

Climate Dynamics manuscript No.
(will be inserted by the editor)

Daily precipitation statistics in a EURO-CORDEX RCM ensemble: Added value of raw and bias-corrected high-resolution simulations

**A. Casanueva · S. Kotlarski · S.
Herrera · J. Fernández · J.M.
Gutiérrez · F. Boberg · A. Colette ·
O. B. Christensen · K. Goergen · D.
Jacob · K. Keuler · G. Nikulin · C.
Teichmann · R. Vautard**

Received: date / Accepted: date

A. Casanueva (✉) · S. Herrera · J. Fernández
Grupo de Meteorología. Dpto. Matemática Aplicada y Ciencias de la Computación. Univ.
de Cantabria, Avda. de los Castros, s/n, 39005 Santander, Spain
Tel.: +34 942201723
E-mail: ana.casanueva@unican.es

S. Kotlarski
Institute for Atmospheric and Climate Science, ETH Zurich, Switzerland

J.M. Gutiérrez
Grupo de Meteorología. Instituto de Física de Cantabria. CSIC-Univ. de Cantabria, Avda.
de los Castros, s/n, 39005 Santander, Spain

F. Boberg · O. B. Christensen
Climate and Arctic Research, Danish Meteorological Institute, Lyngbyvej 100, DK-2100
Copenhagen Ø, Denmark

A. Colette
Institut National de l'Environnement industriel et des risques (INERIS), Verneuil en Ha-
latte, France

K. Goergen
Meteorological Institute, University of Bonn, Auf dem Hügel 20, 53121 Bonn, Germany
(former Dpartement Environnement et Agro-Biotechnologies, Centre de Recherche Public
Gabriel Lippmann, 41 Rue du Brill, 4422 Belvaux, Luxembourg, since January 2015 re-
named to Luxembourg Science and Technology, LIST)

D. Jacob · C. Teichmann
Climate Service Center Germany (GERICS), Hamburg, Germany

D. Jacob
Max Planck Institute for Meteorology, Hamburg, Germany

K. Keuler
Chair of Environmental Meteorology, Brandenburg University of Technology (BTU),
Cottbus-Senftenberg, Germany

G. Nikulin
Swedish Meteorological and Hydrological Institute, Norrköping, Sweden

Abstract Daily precipitation statistics as simulated by the ERA-Interim-driven EURO-CORDEX regional climate model (RCM) ensemble are evaluated over two distinct regions of the European continent, namely the European Alps and Spain. The potential added value of the high-resolution 12 km experiments with respect to their 50 km resolution counterparts is investigated. The statistics considered consist of wet-day intensity and precipitation frequency as a measure of mean precipitation, and three precipitation-derived indicators (90th percentile on wet days —90pWET—, contribution of the very wet days to total precipitation —R95pTOT— and number of consecutive dry days —CDD—). As reference for model evaluation high resolution gridded observational data over continental Spain (Spain011/044) and the Alpine region (EURO4M-APGD) are used. The assessment and comparison of the two resolutions is accomplished not only on their original horizontal grids (approximately 12 km and 50 km), but the high-resolution RCMs are additionally regridded onto the coarse 50 km grid by grid cell aggregation for the direct comparison with the low resolution simulations.

The direct application of RCMs e.g. in many impact modelling studies is hampered by model biases. Therefore bias correction (BC) techniques are needed at both resolutions to ensure a better agreement between models and observations. In this work, the added value of the high resolution (before and after the bias correction) is assessed and the suitability of these BC methods is also discussed. Three basic BC methods are applied to isolate the effect of biases in mean precipitation, wet-day intensity and wet-day frequency on the derived indicators.

Daily precipitation percentiles are strongly affected by biases in the wet-day intensity, whereas the dry spells are better represented when the simulated precipitation frequency is adjusted to the observed one. This confirms that there is no single optimal way to correct for RCM biases, since correcting some distributional features typically leads to an improvement of some aspects but to a deterioration of others.

Regarding mean seasonal biases before the BC, we find only limited evidence for an added value of the higher resolution in the precipitation intensity and frequency or in the derived indicators. Thereby, evaluation results considerably depend on the RCM, season and indicator considered. High resolution simulations better reproduce the indicators' spatial patterns, especially in terms of spatial correlation. However, this improvement is not statistically significant after applying specific BC methods.

Keywords Regional Climate Models · EURO-CORDEX · added value · bias correction · precipitation indices

1 Introduction

Regional Climate Models (RCMs) are sophisticated tools that allow representing physical processes in the atmosphere that are not yet resolved by the coarse resolution of Global Climate Models (GCMs) (Giorgi, 2006; Feser et al, 2011). During the last decade, a huge effort has been made in order to adapt and apply these models to produce regional climate change scenarios in different regions worldwide. As a result, there is nowadays a number of comprehensive datasets developed in projects such as ENSEMBLES (van der Linden and Mitchell, 2009) and CORDEX (Giorgi et al, 2009), which also provide new opportunities for the intercomparison of different models, grid resolutions, boundary conditions, parameterizations (see e.g. Christensen et al, 1997; Jacob et al, 2007; Nikulin et al, 2011; García-Díez et al, 2013; Vautard et al, 2013) and model domains (Teichmann et al, 2013). For instance, the EURO-CORDEX initiative experiment design (European branch of CORDEX, <http://www.euro-cordex.net/>, see Jacob et al (2014) and Kotlarski et al (2014)) considers simulations at two horizontal resolutions, 0.44° and 0.11° . The latter is computationally very costly and its benefits have just recently been questioned by Prein et al (2015).

In principle, higher resolution experiments are able to capture features related to topography or land-sea mask, which are missed by coarser ones (Pryor et al, 2012; Walther et al, 2013); however, the added value of high resolution simulations is not always evident (Chan et al, 2013). Deficiencies may be caused for example by the fact that a given model is often developed and tuned in its low-resolution version (Gibelin and Déqué, 2003), therefore the high resolution cannot systematically improve the model performance. Several studies point out the importance of the right combination of parameterizations and horizontal (e.g. Déqué et al, 2005; Prein et al, 2013) and vertical resolution (Roeckner et al, 2006), highlighting for instance the role of the convection scheme (Kendon et al, 2012). Nevertheless, a single best parameterization for a specific resolution may not exist (Fernández et al, 2007; Jerez et al, 2012; García-Díez et al, 2015), and also depends on the particular application, i.e. the final use of the RCM simulation results. Thus, this situation supports the use of ensembles sampling different parameterizations and other model settings.

Model biases typically hamper the direct application of RCM output in impact studies (see e.g. Christensen et al, 2008; Kotlarski et al, 2014). Therefore, different bias correction (BC) methods were introduced in the literature (see e.g. Panofsky and Brier, 1968; Durman et al, 2001) and they have recently become increasingly popular in their application. BC methods vary from very simple factor scaling (additive or multiplicative, Durman et al (2001); Casanueva et al (2013)), to multi-variable BC techniques for particular combinations of variables (Wilcke et al, 2013; Vrac and Friederichs, 2015) and methods pursuing the correction of more sophisticated bias features, such as temperature-dependent biases (Boberg and Christensen, 2012; Bellprat et al, 2013). BC methods applied to precipitation have traditionally relied on the

46 assumption that models produce more rainy days than the reference observa-
47 tions (drizzle effect). There are also some BC methods dealing with dry-day
48 frequency overestimation, such as the frequency adaptation (Thiemeßl et al,
49 2012) or the Piani et al (2010) method, modified by Argüeso et al (2013) to
50 be used with station data.

51 Against the background outlined above, the aim of this study is to, first,
52 assess the added value of high (12 km) versus low (50 km) resolution RCM sim-
53 ulations regarding daily precipitation statistics. For this purpose, several pre-
54 cipitation derived-indicators (accounting for the mean and extreme regimes)
55 are evaluated in a EURO-CORDEX RCM ensemble. Secondly, since RCMs are
56 prone to systematic biases, BC methods are applied and the question whether
57 a potential added value of the raw high-resolution experiments with respect
58 to their low-resolution counterparts also remains after bias-correction is in-
59 vestigated. The considered BC methods consist of the adjustment of the first
60 moments of the precipitation distribution (mean precipitation, wet-day inten-
61 sity and wet-day frequency), which are applied separately to isolate the effect
62 of biases in precipitation amount and occurrence on precipitation derived indi-
63 cators. Note that we do not intend to provide the optimally bias-corrected data
64 —more sophisticated BC methods correcting the whole precipitation distribu-
65 tion would be needed—, but to attribute indicators’ biases to deficiencies in
66 the precipitation frequency (occurrence) and the intensity (amount). By doing
67 this, the basic precipitation features are investigated in depth to shed light on
68 the limitations and merits of both resolutions and to inform climate scenario
69 end users about undesired effects which may also affect more sophisticated BC
70 methods.

71 Several aspects can be analysed to assess the added value of high resolution
72 RCM simulation results. In this respect, a crucial question is the spatial scale
73 on which the evaluation and intercomparison is carried out. An evaluation on
74 the high resolution grid would penalize the coarse simulations because even
75 a perfect coarse simulation would miss sub-grid-scale features (Prein et al,
76 2015). For this reason, all comparisons are performed on the coarse resolution
77 (50 km), corresponding to the skillful scale of the 12 km experiments (Grasso,
78 2000). We consider this as the ‘fairest’ approach since it compares features
79 resolved by both resolutions. Nevertheless it is important to note that with
80 this choice we do not consider all aspects of the added value, e.g. the more
81 local information provided by the high resolution. See also Di Luca et al (2015)
82 for a comprehensive discussion about various definitions of added value.

83 A further note of caution relates to the fact that our study focuses on
84 several extreme precipitation indicators, such as the contribution of very wet
85 days (R95pTOT), dry spell lengths (maximum number of consecutive dry
86 days, CDD) and percentiles (90th percentile on wet days, 90pWET). These
87 indicators are very sensitive to the definition of a wet day, which is widely
88 discussed throughout the paper.

89 We consider two target areas in Europe: Continental Spain and the Euro-
90 pean Alps, where high-resolution and high-quality gridded observational data
91 sets are available for the evaluation and where previous versions of the same

RCMs have been examined (see e.g. Frei et al, 2003; Herrera et al, 2010). These areas cover a wide range of climatic conditions, from Mediterranean to Alpine climates, and orographic complexity.

Taking into account all the above, the specific objectives of the study are to

- examine the added value of high resolution simulations at the skillful scale (50 km) of the high resolution
- assess the added value of the high resolution simulations before and after bias correction
- provide some hints of possible implications of the results for more sophisticated BC methods.

This work is organized as follows. In Section 2 we present the data used. Section 3 introduces the methodology followed to evaluate the RCMs. The results are given in Section 4. Finally, the conclusion and the summary are given in Section 5.

2 Data

In the present study both high-resolution observational reference and RCM output data were used. All analyses were based on the common period 1989-2008. The study was performed on a seasonal basis, although only winter (DJF) and summer (JJA) results are shown for the sake of conciseness.

2.1 Observational Data

Observations play a major role in the evaluation and bias correction procedure and, as the RCM grid cells represent areal averages, gridded observational products are usually considered for the evaluation.

Over Spain, we used the new EURO-CORDEX-compliant, gridded observational data sets (Spain011/044; Herrera et al, 2015). More than 2700 quality-controlled stations were selected to develop these gridded precipitation data sets with 0.11° and 0.44° horizontal resolution, regular in a rotated longitude-latitude system, covering the period from 1971 to 2010. They were interpolated following trivariate thin plate splines (TPS) and ordinary kriging (AA-3D method in Herrera et al, 2015). This interpolation process is equivalent to the one used to build the European-scale E-OBS data set (Hofstra et al, 2009), considering orography as covariable in the formulation of the TPS. In order to ensure the area-averaged representativeness, the interpolation method was applied at an auxiliary 0.01° horizontal resolution grid and the final grids were obtained by averaging the results into the final resolution.

For the Alps, the Alpine Precipitation Grid Dataset (APGD, Isotta et al, 2013) was used as observational reference. This data set was developed in the

130 framework of EURO4M (European Reanalysis and Observations for Monitor-
 131 ing) for the period 1971-2008 and is a 5km resolution gridded product pro-
 132 vided by MeteoSwiss. The interpolation procedure consists of local regression
 133 (precipitation-elevation regression on independent slopes model) and angular
 134 distance weighting. In this study, the APGD was conservatively remapped onto
 135 the rotated 0.11° and 0.44° RCM grid, therefore the APGD011/044 versions
 136 exactly match the EURO-CORDEX grids (0.11° and 0.44°). We re-gridded
 137 from the original 5km resolution in a Lambert Azimuthal Equal Area Coordi-
 138 nate Reference System to 1km grid in a rotated longitude-latitude system and
 139 afterwards we averaged the values inside every EURO-CORDEX grid-cell in
 140 order to guarantee the representation of areal averages.

141 2.2 Regional Climate Models

142 We evaluated daily precipitation from the EURO-CORDEX RCMs integrated
 143 at horizontal resolutions of 0.11° and 0.44° on rotated grids (Table 1). These
 144 simulations were driven by the ERA-Interim reanalysis (Dee et al, 2011) and
 145 covered the period 1989-2008. We refer the reader to Table 1 in Vautard
 146 et al (2013) and Table 1 in Kotlarski et al (2014) for the model details. In
 147 those tables, WRF311A and WRF311F are referred to as WRF-CRPGL and
 148 WRF-IPSL-INERIS, respectively, and these two WRF setups apply different
 149 combinations of physical parameterization schemes (details in Kotlarski et al,
 150 2014). Most of these simulations are available via the Earth System Grid
 151 Federation (ESGF archive, <http://esgf.org/>) under the CORDEX initia-
 152 tive. Throughout this paper, the individual simulations are referred to as the
 153 name in the second column in Table 1 plus the resolution (e.g. HIRHAM011,
 154 HIRHAM044).

Table 1 EURO-CORDEX RCMs used in the present study. Codes were used to label RCMs in Figure 10. The last column indicates whether or not the respective RCM applies a smoothed surface orography.

Code	RCM	Institution	Orog. smoothed
1	CCLM	COSMO-CLM Community	Yes
2	HIRHAM	Danish Meteorological Institute, Denmark	No
3	RACMO	Royal Netherlands Meteorological Institute, Netherlands	Yes
4	RCA	Swedish Meteorological and Hydrological Institute, Sweden	Yes
5	REMO	Climate Service Center, Germany	No
6	WRF311A	CRP - Gabriel Lippmann, Luxembourg	No
7	WRF311F	Institut Pierre Simon Laplace / Institut National de l'Environnement Industriel et des Risques, France	No

3 Methodology

3.1 Precipitation Indices

Within the framework of the World Meteorological Organization, the Expert Team on Climate Change Detection and Indices (ETCCDI, <http://etccdi.pacificclimate.org/>) deals with the definition of climate indices in order to obtain comparable results worldwide. Based on these definitions we here used seasonal precipitation indices derived from daily precipitation amounts (Table 2).

Table 2 Precipitation and derived indices used in this study.

ID	Indicator	Units
RR	Daily precipitation amount	mm/day
SDII	Simple day intensity index (mean wet day precipitation)	mm/day
RR1	Wet-day frequency	%
90pWET	90th percentile on wet days	mm
R95pTOT	Percentage of total precipitation contributed by 5% most rainy days	%
CDD	Maximum number of consecutive dry days	days

SDII, RR and RR1 account for the mean precipitation regime, whereas 90pWET and R95pTOT are considered extreme indices in the sense that they are related to the tails of the probability distribution function, even though they are not associated to rare events. R95pTOT measures the contribution of heavy precipitation events to total precipitation. For Spain, this indicator clearly separates the different extreme regimes of the Atlantic and Mediterranean climates (see, for example, Fig. 10 in Herrera et al, 2012). CDD quantifies dry spells and is linked to precipitation occurrence. CDD and R95pTOT also present different driving mechanisms: CDD is more related to large-scale atmospheric circulation while R95pTOT has a convective origin and depends more on local processes and moisture fluxes (Casanueva et al, 2014).

As recommended by Orłowsky and Seneviratne (2012), 90pWET and R95pTOT were derived over the entire period, while CDD was calculated for each year and season before computing the median for all years.

Figure 1 shows the seasonal observed values for the indices in Table 2 for both regions as represented by APGD011 and Spain011. Note that in both regions, the spatial pattern has a paramount orographic component, especially in winter.

3.2 Aggregation Procedure

In order to examine the added value of high resolution simulations at their skillful scale (0.44°), a comparison between the evaluation of the coarse and high resolutions with respect to the coarse resolution observations (Spain044/APGD044)

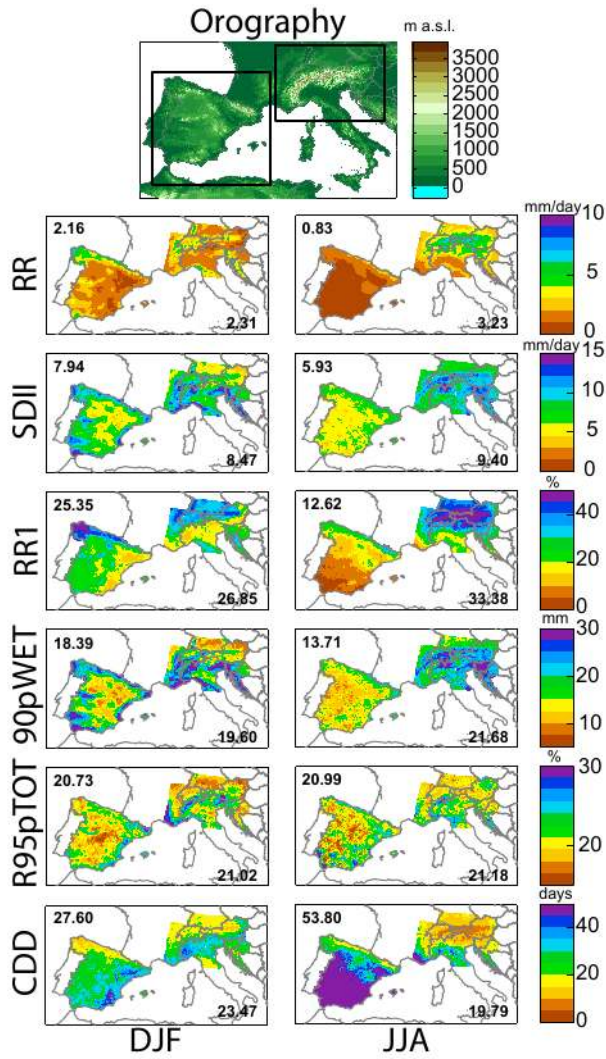


Fig. 1 Seasonal observed distribution of the RR, SDII, RR1, 90pWET, R95pTOT and CDD in winter (left) and summer (right), according to the APGD011/Spain011 datasets. The numbers are the spatial averages in both regions.

185 was performed. Also, the standard evaluation of the high resolution (i.e. RCMs
 186 at 0.11° with respect to Spain011/APGD011) is shown for illustrative pur-
 187 poses. Thus, from now on we refer to the individual resolutions as three in-
 188 dependent datasets: 0.44° (original simulation), 0.11AGG (0.11° simulation
 189 aggregated to 0.44° resolution) and 0.11° (original simulation). By construc-
 190 tion, the 0.44° and 0.11° EURO-CORDEX grids match each other at the grid
 191 cell boundaries, i.e. each 0.44° grid cell contains exactly 16 0.11° grid cells. We

192 obtained the aggregated data (0.11AGG) by spatially averaging the grid boxes
 193 from the 0.11° grid that belonged to each 0.44° grid box (i.e. 16 0.11° grid
 194 cells were spatially averaged for each 0.44° grid box). Firstly, we obtained the
 195 011AGG data from the original 0.11° simulations for each RCM, and secondly
 196 we calculated the derived indices from that aggregated data. By means of
 197 this procedure we address the added value of the high resolution at its skillful
 198 scale (Grasso, 2000), therefore grid-point details (not only related to topogra-
 199 phy and land-sea mask, but also to better resolved local processes) from the
 200 high resolution cannot be discerned, but may be still present after smoothing
 201 them onto the coarse resolution. Note that 0.44° and 0.11AGG were defined
 202 in the same 0.44° EURO-CORDEX grid and, therefore, they can be directly
 203 compared with respect to the same observations (APGD044/Spain044).

204 The comparison of 0.44° and 0.11AGG on the coarse grid can be considered
 205 ‘fair’ since both resolutions are able to resolve the analysed features, however
 206 this is not a unique way to assess the added value. Comparing both resolutions
 207 on the high resolution grid would penalize the coarse resolution since, even for
 208 a perfect simulation, some sub-grid-scale features are missing. According to
 209 Prein et al (2015) added value of high resolution is more evident in the evalu-
 210 ation on the high resolution grid since more fine-scale processes are captured.
 211 Note that a completely fair comparison would also imply to perform the evalu-
 212 ation exercise at the skillful scale of the coarse resolution experiments (for
 213 instance at $4 \times 0.44^\circ = 1.76^\circ$), otherwise one resolution is always punished. This,
 214 however, would considerably smooth the spatial precipitation fields and also
 215 precipitation extremes. We here refrain from doing so but acknowledge that
 216 the 0.11° experiments might be slightly favored in our evaluation setup.

217 3.3 Assessment of simple bias correction techniques

218 Additionally to the evaluation of the raw RCM outputs, we assessed the results
 219 of three simple BC techniques for precipitation frequency and intensity and
 220 evaluated their effect on precipitation indices. The indicators considered in
 221 this work depend on precipitation occurrence and/or amount (see Table 2),
 222 therefore their biases can be attributed to deficiencies in the precipitation
 223 frequency and/or intensity. The three corrections considered were performed
 224 seasonally. The first one was based on mean precipitation (considering rainy
 225 and non-rainy days), while the others isolated the effect of the precipitation
 226 amount (how much) and occurrence (how often), respectively.

227 First, rainfall data were corrected using a multiplicative scaling factor ob-
 228 tained as the quotient of the observed and simulated spatially averaged pre-
 229 cipitation over the specific region (from now on denoted as *global scaling, GS*):

$$RR_{GS} = RR_{RCM} \frac{\overline{\langle RR_{OBS} \rangle}}{\overline{\langle RR_{RCM} \rangle}} \quad (1)$$

230 where RR_{RCM} represents daily RCM precipitation at an individual grid point
 231 and the overline and angle brackets represent temporal and spatial averages,

232 respectively. This scaling reduces the bulky systematic biases present in every
 233 RCM and can be considered as a minimum correction needed at both resolu-
 234 tions. The correction factor is the same for all grid points in a single model,
 235 resolution and season, it does not depend on the grid box but on the spatial
 236 mean precipitation for the respective analysis domain. To some extent, this
 237 bulk correction could mimic a (global) retuning of the model to better fit ob-
 238 servations. Note that this correction would work well for overall too dry or
 239 too wet models, because it implies constant biases across grid points. As the
 240 precipitation spatial pattern presents very high variability and local features,
 241 some further corrections were needed at grid point level.

242 Second, a *local scaling* (LS) was applied at a grid point level considering
 243 the quotient of the observed and simulated wet-day precipitation:

$$RR_{LS} = RR_{RCM} \frac{SDII_{OBS}}{SDII_{RCM}} \quad (2)$$

244 Note that these two corrections, GS and LS, do not alter the values of the
 245 indicators R95pTOT and CDD.

246 Third, in addition to these corrections related to the precipitation amount,
 247 a correction concerning the precipitation frequency was applied. Many indi-
 248 cators (including those in Table 2) consider the wet day frequency (RR1) in
 249 their definitions. Thus an over/underestimation of RR1 would inevitably lead
 250 to biases in the derived indicators. A dry (wet) day is defined as a day with
 251 precipitation below (above) a given threshold. In the recent literature, the
 252 analysis of dry/wet days of observed precipitation and climate model output
 253 normally uses subjectively-selected rainfall thresholds (often 0.1mm or 1mm)
 254 to separate dry and wet days (see e.g. Lázaro et al, 2001; Herrera et al, 2010;
 255 Orłowsky and Seneviratne, 2012). Some studies suggest the use of alternative
 256 wet/dry day thresholds different to the usually accepted such as 10mm (Yoo
 257 et al, 2001), 5 and 10mm (Fdez-Arroyabe Hernández and Martín-Vide, 2012) or
 258 the amount exceeded by 96% of the total rainfall (Aviad et al, 2013). Barring
 259 et al (2006) find an optimal (according to several statistics) wet-day thresh-
 260 old for the whole of Europe of 0.56 and 1.20mm (for two model versions of a
 261 specific RCM), for the reference threshold of 1mm in point observation series.
 262 Selecting a single optimal threshold for the whole of Europe is a compromise,
 263 since this threshold depends on the location. In this study, we estimated an
 264 *adjusted wet-day threshold* P^* at grid point level by selecting the wet-day
 265 threshold value in the RCM which matches the observed wet-day frequency
 266 (RR1) computed with a 1mm threshold:

$$P^* = F_{RCM}^{-1}(F_{OBS}(1mm)) \quad (3)$$

267 where F is the empirical cumulative distribution function (CDF), so F_{RCM}
 268 and F_{OBS} refer to the simulated and observed CDFs, respectively. This value
 269 could be different from grid cell to grid cell and was derived separately for
 270 each RCM, resolution and season. From now on we denote by X_{FA} a given
 271 indicator after the frequency adjustment (FA) is applied, i.e. P^* is used for

272 the indicator calculation instead of 1mm. Every precipitation value above this
273 adjusted threshold was assumed to be a wet day in the RCM simulation,
274 otherwise the day was considered dry. After this adjustment in the wet-day
275 threshold, the observed RR1 is perfectly reproduced by the simulation and the
276 contribution of the frequency to the biases in the derived indicators can be
277 isolated.

278 Bear in mind that FA with very large ($P^* \gg 1\text{mm}$) and close to zero
279 ($P^* \ll 1\text{mm}$) P^* values may lead to non reliable results, especially considering
280 its application to impact studies. In the former case, FA and more sophisti-
281 cated methods adjusting the wet-day frequency —such as quantile mapping
282 (Panofsky and Brier, 1968)— can deal with $P^* \gg 1\text{mm}$, but at the expense
283 of mapping P^* values into 1mm. For $P^* < 1\text{mm}$, the FA itself is not able to
284 provide an optimal correction since the model is drier than observations and
285 it cannot ‘invent’ wet days (Barring et al, 2006). For this reason, sophisticated
286 BC methods have included additional methods such as the frequency adapta-
287 tion by randomly sampling the observational distribution into the simulated
288 first bin (Thiemeßl et al, 2012; Wilcke et al, 2013).

289 The perfect representation of the wet-day frequency does not necessarily
290 lead to reduced biases in precipitation threshold-dependent indicators, espe-
291 cially in the two cases mentioned above. In the following sections we address
292 the effect of considering P^* instead of 1mm on the indicators’ biases at the
293 different resolutions.

294 A fourth correction was considered combining the local scaling and the
295 frequency adjustment. Thus, we locally scaled the daily precipitation after
296 adjusting the wet-day threshold (i.e. Y_{LS} where $Y = X_{FA}$, as Schmidli et al,
297 2006). Results were similar to the LS case, therefore, this correction is not
298 shown in the paper.

299 4 Results

300 4.1 Added value in mean precipitation

301 Mean precipitation consists of the combination of the daily intensity and wet-
302 day frequency. We analyse the contribution of both components separately in
303 order to account for their effect on biases in precipitation-derived indicators
304 (Sect. 4.2). Biases in the precipitation intensity (SDII) are shown in Figures
305 2 and 3 for winter and summer, respectively, using the 1mm threshold for the
306 wet-day definition (same for the wet-day frequency —RR1— in the supple-
307 mentary material, Fig.S1-S2). The fourth column represents the difference of
308 the bias on the 0.44° grid minus 0.11AGG, both in absolute values. Thus, posi-
309 tive differences (greenish colours) show added value of the 0.11AGG and the
310 opposite for negative differences (brownish colours). There is no overall added
311 value of the high resolution simulations aggregated to the 0.44° grid since,
312 depending on the model and season, biases are smaller for one resolution or
313 the other (in agreement with Kotlarski et al, 2014). Due to the averaging

314 procedure, in most cases, 011AGG presents smoother patterns than 0.11° . In
 315 winter (Fig. 2), there is a clear orographic pattern in the bias of both regions
 316 with some improvements of the high resolution for WRF311A and WRF311F,
 317 whereas HIRHAM, RCA and REMO present the highest positive biases at
 318 both resolutions for at least one of the regions. In summer (Fig. 3), there is no
 319 common spatial bias pattern in both regions. CCLM and RCA considerably
 320 reduce biases in the high resolution (especially in the Alpine region) whereas
 321 WRF311A and WRF311F present negligible differences between both resolu-
 322 tions. In both seasons, opposite-sign biases at 0.44° and 0.11° AGG are found in
 323 some areas, more noticeable for CCLM in winter and RCA and REMO (also
 324 HIRHAM for Spain) in summer. This means that the same parameterizations
 325 with different resolutions lead to different precipitation intensities and also
 326 different spatial patterns. REMO stands out in summer, since both regions
 327 present mainly wet biases at 0.44° , but dry biases predominate at 0.11° and,
 328 therefore, at 011AGG. Apparently, some physical schemes seem more resolu-
 329 tion dependent than others (Déqué et al, 2005). Further research about the
 330 sensitivity of seasonal biases to the different schemes (see e.g. García-Díez
 331 et al, 2013) should be performed in the specific RCMs and resolutions. Note
 332 also the patchy spatial pattern in HIRHAM011, RCA011 and REMO011 with
 333 strong, opposite biases in nearby grid boxes, i.e. there is not a gradual change
 334 across the zero bias between opposite sign biases. For HIRHAM and REMO
 335 this could be due to the use of non-smoothed orography (see the fourth col-
 336 umn in Table 1). Previous studies have also associated biases to the excessively
 337 smoothed (Shkolnik and Efimov, 2013) or non-smoothed topography (Polanski
 338 et al, 2010), being the orography another factor to take into account in RCMs
 339 evaluation.

340 Precipitation occurrence is characterized in terms of the wet-day frequency,
 341 which depends on the particular threshold (e.g. 1mm) used to define a wet day.
 342 Figure 4 shows the q-q plot for three selected grid points in RCA011 for winter
 343 daily precipitation (black crosses). The vertical line corresponds to 1mm in
 344 the observations. Therefore the intersection with the q-q plot provides the ad-
 345 justed threshold equivalent to 1mm in the observations (P^* , see Eq.3). On the
 346 left, the model overestimates the wet-day frequency and P^* is around 16mm
 347 (intersection of black crosses with the vertical line). The center panel corre-
 348 sponds to a grid point where P^* approximates 1mm. On the right, the model
 349 presents more dry days than observed ($P^* < 1\text{mm}$). This figure illustrates
 350 that adjusted thresholds are in some cases very far from 1mm (left), but also
 351 close to 0mm (right). Only the points with $P^* \approx 1\text{mm}$ (center) would work
 352 well with the usually accepted 1mm threshold.

353 When defining wet days using P^* instead of 1mm, precipitation intensity
 354 (SDII) is also altered, since it is defined as the mean of the wet-day precip-
 355 itation. For $P^* > 1\text{mm}$, SDII would be shifted towards higher values, since
 356 low-precipitation values in the range (1mm, P^*) are considered dry days. For
 357 $P^* < 1\text{mm}$, dry SDII biases are expected due to the contribution of many
 358 close-to-zero precipitation values regarded as wet days.

359 Figures 5 and 6 show the spatial pattern of P^* computed according to Eq. 3
360 for winter and summer, respectively. In winter, P^* is higher in places with
361 complex orography (the Alps and Pyrenees) and lower in lowlands, especially
362 for HIRHAM and RCA. In Spain, P^* does not reach as large values as in the
363 Alpine region. Smaller P^* is also found in summer, with many zero or close
364 to zero values (e.g. CCLM, HIRHAM, RCA especially in Spain). Notice that
365 the definition of wet days depends on P^* and, therefore, the indicators (SDII,
366 CDD, 90pWET and R95pTOT) are expected to change as the threshold P^*
367 changes. Hence, the evaluation results depend on the value which is chosen as
368 the wet-day threshold, which is usually subjectively adopted (e.g. commonly
369 1mm or 0.1mm) or adjusted with the observations, as done in this study.

370 Thus far, EURO-CORDEX RCMs present biases in SDII and RR1 at both
371 resolutions that will propagate to the derived indices. BC methods allow to
372 statistically correct these deviations, but the underlying physical misrepresenta-
373 tions will remain and may still affect higher-order moments of the corrected
374 variables. Large P^* also affects sophisticated BC methods. For instance, P^*
375 (Figs. 5 and 6) determines the highest (or lowest) value which is mapped into
376 1mm when applying a quantile mapping (Panofsky and Brier, 1968) using the
377 standard 1mm wet-day threshold. The patchy spatial pattern shown before for
378 SDII (Figs. 2 and 3) is also found in the HIRHAM011, RCA011 and REMO011
379 adjusted thresholds (Figs. 5 and 6). This may lead to spatial inconsistencies
380 in sophisticated BC techniques, since two nearby grid points can have very
381 different P^* . These aspects constitute a theoretical discussion and need to be
382 proven and analysed in further studies. No patchy spatial pattern is noticeable
383 at the 011AGG scale, due to the underlying averaging procedure. Therefore,
384 corrections in the frequency should be accomplished at the coarse scale, where
385 no spatial inconsistencies are found and 0.44° and 011AGG present compa-
386 rable results in terms of spatial coherence. Conversely, at 0.11° resolution,
387 WRF and CCLM present more spatially coherent and smoothed patterns and
388 RACMO stands out with P^* close to 1mm, especially over large parts of Spain.

389 4.2 Added value in precipitation derived indicators

390 We now examine the added value of the high resolution experiments for precipitation-
391 derived indicators (low versus high resolution) and account for the effect of
392 the biases in SDII and RR1 on derived indicators by applying BC methods at
393 both resolutions (raw versus corrected data). All indicators considered (Table
394 2) depend on the wet-day threshold. Thus, they are calculated with the 1mm
395 and P^* (FA, Sect.3.3). 90pWET is also affected by the precipitation amount,
396 therefore GS and LS corrections are performed. Biases for 90pWET and CDD
397 (relative) and R95pTOT (absolute) are obtained for the raw and the corrected
398 data for 0.44° and 011AGG.

399 *90pWET*

400 Results for the 90th percentile on wet days (90pWET) are summarized in
 401 Figure 7. The ‘original’ label refers to the case when no correction is per-
 402 formed, i.e. without any scaling and taking 1mm as the wet-day threshold.
 403 Global scaling does not lead to overall conclusions, it usually reduces very
 404 high biases but deteriorates smaller ones. Local scaling strongly reduces the
 405 biases, along with their spatial variability, in every RCM and resolution. This
 406 results in individual grid-point biases of similar magnitude at both resolutions
 407 (median markers close to zero, and similar-sized boxes). This means that any
 408 improvement of a resolution with respect to the other before the correction
 409 does not necessarily lead to an improvement after the local scaling since biases
 410 become comparable for both resolutions (see e.g. WRF331F in Spain). Less
 411 frequently, an improvement before the correction remains (see e.g. RCA in
 412 Spain) or turns into a deterioration (see e.g. RACMO in Spain in winter) after
 413 the LS, although these are subtle changes. The application of the frequency
 414 adjustment (using P^* as the wet-day threshold) leads to similar biases to those
 415 in the original case (using 1mm wet-day threshold). Notice that especially for
 416 $P^* > 1\text{mm}$, changing the threshold yields slightly different percentiles, while
 417 the scaling changes the indicator more rapidly (see dots in Fig. 4, representing
 418 the deciles from the wet day distribution for the original (black), FA (blue)
 419 and LS (red)).

420 Regarding the added value, neither the original nor the corrected data lead
 421 to an overall improvement of one resolution against the other since results are
 422 similar for both resolutions and the best performance is obtained for 0.44° or
 423 0.11AGG depending on the specific case.

424 *R95pTOT*

425 Figure 8 summarizes the absolute biases for the contribution of very wet days
 426 (R95pTOT). In winter, negligible differences are found between resolutions in
 427 both regions and few changes are obtained after the frequency adjustment.
 428 The correction can cause a small improvement (e.g. RCA in Spain) or deterio-
 429 ration (e.g. CCLM in the Alpine region). Hence the R95pTOT involves more
 430 processes (related to precipitation intensity) responsible for the biases that
 431 cannot be attributed to the wet-day frequency. In summer, the 0.44° simu-
 432 lation in the original case in the Alpine region is slightly better or does not
 433 differ much from 0.11AGG except for CCLM. After the frequency adjustment
 434 CCLM, HIRHAM and RCA biases increase dramatically on the coarse reso-
 435 lution. This could be due to the very low P^* values (see Fig. 6), leading to an
 436 increase of this indicator and therefore to very large positive biases.

437 *CDD*

438 For the number of consecutive dry days (CDD, Fig. 9) in winter the frequency
 439 adjustment tends to reduce biases and diminish the differences between reso-

440 lutions in Spain, but slightly benefits 0.11AGG in the Alpine region. Relative
441 biases for summer CDD are very large for CCLM and HIRHAM at both reso-
442 lutions —related to their large negative biases in the frequency (see Fig.S2)—,
443 showing their difficulties to properly represent the lower part of the precip-
444 itation distribution and the temporal coherence. The frequency adjustment
445 substantially reduces these biases, meaning that the spells are better captured
446 when the wet-day frequency is adjusted to the observed one. Unlike CCLM
447 and HIRHAM, RCA does not reduce its large bias in summer in the Alpine
448 region after the frequency adjustment and a large negative bias remains. This
449 is due to the zero values of the wet-day adjusted thresholds that are apparent
450 over approximately one third of the Alpine domain (black grid boxes in Figure
451 6). Since the wet-day threshold is exactly zero, there are no dry spells in these
452 grid boxes, leading to large negative biases of CDD.

453 *Joint discussion*

454 The above results show that 90pWET is more sensitive to the intensity whereas
455 the CDD is affected by the wet-day threshold. Thus, they also present differ-
456 ent sensitivities to the local scaling and frequency adjustment. On the one
457 hand, the frequency adjustment slightly changes the 90pWET (i.e. changing
458 the threshold yields a slightly different percentile). This means that the precip-
459 itation frequency (i.e. the lower tail of the precipitation distribution) does not
460 have a systematic implication for the upper percentiles (i.e. upper tail based
461 indices). In some models the correction can either make it slightly better or
462 worse, but in very dry models the correction can even strongly deteriorate it
463 (e.g. CCLM and RCA in summer). Precipitation intensity, however, plays a
464 major role in 90pWET; when the precipitation distribution is scaled by SDII,
465 the upper percentiles are scaled too (in agreement with Benestad et al, 2012).
466 On the other hand, the frequency adjustment considerably reduces CDD bi-
467 ases. Once the observed and simulated wet-day frequencies are equal, the RCM
468 better captures the dry spell durations. As mentioned before, in some cases
469 (see e.g. RCA in summer in the Alpine region in Fig. 9) frequency adjustment
470 does not reduce CDD biases because the time series autocorrelation and persis-
471 tence (and therefore occurrence) of specific situations are not well represented
472 by the model and the correction is not able to resolve this. For this reason, the
473 frequency adjustment deteriorates biases in that example for 90pWET and
474 R95pTOT too. Unlike 90pWET and CDD, which are more sensitive to the
475 intensity and frequency, respectively, for R95pTOT the frequency adjustment
476 can either deteriorate or not affect the results. The definition of R95pTOT in-
477 cludes both the intensity and the frequency, and correcting for biases of these
478 two aspects can have converse effects.

479 As shown in Figures 7, 8 and 9, the selected indicators are affected by
480 very low summer P^* . In this case, 90pWET (CDD) is lower (higher) because
481 of too many zero-precipitation values. R95pTOT is based on a lower 95th
482 percentile, resulting in a higher quotient of the contribution of very wet days
483 relative to the total wet-day precipitation amount. This is a limitation in the

484 frequency adjustment; when RCMs are too dry, an optimal threshold does not
485 necessarily lead to an improvement because the correction cannot ‘invent’ wet
486 days (Barring et al, 2006). This should be carefully examined since basic bias
487 correction techniques are not able to solve this problem and more sophisticated
488 techniques are required to deal with this issue (Thiemeßl et al, 2012; Wilcke
489 et al, 2013).

490 The correction methods used in this study are not able to correct all in-
491 dicators at a time. The precipitation occurrence affects the indices related to
492 spells rather than the upper-tail percentiles, which are more influenced by
493 precipitation intensity. That result suggests that there is not a single optimal,
494 best way of bias correcting RCMs, since methods adjusting the frequency bet-
495 ter represent the CDD, but can deteriorate the upper tail percentiles. Further
496 analyses should be performed to quantify this result in more sophisticated bias
497 correction methods combining several corrections. Note that we use the same
498 period for the calibration and validation of the BC methods, since we validate
499 aspects that are not directly tackled by the BC procedure. Validation results
500 might look worse if an independent validation period would be chosen.

501 4.3 Added value and bias correction effect on the spatial pattern

502 The correction methods applied in the previous section preserve the temporal
503 structure of the data and in this section we analyse their effect on the spatial
504 pattern. We only show results for the 0.44° and 0.11AGG data sets, since these
505 can be directly compared against the same observations.

506 The ability to represent the spatial structure is summarized by means of
507 Taylor (2001) diagrams (Fig. 10). These show several spatial scores at a time:
508 spatial Pearson correlation coefficient (r), Centered Root Mean Squared Dif-
509 ference (RMSD), standard deviation (std) and biases of spatial averages ($bias$).
510 Arrows join, for a given RCM, the 0.44 score (squares) with the 011AGG (cir-
511 cles). Therefore, arrows pointing towards the observational reference indicate
512 that the high-resolution runs (011AGG) perform better than the coarse ones
513 (0.44°). To summarize the added value of 0.11AGG, the bars in the right pan-
514 els show the number of RCMs (from 0 to 7) in which the 0.11AGG improves
515 with respect to the 0.44° resolution, in terms of the four statistics shown in the
516 Taylor diagram. In these barplots, the numbers of the right Y-axis show the
517 statistical significance of the existence of added value, obtained by a Z-test for
518 proportions. The results are statistically significant only when 6 or 7 RCMs
519 (and symmetrically for 0 and 1) out of 7 improve upon the 0.44° resolution
520 runs. For the raw RCM output (Fig. 10, first row), high-resolution RCMs gen-
521 erally perform better than the coarse counterparts, especially in terms of r and
522 RMSD ($bias$ and std are not conclusive). After the corrections (Fig. 10, second
523 row), all the RCMs present similar validation scores for CDD and 90pWET re-
524 gardless of the resolution (the squares are close to the circles). The proportion
525 of the 0.11AGG RCMs that improves with respect to the coarse ones is not
526 statistically significant after the corrections. R95pTOT deteriorates with the

527 frequency adjustment except for CCLM and RACMO (labels 1 and 3, respec-
528 tively) since these two RCMs present P^* close to 1mm (as shown in Fig.5).
529 This degradation after applying the correction was also shown in the previous
530 section, meaning that this indicator is not favoured by the frequency adjust-
531 ment. The results for summer and for Spain in both seasons lead to similar
532 conclusions and are included in the supplementary material (Fig.S3-S5).

533 5 Conclusions

534 This paper evaluates daily precipitation characteristics in the ERA-Interim-
535 driven EURO-CORDEX RCM ensemble. Experiments at both 0.11° and 0.44°
536 horizontal resolution are considered, and the potential added value of the 0.11°
537 simulations is addressed. For this purpose, high-resolution RCMs are regrid-
538 ded onto the coarse grid by grid cell aggregation (0.11AGG). The analysis is
539 performed in two regions of Europe where high quality gridded observational
540 data sets are available (continental Spain and the Alpine region) consider-
541 ing mean precipitation and derived indicators (90th percentile on wet days
542 —90pWET—, contribution of the very wet days —R95pTOT— and number
543 of consecutive dry days —CDD—).

544 In terms of seasonal biases we find only limited evidence for an added value
545 of the higher resolution in the precipitation intensity, wet-day frequency and
546 derived indicators, since results depend on the RCM, season and indicator and
547 small differences rather randomly favour the 0.44° or the 0.11AGG resolutions.
548 We find added value of the high resolution simulations in the spatial pattern
549 (especially in correlation and RMSD). To adequately represent daily precip-
550 itation statistics, bias correction techniques are needed at both resolutions.
551 Nevertheless, after applying simple bias correction techniques the proportion
552 of the 0.11AGG RCMs that improves with respect to the coarse ones is not
553 statistically significant and there are negligible differences between resolutions.

554 Note that we only partly address the potential added value, since high
555 resolution simulations are not only used with the intention to improve the
556 larger scale processes and features but also in order to provide better local
557 information (i.e. the local departures of the 0.11° relative to the 0.11AGG or
558 0.44° simulation results should be better than a random information). Our
559 validation on the coarse grid can be considered ‘fair’ because it compares
560 features resolved by both resolutions, however it is not the unique way to assess
561 the added value (Di Luca et al, 2015). Pursuing a fairer comparison between
562 resolutions would also require the retuning of the high resolution simulations
563 and the consideration of the 0.44° simulations at their skillful scale. Prein et al
564 (2015) claim that the added value of the high resolution is more evident when
565 the comparison is performed on the high resolution grid but acknowledge that
566 this procedure benefits the high resolution runs.

567 The present work and Prein et al (2015) try to shed light on the added
568 value issue by analysing different precipitation aspects taking into account
569 precipitation-derived indices related to the intensity, frequency and extremes,

570 and from daily to sub-daily scales. As such, both works are complementary to
571 each other. Both show the benefits of the high resolution in spatial patterns,
572 however we find no statistically significant results after bias correction. Prein
573 et al (2015) identify Spain and the Alpine area as the regions with largest net-
574 improved-area fractions (i.e. fraction of grid boxes in which more than 75% of
575 the high resolution simulations improve on the coarse counterparts). However
576 that fraction is never higher than 50% of improvement, meaning that more
577 than half of the grid points in each region is deteriorated or (in most of the
578 cases) is similar in both resolutions. From our results, added value of the high
579 resolution on seasonal mean biases depends on the indicator, RCM and season
580 and the best performance is obtained for 0.44° or 0.11AGG depending on
581 the specific case. Both studies are focused on different indicators for extreme
582 precipitation, thus making it difficult to intercompare them. While Prein et al
583 (2015) find also added value in other aspects such as the sub-daily scale, we
584 focus on the added value of bias corrected simulations, which could be of great
585 interest for the impact community.

586 We apply three simple bias correction methods based on the correction of
587 the first moments of the precipitation distribution. First, results show that
588 scaling by the quotient between observed and simulated spatial mean precipi-
589 tation is not enough to reduce biases in the 90pWET. Second, the local scaling
590 with the wet-day intensity reduces the 90pWET biases dramatically (i.e. cor-
591 recting the mean also corrects the percentiles) and makes both resolutions
592 comparable after the correction. Third, the frequency adjustment improves
593 the lower part of the probability distribution function (better representation
594 of the CDD) but it deteriorates the upper tails (worse or negligible changes
595 in the 90pWET and R95pTOT). Therefore, these corrections do not show an
596 overall improvement which strongly indicates that there is no single optimal
597 way to correct for RCM biases. Users should make a choice for one bias correc-
598 tion method or the other depending on the precipitation metric being assessed
599 (e.g. local scaling is more efficient for percentiles and the frequency adjust-
600 ment for dry spells), but being aware that the same method can at the same
601 time deteriorate another feature of the distribution. This emphasizes the need
602 to investigate more sophisticated bias correction methods that correct several
603 aspects at a time.

604 Large biases remain in the derived indicators after the frequency adjust-
605 ment when the adjusted threshold is zero (see e.g. CCLM in Spain and RCA
606 in the Alpine region in summer). Bias correction has traditionally relied on the
607 assumption that models produce more rainy days than the reference observa-
608 tions and these methods work well for finding optimized thresholds when the
609 climate model overestimates the number of wet days by frequently simulating
610 light rainfall. However, the procedure cannot improve the opposite situation
611 because it cannot ‘invent’ wet days if the model is too dry (in agreement with
612 Barring et al, 2006). This problem is very noticeable in summer and further
613 research is needed (e.g., along the lines of the frequency adaptation from The-
614 meßl et al, 2012) since it is not possible to fully solve it with the basic bias
615 correction techniques applied in the present work.

616 Sophisticated bias correction methods are well prepared to solve any prob-
617 lem in the precipitation Probability Density Function. In this work we identify
618 some shortcomings (e.g. deficiencies in the representation of the wet-day fre-
619 quency) in specific RCM simulations that may have implications for the suit-
620 ability for such methods. For instance, some RCMs at 0.11° resolution present
621 very high P^* (caused by strong biases in the precipitation frequency). Some
622 sophisticated methods (e.g. quantile mapping) map this large P^* onto 1mm
623 and values in the range (1mm, P^*) onto dry days. There are also spatial in-
624 consistencies in some models at the high resolution which might be related to
625 instabilities due the use of non-smoothed orographies and a spatial displace-
626 ment of precipitation structures (Maraun and Widmann, 2015). This could
627 give unreliable results after applying single-site bias correction methods (i.e.
628 point-wise methods, not considering the spatial coherence).

629 This study gives insight into the daily precipitation statistics in the EURO-
630 CORDEX RCM ensemble by analysing each ensemble member individually.
631 Better agreement with the observations is usually found when ensemble aver-
632 ages are validated and, moreover, this improves when only the best performing
633 models are considered (Herrera et al, 2010). Bearing this in mind more efforts
634 should be devoted towards the improvement of the individual models in order
635 to avoid very strong biases. For this purpose, further research about the im-
636 pact of different parameterization schemes on seasonal biases as García-Díez
637 et al (2013) should be performed for the specific RCMs.

638 **Acknowledgements** The authors are grateful to Prof. C. Schär for his helpful comments
639 and E. van Meijgaard for making available the RACMO model data. We acknowledge the
640 observational data providers. Calculations for WRF311F were made using the TGCC su-
641 per computers under the GENCI time allocation GEN6877. The WRF331A from CRP-GL
642 (now LIST) was funded by the Luxembourg National Research Fund (FNR) through grant
643 FNR C09/SR/16 (CLIMPACT). The KNMI-RACMO2 simulations were supported by the
644 Dutch Ministry of Infrastructure and the Environment. The CCLM and REMO simu-
645 lations were supported by the Federal Ministry of Education and Research (BMBF) and
646 performed under the Konsortial share at the German Climate Computing Centre (DKRZ).
647 The CCLM simulations were furthermore supported by the Swiss National Supercomputing
648 Centre (CSCS) under project ID s78. Part of the SMHI contribution was carried out in
649 the Swedish Mistra-SWECIA programme founded by Mistra (the Foundation for Strategic
650 Environmental Research). This work is supported by CORWES (CGL2010-22158-C02) and
651 EXTREMBLES (CGL2010-21869) projects funded by the Spanish R&D programme and
652 the European COST ACTION VALUE (ES1102). A. C. thanks the Spanish Ministry of
653 Economy and Competitiveness for the funding provided within the FPI programme (BES-
654 2011-047612 and EEBB-I-13-06354). We also thank two anonymous referees for their useful
655 comments that helped to improve the original manuscript.

656 References

657 Argüeso D, Evans JP, Fita L (2013) Precipitation bias correction of very high
658 resolution regional climate models. *Hydrology and Earth System Sciences*
659 17(11):4379–4388, DOI 10.5194/hess-17-4379-2013

- 660 Aviad Y, Kutiel H, Lavee H (2013) Empirical models of rain-spells character-
661 istics — A case study of a Mediterranean-arid climatic transect. *Journal of*
662 *Arid Environments* 97:84–91, DOI 10.1016/j.jaridenv.2013.05.015
- 663 Barring L, Holt T, Linderson M, Radziejewski M, Moriondo M, Palutikof JP
664 (2006) Defining dry/wet spells for point observations, observed area aver-
665 ages, and regional climate model gridboxes in Europe. *Climate Research*
666 31(1):35–49, DOI 10.3354/cr031035
- 667 Bellprat O, Kotlarski S, Lüthi D, Schär C (2013) Physical constraints
668 for temperature biases in climate models. *Geophysical Research Letters*
669 40(15):4042–4047, DOI 10.1002/grl.50737
- 670 Benestad RE, Nychka D, Mearns LO (2012) Spatially and temporally con-
671 sistent prediction of heavy precipitation from mean values. *Nature Climate*
672 *Change* 2(7):544–547, DOI 10.1038/NCLIMATE1497
- 673 Boberg F, Christensen JH (2012) Overestimation of Mediterranean summer
674 temperature projections due to model deficiencies. *Nature Clim Change*
675 2(6):433–436, DOI 10.1038/nclimate1454
- 676 Casanueva A, Herrera S, Fernández J, Frías M, Gutiérrez J (2013) Evalu-
677 ation and projection of daily temperature percentiles from statistical and
678 dynamical downscaling methods. *Natural Hazards and Earth System Sci-*
679 *ences* 3:2089–2099, DOI 10.5194/nhess-13-2089-2013
- 680 Casanueva A, Rodríguez-Puebla C, Frías M, González-Reviriego F (2014) Vari-
681 ability of extreme precipitation over Europe and its relationships with tele-
682 connection patterns. *Hydrology and Earth System Sciences* 18:1–17, DOI
683 10.5194/hess-18-709-2014
- 684 Chan SC, Kendon EJ, Fowler HJ, Blenkinsop S, Ferro CAT, Stephenson DB
685 (2013) Does increasing the spatial resolution of a regional climate model
686 improve the simulated daily precipitation? *Climate Dynamics* 41(5-6):1475–
687 1495, DOI 10.1007/s00382-012-1568-9
- 688 Christensen JH, Machenhauer B, Jones RG, Schar C, Ruti PM, Castro M,
689 Visconti G (1997) Validation of present-day regional climate simulations
690 over Europe: LAM simulations with observed boundary conditions. *Climate*
691 *Dynamics* 13(7-8):489–506
- 692 Christensen JH, Boberg F, Christensen OB, Lucas-Picher P (2008) On the
693 need for bias correction of regional climate change projections of tempera-
694 ture and precipitation. *Geophysical Research Letters* 35(20):L20,709, DOI
695 10.1029/2008GL035694
- 696 Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae
697 U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de
698 Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ,
699 Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kållberg P,
700 Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette J, Park
701 B, Peubey C, de Rosnay P, Tavolato C, Thépaut JN, Vitart F (2011) The
702 ERA-Interim reanalysis: configuration and performance of the data assim-
703 ilation system. *Quart J R Meteorol Soc* 137:553–597
- 704 Déqué M, Jones RG, Wild M, Giorgi F, Christensen JH, Hassell DC, Vidale
705 PL, Rockel B, Jacob D, Kjellström E, Castro Md, Kucharski F, Hurk Bvd

- (2005) Global high resolution versus limited area model climate change projections over Europe: quantifying confidence level from PRUDENCE results. *Climate Dynamics* 25(6):653–670, DOI 10.1007/s00382-005-0052-1
- Di Luca A, de Ela R, Laprise R (2015) Challenges in the Quest for Added Value of Regional Climate Dynamical Downscaling. *Current Climate Change Reports* 1(1):10–21, DOI 10.1007/s40641-015-0003-9
- Durman CF, Gregory JM, Hassell DC, Jones RG, Murphy JM (2001) A comparison of extreme European daily precipitation simulated by a global and a regional climate model for present and future climates. *QJR Meteorol Soc* 127(573):1005–1015, DOI 10.1002/qj.49712757316
- Fdez-Arroyabe Hernández P, Martín-Vide J (2012) Regionalization of the probability of wet spells and rainfall persistence in the Basque Country (northern Spain). *International Journal of Climatology* 32(12):1909–1920, DOI 10.1002/joc.2405
- Fernández J, Montávez JP, Sáenz J, González-Rouco JF, Zorita E (2007) Sensitivity of MM5 mesoscale model to physical parameterizations for regional climate studies: Annual cycle. *J Geophys Res* 112:D04,101, doi:10.1029/2005JD00664
- Feser F, Rockel B, von Storch H, Winterfeldt J, Zahn M (2011) Regional climate models add value to global model data: a review and selected examples. *Bulletin of the American Meteorological Society* 92(9):1181–1192
- Frei C, Christensen JH, Déqué M, Jacob D, Jones RG, Vidale PL (2003) Daily precipitation statistics in regional climate models: Evaluation and intercomparison for the European Alps. *Journal of Geophysical Research: Atmospheres* 108(D3):4124, DOI 10.1029/2002JD002287
- García-Díez M, Fernández J, Fita L, Yagüe C (2013) Seasonal dependence of WRF model biases and sensitivity to PBL schemes over Europe. *Quarterly Journal of the Royal Meteorological Society* 139:501–514, DOI 10.1002/qj.1976
- García-Díez M, Fernández J, Vautard R (2015) An RCM multi-physics ensemble over Europe: multi-variable evaluation to avoid error compensation. *Clim Dyn* pp 1–16, DOI 10.1007/s00382-015-2529-x
- Gibelin AL, Déqué M (2003) Anthropogenic climate change over the Mediterranean region simulated by a global variable resolution model. *Climate Dynamics* 20(4):327–339, DOI 10.1007/s00382-002-0277-1
- Giorgi F (2006) Regional climate modeling: Status and perspectives. *Journal de Physique IV (Proceedings)* 139(1):101–118, DOI 10.1051/jp4:2006139008
- Giorgi F, Jones C, Asrar G (2009) Addressing climate information needs at the regional level: the CORDEX framework. *WMO Bulletin* 58(3):175–183
- Grasso L (2000) The differentiation between grid spacing and resolution and their application to numerical modeling. *Bulletin of the American Meteorological Society* 81:579580
- Herrera S, Fita L, Fernández J, Gutierrez JM (2010) Evaluation of the mean and extreme precipitation regimes from the ENSEMBLES regional climate multimodel simulations over Spain. *J Geophys Res* 115(D21):D21,117, DOI 10.1029/2010JD013936

- 752 Herrera S, Gutiérrez J, Ancell R, Pons M, Frías M, Fernández J (2012) Devel-
753 opment and analysis of a 50 year high resolution daily gridded precipitation
754 dataset over Spain (Spain02). *International Journal of Climatology* DOI
755 10.1002/joc.2256
- 756 Herrera S, Fernández J, Gutiérrez (2015) Update of the Spain02 gridded obser-
757 vational dataset for EURO-CORDEX evaluation: Assessing the effect of the
758 interpolation methodology. *International Journal of Climatology* , in print,
759 DOI 10.1002/joc.4391
- 760 Hofstra N, Haylock M, New M, Jones PD (2009) Testing E-OBS European
761 high-resolution gridded data set of daily precipitation and surface tem-
762 perature. *Journal of Geophysical Research: Atmospheres* 114(D21), DOI
763 10.1029/2009JD011799
- 764 Isotta FA, Frei C, Weilguni V, Perčec Tadić M, Lassègues P, Rudolf B, Pavan
765 V, Cacciamani C, Antolini G, Ratto SM, Munari M, Micheletti S, Bonati
766 V, Lussana C, Ronchi C, Panettieri E, Marigo G, Vertačnik G (2013) The
767 climate of daily precipitation in the Alps: development and analysis of a
768 high-resolution grid dataset from pan-Alpine rain-gauge data. *International*
769 *Journal of Climatology* DOI 10.1002/joc.3794
- 770 Jacob D, Bärring L, Christensen OB, Christensen JH, de Castro M, Déqué
771 M, Giorgi F, Hagemann S, Lenderink G, Rockel B, Sanchez E, Schär C,
772 Seneviratne SI, Somot S, van Ulden A, van den Hurk B (2007) An inter-
773 comparison of regional climate models for Europe: model performance in
774 present-day climate. *Climatic Change* 81:31–52
- 775 Jacob D, Petersen J, Eggert B, Alias A, Christensen OB, Bouwer LM, Braun
776 A, Colette A, Déqué M, Georgievski G, Georgopoulou E, Gobiet A, Menut
777 L, Nikulin G, Haensler A, Hempelmann N, Jones C, Keuler K, Kovats S,
778 Kröner N, Kotlarski S, Kriegsmann A, Martin E, Meijgaard Ev, Moseley C,
779 Pfeifer S, Preuschmann S, Radermacher C, Radtke K, Rechid D, Rounsevell
780 M, Samuelsson P, Somot S, Soussana JF, Teichmann C, Valentini R, Vau-
781 tard R, Weber B, Yiou P (2014) EURO-CORDEX: new high-resolution cli-
782 mate change projections for european impact research. *Reg Environ Change*
783 14(2):563–578, DOI 10.1007/s10113-013-0499-2
- 784 Jerez S, Montavez JP, Jimenez-Guerrero P, Gomez-Navarro JJ, Lorente-Plazas
785 R, Zorita E (2012) A multi-physics ensemble of present-day climate regional
786 simulations over the Iberian Peninsula. *Climate Dynamics* pp 1–24, DOI
787 10.1007/s00382-012-1539-1
- 788 Kendon EJ, Roberts NM, Senior CA, Roberts MJ (2012) Realism of rain-
789 fall in a very high-resolution regional climate model. *Journal of Climate*
790 25(17):5791–5806, DOI 10.1175/JCLI-D-11-00562.1
- 791 Kotlarski S, Keuler K, Christensen OB, Colette A, Déqué M, Gobiet A, Go-
792 ergen K, Jacob D, Lüthi D, van Meijgaard E, Nikulin G, Schär C, Te-
793 ichmann C, Vautard R, Warrach-Sagi K, Wulfmeyer V (2014) Regional
794 climate modeling on european scales: a joint standard evaluation of the
795 EURO-CORDEX RCM ensemble. *Geosci Model Dev* 7(4):1297–1333, DOI
796 10.5194/gmd-7-1297-2014

- 797 Lázaro R, Rodrigo FS, Gutiérrez L, Domingo F, Puigdefábregas J (2001)
798 Analysis of a 30-year rainfall record (1967–1997) in semi-arid SE Spain for
799 implications on vegetation. *Journal of Arid Environments* 48(3):373–395,
800 DOI 10.1006/jare.2000.0755
- 801 van der Linden P, Mitchell J (eds) (2009) ENSEMBLES: Climate Change
802 and its Impacts: Summary of research and results from the ENSEMBLES
803 project. Met Office Hadley Centre, FitzRoy Road, Exeter EX1 3PB, UK
- 804 Maraun D, Widmann M (2015) The representation of location by regional cli-
805 mate models in complex terrain. *Hydrol Earth Syst Sci Discuss* 12(3):3011–
806 3028, DOI 10.5194/hessd-12-3011-2015
- 807 Nikulin G, Kjellstrom E, Hansson U, Strandberg G, Ullerstig A (2011) Evalua-
808 tion and future projections of temperature, precipitation and wind extremes
809 over Europe in an ensemble of regional climate simulations. *Tellus Series a-*
810 *Dynamic Meteorology and Oceanography* 63(1):41–55
- 811 Orłowsky B, Seneviratne SI (2012) Global changes in extreme events: regional
812 and seasonal dimension. *Climatic Change* 110(3-4):669–696, DOI 10.1007/
813 s10584-011-0122-9
- 814 Panofsky HA, Brier GW (1968) Some applications of statistics to meteorol-
815 ogy. University Park : Penn. State University, College of Earth and Mineral
816 Sciences
- 817 Piani C, Haerter JO, Coppola E (2010) Statistical bias correction for daily pre-
818 cipitation in regional climate models over Europe. *Theoretical and Applied*
819 *Climatology* 99(1-2):187–192, DOI 10.1007/s00704-009-0134-9
- 820 Polanski S, Rinke A, Dethloff K (2010) Validation of the HIRHAM-
821 simulated Indian summer monsoon circulation. *Advances in Meteorology*
822 2010:e415,632, DOI 10.1155/2010/415632
- 823 Prein AF, Holland GJ, Rasmussen RM, Done J, Ikeda K, Clark MP, Liu CH
824 (2013) Importance of regional climate model grid spacing for the simula-
825 tion of heavy precipitation in the Colorado headwaters. *Journal of Climate*
826 26(13):4848–4857, DOI 10.1175/JCLI-D-12-00727.1
- 827 Prein AF, Gobiet A, Truhetz H, Keuler K, Goergen K, Teichmann C,
828 Fox Maule C, van Meijgaard E, Déqué M, Nikulin G, Vautard R, Colette A,
829 Kjellström E, Jacob D (2015) Precipitation in the EURO-CORDEX 0.11°
830 and 0.44° simulations: high resolution, high benefits? *Climate Dynamics*
831 DOI 10.1007/s00382-015-2589-y
- 832 Pryor SC, Nikulin G, Jones C (2012) Influence of spatial resolution on regional
833 climate model derived wind climates. *J Geophys Res* 117(D3):D03,117, DOI
834 10.1029/2011JD016822
- 835 Roeckner E, Brokopf R, Esch M, Giorgetta M, Hagemann S, Kornblueh L,
836 Manzini E, Schlese U, Schulzweida U (2006) Sensitivity of simulated climate
837 to horizontal and vertical resolution in the ECHAM5 atmosphere model.
838 *Journal of Climate* 19(16):3771–3791
- 839 Schmidli J, Frei C, Vidale PL (2006) Downscaling from GCM precipitation: a
840 benchmark for dynamical and statistical downscaling methods. *International*
841 *Journal of Climatology* 26(5):679–689, DOI 10.1002/joc.1287

- 842 Shkolnik IM, Efimov SV (2013) Cyclonic activity in high latitudes as simulated
843 by a regional atmospheric climate model: added value and uncertainties.
844 *Environmental Research Letters* 8(4):045,007, DOI 10.1088/1748-9326/8/
845 4/045007
- 846 Taylor KE (2001) Summarizing multiple aspects of model performace in a
847 single diagram. *J Geophys Res* 106(D7):7183–7192
- 848 Teichmann C, Eggert B, Elizalde A, Haensler A, Jacob D, Kumar P, Moseley
849 C, Pfeifer S, Rechid D, Remedio AR, Ries H, Petersen J, Preuschmann S,
850 Raub T, Saeed F, Sieck K, Weber T (2013) How does a regional climate
851 model modify the projected climate change signal of the driving GCM: A
852 study over different CORDEX regions using REMO. *Atmosphere* 4(2):214–
853 236, DOI 10.3390/atmos4020214
- 854 Themeßl MJ, Gobiet A, Heinrich G (2012) Empirical-statistical downscal-
855 ing and error correction of regional climate models and its impact on
856 the climate change signal. *Climatic Change* 112(2):449–468, DOI 10.1007/
857 s10584-011-0224-4
- 858 Vautard R, Gobiet A, Jacob D, Belda M, Colette A, Déqué M, Fernández J,
859 García-Díez M, Goergen K, Güntler I, Halenka T, Karacostas T, Katragkou
860 E, Keuler K, Kotlarski S, Mayer S, Meijgaard Ev, Nikulin G, Patarčić
861 M, Scinocca J, Sobolowski S, Suklitsch M, Teichmann C, Warrach-Sagi K,
862 Wulfmeyer V, Yiou P (2013) The simulation of european heat waves from an
863 ensemble of regional climate models within the EURO-CORDEX project.
864 *Clim Dyn* 41(9):2555–2575, DOI 10.1007/s00382-013-1714-z
- 865 Vrac M, Friederichs P (2015) Multivariateintervariable, spatial, and temporal-
866 bias correction. *J Climate* 28(1):218–237, DOI 10.1175/JCLI-D-14-00059.1
- 867 Walther A, Jeong JH, Nikulin G, Jones C, Chen D (2013) Evaluation of
868 the warm season diurnal cycle of precipitation over Sweden simulated by
869 the Rossby Centre regional climate model RCA3. *Atmospheric Research*
870 119:131–139, DOI 10.1016/j.atmosres.2011.10.012
- 871 Wilcke RAI, Mendlik T, Gobiet A (2013) Multi-variable error correction of
872 regional climate models. *Climatic Change* 120(4):871–887, DOI 10.1007/
873 s10584-013-0845-x
- 874 Yoo C, Lee J, Yoon Y (2001) Climatological thresholds of daily rainfall.
875 *Journal of Hydrologic Engineering* 6(5):443–449, DOI 10.1061/(ASCE)
876 1084-0699(2001)6:5(443)

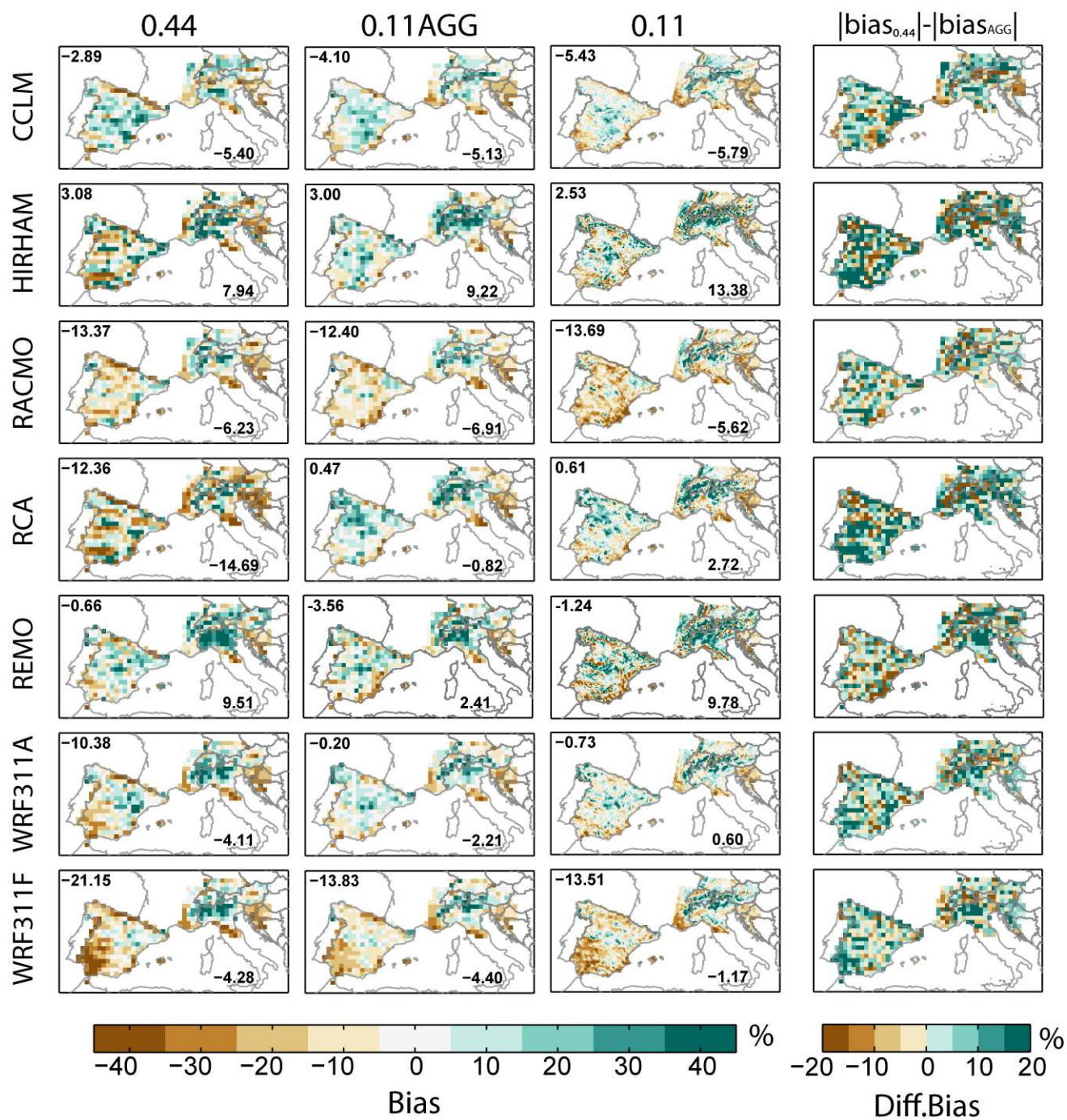


Fig. 2 Winter SDII relative biases (%) for the RCMs (rows) at 0.44° (first column), 0.11AGG (second column) and 0.11° (third column) resolutions. Values in the upper left and lower right corner represent the relative biases of the spatially averaged SDII in both regions. The fourth column shows the difference between the first two columns in absolute values.

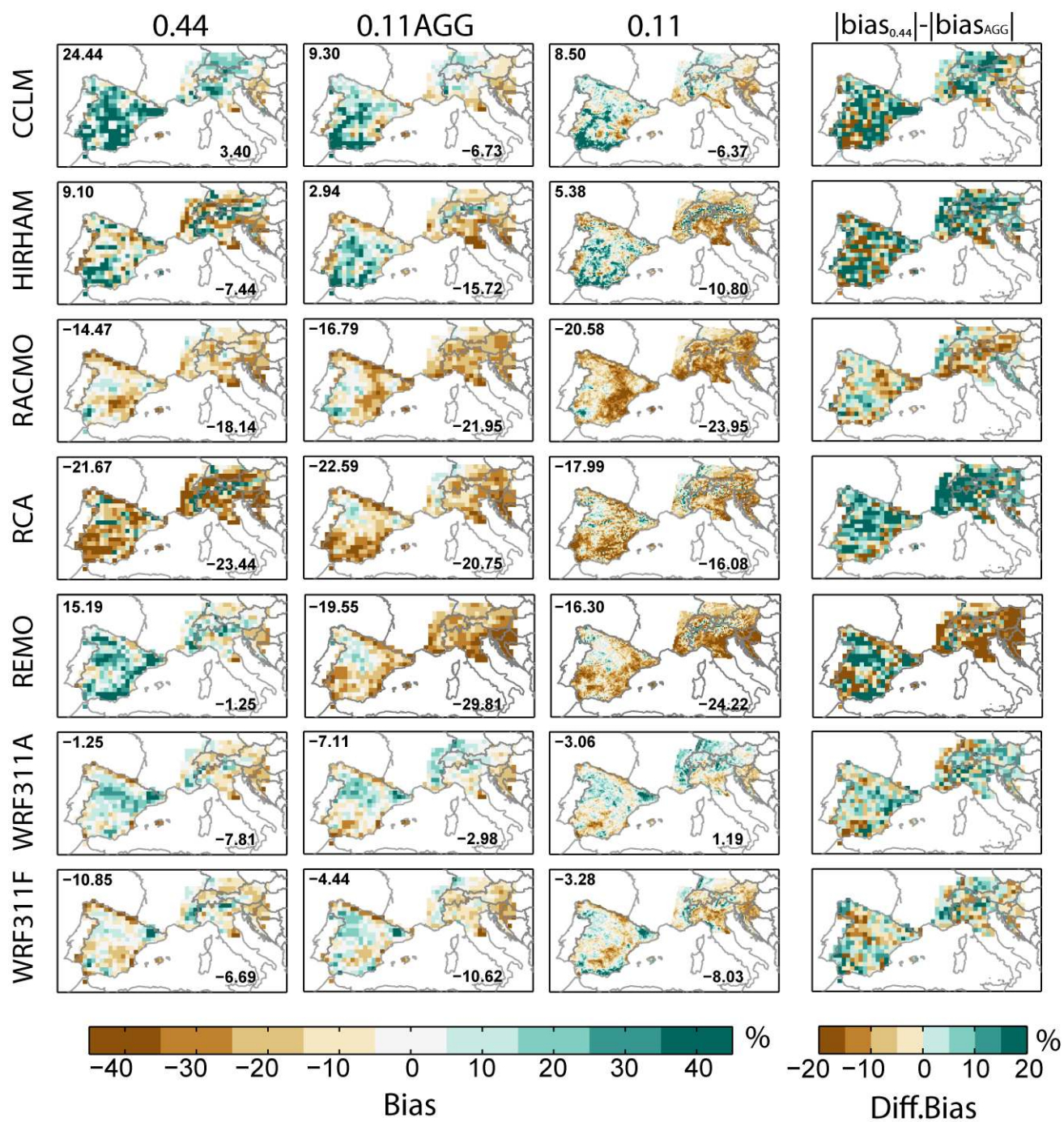


Fig. 3 As Figure 2, but in summer.

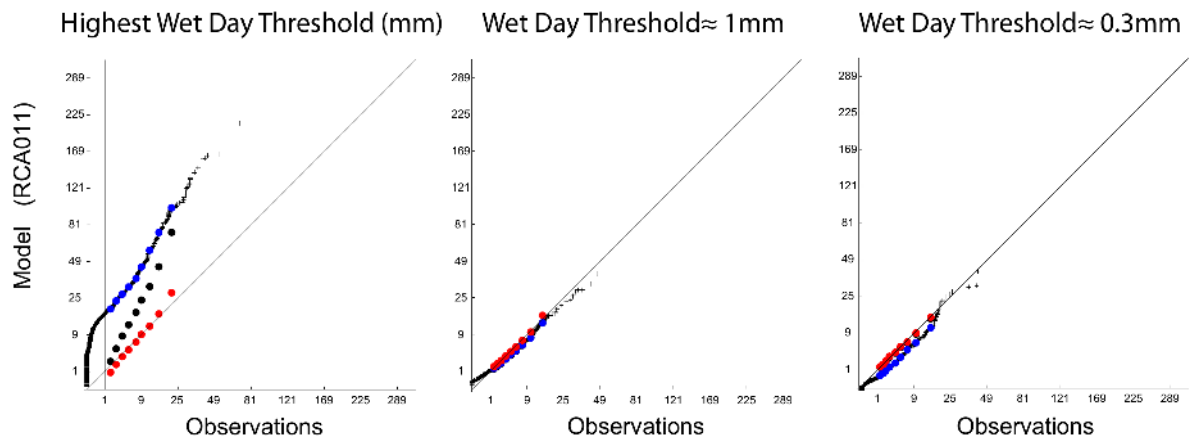


Fig. 4 q-q plots for 3 selected grid points from RCA011 simulations in winter (black crosses). These grid points corresponds to adjusted wet-day thresholds greatly exceeding 1mm (left), around 1mm (center) and under 1mm (≈ 0.3 mm, right). Values are presented in squared root scale with labels in the original units. The vertical line corresponds to 1mm in the observations; its intersection with the q-q plot provides the adjusted threshold equivalent to 1mm in the observations (P^*). Dots show the deciles from the wet day distribution for 1mm threshold (black), the adjusted wet day threshold (blue) and after local scaling (red). 90pWET corresponds to the last dot of each series.

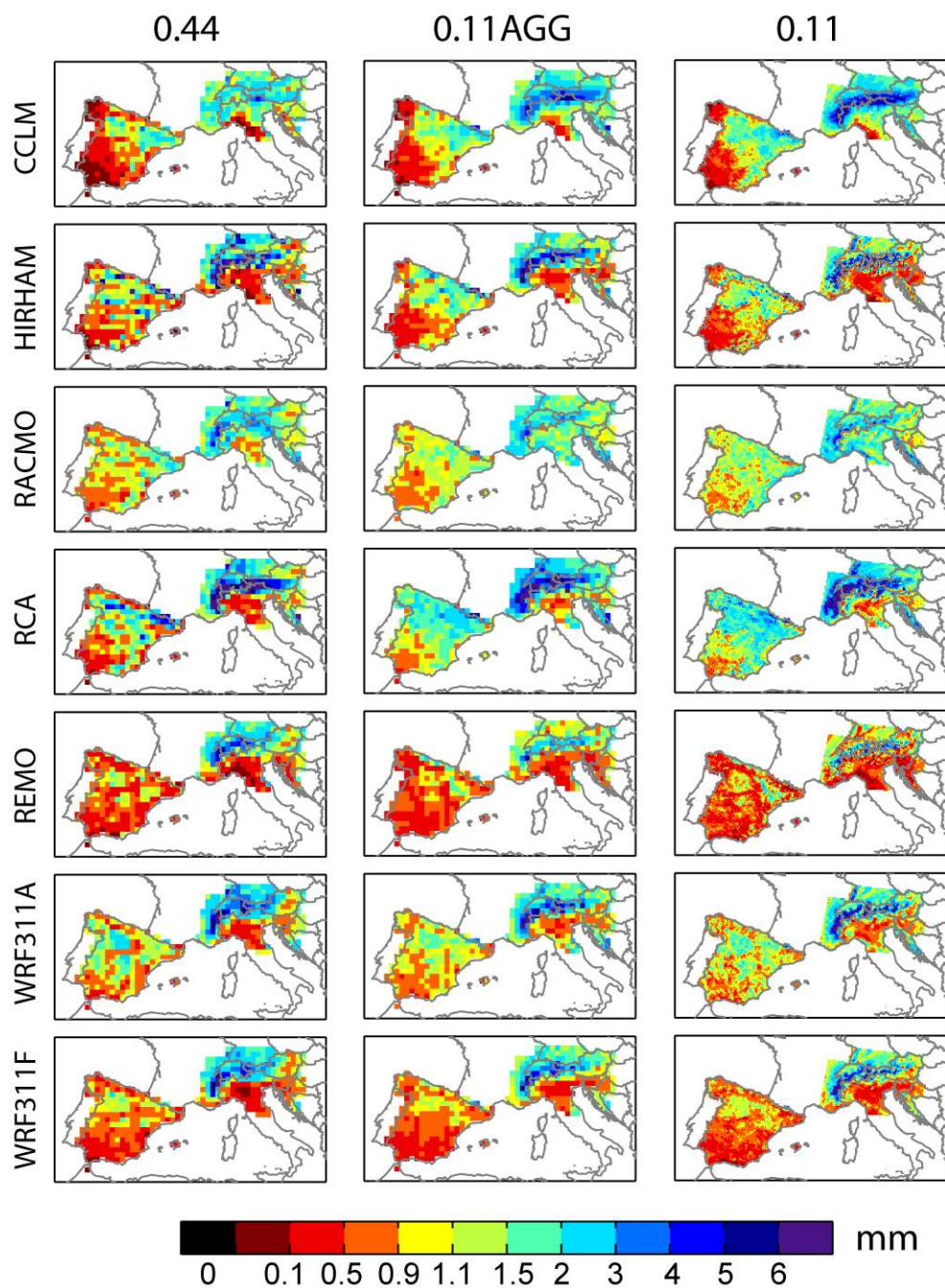


Fig. 5 Wet-day threshold equivalent to 1mm in the observations for all the RCMs (rows) and resolutions (columns) in winter. For a better contrast of spatial differences, values are presented using a non-linear scale. Note that the black color represent $P^* = 0$.

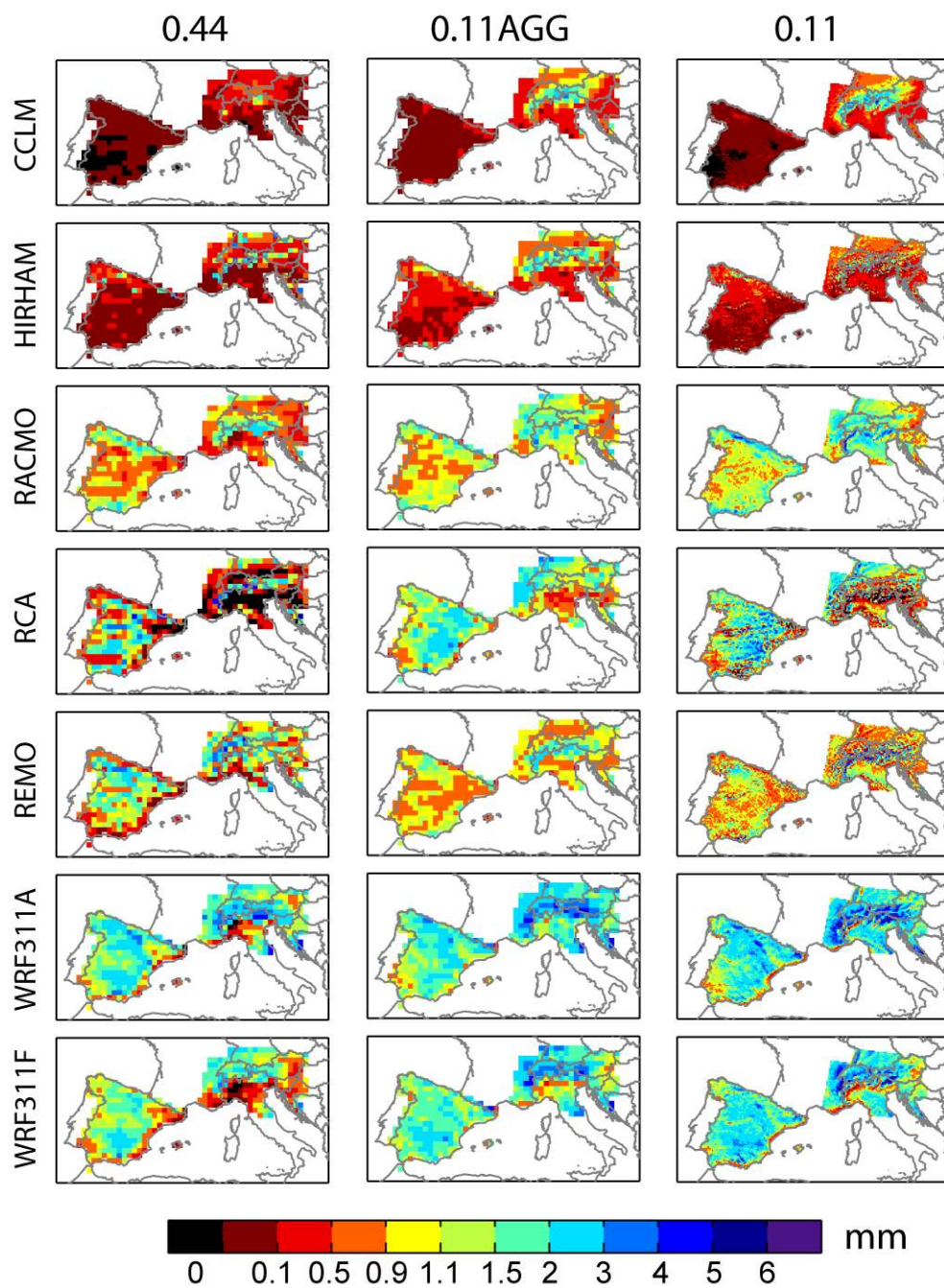


Fig. 6 As Figure 5, but in summer.

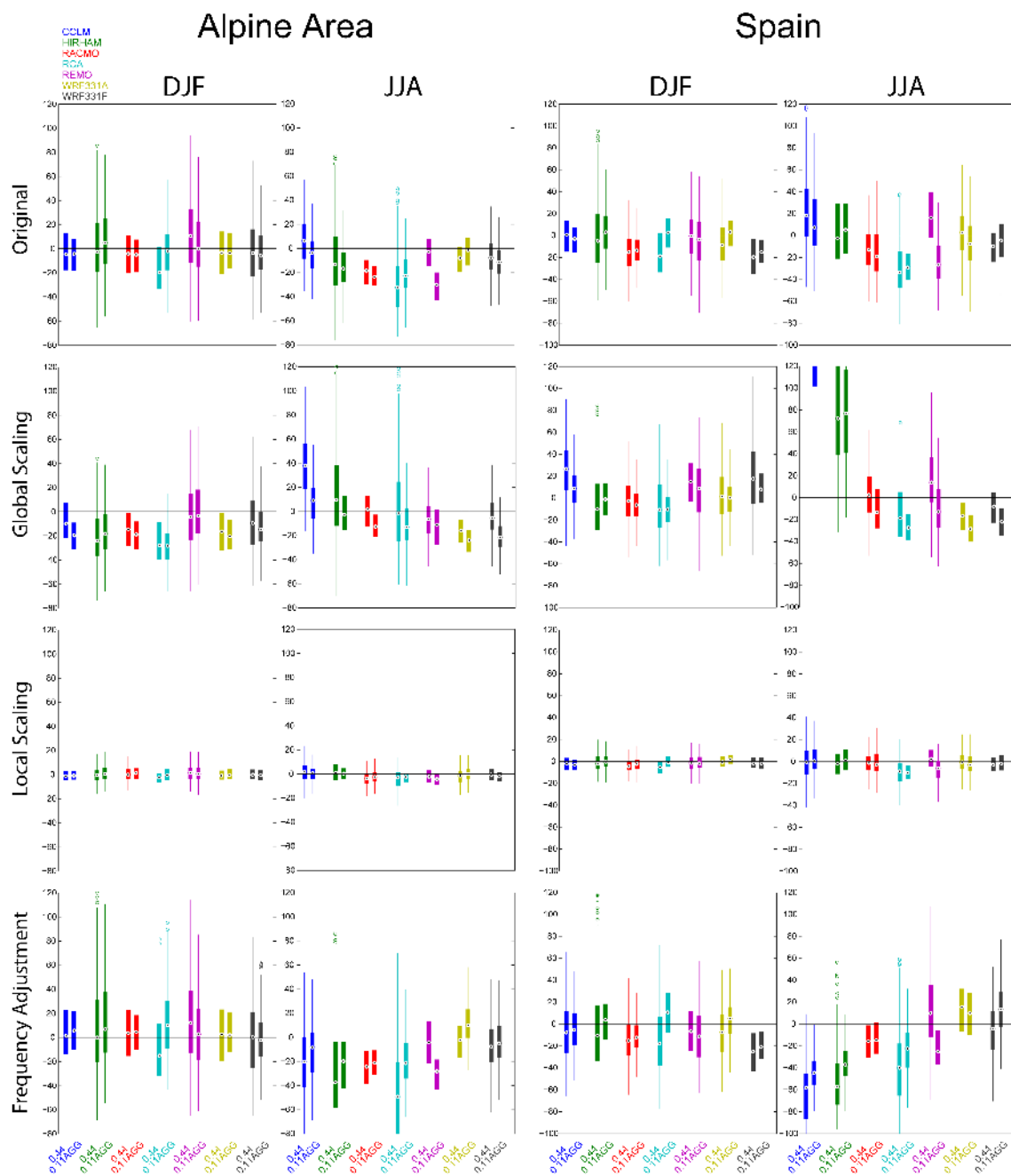


Fig. 7 Boxplots summarizing the spatial distribution of 90pWET relative biases (in %) for winter (DJF) and summer (JJA) for both regions (columns). The 90th percentile is calculated with the standard 1mm fixed threshold as reference (first row). Second to fourth rows correspond to the relative biases in 90pWET when GS, LS and FA corrections are applied, respectively. Each box corresponds to one RCM and resolution (0.44° and 0.11AG per RCM). The boxes show the interquartile range and the median (circle) but, for the sake of clarity, the whiskers extend only to the 5th and 95th percentiles.

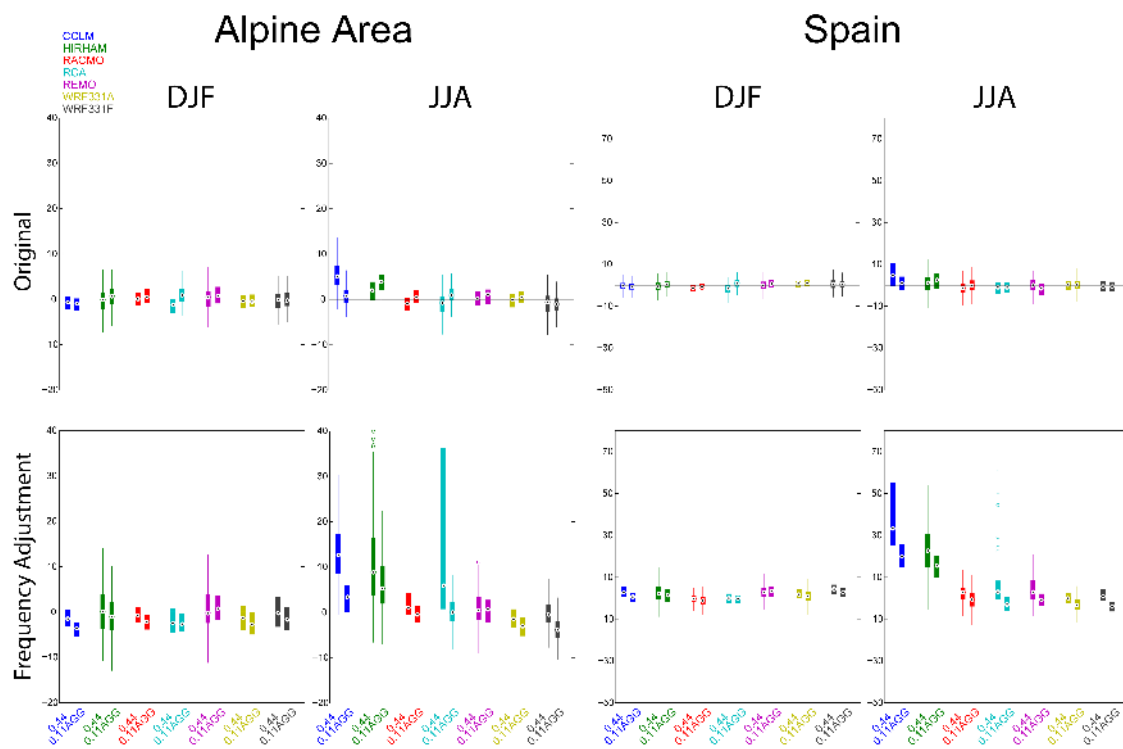


Fig. 8 As Figure 7, but for the absolute biases of R95pTOT (in %). GS and LS corrections are omitted, since they do not affect this index.

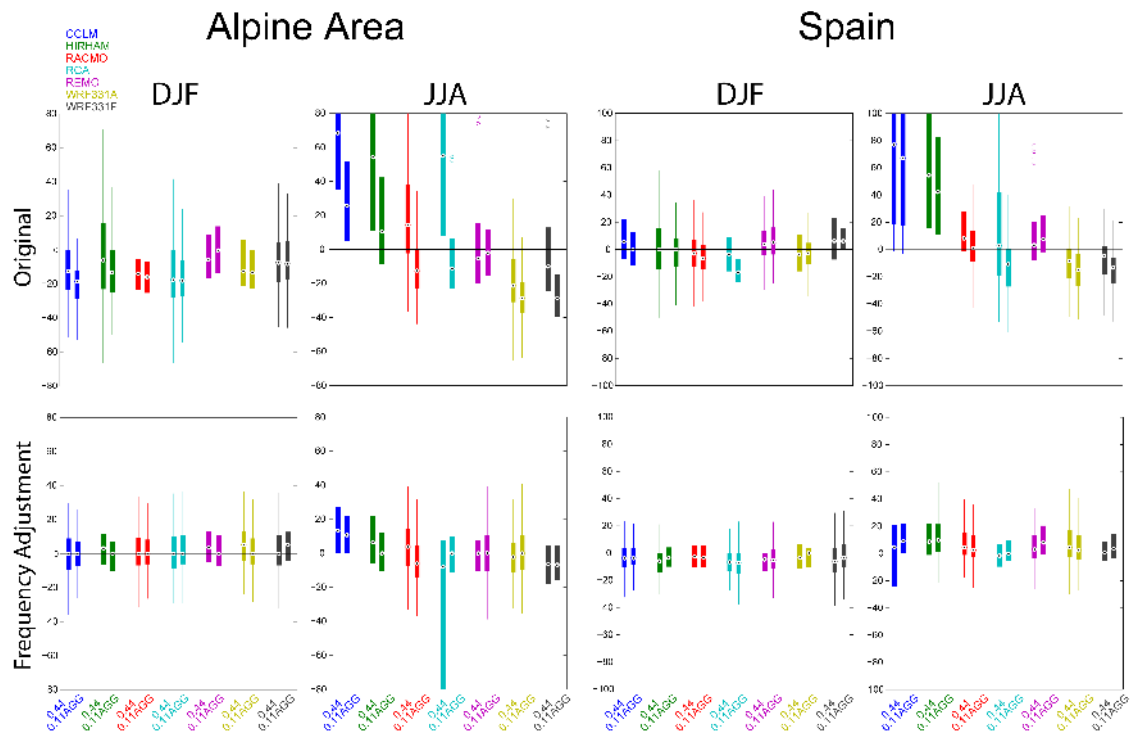


Fig. 9 Boxplots for the spatial distribution of CDD relative biases (in %) for winter (left) and summer (right). The indicator is calculated with the 1mm fixed threshold (first row) and with the adjusted wet-day threshold (second row). See Figure 7 for further details.

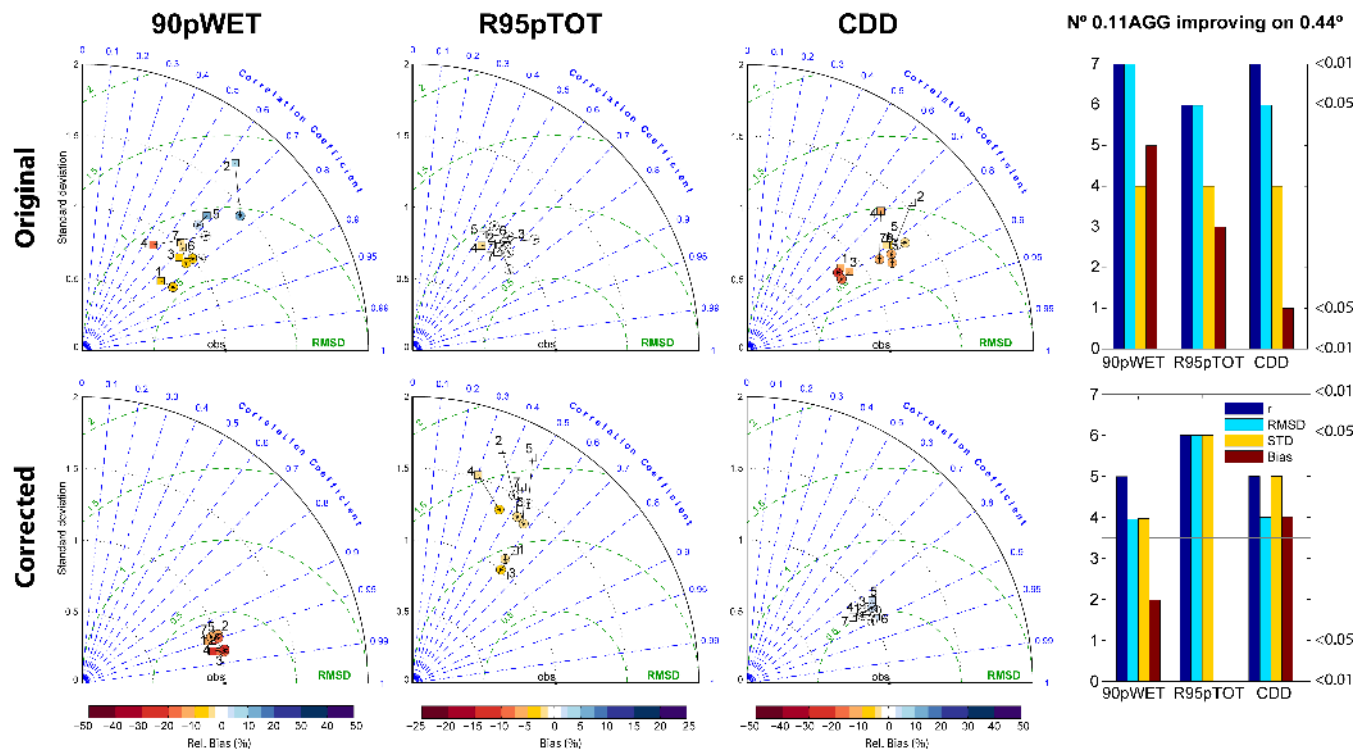


Fig. 10 Taylor diagrams for winter 90pWET, R95pTOT and CDD in the Alpine region. The first row shows the original data using a 1mm fixed wet-day threshold. The second row shows corrected data (LS for 90pWET and FA for R95pTOT and CDD). Squares represent 0.44° resolution and circles 0.11AGG. Their colors correspond to the biases in the spatially-averaged index. The numbers close to the square markers identify the RCMs (see codes in Table 1). The right panel shows barplots of the number of RCMs at 0.11AGG resolution that perform better than the 0.44° resolution in spatial Pearson correlation coefficient (r), centered root mean squared difference (RMSD), variability (std) and bias. The results are statistically significant only when 6 or 7 RCMs improve upon the 0.44° (see text).