical Depth from AERONET. J. Geophys. Res. 106:12067-12097.

Hulme M, Jenkins GJ, Lu X, Turnpenny JR, Mitchell TD, Jones RG, Lowe JA, Murphy JM, Hassell D, Boorman P, McDonald R, Hill S (2002). Climate change scenarios for the United Kingdom: The UKCIP02 Scientific Report. Tyndall Centre for Climate Change Research, School of Environmental Sciences, University of East Anglia, Norwich, UK, 120pp.

Humphrey K, Murphy J (2016). UK Climate Change Risk Assessment Evidence Report: Chapter 1, Introduction. Contributing authors: Harris G, Brown S, Lowe J, McCarthy M, Jevrejeva S, Watts G, Johns D, Bell M. Report prepared for the Adaptation Sub-Committee of the Committee on Climate Change, London. Available from…

https://www.theccc.org.uk/uk-climate-change-risk-assessment-2017/ccra-chapters/introduction/

Huntingford C, Marsh T, Scaife A, Kendon E, Hannaford J, Kay A, Lockwood M, Prudhomme C, Reynard N, Parry S, Lowe J, Screen J, Ward H, Roberts M, Stott P, Bell V, Bailey M, Jenkins A, Legg T, Otto F, Massey M, Schaller N, Slingo J, Allen M (2014). Potential influences on the United Kingdom's floods of winter 2013/14. Nat. Clim. Change 4:769-777.

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

Hurrell J, Meehl GA, Bader D, Delworth TL, Kirtman B, Wielicki B (2009). A unified modeling approach to climate system prediction. Bull. Amer. Met. Soc. 90:1819-1832.

Hurtt GC, Chini LP, Frolking S, Betts R, Feddema J, Fischer G, Fisk JP, Hibbard K, Houghton RA, Janetos A, Jones C, Kindermann G, Kinoshita T, Klein Goldewijk K, Riahi K, Shevliakova E, Smith S, Stehfest E, Thomson A, Thornton P, van Vuuren DP, Wang Y (2011). Harmonization of land-use scenarios for the period 1500–2100: 600 Years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. Clim. Change 109:117-161.

Hyder P, Edwards J, Allan RP, Hewitt HT, Bracegirdle TJ, Gregory JM, Wood R, Meijers A, Mulcahy J, Field P, Furtado K, Bodas-Salcedo A, Williams K, Copsey D, Josey S, Liu C-L, Roberts C, Sanchez C, Ridley J, Thorpe L, Hardiman SC, Mayer M, Berry D, Belcher S (2018). Critical Southern Ocean climate model biases traced to atmospheric model cloud errors. Nature Comms. 9:3625. doi: 10.1038/s41467-018-05634-2

Ingleby B, Huddleston M (2007). Quality control of ocean temperature and salinity profiles – historical and real-time data. J. Mar. Sys. 55:158-175.

Ishii M, Kimoto M, Kachi M (2003). Historical ocean subsurface temperature analysis with error estimates. Mon. Wea. Rev. 131: 51-73.

Jacob D, Petersen J, Eggert B, Alias A, Christensen OB, Bouwer LM, Braun A, Colette A, Déqué M, Georgievski G (2014). EURO-CORDEX: new high-resolution climate change projections for European impact research. Reg. Environ. Change 14:563-578.

Jenkins GJ, Murphy JM, Sexton DMH, Lowe JA, Jones P, Kilsby CG (2009). UK Climate Projections: Briefing Report. Met Office Hadley Centre, Exeter, UK.

Jones CD, and co-authors (2011). The HadGEM2-ES implementation of CMIP5 centennial simulations. Geosci. Model. Dev. 4: 543-570.

Jones P, Kilsby C, Harpham C, Glenis V, Burton A (2010). UK Climate Projections Science Report: Projections of Future Daily Climate for the UK from the Weather Generator. University of Newcastle, UK. Available from http://ukclimateprojections.metoffice.gov.uk/22530

Jones RG, Murphy JM, Noguer M (1995). Simulation of climate change over Europe using a nested regional climate model. I: assessment of control climate, including sensitivity to location of lateral boundary conditions. Q. J. R. Meteorol. Soc. 121: 1413-1449.

Josey SA, Kent EC, Oakley D, Taylor PK (1996). A new global air-sea heat and momentum climatology. Int. WOCE Newsl. 24:3–5.

Joshi MM, Gregory JM, Webb MJ, Sexton DMH, Johns TC (2008). Mechanisms for the land/sea warming contrast exhibited by simulations of climate change. Clim. Dyn. 30:455–465.

Joshi M, Hawkins E, Sutton R, Lowe J, Frame D (2011). Projections of when temperature change will exceed 2ºC above pre-industrial levels. Nat. Clim. Change 1:407-412.

Karmalkar AV, Sexton DMH, Murphy JM, Booth BBB, Rostron JW, McNeall DJ (2018). Efficient design to identify plausible and diverse variants of a model for climate projections. Part II: Development and validation of methodology. Clim. Dyn., submitted.

Karpechko A, Manzini E (2012). Stratospheric influence on tropospheric climate change in the Northern Hemisphere. J. Geophys. Res.:117, D05133.

Karoly DJ, Stott PA (2006). Anthropogenic warming of central England temperature. Atmos. Sci. Lett. 7:81-85.

Kay A, Jones R (2012). Comparison of the use of alternative UKCP09 products for modelling the impacts of climate change on flood frequency. Clim. Change 114:211-230.

Kendon EJ, Roberts NM, Fowler HJ, Roberts MJ, Chan SC, Senior CA (2014). Heavier summer downpours with climate change revealed by weather forecast resolution model. Nat. Clim. Change 4:570-576.

Kendon EJ, Ban N, Roberts NM, Fowler HJ, Roberts MJ, Chan SC, Evans JP, Fosser G, Wilkinson JM (2017). Do convection-permitting regional climate models improve projections of future precipitation change? Bull. Amer. Met. Soc. 98 (1): doi:10.1175/BAMS-D-15-0004.1

Kendon M, McCarthy M, Jevrejeva S, Legg T (2017). State of the UK Climate 2016, Met Office, Exeter, UK.

Kent C, Pope E, Thompson V, Lewis K, Scaife AA, Dunstone N (2017). Using climate model simulations to assess the current climate risk to maize production. Environ. Res. Lett. 12:054012.

Kjellström E, Nikulin G, Strandberg G, Christensen OB, Jacob D, Keuler K, Lenderink G, van Meijgaard E, Schär C, Somot S, Sørland SL, Teichmann C, Vautard R (2018). European climate change at global mean temperature increases of 1.5 and 2°C above pre-industrial conditions as simulated by the EURO-CORDEX regional climate models. Earth Syst. Dynam. 9:459-478.

Knutti R, Allen MR, Friedlingstein P, Gregory JM, Hegerl GC, Meehl GA, Meinshausen M, Murphy JM, Plattner G-K, Raper SCB, Stocker TF, Stott PA, Teng H, Wigley TML (2008). A review of uncertainties in global temperature projections over the twenty-first century. J. Clim. 21: 2651-2663.

Knutti R, Masson D, Gettelman A (2013). Climate model genealogy: Generation CMIP5 and how we got here. Geophys. Res. Let. 40: 1194-1199.

Kotlarski S, Lüthi D, Schär C (2015). The elevation dependency of 21st century European climate change: an RCM ensemble perspective. Int. J. Climatol. 35: 3902–3920.

Lambert FH, Collins M, Harris GR, Murphy JM, Sexton DMH, Booth BBB (2013). Interactions between uncertainties in different climate system components simulated by a fully-coupled general circulation model. Clim. Dyn. 41:3055-3072.

Lean JL (2009). Available at https://www.geo.fu-berlin.de/en/met/ag/strat/forschung/SOLARIS/Input_data/CMIP5_solar_irradiance

Le Quéré C and co-authors (2015). Global carbon budget 2015. Earth Syst. Sci. Data 7:349–396, doi:10.5194/essd-7-349-2015.

Levitus S, Antonov JI, Boyer TP, Baranova OK, Garcia HE, Locarnini RA, Mishonov AV, Reagan JR, Seidov D, Yarosh ES, Zweng MM (2012). World ocean heat content and thermosteric sea level change (0-2000m), 1955-2010. Geophys. Res. Lett. 39: L10603, doi:10.1029/2012GL051106. Download: https://www.nodc.noaa.gov/OC5/3M_HEAT_CONTENT/index.html

Liu Y, Daum PH, Guo H, Peng Y (2008). Dispersion bias, dispersion effect, and the aerosol-cloud conundrum. Environ. Res. Lett. 3:045021, doi:10.1088/1748-9326/3/4/045021.

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

Lowe JA, Howard TP, Pardaens A, Tinker J, Holt J, Wakelin S, Milne G, Leake J, Wolf J, Horsburgh K, Reeder T, Jenkins G, Ridley J, Dye S, Bradley S (2009). UK Climate Projections science report: Marine and coastal projections. Met Office Hadley Centre, Exeter, UK.

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

Ma H-Y, Xie S, Klein SA, Williams KD, Boyle JS, Bony S, Douville H, Fermepin S, Medeiros B, Tyteca S, Watanabe M, Williamson D (2014). On the correspondence between mean forecast errors and climate errors in CMIP5 Models. J. Clim. 27:1781-1798.

Maidens A, Arribas A, Scaife AA, MacLachlan C, Peterson D, Knight J (2013). The influence of surface forcings on prediction of the North Atlantic Oscillation regime of winter 2010/11. Mon. Wea. Rev. 141:3801-3813.

Malavelle FF, Haywood JM, Jones A, Gettelman A, Clarisse L, Bauduin S, Allan RP, Karset IHH, Kristjansson JE, Oreopoulos L, Cho N, Lee D, Bellouin N, Boucher O, Grosvenor DP, Carslaw KS, Dhomse S, Mann GW, Schmidt A, Coe H, Hartley ME, Dalvi M, Hill AA, Johnson BT, Johnson CE, Knight JR, O'Connor FM, Partridge DG, Stier P, Myhre G, Platnick S, Stephens GL, Takahashi H, Thordarson T (2017). Strong constraints on aerosol-cloud interactions from volcanic eruptions. Nature 546:485-491.

Mann GW, Carslaw K, Spracklen DV, Ridley DA, Manktelow PT, Chipperfield MP, Pickering SJ, Johnson CE (2010). Description and evaluation of GLOMAP-mode: a modal global aerosol microphysics model for the UKCA composition-climate model. Geosci. Model Dev. 3:519-551.

Maraun D, Shepherd TG, Widmann M, Zappa G, Walton D, Gutiérrez JM, Hagemann S, Richter I, Soares PMM, Hall A, Mearns LO (2017). Towards process-informed bias correction of climate change simulations. Nat. Clim. Change 7:764-773.

Martin G, Milton S, Senior C, Brooks M, Ineson S, Reichler T, Kim J (2010). Analysis and reduction of systematic errors through a seamless approach to modeling weather and climate. J.Clim. 23:5933-5957.

Massey N, Jones R, Otto FEL, Aina T, Wilson S, Murphy JM, Hassell D, Yamazaki YH, Allen MR (2015). weather@home – development and validation of a very large ensemble modelling system for

probabilistic event attribution. Q. J. R. Meteorol. Soc. 141:1528–1545, doi:10.1002/qj.2455.

Mauritzen C, Zivkovic T, Veldore V (2017). On the relationship between climate sensitivity and modelling uncertainty. Tellus A 69: 1327765, https://doi.org/10.1080/16000870.2017.1327765

McColl L, Palin EJ, Thornton HE, Sexton DMH, Betts R, Mylne K (2012). Assessing the potential impact of climate change on the UK's electricity network. Clim. Change 115:821-835.

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

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

McSweeney CF, Sexton D, Bett P, McDonald R, Thornton H (2018b). Evaluating regional drivers of mean UK temperature and precipitation in a coupled perturbed parameter ensemble. In preparation.

Meijers AJS (2014). The Southern Ocean in the Coupled Model Intercomparison Project phase 5. Phil. Trans. R. Soc. A 372:20130296. http://dx.doi.org/10.1098/rsta.2013.0296

Meinshausen M, Smith SJ, Calvin KV, Daniel JS, Kainuma MLT, Lamarque J-F, Matsumoto K, Montzka SA, Raper SCB, Riahi K, Thomson AM, Velders GJM, van Vuuren D (2011). The RCP greenhouse gas concentrations and their extension from 1765 to 2300. Clim. Change (Special Issue), doi: 10.1007/s10584-011-0156-z.

Menary MB, Roberts CD, Palmer MD, Halloran PR, Jackson L, Wood RA, Müller WA, Matei D, Lee S-K (2013). Mechanisms of aerosol-forced AMOC variability in a state of the art climate model, J. Geophys. Res. Oceans 118: 2087–2096, doi:10.1002/jgrc.20178.

Met Office, CEH and University of Reading (2015). Developing H++ climate change scenarios for heat waves, cold snaps, droughts, floods, windstorms and wildfires. Report prepared for the ASC (Contributing authors: Steven Wade, Michael Sanderson, Nicola Golding, Jason Lowe and Richard Betts, Met Office; Nick Reynard, Alison Kay, Lisa Stewart and Christel Prudhomme, Centre for Ecology and Hydrology (CEH) and Len Shaffrey, Ben Lloyd-Hughes and Ben Harvey, University of Reading).

Morice CP, Kennedy JJ, Rayner NA, Jones PD (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 dataset. J. Geophys. Res. 117: D08101, doi:10.1029/2011JD017187.

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

Murphy JM, Sexton DMH, Jenkins GJ, Boorman PM, Booth BBB, Brown CC, Clark RT, Collins M, Harris GR, Kendon EJ, Betts RA, Brown SJ, Howard TP, Humphrey KA, McCarthy MP, McDonald RE, Stephens A, Wallace C, Warren R, Wilby R, Wood RA (2009). UK Climate Projections Science Report: Climate change projections. Met Office Hadley Centre, Exeter.

Murphy JM, Booth BBB, Boulton CA, Clark RT, Harris GR, Lowe JA, Sexton DMH (2014). Transient climate

changes in a perturbed parameter ensemble of emissions-driven earth system model simulations. Clim. Dyn. 43:2855-2885.

Myhre G, Shindell D, Bréon F-M, Collins W, Fuglestvedt J, Huang J, Koch D, Lamarque J-F, Lee D, Mendoza B, Nakajima T, Robock A, Stephens G, Takemura T, Zhang H (2013). Anthropogenic and Natural Radiative Forcing. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 659–740, doi:10.1017/CBO9781107415324.018.

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

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

Noguer M, Jones R, Murphy J (1998). Sources of systematic errors in the climatology of a regional climate model over Europe. Clim. Dyn. 14: 691-712.

Olson R, Sriver R, Chang W, Haran M, Urban N, Keller K (2013). What is the effect of unresolved internal climate variability on climate sensitivity estimates? J. Geophys. Res. 118:4348-4358.

Oueslati B, Bellon G (2015). The double ITCZ bias in CMIP5 models: interaction between SST, large-scale circulation and precipitation. Clim. Dyn. 44: 585-607.

Palin EJ, Thornton HE, Mathison CT, McCarthy RE, Clark RT, Dora J (2013). Future projections of temperature-related climate change impacts on the railway network of Great Britain. Clim. Change 120:71-93.

Pall P, Aina T, Stone D, Stott P, Nozawa T, Hilberts A, Lohmann D, Allen M. (2011). Anthropogenic greenhouse gas contribution to flood risk in England and Wales in autumn 2000. Nature 470:382-386.

Perez J, Menendez M, Mendez FJ, Losada IJ (2014). Evaluating the performance of CMIP3 and CMIP5 global climate models over the North East Atlantic region. Clim. Dyn. 43:2663-2680.

Perry MC, Hollis DM (2005). The generation of monthly gridded datasets for a range of climatic variables over the UK. Int. J. Climatol. 25:1041–1054, doi: 10.1002/joc.1161.

Perry M, Hollis D, Elms M (2009). The generation of daily gridded datasets of temperature and rainfall for the UK. National Climate Information Centre, Climate Memorandum No. 24, http://www.metoffice.gov.uk/climatechange/science/monitoring/ukcp09/methods.html

Piani C, Weedon GP, Best M, Gomes SM, Viterbo P, Hagemann S, Haerter JO (2010). Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. J. Hydrol. 395:199-215.

Phillips TJ, Potter GL, Williamson DL, Cederwall RT, Boyle JS, Fiorino M, Hnilo JJ, Olson JG, Xie S, Yio JJ (2004). Evaluating parameterizations in general circulation models: Climate simulation meets weather prediction. Bull. Amer. Met. Soc. 85:1903-1915.

Porson A, Clark PA, Harman IN, Best MJ, Belcher SE (2010). Implementation of a new urban energy budget scheme in the MetUM. Part I: Description and idealized simulations. Q. J. R. Meteorol. Soc. 136:1514-1529.

Prestele R, Arneth A, Bondeau A, de Noblet-Ducoudré N, Pugh TAM, Sitch S, Stehfest E, Verburg PH (2017). Current challenges of implementing anthropogenic land-use and land-cover change in models contributing to climate change assessments. Earth. Syst. Dynam. 8:369–386.

Prior J, Kendon M (2011). The UK winter of 2009/2010 compared with severe winters of the last 100 years. Weather 66:4-10.

Prudhomme C, Young A, Watts G, Haxton C, Crooks S, Williamson J, Davies H, Dadson S, Allen S (2012). The drying up of Britain? A national estimate of changes in seasonal river flows from 11 Regional Climate Model simulations. Hydrol. Proc. 2:1115-1118.

Raes F, Seinfeld JH (2009). New directions: Climate change and air pollution abatement: A bumpy road. Atmos. Environ. 43:5132-5133.

Rajczak J, Schär C (2017). Projections of future precipitation extremes over Europe: A multimodel assessment of climate simulations. J. Geophys. Res. Atmos. 122: 10773-10800.

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

Reynolds RW, Rayner NA, Smith TM, Stokes DC, Wang W (2002). An improved in situ and satellite SST analysis for climate. J. Clim. 15:1609-1625.

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

Ringer MA, Andrews T, Webb MJ (2014). Global-mean radiative feedbacks and forcing in atmosphere-only and coupled atmosphere-ocean climate change experiments. Geophys. Res. Lett. 41: 4035–4042, doi:10.1002/2014GL060347.

Roberts CM, Gordon C, Cooper C (1997). The origin of flux adjustments in a coupled model. Mon. Wea. Rev. 125:909-925.

Rodwell M, Palmer T (2007). Using numerical weather prediction to assess climate models. Q. J. R. Meteorol. Soc. 133:129-146.

Rossow WB, Zhang YC (1995). Calculation of surface and top of atmosphere radiative fluxes from physical quantities based on ISCCP data sets. 2, validation and first results. J. Geophys. Res. 100:1167–1197.

Rossow WB, Schiffer RA (1999). Advances in understanding clouds from ISCCP. Bull. Amer. Met. Soc. 80:2261-2287.

Rostron JW, Sexton DMH, Yamazaki K, McSweeney CF (2018). Key biases, responses and parameter sensitivities in the HadGEM3-GA7.05 perturbed parameter ensemble. In preparation.

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

Rummukainen M (2016). Added value in regional climate modelling. WIREs Clim. Change 7:145-159. doi: 10.1002/wcc.378.

Sanchez C, Williams KD, Collins M (2016). Improved stochastic physics schemes for global weather and climate models. Q. J. R. Meteorol. Soc. 142:147–159, doi:10.1002/qj.2640.

Sanderson BM (2011). A multimodel study of parametric uncertainty in predictions of climate response to rising greenhouse gas concentrations. J. Clim. 24:1362-1377.

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

Sanderson BM, Knutti R, Caldwell P (2015b). Addressing interdependency in a multimodel ensemble by interpolation of model properties. J. Clim. 28:5150-5170.

Sanderson MG, Wiltshire A, Betts RA (2012). Projected changes in water availability in the United Kingdom. Water Resources Res. 48:W08512.

Sato M, Hansen JE, McCormick MP, Pollack JB (1993). Stratospheric aerosol optical depths (1850-1990). J. Geophys. Res. 98:22987-22994.

Scaife AA, Copsey D, Gordon C, Harris C, Hinton T, Keeley S, O'Neill A, Roberts M, Williams K (2011). Improved Atlantic winter blocking in a climate model. Geophys. Res. Lett. 38:L23703.

Scaife AA, Spangehl T, Fereday DR, Cubasch U, Langematz U, Akiyoshi H, Bekki S, Butchart N, Chipperfield MP, Gettelman A, Hardiman SC, Michou M, Rozanov E, Shepherd TG (2012). Climate change projections and stratosphere-troposphere interaction. Clim. Dyn. 38:2089-2097.

Scaife AA, Arribas A, Blockley E, Brookshaw A, Clark RT, Dunstone N, Eade R, Fereday D, Folland CK, Gordon M, Hermanson L, Knight JR, Lea DJ, MacLachlan C, Maidens A, Martin M, Peterson AK, Smith D, Vellinga M, Wallace E, Waters J, Williams A (2014). Skillful long-range prediction of European and North American winters. Geophys. Res. Lett. 41:2514-2519.

Selten FM, Branstator GW, Kliphuis M, Dijkstra HA (2004). Tropical origins for recent and future northern hemisphere climate change. Geophys. Res. Lett. 31:L21205.

Seneviratne SI, Corti T, Davin EL, Hirschi M, Jaeger EB, Lehner I, Orlowsky B, Teuling AJ (2010). Investigating soil moisture–climate interactions in a changing climate: A review. Earth-Sci. Rev. 99:125-161. ISSN 0012-8252, https://doi.org/10.1016/j.earscirev.2010.02.004.

Sen Gupta A, Muir LC, Brown JN, Phipps SJ, Durack PJ, Monselesan D, Wijffels SE (2012). Climate drift in the CMIP3 models. J. Clim. 25:4621–4640.

Sen Gupta A, Jourdain NC, Brown JN, Monselesan (2013). Climate drift in the CMIP5 models. J. Clim. 26:8597-8615.

Senior CA, Andrews T, Burton C, Chadwick R, Copsey D, Graham T, Hyder P, Jackson L, McDonald R, Ridley J, Ringer M, Tsushima Y (2016). Idealized climate change simulations with a high-resolution physical model: HadGEM3-GC2. J. Adv. Model. Earth Sys.8:813-830.

Sexton DMH, Harris GR, Murphy JM (2010). UKCP09: Spatially coherent projections. UKCP09 additional product. Available from http://ukclimateprojections.metoffice.gov.uk/22530.

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

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

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

Sexton D, Richardson K, Harris G, Karmalkar A, Murphy J, Brown S, Tinker J (2016). Assessment of the UKCP09 Probabilistic Land Scenarios, Including Comparison Against IPCC CMIP5 Multi-Model Simulations. Hadley Centre Technical Note No. 99, Met Office, Exeter, UK. http://www.metoffice.gov.uk/learning/library/publications/science/climate-science-technical-notes

Sexton DMH, Karmalkar AV, Murphy JM, Williams KD, Boutle IA, Morcrette CJ, Stirling AJ, Vosper SB (2018a). Efficient design to identify plausible and diverse variants of a model for climate projections. Part I: establishing the use of 5-day forecasts to filter out poorly performing ensemble members. Clim. Dyn., submitted.

Sexton DMH, McSweeney CF, Rostron JW, Yamazaki K, Booth BBB, Johnson J, Murphy JM, Regayre L (2018b). A perturbed parameter ensemble of HadGEM3-GC3.05 coupled model projections: Part 1: Selecting the parameter combinations. In preparation.

Shindell DT (2014). Inhomogeneous forcing and transient climate sensitivity. Nat. Clim. Change 4:274-277.

Shiogama H, Stone DA, Nagashim T, Nozawa T, Emori S (2013). On the linear additivity of climate forcing-response relationships at global and continental scales. Int. J. Climatol. 33:2542–2550.

Smeed DA, Josey SA, Beaulieu C, Johns WE, Moat BI, Frajka-Williams E, Rayner D, Meinen CS, Baringer MO, Bryden HL, McCarthy GD (2018). The North Atlantic Ocean is in a state of reduced overturning. Geophys. Res. Lett. 45:1527–1533. https://doi.org/10.1002/2017GL076350

Sørland SL, Schär C, Lüthi D, Kjellström E (2018). Bias patterns and climate change signals in GCM-RCM model chains. Environ. Res. Lett. 13:074017.

Stevens B, Fiedler S, Kinne S, Peters K, Rast S, Müsse J, Smith SJ, Mauritsen T (2017). MACv2-SP: a parameterization of anthropogenic aerosol optical properties and an associated Twomey effect for use in CMIP6. Geosci. Model Dev. 10:433-452.

Stocker T, Qin D, Plattner G-K, Alexander L, Allen S, Bindoff N, Bréon F-M, Church J, Cubasch U, Emori S, Forster P, Friedlingstein P, Gillett N, Gregory J, Hartmann D, Jansen E, Kirtman B, Knutti R, Krishna Kumar K, Lemke P, Marotzke J, Masson-Delmotte V, Meehl G, Mokhov I, Piao S, Ramaswamy V, Randall D, Rhein M, Rojas M, Sabine C, Shindell D, Talley L, Vaughan D, Xie S-P (2013). Technical summary. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment report of the Intergovernmental Panel on Climate Change (Stocker T, Qin D, Plattner G-K, Tignor M, Allen S, Boschung J, Nauels A, Xia Y, Bex V, Midgley, P eds.). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 33–115, doi: 10.1017/ CBO9781107415324.005.

Sun Y, Solomon S, Dai A, Portmann RW (2006). How often does it rain? J. Clim. 19:916–934.

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

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

Tebaldi C, Arblaster JM (2014). Pattern scaling: Its strengths and limitations, and an update on the latest model simulations. Clim. Change 122:459–471.

Titchner HA, Rayner NA (2014). The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations. J. Geophys. Res. Atmos. 119: 2864–2889. doi:10.1002/2013JD020316.

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

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

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

Tucker S, Kendon E, Buonomo E, Jones R, Murphy J and co-authors (2018). The UKCP18 12km RCM: PPE ensemble. Description and evaluation. In preparation.

Uppala SM, and co-authors (2005). The ERA-40 re-analysis. Q. J. R. Meteorol. Soc. 131:2961-3012. doi:10.1256/qj.04.176.

van Genuchten MT (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci. Soc. Am. J. 44: 892–898, https://doi.org/10.2136/sssaj1980.03615995004400050002x

Vial J, Dufresne J-L, Bony S (2013). On the interpretation of inter-model spread in CMIP5 climate sensitivity estimates. Clim. Dyn. 41:3339-3362.

Walters DN, Williams KD, Boutle IA, Bushell AC, Edwards JM, Field PR, Lock AP, Morcrette CJ, Stratton RA, Wilkinson JM, Willett MR, Bellouin N, Bodas-Salcedo A, Brooks ME, Copsey D, Earnshaw PD, Hardiman SC, Harris CM, Levine RC, MacLachlan C, Manners JC, Martin GM, Milton SF, Palmer MD, Roberts MJ, Rodríguez JM, Tennant WJ, Vidale PL (2014). The Met Office Unified Model Global Atmosphere 4.0 and JULES Global Land 4.0 configurations. Geosci. Model Dev. 7:361–386. doi:10.5194/gmd-7-361-2014.

Walters DN, Baran A, Boutle I, Brooks ME, Earnshaw P, Edwards J, Furtado K, Hill P, Lock A, Manners J, Morcrette C, Mulcahy J, Sanchez C, Smith C, Stratton R, Tennant W, Tomassini L, van Weverberg K, Vosper S, Willet M, Browse J, Bushell A, Dalvi M, Essery R, Gedney N, Hardiman S, Johnson B, Johnson C, Jones A, Mann G, Milton S, Rumbold H, Sellar A, Ujiie M, Whiltall M, Williams KD, Zerroukat M (2017). The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations, Geosci. Model Dev., doi:10.5194/gmd-2017-291.

Wan H, Rasch PJ, Zhang K, Qian Y, Yan H, Zhao C (2014). Short ensembles: an efficient method for discerning climate-relevant sensitivities in atmospheric general circulation models. Geosci. Model Dev. 7:1961–1977. doi:10.5194/gmd-7-1961-2014.

Webb MJ, Senior CA, Sexton DMH, Ingram WJ, Williams KD, Ringer MA, McAvaney BJ, Colman R, Soden BJ, Gudgel R, Knutson T, Emori S, Ogura T, Tsushima Y, Andronova N, Li B, Musat I, Bony S, Taylor K (2006). On the contribution of local feedback mechanisms to the range of climate sensitivity in two GCM ensembles. Clim. Dyn 27:17-38.

Webb MJ, Lock AP (2013). Coupling between subtropical cloud feedback and the local hydrological cycle in a climate model. Clim. Dyn. 41: 1923-1939.

Wielicki B, Raj M, Sau S (1996). Clouds and the Earth's Radiant Energy System (CERES): an earth observing system experiment. Bull. Amer. Met. Soc. 77:853–868.

Williams KD, Bodas-Salcedo A, Déqué M, Fermepin S, Medeiros B, Watanabe M, Jakob C, Klein SA, Senior CA, Williamson DL (2013). The Transpose-AMIP II experiment and its application to the understanding of Southern Ocean cloud biases in climate models. J. Clim. 26:3258-3274.

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

Wilson DR, Bushell AC, Kerr-Munslow AM, Price JD, Morcrette CJ (2008). PC2: A prognostic cloud fraction and condensation scheme. I: Scheme description. Q. J. R. Meteorol. Soc. 134:2093-2107, doi:10.1002/qj.333.

Woollings T, Hannachi A, Hoskins B (2010). Variability of the North Atlantic eddy-driven jet stream. Q. J. R. Meteorol. Soc. 136: 856–868.

Wood N, Staniforth A, White A, Allen T, Diamantakis M, Gross M, Melvin T, Smith C, Vosper S, Zerroukat M, Thuburn J (2014). An inherently mass-conserving semi-implicit semi-Lagrangian discretisation of the deep-atmosphere global nonhydrostatic equations. Q. J. R. Meteorol. Soc. 140:1505–1520, doi:10.1002/qj.2235.

Wylie DP, Menzel WP (1999). Eight years of high cloud statistics using HIRS. J. Clim. 12:170–184.

Xie P, Arkin PA (1996). Analyses of global monthly precipitation using gauge observations, satellite estimates and numerical model predictions. J. Clim. 9:840–858.

Yang Y, Zhu J (2011). Equilibrium thermal response timescale of global oceans. Geophys. Res. Lett. 38: L14711.

Yamazaki K, Sexton DMH, Rostron JW, McSweeney CF, Harris GR, Murphy JM (2018). A perturbed parameter ensemble of HadGEM3-GC3.05 coupled model projections: Part 1: Global performance and future changes. In preparation.

Yokohata T, Webb MJ, Collins M, Williams KD, Yoshimori M, Hargreaves JC, Annan JD (2010). Structural similarities and differences in climate responses to CO_2 increases between two perturbed physics ensembles. J. Clim. 23: 1392-1410.

Zappa G, Shaffrey LC, Hodges KI, Sansom PG, Stephenson DB (2013). A multimodel assessment of future projections of North Atlantic and European extratropical cyclones in the CMIP5 climate models. J. Clim. 26:5846–5862.

Zappa G, Shepherd TG (2017). Storylines of atmospheric circulation change for European regional climate impact assessment. J. Clim. 30:6561-6577.

Appendix A (Strand 1)

Representation of uncertainties on 1-20 year time scales in the probabilistic projections

Modification to timescaling in volcano years

Following Sexton and Harris (2015), the timescaling method is designed to exclude the near-term historical response to major volcanic eruptions from the scaled regional predictions derived from NTR. This is because volcanic forcing may give rise to different characteristic response patterns than those driven by changes in greenhouse gas concentrations. This is achieved by smoothing SCM predictions of GMST(t) from one year before till four years after each eruption. Since calibration of the timescaling bias term involves comparing SCM outputs against the response of ESPPE members (which simulate the combined response to anthropogenic and natural forcing agents), the volcanic response appears as a contribution to the bias term. If the response is found to be statistically significant for the variable of interest, low-pass filtering of the calibrated bias term (see section 2.2) is withheld in volcano years, to avoid attenuating the response. In practice the volcanic response is found to be significant for GMST itself, but not for the UK_GCM variables.

Sampling seasonal and annual residuals

The timescaling and downscaling calculations described in section 2.2 both require calibration and subsequent sampling of scaling residuals for individual years, which represent the combined effects of uncertainties due to interannual variability and scaling assumptions. The method, AUTOVAR, is updated from that used by Sexton and Harris (2015), and is described below.

a. **Calibration**

- This involves comparing time series of scaled predictions against outcomes simulated by 57 ESPPE members (in the case of timescaling) or 11 HadRM3-PPE members (downscaling). In the case of timescaling (Stage 2 in section 2.2), residuals are calculated from 1861-2100 for annual mean GMST, CO_2 and OHC, and the UK_GCM variables. In the downscaling case (Stage 3 in section 2.2), the calibration uses residuals from 1961-2099 for climate variables defined on the four sets of target regions for Strand 1 (Fung et al., 2018).

- The methodology is the same in both cases. The residuals for each variable in each year are first standardised, by dividing by their standard deviation (calculated across the available ensemble members). Note that the standard deviations are calculated separately for each year, and typically vary significantly with time (e.g. Sexton and Harris 2015, Supplementary Figs. 1 and 2). We return to this point below.

- For each set of target regions, predictions are made for monthly, seasonal and annual means for each of the 12 climate variables. UK_GCM comprises five HadCM3 grid points, while the 25km grid and regional (country, river-basin and administrative) definitions (Fung et al., 2018) contain 437 and 43 values respectively. This gives a total of 1020 variables for UK_GCM, and a combined total of 97920 variables for the 25km grid and regional data. The datasets of standardised residuals therefore consist of 57 members x 240 years x 1020 variables for timescaling, and 11 members x 139 years x 97920 variables for downscaling.

- For each ensemble member, the set of years and variables is concatenated into a long vector, and a Singular Value Decomposition is then performed on the resulting 2-dimensional data matrix. This yields two sets of orthonormal eigenvectors which, when scaled by the corresponding eigenvalues, describe the characteristic inter-variable and temporal relationships found in the relevant array of residuals (hereafter VAR), and the projections of these onto each ensemble member.

b. Generating sampled outcomes

- When applying timescaling or downscaling during the creation of a realisation of past and future climate, uncertainties relating to 1-20 year time scales are added by sampling the calibrated residuals. The first step is to generate a multi-variable, multi-year sample of standardised residuals. This is done by multiplying VAR by a random number sampled from a normal distribution of zero mean and unit variance.

- For each variable and year in the realisation, the standardised residual is then rescaled by the relevant residual standard deviation. The calibrated standard deviations are low-pass filtered prior to rescaling. This removes high frequency noise due to the limited set of ensemble simulations available for calibration, while retaining any longer term trends (Sexton and Harris, 2015).

- This procedure generates realisations consistent with the time-varying spread of the calibrated residual distributions, as well as their temporal and inter-variable covariances. These properties mainly reflect the characteristics of simulated climate variability in either the ESPPE or HadRM3-PPE, as these are the dominant contributor to the residuals (e.g. Fig. 2.5).

Modifying AUTOVAR distributions to correct for non-gaussian residuals

The distributions of AUTOVAR residuals are assumed Gaussian. However, for some variables corrections are needed to remove errors arising from non-Gaussian characteristics. An example is precipitation, for which changes are provided as percentage anomalies. For the 25km grid box containing Edinburgh, Figure A.1 (green curve, left panel) shows a distribution of seasonal precipitation residuals for summer, for the period 2080-2099. This was obtained by predicting the response of HadRM3-PPE members from their driving AO-PPE-A global simulations (Collins et al., 2011), using the relevant downscaling relationship. This distribution, which largely reflects internal variability in the HadRM3 data, has a positive skew of 0.60. The red curve (left panel) shows the distribution of seasonal percentage anomalies in precipitation in a sample of 3000 downscaled realisations obtained by applying the calibrated downscaling relationships and resampling residuals as described in (b) above. Use of the Gaussian assumption results in significant differences from the green curve.
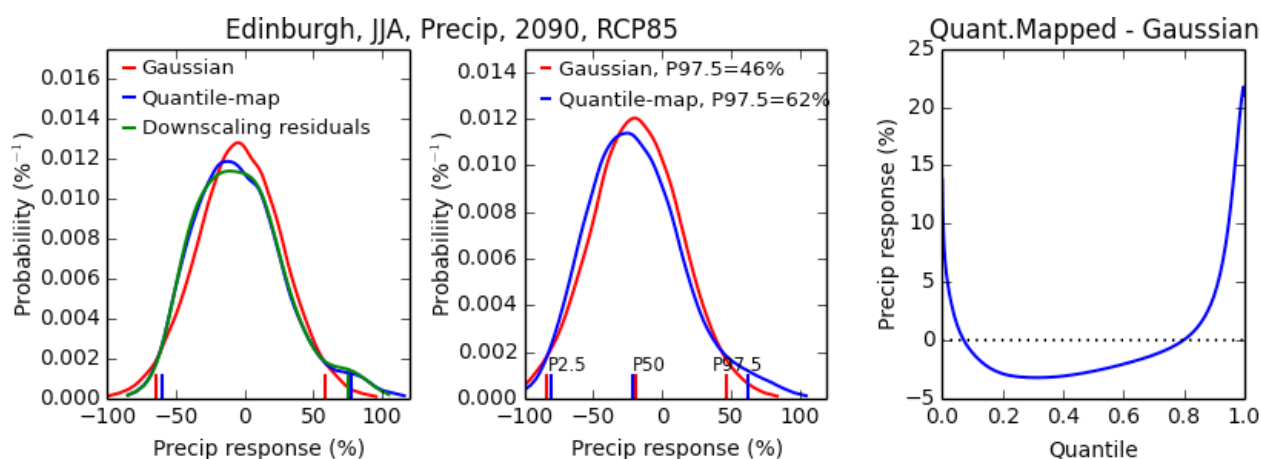

Figure A.1. Left panel shows relative probability of Strand 1 downscaling residuals obtained during calibration (green curve), for seasonal anomalies (%) in summer precipitation during 2080-2099. The example shown is for the box on the UKCP18 25km grid (see Fung et al., 2018) containing Edinburgh. The red curve shows the distribution of residuals for a pooled sample from 2080-99, obtained during creation of a set of 3000 realisations in which the green distribution is resampled by the AUTOVAR procedure of Appendix A, using its standard Gaussian assumption. The blue curve shows residuals obtained when Gaussian AUTOVAR quantiles are remapped to match those of the original distribution of downscaling residuals, using a sliding 21-year window to account for time-dependencies (not shown) in the adjustments. Vertical lines show the 2.5[th] and 97.5[th] quantiles from each distribution, noting that the green lines are not visible because they coincide with the blue lines. The middle panel shows probabilistic projections of precipitation changes (%) for 2080-2099 relative to 1981-2000 under the RCP8.5 scenario, generated using AUTOVAR with Gaussian (red) and quantile-mapped resampling (blue) of residuals. The right panel shows the quantile-dependent differences between the blue and red curves of the middle panel.

We address this using a quantile mapping approach. For a given target year, we create a distribution of 3000 sampled realisations, consistent with the Gaussian assumption. These are then adjusted by mapping them from quantiles of the Gaussian cumulative distribution function onto quantiles of the original, positively skewed distribution of percentage precipitation residuals, for a window (of length 21 years) centred on the target year. The new set of 3000 realisations (Fig. A.1, blue curve, left panel) matches the original distribution of residuals (green curve) much more closely. Projected anomalies are generated by adding the sampled residuals to the relevant downscaled climate changes, relative to the 1981-2000 baseline. The results show a substantial increase in the upper tail of projected changes, compared to that generated in the absence of quantile corrections (Fig. A.1, middle panel, blue cf red curves). The differences in precipitation response can exceed 15% for the upper quantiles (Fig. A.1, right panel), demonstrating that application of quantile mapping reproduces better the wet extremes in summer precipitation found in the HadRM3 simulations.

Appendix B (Strand 1)

Observables used as constraints in the probabilistic projections

Table B.1 Data sources for seasonal climate variables used in CLIM, the observational constraint arising from recent climatological averages that is applied in Strand 1.

Mean climate constraint variables		
Dataset	**Variable(s)**	**Reference**
HadISST1	SST (ts)	Rayner et al. (2003)
HadCRUT3	Screen temperature (tas)	Brohan et al. (2006)
CMAP	Precipitation (pr)	Xie and Arkin (1996)
GPCP	Precipitation (pr)	Adler et al. (2003)
CERES	TOA outgoing shortwave flux (rsut), TOA outgoing longwave flux (rlut), TOA SW cloud radiative effect (cresw) TOA LW cloud radiative effect (crelw)	Wielicki et al. (1996)
ISCCP FD	TOA outgoing shortwave flux (rsut) TOA outgoing longwave flux (rlut) TOA SW cloud radiative effect (cresw) TOA LW cloud radiative effect (crelw)	Rossow and Zhang (1995)
ERBE	TOA outgoing shortwave flux (rsut) TOA outgoing longwave flux (rlut) TOA SW cloud radiative effect (cresw) TOA LW cloud radiative effect (crelw)	Harrison et al. (1990)
HadSLP2	Sea-level pressure (psl)	Allan and Ansell (2006)
ERA40	Sea-level pressure (psl) Relative humidity (hur)	Uppala et al. (2005)
AIRS	Relative humidity (hur)	Aumann et al. (2003)
ISCCP D2	Total cloud (clt)	Rossow and Zhang (1995)
HIRS	Total cloud (clt)	Wylie and Menzel (1999)
da Silva	Surface sensible heat flux (hfss) Surface latent heat flux (hfls)	da Silva et al. (1994)
SOC	Surface sensible heat flux (hfss) Surface latent heat flux (hfls)	Josey et al. (1996)
New SOC	Surface sensible heat flux (hfss) Surface latent heat flux (hfls)	Berry and Kent (2009)

Table B.2. Data sources, variable definitions and time periods used in the historical trend constraint variables applied in Strand 1.

Historical trend constraint variables			
Dataset	Variables	Trend period(s)	References
HadCRUT4	1. Global MST (Mean Surface Temperature). 2. Land MST minus Ocean MST. 3. Northern Hemisphere minus Southern Hemisphere MST. 4. Meridional Temperature Gradient: 52.5°N:67.5°N minus 22.5°N:37.5°N, zonal means.	1958:1987 average minus 1910:1939 average. 1988:2017 minus 1958:1987.	Morice et al. (2012) Braganza et al. (2003).
CO_2: RCP Hist, and ESRL	Global mean CO_2 concentration	2005:2014 minus 1960:1969.	Meinshausen et al. (2011). Dlugokencky and Tans. Booth et al. (2017) Ballantyne et al. (2012) (uncertainty estimates 1958 to 1980). Etheridge et al. (1996) (uncertainty estimates prior to 1958)
EN4	Global ocean heat content to 700m (OHC)	1993:2012 minus 1955:1974	Good et al. (2013)
Domingues	OHC	1993:2012 minus 1955:1974	Domingues et al. (2008)
Levitus	OHC	1993:2012 minus 1955:1974	Levitus et al (2012).
Ishii	OHC	1993:2012 minus 1955:1974	Ishii et al. (2003).

Appendix C (Strand 1)

Credibility checks for the probabilistic projections

The Bayesian methodology used to construct the probabilistic projections requires specification of several sources of statistical uncertainty that broaden the posterior distributions. These include uncertainty in emulation of the equilibrium response to doubled CO_2, discrepancy associated with structural uncertainty (obtained by finding best-analogues of CMIP5-ESM responses in the parameter space of HadCM3), and uncertainty associated with timescaling and downscaling (see section 2.2 for details). It is therefore important to compare the projections with the input climate model data, to test whether any failure in the validity of the methodological assumptions might leads to excessive statistical uncertainty, and hence create inconsistency between the projections and the climate model inputs implying a lack of credibility in the probability distributions. We assess this question by comparing our posterior pdfs with climate model data from the 12 CMIP5-ESM, 57 ESPPE and 280 SLAB simulations that provide inputs to the projections.

For eight different climate variables, Figures C.1 and C.2 compare the pdfs (derived from a sample of 3000 realisations as described in section 2.2) with equivalent changes from the 349 model simulations mentioned above. The figures show 20 year mean changes in response to emissions-driven RCP8.5 forcing for the period 2080-2099 relative to 1981-2000. In order to facilitate the comparison, the 280 SLAB responses are converted to a synthetic ensemble of time-dependent simulations. First, the normalised equilibrium responses to doubled CO_2 (NER, see section 2.2, Stage 1a) are adjusted to allow for the effects of coupled earth system processes omitted from the SLAB ensemble (NTR, section 2.2, Stage 1b). Secondly, NTR is scaled using a projected time series of the transient response of global mean temperature. This is obtained using the simple climate model (SCM) of Stage 2a, by prescribing the physical climate feedback values simulated by the relevant SLAB member, and sampling other required earth system parameters randomly from the 57 sets of calibrated values used by the SCM to replicate ESPPE results.

As a measure of credibility, we estimate the fraction of realisations in the pdf (represented by blue area F in Figs. C.1 and C.2) that lie beyond the most extreme response across all three GCM ensembles. For each variable, we use the larger of the two tail probabilities from either end of the pdf. A high value for F indicates possible statistical inflation, potentially unjustified by the GCM data. Note, however, that we do expect the statistical processing to generate some outcomes beyond the modelled results. This is expected to happen, for example, through emulation of a broader set of possible responses through fuller sampling of parameter space. Finding modest cumulative probabilities for outcomes outside the model range does not, therefore, indicate a lack of credibility. For a given variable, finding larger probabilities (e.g. F>15%) is likely to indicate a general credibility issue if it occurs regularly (across different months, seasons or regions), though this is not necessarily the case if the occurrence is limited to a small minority of cases.
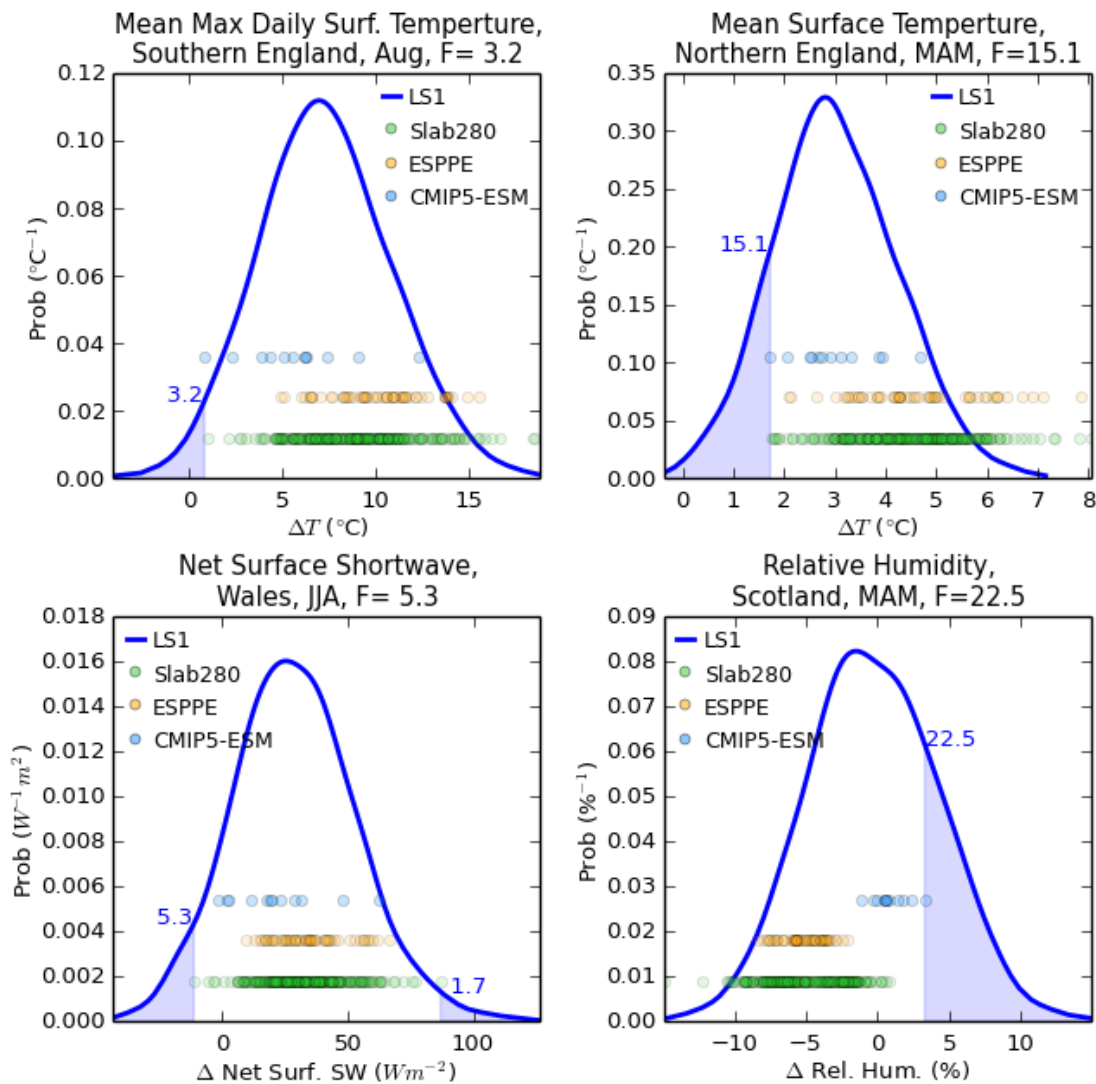
Figure C.1 Blue curves show posterior probability distributions from Strand 1 for 20-year mean changes for 2080-2099 relative to 1981-2000, under the RCP8.5 scenario. These are compared with the responses of the 57 ESPPE, 12 CMIP5-ESM and 280 SLAB ensemble members. Variables are daily maximum surface air temperature (°C) for Southern England (top left), daily average surface air temperature (°C) for Northern England in spring (top right), net surface shortwave radiation (Wm^{-2}) for Wales in summer in August (bottom left), and relative humidity (%) for Scotland in spring (bottom right). In order to enable the comparison, the SLAB equilibrium climate change responses are converted to a synthetic ensemble of time-dependent simulations, as described in Appendix C. These comparisons are made at GCM scales, excluding the final downscaling step. The shaded area represents the cumulative probability for outcomes beyond the most extreme GCM response. A high value for this 'tail probability' (denoted by F), indicates that the extremes of the probability distribution may be influenced by an unrealistic level of statistical inflation. Such cases are flagged as demonstrating a potential lack of credibility.

For example, in Fig C.1 (top left panel) we consider monthly mean daily maximum temperature change for August, for the Southern England grid-point. The large spread in the pdf (5-95% range of about 1 to 13°C) is representative of the underlying GCM responses. Just 3% of the realisations are cooler than the coolest of the GCM responses, and the pdf appears consistent with the input data, and credible.

Fig C.1 (top right panel) makes a similar comparison for daily average surface temperature change in spring, for Northern England. In this case, 15% of the realisations comprising the probabilistic projections are cooler than the coolest GCM response. This is mainly due to the CMIP5-ESM response being, on average, cooler than the response in the SLAB ensemble. The resulting discrepancy distribution of best-analogue differences therefore imparts an adjustment toward a cooler response, and inflates the posterior distribution somewhat. For mean surface temperature change in general, tail-probabilities of this magnitude are typically limited to the winter and spring seasons. Similar discrepancy adjustments were obtained in the UKCP09 projections (Murphy et al., 2009), and the same underlying reason applies: a bias towards high climatological regional snow cover in the HadCM3 simulations leads to larger relative warming under climate forcing, compared to the corresponding multi-model ensemble simulations. The discrepancy adjustment is therefore acting as designed, although the GCM results suggest that less confidence should be placed in relative or cumulative probabilities associated with the cool tail of the probability distribution, compared with warmer outcomes. At the other end of the distribution a fraction of the GCM responses (those above ~6°C) lie outside the pdf. This is not surprising, as such high-end responses lie within the prior distribution for this variable (not shown), but are down-weighted when observational constraints are applied to form the posterior distribution.

Fig. C.1 (bottom left panel) compares the posterior pdf with the GCM responses for the change in net surface shortwave radiation during summer for Wales. There is a consistent signal across all three ensembles for an increase, and this is reflected in the pdf. The probabilistic projections suggest a 5% chance of a lower outcome than the smallest predicted change in the GCM ensembles.

Finally, Fig C.1 (bottom right panel) considers percentage change in spring relative humidity (RH) in Scotland. For this variable, almost 23% of the Strand 1 realisations project a higher outcome than the most extreme GCM response. In this case, there is a strong signal in both the ESPPE and SLAB ensembles for a reduction in RH, in contrast to the CMIP5-ESM responses which on average predict an increase. Indeed, the distributions of ESPPE and CMIP5-ESM changes do not overlap at all. The resulting discrepancy distribution provides a substantial mean adjustment towards a smaller reduction in RH, and contributes to an inflation in spread of the posterior distribution. This violates a basic assumption of the method, namely that the HadCM3 parameter space can provide a range of outcomes that is sufficiently consistent with alternative multi-model results to prevent the discrepancy term from playing too dominant a role in the final pdf. In this example, the method produces a pdf that is not a credible representation of the knowledge incorporated in the climate model simulations.

Fig C.1 (bottom right panel) illustrated a single instance of limited credibility for RH in just one region and season, so we need to look more widely to assess whether multiple issues occur. Identical analysis has been performed for all five HadCM3 UK grid-points, and for 17 monthly, seasonal or annual periods. The results of this are summarized in the second row of Table C.1, which gives the fraction of these 85 variables whose tail-probability F is greater than a particular threshold. For example, 31% of the 85 RH pdfs predict F >15%. The high tail-probability in Fig C.1 (bottom right) is more widely obtained for a large proportion of the RH variables, for the same reasons: a mismatch between HadCM3 and CMIP5 leading to a high structural error component. For this reason, probabilistic projections for RH are not found to be credible, and are not provided in UKCP18.
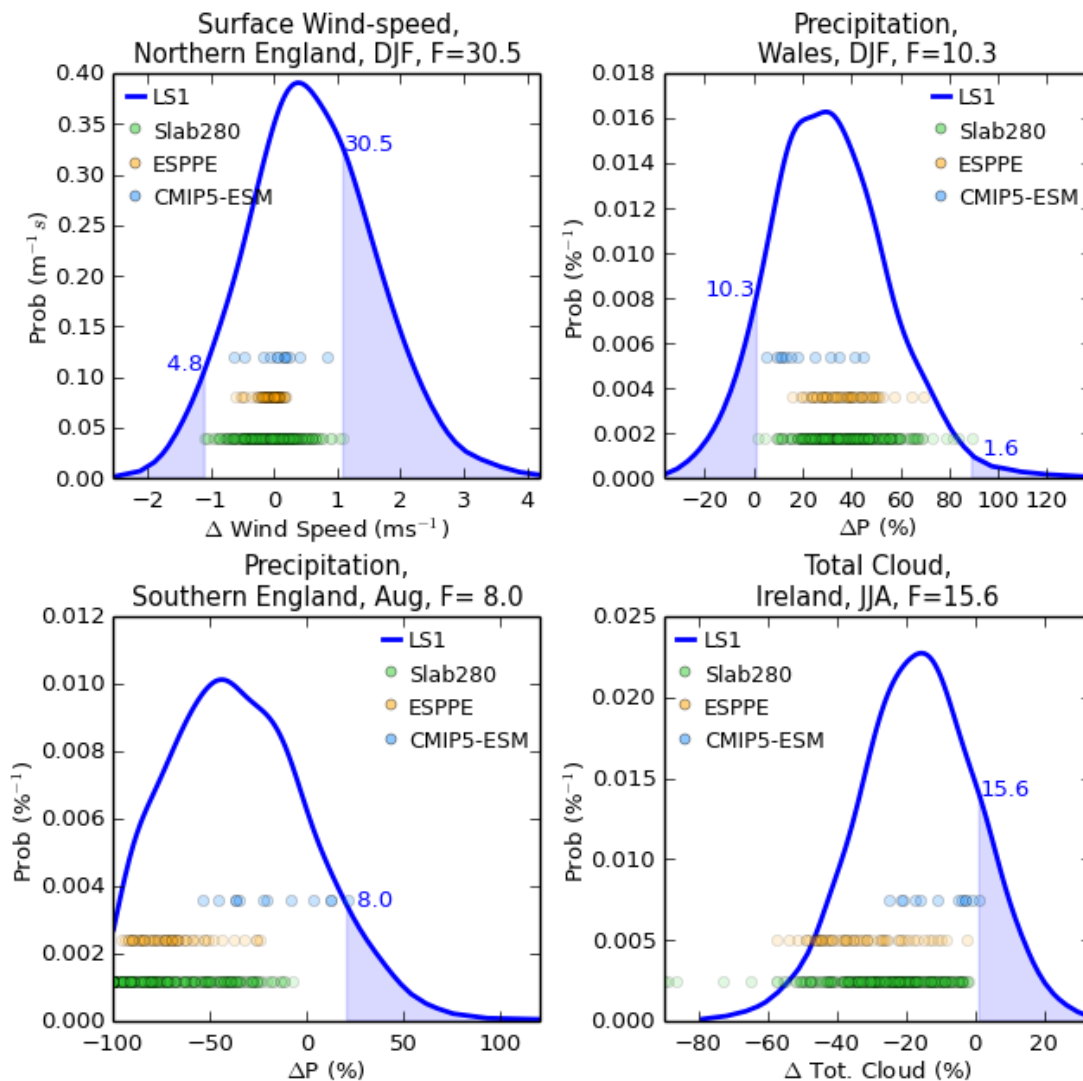
Figure C.2 As Fig. C.1, but for four different variables, consisting of near surface wind speed (ms⁻¹) for Northern England in winter (top left), precipitation (%) for Wales in winter (top right), precipitation (%) for Southern England in August (bottom right), and total cloud cover (%) for Ireland in summer (bottom right).

Figure C.2 continues the analysis in Fig C.1, for four additional variables. Firstly we look at change in winter wind-speed for Northern England (Fig. C.2, top left panel). For this variable, the probabilistic projections suggest a 31% chance of an increase greater than the largest predicted in the GCM simulations. In this case, there are no large differences in the responses for the ESPPE, CMIP5-ESM and SLAB ensembles, so unlike RH, it is not discrepancy driving the inflation of uncertainty. For wind-speed, it is the timescaling errors that are large, leading to large spread and lack of credibility. The reason is that timescaling can sometimes fail for wind-speed, since the equilibrium response is often not representative of the response when the same perturbed atmosphere variant is coupled with a dynamic ocean and used to simulate transient climate change. This violates another basic assumption of the methodology. The issue is illustrated in Fig. C.3 for four typical ESPPE simulations. In the first two panels the equilibrium response is representative of the response in the coupled simulations, and scaling performs well. In the second two panels, the substantial scaled equilibrium responses are not reproduced in the coupled response, leading to a large error. Table C.1 confirms that this problem occurs in a large fraction of cases: 36% of the wind-speed pdfs give a cumulative probability of more than 15% of an outcome beyond the most extreme GCM response. Due to the poor performance of time scaling, and high tail probabilities, pdfs for wind-speed are not provided in UKCP18.
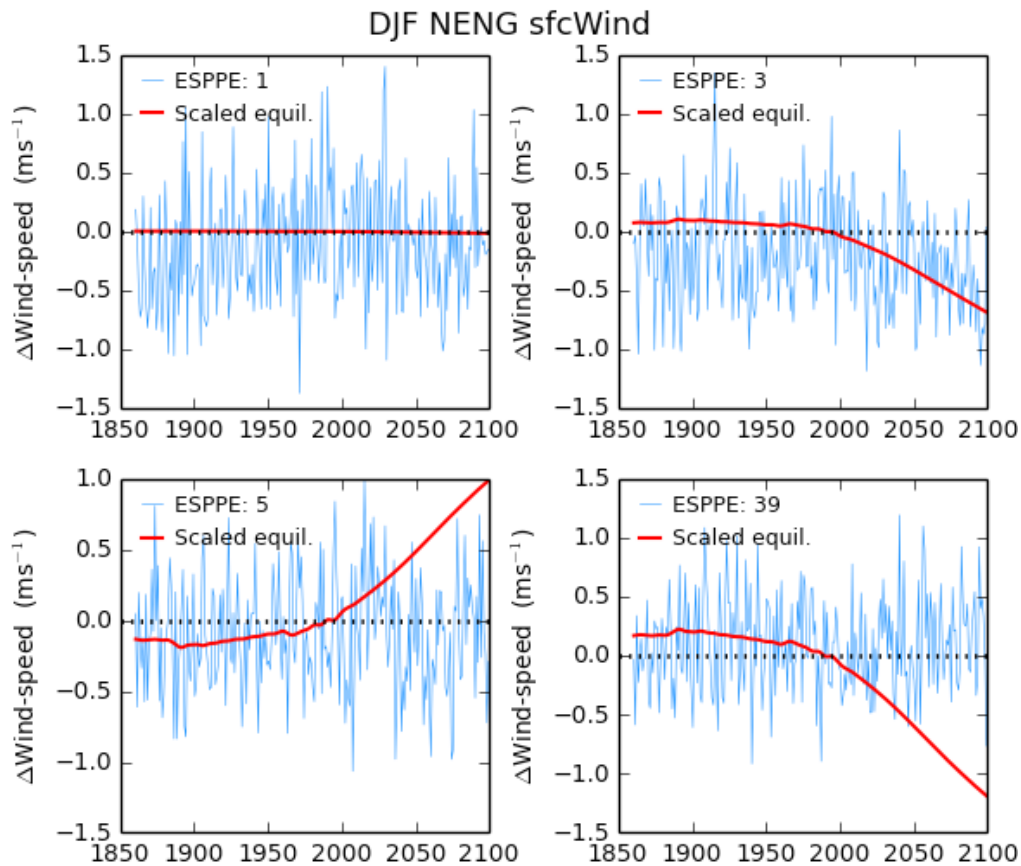
Figure C.3 Surface wind-speed response (ms^{-1}) to RCP8.5 forcing (relative to 1981-2000) in winter for Northern England, from four typical ESPPE simulations (blue). These are compared to scaled time-dependent changes (red), derived from the corresponding simulation of the equilibrium response to doubled CO_2 in the SLAB ensemble.

Figure C.2 (top right) looks at the credibility of the pdf for winter precipitation response for Wales, a case for which a substantial increase is predicted. Amongst the Strand 1 realisations, 10% are lower than the lowest GCM response. Some of this can be attributed to the discrepancy between the CMIP5 and SLAB response, which gives a slight adjustment toward a smaller increase in precipitation. Figure C.2 (bottom left) looks at the pdf for a precipitation variable with strong predicted drying: Southern England in August. Almost all members in the ESPPE and SLAB ensembles predict strong drying, but the picture is more mixed for the CMIP5-ESM ensemble. A majority of these simulations predict drying, although typically more modest than in the PPE, while a minority of simulations predict a small increase. The discrepancy distribution, which results from finding best HadCM3 analogues for these CMIP5 responses, adjusts the posterior towards a less dry response. It also inflates the spread, leading to a cumulative probability of 8% for a larger change than the response of the wettest CMIP5 simulation. The precipitation pdfs are judged to be credible, on the basis that the degree of overlap between the SLAB and CMIP-ESM results is large enough to give credence to the discrepancy estimate. However, there is less confidence in the upper tail of wet responses in the pdf, compared with the bulk of the distribution.

Finally, Fig. C.2 (bottom right) considers percentage change in summer total cloud amount for the Ireland grid-point. As in the summer precipitation case, the HadCM3 simulations predict larger reductions than the CMIP5-ESM simulations, resulting in adjustment of the posterior toward a smaller reduction. For this example, 16% of the Strand 1 realisations give a change greater than the most extreme GCM response.

Table C.1 summarizes the tail-probabilities for different thresholds, for all twelve of the core UKCP18 climate variables identified in Fung et al. (2018). The fractions of occurrence of high tail probabilities are a useful summary measure to indicate potential credibility issues. However, they are not the sole criterion. For relative humidity and wind-speed we find widespread evidence of high structural error or poor timescaling performance (similar to the specific examples discussed above). It is this, in conjunction with the high frequency of large F values, which confirms the methodology cannot produce credible pdfs for these variables. In Table C.1 there is a clear gap between the tail-probability statistics for these two variables, and the others. For the other variables we assess that the methodological assumptions are generally satisfied, hence we provide the probabilistic projections.

Table C.1 Fraction of regional, monthly and seasonal 20-year mean probabilistic projections of changes for 2080-2099 relative to 1981-2000 (under the RCP8.5 scenario) for which the 'tail-probability' (F, see Fig. C.1) is greater than a specified threshold. The rows representing the 12 candidate climate variables are ordered by the fraction of exceedances for the 15% threshold. For example, 9% of the 85 surface air temperature projections (comprised of five UK regions times 17 monthly, seasonal, or annual variables) show F >15%, indicating a cumulative probability of more than 0.15 of an outcome beyond the most extreme response across the ESPPE, SLAB and CMIP5-ESM simulations. A relatively large fraction of the relative humidity and wind-speed variables have tail-probabilities greater than 15 or 20%, indicating reduced credibility. Therefore, UKCP18 does not provide probabilistic projections for these variables, shown in red.

Variable	Acronym	F > 10%	F > 15%	F > 20%
Wind speed at 10m	sfcWind	0.64	0.36	0.20
Relative humidity	hurs	0.64	0.31	0.07
Cloud area fraction	clt	0.49	0.19	0.01
Pressure at mean sea level	psl	0.33	0.11	0.06
Specific humidity	huss	0.31	0.11	0.05
Diurnal maximum air temperature	tasmax	0.18	0.11	0.01
Net surface short wave flux	rss	0.29	0.09	0.01
Diurnal average air temperature	tas	0.25	0.09	0.05
Net surface long wave flux	rls	0.38	0.08	0.01
Downward surface shortwave flux	rsds	0.29	0.08	0.01
Precipitation rate	pr	0.26	0.07	0.01
Diurnal minimum air temperature	tasmin	0.13	0.06	0.00

Having examined the credibility of the projections at GCM scales, we move on to compare the downscaled projections with the HadRM3-PPE output used to calibrate the downscaling relationships. In this case, our focus is placed specifically on variability in precipitation in the probabilistic projections. This is because precipitation is a bounded variable (negative values are impossible), which creates challenges in representing its variability statistically. The example in Fig. C.4 (left panel) shows selected probability levels (grey plume) for annual precipitation anomalies (in %) for London in August, in response to A1B forcing. Also shown are 50 randomly-selected realisations from the pdfs. These are compared to the responses of the 11 HadRM3-PPE members.

In UKCP18, the statistical calculations are applied to precipitation data expressed as percentage anomalies relative to the baseline period, consistent with the units used in the final projections. Figure C.4 (right panel) makes the same comparison, but this time using the UKCP09 approach. In UKCP09, projections of negative precipitation were avoided by applying a logarithmic transformation prior to all statistical processing in the methodology, followed by application of the inverse transform to recover projections expressed as percentage changes. In Fig C.4 the five highest monthly anomalies in the HadRM3-PPE data are of the order of 300%, whereas the highest projected anomalies are over 1000% when the log transform is used. These are clearly not credible. The higher probability levels of the projected time-evolving distribution are also substantially inflated relative to the underlying HadRM3-PPE data, when the log transform is used.
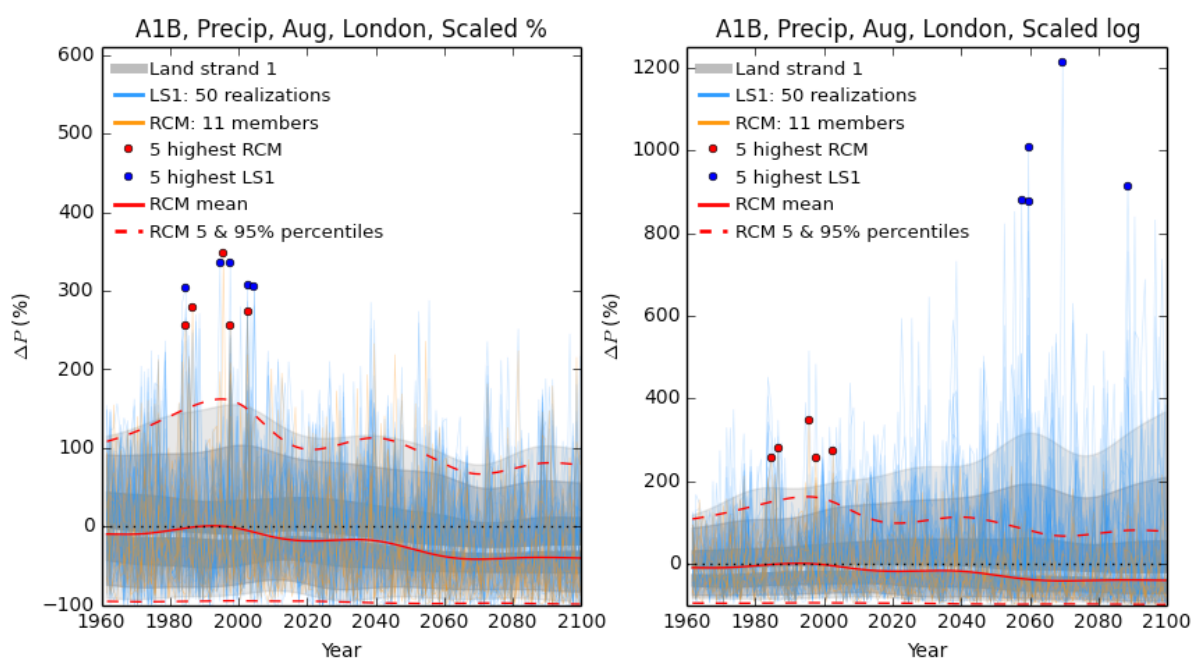
Figure C.4 Annual change in precipitation (%) for London in August for the A1B scenario, relative to 1981-2000. The white line shows the median of the probabilistic projections, and grey shading shows the 5%, 10%, 25%, 75%, 90% and 95% probability levels. Blue lines show 50 randomly selected realisations from the pdfs. These are compared with the response of the 11 HadRM3-PPE simulations for this variable. The solid orange curve shows the median of the 11 simulations, and the dashed lines the 5th and 95th percentiles of their time-evolving frequency distribution of anomalies. The red and blue dots show the five highest instances of wet months from HadRM3-PPE and the probabilistic projections, respectively. The left panel shows the UKCP18 results, in which the statistical calculations in the Bayesian methodology are performed on precipitation data expressed as percentage anomalies. The right panel shows results obtained using the UKCP09 approach, in which a logarithmic transformation is applied before the statistical processing, with the inverse transform applied at the end to obtain projections expressed as percentage changes.

There are several factors that combine to produce these effects. Firstly, we find that it is only at the annual time scale, and in the presence of strong drying that this inflation of the wet tail becomes substantial. Secondly, the method used to sample residual uncertainties in the downscaling (Appendix A), the dominant component of which is due to internal variability, is assumed to be independent of the magnitude of the projected climate change signal. This assumption is necessary because HadRM3-PPE is not large enough to allow the residual uncertainty to be predicted as a function of location in parameter space. In cases of strong summer drying, however, this assumption does not hold well, contributing to inflation of statistical uncertainties at the wet end of the distribution when the log transform is used. The impact of this is increased by a third factor, which is that the discrepancy adjustment for summer precipitation changes invariably gives projections that are on average less dry. Following application of the inverse exponential transform, the net result is to inflate the accumulated statistical uncertainty in the wet tail, leading to unrealistically wet months during summer (e.g. over 1000% of the baseline value) that are not supported by the HadRM3-PPE data. As noted above, this sensitivity to use of the log transform only arises when strong drying is present. For example, there is little difference between projections made with or without a log transform for winter precipitation, when strong drying is not predicted.

Direct scaling of percentage change as in Fig C.4a results in probabilistic projections that are consistent with the HadRM3-PPE simulations, due to elimination of the excessive statistical uncertainty found in the wet tail using the UKCP09 approach. For all variables to which log transforms were applied in UKCP09 (precipitation, total cloud amount and specific humidity, noting that relative humidity was eliminated by the credibility check at GCM scales), the probabilistic projections in UKCP18 are produced by processing the data as percentage change throughout the calculations.

Note, however, that dropping use of the log transform has the disadvantage that precipitation anomalies projected in UKCP18 are not automatically bounded to be above -100%. Therefore, clipping at this value is applied to the 3000 realisations comprising the final projections. This would lead to apparent 'spikes' near -100% in pdfs calculated from the realisations. Therefore, the pdfs are also clipped, as in Fig C.2 (bottom left panel). For the 20-year mean changes shown in this example, fewer than 0.25% of the sampled realisations are clipped.

In addition to clipping variables expressed as percentage changes at -100%, a second level of clipping is applied to all variables, in order to ensure physically credible changes. For each year from 1961 to 2099, the 3000 realisations are clipped at the 1% and 99% probability levels of the distribution. Therefore, distributions calculated from the realisations will present 'spikes' at the lower and upper extremes, in all cases. Similar clipping of realisations (usually at 1% and 99%) was also applied in the UKCP09 projections. We do this to avoid provision of implausible extremes in which we have reduced confidence. In general, users should not place confidence in relative probabilities of specific outcomes in the extreme tails of the distributions, because these are typically where limitations in the statistical assumptions in the method will have their greatest impact. More confidence can be placed on cumulative probabilities for thresholds away from the extreme tails. For example, the 10% to 90% probability range of the cumulative distribution function can normally be taken as trustworthy. More guidance on interpretation of the probabilities is available from the UKCP18 website (see https://ukclimateprojections-ui.metoffice.gov.uk/).

Appendix D (Strand 2)

Supporting information

Table D.1. Parameters in the atmosphere, land and aerosol components of the GC3.05 climate model, perturbed in the Strand 2 GC3.05-PPE simulations. Middle and right columns describe the physical role of each parameter, and the main effects of varying its value.

Parameter	Role	Signature effect
CONVECTION		
Deep entrainment amplitude (ent_fac_dp)	Controls the shape of the mass flux and the sensitivity of deep convection to relative humidity to deep entrainment.	Increased values lead to the reduction of convection depth and to some extent suppression of active precipitating convection.
Mid entrainment amplitude (ent_fac_md)	Controls the shape of the mass flux and the sensitivity of mid-level convection to relative humidity to mid-level entrainment.	Increased values lead to the reduction of convection depth and to some extent suppression of active precipitating convection.
Mixing detrainment (amdet_fac)	Controls the rate of humidification of the atmosphere, and the shape of convective heating profile.	Increases large-scale humidity and temperature profiles.
Coefficient for adaptive detrainment (r_det)	Decrease of mass flux with height under decreasing parcel buoyancy. Tends to oppose this buoyancy reduction and thus raises the termination height of the convection.	Larger rdet gives deeper convection but also changes the height distribution.
Convective core radiative effects (cca_md_knob, cca_dp_knob). These two parameters are set equal to each other in the model, so the same perturbations are applied to both.	Control how much deep and mid-level convective core gets seen by radiation.	Increasing the values will mean convective cores have more of a radiative impact (i.e. more reflection of shortwave (SW) radiation and more longwave (LW) emission from a cold cloud top).
Shallow convective core radiative effects (cca_sh_knob)	Control how much shallow convective core gets seen by radiation.	Increasing the values will mean convective cores have more of a radiative impact (i.e. more reflection of SW and more LW emission from a cold cloud top).
Maximum condensate (mparwtr)	The maximum condensate a convective parcel can hold before it is converted to precipitation.	See left.
Minimum critical cloud condensate (qlmin)	The minimum value of the function that defines the maximum amount of condensate a convective parcel can hold before it is converted to precipitation.	Reducing it cools the troposphere, increasing it warms the troposphere, by decreasing/increasing the amount of high cloud.
GRAVITY WAVE DRAG (GWD)		

Parameter	Role	Signature effect
Critical Froude number (gwd_frc)	Determines the cut-off mountain height and the depth of the blocked flow layer around sub-grid mountains.	Affects tropospheric and stratospheric winds and mean sea-level pressure (MSLP). Increases in gwd_frc will slow low-level winds and increase MSLP.
Flow blocking drag coefficient (fbcd)	Determines the size of the low-level drag associated with flow blocking effects by sub-grid mountains.	Affects tropospheric winds and MSLP.
Inverse critical Froude number for wave saturation (gwd_fsat)	Determines the amplitude at which mountain waves generated by sub-grid orography will break, and exert a drag on the flow. As gwd_fsat is reduced, smaller amplitude waves will break, typically leading to wave breaking (and drag) at lower altitudes.	Affects tropospheric and stratospheric winds and MSLP.
Mountain wave amplitude (gsharp)	Determines the amplitude of the mountain waves generated by sub-grid orography, and hence the size of the orographic gravity wave stress.	Affects tropospheric and stratospheric winds and MSLP. Increases in Gsharp will lead to slower winds in the upper troposphere and above, and changes to temperature biases (through thermal wind balance).
Drag coefficient for turbulent form drag (orog_drag_param)	Determines the size of the form drag exerted on flow by small-scale sub-grid hills.	Affects boundary-layer winds and MSLP. Larger values of the drag coefficient will gives lower boundary layer winds.
Scaling factor applied to sigma, the standard deviation of sub-grid mountain heights (nsigma)	Multiplies sigma(x, y) to determine the local assumed sub-grid orography height which is used in the GWD scheme. This effects the calculation of the Froude number, which then influences the magnitude of the parameterized flow blocking and mountain wave drag.	Larger (smaller) values of nsigma will result in increases (decreases) in the drag. The relative changes in flow blocking and mountain wave drag will be regionally dependent.
BOUNDARY LAYER (BL)		
Flux profile parameter (g0)	Used in the definition of stability functions.	Increasing implies smaller stability function, less BL mixing, e.g. less wind turning, shallower BL.
Critical Richardson number (ricrit = 10.0 / g0)	Value of Richardson number below which air becomes dynamically unstable and turbulent.	Reducing lowers stable BL top, smaller mixing length, less BL mixing.
Cloud-top entrainment rate (a_ent_1)	Parameter used in entrainment rate calculation.	Increasing gives larger entrainment rate at boundary layer top, deeper and warmer mixed layer, potentially quicker break up of cloud.

Parameter	Role	Signature effect
Cloud-top diffusion (g1 = 0.85 * a_ent_1 / 0.23	Parameter in cloud top diffusion coefficient calculation.	Increasing gives larger top-down diffusivity profile, better mixed cloud layer, possibly less decoupling and more stratocumulus (Sc).
Threshold fraction of the cloud layer depth (zhloc_depth_fac)	Fractional height into cloud layer for which Ri-based BL depth can diagnose shear dominated layer.	Higher value leads to more cumulus capped BLs and fewer shear dominated BLs, lower cloud fraction in cold air outbreaks.
Neutral mixing length (par_mezcla)	Mixing length for fluid parcels under neutral stability conditions.	Reducing implies smaller stability function, less BL mixing, e.g. less wind turning, shallower BL.
Minimum value of mixing length (lambda_min = 40 * par_mezcla / 0.15)	Minimum mixing length.	Reducing implies smaller stability function, less BL mixing, e.g. less wind turning, shallower BL.
Decoupling threshold for cloudy BLs (dec_thres cld, dec_thres_cloud2cu = 0.5 * dec_thres_cld)	Decoupling threshold for cloudy BLs.	Larger value makes decoupling less likely, shifts to more well-mixed boundary layers
Mixing factor applied to the in-cloud water content of forced cumulus clouds (forced_cub_fac)	Determines the fraction of the diagnosed adiabatic water content of forced cumulus clouds which is allowed to remain. 0 means no forced cumulus clouds, 1 means maximum possible water content based on an adiabatic parcel ascent, within 0-1 means mixing between clear and cloudy air.	Increasing the value will give more water in shallow convective regions, increased reflected shortwave radiation.
CLOUD AND CLOUD RADIATION		
Cloud erosion rate (dbsdtbs_turb_0)	Determines the rate with which un-resolved sub-grid motions mix clear and cloudy air and hence remove liquid condensate and evaporate liquid cloud fraction.	Modifies the radiative properties of the cloud (especially in regions of liquid only cloud e.g. sub-tropical maritime Sc) and also affects the in-cloud liquid water content and hence how the precipitation formation processes will evolve.
Scaling to make sub-grid cloud condensate variance to cloud cover and convective activity two dimensional. (two_d_fsd_factor)	Makes the cloud water variability around the grid box average a two dimensional relationship, based on a 1-d empirical relationship derived from CloudSat observations (see http://cloudsat.atmos.colostate.edu/)	This changes the cloud-radiation interaction and hence the LW/SW radiative balance. Increasing this parameter from 1.4 to 1.5 reduces the amount of outgoing SW at the top of the atmosphere by around 1.2 Wm^{-2}, and increases the outgoing LW radiation by 0.4 Wm^{-2}.
Decorrelation scale pressure. (dp_corr_strat)	Determines the vertical overlap between clouds in the sub-column in the cloud generator used to calculate the radiative impact of clouds.	Alters the cloud radiative effect. High values mean the cloud is more maximally overlapped and its radiative effect reduced. Impacts are seen most clearly in convective regions.

Parameter	Role	Signature effect
Ice width (ice_width)	Determines the amount of ice water content (as a fraction of qsat-liquid) that corresponds to a factor of two reduction in the width of the vapour distribution in the liquid-free part of the grid box.	Changes the ice water content and ice cloud fraction, hence impacting SW and LW properties and modifying the radiative balance.
CLOUD MICROPHYSICS		
Cloud-rain correlation coefficient (c_r_correl)	Determines the sub-grid correlation between cloud and precipitation. A high value means that regions of high cloud water are correlated with regions of high precipitation, a small or negative number means they are un- or anti-correlated	Increasing the value will result in more warm rain, reducing the water content of stratocumulus clouds and the reflected SW.
Ice fall speed (m_ci)	Scaling factor for the ice fall speed	Increasing fall speed will decrease ice water content.
Precursor coefficient in the mass-diameter relationship for ice (m = ai × Dbi) (ai)	Changing ai has the effect of changing the density of the ice.	Increasing ai will produce a narrower particle size distribution (PSD) and so the mass weighted fall speed will be lower and hence the cloud ice content should increase.
x1r	Controls shape of PSD for raindrops.	Increasing x1r will decrease rain rate
Aspect ratio of ice particles (ar)	Used to calculate the depositional capacitance of ice crystals which affects how efficiently they grow by depleting water vapour. Ice particles are assumed to be spheroidal for the purposes of calculating deposition rates. Capacitance is maximal for spheres (ar=1) and reduces for oblate (ar<1) and prolate (ar>1) spheroids.	Higher capacitance will lead to a lower relative humidity. The ice water contents will only be weakly affected. In mixed-phase regions, the liquid water contents will increase.
Vertical scale in mixed phase turbulent production of supercooled liquid water (mp_dz_scal)	Vertical length scale over which the turbulence acts to produce supercooled water.	Increasing mp_dz_scal will lead to an increase in liquid water path.
AEROSOLS		
Anthropogenic SO$_2$ emission flux (ps_anth_so$_2$_emiss)	Direct scaling of emissions flux.	Increasing this leads to higher aerosol concentrations in source regions.
Dry deposition rate of SO$_2$ (ps_dry_so2_veloc)	Scaling factor for dry deposition rate calculated in the model, which removes SO$_2$ from lowest levels through deposition according to land surface type and prevailing wind speed.	Increasing this will reduce low level SO$_2$ concentrations

Parameter	Role	Signature effect
Scaling of the standard deviation used to define the pdf of updraught velocity (ps_sigma_updraught)	Relates the activation of aerosols to cloud droplets to the standard deviation used to define the pdf of updraught velocity.	Increasing this produces higher cloud droplet number concentration.
Scaling of emission flux from biomass burning (biom_aer_ems_scaling)	Direct scaling of emissions flux.	Increasing this directly affects black carbon/organic carbon aerosol concentrations proportionately.
Scaling of emission flux from sea spray (ps_natl_ss_emiss)	Direct scaling of emissions flux.	Increasing this directly affects hygroscopic aerosol concentrations.
Dimethyl-sulphide emission flux (ps_natl_dms_emiss)	Direct scaling of emissions flux.	DMS is a precursor gas for sulphate production via oxidisation.
Scavenging rate in the coarse and accumulation modes (ps_acc_cor_scav)	Scaling of the scavenging rate calculated in the model.	Increasing this will reduce concentrations of aerosols in coarse and accumulation mode.
pH of cloud drops (ps_cloud-ph)	This controls in-cloud SO_4 production dependent on SO_2 availability.	An increase in cloud pH leads to faster SO_2 oxidation by ozone in cloud water, so more SO_4 production.
LAND SURFACE AND SNOW		
Maximum wind speed used in Coupled Ocean-Atmosphere Response Experiment (COARE) algorithm (u10_max_coare)	This is the highest wind speed used in calculating the Charnock coefficient in the COARE algorithm.	Higher values allow the sea surface to become rougher in strong winds.
Grain size of fresh snow (r0)	The grain size of fresh snow is set to this value, which affects the albedo of snow.	Higher values make the snow less reflective.
Fresh snow density (rho-snow_fresh)	The density of fresh snow.	Lower densities reduce the thermal conductivity of snow, leading to colder surface temperatures.

Parameter	Role	Signature effect
Upper value about 4K above Topt, the optimal temperature for photosynthesis (tupp io)	Topt determines the turn-over point for temperature, above which further increases in temperature will drive a decline in photosynthesis.	In tropical and sub-tropical regions the optimal temperature would be expected to have the biggest impact on plant functioning, with low values for this parameter leading to greater temperature dependence of photosynthesis.
Maximum ratio of internal to external CO$_2$ (f0_io)	Controls the gradient of CO_2 between plant stomata and the ambient air.	See left.
Top leaf Nitrogen concentration in kg N per kg C (nl0-io)	Defines the top leaf ratio of nitrogen to carbon. Plant photosynthesis (Vcmax) is defined in the model to be proportional to the leaf nitrogen concentration.	Higher ratios are associated with higher photosynthesis.
Root depth (rootd_ft_io)	Controls the depth (in model soil levels) to which soil moisture is available.	Larger values equate to greater depths in the soil, and more resilience to short timescale droughts.
Scaling factor for critical and saturation levels for soil moisture towards wilt level (psm)	This pair of parameters control the critical and saturated volumetric soil moisture thresholds. The critical threshold controls the level above which evapotranspiration is no longer soil moisture dependent.	Higher values lead to larger soil moisture regimes where soil moisture limits this evapotranspiration, with its consequent implications for moisture and surface energy fluxes.

Differences between GC3.05 and GC3.1 coupled ocean-atmosphere model configurations

As noted in section 3.2, a number of changes were included in GC3.1 to reduce the strong negative forcing due to anthropogenic aerosol emissions found in GC3.0 (Williams et al., 2018). Changes were made to the atmosphere model (Walters et al., 2017) and the sea-ice component (Williams et al., 2018). The main change was inclusion of the Liu et al. (2008) parameterisation of the impact of cloud droplet number on the droplet size distribution. This was included in GC3.05, as was an update to the refractive index of black carbon (Bond and Bergstrom, 2006).

In GC3.1, changes were made to the values of three parameters (see Table D.1) that are amongst those perturbed in GC3.05-PPE:

- The scaling factor for natural marine emissions of dimethylsulphide (parameter ps_natl_dms_emiss) was increased from 1.0 to 1.7, to account for the omission in GLOMAP-mode of natural emissions of organic aerosol. The altered value of 1.7 lies near the upper end of the expert prior distribution of values explored in development of GC3.05-PPE (Fig. 3.2).

- The value of mp_dz_scal, the vertical length scale for turbulent production of supercooled liquid in mixed phase clouds, was increased from 1.0 in GC3.0 to 2.0 in GC3.1. The latter value is close to the median of the prior distribution for GC3.05-PPE (Fig. 3.2).

- The value of cca_sh_knob was increased from 0.2 to 0.5, increasing the radiative effects of shallow convective cores. The altered value lies close to the upper end of the prior distribution for this parameter (Fig. 3.2).

Additional changes introduced in GC3.1 were not included in GC3.05, because they were not available at the time the GC3.05 configuration was finalised. These were:

- Inclusion of more detailed look-up tables for aerosol optical properties, allowing more accurate spectral resolution of solar absorption by aerosols.

- Improvements to the calculation of turbulent kinetic energy (TKE) in the parameterisation of boundary layer processes. These included correction of an error in the indexing of vertical levels, a reduction in the minimum value of TKE and introduction of an explicit estimate of TKE in cumulus clouds. The net effect of these changes is to reduce cloud droplet number concentrations.

- The specification of emissions of sulphur dioxide from continuously degassing volcanoes and explosive eruptions was updated to use the dataset of Dentener et al. (2006).

- The coefficient controlling the drag exerted by the ocean on sea ice was doubled in GC3.1, improving the simulation of ice thicknesses by increasing values along the North Greenland coast and reducing values in central regions of the Arctic Ocean. An increase was implemented to the thickness of ice within which the bare-ice albedo is reduced to account for the visibility of the underlying darker ocean.

Appendix E (Observational datasets, Strands 2 and 3)

Table E.1 gives acronyms for some of the key variables evaluated in the report, and Table E.2 shows the sources of observational data used for evaluation in Strands 2 and 3. Datasets used as formal observational constraints in Strand 1 are shown separately, in Tables B.1 and B.2. These Tables give an overarching view of the variables and periods considered across the project as a whole, with specific uses, periods and time sampling explained in the main text and figures. For example, Strand 3 uses 6-hourly atmospheric wind, temperature and specific humidity fields from ERA-Interim throughout the atmosphere, whereas Strand 2 uses daily or seasonal fields on selected pressure levels to screen potential GC3.05-PPE members. In some cases evaluation diagnostics are derived by further processing of the data listed below, for example by usage of clear- and all-sky radiative fluxes to obtain values of cloud radiative effect in Figure 3.3. The periods listed in Table E.2 are those used in this report.

Table E.1. Data acronyms (following CMIP5, or as defined in main text) used in Table E.2.

Variable	Acronym
Surface air temperature	tas
Precipitation rate	pr
Eastward wind on atmospheric pressure levels	ua
Northward wind on atmospheric pressure levels	va
Air temperature on atmospheric pressure levels	ta
Relative humidity on atmospheric pressure levels	hur
Specific humidity on atmospheric pressure levels	hus
Geopotential height on atmospheric pressure levels	za
Air pressure at sea level	psl
Outgoing short- and long-wave radiative flux at the top of the atmosphere	rsut, rlut
Outgoing short- and long-wave radiative flux at the top of the atmosphere under clear-sky conditions	rsutcs, rlutcs
Surface downwelling short- and long-wave radiative flux	rsds, rlds
Surface downwelling short- and long-wave radiative flux under clear-sky conditions	rsdscs, rldscs
Cloud radiative effect, and its shortwave and longwave components	cre, cresw, crelw
Global mean surface 0temperature	GMST
Sea surface temperature	SST

Table E.2. Observational datasets used to evaluate projections in Strands 2 and 3.

Dataset	Period	Variable(s)	Reference	Strand(s)
ERA-Interim	1980-2009	ua, va, za, ta, tas, hur, hus, psl, cre, cresw, crelw	Dee et al. (2011)	2,3
CERES	2000-2009	rlut, rsut, rlutcs, rsutcs, rsds, rlds	Loeb et al. (2009)	2
ERBE	2000-2009	rlut, rsut	Harrison et al. (1990)	2
ISCCP D2	2005-2009	Cloud cover in thin, medium-thick and thick categories at low, middle and high levels of the atmosphere	Rossow and Schiffer (1999)	2
GPCP	1980-2009	pr	Adler et al. (2003)	2
CMAP	1980-2009	pr	Xie and Arkin (1996)	2
HadSLP2	1900-2005	psl	Allan and Ansell (2006)	2
HadCRUT4	1900-2017	GMST	Morice et al. (2012)	2
NCIC	1910-2017	tas and pr, UK	Perry et al. (2009)	2,3
E-OBS	1981-2000	tas and pr, Europe. Version 15 is used for pr and version 14 for tas.	Haylock et al. (2008)	2,3
HadISST1	1900-2005	SST	Rayner et al. (2003)	2
HadISST2	2005-2009	SST	Titchner and Rayner (2014)	2
Reynolds	1982-2001	SST	Reynolds et al. (2002)	3

Acknowledgment

We thank the UKCP18 Peer Review Panel, chaired by Sir Professor Brian Hoskins, for their continued valuable input during the project, and specifically in reviewing and commenting on the final reports. The Peer Review Panel consisted Professor Mat Collins, Professor Jim Hall, Dr Ed Hawkins, Professor Gabi Hegerl, Dr Erik Kjellström, Professor Christoph Schär, Professor Ted Shepherd, Dr Claudia Tebaldi, Professor Dr Bart Van Den Hurk and Professor Sybren Drijfhout (University of Southampton).

The UKCP18 Project is part of the Met Office Hadley Centre Climate Programme funded by BEIS and Defra.

Version	Issue Date	Review Date	Reviewer	Change Description
1.0	November 18	November 19	Met Office	Launch version
2.0	March 19	November 19	Met Office	Updated to account for new methodology for production of UK and regional aggregate information in strand 1

00830/d