



Weighted multi-model ensemble projection of extreme precipitation in the Mediterranean region using statistical downscaling

Luzia Keupp^{1,2} · Elke Hertig² · Irena Kaspar-Ott² · Felix Pollinger¹ · Christoph Ring¹ · Heiko Paeth¹ · Jucundus Jacobeit²

Received: 20 February 2018 / Accepted: 25 March 2019 / Published online: 22 May 2019
© Springer-Verlag GmbH Austria, part of Springer Nature 2019

Abstract

Projections of seasonal extreme precipitation changes in eight Mediterranean subregions between the end of the twentieth and the end of the twenty-first century are analyzed using weighted multi-model ensembles. Weights are based on the performance of predictor variables in the scope of statistical downscaling. Two indices of precipitation scarcity as well as two indices of heavy precipitation are downscaled from global climate model data of the Coupled Model Intercomparison Project phase 3 and 5 (CMIP3, CMIP5) multi-model ensembles, considering two emission scenarios each. Based on the performance with regard to observations of extreme precipitation as well as inter-model consistency, three weighting metrics are calculated and subsequently applied to each ensemble. While meteorological droughts are projected to increase in most cases, the tendency is less pronounced for heavy precipitation events and mostly points towards reduction. The weighting does not affect the multi-model mean changes, but induces a decrease of ensemble spread (although mostly not significant), implying a decrease of model uncertainty. As the ensemble and scenario considered have minor effect on the findings and also the differences between seasons and subregions are not marked, there is strong evidence for enhanced droughts in the Mediterranean region, implying major socio-economic and ecological consequences.

Keywords Climate change · Model weighting · Precipitation extremes · Mediterranean · Statistical downscaling · GLM · CMIP3 · CMIP5

1 Introduction

Although the number of general circulation models (GCMs) and their complexity is generally increasing, no perfect model exists, as different aspects of each model lead to shortcomings due to lack in process knowledge and resolvability: initial state and boundary conditions, parameterization as well as structural deficiencies (e.g., Knutti et al. 2010a; Tebaldi and Knutti 2007, Reichler et al. 2008). To compensate for this, a preferably large multi-model ensemble (MME) should be taken into account when analyzing climate projections as it comprises the

most realistic extent of uncertainty possible. This holds true despite interdependencies among the models and the MME being an ensemble of opportunity (non-systematic choice) (Collins 2007; Collins et al. 2013; Knutti et al. 2010a; Tebaldi and Knutti 2007). Encouraging this approach is that multi-model means (MMM) commonly yield improvements regarding reproducing observations when compared to single model results (Gleckler et al. 2008; Flato et al. 2013; Knutti et al. 2010a; Reichler et al. 2008; Randall et al. 2007; Sillmann et al. 2013a; Weigel et al. 2010). In this paper, the term model uncertainty is defined as the spread between different model simulations (cp. e.g., Hawkins and Sutton 2011), which can be described by the standard deviation of the MME results.

Besides the widespread practice of equal weighting, weights can be used in order to accentuate models performing better with respect to the studied topic. Although climate models are established in a way to fit past and present-day climate best way possible and a transferability to the future is indeed debatable, comparisons

✉ Luzia Keupp
luzia.keupp@uni-wuerzburg.de

¹ Institute of Geography and Geology, University of Würzburg, Würzburg, Germany

² Institute of Geography, University of Augsburg, Augsburg, Germany

with observations are a common way to evaluate them and construct performance metrics (Tebaldi and Knutti 2007; Knutti et al. 2010b; Reichler et al. 2008; Giorgi and Mearns 2002; Ring et al. 2016; Ring et al. 2017). It has to be noted that weighting metrics should be chosen with care (e.g., Weigel et al. 2010).

The present study uses a novel approach to evaluate model skill and to develop respective weights: the weighted MME is based on the performance of predictor variables within perfect prog (PP) statistical downscaling for the assessment of extreme precipitation in the Mediterranean area. The rationale behind this is that GCMs are not able to simulate regional precipitation (extremes) in a satisfying quality yet (e.g., Fowler et al. 2007; Mueller and Seneviratne 2014; Maraun et al. 2010; Sillmann et al. 2013a, b; Sun et al. 2006; Trigo et al. 2001), due to, for example, lacks in the representation of uplift mechanisms (Trenberth et al. 2003; Sun et al. 2006; Stephens et al. 2010) or small-scale surface heterogeneities not being resolved (e.g., Reichler et al. 2008; Flato et al. 2013). Thus, instead of using GCM precipitation, we estimate it from well-simulated atmospheric variables via statistical downscaling (Trigo et al. 2001; Sillmann et al. 2013b). Statistical downscaling is computationally inexpensive and has been shown to yield improvements when compared to GCM outputs.

In contrast to dynamical downscaling, where a higher resolution regional climate model is nested within a GCM (Fowler et al. 2007; Maraun et al. 2010), in statistical downscaling, local climate is estimated using empirically assessed relationships to large-scale climate (Flato et al. 2013; Fowler et al. 2007; Maraun et al. 2010; Maraun and Widmann 2018; von Storch et al. 1993). Three groups of statistical downscaling methods can be distinguished: In perfect prognosis (PP) downscaling, the calibration of the statistical model is performed using observational (or reanalysis) data of large-scale predictors and local-scale predictands. The large-scale (free-)atmospheric predictors are assumed to be “perfectly” simulated by the GCM, justifying the application on model data (Maraun et al. 2010; Maraun and Widmann 2018; von Storch et al. 1993; von Storch 1995). Model output statistics (MOS) use a predictor-predictand link established from simulated predictors and observed predictands. Therefore, the statistical model can only be applied to the climate model which it was calibrated with, but systematic model biases can be automatically corrected (depending on the specific method) (Maraun et al. 2010; Maraun and Widmann 2018; von Storch 1995). As no temporal correspondence between simulations and observations exists, only distribution-wise calibration using the same variable as predictor and predictand is possible for MOS in the context of climate modeling (Maraun and Widmann 2018). Weather generators, stochastic models

generating random time series with statistical characteristics mimicking those of observed climate, form the third group (Maraun et al. 2010; Maraun and Widmann 2018).

In the perfect-predictor experiment of the COST action VALUE project (Maraun et al. 2015), statistically downscaled daily precipitation using GLM-based perfect prognosis methods was found to show good results regarding spells as well as inter-annual variability and the annual cycle. Also concerning the mean of relative wet-day frequency and mean wet-day precipitation, downscaling improvements were shown (Gutiérrez et al. 2018; Hertig et al. 2018; Maraun et al. 2017). However, the quality of the simulated results always depends on the knowledge and care of the user, in particular the choice of predictors and whether the GCM simulates the predictors realistically.

The Mediterranean region is chosen as study area as Diffenbaugh and Giorgi (2012) encouraged the study of climate response in this region being a hot spot of climate change. This is particularly important with regard to extremes as they exert profound socio-economic and ecological impact. Water availability and its temporal distribution is critical for the future development of the Mediterranean area (e.g., Bolle 2003). With dry events on the one side, contributing to water scarcity, vegetation stress, wildland fires, and erosion as well as high precipitation extremes on the other side, causing floods and erosion, this topic is highly relevant. Additionally, climate change often becomes visible particularly in modifications of extreme event features (e.g., Flato et al. 2013; Trenberth et al. 2007).

In Hertig et al. (2012), statistically and dynamically downscaled precipitation extremes in the Mediterranean region for autumn and winter performed similar, markedly exceeding the direct GCM output. Also, Lavaysse et al. (2012) found for eight stations in the French Mediterranean area, that PP-based statistically downscaled GCM results generally match rainfall observations including extremes better than direct ones. In a study, regarding precipitation extremes in the Mediterranean, regional climate models were shown to have major deficiencies (Cornes et al. 2013).

As concluded in a previous study, future precipitation variability over the Mediterranean area can be ascribed in large parts to model uncertainty, particularly over land areas. By correcting the single model results with respect to observational data in the late twentieth century, the inter-model spread was found to be reduced (Paeth et al. 2017). Also, other studies found a decrease in the model spread by applying performance metrics (e.g., Duan and Phillips 2010; Flato et al. 2013; Knutti et al. 2017; Räisänen and Ylhäisi 2012). Generally, the attribution of inter-model differences to uncertainties in precipitation extreme changes is higher than the one of differing scenarios (Seneviratne et al. 2012), making reasonable weighting an important contribution to decrease projection uncertainty. This is also aimed here.

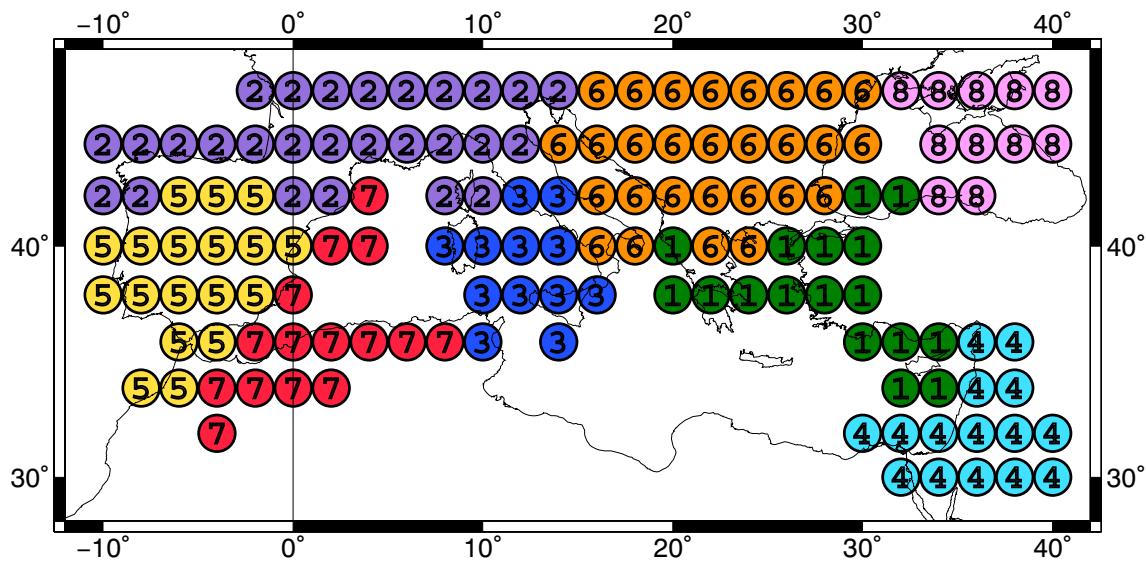


Fig. 1 Study region and subregions: Aegean (1), North Atlantic (2), Tyrrhenian Sea riparians (3), Near East (4), Iberian Peninsula (5), Balkans (6), Maghreb (7), Eastern Black Sea (8)

We therefore follow the advice of Hertig et al. (2014) to use MMEs for deriving downscaling performance-based weights in order to gain more reliable results.

In the present contribution, the studied phenomenon is extreme precipitation and its change until the end of the twenty-first century in the Mediterranean region, implying precipitation intensity, number of threshold shortfalls, number of quantile exceedances, rainfall amount from these quantile exceedances, as well as maximum length of dry spells. Therefore, four indices are used for describing and studying the phenomenon, all downscaled from atmospheric predictors, meaning that no GCM precipitation output is used in this study. The performance measures are the coefficient of determination (for considering the

downscaling skill) as well as three weighting metrics based on the bias, the bias considering natural variability and the latter combined with single model performance relative to average performance (for assessing the skill of each GCM to simulate input for regional climate change projection). Finally, the performance-based weighting metrics are applied to the change in index values from the end of the twentieth to the end of the twenty-first century and the results are examined.

Limitations arise from the stationarity assumption regarding the predictor-predictand relationships (which is tried to be accounted for using a bootstrapping approach implying random sampling) as well as the choice of reference data.

Table 1 Considered CMIP3 models (the spelling in particular regarding the case, is adopted from the datasets)

Model name	Research unit
bccr_bcm2	Bjerknes Centre for Climate Research, Norway
ccsm3	National Center for Atmospheric Research, USA
cnrm_cm3	Météo-France / Centre National de Recherches Météorologiques, France
csiro_mk3_5	CSIRO Atmospheric Research, Australia
gfdl_cm2_0	US Dept. of Commerce / NOAA / Geophysical Fluid Dynamics Laboratory, USA
gfdl_cm2_1	US Dept. of Commerce / NOAA / Geophysical Fluid Dynamics Laboratory, USA
giss1 (e)	NASA / Goddard Institute for Space Studies, USA
ingv_echam4	Instituto Nazionale di Geofisica e Vulcanologia, Italy
inm_cm3_0	Institute for Numerical Mathematics, Russia
ipsl_cm4	Institute Pierre Simon Laplace, France
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
mpi_echam5	Max Planck Institute for Meteorology, Germany
mri_cgcm2_3_2a	Meteorological Research Institute, Japan

Table 2 Considered CMIP5 models (the spelling in particular regarding the case, is adopted from the datasets)

Model name	Research unit
ACCESS1-0	Commonwealth Scientific and Industrial Research Organization and Bureau of Meteorology, Australia
ACCESS1-3	
CanESM2	Canadian Centre for Climate Modelling and Analysis, Canada
CMCC-CM	Centro Euro-Mediterraneo per i Cambiamenti Climatici , Italy
CMCC-CMS	
CNRM-CM5	Centre National de Recherches Météorologiques / Centre Européen de Recherche et Formation Avancée en Calcul Scientifique, France
GFDL-CM3	
GFDL-ESM2G	NOAA Geophysical Fluid Dynamics Laboratory, USA
GFDL-ESM2M	
GISS-E2-H	NASA Goddard Institute for Space Studies, USA
GISS-E2-R	
HadGEM2-AO	National Institute of Meteorological Research / Korea Meteorological Administration, Korea
HadGEM2-CC	
HadGEM2-ES	Met Office Hadley Centre (additional HadGEM2-ES realizations contributed by Instituto Nacional de Pesquisas Espaciais), UK
inmcm4	
IPSL-CM5A-LR	Institute for Numerical Mathematics, Russia
IPSL-CM5A-MR	
IPSL-CM5B-LR	Institute Pierre Simon Laplace, France
MIROC5	
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
MIROC-ESM-CHEM	
MPI-ESM-LR	Max Planck Institute for Meteorology (MPI-M), Germany
MPI-ESM-MR	
MRI-CGCM3	Meteorological Research Institute, Japan
NorESM1-M	
NorESM1-ME	Norwegian Climate Centre, Norway

The article is structured as follows: Section 2 presents the region of interest as well as the observational respectively reanalysis and model data employed. In the subsequent chapter, the methodology is outlined: the calculation of the predictands used for assessing extreme precipitation (3.1), the selection and preparation of atmospheric predictors for simulating precipitation extremes (3.2), the model construction for the statistical downscaling (3.3), the simulation itself (3.4), and finally the quantification of model performance using metrics (3.5). The results are described in Section 4 regarding GLM model construction (4.1), weights and ranks (4.2), extreme precipitation change in the MMM (4.3.1), distribution (4.3.2), and the effect of weighting (4.3.3). Rounding off, a discussion of the results (Section 5) as well as a summary of the study and concluding remarks (Section 6) are given.

2 Study area and data

2.1 Study area

The Mediterranean region (12° W–40° E, 28–46° N) is divided into subregions based on similarity in precipitation variability independent of the time of the year. This is carried out using a s-mode varimax-rotated principal component analysis (PCA) of annual precipitation sums followed by a differentially initialized *k*-means (DKM) cluster analysis.

Eight subregions are delineated by this procedure which are shown in Fig. 1: Aegean (1), North Atlantic (2), Tyrrhenian Sea riparians (3), Middle East (4), Iberian Peninsula (5), Balkans (6), Maghreb (7), and Eastern Black Sea (8).

Table 3 Number of dominant PCs and their cumulated explained variance

Variable	Abbreviation	Winter		Spring		Summer		Autumn	
		PCs	EV	PCs	EV	PCs	EV	PCs	EV
Relative humidity at 700 hPa	hur700	9	84	10	84	9	83	11	87
Relative humidity at 850 hPa	hur850	8	83	6	87	9	87	8	88
Specific humidity at 700 hPa	hus700	6	83	8	89	9	79	8	90
Specific humidity at 850 hPa	hus850	7	86	7	89	8	79	7	88
Meridional wind component at 700 hPa	ua700	8	89	10	88	9	81	11	88
Meridional wind component at 850 hPa	ua850	7	85	9	84	8	76	10	82
Zonal wind component at 700 hPa	va700	11	87	9	84	11	79	9	80
Zonal wind component at 850 hPa	va850	10	86	9	80	11	76	8	79
Geopotential height at 500 hPa	zg500	10	93	8	94	8	88	6	92
Geopotential height at 700 hPa	zg700	8	92	8	94	9	89	9	94
Sea level pressure	psl	7	92	8	91	8	87	8	91
Sum/mean		91	87	92	88	99	82	95	87

2.2 Data

For constructing the downscaling model, daily E-OBS precipitation data (version 14, $0.25^\circ \times 0.25^\circ$, Haylock et al. 2008) as well as monthly NCEP-NCAR reanalysis data ($2.5^\circ \times 2.5^\circ$, Kalnay et al. 1996) of large-scale atmospheric variables (see Section 3.2) are used.

GCM predictor data of the Coupled Model Intercomparison Project phase 3 and phase 5 (CMIP3 Meehl et al. 2007, CMIP5 Taylor et al. 2012) serve as input for the simulations. For each, the historical experiment (CMIP3 20c3m) as well as two scenarios of twenty-first climate are used (A1B, A2 Moss et al. 2008 and RCP4.5, RCP8.5 van Vuuren et al. 2011, respectively). By doing so, intermediate as well as peak pathways are covered for both ensembles. The first run (CMIP5 r1i1p1) of those models is selected for which all predictor variables are available for both scenarios (see Section 3.2), resulting in 13 models for CMIP3 (Table 1) and 26 for CMIP5 (Table 2). No direct precipitation output of GCMs is used.

All data is interpolated onto a regular $2^\circ \times 2^\circ$ grid. For precipitation, the spatial mean time series of the eight subregions is used.

3 Methodology

3.1 Predictands: indices of extreme precipitation

Extreme events are defined by the Intergovernmental Panel on Climate Change (IPCC) Special Report on Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaption (SREX, IPCC 2012a, b; Seneviratne et al. 2012) as “the occurrence of a value of a weather or climate

variable above (or below) a threshold near the upper (or lower) ends of the range of observed values of the variable.”

We examine them with regard to precipitation in the Mediterranean using four indices (cp. Nicholls and Murray 1999; Klein Tank et al. 2009; Hertig et al. 2014; Seneviratne et al. 2012). Moderately heavy precipitation, which is defined as exceedance of the 95th percentile, is considered with respect to number of days (R95n) and amount (R95am). The percentile calculation is based on the reference period 1961–1990 with only events of more than 1 mm being taken into account. As typical precipitation amounts and distributions are different for each region, percentiles are the adequate way to examine deviations from base period climate in the given context (Klein Tank et al. 2009). The low precipitation extreme is investigated using the number of dry days (R1mmn) as well as the maximum number of consecutive dry days (MCD), with dryness being defined as a scarcity of precipitation with 1 mm at maximum per day. These events can also be referred to as meteorological droughts (c.p. IPCC 2012a). All indices are calculated per single-monthly season (winter/DJF = December, January, February, etc.) using 1950–1999 E-OBS data (for the first DJF, January values are repeated).

3.2 Predictors: selection and preparation

The predictors should be chosen with care in a way that the physical background of extreme precipitation is considered. Mediterranean precipitation is caused by the interplay of dynamic and thermodynamic processes. Therefore, large-scale atmospheric circulation as well as smaller-scale thermodynamic activity are taken into account.

The first stimulus is represented by geopotential height (zg at 500 and 700 hPa), meridional and zonal component

of wind velocity (ua , va at 700 and 850 hPa), as well as sea level pressure (psl) in the Northeastern Atlantic domain (50° W– 50° E, 24 – 66° N). These variables (zg , ua , and va in different layers) are commonly used for statistical downscaling of mean precipitation (e.g., Xoplaki et al. 2004; Tatli et al. 2004; Lutz et al. 2012) and extremes (e.g., Cavazos and Hewitson 2005; Haylock and Goodess 2004; Hertig et al. 2014; Hertig and Trambly 2017; Merckenschlager et al. 2017; Monjo et al. 2016; Nuissier et al. 2011; Toreti et al. 2010; Vrac and Yiou 2010) in the Mediterranean region. The size of the domain allows the depiction of the most important synoptic-scale weather systems that affect the Mediterranean precipitation like the North Atlantic Oscillation, the Scandinavian pattern, the East Atlantic pattern, and the East Atlantic/West Russia pattern.

Relative and specific humidity (hur , hus at 700, and 850 hPa) act as thermodynamic indicators inside the study area (12° W– 40° E, 28 – 46° N). The high relevance of including humidity for downscaling of extreme precipitation was pointed out by, e.g., Cavazos and Hewitson (2005) and Hertig et al. (2014). The incorporation of other physically meaningful variables like the vertical wind component is refrained from due to insufficiencies in simulation.

In preparation for the downscaling, the NCEP-NCAR reanalysis data of each of the named variables for 1950–1999 is condensed by means of a s-mode varimax-rotated PCA using correlation matrix and latitudinal weighting for each season. In order to choose the adequate number of PCs to use, the dominance criteria after Jacobeit (1993) are applied. Additionally, the loading of the respective PC needs to exceed the one of the next higher PC by more than one standard deviation and the dominance has to be present for at least eight grid cells (cp. Philipp et al. 2007). Finally, the explained variance of the PC has to cover no less than 4% of the overall variance (cp. Kaspar-Ott et al. 2019).

The PC scores of each variable thus represent the time component of the most important modes of variability in the respective variable over the analyzed domain and are therefore used as predictor time series. Table 3 shows the resulting number of dominant PCs and their explained variance.

3.3 Statistical downscaling

Since the chosen indices are neither continuous nor normally distributed, generalized linear models (GLMs) are used in order to simulate extreme precipitation indices based on the predictor time series. As depicted in Eq. 1, for GLMs, the expectation value E of the predictand Y_t at time step t is equivalent to the mean μ_t of the distribution of

Y at time t , which is the inverse of the link function g applied to the linear predictor η_t (the linear combination of the predictors/covariates, i.e., systematic component). Finally, η_t is the sum of the values of the j th predictors for observation t , x_{tj} , multiplied by the respective parameter β_j over all p covariates (McCullagh and Nelder 1989).

$$E(Y_t) = \mu_t = g^{-1}(\eta_t) = g^{-1} \sum_{j=1}^p x_{tj} \beta_j \quad (1)$$

For the count variables R95n, R1mmn, and MCD, a log-linear model is used, implying multiplicative systematic effects and Poisson-distributed errors (variance function $V(\mu) = \mu$, canonical link $\eta_t = \log(\mu_t)$). Classical linear models on the contrary are suitable for continuous data with constant error variances and have a normal error distribution as well as additive systematic effects ($V(\mu) = 1$, $\eta_t = \mu_t$). The application of the log-linear model assures positive model results even for negative predictor values, as μ_t and therefore $E(Y_t)$ is derived by applying the inverse of the link function to η_t , which is the exponential function (McCullagh and Nelder 1989).

R95am is defined in case of non-zero R95n only (an amount of precipitation over the 95th percentile can only occur in seasons with at least one precipitation event passing the percentile) and hence has to be a non-zero positive value. Therefore, all zero values of R95am are removed before conducting the downscaling. For the simulations, information on zero values is obtained from rounded R95n simulation values. Thus, a gamma distribution is used for modeling the non-negative continuous R95am data ($V(\mu) = \mu^2$). Although the canonical link function for this GLM family is reciprocal, a log link ($\eta_t = \log(\mu_t)$) and consequently a multiplicative exponential model is used, as it “is usually more adequate for both modeling and interpretation” (Fahrmeir et al. 2013, cp. Hertig et al. 2014).

As the variance function $V(\mu)$ and the link function g are specified by the model family used (see above), the model is constructed by means of maximum likelihood estimation of the parameters β_j .

The selection of key predictors and construction of the final GLM is performed based on Hertig et al. (2014) and Kaspar-Ott et al. (2019), considering the possibility for collinearity and a temporally unsteady predictor-predictand relationship:

For each region, season, and index, 100 bootstrap iterations (drawing without replacement) of calibration are carried out using two thirds of the data, the remaining third being deployed for validation. With this approach, possible non-stationarities in the relationship between atmospheric variables and precipitation extremes are taken into account,

Table 4 Number of key predictors and determination coefficient of the final GLMs

	DJF		MAM		JJA		SON	
	r^2	n	r^2	n	r^2	n	r^2	n
R95n								
reg1	0.46	15	0.41	21	–	–	0.44	20
reg2	0.45	16	0.48	20	0.41	24	0.48	21
reg3	0.31	16	0.37	24	–	–	0.49	22
reg4	0.41	14	0.48	20	–	–	0.66	20
reg5	0.59	14	0.51	23	0.49	18	0.60	21
reg6	0.37	15	0.35	26	0.36	27	0.37	21
reg7	0.53	15	0.46	22	–	–	0.44	21
reg8	0.33	15	0.33	27	–	–	0.59	21
R95am								
reg1	0.65	12	0.61	15	(0.95)	(8)	0.41	12
reg2	0.43	15	0.51	19	0.41	21	0.40	19
reg3	0.38	14	0.59	20	(0.94)	(10)	0.74	17
reg4	0.56	14	0.89	6	–	–	0.79	6
reg5	0.81	13	0.41	14	0.85	9	0.60	12
reg6	0.28	6	0.54	22	0.60	22	0.46	17
reg7	0.55	12	0.72	13	(0.83)	(8)	0.59	16
reg8	0.46	15	–	–	(0.67)	(23)	0.76	16
R1mmn								
reg1	0.79	21	0.75	29	0.65	26	0.80	27
reg2	0.59	20	0.77	26	0.65	26	0.77	26
reg3	0.69	21	0.75	29	0.64	28	0.75	29
reg4	0.64	21	0.73	29	0.77	27	0.70	28
reg5	0.68	20	0.75	29	0.71	26	0.72	27
reg6	0.80	21	0.76	29	0.77	24	0.79	25
reg7	0.49	21	0.72	28	0.73	29	0.74	28
reg8	0.71	21	0.62	28	0.70	27	0.69	28
MCD								
reg1	0.66	20	0.57	27	0.53	25	0.51	25
reg2	0.48	20	0.61	26	0.48	25	0.64	27
reg3	0.59	19	0.49	27	0.52	26	0.61	25
reg4	0.50	20	0.69	28	0.36	25	0.72	26
reg5	0.49	19	0.55	27	0.46	25	0.60	26
reg6	0.44	20	0.73	27	0.67	27	0.71	26
reg7	0.32	20	0.48	27	0.44	25	0.63	26
reg8	0.56	21	0.50	28	0.58	26	0.55	27

R95n and R95am number of days with respectively amount of precipitation above the 95th percentile of 1961–1990, R1mmn and MCD number of days respectively maximum number of consecutive days with a maximum precipitation of 1 mm. For situations where no transfer function for R95n is found, R95am values are shown in brackets, as R95n is needed for the calculation of R95am from model output

the necessity for this having been emphasized by several studies (e.g., Hertig and Jacobeit 2015; Merkenchlager et al. 2017). In a first step, GLMs are constructed leaving out one predictor at a time, followed by the calculation

of the mean-squared error (MSE) for all combinations of $n - 1$ predictors. Next (step 2), the correlation coefficients between the predictors are computed. In case of absolute correlations of 0.5 or higher, the predictor having lower

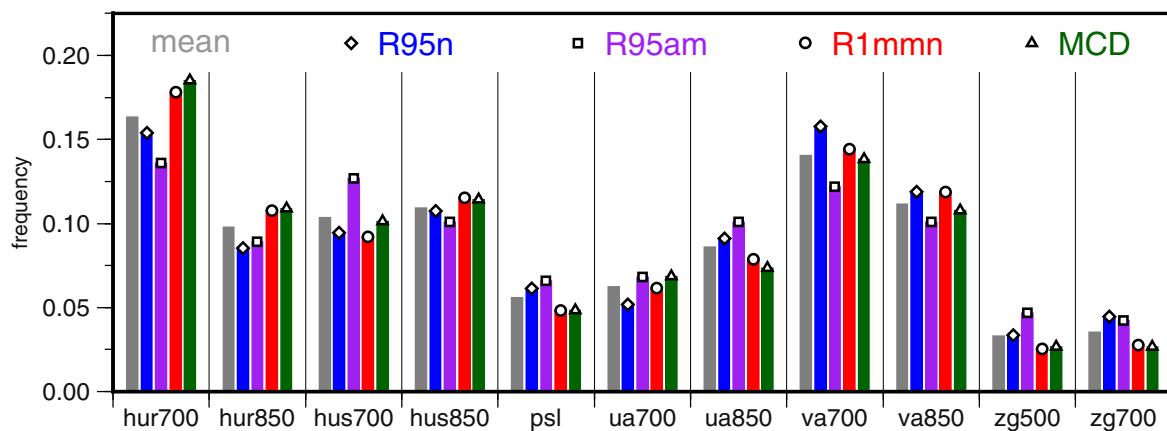


Fig. 2 Relative frequency of predictor variables. Indices: R95n and R95am number of days with respectively amount of precipitation above the 95th percentile of 1961–1990, R1mmn and MCD number of

days respectively maximum number of consecutive days with a maximum precipitation of 1 mm. For abbreviation of predictor variables, see Table 3 or Section 3.1

importance for GLM quality is discarded. This is the one for which the omission in step 1 leads to the lower MSE value. In case of failure in step 1 (all estimated models having equal MSE due to overfitting), the GLMs are built with single predictors (instead of all minus one at a time), implicating that those predictors with higher MSE values are rejected in step 2. Afterwards (step 3), the GLM generation procedure is repeated as in step 1 but only with the remaining predictors. The last construction phase (step 4) involves the construction of GLMs consisting of five up to all remaining predictors (the most important ones as deduced from MSEs of step 3) and the one model with the highest coefficient of determination (r^2) is used for the particular bootstrap iteration. For each iteration, r^2 and mean-squared skill score (MSSS, Eq. 2, e.g., Wilks 2011) for the calibration sample as well as the projection using the validation sample determine the skill of the respective iteration. If one of the MSSSs is non-positive, a zero skill is ascribed.

$$\begin{aligned} \text{MSSS} &= \frac{\text{MSE} - \text{MSE}_{\text{ref}}}{\text{MSE}_{\text{perf}} - \text{MSE}_{\text{ref}}} = 1 - \frac{\text{MSE}}{\text{MSE}_{\text{ref}}} \\ &= 1 - \frac{\text{MSE}}{\frac{nt-1}{nt} \cdot s_{\text{ref}}} \end{aligned} \quad (2)$$

(MSE MSE of simulation, MSE_{ref} and s_{ref} MSE resp. variance of reference sample (E-OBS predictand time series), MSE_{perf} MSE of perfect forecast (= 0), nt number of timesteps)

After finishing the bootstrapping, the 25% best models are used, as they represent the upper quartile of the model quality range. From these, the average optimal number (AON) of predictors—the arithmetic mean of

number of predictors from the 25% best models for each bootstrap iteration—is derived and the AON most frequent predictors are assigned as the final key predictors. In case of equivocality, predictors with minimal minimum significance level are employed.

3.4 Simulation of extreme precipitation from GCM predictor data

With the key predictors determined, the GCM data can be employed for simulating extreme precipitation. Single-monthly seasonal data of each variable at each grid cell is z-transformed using mean and standard deviation of the reanalysis data in the respective situation in order to allow for bias calculation (see below). As for the NCEP-NCAR variables, a latitudinal-weighted 2D matrix is constructed. However, it does not serve as PCA input but is projected onto the loading patterns of the reanalysis PCs. From these time series, extreme precipitation is simulated, pursuing two objectives.

First, the deviation between simulations based on reanalysis and GCM data has to be estimated. For this purpose, the output of the afore-described procedure is used to drive the respective final GLM without further adaption and for the longest time period available (1950–1999). The second target is analyzing the change in precipitation extremes from the end of the twentieth to the end of the twenty-first century (subsequently also named delta). In order to remove the direct reference to the reanalysis data and enable the calculation of a delta, data for the periods 1970–1999 as well as 2070–2099 (two scenarios) are thus standardized conjointly, before applying the GLM. However, this also implies that historical values depend on future ones in a way that they should not be considered in

an isolated manner. As mentioned before, R95am is only computed for situations with R95n values being non-zero.

3.5 Simulation quality and model weight generation

Various performance metrics have been used so far, based on, e.g., 2×2 contingency table approaches (applied, e.g., by Ring et al. 2017), probability density functions (e.g., Kjellström et al. 2010), Bayesian statistics (e.g., Tebaldi et al. 2004), or reliability ensemble averaging (Giorgi and Mearns 2002).

In order to evaluate simulation quality of the GCMs and derive model weights, three metrics are applied using the period 1950–1999 (cp. Section 3.4). All weights are calculated separately for the two CMIP ensembles, each index and situation (region and season). As already noted in Section 2.2, no direct GCM precipitation output is used, the weighting is based on precipitation indices as downscaled from large-scale predictors.

The most basic measure is the absolute difference of the means (ADM) of two corresponding time series (Eq. 3, time series x of reanalysis and time series y_i of model i).

$$\text{ADM}_i = |\bar{x} - \bar{y}_i| \quad (3)$$

In combination with the natural variability (ϵ), ADM can be transformed to the model performance component of the reliability ensemble averaging (REA) technique by Giorgi and Mearns (2002) (Eq. 4). Natural variability is estimated from reanalysis-based simulated extreme precipitation indices, using the range of floating 50a-averages spanning 1948–2016. Although this estimation method was adapted from Giorgi and Mearns (2002) and the largest possible time span was used, the contribution of forced vs. internal variability is not clear. Divided by raw ADM weights, with 1 (raw $\text{ADM} \leq$ natural variability) as fixed maximum value, the raw REA bias weights (subsequently called REA) result. Owing to this definition, ADM and REA differ only for situations where the bias of at least one model does not exceed natural variability.

$$\text{REA}_i = \frac{\epsilon}{\text{ADM}_i} (\text{max. } 1) \quad (4)$$

The model performance criterion (REA) can be extended by model convergence (D_i), implicating the distance of the change simulated using a particular model (Δ_{model}) from the REA multi-model mean (MMM) of the delta ($\Delta_{\text{MMM}_{\text{weighted}}}$). This deviation is iterated with the starting values calculated in the following way: A first distance estimate is calculated as the absolute difference of the model delta and the MMM delta ($\Delta_{\text{model}} - \Delta_{\text{MMM}}$). Afterwards, a first-weighted MMM (wMMM) is compiled using the ratio of the natural variability (see above) and the distance as weight ($\Delta_{\text{model}} - \Delta_{\text{MMM}_{\text{weighted}}}$). This procedure is

repeated using the wMMM and the weighted distance until the resulting weights converge. The product of the two components is named REA-B \times D (Bias \times Distance, Eq. 5). As the divergence from the MMM depends on the future, separate weights eventuate for each scenario.

$$\text{REA-B} \times D_i = \text{REA}_i \cdot \frac{\epsilon}{D_i} \quad (5)$$

The resulting raw weights are standardized in a way that their sum is 1 and the highest weight is ascribed to the model with the lowest deviation. This is done in the following way:

As for ADM, higher raw weights (i.e., higher differences between the model and the reanalysis time series) signify lower model performance, the order of weights is modified first. This is done by the calculation of the ratio of the minimum of all raw weights and each single raw weight ($W_i = \min(W) / W_i$). Raw weights of REA and REA-B \times D are already calculated in a way such that higher raw weights correspond to better model performance, due to constructing a ratio with ϵ as dividend (see Eqs. 4, 5). Afterwards, for all three metrics, the weights for each model are divided by the sum of all model weights ($W_i / \Sigma W$). For the derivation of one metric from the other, raw weights are used.

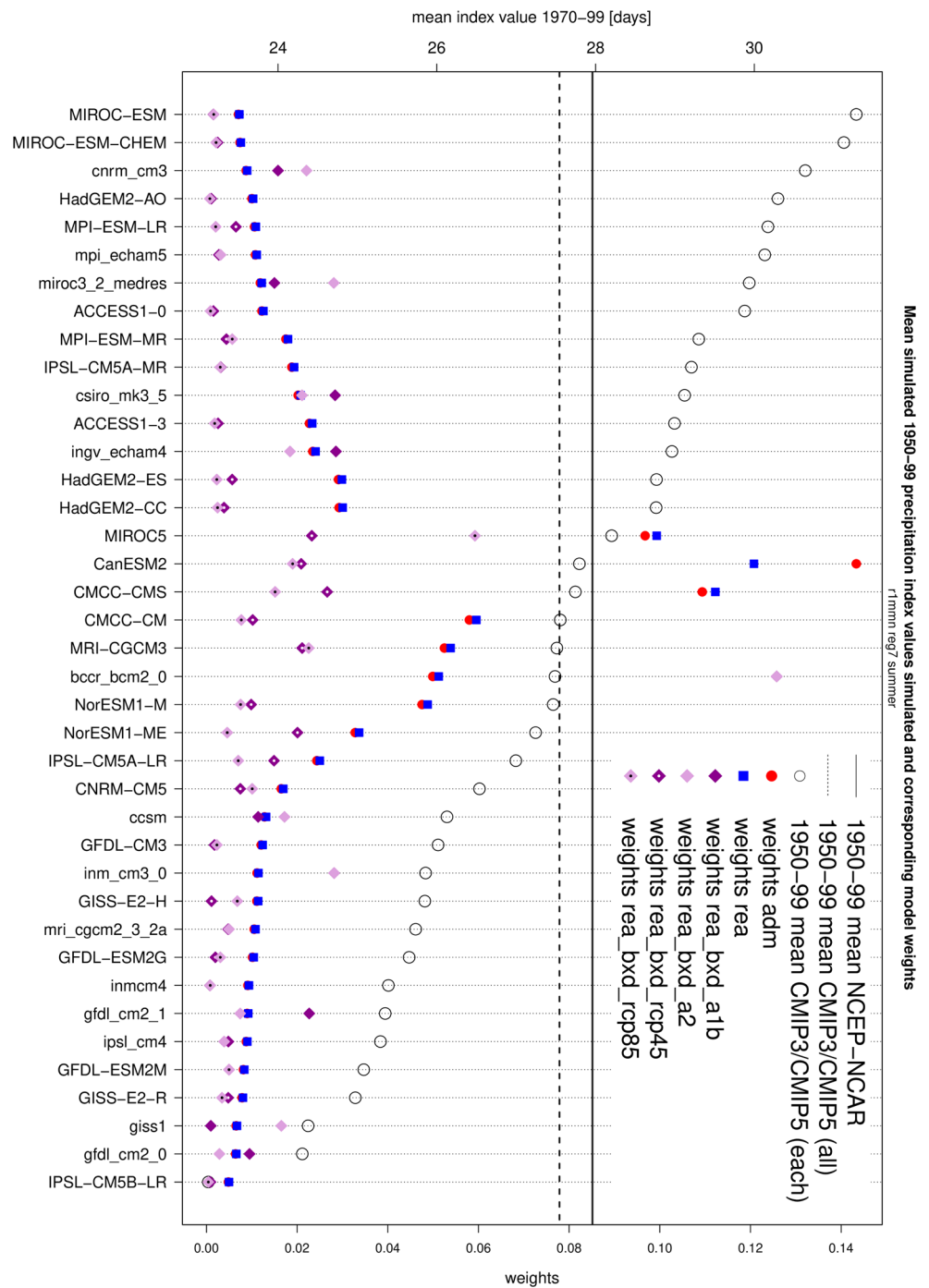
4 Results

4.1 Construction of downscaling models

The construction of GLMs did not result in a valid model for the following situations: R95n JJA regions 1, 3, 4, 7, 8; R95am MAM region 8, JJA region 4. Due to R95am zero values being derived from R95n, non-valid R95n situations inhibit the simulation of R95am as well. Regarding region 4 (Near East), the modeling incapacity (calibration) can be attributed to the lack of high precipitation events in the observation-based time series (3 of 150 months with 1 day each). For the other situations, weak validation results (i.e., zero skill for all iterations due to negative MSSS) lead to the dismissal. High determination coefficients for R95am in situations where no models for R95n result, can be attributed to the low values in these situations which are common in calibration and validation periods. As mentioned above, R95am results of situations without results for R95n are not used. With exception of these cases, precipitation extremes are projected from reanalysis data for each index, region, and season.

The number of key predictors as well as the coefficient of determination of the final GLMs from 100 bootstrap iterations is shown in Table 4, whereas the relative frequency of predictor variables averaged over all regions and seasons is displayed in Fig. 2. Highest coefficients (0.71 on average), but also most predictors (26 on average)

Fig. 3 Mean-simulated 1950–1999 index values and corresponding model weights (combined for CMIP3 and CMIP5), exemplary for R1mmn (number of days with a maximum precipitation of 1 mm), region 7, summer. Weights of each type (ADM absolute difference of the means, REA ADM with inclusion of a natural variability criterion, REA-B×D REA extended using the difference of change between the model and the multi-model mean (differing for each emission scenario), see Section 3.5) sum up to 1



result for R1mmn, lowest for R95n (r^2 0.45) and R95am (14 predictors), respectively. Averaged over all indices, the smallest number of key predictors as well as lowest r^2 are determined for winter. Regarding regions, no clear differentiation is evident.

No single most important predictor variable can be detected, but overall 700 hPa relative (dominating in MAM and SON except for R95n with JJA and SON instead) and specific (JJA) humidity as well as the zonal wind

component (DJF 700 hPa, R95n also MAM 850 hPa, mean 700 hPa) play major roles. Considering all indices, geopotential height is of minor relevance (each level rarely exceeding 5%) with the highest frequencies for R95am (for a discussion of this, see Section 5).

Altogether, relative humidity as well as 850 hPa specific humidity exert more importance for the low than for the high precipitation indices. For the latter, a slightly higher relation to psl and zg is found. R95n shows higher frequencies of

mri_cgcm2_3_2a	5.7	6	6.5	5.1	5.5
mpi_echam5	5.9	5.1	5.5	6.3	6.3
gfdl_cm2_1	6.4	6.8	6.5	6.4	6
csiro_mk3_5	6.4	7.7	5.7	6.5	5.9
inm_cm3_0	6.5	6.9	7.1	5.9	6.4
miroc3_2_medres	6.6	7.3	5.9	6.3	7
bccr_bcm2_0	6.7	7	7	6.5	6.6
cnrm_cm3	7.1	7.4	6.8	6.2	7.8
ingv_echam4	7.3	6.8	7.2	7.4	7.8
gfdl_cm2_0	7.5	7.6	7.4	8.1	6.9
giss1	7.5	7.3	7.2	8.3	7.3
ccsm	7.7	7.2	7.4	7.9	8
ipsl_cm4	8.8	6.9	9.8	9.6	8.9

(a) CMIP3

CMCC-CM	10.8	10.9	11.7	10.5	10.4
CMCC-CMS	11	12.7	10.9	9.9	10.9
MPI-ESM-MR	11.1	12.3	10.4	10.8	10.8
CNRM-CM5	12	10.1	10.3	14.1	13
MIROC5	12	11.8	10.8	12.6	12.6
ACCESS1-0	12.2	12.6	11.4	13.7	10.9
HadGEM2-CC	12.2	10.6	13.5	13.8	10.9
HadGEM2-ES	12.4	11.6	13.9	12.1	12.1
MPI-ESM-LR	12.5	14	11.8	11.9	12.5
NorESM1-M	12.7	14.3	14.4	11.3	11.6
HadGEM2-AO	12.7	12.7	12.5	13.3	12.4
NorESM1-ME	12.9	15.4	15.1	11.1	10.9
GFDL-CM3	13	12.3	13.8	13	12.8
CanESM2	13.3	15.2	14.2	11.3	13.1
inmcm4	13.4	12.4	13.3	13.2	14.4
IPSL-CM5A-MR	14.1	13.9	14	12.7	15.9
MRI-CGCM3	14.2	12.7	13.9	15.5	14.4
GFDL-ESM2G	14.2	13.6	15.7	13.5	14.4
ACCESS1-3	14.7	16.1	15	13.9	13.9
GFDL-ESM2M	14.7	14.6	14.8	15.1	14.3
MIROC-ESM	14.7	16	13.9	13.4	15.6
IPSL-CM5A-LR	14.8	11.5	15.9	16	15.5
MIROC-ESM-CHEM	15	14.9	13.9	14.4	16.6
GISS-E2-R	15.5	15.1	13.4	17.3	15.8
GISS-E2-H	15.6	16	13	16.6	16.6
IPSL-CM5B-LR	17.2	15.7	17.3	18.3	17.2

(b) CMIP5

Fig. 4 Model ranking averaged over all subregions and seasons. Colors are split in intervals of one third (green best 33.3%, red worst 33.3%, yellow in between). Boxes encompass lowest and highest ranks. Indices: R95n and R95am number of days with respectively amount of precipitation above the 95th percentile of 1961–1990, R1mmn and MCD number of days respectively maximum number of consecutive days with a maximum precipitation of 1 mm

va than the other indices, R95am of ua. Regarding the dry extreme, hur700 is chosen particularly often in autumn (26 resp. 29%) and spring (20 resp. 20%), for winter va700 is most common (22 resp. 24%), for summer—the season most balanced regarding predictor frequencies—both hus predictors (15 and 14 resp. 18 and 16%). Concerning the high precipitation extremes, the similarities are less pronounced, with the highest R95n predictor frequencies resulting for va700 in winter (23%), hur700 and hus700 in summer (19%), hur700 in autumn (21%), and va in spring but quite similar frequencies for hur700, hus as well as ua850. For R95am, which generally shows the least differences between the frequencies of the single predictor variables, hus700 stands out in summer (24%), va700 in winter (16%), while for the other seasons hur700 is most often used.

All humidity variables combined make up between 46 and 51% of the predictors in the mean of all situations, with specific humidity being less important than relative humidity. In single-seasonal view, an exception can be found in summer for all indices but R95n (same frequency for R95n and R95am). Most of the remaining predictors are composed of horizontal wind components (39–41% on average, with va occurring more often than ua; regarding single seasons ua is more frequent for R95am summer, and MCD autumn).

Summed up, the predictors representing the large-scale atmospheric (dynamic) as well as those representing the thermodynamic processes of extreme precipitation generation account for important shares of the constructed GLMs. Nevertheless, the dynamic ones slightly outnumber the thermodynamic ones when averaged over all situations and indices. While this holds true for all seasons of R95am, thermodynamic predictors prevail in summer for R95n and R1mmn, as well as in summer and autumn for MCD.

4.2 Model weights and ranks

Mean-simulated R1mmn values based on each CMIP model and NCEP-NCAR reanalysis data as well as the resulting model weights are displayed in Fig. 3 for JJA in the Maghreb region. Note that CMIP3 and CMIP5 models are combined here, implicating differing final weights when compared to a single-ensemble assessment, as the raw weights of both ensembles are standardized together (see Section 3.5). For ADM and REA, those models (open circles) deviating least from the reanalysis index values (solid line) are regarded as the most trustworthy ones and therefore assigned the highest weights. The lowest difference is simulated by CanESM2 (0.16 days), resulting in a final weight of 0.14 for ADM. In contrast, 4.94 less days of low precipitation are modeled using IPSL-CM5B-LR when compared to NCEP-NCAR, the weight thus being

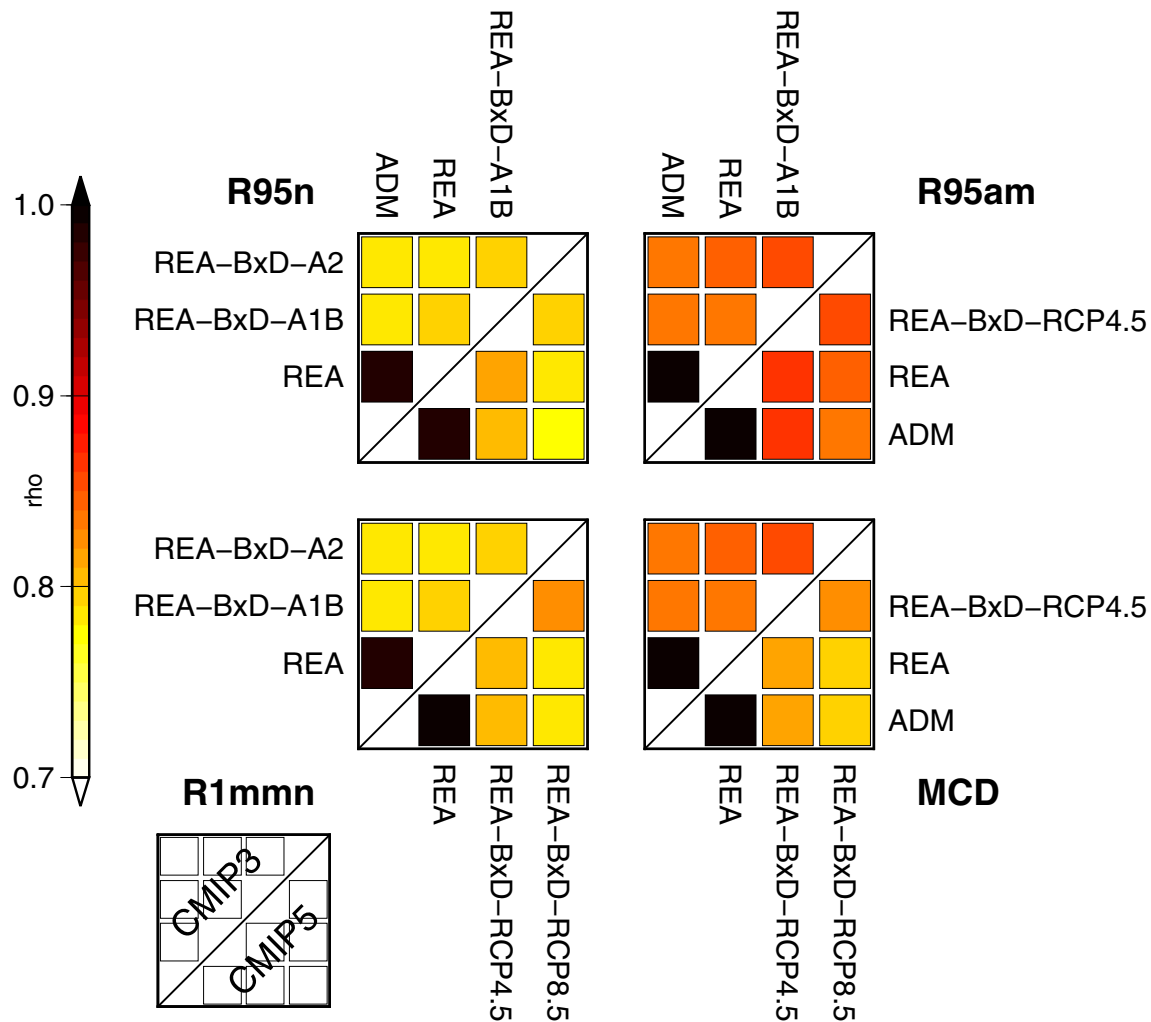


Fig. 5 Correlation coefficients between ranks produced by the three metrics for all situations (subregions and seasons). Indices: R95n and R95am number of days with respectively amount of precipitation above the 95th percentile of 1961–1990, R1mmn and MCD number of days respectively maximum number of consecutive days with a maximum precipitation of 1 mm. Weights: ADM absolute difference of

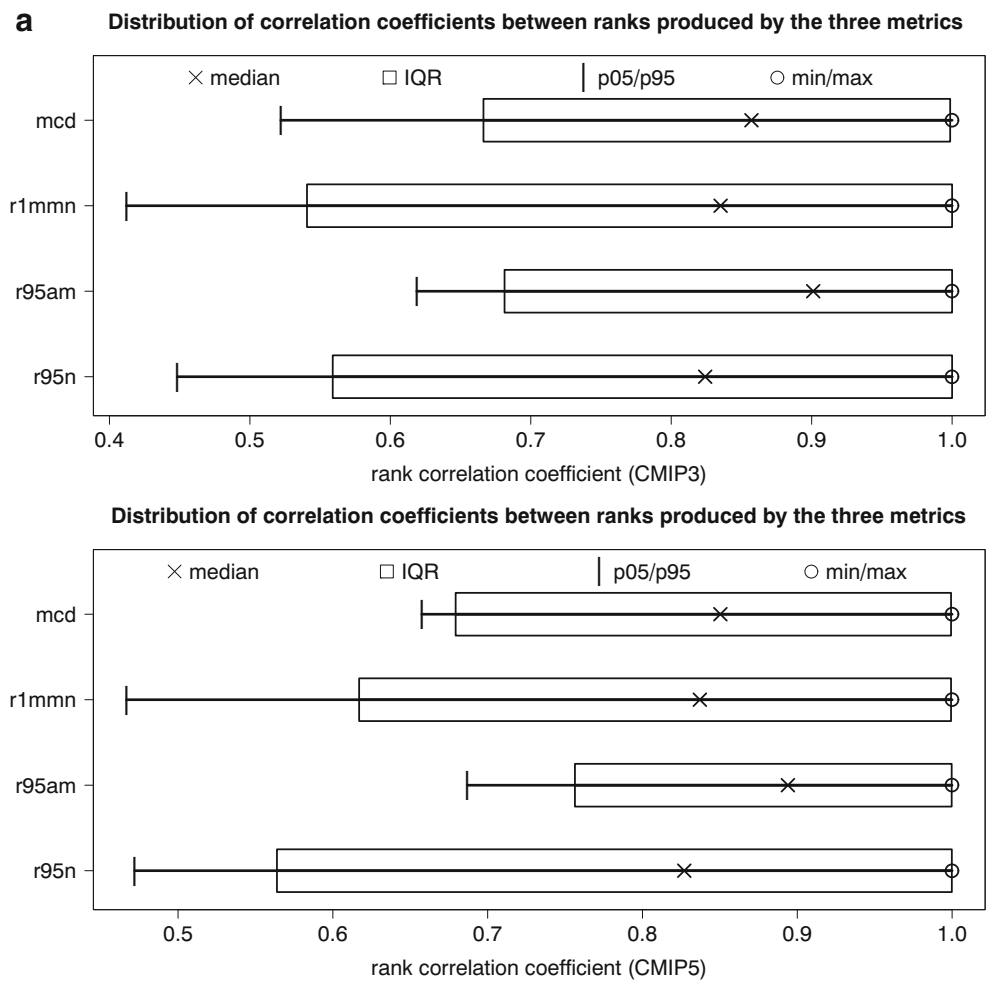
the means, REA ADM with inclusion of a natural variability criterion, REA-B×D REA extended using the difference of change between the model and the multi-model mean (differing for each emission scenario), see Section 3.5

0.00. As REA includes the natural variability (0.20 in the exemplary situation), the raw REA weight becomes 1, standardized 0.11. As the IPSL-CM5B-LR REA weight only differs in the fourth decimal place, REA superimposes ADM in the illustration. While ADM and REA results are rather similar owing to their definitions, the inclusion of the model divergence in REA-B×D has big influence on the quality measure. For this metric, each model is considered for both scenarios of future climate (CMIP3 models A1B and A2, CMIP5 RCP4.5, and RCP8.5). The best-rated model for REA-B×D is bccr-bcm2-0 (ranks 1 and 2 as well as equal weights for A1B and A2, respectively), whereas the RCP8.5 version of IPSL-CM5B-LR is once again judged least plausible (rank 78; RCP4.5 rank 74).

These results are just an example for one index and situation. An overview is possible via the use of averaged ranking lists. REA-B×D is dependent on the emission scenario, whereas the CMIP3 and CMIP5 scenarios do not conform with each other. Thus, a summarization with combined ensembles is only possible on seasonal and regional level, but cannot be performed on metric or index level.

Averaged over all subregions ($n = 8$), seasons ($n = 4$), scenarios (REA-B×D only), and metrics, the ranks depicted in Fig. 4a result for CMIP3. Altogether, the best results are achieved by mri-cgcm2-3-2a, while ipsl-cm4 performs worst. For R95n, mpi-echam5 obtains the first rank (even for each metric), while csiro-mk3.5 comes last (if not

Fig. 6 Boxplots showing the distribution of the correlation coefficients between ranks produced by the three metrics for single situations. Indices: R95n and R95am number of days with respectively amount of precipitation above the 95th percentile of 1961–1990, R1mmn and MCD number of days respectively maximum number of consecutive days with a maximum precipitation of 1 mm



noted otherwise: only on average, single metrics differ). The model mpi-echam5 also occupies the best rank for R95am, opposed by ipsl-cm4 (worst rank for all metrics). The remaining indices agree with this classification and both assign the first place to mri-cgcm2-3-2a.

Concerning CMIP5 (Fig. 4b), CMCC-CM is classified to have the overall highest, IPSL-CM5B-LR the overall lowest performance. CNRM-CM5 occupies the best, ACCESS1-3 the worst rank with regard to R95n. For all other indices, IPSL-CM5B-LR shows the lowest simulation quality (all metrics for R95am), and CNRM-CM5 (R95am), CMCC-CMS (R1mmn) respectively CMCC-CM (MCD) are judged best.

From this, it becomes apparent that model families to some extent show similarities with respect to the assessed simulation quality. While the MPI models usually attain ranks in the upper third, the IPSL family models are classified weak. However, due to being result of averaging quite heterogeneous data regarding seasons and regions, the differences between the gradings described in the paragraphs above are not very pronounced.

As depicted in Fig. 5, the correlations between the ranking of the different metrics over all situation for each index are highest for ADM-REA (0.98–1.00). For the two REA-B×D indices, the coefficients obtain values between 0.73 (R95n CMIP3) and 0.86 (R95am CMIP5). Between ADM or REA and the REA-B×D indices, they are rather similar for one individual index and ensemble, spanning 0.73 (R95n ADM to REA-B×D-A1B CMIP3) to 0.87 (R95am ADM/REA to REA-B×D-RCP4.5 CMIP5). Over all indices, ADM and REA correlate by 0.99, the other metrics between 0.78 and 0.81. A high consensus is also found regarding the individual situations, the median of the rank correlation coefficients being between 0.8 and 0.9, once again with highest values for R95am (see Fig. 6). In contrast to this (not shown), the ranking list of any index is not highly correlated to the ranking list of any other index, irrespective of considering a single metric or all combined. Highest rank correlation coefficients occur for the low precipitation indices R1mmn and MCD (0.30–0.34), the high precipitation indicators correlate by 0.09 to 0.20, regarding the remaining index relations the values fall below 0.16.

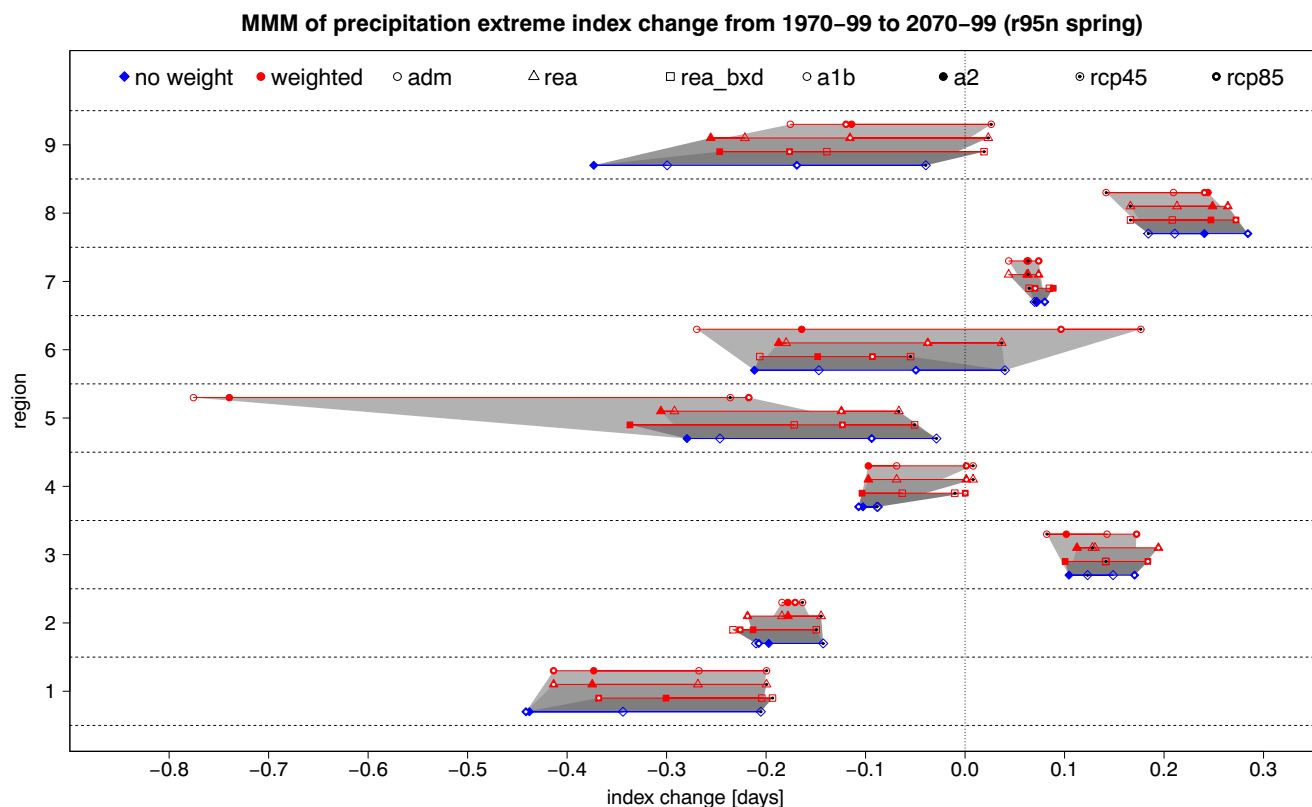


Fig. 7 Multi-model mean of extreme precipitation change for different weighting methods and scenarios. R95n (number of days with precipitation above the 95th percentile of 1961–1990) MAM. Weights: ADM absolute difference of the means, REA ADM with inclusion

of a natural variability criterion, REA-B×D REA extended using the difference of change between the model and the multi-model mean (differing for each emission scenario), see Section 3.5

4.3 Extreme precipitation change in the Mediterranean region

The estimated change of Mediterranean extreme precipitation reveals a high inter-model spread regarding all indices, situations, and scenarios (not shown). In most cases, there is also no consensus regarding the sign of change. This underlines the need for reasonable weighting and averaging of the single model results of the ensembles.

4.3.1 Multi-model mean

An example for the weighted and unweighted MMM results of downscaled R95n for spring is shown in Fig. 7. As can be seen, the difference between scenarios as well as the effect of the metric depends on the region considered. In some cases, the order of the results of the scenarios is modified by weighting, sometimes even the sign of change is reversed, while for other situations no major influence can be made out. Also, the spread between the values of the particular scenarios depends on situations as well as weighting metric. Furthermore, there is no overall consensus regarding the

sign of change. While for region 1, 2, and 5 all metrics and scenarios lead to negative MMMs (less events of high precipitation at the end of the twenty-first century), for regions 3, 7, and 8, they agree on increased frequencies. Regarding regions 4, 6, and the study area as a whole, signs differ with positive changes only occurring based on CMIP5. These observations of dependency on situation, metric, and scenario pertain in general.

Figures 8, 9, 10, and 11 show the sign of simulated MMM change for all indices and scenarios. Each circle represents one region, its sectors the seasons, and the different radii depict the unweighted (smallest) and weighted cases. Overall, positive deltas (more extreme events, red) eventuate in 62% of all cases. They also dominate for the low precipitation indices (R1mmn 83%, MCD 73%; Figs. 8, 9), while for R95n and R95am negative values (blue) slightly prevail (54%, 62%; Figs. 10 and 11). Regarding scenarios and weighting metrics (including unweighted), no differentiation is visible (59–64% positive cases), whereas for the seasons, the percentage decreases from summer (74%) over autumn and spring to winter (48%). Most cases with more extreme precipitation are found for regions 2 and

Fig. 8 Sign of simulated multi-model mean change between the end of the twentieth and the end of the twenty-first century for R95n (number of days with precipitation above the 95th percentile of 1961–1990)

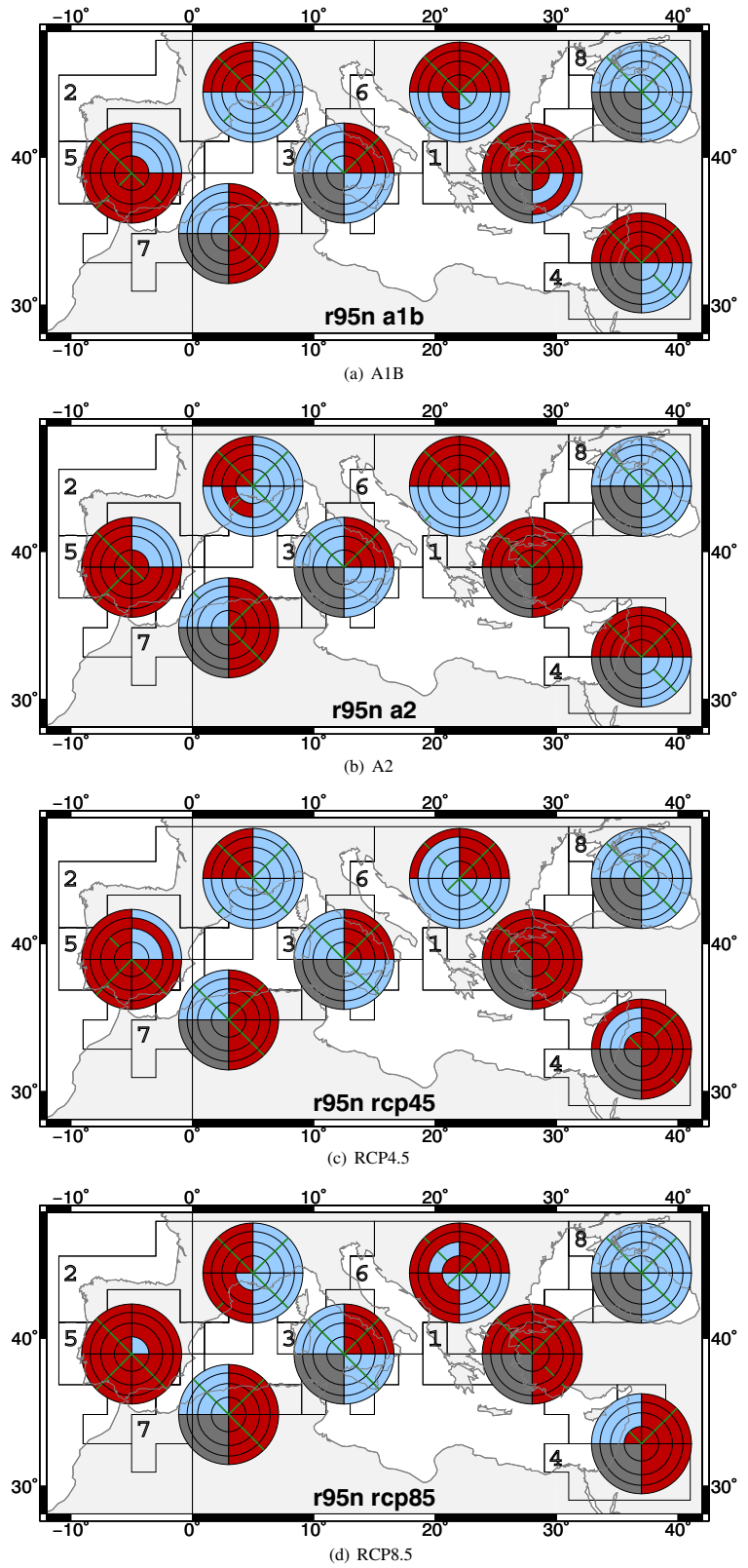


Fig. 9 Same as Fig. 8 for R95am (amount of precipitation above the 95th percentile of 1961–1990)

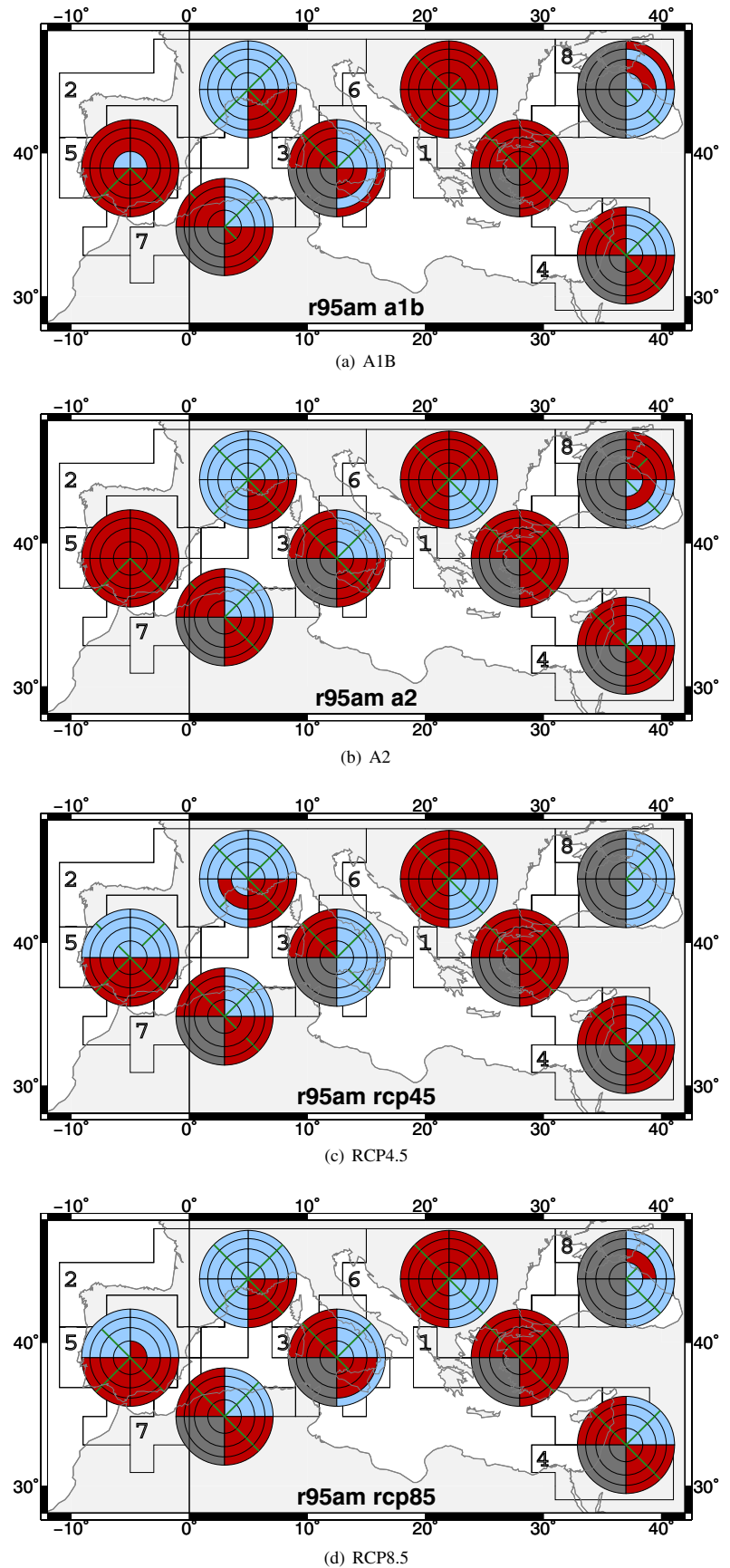


Fig. 10 Same as Fig. 8 for R1mmn (number of days with a maximum precipitation of 1 mm)

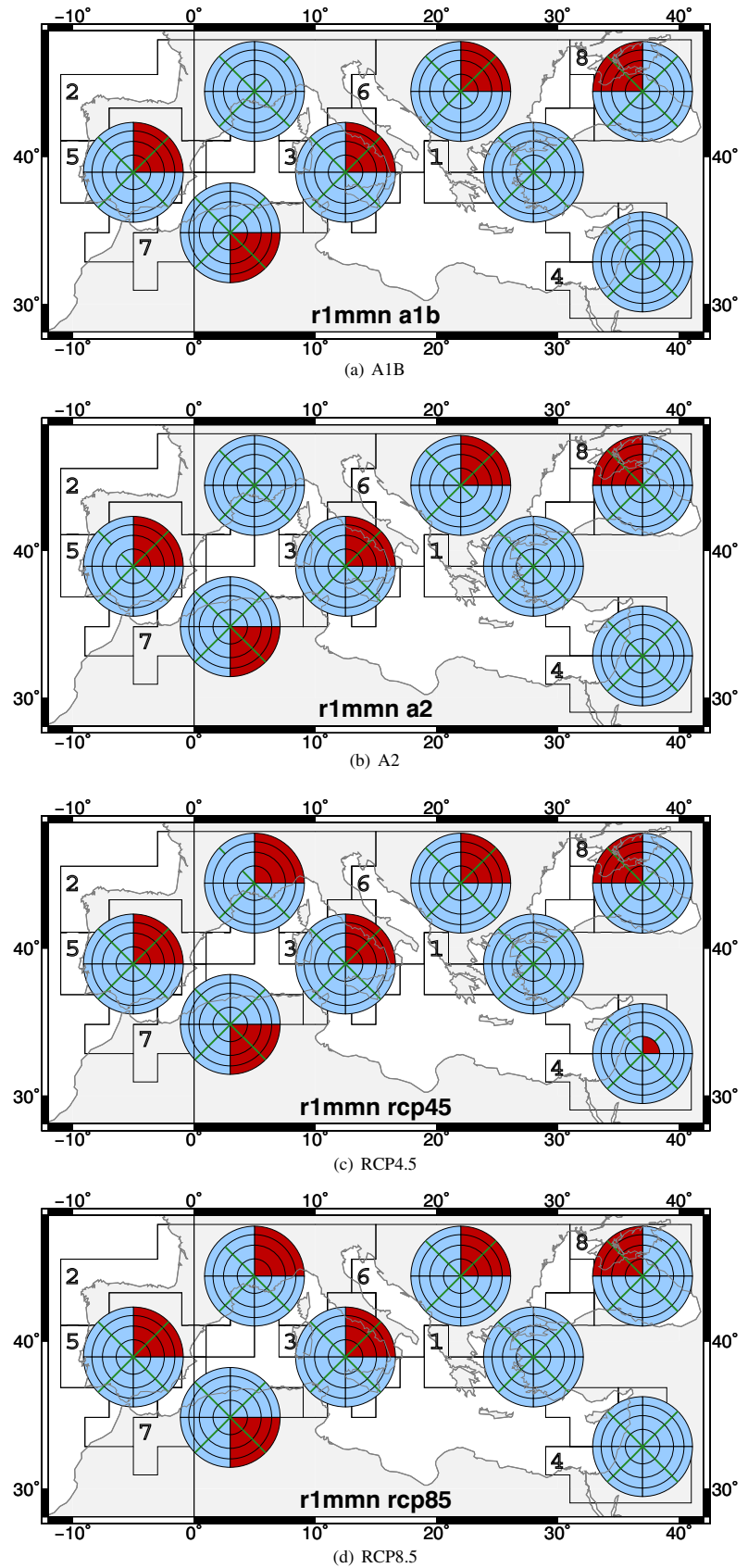
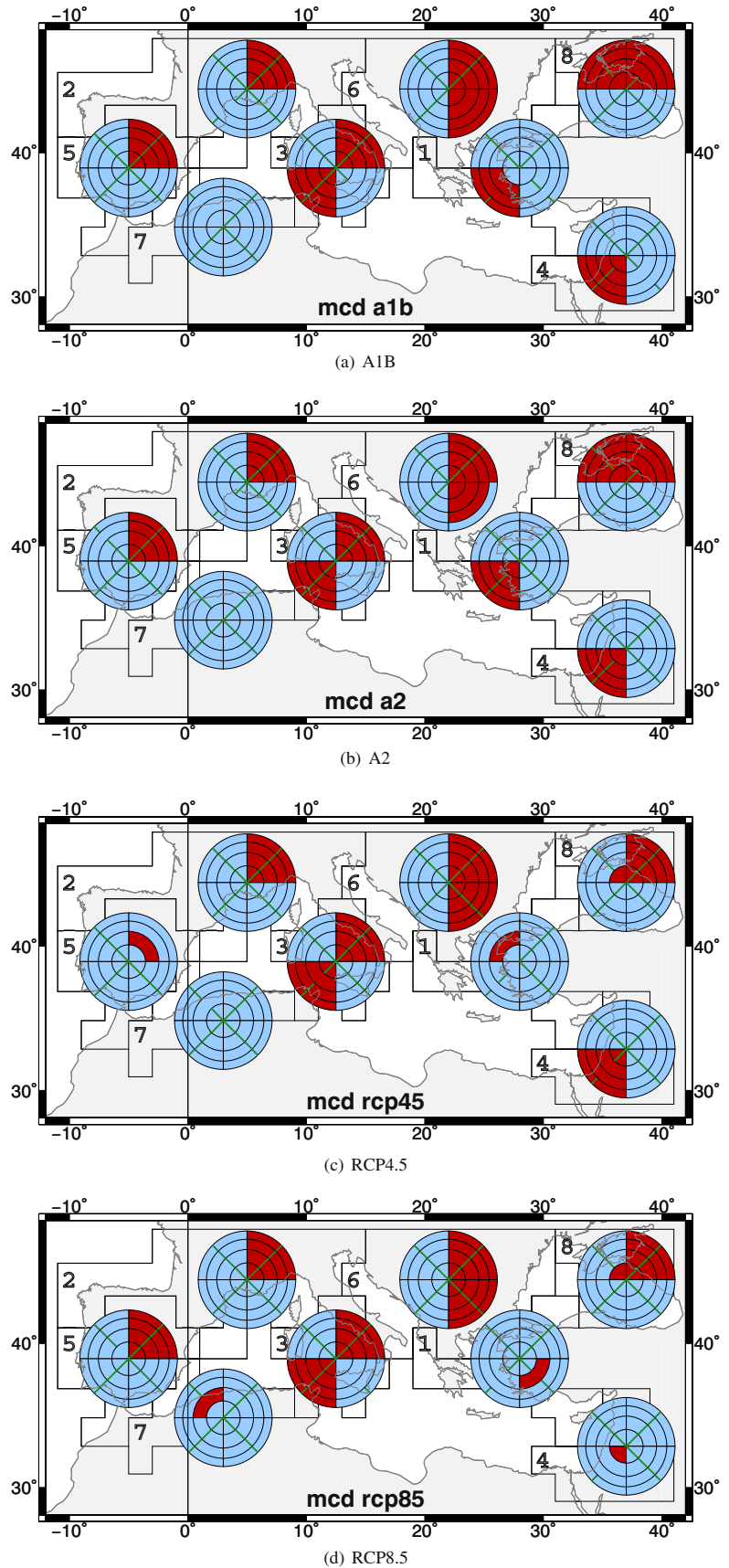


Fig. 11 Same as Fig. 8 for MCD (maximum number of consecutive days with a maximum precipitation of 1 mm)



8 (76%), while for regions 5 and 6, no clear tendency is detectable (49 and 50%). In 87% of all cases, the change in extreme precipitation between the end of the twentieth and the end of the twenty-first century is significant (two-sided t test, $\alpha = 5\%$; green line). For all metrics, regions, seasons and scenarios, the shares are between 78 and 90%, concerning the indices R1mmn stands out with 96%, for R95n the least number of significant cases results (77%). Sixty-three percent of the significant cases show positive signs, once again with no apparent differentiations for metrics and scenarios and highest frequencies for summer and region 8 opposed by lowest frequencies in winter as well as for regions 5 and 6. The tendency of sign for the indices is enhanced when only considering significant deltas. Region 1 shows solely significant deltas with positive sign for R1mmn, with negative sign for R95am. For the summer season in all subregions exclusively significant positive changes occur for R1mmn. Thus, it becomes apparent, that scenario and metrics are of minor importance regarding significance and sign of index change (cp. Section 4.3.3). Some differences regarding the sign show up (regions 8 and summer most, region 6 and winter least positive values) with respect to region and seasons, but the index considered is the most important factor. Most prominent is the obvious increase projected for low precipitation events, particularly for the summer half of the year and for R1mmn. Regarding high precipitation, no clear tendency for the future is visible.

4.3.2 Distribution

The described MMMs are a measure of central tendency for the single-model results. A kernel density estimation of the latter is depicted in Fig. 12 for R1mmn summer in region 2 as example. Each subfigure shows the density distributions for all scenarios including a combined one as well as the individual model values on the bottom for one weighting method or the unweighted case. Here, the MMM is always positive, indicating increased occurrence of dry events for the end of the twenty-first century. While the mean value (right upper corner of each subfigure) shows only small differences between unweighted (upper left) and weighted distributions, the standard deviation is mainly reduced, most notably regarding REA-BxD (lower right subfigure). The results of this situation depict the general outcomes.

4.3.3 Metric influence

The general effect of the metric on mean and standard deviation is illustrated in Fig. 13. With respect to the multi-model mean (abscissa), the direction of influence is quite balanced, with a few more positive cases for high and a few more negative cases for low precipitation indices. Also, when going more into detail, no noticeable differentiation appears, with proportions of positive cases between 42 and 62% regardless of metric, scenario, region,

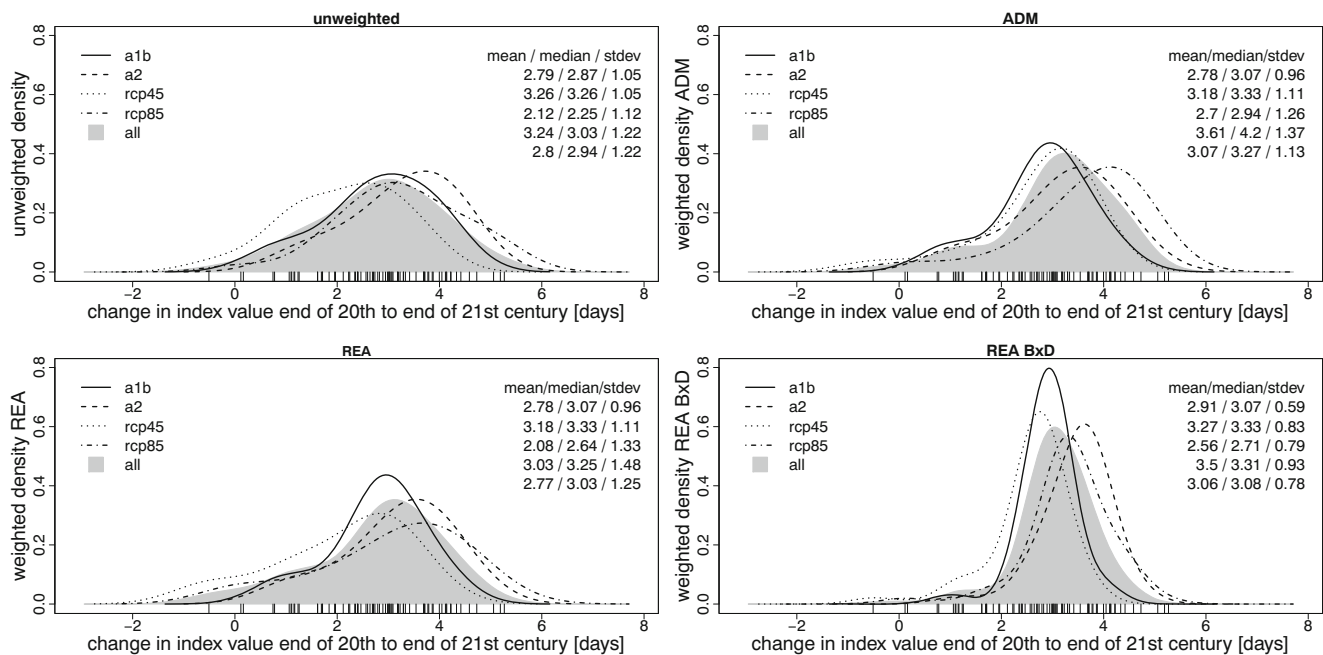


Fig. 12 Example for possible effects of weighting on the shape and position of a multi-model ensemble kernel density distribution. Estimated kernel density distribution for all weighting metrics and scenarios R1mmn (number of days with a maximum precipitation of 1 mm) region 2 JJA. Weights: ADM absolute difference of the means,

REA ADM with inclusion of a natural variability criterion, REA-BxD REA extended using the difference of change between the model and the multi-model mean (differing for each emission scenario), see Section 3.5

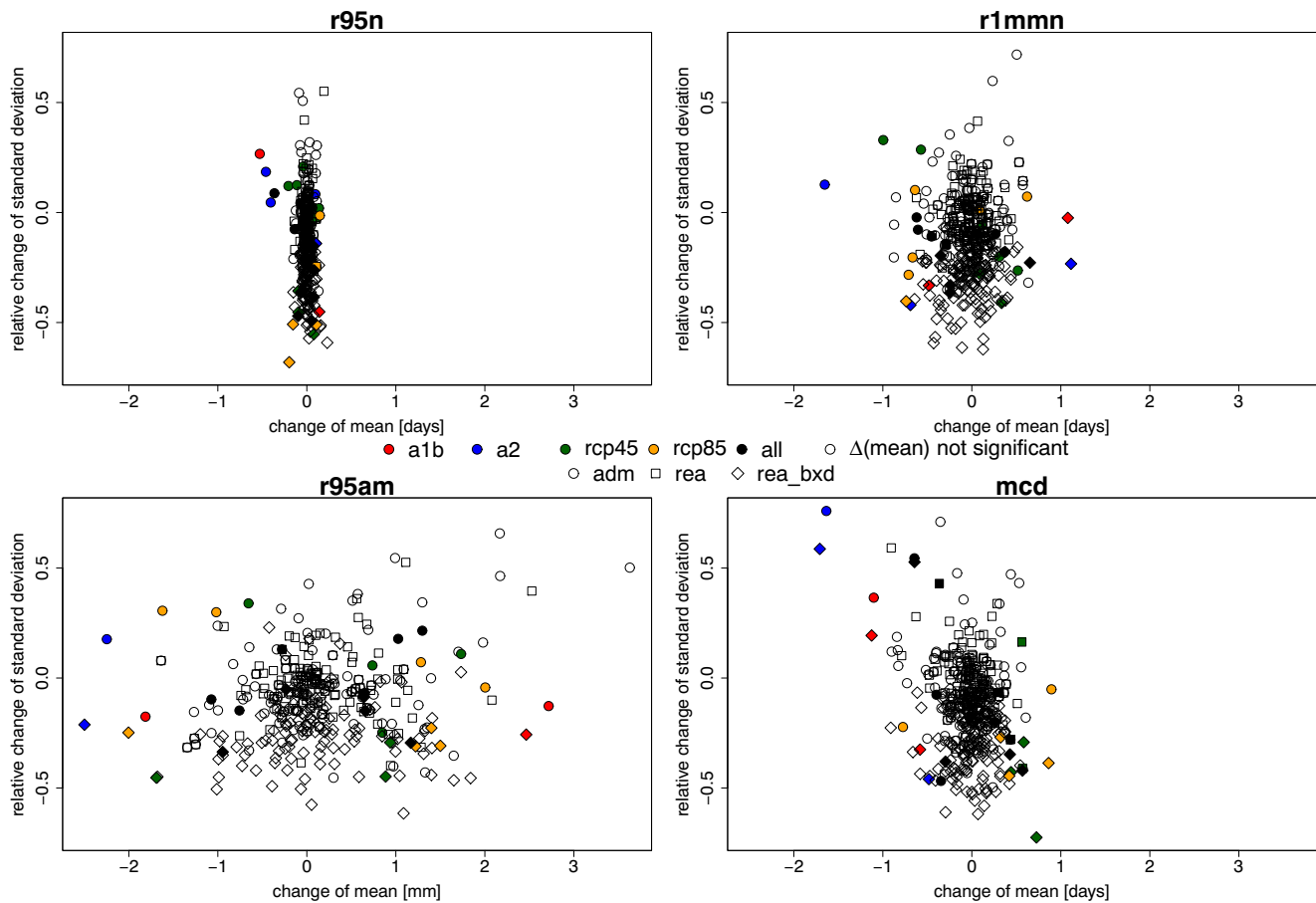


Fig. 13 Change of mean and standard deviation after weighting for all metrics and situations (subregions and seasons). Indices: R95n and R95am number of days with respectively amount of precipitation above the 95th percentile of 1961–1990, R1mmn and MCD number of days respectively maximum number of consecutive days with a maximum precipitation of 1 mm. Index value unit R95am: mm,

remaining indices: days. Metrics: ADM absolute difference of the means, REA ADM with inclusion of a natural variability criterion, REA-B×D REA extended using the difference of change between the model and the multi-model mean (differing for each emission scenario), see Section 3.5. Only situations where the metric influence on the MMM is significant are depicted with color-filled symbols

or season. Furthermore, a significant change of MMM by application of weights exists only for 5% of all cases (colored symbols). Here, some small disparity can be detected with proportionally more significant cases for the CMIP5 compared to the CMIP3 scenarios (8 resp. 7% and 3 resp. 3%) as well as winter (3%) compared to spring and summer (8%). Regarding the regions, regions 2 and 8 (2 and 3%) oppose region 4 (10%). However, most important for the significance of weighting is the metric itself. In merely 1% of all cases, REA exerts significant influence, as against 8 and 7% for ADM and REA-B×D, respectively. The index is of minor importance. Within the significant situations, no direction prevails in general (55% positive), only for some regions (e.g., region 2, 80% negative; region 4, 76% positive) and scenarios.

For the standard deviation as a measure of confidence (ordinate), a reduction is perceptible in 74% of all cases. While index and scenario chosen show hardly any effect,

and also the influence of region and season is not very strong (region 3 (80%) cp. to regions 2 and 8 (66%), summer (66%) cp. to remaining seasons 76–77%), the choice of metric is crucial. REA-B×D leads to a decrease in standard deviation in almost all cases (97%) when compared to the unweighted ones, for ADM and REA this occurs less often (64 and 63%, respectively). However, the influence of the metric on the inter-model confidence is significant in only 9% of all cases, more frequently for the CMIP5 than for the CMIP3 scenarios (11 resp. 12% cp. to 6 resp. 7%) and especially rarely in winter (4%). As for the direction, REA-B×D (25% significant) stands out against both other metrics (2 and 0%, respectively). Of these significant weighting effects, nearly all result in smaller standard deviations (98%). Only regarding ADM, seven of nine cases are negative, for the remaining metrics all significant cases are positive. To sum up, it can be stated that the application of weights does rarely induce a broadening of the probability

density of the multi-model ensemble. Instead, in most cases, a narrowing can be identified, particularly when considering the significant cases. This is mainly driven by the REA-B×D metric, ADM and even more REA exerting less influence.

Thus, the location of the ensemble distribution is hardly affected by the weighting metric applied, while the width is mostly decreased. However, the latter strongly depends on the metric.

5 Discussion

Construction of downscaling models The relative frequency of predictor variables (Fig. 2) exhibits rather low values for sea-level pressure and particularly geopotential height (the three variables for which the dominant PCs account for the highest shares of explained variance, cp. Table 3). This is unexpected as previous studies exposed *zg* as important predictor for extreme precipitation in the Mediterranean, while *psl* was not used in this context (Hertig et al. 2012, 2014, Hertig and Trambly 2017). However, the GLM construction approach involves the direct comparison of possible predictors using correlation analysis. For correlation coefficients between two variables exceeding 0.5, the one with the lower importance regarding model quality is dismissed. Due to this, pressure-related variability can partly be represented by other variables as specific pressure situations lead to typical wind and also humidity patterns. Furthermore, due to complex topographic patterns and a high influence of land-sea contrasts in the Mediterranean area, large-scale circulation patterns can be highly modified on a regional level (e.g., Bolle 2003). The prevalence of wind over humidity predictors in winter for all indices can be explained by the higher importance of advection when compared to local evaporation in this time of the year, while during the climatologically dry seasons, the pattern is reversed (Drumond et al. 2017; Gómez-Hernández et al. 2013). As relative humidity is more decisive with regards to the occurrence of precipitation, the somewhat higher predictor frequency of relative when compared to specific humidity is reasonable.

Model weights and ranks While the model ranks are quite heterogeneous concerning individual seasons and regions, the agreement between the four metrics used is rather strong (minimum rank correlation coefficient 0.73). This was also observed by Ring et al. (2017) regarding mean and trend of precipitation as well as temperature in the Mediterranean subregions and seven global regions. As the metrics are based on each other, this is to be expected. However, the weights and therefore the ranks highly differ between the indices considered. From this

we conclude, that model performance for a specific aspect of extreme precipitation cannot be transferred onto each other. Several studies showed that model performance is to be understood in the context of the variable and also spatial (and temporal) location employed for evaluation (e.g., Gleckler et al. 2008; Kjellström et al. 2010; Ring et al. 2016; Ring et al. 2017). Although the metrics considered yield similar results, other metrics might lead to differing weights and ranks. Apart from not having included structural model interdependencies, the model performance is only assessed with respect to the mean while higher order moments like variability are not incorporated. Furthermore, there are various other performance metrics like e.g. 2×2 contingency table approaches (applied e.g. by Ring et al. 2017) or Bayesian statistics (e.g., Tebaldi et al. 2004) which could be employed. In this context, it should be noted that the length of the time period used for evaluating the GCMs might allow the dominance of multi-decadal natural variability in the form of circulation modes like the North Atlantic Oscillation, which are found to be insufficiently reproduced in the Mediterranean (Paxian et al. 2014). This could also influence the projections of extreme precipitation changes.

Extreme precipitation change in the Mediterranean region

The results of the index delta in the MMM between the end of the twentieth and the end of the twenty-first century regarding the increase in dry extremes in the Mediterranean area was also found in previous studies (e.g., Collins et al. 2013; Christensen et al. 2007; IPCC 2012b; Polade et al. 2015; Seneviratne et al. 2012; Sillmann et al. 2013a), and supported by Hertig and Trambly (2017) with respect to the October to December period (three run ensemble of one GCM RCP8.5).

However, Sillmann et al. (2013a) also detected a small increase in the amount of precipitation from events in excess of the 95th percentile, which contradicts the slight decrease for this index in the study on hand. Nevertheless, a reduction in high precipitation extremes was observed by Hertig et al. (2012) for the Mediterranean region and by Trambly et al. (2012) for Morocco, which could be explained by a lack of water availability despite higher water storage capacity.

The atmospheric water holding capacity increases with rising temperature following the Clausius-Clapeyron relation (e.g., Trenberth et al. 2003; Allen and Ingram 2002). However, high precipitation events also depend on water availability itself, moist-adiabatic temperature lapse rate as well as upward velocity.

The former in combination with convergence should lead to higher intensities but rarer or shorter events, both influencing the overall amount. Yet, the remaining factors complicate the resulting high precipitation properties (Collins et al. 2013; Seneviratne et al. 2012). Emori

and Brown (2005) concluded that for many subtropical regions, “strong upward motion” will occur less frequently under changed climate. The results regarding extreme precipitation change could be explained by a reduction in cyclonal activity of Atlantic origin in the context of a possible strengthening of the North Atlantic Oscillation and poleward expansion of the Hadley Cell descent as indicated in the current IPCC assessment report (Collins et al. 2013; Christensen et al. 2013).

On the Iberian Peninsula for 1986–2005, consecutive dry days were found to increase in summer as well as the whole year regarding the southern parts, whereas the northwestern regions showed decreases (Bartolomeu et al. 2016). For the second half of the twentieth century positive trends of low precipitation occurred as well as weaker negative trends of high precipitation extremes (Rodrigo 2010). Therefore, the findings of the study on hand support and extend the latter observations.

The detected increase in occurrence and length of meteorological droughts, particularly in the summer half of the year (mostly significant deltas, all regions regarding reg. frequency, most of the regions reg. max. duration) is in concordance with the generally projected drying in the Mediterranean area (e.g., Collins et al. 2013), implying increased drought stress, wildfire jeopardy and also water supply shortfalls. Therefore, a reasonable dealing with water resources in the Mediterranean region becomes even more important than today.

In this context it has to be noted that La Jeunesse et al. (2016) found low awareness of stakeholders regarding climate change impacts on water availability in the Mediterranean area. This highlights the need for informing about climate change and its impact on regional and local levels.

Metric influence Since the weighting does hardly exert any statistically significant influence on the mean of the ensemble for each scenario and the direction of influence is rather balanced, the results regarding the MMMs can be seen as robust concerning inter-model differences. Also previous studies did not reveal large metric influences: Comparing weighted and unweighted ensemble means of regional climate models regarding precipitation in Europe only small improvements resulted from metric application (Kjellström et al. 2010; Sánchez et al. 2009). A moderate effect regarding CMIP5 models and North American precipitation was detected by Sanderson et al. (2016), while regarding precipitation extremes resulting from regional climate models in Morocco, no major weighting effect was found by Tramblay et al. (2012).

Nevertheless, the model uncertainty is reduced in about three quarters of all cases, although the influence is rarely significant (9%). A reduction in standard deviation is most

frequent for REA-B×D. This can be explained by the construction of this metric involving a measure of distance between the single and multi-model values of the delta (see Section 3.5). For the other metrics, a negative effect prevails as well. In spite of the lack of significance, this results lead in the same direction as the findings of previous studies (e.g., Paeth et al. 2017; Flato et al. 2013; Seneviratne et al. 2012).

Limitations Some limitations hold true for the study on hand. The use of statistical downscaling for the simulation of changes in extreme precipitation implies the assumption of stationary predictor-predictand relationships. While it is not possible to proof this for the future, the downscaling approach used here involves several bootstrap iterations with random sampling of two thirds of the data, the remaining third being employed for validation. Hereby, a more robust model can be constructed.

A common problem in statistical downscaling is the extrapolation of the predictor-predictand relationships to unknown future climate states. Especially moisture variables are known to underlie substantial changes (Wilby et al. 2004). However, changes of the chosen predictor values in our study are not outside the range of the observed values in the vast majority of cases, indicating the applicability of the predictors. But it should be noted that even if the changes of the individual predictors lie within the range of natural variability of the known climate, the multivariate approach may add additional and unknown uncertainties.

Also, the potential nonstationarity of the GCM biases is an important area of ongoing research. Because our study avoids the use of direct CMIP precipitation output, the potential bias of these data and its behavior in future scenarios have no influence on our results. However, the biases of our predictor variables may underlie a time dependent magnitude. In what sense potential changing biases of the large-scale atmospheric drivers might influence the downscaled predictand through the MLR equations is a very interesting question and should be subject of further research.

A further limitation of the presented results is the observational data employed. The choice of reference data as well as unavoidable errors inside the same can impose influence on the evaluation results. This particularly holds true with regard to extreme precipitation (Gleckler et al. 2008; Gómez-Navarro et al. 2012; Flato et al. 2013). With respect to the indices, the calculation is conducted in the form of single-monthly seasons. While this is of no effect for R95n, R95am, and R1mmn, cutting the MCD index at the end of every month could reduce the maximum number of consecutive dry days. Yet, the index is defined consistently for all analyses, ensuring comparability.

Weighting method The ideal solution for climate model weighting is an intensively discussed research field. Not only the overall question whether to weight or not is debatable but also the importance of different requirements for a sound weighting metric is investigated. The weighting approach in this study has its focus on the avoidance of the direct use of precipitation as it presents an insufficient modeled variable for the weighting process, instead, a more complex and as much as possible robust multivariate downscaling approach is used as basis for the weighting. However, recent studies propose additional conditions for optimal GCM weighting. For example, Knutti et al. (2017) propose a weighting scheme that accounts also for model interdependencies which ensures that GCMs with similar background (duplicated code, resolution experiments, etc.) are down-weighted. Another research topic focusses on so-called emergent constraints which is able to reduce uncertainties in climate change projections by finding “relationships across an ensemble of models, between some aspect of the Earth system sensitivity and an observable trend or variation in the current climate” (Wenzel et al. 2014). However, this method is highly susceptible to subjective decisions of the researcher, the quality of the observed data and the understanding of the physical processes. As Caldwell et al. (2018) show, recent studies using emergent constraints on equilibrium climate sensitivity cast more doubt than they confirm. An application on extreme precipitation might be even more challenging.

6 Summary and conclusions

In this study, seasonal changes of extreme precipitation in eight Mediterranean subregions between 1970–1999 and 2070–2099 are analyzed using weighted multi-model ensembles based on statistical downscaling. Considering two indices for precipitation scarcity with a daily threshold of 1 mm as well as two indices of heavy precipitation delineated by the 95th percentile, precipitation extremes are projected to increase in 65% of all cases (cp. Figs. 8, 9, 10, and 11: all put together, 65% of all areas are filled in light-blue color, signifying a positive sign of change). However, this positive delta is caused by a higher-magnitude increase for dry events (R1mmn 83 %, MCD 73 %) and a lower-magnitude decrease for high precipitation (R95n 54%, R95am 62%). Most of the changes are statistically significant. For these general results, subregion and season exert influence in some cases, while scenario and metric are of minor importance. All in all, a rise in precipitation scarcity is simulated while for heavy precipitation the results are more balanced with negative tendencies.

Regarding the MMMs, the application of weights shows evened impacts which are hardly significant. However, when considering the standard deviation as a measure of model uncertainty, a decrease can be found in 74% of all cases (cp. ordinate of Fig. 13: for all four sub-figures (indices) put together, 74% of all symbols are located below 0, particularly for the REA-B×D metric shown as rhombus), albeit rarely significant (9%). Although this lack of significance lowers the validity of this results, a tendency towards enhanced model consensus by weight application is visible.

This rather robust projection of increased meteorological drought frequency and duration combined with a small decrease in heavy precipitation event frequency and amount is in line with the findings of a general rise in dryness in the Mediterranean region (e.g., Collins et al. 2013) and implies large socio-economic and ecological impacts.

Further studies could include a more detailed look by studying intra-regional heterogeneities on a grid cell scale. Additionally a comparison with non-downscaled results as well as the application of the downscaling-based metrics to deltas calculated from direct GCM precipitation output might enhance the gain of insight. Finally, to contrast the results to those yielded by other model performance metrics could bring further insights.

Acknowledgements We acknowledge the respective modeling groups as well as the U.S. Department of Energy’s Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the World Climate Research Programme’s Working Group on Coupled Modelling (WGCM) for producing, coordinating, and making available the CMIP3 and CMIP5 multi-model datasets. 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>). Further, NCEP-NCAR Reanalysis data was provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at <http://www.esrl.noaa.gov/psd/>. Additionally, we appreciate the helpful comments of two anonymous reviewers.

Funding information 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) under contract PA 1194/10-1 and JA 831/11-1.

References

- Allen MR, Ingram WJ (2002) Constraints on future changes in climate and the hydrologic cycle. *Nature* 419(6903):224–232. <https://doi.org/10.1038/nature01092>
- Bartolomeu S, Carvalho M, Marta-Almeida M, Melo-Gonçalves P, Rocha A (2016) Recent trends of extreme precipitation indices in the Iberian Peninsula using observations and WRF model results. *Phys Chem Earth Parts A/B/C* 94:10–21. <https://doi.org/10.1016/J.PCE.2016.06.005>
- Bolle HJ (2003) Climate, climate variability, and impacts in the Mediterranean Area: an overview. In: *Mediterranean climate*.

- Springer, Berlin, pp 5–86. https://doi.org/10.1007/978-3-642-55657-9_2
- Caldwell PM, Zelinka MD, Klein SA (2018) Evaluating emergent constraints on equilibrium climate sensitivity. *J Climate* 31(10):3921–3942. <https://doi.org/10.1175/JCLI-D-17-0631.1>
- Cavazos T, Hewitson BC (2005) Performance of NCEP – NCAR reanalysis variables in statistical downscaling of daily precipitation. *Climate Res* 28(2):95–107. <https://doi.org/10.3354/cr028095>
- Christensen J, Hewitson B, Busuioac A, Chen A, Gao X, Held I, Jones R, Kolli R, Kwon WT, Laprise R, Magaña Rueda V, Mearns L, Menéndez C, Räisänen J, Rinke A, Sarr A, Whetton P (2007) Regional climate projections. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) *Climate change 2007. The physical science basis. Contribution of working Group I to the fourth assessment report of the intergovernmental panel on climate change*. Cambridge University Press, chap 8, Cambridge, pp 847–940
- Christensen JH, Krishna Kumar K, Aldrian E, An SI, Cavalcanti IFA, de Castro M, Dong W, Goswami P, Hall A, Kanyanga JK, Kitoh A, Kossin J, Lau NC, Renwick J, Stephenson DB, Xie SP, Zhou T (2013) Climate phenomena and their relevance for future regional climate change. In: Stocker TF, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds) *Climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change*. University Press, Cambridge
- Collins M (2007) Ensembles and probabilities: a new era in the prediction of climate change. *Philos Trans Series Math Phys Eng Sci* 365(1857):1957–70. <https://doi.org/10.1098/rsta.2007.2068>
- Collins M, Knutti R, Arblaster J, Dufresne JL, Fichet F, Friedlingstein P, Gao X, Gutowski W, Johns T, Krinner G, Shongwe M, Tebaldi C, Weaver A, Wehner M (2013) Long-term climate change: projections, commitments and irreversibility. In: Stocker T, Qin D, Plattner GK, Tignor M, Allen S, Boschung J, Nauels A, Xia Y, Bex V, Midgley P (eds) *Climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change*. Cambridge University Press, Cambridge, pp 1029–1136. chap. 12
- Cornes R, Goodess C, Casanueva A, Coppola E, Dell’Aquila A, Dubois C, Frías MD, Gutiérrez JM, Herrera S (2013) Deliverable D3.7 Extremes: downscaling of percentiles and extremes events indicators
- Diffenbaugh NS, Giorgi F (2012) Climate change hotspots in the CMIP5 global climate model ensemble. *Clim Change* 114(3–4):813–822. <https://doi.org/10.1007/s10584-012-0570-x>
- Drumond A, Gimeno L, Nieto R, Trigo RM, Vicente-Serrano SM (2017) Drought episodes in the climatological sinks of the Mediterranean moisture source: the role of moisture transport. *Global Planet Change* 151:4–14. <https://doi.org/10.1016/j.gloplacha.2016.12.004>
- Duan Q, Phillips TJ (2010) Bayesian estimation of local signal and noise in multimodel simulations of climate change. *J Geophys Res Atmos* 115(18):D18123. <https://doi.org/10.1029/2009JD013654>
- Emori S, Brown SJ (2005) Dynamic and thermodynamic changes in mean and extreme precipitation under changed climate. *Geophys Res Lett* 32(17):1–5. <https://doi.org/10.1029/2005GL023272>
- Fahrmeir L, Kneib T, Lang S, Marx B (2013) *Regression: models, methods and applications*. Springer, Berlin. <https://doi.org/10.1007/978-3-642-34333-9>
- Flato G, Marotzke J, Abiodun B, Braconnot P, Chou S, Collins W, Cox P, Driouech F, Emori S, Eyring V, Forest C, Gleckler P, Gouillyard E, Jakob C, Kattsov V, Reason C, Rummukainen M (2013) Evaluation of climate models. In: Stocker T, Qin D, Plattner GK, Tignor M, Allen S, Boschung J, Nauels A, Xia Y, Bex V, Midgley P (eds) *Climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change*. Cambridge University Press, Cambridge, pp 741–866. chap 9
- Fowler HJ, Blenkinsop S, Tebaldi C (2007) Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. *Int J Climatol* 27(12):1547–1578. <https://doi.org/10.1002/joc.1556>
- Giorgi F, Mearns LO (2002) Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the “Reliability Ensemble Averaging” (REA) method. *J Clim* 15(10):1141–1158. [https://doi.org/10.1175/15210.1175/1520-0442\(2002\)015<1141:COAURA>2.0.CO;2](https://doi.org/10.1175/15210.1175/1520-0442(2002)015<1141:COAURA>2.0.CO;2)
- Gleckler PJ, Taylor KE, Doutriaux C (2008) Performance metrics for climate models. *J Geophys Res Atmos* 113(6):D06104. <https://doi.org/10.1029/2007JD008972>
- Gómez-Hernández M, Drumond A, Gimeno L, Garcia-Herrera R (2013) Variability of moisture sources in the Mediterranean region during the period 1980–2000. *Water Resour Res* 49(10):6781–6794. <https://doi.org/10.1002/wrcr.20538>
- Gómez-Navarro JJ, Montávez JP, Jerez S, Jiménez-Guerrero P, Zorita E (2012) What is the role of the observational dataset in the evaluation and scoring of climate models? *Geophys Res Lett* 39(24):L24701. <https://doi.org/10.1029/2012GL054206>
- Gutiérrez JM, Maraun D, Widmann M, Huth R, Hertig E, Benestad R, Roessler O, Wibig J, Wilcke R, Kotlarski S, San Martín D, Herrera S, Bedia J, Casanueva A, Manzanar R, Iturbide M, Vrac M, Dubrovsky M, Ribalaygua J, Pórtolos J, Räty O, Räisänen J, Hingray B, Raynaud D, Casado MJ, Ramos P, Zerener T, Turco M, Bosshard T, Štěpánek P, Bartholy J, Pongracz R, Keller DE, Fischer AM, Cardoso RM, Soares PMM, Czernecki B, Pagé C (2018) An intercomparison of a large ensemble of statistical downscaling methods over Europe: results from the VALUE perfect predictor cross-validation experiment. *International Journal of Climatology*. <https://doi.org/10.1002/joc.5462>
- Hawkins E, Sutton R (2011) The potential to narrow uncertainty in projections of regional precipitation change. *Climate Dynam* 37(1–2):407–418. <https://doi.org/10.1007/s00382-010-0810-6>
- Haylock MR, Goodess CM (2004) Interannual variability of European extreme winter rainfall and links with mean large-scale circulation. *Int J Climatol* 24(6):759–776. <https://doi.org/10.1002/joc.1033>
- Haylock MR, Hofstra N, Klein Tank AMG, Klok EJ, Jones PD, New M (2008) A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *J Geophys Res* 113(D20):D20119. <https://doi.org/10.1029/2008JD010201>
- Hertig E, Jacobeit J (2015) Considering observed and future nonstationarities in statistical downscaling of Mediterranean precipitation. *Theor Appl Climatol* 122(3–4):667–683. <https://doi.org/10.1007/s00704-014-1314-9>
- Hertig E, Trambly Y (2017) Regional downscaling of Mediterranean droughts under past and future climatic conditions. *Global Planet Change* 151:36–48. <https://doi.org/10.1016/j.gloplacha.2016.10.015>
- Hertig E, Paxian A, Vogt G, Seubert S, Paeth H, Jacobeit J (2012) Statistical and dynamical downscaling assessments of precipitation extremes in the Mediterranean area. *Meteorol Z* 21(1):61–77. <https://doi.org/10.1127/0941-2948/2012/0271>
- Hertig E, Seubert S, Paxian A, Vogt G, Paeth H, Jacobeit J (2014) Statistical modelling of extreme precipitation indices for the Mediterranean area under future climate change. *Int J Climatol* 34(4):1132–1156. <https://doi.org/10.1002/joc.3751>
- Hertig E, Maraun D, Bartholy J, Pongracz R, Vrac M, Mares I, Gutiérrez JM, Wibig J, Casanueva A, Soares PMM (2018) Comparison of statistical downscaling methods with respect to

- extreme events over Europe: validation results from the perfect predictor experiment of the COST Action VALUE. *International Journal of Climatology*. <https://doi.org/10.1002/joc.5469>
- IPCC (2012a) Glossary of terms. In: Field C, Barros V, Stocker T, Qin D, Dokken D, Ebi K, Mastrandrea M, Mach K, Plattner GK, Allen S, Tignor M, Midgley P (eds) *Managing the risks of extreme events and disasters to advance climate change adaptation. A special report of working groups I and II of the intergovernmental panel on climate change*. Cambridge University Press, Cambridge, pp 555–564
- IPCC (2012b) Summary for policymakers. In: Field C, Barros V, Stocker T, Qin D, Dokken D, Ebi K, Mastrandrea M, Mach K, Plattner GK, Allen S, Tignor M, Midgley P (eds) *Managing the risks of extreme events and disasters to advance climate change adaptation. A special report of working groups I and II of the intergovernmental panel on climate change*. Cambridge University Press, Cambridge, pp 3–22
- Jacobeit J (1993) Regionale Unterschiede im atmosphärischen Zirkulationsgeschehen bei globalen Klimaveränderungen. *Die Erde* 124:63–77
- Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Ropelewski C, Wang J, Leetmaa A, Reynolds R, Jenne R, Joseph D (1996) The NCEP/NCAR 40-year reanalysis project. *Bull Am Meteorol Soc* 77(3):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)
- Kaspar-Ott I, Hertig E, Kaspar S, Pollinger F, Ring C, Paeth H, Jacobeit J (2019) Weights for general circulation models from CMIP3/CMIP5 in a statistical downscaling framework and the impact on future Mediterranean precipitation. *International Journal of Climatology*. <https://doi.org/10.1002/joc.6045>
- Kjellström E, Boberg F, Castro M, Christensen J, Nikulin G, Sánchez E (2010) Daily and monthly temperature and precipitation statistics as performance indicators for regional climate models. *Climate Res* 44(2-3):135–150. <https://doi.org/10.3354/cr00932>
- Klein Tank AM, Zwiers FW, Zhang X (2009) Guidelines on analysis of extremes in a changing climate in support of informed decisions for adaptation. *Tech. Rep. 72 WMO World Climate Data and Monitoring Programme*, Geneva, Switzerland
- Knutti R, Abramowitz G, Collins M, Eyring V, Gleckler P, Hewitson B, Mearns L (2010a) Good practice guidance paper on assessing and combining multi model climate projections. *IPCC Expert Meeting on Assessing and Combining Multi Model Climate Projections*, p 15
- Knutti R, Furrer R, Tebaldi C, Cermak J, Meehl GA (2010b) Challenges in combining projections from multiple climate models. *J Climate* 23(10):2739–2758. <https://doi.org/10.1175/2009JCLI3361.1>
- Knutti R, Sedláček J, Sanderson BM, Lorenz R, Fischer EM, Eyring V (2017) A climate model projection weighting scheme accounting for performance and interdependence. *Geophys Res Lett* 44(4):1909–1918. <https://doi.org/10.1002/2016GL072012>
- La Jeunesse I, Cirelli C, Aubin D, Larrue C, Sellami H, Afifi S, Bellin A, Benabdallah S, Bird DN, Deidda R, Dettori M, Engin G, Herrmann F, Ludwig R, Mabrouk B, Majone B, Paniconi C, Soddu A (2016) Is climate change a threat for water uses in the Mediterranean region? Results from a survey at local scale. *Sci Total Environ* 543:981–996. <https://doi.org/10.1016/j.scitotenv.2015.04.062>
- Lavaysse C, Vrac M, Drobinski P, Lengaigne M, Vischel T (2012) Statistical downscaling of the French Mediterranean climate: assessment for present and projection in an anthropogenic scenario. *Nat Hazards and Syst Sci* 12(3):651–670. <https://doi.org/10.5194/nhess-12-651-2012>
- Lutz K, Jacobeit J, Philipp A, Seubert S, Kunstmann H, Laux P (2012) Comparison and evaluation of statistical downscaling techniques for station-based precipitation in the Middle East. *Int J Climatol* 32(10):1579–1595. <https://doi.org/10.1002/joc.2381>
- Maraun D, Wetterhall F, Ireson AM, Chandler RE, Kendon EJ, Widmann M, Brienen S, Rust HW, Sauter T, Themeßl M, Venema VKC, Chun KP, Goodess CM, Jones RG, Onof C, Vrac M, Thiele-Eich I (2010) Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user. *Rev Geophys* 48(3):RG3003. <https://doi.org/10.1029/2009RG000314>
- Maraun D, Widmann M, Gutiérrez JM, Kotlarski S, Chandler RE, Hertig E, Wibig J, Huth R, Wilcke RA (2015) VALUE: a framework to validate downscaling approaches for climate change studies. *Earth's Fut* 3(1):1–14. <https://doi.org/10.1002/2014EF000259>
- Maraun D, Huth R, Gutiérrez JM, Martín DS, Dubrovsky M, Fischer A, Hertig E, Soares PMM, Bartholy J, Pongrácz R, Widmann M, Casado MJ, Ramos P, Bedia J (2017) The VALUE perfect predictor experiment: evaluation of temporal variability. *International Journal of Climatology*. <https://doi.org/10.1002/joc.5222>
- Maraun D, Widmann M (2018) *Statistical downscaling and bias correction for climate research*, 1st edn. Cambridge University Press, Cambridge. <https://doi.org/10.1017/9781107066052>
- McCullagh PJ, Nelder JA (1989) *Generalized linear models*. Chapman & Hall/CRC, London
- Meehl G, Covey C, Delworth T, Latif M, McAvaney B, Mitchell J, Stouffer R, Taylor K (2007) The WCRP CMIP3 multi-model dataset: a new era in climate change research. *Bull Am Meteorol Soc* 88:1383–1394
- Merkenschlager C, Hertig E, Jacobeit J (2017) Non-stationarities in the relationships of heavy precipitation events in the Mediterranean area and the large-scale circulation in the second half of the 20th century. *Global Planet Change* 151:108–121. <https://doi.org/10.1016/j.gloplacha.2016.10.009>
- Monjo R, Gaitán E, Pórtoles J, Ribalaygua J, Torres L (2016) Changes in extreme precipitation over Spain using statistical downscaling of CMIP5 projections. *Int J Climatol* 36(2):757–769. <https://doi.org/10.1002/joc.4380>
- Moss R, Babiker M, Brinkman S, Calvo E, Carter T, Edmonds J, Elgizouli I, Emori S, Erda L, Hibbard K, Jones R, Kainuma M, Kelleher J, Lamarque JF, Manning M, Matthews B, Meehl J, Meyer L, Mitchell J, Nakicenovic N, O'Neill B, Pichs R, Riahi K, Rose S, Runci P, Stouffer R, van Vuuren D, Weyant J, Wilbanks T, van Ypersele JP, Zurek M (2008) Towards new scenarios for analysis of emissions, climate change, impacts, and response strategies. *Technical Summary*. IPCC Technical Meeting Report 19–21 September 2007. Genf: IPCC, Noordwijkerhout
- Mueller B, Seneviratne SI (2014) Systematic land climate and evapotranspiration biases in CMIP5 simulations. *Geophys Res Lett* 41(1):128–134. <https://doi.org/10.1002/2013GL058055>
- Nicholls N, Murray W (1999) Workshop on indices and indicators for climate extremes: Asheville, NC, USA, 3–6 June 1997 Breakout Group B: precipitation. *Clim Change* 42(1):23–29
- Nuissier O, Joly B, Joly A, Ducrocq V, Arbogast P (2011) A statistical downscaling to identify the large-scale circulation patterns associated with heavy precipitation events over southern France. *Q J Roy Meteorol Soc* 137(660):1812–1827. <https://doi.org/10.1002/qj.866>
- Paeth H, Vogt G, Paxian A, Hertig E, Seubert S, Jacobeit J (2017) Quantifying the evidence of climate change in the light of uncertainty exemplified by the Mediterranean hot spot region. *Global Planet Change* 151:144–151. <https://doi.org/10.1016/j.gloplacha.2016.03.003>
- Paxian A, Hertig E, Vogt G, Seubert S, Jacobeit J, Paeth H (2014) Greenhouse gas-related predictability of regional climate model

- trends in the Mediterranean area. *Int J Climatol* 34(7):2293–2307. <https://doi.org/10.1002/joc.3838>
- Philipp A, Della-Marta PM, Jacobeit J, Fereday DR, Jones PD, Moberg A, Wanner H (2007) Long-term variability of daily North Atlantic-European pressure patterns since 1850 classified by simulated annealing clustering. *J Climate* 20(16):4065–4095. <https://doi.org/10.1175/JCLI4175.1>
- Polade SD, Pierce DW, Cayan DR, Gershunov A, Dettinger MD (2015) The key role of dry days in changing regional climate and precipitation regimes. *Sci Rep* 4(1):4364. <https://doi.org/10.1038/srep04364>
- Räisänen J, Ylhäisi JS (2012) Can model weighting improve probabilistic projections of climate change? *Climate Dynam* 39(7–8):1981–1998. <https://doi.org/10.1007/s00382-011-1217-8>
- Randall D, Wood R, Bony S, Colman R, Fichet T, Fyfe J, Kattsov V, Pitman A, Shukla J, Srinivasan J, Stouffer R, Sumi A, Taylor K (2007) Climate models and their evaluation. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) *Climate change 2007. The physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change*. Cambridge University Press, Cambridge, pp 589–662. chap 8
- Reichler T, Kim J, Reichler T, Kim J (2008) How well do coupled models simulate today's climate? *Bull Am Meteorol Soc* 89(3):303–311. <https://doi.org/10.1175/BAMS-89-3-303>
- Ring C, Mannig B, Pollinger F, Paeth H (2016) Uncertainties in the simulation of precipitation in selected regions of humid and dry climate. *Int J Climatol* 36(10):3521–3538. <https://doi.org/10.1002/joc.4573>
- Ring C, Pollinger F, Kaspar-Ott I, Hertig E, Jacobeit J, Paeth H (2017) A comparison of metrics for assessing state-of-the-art climate models and implications for probabilistic projections of climate change. *Climate Dynamics*. <https://doi.org/10.1007/s00382-017-3737-3>
- Rodrigo FS (2010) Changes in the probability of extreme daily precipitation observed from 1951 to 2002 in the Iberian Peninsula. *Int J Climatol* 30(10):1512–1525. <https://doi.org/10.1002/joc.1987>
- Sánchez E, Romera R, Gaertner MA, Gallardo C, 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. *Atmos Sci Lett* 10(4):241–248. <https://doi.org/10.1002/asl.230>
- Sanderson BM, Wehner M, Knutti R (2016) Skill and independence weighting for multi-model. <https://doi.org/10.5194/gmd-2016-285>
- Seneviratne SI, Nicholls N, Easterling D, Goodess CM, Kanae S, Kossin J, Luo Y, Marengo J, McInnes K, Rahimi M, Reichstein M, Sorteberg A, Vera C, Zhang X, Rusticucci M, Semenov V, Alexander LV, Allen S, Benito G, Cavazos T, Clague J, Conway D, Della-Marta PM, Gerber M, Gong S, Goswami BN, Hemer M, Huggel C, van den Hurk B, Kharin VV, Kitoh A, Tank AMK, Li G, Mason S, McGuire W, van Oldenborgh GJ, Orłowsky B, Smith S, Thiaw W, Velegrakis A, Yiou P, Zhang T, Zhou T, Zwiers FW (2012) Changes in climate extremes and their impacts on the natural physical environment. In: Field C, Barros V, Stocker T, Qin D, Dokken D, Ebi K, Mastrandrea M, Mach K, Plattner GK, Allen S, Tignor M, Midgley P (eds) *Managing the risks of extreme events and disasters to advance climate change adaptation*. Cambridge University Press, Cambridge, pp 109–230. <https://doi.org/10.1017/CBO9781139177245.006>. chap 3
- Sillmann J, Kharin VV, Zhang X, Zwiers FW, Bronaugh D (2013a) Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *J Geophys Res Atmos* 118(4):1716–1733. <https://doi.org/10.1002/jgrd.50203>
- Sillmann J, Kharin VV, Zwiers FW, Zhang X, Bronaugh D (2013b) Climate extremes indices in the CMIP5 multimodel ensemble: part 2. Future climate projections. *J Geophys Res: Atmos* 118(6):2473–2493. <https://doi.org/10.1002/jgrd.50188>
- Stephens GL, L'Ecuyer T, Forbes R, Gettelmen A, Golaz JC, Bodas-Salcedo A, Suzuki K, Gabriel P, Haynes J (2010) Dreary state of precipitation in global models. *J Geophys Res: Atmos* 115(D24):D24211. <https://doi.org/10.1029/2010JD014532>
- Sun Y, Solomon S, Dai A, Portmann RW (2006) How often does it rain? *J Climate* 19(6):916–934. <https://doi.org/10.1175/JCLI3672.1>
- Tatli H, Nüzhet Dalfes H, Sibel Menteş C (2004) A statistical downscaling method for monthly total precipitation over Turkey. *Int J Climatol* 24(2):161–180. <https://doi.org/10.1002/joc.997>
- Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. *Bull Am Meteorol Soc* 93(4):485–498
- Tebaldi C, Knutti R (2007) The use of the multi-model ensemble in probabilistic climate projections. *Philos Trans R Soc A: Math Phys Eng Sci* 365(1857):2053–2075. <https://doi.org/10.1098/rsta.2007.2076>
- Tebaldi C, Mearns LO, Nychka D, Smith RL (2004) Regional probabilities of precipitation change: a Bayesian analysis of multimodel simulations. *Geophys Res Lett* 31(24):L24213. <https://doi.org/10.1029/2004GL021276>
- Toreti A, Xoplaki E, Maraun D, Kuglitsch FG, Wanner H, Luterbacher J (2010) Characterisation of extreme winter precipitation in Mediterranean coastal sites and associated anomalous atmospheric circulation patterns. *Nat Hazards Earth Syst Sci* 10(5):1037–1050. <https://doi.org/10.5194/nhess-10-1037-2010>
- Tramblay Y, Badi W, Driouech F, El Adlouni S, Neppel L, Servat E (2012) Climate change impacts on extreme precipitation in Morocco. *Global Planet Change* 82–83:104–114. <https://doi.org/10.1016/j.gloplacha.2011.12.002>
- Trenberth KE, Dai A, Rasmussen RM, Parsons DB, Trenberth KE, Dai A, Rasmussen RM, Parsons DB (2003) The changing character of precipitation. *Bull Am Meteorol Soc* 84(9):1205–1217. <https://doi.org/10.1175/BAMS-84-9-1205>
- Trenberth K, Jones P, Ambenje P, Bojariu R, Easterling D, Klein Tank A, Parker D, Rahimzadeh F, Renwick J, Rusticucci M, Soden B, Zhai P (2007) Observations: surface and atmospheric climate change. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt K, Tignor M, Miller H (eds) *Climate change 2007. The physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change*. Cambridge University Press, Cambridge, pp 235–336. chap 3
- Trigo RM, Palutikof JP, Trigo RM, Palutikof JP (2001) Precipitation scenarios over Iberia: a comparison between direct GCM output and different downscaling techniques. *J Clim* 14(23):4422–4446. [https://doi.org/10.1175/1520-0442\(2001\)014<4422:PSOIAO>2.0.CO;2](https://doi.org/10.1175/1520-0442(2001)014<4422:PSOIAO>2.0.CO;2)
- Vrac M, Yiou P (2010) Weather regimes designed for local precipitation modeling: application to the Mediterranean basin. *J Geophys Res Atmos* 115(12):D12103. <https://doi.org/10.1029/2009JD012871>
- van Vuuren DP, Edmonds J, Kainuma M, Riahi K, Thomson A, Hibbard K, Hurtt GC, Kram T, Krey V, Lamarque JF, Masui T, Meinshausen M, Nakicenovic N, Smith SJ, Rose SK (2011) The representative concentration pathways: an overview. *Clim Change* 109:5–31
- von Storch H (1995) Inconsistencies at the interface of climate impact studies and global climate research. *Meteorol Z* 4(2):72–80. <https://doi.org/10.1127/metz/4/1992/72>
- von Storch H, Zorita E, Cubasch U, von Storch H, Zorita E, Cubasch U (1993) Downscaling of global climate change estimates to regional

- scales: an application to Iberian rainfall in wintertime. *J Clim* 6(6):1161–1171. [https://doi.org/10.1175/1520-0442\(1993\)006<1161:DOGCCE>2.0.CO;2](https://doi.org/10.1175/1520-0442(1993)006<1161:DOGCCE>2.0.CO;2)
- Wilby RL, Charles SP, Zorita E, Timbal B, Whetton P, Mearns LO (2004) Guidelines for use of climate scenarios developed from statistical downscaling methods. Supporting material of the intergovernmental panel on climate change, available from the DDC of IPCC TG CIA, 27
- Weigel AP, Knutti R, Liniger MA, Appenzeller C (2010) Risks of model weighting in multimodel climate projections. *J Climate* 23(15):4175–4191. <https://doi.org/10.1175/2010JCLI3594.1>
- Wenzel S, Cox PM, Eyring V, Friedlingstein P (2014) Emergent constraints on climate carbon cycle feedbacks in the CMIP5 Earth system models. *Geophys Res Biogeosci* 119:794–807. <https://doi.org/10.1002/2013JG002591>
- Wilks DS (2011) *Statistical methods in the atmospheric sciences*. Academic Press
- Xoplaki E, González-Rouco JF, Luterbacher J, Wanner H (2004) Wet season Mediterranean precipitation variability: influence of large-scale dynamics and trends. *Climate Dynam* 23(1):63–78. <https://doi.org/10.1007/s00382-004-0422-0>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Reproduced with permission of copyright owner. Further reproduction prohibited without permission.