Mediterranean climate change projections from CMIP5 and CMIP6

Author: Josep Cos Espuña* and Supervisors: Francisco J Doblas-Reyes[1], francisco.doblas-reyes@bsc.es; Martin Jury[1], martin.jury@bsc.es; Bernat Codina[2], bcodina@ub.edu
[1]*Earth Sciences Department, Barcelona Supercomputing Center (BSC), Barcelona, Spain. and*
[2]*Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.*

Abstract: The Mediterranean has been identified as a climate change hotspot due to increased warming trends and precipitation decline. To estimate the impacts of the ongoing climate change on the region, projections of various CMIP5 and CMIP6 experiments and scenarios are compared. The changes in temperature and precipitation for the 21st century are studied under scenarios RCP2.6/SSP1-2.6, RCP4.5/SSP2-4.5 and RCP8.5/SSP5-8.5 as well as the high resolution HighResMIP experiments. Additionally, to give robust estimates of projected changes a model weighting scheme is applied, accounting for historical performance and inter-independence of the multi-member multi-model ensembles. Results indicate a significant and robust warming over the Mediterranean during the 21st century over all the ensembles and experiments. Nevertheless, the amplified Mediterranean warming with respect to the global average is only found for summer. Projected changes vary between CMIP5 and CMIP6, with the latter projecting a stronger warming. In contrast to temperature, precipitation changes show a higher level of uncertainty and spatial heterogeneity. However, for the high emission scenario, a robust decline in precipitation is projected for large parts of the region during summer. Results applying the model weighting scheme indicate reductions in CMIP6 and increases in CMIP5 warming trends, making the differences between the two ensembles smaller.

I. INTRODUCTION

The Mediterranean region (10°W, 40°E, 30°N, 45°N) (*Iturbide et al.* 2020) is located between the humid and mild European continental climate and the dry and warm arid North-African climate (*Cramer et al.* 2018). The contrast between these climates is explained by the characteristics of the general atmospheric circulation in midlatitudes, the impact of the surrounding oceans and the interaction with the land surface (*Boé and Terray* 2014).

Global warming is not homogeneous and as previous research suggests (*Lionello and Scarascia* 2018), the Mediterranean region is a hotspot of climate change. The complex and diverse socioeconomic situation of the countries located around the Mediterranean Sea, with strong vulnerabilities to climate change and variability, requires adaptation to the consequences of the already changing climate (*Barros et al.* 2014). The observed increase in temperatures during the recent past is expected to continue and be larger than the planetary mean temperature increase (*Lionello and Scarascia* 2018). Additionally, declines in precipitation totals have been detected during the late 20th century and beginning of the 21st (*Longobardi and Villani* 2010), although there are studies arguing that longer periods should be used to assess historic precipitation trends to avoid influences from interdecadal and interannual variability (*Peña-Angulo et al.* 2020).

In the present study, though, the physical dynamics driving future changes in the Mediterranean region will not be assessed. There are many studies linking characteristics of the Mediterranean climate change to dynamical sources such as the changes in large-scale, upper-tropospheric flow and the reduction of the regional land–sea temperature gradient in winter (*Tuel and Eltahir* 2020) or to thermodynamic processes such as the lapse rate change and land-ocean warming contrast in summer (*Brogli et al.* 2019).

Future climate change is estimated using numerical models. Global climate models (GCMs) are used to project the future state of the climate system by accounting for the physical processes in each climate subsystem (atmosphere, hydrosphere, biosphere, cryosphere, and land-surface) and the interactions between them. Models need to account for the variation in greenhouse gas (GHG) concentrations altered by anthropogenic emissions that produce changes in the atmospheric radiative forcing, but also whenever possible changes in natural forcings (mainly solar irradiance and volcanic aerosols). Different institutes develop different GCMs using slightly different assumptions but always the same physical principles. This offers the opportunity of performing climate change experiments leading to more robust estimates, because different models allow sampling model uncertainty (*Tebaldi and Knutti* 2007). To make models comparable, intercomparison projects have been organised by the international community where a number of GCMs perform experiments under common conditions (e.g using standardized scenarios of GHG emissions and concentrations) (*Meinshausen et al.* 2011, *Riahi et al.* 2017). One of the primary community efforts in this regard are the Coupled Model Intercomparison Projects (CMIPs), where both experiment descriptions and scenarios are prescribed. In the present study the latest two CMIPs are considered and their similarities and differences explored (*Eyring et al.* 2016, *Taylor et al.* 2012).

The simulations from these CMIP experiments, which

* Electronic address: jcosespu7@alumnes.ub.edu

contain results from a large number of models each, show a large spread in projected changes due to the differences in model design and internal variability. Weighting single model runs using information from experiments simulating the observed past (historical period), aims at constraining the climate projection uncertainty and obtaining a more accurate estimate of regional climate change signals. Past approaches have used emergent constraints (*Tokarska et al.* 2020), performance-based model subsets (*Herger et al.* 2019) or weighted models accounting for performance and independence (*Knutti et al.* 2017, *Lorenz et al.* 2018). The latter approach is explored in this study.

The main aim of this study is to evaluate and quantify climate change over the Mediterranean region throughout the 21^{st} century using temperature and precipitation data available from the most recent CMIPs, CMIP5 and CMIP6. Furthermore, it quantifies disagreements between projections of the two CMIPs, assesses the influence of model resolution and tries to improve projections using a weighting method based on the assumptions of model performance and model independence.

Section II describes the climate models and observational data used, explains the methods used to quantify climate change, to weight the projections and to assess model performance. The Mediterranean hotspot, the projected changes (weighted and unweighted) and uncertainties are presented in section III, while these results are discussed in section IV. Section V concludes and presents questions that require further investigation.

II. DATA & METHODS

A. Model data

This study is based on the historical and climate projection experiments of CMIP5 and CMIP6. The historical experiments span from 1850 to 2005 in CMIP5 (*Taylor et al.* 2012) and from 1850 to 2014 in CMIP6 (*Eyring et al.* 2016). The future projections continue from the end of the historical run until 2100. The variables used are monthly mean near-surface air temperature (TAS), precipitation rate (PR) and sea-level pressure (PSL). The latter is used to weight single models (see section II).

Increasing computational power allows for continuing increases in model resolutions, thereby enabling models to explicitly resolve processes at finer scales and improving the data basis for adaptation measures. The High Resolution Model Intercomparison Project (HighResMIP), a CMIP6 endorsed MIP (*Haarsma et al.* 2016), aims at comparing lower and higher resolution versions of the same model. The HighResMIP historical period spans from 1950 to 2014 and the future period from 2015 to 2050. This smaller ensemble is considered to assess how a higher model resolution affects projections over the orographically and geographically complex Mediterranean region.

Three radiative forcing scenarios are used accounting for uncertainty in future emissions: Representative Concentration Pathways (RCPs) 2.6, 4.5 and 8.5 for CMIP5 and Shared Socioeconomic Pathways (SSPs) 1-2.6, 2-4.5 and 5-8.5 for CMIP6. The 2.6, 4.5 and 8.5 values (in Wm^{-2}) represent anthropogenic global radiative forcing increases by 2100 compared to the pre-industrial era for both RCP and SSP (*Meinshausen et al.* 2011). SSP1, SSP2 and SSP5 associate each radiative forcing to a shared socioeconomic pathway, SSP1 being based on sustainability, SSP5 based on a fossil-fueled development and SSP2 representing a middle of the road scenario (*O'Neill et al.* 2016). It should be borne in mind that, while the three studied RCPs and SSPs share the same forcing at the end of the century, GHG and aerosol concentrations during the century differ and this makes them not directly comparable (*Wyser et al.* 2020). Results from CMIP5 and CMIP6 sharing the same 2100 radiative forcing will be displayed together for simplicity. HighResMIP only uses SSP5-8.5 for future projections.

A summary of the initial-condition runs of the models used for every scenario is found in Appendix A. In order to clarify the nomenclature, in the present study "models" is used to refer to the model as a whole and "members" is used for the model's initial-condition members. Some of the models have more than one member, meaning that the multi-model ensemble is a super-ensemble.

B. Observational data

Observational data are used to compare model experiments to the observed past and to derive performance weights of ensemble members. To account for observational uncertainty multiple observational products are used containing both reanalysis (ERA5 and JRA55) and observations (GPCC, CRU and E-OBS). A more detailed summary of the datasets used can be found in Appendix A.

Some listed datasets have constraints and are not used consistently in this study. JRA55 is not shown in the time series as it was shown to overestimate Mediterranean precipitation during the period 1958-1978 (*Tsujino et al.* 2018). E-OBS is not used when plotting time series or box plots as it has many missing values due to the low density of measurement stations in the southern part of the Mediterranean region.

C. Methods

The analyses have been performed using the Earth System Model Evaluation Tool (ESMValTool). ESMValTool is a tool developed by the community facilitating the processing of generic climate datasets, allowing for reproducibility of results (*Righi et al.* 2020). Tools from the ESMValCore (`https://github.com/ESMValGroup/`

`ESMValCore`) Preprocessor modules have been used to compute the diagnostics.

As TAS and PR are only assessed over land , a land-sea mask is applied at the beginning of the data preprocessing. The time dimension is split into different historical and future periods to compute the diagnostics. The baseline periods (1980-2014 and 1986-2005) are used as a reference to assess the performance of the models against observations. The longer reference period (35 years) is used to show historical trends, as 20-year trends are considered to be too heavily influenced by internal variability (*Merrifield et al.* 2020, *Peña-Angulo et al.* 2020). The shorter 1986-2005 period (from *Collins et al.* (2013)) is used to compute deltas i.e. the change between the historical period and a future period. This historical period is used instead of 1995-2014 (*Brunner et al.* 2019) to avoid CMIP5's historical (which ends in 2005) and future projection runs to overlap. Three future periods are considered to assess changes: near-term (2021-2040), mid-term (2041-2060) and long-term (2081-2100). Only the near-term period is available for HighResMIP as projection experiments have been performed only until 2050.

To be able to compare models and observational references all datasets are regridded to a 1°x1° grid over the region (10°W, 40°E, 25°N, 50°N) using a horizontal conservative interpolation method. Additionally, when computing climatologies the surface temperature is corrected for height differences between the model's orography and the evaluation grid (using the standard lapse rate, 6.49 K/m) (*Dennis* 2014).

Climate change varies for different seasons. Therefore, seasonal means for four seasons (DJF, MAM, JJA, SON) are computed by averaging the monthly data accordingly.

The most recurrent diagnostics used are anomalies and trends. Anomalies are computed subtracting the reference period 20-year climatology. Trends are computed following a linear ordinary least square regression fit with time as the independent variable. Projected changes, also referred to as deltas, are computed by calculating the difference between time-averaged values for a future period and the time-averaged values for the reference period.

To illustrate how differently models perform between them, and depending on the diagnostic, the historical period root mean squared error (RMSE) between each model of the super-ensemble and observational references is computed. The RMSE is normalized with the multi-model ensemble RMSE median of the diagnostic (*Gleckler et al.* 2008). This way a metric is obtained of how well a model performs (according to a reference dataset) compared to the multi-model average performance. A better than average performance is indicated by negative values, while a worse than average performance by positive ones.

To assess the uncertainty and robustness of changes in the multi-model ensemble two dimensions are considered. The degree of agreement between the models and the statistical significance of the variable changes are used. The agreement is measured using a threshold of 80% of the models showing the same sign as the multi-model ensemble mean (*Collins et al.* 2013). A mean change is considered significant if it is beyond the threshold of a two-tailed t-Student test at the 95% confidence level. The t-statistic is computed with the historical and future period ensemble means and their inter-model distribution.

An approach to improve future projections ensembles is to give more weight to models with good performance. Assuming that an observational ensemble is a good representation of the climate, historical model runs are compared against it and more weight is given to those members of the multi-model ensemble that perform better i.e. weighting by performance.

Another aspect to take into account when weighting a multi-model is the independence between models. Some models share similarities in their formulation and it would not be fair to give equal weight to dependent and independent models as some similar model structures would be over-represented. To account for this an independence weighting method is also used.

From the work by *Knutti et al.* (2017), *Lorenz et al.* (2018), *Brunner et al.* (2019) and *Merrifield et al.* (2020) the equation (1) is used to obtain the weight w_i that each member i will have in the projections ensemble. The weight accounts for the distance D_i from member i to the observational ensemble, obtained with the RMSE, and for the distance S_{ij} between member i and every other model member j from the multi-model ensemble. σ_s and σ_d are constant independence and performance shape parameters, respectively. The observational ensemble is computed with the observational data mean.

$$w_i = \frac{e^{-\left(\frac{D_i}{\sigma_d}\right)^2}}{1 + \sum_{j \neq i}^{M} e^{-(\frac{S_{ij}}{\sigma_s})^2}} \tag{1}$$

The weighting method needs to account for different diagnostics (trends, anomalies, variabilities...) in order to better represent the model characteristics and to avoid weighting models that could match, just by chance, the independence and performance criteria of a single diagnostic. In this study, the diagnostics used to evaluate the distances D_i and S_{ij} are not the same, as suggested by *Merrifield et al.* (2020). When evaluating performance the aim is to give weight to models that resemble in a more faithful way the observations, while independence weighting aims to clearly identify members that behave similarly. In the present study TAS and PSL have been used as diagnostic variables (*Merrifield et al.* 2020). The diagnostics used to weight performance, both for TAS and PR future projections, are TAS-ANOM, TAS-STD, TAS-TREND, PSL-ANOM, PSL-STD; the diagnostics to weigh independence are the 20-year climatologies TAS-CLIM, PSL-CLIM, also for temperature and precipitation projections. The reference period over which the diagnostics are computed is 1980-2014 (*Brunner et al.* 2020).

In general terms, the shape parameters are the refer-

ence values that inform if a distance is enough to downweight a model (σ_d) and if two models are dependent or independent (σ_s). Each scenario and season and multimodel ensemble has associated its own shape parameters.

Appendix B explains in further detail the diagnostics used and the physical meaning of the shape parameters and the methods used to compute them.

III. RESULTS

With the available amount of data and the diagnostics described a large number of analyses have been performed. Only a few of them are shown in this thesis. However, the rest have been made available in a shiny app at `https://earth.bsc.es/shiny/medprojections-shiny_app/` for the sake of completeness. This section refers to some figures not shown in the text but available in the shiny app for consultation if desired.

A. The Mediterranean as a climate change hotspot

A first approach to the future Mediterranean climate is to show how its projections compare to global mean projected changes.

Figure 1 illustrates the Mediterranean hotspot for the highest emissions scenario. It shows TAS and PR differences between the global mean and the Mediterranean region anomalies during the mid and long-term with respect to 1986-2005. The Mediterranean region shows a higher annual temperature increase than the global mean. Nevertheless, when accounting for seasonal differences, the highest amplification is visible from summer. The amplification during winter is small and even negative in the north-west part of the domain. While projections agree on a global precipitation increase for the long-term period (*Lionello and Scarascia* 2018), the Mediterranean region shows smaller precipitation increases.

The TAS and PR differences show an increase in magnitude from the mid-term to the long-term period, indicating that the Mediterranean climate changes at a faster rate than the global mean for the highest emission scenario. The 2.6 Wm^{-2} scenario shows, instead, a dampening of the hotspot from the mid-term to the long-term.

The combination of projected increased warming and an anomalous precipitation decline makes the region a hotspot for climate change (*Lionello and Scarascia* 2018).

B. Model performance over the historical period

The ensembles used in the present study contain many models and they all perform differently for different diagnostics. Figure 2 aims to justify the use of multi-model ensembles rather than single models. The normalised RMSE from diverse diagnostics (rows) is shown for all the models in CMIP6 (columns). The boxes are diagonally split to show the errors with respect to two observational datasets: JRA55 and CRU. Negative values indicate better than median performance, positive values worse than median performance.

The models' performance against the diverse diagnostics is very different as shown in Figure 2. Some models perform better over most of the diagnostics (e.g ACCESS-ESM1-5, KACE1-0-G or UKESM-0-LL). Some models seem to perform well for some diagnostics or variables and bad for others. Two examples are MIROC6, which shows low performance mainly for surface temperature climatologies, and IPSL-CM6A-LR which has a low performance on precipitation climatologies. Others, like CNRM-CM6-1-HR, perform worse than the RMSE median for most diagnostics.

Some models seem to share characteristics. The models from the institutions MIROC, INM and CNRM-CERFACS seem to have similar responses to certain diagnostics. The potential dependency between models raises another concern when dealing with multi-model ensemble projections. Dependent models bear the risk of producing a biased mean projection as more weight could be given to a certain way of modelling the climate.

Interestingly, the multi-model-mean and multi-model-median performances are better than the single model performance median. Showing that while there are models that perform better than the ensemble mean, it is still better to use the mean than single models when considering multiple diagnostics. The fact that the multi-model ensemble performance is consistently better for all diagnostics shows how single model errors partly cancel off (*Tebaldi and Knutti* 2007). This is one of the pillars behind the Coupled Model Intercomparison Project: to get a better representation of the climate by combining different models.

C. Unweighted future projections

The historical temperature and precipitation trends are compared with the observational ensemble to verify the CMIPs and HighResMIP performance. As seen in Figure 3 (a) and (c), the spread of the multi-model ensembles trend spans the observational ensemble trends. Mostly, for the rest of variables and seasons, the observations also fall inside the multi-model ensemble historical runs (see shiny app). The multi-model ensemble spread shown in the historical temperature trends is notably larger than the observational ensemble. For precipitation, spreads are larger for all ensembles and usually show negative and positive trends (e.g CMIP5 5th to 95th precipitation trend percentiles go from -0.06 to 0.06 $mm\,day^{-1}\,decade^{-1}$). Qualitatively, for HighResMIP temperature trends, high resolution experiments tend to be less dispersed than the lower resolution ones, although the ensemble size is small. Precipitation trends do not show such a behaviour.

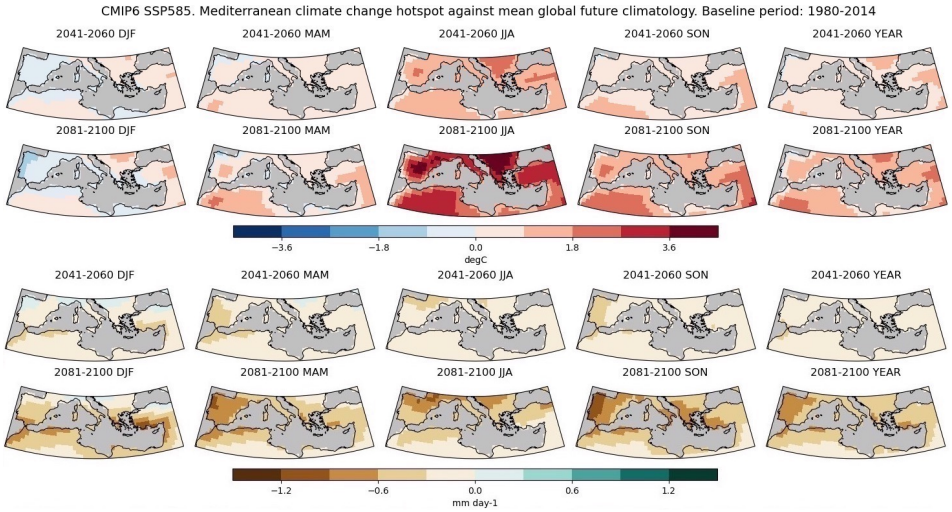

Figure 1: Mediterranean temperature (upper two rows) and precipitation (lower two rows) change differences with respect to the mean global change. The changes are computed with respect to the 1986-2005 period for the mid-term (1^{st} and 3^{rd} row) and late-term periods (2^{nd} and 4^{th} row). The differences are shown for each season and annual means (columns) for the CMIP6 SSP5-8.5 scenario.

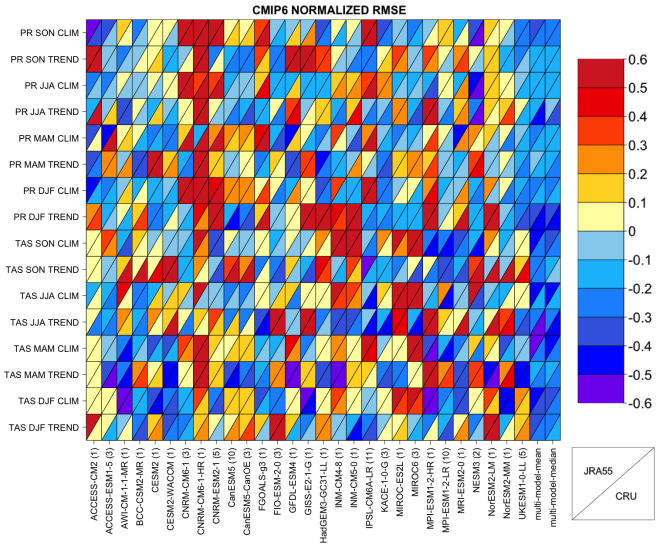

Figure 2: Normalized root mean squared error of diagnostics (rows) between each CMIP6 model (columns) and observational references. Each cell is divided in two triangles to display the RMSE against JRA55 and CRU. The last two columns show the RMSE of multi-model ensemble mean and median. The diagnostics are TAS and PR climatologies and trends over the Mediterranean region for the four different seasons.

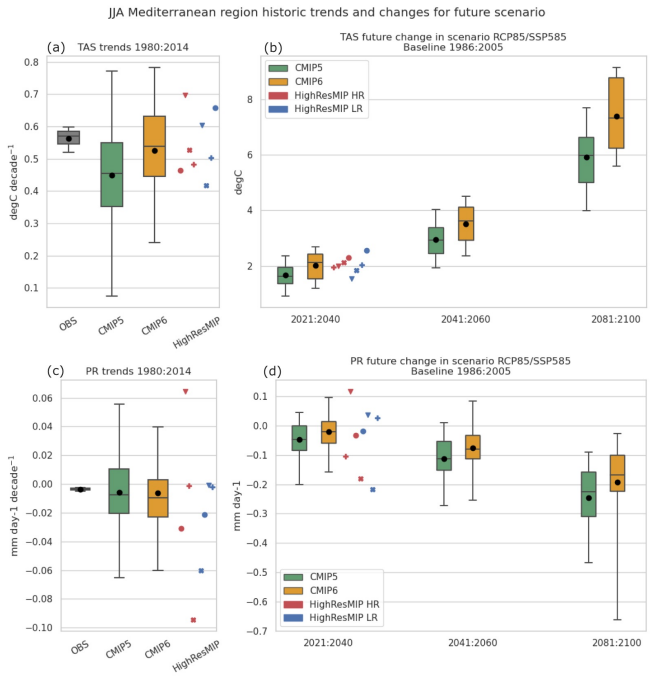

Figure 3: Summer temperature (a) and precipitation (c) historical trends of the observational ensemble, CMIP5, CMIP6 and HighResMIP. Summer temperature (b) and precipitation (d) projected changes for the near, mid and long-term periods with respect to the baseline period. In the box plots, the black horizontal line represents the median and the black dot is the mean. The interquartilic range (IQR) and whiskers are defined by the 25th-75th and 5th-95th percentiles, respectively. HighResMIP models are displayed as markers, enabling for a comparison of the HR (red) and LR (blue) models within the experiment. The same markers are used for two different resolution runs of the same model (see Appendix A)

Figure 3 (b) and (d), show the multi-model ensemble summer TAS and PR projected changes displayed for the three future periods. Regarding temperature, the CMIP6 ensemble shows larger changes than CMIP5 throughout the century. Uncertainties for the end of the century seem to grow larger for CMIP6 as its IQR is larger than CMIP5. The projected change at the end of the century by CMIP6 is over 7ºC averaged for the whole Mediterranean region with 90% ranging from 5.5 to 9.5ºC. CMIP5 shows a mean increase of 6ºC by the end of the century and 90% ranging from over 4ºC to

7.5°C. For the rest of the seasons and scenarios, CMIP6 shows bigger temperature changes and broader IQR and whisker ranges than CMIP5 (see shiny app). Summer remains the season with the largest temperature increase. As expected, the projection spread broadens for each future period as models diverge. The winter 90% range temperature change goes from 3.2 to 6.8°C for CMIP6 and from 2.7 to 5°C in CMIP5 (see shiny app). HighResMIP HR projections show a smaller spread, converging to values close to the CMIP6 ensemble mean (for all seasons). Conclusions must be drawn carefully, though, as the HighResMIP ensemble is very small.

Regarding precipitation projections, CMIP5 draws a more pessimistic picture in summer with decreases of 0.25 $mm\,day^{-1}$, which means 23 mm less during the summer season at the end of the century, with a 90% range from -0.45 to -0.1 $mm\,day^{-1}$. The rest of the seasons and the middle of the road scenario project, generally, mean precipitation loss for the mid and long-term (see the shiny app). The sustainability scenario shows no robust change. The large uncertainties inherent to precipitation projections make the assessment of the results harder as 25% of the multi-model distribution is showing precipitation increase. The winter 90% range precipitation changes go from -0.35 to 0.03 $mm\,day^{-1}$ for CMIP6 and from -0.45 to 0.07 $mm\,day^{-1}$ in CMIP5 (see shiny app). There are no obvious patterns that distinguish the HighResMIP HR and LR precipitation experiments throughout seasons.

The spatial distribution of the changes projected by the high emission scenario for JJA TAS and PR is shown in Figures 4-5.

Summer TAS changes are robust and significant over all the 21st century periods in the Mediterranean region (see Figure 4). HighResMIP shows statistically non-significant changes in the region's middle latitudes (note that the ensemble only has seven models). As previously said, CMIP6 shows larger warming than CMIP5. Nevertheless, there is good agreement between both CMIP experiments in the spatial distribution of the temperature increase over the Mediterranean region. Eastern Europe, the Balkans and the Iberian peninsula would be the regions with the largest mean summer temperature increases, around 8-9°C.

The middle of the road scenario (4.5 Wm^{-2} forcing) projects, generally, a robust and significant temperature increase for all the century, but with a tendency to reduce the rate of increase. CMIP6 shows the most important warming (mean of 4.5-5°C) in the Iberian and Balkan peninsulas. The sustainable scenario (2.6 Wm^{-2} forcing) warming is generally robust and significant and shows a temperature plateau beginning between 2040 and 2050 for both CMIPs. Nevertheless, CMIP6 is projected to reach an average increase of just over 2°C while CMIP5 averages approximately 1.5°C. The warming spatial patterns and magnitudes for the mid and long-term periods remain very similar. The near-term projected changes are similar for all scenarios.

The temperature changes during winter are smaller than during summer, but still significant. CMIP6 projects larger changes than CMIP5, but the spatial distribution of the magnitudes is similar for both. The north-east Mediterranean region is the one with the largest projected warming in winter. The near-term HighResMIP shows a slightly larger TAS increase than CMIP6 in eastern Europe (see shiny app).

Regarding precipitation in summer, the end of the century changes projected by the high emission scenario indicates a robust and significant precipitation decline for all the Mediterranean region (see Figure 5). Note that neither robust nor significant changes are projected in the southern and eastern region mainly because, according to the CRU dataset climatology, precipitation is already non-existent or very low during summer. Both CMIP experimetns agree that the region most affected by the precipitation decline is the one spanning from the Pyrenees to the north of the Alps, with long-term changes ranging from -0.6 to -0.9 $mm\,day^{-1}$ for the high emission scenario. Despite lower forcing scenarios projecting non-significant changes (except the west Mediterranean for long-term SSP2-4.5) and fewer regions showing inter-model agreement, the change points to a general precipitation decline. The HighResMIP projections for the near-term are in good agreement with the CMIP6 spatial pattern and magnitudes through most of the seasons (note that most of the grid points are not robust).

Winter precipitation changes in a different way than in summer (see shiny app). The southern part of the Mediterranean region sees a robust and significant precipitation change by the end of the century, with maximum declines of -0.4 to -0.6 $mm\,day^{-1}$ over south Turkey and the north African coastline. The northern Mediterranean region lies in a transition zone, as Western Europe is projected to see an increase in precipitation, leading to increasing uncertainties in the Iberic, Italian and Balkan peninsulas projections. CMIP6 ensemble shows more robust and significant changes throughout the region and projects broader 5-95th percentile ranges. This behaviour is also found in the rest of the scenarios (see shiny app).

The amplitude of the variations in the observed winter precipitation time series is wider than the simulated 90% range, suggesting that models fail to capture the magnitude of precipitation variations during winter.

D. Weighting projections

New projection distributions are obtained from applying the weighting for performance and independence method to the CMIP5 and CMIP6 multi-model ensembles. Figure 6 shows TAS boxplots comparisons between the weighted and unweighted ensembles during summer for the three future periods. Both ensembles display a reduction in their IQRs with the weighting. CMIP5 mean and median projections increase in the weighted

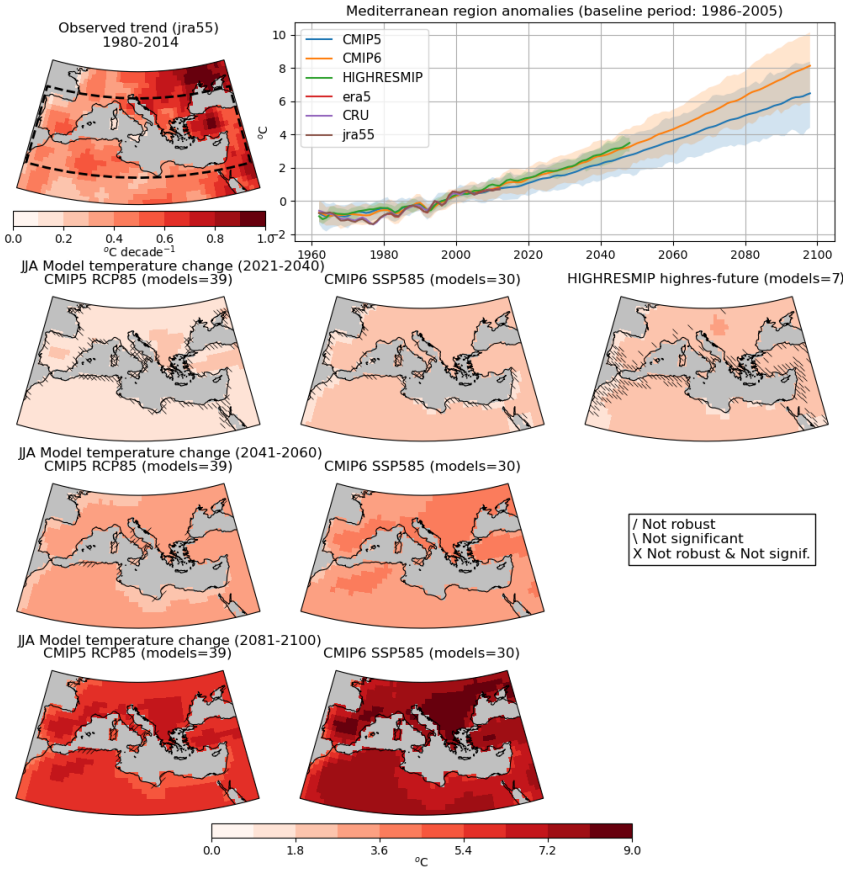

Figure 4: Summer temperature change fields for the three relevant future periods (rows) according to the CMIP5, CMIP6 and HighResMIP ensembles (columns) following the high emission scenario. The time series shows the evolution of the anomaly with respect to the baseline period on the Mediterranean region for the observational datasets and the multi-model ensembles. Solid lines indicate the one-member-per-model ensemble mean and the shaded region indicates the 5th-95th percentiles. JRA55 historical (1980-2014) trend is shown along with the Mediterranean region used (dashed line).

ensemble while CMIP6 decrease, bringing the two experiment means closer together. The upper range (75-95th percentiles) in the CMIP6 weighted ensemble increases while the lower (5-25th percentiles) decreases. Also, the CMIP6 weighted ensemble displays a more skewed distribution towards lower temperature increases. CMIP5 evolves to a much more constrained ensemble, showing an IQR reduction of 1.15°C.

The winter weighted temperature projections show some similar responses to the weighting as summer weighted projections: CMIP5 signal increases while CMIP6 decreases. However, the differences between the weighted and the unweighted means are smaller. CMIP5, weighted and unweighted, IQR and whisker range are pretty similar in the long-term period. This suggests that uncertainties in the temperature changes are well represented in winter by the original CMIP5 ensemble.

It is especially interesting to see how the large IQR projected by the CMIP6 models at the end of the century is reduced by half, once weighted. Nevertheless, even if the probabilities of a high-warming future decrease, such temperature changes are still considered plausible by the weighted ensemble.

IV. DISCUSSION

Projections obtained from climate multi-model ensembles are a game of uncertainties. Different modelling methods and forcings that can only be estimated (e.g GHG emissions, land use...) lead to differences in the model results (*Tebaldi and Knutti* 2007). In the present study, to consider as many uncertainties as possible of the projected climate change in the Mediterranean region, different radiative forcing scenarios and multi-model ensembles have been used. A weighting method, constraining the projections, has been applied to reduce uncertainty in the results.

From temperature projections, it is found that changes over the 21st century are larger when larger radiative forcing scenarios are applied, consistent with basic radiative forcing theory (*Wallace and Hobbs* 2006). Compared to the global warming signal, Mediterranean temperature changes are bigger in summer and similar in winter. This is a general behaviour throughout scenarios and multi-model ensembles, meaning that summer warming amplification is independent of the scenario and ensemble considered. Previous studies also identify the Mediterranean

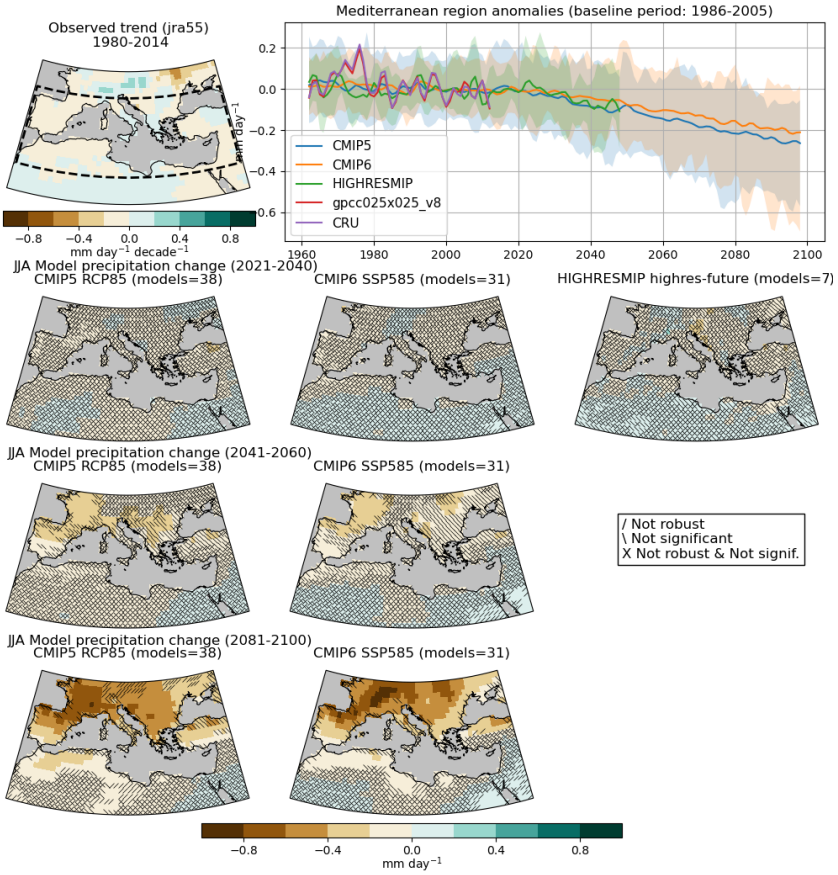

Figure 5: Same as Figure 4 for summer precipitation

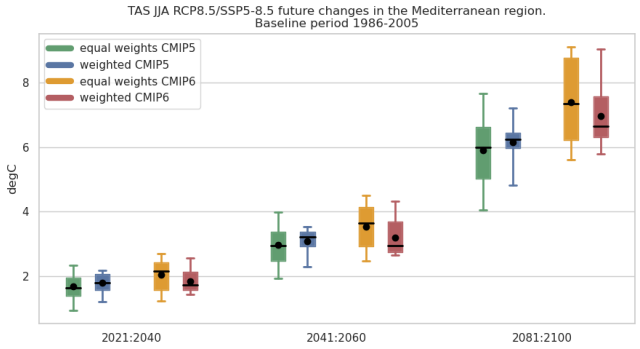

Figure 6: Weighted (blue, red) and non-weighted (green, yellow) summer temperature change with respect to 1986-2005 for CMIP5 and CMIP6 following the high emission scenario.

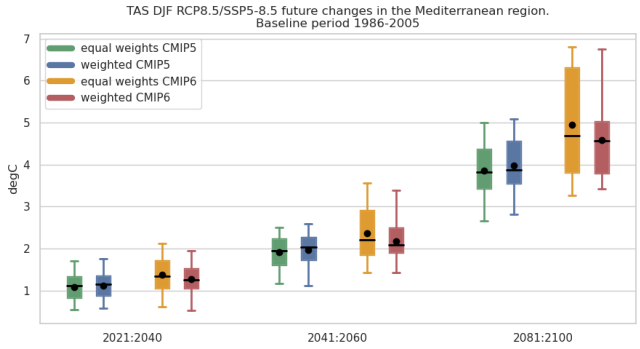

Figure 7: Same as Figure 6 for winter

warming amplification (*Lionello and Scarascia* 2018, *Zittis et al.* 2019), but it must be stressed that it does not apply to the winter season. From the projections, the magnitude of the temperature changes is what remains uncertain and highly dependent on the emission scenario that humanity shall follow.

There is confidence in a precipitation decline for the higher radiative forcing scenario over all the Mediterranean region during summer and only in the southern part during winter. Conclusions must be drawn carefully from precipitation as the spread in the multi-model ensembles is large. For other scenarios and seasons, precipitation changes tend to show declines, although results remain uncertain due to low robustness and significance over most of the region and high spread ranges in the ensemble results.

The signal magnitudes vary over the Mediterranean region, for temperature but more notably for precipitation. This indicates that it would be useful to divide the Mediterranean region into smaller subregions to assess,

more accurately, the area-averaged changes and their implications.

Non-significant grid points in the coastlines are due to the differences in the original grid resolutions between models. When regridding, information is lost in the complex coastline regions as coarse models can not represent them, causing some grid points in the ensemble to be considered land by some models and ocean by others. When the significance test is computed the degrees of freedom at these grid points are reduced as not all models are represented, leading to results with non-significant signals.

As suggested in *Tebaldi and Knutti* (2007), historical model temperature trends are a good indicator of projected changes. The results in Figure 3 show a larger temperature trend in CMIP6 than CMIP5 and projections show larger changes for the former. This remains true for the rest of scenarios and seasons (see shiny app), suggesting that the historical ensemble temperature trend is a good indicator of the magnitude of future warming.

The CMIP6 ensemble is known to have a larger range of climate sensitivity i.e. radiative forcings generate larger signals and at a faster rate (*Hausfather* 2019). The higher sensitivity can be due to the way models are designed or the way the radiative forcing scenario is defined. Even if the scenarios are named after the radiative forcing by the end of the century (in Wm^{-2}), the GHG concentration curves are different (*Meinshausen et al.* 2011, *Riahi et al.* 2017). As suggested by *Wyser et al.* (2020), when the same model is run with the GHG concentrations from equal 2100 radiative forcing SSPs and RCPs (2.6, 4.5 and 8.5 Wm^{-2}), the first causes larger temperature chang es. Another CMIP6 aspect that can affect climate sensitivity is that models have improved the aerosols and clouds formulation (*Hausfather* 2019). While the higher sensitivity to radiative forcing is true for some CMIP6 models, not all of them show this behaviour. Assessing the weighted temperature ensemble, it is found that the CMIP6 distribution shifts to lower changes, meaning that models giving larger TAS signals have been down-weighted, reducing the differences between CMIP6 and CMIP5 experiments. Studies have shown that some CMIP6 models with higher warming signals are poorly representing the historical climate (*Tokarska et al.* 2020).

The projections' weighting also modifies the ensemble's distribution. Generally, the weighting method reduces the IQR and/or the 5-95th percentile range, reducing future changes uncertainty. Nevertheless, uncertainty is also reduced when the ranges widen, as it would indicate that the spread was under-estimated in the original ensemble.

Precipitation weighted projections have not been shown as there is no proof that the diagnostics assessing performance are relevant to evaluate the precipitation response of models. Also, from the weighting methodology, it could be argued that trends have been computed over a too short historical period of 35 years (*Merrifield et al.* 2020, *Peña-Angulo et al.* 2020).

V. CONCLUSIONS

This study has aimed to interpret the projected temperature and precipitation changes by CMIP5, CMIP6 and HighResMIP GCM ensembles in the Mediterranean region. To tackle the uncertainty issues of projection ensembles, different scenarios and seasons have been assessed and a weighting method that accounts for historical performance and inter-independence of the ensemble models has been applied.

The Mediterranean is expected to be a climate hotspot due to the projected smaller or negative precipitation trends compared to the global signal. However, the amplified warming of the Mediterranean is not found for winter, it especially affects temperature during the summer season.

Conclusions must be drawn carefully from projection ensembles as they are composed of models performing very differently and models dependent on each other. Nevertheless, the ensemble mean has been shown to give better performance across different diagnostics than single models. There is high confidence that significant and robust warming is going to affect the Mediterranean region during the 21st century due to anthropogenic radiative forcing. It is expected to affect more notably the Iberian and Balkan peninsulas during summer and the Balkans during winter. In contrast to temperature, precipitation changes show a higher level of uncertainty and spatial heterogeneity, although there is high confidence that a robust and significant decline in precipitation will affect the whole region during summer and the southern part during winter at the end of the century if the high emission scenario is followed. The combination of warming and precipitation decline could put the region under strain, especially the south, which is less prepared to adapt to the already changing climate.

A weighting method has been applied to reduce uncertainties induced by models poorly representing aspects of the historical climate or by dependent models being overrepresented in the ensemble. It is concluded that CMIP6 over-estimates warming in the Mediterranean and its distribution uncertainty. CMIP5 slightly under-estimates warming and generally overestimates inter-model spread. This is relevant as it homogenizes the output from the last two CMIP phases, reducing future uncertainties of climate change.

Further work needs to be done on the weighting method to identify the relevant diagnostics that best assess historical model precipitation performance. Inhomogeneities in the fields have been seen in the Mediterranean region, therefore it would be a good consideration to divide it into sub-regions to extract more relevant information from the area-averaged distributions. In the same line, using RCMs (like MED-CORDEX) and longer high-resolution global model runs should be considered as the scales in which GCMs work might be too large for the complex orography and geometry of the Mediterranean region.

ACKNOWLEDGMENTS

I'd like to thank Francisco J Doblas-Reyes, Martin Jury and Raül Marcos for the support and the discussions that made me go forward. I acknowledge diverse people from BSC Earth Sciences Department that helped me with data handling and data processing issues. Finally I'd like to thank Bernat Codina for his valuable suggestions.

REFERENCES

Barros, V., et al., IPCC, 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, *Tech. rep.*, 2014.

Boé, J., et al., Land-sea contrast, soil-atmosphere and cloud-temperature interactions: Interplays and roles in future summer European climate change, *Climate Dynamics*, *42*, 683–699, 2014.

Brogli, R., et al., The role of hadley circulation and lapse-rate changes for the future European summer climate, *Journal of Climate*, *32*, 385–404, 2019.

Brunner, L., et al., Quantifying uncertainty in European climate projections using combined performance-independence weighting, *Environmental Research Letters*, *14*, 124–010, 2019.

Brunner, L., et al., Reduced global warming from CMIP6 projections when weighting models by performance and independence, *Earth System Dynamics Discussions*, *11*, 995–1012, 2020.

Collins, M., et al., Long-term climate change: Projections, commitments and irreversibility, Climate change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, *Tech. rep.*, 2013.

Cornes, R. C., et al., An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets, *Journal of Geophysical Research: Atmospheres*, *123*, 9391–9409, 2018.

Cramer, W., et al., Climate change and interconnected risks to sustainable development in the Mediterranean, *Nature Climate Change*, *8*, 972–980, 2018.

Dennis, S., The Climate Data Guide: Regridding Overview, 2014. url: `https://climatedataguide.ucar.edu/climate-data-tools-and-analysis/regridding-overview`.

Eyring, V., et al., Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, *Geoscientific Model Development*, *9*, 1937–1958, 2016.

Gleckler, P. J., et al., Performance metrics for climate models, *Journal of Geophysical Research Atmospheres*, *113*, 1–20, 2008.

Haarsma, R. J., et al., High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6, *Geoscientific Model Development*, *9*, 4185–4208, 2016.

Harris, I., et al., Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset, *Scientific Data*, *7*, 1–18, 2020.

Hausfather, Z., CMIP6: the next generation of climate models explained, 2019. url: `https://www.carbonbrief.org/cmip6-the-next-generation-of-climate-models-explained`.

Hawkins, E., et al., The potential to narrow uncertainty in projections of regional precipitation change, *Climate Dynamics*, *37*, 407–418, 2011.

Herger, N., et al., Ensemble optimisation, multiple constraints and overconfidence: a case study with future Australian precipitation change, *Climate Dynamics*, *53*, 1581–1596, 2019.

Hersbach, H., et al., The ERA5 global reanalysis, *Quarterly Journal of the Royal Meteorological Society*, *146*, 1999–2049, 2020.

Iturbide, M., et al., An update of IPCC climate reference regions for subcontinental analysis of climate model data: Definition and aggregated datasets, *Earth System Science Data Discussions*, *12*, 1–16, 2020.

Knutti, R., et al., A climate model projection weighting scheme accounting for performance and interdependence, *Geophysical Research Letters*, *44*, 1909–1918, 2017.

Kobayashi, S., et al., The JRA-55 reanalysis: General specifications and basic characteristics, *Journal of the Meteorological Society of Japan*, *93*, 5–48, 2015.

Lionello, P., et al., The relation between climate change in the Mediterranean region and global warming, *Regional Environmental Change*, *18*, 1481–1493, 2018.

Longobardi, A., et al., Trend analysis of annual and seasonal rainfall time series in the Mediterranean area, *International Journal of Climatology*, *30*, 1538–1546, 2010.

Lorenz, R., et al., Prospects and Caveats of Weighting Climate Models for Summer Maximum Temperature Projections Over North America, *Journal of Geophysical Research: Atmospheres*, *123*, 4509–4526, 2018.

Meinshausen, M., et al., The RCP greenhouse gas concentrations and their extensions from 1765 to 2300, *Climatic Change*, *109*, 213–241, 2011.

Merrifield, A. L., et al., An investigation of weighting schemes suitable for incorporating large ensembles into multi-model ensembles, *Earth System Dynamics*, *11*, 807–834, 2020.

O'Neill, B. C., et al., The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6, *Geoscientific Model Development*, *9*, 3461–3482, 2016.

Peña-Angulo, D., et al., Long-term precipitation in Southwestern Europe reveals no clear trend attributable to anthropogenic forcing, *Environmental Research Letters*, *15*, 094–070, 2020.

Riahi, K., et al., The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: An overview, *Global Environmental Change*, *42*, 153–168, 2017.

Righi, M., et al., Earth System Model Evaluation Tool (ESMValTool) v2.0 - technical overview, *Geoscientific Model Development*, *13*, 1179–1199, 2020.

Schamm, K., et al., Global gridded precipitation over land: A description of the new GPCC First Guess Daily product, *Earth System Science Data*, *6*, 49–60, 2014.

Taylor, K. E., et al., An overview of CMIP5 and the experiment design, *Bulletin of the American Meteorological Society*, *93*, 485–498, 2012.

Tebaldi, C., et al., The use of the multi-model ensemble in probabilistic climate projections, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *365*, 2053–2075, 2007.

Tokarska, K. B., et al., Past warming trend constrains future warming in CMIP6 models, *Science Advances*, *6*, 1–14, 2020.

Tsujino, H., et al., JRA-55 based surface dataset for driving ocean–sea-ice models (JRA55-do), *Ocean Modelling*, *130*, 79–139, 2018.

Tuel, A., et al., Why Is the Mediterranean a Climate Change Hot Spot?, *Journal of Climate*, *33*, 5829–5843, 2020.

Wallace, J. M., et al., *Atmospheric science: An introductory survey*, Amsterdam: Elsevier Academic Press, 2006.

Wyser, K., et al., Warmer climate projections in EC-Earth3-Veg: The role of changes in the greenhouse gas concentrations from CMIP5 to CMIP6, *Environmental Research Letters*, *15*, 054–020, 2020.

Zittis, G., et al., A multi-model, multi-scenario, and multi-domain analysis of regional climate projections for the Mediterranean, *Regional Environmental Change*, *19*, 2621–2635, 2019.

Appendix A: Models and observational data summary

The initial-condition runs from the models used in the study, for the three radiative scenarios are summarized in Table I. In order to identify the area-averaged differences in the Mediterranean region between the HR and LR HighResMIP runs in Figure 3, each model has a distinctive marker as seen in Table II. The observational datasets used are summarized in Table III.

Appendix B: Diagnostics, σ_d and σ_s of the weighting method

The weighting method described in equation (1) uses a specific set of diagnostics and two constant shape parameters. The current section aims to describe in more detail the methodology behind the weighting of the projection ensembles.

The chosen diagnostics and the variables used for the diagnostics must be relevant to the weighted projections ensemble. The aim is to obtain weighted temperature and precipitation future projections from a multi-model ensemble. There is evidence of the relevance of historical TAS and PSL to projected temperature (*Brunner et al.* 2020, *Merrifield et al.* 2020). For precipitation projections, the diagnostics should assess temperature and dynamics of the atmosphere, since precipitation is affected by both (*Tebaldi and Knutti* 2007). For simplicity, it is assumed that TAS and PSL could also be used to weight precipitation ensembles. Temperature and precipitation ensembles from the same project and season and scenario will then share the weights for each model. The reference period used is 1980-2014 as defined in *Brunner et al.* (2020). Note that CMIP5 reference periods will include historical runs data (1980-2005) and the first years of future projections (2006-2014).

The time aggregations used for the performance assessment are trends, anomalies and variability. Temperature historical trends have a high correlation and evident physical link with future warming (*Tebaldi and Knutti* 2007). The trend is calculated by fitting a linear time trend during the reference period to each grid point (TREND); the climatologic anomaly is computed by subtracting the area averaged climatology to each grid point's climatology of the reference period (ANOM) and the variability is represented by the standard deviation over the time coordinate for each grid point (STD). TREND is not relevant for PSL and it is not computed (Merrifield2020).

The aim when assessing a model independence is to clearly identify members that behave similarly. Using relatively long periods for climatologies is a good approach as it minimizes the internal variability, which ideally is the main factor distinguishing between two members of the same model differ (*Hawkins and Sutton* 2011). Biases in models are well represented by the CLIM diagnostic. Therefore, to compute the distances S_{ij}, CLIM is used both for surface temperature and sea level pressure (TAS-CLIM and PSL-CLIM).

As explained in the main text, the shape parameters act as thresholds on the performance and independence values obtained from the RMSE D_i and S_{ij}, respectively. Low values of σ_d will make the performance weighting very strict as low values of D_i will be needed to make the equation (1) numerator's exponent small enough for the member i to receive performance weight. High values of σ_d will allow greater distances between models and observations.

Too small values of σ_s will make all models seem independent as the distance to consider two members dependent will have to be very small. This would make all models receive similar weights. Too large values of σ_s will make most models seem dependent as large distances S_{ij} would still be considered small enough for models to be dependent. Therefore making models receive similar weights. An optimal σ_s that is neither too large nor too small, and capable of discriminating independence must be found (*Knutti et al.* 2017).

The information given by the ensemble is needed to make a best estimation of both shape parameters. The parameter σ_s feeds on the information given by models with more than one member. Initial-condition runs from the same model should be considered completely dependent as the physics behind are the same. Nevertheless, internal variability makes the runs differ. With this in mind, the independence weighting should identify when members (initial-condition runs) of the same model are added or subtracted from the ensemble. If the denominator of equation (1) (independence weighting) is computed for an ensemble with only one member per model (w_j^{ind}) and then E_j members (all the available members) of only one model j are added to the ensemble, the independence weights of model j ($\tilde{w}_j^{ind}$) are expected to decrease. On the other hand, adding members of a model j to the ensemble should have a minimal effect on the independence weights of the rest of models i represented by only one member in the ensemble.

To find the optimal σ_s an iterative process for a range of σ_s candidates is conducted with the aim of minimizing the sum $\epsilon_1 + \epsilon_2$, where:

$$mean_j \left[w_j^{ind}(\sigma_s) + E_j - \tilde{w}_j^{ind}(\sigma_s) \right]^2 = \epsilon_1$$
$$mean_j \left\{ mean_i \left[w_{i \neq j}^{ind}(\sigma_s) - \tilde{w}_{i \neq j}^{ind}(\sigma_s) \right]^2 \right\} = \epsilon_2 \ \forall \mathrm{j}$$

Small values of σ_d would discriminate more strictly the model's performance, but this could lead to an overconfident weighting, giving too narrow spreads for the projections. To find the ideal σ_d, as suggested in *Knutti et al.* (2017), perfect model tests are conducted for a range of possible σ_d (from 10% to 200% of the median distance D_i between the models and the observational reference).

Table I: Summary of the CMIP5. CMIP6 and HighResMIP multi-model ensemble used in this study. The model initial-conditions members are listed according to their emission scenarios.

CMIP5	**RCP2.6**	**RCP4.5**	**RCP8.5**	**CMIP6**	**SSP1-2.6**	**SSP2-4.5**	**SSP5-8.5**
ACCESS1-0	-	r1i1p1	r1i1p1	ACCESS-CM2	r1i1p1f1	r1i1p1f1	r1i1p1f1
ACCESS1-3	-	r1i1p1	r1i1p1	ACCESS-ESM1-5	r(1-3)i1p1f1	r(1-3)i1p1f1	r(1-3)i1p1f1
BCC-CSM1-1	r1i1p1	r1i1p1	r1i1p1	AWI-CM-1-1-MR	r1i1p1f1	r1i1p1f1	r1i1p1f1
BCC-CSM1-1-M	r1i1p1	r1i1p1	r1i1p1	BCC-CSM2-MR	r1i1p1f1	r1i1p1f1	r1i1p1f1
BNU-ESM	r1i1p1	r1i1p1	r1i1p1	CanESM5	r(1-10)i1p1f1	r(1-10)i1p1f1	r(1-10)i1p1f1
CanESM2	r(1-5)i1p1	r(1-5)i1p1	r(1-5)i1p1	CanESM5-CanOE	r(1-3)i1p1f1	r(1-3)i1p1f1	r(1-3)i1p1f1
CCSM4	r(1-5)i1p1	r(1-5)i1p1	r(1-6)i1p1	CESM2	r1i1p1f1	r1i1p1f1	r1i1p1f1
CESM1-BGC	-	-	r1i1p1	CESM2-WACCM	r1i1p1f1	r1i1p1f1	r1i1p1f1
CESM1-CAM5	r(1-3)i1p1	r(1-3)i1p1	r(1-3)i1p1	CMCC-CM2-SR5	-	-	r1i1p1f1
CMCC-CESM	-	-	r1i1p1	CNRM-CM6-1	r(1-6)i1p1f2	r(1-3)i1p1f2	r1i1p1f2
CMCC-CM	-	r1i1p1	r1i1p1	CNRM-CM6-1-HR	r1i1p1f2	r1i1p1f2	r1i1p1f2
CMCC-CMS	-	r1i1p1	r1i1p1	CNRM-ESM2-1	r1i1p1f2	r(1-5)i1p1f2	r1i1p1f2
CNRM-CM5	r1i1p1	r1i1p1 (only pr)	r(1-2,4,6,10)i1p1	FGOALS-g3	r1i1p1f1	r1i1p1f1	r1i1p1f1
CSIRO-Mk3-6-0	r(1-10)i1p1	r(1-10)i1p1	r(1-10)i1p1	FGOALS-f3-L	r1i1p1f1	r1i1p1f1	r1i1p1f1
EC-Earth	r(2,12)i1p1	r(2,9,12)i1p1	r(2,8,9,12)i1p1	FIO-ESM-2-0	r(1-3)i1p1f1	r(1-3)i1p1f1	r(1-3)i1p1f1
FGOALS-s2	-	r1i1p1	r(1-3)i1p1	GFDL-ESM4	r1i1p1f1	r1i1p1f1	r1i1p1f1
FGOALS-g2	-	-	r1i1p1 (no pr)	GISS-E2-1-G	r1i1p3f1	r(1,3)i1p3f1	r1i1p3f1
FIO-ESM	r(1:3)i1p1	r(1-3)i1p1	r(1-3)i1p1	HadGEM3-GC31-LL	r1i1p1f3	r1i1p1f3	r(1-3)i1p1f3
GFDL-CM3	r1i1p1	r1i1p1	r1i1p1	INM-CM4-8	r1i1p1f1	r1i1p1f1	r1i1p1f1
GFDL-ESM2G	r1i1p1	r1i1p1 (no pr)	r1i1p1	INM-CM5-0	r1i1p1f1	r1i1p1f1	r1i1p1f1
GFDL-ESM2M	r1i1p1	r1i1p1 (no pr)	r1i1p1	IPSL-CM6A-LR	r(1-4,6)i1p1f1	r(1-6,10,11,14,22,25)i1p1f1	r1i1p1f1
GISS-E2-H	r1i1p1	r(1-5)i1p1	r(1-2)i1p1	KACE-1-0-G	r(1-2)i1p1f1	r(1-3)i1p1f1	r1i1p1f1
GISS-E2-H-CC	-	r1i1p1 (no pr)	r1i1p1	MIROC-ES2L	r1i1p1f2	r1i1p1f2	r1i1p1f2
GISS-E2-R	r1i1p1	r(2,5,6)1i1p3	r(1-2)i1p1	MIROC6	r(1-3)i1p1f1	r(1-3)i1p1f1	r(1-3)i1p1f1
GISS-E2-R-CC	-	r1i1p1 (no pr)	r1i1p1	MPI-ESM1-2-HR	r1i1p1f1	r1i1p1f1	r1i1p1f1
HadGEM2-AO	r1i1p1	r1i1p1 (only pr)	r1i1p1	MPI-ESM1-2-LR	r(1-10)i1p1f1	r(1-10)i1p1f1	r(1-10)i1p1f1
HadGEM2-CC	-	r1i1p1	r1i1p1	MRI-ESM2-0	r1i1p1f1	r1i1p1f1	r1i1p1f1
HadGEM2-ES	r(1-4)i1p1	r(1-4)i1p1	r(1-4)i1p1	NESM3	r(1-2)i1p1f1	r(1-2)i1p1f1	r(1-2)i1p1f1
INMCM4	-	r1i1p1	r1i1p1	NorESM2-LM	r1i1p1f1	r1i1p1f1	r1i1p1f1
IPSL-CM5A-LR	r(1-4)i1p1	-	r(1-4)i1p1	NorESM2-MM	r1i1p1f1	r1i1p1f1	r1i1p1f1
IPSL-CM5A-MR	r1i1p1	r1i1p1	r1i1p1	UKESM1-0-LL	r(1-4,8)i1p1f2	r(1-4,8)i1p1f2	r(1-4,8)i1p1f2
IPSL-CM5B-LR	-	r1i1p1	r1i1p1	**HighResMIP**	**SSP5-8.5**		
MIROC-ESM	r1i1p1	r1i1p1	r1i1p1	CMCC-CM2-HR4	r1i1p1f1		
MIROC-ESM-CHEM	r1i1p1	r1i1p1	r1i1p1	CMCC-CM2-VHR4	r1i1p1f1		
MIROC5	r(2-3)1i1p1	r(2-3)i1p1	r(2-3)i1p1	CNRM-CM6-1-HR	r1i1p1f1		
MPI-ESM-LR	r(1-3)i1p1	r(1-3)i1p1	r(1-3)i1p1	EC-Earth3P-HR	r2i1p2f1		
MPI-ESM-MR	r1i1p1	r(1-3)i1p1	r1i1p1	HadGEMGE3-GC31-HH	r1i1p1f1		
MPI-CGCM3	-	r1i1p1	r1i1p1	HadGEMGE3-GC31-HM	r1i1p1f1		
NorESM1-M	r1i1p1	r1i1p1	r1i1p1	HadGEMGE3-GC31-MM	r1i1p1f1		

Table II: HighResMIP markers used in Figure 3 to highlight differences in resolution from the same model.

Marker	LR/HR Models
▼	CMCC-CM2-SR5 / CMCC-CM2-VHR4
✖	CNRM-CM6-1 / CNRM-CM6-1-HR
✚	EC-Earth3P / EC-Earth3P-HR
●	HadGEM3-GC31-LL / HadGEM3-GC31-HM

Table III: Observational references summary.

Name	Type	Institute	Variables	Reference
JRA55	Reanalysis	Japan Weather Association (JWA)	TAS, PR, PSL	(*Kobayashi et al.* 2015)
ERA5	Reanalysis	European Centre for Medium-Range Weather Forecasts (ECMWF)	TAS, PSL	(*Hersbach et al.* 2020)
CRU (v4.04)	Gridded observations	University of East Anglia (UEA)	TAS, PR	(*Harris et al.* 2020)
GPCC (v2018)	Gridded observations	Deutscher Wetterdienst (DWD)	PR	(*Schamm et al.* 2014)
E-OBS (v20)	Gridded observations	European Climate Assessment & Dataset (ECAD)	PR, TAS	(*Cornes et al.* 2018)

Consecutively, all models are considered as the reference while the rest are weighted following equation (1) with D_i being the distance between the perfect model and model i. A σ_d is considered to be big enough to avoid giving overconfident weighting when 80% of the perfect models fall in between the 10th and 90th percentiles of the weighted ensembles they produced. The diagnostics used in the test are the same as the ones that were presented previously but computed for the future periods (2041-2060 and 2081-2100) as the aim is to base σ_d on the ensemble projection uncertainties. The average σ_d between both periods is used for the rest of the study.