AGU
ADVANCING
EARTH AND
SPACE SCIENCE

JGR Atmospheres

Evaluation of the Large EURO-CORDEX Regional Climate Model Ensemble

Robert Vautard[1], Nikolay Kadygrov[1], Carley Iles[2], Fredrik Boberg[3], Erasmo Buonomo[4], Katharina Bülow[5], Erika Coppola[6], Lola Corre[7], Erik van Meijgaard[8], Rita Nogherotto[6], Marit Sandstad[9], Clemens Schwingshackl[9], Samuel Somot[10], Emma Aalbers[8], Ole B. Christensen[3], James M. Ciarlo[6,11], Marie-Estelle Demory[12], Filippo Giorgi[6], Daniela Jacob[5], Richard G. Jones[4], Klaus Keuler[13], Erik Kjellström[14], Geert Lenderink[8], Guillaume Levavasseur[1], Grigory Nikulin[14], Jana Sillmann[9], Cosimo Solidoro[11], Silje Lund Sørland[12], Christian Steger[15], Claas Teichmann[5], Kirsten Warrach-Sagi[16], and Volker Wulfmeyer[16]

[1]Institut Pierre-Simon Laplace, Paris, France, [2]Laboratoire des Sciences du Climat et de l'Environnement, Institut Pierre-Simon Laplace, Gif-sur-Yvette, France, [3]Danish Meteorological Institute, Copenhagen, Denmark, [4]Met Office Hadley Center, Exeter, UK, [5]Climate Service Center Germany (GERICS), Helmholtz-Zentrum Geesthacht, Hamburg, D-20095, Germany, [6]The Abdus Salam International Center for Theoretical Physic (ICTP), Trieste, Italy, [7]Meteo-France, Toulouse, France, [8]KNMI, Royal Netherlands Meteorological Institute, De Bilt, The Netherlands, [9]Center for International Climate Research (CICERO), Oslo, Norway, [10]CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France, [11]National Institute of Oceanography and Experimental Geophysics-OGS, Trieste, Italy, [12]Institute for Atmospheric and Climate Science, ETH Zurich, Zürich, Switzerland, [13]Atmospheric Processes, Brandenburg University of Technology Cottbus-Senftenberg, Cottbus, Germany, [14]SMHI: Swedish Meteorological and Hydrological Institute, Norrköping, Sweden, [15]Deutscher Wetterdienst, Offenbach, Germany, [16]Institute of Physics and Meteorology, University of Hohenheim, Stuttgart, Germany

Abstract The use of regional climate model (RCM)-based projections for providing regional climate information in a research and climate service contexts is currently expanding very fast. This has been possible thanks to a considerable effort in developing comprehensive ensembles of RCM projections, especially for Europe, in the EURO-CORDEX community (Jacob et al., 2014, 2020). As of end of 2019, EURO-CORDEX has developed a set of 55 historical and scenario projections (RCP8.5) using 8 driving global climate models (GCMs) and 11 RCMs. This article presents the ensemble including its design. We target the analysis to better characterize the quality of the RCMs by providing an evaluation of these RCM simulations over a number of classical climate variables and extreme and impact-oriented indices for the period 1981–2010. For the main variables, the model simulations generally agree with observations and reanalyses. However, several systematic biases are found as well, with shared responsibilities among RCMs and GCMs: Simulations are overall too cold, too wet, and too windy compared to available observations or reanalyses. Some simulations show strong systematic biases on temperature, others on precipitation or dynamical variables, but none of the models/simulations can be defined as the best or the worst on all criteria. The article aims at supporting a proper use of these simulations within a climate services context.

Plain Language Summary This study analyses the ability of the unprecedently large ensemble of 55 regional climate simulations to properly simulate the climatology of several variables, extremes, and impact-oriented indices over the European continent. This analysis should guide the use of regional climate projections in climate services development.

1. Introduction

Regional climate change projections are widely used for research and applications in order to describe the future of climatic conditions and subsequent impacts at a scale that is presumably better suited than coarser resolution global climate projections (Giorgi, 2019; Rummukainen, 2016). However, comprehensive assessments of regional climate simulations are still hard to establish. Regional climate projections generally use limited area models, or regional climate models (RCMs), with resolutions currently ranging from a few kilometers to about 50 km, downscaling GCMs. Because they use higher resolutions (and are sometimes

Formal analysis: Robert Vautard, Nikolay Kadygrov, Carley Iles, Fredrik Boberg, Erasmo Buonomo, Katharina Bülow, Erika Coppola, Lola Corre, Erik van Meijgaard, Rita Nogherotto, Marit Sandstad, Clemens Schwingshackl, Samuel Somot, Emma Aalbers, Ole B. Christensen, James M. Ciarlo, Marie-Estelle Demory, Klaus Keuler, Erik Kjellström, Silje Lund Sørland, Christian Steger, Kirsten Warrach-Sagi, Volker Wulfmeyer

Investigation: Robert Vautard, Carley Iles, Erasmo Buonomo, Erika Coppola, Lola Corre, Erik van Meijgaard, Samuel Somot, Ole B. Christensen, Richard G. Jones, Silje Lund Sørland

Methodology: Samuel Somot

Software: Guillaume Levavasseur

Validation: Robert Vautard, Nikolay Kadygrov, Jana Sillmann

Writing – original draft: Robert Vautard, Fredrik Boberg, Erasmo Buonomo, Erik van Meijgaard, Samuel Somot, Emma Aalbers, James M. Ciarlo, Klaus Keuler, Erik Kjellström, Christian Steger, Kirsten Warrach-Sagi

Writing – review & editing: Richard G. Jones

configured for specific regions), the quality of RCM simulations is expected to be better than those from their driving GCMs. A number of studies have indeed shown this added value for some variables and their extremes, especially in complex-terrain areas, although a higher skill than global climate models (GCMs) in describing climate variables over continental plains remains more difficult to demonstrate (Fantini et al., 2018; Prein et al., 2016; Ruti et al., 2016; Torma et al., 2015). In practice, RCMs also have their own biases and therefore systematic errors in regional climate simulations result from both model types.

Over Europe, a large effort in downscaling GCM projections from the 5th Coupled Model Intercomparison Project (CMIP5, Taylor et al., 2012) has been developed for about a decade within the framework of CORDEX (Jones et al., 2011). This is thanks to a coordinated effort from the European modeling institutes, European projects such as IMPACT2C (https://www.atlas.impact2c.eu/en/about/about-impact2c/), the Copernicus Climate Change Service (C3S; https://climate.copernicus.eu/), and several European and national projects support, for example, the German initiative ReKliEs-De (http://reklies.hlnug.de). EURO-CORDEX covers Europe at 0.11° (12.5 km) and 0.44° (50 km) resolutions (Jacob et al., 2014, 2020; Kotlarski et al., 2014), with the purpose to explore the added value of ~10-km resolution RCMs and to develop a large ensemble allowing to assess robust future climate change which can also feed climate service activities. Note that other experiments such as Med-CORDEX (Ruti et al., 2016; Somot et al., 2018), focusing on the Mediterranean area, also cover large portions of Europe.

The EURO-CORDEX community endeavor has achieved several goals, but major challenges remain (Jacob et al., 2020). One is to better identify, characterize, and understand the origins and drivers of the biases and uncertainties in regional climate simulations and projections. For instance, simulation biases are a combination of both GCM and RCM biases arising from a number of sources. Understanding these sources requires a large ensemble and a GCM-RCM "matrix" (the ensemble of possible combinations for downscaling every GCM with every RCM) that is as full as possible (Christensen & Kjellström, 2020; Déqué et al., 2012; Evin et al., 2019). To date, this possibility has remained limited at the higher resolution (0.11°) due to the computational burden necessary to carry out the simulations, and as a result the distribution of GCM-RCM simulations across the matrix remains characterized by large imbalances (Fernández et al., 2019). Different earlier model evaluation studies have analyzed reduced sets of EURO-CORDEX simulations For example, Kotlarski et al., 2014 used a set of nine simulations at 0.11° resolution using seven different RCMs. Previously, evaluations of RCM-based large ensembles over Europe were conducted at lower spatial resolution, for example, the PRUDENCE project ensemble of 15 regional climate simulations at 50 km (Jacob et al., 2007) or later the ENSEMBLES project ensemble at 25 km (Kjellström et al., 2010; Lenderink, 2010; Lorenz & Jacob, 2010; Sanchez-Gomez et al., 2009, 2011).

Recent progress in the number of simulations carried out in a coordinated effort, following the first publications of the EURO-CORDEX simulations (Jacob et al., 2014; Vautard et al., 2013), was fostered by the Copernicus Climate Change Service (C3S). As of October 2019, a total of 55 combinations of 11 RCMs run at ~0.11° grid spacing with boundary conditions from eight GCMs is available, an unprecedented number, with a greater portion of the GCM-RCM matrix being filled (Christensen & Kjellström, 2020). Some of these simulations also use the same RCM to downscale several ensemble members of a given GCM to better enable the assessment of internal variability at high resolution (e.g., Aalbers et al., 2018). Despite this progress, the GCM-RCM matrix covering several scenarios for radiative forcing is not complete, and a filling strategy has been designed. Note that the largest ensemble of GCM-driven RCMs over Europe before EURO-CORDEX was completed by the ENSEMBLES project with a similar number of GCMs (seven from CMIP3 including three versions of the MetOffice GCM with different climate sensitivities), more RCMs (16), but for a total of only about 20 simulations (Déqué, 2010; Déqué et al., 2012; Kjellström et al., 2013). Indeed, the maximum number of GCMs used per RCM was three (RCA3 and HIRHAM5) in ENSEMBLES, whereas it is seven in EURO-CORDEX (see Table 1) without accounting for multiple members from the same GCM.

This article takes the opportunity of the latest developments in the EURO-CORDEX initiative to (i) present the most recent GCM-RCM EURO-CORDEX ensemble, (ii) evaluate the main biases of the ensemble for an extended set of variables and indices important for socio-economic sectors, and (iii) assess the origins of the biases and the respective contributions of GCMs and RCMs. We focus here on the analysis of climate simulations for the recent past as compared with observations, including seasonal means,

Table 1

Simulations Analyzed in This Study, Which Cover Both Historical and RCP8.5 Periods. Cells highlghted gray are combinations of GCM and RCM that have a simulation. Footnotes and ensemble member specified whe several possibilities were available.

RCM-GCM	CNRM r1	ECEARTH r12, r1, and r3	HADGEM r1	MPI r1, r2, and r3	NORESM r1	IPSL r1	CANESM r1	MIROC r1
CCLM		r12		r1			[4]	[4]
HIRHAM		r12, r1, and r3						
RACMO		r12, r1, and r3		r1				
RCA		r12, r1, and r3		r1 and r3				
REMO		r12		$r1_2$, $r2_2$, and r3				
WRF361H		r12[4,5,6,7]	[4,5,6,7,8]	r1[4,5,6,7]				
WRF381P	[9]		[9]		[9]	[9]		
ALADIN53	[3,4]							
ALADIN63								
REGCM			[1]	$r1_1$				
COSMO-crCLIM		r12		r1, r2, and r3				

Note. The name of GCM simulations is given as "GCM (rR)" where "R" is the model ensemble realization. Footnotes, given as superscripts in the table: [1] This simulation does not have sea level pressure available; [2] this REMO version is REMO2009, while other versions are REMO2015; [3] this simulation does not have daily surface wind available; [4] this simulation does not have daily maximum surface wind available; [5] this simulation does not have daily sea level pressure fields; [6] this simulation does not have surface solar downward radiation (rsds) field analyzed; [7] this simulation does not have WBGT index analyzed; [8] this simulation does not have total runoff; and [9] this simulation does not have evaporation or total soil moisture content.

extremes, and a few hazard indices. A companion paper (Coppola et al., 2021) assesses how the same ensemble behaves in future projections for the representative concentration pathways (RCPs) RCP2.6 and RCP8.5 (Moss et al., 2010).

Our purpose is to increase understanding of regional climate modeling skill and limitations to feed future research. This information is also important to guide appropriate interpretation by users of climate services of results from the simulations. However, the main purpose of this article is not to provide detailed recommendations to climate services, which would require a different approach taking into account sector-specific needs, nor to document individual simulation biases or to understand their origin.

Section 2 presents the GCM-RCM matrix and the status of the EURO-CORDEX ensemble. Section 3 presents the indices analyzed and the observations used as a reference for the model assessment. Section 4 presents the analysis of mean biases of temperature, precipitation, dynamics, and radiation, while section 5 analyses extremes and hazard indices. Bias partitioning analysis can be found in sections 4 and 5. A discussion including suggested consequences for future research and climate services closes the article (section 6).

2. The 55-Model GCM-RCM EURO-CORDEX Ensemble

2.1. The GCM-RCM Matrix

One of the goals of EURO-CORDEX is to assess the uncertainty of regional climate projections using a large set of simulations over Europe at a high resolution of 0.11°. In particular, an experimental protocol allowing an assessment of how the choice of RCM and GCM models contributes to the spread in the resulting ensemble has been set up. The general strategy is to construct sub-matrices in the three-dimensional space of scenarios, GCMs and RCMs, which are complete and hence enable targeted analyses. However, we focus here on the overall assessment of the ensemble, using all simulations available as of October 2019, with minimal variable availability requirements (see below). This set of simulations was carried out since 2011 in various national and European projects and was recently enhanced as part of the Copernicus Climate Change Service program. This led to a focus on a reduced set of eight GCMs as detailed below, which roughly covers the range of CMIP5 equilibrium climate sensitivities as assessed by the AR5 WGI IPCC report (IPCC, 2013). Note however that the Euro-CORDEX initiative and the related C3S project are still active. Therefore, new 0.11° projection simulations will become available in addition to the 55 simulations analyzed in the current study.

2.2. The Ensemble of Simulations

We consider the ensemble of all EURO-CORDEX simulations including at least daily mean, minimum and maximum 2-m temperature, and daily precipitation amount, provided that they include a historical period (1950–2005) and future conditions following the RCP8.5 scenario. The ensemble contains 55 different simulations, each resulting from the downscaling of a GCM by an RCM. In total, the ensemble includes 11 RCMs downscaling 8 GCMs. The simulations are shown as a matrix in Table 1.

Some RCMs have been used to downscale several realizations of the same GCM, allowing us to estimate the effect of simulated natural variability on climate change responses, although this aspect is not presented here. Such is the case for HIRHAM5, RACMO22E, and RCA4 downscaling EC-Earth, and for COSMO-crCLIM, RCA4, and REMO downscaling the MPI GCM. For some RCMs, different versions of the same model were used. For REMO, REMO2009 is used for two simulations and REMO2015 for the others, but they were combined together as the model differences were minor. Two Weather Research and Forecasting (WRF) versions were used (WRF361H and WRF381P) but were considered as two different models due to a number of differences in parameterizations and implementation. García-Díez et al. (2015) showed that the spectrum of possible parameterizations in WRF could lead to a very large model spread. The ALADIN RCM was also used in two different versions with more than 10 years of model development between the Version 5.3 (hereafter ALADIN53, Colin et al., 2010) used at the beginning of EURO-CORDEX and the recent Version 6.3 (ALADIN63, Daniel et al., 2019). Two versions of COSMO-CLM are also part of the model ensemble: COSMO-CLM4–8-17 (named CCLM in the following, Rockel et al., 2008) and an accelerated version, COSMO_crCLIM-v1–1 (hereafter COSMO-crCLIM, Leutwyler et al., 2016), which should also be treated as separate models.

RCM characteristics are briefly described in the references mentioned in supporting information Table S1. For parts of the simulations, some fields other than temperature and precipitation were not available (see footnotes in Table 1). Additional issues arose in the course of the EURO-CORDEX runs but were not considered to influence the results. They are listed in the so-called Euro-CORDEX Errata Table (https://www.euro-cordex.net/078730/index.php.en). For instance, RCM simulations using model-level data from the historical CNRM-CM5 GCM run had lateral boundary conditions that were taken from another simulation member by mistake. Therefore, their atmosphere lateral boundary conditions are not consistent with sea surface temperature and sea ice cover surface boundary conditions. This was however found to not change model climatologies. Therefore, simulations (v1) were kept unless a new simulation (v2) using consistent pressure-level boundary conditions was carried out to replace it. Early simulations from WRF331F with a known problem of sea surface temperature (SST) interpolation along some coastal areas were removed from the ensemble and replaced by simulations with the new version WRF381P.

In this article we use simplified simulation naming, described in supporting information Table S1 in order to shorten references to models. The simplified naming takes the form "<GCM>r<n>-<RCM>," where GCM is a simplified model name, for example, "HADGEM" for HadGEM2-ES, n is the member number, and RCM is the simplified RCM name, for example, "CCLM."

3. Evaluation Framework

We focus on a subset of variables and indices linked to some economic sectors, intending to provide a broad view of the ensemble's capacity to represent useful information for further use in science and decision making. Note that for some applications, model outputs will need to be bias adjusted (Maraun, 2016); however, bias adjustment is beyond the scope of this article and therefore the evaluation is conducted on the raw model output.

We use the most recent World Meteorological Organization (WMO) 30-year reference period 1981–2010 for the comparison of model outputs with observations. Since this period exceeds the CMIP5 historical period ending in 2005, we concatenated RCP8.5 scenario simulations to historical simulations to complete the data. We do not expect the scenario selection to affect the results of the evaluation because the overall temperature development is very similar across scenarios in the first years of the scenario period (starting 2006).

Models are assessed by calculating biases as maps or by averaging biases over the so-called PRUDENCE REGIONS (Rockel & Woth, 2007), which are described in supporting information Figure S1 (BI = British

Table 2
Variables and Indices Used in This Study With Reference Data Sets, Evaluation Period, and Concerned Economic Sectors for the Indices

Variable	Acronym	Essential climate variable (ECV) name in climate and forecasts convention	Observation data set	Season	Remark
daily mean temperature	TG	tas	E-OBS17 022	Year, DJF and JJA	
daily max temperature	TX	tasmax	E-OBS17 022	Year, DJF and JJA	
daily min temperature	TN	tasmin	E-OBS17 022	Year, DJF and JJA	
daily precipitation amount	RR	Pr	E-OBS17 022	Year, DJF and JJA	
daily mean surface wind		sfcWind	ERA5	Year	
sea level pressure		psl	ERA5	Year	
surface solar radiation		rsds	HELIOSAT	Year	
evaporation		evspsbl	/	DJF and JJA	
Runoff		mrro	/	DJF and JJA	
soil moisture		mrso	/	DJF and JJA	
Index		ECV involved CF variable name	Observation data set used for assessment	Link to sector	Category of extreme or impact-oriented index
Yearly maximum of TX	TXx	Tasmax	E-OBS17 022	All	Warm
#days/year TX > 35°C	TX35	Tasmax	E-OBS17 022	agriculture, health	Warm
#days/year wet bulb globe temperature > 31°C	WBGT31	tasmax, huss, and psl	ERA5	Health	Warm
Length of frost-free period	LFFP	Tasmin	E-OBS17 022	agriculture, ecosystems	Warm
Growing degree days > 5°C	GDD	Tas	E-OBS17 022	agriculture, ecosystems	Warm
Cooling degree day > 22°C	CDD	tas, tasmin, tasmax	E-OBS17 022	Energy	Warm
Yearly minimum of TN	TNn	Tasmin	E-OBS17 022	All	Cold
#Frost days/year	FD	Tasmin	E-OBS17 022	agriculture, ecosystems	Cold
Heating degree day < 15.5°C	HDD	tas, tasmin, and tasmax	E-OBS17 022	Energy	Cold
Days with RR > 99th percentile of daily amounts for all days	R99a	Pr	E-OBS17 022	Flood	Wet
Highest 1-day precipitation amount	RX1d	Pr	E-OBS17 022	Flood	Wet
6-month Standardized Precipitation Index	DF6	Pr	E-OBS17 022	agriculture	Dry
annual maximum surface wind	SWXx	sfcWindmax	ERA5	insurance	Storm

Isles; IP = Iberian Peninsula; FR = France; ME = Mid-Europe; SC = Scandinavia; AL = Alps; MD = Mediterranean; and EA = eastern Europe). The choice of regions is always subjective and should in principle be based on considerations of climate homogeneity or on national boundaries, if national climate services are targeted. However, many studies now have used the PRUDENCE regions and results can be inter-compared by readers, hence our choice here.

Biases are calculated by simple differences between model outputs and a reference observational or reanalysis data set (see section 3.2 and Table 2). In some cases, biases are expressed as % of observations. We also compare the mean RCM biases with the mean biases of the driving GCMs in two ways: (i) by a simple average over the driving GCM ensemble and (ii) by a weighted average counting a GCM as many times as it is used to drive a RCM. This distinction is made to investigate the potential effect of the imbalanced weight of some GCMs in the biases.

3.1. Investigated Variables and Indices

The indices we consider are chosen to represent basic variables and processes (temperature, precipitation, surface solar radiation, near-surface wind speed, and mean sea level pressure, MSLP) and a few more elaborated indices for key sectors (agriculture, health, energy, infrastructure, and water management) and are presented in Table 2. Some of these indices are borrowed from the classical ETCCDI catalog (Zhang et al., 2011). However, we also added indices relating to extremes and sector impacts that were not clearly covered, such as energy or health. An intercomparison of water balance variables (evaporation, runoff, soil moisture, and precipitation) is also carried out. Three extreme heat indices are considered: annual maximum temperature (TXx), the number of days with maximum temperature exceeding 35°C (TX35), a typical threshold for assessing agricultural yield impacts (Deryng et al., 2014), and the number of days with wet bulb globe temperature (WBGT) greater than 31°C the latter measuring the combined effect of heat and humidity

on human health. We use a simplified WBGT version (see, e.g., Fischer & Knutti, 2013) which represents conditions in the shade. Above a WBGT value of 31°C frequent rest periods even for light work activity are recommended (Kjellstrom et al., 2009). Two standard indices for agriculture and ecosystems are used (growing degree-day above 5°C, GDD, and length of frost-free period, LFFP), both characterizing climate effects on plant phenology. Two classical energy-demand indices are also used (cooling degree-day, CDD above 22°C, and heating degree-day below 15.5°C, HDD, using the definition of Spinoni et al., 2015), each characterizing energy demand in summer or winter. For precipitation, we analyze two extreme indices (the 99th percentile of all-day daily precipitation amount, R99a, and the annual maximum of daily amount, RX1d), while for drought occurrence we use the drought spell frequency based on the Standardized Precipitation Index (SPI) cumulated over 6 months, using the approach of Spinoni et al. (2014). A drought spell starts when the cumulative precipitation falls below one standard deviation (STDE) for at least two consecutive months and ends when it goes back above normal. We express the index as the frequency of drought spells per decade, DF6. As wind storm index we use the annual maximum of daily maximum instantaneous wind speed (SWXX). This selection of indices is somewhat weighted toward temperature indices but illustrates a wide range of potential impacts.

3.2. Observations

The observations used here for precipitation (pr) and temperature (tas, tasmin, and tasmax) are taken from the E-OBS17 0.22° data set (Haylock et al., 2008) and are interpolated onto a common 0.11° resolution grid either by using a bilinear interpolation (for temperature) or by nearest-neighbor interpolation (for precipitation). This choice was made in order not to lose any effect of the high-resolution simulations compared to the GCMs. While for temperature we do not expect that the lower resolution of the observations would affect results except in mountainous areas, we expect the observational resolution to affect precipitation. Thus, for precipitation in mountainous regions, we also use a number of high-resolution gridded data sets (Fantini et al., 2018) as an alternative for the relatively coarse resolution E-OBS data. High-resolution data sets, albeit only available for relatively small regions, are often compiled from a higher density station network than E-OBS, which reduces the chance of undersampling (Prein & Gobiet, 2017). Also, different interpolation methodologies generally lead to different gridded precipitation products (Herrera et al., 2016). Here we use high-resolution precipitation data sets covering the Alps (EURO4 M-APGD; Isotta et al., 2014), Norway (KLIMAGRID; Mohr, 2009), Scotland and Wales (U.K. gridded data set; Perry & Hollis, 2005), Spain and Portugal (Herrera et al., 2019), and the Carpathians (CARPATCLIM; Szalai et al., 2013). Each data set is confined to a single PRUDENCE region, respectively AL, SC, BI, IP, and EA. They all span the period 1981–2010, apart from EURO4-M-APGD which ends in 2008 (1979–2008 was taken). Conservative remapping is used to regrid the high-resolution data set onto the EUR-11 grid. The regridded pattern is then substituted for the E-OBS pattern; for subregions within a PRUDENCE region that are not covered by the high-resolution data set, the E-OBS values are retained.

For the dynamical variables (MSLP and surface wind) and wet bulb globe temperature the new ERA5 reanalysis (C3S, 2017) is used as reference over the period 1981–2010. These reanalyses are model outputs constrained by observations, which therefore can also include model biases. For radiation, the surface short-wavec downwelling radiation (rsds) annual means are evaluated against the Surface Solar Radiation Data Set—HELIOSAT (Müller, Pfeifroth, Träger-Chatterjee, Cremer, et al., 2015), which is a remote-sensed surface radiation data set. The latter is interpolated to the common EURO-CORDEX horizontal rotated 0.11° grid for the period 1983–2012 (not available before). For this variable, the mean absolute bias for monthly means is equal to 5.5 W/m^2 in the HELIOSAT product (Müller, Pfeifroth, Träger-Chatterjee, Trentmann, et al., 2015). Therefore, only absolute biases higher than this value are considered as significant.

4. Mean Biases for Main Climate Variables

4.1. Temperature

We first analyze the overall temperature bias of the ensemble. Figure 1 (top row) shows the median and 5th and 95th percentiles of the temperature biases among the 55 RCM simulations for the winter season. Note that these percentile maps do not necessarily come from the same model at all grid points, and, provided that all models are not biased in the same way, we expect the 5th percentile map to include essentially negative
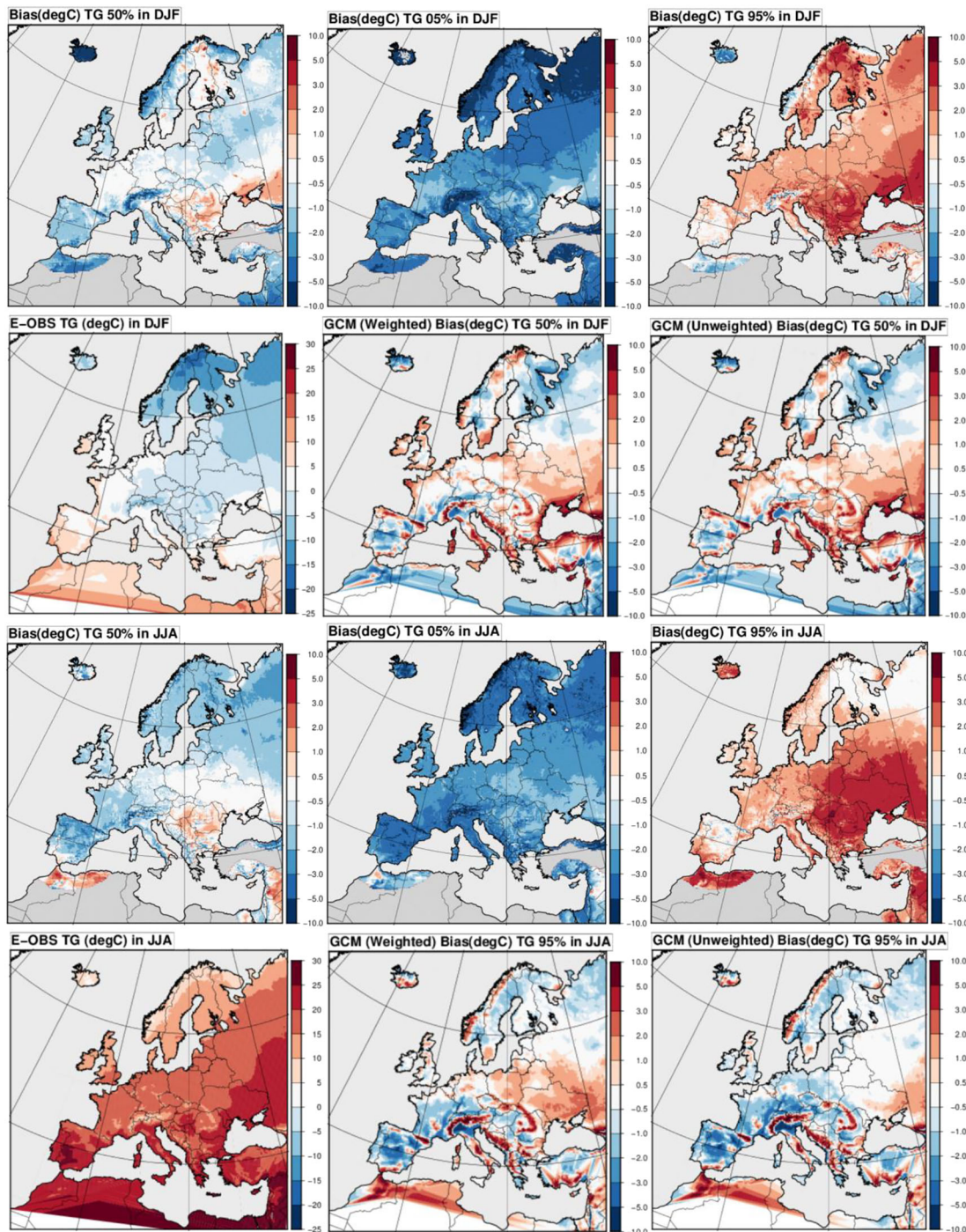
Figure 1. Distribution of temperature biases (°C) for the winter season (top two rows) and the summer season (bottom two rows). First row, left: median of the mean temperature bias among the 55 RCM simulations, middle: 5th percentile of the RCM temperature bias distribution, and right: 95th percentile of the RCM temperature bias distribution; second row, left: E-OBS mean observed temperature, middle: median of the biases of the eight GCMs used 55 times (including repetitions), and right: median of the eight GCMs biases without repetition; bottom two rows: same as top two rows for the summer season.

numbers and the 95th percentile map to include positive numbers. Median biases remain within ±2°C, except in scattered areas such as North Africa, where the observations are likely less reliable, and over high mountains (Norway, Alps), where the altitude difference between model and observational data sets was not considered when estimating biases. However, a group of models (e.g., RACMO, WRF381P, and RCA) has strongly negative biases over Scandinavia (see Figure 2) and northeastern Europe, which is reflected by
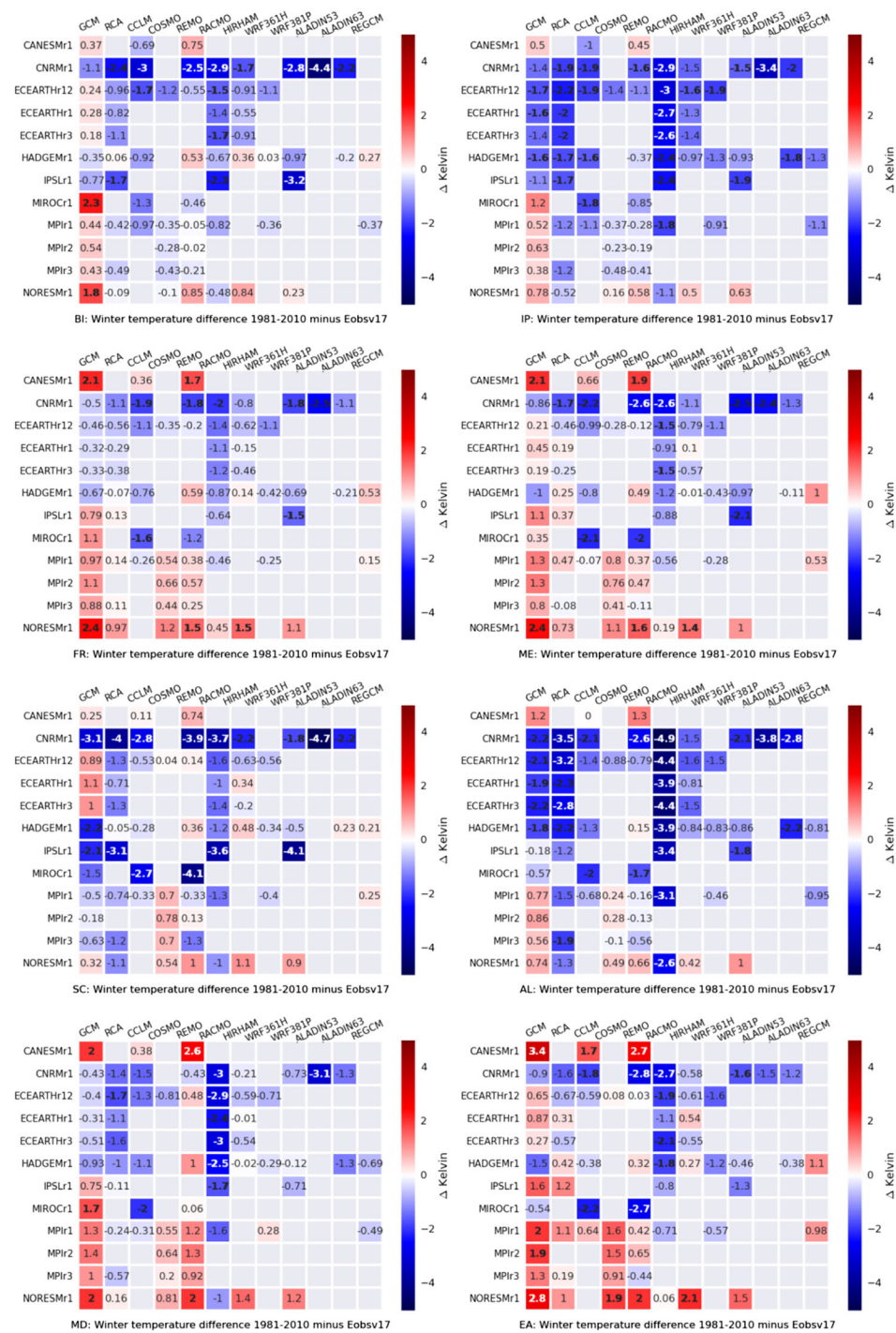
Figure 2a. Temperature biases (°C) per season and per PRUDENCE region (see supporting information Figure S1) with respect to E-OBS. Bold numbers indicate biases greater than the average standard deviation of EOBS. Standard deviation is provided along with E-OBS v17 and was averaged over days and regions. Rows represent different GCMs. The first column indicates the bias of each GCM, the other columns the biases of the RCMs. (a) Winter (DJF) and (b) summer (JJA).
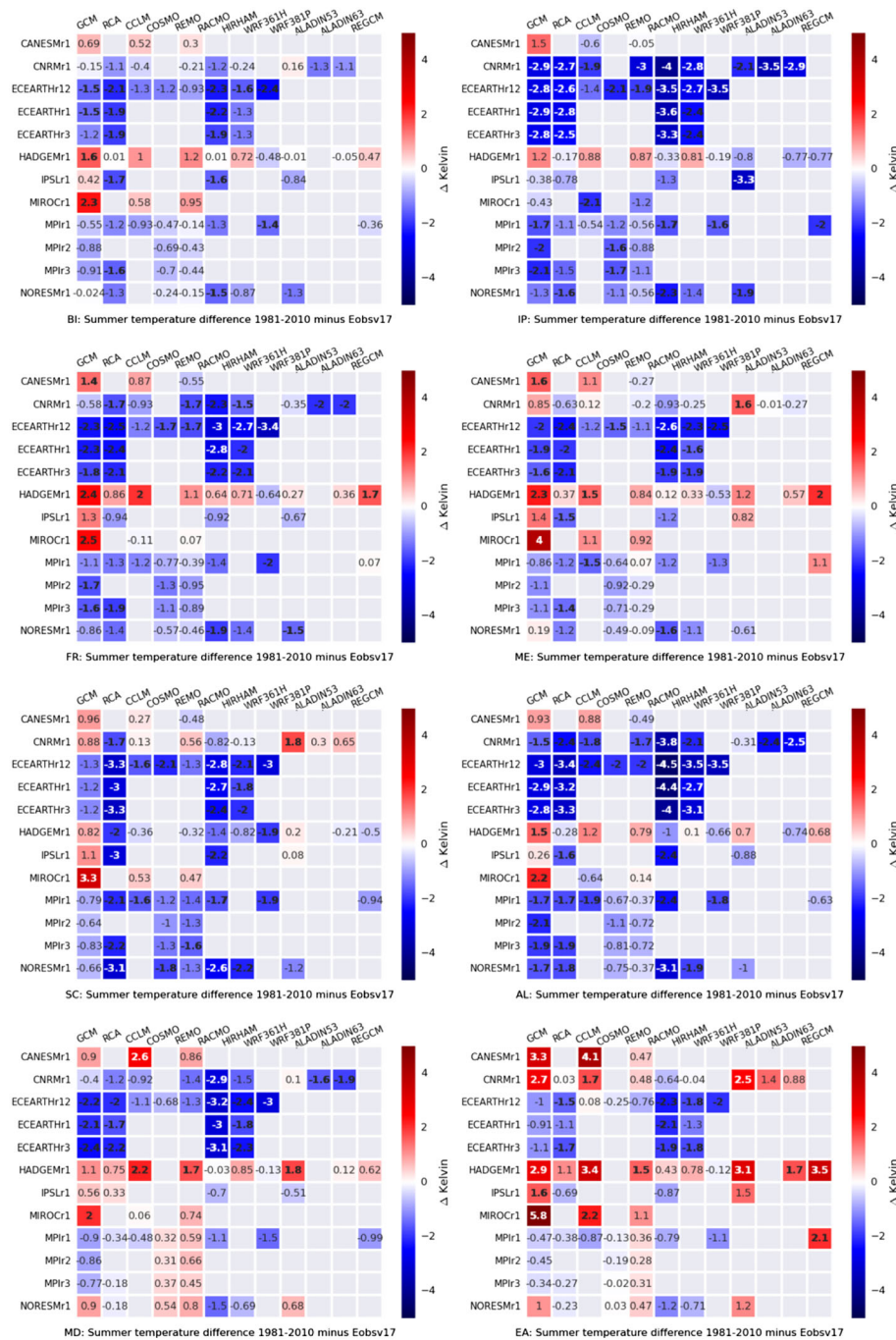
(b)

Figure 2b. (continued)

the 5th percentile of mean bias being lower than −5°C (Figure 1). This could be linked to a bias over snow-covered areas in winter, indicating shortcomings in snow-atmosphere interaction as found in previous studies (García-Díez et al., 2015). By contrast a group of models has a strong positive bias over southeastern Europe, mostly driven by a few GCMs (Figure 2) in this case (CANESM, IPSL, MPI, and NORESM). In summer, a weak median negative bias (<2°C) is present over a large area covering the Iberian Peninsula, western Europe, and Scandinavia. This is largely driven by negative GCM biases over the same areas, particularly

marked for ECEARTH, MPI, and NORESM. A median positive bias is present over eastern Europe, driven by a group of GCMs (CANESM, CNRM, HADGEM, and MIROC) having a strong mean positive bias over the region in the range 2.5–6°C.

Essentially the same conclusions hold for the mean maximum daily temperature (not shown). In this case the ensemble median bias is stronger, with substantial underestimations in the PRUDENCE region boxes BI, IP, SC, and AL (not shown). Regarding minimum daily temperature (not shown), biases appear to be more tied to the RCM, pointing to local land-atmosphere interactions combined with surface, boundary layer, and low-cloud processes driving much of the bias structure. For instance, RACMO exhibits systematic negative biases and REMO almost systematic positive biases, illustrating how the biases patterns are often very specific for each RCM. Ensemble mean biases are less pronounced, as biases are more diverse among models and thus tend to cancel out.

One question is how these biases compare with observational errors coming from various sources, including the gridding scheme, which was shown to be significant (Cornes et al., 2018; Prein & Gobiet, 2017). E-OBS v17 data also include an estimate of the error STDE for each day and each grid point. The STDE is typically estimated to be within 1.3–2°C on average at the grid point level (not shown), both in summer and winter, which is mostly higher than the ensemble mean bias, except for some small regions. However, on a model-by-model basis this is not the case. Figure 2 shows in which regions and for which models the mean bias is higher than the STDE in absolute value. In winter, several RACMO, ALADIN53, and RCA simulations and CNRMr1 downscalings show a significant negative bias in several regions, while NORESM downscalings show a significant positive bias. In summer, strong negative biases are found for downscalings of the EC-EARTH and CNRM models, while several HadDGEM downscalings have a positive bias higher than the STDE. RACMO keeps a strong negative bias. While these comparisons are interesting at a qualitative level, caution should be taken when comparing the STDE with the mean bias amplitudes, as the observation error does not include only a systematic component; that is, it is not strictly a bias.

In order to better understand the drivers of the biases, we calculate, for each variable (mean, maximum, and minimum daily temperatures), the average over all GCMs and PRUDENCE regions, of the variance of GCM-RCM simulation biases averaged within groups driven by the same GCM:

$$\overline{V}_i = \frac{1}{N_{Region}}\sum_{Region\,k}\frac{1}{N_{RCM}}\sum_{RCM\,j}\left(B_{ijk}-\overline{B}_{ik}\right)^2, \tag{1}$$

where B_{ijk} is the mean bias of the RCM #j downscaling GCM i in region k and $\overline{B}_{ik}$ is the mean bias over all RCMs downscaling GCM #i. This mean variance is normalized by the overall bias variance V_{tot} considering all simulations and regions, to form a "mean within-GCM normalized variance" (WGNV).

$$WGNV_i = \frac{\overline{V}_i}{V_{tot}} \tag{2}$$

If the GCM-RCM bias is a function of the GCM only, the normalized variance should vanish. The average variance is weighted by the number of simulations within each GCM-driven group. A symmetric operation is done with within-RCM sub-ensembles instead of GCM sub-ensembles to calculate a "mean within RCM normalized variance" (WRNV).

$$WRNV_j = \frac{1}{V_{tot}}\frac{1}{N_{Region}}\sum_{Region\,k}\frac{1}{N_{GCM}}\sum_{GCM\,i}\left(B_{ijk}-\overline{B}_{jk}\right)^2, \tag{3}$$

where now $\overline{B}_{jk}$ is the mean bias over the ensemble of GCMs downscaled by RCM #j. Then WRNV is plotted versus WGNV in Figure 3. Only temperature is discussed here, while other variables are discussed in their relevant subsections. In this graph, when a point lies near the diagonal, one expects an equal contribution of RCM versus GCM in driving the bias, while when the point lies above (below) the diagonal, the GCM (RCM) contribution to the bias dominates.

For mean and maximum daily temperature, we find that the driving GCM is contributing more than the RCM itself, but for daily minimum temperature, the most important contribution is the RCM. In
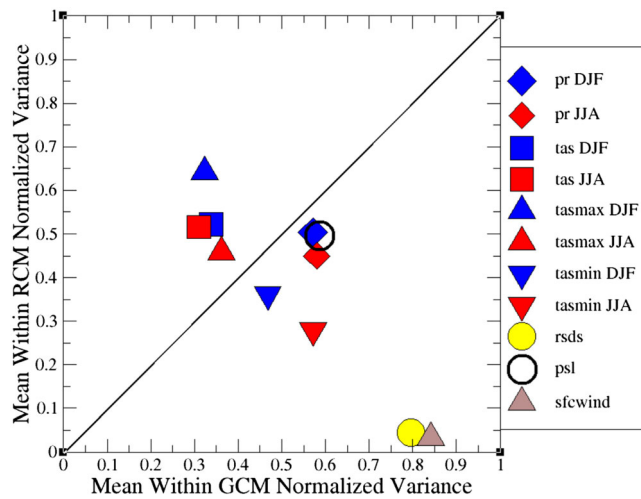
Figure 3. Analysis of the variance of biases by grouping either by GCM or RCM: Ratio of the normalized within RCM bias variance versus the normalized within GCM variance, averaged over all PRUDENCE regions and for a few variables and seasons (blue = winter, red = summer, square = mean temperature, triangle up [down] = max [min] temperature, diamond = precipitation, and circle = yearly mean radiation). The WRNV is the weighted average over all RCMs of the variance of the ensemble of GCM-RCM biases with the same RCM; the within GCM normalized variance is the weighted average over all GCMs of the variance of the ensemble of GCM-RCM biases with the same GCM. A zero RCM variance (e.g., for rsds) means that the bias is the same for all GCMs downscaled by this RCM.

general, during nighttime, the atmosphere is stratified, making model biases substantially dependent on stable boundary layer parameterizations, boundary layer turbulence, low-level clouds, and the description of the land surface. This dependence is expected to be larger in summer than in winter due to lighter winds inducing less mixing than in winter, which is consistent with more RCM-driven biases during summer. By contrast, for daily maximum temperatures in fully developed diurnal boundary layers, one expects that the surface temperature influence is mainly driven by large-scale processes (and therefore the GCM), especially in winter, which is consistent with our bias analysis. However, maximum temperatures are also expected to depend on soil-atmosphere interactions.

These results however exhibit variability among European regions. The WGNV and WRNV have been calculated also without averaging across regions (supporting information Figure S2). Daily average and maximum temperatures have large-scale GCM dominating contributions in flat areas, while in the Alps and Mediterranean a reverse situation is found. For minimum temperature the RCM contribution dominates over the GCM contribution in most cases, and it is systematically larger than for average temperatures.

4.2. Precipitation

Spatial maps of the difference between model precipitation and E-OBS data are depicted in Figure 4 for the winter (December–February, DJF) and summer (June–August, JJA) seasons. The median bias and the 5% and 95% bias of the model distribution are shown. In the winter season there is a widespread overest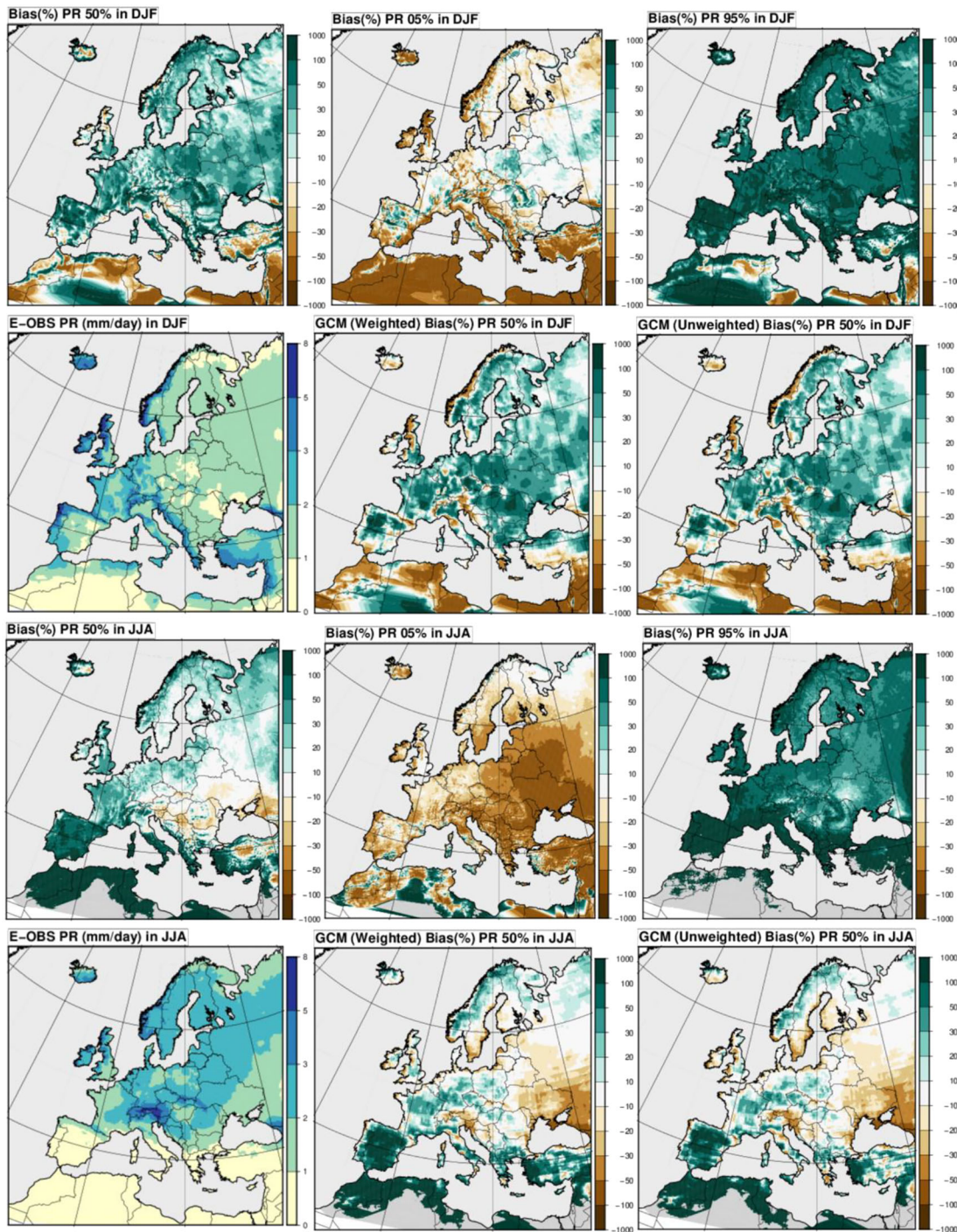imation across virtually the entire European land mass, with relative median bias values reaching 50% (a factor of 1.5) over a substantial number of regions (e.g., mid-northern Iberian Peninsula, Central France, Alpine regions, Po Valley, and eastern Europe). Even the lower tail (5%) of the bias distribution is positive in certain regions (e.g., area in Spain, France, Poland, the Alps, and Romania) meaning that at least 95% of the simulations have a positive bias. The upper tail (95%) of the bias distribution exceeds 100% in a large number of regions (a bias of a factor of 2 at least). Such biases found here are consistent with previous analyses (e.g., from Kotlarski et al., 2014), but they have to be considered in view of the fact that the E-OBS data do not include an undercatch gauge correction, which in the winter can reach up to 35%, especially in mountain environments (e.g., Adam & Lettenmaier, 2003).

In summer, median biases are also positive but smaller than in winter, with the exception of the Mediterranean region. Here rain is essentially absent in observations, while the majority of simulations produce rain through predominantly (parameterized) convective events. In other areas precipitation is overestimated by 30% or less. Over central and eastern Europe there is no systematic bias in summer precipitation, while over areas closer to the Atlantic Ocean and the North Sea there is a small but distinct positive bias. A small negative bias is found over Ukraine and Southern Russia areas. The lower tail of the bias distribution is negative virtually everywhere with values ranging from slightly less than 0 in Scandinavia and western Europe to nearly 100% in the Balkan, Ukraine, and Southern Russia; exceptions are a few scattered regions in Spain and Italy where even the 5% bias is positive. The upper tail of the bias distribution is positive everywhere, with values ranging from 10% in South Germany and some regions in the Balkan and Ukraine to 100% or beyond in the Mediterranean, western Europe, and Norway.

Median bias spatial patterns in GCMs, both weighted and unweighted, are qualitatively similar to those inferred from the RCMs, both in winter and in summer, but the bias values are generally slightly smaller (Figure 4). An exception occurs in mountainous regions (Alps, Pyrenees, Western Norway, Western Scotland, ...) where median biases in GCMs are distinctly smaller than in RCMs or, in particular in winter, even negative. The latter can most likely be attributed to the difference in horizontal resolution, with much lower orography in GCMs. While all RCMs are operated at a nominal 12-km resolution, the finest resolution

Figure 4. Same as Figure 1 for precipitation biases; units: biases in %, climatologies in mm/day.

among the GCMs is 125 km. Again, in these regions, as during winter the undercatch bias in observations can be quite large.

In Figure 5 the mean bias in precipitation amount (%) for each GCM-RCM combination averaged across each of the eight PRUDENCE regions (BI, IP, FR, ME, SC, AL, MD, and EA) is summarized in tabular form. In winter, in all regions except the British Islands (BI), all RCMs show positive biases, with the exception of RegCM for a few regions. For BI a few more RCMs (CCLM, REMO, ALADIN63, and REGCM) have a negative bias (RegCM max −1.1 mm/day). Regarding the PRUDENCE regions, the largest positive biases are seen
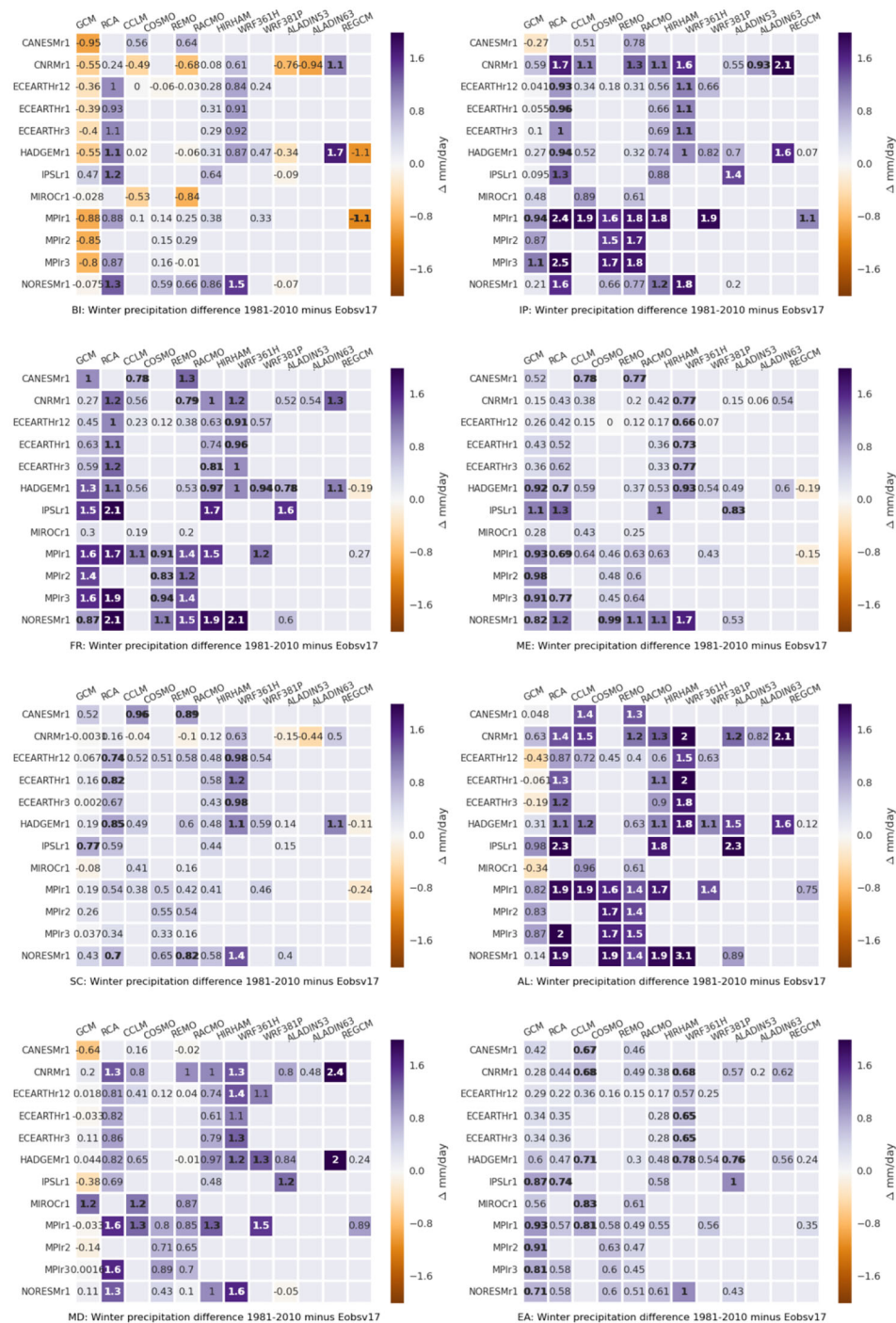
Figure 5a. Same as Figure 2 for precipitation biases as compared to E-OBS. (a) DJF and (b) JJA.

for the Alps (AL) and, to a lesser extent, the Iberian Peninsula (IP), France (FR), and the Mediterranean (MD). Stratified on the driving GCM, the largest biases are found for the simulations forced by MPI, NorESM, and CNRM. When based on RCM, the largest biases are found for RCA (max 2.5 mm/day in IP), CCLM (max 1.9 mm/day in IP and AL), COSMOcrCLIM (max 1.9 mm/day in AL), REMO (max 1.8 mm/day in IP), RACMO (max 1.9 mm/day in FR and AL), HIRHAM (max 3.1 mm/day in AL),
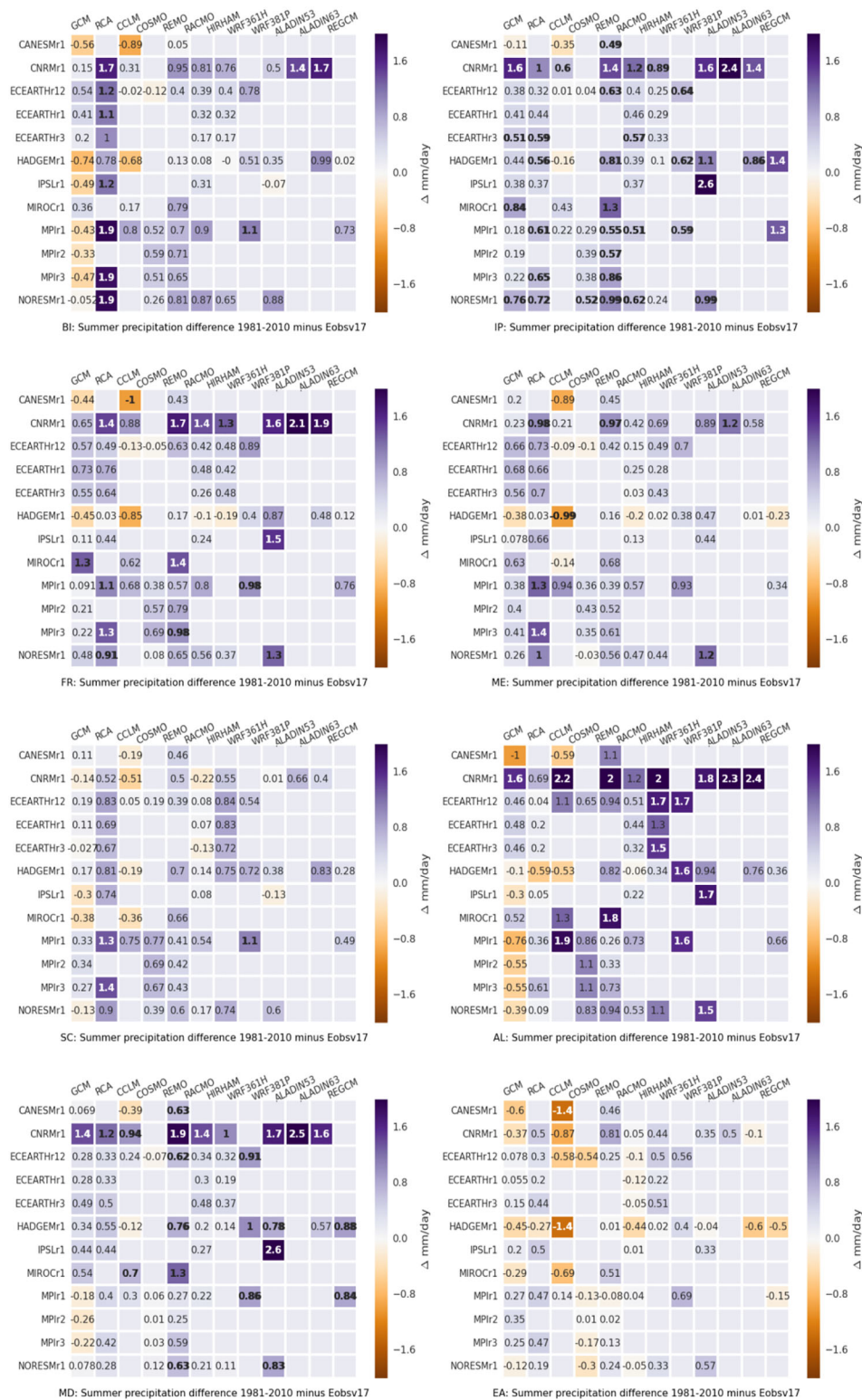
(b)

Figure 5b. (continued)

WRF361H (max 1.9 mm/day in IP), WRF381P (max 2.3 mm/day in AL), ALADIN63 (max 2.4 mm/day in MD), and RegCM (max 0.89 mm/day in MD). The GCMs have largest positive biases in FR and largest negative biases in BI. The latter obviously relates to the large relative negative bias in GCM-precipitation for Western Scotland, which dominates the bias in the absolute amount.

In summer, the picture is somewhat more mixed. For IP and MD all GCM-RCM combinations show a positive bias, confirming Figure 4, with the exception of a few combinations including CCLM. For the other regions (BI, ME, SC, FR, and AL) gradually more combinations have a negative bias. Most negative biases and also the largest values (CCLM max −1.4 mm/day) are found for EA, again in line with Figure 4. Stratified on the driving GCM, we note the large positive biases in all regions but SC and EA for combinations forced by CNRM. This applies to a lesser extent to MPI. Based on RCM, the largest positive biases are seen for ALADIN63 (max 2.4 mm/day in AL, 2.6 in MD and IP), RCA (max 1.9 mm/day in BI, 1.4 mm/day in FR and ME), CCLM (max 2.2 mm/day in AL), REMO (max 2.0 mm/day in AL, 1.9 in MD, and 1.7 in FR), WRF361H (max 2.0 mm/day in AL), and WRF381P (1.7 mm/day in AL).

The overall picture is that RCMs in general tend to overestimate precipitation but more so in winter than in summer and probably amplified by the lack of undercatch gauge correction in E-OBS. GCMs show the same behavior but generally with smaller bias values. An exception occurs in mountainous regions where RCM biases are larger than in other regions, both in a relative and an absolute sense. GCMs, on the other hand, have smaller or even negative biases in mountainous regions. The analysis of variance of Figure 3, carried out in the same way as for temperature, shows that mean precipitation are half driven by GCMs and half by RCMs. This can be understood as both large-scale drivers and local physical parameterizations (e.g., convection and microphysics) are important in determining the biases. The variability across regions of these contributions is large (supporting information Figure S2).

In winter, the ensemble mean precipitation biases are generally of the same order as the average STDE of the observation error (typically 0.5–1 mm/day over plain areas) or higher in many places (not shown), indicating a significant systematic bias by most models. Except for the British Isles and Scandinavia, the biases are found significant for many models when data are regionally averaged (Figure 5a). We remark the almost systematically significant biases of MPI downscalings for IP, FR, and AL, for the HIRHAM RCM in most regions. In summer, biases are less significant except in some models or regions: downscalings of CNRMr1 in MD, FR, and AL, WRF381P over most regions, and RCA simulations over BI and IP.

We have examined how the use of gridded high-resolution data sets (see section 3.2) affects the interpretation of Figure 4. The general finding is that replacing part of the E-OBS data by high-resolution data always leads to larger estimates of observed 30-year mean precipitation, irrespective of PRUDENCE region and season. For DJF the absolute/relative increases are in mm/day/%, 0.03/0.9 (AL), 0.46/24 (SC), 0.26/7.7 (BI), 0.42/21 (IP), and 0.06/5.6 (EA), while for JJA they are 0.38/12 (AL), 0.24/9.8 (SC), 0.17/7.2 (BI), 0.17/24 (IP), and 0.11/5.0 (EA). Considerig these numbers, all negative biases found in Figure 5 become more negative, while all positive biases become less positive, or even slightly negative. Therefore, the simulations have generally lower biases than shown in Figure 5. However, while the magnitude of the adjustment induced by the high-resolution data appears modest, it should be kept in mind that none of the five PRUDENCE boxes are fully covered by the high-resolution data sets, with only three regions (AL, BI, and IP) covered by more than 75% of the area. For the other two regions (SC and EA) the coverage is between 20 and 30%, yet the parts that are covered have sizeable contributions to the mean precipitation averaged over the full PRUDENCE region.

4.2.1. Water Balance

Seasonal means (DJF and JJA) of precipitation (pr), total runoff (mrro), evaporation (evap), and soil moisture (mrso) for the period 1981–2010 have been calculated for up to 55 RCM simulations (55 for precipitation, 54 for total runoff, and 51 for evaporation and soil moisture). They are presented here for model intercomparison purposes only, due to lack of suitable observations for the latter three variables. Absolute values averaged over the eight sub-regions for *pr*, *evap*, *mrro*, and *mrso* are shown in Figure 6. For precipitation, we see some well-known patterns across the matrix. Regions IP and MD are dry in JJA, while BI and AL are wet in both seasons. As expected, evaporation is low for all models in DJF and much higher in JJA for all regions except IP and MD, where evaporation is limited by low soil moisture amounts.
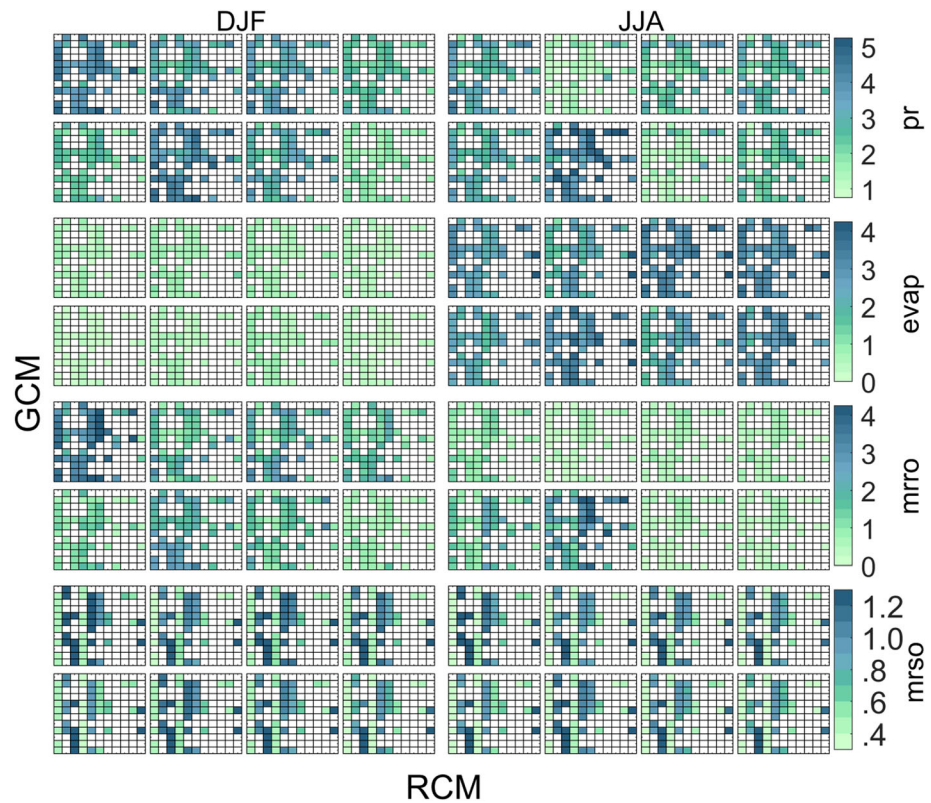
Figure 6. The 1981–2010 means for 55 RCM simulations divided into a GCM/RCM matrix where the rows (GCMs) and columns (RCMs) match those in Table 1. The left column of panels is for DJF and right column for JJA. The variables shown are precipitation (first row, units mm/day), evaporation (second row, units mm/day), runoff (third row, units mm/day), and soil moisture (fourth row, units m). For each variable and season, the data are divided into eight PRUDENCE regions: British Isles, Iberian Peninsula, France, and Mid-Europe in the first row and Scandinavia, Alps, the Mediterranean, and eastern Europe in the second row.

For total runoff, we see high values over BI in DJF and AL in JJA, while for total soil moisture content, we note some large differences across RCMs. RCA, REMO, ALADIN53, and ALADIN63 show small values for all sub-regions and both seasons while CCLM, RACMO22, REGCM, and COSMOcrCLIM show large values, especially for sub-region IB. This can probably be ascribed to different definitions on the soil moisture in the models. For instance, SMHI-RCA, with relatively small values of total soil moisture content, has a field capacity limit near 0.3 m^3/m^3 and a maximum total soil column of about 2.5 m giving a maximum water holding capacity of about 750 mm (Samuelsson et al., 2015). On the other hand, KNMI-RACMO, with relatively large values of total soil moisture content, has a total soil column of 2.89 m while the field capacity limit is in a range 0.2–0.6 m^3/m^3 giving a maximum water holding capacity in the range 600–1,700 mm (van Meijgaard et al., 2012).

4.3. Dynamics

Dynamics here is characterized by two complementary variables. The near-surface wind speed is key for several sectors such as agriculture and energy, and the MSLP allows to characterize the mean large-scale, low-level circulation.

4.3.1. Surface Winds

Mean near-surface winds are positively biased when compared to the ERA5 reanalysis in a fairly homogeneous way (Figure 7). Only a few models show slightly negative biases, for Scandinavia (CNRMr1-ALADIN63, CNRMr1-CCLM, and MIROCr1-CCLM); otherwise, the range of biases is positive, reaching more than 2 m/s over the Alps or the Iberian Peninsula (e.g., for WRF381P or REGCM simulations). Over mountain ranges we expect large differences due to resolution. However, the systematic nature of the positive bias for all models and regions (including plain areas) could potentially be due to a negative bias of the
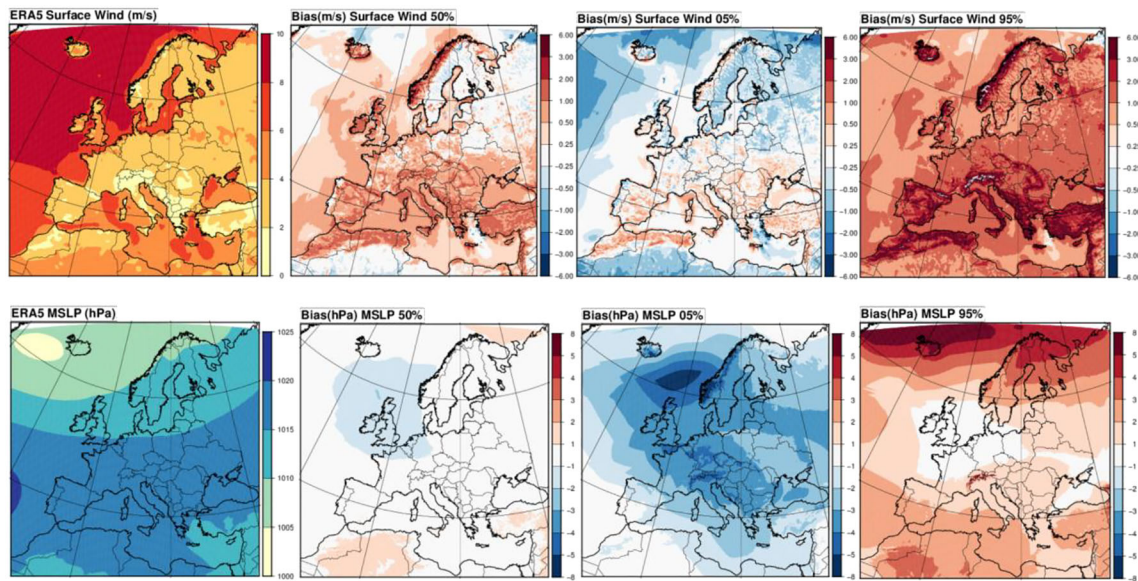
Figure 7. Biases for surface wind speed (top row) and MSLP (bottom row); from left to right: mean values from the ERA5 reanalysis, median RCM bias, and 5% and 95% of the bias distribution.

surface wind speed in the ERA5 reanalysis, which has been noted in previous studies (e.g., Betts et al., 2019). The bias is essentially driven by the RCM in all regions (supporting information Figure S2) as shown by the variance analysis (see Figure 3), and it is probable that differences in the boundary and surface layer schemes account for most of the biases. Finally, note that this analysis has been conducted on an annual basis in order to keep the article concise; however, yearly estimates may hide stronger or weaker seasonal biases.

4.3.2. MSLP

Median and extreme MSLP biases across the ensemble are small, as compared to the ERA5 reanalysis, and do not exceed 2 hPa (Figure 7). In extreme cases, models have negative MSLP biases exceeding 5 hPa over northwestern Europe and across central Europe. This appears as a rather general property of most models since the upper extreme biases almost vanish over these areas. Extreme positive biases exceeding 5 hPa are also present over the northernmost latitudes. Other significant positive biases can be found for latitudes below about 40°N.

MSLP biases seem to have both GCM and RCM as driving factors (Figure 3). Some RCMs have a significant bias, for instance WRF381P over Mid-Europe, with a range from −8.3 to −1.7 hPa. REMO, CCLM, and RCA simulations also have significant negative biases below −3 hPa in a few regions. The reason for these biases remains difficult to establish. However, internal variability may not be symmetrical, and convection-related low-pressure systems which would not be present in GCMs and would develop within the RCM domain could explain the biases in some cases. MSLP and surface wind biases may not be independent: for instance, deeper lows are expected to induce higher winds.

4.3.3. GCM-RCM Pattern Correlation

RCM integrations depend on the consistency of the "one-way" nesting approach, that is, by assuming that the large-scale GCM dynamical forcing imposed at the boundary of the RCM domain is not strongly altered by the interaction with the local-scale forcings and internal dynamics and physics of the RCM.

An assessment of the consistency between GCM and RCM large-scale circulation is most relevant at the daily time scale, since this is the scale at which RCMs produce valuable output which could not be obtained directly from their driving GCMs. Day-to-day consistency of large-scale circulation patterns between RCMs and their driving GCMs has been assessed from the correlation of spatial patterns of daily averaged MSLP. Part of this correlation is due to the existence of near-stationary spatial features over the domain, resulting in correlations different from zero even for independent GCM/RCM realizations. This problem could be solved by estimating correlations from anomalies, but the interpretation of the resulting correlations would be more difficult. We used all RCMs except ALADIN53 due to the inconsistencies mentioned above for this

model. The period 1976–2005 is used for the RCM and GCM MSLP interpolated onto a 0.75° regular grid via a bilinear interpolation. This process is essentially equivalent to filtering out the largest scales from both sets of simulations. The results are analyzed by seasons and summarized as box plots (Figure 8).

The analysis shows the expected result of a very high correlation in winter for all the models, with all the median values larger than 0.9, while in summer the inter-model spread is larger (0.6–0.9 for medians). The results of the intermediate seasons, not shown, are mid-way between the winter and summer values. The seasonal differences for a given RCM simulation can be explained by the strength of the dynamical forcing from the GCM in each season, which is maximum in winter and ultimately determines the intensity of the constraint on the circulation over the RCM domain (e.g., Sanchez-Gomez et al., 2009).

The figures in this study also show rather large model differences, in particular in the lower tail of the correlation distributions; however, these differences are not necessarily an indication of different RCM skill in reproducing the driving large-scale circulation. Factors which may contribute to these differences include the size of the RCM domain or nudging zone (which might explain the results from the WRF381P integrations for which a larger domain was used with a small nudging zone), the GCM resolution, and the strength of the driving circulation (e.g., Laprise et al., 2008). It is worth noting that there are differences between the distributions of correlations; one of the ECEARTH realizations (r3) has systematically lower correlations compared to the other two realizations (r1 and r12), which can be explained with the lower quality of MSLP in the former. In addition, the MSLP is obtained from the surface pressure through a vertical extrapolation from surface pressure and temperature, and different methods can be used for such procedure, possibly affecting the correlations (e.g., Pielke & Cram, 1987, for HadGEM2-ES).

4.4. Surface Solar Radiation

Figure 9 shows the observation pattern (bottom left) and the multi-model ensemble bias distribution for the annual-mean surface solar downward radiation. Note that the ECEARTH r1 et r3 simulations have been replaced by ECEARTHr12 to compute the GCM median bias, as rsds data are not available for these two simulations.

The RCM median bias map shows an overall good behavior, with low maximum positive biases (typically between +10 and +20 W/m^2) in the northeast of the domain and maximum negative biases in the southwest (typically between −10 and −20 W/m^2). The southwest biases cover a region with few clouds but large dust aerosol optical depth (AOD). This suggests that the models have too much cloudiness or too strong dust aerosol effects. Note that most RCMs do not include aerosols, and the annual-mean aerosol direct radiative forcing at the surface has been evaluated to [−19; −15] W/m^2 over Northern Africa and to [−15; −10] W/m^2 over southern Europe (Nabat et al., 2012, 2015).

The RCM median bias map is consistent with the GCM median bias map except for some specific areas. Indeed, the RCM ensemble degrades the behavior with respect to the GCM ensemble above some mountain ranges (Alps, Turkish mountains, southwest Norway, and Iceland) but improves it near the northern coast of Europe and the United Kingdom. The southwestern negative bias is also more pronounced in the RCM ensemble. The explanation of the RCM strong positive biases over the mountains remains unclear and may be also due to limitations of the HELIOSAT reference data set. Overall, there is no obvious improvement in the Euro-CORDEX ensemble compared to the driving CMIP5 ensemble for the bias in annual-mean surface solar downward radiation.

Looking at individual RCM biases, we conclude that, at the first order, biases in surface solar downward radiation in RCM simulations are related to the choice of the RCMs. We notice only a weak influence of the forcing GCM contrary to most of the other variables. Moreover, the influence of the member choice for a given GCM is not detectable. This individual model analysis is confirmed by Figure 3 in which the position of the yellow circle clearly shows that the variance of the rsds bias is dominated by the RCM choice. Supporting information Figure S2 shows that such is the case for all regions. The rsds variable indeed shows the minimum WRNV among all other analyzed variables.

Two RCMs (WRF381P and REGCM) present particularly strong positive biases over land regardless of the geographical area and the driving GCM, suggesting a lack of low-level clouds or too weak aerosol effects. RCA4 also exhibits mainly positive biases. REMO and HIRHAM5 have mainly negative biases, clearly
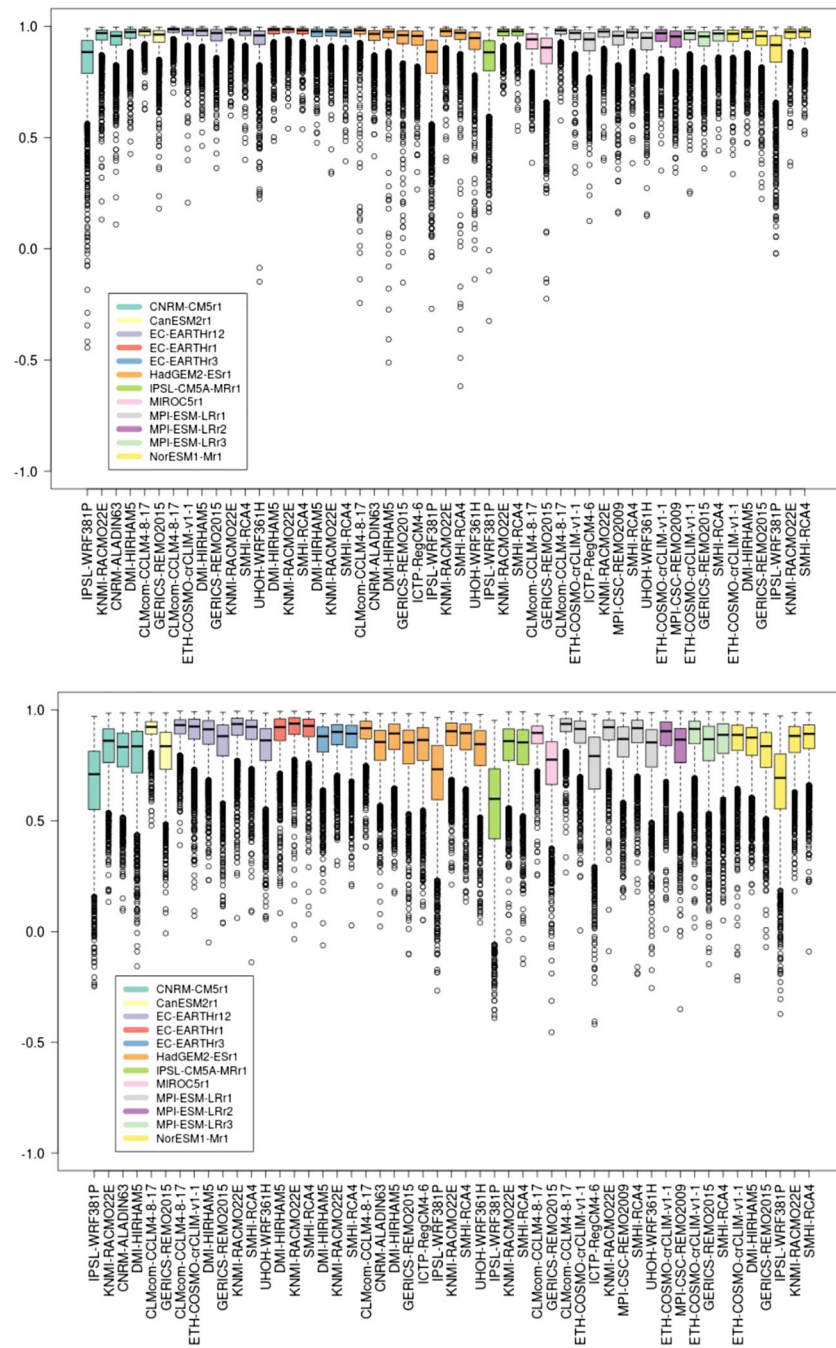
Figure 8. Distribution of mean sea level pressure spatial correlation between RCMs and driving GCMs, for simulation in the historical period (1976–2005). (top panel) Winter (DJF) and (bottom panel) summer (JJA). *X* axis lists the RCM name, and GCMs are identified by different colors, reported in the legend. The boxes are constructed from the 25th and the 75th percentiles, with the median shown as a line inside it, while the whiskers extend to data up a distance of 1.5 times the interquartile range. All the correlations outside the whiskers are shown as points. Note that the WRF361H simulations are included here using a simplified calculation of the sea level pressure from the surface pressure and temperature.

stronger over the southern Europe and the Mediterranean area (even stronger than -40 W/m^2). RACMO22E, ALADIN53, ALADIN63, and CCLM show no strong domain-average bias but a north-south contrast with positive biases in the north and negative in the south close to the median feature in Figure 9. It is noticeable that RCA4 and REGCM show a land-sea contrast in their biases (not shown),
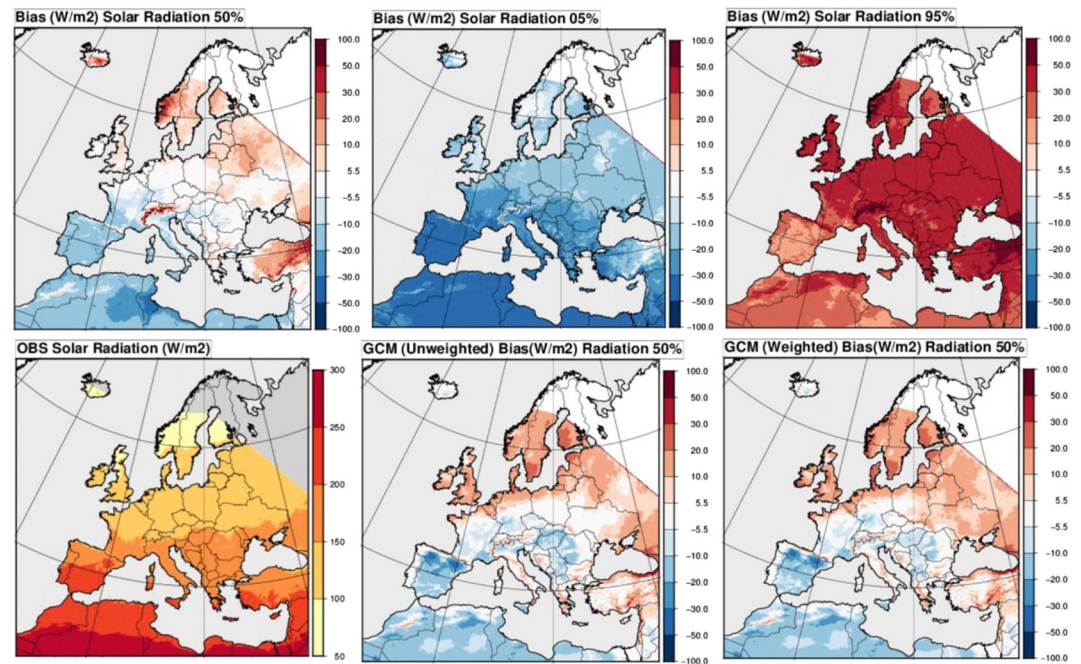
Figure 9. Same as Figure 1 for the annual-mean surface solar downward radiation (rsds) (six panels). White-colored areas (-5.5 to $+5.5$ W/m^2) correspond to areas where the mean bias is lower than the mean absolute monthly observational error (Müller, Pfeifroth, Träger-Chatterjee, Trentmann, et al., 2015).

which remains to be explained. For the Mediterranean Sea, we can compare the current model evaluation with the work by Sanchez-Gomez et al. (2011) for the ENSEMBLES RCMs (ERA40-driven runs). Since this assessment, HIRHAM and REMO did not improve, as they were already showing strong negative biases, REGCM degraded its bias whereas RACMO and CCLM improved and ALADIN and RCA kept their good behavior over the area.

Over continental Europe, the biases obtained here are difficult to compare with previous RCM assessments (Bartók et al., 2017; Nabat et al., 2015) because of different methodologies but the order of magnitude of the RCM biases are similar. It is worth mentioning that Bartók et al. (2017) found that RCMs have generally larger biases than their driving GCMs, which is not true in our evaluation. Besides, evaluation of the past trend of surface solar downward radiation has not been conducted here. However, a clear underestimation of the observed past trend from the 1980s has been reported in the literature for RCMs not taking into account the aerosol past evolution over Europe (Bartók et al., 2017; Boé et al., 2020; Nabat et al., 2014). Finally, an inconsistency in the GCM and RCM future responses for the surface shortwave downward radiation has been described recently (Bartók et al., 2017; Boé et al., 2020; Gutiérrez et al., 2020). This aspect is specifically discussed in the companion paper (Coppola et al., 2021) dedicated to the RCM future projections.

5. Extreme and Impact-Oriented Indices

In this section we assess the biases of impact-oriented indices, as defined in Table 2, organized by broad categories of hazards or involved sectors. The observation mean over the 30-year reference period, together with the median ensemble bias and extreme biases (5% and 95% of the bias distribution), is shown as maps in Figure 10. The detailed information per region and index is provided in supporting information Figures S4–S16.

For the maximum annual temperature (TXx), as well as for mean summer temperature, a general cold bias is found across Europe, except in a few areas in southeastern Europe, but large biases exceeding 5°C in absolute values are found in some cases. Cold biases below -5°C are found for several simulations made by RCA over Scandinavia, HIRHAM over France and central Europe, and IPSLr1-WRF381P over the Iberian Peninsula. Large positive biases are found in eastern Europe for two simulations of CCLM. The number of days with maximum temperature exceeding 35°C (TX35) has a median low bias over the Iberian
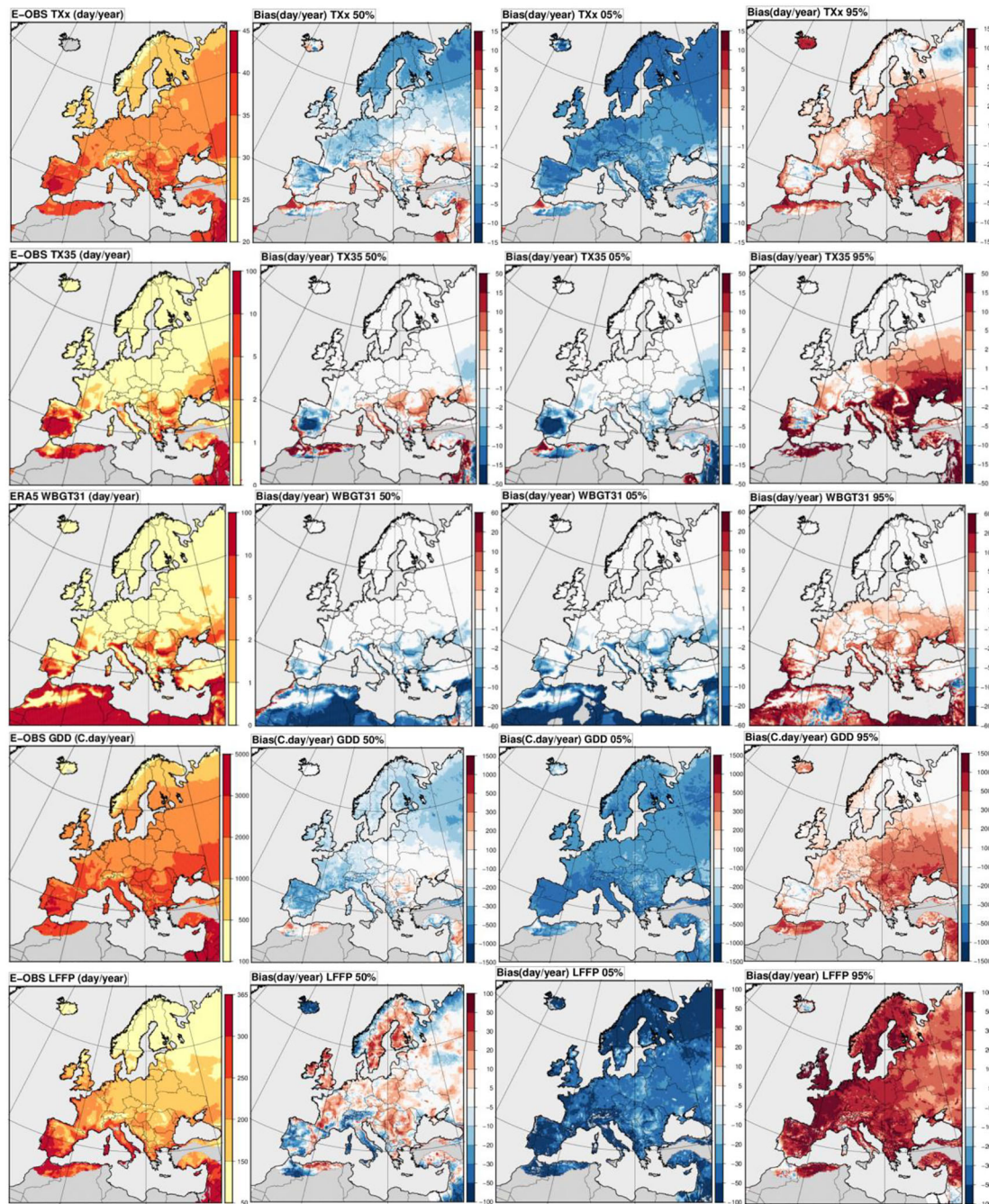
Figure 10. Distribution of biases for the indices described in Table 2; each row includes from left to right: mean observation, median bias and 5–95% biases. Each row stands for an index, and indices are presented in the same order as in Table 2.

Peninsula. In contrast, over the Eastern Mediterranean areas and eastern Europe, general positive biases are found for TX35, sometimes very strong (more than 10 days) such as for REMO or CCLM simulations. WBGT exceeds 31°C only in southern Europe on a substantial number of days each year, particularly in the regions around the Mediterranean and Black Sea. Median WBGT31 exceedances in RCMs are lower than in ERA5 everywhere, except for some coastal regions, indicating that RCMs systematically underestimate health effects due to extreme heat. Some RCM simulations (particularly REMO) overestimate the WBGT occurrence and show exceedances over a much bigger area than ERA5, including a large part of central Europe.
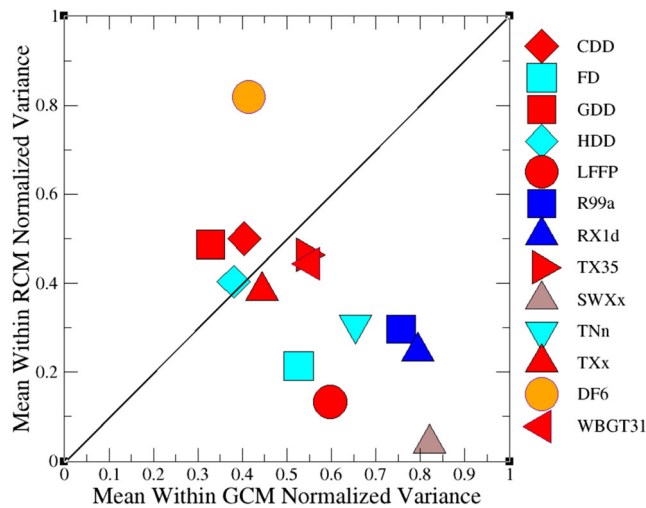
Figure 11. Same as Figure 3 (variance analysis) but for the extreme and impact-oriented indices.

For GDD and CDD, which are driven by warm summer temperatures, similar bias patterns as for TXx and mean summer temperatures are found with a general negative bias across Europe consistent with the summer temperature bias pattern. Strong positive biases are found over eastern Europe (in particular for CCLM and REGCM), and strong negative biases for RACMO, HIRHAM, ALADIN53, RCA, WRF361H, and WRF381P in the Mediterranean regions (IP and MD). ECEARTH-driven simulations clearly induce a cold bias while HADGEM-driven simulations induce a warm bias. For LFFP, which is mostly driven by winter temperatures, median biases generally do not exceed 20 days, and there is a general positive bias across central Europe, together with negative in Western Mediterranean areas. REMO has a systematic warm bias of more than 50 days over the British Isles, while other models have substantial negative biases (such as RACMO and ALADIN53) in AL, IP, and MD.

For TNn, median biases are moderate and do not follow the mean winter temperature bias patterns, while extreme biases reach more than 5° in absolute values. Only the REMO, RACMO, and RCA simulations produce cold biases with amplitudes larger than 5°, and CCLM, COSMOcrCLIM, and REMO produce warm biases of the same amplitude or more over AL, MD, and EA, while large warm biases are found for several CCLM and COSMOcrCLIM simulations. FD and HDD indices are largely overestimated in northern and eastern Europe as a consequence of the general cold bias (see Figure 1). For Frost days, large positive biases (>50 days) are found over the BI with ALADIN53, RACMO, and IPSLr1-WRF381P.

For the two heavy precipitation indices we find a general wet bias following the mean precipitation. This bias is for some regions quasi-systematic, especially in Mediterranean areas where even the low tail of the distribution is positive. When the gridded high-resolution data sets are used instead of the E-OBS data, the effect on the indices in all PRUDENCE regions is positive, as we saw for seasonal mean precipitation. Relative increases are in the same order for SC, IP, and EA, but considerably larger for AL and BI. For RX1d the absolute/relative increases in mm/day/% are as follows: 13/24 (AL), 4.8/16 (SC), 6.5/20 (BI), 9.1/27 (IP), and 2.0/6.8 (EA). Compared to RX1d, the increases in R99a are slightly smaller: 5.3/16 (AL), 3.0/17 (SC), 3.3/16 (BI), 4.6/23 (IP), and 1.0/6.1 (EA). These numbers indicate that the substantial model overestimation of heavy precipitation indices when assessed with E-OBS data can be reduced by resorting to high-resolution precipitation data sets, which have a denser station network, especially in topographically complex regions. The issue of undercatch gauge correction in observations also needs to be considered when interpreting the biases in precipitation-based indices.

Drought frequency, as measured by a dry spell frequency exceeding 6 months (DF6), exhibits a median dry bias in northern Europe (too many spells) and a wet bias (too few spells) in southern Europe. Large dry biases of more than 1 spell/decade are found over Scandinavia for about 10 simulations, especially from the driving GCMs HADGEMr1, ECEARTHr12, CNRMr1, and NORESEMr1, and wet bias of less than −1 spells/decade are limited to IPSLr1-WRF381P.

Regarding storm conditions as measured by the maximum daily maximum wind speed (SWXx), we find, as for the mean surface wind, general positive biases, especially over mountainous and Mediterranean areas. The WRF381P simulations have particularly high biases exceeding 10 m/s over the Alps and the Mediterranean, reaching a factor of about 2 as compared to ERA5 climatological values. Other simulations (from RCA, REGCM, and HIRHAM) also have large biases. The least biased simulations are those using ALADIN63, CCLM, COSMOcrCLIM, and REMO.

In order to assess the contribution of the driving GCM versus RCM in the indices biases, we performed the same analysis as for the previous variables and present the results in Figure 11 for the overall average and supporting information Figure S3 for average over each PRUDENCE region. For GDD, HDD, CDD, and DF6, which are indices depending on seasonal or long-duration, large-scale phenomena, the contribution of the GCMs to the biases is dominant, although this occurs systematically over all individual regions only for DF6. For indices depending on cold conditions (LFFP, FD, and TNn), and for extreme precipitation,

the contribution of RCMs to the biases is dominant over all regions (except for TNn in ME), probably due to the strong dependence on physical parameterizations. The effect of RCM physics schemes is also large for the storm index, which depends on surface and boundary layer parameterizations.

6. Synthesis and Discussion

We have analyzed the performance of the large EURO-CORDEX ensemble that was recently developed in order to provide a large set of GCM-RCM simulations for use in climate change research and climate services. This ensemble consists of 55 simulations combining 8 GCMs and 11 RCMs. We focused on biases for the most important climate variables (temperature, precipitation, wind, radiation, and sea level pressure) and a variety of extreme and impact-oriented indices.

6.1. Synthesis of Results

Results show, above all, that for many climatological aspects, simulations reproduce fairly well the recent past climate despite some biases. Simulations can have both small biases for a set of variables and large biases for others. The same diversity of skill holds across different regions. In particular, it seems pointless to try to determine the overall best or worst GCM-RCM pair as the skills strongly depend on the region and on the variable. Based on the above simulation evaluations, we judge that none of the driving GCMs, Euro-CORDEX RCMs, or GCM-RCM pairs can be flagged as being implausible. However, there are general systematic biases in the ensemble, and simulations have collective deviations from observations. The median of the simulations is generally colder and wetter than observed across Europe, except over southeastern Europe, but deviations from observations are mostly within observational uncertainties, although some models present significant biases. In general, RCMs do not improve significantly over GCMs in terms of mean biases, but the added value of RCM downscaling was not the focus of this study and hence it was not examined in detail.

Simulations also generally have a systematic negative sea level pressure bias over western Europe and are systematically too windy close to the surface. The latter could be a consequence of the former, but near-surface wind speeds depend strongly on the surface parameterization and resolution in both models and reanalyses used for reference, so that robust conclusions are difficult to draw. For radiation, simulations can have substantial biases reaching 30–50 W/m^2 in absolute values, but the median biases in most regions (except mountains) are within 10 W/m^2, as for the GCMs.

Each GCM-RCM combination has its own bias, and the contribution of GCMs and RCMs to the biases is examined through an analysis of variance of the bias spread for each variable and index. The GCM and RCM contributions depend on the variable and season, and in particular for some variables, such as radiation and surface wind, the RCM internal physics represents the main source of bias. For other variables, for example, long-term drought frequency biases are essentially driven by the large-scale boundary conditions of the model.

6.2. Questions Raised by This Analysis and Pending Issues

A few questions are raised by this study. First of all, a few key variables still have unexplained systematic biases. For example, sea level pressure shows up to 5-hPa underestimation in the central regions of the domain (see Figure 7), which induces slightly too cyclonic weather over western Europe, and could also partly explain the excessively cold, wet and windy weather. There are contributions from both RCMs and GCMs in the variability of sea level pressure biases (see Figure 3). Systematic positive precipitation and wind speed biases also remain largely unexplained, with potentially large contributions from observation or reanalysis errors.

In many cases, a major difficulty arises from uncertainties in the observations or reanalyses used for model assessment. For example, gridded observations suffer from large uncertainties due to the gridding scheme, station density, scale of variability, and measurement errors such as the gauge undercatch of precipitation. Similarly, reanalyses suffer from model biases or low spatial resolution. These shortcomings are not new but are of particular importance when evaluating high-resolution ensemble climatologies.

Our study itself remains relatively limited in scope, mostly focusing on mean biases. Further studies could investigate other simulation aspects, such as historical trends, higher-order statistics or specific processes,

where in fact the added value of downscaling can be better explored. Our study also does not assess the coverage by the EURO-CORDEX ensemble of the uncertainty related to the full GCM-RCM matrix. The present GCM-RCM matrix remains largely incomplete, and some RCMs and GCMs are overrepresented compared to others. The limited number of GCMs and difference in representation may bias the potential response to anthropogenic forcing. Despite this, we note the potential of the ensemble to further elucidate aspects of GCM-RCM matrix design as the current matrix does contain large completely filled sub-matrices that can be used for testing sampling strategies (Christensen & Kjellström, 2020).

6.3. Consequences for Impact Studies and Climate Services

This article was mostly designed to assess the main characteristics of the EURO-CORDEX ensemble for the recent climate period, and the emphasis was to investigate a range of variables and impact-relevant indices to provide a relatively exhaustive view of the model performance in relation to potential applications. However, our results should be complemented by further targeted studies in order to derive recommendations for climate services. Yet, our results can inform applications in several ways.

First, generic sectoral indicators are analyzed here, which give a broad idea of the range of uncertainty to expect for specific applications before fine-tuning the analysis with more specific indices. In some cases, users can be reassured at a glance that models do not have systematic biases, such as for cold extremes (e.g., TNn, FD, and HDD in central Europe), which are of relevance to the electricity production and transmission. In others, users can be alerted by the presence of significant differences with observations, such as for wind speed, so that appropriate processing of the data is carried out (e.g., bias correction). Also, the range of biases and uncertainties can be used as a reference for comparison in more specific applications.

The ensemble also offers a large number of simulations, providing more opportunities to identify simulations with realistic behavior for the region or application considered or at least eliminate unrealistic ones. However, too strict criteria or a too narrow ensemble of indices could lead to either undersampling potential responses to climate change or to select simulations that would overfit these criteria and remain unrealistic in other aspects. Hence, it is important that the selection of models covers a range of aspects (for instance, both means and extremes, several variables). In practice, climate projections sometimes need to be bias adjusted. However, bias adjustment can in principle only be applied when biases are not too large. The evaluation provided here and a quasi-objective sub-ensemble selection can be a first step to bias adjustment.

As an example of how simulations can be ranked for specific applications, we use all the 13 indices considered here for all variables, which yields 24 metrics to evaluate model biases and rank the models. Based on these 24 metrics, rank models based on two criteria: (i) The counts of indices x regions for which each simulation is in the "best half" of the simulations ensemble in terms of absolute bias and (ii) the count of indices x regions for which each model is in the "worst 5%" of the ensemble. The first criterion describes the ability of the model to broadly simulate with success most indices and regions while the second describes the number of low-tail rankings of the simulation. The results are shown in Figure 12.

For the "best half" criterion, we do not see any specific deficiency of individual RCMs or GCMs. The HADGEM GCM is often found in the upper part of the distribution and ECEARTH in the lower part, but RCMs have a well-distributed ranking, although RCA, HIRHAM, and WRF361H are often found in the lower part of the ranking, and CCLM and COSMOcrCLIM in the upper part. This is due to their biases on temperature and related indices which have a large weight in the selection of indices. For the "worst 5%" criterion, a fairly different ranking is obtained (ranking has to be taken in reverse order), with a group of simulations having relatively frequently their biases among the highest two values (5%) of the ensemble. Such is the case of some simulations using WRF381P (systematically larger bias on wind speed and storm index as well as a larger MSLP and radiation bias). The ranking uses a somewhat arbitrary but large, spectrum of indices, selection of seasons, and regions. The result should therefore be taken as illustrative and not universal. The lack of universality of such ranking can be clearly seen if a specific region is selected.

6.4. Concluding Remarks

Our analysis provides a robust assessment of the present EURO-CORDEX regional climate simulations ensemble. Given the large size of the ensemble, our conclusions are unlikely to change until a new
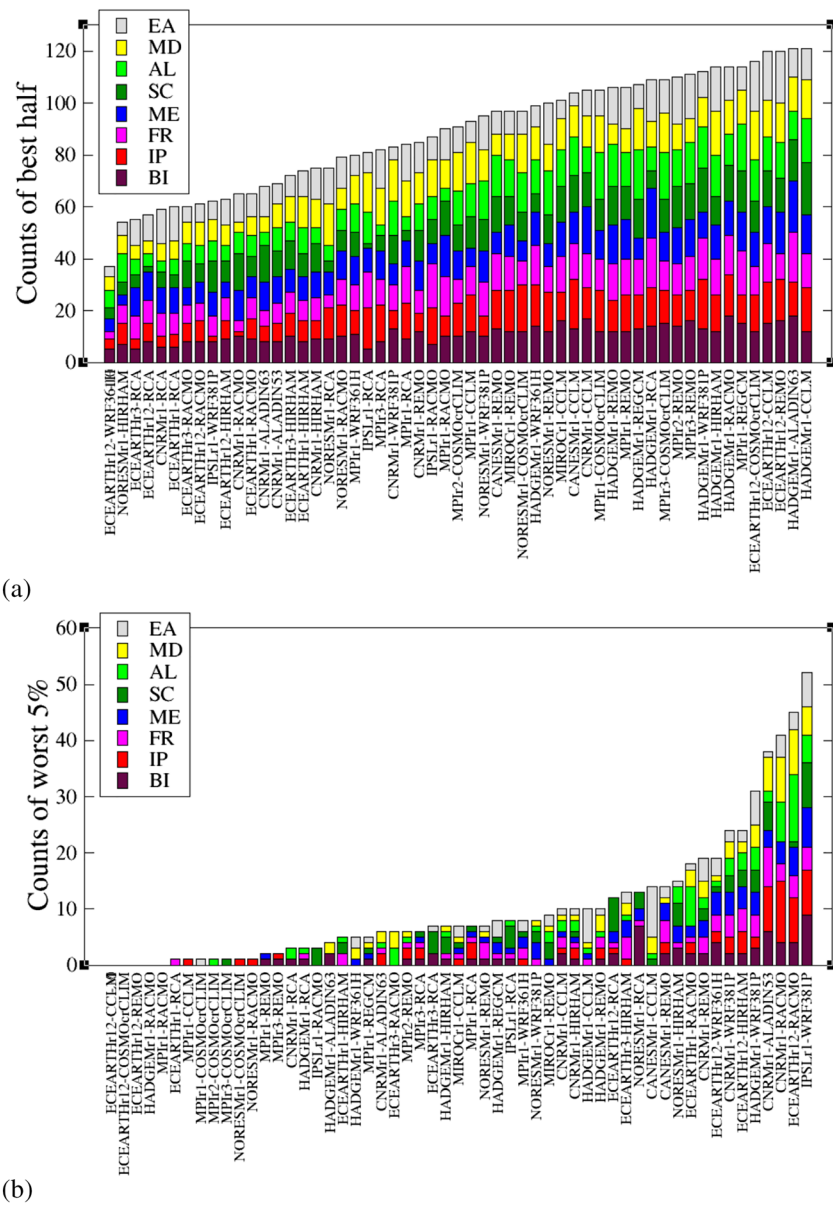
Figure 12. (a) (upper panel) Counts of the number of times among 8 regions and 24 indices for which each model is ranked in the "best half." (b) (lower panel) Same as (a) but for the number of times where each model is in the worst 5%.

generation of models are in place, such as convection-permitting models (Coppola et al., 2018) or fully-coupled Regional Climate System Models (Ruti et al., 2016; Somot et al., 2018). The EURO-CORDEX ensemble will continue to increase in size, including more GCMs or RCMs, but our results will most probably persist. Although this would deserve a dedicated investigation, it is unlikely that CMIP6 boundary conditions will significantly alter our results concerning recent past climate. Therefore, our work can be considered as an assessment of state-of-the-art RCMs for the European region and a reference for users of this ensemble.

A companion paper (Coppola et al., 2021) is providing an analysis of future projections, including comparison between EURO-CORDEX, CMIP5, and CMIP6 ensembles. This ensemble will generally provide one of the key resources for the European Vulnerability, Impacts and Adaptation (VIA) research community and for climate service activities. It is therefore important that users understand not only the value of this resource but also the limitations inherent to this ensemble, as identified in our bias analysis. Further work should investigate a wider range of variables and statistics for targeted applications.

Data Availability Statement

Climate simulations were obtained from the Earth System Grid Federation portals (e.g., https://esgf-node.ipsl.upmc.fr/projects/esgf-ipsl/) and from the DMI servers (http://prudence.dmi.dk/data/temp/PRINCIPLES/). E-OBS data sets were obtained from the KNMI server (https://eca.knmi.nl/). ERA5 re-analyses were obtained from the ECMWF server (http://www.ecmwf.int). High precipitation data sets were obtained from several public servers and are described in Fantini et al. (2018) (https://www.meteos-wiss.admin.ch/home/search.subpage.html/en/data/products/2015/alpine-precipitation.html, ftp://ftp.met.no/projects/klimagrid, https://github.com/ukcp-data/UKCP18_CVs, http://www.aemet.es/en/datos_abier-tos/AEMET_OpenData, and http://www.carpatclim-eu.org/pages/download/).

References

Aalbers, E. E., Lenderink, G., van Meijgaard, E., & van den Hurk, B. J. J. M. (2018). Local-scale changes in mean and heavy precipitation in Western Europe, climate change or internal variability? *Climate Dynamics*, *50*(11-12), 4745–4766. https://doi.org/10.1007/s00382-017-3901-9

Adam, J. C., & Lettenmaier, D. P. (2003). Adjustment of global gridded precipitation for systematic bias. *Journal of Geophysical Research*, *108*(D9), 4257. https://doi.org/10.1029/2002JD002499

Bartók, B., Wild, M., Folini, D., Lüthi, D., Kotlarski, S., Schär, C., et al. (2017). Projected changes in surface solar radiation in CMIP5 global climate models and in EURO-CORDEX regional climate models for Europe. *Climate Dynamics*, *49*(7–8), 2665–2683. https://doi.org/10.1007/s00382-016-3471-2

Betts, A. K., Chan, D. Z., & Desjardins, R. L. (2019). Near-surface biases in ERA5 over the Canadian prairies. *Frontiers in Environmental Science*, *7*, 129. https://www.frontiersin.org/article/10.3389/fenvs.2019.00129

Boé, J., Somot, S., Corre, L., & Nabat, P. (2020). Large differences in summer climate change over Europe as projected by global and regional climate models: Causes and consequences. *Climate Dynamics*, *54*(5-6), 2981–3002. https://doi.org/10.1007/s00382-020-05153-1

C3S: Copernicus Climate Change Service (C3S) (2017). ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS), Nov. 2019. https://cds.climate.copernicus.eu/cdsapp#!/home

Christensen, O. B., & Kjellström, E. (2020). Partitioning uncertainty components of mean climate and climate change in a large ensemble of European regional climate model projections. *Climate Dynamics*, 1–16.

Colin, J., Déqué, M., Radu, R., & Somot, S. (2010). Sensitivity study of heavy precipitations in limited area model climate simulation: Influence of the size of the domain and the use of the spectral nudging technique. *Tellus A*, *62*(5), 591–604. https://doi.org/10.1111/j.1600-0870.2010.00467.x

Coppola, E., Nogherotto, R., Ciarlò, J. M., Giorgi, F., van Meijgaard, E., Iles, C., et al. (2021). Assessment of the European climate projections as simulated by the large EURO-CORDEX regional climate model ensemble. *Journal of Geophysical Research: Atmospheres*, *125*, e2019JD032356. https://doi.org/10.1029/2019JD032356

Coppola, E., Sobolowski, S., Pichelli, E., Raffaele, F., Ahrens, B., Anders, I., et al. (2018). A first-of-its-kind multi-model convection permitting ensemble for investigating convective phenomena over Europe and the Mediterranean. *Climate Dynamics*, *55*(1–2), 3–34. https://doi.org/10.1007/s00382-018-4521-8

Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M., & Jones, P. D. (2018). An ensemble version of the E-OBS temperature and precipitation data sets. *Journal of Geophysical Research: Atmospheres*, *123*, 9391–9409. https://doi.org/10.1029/2017JD028200

Daniel, M., Lemonsu, A., Déqué, M., Somot, S., Alias, A., & Masson, V. (2019). Benefits of explicit urban parameterization in regional climate modeling to study climate and city interactions. *Climate Dynamics*, *52*(5–6), 2745–2764. https://doi.org/10.1007/s00382-018-4289-x

Déqué, M. (2010). Weighted frequency distributions express modelling uncertainties in the ENSEMBLES regional climate experiments. *Climate Research*, *44*(2–3), 195–209. https://doi.org/10.3354/cr00866

Déqué, M., Somot, S., Sanchez-Gomez, E., Goodess, C. M., Jacob, D., Lenderink, G., & Christensen, O. B. (2012). The spread amongst ENSEMBLES regional scenarios: Regional climate models, driving general circulation models and interannual variability. *Climate Dynamics*, *38*(5–6), 951–964. https://doi.org/10.1007/s00382-011-1053-x

Deryng, D., Conway, D., Ramankutty, N., Price, J., & Warren, R. (2014). Global crop yield response to extreme heat stress under multiple climate change futures. *Environmental Research Letters*, *9*(3), 034011. https://doi.org/10.1088/1748-9326/9/3/034011

Evin, G., Hingray, B., Blanchet, J., Eckert, N., Morin, S., & Verfaillie, D. (2019). Partitioning uncertainty components of an incomplete ensemble of climate projections using data augmentation. *Journal of Climate*, *32*(8), 2423–2440. https://doi.org/10.1175/JCLI-D-18-0606.1

Fantini, A., Raffaele, F., Torma, C., Bacer, S., Coppola, E., Giorgi, F., et al. (2018). Assessment of multiple daily precipitation statistics in ERA-Interim driven Med-CORDEX and EURO-CORDEX experiments against high resolution observations. *Climate Dynamics*, *51*(3), 877–900. https://doi.org/10.1007/s00382-016-3453-4

Fernández, J., Frías, M. D., Cabos, W. D., Cofiño, A. S., Domínguez, M., Fita, L., et al. (2019). Consistency of climate change projections from multiple global and regional model intercomparison projects. *Climate Dynamics*, *52*(1–2), 1139–1156. https://doi.org/10.1007/s00382-018-4181-8

Fischer, E. M., & Knutti, R. (2013). Robust projections of combined humidity and temperature extremes. *Nature Climate Change*, *3*(2), 126–130. https://doi.org/10.1038/nclimate1682

García-Díez, M., Fernández, J., & Vautard, R. (2015). *Climate Dynamics*, *45*(11–12), 3141–3156. https://doi.org/10.1007/s00382-015-2529-x

Giorgi, F. (2019). Thirty years of regional climate modeling: Where are we and where are we going next? *Journal of Geophysical Research: Atmospheres*, *124*, 5696–5723. https://doi.org/10.1029/2018JD030094

Gutiérrez, C., Somot, S., Nabat, P., Mallet, M., Corre, L., van Meijgaard, E., et al. (2020). Future evolution of surface solar radiation and photovoltaic potential in Europe: Investigating the role of aerosols. *Environmental Research Letters*, *15*(3), 034035. https://doi.org/10.1088/1748-9326/ab6666

Haylock, M. R., Hofstra, N., Tank, A. M. G., Klok, E. J., Jones, P. D., & New, M. (2008). A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *Journal of Geophysical Research*, *113*, D20119. https://doi.org/10.1029/2008JD010201

Herrera, S., Cardoso, R. M., Soares, P. M., Espírito-Santo, F., Viterbo, P., & Gutiérrez, J. M. (2019). Iberia01: A new gridded dataset of daily precipitation and temperatures over Iberia. *Earth System Science Data*, *11*, 1947–1956. https://doi.org/10.5194/essd-11-1947-2019

Herrera, S., Fernández, J., & Gutiérrez, J. M. (2016). Update of the Spain02 gridded observational dataset forEURO-CORDEX evaluation: Assessing the effect of the interpolation methodology. *International Journal of Climatology*, *36*(2), 900–908. https://doi.org/10.1002/joc4391

IPCC (2013). In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, et al. (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* (p. 1535). Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Isotta, F. A., Frei, C., Weilguni, V., Perčec Tadić, M., Lassègues, P., Rudolf, B., et al. (2014). The climate of daily precipitation in the Alps: Development and analysis of a high-resolution grid dataset from pan-Alpine raingauge data. *International Journal of Climatology*, *34*(5), 1657–1675. https://doi.org/10.1002/joc.3794

Jacob, D., Bärring, L., Christensen, O. B., Christensen, J. H., de Castro, M., Déqué, M., et al. (2007). An inter-comparison of regional climate models for Europe: Design of the experiments and model performance. *Climatic Change*, *81*(Suppl. 1), 31–52. https://doi.org/10.1007/s10584-006-9213-4

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O. B., Bouwer, L. M., et al. (2014). EURO-CORDEX: New high-resolution climate change projections for European impact research. *Regional Environmental Change*, *14*(2), 563–578. https://doi.org/10.1007/s10113-013-0499-2

Jacob, D., Teichmann, C., Sobolowski, S., Katragkou, E., Anders, I., Belda, M., et al. (2020). Regional climate downscaling over Europe: Perspectives from the EURO-CORDEX community. *Regional Environmental Change*, *20*(2), 51. https://doi.org/10.1007/s10113-020-01606-9

Jones, C., Giorgi, F., & Asrar, G. (2011). The Coordinated Regional Downscaling Experiment: CORDEX–an international downscaling link to CMIP5. *CLIVAR exchanges*, *16*(2), 34–40.

Kjellström, E., Boberg, F., Castro, M., Christensen, J. H., Nikulin, G., & Sanchez, E. (2010). On the use of daily and monthly temperature and precipitation statistics as a performance indicator for regional climate models. *Climate Research*, *44*(2–3), 135–150. https://doi.org/10.3354/cr00932

Kjellström, E., Thejll, P., Rummukainen, M., Christensen, J. H., Boberg, F., Christensen, O. B., & Fox Maule, C. (2013). Emerging regional climate change signals for Europe under varying large-scale circulation conditions. *Climate Research*, *56*(2), 103–119. https://doi.org/10.3354/cr01146

Kjellstrom, T., Holmer, I., & Lemke, B. (2009). Workplace heat stress, health and productivity – An increasing challenge for low and middle-income countries during climate change. *Global Health Action*, *2*(1), 2047. https://doi.org/10.3402/gha.v2i0.2047

Kotlarski, S., Keuler, K., Christensen, O. B., Colette, A., Déqué, M., Gobiet, A., et al. (2014). Regional climate modeling on European scales: A joint standard evaluation of the EURO-CORDEX RCM ensemble. *Geoscientific Model Development*, *7*(2), 1297–1333. https://doi.org/10.5194/gmd-7-1297-2014

Laprise, R., de Elía, R., Caya, D., Biner, S., Lucas-Picher, P., Diaconescu, E., et al. (2008). Challenging some tenets of Regional Climate Modelling. *Meteorology and Atmospheric Physics*, *100*(1–4), 3–22. https://doi.org/10.1007/s00703-008-0292-9

Lenderink, G. (2010). Exploring metrics of extreme daily precipitation in a large ensemble of regional climate model simulations. *Climate Research*, *44*(2–3), 151–166. https://doi.org/10.3354/cr00946

Leutwyler, D., Fuhrer, O., Lapillonne, X., Lüthi, D., & Schär, C. (2016). Towards European-scale convection-resolving climate simulations with GPUs: A study with COSMO 4.19. *Geoscience Model Development*, *9*(9), 3393–3412. https://doi.org/10.5194/gmd-9-3393-2016

Lorenz, P., & Jacob, D. (2010). Validation of temperature trends in the ENSEMBLES regional climate model runs driven by ERA40. *Climate Research*, *44*(2–3), 167–177. https://doi.org/10.3354/cr00973

Maraun, D. (2016). Bias correcting climate change simulations—A critical review. *Current Climate Change Reports*, *2*(4), 211–220. https://doi.org/10.1007/s40641-016-0050-x

Mohr, M. (2009). Comparison of versions 1.1 and 1.0 of gridded temperature and precipitation data for Norway. Technical report, Norwegian Meteorological Institute, Report No. 19

Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., van Vuuren, D. P., et al. (2010). The next generation of scenarios for climate change research and assessment. *Nature*, *463*(7282), 747–756. https://doi.org/10.1038/nature08823

Müller, R., Pfeifroth, U., Traeger-Chatterjee, C., Trentmann, J., & Cremer, R. (2015). Digging the METEOSAT Treasure-3 decades of solar surface radiation. *Remote Sensing*, *7*(6), 8067–8101. https://doi.org/10.3390/rs70608067

Müller, R., Pfeifroth, U., Träger-Chatterjee, C., Cremer, R., Trentmann, J., & Hollmann, R. (2015). Surface solar radiation data set—Heliosat (SARAH)—Edition 1. EUMETSAT satellite application facility on climate monitoring (CM SAF). https://doi.org/10.5676/EUM_SAF_CM/SARAH/V001

Nabat, P., Solmon, F., Mallet, M., Kok, J. F., & Somot, S. (2012). Dust emission size distribution impact on aerosol budget and radiative forcing over the Mediterranean region: A regional climate model approach. *Atmospheric Chemistry and Physics*, *12*, 10,545–10,567. https://doi.org/10.5194/acp-12-10545-2012

Nabat, P., Somot, S., Mallet, M., Sanchez-Lorenzo, A., & Wild, M. (2014). Contribution of anthropogenic sulfate aerosols to the changing Euro-Mediterranean climate since 1980. *Geophysical Research Letters*, *41*, 5605–5611. https://doi.org/10.1002/2014GL060798

Nabat, P., Somot, S., Mallet, M., Sevault, F., Chiacchio, M., & Wild, M. (2015). Direct and semi-direct aerosol radiative effect on the Mediterranean climate variability using a coupled Regional Climate System Model. *Climate Dynamics*, *44*(3-4), 1127–1155. http://rd.springer.com/article/10.1007%2Fs00382-014-2205-6

Perry, M., & Hollis, D. (2005). The generation of monthly gridded datasets for a range of climatic variables over the UK. *International Journal of Climatology*, *25*(8), 1041–1054. https://doi.org/10.1002/joc.1161

Pielke, R. A., & Cram, J. M. (1987). An alternate procedure for analyzing surface geostraphic winds and pressure over elevated terrain. *Weather and Forecasting*, *2*(3), 229–236. https://doi.org/10.1175/1520-0434(1987)002<0229:AAPFAS>2.0.CO;2

Prein, A., Gobiet, A., Truehetz, H., Keuler, K., Goergen, K., Teichmann, C., et al. (2016). Precipitation in the EURO-CORDEX 0.11° and 0.44° simulations: High resolution, high benefits? *Climate Dynamics*, *46*(1–2), 383–412. https://doi.org/10.1007/s00382-015-2589-y

Prein, A. F., & Gobiet, A. (2017). Impacts of uncertainties in European gridded precipitation observations on regional climate analysis. *International Journal of Climatology*, *37*(1), 305–327. https://doi.org/10.1002/joc.4706

Rockel, B., Will, A., & Hense, A. (2008). The regional climate model COSMO-CLM (CCLM). *Meteorologische Zeitschrift*, *17*(4), 347–348. https://doi.org/10.1127/0941-2948/2008/0309

Rockel, B., & Woth, K. (2007). Extremes of near-surface wind speed over Europe and their future changes as estimated from an ensemble of RCM simulations. *Climatic Change*, *81*(S1), 267–280. https://doi.org/10.1007/s10584-006-9227-y

Rummukainen, M. (2016). Added value in regional climate modeling. *Wiley Interdisciplinary Reviews: Climate Change*, *7*, 145–159.

Ruti, P. M., Somot, S., Giorgi, F., Dubois, C., Flaounas, E., Obermann, A., et al. (2016). MED-CORDEX initiative for Mediterranean climate studies. *Bulletin of American Meteorological Society*, *97*(7), 1187–1208. https://doi.org/10.1175/BAMS-D-14-00176.1

Samuelsson, P., Gollvik, S., Jansson, C., Kupiainen, M., Kourzeneva, E., & van de Berg, W. J. (2015). The surface processes of the Rossby Centre regional atmospheric climate model (RCA4). *Meteorologie*, 157.

Sanchez-Gomez, E., Somot, S., & Déqué, M. (2009). Ability of an ensemble of regional climate models to reproduce the weather regimes during the period 1961–2000. *Climate Dynamics*, *33*(5), 723–736. https://doi.org/10.1007/s00382-008-0502-7

Sanchez-Gomez, E., Somot, S., Josey, S. A., Dubois, C., Elguindi, N., & Déqué, M. (2011). Evaluation of Mediterranean Sea water and heat budgets simulated by an ensemble of high resolution regional climate models. *Climate Dynamics*, *37*(9–10), 2067–2086. https://doi.org/10.1007/s00382-011-1012-6

Somot, S., Ruti, P., Ahrens, B., Coppola, E., Jordà, G., Sannino, G., & Solmon, F. (2018). Editorial for the Med-CORDEX special issue.

Spinoni, J., Naumann, G., Carrao, H., Barbosa, P., & Vogt, J. (2014). World drought frequency, duration, and severity for 1951–2010. *International Journal of Climatology*, *34*(8), 2792–2804. https://doi.org/10.1002/joc.3875

Spinoni, J., Vogt, J., & Barbosa, P. (2015). European degree-day climatologies and trends for the period 1951–2011. *International Journal of Climatology*, *35*(1), 25–36. https://doi.org/10.1002/joc.3959

Szalai, S., Auer, I., Hiebl, J., Milkovich, J., Radim, T. Stepanek, P., et al. (2013). Climate of the Greater Carpathian Region. Final technical report. http://www.carpatclim-eu.org

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

Torma, C., Giorgi, F., & Coppola, E. (2015). Added value of regional climate modeling over areas characterized by complex terrain-precipitation over the Alps. *Journal of Geophysical Research: Atmospheres*, *120*, 3957–3972. https://doi.org/10.1002/2014JD022781

van Meijgaard, E., Van Ulft, L. H., Lenderink, G., de Roode, S. R., Wipfler, L., Boers, R., & Timmermans, R. M. A. (2012). Refinement and application of a regional atmospheric model for climate scenario calculations of western Europe, climate changes spatial planning publication: KvR 054/12, the Programme office climate changes spatial planning, Nieuwegein, the Netherlands, ISBN/EAN 978–90–8815-046-3, 44 pp. http://climexp.knmi.nl/publications/FinalReport_KvR-CS06.pdf

Vautard, R., Gobiet, A., Jacob, D., Belda, M., Colette, A., Déqué, M., et al. (2013). The simulation of European heat waves from an ensemble of regional climate models within the EURO-CORDEX project. *Climate Dynamics*, *41*, 2555–2575.

Zhang, X., Alexander, L., Hegerl, G. C., Jones, P., Tank, A. K., Peterson, T. C., et al. (2011). Indices for monitoring changes in extremes based on daily temperature and precipitation data. *WIREs Climate Change*, *2*(6), 851–870. https://doi.org/10.1002/wcc.147

References From the Supporting Information

Christensen, O. B., Christensen, J. H., Machenhauer, B., & Botzet, M. (1998). Very high-resolution regional climate simulations over Scandinavia—Present climate. *Journal of Climate*, *11*(12), 3204–3229. https://doi.org/10.1175/1520-0442(1998)011<3204:VHRRCS>2.0.CO;2

Chylek, P., Li, J., Dubey, M. K., Wang, M., & Lesins, G. (2011). Observed and model simulated 20th century Arctic temperature variability: Canadian Earth System Model CanESM2. *Atmospheric Chemistry and Physics Discussions*, *11*(8), 22,893–22,907. https://doi.org/10.5194/acpd-11-22893-2011

Collins, W. J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., et al. (2011). Development and evaluation of an Earth-System model-HadGEM2. *Geoscientific Model Development*, *4*(4), 1051–1075. https://doi.org/10.5194/gmd-4-1051-2011

Dufresne, J.-L., Foujols, M.-A., Denvil, S., Caubel, A., Marti, O., Aumont, O., et al. (2013). Climate change projections using the IPSL-CM5 Earth system model: From CMIP3 to CMIP5. *Climate Dynamics*, *40*(9–10), 2123–2165. https://doi.org/10.1007/s00382-012-1636-1

Giorgi, F., Coppola, E., Solmon, F., Mariotti, L., Sylla, M. B., Bi, X., et al. (2012). RegCM4: Model description and preliminary tests over multiple CORDEX domains, CR 52:7–29 (2012) - https://doi.org/10.3354/cr01018

Hazeleger, W., Severijns, C., Semmler, T., S tefanescu, S., Yang, S., Wang, X., et al. (2010). EC-Earth: A seamless Earth-system prediction approach in action. *Bulletin of the American Meteorological Society*, *91*(10), 1357–1364. https://doi.org/10.1175/2010BAMS2877.1

Jacob, D., Elizalde, A., Haensler, A., Hagemann, S., Kumar, P., Podzun, R., et al. (2012). Assessing the transferability of the regional climate model REMO to different coordinated regional climate downscaling experiment (CORDEX) regions. *Atmosphere*, *3*, 181–199. https://doi.org/10.3390/atmos3010181

Keuler, K., Radtke, K., Kotlarski, S., & Lüthi, D. (2016). Regional climate change over Europe in COSMO-CLM: Influence of emission scenario and driving global model. *Meteorologische Zeitschrift*, *25*, 121–136. https://doi.org/10.1127/metz/2016/0662

Kjellström, E., Bärring, L., Nikulin, G., Nilsson, C., Persson, G., & Strandberg, G. (2016). Production and use of regional climate model projections—A Swedish perspective on building climate services. *Climate Service*, *2–3*, 15–29. https://doi.org/10.1016/j.cliser.2016.06.004

Voldoire, A., Sánchez-Gómez, E., Salas y Mélia, D., Decharme, B., Cassou, C., Sénési, S., et al. (2012). The CNRM-CM5.1 global climate model: Description and basic evaluation. *Climate Dynamics*, *40*(9–10), 2091–2121. https://doi.org/10.1007/s00382-011-1259-y

Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., Okajima, H., et al. (2011). MIROC-ESM 2010: Model description and basic results of CMIP5-20c3m experiments. *Geoscientific Model Development*, *4*(4), 845–872. https://doi.org/10.5194/gmd-4-845-2011