Deliverable D2.3

(GRANT AGREEMENT 776613)

Production of uncertainty quantifications/PDFs, including separation into different components.

Deliverable Title	*Production of uncertainty quantifications/PDFs, including separation into different components*	
Brief Description	*European Projections in the light of CMIP6 (including constrained ranges). Dissemination of new European Projections (via the Atlas) with scientific insight into the commonalities and differences, including guidance in adopting and using a particular methodology.*	
WP number	2	
Lead Beneficiary	*Met Office*	
Lead Contributors	*Ben Booth, Met Office and Said Qasmi, CNRM*	
Contributors	*Lukas Brunner, ETHZ* *Andrew Ballinger, UEDIN* *Erika Coppola, ICTP* *Glen Harris, Met Office* *Gabriele Hegerl, UEDIN* *Peter Kalverla, eScience Centre* *Yang Lui, eScience Centre* *Carol McSweeney, Met Office* *Rita Nogherotto, ICTP* *Christopher O'Reilly, Oxford/Reading University* *Tamzin Palmer, Met Office* *Said Qasmi, CNRM* *Aurelian Ribes, CNRM* *Hylke de Vries, KNMI*	
Creation Date Version Number Version Date	*08/10/2021* *1.0* *08/10/2021*	
Deliverable Due Date Actual Delivery Date	*30/11/2021* *30/11/2021*	
Nature of the Deliverable	x	*R – Report*
		P - Prototype
		D - Demonstrator
		O – Other
Dissemination Level/ Audience	x	*PU – Public*
		PP - Restricted to other programme participants, including the Commission services
		RE - Restricted to a group specified by the consortium, including the Commission services
		CO - Confidential, only for members of the consortium, including the Commission services

Version	Date	Modified by	Comments
0.1	08/10/2021	B. Booth	Transferred the contents from the collaboration hackmd.io document.
1.0	11/10/2021	B. Booth	Added context (Section 3) and finalised changes to the text and figures. Ready for Internal review
1.1	04/11/2021	B. Booth	Minor edits, including revised figures in Sections 8 and 9
1.2	24/11/2021	B. Booth and S. Qasmi	Revised document in light of comments made by our two internal reviewers

Contents

1. Executive Summary

This deliverable captures progress and outputs from the probabilistic projection methodologies, developed or assessed in EUCP. It includes links to the dissemination of WP2 climate projections and assessment of their various strengths and limitations. Here we outline some of the new top level science messages from this work:

- The availability of new CMIP6 climate simulations appear to change the European Climate Projection picture, pointing towards warmer summer conditions across Europe without the wetter Northern and Central European summer projections seen in earlier climate models.
- All the constraint methods delivering local probabilistic projections, mainly applied to the CMIP6 models, agree on a common picture of the mean climate over the 2041-2060 period, both in summer and winter. This is a remarkable result, particularly because the application of the methods here did not benefit from the same common frameworks (as used in deliverable D2.2).
- The constraint methodologies applied to the new CMIP6 simulations tend to down weight the warmer end of the climate projection ranges.
- The diversity of the proposed probabilistic methods allows us to respond to a large spectrum of user needs (who would require, for example, a comprehensive estimate of climate uncertainty, or spatial consistency over large regions etc.).
- A new out-of-sample "perfect model" analysis provides the first ever objective assessment and comparison of skill in available probabilistic climate projection methodologies. This approach demonstrates that using observational constraints provides a clear benefit for temperature (for all methodologies) over using just raw climate model ensemble data. It also exposes limitations of a couple of individual methods, with two showing degradation of skill in either Mediterranean temperature or central European and Mediterranean rainfall projections.
- There is insufficient climate change signal in the rainfall to provide a reliable constraint on future rainfall changes based on the observations to date. The out-of-sample analysis is the first to show that there are no consistent improvements over taking the raw climate model spread - though those methods that provide weights for individual models may still have value in excluding particularly poor simulations, even where doing so does not reduce the overall projection range.
- The out-of-sample analysis indicates that a multi-method constraint performs better than each individual constraint.
- Insights from the out-of-sample analysis provide useful context for potential users, by highlighting particular methodologies with potential issues in the Mediterranean and by identifying which variables are likely to benefit most from such constraints (temperature, and variables closely linked to temperature change).

2. Project Objectives

WITH THIS DELIVERABLE, EUCP HAS CONTRIBUTED TO THE ACHIEVEMENT OF THE FOLLOWING OBJECTIVES (DESCRIPTION OF ACTION, SECTION 1.1):

No.	Objective	Yes	No
1	Develop an ensembles climate prediction system based on high-resolution climate models for the European region for the near-term (~1-40 years)		no
2	Use the climate prediction system to produce consistent, authoritative and actionable climate information	yes	
3	Demonstrate the value of this climate prediction system through high impact extreme weather events in the near past and near future		no
4	Develop, and publish, methodologies, good practice and guidance for producing and using EUCP's authoritative climate predictions for 1-40 year timescales	yes	

3. Overview

3.1 Motivation

- Not all current climate simulations are equally likely, and yet measures of more skillful projections are not routinely used in current climate projections.
- There is, therefore, interest in exploring the range of current methods that use evaluation of simulated historical skill vs observations to assess the more likely future response
- These methods are often effective at ruling out poor simulations. The impact of doing so, however, may not necessarily have a large impact of the future range. It is useful to identify where it does, as this points to more confident information for future projected changes.

This deliverable aims to capture the work carried out under WP2's Task 2.2 (Production of Uncertainty Quantifications (UQs)/Probability Density Functions (PDFs)). There have been three previous deliverables that have informed the reporting here. D2.1 was a report on the evaluation and combination of potential observational/emergent constraints relevant to European climate and D2.2 was an evaluation report on different methods to produce UQs/PDFs. Due to the inter-linking nature of the work that informed both of these, they were submitted as merged reports. The previous deliverables explored the impact of using observations to down weight poor climate projections in 6 different potential methodologies (that have either been used in European projections previously or developed under EUCP). Furthermore, the use of observational constraints across the prediction/projection timeline has been explored in Deliverable 5.1, published in Hegerl et al., 2021.

The work in these deliverables represented the first systematic intercomparison of probabilistic methods, which exposed both where there was common agreement on the inferences (which might imply a more robust result) and where the inferences disagreed across methodologies (highlighting lack of consensus). The current deliverable (D2.3) extends this work in a number of important ways:

- **To bring the climate projections up to date, in light of CMIP6**:
- **Make observationally constrained projection data available**: There are a number of ways that WP2 has worked to do this, beyond the publication of papers. This includes an Atlas of European Climate Projections and (downloadable) access to this underlying data for all WP2 methodologies. We have also strongly encouraged groups to release source code and public implementations (e.g. ESMValTools).
- **To provide clearer information on adopting an observationally calibrated approach**: This work package looked at climate projections from 6 different methodological approaches, with the aim to provide an assessment of the added value and shed light on potential differences in outcomes. The expectation was to use the outcome as a basis of selecting a

targeted, reliable methodology that could underpin European climate projection advice. However, the outcomes of this assessment left it as an open question to the reader what method to adopt, which for almost all potential users of this data falls beyond their expertise. The challenge is that these methodologies have each been developed against different use cases and each has distinct strengths. This deliverable aims to address the challenge of selecting and applying methodologies to observational constraint climate projections in a number of ways: We discuss a number of hypothetical user cases which explore how different users can make use of the methods presented here, considering their expertise and need. Secondly, we expand and explore ways in which these methodologies are different (which informs the former). Lastly, we present a new perfect model analysis that provides objective assessment of the skill of individual methodologies.

Framing of the deliverable in this way, has been influenced by the feedback we have received from our External EUCP reviewers.

3.3 Deliverable Outline

This deliverable aims to capture the work carried out under WP2's Task 2.2 (Production of Uncertainty Quantifications (UQs)/Probability Density Functions (PDFs)). It is not intended to be read linearly, from cover to cover, but instead as a reference to particular analysis and work strands. This brief description here is intended to provide a plain language guide.

This task has either produced or is the final stages of producing a number of key outputs. These outputs are described in the following section (**Section 4**).

One of the key new developments that this deliverable captures is the arrival of new CMIP6 simulations (that have come online, in sufficient number to now be used as a basis for updated climate projections). CMIP6 offers a different picture of how Europe will change, compared to previous generations of climate model projections. We discuss these differences and what they may or may not mean in practice, in **Section 5**.

Section 6 documents the application of several our WP2 methodologies to the new CMIP6 data. This section is fairly technical and intended for those who wish to understand the details of the underlying WP2 methodologies.

One of the key outputs from the Task 2.2 activity has been the production of new climate projection maps of Europe (that will be made publicly available). **Section 7** presents an overview of these new projections.

There are differences evident between European climate projections produced by our 6 WP2 methodologies. Some of the more influential differences are discussed in **Section 8** alongside some of their implications.

Section 9 presents explorations of a number further application of these methods. The first of this is the output from a systematic out-of-sample blind test of each of our methodologies (in **Section 9.1**). This provides an objective assessment of whether calibrating available climate simulations using observations adds value or not – implications of which influence potential user applications (discussed in Section 10). **Section 9.2** introduces the application of the first observational constraints on the range of extreme precipitation change (using one of our WP2 methodologies). The new European climate projection maps produced in the new Atlas (sections 4 and 7) are based on 20-year climate mean changes. **In Section 9.3** we illustrate how internal variability can change this projection picture at small spatial or temporal scales. Finally, we explore in **Section 9.4** how some of the new constrained ranges can provide a wider context for the new Convective Permitting Regional Climate Models produced in EUCP (in work package 3).

The last section (**Section 10**) focuses on the adoption of WP2 observationally calibrated climate projection methodologies by potential users and aims to draw out the various strengths and limitations of availability methodologies. Firstly, in **Section 10.1** we provide a precis of the out-of-sample analysis, pointing to the relative strengths of available methods and where calibration with observations does or does not add value. There are different barriers to apply WP2 methodologies to new applications. In **Section 10.2**, we provide an indication of the challenge involved in doing so, for those methods which have made their code available. In **Section 10.3** we discuss spatial and physical coherence, which is a feature of a subset of projections and we briefly summarise the available outputs and added strengths of the available methods in **Section 10.4**. In the final sub-section, **Section 10.5**, we use four hypothetical users, and their climate information needs, to provide some signposting to how available data from WP2 might be used.

3.4 Observationally constrained ranges and the IPCC

During the course of EUCP, methods that constrain climate projections based on their ability to capture observed changes have gained greater acceptance. The recent adoption of such methods in the latest IPCC report, is perhaps the best illustration of this. Quoting from the Summary for Policy Makers (SPM):

> For the first time in an IPCC report, assessed future changes in global surface temperature, ocean warming and sea level are constructed by combining multi-model projections with observational constraints based on past simulated warming, as well as the AR6 assessment of climate sensitivity.
>
> Summary for Policy Makers, AR6, IPCC

Two of the WP2 methods that constrained the future projection range, based on observations of past changes contributed are featured in the AR6 assessment. Accounting for model dependence and performance was also discussed as an important consideration in the latest IPCC report (AR6, CH4, Box4.1). The Climate model Weighting by Independence and Performance (ClimWIP) method (Knutti et al. 2017, Brunner et al. 2019, 2020a, b), that was partly developed within the framework of EUCP, was, among other approaches, highlighted by the IPCC for taking into account model independence. As the method and source code of ClimWIP are freely available they can also be used for other studies, such as the one by Liang et al. (2020), which was used to constrain future projections in IPCC (AR6, CH4, Fig. 4.11).

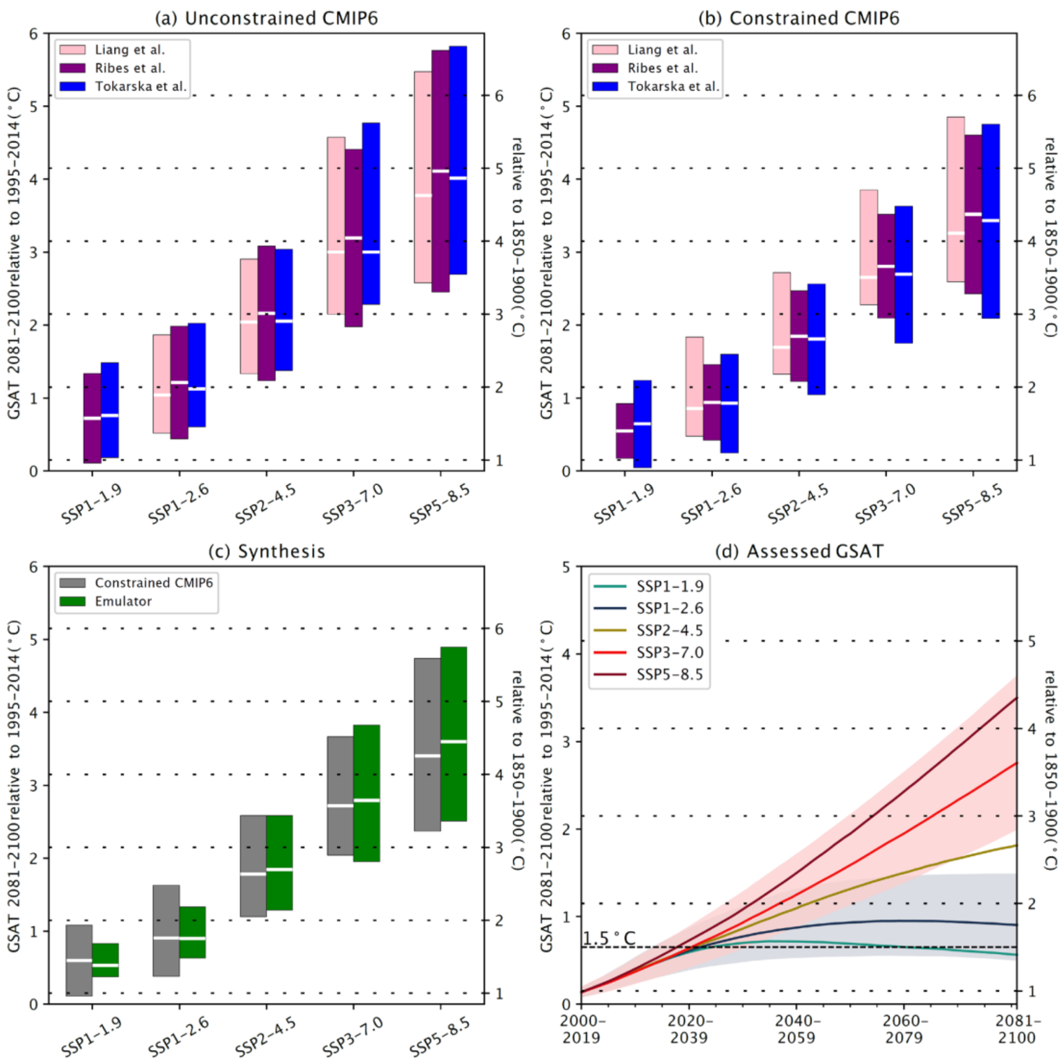


Figure 4.11 (<u>reproduced from Chapter 4, AR6, IPCC assessment</u>): Multiple lines of evidence for GSAT changes for the long-term period, 2081–2100, relative to the average over 1995–2014, for all five priority scenarios. The unconstrained CMIP6 5–95% ranges (coloured bars) in (a) differ slightly because different authors used different subsamples of the CMIP6 archive. The constrained CMIP6 5–95% ranges (coloured bars) in (b) are smaller than the unconstrained ranges in (a) and differ because of different samples from the CMIP6 archive and because different observations and methods are used. In (c), the average of the ranges in (b) is formed

(grey bars). Green bars in (c) show the emulator ranges, defined such that the best estimate, lower bound of the very likely range, and upper bound of the very likely range of climate feedback parameter and ocean heat uptake coefficient take the values that map onto the corresponding values of ECS and TCR of Section 7.5 (see BOX 4.1). The time series in (d) are constructed by taking the average of the constrained CMIP6 ranges and the emulator ranges. The y-axes on the right-hand side are shifted upward by 0.85°C, the central estimate of the observed warming for 1995–2014, relative to 1850–1900 (Cross-Chapter Box 2.3, Table 1). Further details on data sources and processing are available in the chapter data table (Table 4.SM.1)

The second WP2 methodology, KCC (Ribes et al, 2020, Qasmi et al, 2021), was one of three methods used in the IPCC figure 4.11 illustration (see figure above). KCC was formally known as HistC in the previous EUCP deliverables D2.1 and D2.2 (and referred to in the figure by "Ribes et al").

This AR6 quote is not quite accurate. In fact, another WP2 probabilistic methodology, ASK (Allen et al., 2000; Stott and Kettleborough, 2002), was used in the IPCC AR4 to constrain temperature projections together with the model range (Knutti et al., 2008) and hence was the first methodology used in this context. However, it is now gratifying to see these methods hold a higher profile place.

One of the factors that prompted the IPCC lead authors to adopt these methods has been the differences between what the AR6 report deems the likely range to be for global climate sensitivity, and the new CMIP6 ensemble that appears to sample a large number of realisations in the upper end of this likely range. Weighting these new CMIP6 simulations by observations helps to reconcile these apparent contradictory lines of evidence. The new CMIP6 simulations, and the methods to constrain these new projections ranges, both have implication for the European climate outlook. We cover both of these aspects in the following sections.

4. Work package outputs and impact

This deliverable documents climate projections from 6 methodologies (which are referred to in this deliverable by their institutional acronym). The Methodology names are also referred to here, in the links to the source code and public implementations. The acronyms for both institutes and methodologies are documented in Tables 1 and 2

Four of these methods have been applied to the new CMIP6 simulations and are documented in Table 1. The first three of these use observations to constrain the simulation range. The fourth methodology (ASK) adopts an independent and complementary approach. It uses information from a formal Detection and Attribution of the human fingerprint in observed climate change (using a CMIP6 based estimate) to constrain the range of future change that is consistent with this.

Table 1: List of the contributing groups and their methods for constraining climate projections, updated to use CMIP6 as the basis.

Contributing Institute	**Institute Acronym**	**Method name**
International Centre for Theoretical Physics	ICTP	REA
Centre National de Recherches Météorologiques/ Le Centre National de la Recherche Scientifique,	CNRM/CNRS	KCC (formerly HistC)
Eidgenössische Technische Hochschule (ETH) Zürich	ETH Zurich (ETHZ)	ClimWIP
University of Edinburgh	UEdin	ASK

This deliverable also present climate projections from two other methodologies, in Table 2. These draw on different, bespoke, climate model ensembles that are considered state-of-the-art even though they are not tied to the CMIP cycles.

Table 2: List of the contributing groups and their methods for constraining climate projections, that provide updated projections for the new regions (e.g. Atlas).

Contributing group	**Institute Acronym**	**Method name**
University of Oxford	UOxf	CALL
UK Met Office	Met Office	UKCP

Terminology

In this deliverable, all the methods described make use of the observational record to identify future projections that are most consistent with the observed climate. Methodologies will variously describe this as “weighting” (ClimWIP, REA, UKCP) or more generally as a “observational constraint” (ClimWIP, REA, UKCP, KCC) where the observations help narrow down the range. The climate projections from the detection and attribution approach (ASK) and from CALL are able to shift distributions of projected change in light of the observational comparison, so are not strictly speaking with a “weighting” or a “constraint”. All methods are a way of “calibrating” the projections against observations. In this deliverable, we do not dwell on these differences and use “constraint” in a more general sense, when we are actually discussing “calibrations”.

4.1. Task 2.2 outputs

Atlas

We have been developed a Climate Projection Atlas to provide public access to seasonal mean European climate projection maps and underlying data, from WP2 methodologies. This provides complimentary information to the IPCC Atlas. The recent IPCC WG1 report concludes that (globally) the range of projected global climate changes are broadly inline with earlier assessments, whereas the new CMIP6 simulations include a large number of projections which are warmer than earlier assessments. This is because the WG1 report synthesizes multiple lines of evidence, including output from methodologies to constrain climate projection based on historical observations (such as those presented here) that suggest these new warmer projections are samples of less likely (but still plausible) changes. However, this synthesis analysis is not reflected in the IPCC Atlas which just present the raw CMIP6 range. The EUCP Climate Projection Atlas presents both raw and constrained climate model ranges. This provides the first publicly available data portal for observationally constrained climate projections, in this case for the European domain.

Top level messages from this Atlas are analysed in Section 7. This presents readers with maps of projected changes in mid-century European temperature and precipitation, for both summer and winter. This Atlas has been developed in conjunction with Peter Kalverla and Yang Lui at the e-Science Centre (WP6)

Data

Data the underpins the Climate Projection Atlas is available for public download (via the download link on the Atas). This download is not yet live, as we are still in the process of standardizing the outputs from our different Methods.

Papers

Projections from ALL groups:

Brunner, L., McSweeney, C., Ballinger, A. P., Befort, D. J., Benassi, M., Booth, B., Coppola, E., de Vries, H., Harris, G., Hegerl, G. C., Knutti, R., Lenderink, G., Lowe, J., Nogherotto, R., O’Reilly, C., Qasmi, S., Ribes, A., Stocchi, P., & Undorf, S. (2020). Comparing Methods to Constrain Future European Climate Projections Using a Consistent Framework. Journal of Climate, 33(20), 8671–8692. https://doi.org/10.1175/JCLI-D-19-0953.1

CNRM/CNRS

Ribes, A. et al. (2021): Making climate projections conditional on historical observations. Science Advances, 7, eabc0671, DOI: 10.1126/sciadv.abc0671

Qasmi, S. and Ribes, A. (in revision): Reducing uncertainty in local climate projections Nature Communications DOI: 10.21203/rs.3.rs-364943/v1

ETHZ

Brunner, L., Lorenz, R., Zumwald, M., & Knutti, R. (2019). Quantifying uncertainty in European climate projections using combined performance-independence weighting. Environmental Research Letters, 14(12), 124010. https://doi.org/10.1088/1748-9326/ab492f

Brunner, L., Pendergrass, A. G., Lehner, F., Merrifield, A. L., Lorenz, R., & Knutti, R. (2020). Reduced global warming from CMIP6 projections when weighting models by performance and independence. Earth System Dynamics, 11(4), 995–1012. https://doi.org/10.5194/esd-11-995-2020

Merrifield, A. L., Brunner, L., Lorenz, R., Medhaug, I., & Knutti, R. (2020). An investigation of weighting schemes suitable for incorporating large ensembles into multi-model ensembles. Earth System Dynamics, 11(3), 807–834. https://doi.org/10.5194/esd-11-807-2020

Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E. M., Brunner, L., Knutti, R., & Hawkins, E. (2020). Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. Earth System Dynamics, 11(2), 491–508. https://doi.org/10.5194/esd-11-491-2020

UOxf

O'Reilly, C. H., Befort, D. J., & Weisheimer, A. (2020). Calibrating large-ensemble European climate projections using observational data. Earth System Dynamics, 11(4), 1033-1049. https://doi.org/10.5194/esd-11-1033-2020

UEdin

Hegerl, G., Ballinger, A., Booth, B., Borchert, L. F., Brunner, L., Donat, M., Doblas-Reyes, F., Harris, G., Lowe, J., Mahmood, R., Mignot, J., Murphy, J., Swingedouw, D. & Weisheimer, A., (2021): Towards consistent observational constraints in climate predictions and projections. Frontiers in Climate.doi:10.3389/fclim.2021.678109.

UKCP

Murphy, J.M., Harris, G.R., Sexton, D.M.H., Kendon, E.J., Bett, P.E., Clark, R.T., Eagle, K.E., Fosser, G., Fung, F., Lowe, J.A., McDonald, R.E., McInnes, R.N., McSweeney, C.F., Mitchell, J.F.B., Rostron, J.W., Thornton, H.E., Tucker, S., Yamazaki, K. (2018) UKCP18 Land Projections: Science Report, Met Office Hadley Centre, Exeter, U.K., https://www.metoffice.gov.uk/pub/data/weather/uk/ukcp18/science-reports/UKCP18-Land-report.pdf

Lowe J.A., D. Bernie, P. Bett, L. Bricheno, S. Brown, D. Calvert, R. Clark, K. Eagle, T. Edwards, G. Fosser, F. Fung, L. Gohar, P. Good, J. Gregory, G. Harris,T. Howard, N. Kaye, E. Kendon, J. Krijnen, P. Maisey, R. McDonald, R. McInnes, C. McSweeney, J.F.B. Mitchell, J. Murphy, M. Palmer, C. Roberts, J. Rostron, D. Sexton, H. Thornton, J. Tinker, S. Tucker, K. Yamazaki, S. Belcher (2018) UKCP18 Science Overview Report, Met Office Hadley Centre, Exeter, U.K., https://www.metoffice.gov.uk/pub/data/weather/uk/ukcp18/science-reports/UKCP18-Overview-report.pdf

**NOTE, NEITHER OF THE ABOVE TWO PUBLICATIONS WERE DEVELOPED UNDER EUCP FUNDING, BUT BOTH ARE INCLUDED HERE AS REFERENCES TO THE UNDERLYING SCIENCE THAT CONTRIBUTES TO THIS METHODOLOGY, AND (IN THE CASE OF THE 2ND LINK) AS AN EXAMPLE OF USING AND APPLYING OUTPUTS FROM THE PROBABILISTIC PROJECTIONS. ALL OTHER REFRENCES REFER TO WORK WHERE EUCP FUNDING HAS CONTRIBUTED TO THE PUBLISHED RESULTS.*

Method code

CNRM/CNRS

KCC R package: https://doi.org/10.5281/zenodo.5233947
Jupyter Notebook: https://gitlab.com/saidqasmi/kcc_notebook
Shiny application: https://saidqasmi.shinyapps.io/KCC-shinyapp/

ETHZ

ClimWIP development version: https://github.com/lukasbrunner/ClimWIP/

ClimWIP used for Brunner et al. (2020): https://doi.org/10.5281/zenodo.4073039

Documentation of the ESMValTool implementation of ClimWIP: https://docs.esmvaltool.org/en/latest/recipes/recipe_climwip.html

ICTP

REA method: http://doi.org/10.5281/zenodo.3890966

5. How does CMIP6 change the European Projection picture?

Climate projections are based on currently available climate model simulations. New global climate model simulations from the Coupled Model Intercomparison Project 6 (CMIP6) first became available within the lifetime of the EUCP projection. Whilst timescales mean that much of our regional projection are based on previous generation global modelling (such as the new convective permitting simulations in WP3), the new CMIP6 simulations provide an opportunity to reassess current climate projections in light of any changes in the large-scale European climate drivers that they represent. Each climate model generation captures the impact of advances in the underlying climate modelling capabilities (for example, CMIP6 includes more comprehensive inclusion of prognostic cloud properties and their interactions with aerosols (Meehl et al, 2020) and includes higher resolution modelling). New simulations, with their potential to shed new insights in how climate change may manifest itself in our regions, prompts us to re-evaluate our existing climate projections. For example, as has been done in Australia (Grose et at, 2020).

The impact of new CMIP6 data for European Climate Projections was analysed as a core part of this work package, and published earlier this year. Unlike differences in previous climate model generations (such as differences between CMIP3 and CMIP5) CMIP6 appears to offer a different outlook for future European Summer climate changes.

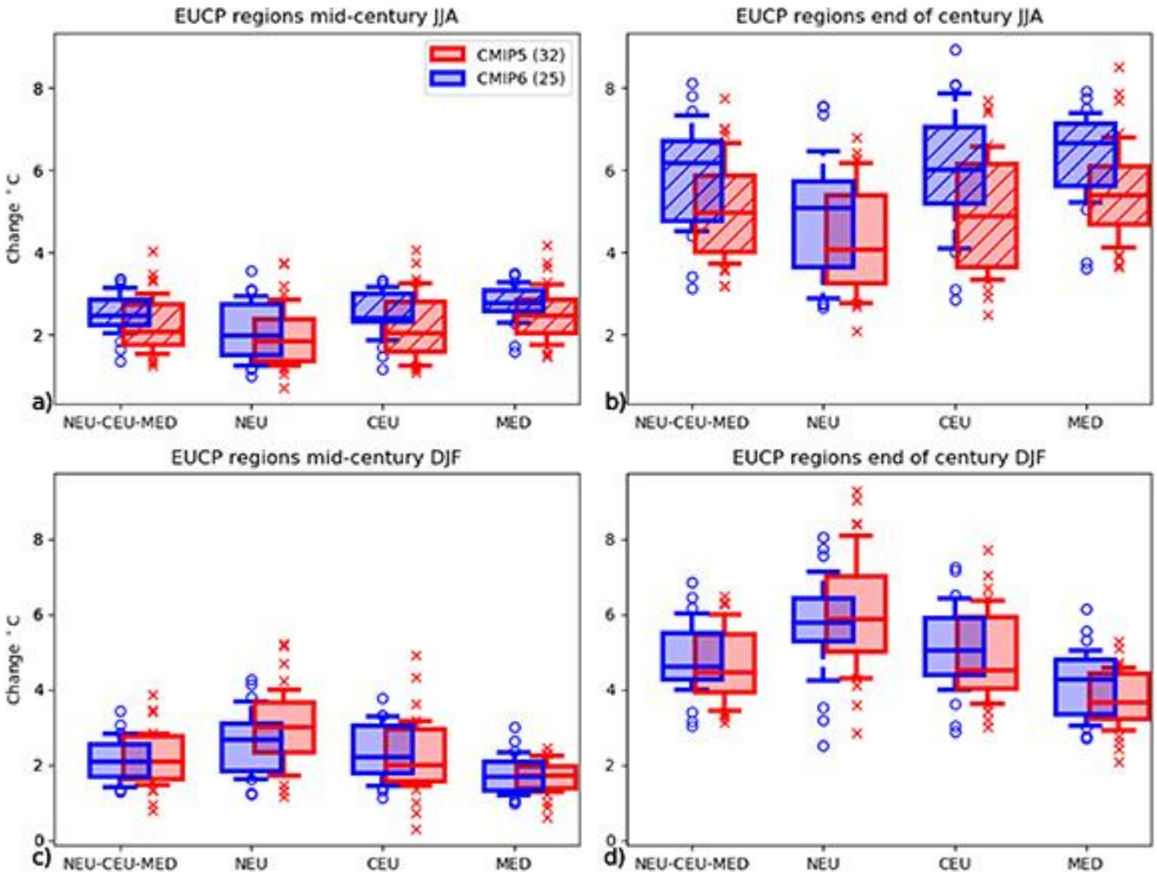


Figure 1 *(reproduced from Palmer et al, 2021). Projections of average summer (JJA) and winter (DJF) temperature change for CMIP5 and CMIP6 ensembles in their respective RCP8.5 scenarios. Baseline: 1995–2014. Mid-century: 2041–2060. End of century: 2081–2100. Boxes show the interquartile range and whiskers are at 10th and 90th percentiles. Hatching is shown where the MIPs were found to be significantly different.*

As can be seen from Figure 1, CMIP6 projects stronger summer warming compared to CMIP5. Many of these changes are linked to greater warming in the CMIP6 global projections. The AR6 report (Lee et al, 2021) attributes this greater warming to:

> *"About half of the increase in simulated warming has occurred because higher climate sensitivity is more prevalent in CMIP6 than in CMIP5; the other half arises from higher ERF in nominally comparable scenarios (e.g., RCP8.5 and SSP5-8.5; medium confidence). … For SSP5-8.5, higher climate sensitivity is the primary reason behind the upper end of the warming being higher than in CMIP5"*

The increased sampling of larger climate sensitivities in CMIP6 simulations is the largest factor, and for Northern Europe and the Mediterranean explains the largest difference in the European temperature projections (Figure 1). However, this does not suggest that our current projection advice for these European regions should necessarily change. This is because the higher climate sensitivities that are now more thoroughly sampled in CMIP6, are not considered any more likely that previous IPCC assessments. So, the upper end of the projections increases because larger climate sensitivities are sampled, not because we think these stronger responses are more likely. AR6 (Forster et al, 2021):

> *On average, CMIP6 models have higher mean ECS [Equilibrium Climate Sensitivity] and TCR [Transient Climate Response] values than the CMIP5 generation of models. They also have higher mean values and wider spreads than the assessed best estimates and very likely ranges within this Report.*

The larger warming samples in CMIP6 are not necessarily without merit. Arguably, previous CMIP5 based projections were poorly based to sample higher impact, but less likely changes. CMIP6 suffers not with the issue that all the high-end responses are extremely unlikely, but that it has large number of samples for less likely high-end changes. Even the extreme end of this distribution may still have value for any assessment that might be aimed at particularly risk adverse users. As AR6 notes (Lee et al, 2021):

> While high-warming storylines – those associated with GSAT [Global Surface Air Temperature] levels above the upper bound of the assessed very likely range – are by definition extremely unlikely, they cannot be ruled out.

However, CMIP6 presents a different picture for Central Europe. Here, the stronger warming in CMIP6 projections is not solely due to the just stronger global warming. As such, the stronger warmings in Central Europe cannot be as easily dismissed. Palmer et al, 2021 show that roughly 40% of the enhanced central European warming is due to differences in the regional response to global change (See Figure 2). This enhanced regional sensitivity may imply that existing climate projections, for the central European region, are under-stating the future regional warming.

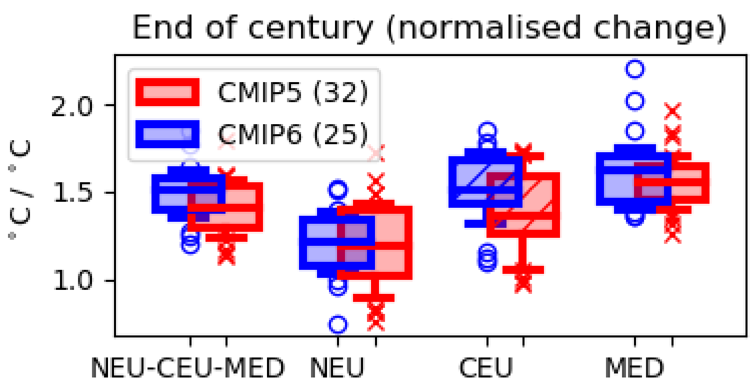
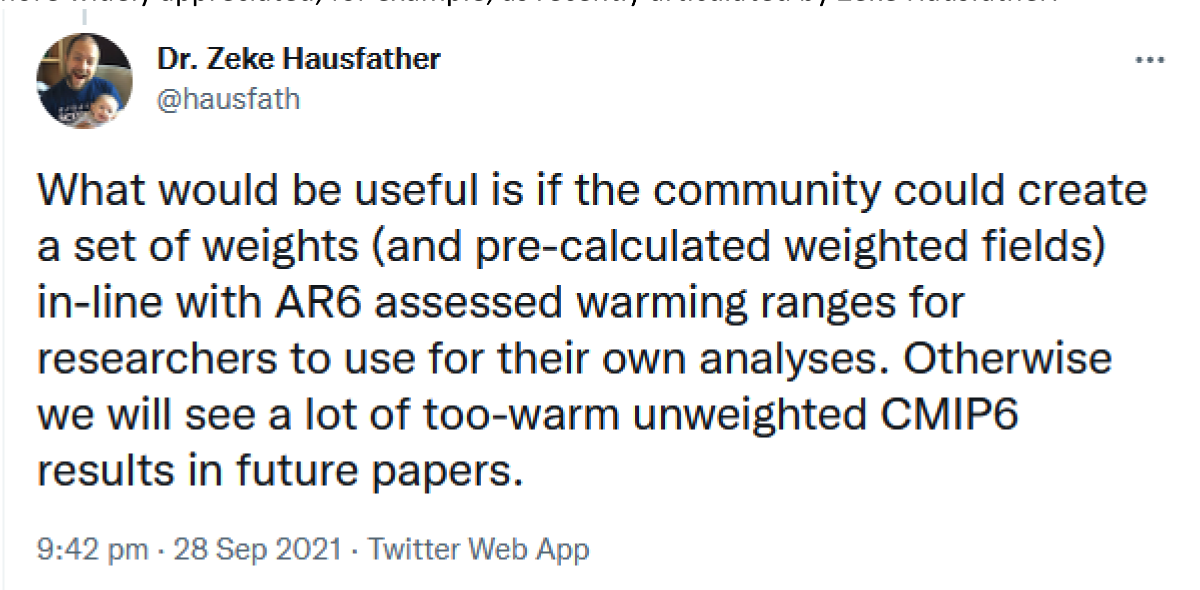


Figure 2 *(taken from Figure 5 in Palmer et al, 2021). Projections of average summer normalised temperature change (change per °C) for CMIP5 and CMIP6 ensembles. Where differences exist between CMIP5 and CMIP6, this points to enhanced regional sensitivity that is not linked to reported differences in global sensitivity.*

There is evidence that the precipitation projections also differ. Again, particularly in summer. CMIP6 does not sample the wetter CMIP5 summer rainfall responses in Northern and Central Europe suggesting, perhaps, that these are less likely. Indeed, CMIP6 points to a clearer summer drying signal in Central Europe. For the Mediterranean region, CMIP6 suggests that the summer drying is smaller for a certain magnitude of warming while a number of stronger magnitude warmings are now represented. So, on the face of it the range of projected summer has not changed with the new CMIP6 simulations, though we might want to scale back some of the stronger drying signals if we felt the stronger warming in these models was less likely. More details of how CMIP6 changes the rainfall projections can be found Palmer et al, 2021.

This analysis, that was carried out as part of WP2, provides the basis for the next sections. Interest in observationally weighted climate projections goes beyond our desire to reduce climate project uncertainties/spread. The latest AR6 IPCC assessment suggests that stronger projected global warming in CMIP6 (which tends to increase European summer projections) is merely more comprehensively sampling the lower likelihood but high impact responses, so this projected range should not necessarily be directly translated into impacts. This challenge is starting to become more widely appreciated, for example, as recently articulated by Zeke Hausfather:



The observationally weighted approaches developed and applied under WP2 provide a way that this challenge can be addressed. In the following section, we outline the impact of observational constraints when they are applied to the new CMIP6 climate projections for Europe.

6. Observational constraints on CMIP6 - new results from the contributing groups

All the 6 WP2 methodologies have been documented before (Brunner, et al, 2020a, and Deliverable D2.2). **This section summarises the application of those methodologies that were applied to CMIP6 (done by CNRS, ICTP, ETH and UEdin) for the first time, as well as new insights that arise from this application.** The two other methodologies (neither of which are tied to CMIP cycles) also provide new Climate Projection data for the Atlas Data from the UKCP method (Met Office) is in the Atlas, but the method is that which has already been documented in D2.3 and Brunner et al, 2020a. As such UKCP is only briefly discussed in this section, and the reader is directed to an appendix for the fuller description of this method.

CNRS

CNRM/CNRS has contributed to provide constrained projections using the Kriging for Climate Change (KCC) method with both CMIP5 and CMIP6 ensembles.

A constraint based on observed warming to date is implemented by considering past observations. The comprehensive description of the KCC method is available in Qasmi and Ribes (in revision), and can be summarized in three steps:

- estimate the response of each CMIP model to all external forcings over the historical period;
- derive a multi-model distribution which characterizes the model uncertainty in this forced response;
- sub-select trajectories which are consistent with available observations, according to the Bayesian theory.

This procedure has been applied for annual Global Mean Surface Temperature (GMST) in Ribes et al. (2021). We show that uncertainty in GMST can be considerably reduced by using the information provided by the recent observed warming: typically, by a factor of 3 in the past and in the near term (i.e., by 2040) and a factor of 2 in the long term (late 21st century). Figure 3 illustrates the constraints obtained for the CMIP6 and CMIP5 ensembles. On average, the unconstrained CMIP6 ensemble projects a higher warming than CMIP5, consistent with a higher sensitivity. The CMIP6 ensemble also exhibits much wider uncertainty ranges, leading to a lower bound on 2100 projected warming lower than that of CMIP5. After applying the observational constraint, ranges agree remarkably well in the near term (e.g., before 2040), suggesting that observations play a dominant role in this time frame. CMIP5- and CMIP6-constrained ranges also exhibit comparable width for all scenarios and periods considered.

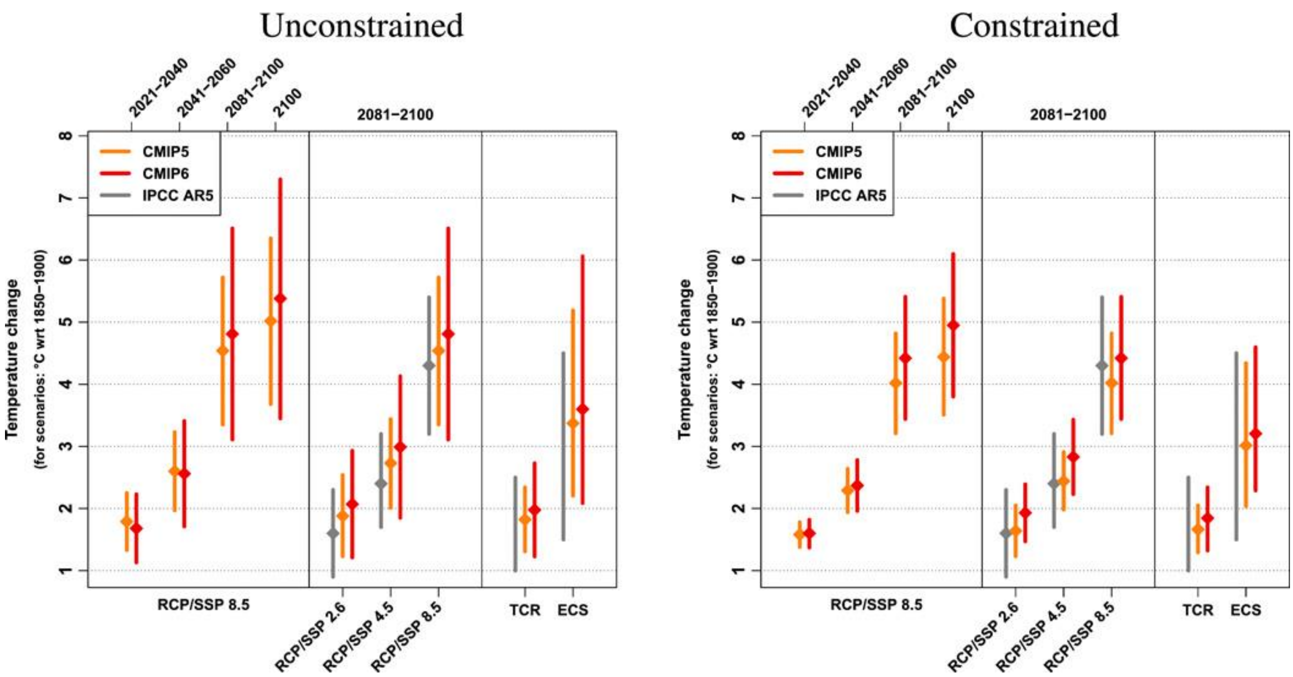


***Figure 3**: Comparison of warming ranges from the CMIP6 ensemble to CMIP5 and IPCC AR5. Comparison of the unconstrained (left) and constrained (right) projected warming ranges in response to various emission scenarios and idealized experiments, as estimated using the CMIP5 [17 models, Representative Concentration Pathway (RCP) scenarios] and CMIP6 ensembles (22 models, SSP scenarios) to those assessed in the IPCC AR5. Comparison is carried at different dates for a scenario (8.5 W m^{-2}; SSP5-8.5 in CMIP6, RCP8.5 in CMIP5, and IPCC AR5) over the 2081–2100 period for various emission scenarios and for the TCR and the ECS. Numbers for RCP scenarios in IPCC AR5 are taken from table 12.2 in and shifted by +0.61°C (in agreement with the caption of that table) to account for the discrepancy in reference periods. TCR/ECS ranges from IPCC AR5 are assessed likely range (implying 66% confidence). Reprinted from Ribes et al (2021).*

We have transposed these new results on the constrained GMST findings to regional and local scales in Qasmi and Ribes (in review). The KCC method has been extended to take into account global observations as well as local observations available at a grid cell scale. Over Europe, the CMIP6 annual temperature projections constrained by both global and local observations exhibit a reduction of the uncertainty of 50 % in average by 2050, and a downward revision of the projected warming by about 0.5°C (Qasmi and Ribes, in revision). Overall, similar results are obtained for the temperature projections aggregated over the European SREX regions. A downward revision of 0.5 °C is associated with the NEU region for both CMIP5 and CMIP6 ensembles (Figure 4). The difference between the two ensembles is more marked for the CEU and MED regions. For the MED region, the projected warming is revised upward for CMIP6 and downward for CMIP5, while the downward revision for the CEU region in greater for CMIP5 (1°C) than for CMIP6 (0.3°C). The higher ratio between local and global warming in CMIP6 than in CMIP5 would explain this difference. We find that a constraint by the sole GMST observations leads to a downward revision of the projections (Qasmi and Ribes, in revision), while the use of the sole local observations provides a similar picture for both CMIP5 and CMIP6. Impacts of local and global observations are discussed in Section 8 and 10.3.

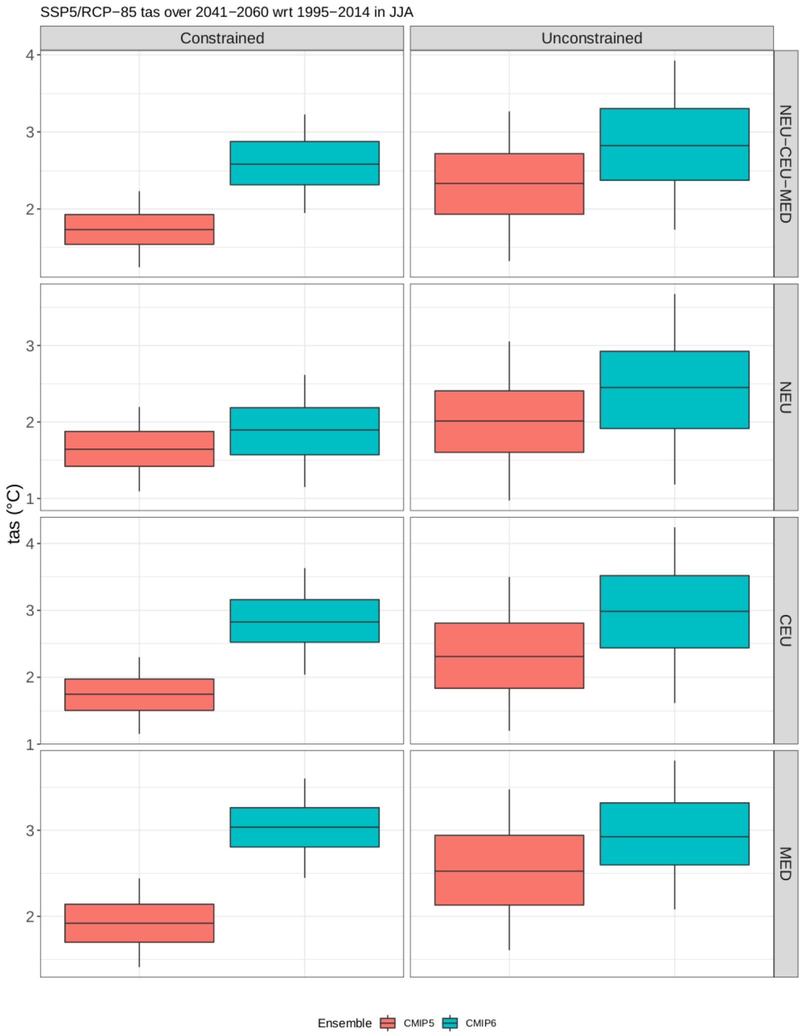


***Figure 4**: Summer (July–August) temperature change 2041–60 relative to 1995–2014 for the combined European region (NEU-CEU-MED) as well as the three European SREX regions. The red (blue) boxes give the unconstrained and constrained distributions for CMIP5 (CMIP6). The top (bottom) of the box represents the 25th (75th) percentile of the distribution and the upper (lower) whisker represents the 95th (5th) percentile.*

A perfect model evaluation of these results has been conducted for all these results in the associated papers, indicating a significant added value of the constraints.

ICTP

ICTP's work has focused on applying the Reliability Ensemble Averaging (REA) method to mean seasonal temperature and precipitation changes in three different European spatial regimes. REA uses two general "reliability criteria" to assess, mostly in a qualitative way, the reliability of regional climate change simulations (e.g., Kattenberg etal. 1996; Giorgi et al. 2001b).
The first is based on the ability of models to reproduce different aspects of present-day climate: the better a model performance in this regard, the higher the reliability of the climate change simulation. We refer to this as the "model performance" criterion.

The second criterion is based on the convergence of simulations by different models for a given forcing scenario, greater convergence implying higher reliability of robust signals that are little sensitive to the differences among models. We refer to this as the ‘‘model convergence’’ criterion. For each j-model, weights are calculated as:

$$R_j = \left[\left(\frac{\epsilon}{|B_j|} \right)^m + \left(\frac{\epsilon}{|D_j|} \right)^n \right]^{\frac{1}{mn}}$$

and are functions of the model bias (Bj) (the higher the bias the lower the model reliability) and of the distance (Dj) of the change calculated by a given model from the REA average change, that is, the higher the distance the lower the model reliability. The parameter ϵ in the equation above is a measure of natural variability in 30-yr average regional temperature and precipitation. The parameters m and n can be used to weigh each criterion. For most calculations in this work, m and n are assumed to be equal to 1. In order to calculate ϵ, we compute timeseries of observed regionally averaged temperature and precipitation for the historical period over each region of interest. We then compute the variance of the series after linearly detrending the data (to remove century-scale trends).

For a detailed description of the method, we refer to Giorgi and Mearns (2002).

REA method is applied for the periods 2041-2060 and 2081-2100 relative to the reference period 1995-2014. Regional ensemble results of 55 scenario simulations for the RCP8.5 and RCP2.6 at 0.11-degree resolution over the common EURO-CORDEX domain, using 8 GCMs and 11 RCMs, are compared with the driving CMIP5 global models and with 15 CMIP6 global models at 2.5 degrees resolution (see Table 3). Observations used for the bias calculation are the E_OBS version 19.0e.

In order to investigate the difference in constraining projections from CMIP6 and CMIP5, REA was applied for the period 2041-2060 relative to the period 1995-2014. A selection of CMIP5 global models is compared to a selection of CMIP6 global models (see Tables 3). Observations used for the bias calculation of temperature and precipitation are the E_OBS version 19.0e.
Figure 5 illustrates summer (JJA) temperature changes in four of the European subregions used in Brunner et al. (2019) based on the three European SREX regions: Northern Europe (NEU), Central Europe (CEU), Mediterranean (MED), and the combined European region (EUR: NEU + CEU + MED). The temperature changes are shown for the CMIP5 (in orange) and CMIP5 (violet) ensembles. For all the regions the CMIP6 ensemble projects a higher warming than CMIP5, consistent with a higher climate sensitivity (AR6 IPCC). The CMIP6 ensemble exhibits a much wider uncertainty range with respect to CMIP5 over Northern Europe, and a reduction of the spread in the Mediterranean. The reduction of the uncertainties evident for the CMIP5 ensemble in Central Europe and CMIP6 in Northern Europe but does not have an impact in other regions.

The different behaviour of the two ensembles could be attributed to their different resolution and process parameterization that could lead for example to a decrease of spread in the Mediterranean region for the CMIP6 ensemble and an increase in the Northern European region.

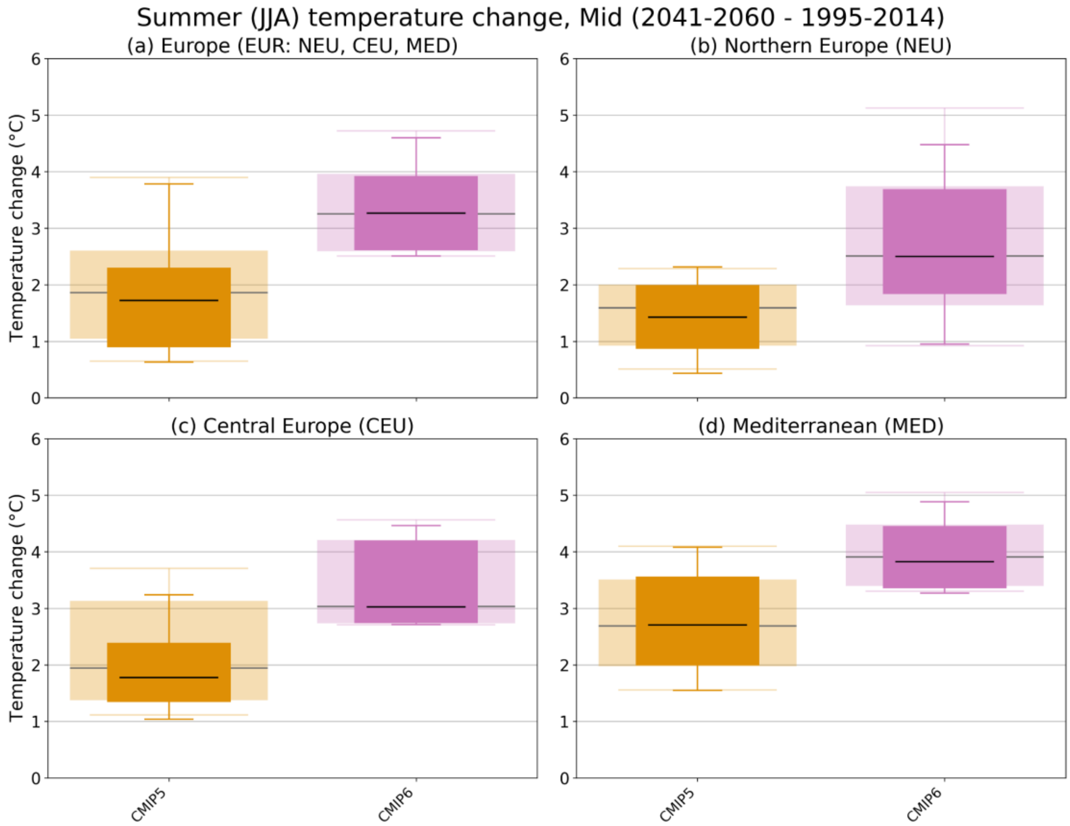


***Figure 5**: Summer (JJA) temperature change 2041-2060 relative to 1995–2014 for the combined European region (NEU-CEU-MED) as well as the three European SREX regions. The lighter boxes give the unconstrained distributions; the darker boxes give the constrained distributions. The colours indicate different ensembles: CMIP5 (orange) and CMIP5 (violet). The top (bottom) of the box represents the 75th (25th) percentile of the distribution and the upper (lower) whisker represents the 90th (10th) percentile.*

The difference between the two ensembles is more marked for the precipitation change. Figure 6 shows the summer precipitation change (%) obtained for the CMIP6 and CMIP5 ensembles for the same regions as in Figure 5. CMIP5- and CMIP6- median projected precipitation changes exhibit comparable values except for the Mediterranean region where the decrease of precipitation for CMIP6 is lower (-8%) with respect to that projected by the CMIP5 (-18%). The CMIP6 ensemble exhibits much wider uncertainty ranges over the Mediterranean region while the spread decreases for the Central European region and Northern Europe. This could be addressed to CMIP6's better representation of the storm tracks, as shown in Priestley et al., 2020, which thereby may reduce precipitation biases. For both ensembles the constraint method does not seem to have an impact and further investigation is needed for better understanding the reason.

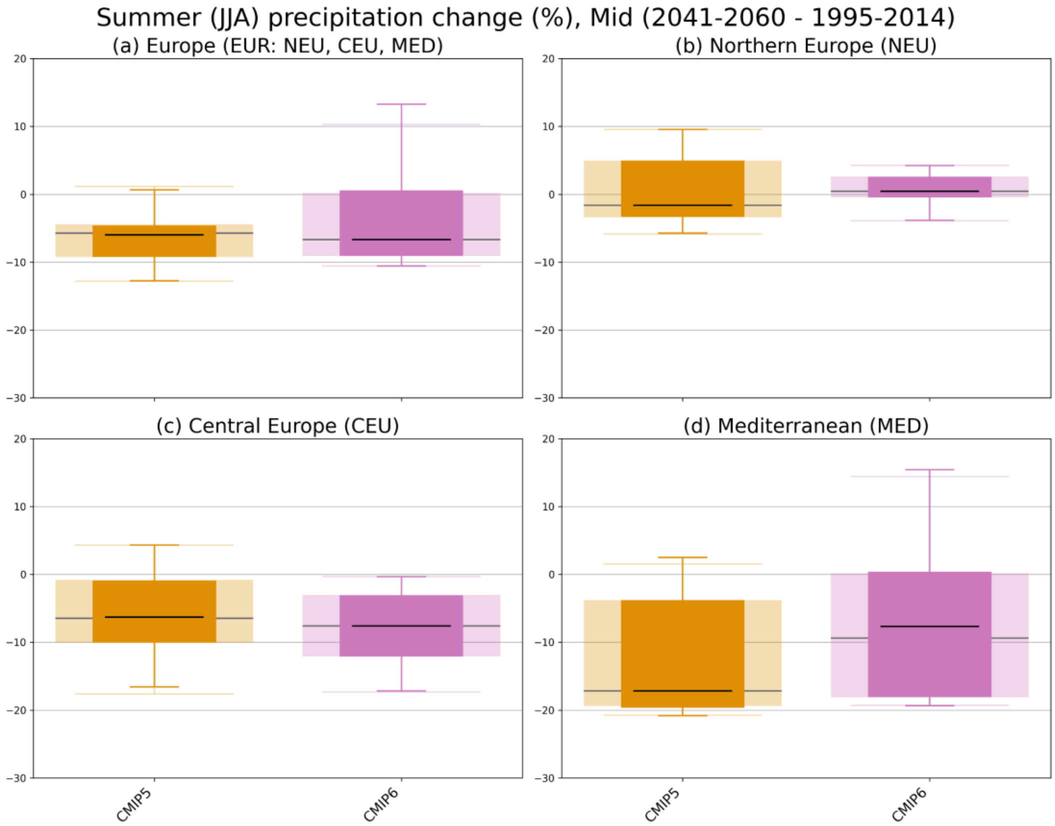


***Figure 6**: Summer (JJA) precipitation change (%) 2041-2060 relative to 1995–2014 for the combined European region (NEU-CEU-MED) as well as the three European SREX regions. The lighter boxes give the unconstrained distributions; the darker boxes give the constrained distributions. The colours indicate different ensembles: CMIP5 (orange) and CMIP6 (violet). The top (bottom) of the box represents the 75th (25th) percentile of the distribution and the upper (lower) whisker represents the 90th (10th) percentile.*

ETH Zurich

1. *Constraining global temperature projections from CMIP6*

Several CMIP6 models show considerably stronger global temperature increases than projected by the previous-generation CMIP5 models. The investigation and interpretation of the resulting higher global multi-model mean warming is relevant also for projecting European climate. ETH Zurich contributes a study which applies the Model Weighting by Independence and Performance (ClimWIP) approach (e.g., Brunner et al. 2019, 2020a, Merrifield et al. 2020) with the aim of constructing constrained distributions of global temperature change from CMIP6 (Brunner et al. 2020b).

Our main finding is that models, which show strong warming in future scenarios (SSP1-2.6 and SSP5-8.5), receive systematically lower weights based on their past performance and on their independence from other models (see Brunner et al. 2020b for more details). This leads to reduced mid-century (2041-2060) global warming by about −0.15 °C in the multi-model mean for both scenarios. The likely (i.e., 66%) model range is reduced by 36% and 19% for SSP1-2.6 and SSP5-8.5, respectively.

The study also investigates the skill of the weighting method using 'perfect model' approaches as suggested in Brunner et al. (2020a). The global mean Continuous Ranked Probability Skill Score (CRPSS) improves by 10%-20% using 27 previous-generation CMIP5 models as pseudo-observations. Relevant for European projections, skill consistently increases across Europe even when using global mean temperature change as target. Figure 7 shows a map of the CRPSS for the middle of the century under SSP5-8.5 (CMIP5 and RCP8.5 was used for the 'perfect models'). Weighted projections are more skilful than the unweighted baseline in all of Europe, with highest skill found in Central Europe and the Mediterranean. A more detailed analysis of the effect of global and region metrics to inform the weights can be found in the next section.

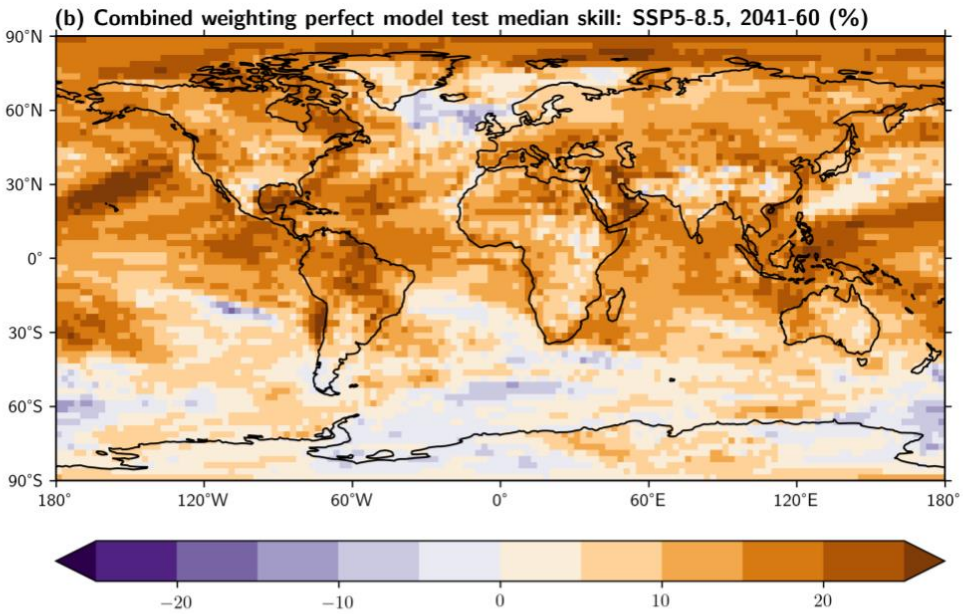


***Figure 7**: Map of the Continuous Ranked Probability Skill Score for weighting CMIP6 projections for the period 2041-2060 under SSP5-8.5. Shown is the median skill change across 27 perfect models from CMIP5 using RCP8.5. Reprinted from Brunner et al. (2020b)*

2. The effect of model weighting across CMIP5 and CMIP6

In earlier work ETH Zurich applied the ClimWIP method to CMIP5 projections of temperature and precipitation in different European sub-regions (see Brunner et al. 2019, 2020a, as well as deliverables D2.1 and D2.2). Work is ongoing to analyse how unweighted and weighted multi-model projections differ between CMIP5 and CMIP6. Special focus is given to the effect of the constraining using metrics identified as relevant for Europe and CMIP5 in Brunner et al. (2019): climatologies of precipitation, temperature, shortwave downwelling and upwelling radiation as well as standard deviation of upwelling shortwave and downwelling longwave radiation. As an additional metric temperature trend (tasTREND) is introduced and investigated because historical warming has been identified as highly relevant in CMIP6. Here we limit our discussion mainly to the differences between CMIP5 and CMIP6.

Figure 8 shows summer (JJA) temperature changes in four of the European sub-regions used in Brunner et al. (2019) based on the three European SREX regions: Northern Europe (NEU), Central Europe (CEU), Mediterranean (MED), and the combined European region (EUR: NEU + CEU + MED). Adding temperature trend as metric to inform the weighting leads to a somewhat diverse picture for CMIP5 (second and third box in each panel of figure 8): in NEU and CEU the regional temperature trend tends to widen the original constraining based on the metrics from Brunner et al. (2019; first box), while global trend predominantly leads to additional constraining in NEU and to a shift of the distribution in CEU. In MED adding regional or global temperature trend leaves the original constrained distribution mostly unchanged.

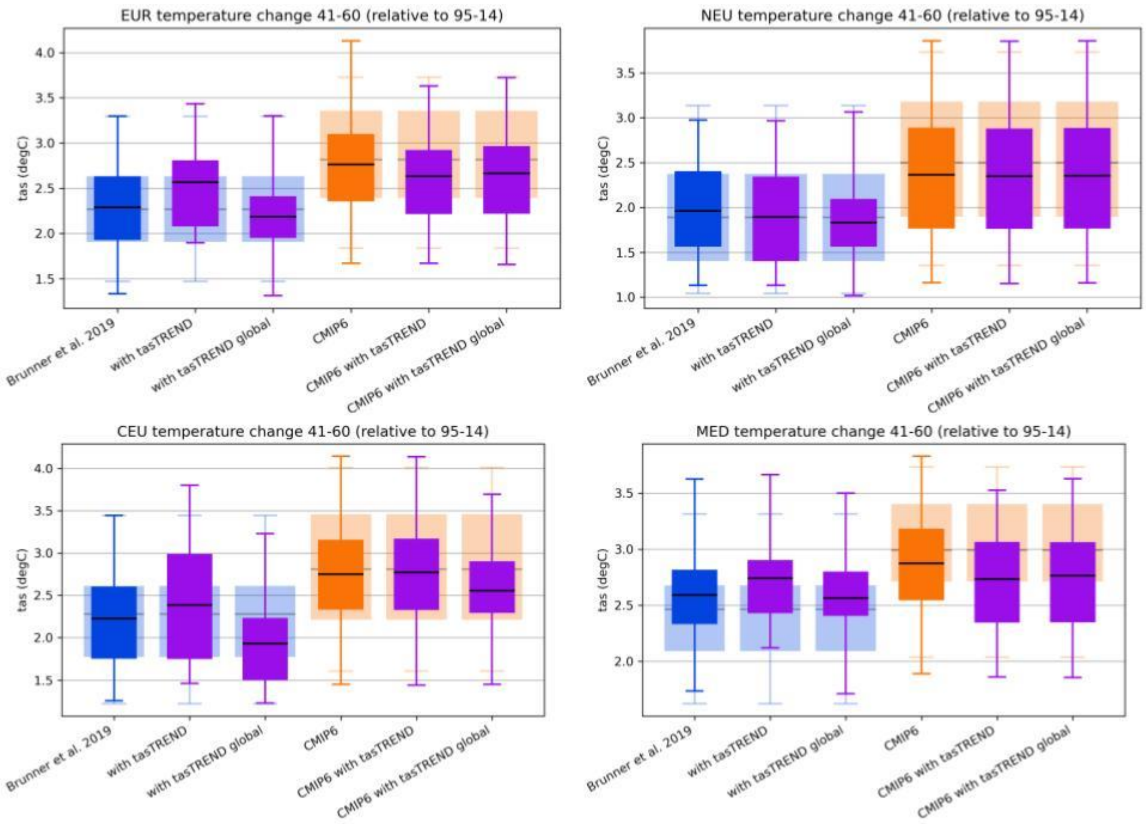


***Figure 8**: Distributions of summer (JJA) temperature change (2041-2060 minus 1995-2014) from CMIP5 (leftmost three boxes in each panel) and CMIP6 (rightmost three boxes in each panel). Shown is the unweighted distribution in the background (light blue for CMIP5 and light orange for CMIP6) as well as the weighted distribution in the foreground: dark blue/orange for CMIP5/6 constrained with the original metrics from Brunner et al. (2019) and violet for CMIP5/6 constrained using temperature trend in addition.*

For CMIP6 a more homogeneous picture arises with all three combinations of metrics (original, original + regional trend, and original + global trend) clearly constraining the upper end of the likely range for all four regions. Generally speaking, this means weighting brings the distributions from CMIP5 and CMIP6 closer together for most cases. More specifically, for MED, the region with the largest differences between the raw distributions of CMIP5 and CMIP6, the effect of the weighting leads to the strongest reduction in differences in particular when also considering temperature trend as a metric. The same holds for NEU and CEU when considering only local information (i.e., all cases except the ones drawing on global trend as metric).

Work is ongoing to further investigate the role of different metrics (including temperature trends) when constraining global and regional projections from CMIP5 and CMIP6 (see, e.g., section 8). Similar to the global case CMIP6 projects considerably more summer warming in Europe than CMIP5 also regionally without any weighting (see also Palmer et al. 2021). Our initial results suggest that weighting models can help to bring temperature projections from the two generations closer together depending on the region and metrics used.

University of Edinburgh (UEdin)
UEdin applies the ASK approach (Allen-Stott-Kettleborough, based on Allen et al, 2000 and Stott and Kettleborough, 2002). The method assumes that the true observed climate response (y_{obs}) to historical forcing(s) is a simple linear combination of one or more (n) individual forcing fingerprints (X_j). The fingerprints are scaled to match the signal in observations by their respective scaling factors (β_j), accounting for noise/uncertainty in both the observations (ε_{obs}), and in the modelled response to each of the forcings (ε_j):

$$y_{obs} = \sum_{j=1}^{n} \beta_j \left(X_j - \varepsilon_j\right) + \varepsilon_{obs} \qquad (1)$$

This method has been applied to constrain IPCC projections (Knutti et al., 2008) along with other methods. A recent application to global data of surface temperature and ocean heat content was used to estimate the equilibrium climate sensitivity from the historical period, and showed that some models with stronger warming may be outside the observationally consistent range of greenhouse warming (Tokarska et al., 2020)

In this study the method is applied to seasonal changes in Europe, disregarding global scale temperature change (which may help to better constrain warming) in order to derive constraints specific to Europe. The method utilises CMIP6 model simulations run with historical forcings, and Detection and Attribution MIP (DAMIP; Gillett et al., 2016) single-forcing simulations over the same historic period. For the future projections, historical simulations are extended with CMIP6 Scenario-MIP Shared Socioeconomic Pathway (SSP; Gidden et al., 2019) simulations. Observations are retrieved from the gridded E-OBS v19.0e dataset (Haylock et al., 2008), with monthly values computed from the daily data. The monthly surface air temperature fields from the observations, along with each of the CMIP6 model ensemble members, were spatially regridded to a regular 2.5° x 2.5° latitude-longitude grid.

The model fingerprints {XHist,XGHG,XNat} are constructed by taking an unweighted average of the multimodel ensemble members (the mean of individual model-ensemble means) from the all-forcings historical, and single-forcing GHG-only and Natural-only simulations, respectively. Each fingerprint comprises the conjoined annual time series of the multi-model mean over the three area-averaged European SREX regions (NEU, CEU, and MED), with each of the time series first normalised by an estimate of that region's internal variability (standard deviation from piControl runs). The different scaling factors are then estimated through a total least squares (TLS) regression. We explore three different linear combination of the model fingerprints:

$$
\begin{aligned}
&\text{a)} \quad y_{obs} = \beta_{a1}(X_{Hist} - \varepsilon_{Hist}) + \varepsilon_{obs}, \qquad\qquad (2)\\
&\qquad\qquad \{\hat{\beta}_{ALL} = \beta_{a1}\}.\\
&\text{b)} \quad y_{obs} = \beta_{b1}(X_{Hist} - \varepsilon_{Hist}) +\\
&\qquad\qquad \beta_{b2}(X_{GHG} - \varepsilon_{GHG}) + \varepsilon_{obs},\\
&\qquad\qquad \{\hat{\beta}_{GHG} = \beta_{b1} + \beta_{b2}\, ;\ \hat{\beta}_{OTH} = \beta_{b1}\}.\\
&\text{c)} \quad y_{obs} = \beta_{c1}(X_{Hist} - \varepsilon_{Hist}) +\\
&\qquad\qquad \beta_{c2}(X_{Nat} - \varepsilon_{Nat}) + \varepsilon_{obs},\\
&\qquad\qquad \{\hat{\beta}_{ANT} = \ \beta_{c1}\, ;\ \hat{\beta}_{NAT} = \beta_{c1} + \beta_{c2}\}.
\end{aligned}
$$

The greenhouse gas scaling factor estimate ($\hat{\beta}_{GHG}$ from Eq.2b) provides a simple scaling that can be used as a future constraint to CMIP6 projections; we label this application, ASK-GHG. Likewise, the estimate of the anthropogenic scaling factor ($\hat{\beta}_{ANT}$ from Eq.2c, which includes all forcings except those in the natural-only simulations) can also be used as a future constraint, labelled ASK-ANT. The all-forcings scaling factor ($\hat{\beta}_{ALL}$ from Eq.2a) is derived straightforwardly using only the historical CMIP6 simulations (not limited to DAMIP models), and hence incorporates a larger number of models into the fingerprint (reducing the noise); this constraint (ASK-ALL) was applied to precipitation projections in the perfect model study discussed below.

A confidence interval for each of the scaling factors describes the range of magnitudes of the model response that are consistent with the observed signal. A forced model response is detected if the range of scaling factors are significantly greater than zero and can be described as being consistent with observations if the range of values contains the magnitude of one. For this study the confidence intervals are estimated by adding samples (of the same length) from the piControl simulations to both the noise-reduced fingerprints and observations and recomputing the TLS regression (10,000 times) in order to build a distribution of scaling factors, from which the 5th-95th percentile range is computed. The uncertainty range arising from the scaling factor only consists of the fingerprint times the range of scaling factors but constrains the forced signal only. In order to arrive at a range that includes internal climate variability, it needs to be added on (which has been done here by Monte Carlo analysis).

UKCP

UKCP applies a Bayesian approach to produce probabilistic projections. In contrast to other methods that use the empirical spread of CMIP5 projections to represent prior model uncertainty, UKCP uses a statistical emulator trained on a single-model perturbed physics ensemble. This provides a more systematic and comprehensive sampling of climate responses by allowing a larger sample size in the emulated ensemble and structured sampling of uncertainties. Further, by basing the simulations on the emission-driven representative concentration pathway 8.5 (RCP8.5) scenario simulations (as opposed to the concentration-driven used in the other methods) and drawing from a second perturbed physic ensemble of Earth system model variants, this method samples additional uncertainties associated with the carbon cycle. This inclusion of additional uncertainties differentiates UKCP from the other methods described here. To further sample the structural error component associated with using a single perturbed physics model, CMIP5 Earth system model simulations are used to define an additional "discrepancy" term (Sexton et al. 2012). This methodology means that unconstrained distributions are wider than for the other methods.

Observational constraints are applied by weighting sampled outcomes by likelihood weights calculated from multivariate distances to observations. The observations comprise 12 climate variables reduced in dimensionality to 6 leading eigenvectors. In addition, historical trends for several climate indicators are also considered in the set of observational constraints, including the Braganza indices based on global mean surface temperature (Braganza et al. 2003), heat content change in the top 700 m of the oceans, and change in atmospheric CO2 concentration over a recent 45-yr period (Booth et al. 2017). The separate components of the method are validated and the additional statistical uncertainties that arise are included at each stage. These include equilibrium response emulation error, error in converting from equilibrium to transient response, time-scaling error (including inherent model internal variability), and structural error estimates.

A full description of the method is included in the Appendix

7. The Climate Projection Atlas

All contributing groups (see Tables 1 and 2) provided constrained climate projection maps for the 2041-2060 lead time based on the methods originally described Brunner 2020a, or updated applications described in Section 7. Here we show maps of the projected changes (which are available for potential users to access from the Atlas).

Figure 9 shows the constrained probabilistic projections from the CMIP6 ensemble (see Table 3), or state-of-the-art bespoke climate ensembles (CALL and UKCP methods), for the SSP5-8.5 scenario for summer temperature (June-July-August, JJA) averaged over the 2041-2060 period. All the methods indicate an increase in temperature over the whole of Europe, with a median warming over the Mediterranean basin of 3.5°C compared to the 1995-2014 period, while the median warming in Northern Europe is more moderate, with a 2°C change. Note that all methods tend to provide similar median warming patterns. Larger differences between methods are obtained regarding the tails of the probabilistic projections, i.e., for the 10th and 90th percentiles. As not all methods characterise the prior uncertainty in the same way (different model ensembles, inclusion of the carbon cycle in the UKMO method), the magnitude of the relative warming at the 10th and 90th percentiles partly depends on the prior spread of uncertainty. Temperature changes associated with the 90th (10th) percentile range from 4°C to over 5°C (0.5°C to 3°C). Differences in methodology also contribute to explain the discrepancies (see Section 8).

Similar results are obtained for the constrained winter temperature projections (December-January-February, DJF). Figure 10 shows remarkable agreement between the methods on the median warming, which is more pronounced over North-Eastern Europe, ranging from 3°C to 4°C depending on the location. The warming associated with the tails of the distribution differs more, for similar reasons as those found in European summer (differences in their prior distributions and methodological differences).

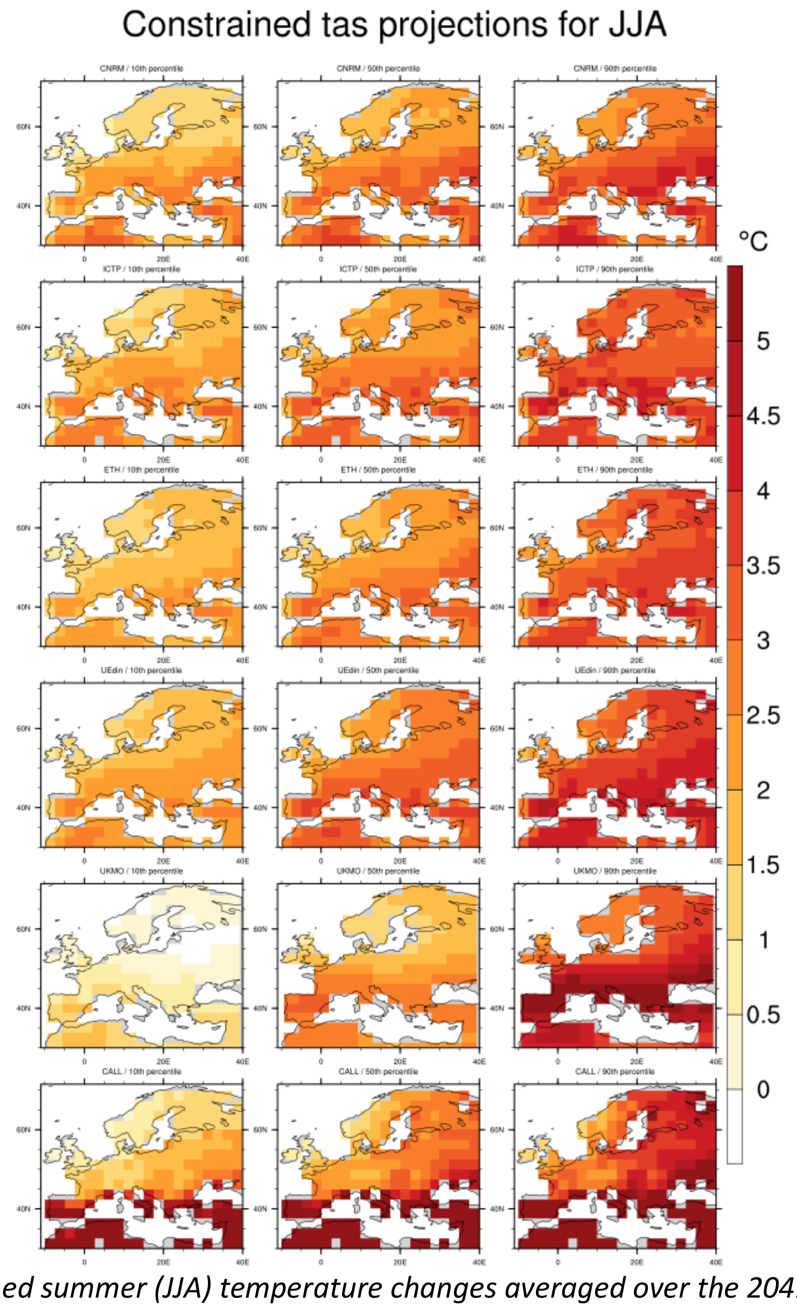


Figure 9*: Constrained summer (JJA) temperature changes averaged over the 2041-2060 period with reference to the 1995-2014 period. Models come from the CMIP6 ensemble (see Table 3) for the SSP5-8.5 scenario. Each row corresponds to one method. Each column corresponds to a given percentile (10th, 50th, 90th) of the distribution.*

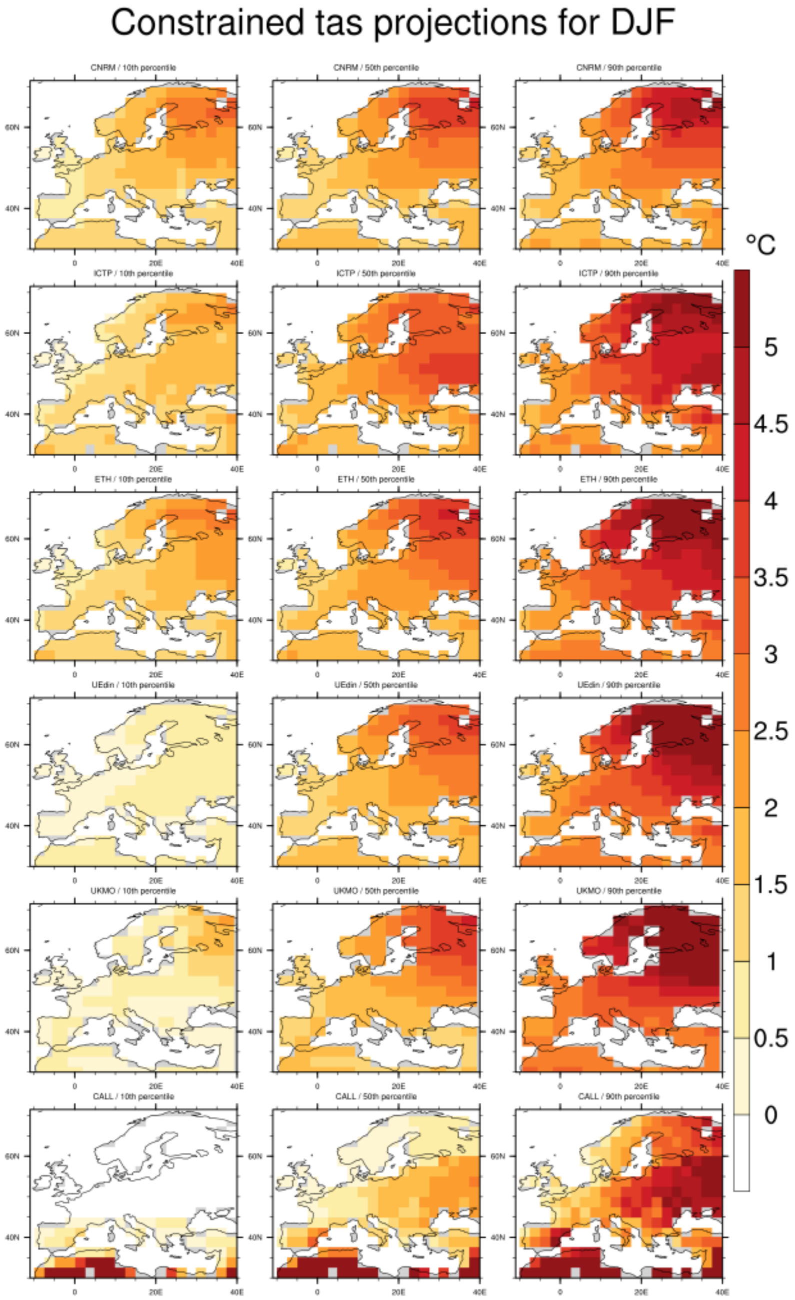


Figure 10*: Same as Figure 9 but for winter (DJF).*

Three groups provided constrained probabilistic projections for precipitation. In summer, the methods agree on a median projection with a relative decrease over the Mediterranean basin and central Europe of -10% to -20% and an increase of 10% over Scandinavia (Figure 11). The upper (lower) end of the constrained distributions indicates a generalised increase over the whole of Europe of between 10% and 50% (-10% and 60%). In winter, the median projections also indicate a dipole of precipitation anomalies, with an area of increase (decrease) north (south) of 45°N of about 20% (Figure 12). For the 90th percentile, an average increase of 30% is projected over the whole of Europe. For the 10th percentile, a decrease in precipitation of -20% to -40% is obtained over southern Europe depending on the methods.

As for temperature, the differences between the methods are partly explained by different prior distributions.

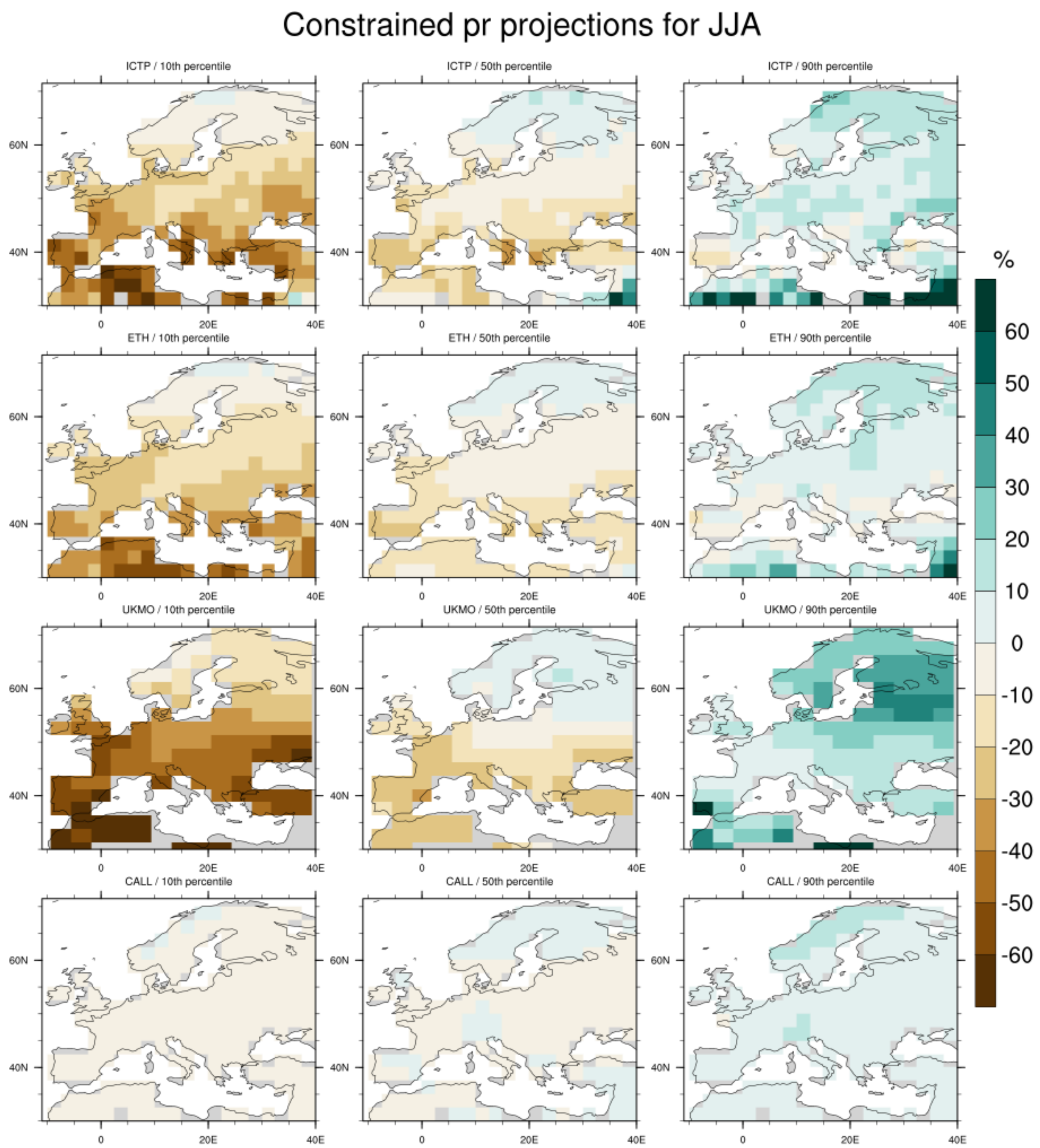


Figure 11*: Same as Figure 9 but for relative precipitation changes.*

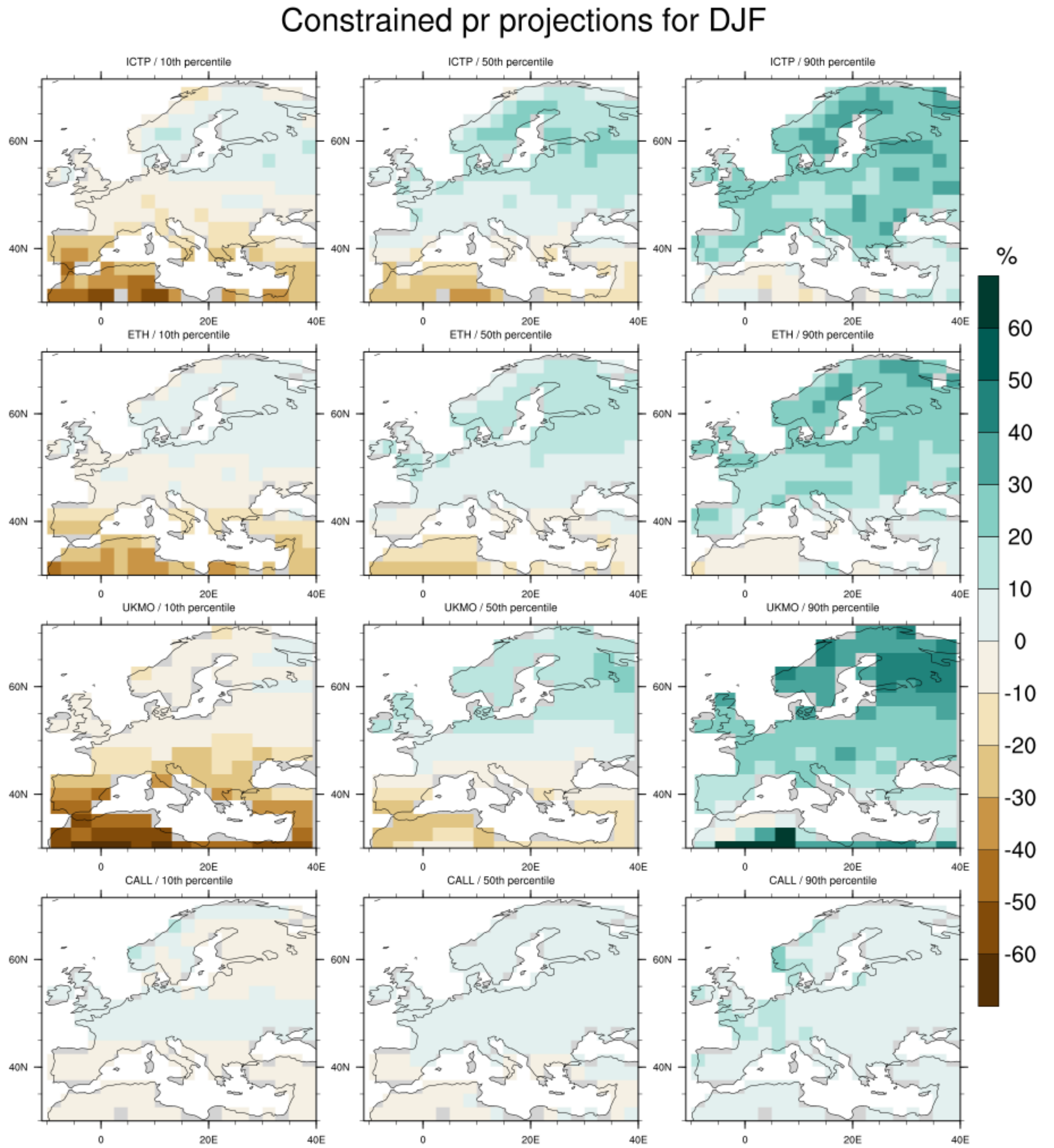


***Figure 12**: Same as Figure 9 but for winter (DJF).*

8. Assessing current probabilistic projections

As evident from the new Atlas of European Climate Projections, there are differences in projected changes produced by the different WP2 methodologies. In this section we have a more focused look at differences in the underlying approaches and, where possible, look at how these differences influence outputs.

One of the main challenges in providing climate projections is based on the ability of any given method to sample the current range of climate model responses. In the earlier focused inter-comparison of probabilistic methods (Brunner et al, 2020a), the participating groups worked very hard to standardise the climate simulations across the methods. We have not had that same luxury with the Atlas data, and so some of the differences in projection arise purely because different methodologies have managed to account for a different subset of potentially available models. This is an endemic challenge in making projections - in that quality control of underlying data; date in the CMIP model generation cycle that analysis starts; and even factors like relative stability of different regional ESGF data portals limits how extensive these samples are. We've summarised the CMIP6 models used in the Atlas here.

	CNRM/CNRS	ICTP	ClimWIP	UEdin
ACCESS-CM2	x		x	
ACCESS-ESM1	x		x	x
AWI-CM-1-1-MR	x		x	
BCC-CSM2-MR	x	x	x	x
CanESM5-CanOE	x		x	
CanESM5	x	x	x	x
CESM2-WACCM	x		x	
CESM2	x		x	
CIESM	x			
CMCC-CM2-SR5	x			
CMCC-ESM2	x			
CNRM-CM6-1-HR	x		x	
CNRM-CM6-1	x	x	x	x
CNRM-ESM2-1	x	x	x	
E3SM-1-1	x			
EC-Earth3-Veg	x	x	x	
EC-Earth3	x	x	x	
FGOALS-f3-L	x		x	
FGOALS-g3	x		x	
FIO-ESM-2-0	x		x	
GFDL-CM4	x	x		
GFDL-ESM4	x		x	x
GISS-E2-1-G	x		x	
HadGEM3-GC31-LL	x		x	x
HadGEM3-GC31-MM	x			
INM-CM4-8	x		x	
INM-CM5-0	x		x	
IPSL-CM6A-LR	x	x	x	x
KACE-1-0-G	x		x	
KIOST-ESM	x			

	CNRM/CNRS	ICTP	ClimWIP	UEdin
MCM-UA-1-0	x		x	
MIROC6	x	x	x	x
MIROC-ES2L	x		x	
MPI-ESM1-2-HR	x		x	
MPI-ESM1-2-LR	x		x	
MRI-ESM2-0	x	x	x	x
NESM3	x	x	x	
NorESM2-LM	x			x
NorESM2-MM	x		x	
TaiESM1	x			
UKESM1-0-LL	x	x	x	

Table 3: List of the available CMIP6 Models used by each group to constrain temperature projections from the SSP5-8.5 simulations. Note that the UKCP and CALL methods do not rely on CMIP6 models (see the description of these methods in Section 6). UKCP's prior distribution is derived from the UKCP perturbed parameter ensemble + a subset of CMIP5 models. CALL's prior distribution is derived from the CESM Large Ensemble.

In addition to the different prior distributions, alternative methodological approaches explain some of the differences between the constrained projections analysed in the Atlas section (Section 7). For each method, the main assumptions as well as the technical features are listed in Table 4. This section aims to document the impact of these features on the probabilistic projections, and how they can be compared with each other, although this constitutes a difficult exercise due to the different prior distributions used.

	UKCP	ClimWIP	ASK	REA	HistC	CALL
Assumes truth centered				✓		
Constrained range can lie beyond unconstrained range			✓			✓
Multiple estimates of observations are used in weights/constraint	✓	✓				
Spatial scale at which constraint or performance weighting is calculated	Global + large scale	Same as target	Europe	Local	Global + local	Same as target
Multiple variables used to weight each target variables	✓	✓				
Observation uncertainty	✓	✓				
Includes estimate of internal variability	✓	✓		✓		✓
Carbon cycle	✓					
Model uncertainty (parameter)	✓					
Model uncertainty (structural)	✓	✓	✓	✓	✓	
Method error	✓	✓				
Outputs are spatially coherent	✓	✓	✓			
Outputs are physically coherent	✓	✓				

Table 4: Key characteristics of the different methods. Reprinted from Brunner et al (2020a).

We briefly highlight and discuss some of these methodological design decisions and what their implications are for the projections.

- UKCP and ClimWIP methods both use multiple variables (i.e. other than just a single variable, be that temperature or precipitation) to constrain temperature and precipitation projections. This approach is useful for identifying physically coherent responses across variables, as discussed in Section 10.3

- Some methods constrain variables at different spatial scales to derive constrained temperature or precipitation projections locally. At one end of this spectrum, UKCP uses observations of global climatological fields, ocean heat content and CO2 trends as well as trends of large regional temperature data. So, a good projection under the UKCP framework is a measure of global performance, which may or may not be strongly related to the local changes in Europe. At the other end of this spectrum, REA is based solely on local observations, and as such is likely to minimise the bias locally (Figure 13) but may have more difficulty in providing a physically and spatially consistent climate projection at larger spatial scales. The ASK implementation in the Atlas, uses trends over the wider European region and so sits somewhere between the two. The Atlas implementation of both ClimWIP and CALL makes use of European wide observations. KCC's Atlas implementation uses local observations. However, both ClimWIP and KCC Atlas constraints include information on Global Mean Surface Temperature (GMST). A direct implication of including larger scale information is that these methods provide climate projections characterised by more consistent spatial structures over larger regions of Europe. As an example, several papers have shown that constraining projected GMST by observations lead to a downward revision of the projected global warming (Ribes et al 2021, Tokarska et al 2020, Brunner et al. 2020), and as European temperature is highly correlated with GMST, the inclusion of GMST in these methods leads to a downward revision of warming over Europe (Figure 13 and 16), especially the Northern regions which appear to be sensitive to this metric (see Figure 8 from ClimWIP and Figure 15). Depending on how the local constraint is treated in addition to the large-scale constraints, some projections may nevertheless be locally biased because they do not give sufficient weight to in situ observations. For an end-user and his/her specific needs, these aspects are fundamental and condition the choice of one or more methods.
- All methods except CALL and UKCP sample the uncertainty (i.e. the inter-model spread) from models derived from CMIP5 or CMIP6. These models are often qualified as "ensemble of opportunities", which poorly sample climate uncertainty (Tebaldi and Knutti, 2007). The UKCP method includes, in addition to a sample of CMIP5 models, an ensemble of perturbed physics simulations that quantify parameter uncertainty, particularly associated with the carbon cycle. This additional uncertainty contributes to explain a larger inter-model spread in the prior distributions compared to the other methods. As a consequence, the associated constrained projections are also different, especially for the tails of the probabilistic projections. This method might therefore be suitable for a user who is sensitive to rare climate events.
- The REA and ClimWIP methods are based on weighting historical and future climate simulations according to (i) their performance regarding the historical observed climate. ClimWIP also accounts for (ii) the level of dependence between the CMIP models from which these simulations were derived. This approach results in the observationally constrained probabilistic projections always being a subset of the unconstrained projections. The UKCP and KCC methods are based on Bayesian statistical techniques that calculate a probability of occurrence (e.g. of future warming) conditional on the observations. This approach leads to a spread in the probabilistic projections that is generally either reduced or at worst identical to the prior spread.

- The CALL and ASK methods are methods of scaling simulations on past observations with regression factors that can lead to projected future changes being outside the a priori simulated changes. The cases for which projected changes fall outside the unconstrained ranges (i.e. ASK, CALL) present potential strengths and reasons to be cautious about how these changes should be interpreted. The strength of these methods is that they can potentially compensate for real-world processes that are absent or not well represented in the simulations. For example, there are cases where observed historical precipitation changes are larger than those simulated. In these cases, the methods may be better able to compensate for any simulated underestimation of the actual climate response, as long as those biases persist into the future. On the other hand, caution is also required as the available simulations represent the range of responses arising from our current physical understanding of these processes, and as large ranges can arise from ASK if the signal-to-noise ratio is not yet sufficiently high to provide a strong constraint. Thus, methods that adjust projected changes outside these ranges may provide responses that may be less physically realistic. Generally, we recommend additional analysis particularly in cases where the range is shifted outside the unconstrained model range to a large extent (e.g. supporting significantly larger or smaller signals).

The analyses of change of constrained minus unconstrained projections shown below do not contain the ASK and CALL results, as in that case the constraint arises from the magnitude of the Europe-wide signal only and hence its uncertainty composed of magnitude uncertainty on the signal with added noise can lead to misleading impressions of the geographical pattern.

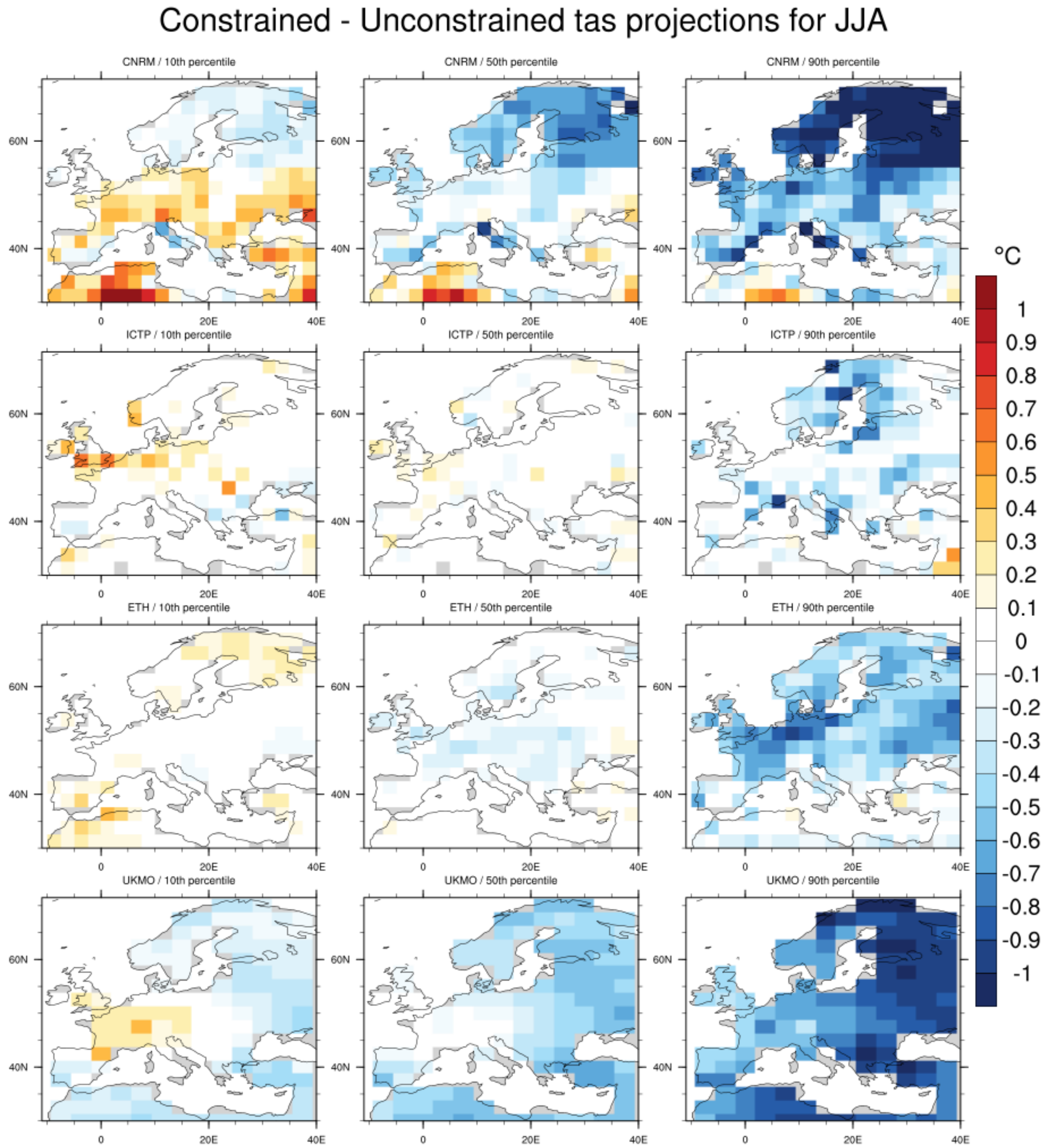


Figure 13*: Difference between the constrained and unconstrained summer (JJA) temperature projections averaged over the 2041-2060 period with reference to the 1995-2014 period. Each row corresponds to one method. Each column corresponds to the difference between percentiles of the distribution.*

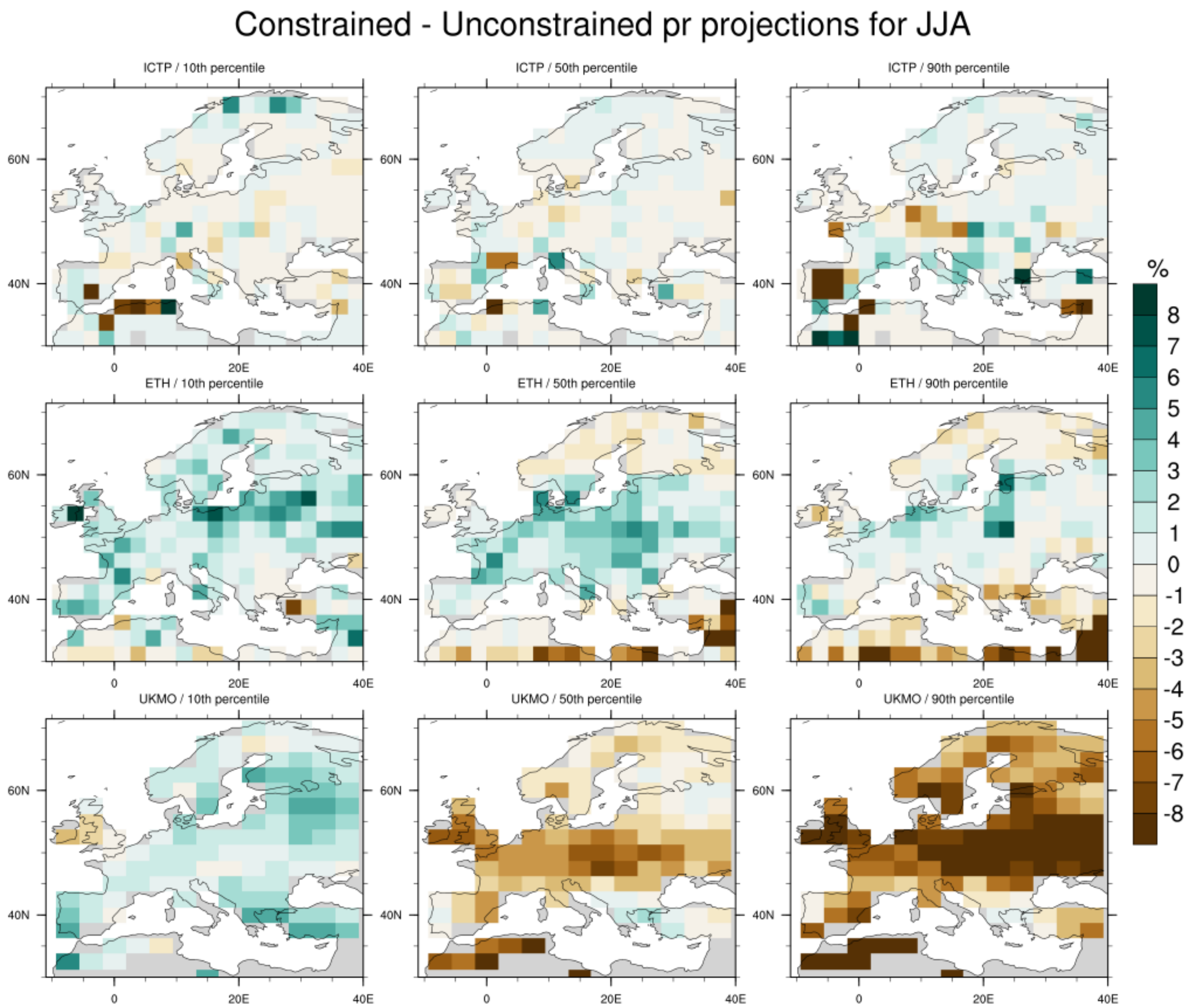


Figure 14*: Same as Figure 13 but for precipitation.*

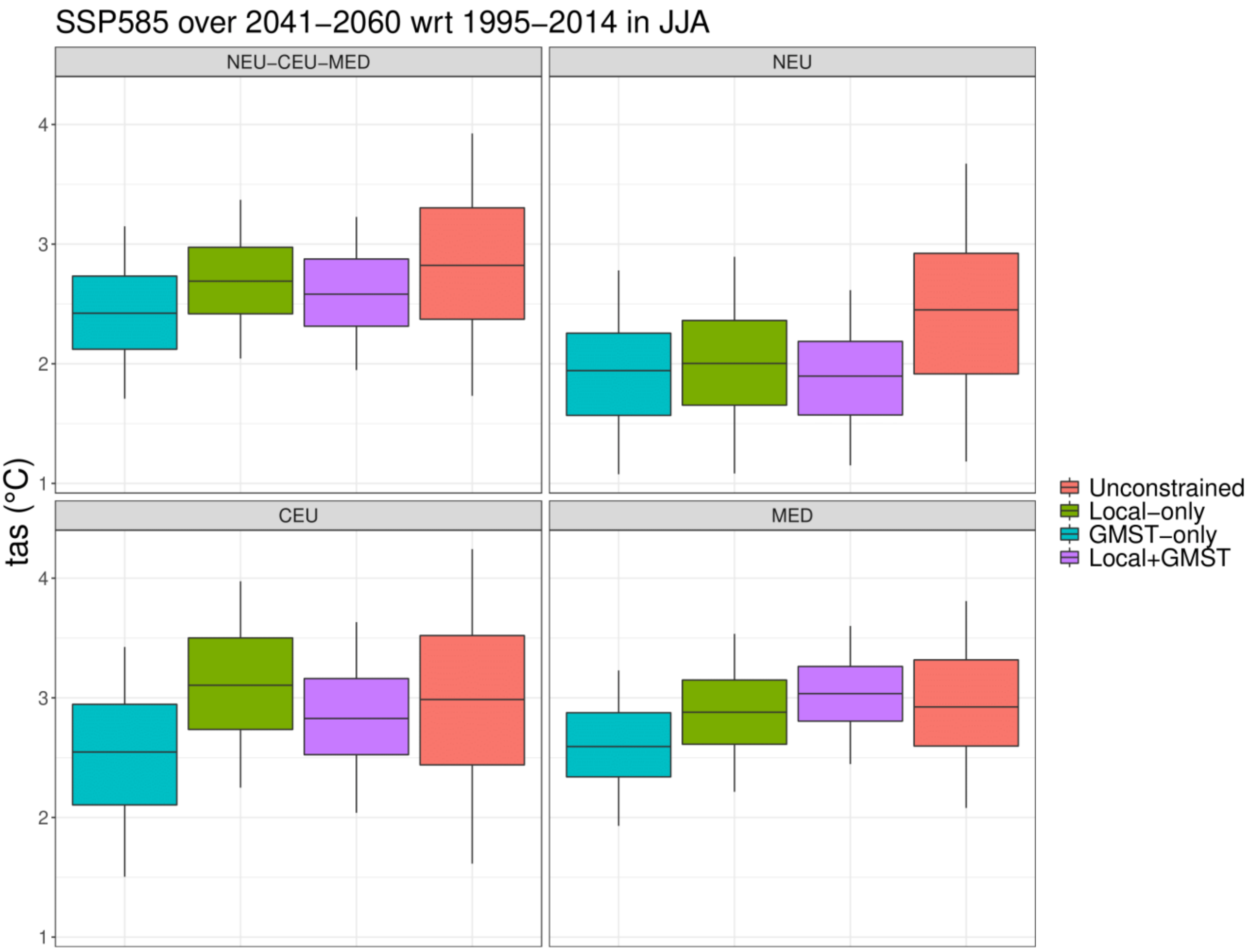


***Figure 15**: Summer (July–August) temperature change 2041–2060 relative to 1995–2014 for the combined European region (NEU-CEU-MED) as well as the three European SREX regions. Red boxplots correspond to the unconstrained distributions. The blue, green and magenta boxplots correspond the constrained distributions using the KCC method based on the use of local temperature time series only, global temperature time series, and both time series, respectively. The top and bottom of the box represents the 25th and 75th percentile of the distribution, respectively. The upper and lower whiskers represent the 90th and 10th percentiles, respectively. This is to be compared with the analogous figure 8 from ETH Zurich.*

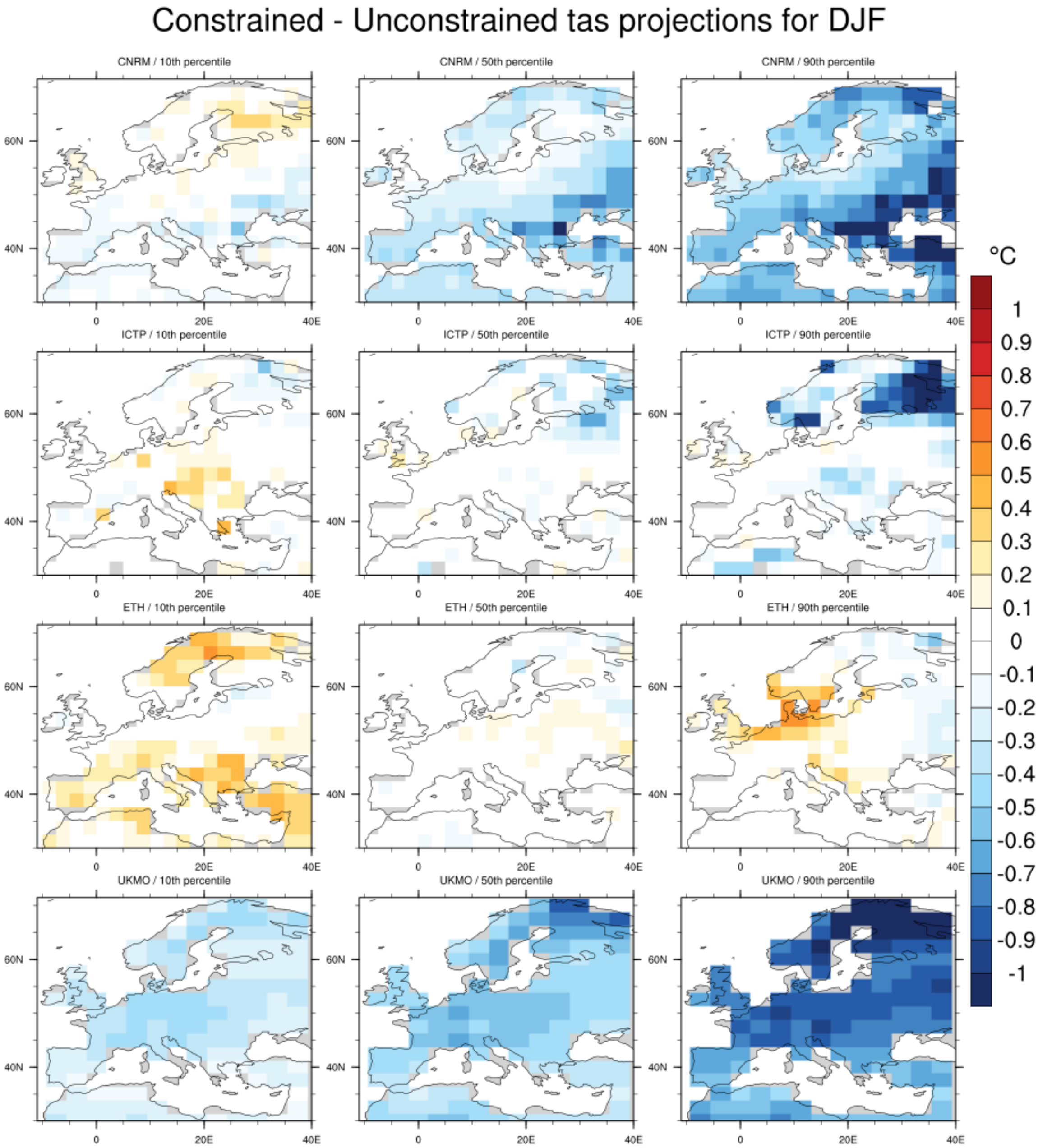


***Figure 16**: Same as Figure 13 but for winter (DJF).*

9. Understanding and using our constrained projections

9.1 Perfect model analysis

A major achievement of EUCP work-package 2 has been to develop and bring together a number of different methods for constraining future climate projections using the observational record. In an earlier study (Brunner et al., 2020) these different methods were compared using a consistent testing framework (i.e., variable, region, time period etc.). Whilst these constrained projections displayed some overlap and common characteristics, there were also some distinct differences. However, how these different projections should be interpreted is not clear. For example, are some of the methods more accurate and reliable than others?

To address this question, we designed an out-of-sample "perfect model" experiment to test the different constraining methods. In this experiment we took advantage of newly available CMIP6 archive of coupled climate simulations to act as pseudo-observations, to test the methods which were all previously used CMIP5-era model data; for this reason, we refer to this as an out-of-sample test. The pseudo-observational datasets were produced by regridding the required variables from the CMIP6 model simulations and restricting them to a common observational period (see schematic in Figure 17). This data was then uploaded to the Zenodo online repository and the different groups used these pseudo-observations to constrain their CMIP5-era model projections. For this test we used the historical and RCP 8.5 scenario experiments from the CMIP5-era models and the historical and SSP5-8.5 scenario for the pseudo-observations.

Whilst we refer to this analysis as a perfect model study, in some senses it represents a steeper test of the various methodologies. The pseudo-observations are drawn from CMIP6, an ensemble that includes better representation of key processes (such as super-cooled cloud droplets and wider representation of aerosol-cloud interactions, for example) than the CMIP5 ensemble that most methods use as the basis for their projections. In addition, the pseudo-observations include those from a number of CMIP6 simulations which fall outside (on the warm end) of the CMIP5 ranges (see Palmer et al, 2021). These factors led to a tougher test of the methods than the "perfect model" name for this approach usually implies, and is more along the line of 'imperfect model tests' (e.g., Schurer et al., 2018). The assessment was done in this way, as it goes part of the way to replicating some of the differences between the real world and the necessarily simplified representations that we use in climate projections. At the same time several of the 'perfect models' from CMIP6 are direct successors of their CMIP5 ancestors and are therefore not entirely independent (Brunner et al. 2020b; Brunner et al. 2022 - in preparation) as was shown to be the case for CMIP3 and 5 (Knutti et al. 2013).

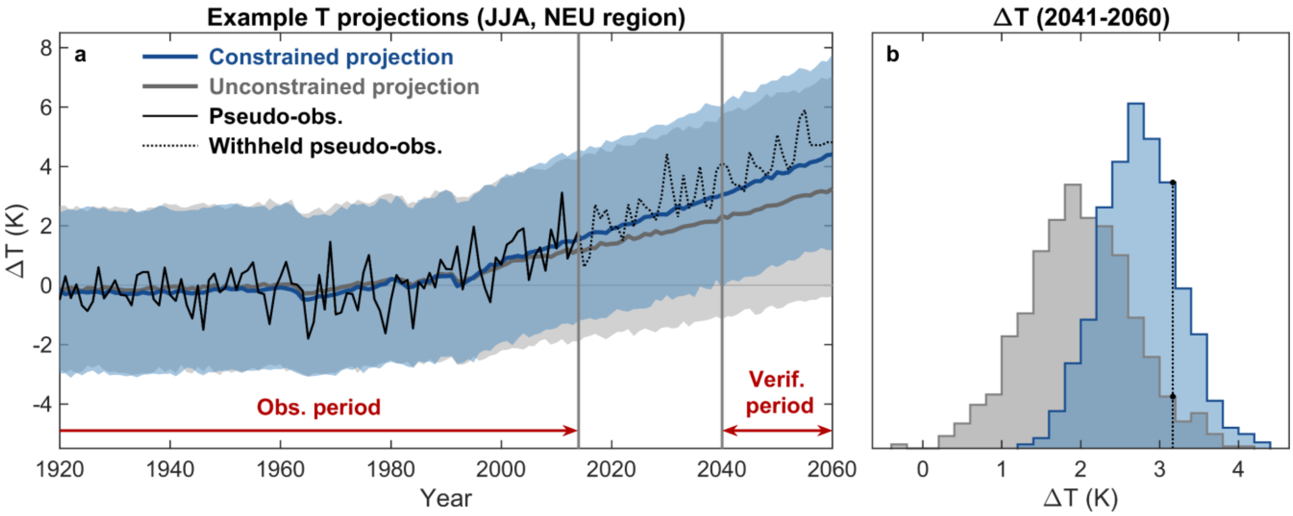


Figure 17*: Schematic of the out-of-sample "perfect model" analysis of the observational constraints. Pseudo-observations are taken from the observational period (up to 2014) and data from the future verification period was withheld prior to analysis of the results. The ability of the constrained and unconstrainted projections can then be compared to the withheld future response (right panel).*

An example of projected changes for the pseudo-observational datasets is shown in Figure 18. Whilst this plot is somewhat overwhelming in detail, it is shown here to demonstrate the type of data that has been produced in the collaborative activity. For each of the 126 pseudo-observational datasets, each of the methods have provided a probabilistic range of projections for the 2041-2060 climatology with respect to the 1995-2014 baseline period - shown here for summer surface-air temperature in the Northern Europe (NEU) region. The range of the projections is shown by the box-whisker plots and for each method the unconstrained (i.e., underlying) projection shown in lighter colours and the constrained projection is shown in darker colours. Also shown are the verification from the actual pseudo-observational dataset in each case (black horizontal lines). The most striking thing about Figure 18 is the diversity in predictions across the different methods and the different pseudo-observational datasets. Some of the methods seem to show larger changes, such as method B here, whereas other are generally fairly modest, such as Method A.

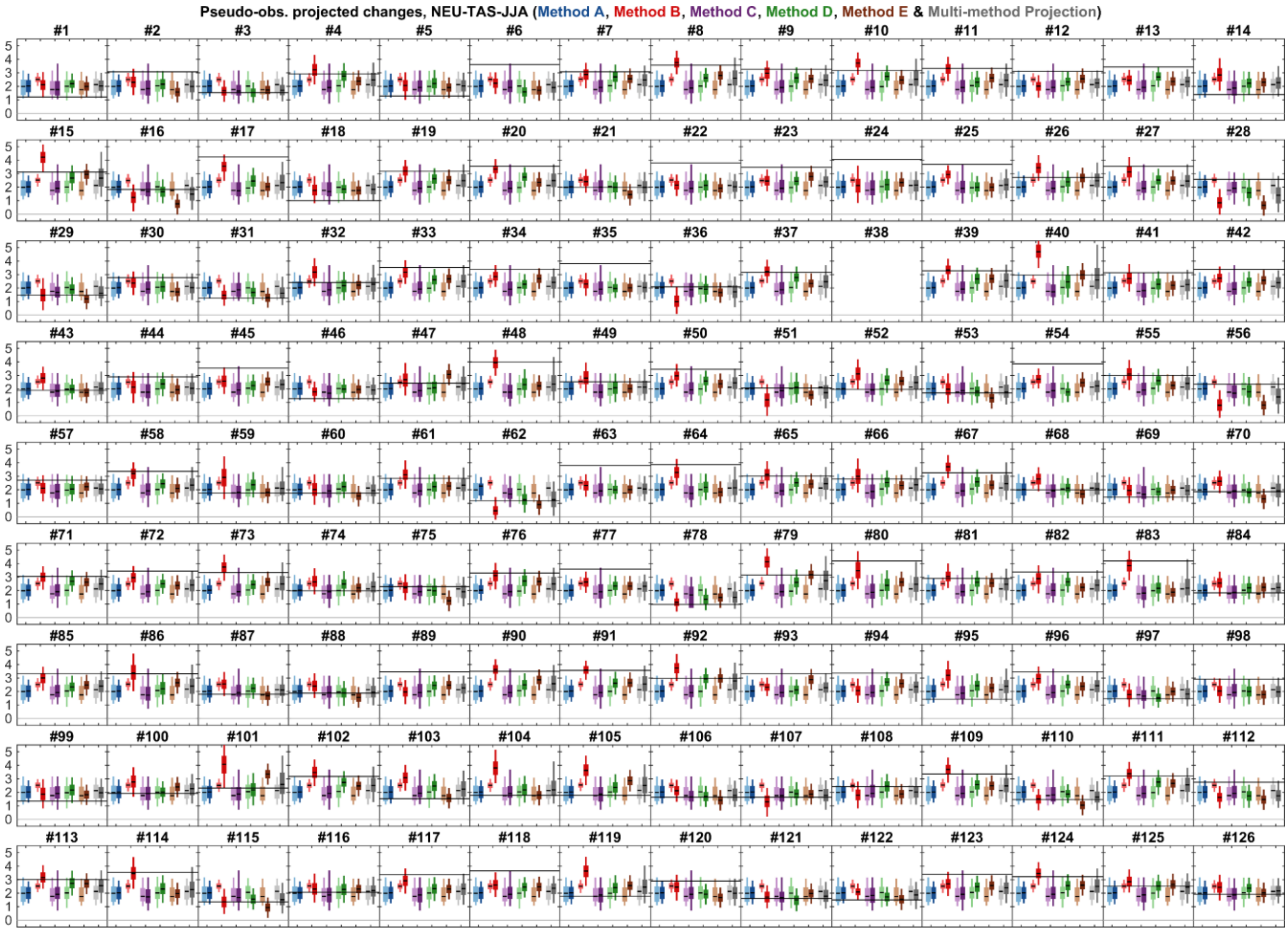


Figure 18*: Predicted changes for 2041-2060 summer surface air temperature in the Northern Europe region (i.e. the "NEU" S-REX region). Probabilistic predictions are shown for five of the contributing methods from WP2, along with the multi-method projection. The box, whiskers and centre line show the interquartile range, 5-95% range and the median of the projected change, respectively. The unconstrained projections are shown in lighter colours and the constrained projections are shown in darker colours. The horizonal black lines in each panel show the actual change for each pseudo-observational dataset, which was withheld during the constraining phase.*

To establish the accuracy and reliability of the projections we have employed a number of verification metrics, which are shown in Figure 19 for the summer surface-air temperature in the European SREX regions. The verification metrics shown are:

- Root-mean square error (RMSE), which measures the accuracy of the ensemble mean projection with respect to the actual future change in the pseudo-observations.
- Spread/Error, which measure the spread across the future change (measured by the standard deviation of the projected changes) with the ensemble mean error. For temperature, all methods (both constrained and unconstrained) are over-confident (with a ratio of less than one). By construction a subset of the perfect model data is chosen (a priori) to be outside the unconstrained ranges of the methodologies used here - which contributes to this apparent over-confidence. We would expect the methods not to show as pronounced limitation, should the real world fall within their unconstrained ranges.
- Continuous Ranked Probability Score (CRPS) which is measure of the accuracy of the probabilistic projections - this is lower when the projections are more accurate.

There is a lot of detail in the verification statistics shown for temperature projections (in Figures 19 and 20) and for rainfall (Figure 21). These plots provide an objective measure of the methodological skill and show that any benefit is very much variable dependent (depending on whether it is for temperature or precipitation).

For summer temperature, all of the constraining methods act to improve the projections compared to their respective underlying unconstrained projection. For example, all methods act to reduce or maintain the RMSE in Northern and Central European regions. Similarly, Spread/Error metrics can be improved, and the methods (largely) produce better Continuous Ranked Probability Scores (CPRS). This analysis suggests that some methods provided better skill than others, but all methods showed improvements over using raw distributions of climate models.

There is a different picture for Mediterranean climate projections. Two of the four methodologies showed worse skill than the raw climate model data - one substantially worse. This is problematic, as it suggests that the constrained Mediterranean responses is actually worse that just taking the off-the-shelf climate projections for these regions. The other two methods showed marked improvements in skill, however. At this stage, it is unclear which aspects of the observationally based constraint led to the degradation in skill in two of the methods. This is an aspect that will warrant further investigation.

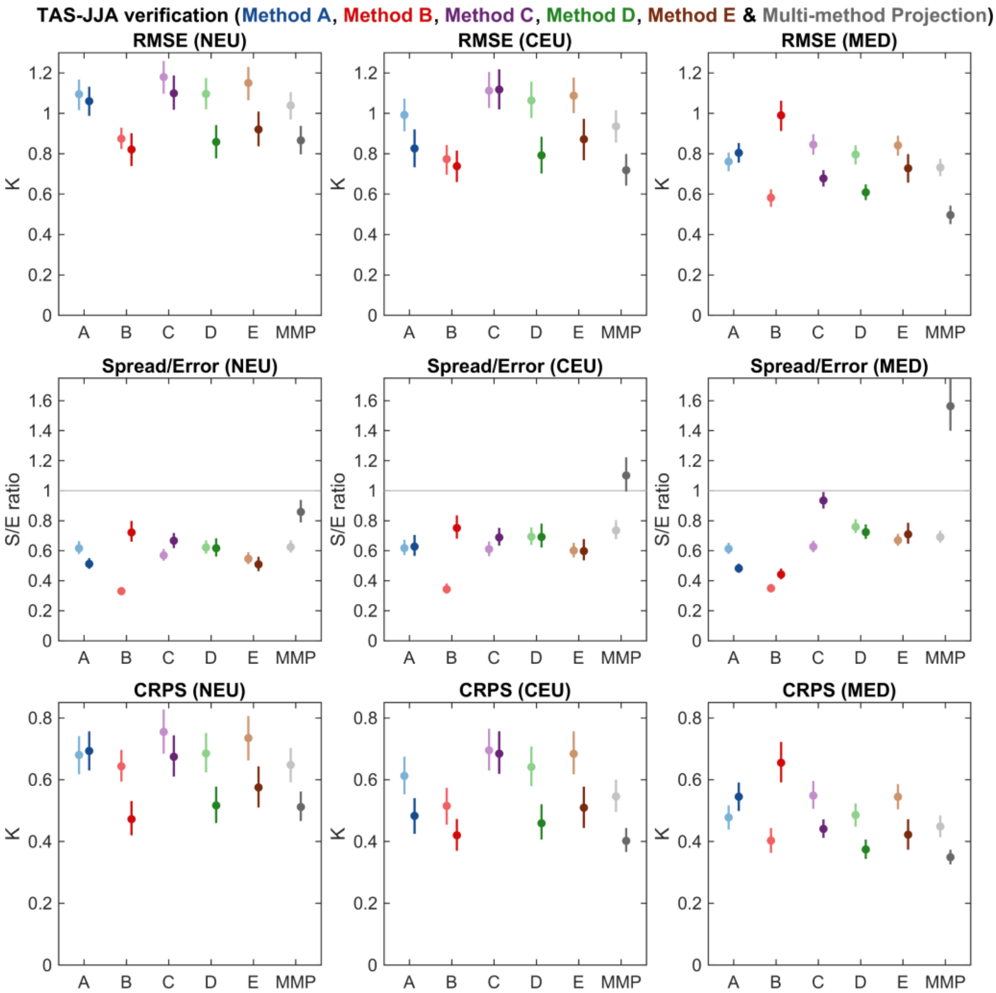


Figure 19*: Verification of the projections applied to the 125 pseudo-observational datasets for the summer surface-air temperature in the three SREX regions covering mainland Europe. Shown are the Root-mean square error (RMSE; top row), spread/error ratio (second row) and continuous ranked probability score (CRPS; bottom row). The unconstrained projections are shown in lighter colours and the constrained projections are shown in darker colours. The vertical bars indicate the 95% confidence interval of the statistics based on a bootstrap resampling (with replacement) of the 125 prediction/verification pairs.*

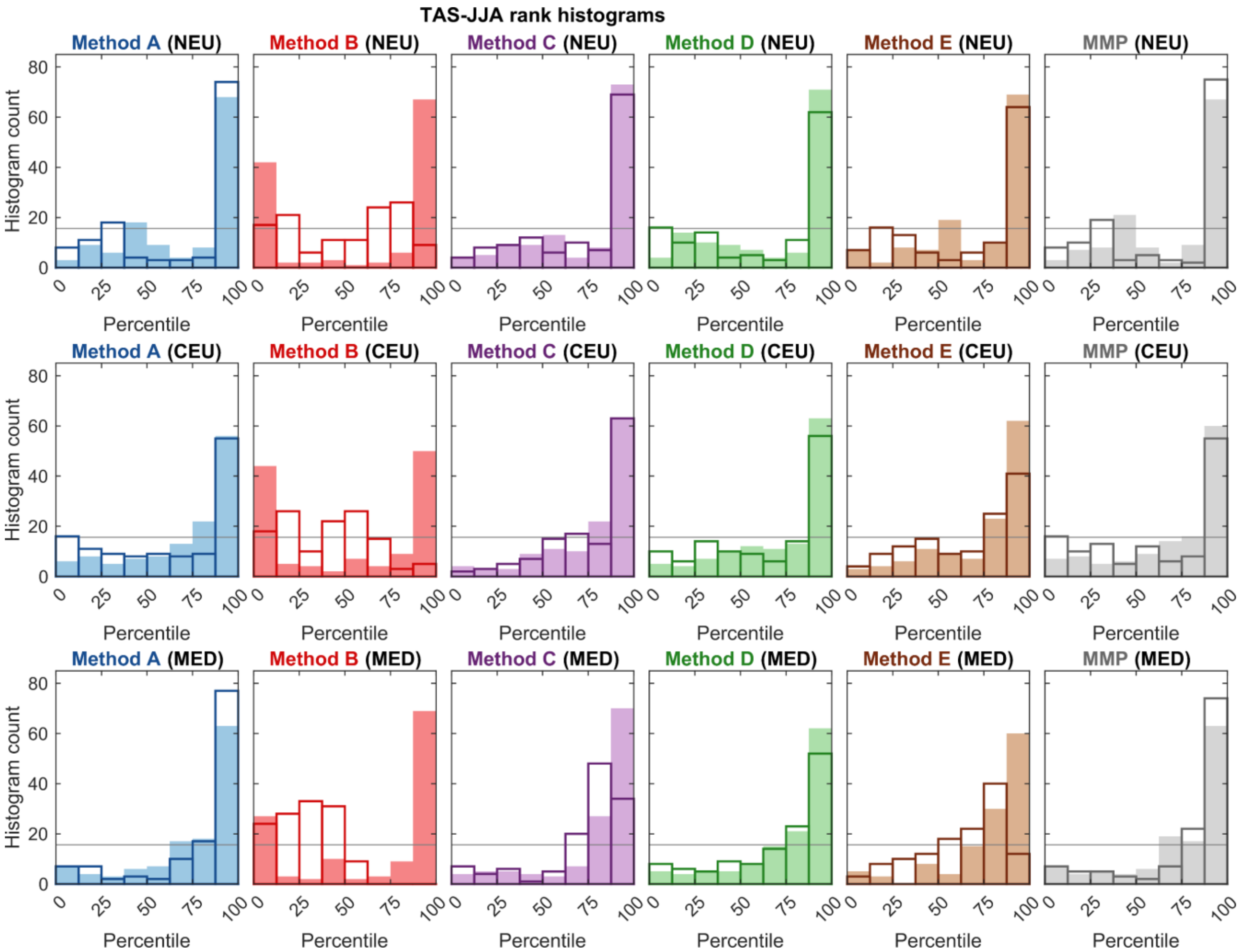


Figure 20: *Rank-histograms of the projections applied to the 125 pseudo-observational datasets for the summer surface-air temperature in the three SREX regions covering mainland Europe. The unconstrained projections are shown in solid lighter colours and the constrained projections are shown in the darker outlined histograms. The horizontal grey lines indicate the number of counts expected for a flat rank-histogram, which would be considered a perfectly reliable probabilistic prediction in this metric.*

For summer rainfall projections, no single method shows consistent improvement in skill (in any of the three metrics illustrated here) for all three European regions. There is some suggestion that the methodologies reduce the under-confidence (Spread/Error) of the Northern European projections (which is encouraging) but the Central European and Mediterranean projections show no clear improvement or degradation across the four methodologies. The best that can be said about the regional rainfall projections is that at least the observational constraints do not make the projection worse, in any obvious and systematic way across all metrics and regions.

Precipitation is a more challenging variable to make projections for due, in part, to its much larger variability. Climate change signals take a longer time to emerge both in terms of historical records and future changes. The skill assessed here is based on mid-century projections. As such this lack of any strong skill may not be reflected in end of century constrained rainfall projections. Nevertheless, the lack of skill in producing better rainfall projections compared to increased skill in temperature projections, is interesting insight. We have not had the tools before (certainly across a range of potential methodologies) to assess the relative benefits of observational constraints for different variables. This suggests that observational constraints (at least for some methodologies) have a strong benefit for projections of regional temperature (and variables strongly tied to this) but has a less obvious benefit to projections of regional precipitation change.

One notable aspect of the verification of the calibrated projections for European summer is the performance of the multi-method projection (MMP). The MMP the average of the projection data provided from the five methodologies considered here. In almost all metrics and regions the MMP performs better than the majority of the individual calibration methods - which is the case for both temperature and rainfall. This is particularly notable in the RMSE and CRPS. The reason for the relatively impressive performance is not yet completely clear but there is some understanding that we can glean from the results here. Since the methods are all different and these constraints demonstrate differing levels of improvement across the different regions, we can conclude that the constraints of the methods are somewhat independent. As the different methods mostly improve the accuracy of the projections, combining these independent improvements results in greater levels of accuracy that most of the individual methods. In addition, the reliability of the MMP is typically improved with respect to the individual methods. Specifically, most of the individual methods are found to be overconfident (spread/error < 1) even after constraining, however, when combined in the MMP the overconfident projections produce a broader probabilistic projection that is more reliable in the NEU and CEU regions. Similar findings have previously been demonstrated in the context of multi-model seasonal forecasting ensembles, that are typically overconfident but are more reliable and skilful when combined (Hagedorn et al., 2015), but this is the first such demonstration in the context of longer term climate projections. A notable exception for the improvement of the MMP is the spread/error in the Mediterranean region, which might be linked to the very poor performance of method B in this region, but more analysis is ongoing with respect to this issue.

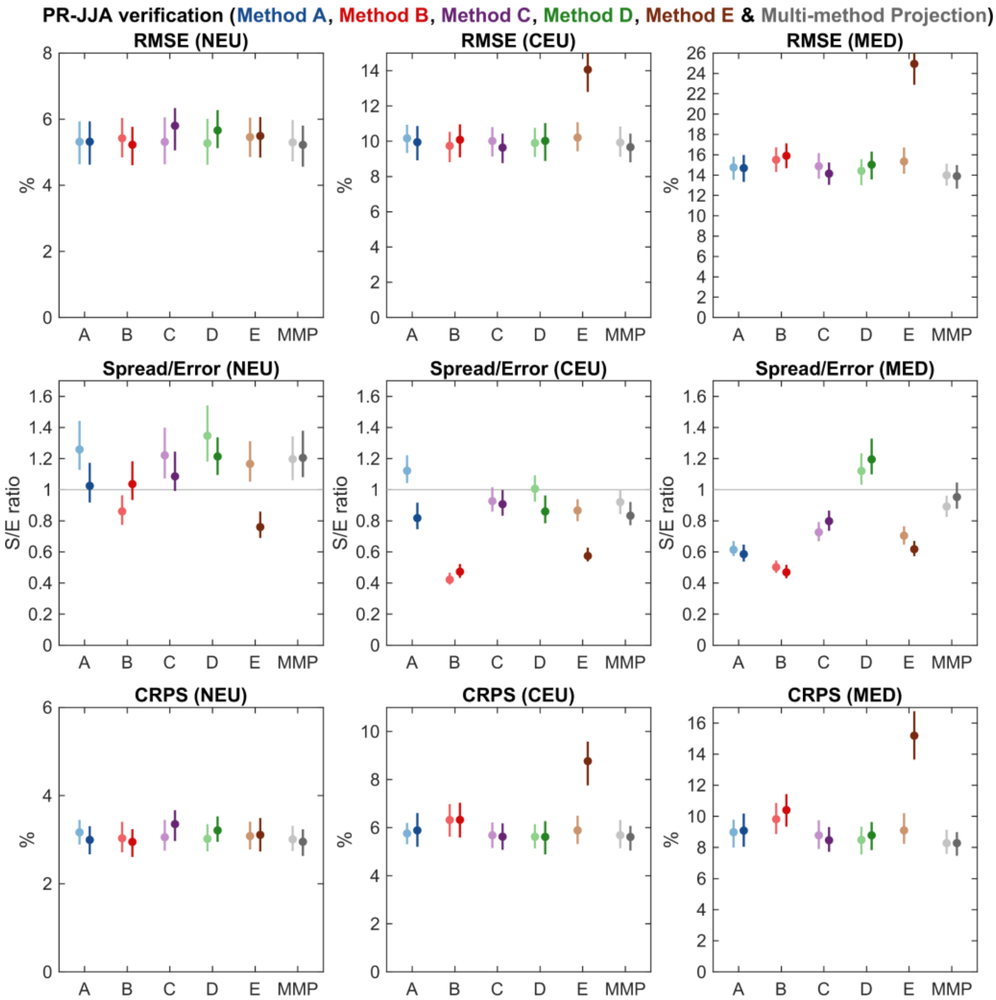


Figure 21*: Verification of the projections applied to the 125 pseudo-observational datasets for the summer precipitation in the three SREX regions covering mainland Europe. Shown are the Root-mean square error (RMSE; top row), spread/error ratio (second row) and continuous ranked probability score (CRPS; bottom row). The unconstrained projections are shown in lighter colours and the constrained projections are shown in darker colours. The vertical bars indicate the 95% confidence interval of the statistics based on a bootstrap resampling (with replacement) of the 125 prediction/verification pairs.*

The improvements in the MME for rainfall (whilst modest and not uniform) is encouraging, given the lack of improvement in projection skill found in individual methodologies. The RMSE and CRPS errors tend to improve at the expense of a small increase in over-confidence. Whilst the causes of the improvements are challenging to establish, the improvement provides encouragement that more skilful mid-term (30-40 year) rainfall projections are possible and worth working for.

9.2 Constraining extreme precipitation projections

Heavy rainfall events have important human and economic impacts over the European continent, where pluvial flooding is one of the main natural hazards (Llasat et al. 2013). The concern about the potential climate change impacts on extreme events over Europe is increasing. Many studies show an increase of precipitation extremes associated with the temperature increase in future climate projections (Min et al 2009, Donat et al 2016, Coppola et al 2020). While models mostly agree in the increase of precipitation over Northern Europe, over the Mediterranean region the extreme precipitation changes results from the competition between a drying linked with a poleward shift of the circulation (Pfahl et al. 2017) and a thermodynamic effect leading to an increase of precipitable water content in the atmosphere (Drobinski et al. 2016). The signal over this region is less confident between models and the application of weighting methods can help in exploring possible reductions of uncertainties between different models.

Figure 22. *EURO-CORDEX P99 change mid (2041-2070) far (2071-2100) relative to 1981–2010 for for (a) the combined European region as well as (b)–(d) the three European SREX regions. The lighter boxes give the unconstrained distributions; the darker boxes give the constrained distributions. The colours indicate different scenarios: far future change RCP8.5 (red), mid future change RCP8.5 (orange), far future change RCP2.6 (green) and mid future change 2.6 (violet).*

The REA method is applied to the 99th percentile of precipitation (hereafter referred to P99). The aim is to explore how the method weights model's representation of extreme events of precipitation. It is applied for the periods 2041-2070 and 2071-2100 relative to the reference period 1981-2010. Regional ensemble results of 55 scenario simulations for the RCP8.5 and RCP2.6 at 0.11 degree resolution over the common EURO-CORDEX domain, using 8 GCMs and 11 RCMs, are compared with the driving CMIP5 global models at 2.5 degrees resolution. Observations used for the bias calculation are the E_OBS version 19.0e, with a 0.25° by 0.25° grid.
Weights are calculated as:

$$R_j = \left\{ \left[\frac{1}{abs(B_j)} \right]^m \cdot \left[\frac{1}{abs(D_j)} \right]^n \right\}^{\frac{1}{m \times n}}$$

and are functions of the model bias (B_j) (the higher the bias the lower the model reliability) and of the distance (D_j) of the change calculated by a given model from the REA average change, that is, the higher the distance the lower the model reliability. A new approach including weights which take into account the natural variability of the extreme event is still in progress.
Figure 22 and 23 show CORDEX and CMIP5 changes projections of the European subregions used in Brunner et al. (2019) based on the three European SREX regions: Northern Europe (NEU), Central Europe (CEU), Mediterranean (MED), and the combined European region (EUR: NEU + CEU + MED). For both the CORDEX and CMIP5 ensembles the P99 index shows an increase in extreme precipitation events over Northern and Central European regions, of 20(10)% for the far(mid) century period. CMIP5's projections present a wider range over the Mediterranean probably due to their low resolution that leads to discrepancies between global models land-sea masks, which play an important role when the water cycle is primarily determined by sea-to-land transport. EURO-CORDEX projections strongly reduces the uncertainty over the Mediterranean: the higher resolution allows a better representation of the instability at low levels (Fantini et al 2016). EURO-CORDEX projections show more variability over NEU and CEU, where complex topographic regions are included that play an important role in the development of precipitation.

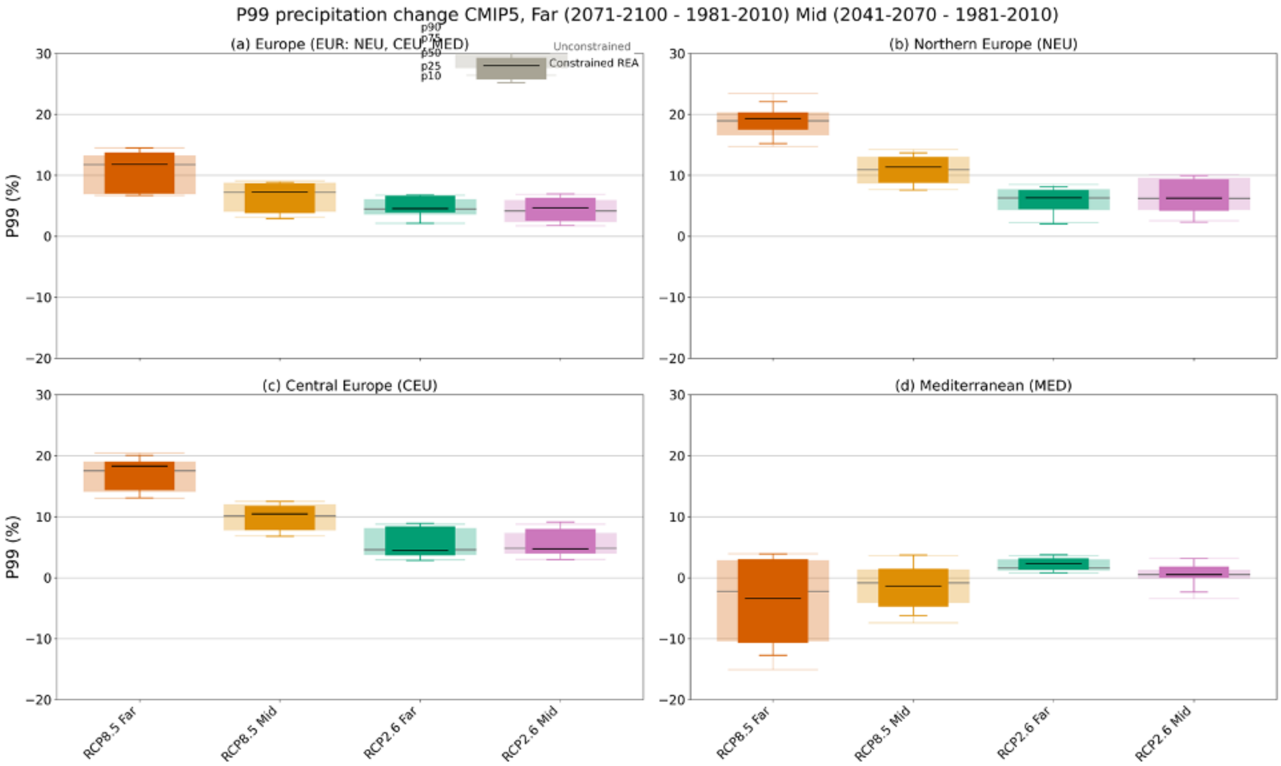


Figure 23*. CMIP5 P99 change mid (2041-2070) far (2071-2100) relative to 1981–2010 for for (a) the combined European region as well as (b)–(d) the three European SREX regions. The lighter boxes give the unconstrained distributions; the darker boxes give the constrained distributions. The colors indicate different scenarios: far future change RCP8.5 (red), mid future change RCP8.5 (orange), far future change RCP2.6 (green) and mid future change 2.6 (violet).*

To investigate the role of the weighting in reducing the uncertainty of the projections we show in Figure 24 the difference between the constrained and unconstrained P99 median value for both the EURO-CORDEX and CMIP5 ensembles. The difference is non-zero either in areas where models have higher biases or are outliers, therefore less reliable according to REA's method. We see that over complex topographic regions such as the Alps and the Scandinavian mountains, areas that are particularly vulnerable to extreme precipitation (Frei et al 2000), EURO-CORDEX models' distribution is more constrained. This is likely due to the fact that over the orography E_OBS extreme precipitation is smoothed by the spatial interpolation, thereby increasing precipitation biases.

The CMIP5 ensemble shows a disagreement in the representation of the P99 change over the Mediterranean, probably due to the land-sea mask representation, while gives a more homogeneous signal over the orography.

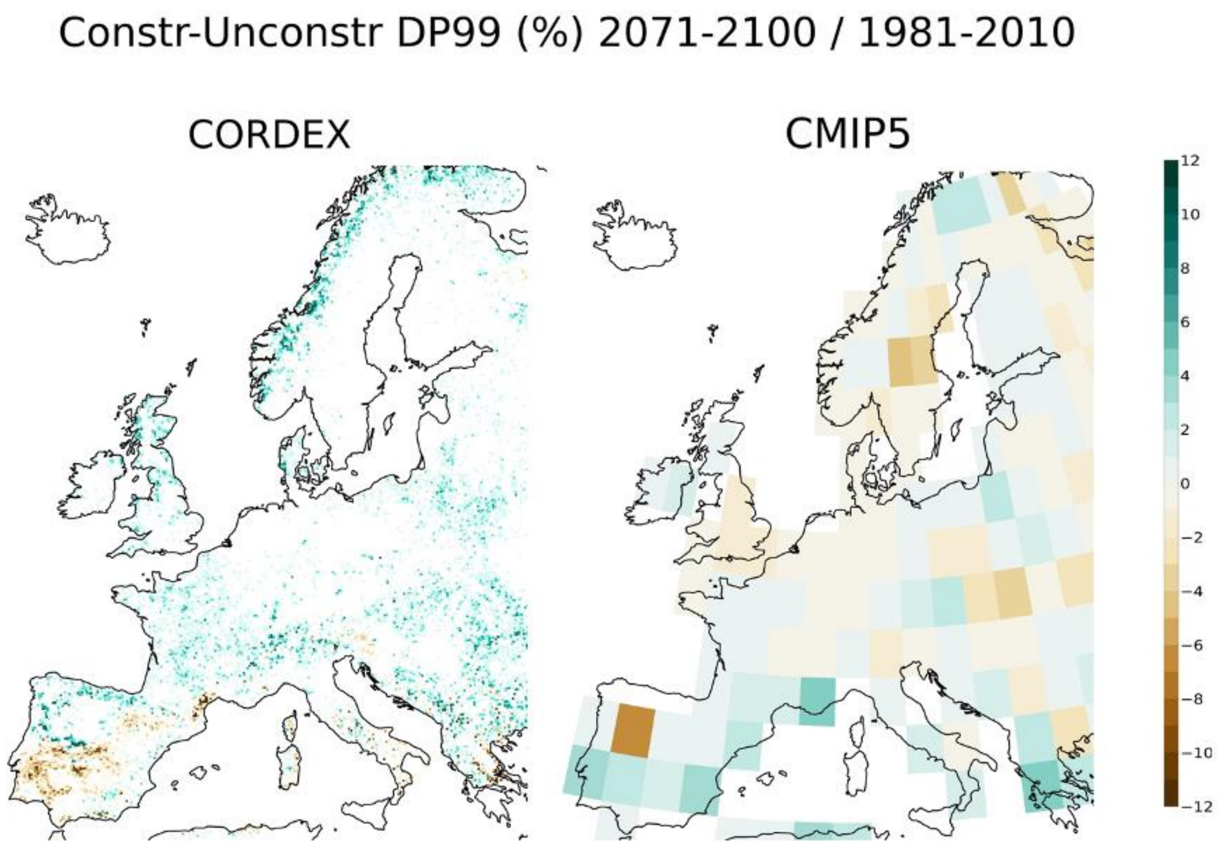


Figure 24. *EURO-CORDEX (left) and CMIP5 (right) difference between Constrained and Unconstrained P99 ensemble median change (%).*

9.3 Climate Projections in the context of internal variability.

Most climate projections focus on large-scale regions and time mean responses (e.g., 20 year mean climate conditions found in the EUCP Atlas) as these help to illustrate the climate change signal. However, what is actually experienced may be enhanced or mitigated by internal variability and is often not typical of the central estimate from the projected changes of the mean climate. The role of internal variability becomes more important on shorter time and smaller spatial scales. This context is important to consider where impacts are influenced by both climate variability and climate change.

Internal variability modifies the picture of a changing climate in a number of ways. Here we show how internal variability can lead to outcomes that might, at first glance, not be obvious from the projected time mean climate changes. We illustrate this by looking at the differences between long term climate mean projected changes (say, based 10- or 20-year averages) and projections for a single season, for a single climate model scale grid box. These would not be as evident for larger spatial or temporal scales but may be expected to be more pronounced for any applications that looks at smaller spatial scales (for example using CORDEX or convective permitting simulations). Generally, we find that:

- Internal variability broadens the range of potential outcomes for projections at smaller geographical or temporal periods. The amplitude of internal variability differs strongly across variables. Precipitation, for example, may vary several tenths of percent on decadal timescales. The amplitude of variability may also increase under climate change.
- Where climate change projections provide a confident picture of a warming or drying/wetting climate, internal variability is still able to throw up changes of opposite sign, for individual seasons. This is often underestimated by the users of climate change information.
- Where the seasonal distribution is skewed from gaussian (such as winter temperature or summer rainfall), the time mean climate change signal may hide even larger changes in the typical seasonal climate response.

Example 1: Mediterranean drying

The role of time averaging in isolating any climate change signal is illustrated by the example of Mediterranean drying. Many, if not most climate projections show that the mean precipitation in this region will decrease in the future. However, different climate models (and even different realizations with the same model) give quantitatively different results. The primary reason why different realizations of the same model produce different results, is the role of internal variability. Figure 25 below shows the future change of summer precipitation in the Mediterranean region, as simulated by two integrations with the regional climate model RACMO for the period 1950-2100 (using the RCP8.5 scenario from 2005). The two realisations differed only in the initial conditions of their driving GCM data.

A common approach to compute (future, or historic) trends, is by subtracting the mean of a given period from that of a reference period. Especially in (high-resolution) regional climate modelling this is an often-used approach. For example, most of the convection permitting simulations (WP3) are only 10 years in length. Figure 25 shows what happens if one uses the time-slice difference approach to compute the trend. Here, the period 1995-2014 is taken as a reference period. For each N-year period (N>3, vertical axis) centred at year x (horizontal axis) the colour indicates the relative difference between the precipitation in that period and that of the reference period. Blue colours indicate increases (e.g., 0.4 corresponds to a 40% increase), orange/brown colours indicate drying.

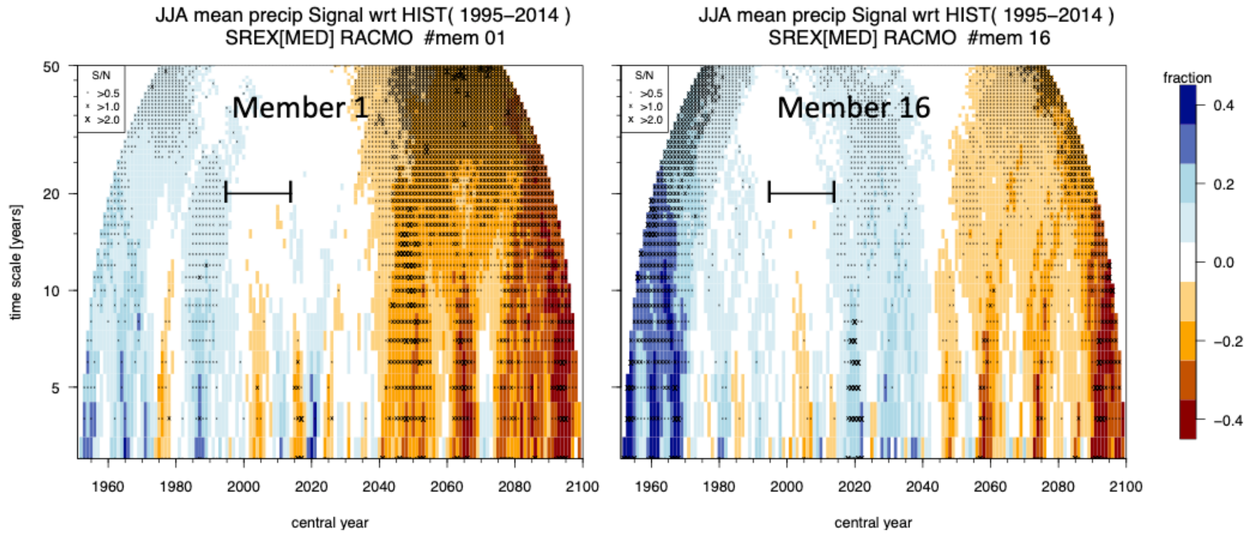


***Figure 25**. The trend in summer Mediterranean precipitation as a function of time averaging (Y) and central year (X), derived from two climate-change simulations of a business as usual scenario with the same model. Trends are computed as time-slice differences wrt the 1995-2014 mean (black horizontal line). Hatching is used to indicate regions where signal to noise exceeds a given value (see top-left legend). The climate change signal emerges earlier at the longer temporal scales (vertical axis), where we see a clear wet to dry signal as we move from pre-present-day to the future. As the temporal scales (vertical axis) approach the annual timescale, we see the emergence of greater variability around the more general inter-decadal trends. At these shorter timescales, this variability can substantially enhance or even entirely counteract the underlying drying trend. Whilst the particular realisation of internal variability would change depending on the manifestation within a given simulation, then general characteristics of this plot would be expected to hold.*

At longer time scales (vertical axis), there is a clear emergence of the drying signal from past to present to future in each member. At shorter temporal scales, however, we see a much greater role for natural variability, even on a relatively large domain like the Mediterranean. Had we taken a smaller location (like a province, or a small country) the figure would be much “noisier”. The conventional focus of climate projections on 20 years (or longer time averages) is to eliminate this ‘noise’ (internal variability) and isolate the climate change signal. However, many of the climate change impacts are linked to the interaction between natural variability and climate change. In this example, relatively wet periods occur still later in the century, despite general drying trends. In addition, in the first realisation (left) some of the driest conditions are already encountered prior to 2070. The internal variability role in the real world is not possible to predict beyond the next decade, however, this illustration of a possible manifestation is useful as it exposes many of the challenges of interpreting projections in the presence of natural variations in climate.

Another often overlooked aspect is that the amplitude of internal variability is usually underestimated when diagnosed from time-slice experiments. In figure 25, this is illustrated by the 'significance' crosses. At shorter time scales these occur at random locations in the graph (but at different years in the two realizations). Statistically this phenomenon is referred to as "regression to the mean": if because of some natural fluctuation the reference climate was abnormally wet, it is likely that a future climate will be drier (even in a stationary climate). Although the trend is statistically robust and different from zero, a large part of it relates to internal fluctuations. The problem is that without long transient integrations, or an ensemble of simulations, it is impossible to find out what is the true (physical) climate change signal. Finally, note that the amplitude of internal variability may also change over time. Also, this is seen in the drying Mediterranean area. The tendency towards systematic drying, is accompanied by a (relative) increase in the amplitude of internal variability.

Example 2: pdf changes at shorter time scales

Two further examples are now given, to show the impact of internal variability on smaller-scale regions at different time scales. Specifically, we discuss projections of the winter temperature changes in a 2.5°x2.5° region in north west Romania (centred around Sibiu) and summer rainfall changes in 2.5°x2.5° Madrid region (Spain).

We look at both 20 year averaged and single season climate projections, for a selection of (weighted) CMIP6 models and the full pdfs from the UKCP18 methodology. Comparing the 20 year with the single season projections in the CMIP6 models (box and whisker plots in the two following figures, 26 and 27) highlights the wider spread of responses when considering changes for only a single season. The orange box and whiskers correspond to the CMIP6 REA Atlas data for these two regions, seasons, and variables. The REA is based on 15 GCM simulations, which is probably enough to get a handle one the 20 year mean change. However, the REA CMIP6 estimates of change for these single seasons are evidently noisy, which illustrates the need to

Clearer insights can be seen in the full pdf estimates from the UKCP methodology (pdfs in figures 26 and 27). The impact of wider season to season variability is immediately evident from the wider distribution of projected temperature and precipitation change. Projections from the single season context provide useful insights that we should bear in mind when presenting and interpreting the more convectional 20 year mean climate projections (such as those from the Atlas). The 20-year climate warming signal is fairly unambiguous, with expectations that the average winter between 2041 and 2060 will be warmer (95% of the time). However, we'd expect a 2050 winter to actually be cooler than the present day (1995 to 2014) baseline 18% of the time. We see a similar picture with rainfall, with the climate mean projections suggesting only a 18% chance that the 20-year average summer would be wetter than the present-day baseline. Projections for the 2050 summer season, whilst consistent in capturing this same tendency to dry, suggest that an individual season may be almost twice as likely (31%) to be wetter than the present-day baseline.

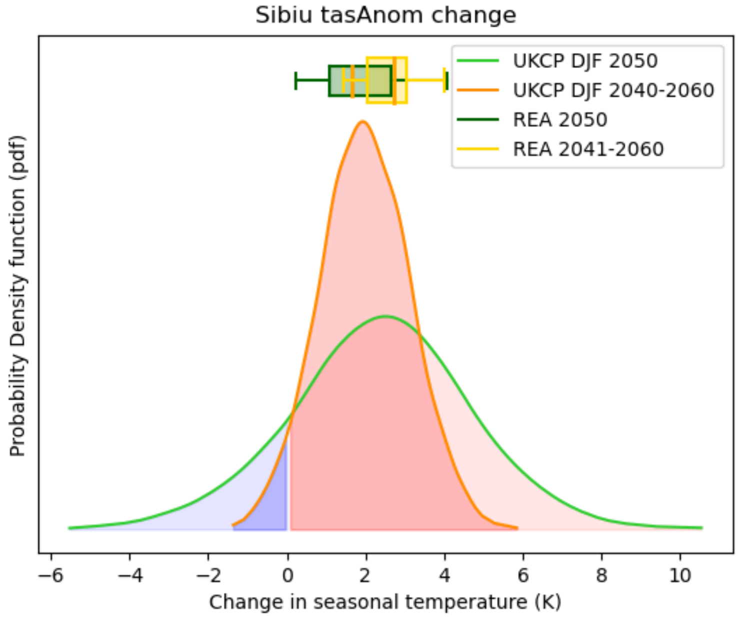


***Figure 26**: Difference between 20 year mean projections and projections for a single season, for north west Romania (Sibiu) winter temperature changes. The plot shows a clear winter warming signal in the 20-year mean (orange pdf). However, the projections for the 2050 winter season (green pdf) reveal that there is a sizable chance that individual winters may be colder than the present-day baseline. These projections are all relative to a 1995 to 2014 baseline state. The box and whiskers (showing 10,25, 50, 75 and 90 percentiles) are based on weighted estimates of 15 CMIP6 climate models (using the REA methodology) and the pdfs are estimates of the same quantities using the full probabilities from the UKCP18 methodology.*

For many locations and variables, internal variability will just broaden the range of what we can expect as we make projections for shorter timescales or smaller geographical locations. However, where the inter-seasonal variability is not symmetrically distributed this also has implications for what we'd expect from a typical seasonal change (as opposed to a longer-term averaged change). Summer rainfall in Mediterranean regions is a good example of this. In the Madrid summer rainfall figure (figure 27), we can see that the distribution of projected changes for the 2050 summer is strongly skewed, with most of the distribution showing a drying but with a long-wet tail. Whilst the mean estimate of summer drying is -38% compared to the baseline, the median is -43% drier. So, the more typical summer will be drier than the climate mean suggests because the climate mean estimate also includes the possibilities of rare exceptionally wet summers.

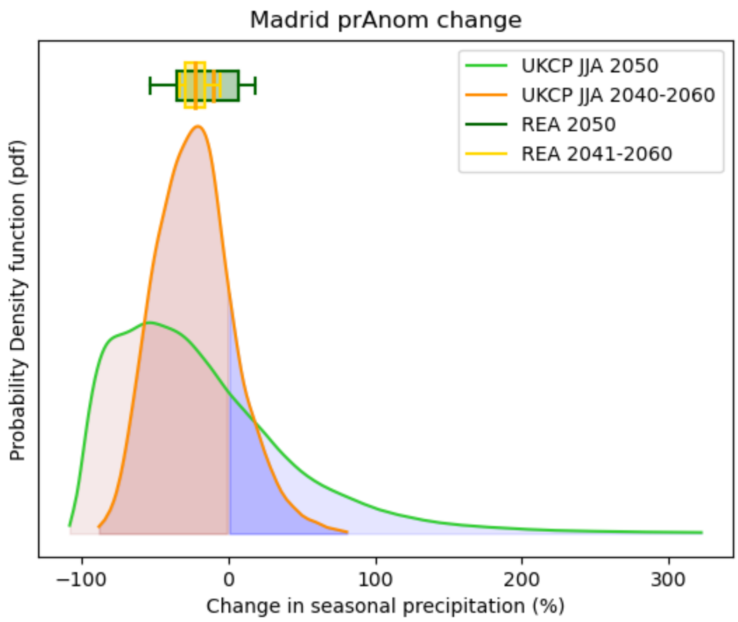


Figure 27: *Difference between 20 year mean projections and projections for a single season, for Madrid summer rainfall changes. The plot shows a clear summer drying signal in the 20-year mean. However, the projections for the 2050 winter season reveal that the typical summers are likely to be drier than this climatological change suggests and at the same time we are unable to rule out low likelihood but high impact wet summer seasons (with rainfall as much as 200% above the present-day baseline). These projections are all relative to a 1995 to 2014 baseline state. The box and whiskers (showing 10,25, 50, 75 and 90 percentiles) are based on weighted estimates of 15 CMIP6 climate models (using the REA methodology) and the pdfs are estimates of the same quantities using the full probabilities from the UKCP18 methodology.*

Take away messages from comparing the more typical climate mean projections of changes with projections that account for internal variability is that we'd expect a greater range of change at smaller temporal and spatial scales, including the possibility of changes of the opposite sign. Where internal variability produces skewed distributions, the typical seasonal response (as measured by the median) can often be consistent with larger magnitude changes than the headline messages from the longer-term climate mean projections.

9.4 Using WP2 projections to provide a context for WP3 projections

The aim, here, is to provide an indication of how representative of the broader potential range of projected changes (in mean seasonal temperature and rainfall) the available Convective Permitting Models (CPMs) are. We have (comparatively) data rich regions (the Alpine domain with 7 CPMs) and regions with sparser CPM model availability (where only 2 or more are available). This section is intended to showcase the CPM responses compared to the wider pdfs.

There are several things to look for in the following analysis. Firstly, the comparison of climate models and the probabilistic projections provides an initial impression on how well the available modelling captures the range of the wider potential climate responses. The UKCP probabilistic projections, used in the follow examples, accounts for additional sources of projection uncertainty that is not accounted for in the GCM to CPM modelling. These include carbon cycle feedbacks (which tend to produce stronger warming responses) and the parametric uncertainties (that can expand the uncertainty distribution beyond what the CMIP ensemble is able to capture). So, there are good reasons why the GCM to CPM simulations wouldn't be expected to span the full potential probabilistic response space, but how well they do at this is important context for users of this CPM data. This context can inform users of this regional modelling data of where responses may under (or over) estimate the magnitude of some changes. This comparison makes most sense when comparing the climate response of the parent GCM that is used to drive the regional models. This is because the observational constraints that inform the probabilistic projections are all done on large scale global climate models (GCMs). One of the motivations for carrying out high resolution regional climate models, is that their ability to start explicitly resolving convective cloud processes may lead to changes in their projected responses.

This leads on to the second aspect to look out for. The following figures (Figures 28 to 32) show the change in the projected temperature and precipitation between parent GCM and coarse scale RCM, as well as coarse scale RCM and convective permitting CPM. Should there be a common shift in responses at either of those steps, then this would imply that resolving the higher resolutions, changes the climate projections (a process that the probabilistic projections are unable to currently account for). Benefits of resolving processes may well be confounded by small ensemble sizes, presence of biases in some of the models, etc - so the absence of such consistent shifts can't necessarily be interpreted as evidence that resolving those scales does not have implications for current climate projections. Alternatively, many of the documented benefits of resolving these higher resolutions in CPMs, is in higher frequency changes. So, we may just not see these changes in the large scale seasonal mean 10-year averaged projections presented here. As we will see, we don't observe much evidence of consistent shifts in the climate responses between high resolution regional modelling and their courser scale parent simulations, in the examples we explore here.

Lastly, it is useful to look for any underlying structure between projected changes in temperature and rainfall in the probabilistic projections. There can often be some quite strong correlations, which we might not always expect the small number of available CPM simulations to pick up. Typically, warmer conditions can often be associated with summer drying in southern Europe. Whereas wetter winter conditions may be associated with stronger warming in some northern European regions. Where these correlations exist, they can be important for some users of climate change who have exposure to changes in both variables. So, it is useful to note where available regional models can capture these underlying relationships. We'll discuss a few examples below.

9.4.1 Context for regions with limited CPM modelling

Convective Permitting regional climate Models (CPMs) are very much a novel modelling tool and due to their great computational expense, they tend to be run for only limited domains. EUCP has expanded the number of CPM so that all mainland European regions have projections from some of these models. However, for most regions we have limited samples which are each drawn from the wider range of climate change responses. We show here comparisons of the regional modelling and the probabilistic projections, for two seasons, for the Central Eastern Europe region, as an illustration of some of the challenges in having only a limited sample of simulations. The top-level inferences from the following comparison are:

- Summer (JJA): the chance sampling of these two GCM-RCM-CPM modelling sets does a reasonable job of sampling summer conditions. Lower likelihood but higher impact change may be possible.
- Winter (DJF): fails to sample winter drying (though this is modest, and may well just be internal variability) and misses higher end warming (especially if only the CPM is used).

For summer, the two convective permitting regional models (CPMs) available for the Central Eastern European domain do a reasonable job of spanning the range of projected summer conditions. Both sit within the 25th percentiles from the UKCP probabilistic methodology, so neither can be said to explore the less likely but still possible parts of the projected response space. Between them they capture both a stronger warming with modest drying and more modest warming with less change in mean summer rainfall. The UKCP probabilistic projections expose the correlation between warmer temperature increases and the tendency of the projections to produce more of a drying signal. These kinds of relationships can be important for users who have exposure to changes in both variables, and it is often hard to impossible to identify such relationship from the small handful of high-resolution simulations that are typically available. In this instance, the chance sampling has happened to have done well with the warmer projection also capturing drier conditions that are more typical of responses of this magnitude.

The estimates from the CMIP5 are narrower than those from UKCP method, which is related to UKCP including additional uncertainties due to decadal variability; and inclusion of wider uncertainties (namely carbon cycle and bespoke uncertainty modelling as well as statistical uncertainties). Whilst neither CPM explores the whole CMIP5 temperature distribution, they do a reasonable job of spanning the inter-quartile range temperature range and represent moistening and drying just either side of the CMIP5 range.

Another thing to note is that there is not sufficient consistency in the shifts of the two CPMs to suggest that resolving processes at this scale significantly changes the projections, as this seasonal mean temporal scale.

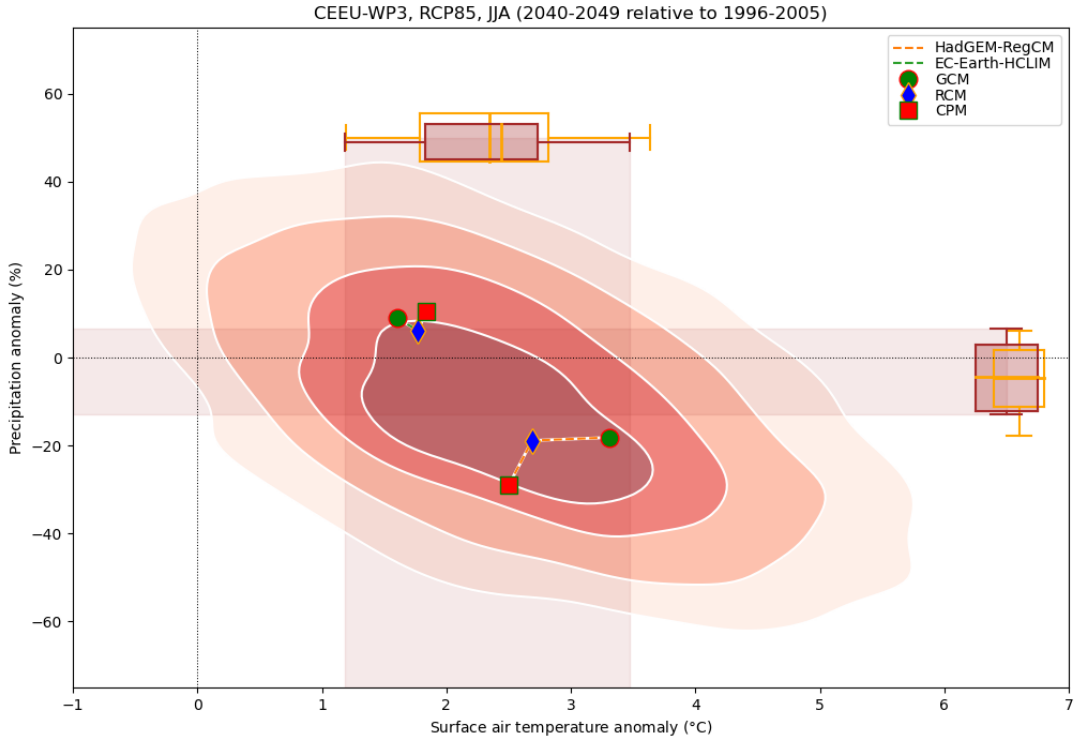


***Figure 28**: **Summer Climate Projections for Central Eastern Europe**. Contours represent 90th, 75th, 50th and 25th percentiles for constrained joint projections from the UKMO probabilistic projections of temperature and rainfall change. The points represent projected changes from the two available Convective Permitting regional Models (CPMs) that produce projections that cover the whole of this region (red squares). The lines connect these CPM responses to the responses of their RCM and GCM parent models. The box and whisker plots show estimated changes from the ClimWIP methodology, with the raw CMIP5 range and the observationally constrained range show in open orange and filled brown, respectively, for the 90th, 75th, 50th, 25th and 10th percentiles*

For winter, these two CPMs represent a more limited sample of changes in the Central Eastern European domain. The RCM and CPM do not reproduce the stronger warming of their parent GCM and fall more in line with HCLIM, projecting a warming of roughly 1.5 K. The CPMs, therefore, provide only a limited sample of what would be considered the plausible range of temperature change (given large scale observational constraints). Interestingly, the observational constraints in ClimWIP tend to suggest that the larger temperature future responses are more consistent with observed changes, in this region. Both CPM simulations capture a winter enhancement of precipitation, on the upper end or outside the CMIP5 range, though it is considered consistent with the wider UKCP uncertainty range. Both CMIP5 and the UKCP methodology suggest that a winter drying is consistent with large fraction of the project changes, and this is not sampled in the available CPM modelling. It also suggests that, unlike in summer, there is not a strong relationship between winter warming and changes in precipitation (which is again, useful context).

So, the winter CPM simulations represent similar, limited, samples of the wider underlying uncertainties. Projections that are either considerably warmer or even produce a net drying are both possible, but not represented here.

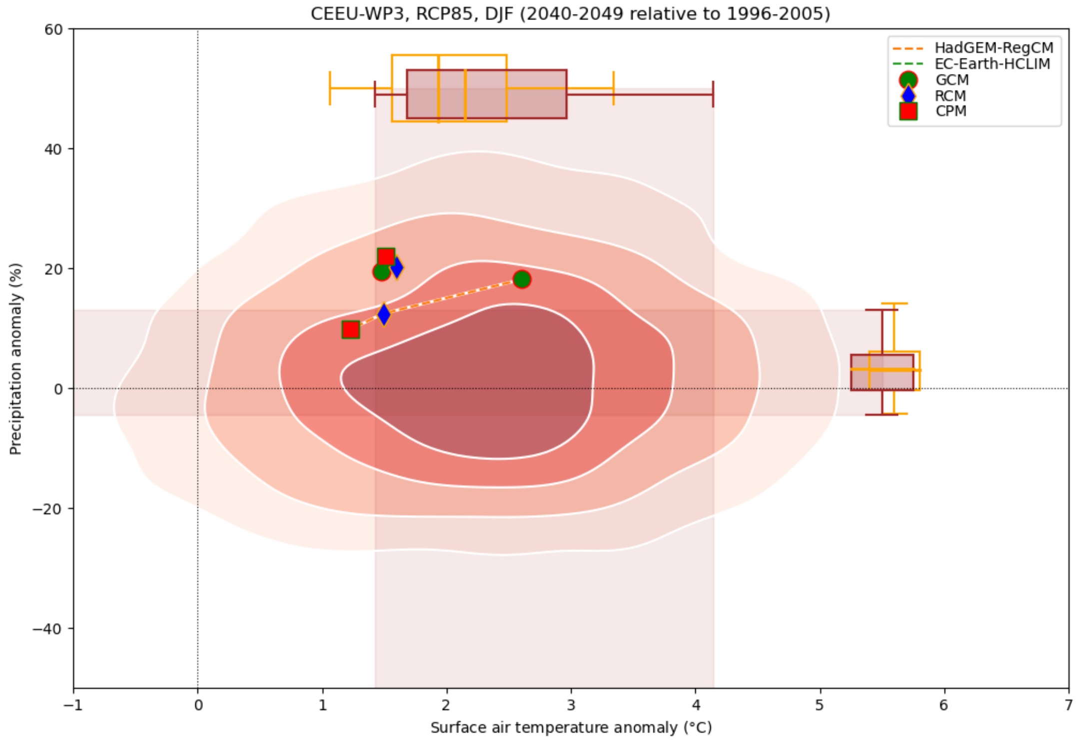


***Figure 29**: **Winter Climate Projections for Central Eastern Europe**. Contours represent 90th, 75th, 50th and 25th percentiles for constrained joint projections from the UKMO probabilistic projections of temperature and rainfall change. The points represent projected changes from the two available Convective Permitting regional Models (CPMs) that produce projections that cover the whole of this region (red squares). The lines connect these CPM responses to the responses of their RCM and GCM parent models. The box and whisker plots show estimated changes from the ClimWIP methodology, with the raw CMIP5 range and the observationally constrained range show in open orange and filled brown, respectively, for the 90th, 75th, 50th, 25th and 10th percentiles*

9.4.2 Context for the Alpine region

This region is the most well served by current convective permitting modelling. Shown below are the GCM, RCM and CPM responses compared against current probability projections.

- DJF: The models sample temperature changes only in bottom half of the potential temperature distribution and none of the models capture sample winter drying. The latter may be less relevant if the precipitation range is largely internal variability (but it may be because the drying is driven largely by the PPE, and hence not captured by the CPMs and their CMIP5 based drivers)
- JJA: ICTP's HadGEM2-ES – RegCM modelling compensates for the otherwise cool biased sample. With this model, some of the warmer and drier responses are captured but the

ensemble still fails to capture some the higher end, albeit less likely, strong warming and drying.

- SON: The CPM sampling is better placed to capture the range of responses. RegCM is positioned to capture higher end warming; GERIC's REMO CPM captures strong drying with modest warming and COSMO captures moderate warming and moistening.

In principle, potential users of CPM projections are much better served in the Alpine domain. Here we show 7 EUCP CPM simulations (with their parent RCMs and GCMs) for winter, summer and autumn seasons. As can be seen from the winter figure (Figure 30), below, this larger number of models does not always translate into wider sampling of potential climate responses. None of the CPM models are able to sample temperature in the upper half of the UKCP distribution nor sample the potential for this region to experience winter drying. Compared to the CMIP5 ensemble they do reasonably well. Again, they tend to miss the upper end of the CMIP5 observationally constrained distribution. Interestingly 3 of the 4 GCM projections, that are used as boundary conditions, fall below the lower 10 percentile of the CMIP5 range. The regional modelling has tended to shift up the sampling of lower temperature responses somewhat, but it is worth noting that the chance sampling still tends to over-sample cooler winter temperature responses compared to what might be expected from the constrained range from either probabilistic methodology presented here. Similarly, the winter CPMs do not encompass the potential drying signal in both the CMIP5 and UK probabilistic methodology.

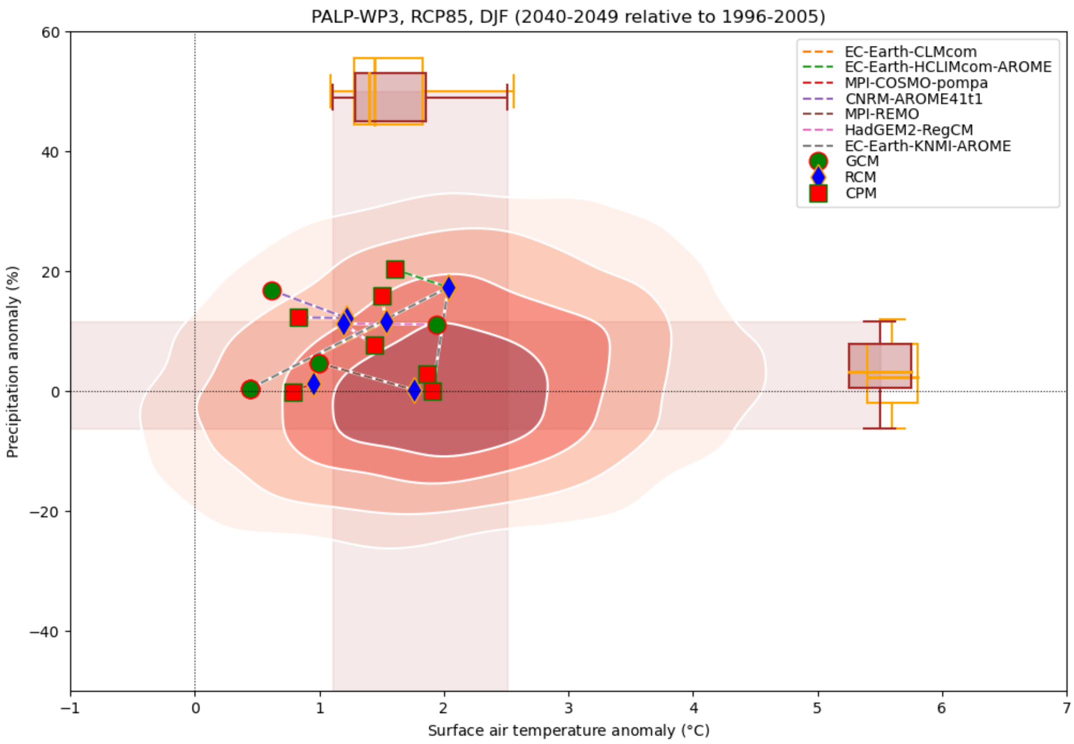


***Figure 30*: *Winter Climate Projections for the Alpine domain*.** *Contours represent 90th, 75th, 50th and 25th percentiles for constrained joint projections from the UKMO probabilistic projections of temperature and rainfall change. The points represent projected changes from the seven available Convective Permitting regional Models (CPMs) that produce projections that cover the whole of*

this region (red squares). The lines connect these CPM responses to the responses of their RCM and GCM parent models. The box and whisker plots show estimated changes from the ClimWIP methodology, with the raw CMIP5 range and the observationally constrained range show in open orange and filled brown, respectively, for the 90th, 75th, 50th, 25th and 10th percentiles

Looking at the autumn/fall season in the Alpes (Figure 31), the CPMs appear to be in a better position in terms of sampling underlying range of projections. Individual CPM members sample stronger warming, modest warming with a drying signal and modest warming with a moistening signal. A couple of CPMs produce small warming with quite strong increases in Autumn rainfall, which (according to the probabilistic projections) would be considered less likely. Whilst the underling probability distribution suggests these responses are not implausible, neither has a particularly strong weight. Caution needs to be exercised before such models can be considered as outliers, as it may well be that resolving regional scale processes at higher resolution has led to the more pronounced increase in Autumn rainfall. Rainfall changes in the Autumn season (September to November) are of particular interest because there is now some suggestion from the CPMs that climate change may extend summer convective rainfall events into the autumn season. The probabilistic projections are based on lower resolution global modelling that does not capture these processes. Nevertheless, comparisons like this provide a useful context to frame such questions.

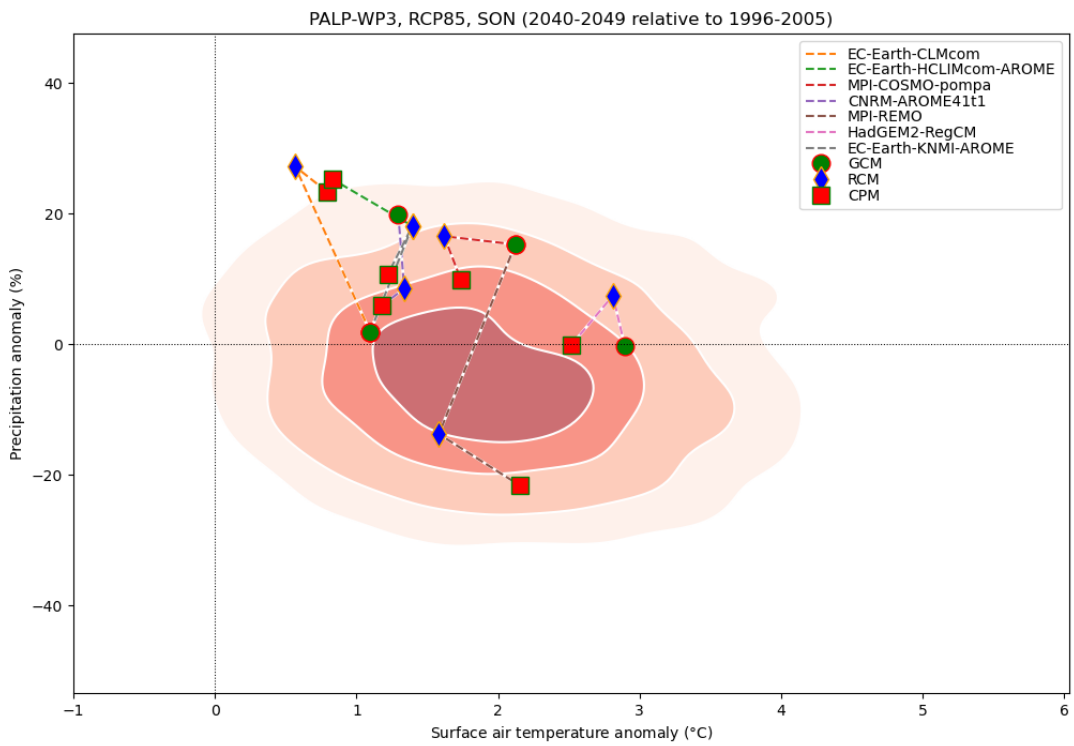


***Figure 31**: **Autumn Climate Projections for the Alpine domain**. Contours represent 90th, 75th, 50th and 25th percentiles for constrained joint projections from the UKMO probabilistic projections of temperature and rainfall change. The points represent projected changes from the seven available Convective Permitting regional Models (CPMs) that produce projections that cover the*

whole of this region (red squares). The lines connect these CPM responses to the responses of their RCM and GCM parent models.

The CPM ensemble represents a better sample of potential summer changes (Figure 32) than the winter season, according to the probabilistic methods. Here (like in the Central Eastern European domain) we see that there is a strong relationship between greater warming and greater summer drying. The available CPM ensemble captures much of this underlying structure, and the inclusion of the RegCM regional model does much to compensate for the otherwise cool bias to the sample represented by the other CPMs. Interestingly, ClimWIP suggests that CMIP5 simulations with stronger future warming are more consistent with observed changes. Five of the seven CPMs produce a summer warming that is cooler than the lower 25 percentile of the constrained CMIP5 range. Whilst there are limitations to the CPM sampling of the climate projection space (for example, we may expect the potential for stronger summer warming and drying than the CPM sample is able to capture) the available members do span some of the diversity in potential responses.

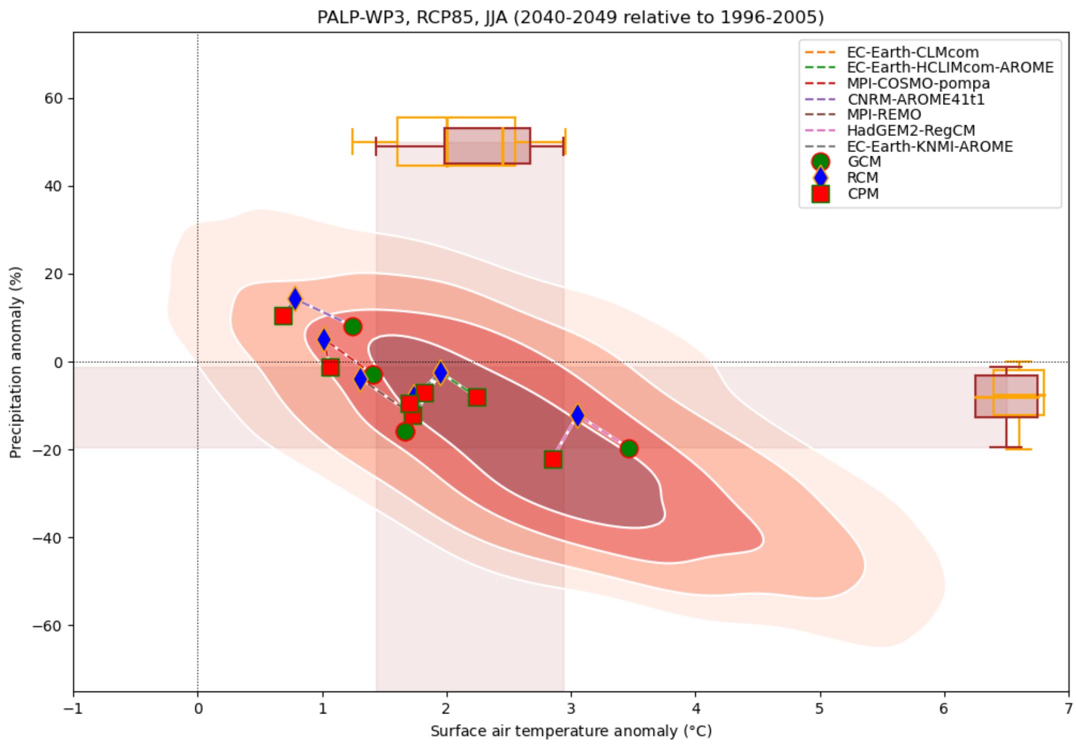


***Figure 32**: **Summer Climate Projections for the Alpine domain**. Contours represent 90th, 75th, 50th and 25th percentiles for constrained joint projections from the UKMO probabilistic projections of temperature and rainfall change. The points represent projected changes from the seven available Convective Permitting regional Models (CPMs) that produce projections that cover the whole of this region (red squares). The lines connect these CPM responses to the responses of their RCM and GCM parent models. The box and whisker plots show estimated changes from the ClimWIP methodology, with the raw CMIP5 range and the observationally constrained range show in open orange and filled brown, respectively, for the 90th, 75th, 50th, 25th and 10th percentiles*

9.4.2 Applications for using PDFs to inform RCM driving data selections

The observationally constrained projection can be used beyond the context they provide to currently available climate projections. Exploring some of the challenges in selecting future simulations for downscaling, as an example, they can also be used as a visual guide of potential candidates for further regional climate simulations (or statistical downscaling) that can complement existing data.

Figure 33 shows how the CMIP6 models position themselves with respect to the ranges constrained by 3 methods for summer mean temperature change over different European regions. Putting these three constrained ranges together allows us to identify a subset of models that can be selected as driving GCMs for regional simulations produced by the RCMs. A relevant strategy for an optimal selection would be to sample the climate sensitivity (in the broad sense) within the constrained ranges. For example, the ACCESS-ESM1-5 model (number 2 in Figure 33) could be selected as characteristic of strong warming over Europe, while the MPI-ESM1-2-LR model (number 35) would be more representative of weak warming. Models that are incompatible with constrained ranges could also be useful for estimating climate risk in a comprehensive way, by considering them as low likelihood, but high impact realisations (i.e. beyond the 10% and 90% thresholds). This is for example the case for the UKESM1-0-LL (number 41) or FGOALS-g3 (number 19) models. Such a modelling strategy is likely to be adopted for the next CORDEX phase.

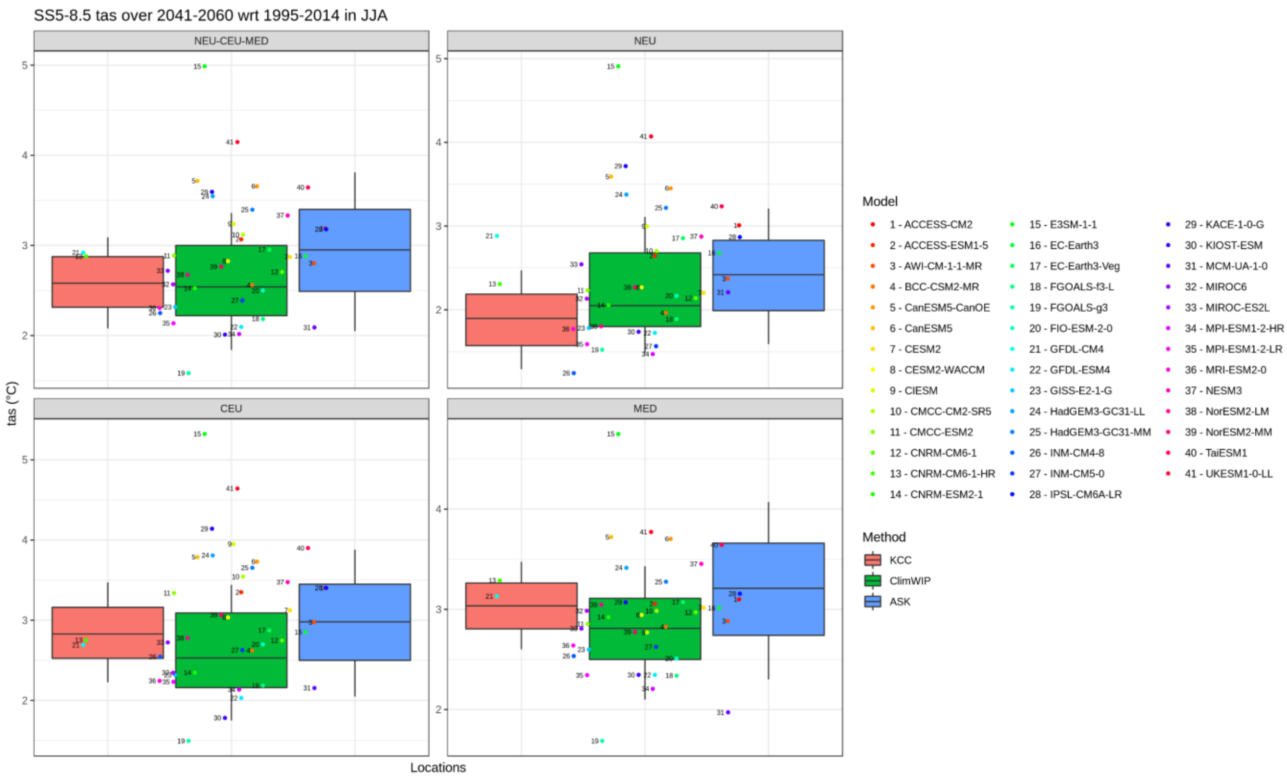


Figure 33: Comparison of observationally constrained ranges for summer temperature, and currently available CMIP6 simulations. *Constrained temperature changes (relative to 1995–2014) from the KCC, ClimWIP and ASK methods for the SSP5-8.5 scenario over the 2041-2060 period for*

the European (NEU-CEU-MED) region and each SREX subregion. The dots show individual models as labeled. The horizontal dot position and its color is arbitrary.

This contextualisation exercise was also done for the CMIP5 models and CPMs used by WG3. In the examples, below, we look at where the range of CMIP5 simulations sit within the constrained ClimWIP and UKCP probabilistic projections.

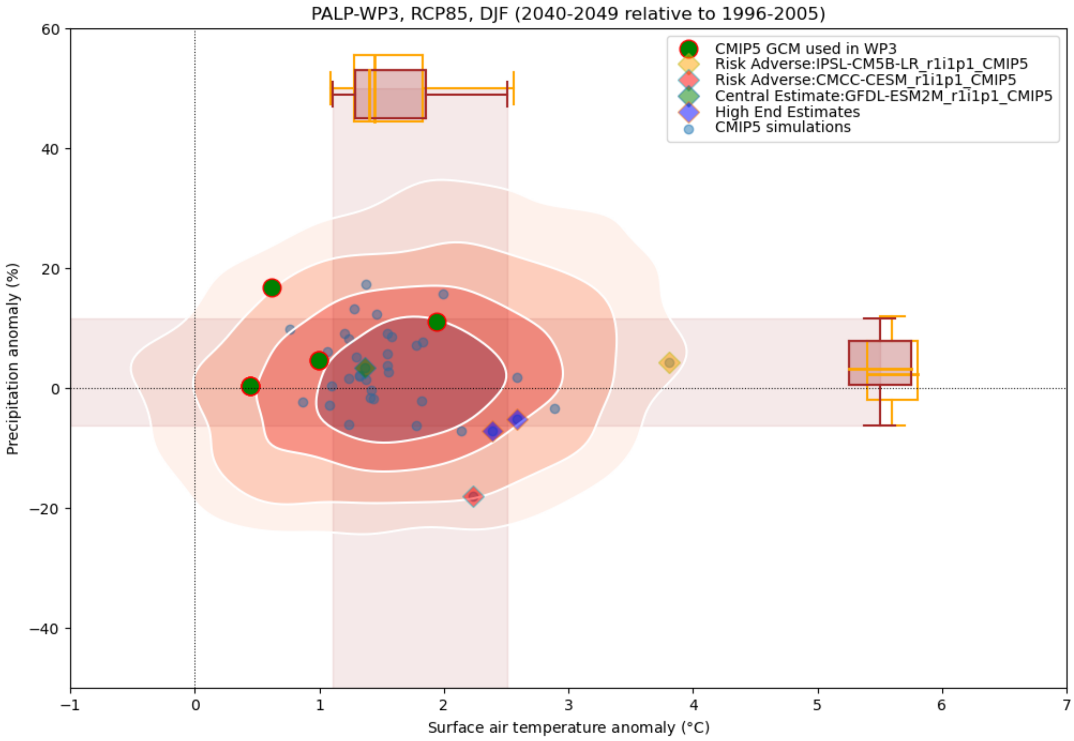


Figure 34: Exploring potential GCM candidates as boundary conditions for future (statistical) downscaling in the Alpine region. As Figure 30 but including only CMIP5 simulations used for existing regional modelling boundary conditions (green circles) or the wider CMIP5 simulations (blue dots). This illustrates the distribution of climate responses across the available CMIP5 simulations. In this case, the CMIP5 simulations, that have been used, sit within the cooler and wetter part of the wider distribution. Potential candidates for use as a central estimate (green diamond) or high-end changes, within the ClimWIP constrained 10-90 percentiles (blue diamonds) are shown. Lower likelihood, but still plausible CMIP5 projections are identified to inform more risk-adverse users (yellow and red diamonds).

In figures 34 and 35 we can illustrate how limited a sampling of the potential climate response space is over the Global Climate Model (GCM) simulations that have been chosen for the RCM and CPM downscaling simulations. As shown in the earlier Figures 29 and 30, this limitation is more apparent in winter than it is in summer. Here, we use the ClimWIP constrained ranges to identify alternative GCMs that could be used to provide regional model boundary conditions to complement the existing choices. We use proximity to ClimWIP 50% and 10/90% percentiles to have highlighted potential candidates for better estimates of central or high-end changes, respectively (identified in the figure legends).

This work is important because the current lack of sampling of warmer and drier future winter changes in these two regions is likely to mask larger impacts on the overall water cycle than the current regional climate models suggest. If the real world also shows winter drying and stronger warming then this would imply further pressures on the water availability, as there would be reduced capacity to recharge water reserves during the winter and the implied reduction in winter snowpack would further reduce river flows in spring and early summer. Identifying simulations that explore this narrative should inform more robust decision making in light of the current range of potential future changes.

We have also identified a few less likely but still plausible higher impact climate projections that might be more relevant for risk adverse users of climate projection information (indicated in the figure 33 and 34 legends)

We have illustrated the joint temperature and precipitation approach, here, with CMIP5 era simulations. The potential to utilize CMIP5 for subsequent downscaling is limited due to their sparse availability of lateral boundary conditions (needed to drive regional climate simulations) and the focus has moved on to newer CMIP6 simulations for much of the further downscaling work. However, this section illustrates an approach that could equally be applied to the more recent CMIP6 simulations and the identification of CMIP5 realisations may still have relevance should there be an appetite to pool realisations from both climate model ensembles (particularly if statistical, rather than dynamical, downscaling approaches were adopted). This work illustrates how the constrained climate projections produced here can be both used to provide context or to provide a guide to selecting GCMs for downscaling that retain the diversity of the climate responses of global simulations.

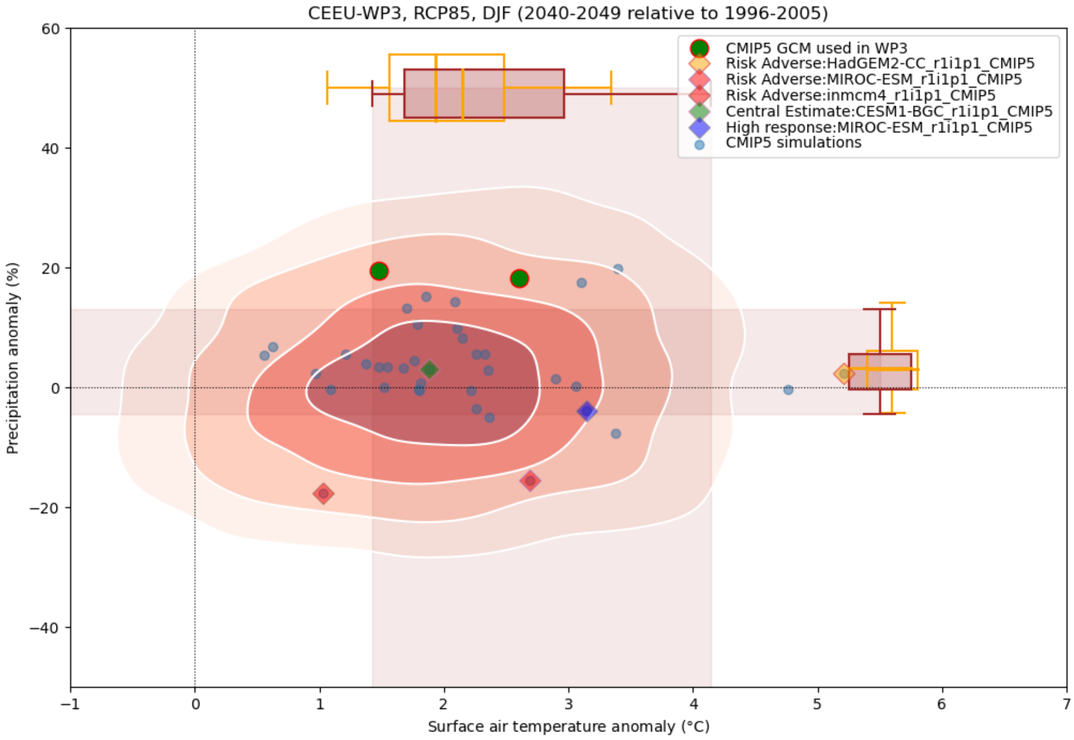


Figure 35: Exploring potential GCM candidates as boundary conditions for future (statistical) downscaling in the Central Eastern European region. As Figure 29 but including only CMIP5 simulations used for existing regional modelling boundary conditions (green circles) or the wider CMIP5 simulations (blue dots). This illustrates the distribution of climate responses across the available CMIP5 simulations. In this case, the CMIP5 simulations, that have been used, sit at the wet extreme of the wider distribution. Potential candidates for use as a central estimate (green diamond) or high-end changes, within the ClimWIP constrained 10-90 percentiles (blue diamond) are shown. Lower likelihood, but still plausible CMIP5 projections are identified to inform more risk-adverse users (yellow and red diamonds).

10. Making use of available WP2 outputs.

WP2 has followed a number of approaches to make the outputs of this work more widely accessible. Maps of projected European climate change, that are constrained by observations, are provided in the Atlas alongside links to download the underlying data (in netCDF format). In addition, the source code for a number of these methodologies has been released in the public domain.

Here we have tried to provide some points to aid potential users in identifying appropriate outputs for their applications.

10.1 Estimates of methodological skill

The systematic assessment of out-of-sample projection skill provides useful context for potential users. In the following table (Table 5), we name the methods labelled in that study and briefly discuss what these findings imply about the strengths of each approach.

Table 5: Summary of the implied skill, of each method, from the out-of-sample analysis (Section 9.1)

Label used in Out-of-sample analysis	Method Acronym	Summer Temperature	Summer Rainfall
A	REA	Good improvement in CEU. Mixed improvements in the other regions - with projections tending to be more over-confident	Little consistent evidence of skill
B	CALL	One of the best in NEU and CEU across a number of metrics. However, serious problems with its ability in MED which may raise a red flag for some users.	Little consistent evidence of skill
C	ClimWIP	Observations improve projections in all regions, particularly strong improvements in the MED; Does better than most in reducing over-confidence	Little consistent evidence of skill
D	KCC	Good improvement over the unconstrained projections. Either the best (MED) or one of the best constrained projections	Little consistent evidence of skill
E	ASK	Observations improve projections in all regions.	ASK represents poorer projections for CEU and MED rainfall change, compared to taking the CMIP6 range

<u>Summer Temperature</u>: Firstly, it is worth summarising again that all methods showed benefits of observational constraints for European summer temperature. The means that we can be confident that observationally constrained summer temperature projections will be more skilful that just taking available climate simulations, regardless of the methodology applied.

The exceptions to this are in the Mediterranean, where two methodologies (REA and CALL) show a degradation in climate projection skill. This is only a modest impact for REA, but this degradation is large for the CALL method - which raises a red flag for using this particular method in this region (and potentially raises questions about reliability in others).

Summer Rainfall: There is less evidence that observational constraints lead to consistent improvements in summer rainfall ranges, across all regions. Though, with one exception (below), does not consistently degrade projections (compared to just taking raw output). These new results do not suggest that there is a compelling reason to adopt one of the observationally calibrated methodologies (unlike for summer temperature, where there is) but if this is being done as part of a wider multi-variate assessment, then there is still value in doing so.

The exception is the projections produced by the ASK methodology. This shows examples of large degradation of skill compared to raw CMIP5 output in the CEU and MED regions, probably related to the low signal-to-noise ratio violating the premise of the method to use a well constrained detectable signal as constraint. The team behind this method recognise the challenge where there is not yet a sufficient emergent climate signal (other methods fall back to the unconstrained model data, in cases like this), which is one of the reasons why ASK precipitation projections do not form part of the Atlas.

Value of model weights: It is worth noting that even where calibration of methods with observations does not lead to reduced ranges in climate projections (such as summer rainfall) there can still be benefits from doing so. The methods providing model weights (REA and ClimWIP) still down weight poor simulations even where this does not affect the headline ranges - so making use of these model weight still has a value for applications that need to identify "good" simulations for downstream impacts/analysis.

10.2 Applying methodologies to new bespoke applications

All the methodologies require a level of technical expertise to apply the source code to new bespoke applications. So, taking this route will likely only be appropriate for a small subset of potential users. Furthermore, there is a big range in complexity (such as the need for expert judgement and required data inputs) across the various WP2 methodologies. We try and briefly summarise those difference in this section, to provide a rough signposting of how difficult each approach may be to apply.

- KCC Application: the method requires observation time series of local and global temperature, and an estimation of instrumental uncertainty. The method also requires calculating the member ensemble mean in each CMIP model before prior calculations. Users need to be comfortable with large sample outputs (tens of thousands).
- ClimWIP application: the method has been added to the Earth System Model Evaluation Tool (ESMValTool), including example cases and an extensive documentation to allow easy reusability. The method requires observations as well as model data, which will be handled

by ESMValTool once set up. The output weights can easily be understood and applied by most potential technical users. Several tuning parameters require expert knowledge but for many applications the user can fall back to provided values based on published literature. Two real examples illustrate a potential application of ClimWIP by new users:

  - After presenting the ClimWIP-ESMValTool implementation at a conference (https://doi.org/10.5194/egusphere-egu21-9387)ETHZ was approached directly a by a group interested in applying the method. In a next step a new setup was developed in a cooperation between TU Delft and ETH Zurich and weights were subsequently provided by ETH. The resulting study will be submitted in the near future (Gründemann et al. in preparation).
  - WP5.5 within EUCP have set up ClimWIP for the UK JASMIN system for a different application. The public GitHub discussion platform provided in the frame of ESMValTool enabled easy bug fixing and communication (https://github.com/ESMValGroup/ESMValTool/issues/2320) and the new case is now running.

- REA Application: perhaps the most straightforward application. REA requires observation time series as well as model data. The outputs can be readily understood by most potential users (e.g. relative weights on available models, based on how consistent they were to observed estimate and how outliers model-simulated changes are compared to the other models).
- ASK Application: The observational requirements from the ASK method are more akin to HistC (longer historical timeseries) but expert judgement is need in its application (with both familiarity detection and application approaches and their code implementations and needing make expert choices such as whether the greenhouse gas or total anthropogenic components are chosen as the basis of the future constraint; which of several available methods to use, and if to use signal-to-noise optimization or not - if it is chosen, the demands on expertise are substantial, and the user really does need to be familiar with the existing methods). ASK Outputs: Like HistC, users need to be comfortable with interpretation and processing of large sample output; and user needs to be able to interpret the combination of scaling factor for signal amplitude with the future simulation, in order to arrive at uncertainty ranges for the signal only. For full uncertainty ranges, noise needs to be added and sampled as well. In conclusion, ASK is a method that requires significant expertise, or should be done in collaboration with an expert.
- CALL Application: The observational requirements are similar to ASK and KCC in that long observational datasets are required. The method can be applied to any large ensemble dataset that has been produced with a suitable external forcing and has been tested on the CESM and MPI-GE large ensembles. The basic algorithm only requires seasonal values of the target variable on a grid (though the algorithm can also be extended to calibrate the dynamical and thermodynamic components separately; see O'Reilly et al. (2020)) CALL Outputs: Like ASK and KCC, users need to be comfortable with interpretation and processing of large sample output.
- UKMO method is not available as source code – only the maps of the projections and the underlying (netCDF) data produced by the method will be available.

10.3 Coherency of projections across geographical space and variables

The various climate projection methodologies represent different approaches to spatial and physical coherency of their outputs. Being aware of these differences can inform which method is more appropriate for particular applications.

We can illustrate these differences using weights from two methods, REA and ClimWIP. REA provides a weight for a particular model based on the specific location and variable that it is applied for. So, in the REA Atlas data the weights are calculated for each individual grid box and for temperature and precipitation individually. In contrast ClimWIP calculates its weights for individual models based on their performance over the whole domain and for multiple observational variables, at the same time. The consequences of this can be seen in Figure 36, which illustrates what the weights look like for a single climate model.

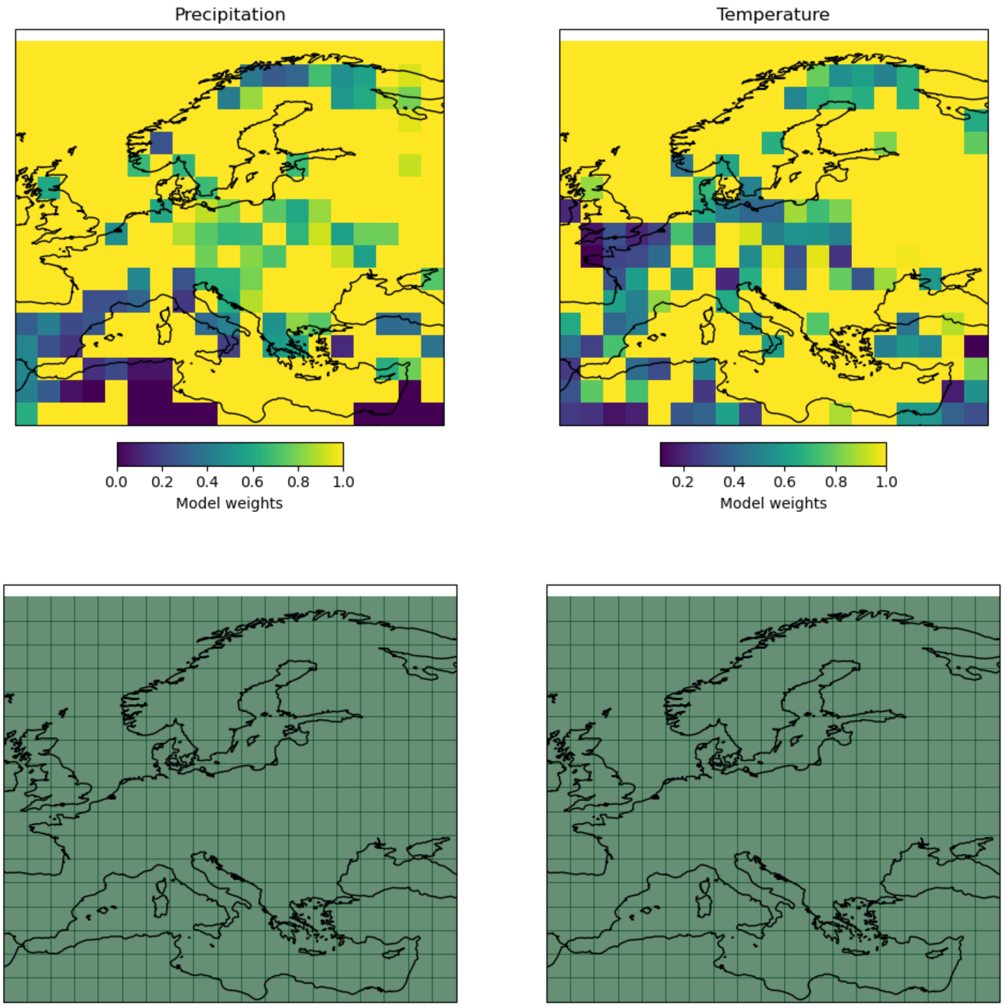


Figure 36*: The comparison of weights for a single CMIP6 model (CNRM-CM6-1) based on the REA (top row) and ClimWIP (bottom row) for summer precipitation (left column) and temperature (right column).*

This particular model can be given either a low or high weight in REA, based on the variable and location, whereas a single weight is given to this model in ClimWIP. Both methods are comparable in terms of their ability to capture more skilful climate projections (see Perfect Model analysis) but the differences in spatial and physical coherency can have implications for studies that want to identify particular models for downstream assessment and impact.

For example, REA will give this particular model a low weight for summer rainfall changes in southern France, but a high weight in central France. Similarly, REA will down weight this model in northern France for summer temperature change, but central eastern France will still be weighted highly. In both cases, the weighting is picking up where there is a large bias, or not, for a particular region and variable. This can be useful should a particular application want to focus on the local scale bias (which can often be an important consideration for using climate model outputs to drive other impacts models). It can prove problematic, however, should an application want to extend the analysis to more than one variable (because formally highly weighted simulations may no longer be highly weighted) or when considering changes over larger regions. There is nothing to stop REA being applied to larger regions (as was done in the Perfect Model analysis, for example) or including a weighting based on more than one observational variable. However, what has been done for the Atlas data is weighting of available projections at a grid box and single variable basis. In contrast ClimWIP provides a single weight to a particular model, over an entire domain, based on multiple metrics (i.e., informing variables), and potentially also for different target variables. The strength of this approach is that is can be used to identify spatially and physically coherent realisations with high (or low) weights. It can also be used to weight target variables for which no observations are available by basing the weights on other variables (which a plausible impact on the target) for which observations are available. To give a few examples of where this might be useful, climate projections with stronger warming tend to be linked with enhanced winter rainfall in parts of Northern Europe but stronger reductions in summer rainfall in parts of the Mediterranean. The underlying climate model simulations capture these relationships and by providing a single weight for a model (across space and variables) enables these relationships to be retained in the weight projections. Another example might be selecting a plausible high end warming model that captures how temperature changes in both the Mediterranean and the Baltic. The trade-off of this approach is that it provides a weighted metric based on the whole large-scale region and for a number of variables, so it will not necessarily capture whether a particular model reproduces past changes at smaller scales or for single variables.

ClimWIP and UKCP methodologies both provide spatially and physically coherent projections (i.e., they retain the information in the underlying climate simulations on the spatial and physical correlations). REA, ASK and KCC outputs (as implemented in the Atlas) are all based on evaluation against single variables, individually. Spatially, REA provides local weights; ASK is based on pan-European information (so are spatially coherent at the larger scale) and KCC uses both a local and a global constraint, which puts it somewhere in between.

10.4 Summary information

Table 6 contains a top-level summary on the available outputs and a number of key characteristics that identify particular strengths of each method.

Table 6: *This table indicates what information is publicly available, for each methodology as well as summarising properties of the method which might influence whether a particular method is appropriate for a potential application.*

Method	**Which products are available?**			**What are the key characteristics of the method?**		
	Code/methods for user to apply in bespoke application	*Pre-calculated results – maps of upper and lower bounds at each grid point for Europe, based on CMIP5*	*Pre-calculated results: Model weights*	*Method is multi-variate (i.e.the model weightings apply across several variables. This is important where we are interested in how changes in several variables (e.g. temperature and precipitation) change together.*	*Method includes additional uncertainties (parameter uncertainty, carbon cycle uncertainties) and capture more high-end, but low likelihood outcomes*	*Method can result in wider uncertainty range than original model spread, if supported by observed change*
UKCP	x	√	x	√	√	x
REA	√	√	√	x	x	x
ClimWIP	√	√	√	√	x	x
ASK	x	√	x	x	x	√
KCC	√	√	x	x	x	x
CALL	x	(√)	x	x	x	√

10.5 Hypothetical cases – to expose potential decision making on which methods to adopt

The Atlas presents a large dataset of climate projections, based on a variety of different currently available methodologies – which arises out of the different needs and requirements across users. We recognise that for many users, the range of Atlas outputs can present a barrier to adopting the outputs within their own decision making. Current users of the EUCP constrained climate projections include the recent IPCC WG1 report, climate scientists working on multiple lines of evidence and those involved in planning for future CORDEX simulations. However, we believe that this data should have wider relevance and usage outside the immediate climate scientist circles. To address this, here we present four hypothetical user cases, each of which provide some signposting for how Atlas projections might be more widely deployed, depending on a particular user's familiarity with the datasets and their own requirements.

Hypothetical User 1

Xinyi Ó Broin
Government employee
Persona Icon Vectors by Vecteezy

About the User: This user works in a government department of a European nation, with a focus on long term local development planning. They have a working familiarity with headline messages coming out of climate science (such as produced in IPCC reports). They do not necessarily know how these headline messages relate to their particular part of Europe nor do they have the understanding of the underlying science that informs climate projections nor the technical capabilities or time to analyse them.

Their interest in climate data: This user's role is to raise awareness about potential climate change impacts as part of wider longer-term planning. This is to inform the engagement with local authorities within the country. This country is one of many European nations that does not yet have established National Climate Scenarios. They would like to make use of graphical representation that represents the magnitude of climate change within their region. This is not intended as a decision-making tool but as a way to raise awareness of potentially climate change impact. As such, they are looking for an off the shelf visualisation based on data and analysis that would have credibility in scientific circles.

Their decision process: The EUCP Atlas provides maps of projected European climate change for a time period centred on 2050. This timescale matches the roughly 30-50 year planning horizon of the infrastructure planning that this user is hoping to inform. They could opt to just represent the central estimate from the latest generation of climate simulations (by opting for an unconstrained CMIP6 estimate from one of the methods in the percentile maps in the Atlas). However, they decide to use visualisations from one of the estimates that weighted projected future changes against how consistent their historical simulations were with observed climate changes by opting for a constrained CMIP6 estimate from one of the methods in the percentile maps in the Atlas. This user is not interested in the particular strengths and limitations of individual climate change methodologies and is initially uncertain about which visualisation to choose. They are encouraged that central estimates from all the represented methods agree on the broad messages around most likely changes. The focus on awareness raising, rather than informing decisions, makes most of the quantitative differences between methodologies less relevant to this user. In the end they choose one which is well referenced, so they can point to these should the details be challenged at a later stage.

For most of the cases, this user opts to make use of just central estimates (50 percentile maps in the Atlas). However, in discussion around planning for future water availability this user also opts to include maps of lower likelihood but high-end changes as the projections suggest quite substantial reductions in summer rainfall may be possible in this particular user's country (10 percentile rainfall maps in the Atlas)

Hypothetical User 2

About the User: This user is a climate scientist working on storm surge modelling as part of a wider EU consortium project. She works with output from climate simulations as a core part of her work and has colleagues who have contributed to the development of these models. She is technically literate and able to handle, process and visualise climate model output as part of her day-to-day work.

Sheridan Lager
Climate scientist
Persona Icon Vectors by Vecteezy

Their interest in climate data: She is currently tasked with running a number of coupled regional surge-wave models of the North Sea (to explore changes in storm surges). These are computational very expensive to run, and she can only afford a run a small handful. So, her initial task is to select a subset of climate models to use as input to the surge-wave models. She wants this selected subset of climate simulations to be able to capture central, low- and high-end estimates of projected future climate change - as each of these speaks to different messages. Whilst the central estimate indicates the more likely projected change, she also wants to inform what the minimum and maximum adaptions pathways might need to look like.

Their decision process: Initially, she makes use of the output from the various WP2 methodologies (see the Atlas) to explore projected low (10 percentile), central (50 percentile) and high end (90 percentile) changes. She can compare these constrained ranges with the changes in the individual climate models that she has already computed (such as is done in Figure 33). This enables her to identify candidate models that might fit her criteria and span the range she requires.

She found that all the CMIP6 based methods tended to down weight the upper temperature change (often use as a metric of the magnitude of climate change) bound compared to the standard CMIP6 dataset. This helped her identify models with slightly smaller magnitude responses that would better represent a plausible high-end scenario. The impact of observational constraints was more modest for the central and low end estimated changes. However, the observational constraints are also effective at ruling out poor simulations even where doing so does not impact the overall range of projected changes. Two of the methodologies (REA and ClimWIP) both provide weights for each model that they use in their assessment (which are provided in the Atlas, via the data download option). These weights can be used as a measure of plausibility of each individual climate simulation. In the end, she chooses to use weights from the ClimWIP method for two reasons. The first was because ClimWIP provides a measure of model plausibility across a range of observational variables (including circulation) whereas the off-the-shelf implementation of REA (see the Atlas) based these weights on only observations of temperature and precipitation for projections of each of these variables, respectively. For the storm surge modelling, the plausibility of the circulation and winds are the most important variables and whilst ClimWIP does not target these two variables in isolation, its wider assessment of plausibility across a number of variables can be hoped to rule out generally poor climate simulations. Secondly and relatedly, ClimWIP retains coherency across variables and geographical locations. In contrast REA's weights are based only on how well a particular simulation reproduced observations for the target variable, and so REA may not give the same measure of good or poor models if it was to be applied to a different variable.

So, she ended up using the pdf output from all CMIP6 based methods (see the Atlas) to identify potential candidates for boundary conditions for her surge-wave models, that represented central, low- and high-end climate change. She extended this analysis by using individual model weights from ClimWIP to ensure that the choice of candidates would not be ruled out as poor comparisons compared to the observed changes. The implementation of the ClimWIP methodology in ESMValTool was seen as an added advantage is it would enable this method to be rerun for new climate model data, should the requirements for new storm surge modelling emerge later in her project.

Hypothetical User 3

Thanh Keil
Hydrological modeller
Persona Icon Vectors by Vecteezy

About the User: This user is a hydrological modeller, working to link climate change impact in a small-scale regional river basin in Europe. Her role has many technical and scientific challenges, and she is comfortable processing analysing geospatial data but is not overly familiar with climate simulations. Her experience with them comes from reading hydrological impact papers, where they are often used as a boundary condition.

Their interest in climate data Like user 2, she is looking to use climate model output to drive her hydrological modelling. Most of her applications are envisaged to revolve around a central estimate climate scenario of the hydrological impacts. This is intended to be used for detailed impacts analysis by users of her hydrological outputs. The presence of biases in rainfall in the climate model data used as boundary conditions is an issue for her approach. Information on where the projected changes in rainfall in the central estimate scenario, sit within the wider uncertainty ranges would be important context for users of her data. Her users will be able to relate that to their own risk tolerance for their own dynamic adaptation strategies.

Their decision process Initially she uses the Atlas to assess central estimates for projected climate change looking at precipitation (as the main variable of interest) and temperature changes (as a metric of the magnitude of regional climate change). These estimates qualitatively agreed on the broad changes, whilst differences exist in the magnitude of some regional details. Next, she wanted to make use of model weights to help identify an individual climate model configuration to use as her baseline estimate. Two methods provide weights to individual users, REA and ClimWIP. Due to issues that would be introduced should there be large historical biases, between the driving boundary data (climate model used as the input) and observations, she opts to use REA as the model weights explicitly target present day rainfall bias. Using these weights, she identifies one of the more highly weighted climate models that lies close to the (multi-method) central estimates, which she can then pull off the Climate Model Intercomparison Data Archive

Having produced her own hydrological impact simulations (based on a central climate change estimate scenario) and started engagement with users of her data, she returns to the Atlas as a way to provide context on her top-level messaging. The climate simulation that she is using as her central estimate of climate change shows -10 to -20% rainfall reductions in the summer season. First, she wants to look at the level of agreement in the central estimates and notes how her chosen central estimate compares to central estimates from the range of WP2 methods. Secondly, she looks at how well the methods agree on the upper bounds. On the lower bounds the methods suggest a range between neutral and even a small moistening, but she can discount these differences as they lie well within the historical variability. The upper bounds (10 percentile rainfall maps in the Atlas) suggests stronger drying is possible, with reductions up to -50%, but there is less agreement on the magnitude of this reduction across methodologies. This message forms an important context strand in her communication of her own results. Users of her data will relate these messages to their own risk tolerance, which can inform their own dynamic adaption strategies. Flexible adaption strategies are often based around a plan made for a central estimate of change, but the high end is used to (a) scan for options/risks and (b) used for monitoring to assess whether a plan B would be needed. The more risk adverse may well come back at a later stage to request new hydrological scenarios that explore this high-end impact.

Hypothetical User 4

About the User This user works as a financial analyst. Highly numerate and technically competent, he has a broad range of analytical skills, toolsets and expertise but lacks a background in climate science.

Rajinder Järvinen
Financial analyst
Persona Icon Vectors by Vecteezy

Their interest in climate data: The company is looking to update its own guidance on the management of its extensive building portfolio, including questions of divestment and further investment. These assets are distributed across both eastern and western Europe. They are looking for guidance and/or data so that their own analyst team can stress test their building portfolio. They have their own internal estimate for the vulnerability of their assets to climate. These assets could be particularly exposed to either flooding, high temperatures or windstorms. Costs are disproportionately skewed for higher end impacts, and so any cost-benefit analysis has tended to be sensitive to estimates of where the plausible upper bounds are.

Their decision process: The analysts have previously taken climate model data straight from the current climate model archives. CMIP6 has provided a bit of shock to their assessment process as it includes a larger number of simulations that are above what would previous have been seen as the upper bound (based on earlier climate model archives), and as such appears to change the risk assessment across their portfolio.

The CMIP6 based methodologies are useful in that they suggest that some of the high impact summer temperature responses in CMIP6, whilst still plausible, are less likely. There appears to be some consensus about this from the different methods, though these methods disagree, spatially, on where they down weight the stronger climate change signals (see Figures 13 and 16 in Section 8). There is less consensus in the CMIP6 based methods on the changes in rainfall (Figure 14) - with clear but noisy signals in REA compared to little evidence of information (one way or another) in ClimWIP. Because he is looking for information to inform their internal pan-Europe risk assessment, he opts to use ClimWIP in favour of REA because it retains spatial coherency (see discussion of spatial coherency in Section 10.3).

The Atlas also provides a new strand of evidence that his previous experience with raw climate model data has not exposed him to. The Atlas includes climate projection data from the UKCP methodology, which produces wider probability distributions for projected European changes. This is because it also accounts for both wider modelling and carbon cycle uncertainties that are not accounted for the main experimental design in the current climate model archive. Including these additional uncertainties implies that some of the higher end impacts may be more likely than projections based solely on CMIP5 or CMIP6 data (including associated probabilistic methodologies) would suggest. This again, provides new context for their existing 'in-house' risk assessments.

11. Lessons Learned, links built, and challenges faced

Deliverable 2.3 represents the cumulation of the WP2 Task 2.2 activity over the course of the project. Much of the work reported here, post-dates the previous Deliverables (D2.1 and D2.2, November 2019) and has occurred over the covid period. This has influenced our work in a number of ways, presenting both challenges and opportunities.

Perhaps the greatest achievement of this work package has been its ability to systematically analyse output of our six observationally calibrated climate projection approaches (summarized in Brunner, 2020a and Deliverable D2.2). This was the first time an approach like this has been achieved on this scale and exposed differences (and commonalities) in climate projections produced by different potential approaches, that European users might chose to adopt. This was achieved because we were able to sit down as a group and plan together how we'd approach this over several days (at two workshops in ICTP and SMHI). The arrival of Covid has meant that we did not have this luxury, in the 2nd half of the EUCP program. This limited the kind of brain storming activities, which are much harder to reproduce using online video-conferencing tools (where people's focused attention can be hard to maintain for long periods, within a virtual environment).

We have still achieved a lot, over this second period, but it has not been as coordinated. For example, we were able to control for a common set of Climate Simulations in the pre-Covid (Brunner 2020a) published analysis, whereas different groups contributing to the Atlas' Climate Projections presented here (See Sections 4 and 7) did not have the same level playing field. The respective groups started their analysis at different stages, capturing different samples of underlying CMIP6 models (that affected the number of simulations they were able to capture, between 15 to 41).

As well as the challenges there have been benefits to covid driven changes in how we work. We were able to set up and maintain regular Science Coffee talks, where WP2 (and interested parties from other work packages) were able to share science, as their work progressed. For many of us, this regular informal and formal contact brought a sense of a WP2 group that spanned our geographical locations (and was just as important as the institutional working groups might have been prior to Covid). These changed the dynamic (from the original dynamic of PIs disseminating and coordinating work in the various institutes, to a more open and transparent one where work was more visible across the work package). This particularly benefited postdocs, working in the work package, by giving them more visibility and agency.

Another success has been seeing how publication of both data outputs and code has facilitated wider engagement with our work, outside the work package. The publication of both have gone beyond the expectations of many involved and has required considerable work. However, having done so, it has changed the nature of many of our interactions. For example, we have been talking with Task 5.5 for some time about their work looking at using information from multiple strands of evidence to provide context. However, it was only when our data became available via the public facing data portal (linked via the Atlas) that we saw wider adoption of WP2 outputs in their approach. In principle, this data was always available to them, but the informal nature of any exchange was a larger barrier than either party recognized.

11.1 Links with the wider consortium

The work documented here has connections with wider EUCP activities, on a number of important strands. Within WP2, there are connection between the probabilistic methodologies developed, explored, and published here, with the Climate Narrative approaches explored in D2.4 deliverable. For example, the approach taken by Anna Merrifield at ETH of using ClimWIP (one of our probabilistic methods) as a tool to identify small sub-selections of available climate simulations that users could adopt that are both plausible and explore the diversity of potential future responses. There also links between the probabilistic projections and approaches to identifying climate narrative scenarios, being developed within WP5's multiple lines of evidence (see below).

The two EUCP strands that look at predictability on near (WP1) and medium (WP2) term timescales have very much focused on their own science questions. However, much of WP5's role is to explore how we can blend or combine information from these two strands. This deliverable draws on earlier finalised work in deliverable D5.1. D5.1 explored the information arising from observations in both timescales, with the added value of the observational data on projections, presented here, representing an important part of this work. There are further links with WP5, both in providing context for the spatial blending of climate projection data (T5.4) and contributing an important strand of evidence in developing contextual narratives around future climate change (T5.5).

There has also be links made with WP4 who are working with the user case studies, where we have explore using WP2 performance weights (from ClimWIP) in estimates of European hydrological impacts (Sperna Weiland et al, 2021). With WP6 we have co-developed the European Climate Projection Atlas (described in this deliverable). They have helped us move from addressing the underly science questions to being able to present and publish outputs from our work (both via the data portal and via code availability, software and ESMValTools implementations). With WP6's coordination, we have been engaging feedback from the Multi-User forum on the Atlas implementation, as a way to improve accessibility.

12. References

- Allen, M. R., P. A. Stott, J. F. Mitchell, R. Schnur, and T. L. Delworth, 2000: Quantifying the uncertainty in forecasts of anthropogenic climate change. Nature, 407, 617–620, https://doi.org/10.1038/35036559;
- Brunner, L., Lorenz, R., Zumwald, M., & Knutti, R. (2019). Quantifying uncertainty in European climate projections using combined performance-independence weighting. Environmental Research Letters, 14(12), 124010. https://doi.org/10.1088/1748-9326/ab492f
- Brunner, L., McSweeney, C., Ballinger, A. P., Befort, D. J., Benassi, M., Booth, B., Coppola, E., de Vries, H., Harris, G., Hegerl, G. C., Knutti, R., Lenderink, G., Lowe, J., Nogherotto, R., O'Reilly, C., Qasmi, S., Ribes, A., Stocchi, P., & Undorf, S. (2020a). Comparing Methods to Constrain Future European Climate Projections Using a Consistent Framework. Journal of Climate, 33(20), 8671–8692. https://doi.org/10.1175/JCLI-D-19-0953.1

- Brunner, L., Pendergrass, A. G., Lehner, F., Merrifield, A. L., Lorenz, R., & Knutti, R. (2020b). Reduced global warming from CMIP6 projections when weighting models by performance and independence. Earth System Dynamics, 11(4), 995–1012. https://doi.org/10.5194/esd-11-995-2020
- Coppola, E., Nogherotto, R., Ciarlo', J. M., Giorgi, F., van Meijgaard, E., Kadygrov, N., … & Wulfmeyer, V. (2021). Assessment of the European climate projections as simulated by the large EURO-CORDEX regional and global climate model ensemble. Journal of Geophysical Research: Atmospheres, 126(4), e2019JD032356.
- Donat, M. G., Lowry, A. L., Alexander, L. V., O'Gorman, P. A., & Maher, N. (2016). More extreme precipitation in the world's dry and wet regions. Nature Climate Change, 6(5), 508-513.
- Drobinski, P., Alonzo, B., Bastin, S., Silva, N. D., & Muller, C. (2016). Scaling of precipitation extremes with temperature in the French Mediterranean region: What explains the hook shape?. Journal of Geophysical Research: Atmospheres, 121(7), 3100-3119.
- Fantini, A., Raffaele, F., Torma, C., Bacer, S., Coppola, E., Giorgi, F., Verdecchia, M. (2016). Assessment of multiple daily precipitation statistics in ERA-Interim driven Med-CORDEX and EURO-CORDEX experiments against high resolution observations. Clim Dyn 51, 877–900 (2018). https://doi.org/10.1007/s00382-016-3453-4
- Frei, C., Davies, H. C., Gurtz, J., & Schär, C. (2000). Climate dynamics and extreme precipitation and flood events in Central Europe. Integrated Assessment, 1(4), 281-300.
- Forster, P., T. Storelvmo, K. Armour, W. Collins, J. L. Dufresne, D. Frame, D. J. Lunt, T. Mauritsen, M. D. Palmer, M. Watanabe, M. Wild, H. Zhang, 2021, The Earth's Energy Budget, Climate Feedbacks, and Climate Sensitivity. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu and B. Zhou (eds.)]. Cambridge University Press. In Press.
- Grose, M.R., Narsey, S., Delage, F.P., Dowdy, A.J., Bador, M., Boschat, G., Chung, C., Kajtar, J.B., Rauniyar, S., Freund, M.B. and Lyu, K., 2020. Insights from CMIP6 for Australia's future climate. Earth's Future, 8(5), p.e2019EF001469.
- Hagedorn, R., Doblas-Reyes, F. J., & Palmer, T. N. (2005). The rationale behind the success of multi-model ensembles in seasonal forecasting—I. Basic concept. Tellus A: Dynamic Meteorology and Oceanography, 57(3), 219-233.
- Hegerl, G., Ballinger, A., Booth, B., Borchert, L. F., Brunner, L., Donat, M., Doblas-Reyes, F., Harris, G., Lowe, J., Mahmood, R., Mignot, J., Murphy, J., Swingedouw, D. & Weisheimer, A., Towards consistent observational constraints in climate predictions and projections, (2021) Frontiers
- Knutti, R., M. R. Allen, P. Friedlingstein, J. M. Gregory, G.C. Hegerl, G. A. Meehl, M. Meinshausen, J. M. Murphy, G.-K. Plattner, S. C. B. Raper, T. F. Stocker, P. A. Stott, H. Teng and T. M;. L. Wigley (2008): A review of uncertainties in global temperature projections over the twenty-first century. Journal of Climate, 21, 2651-2663.
- Lee, J. Y., J. Marotzke, G. Bala, L. Cao, S. Corti, J. P. Dunne, F. Engelbrecht, E. Fischer, J. C. Fyfe, C. Jones, A. Maycock, J. Mutemi, O. Ndiaye, S. Panickal, T. Zhou, 2021, Future Global Climate: Scenario-Based Projections and Near-Term Information. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E.

Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu and B. Zhou (eds.)]. Cambridge University Press. In Pres

- Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E. M., Brunner, L., Knutti, R., & Hawkins, E. (2020). Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. Earth System Dynamics, 11(2), 491–508. https://doi.org/10.5194/esd-11-491-2020
- Llasat, M. C., Llasat-Botija, M., Petrucci, O., Pasqua, A. A., Rosselló, J., Vinet, F., & Boissier, L. (2013). Towards a database on societal impact of Mediterranean floods within the framework of the HYMEX project. Natural Hazards and Earth System Sciences, 13(5), 1337-1350.
- Lowe J.A., D. Bernie, P. Bett, L. Bricheno, S. Brown, D. Calvert, R. Clark, K. Eagle, T. Edwards, G. Fosser, F. Fung, L. Gohar, P. Good, J. Gregory, G. Harris,T. Howard, N. Kaye, E. Kendon, J. Krijnen, P. Maisey, R. McDonald, R. McInnes, C. McSweeney, J.F.B. Mitchell, J. Murphy, M. Palmer, C. Roberts, J. Rostron, D. Sexton, H. Thornton, J. Tinker, S. Tucker, K. Yamazaki, S. Belcher (2018) UKCP18 Science Overview Report, Met Office Hadley Centre, Exeter, U.K., https://www.metoffice.gov.uk/pub/data/weather/uk/ukcp18/science-reports/UKCP18-Overview-report.pdf
- Meehl, G.A., Senior, C.A., Eyring, V., Flato, G., Lamarque, J.F., Stouffer, R.J., Taylor, K.E. and Schlund, M., 2020. Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth system models. Science Advances, 6(26), p.eaba1981.
- Merrifield, A. L., Brunner, L., Lorenz, R., Medhaug, I., & Knutti, R. (2020). An investigation of weighting schemes suitable for incorporating large ensembles into multi-model ensembles. Earth System Dynamics, 11(3), 807–834. https://doi.org/10.5194/esd-11-807-2020
- Min, S. K., Zhang, X., Zwiers, F. W., Friederichs, P., & Hense, A. (2009). Signal detectability in extreme precipitation changes assessed from twentieth century climate simulations. Climate Dynamics, 32(1), 95-111.
- Murphy, J.M., Harris, G.R., Sexton, D.M.H., Kendon, E.J., Bett, P.E., Clark, R.T., Eagle, K.E., Fosser, G., Fung, F., Lowe, J.A., McDonald, R.E., McInnes, R.N., McSweeney, C.F., Mitchell, J.F.B., Rostron, J.W., Thornton, H.E., Tucker, S., Yamazaki, K. (2018) UKCP18 Land Projections: Science Report, Met Office Hadley Centre, Exeter, U.K., https://www.metoffice.gov.uk/pub/data/weather/uk/ukcp18/science-reports/UKCP18-Land-report.pdf
- O'Reilly, C. H., Befort, D. J., and Weisheimer, A.: Calibrating large-ensemble European climate projections using observational data, Earth Syst. Dynam., 11, 1033–1049, https://doi.org/10.5194/esd-11-1033-2020, 2020.
- Palmer, T.E., Booth, B.B.B. and McSweeney, C.F., 2021. How does the CMIP6 ensemble change the picture for European climate projections?. Environmental Research Letters, 16(9), p.094042.
- Pfahl, S., O'Gorman, P. A., & Fischer, E. M. (2017). Understanding the regional pattern of projected future changes in extreme precipitation. Nature Climate Change, 7(6), 423-427.
- Ribes, A. et al. (2021): Making climate projections conditional on historical observations. Science Advances, 7, eabc0671, DOI: 10.1126/sciadv.abc0671
- Stott, P. A., and J. A. Kettleborough, 2002: Origins and estimates of uncertainty in predictions of twenty-first century temperature rise. Nature, 416, 723–726, https://doi.org/10.1038/416723a
- Schurer A., Hegerl GC, Ribes A, Polson D, Morice C and Tett SFB (2018): Estimating the transient climate response from observed warming. J Climate, 31, 8645-8663.)
- Sperna Weiland, Frederiek C., et al. "Estimating Regionalized Hydrological Impacts of Climate Change Over Europe by Performance-Based Weighting of CORDEX Projections." Frontiers in

Water 3 (2021): 713537. Tebaldi C. and Knutti R. (2007): The use of the multi-model ensemble in probabilistic climate projections. Phil. Trans. R. Soc. A., 365, 2053–2075

- Tokarska K., Hegerl G.C., Schurer A.P., Forster P. and Marvel K. (2020): Observational Constraints on the effective climate sensitivity from the historical record. Environ. Res. Lett. 15 (2020) 034043 https://iopscience.iop.org/article/10.1088/1748-9326/ab738f/pdf
- Qasmi, S. and Ribes, A. (in revision): Reducing uncertainty in local climate projections Nature Communications DOI: 10.21203/rs.3.rs-364943/v1

Appendix

A Bayesian Method for Probabilistic Climate Projections for Europe

G.R. Harris, J.M. Murphy, D.M.H. Sexton, B.B.B. Booth
Met Office Hadley Centre, Exeter EX1 3PB, UK

<u>Short summary of method</u>

A Bayesian approach has been used to produce probabilistic projections of the European climate response for the RCP85 emission-driven scenario. To achieve this, a large Monte-Carlo integration over defined ranges of uncertain model parameters is performed, weighting the sampled outcomes in the integration by relative likelihood (estimated using a set of specified observables) to produce observationally constrained pdfs. In order to make the estimation computationally feasible, predicted outcomes are sampled from a fast statistical emulation of the equilibrium response to doubled CO_2, calibrated to the response of a large perturbed parameter ensemble (PPE) of simulations based on a single model. A scaling approach calibrated to the transient response of a second PPE of earth system model variants allows the transient response to be inferred, while also enabling sampling of other uncertainties associated with the carbon cycle, ocean and aerosol components of the climate system. A third ensemble of CMIP5 simulations is also used to sample the structural error component associated with the PPE base model.

<u>Key Publications</u>

Murphy *et al.* (2018), Harris *et al.* (2013), Sexton *et al*. (2012), Sexton and Harris (2015).

<u>Details of Method</u>

The methodology used here is based on the Bayesian statistical framework that underpinned a previous set of UK Climate Projections (UKCP09, Murphy *et al*. 2009). Details of the approach are given in Sexton *et al.* (2012) and Harris *et al*. (2013). This methodology has been subsequently updated and used to produce a new set of probabilistic climate projections for the United Kingdom (Murphy *et al*. 2018). Key elements of the method, including recent updates are summarized below (see preceding references for additional detail).

Modeling uncertainty is explored using the Perturbed Parameter Ensemble (PPE) approach (Collins *et al*. 2006, Murphy *et al*. 2007), where key parameters controlling processes in the atmosphere, surface, ocean, aerosol and land carbon cycle components of the climate system are varied within expert-defined “prior” parameter spaces for a single climate model. Here, variation in historical climate and the equilibrium response to a doubling of CO_2 concentrations is explored using a relatively large 280 member PPE based on the atmosphere-mixed-layer (hereafter SLAB) configuration of the HadCM3 model (Pope *et al*. 2000, Johns *et al*. 2003). Multivariate regression relationships trained on the 280 PPE variants are then used to construct a statistical emulator for the historical and future equilibrium responses. In contrast to use of multi-model ensemble (MME) output, emulation based on PPE provides a systematic and comprehensive sampling of climate response for untried parameter combinations, and makes use of a Bayesian approach both possible and practicable.

Users are impacted by the transient climate response to scenarios of climate forcing rather than the equilibrium response, so a second emulation stage is implemented. Assuming the climate response is proportional to global mean surface temperature (GMST) change (Santer *et al*. 1990), scaling techniques are then used to translate the emulated equilibrium response into estimates of past and future climate response, for any point in parameter space. Given emulated predictions for the climate feedbacks, a Simple Climate Model (SCM) (Harris *et al*. 2013, Suppl. Mat.) is used to predict the transient global temperature response and perform the scaling.

Uncertainty in modeling the rate of CO_2 uptake by the land and ocean biogeochemical systems contribute substantially to uncertainties in projections of GMST, which in turn influence regional response (Knutti *et al*. 2008, Booth *et al*. 2012). Here we therefore use an emission-driven approach (rather than concentration-driven) in order to take fuller account of known limitations in the current modeling of earth system processes. To this end, a 57-member PPE, based on variants of the HadCM3C Earth System Model (ESM) has been used to explore modeling uncertainty in response to RCP85 forcing, accounting for interactions between different earth system components (Lambert *et al*. 2013, Murphy *et al*. 2014). This ensemble (hereafter the ESPPE) is used to calibrate the SCM and provide prior distributions for key earth system processes, such as ocean heat uptake and climate-carbon feedbacks, as well as providing adjustment of potential differences between SLAB and transient regional response.

To produce probability distributions for climate response, Monte-Carlo integration is performed over the prior space of model and SCM parameters. Emulation is used to estimate response, allowing large sample sizes of order 10^6 to be produced, hence improving the coverage of parameter space. Some parameter choices perform better than others when comparing, for example, their predictions of historical climate. Within the Bayesian framework, we include a specified multivariate set of observables in the set of predicted variables, and estimate a “likelihood” for each parameter set (model variant), given the observed data. Each variant is weighted by relative likelihood in the Monte-Carlo integration, to produce updated posterior predicted probability distributions for climate response (Harris *et al*. 2013). The observational data used for likelihood estimation is specified below.

One important component of the method is the recognition that models are imperfect. Given that some parameter choices are better than others at reproducing observations, it is reasonable to assume that there exists a “best input” set of parameter choices that provides the best simulation of true climate. However, due to model imperfection, even the best-input models will possess an irreducible structural error component (termed “discrepancy” in Sexton *et al*. 2012) that cannot be eliminated. We estimate the effects of such structural errors by using output from other independent climate model simulations, searching the prior space for best-input parameter sets that best reproduce output for selected CMIP5 models (Taylor *et al*. 2012). Since we account for carbon-cycle modeling uncertainty, emission-driven CMIP5-ESMs are used for this purpose. We do not use observations to estimate structural error because (a) we don’t have observations of the future, and (b) in order to avoid any issues associated with double-counting, since the observations are also used to constrain the projections.

We are obliged to develop the complex projection methodology described above since there is insufficient computational resource to produce sufficiently large, fully-coupled transient ensembles of scenarios projections to provide robust, comprehensive and yet plausible samples of projections for future climate change. We are careful therefore to validate the components of the method, and include the additional statistical uncertainties that arise from the different steps. These include: equilibrium response emulation error, error in converting from equilibrium to transient response, time-scaling error (including inherent model internal variability), and structural error estimates.

<u>Developments going from ENSEMBLES to EUCP</u>

The methodology described in Harris *et al*. (2013) was developed for the UKCP09 projections released in 2009. Subsequently, the same methodology was used to produce European projections (Harris *et al*. 2010) as part of the ENSEMBLES project (van der Linden and Mitchell 2009). More recently, the method has been developed to provide the UKCP18 projections (Murphy *et al*. 2018), and now projections for EUCP. It is useful to review and highlight the main differences in methodology and scope compared to these earlier projections.

- A new 57-member ESPPE ensemble of HadCM3C earth system models has been used to calibrate the projections. This simplifies the method by allowing the number of PPE inputs to be reduced from seven to three, and 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.
- The ESPPE ensemble allows full incorporation of the land carbon cycle into the Bayesian methodology, with definition of a prior parameter space, and likelihood estimation that includes the carbon cycle response.
- Rather than using the older slab-ocean CMIP3 simulations (Meehl *et al*. 2007) to estimate the structural error component, we now use 12 more recent CMIP5-ESM simulations (Taylor *et al*. 2012).
- Structural errors in predicting non-linear aspects of the transient GMST response (such as temporal effects which cannot be explained by use of a fixed climate feedback parameter) have additionally been accounted for.
- Additional 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).
- The methodology has been extended to present the probabilistic projections for individual years, rather than the 20 or 30-year averages of ENSEMBLES and UKCP09. This is based on the method of Sexton and Harris (2015).
- For UKCP18, we scaled percentage change in precipitation to predict the transient response. For the southern regions of Europe that are predicted to experience stronger summer drying compared to the UK (especially under RCP85 forcing toward the end of the century), this approach in some instances leads to unphysical scaled projections of less than -100%. We have therefore employed a mixed approach here, using the log transform for realizations with strong drying, while for realizations with weak drying or increased precipitation, linear scaling of percentage change is used (Watterson 2008).
- Unlike the UKCP09 and UKCP18 projections, here we do not statistically downscale the projections to a higher resolution of 25km. In this respect, the EUCP projections are like the ENSEMBLES projections, with a finest grid-point resolution of 2.5°×3.75°.
-

Data

Three ensembles are used to calibrate the methodology:

- 280 1×CO_2 and 2×CO_2 PPE equilibrium simulations with the HadSM3 model (Pope *et al*. 2000, Sexton *et al*. 2012)
- 57 PPE simulations with the coupled atmosphere-ocean earth system model HadCM3C (Lambert *et al*. 2013, Murphy *et al*. 2014), simultaneously varying a total of 54 parameters in the atmosphere/land surface, ocean, sulfur cycle and terrestrial carbon cycle components. Flux-adjustments are used (Collins *et al*. 2011), and emission-driven RCP85 scenarios from 1860 to 2100 have been produced.
- 12 CMIP5-ESM emission-driven RCP85 simulations. A total of 15 models in the CMIP5 database were initially considered, but three were omitted due to lack of data for some variables, or due to a lack of independence. The 12 models are: bcc-csm1-1, bcc-csm1-1-m, BNU-ESM, CanESM2, CESM1-BGC, GFDL-ESM2G, HadGEM2-ES, inmcm4, IPSL-CM5A-LR, MIROC-ESM, MPI-ESM-LR, and MRI-ESM1. All output is regridded to the 2.5°×3.75° HadCM3 grid, and this is the highest resolution for which we can produce probabilistic projections for EUCP.

Observational data from a variety of sources is employed (see next section). The SCM used to perform the scaling is described in the Supplementary Material for Harris *et al*. (2013).

Observational Constraints Applied

Observational constraints are applied by weighting sampled outcomes by likelihood weights calculated from the multivariate distance between emulated estimates of a set of historical variables and verifying observations. These observations include the same set of seasonal climatological spatial fields used for UKCP09 (Sexton *et al*. 2012) comprising of twelve climate variables: sea surface temperature, screen temperature, precipitation, TOA outgoing shortwave flux, TOA outgoing longwave flux, TOA shortwave cloud radiative effect, TOA longwave cloud radiative effect, sea-level pressure, relative humidity, total cloud, surface sensible heat flux, and surface latent heat flux. Data sources and references for each of these are listed in Table B.1 of Murphy *et al*. (2018). The data amounts to about 175,000 observables and it is necessary to reduce its dimensionality, in order to remove dependencies between variables, and to make the multivariate statistical calculations computationally feasible. This is done by identifying the six leading eigenvectors of these climatological variables in the SLAB ensemble, and emulating the amplitudes for these. They are then compared with amplitudes for these observables projected onto the same set of eigenvectors, and used to estimate associated likelihood weights.

In addition, historical trends for several climate indicators are also included in the set of observational constraints. These include the Braganza indices based on GMST (Braganza *et al*. 2003), heat content change in the top 700m of the oceans, and change in atmospheric CO_2 concentration over a recent 45 year period (Booth *et al*. 2017). Observational data sources and references are listed in Table B.2 of Murphy *et al*. (2018).

Key Assumptions

- The Bayesian statistical framework as presented in Rougier (2007) is assumed as the basis for our methodology.
- The true climate is assumed to lie within the spread of sampled prior outcomes, constructed from PPE and MME output.

- It is assumed that the structural error of the base model for the PPE can be estimated by taking MME simulations as proxies for the true climate and using our best input emulations for these models to specify structural error.
- There is a degree of subjectivity and judgment in specifying the observational data used to constrain the projections. Alternative variables, data sources and processing of the observational data could be used or implemented.
- It is assumed that models that are good at simulating historical climate and historical trends will be good at predicting future climate. Weighting by likelihood tends to support projections that match these observations (although imperfectly since constraints are multivariate).
- The patterns of equilibrium response are assumed to be representative of the fully-coupled response patterns. This assumption is validated in a subset of cases, and adjustment applied. Note that in the case of 10m UK wind-speed response, this assumption was found not to apply in a substantial number of cases, thus invalidating use of the method for this particular variable.
- Transient responses are assumed to scale linearly with global temperature response. This assumption is tested for the subset of ESPPE simulations, and any potential non-linearity in the residual error is resampled and included in the projections.
- For the purposes of time scaling, climate feedbacks are assumed to be constant and not evolve with climate state.
- The variability in the ESPPE, which is used to specify variability in the probabilistic projections, is assumed to be a good representation of variability in the true climate. Sensitivity to this assumption was tested in Sexton and Harris (2015) for the UK regions, by implementing a rescaling of variability to match observed variability. Conclusions were not altered substantially by this. This approach has not been implemented yet for the EUCP projections, but could potentially be applied later in the project.

<u>Limitations</u>

- The predicted pdfs can only cover the range spanned by the prior distribution, so if the real world response is outside this, it will not be captured in the pdfs (see next bullet).
- The inclusion of the structural error component can mitigate against overconfidence and lack of spread, but does not account for the unknown effects of errors common to all models (e.g. due to resolution, or missing processes). Note though that this issue of common model biases applies to all model-based projections, and not just our Bayesian methodology.
- If the overlap between the PPE and MME range of response is small, then our best-input emulations are likely to be poor, leading to large structural error. Structural error adjustments may then dominate the modeling uncertainty component, leading to sensitivity in the pdfs to the small sample of MME outcomes (i.e. a lack of robustness). Even if structural errors are relatively small, since the number of MME simulations is limited (12 in this study), the structural error estimation may be sensitive to choice of models.
- Large structural error adjustments along with accumulated statistical uncertainty may lead to broad pdfs and outcomes in the tails that are not supported by any simulations, reducing their credibility. For example, for UKCP18 a large majority of ESPPE variants predicted reductions in UK relative humidity, while most CMIP5-ESMs predicted the

converse, an increase. The resulting large structural error adjustment led to relative humidity pdfs that lacked credibility, so these were not provided (Murphy *et al*. 2018). Similar screening checks were done for all UKCP18 variables, and most passed the tests, including all precipitation and temperature variables. Corresponding checks will be carried out subsequently for the EUCP precipitation and temperature projections provided here.

- Not all sources of uncertainty have been explored in our PPE, including modeling uncertainty in the ocean biogeochemistry component that influences ocean CO_2 uptake, uncertainty associated with forcing from minor gases (methane, nitrous oxide, ozone, CFCs), and variations in solar and volcanic forcing.
- The pdfs are conditional, and depend on many things, including the models and understanding available at the time of their construction, prior assumptions regarding choice and range of uncertain model parameters, limited sample sizes for the PPE and MME ensembles, the choice of observational constraints, choices for input emissions and forcings, and methodological assumptions (e.g., emulation and scaling techniques, use of flux adjustments in the ESPPE, method for structural error estimation, etc). New models and observations, and updated methods will lead to evolution in the predicted pdfs.

Example Projections

The methodology has been used to produce all Tier I projections (large medium and small spatial aggregations for the JJA season). In addition, a set of Tier II projections has been made for the additional WP3 domains, for all four seasons, although no projections are available yet for tasmin and tasmax. Since projections are made on annual timescales, alternative baselines and future time-averaging periods are straightforward to produce. These projections are still provisional and require further calibration and validation, so although close to final, they are still subject to future potential revision. The final output of the methodology is a large sample (3000 here) of realizations constrained by observations that combine modeling uncertainties in the mean signal with internal variability from the input ESPPE simulations. The sample realizations can then be combined to produce probability distributions for projected climate response.

Figure 1 below shows an example of the projections: the summer surface air temperature response to the RCP85 emissions scenario for the three European SREX regions, plus the combined European region. The three coloured realizations are individual examples of the 3000 realizations in the full sample, which are represented in the Figure by selected quantiles of the distribution of response, evolving as grey plumes of uncertainty through the 21st century. There is nothing special about the three coloured realizations, except they have been selected to show a spread in response, picking EUR-SREX realizations close to the 10%, 50% and 90% quantiles at the end of the century. Summer warming is greater in southern regions of Europe compared to the north; for example, a median warming of 7.8°C is predicted for MED-SREX for RCP85, compared to 5.6°C for NEU-SREX.

Figure 2 presents a second example of the probabilistic projections, in this case pdfs for summer precipitation change for the four small WP2 regions for the 20 year mean period 2041-2060, relative to the 1995-2014 baseline, in response to RCP85 forcing. The region aggregates here actually represent 2.5°×3.75° grid-boxes (resolution of our input PPE) closest to the 2.5°×2.5° aggregate regions defined for WP2. For each variable the 3000 realizations are averaged over the selected 20-year future period, and a pdf is fitted to the resulting sample data using kernel density estimation

(KDE) techniques to reduce the effects of sampling noise. The probability distributions are fitted here using the python scipy package 'gaussian_kde' (Jones *et al*. 2001), assuming a normal density function for the kernel and the bandwidth specified using Scott's Rule (Scott, 2015). The pdfs In Figure 2 are compared with climate model data from the SLAB, ESPPE and CMIP5-ESM ensembles that are input to the methodology. The SLAB equilibrium responses have been converted to equivalent transient response by scaling (randomly sampling carbon cycle feedbacks from the ESPPE prior parameter space).

We note that the ESPPE typically predicts a drier response (e.g. for Transylvania) compared to the CMIP5-ESM distribution of responses. The structural error implied by this adjusts the posterior pdfs toward a somewhat less dry response than is obtained in the SLAB and ESPPE ensembles. The correspondence between the GCM data and the pdfs helps confirm that methodological assumptions have not lead to excessive statistical inflation (although some is unavoidable), providing confidence that the pdfs provide a reasonable expression of the uncertainties implied by the underlying set of model simulations. It can be noted too that the spread in response in the 12 member CMIP5-ESM ensemble is substantially smaller than that sampled in the two PPEs. There is a clear signal toward greater summer drying of climate in the southern regions of Europe compared to Northern regions; e.g., the median response for Madrid is -24%, compared a median increase of 5% for Svealand. Uncertainty is large however, and although a majority of realizations give reductions in rainfall, increases cannot be ruled out.

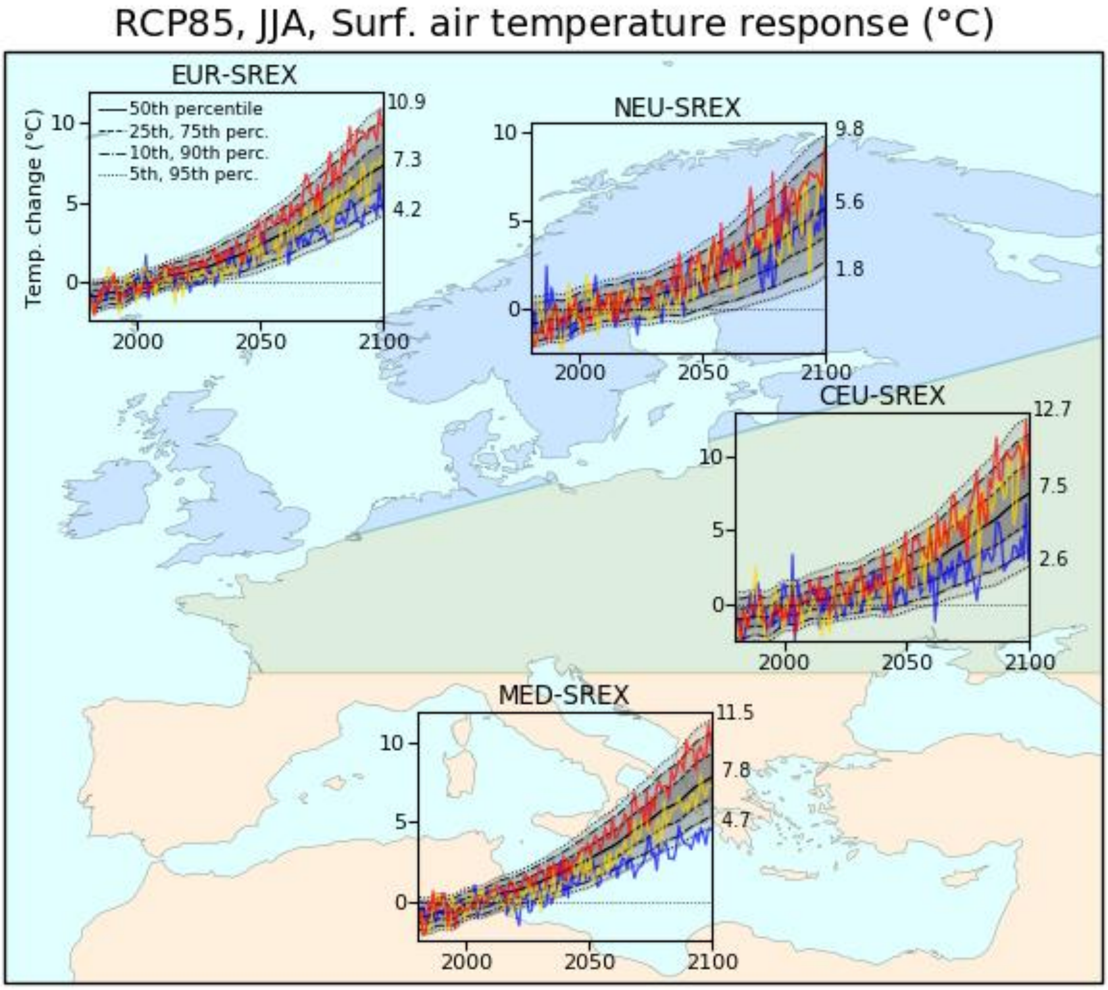


Figure 1 *Projections for individual seasons in response to historical emissions followed by the RCP85 emission-driven scenario for summer temperature change for the SREX regions of Northern Europe (NEU-SREX, blue region), Central Europe (CEU-SREX, green region), the Mediterranean (MED-SREX, orange region) ,and the combined all-Europe region (EUR-SREX). Grey shading and lines show percentiles of anomalies in the variables relative to 1995–2014, calculated from 1-year mean PDFs constructed from 3000 sample realizations. Coloured lines show three individual realizations of year-to-year variation sampled from the 1-year pdfs so that simulated temporal correlations are captured. The three colours represent typical warm (red), medium (gold) and cool (blue) cases, and correspond to the same realization in each of the different plots.*

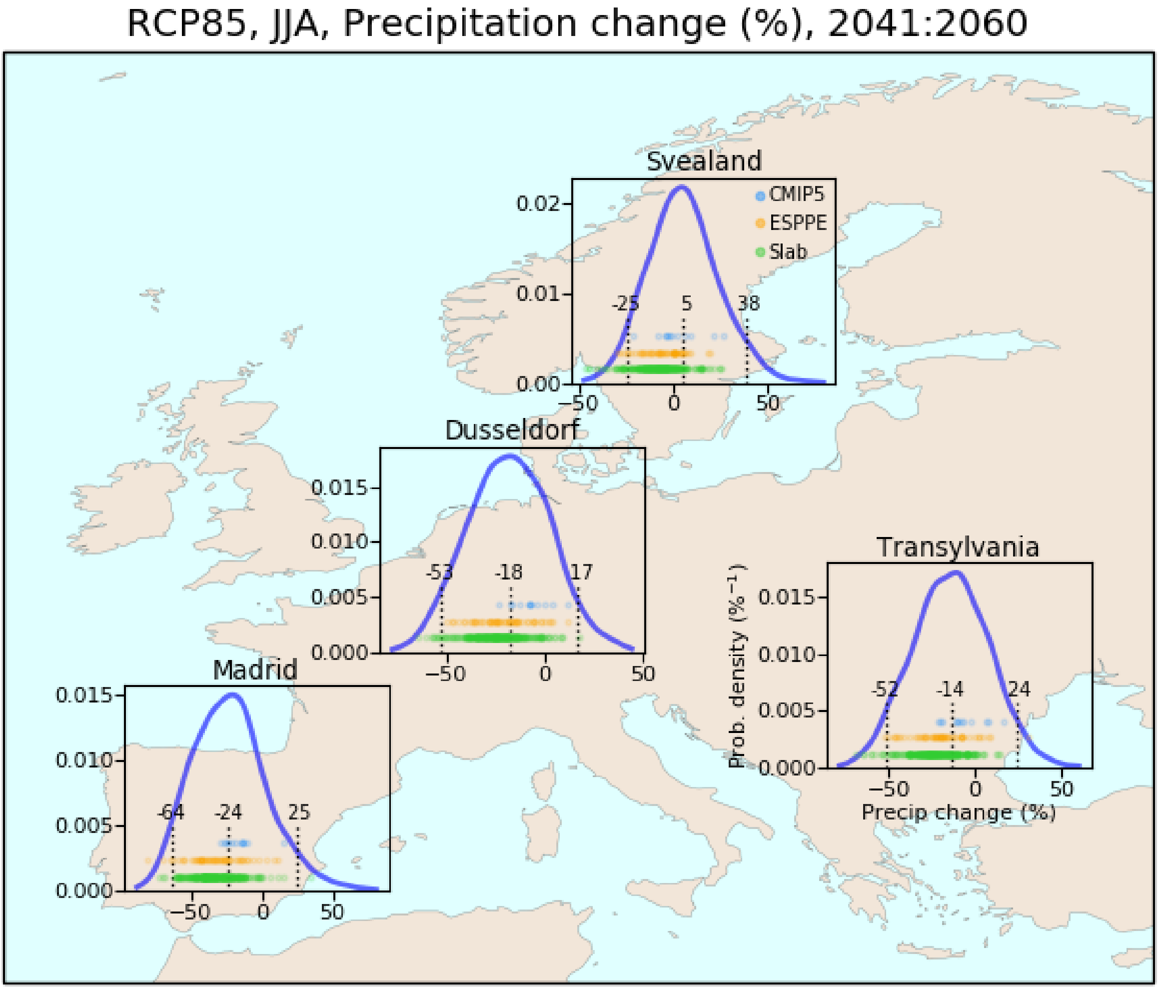


Figure 2 *Probability density functions (blue curves) for the 20-year mean percentage change in precipitation relative to 1995-2014 for the RCP85 emission-driven scenario for the four small WP2 regions. Here we are limited to the by PPE resolution to the closest 2.5°×3.75° grid boxes, rather than the 2.5°×2.5° definitions. Positions and values of the 5th, 50th and 95th percentiles are indicated by the labeled dotted vertical lines. Coloured points correspond to data from the SLAB, ESPPE and CMIP5-ESM ensembles that are input to the methodology (see text).*

References

Booth BBB, Jones CD, Collins M, Totterdell IJ, Cox PM, Sitch S, Huntingford C, Betts RA, Harris GR, Lloyd J (2012) High sensitivity of future global warming to land carbon cycle processes, *Environ. Res. Lett.*, 7, 24002, doi:10.1088/1748-9326/7/2/024002
Booth BBB, Harris GR, Murphy JM, House JI, Jones CD, Sexton DMH, Sitch S (2017), Narrowing the range of future climate projections using historical observations of atmospheric CO_2. *J. Clim.* 30, 3039-3053, https://doi.org/10.1175/jcli-d-16-0178.1
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.
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. doi:10.1007/s00382-006-0121-0.
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, doi 10.1007/s00382-010-0808-0.
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, https://doi.org/10.5194/nhess-10-2009-2010
Harris GR, Sexton DMH, Booth BBB, Collins M, Murphy JM (2013), Probabilistic Projections of Transient Climate Change, *Clim Dyn* 40: 2937. https://doi.org/10.1007/s00382-012-1647-y
Johns TC, Gregory JM, Ingram WJ, Johnson CE, Jones A, Lowe JA, Mitchell JFB, Roberts DL, Sexton DMH, Stevenson DS, Tett SFB, Woodage MJ (2003) Anthropogenic climate change for 1860 to 2100 simulated with the HadCM3 model under updated emissions scenarios. *Clim Dyn* 20:583–612
Jones E, Oliphant E, Peterson P, *et al*. SciPy: Open Source Scientific Tools for Python, 2001-, http://www.scipy.org/ [Online; accessed 2019-05-13]. For Gaussian kernel density estimation: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.gaussian_kde.html
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.
Lambert FH, Harris GR, Collins M, Murphy JM, Sexton DMH, Booth BBB (2013) Interactions between perturbations to different Earth system components simulated by a fully-coupled climate model. *Clim Dyn* 41: 3055. https://doi.org/10.1007/s00382-012-1618-3
van der Linden P., and J.F.B. Mitchell (eds.) 2009: ENSEMBLES: Climate Change and its Impacts: Summary of researchand results from the ENSEMBLES project. Met Office Hadley Centre, FitzRoy Road, Exeter EX1 3PB, UK. 160pp. http://ensembles-eu.metoffice.com/docs/Ensembles_final_report_Nov09.pdf

Meehl GA, Covey C, Delworth T, Latif M, McAvaney B, Mitchell JFB, Stouffer RJ, Taylor KE (2007) The WCRP CMIP3 multi-model dataset: a new era in climate change research. *Bull Am Meteorol Soc* 88:1383–1394
Murphy JM, Booth BBB, Collins M, Harris GR, Sexton DMH, Webb MJ (2007) A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles. *Philos Trans R Soc A* 365:1993–2028
Murphy, J.M., Sexton, D.M.H., Jenkins, G.J., Boorman, P.M., Booth, B.B.B., Brown, C.C., Clark, R.T., Collins, M., Harris, G.R., Kendon, E.J., Betts, R.A., Brown, S.J., Howard T.P., Humphrey, K.A., McCarthy, M.P., McDonald, R.E., Stephens, A., Wallace, C., Warren, R., Wilby, R., Wood, R.A. (2009). UK Climate Projections Science Report: Climate change projections. Met Office Hadley Centre, Exeter, U.K., https://webarchive.nationalarchives.gov.uk/20181204111026/http://ukclimateprojections-ukcp09.metoffice.gov.uk/22530#projections
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. https://doi.org/10.1007/s00382-014-2097-5
Murphy, J.M., Harris, G.R., Sexton, D.M.H., Kendon, E.J., Bett, P.E., Clark, R.T., Eagle, K.E., Fosser, G., Fung, F., Lowe, J.A., McDonald, R.E., McInnes, R.N., McSweeney, C.F., Mitchell, J.F.B., Rostron, J.W., Thornton, H.E., Tucker, S., Yamazaki, K. (2018) UKCP18 Land Projections: Science Report, Met Office Hadley Centre, Exeter, U.K., https://www.metoffice.gov.uk/pub/data/weather/uk/ukcp18/science-reports/UKCP18-Land-report.pdf
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.
Pope VD, Gallani ML, Rowntree PR, Stratton RA (2000) The impact of new physical parametrizations in the Hadley Centre climate model - HadAM3. Clim Dyn 16:123-146.
Rougier J (2007) Probabilistic inference for future climate using an ensemble of climate model evaluations. *Climatic Change* 81:247, https://doi.org/10.1007/s10584-006-9156-9
Santer BD, Wigley TML, Schlesinger ME, Mitchell JFB (1990) Developing climate scenarios from equilibrium GCM results. Max-Planck Institute for Meteorology Report Number 47, Hamburg
Scott DW, Multivariate Density Estimation: Theory, Practice, and Visualization, Second Edition, John Wiley & Sons, 2015, doi:10.1002/9781118575574
Sexton DMH, Harris GR (2015) The importance of including variability in climate change projections used for adaptation. *Nature Climate Change* 5, 931-936, https://doi.org/10.1038/nclimate2705
Sexton DMH, Murphy JM, Collins M, Webb MJ (2012), Multivariate probabilistic projections using imperfect climate models, Part I: Outline of methodology, *Clim Dyn*, 38: 2513, https://doi.org/10.1007/s00382-011-1208-9
Taylor KE, Stouffer RJ, Meehl GA (2012). An overview of CMIP5 and the experiment design. Bull*. Am. Met. Soc.* 93: 485-498.
Trenberth KE (2015). Has there been a hiatus? *Science* 349:691-692.
Watterson IG (2008) Calculation of probability density functions for temperature and precipitation change under global warming. J Geophys Res 113:D12106. doi: 10.1029/2007JD009254