



## RESEARCH ARTICLE

# Weights for general circulation models from CMIP3/CMIP5 in a statistical downscaling framework and the impact on future Mediterranean precipitation

Irena Kaspar-Ott<sup>1</sup>  | Elke Hertig<sup>1</sup>  | Severin Kaspar<sup>1</sup> | Felix Pollinger<sup>2</sup> | Christoph Ring<sup>2</sup> | Heiko Paeth<sup>2</sup> | Jucundus Jacobeit<sup>1</sup>

<sup>1</sup>Institute of Geography, University of Augsburg, Augsburg, Germany

<sup>2</sup>Institute of Geography and Geology, University of Wuerzburg, Germany

**Correspondence**

Irena Kaspar-Ott, Institute of Geography, University of Augsburg, Alter Postweg 118, 86159 Augsburg, Germany.

Email: irena.kaspar-ott@geo.uni-augsburg.de

**Funding information**

German Research Foundation

This study investigates the projected precipitation changes of the 21st century in the Mediterranean area with a model ensemble of all available CMIP3 and CMIP5 data based on four different scenarios. The large spread of simulated precipitation change signals underlines the need of an evaluation of the individual general circulation models in order to give higher weights to better and lower weights to worse performing models. The models' spread comprises part of the internal climate variability, but is also due to the differing skills of the circulation models. The uncertainty resulting from the latter is the aim of our weighting approach. Each weight is based on the skill to simulate key predictor variables in context of large and medium scale atmospheric circulation patterns within a statistical downscaling framework for the Mediterranean precipitation. Therefore, geopotential heights, sea level pressure, atmospheric layer thickness, horizontal wind components and humidity data at several atmospheric levels are considered. The novelty of this metric consists in avoiding the use of the precipitation data by itself for the weighting process, as state-of-the-art models still have major deficits in simulating precipitation. The application of the weights on the downscaled precipitation changes leads to more reliable and precise change signals in some Mediterranean sub-regions and seasons. The model weights differ between sub-regions and seasons, however, a clear sequence from better to worse models for the representation of precipitation in the Mediterranean area becomes apparent.

**KEYWORDS**

CMIP3, CMIP5, Mediterranean area, precipitation, statistical downscaling, weights

## 1 | INTRODUCTION

The growing amount of available output from climate models over the last years also increases the need to evaluate the quality and skill of the offered data. The newest ensemble of state-of-the-art general circulation models and earth system models from the Coupled Model Intercomparison

Project Phase 5 (CMIP5, Taylor *et al.*, 2012), for example, allows the analysis of more than 50 different models. Researchers like climatologists, hydrologists and impact modellers often need to choose a sub-ensemble of models for their work. Thus, there were several attempts over the last years to figure out the skills of climate model data and to form superior sub-ensembles.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. International Journal of Climatology published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

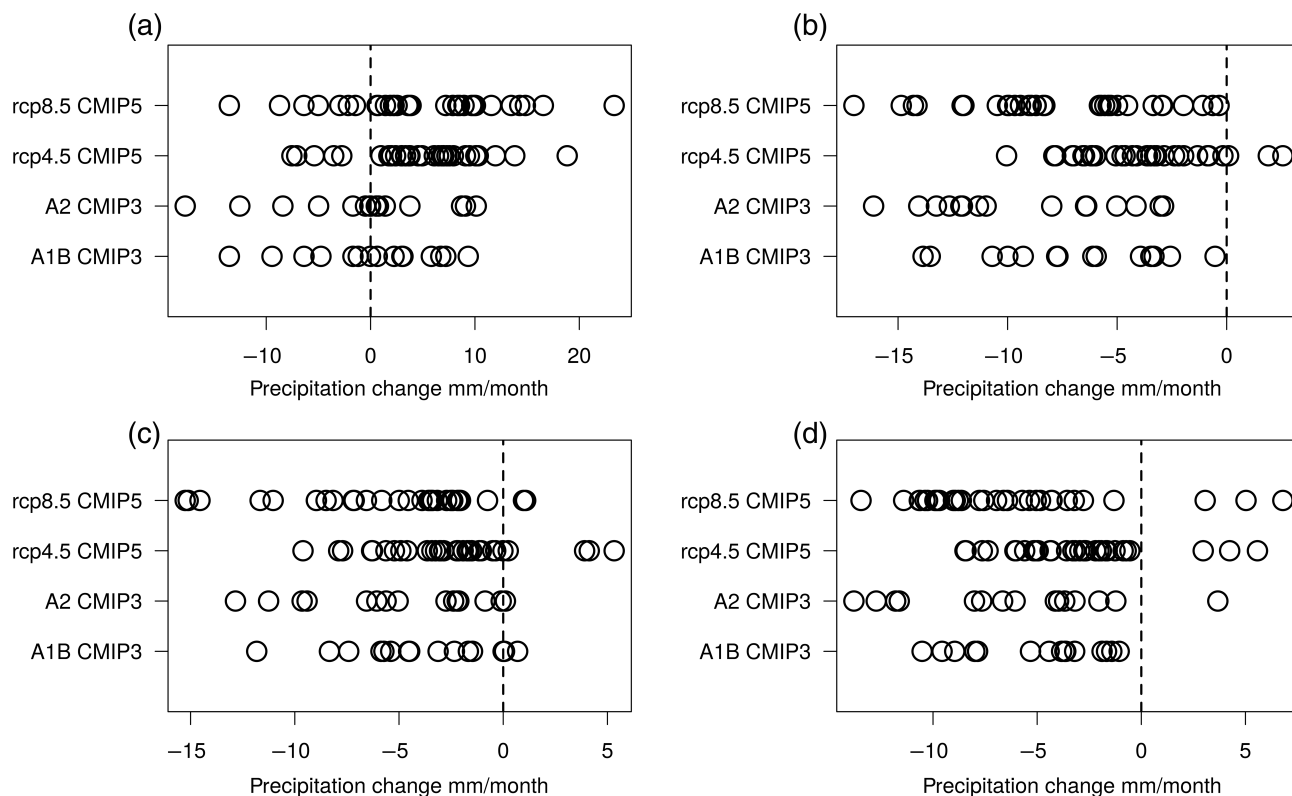
So far, a large number of different metrics for model weighting were tested, first of all performance-based metrics like the reliability ensemble averaging (Giorgi and Mearns, 2002), approaches with Bayesian statistics (Tebaldi *et al.*, 2005), the comparison of probability densities (Boberg *et al.*, 2009; Kjellström *et al.*, 2010) and approaches that analyse trends, extremes or the large- and meso-scale atmospheric circulation (Christensen *et al.*, 2010).

However, scientists often encounter similar problems, like the necessity to make subjective decisions during the model ranking process including the pre-processing of the data, the choice of the metrics and their useful combination (Gleckler *et al.*, 2008). Questionable features are also the inter-model similarities, the dependency between weighting metrics and the quality of the observational or reanalysis data used for comparison (Christensen *et al.*, 2010; Knutti *et al.*, 2017). Another important issue is the consistency of model rankings over time and the applicability of weights on climate projections (Whetton *et al.*, 2007; Reifen and Toumi, 2009; Macadam *et al.*, 2010; Räisänen *et al.*, 2010; Räisänen and Ylhäisi, 2012). A weighting scheme that is based on incorrect assumptions may even cause a decline of the model ensemble's performance (Weigel *et al.*, 2010).

The goal of our study is to generate weights for all available CMIP3 and CMIP5 models with respect to the precipitation changes in the Mediterranean area—a so-called hot

spot of climate change (Giorgi, 2006)—at the end of the 21st century. The meaningful weighting of precipitation change signals is expected to produce more reliable and in the best case more precise results. The need of less manifold projected precipitation changes is underlined by the examples in Figure 1, showing the extreme spread of projected precipitation changes in the study area. The simulated changes of the single ensemble members differ extremely, even from negative to positive change signals within the same sub-region and season. The spread comprises part of the internal climate variability, but is also due to the differing skills of the circulation models. The uncertainty resulting from the latter is the aim of our weighting approach.

Precipitation itself is not used to weight models in our study, as state-of-the-art models still have major deficits in simulating precipitation due to its complex generation mechanisms and the heterogeneous spatial and temporal distributions (e.g., Trigo and Palutikof, 2001; Mueller and Seneviratne, 2014). However, more trustworthy modelled climatological elements, representing the large and medium scale atmospheric patterns that cause precipitation, are analysed. Thus, key predictors for the Mediterranean rainfall are identified with reanalysis data in a statistical downscaling framework. The biases of the identified key predictors are used as weights for the climate models. In this way, the problematic effects of the insufficient simulation of



**FIGURE 1** Precipitation changes in mm/month from 1950–1999 to 2070–2099 for four scenario runs from the single ensemble members of CMIP3 (A1B, A2) and CMIP5 (rcp4.5, rcp8.5) in (a) sub-region 8 in winter (DJF), (b) sub-region 3 in spring (MAM), (c) sub-region 1 in summer (JJA) and (d) sub-region 7 in autumn (SON). For definitions of sub-regions see Section 2.1 and Figure 2

precipitation in the models are avoided. Only the change signal from the downscaled model data is considered and therefore precipitation biases do not affect the weighting results.

The remainder of this paper is structured as follows: Section 2 describes the data and methods, with the definition of Mediterranean sub-regions in Section 2.1, the predictor screening in Section 2.2, the downscaling approach in Section 2.3, and the predictor description of the climate model data in Section 2.4. The preparation of model weights is presented in Section 3, followed by the application of these weights on the Mediterranean precipitation changes in Section 4. The continuance of model weights in time and space is discussed in Section 5. The summary of this work and conclusions are given in Section 6.

## 2 | DATA AND METHODS

### 2.1 | Regionalization of precipitation variability in the Mediterranean region

Mediterranean sub-regions for the target variable precipitation should meet the following conditions: They are determined by similar precipitation variations within a region and they should be valid for all months and seasons. For this purpose, the daily EOBS precipitation data (version 12, Haylock *et al.*, 2008) with a  $0.25^\circ \times 0.25^\circ$  grid size from 1950 to 2010 in the domain  $12^\circ\text{W}$ – $40^\circ\text{E}$  and  $28^\circ\text{N}$ – $46^\circ\text{N}$  was used. The data set was conservatively interpolated to a  $2^\circ \times 2^\circ$  grid. Grid boxes with more than 12 missing months were removed beforehand with a missing month being defined as one with more than two missing days (Hertig and Jacobeit, 2014). Grid boxes with large numbers of missing months are mainly located in eastern Turkey and in North Africa.

Regionalization was done using a s-mode Varimax rotated principal component analysis (PCA) based on a correlation matrix with annually aggregated precipitation sums in the Mediterranean domain. To take time-varying

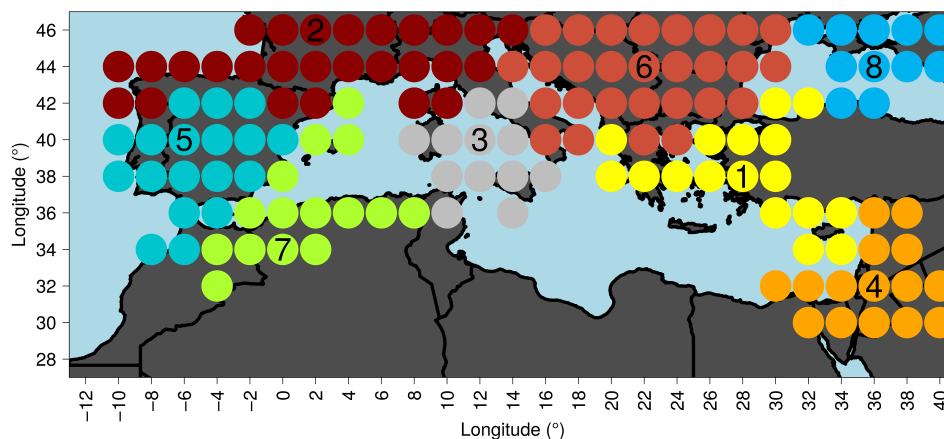
climatological states into account, the precipitation time series from 1950 to 2010 are bootstrapped 2000 times, whereby 30 values are drawn at random and processed with the PCA. The best result is achieved with eight principal components, identified by the dominance criteria (Jacobeit, 1993). To unite the 2000 PCA results, a differentially initialized k-means (DKM) cluster analysis is used (Enke and Spekat, 1997). Figure 2 shows the result of the regionalization process with following sub-regions: *Greece-Turkey* (1), *North-western area* (2), *Tyrrhenian Sea riparians* (3), *Eastern Mediterranean* (4), *Iberian Peninsula* (5), *Balkans* (6), *Maghreb* (7) and *Eastern Black Sea* (8).

The seasonal (single-monthly) and spatially averaged precipitation time series of these eight sub-regions serve as predictand variables within the downscaling framework of this study.

To test the robustness of the regionalization results, further analyses with seasonal-split EOBS data and also reanalysis data (NCEP-NCAR) were made. Differences between the regionalized areas of annual and seasonal EOBS and the reanalysis data appear, however the coarse structure of the sub-regions remains. To proof the robustness of the sub-regions, further, we correlated the spatially averaged time series based on the regionalization results of annual/seasonal EOBS and the reanalysis. The correlation matrices showed that the resulting precipitation time series of the same sub-regions can be clearly assigned and correspond well to each other (Spearman correlation, significant with  $p < 0.01$ ). This justifies the use of the annually-based regionalization result from the EOBS data for the further research purpose.

### 2.2 | Predictor screening

Literature of downscaling approaches within the Mediterranean region (e.g., Tatli *et al.*, 2004; Xoplaki *et al.*, 2004; Palatella *et al.*, 2010; Lutz *et al.*, 2012; Hertig and Jacobeit, 2013; Hertig *et al.*, 2014; Jacobeit *et al.*, 2014) provides a first selection of possible important predictor variables for



**FIGURE 2** Area of investigation with sub-regions of homogeneous precipitation variability. Greece-Turkey (1), North-western area (2), Tyrrhenian Sea riparians (3), Eastern Mediterranean (4), Iberian Peninsula (5), Balkans (6), Maghreb (7) and Eastern Black Sea (8)

the predictand precipitation: geopotential heights at 700 and 500 hPa (zg700, zg500), the atmospheric layer thickness between 925 and 500 hPa (thick500–925) as a proxy for atmospheric temperature of this layer, sea level pressure (psl), zonal and meridional wind velocities at 700 hPa (ua700, va700), specific humidity at 850 and 700 hPa (hus850, hus700) and relative humidity at 850 and 700 hPa (hur850, hur700).

For the following analyses, NCEP-NCAR reanalysis data (Kalnay *et al.*, 1996) for the period 1950–1999, re-gridded to a  $2^\circ \times 2^\circ$  spatial resolution, is used. Although there are some other well-established reanalysis products, NCEP-NCAR reanalysis seemed to be the best alternative for our project, as it provides long-term data (compared to, e.g., ERA-Interim) which are required for the solid calibration and validation of the multiple linear regressions and furthermore does not solely base on the assimilation of surface variables (like for instance NOAA 20CR or ERA-20C).

The variables representing the large-scale atmospheric circulation (zg, thick, psl, ua, va) cover the North-Atlantic/European sector from  $50^\circ\text{W}$  to  $50^\circ\text{E}$  and from  $24^\circ\text{N}$  to  $66^\circ\text{N}$ . To capture the more regional influence of humidity on precipitation, a smaller domain is chosen for hur and hus ( $12^\circ\text{W}$ – $40^\circ\text{E}$  and  $28^\circ\text{N}$ – $46^\circ\text{N}$ ). To reduce these variables to a set of manageable and physically meaningful predictors, seasonal (single-monthly) s-mode PCAs are computed for each variable. The PCAs involve the whole period from 1950 to 1999, are based on the correlation matrix, and their results are Varimax rotated. The input time series are weighted for their latitudes. The number of extracted PCs is determined by the dominance criteria after Jacobeit (1993) and Philipp (2003) with additional conditions: for at least one variable, a particular PC must dominate all other ones by at least more than one standard deviation of the corresponding loadings, the PC must also dominate the other ones at least at eight grid points to have a skilful scale, and the PC must explain at least 4% of the overall variance. The resulting PC loadings indicate the location of variability

centres within the domain, the PC scores represent the corresponding time series and are used as predictor time series in this study.

Table 1 shows the resulting number of PCs and the total explained variances for the four seasons winter (DJF), spring (MAM), summer (JJA) and autumn (SON).

### 2.3 | Downscaling approach

To determine the most important atmospheric predictor variables from the reanalysis data for the eight precipitation sub-regions (= predictands, i.e., the seasonal and spatially averaged precipitation time series from EOBS defined in Section 2.1), multiple linear regressions (MLRs) are computed considering some crucial points: The relationships between predictor and predictand variables might be unsteady in time (1), the 10 chosen predictors from Section 2.2 and their PCs might be correlated (2), and the requirements of the MLR method like normally distributed input data as well as normal distribution, no autocorrelation and homoscedasticity of the residuals must be fulfilled (3).

To meet the first point (1), a bootstrapping method is applied, where 90 out of 150 time steps (3 months per season for 50 years are available at all) are randomly taken to calibrate the MLR. The remaining 60 values are used for the model validation. This random selection including calibration and validation is repeated 800 times to generate downscaling models that are valid for the whole investigation period—independent from non-stationarities within the predictor-predictand relationships. To cope with possibly dependent predictor variables (point 2), a stepwise reduction of predictors was developed.

First, all PCs of the predictor variables from NCEP-NCAR serve as input for the MLR, whereas successively one by one is set to zero. The predicted time series from all regression models, calculated using the adjusted input, are compared with the original validation time series and the Mean Squared Error (MSE) is computed. It is assumed that

**TABLE 1** Number of principal components (PCs) and explained variances (EVs) for the possible predictor variables of NCEP-NCAR from 1950 to 1999

	DJF		MAM		JJA		SON	
	PCs	EV %	PCs	EV %	PCs	EV %	PCs	EV %
zg700	7	90	9	95	9	89	6	88
zg500	8	91	8	94	8	88	5	89
thick	9	91	5	89	8	88	5	93
psl	7	91	7	88	8	86	8	90
ua700	7	87	8	82	8	77	8	81
va700	9	86	9	82	9	71	7	75
hus850	4	75	7	89	6	76	7	89
hus700	7	85	6	85	6	69	7	87
hur850	6	78	6	84	8	86	5	79
hur700	8	84	8	79	8	80	7	80

*Note.* Chosen predictors are geopotential heights (zg) at 700 and 500 hPa, atmospheric layer thickness between 925 and 500 hPa (thick), sea level pressure (psl), zonal and meridional wind components (ua, va) at 700 hPa, specific (hus) and relative (hur) humidity at 850 and 700 hPa.

the predictor set to zero with the highest MSE has the greatest influence on the predictand.

In a second step, all predictors are correlated and each couple of correlated predictors that exceed  $r = 0.3$  is checked for the MSE from the first step. The predictor with the higher MSE remains as important one, the predictor with the lower MSE is removed from the predictors list. The remaining predictors are uncorrelated (below the defined threshold of 0.3, according to a level of significance of 99.6%).

In a third step, the remaining uncorrelated predictors are again, one by one, set to zero, the MLRs are calculated and the respective MSEs between predicted and original precipitation time series are determined.

To further reduce the number of predictors and to find out the optimal set of predictors, MLRs are calculated with at least five up to the maximum number of remaining uncorrelated predictors, containing the most important predictors resulting from the MSEs from step three. The best validation result, given by the  $R^2$  between original and predicted precipitation time series, determines the optimal set of predictor variables.

Now, on the one hand, the most important predictors for the appropriate region and season are derived by the difference of the MSE, and on the other hand, the optimal set of predictors is determined by the highest  $R^2$  of the validation periods.

The  $R^2$  and the MSSS (Mean Squared Error Skill Score) are calculated for the calibration and the validation periods. Additionally, the residuals are checked for normal distribution (Shapiro test), autocorrelation (Durban-Watson test) and heteroskedasticity (Breusch-Pagan test), which fulfils the third requirement (3).

The stepwise reduction of predictors and the calculation of a final MLR are repeated 800 times with randomly picked calibration and validation values from 1950 to 1999. All models that do not fulfil all requirements of the MLR (with  $p$ -values  $>0.05$  and the null hypothesis that the data is

normally distributed/not autocorrelated/no heteroskedasticity) are neglected, and from the remaining regression models, only 25% of the best models are used to determine the final predictor variables. Thereby, the performance of the MLRs is calculated by the sum of  $R^2$  and MSSS of the calibration and validation results.

Now, the frequency of predictor values which pass the predictor reduction process and which are contained in the best MLRs, determines the importance of the single predictor variables. The average optimal number of predictors, resulting from the bootstrapping process, determines the final number of key predictor variables (e.g., if the average optimal number of predictors is 10, the 10 most frequent and thus most important predictors become the key predictors).

For all seasons and sub-regions, valid regression models could be found, except for region 4 (*Eastern Mediterranean*) in summer (JJA) and autumn (SON), and for region 7 (*Maghreb*) in summer. The precipitation sums are too small in the southernmost sub-regions in summer/autumn, so that the calculation of valid MLRs was not possible.

For all successfully validated regions and seasons, the final MLRs can be calculated with the key predictors from the reanalysis data.

Table 2 shows an example of the key predictors and the corresponding goodness of the MLR model ( $R^2$ ) for the winter season for all eight sub-regions.

Table 3 shows the relative frequency of predictor variables for all four seasons. Pressure variables and in particular the geopotential heights at 500 hPa dominate the chosen key predictors. Over the year, they contribute more than 50% of the key predictors. The second important group of atmospheric predictors is the moisture variables of relative and specific humidity at the 700 and 850 hPa levels with an average frequency of almost 30% over the year. The most frequently chosen moisture predictor is the relative humidity of 700 hPa, followed by the specific humidity at the same pressure level. However, the importance of moisture as

**TABLE 2** Final key predictors and the  $R^2$  of the corresponding MLR for winter (DJF) in alphabetical order

	<b>Region 1</b>	<b>Region 2</b>	<b>Region 3</b>	<b>Region 4</b>	<b>Region 5</b>	<b>Region 6</b>	<b>Region 7</b>	<b>Region 8</b>
$R^2$	57	73.6	62.2	56.2	76.6	74.8	59	65.5
1	hus700_4	hus700_6	hur700_6	hur700_5	hur700_6	hus700_6	hus700_3	ua700_6
2	hus700_6	hus850_2	ua700_6	hur850_1	hus700_3	hus850_2	hus700_7	va700_8
3	ua700_6	thick_6	va700_7	hus700_7	hus700_6	zg500_1	ua700_6	va700_9
4	va700_7	va700_7	va700_9	ua700_6	hus850_2	zg500_2	va700_7	zg500_1
5	va700_9	va700_8	zg500_2	va700_9	thick_6	zg500_3	zg500_1	zg500_2
6	zg500_2	zg500_2	zg500_3	zg500_3	zg500_1	zg500_4	zg500_3	zg500_3
7	zg500_3	zg500_3	zg500_4	zg500_4	zg500_2	zg500_5	zg500_4	zg500_5
8	zg500_5	zg500_4	zg500_5	zg500_5	zg500_3	zg500_6	zg500_6	zg500_6
9	zg500_6	zg500_5	zg500_6	zg500_7	zg500_5	zg500_7	zg500_7	zg500_7
10	zg700_3	zg500_6	zg500_7		zg500_7	zg700_3	zg700_2	zg500_8
11								zg700_1

Note. The number behind the variables and levels denotes the number of the PC. For variable abbreviations see Table 1.

TABLE 3 Percentage frequencies of predictor variables from the final MLR setups for the four seasons

%	zg500	zg700	Psl	Thick500-925	va700	ua700	hur700	hur850	hus700	hus850
DJF	53.8	5.0	0.0	2.5	12.5	6.3	3.8	1.3	11.3	3.8
MAM	44.3	11.4	2.5	1.3	13.9	0.0	10.1	1.3	8.9	6.3
JJA	47.8	4.3	2.2	3.3	4.3	0.0	20.7	9.8	3.3	4.3
SON	40.3	9.1	3.9	0.0	7.8	10.4	13.0	0.0	13.0	2.6
All	46.6	7.3	2.1	1.8	9.5	4.0	12.2	3.4	8.8	4.3

Note. For variable abbreviations see Table 1.

predictor variable varies through the seasons. It is most important in summer providing almost 40% of all predictor variables and less important in winter with about 20% at all. The table also shows that specific humidity gains in importance in autumn and in winter compared to relative humidity. The third frequent predictors are the zonal and meridional wind velocities, whereby the meridional wind component is more important than the zonal one, except for autumn. Similar to the atmospheric moisture, the importance of the wind as a predictor for Mediterranean precipitation underlies a seasonal variation, as it is most important in autumn and winter and less significant in spring and especially in summer. The atmospheric layer thickness plays only a marginal role within the key predictor setup and has its greatest influence in summer with 3.3%.

## 2.4 | Downscaling historical and future precipitation of CMIP3 and CMIP5

The GCMs that will be weighted in this paper belong to the third (Meehl *et al.*, 2007) and fifth (Taylor *et al.*, 2012) Coupled Model Intercomparison Projects (CMIP) from the Intergovernmental Panel on Climate Change (IPCC). All available data from CMIP3 and CMIP5 are used in this study. The availability of atmospheric variables limits the number of investigated models to 22 from CMIP3 and 29 from CMIP5. Only the first realization of each model is used. The chosen data include the second half of the 20th century (1950–1999, called historical in CMIP5 and 20c3m in CMIP3) as well as the end of the 21st century (2070–2099) for four future climate scenarios (rcp4.5 and rcp8.5 for CMIP5, A1B and A2 for CMIP3). Tables 4 and 5 list the CMIP models used in this study.

Equivalent to the EOBS data and the NCEP-NCAR reanalysis, the CMIP model data is re-gridded to a  $2^\circ \times 2^\circ$  grid. The fields of the modelled atmospheric variables are normalized over the investigated time period 1950–1999/2070–2099 (separately for each scenario) and then projected onto the loading patterns of the PCAs from NCEP-NCAR (from Section 2.2). The key predictors, defined by the reanalysis and observed data in Section 2.3, are used to drive the MLRs. The resulting downscaled precipitation data from CMIP3 and CMIP5 models for the historical and the four future scenario runs are used to calculate the precipitation change signals in the eight Mediterranean sub-regions.

## 3 | GENERATING MODEL WEIGHTS

In contrast to the previous pre-processing of the CMIP predictors in Section 2.4, the fields of the historical modelled atmospheric variables are normalized by the corresponding mean and standard deviation of the NCEP-NCAR

TABLE 4 CMIP3 models used in this study

CMIP3		
No.	Model name	Institution
1	bcc-csm1-1	Beijing Climate Center, China
2	bcc-csm1-1-m	Beijing Climate Center, China
3	bccr-bcm2-0	Bjerknes Centre for Climate Research, Norway
4	cnrm-cm3	Météo-France/Centre National de Recherches Météorologiques, France
5	csiro-mk3-0	CSIRO Atmospheric Research, Australia
6	csiro-mk3-5	CSIRO Atmospheric Research, Australia
7	gfdl-cm2-0	US Dept. of Commerce/NOAA/Geophysical Fluid Dynamics Laboratory, USA
8	gfdl-cm2-1	US Dept. of Commerce/NOAA/Geophysical Fluid Dynamics Laboratory, USA
9	giss-model-e-h	NASA/Goddard Institute for Space Studies, USA
10	giss-model-e-r	NASA/Goddard Institute for Space Studies, USA
11	iap-fgoals1-0-g	LASG/Institute of Atmospheric Physics, China
12	ingv-echam4	Instituto Nazionale di Geofisica e Vulcanologia, Italy
13	inmcm3-0	Institute for Numerical Mathematics, Russia
14	ipsl-cm4	Institut Pierre Simon Laplace, France
15	miroc3-2-hires	Center for Climate System Research (The University of Tokyo), National Institute for Environmental Studies, and Frontier Research Center for Global Change (JAMSTEC), Japan
16	miroc3-2-medres	Center for Climate System Research (The University of Tokyo), National Institute for Environmental Studies, and Frontier Research Center for Global Change (JAMSTEC), Japan
17	mpi-echam5	Max Planck Institute for Meteorology, Germany
18	mri-cgcm2-3-2a	Meteorological Research Institute, Japan
19	ncar-ccsm3-0	National Center for Atmospheric Research, USA
20	ncar-pcm1	National Center for Atmospheric Research, USA
21	ukmo-hadcm3	Hadley Centre for Climate Prediction and Research/Met Office, UK
22	ukmo-hadgem1	Hadley Centre for Climate Prediction and Research/Met Office, UK

TABLE 5 CMIP5 models used in this study

CMIP5		
No.	Model name	Institution
1	ACCESS1-0	Commonwealth Scientific and Industrial Research Organization (CSIRO) and Bureau of Meteorology (BOM), Australia
2	ACCESS1-3	Commonwealth Scientific and Industrial Research Organization (CSIRO) and Bureau of Meteorology (BOM), Australia
3	CanESM2	Canadian Centre for Climate Modelling and Analysis, Canada
4	CCSM4	University of Miami-RSMAS, USA
5	CESM1-CAM5	Community Earth System Model Contributors, USA
6	CMCC-CM	Centro Euro-Mediterraneo per I Cambiamenti Climatici, Italy
7	CMCC-CMS	Centro Euro-Mediterraneo per I Cambiamenti Climatici, Italy
8	CNRM-CM5	Centre National de Recherches Météorologiques/Centre Européen de Recherche et Formation Avancée en Calcul Scientifique, France
9	CSIRO-Mk3-6-0	Commonwealth Scientific and Industrial Research Organization in collaboration with Queensland Climate Change Centre of Excellence, Australia
10	FGOALS-g2	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences and CESS, Tsinghua University, China
11	GFDL-CM3	NOAA Geophysical Fluid Dynamics Laboratory, USA
12	GFDL-ESM2G	NOAA Geophysical Fluid Dynamics Laboratory, USA
13	GFDL-ESM2M	NOAA Geophysical Fluid Dynamics Laboratory, USA
14	GISS-E2-H	NASA Goddard Institute for Space Studies, USA
15	GISS-E2-R	NASA Goddard Institute for Space Studies, USA
16	HadGEM2-AO	National Institute of Meteorological Research/Korea Meteorological Administration, Korea
17	HadGEM2-CC	Met Office Hadley Centre, UK
18	HadGEM2-ES	Met Office Hadley Centre (additional HadGEM2-ES realizations contributed by Instituto Nacional de Pesquisas Espaciais), UK
19	IPSL-CM5A-LR	Institut Pierre-Simon Laplace, France
20	IPSL-CM5A-MR	Institut Pierre-Simon Laplace, France
21	IPSL-CM5B-LR	Institut Pierre-Simon Laplace, France
22	MIROC5	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology, Japan
23	MIROC-ESM	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies, Japan
24	MIROC-ESM-CHEM	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies, Japan
25	MPI-ESM-LR	Max Planck Institute for Meteorology, Germany
26	MPI-ESM-MR	Max Planck Institute for Meteorology, Germany
27	MRI-CGCM3	Meteorological Research Institute, Japan
28	NorESM1-M	Norwegian Climate Centre, Norway
29	NorESM1-ME	Norwegian Climate Centre, Norway

reanalysis (grid box by grid box) and then projected onto the loading patterns of the PCAs from NCEP-NCAR (from Section 2.2). Thus, the bias of the modelled atmospheric variables in comparison to the reanalysis data is represented in the resulting projected PC scores of the climate models. These biases are used to determine the weights for each investigated CMIP model. Therefore, the appropriate key predictor PC scores are used to drive the final MLRs that were defined by the reanalysis and observed data in Section 2.3. As the projected PCs reflect the biases of the original modelled atmospheric fields (through the normalization with the reanalysis values), the resulting mean of the CMIP precipitation time series differs from the original mean of the fitted values from NCEP-NCAR according to the bias of the input key predictors. The absolute differences between the means of the two time series are used as basis for the model weights. Large

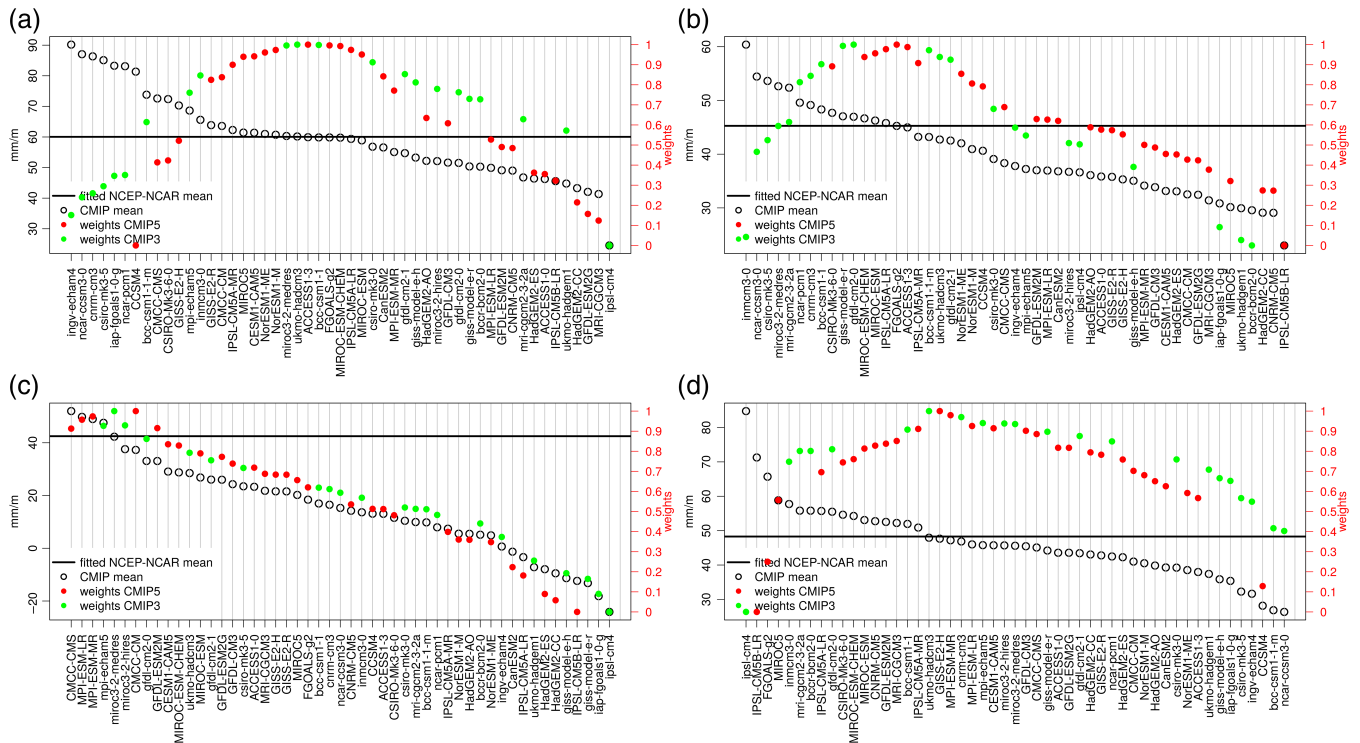
differences of the means get lower weights than smaller differences.

This is done separately for CMIP3 and CMIP5. The basic weights are then normalized between 0 and 1. As a result, each model has one weight per sub-region and season.

Figure 3 shows the mean output values from the MLRs and the corresponding weights for the example of the *Eastern Black Sea* (8) sub-region.

#### 4 | APPLICATION OF MODEL WEIGHTS TO MEDITERRANEAN DOWNSCALED PRECIPITATION CHANGES

The application of the weights on the model data leads to a modified distribution of precipitation change signals which



**FIGURE 3** Mean CMIP output of precipitation sums from the MLRs, which were set up with the reanalysis, and corresponding model weights for sub-region 8 in the historical period (1950–1999) in (a) DJF, (b) MAM, (c) JJA and (d) SON. Model weights reflect the absolute difference between modelled mean values and the mean of NCEP-NCAR, normalized between 0 and 1

are potentially more reliable than the unweighted changes, as the better performing models on current climate have higher impact on the mean change signal than the worse performing models. Figures 4 and 5 show the relative precipitation changes at the end of the 21st century (2070–2099) compared to the historical period (1950–1999) for the weighted multi-model means (wMMM). The results for all sub-regions and seasons that have valid MLRs are shown for the four future scenario runs from the CMIP ensembles (A1B, A2, rcp4.5, rcp8.5). In winter, precipitation is projected to increase almost in the whole study area except for the *Eastern Mediterranean* (4). The largest increase (up to 70%) is expected to be at the *Balkans* (6), followed by the *Eastern Black Sea* (8) and the *Iberian Peninsula* (5). The increasing precipitation amounts persist in spring and summer at the northeastern parts of the investigated area: In spring positive changes occur at the *Balkans* (6) and the *Eastern Black Sea* (8), in summer slightly positive values persist only in the latter region. All other sub-regions are projected to have decreased precipitation amounts at the end of the 21st century from spring to autumn. Particularly the most southern part with the *Iberian Peninsula* (5), the *Maghreb* (7) and the *Eastern Mediterranean* (4) are subject to the strongest precipitation decreases with up to 70%. The four different scenario runs thereby show a consistent change pattern.

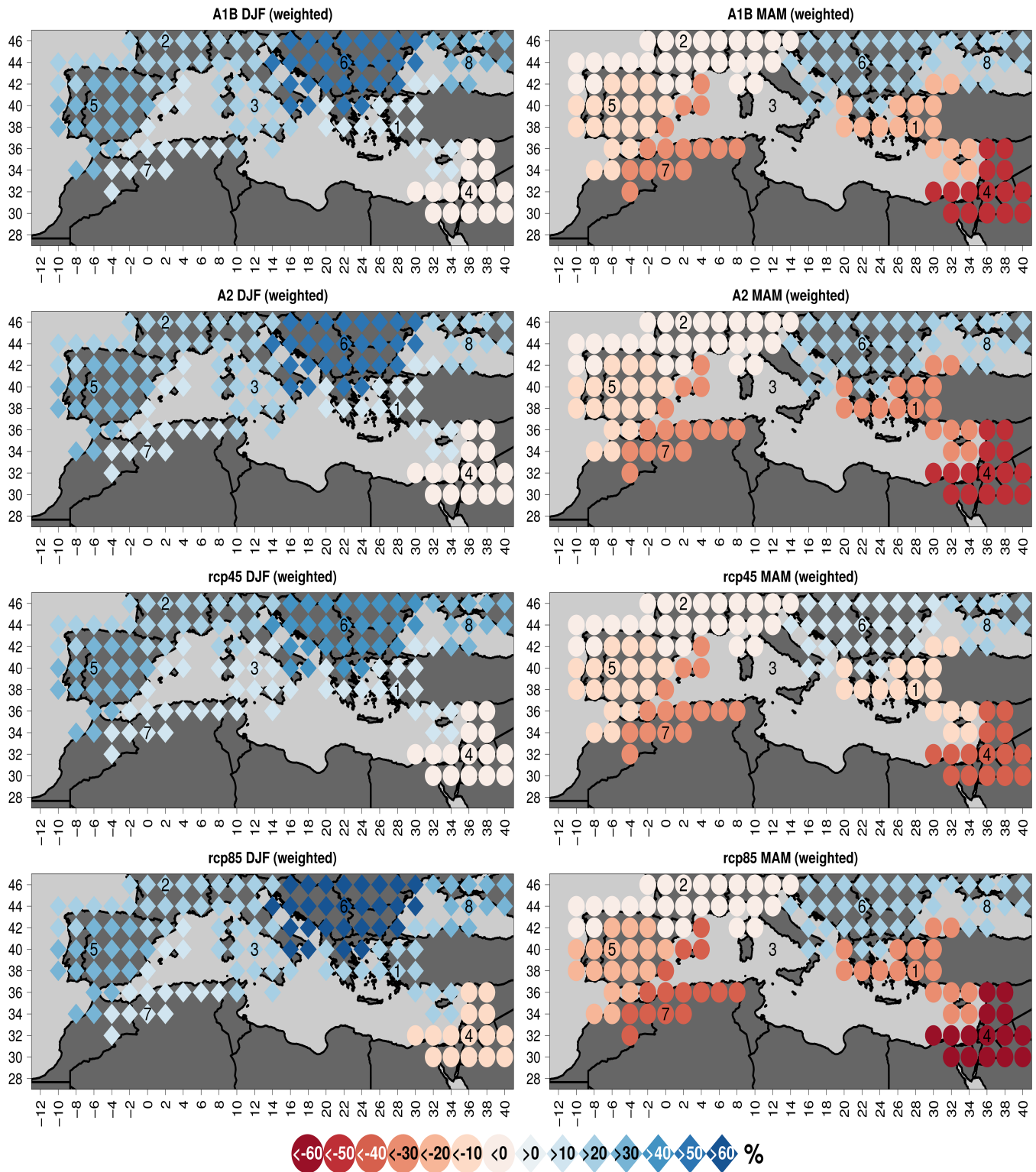
The difference between unweighted and weighted precipitation changes are shown in Figure 6 for ensemble

means and the changes of the kernel density estimates within the model ensemble exemplarily for the *Eastern Black Sea* (8) in winter in Figure 7. These figures illustrate on the one hand the diversity and spread of the climate model ensemble and on the other hand the impact of the weights on the ensemble member distribution. For the *Eastern Black Sea* (8) sub-region in winter (see Figures 6 and 7) and autumn (see Figure 6d) the weighting of CMIP3 leads to more positive projections, and the weighting of CMIP5 leads to reduced positive change signals, thus, the precipitation scenarios become overall more precise. The same effect of the weighting can also be seen in the *Greece-Turkey area* (sub-region 1) in winter (see Figure 6a).

In the *Eastern Mediterranean* (4) area, the weighting process intensifies the negative precipitation changes in winter uniformly among the scenarios and the positive precipitation changes in the sub-regions *Tyrrhenian Sea riparians* (3) and the *Balkans* (6) are shifted towards diminished and intensified change signals, respectively. However, the absolute effect of the weights is small compared to the variability of the ensemble and the weights do not cause visibly more precise change signals for all sub-regions in winter, for example, at the *Iberian Peninsula* (5).

For the other seasons, there are also uniform shifts of the change signals over all four scenarios (in MAM in all sub-regions, in JJA in sub-regions 1, 2, 6 and 8, and in SON in sub-regions 3 and 7) and a reduction of the spread



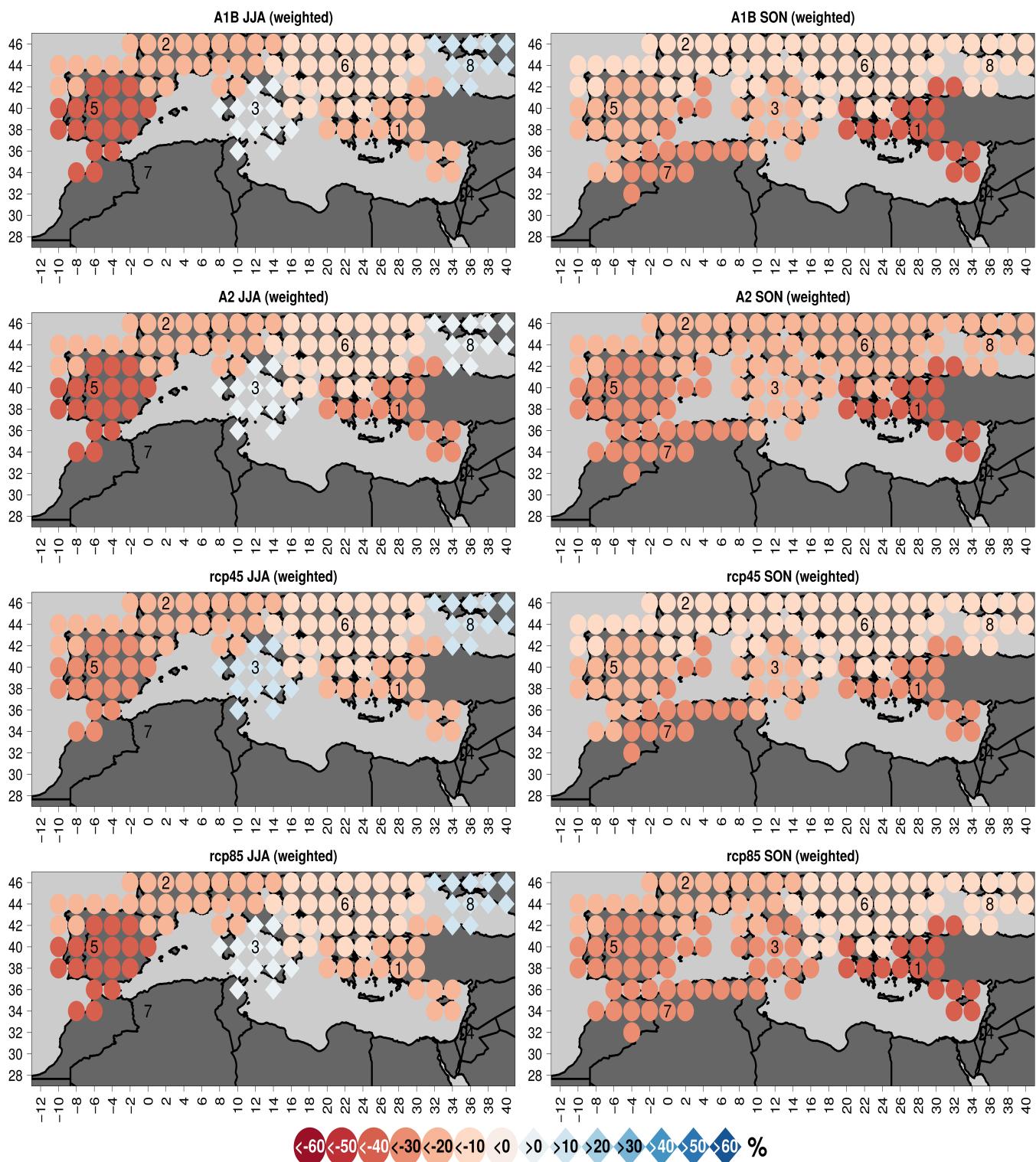


**FIGURE 4** Downscaled and weighted precipitation changes in % from 1950–1999 to 2070–2099 for the scenarios A1B, A2, rcp4.5 and rcp8.5 in DJF and MAM

and thus a more precise signal of the future precipitation change (in JJA in sub-region 5, and in SON in sub-regions 2 and 8).

Characteristics of the modification of the change signals' distributions for all scenarios and seasons can be seen in Figure 8. The change of the mean on the x-axis shows the direction of diminished or intensified precipitation changes

after weighting the climate models, whereas the change of the standard deviations (SD) on the y-axis indicates a broadening or narrowing of the inter-model ensembles' spread. The changes of the ensemble spreads in winter and autumn are quite uniformly distributed, however, increasing ensemble ranges dominate in spring and decreasing ensemble ranges in summer.

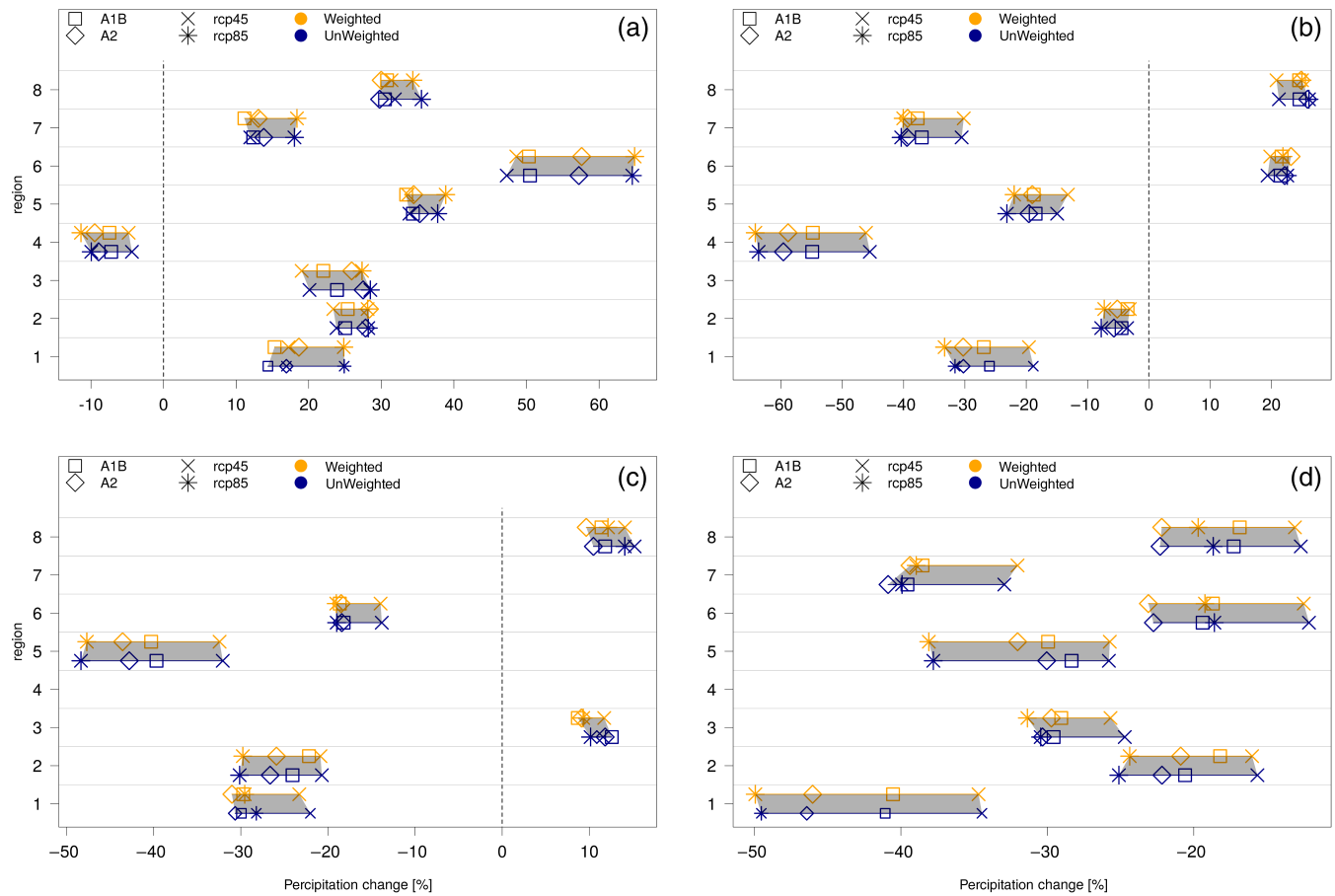


**FIGURE 5** Downscaled and weighted precipitation changes in % from 1950–1999 to 2070–2099 for the scenarios A1B, A2, rcp4.5 and rcp8.5 in JJA and SON

Considering the distributions of the single scenarios and the changes of the wMMM in Figures 6 and 8, the intensified negative precipitation change signal of the *Eastern Mediterranean* (4) sub-region in winter, the diminished positive changes at the *Tyrrhenian Sea riparians* (3) in winter and the confirmation of the negative change signals at the *Iberian Peninsula* (5) and the *Balkans* (6) in

summer become even more reliable through a smaller spread of the single scenario ensembles combined with the accordance or convergence of the scenario wMMM change signals.

Overall, compared to the ensemble spread, the impact of the weighting on the ensemble is rather small and coincidental changes cannot be ruled out.

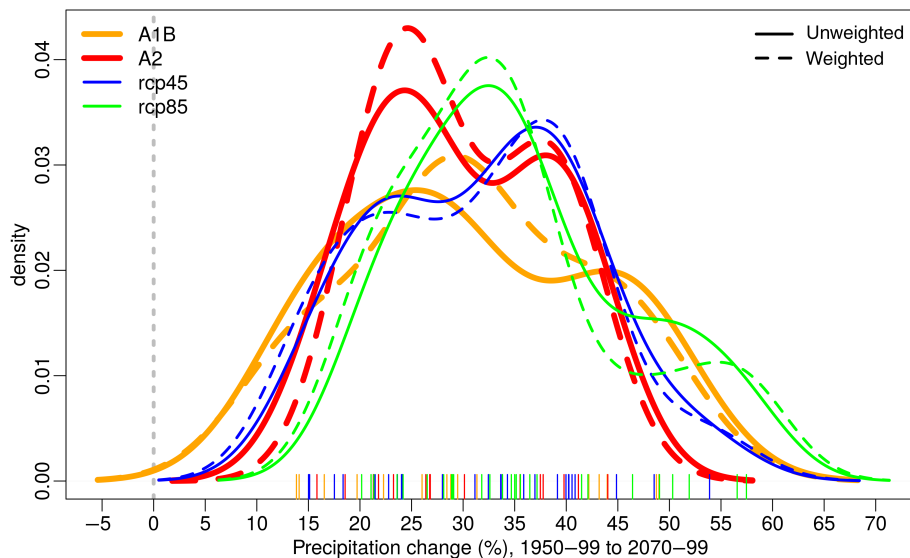


**FIGURE 6** Downscaled mean precipitation changes from 1950–1999 to 2070–2099 with (orange) and without (blue) weights for the scenarios A1B, A2, rcp4.5 and rcp8.5 in % in (a) DJF, (b) MAM, (c) JJA and (d) SON

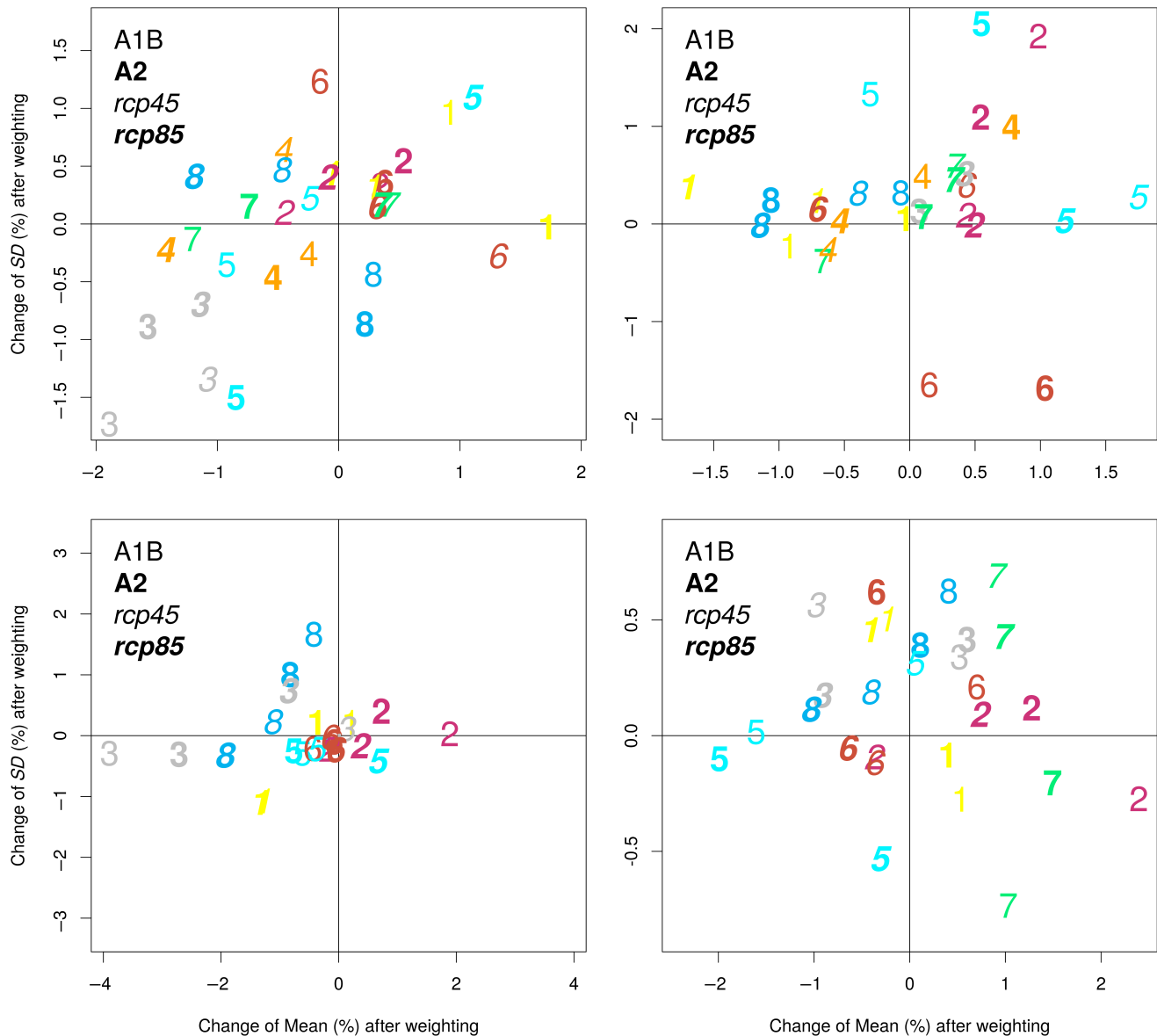
### 5 | CONTINUANCE OF MODEL WEIGHTS IN TIME AND SPACE

The presented weighting metric results in eight (sub-regions) by four (seasons) different assessments of model performance for the entire Mediterranean area with respect to their

skill in representing Mediterranean precipitation. In view of the multitude of weighting results, the question arises whether these weights are comparable among different seasons and sub-regions or whether the single model rankings are rather unique. Therefore, ranking orders are set up to prove the general applicability of the weighting results.



**FIGURE 7** Kernel density estimates of unweighted and weighted precipitation changes in % from 1950–1999 to 2070–2099 for the four scenarios A1B, A2, rcp4.5 and rcp8.5 at the Balkan sub-region (8) in winter



**FIGURE 8** Change of the multi-model mean and the multi-model SD after the weighting procedure in % for the four scenarios A1B (non-italic, normal), A2 (non-italic, bold), rcp4.5 (italic, normal) and rcp8.5 (italic, bold), separated by the four seasons (a) DJF, (b) MAM, (c) JJA and (d) SON. The numbers denote the sub-regions (see Figure 2)

Figures 9 and 10 show ranking lists, aggregated over all sub-regions (left panels) and over all seasons (right panels) for CMIP3 and CMIP5, respectively.

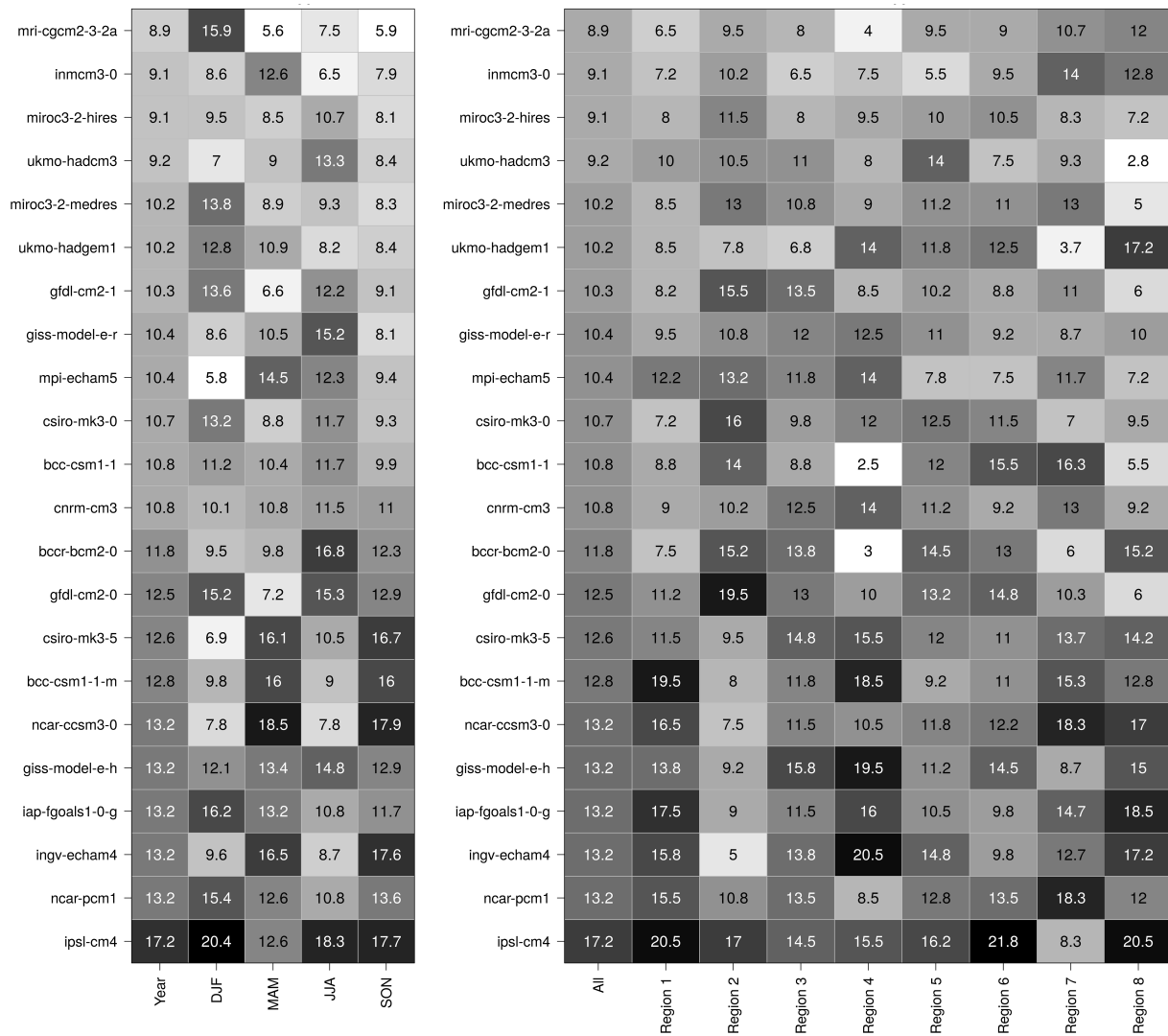
The tables can be read as follows: For example, in the CMIP5 ensemble of Figure 10 the ACCESS1-0 model is the overall best one in representing precipitation in the Mediterranean area, because it has the lowest overall rank number, averaged among all sub-regions and seasons. However, its good performance is not equally distributed over all seasons and sub-regions, as it shows weaknesses in winter and spring and especially in the sub-region of the *Eastern Mediterranean* (4). The second best CMIP5 model CESM1-CAM5 has very good ratings in winter, spring and summer seasons, but the performance decreases in autumn.

Additionally, some model family specific features can be detected. For example, the all in all well rated HadGEM2

model family with three ensemble members in CMIP5 is distinguished by pronounced deficiencies in the *Eastern Black Sea* (8) sub-region as well as the previous model ukmo-hadgem1 from CMIP3, whereas the even older ukmo-hadcm3 model is characterized by the best ranking result for this sub-region in CMIP3.

Another distinctive feature shows the French CMIP5 model family of IPSL-CM5. The IPSL-CM5A models, which are the direct followers of the worst ranked CMIP3 model ipsl-cm4, are noticeable through their below-average rankings in sub-regions 4 and 7, whereas IPSL-CM5B with a new atmospheric model and very different physical parameterizations achieves clearly better rankings in these two dry and southernmost Mediterranean areas.

Looking at the lower ends of both tables, apparently some models show very good performance for the *North-western*



**FIGURE 9** Ranking lists of all used CMIP3 models, resulting from the weighting metric used in this study. Single ranking results are aggregated over all sub-regions (left panel) and over all seasons (right panel). Bright background colour indicates better ranking results than a dark background. The numbers denote the overall ranking result

area (sub-region 2) in comparison to the other sub-regions. Particularly affected are the CanESM2, CCSM4, GISS-E2-H and GISS-E2-R models from CMIP5 and the bcc-csm1-1-m, ncar-ccsm3, giss-model-e-h, iap-fgoals1-0-g, ingv-echam4 and the ncar-pcm1 models from CMIP3. This may be due to the chosen key predictors for this sub-region in summer and/or autumn for the calculation of the MLR (not shown). The by far most important key predictors for the MLRs of these two seasons in the *North-western area* (sub-region 2) are moisture variables (primarily relative humidity at 700 hPa). Usually, geopotential heights play the major role within the set of key predictors and thus the weighting of the single models. However, in this case the named models may profit from an improved moisture modelling.

Despite the variability within the ranking of one model in time and space, sub-groups of better and worse models for the representation of Mediterranean precipitation are clearly recognizable, indicated by the range of the overall averaged ranks at the first columns in Figures 9 and 10. In

the CMIP3 ensemble, mri-cgcm2-3-2a, inmcm3-0 and miroc3-2-hires and for the CMIP5 ensemble, ACCESS1-0, CESM1-CAM5 and MIROC-ESM show the overall best results, whereas ingv-echam4, ncar-pcm1 and ipsl-cm4 in CMIP3 and GISS-E2, MRI-CGCM3 and IPSL-CM5B-LR from CMIP5 are the most inappropriate models to investigate monthly precipitation in the Mediterranean area.

The varying skill of models between different regions and seasons, but nonetheless an obvious overall ranking was also recognized by former studies (Gleckler *et al.*, 2008; Sánchez *et al.*, 2009; Kjellström *et al.*, 2010; Masson and Knutti, 2011; Räisänen and Ylhäisi, 2012).

## 6 | SUMMARY AND CONCLUSIONS

We developed a new weighting approach for general circulation and earth system models based on large and medium scale atmospheric predictor information for the

ACCESS1-0	12.1	15.5	15.5	5.7	10	ACCESS1-0	12.1	12.8	10.2	8.8	18.5	11.8	15.2	6	15.5
CESM1-CAM5	12.5	9.9	12.2	10.8	17.1	CESM1-CAM5	12.5	18.2	13.8	14	4	14.8	6.5	15.3	9.8
MIROC-ESM	12.8	11.5	12.5	12.7	14.6	MIROC-ESM	12.8	12	16.5	14	9.5	9.5	9.8	24	8
CSIRO-Mk3-6-0	13	9.2	17.1	13.5	12.3	CSIRO-Mk3-6-0	13	8.8	15	14.2	13	11	8.8	18.3	16.5
HadGEM2-ES	13.2	17.2	14.8	9.5	9.9	HadGEM2-ES	13.2	13	8.5	10	13	13	18	5.7	22.2
MPI-ESM-LR	13.2	10.6	13.4	15	14.3	MPI-ESM-LR	13.2	7.2	18.5	18.2	10.5	16	11	13.7	9
MPI-ESM-MR	13.3	9.9	13.9	16.5	13.7	MPI-ESM-MR	13.3	9	19.5	16.8	11	13.8	10.5	16	9.2
HadGEM2-CC	13.4	17.9	13.6	10	11	HadGEM2-CC	13.4	10.2	11.5	7.5	17	16.2	17.2	3	23.8
HadGEM2-AO	13.5	15.5	13.6	9.5	14.4	HadGEM2-AO	13.5	16.5	8.5	13	12.5	12	15	10.3	18.8
IPSL-CM5A-MR	13.6	18.8	13.4	12.5	8.9	IPSL-CM5A-MR	13.6	14	15.8	12.5	17.5	6.8	12.5	23.7	10.5
CMCC-CMS	13.6	11.4	13	16.7	14.3	CMCC-CMS	13.6	12.2	18.8	15.8	12	11.8	7.5	20.7	11.2
MIROC-ESM-CHEM	13.8	11.5	13.6	14.3	16	MIROC-ESM-CHEM	13.8	9.5	18.5	15.5	9	12	14	24	7.8
GFDL-CM3	13.9	19.6	13.5	12.5	8.9	GFDL-CM3	13.9	7	17.5	17.8	17	13.2	13.2	13.7	13
CMCC-CM	13.9	10.6	15.6	12.5	16.9	CMCC-CM	13.9	17.2	16	13.5	17.5	11.2	10	13.7	13.8
NorESM1-M	13.9	8.8	16.1	14.7	16.7	NorESM1-M	13.9	17	14.2	10.2	6	19	11	16.3	14.2
NorESM1-ME	14	10	13.9	14.7	18.3	NorESM1-ME	14	19.5	12.2	10.2	1.5	19.8	11.5	16.3	15.5
ACCESS1-3	14.2	11.1	13.5	11.7	20.7	ACCESS1-3	14.2	15.5	7.5	13.2	13.5	21.5	16.2	14	11.8
MIROC5	14.5	9.5	14.2	21.3	14.6	MIROC5	14.5	7.2	20.8	10.8	9.5	16.5	19.2	9	19
CNRM-CM5	15.7	18.9	13	20	11.3	CNRM-CM5	15.7	13.5	17.8	13.5	18.5	16.2	15	12.7	18.8
FGOALS-g2	16.4	9.9	12.8	20	25	FGOALS-g2	16.4	11.2	24	16.5	19.5	15.8	22	11	11.5
IPSL-CM5A-LR	16.6	20.6	11.8	21	13.6	IPSL-CM5A-LR	16.6	18	19.5	14.5	17.5	15.8	16.5	18.3	13.2
GFDL-ESM2G	16.8	23.1	16	14	12.7	GFDL-ESM2G	16.8	10.8	20.2	20	15.5	17.5	17	13.7	18
GFDL-ESM2M	16.8	20.5	14.5	17.2	15	GFDL-ESM2M	16.8	16	17.2	19.5	20.5	16	22	13.3	11
CanESM2	16.9	15.4	19.5	14.8	17.4	CanESM2	16.9	22	9.5	13.8	28	13	17.8	19	18.2
CCSM4	17.3	11.6	20	12.7	24.7	CCSM4	17.3	23	8	18.8	20	18.2	9.5	22.3	21.2
GISS-E2-H	17.5	20	20.2	17	11.9	GISS-E2-H	17.5	22.8	7.8	22	25.5	15.8	20	17.7	12.5
GISS-E2-R	18.1	20.6	18.6	16.5	16.1	GISS-E2-R	18.1	24	8.2	23.2	27.5	13	19.5	20	14.8
MRI-CGCM3	18.6	22.1	18.5	19.7	13.7	MRI-CGCM3	18.6	23.5	16.5	15.5	15	22.2	21.2	13.3	18.2
IPSL-CM5B-LR	22.1	23.8	16.6	28.2	21.1	IPSL-CM5B-LR	22.1	23.2	23	21.8	15	21.8	27.2	10	28
	Year	DJF	MAM	JJA	SON		All	Region 1	Region 2	Region 3	Region 4	Region 5	Region 6	Region 7	Region 8

FIGURE 10 As Figure 9, but for CMIP5

Mediterranean precipitation within a statistical downscaling framework. The spread of modelled precipitation sums and also the spread of projected precipitation changes at this hot spot of climate change underline the need of a particular performance evaluation of state-of-the-art general circulation models (here all available models of CMIP3 and CMIP5) to gain more trustworthy ensembles.

The novelty of this weighting metric consists in avoiding the use of the precipitation bias by itself as a weighting basis, as the modelling of precipitation amounts and their spatial distribution is still a highly insufficient subject. Therefore, significant and robust predictor-precipitation-relationships are derived by means of particular MLRs. An extensive and elaborated predictor screening with a

bootstrapping technique was developed to ensure almost time-independent regression models and their applicability to the future under the assumption of stationary biases. Because our study avoids the use of direct CMIP precipitation output, the potential bias of these data and its behaviour in future scenarios have no influence on our results. However, the biases of our predictor variables may underlie a time dependent magnitude as described by Christensen and Boberg (2012) for temperature. In what sense potential changing biases of the large scale atmospheric drivers might influence the downscaled predictand through the MLR equation is a very interesting question and should be subject of further research.

In our study, the biases in the historical period of the identified key predictors, including pressure, moisture, wind

and temperature proxy variables at different atmospheric levels, determine the specific climate model weights. This is done separately for the four seasons and eight Mediterranean sub-regions, which are characterized by homogeneous precipitation variabilities.



The application of the model weights to the downscaled precipitation changes at the end of the 21st century for four different climate scenarios (A1B, A2, rcp4.5 and rcp8.5) leads in some cases to either a shift or a concretion of the change signals and thus to more reliable results. However, at a few sub-regions and seasons, the weighting resulted in a broadening of the change signal distribution. Overall, the changes are due to the chosen type of weights (normalization between 0 and 1) very small.

Finally, the 8 (sub-regions) by 4 (seasons) different model rankings were analysed with respect to their continuance in space and time. Thereby it is not possible to gain one homogeneous and universal ranking of models over all sub-regions and seasons, however, a clear sequence from better to worse models for the representation of precipitation in the Mediterranean area becomes apparent.

## ACKNOWLEDGEMENTS

We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling groups. We also acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (<http://ensembles-eu.metoffice.com>) and the data providers in the ECA&D project (<http://www.ecad.eu>). Furthermore, NCEP Reanalysis data were provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at <http://www.esrl.noaa.gov/psd/>. Financial support was provided by the DFG (German Research Foundation) within the project COMEPRO (Comparison of Metrics for Probabilistic Climate Change Projections of Mediterranean Precipitation).

## ORCID

Irena Kaspar-Ott  <https://orcid.org/0000-0002-5151-201X>  
Elke Hertig  <https://orcid.org/0000-0002-6934-9468>

## REFERENCES

- Boberg, F., Berg, P., Thejll, P., Gutowski, W.J. and Christensen, J.H. (2009) Improved confidence in climate change projections of precipitation evaluated using daily statistics from the PRUDENCE ensemble. *Climate Dynamics*, 32, 1097–1106. <https://doi.org/10.1007/s00382-008-0446-y>.
- Christensen, J.H. and Boberg, F. (2012) Temperature dependent climate projection deficiencies in CMIP5 models. *Geophysical Research Letters*, 40(10), 2307–2308. <https://doi.org/10.1029/2012GL053650>.
- Christensen, J.H., Kjellström, E., Giorgi, F., Lenderink, G. and Rummukainen, M. (2010) Weight assignment in regional climate models. *Climate Research*, 44, 179–194.
- Enke, W. and Spekat, A. (1997) Downscaling climate model outputs into local and regional weather elements by classification and regression. *Climate Research*, 8, 195–207.
- Giorgi, F. (2006) Climate change hot-spots. *Geophysical Research Letters*, 33, 1–4. <https://doi.org/10.1029/2006GL025734>.
- Giorgi, F. and Mearns, L. (2002) Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the “reliability ensemble averaging” (REA) method. *Journal of Climate*, 15, 1141–1158. [https://doi.org/10.1175/1520-0442\(2002\)015<1141:COAURA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1141:COAURA>2.0.CO;2).
- Gleckler, P.J., Taylor, K.E. and Doutriaux, C. (2008) Performance metrics for climate models. *Journal of Geophysical Research*, 113, D06104. <https://doi.org/10.1029/2007JD008972>.
- Haylock, M.R., Hofstra, N., Klein Tank, A.M.G., Klok, E.J., Jones, P.D. and New, M. (2008) A European daily high-resolution gridded dataset of surface temperature and precipitation. *Journal of Geophysical Research: Atmospheres*, 113, D20119. <https://doi.org/10.1029/2008JD10201>.
- Hertig, E. and Jacobeit, J. (2013) A novel approach to statistical downscaling considering nonstationarities: application to daily precipitation in the Mediterranean area. *Journal of Geophysical Research: Atmospheres*, 118, 520–533. <https://doi.org/10.1002/jgrd.50112>.
- Hertig, E. and Jacobeit, J. (2014) Considering observed and future nonstationarities in statistical downscaling of Mediterranean precipitation. *Theoretical and Applied Climatology*, 122, 667–683. <https://doi.org/10.1007/s00704-014-1314-9>.
- Hertig, E., Seubert, S., Paxian, A., Vogt, G., Paeth, H. and Jacobeit, J. (2014) Statistical modelling of extreme precipitation indices for the Mediterranean area under future climate change. *International Journal of Climatology*, 34, 1132–1156. <https://doi.org/10.1002/joc.3751>.
- Jacobeit, J. (1993) Regionale Unterschiede im atmosphärischen Zirkulationsgeschehen bei globalen Klimaveränderungen. *Die Erde*, 124(1), 63–77.
- Jacobeit, J., Hertig, E., Seubert, S. and Lutz, K. (2014) Statistical downscaling for climate change projections in the Mediterranean region: methods and results. *Regional Environmental Change*, 14(5), 1891–1906.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K., Ropelewski, C., Wang, J., Jenne, R. and Joseph, D. (1996) The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77, 437–471. [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2).
- Kjellström, E., Boberg, F., Castro, M., Christensen, H.J., Nikulin, G. and Sánchez, E. (2010) Daily and monthly temperature and precipitation statistics as performance indicators for regional climate models. *Climate Research*, 44, 135–150.
- Knutti, R., Sedláček, J., Sanderson, B., Lorenz, R., Fischer, E.M. and Eyring, V. (2017) A climate model projection weighting scheme accounting for performance and interdependence. *Geophysical Research Letters*, 44, 1909–1918. <https://doi.org/10.1002/2016GL072012>.
- Lutz, K., Jacobeit, J., Philipp, A., Seubert, S., Kunstmann, H. and Laux, P. (2012) Comparison and evaluation of statistical downscaling techniques for station-based precipitation in the Middle East. *International Journal of Climatology*, 32, 1579–1595. <https://doi.org/10.1002/joc.2381>.
- Macadam, I., Pitman, A.J., Whetton, P.H. and Abramowitz, G. (2010) Ranking climate models by performance using actual values and anomalies: implications for climate change impact assessments. *Geophysical Research Letters*, 37, L16704. <https://doi.org/10.1029/2010GL043877>.
- Masson, D. and Knutti, R. (2011) Spatial-scale dependence of climate model performance in the CMIP3 ensemble. *Journal of Climate*, 24(11), 2680–2692. <https://doi.org/10.1175/2011JCLI3513.1>.
- Meehl, G.A., Covey, C., Delworth, T., Latif, M., McAvaney, B., Mitchell, J.F.B., Stouffer, R.J. and Taylor, K.E. (2007) The WCRP CMIP3 multi-model dataset: a new era in climate change research. *Bulletin of the American Meteorological Society*, 88, 1383–1394.
- Mueller, B. and Seneviratne, S.I. (2014) Systematic land climate and evapotranspiration biases in CMIP5 simulations. *Geophysical Research Letters*, 41, 128–134. <https://doi.org/10.1002/2013GL058055>.
- Palatella, L., Miglietta, M.M., Paradisi, P. and Lionello, P. (2010) Climate change assessment for Mediterranean agricultural areas by statistical downscaling. *Natural Hazards and Earth System Sciences*, 10, 1647–1661. <https://doi.org/10.5194/nhess-10-1647-2010>.
- Philipp, A. (2003). *Zirkulationsdynamische Telekonnektivität des Sommerwinterschlags im südhemisphärischen Afrika*. PhD Thesis, Bayerische Julius-Maximilians-Universität Würzburg, [https://opus.bibliothek.uni-wuerzburg.de/files/686/Philipp\\_A.\\_-\\_Dissertation.pdf](https://opus.bibliothek.uni-wuerzburg.de/files/686/Philipp_A._-_Dissertation.pdf).

- Räisänen, J. and Ylhäisi, J. (2012) Can model weighting improve probabilistic projections of climate change? *Climate Dynamics*, 39, 1981–1998.
- Räisänen, J., Ruokolainen, L. and Ylhäisi, J. (2010) Weighting of model results for improving best estimates of climate change. *Climate Dynamics*, 35, 407–422.
- Reifen, C. and Toumi, R. (2009) Climate projections: past performance no guarantee of future skill? *Geophysical Research Letters*, 36, L13704. <https://doi.org/10.1029/2009GL038082>.
- Sánchez, E., Romera, R., Gaertner, M.A., Gallardo, C. and Castro, M. (2009) A weighting proposal for an ensemble of regional climate models over Europe driven by 1961–2000 ERA40 based on monthly precipitation probability density functions. *Atmospheric Science Letters*, 10, 241–248. <https://doi.org/10.1002/asl.230>.
- Tatli, H., Nüzhet Dalfes, H. and Sibel, M.Ş. (2004) A statistical downscaling method for monthly total precipitation over Turkey. *International Journal of Climatology*, 24, 161–180. <https://doi.org/10.1002/joc.997>.
- Taylor, K.E., Stouffer, R.J. and Meehl, G.A. (2012) An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93, 485–498.
- Tebaldi, C., Smith, R.L., Nychka, D. and Mearns, L.O. (2005) Quantifying uncertainty in projections of regional climate change: a Bayesian approach to the analysis of multimodel ensembles. *Journal of Climate*, 18, 1524–1540.
- Trigo, R. and Palutikof, J. (2001) Precipitation scenarios over Iberia: a comparison between direct GCM output and different downscaling techniques. *Journal of Climate*, 14, 4422–4446. [https://doi.org/10.1175/1520-0442\(2001\)014<4422:PSOIAAC>2.0.CO;2](https://doi.org/10.1175/1520-0442(2001)014<4422:PSOIAAC>2.0.CO;2).
- Weigel, A.P., Knutti, R., Liniger, M.A. and Appenzeller, C. (2010) Risks of model weighting in multimodel climate projections. *Journal of Climate*, 23, 4175–4191.
- Whetton, P., Macadam, I., Batholds, I. and O'Grady, J. (2007) Assessment of the use of current climate patterns to evaluate regional enhanced greenhouse response patterns of climate models. *Geophysical Research Letters*, 34, L14701. <https://doi.org/10.1029/2007GL030025>.
- Xoplaki, E., González-Rouco, J.F., Luterbacher, J. and Wanner, H. (2004) Wet season Mediterranean precipitation variability: influence of large-scale dynamics and trends. *Climate Dynamics*, 23, 63–78. <https://doi.org/10.1007/s00382-004-0422-0>.

**How to cite this article:** Kaspar-Ott I, Hertig E, Kaspar S, *et al.* Weights for general circulation models from CMIP3/CMIP5 in a statistical downscaling framework and the impact on future Mediterranean precipitation. *Int J Climatol.* 2019;39: 3639–3654. <https://doi.org/10.1002/joc.6045>