Regional climate modeling on European scales: A joint standard evaluation of the EURO-CORDEX RCM ensemble

S. Kotlarski[1], K. Keuler[2], O. B. Christensen[3], A. Colette[4], M. Déqué[5], A. Gobiet[6], K. Goergen[7,8], D. Jacob[9,10], D. Lüthi[1], E. van Meijgaard[11], G. Nikulin[12], C. Schär[1], C. Teichmann[9,10], R. Vautard[13], K. Warrach-Sagi[14], V. Wulfmeyer[14]

[1] {Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland}

[2] {Chair of Environmental Meteorology, Brandenburg University of Technology (BTU), Cottbus-Senftenberg, Germany}

[3] {Danish Meteorological Institute, Copenhagen, Denmark}

[4] {Institut National de l'Environnement Industriel et des Risques (INERIS), Verneuil-en-Halatte, France}

[5] {Météo-France/CNRM, CNRS/GAME, Toulouse, France}

[6] {Wegener Center for Climate and Global Change, University of Graz, Graz, Austria}

[7] {Centre de Recherche Public – Gabriel Lippmann, Belvaux, Luxembourg}

[8] {Meteorological Institute, University of Bonn, Germany}

[9] {CSC Climate Service Center, Hamburg, Germany}

[10] {Max Planck Institute for Meteorology, Hamburg, Germany}

[11] {KNMI Royal Netherlands Meteorological Institute, De Bilt, The Netherlands}

[12] {Swedish Meteorological and Hydrological Institute, Norrköping, Sweden}

[13] {Laboratoire des Sciences du Climat et de l'Environnement, IPSL, CEA/CNRS/UVSQ, Gif-sur-Yvette, France}

[14] {Institute of Physics and Meteorology, University of Hohenheim, Stuttgart, Germany}

Correspondence to: S. Kotlarski (sven.kotlarski@env.ethz.ch)

Abstract

EURO-CORDEX is an international climate downscaling initiative that aims to provide high-resolution climate scenarios for Europe. Here an evaluation of the ERA-Interim-driven EURO-CORDEX regional climate model (RCM) ensemble is presented. The study documents the performance of the individual models in representing the basic spatio-temporal patterns of the European climate for the period 1989-2008. Model evaluation focuses on near-surface air temperature and precipitation, and uses the E-OBS dataset as observational reference. The ensemble consists of 17 simulations carried out by seven different models at grid resolutions of 12 km (nine experiments) and 50 km (eight experiments). Several performance metrics computed from monthly and seasonal mean values are used to assess model performance over eight sub-domains of the European continent. Results are compared to those for the ERA40-driven ENSEMBLES simulations.

The analysis confirms the ability of RCMs to capture the basic features of the European climate, including its variability in space and time. But it also identifies non-negligible deficiencies of the simulations for selected metrics, regions and seasons. Seasonally and regionally averaged temperature biases are mostly smaller than 1.5 °C, while precipitation biases are typically located in the +/- 40% range. Some bias characteristics, such as a predominant cold and wet bias in most seasons and over most parts of Europe and a warm and dry summer bias over southern and south-eastern Europe reflect common model biases. For seasonal mean quantities averaged over large European sub-domains, no clear benefit of an increased spatial resolution (12 km vs. 50 km) can be identified. The bias ranges of the EURO-CORDEX ensemble mostly correspond to those of the ENSEMBLES simulations, but some improvements in model performance can be identified (e.g., a less pronounced southern European warm summer bias). The temperature bias spread across different configurations of one individual model can be of a similar magnitude as the spread across different models, demonstrating a strong influence of the specific choices in physical parameterizations and experimental setup on model performance. Based on a number of simply reproducible metrics, the present study quantifies the currently achievable accuracy of RCMs used for regional climate simulations over Europe and provides a quality standard for future model developments.

Keywords

Regional climate models, climate change, Europe, ENSEMBLES, EURO-CORDEX, CORDEX, model evaluation

1 Introduction

Assessing the impacts of expected 21st century climate change and developing response strategies requires local to regional scale information on the nature of these changes, including a sound assessment of inherent projection uncertainties. Driven by a suite of IPCC assessment reports and accompanied by increasing public awareness of ongoing climate change, the past decades have seen a rapid development in the corresponding methods for climate scenario generation. Part of this evolution has been the development and the refinement of climate downscaling techniques which aim at translating coarse resolution information as obtained from global climate models (GCMs) into regional and local scale conditions (e.g., Hewitson and Crane, 1996; Wilby and Fowler, 2011). While statistical downscaling methods attempt to bridge the scale gap by applying empirically derived transfer functions between the coarse resolution climate model output and local weather conditions (e.g., Benestad et al., 2008; Fowler et al., 2007; Maraun et al., 2010; Themeßl et al., 2012; Widmann et al., 2003), dynamical downscaling employs high-resolution regional climate models (RCMs) nested into global model output (e.g., Giorgi, 2006; Laprise, 2008; McGregor, 1997; Wang et al., 2004). This technique allows for a considerably higher spatial resolution over the domain of interest and, hence, for a more realistic representation of important surface heterogeneities (such as topography, coast lines, and land surface characteristics) and of meso-scale atmospheric processes. Dynamical downscaling has originally been developed for the purpose of numerical weather prediction and was first applied in a climate context in the late 1980s and early 1990s (Dickinson et al., 1989; Giorgi, 1990). Since then, considerable efforts were put into further methodological and technical developments, and ever increasing computational resources facilitated simulations of multi-decadal length. Large collaborative research projects such as MERCURE (e.g., Hagemann et al., 2004), PRUDENCE (Christensen et al., 2007), NARCCAP (Mearns et al., 2009), and ENSEMBLES (van der Linden and Mitchell, 2009) constituted major milestones in both regional model development and the usage of regional climate scenarios by the climate impact, adaptation and vulnerability community. Dynamical downscaling of GCM output can

today be considered as a well-established standard technique for the generation of regional climate change scenarios. Recent climate scenario products tailored for use in climate impact assessment, such as (1) the CH2011 Swiss climate change scenarios (CH2011, 2011), (2) the German climate impacts and adaptation initiative (Jacob et al., 2008), (3) the German 'Consortium Runs' (Hollweg et al., 2008), (4) the Styrian STMK12 scenarios in the Eastern Alps (Gobiet et al., 2012), (5) the French high-resolution climate scenarios (Lemond et al., 2011; Vautard et al., 2013a), or (6) the climate change scenarios for the Netherlands (van den Hurk, 2007) are to large parts based on the analysis of RCM ensembles. Concerning the interplay between dynamical and statistical downscaling, recent climate impact applications suggest that a combination of the two approaches is optimal (e.g., Bosshard et al., 2013; Paeth, 2011). Apart from their role in climate scenario development, RCMs also became important tools to advance the understanding of regional-scale climate processes and associated feedbacks (e.g., Fischer and Schär, 2009; Hohenegger et al., 2009; Langhans et al., 2013; Seneviratne et al., 2006).

An integral part of regional model development is the evaluation and quantification of model performance by comparison against observation-based reference data. For this purpose, the standard procedure is to carry out evaluation experiments for the recent decades in a *perfect boundary setting*, i.e., applying re-analysis products as lateral boundary forcing for the regional model. Although atmospheric re-analyses, themselves, are based on imperfect models and considerable differences can exist between different re-analysis products with corresponding impacts on downscaling results (Brands et al., 2012) this technique allows to isolate model biases introduced by the nesting procedure and/or the RCM formulation from biases introduced by a potentially erroneous large-scale forcing. Model evaluation in a perfect boundary context is an important component of RCM development. It highlights areas of model deficiencies, without necessarily uncovering the physical reasons for the found biases though. It is furthermore the basis for model calibration efforts (e.g., Bellprat et al., 2012b) and can be used for weighting individual RCMs in multi-model ensembles (Christensen et al., 2010, and further studies in that *Climate Research* special issue) or for excluding models with identifiable severe shortcomings. A proper and physically consistent representation of the present-day climate by RCMs is generally considered as a pre-requisite for their ability to capture the response of regional climates to enhanced greenhouse gas conditions. As such, model evaluation results are an important piece of information provided to end users of regional climate projections.

A large number of previous studies have been concerned with RCM evaluation. Both perfect-boundary settings and GCM-driven setups, in which RCMs potentially inherit biases from the large-scale boundary forcing, were considered. Over Europe, comprehensive evaluations were carried out in the frame of large research projects such as PRUDENCE and ENSEMBLES. Similar but typically less comprehensive evaluation efforts have been conducted outside of Europe (e.g., Evans and McCabe, 2010; Kim et al., 2013; Lucas-Picher et al., 2013; Nikulin et al., 2012; Paeth et al., 2005). Various aspects of model performance were covered, including long-term mean climatological distributions of temperature and precipitation (the two main parameters required by climate impact modelers; e.g., Bergant et al., 2007; Böhm et al., 2008; Holtanova et al., 2012; Jacob et al., 2007; Jacob et al., 2012; Jaeger et al., 2008; Kotlarski et al., 2005), but also explicitly addressing meso-scale structures (Coppola et al., 2010) and frequency distributions of these two parameters (Déqué and Somot, 2010; Kjellström et al., 2010; Warrach-Sagi et al., 2013) as well as temperature trends (Lorenz and Jacob, 2010) and temperature variability (Fischer et al., 2012; Vidale et al., 2007). Elevation dependencies of near-surface air temperature and precipitation were evaluated by Kotlarski et al. (2012). Given the high impact potential, further studies were concerned with the evaluation of extreme precipitation (Frei et al., 2006; Hanel and Buishand, 2012; Herrera et al., 2010; Lenderink, 2010; Maraun et al., 2012; Rajczak et al., 2013; Wehner, 2013) and temperature (Fischer et al., 2007; Vautard et al., 2013b) as well as extreme wind speeds and related loss potentials (Donat et al., 2010; Kunz et al., 2010). Menut et al. (2013) proposed an evaluation of the key climate parameters driving the onset of air pollution episodes. In order to enhance process understanding and to reveal potential reasons for biases in atmospheric quantities, also surface energy fluxes (Hagemann et al., 2004; Lenderink et al., 2007; Markovic et al., 2008) and non-atmospheric state parameters such as terrestrial water storage (Greve et al., 2013; Hirschi et al., 2007) and snow cover (Räisänen and Eklund, 2012; Salzmann and Mearns, 2012; Steger et al., 2013) have been evaluated. In Europe, several studies explicitly focused on RCM evaluation over the Alps, a region subject to a complex topography and a strong spatial variability of near-surface climates (Frei et al., 2003; Haslinger et al., 2013; Kotlarski et al., 2010; Prömmel et al., 2010; Smiatek et al., 2009; Suklitsch et al., 2008; Suklitsch et al., 2011).

In summary, the mentioned studies show that current RCMs are able to reproduce the most important climatic features at regional scales, particularly if driven by perfect-boundary conditions, but that important biases remain. Some of these deficiencies are specific to

individual models. Others seem to be a common and more systematic feature across different RCMs, such as a dry and warm summer bias in south-eastern Europe (Hagemann et al., 2004) and an overestimation of interannual summer temperature variability in central Europe (Fischer et al., 2012; Jacob et al., 2007; Lenderink et al., 2007). Model biases typically depend on the region analyzed (Jacob et al., 2007; Jacob et al., 2012; Rockel and Geyer, 2008), are partly related to parametric uncertainty and choices in model configuration (e.g., Awan et al., 2011; Bellprat et al., 2012a; de Elia et al., 2008; Evans et al., 2012) and can be affected by internal variability (de Elia et al., 2008; Roesch et al., 2008) as well as by uncertainties of the observational reference data themselves (Bellprat et al., 2012a; Kotlarski et al., 2005; Kyselý and Plavcová, 2010). For certain quantities and seasons a higher grid resolution seems to be associated with reduced biases (Déqué and Somot, 2008; Herrmann et al., 2011; Rauscher et al., 2010; Warrach-Sagi et al., 2013). Concerning the use of RCM projections for climate impact assessment, recent studies suggest a non-stationarity of model biases (Bellprat et al., 2013; Boberg and Christensen, 2012; Buser et al., 2009; Christensen et al., 2008; Ehret et al., 2012; Maraun, 2012), questioning the widely used constant-bias assumption when interpreting simulated climate change signals and challenging bias correction techniques.

While RCM projections from projects such as PRUDENCE and ENSEMBLES are widely used by the climate impact community and are considered as state-of-the-art, the next generation of regional climate projections is already under way in the frame of the CORDEX initiative (Giorgi et al., 2009). CORDEX aims to provide an internationally coordinated framework to compare, improve and standardize regional climate downscaling methods, covering both dynamical and empirical-statistical approaches. As part of this effort, model evaluation activities in the individual modeling centers are harmonized and a new generation of regional climate projections for land-regions worldwide based on new CMIP5 GCM projections will be produced. First joint evaluations of CORDEX RCM experiments have recently been published by Nikulin et al. (2012) and Vautard et al. (2013b). EURO-CORDEX, the European branch of CORDEX (Jacob et al., 2013), provides regional climate projections for Europe at grid resolutions of about 12 and 50 km, applying an ensemble of RCMs in their most recent versions, driven by the latest GCM projections, thereby complementing the already available PRUDENCE and ENSEMBLES data with unprecedented high resolution experiments. In its initial phase EURO-CORDEX focuses on model evaluation for present-day climate in a perfect boundary setting. Several aspects of

model performance are analyzed by project partners in a series of ongoing studies. The present work is primarily concerned with evaluating the “standard” variables near-surface air temperature (simply referred to as temperature hereafter) and precipitation on European scales and based on monthly and seasonal mean values. These two quantities are typically evaluated by the individual modeling centers in the course of model development and tuning, and European-scale observational reference data exist. Furthermore, temperature and precipitation change signals are used by many climate impact assessments, and the ability of RCMs to reproduce these quantities is a useful information for a wide range of end users. In order to include dynamical aspects, we additionally evaluate the representation of the large-scale mean sea-level pressure. Although simulations carried out at grid resolutions of both 12 and 50 km are analyzed, we do not specifically aim to investigate the added value of a higher resolution. This would require reliable observation-based datasets at European scale with equivalent resolution, which are not available. Added value assessments are therefore allocated to a suite of accompanying studies evaluating aspects such as extreme precipitation characteristics over sub-domains of the European continent where corresponding reference data exist (see Section 5.1 for further details). The primary aims of the present study are (1) to document the skill of the EURO-CORDEX RCM ensemble in reproducing the present-day European temperature and precipitation climate when driven by realistic boundary conditions, (2) to quantify modeling uncertainties originating from model formulation, (3) to assess a possible progress with respect to the precursor project ENSEMBLES, and (4) to highlight areas of necessary model improvements. For this purpose, we will apply several evaluation metrics covering a range of aspects of model performance. Our study provides a general overview on model performance and is of rather descriptive nature; it does not aim to ultimately explain biases of individual models. We leave these more detailed investigations to a range of follow-up studies which will address specific aspects of model performance.

The study is organized in the following way: After introducing the RCM ensembles and the observational reference data in Chapter 2, Chapter 3 outlines the evaluation methods applied and introduces the individual performance metrics. Chapter 4 then presents the evaluation results for the EURO-CORDEX ensemble and relates them to the previous ENSEMBLES experiments. The results are further discussed in Chapter 5, highlighting the basic model capabilities identified as well as remaining deficiencies in the simulation of the European climate. Chapter 6 finally concludes the study and provides an outlook on future evaluation activities in the EURO-CORDEX framework.

2 Data

2.1 RCM data

We evaluate a set of 17 RCM simulations carried out in the frame of EURO-CORDEX. In total, six different RCMs plus the global ARPEGE model were applied by nine different institutions at grid resolutions of about 12 km (0.11° on a rotated grid) and 50 km (0.44° on a rotated grid). Eight out of the nine 0.11° experiments have a corresponding partner at 0.44° grid spacing, carried out with the identical model version and the identical choice of parameterizations (with the exception of REMO, where rain advection is used for the 0.11° experiments but not for 0.44°). All simulations cover the period 1989-2008 and are driven by the ERA-Interim re-analysis (Dee et al., 2011), providing the required atmospheric lateral boundary conditions and sea surface temperatures and sea ice cover over ocean surfaces. The ERA-Interim boundary conditions can be considered to be of very high quality (Dee et al., 2011), particularly in the Northern Hemispheric extratropics where re-analysis uncertainty is negligible (Brands et al., 2013). The prescribed surface forcing over land (e.g., topography, vegetation characteristics, soil texture) is model-specific and can differ between the experiments. For instance, three out of the nine RCM setups analyzed (CLMCOM, KNMI, SMHI) apply a considerable smoothing to surface orography in order to avoid steep orographic grid-cell-to-grid-cell gradients. The ensemble includes three different configurations of the WRF model that differ mainly in the choice of physical parameterization schemes for radiation transport, microphysics and convection (see Table 1). The individual regional model domains can slightly differ from each other, but all models fully cover the focus domain required for EURO-CORDEX experiments (Fig. 1) and apply an additional lateral sponge zone of individual width for boundary relaxation. A special case is CNRM's ARPEGE model which is a global spectral model with a stretched horizontal grid. ARPEGE was applied here in a special regional setup in which the model runs freely within the common EURO-CORDEX domain (Fig. 1) at resolutions of about 12 km and 50 km, but is relaxed towards ERA-Interim outside of this domain. The EURO-CORDEX ARPEGE experiments can therefore be considered as RCM simulations with a global sponge zone.

An overview on all models and all experiments is provided by Table 1. The set of analyzed experiments corresponds to the currently available ERA-Interim driven EURO-CORDEX ensemble, which might be subject to future extensions. Throughout this paper, the individual simulations will be identified by the acronym of the institution plus the horizontal grid

resolution (*11* for 0.11° and *44* for 0.44°). For instance, the REMO 2009 experiment carried out at 0.11° by CSC will be referred to as *CSC-11*. The entire EURO-CORDEX 0.11° and 0.44° ensembles, consisting of nine and eight different experiments, respectively, will be referred to as *EUR-11* and *EUR-44*. Experiments that were not carried out on the standard 0.11° and 0.44° rotated grids but with comparable grid spacings (e.g., CNRM-11 and CNRM-44) were mapped onto the standard grids applying the nearest-neighbor interpolation method.

For comparing the performance of the EURO-CORDEX ensembles to that of the precursor project ENSEMBLES we additionally consider 16 RCM experiments carried out within the frame of ENSEMBLES with a horizontal grid resolution of about 25 km (0.22° on a rotated grid). These experiments cover a similar domain and were driven by the ERA40 re-analysis (Uppala et al., 2005) for the period 1961-2000. In the present study only the 20-year period 1981-2000 is considered, including the 12 years 1989-2000 that overlap with the EURO-CORDEX ensembles. The application of different large-scale driving fields in ENSEMBLES (ERA40) and EURO-CORDEX (ERA-Interim) can be expected to introduce slight inconsistencies in the intercomparison. The overall effect, however, is presumably small (see Lucas-Picher et al., 2013 for an example over North America). Following the naming convention Institution-Model according to http://ensemblesrt3.dmi.dk/extended_table.html the 16 ENSEMBLES experiments considered are: C4I-RCA3, CHMI-Aladin, CNRM-Aladin, DMI-HIRHAM, EC-GEMLAM, ETHZ-CLM, HC-HadRM3Q0, HC-HadRM3Q3, HC-HadRM3Q16, ICTP-RegCM, KNMI-RACMO, METNO-HIRHAM, MPI-REMO, OURANOS-CRCM, SMHI-RCA and UCLM-PROMES. This ensemble will be referred to as *ENS-22* in the following.

2.2 Observations

As observational reference for evaluating simulated temperature and precipitation we use version 7 of the daily gridded E-OBS dataset (Haylock et al., 2008). E-OBS covers the entire European land surface and is based on the ECA&D (European Climate Assessment and Data) station dataset plus more than 2000 further stations from different archives. It is available at four different resolutions; we here use the rotated 0.22° version which applies the same grid rotation as most of the EURO-CORDEX and ENSEMBLES experiments. The E-OBS 0.22° grid corresponds to a horizontal resolution of about 25 km and exactly matches the grid of the 0.22° ENSEMBLES simulations. Each E-OBS 0.22° grid cell contains four cells of the rotated 0.11° EURO-CORDEX grid, and four E-OBS 0.22° cells exactly match one rotated

0.44° EURO-CORDEX cell. Several previous studies have questioned the quality of E-OBS in regions of sparse station density, and in particular regarding daily extremes (Bellprat et al., 2012a; Herrera et al., 2012; Hofstra et al., 2009; Hofstra et al., 2010; Kyselý and Plavcová, 2010; Maraun et al., 2012; Rajczak et al., 2013) and its effective spatial resolution (e.g., Hanel and Buishand, 2011; Kyselý and Plavcová, 2010). Since the density of the station network is rather low over a considerable part of Europe, the gridding procedure tends to smooth the spatial variability of both temperature and precipitation, and over many regions the effective resolution of E-OBS is presumably lower than the nominal 0.22° grid spacing. For individual sub-regions of the European continent more accurate datasets might exist which are based on a larger number of observation stations. The clear advantage of E-OBS is its spatial (entire European land surface) and temporal (1950-2012) coverage, which makes it ideal for an approximate evaluation of RCM-simulated temperature and precipitation characteristics over Europe. As observational uncertainties are not explicitly considered here, potential inaccuracies of E-OBS should however be kept in mind when interpreting the evaluation results. In addition to the issues mentioned above, this applies also to E-OBS precipitation sums which do not reflect the systematic undercatch of rain gauge measurements (which on average can be of the order of 4% to 50% depending on the season and region; e.g. Frei et al., 2003; Rubel and Hantel, 2001; Sevruk 1986) and very probably underestimate true precipitation. To account for this inaccuracy of the observational reference, we deliberately highlight precipitation biases between 0 and +25% in some of the analyses. Wet biases in this range could be explained by a mean systematic rain gauge undercatch of up to 20% of true precipitation (i.e., neglecting any seasonal and site-specific variation of the measurement error). Furthermore, note that E-OBS is only available at a maximum spatial resolution of 0.22°. The 0.11° EURO-CORDEX experiments can therefore only be evaluated on the coarser E-OBS grid and an in-depth added-value analysis of the 0.11° experiments compared to the 0.44° simulations is not possible within this framework. For the evaluation of the spatial pattern of the simulated mean sea-level pressure, the driving re-analysis ERA-Interim itself is used as reference, i.e., the analysis reveals to what extent the individual RCMs distort the large-scale flow imposed by the boundary conditions.

3 Methods and metrics

3.1 Regional analysis

In order to capture the spatial variability of model performance over Europe, the individual evaluation metrics (see below) were applied to eight different sub-domains of the European continent (Fig. 1): The Alps (AL), the British Isles (BI), Eastern Europe (EA), France (FR), the Iberian Peninsula (IP), the Mediterranean (MD), Mid-Europe (ME), and Scandinavia (SC). These domains have been specified in the frame of the PRUDENCE project (Christensen et al., 2007) and have since then been widely used for RCM evaluation and analysis of climate change signals (e.g., Bellprat et al., 2012b; Christensen et al., 2008; Kotlarski et al., 2012; Lenderink, 2010; Lorenz and Jacob, 2010). They represent comparatively homogeneous climatic conditions, although pronounced climatic gradients can exist within individual sub-domains. The Alpine domain AL, for instance, covers both high-elevation regions along the Alpine ridge and the low-lying Po Valley in northern Italy. Still, the decomposition of the EURO-CORDEX domain into these eight sub-domains allows representing important large-scale climatic gradients (e.g., the transition from maritime climates in the West to continental climates in the East). In the main part of this study the results for only four sub-domains are shown, sampling a wide range of climatic settings (EA, IP, ME, SC). For completeness, figures for the remaining sub-domains (AL, BI, FR, MD) are presented in Appendix B.

3.2 Evaluation metrics

Besides the analysis of seasonal mean biases at grid point scale for the EUR-11 ensemble and the entire EURO-CORDEX domain, we apply several evaluation metrics to monthly, seasonal (winter: DJF, spring: MAM, summer: JJA, autumn: SON) and annual mean values of temperature and precipitation for all experiments of the EUR-11, EUR-44 and ENS-22 ensembles. These metrics are well-established distance measures that assess the quality of (regional) climate simulations by comparison against a gridded observational reference. They represent spatial and temporal bias characteristics and demonstrate the unavoidable spread of model performances in the reproduction of present-day regional climate. As our aim is not to produce an overall skill score that could be used for model weighting but to document different aspects of model performance, the metrics are presented individually and are not combined into some final performance score. The short evaluation period, leading to a sample

size of only 20 seasonal/annual means, also hampers a sound analysis of statistical robustness. We therefore explicitly refrain from assessing the statistical significance of the detected model biases and do also not address any trends of climate parameters. The following metrics are used (exact mathematical formulations are provided in Appendix A; the term "climatological" refers to mean values over the 20-year period 1989-2008):

BIAS: The difference (model - reference) of spatially averaged climatological annual or seasonal mean values for a selected sub-region (relative difference for precipitation).

95%-P: The 95th percentile of all absolute grid cell differences (model - reference) across a selected sub-region based on climatological annual or seasonal mean values (relative difference for precipitation).

PACO: The spatial pattern correlation between climatological annual or seasonal mean values of model and reference data across all grid points of a selected sub-region.

RSV: Ratio (model over reference) of spatial standard deviations across all grid points of a selected sub-region of climatological annual or seasonal mean values.

TCOIAV: Temporal correlation of interannual variability between model and reference time series of spatially averaged annual or seasonal mean values of a selected sub-region.

RIAV: Ratio (model over reference) of temporal standard deviations of interannual time series of spatially averaged annual or seasonal mean values of a selected sub-region.

CRCO: Spearman rank correlation between spatially averaged monthly values of model and reference data of the climatological mean annual cycle of a selected sub-region.

ROYA: Ratio (model over reference) of yearly amplitudes (differences between maximum and minimum) of spatially averaged monthly values of the climatological mean annual cycle of a selected sub-region.

3.3 Regridding

Several evaluation metrics require a grid-cell-by-grid-cell comparison between models and observations. Consequently, a remapping of either the EURO-CORDEX RCM output or of E-OBS to a common reference grid was necessary prior to the analysis. In order to ensure a fair evaluation, our strategy was to always use the coarser grid as reference, except for mean sea-level pressure (see below). This means that (1) the evaluation of the EUR-11 ensemble was

carried out on the coarser 0.22° E-OBS grid, and that (2) the EUR-44 experiments were evaluated on their native 0.44° model grid. In the first case, the model data were conservatively projected onto the 0.22° E-OBS grid. In the second case, E-OBS 0.22° was conservatively projected onto the 0.44° model grid (rather than directly applying the rotated 0.44° version of E-OBS). Conservative projection in this context means that the value of a target grid cell is calculated by an area-weighted average of all overlapping grid cells of the original grid, conserving area mean values. In the special case where four EUR-11 grid cells exactly fit into one E-OBS 0.22° cell, and four E-OBS cells fill one EUR-44 grid cell the projection results in a simple arithmetic four-point-average of the finer grid. Additionally, an elevation correction was carried out for temperature assuming a uniform temperature lapse rate of 0.0064 K m^{-1}, using the E-OBS topography as reference in the first case and the rotated 0.44° topography of the COSMO-CLM RCM as reference in the second case. For most experiments of the ENS-22 ensemble no regridding was necessary since both the RCM output and E-OBS are defined on the same rotated 0.22° grid. The few ENSEMBLES experiments that have not been carried out on the rotated 0.22° standard grid were conservatively remapped onto that grid. An elevation correction for temperature was applied in all cases.

Because mean sea-level pressure has a large-scale structure and no quantitative grid cell metrics were calculated for this variable, the comparison between the EUR-11 simulations and the ERA-Interim reference data has also been carried out on the 0.22° E-OBS grid. For visualizing the spatial pattern of temporal mean biases, the coarser ERA-Interim geographic grid was therefore projected onto the finer (rotated) E-OBS grid.

4 Results

4.1 Spatial bias pattern

Figures 2 to 4 provide an overview on the spatial distribution of the 20-year mean winter (DJF) and summer (JJA) model biases of the EUR-11 ensemble for temperature, precipitation and mean sea-level pressure. For temperature and in agreement with previous studies (see Chapter 1), this evaluation indicates a good reproduction of the spatial temperature variability by the RCMs, including the North-South temperature gradient and elevation effects (Fig. 2). Still, important biases can occur in individual experiments. In wintertime temperatures are

typically underestimated over large parts of the domain. Largest negative biases exceeding -3 °C are found in north-eastern Europe (IPSL-INERIS, CRP-GL, CSC), in Norway (CNRM, KNMI) and along the Alpine ridge (IPSL-INERIS, CRP-GL, CNRM, CSC, SMHI, KNMI). Only two models show a strong warm bias of more than +3 °C over parts of Scandinavia (UHOH) and north-eastern Europe (CNRM). CSC and IPSL-INERIS overestimate winter temperatures in the South-East. For a number of RCMs the cold temperature bias, which is widespread in winter, is also found in summer (SMHI, KNMI, DMI). These cold biases, however, are generally less pronounced than in winter and most models have a tendency to overestimate summer temperature in the South-East. CLMCOM and CSC show a pronounced warm summer bias over most parts of southern Europe. A notable feature of the temperature evaluation is the fact that the bias range spanned by the three WRF experiments alone (IPSL-INERIS, UHOH, CRP-GL) nearly corresponds to the bias range of the entire EUR-11 ensemble. This is especially true in wintertime, but does not apply to the southern European warm summer biases which are largest in CLMCOM, CSC and DMI. A further conspicuous feature of Fig. 2 is the pronounced small-scale spatial variability of temperature biases in CNRM which is apparently related to orographic patterns.

Concerning mean seasonal precipitation, the evaluation indicates a wet wintertime bias of most models over most parts of Europe (Fig. 3). Biases of more than 50% are obtained over the central and eastern regions. In contrast, winter precipitation amounts over parts of southern Europe (Portugal, northern Italy) are underestimated in most cases. CNRM shows a dry wintertime bias over large parts of the study area. In summer, most experiments overestimate precipitation sums in northern and north-eastern Europe, while three models show a pronounced dry bias in the Mediterranean region (CNRM, CLMCOM, DMI). Again, CNRM considerably underestimates precipitation over most of Europe and, as for temperature, the precipitation bias shows a pronounced variability in space. In contrast to temperature, the three WRF experiments mostly agree in their precipitation bias pattern in winter with a widespread overestimation. In summer, UHOH underestimates precipitation over parts of northern Europe and, hence, shows a slightly different behavior than CRP-GL and IPSL-INERIS which overestimate summer precipitation over the whole analysis domain. Southern European summer precipitation is considerably overestimated by all WRF experiments. A possible reason for the different behavior of UHOH compared to CRP-GL and IPSL-INERIS with respect to summer precipitation over parts of northern Europe is the choice of different microphysics schemes (two-moment scheme in UHOH, one-moment

scheme in IPSL-INERIS and CRP-GL). All models, except CNRM, show a pronounced wet bias along the eastern boundary which may indicate problems with the lateral boundary conditions of the limited area models (e.g., inconsistent velocity and humidity gradients between the RCMs' regional solutions and the ERA-Interim boundary forcing in the lateral sponge zone). In contrast, CNRM uses a global grid and - per definition - a very large sponge zone with a comparatively weak relaxation which likely provides a smoother transition of the prescribed outer boundary conditions into the inner model domain and avoids spurious boundary effects. No difference in the spatial variability of precipitation biases between RCMs that apply a strong smoothing of surface orography (CLMCOM, KNMI, SMHI) and those applying a non-filtered orography (all others) can be identified. This might partly be related to the averaging of simulated precipitation at 0.11° to the 0.22° E-OBS grid prior to the analysis.

To complete the overview on the spatial pattern of model biases and to provide a better handle on dynamical aspects of bias characteristics, Figure 4 presents an evaluation of mean winter and summer mean sea-level pressure. In both seasons the RCMs reproduce the large-scale pattern of mean sea-level pressure fairly well and biases typically do not exceed 3 hPa. The bias pattern is generally smooth and has a large-scale structure in most cases. Exceptions are (a) the SMHI model, which shows a small-scale but strong overestimation in the north-western corner of the analysis domain and an underestimation over continental Europe in winter, leading to a reduced meridional pressure gradient, and (b) the WRF experiments (IPSL-INERIS, UHOH and CRP-GL) which underestimate mean sea-level pressure over continental Europe in both seasons and, in the case of UHOH, also in the north-western corner in summer. A particular feature of the WRF experiments is their agreement on a pronounced negative bias over mountainous terrain in winter (Scandinavian Alps, European Alps, Carpathians, Balkan Mountains) and the small-scale structure of the bias pattern, which is not found in the other models (except for positive summer biases over mountainous regions in CNRM and KNMI). This indicates a contribution of the model-specific method to reduce simulated surface pressure to mean sea level, and the pronounced biases in the mentioned regions should not be over-interpreted. Still, the underestimation of mean sea-level pressure by several hPa over large parts of continental Europe particularly in wintertime seems to be a robust feature of the WRF experiments and is also described by Mooney et al. (2013) in a sensitivity study of WRF in Europe.

4.2 Temporal and spatial means

The regionally averaged biases in mean seasonal and annual temperature and precipitation of both the EUR-11 and the EUR-44 ensemble are summarized in Figures 5 and 6 (and Figures B1 and B2). For temperature the analysis reveals a cold bias of up to -2 °C for most models, most seasons and most sub-domains. Exceptions are the CSC simulations that mostly show a slight warm bias as well as the tendency of both ensembles to overestimate summer temperatures over southern and south-eastern Europe (sub-domains EA, IP and MD). While CNRM, KNMI and SMHI are mostly located at the cold end of the model range, temperatures in CLMCOM and CSC are in many cases higher than in the rest of the ensemble. No obvious benefit of the higher resolution (EUR-11 vs. EUR-44) is apparent. The 0.11° experiment of a given model performs worse or better than the corresponding 0.44° experiment depending on season and sub-domain. A systematic difference between both resolutions can be detected only for SMHI and KNMI where the higher resolution tends to produce lower temperatures in all seasons and regions compared to the coarse-resolution setup.

A slightly different result is obtained for regionally averaged precipitation biases which are positive in most cases and, for many models, tend to be larger in the 0.11° experiments due to higher precipitation sums compared to the 0.44° versions. This is especially true for the SMHI model which shows a much stronger overestimation of precipitation at 0.11° grid resolution compared to 0.44° across all seasons and sub-domains. Special cases are the British Isles (BI) with a dry bias in many experiments in winter, summer and autumn (especially of the 0.44° versions) as well as sub-domains AL, EA, FR and IP with a dry summer bias in many experiments. The precipitation biases of the three WRF experiments (CRP-GL, IPSL-INERIS, UHOH) are in many cases close to each other and do not sample the full range of model uncertainty. In general, the precipitation bias reaches from -40% to +80%. Only the UHOH model shows exceptionally high deviations larger than +140% in summer for regions IP and MD. Again, CNRM shows a special behavior and is often found at the dry end of the model range. For individual seasons and sub-domains, wet model biases are mostly smaller than 25% and could, in principle, be explained by an observational undercatch of up to 20% of true precipitation.

As the BIAS metric represents model biases averaged over a given sub-region compensating effects might arise, i.e., a small BIAS value might be the result of large negative and large positive biases over different parts of a given sub-domain compensating each other. To

identify such effects, the 95%-P metric explores the 95th percentile of absolute biases at grid point scale within each sub-domain. For temperature (Figs. 7 and B3) this metric mostly lies within the 1 °C to 3 °C range. Larger values are obtained for the topographically more structured sub-domains SC and AL, which might partly be a result of the simplifying assumption of a temporally and spatially constant lapse rate used for elevation correction (see Section 3.3). The 95%-P metric does not strongly modify the ranking of the models/experiments, i.e., models/experiments that show a small (large) BIAS typically also show a small (large) 95%-P. Hence, the spatially averaged BIAS metric already provides a fairly good impression of model performance and is not too much affected by compensating effects. Again, an exception to this is CNRM-11 which typically shows a noticeable behavior with large 95%-P values, while the BIAS metric for this experiment is not as special (though it typically also shows the largest biases). No systematic improvement of the 0.11° experiments with respect to their 0.44° counterparts can be identified for 95%-P. In case of SMHI and CNRM the higher resolution models - representing stronger variations of topography - produce even larger peak deviations in sub-domains SC and AL than their coarser resolved counterparts. For precipitation (Figs. 8 and B4), 95%-P mostly lies in the 50% to 100% range but can be considerably larger (up to 400%) for the southern European sub-domains IP and MD. The latter can be explained by the relative definition of 95%-P and the small precipitation sums in these regions especially during summer (cf. Fig. 3). This can lead to a large relative overestimation of precipitation by a particular model, although the absolute biases are small. Large 95%-P values are also obtained for the European Alps (AL) especially for DMI and SMHI, which is the result of a pronounced overestimation of precipitation along the Alpine ridge in combination with a strong dry bias over the low-lying Po valley south of the Alps (cf. Fig. 3). Especially for DMI these compensating effects of diverging precipitation biases within sub-domain AL are not apparent from the BIAS metric (Fig. B2) but only from 95%-P (Fig. B4). For all sub-domains, 95%-P values are typically larger than 25% and, in case these values correspond to wet model biases, cannot be explained by an observational undercatch of up to 20% of true precipitation.

4.3 Spatial variability

The performance of the EUR-11 and EUR-44 ensembles with respect to the spatial variability of mean winter and mean summer temperature and precipitation within individual sub-domains (i.e., at grid box scale) is explored by the Taylor diagrams of Figures 9 and 10 (and

Figures B5 and B6 for the remaining sub-domains). The analysis for temperature (Figs. 9 and B5) indicates a high pattern correlation (PACO) for all experiments and most sub-domains, with values typically larger than 0.9. Smaller correlations down to 0.8 are obtained for sub-domain ME in summertime mainly by simulations of the EUR-44 ensemble. Concerning the spatial standard deviation both ensembles have a clear tendency to an overestimation, particularly in summertime and by up to 50%. RSVs larger than 1.5 are obtained for CNRM and SMHI in a few cases. Wintertime RSVs are typically smaller and the spatial variability is often underestimated (RSV < 1). The systematic difference between summer and winter RSVs over many sub-domains leads to a clustering of the respective markers for summer (triangles) and winter (circles) in these regions (EA, IP, SC, FR). The pronounced overestimation of spatial temperature variability by CNRM-11 over most parts of Europe is very likely related to the large spatial variability of the mean seasonal model bias (cf. Section 4.1). For most experiments and most sub-domains the centered root-mean-square difference between simulation and observational reference amounts to less than 50% of the observed spatial standard deviation. Overall systematic differences in model skill between the 0.11° and the 0.44° versions (filled markers compared to non-filled markers) are not found.

Similarly to temperature, the spatial variability of mean winter and mean summer precipitation is typically overestimated by the experiments (Figs. 10 and B6), RSVs are mostly located between 1 and 2. A stronger overestimation is found for the Mediterranean sub-domain (MD) and in particular for the DMI model with RSVs of up to 4. Compared to temperature, the spatial pattern correlation of mean seasonal precipitation is much lower and PACO typically amounts to between 0.4 and 0.9 only. Whether a better performance is obtained for winter or summer (circles compared to triangles) considerably depends on the sub-domain. There is no apparent systematic difference in model skill between the high and the low resolution versions (filled compared to non-filled markers). The centered root-mean-square difference between models and observations, expressed in units of the observed standard deviation, is typically found in the range between 50% and 200% (RSVs between 0.5 and 2).

4.4 Inter-annual variability

The Taylor diagrams of Figures 11 and 12 (and Figures B7 and B8 for further sub-domains) combine the parameters TCOIAV and RIAV which assess the model performance with respect to the temporal (inter-annual) variability of mean winter and mean summer

temperature and precipitation, based on regional averages over each sub-domain. For winter temperature, temporal correlations are mostly larger than 0.9 while the results are worse for the summer season (Figs. 11 and B7). Summer TCOIAVs are typically larger than 0.6, but values down to 0.3 are obtained for the 0.11° WRF experiments (CRP-GL-11, IPSL-INERIS-11, UHOH-11) in several sub-domains. Although we do not have a definite explanation, this could be linked to the high sensitivity of simulated summer temperatures to the selection of the convection scheme (Vautard et al., 2013b). CLMCOM and CNRM, on the other hand, show a very good performance in all seasons and all sub-domains (TCOIAVs mostly larger than 0.9). For CNRM, this particularity could again be related to the special setup of this global model. The large relaxation zone and the continuous nudging of the model's solution towards ERA-Interim could help to maintain a correct chronology of synoptic events (i.e., of events that might partly be lost by limited area models due to their confined relaxation zone and an update of the boundary forcing at typically 6-hourly intervals only). Regarding the RIAV metric, both ensembles tend to overestimate the magnitude of inter-annual temperature variability, in particular during summertime. Except for Scandinavia (SC), where summer RIAVs are mostly smaller than 1, all sub-domains are affected and summer temperature variability is in some cases overestimated by more than 50% (RIAV larger than 1.5). For most cases, the centered root-mean-square difference between simulated and observed mean seasonal temperatures is smaller than the observed temporal standard deviation (normalized RMS distance smaller than 1). No systematic improvement of an increased resolution (EUR-11 versus EUR-44 ensemble) is apparent; in some cases the switch from 0.44° to 0.11° can even deteriorate the model performance (compare non-filled and filled symbols of the same color and the same marker type).

Similar to mean seasonal temperature, temporal correlations for precipitation are large in wintertime (mostly above 0.8) but systematically smaller in summer (Figs. 12 and B8). Again, a number of 0.11° WRF experiments show very low correlations in summertime. TCOIAVs are partly smaller than 0.3, suggesting inaccuracies in the representation of convective processes and their triggering mechanisms in this model. Concerning the inter-annual variability of precipitation, model performance shows a large spread. RIAV values are centered around 1 for sub-domains IP, SC and FR, but both ensembles typically overestimate the inter-annual precipitation variability in both seasons (AL, EA, MD, SC) or in summer only (ME). Only sub-domain BI shows a general underestimation of inter-annual precipitation variability (by up to 50%). As for temperature, the centered root-mean-square difference for

mean seasonal precipitation does typically not exceed the standard deviation of the observations (except AL) and, again, no obvious benefit of an increased grid resolution can be identified.

4.5 Mean annual cycle

The parameters CRCO and ROYA assess the model performance with respect to the mean annual cycle at monthly resolution, averaged over each sub-domain. Not astonishingly, the rank correlation for temperature (Fig. 13, left panel) is high in all experiments (CRCOs larger than 0.95) reflecting a proper representation of the temperature variation throughout the year by the RCMs, mainly driven by the annual cycle of air temperature and SST in the imposed large-scale forcing and of top-of-the-atmosphere incoming solar radiation. Concerning the ratio of amplitudes (Fig. 13, right panel) most experiments systematically overestimate the intensity of the mean annual cycle of temperature (ROYAs larger than 1). Exceptions are the British Isles where a majority of experiments underestimates the mean annual amplitude as well as the WRF experiments (CRP-GL, IPS-INERIS, UHOH) which systematically underestimate the annual amplitude over most parts of Europe. These results are closely related to the seasonal variability of the temperature bias in Figures 5 and B1. In most cases temperature biases are positive in summer and negative in winter (or less negative in summer than in winter), leading to an over-pronounced annual cycle. For SC, cold winter and cold summer biases are typically close to each other. This causes a negative shift of the annual cycle with only a minor influence on the annual variation. For sub-domain BI, in contrast, many simulations tend to underestimate summer temperatures more than winter temperatures, resulting in a flattening of the annual cycle. This is also the case for most regions in the WRF simulations, especially for IPSL-INERIS and UHOH. For ROYA, most outliers are members of the EUR-44 ensemble, i.e. an increased model resolution seems to be associated with a slightly better performance. For individual models and sub-domains this might, however, not be true.

Regarding the mean annual cycle of precipitation, the model performance is generally worse than for temperature (Fig. 14). While most experiments show a rank correlation CRCO larger than 0.7 in sub-domains BI, IP, SC and MD, correlations are typically much lower in FR, ME, AL and EA. In ME and EA rank correlations close to zero or even negative are obtained, indicating a deficient representation of the mean annual cycle of precipitation. In these regions, the spread of the individual experiments is, however, very large and most simulations

actually have correlations larger than 0.5. Whether the annual amplitude of area-averaged precipitation is over- or underestimated (ROYA metric) strongly depends on the region and the experiment. While the annual amplitude is generally too small over the British Isles (BI), the majority of models overestimates the annual amplitude over France (FR), the Alps (AL) and the Mediterranean region (MD). No systematic difference in model skill between the EUR-11 and the EUR-44 ensemble can be identified. For SMHI, IPSL-INERIS and CRP-GL the ROYA values of the 0.11° simulations are generally larger than in the 0.44° case, but only better in three out of eight sub-domains. The high resolution experiments of KNMI and DMI show better ROYA values than their low resolution counterparts in seven regions whereas the CSC and CNRM simulations perform better in six regions at 0.44° grid spacing. With respect to CRCO six models perform better with the higher resolution in at least six regions. Again, CSC and CNRM produce a better skill in six regions with the coarser resolution.

4.6 EURO-CORDEX versus ENSEMBLES

The gray bars and markers in Figures 5 to 14 (and Figures B1 to B8) represent the ENS-22 ensemble and allow relating the performance of EUR-11 and EUR-44 to the performance of the previous ENSEMBLES experiments. Note that the latter ensemble was driven by a different re-analysis (ERA40 instead of ERA-Interim), has been evaluated over a different period of time (1981-2000 instead of 1989-2008), has a larger ensemble size (16 instead of 9 and 8 experiments for EUR-11 and EUR-44, respectively) and includes models that are not part of EUR-11 and EUR-44.

For temperature, a comparison of the BIAS ranges (Figs. 5 and B1) indicates an improvement in EUR-11 and EUR-44 concerning the strong overestimation of summer temperatures over the southern and south-eastern parts of Europe (EA, IP, FR, MD), but also over central Europe (ME, AL). Regionally averaged summer temperature biases in EUR-11 and EUR-44 are typically smaller than 1.5 °C compared to strong warm biases of some ENSEMBLES experiments. However, the cold biases of SMHI, KNMI and CNRM do partly exceed those of the ENSEMBLES models by some tenths of a degree (AL, BI, MD). Considering the larger ensemble size of ENS-22, the overall bias range seems to be comparable. As for temperature 95%-P (Figs. 7 and B3), both EUR-11 and EUR-44 mostly improve on ENS-22 except for CNRM and partly SMHI, KNMI and CRP-GL which can be subject to strong biases on the grid cell scale in sub-domains EA, IP, SC, BI, MD and especially in AL.

Due to some wet and dry outliers of the EUR-11 and EUR-44 ensembles in individual sub-domains and seasons, the range and the magnitude of the precipitation BIAS of the EURO-CORDEX simulations are partly larger than in ENS-22. This particularly concerns sub-domains EA, ME, BI and FR. The same is true for precipitation 95%-P (Figs. 8 and B4). Some improvements with predominantly smaller values are apparent for sub-domain IP while, on the other hand, several EUR-11 and EUR-44 simulations show larger biases in sub-domains ME, BI and FR compared to ENSEMBLES.

Regarding the reproduction of the spatial variability of temperature (Figs. 9 and B5) and precipitation (Figs. 10 and B8), EUR-11 and EUR-44 often slightly improve on ENS-22 (markers closer to the 1-1 location). Again, exceptions are CNRM and to some extent also SMHI which partly show a pronounced overestimation of the spatial standard deviation of temperature beyond the ENS-22 range. Some features like the higher spatial correlation of winter precipitation (Fig. 10) and the smaller spatial temperature variability (Fig. 9) in SC are concordantly reproduced by all three ensembles. The temporal variability of temperature (Figs. 11 and B7) is slightly improved with respect to ENSEMBLES in summertime, mainly due to a less pronounced overestimation of interannual variability (RIAVs closer to one in many sub-domains). No clear difference between EUR-11 and EUR-44 on one hand and ENS-22 on the other hand is obvious for metrics describing the interannual variability of precipitation (Figs. 12 and B8). Again, the seasonal separation/clustering - like for temperature in EA and FR and for precipitation in EA, IP, ME, SC and FR - is similar in all ensembles.

The rank correlations of the mean annual cycle of temperature averaged over the individual sub-domains are large in all three ensembles (Fig. 13, left panel). For precipitation (Fig. 14, left panel), the performance of the EUR-11 and EUR-44 ensembles is comparable to ENS-22 except for some poor-performing outliers in sub-domains BI (CNRM), FR (IPSL-INERIS) and ME (CNRM). It is worth mentioning that the regions with the largest range of CRCO in ENS-22 (ME and EA) present also the largest ranges in EUR-11 and EUR-44. The ranges in sub-domains IP, SC and MD are, however, considerably reduced in the EUR-11 and EUR-44 ensembles. Regarding the ROYA metric, i.e. the ratio of amplitudes of the mean annual cycle (Figs. 13 and 14, right panels), EUR-11 and EUR-44 show a similar skill as ENS-22, but with a tendency towards an underestimation of the amplitude of the annual cycle by some

experiments in selected sub-domains (IPSL-INERIS and UHOH for temperature; UHOH, CLMCOM, CSC, SMHI for precipitation).

5 Discussion

5.1 The overall picture

The evaluation of the EURO-CORDEX ensembles largely confirms RCM bias characteristics identified by previous studies based on the ENSEMBLES data. This concerns both the general magnitude as well as the sign of model biases. Improvements with respect to ENSEMBLES are a reduced overestimation of southern and south-eastern European summer temperatures, a less pronounced overestimation of interannual summer temperature variability as well as a slightly better representation of the spatial climatic variability within the sub-domains. In some cases, however, individual EURO-CORDEX experiments are subject to bias magnitudes beyond the range found for ENSEMBLES. This especially concerns the CNRM model which is subject to a strong spatial variability of model biases on the grid cell level and a pronounced cold and dry bias over many parts of Europe. CNRM‘s summer dry bias, however, is not due to shortcomings in the physical parameterizations, but is a consequence of the specific design of the CNRM experiments. Further simulations in which the relaxation outside Europe is weaker (6 hours instead of 10 min e-folding time) do not show it. The reason might be an over-drying of the atmosphere in the relaxation area (the rest of the globe) where a permanent spin-up of temperature and moisture relating to the mismatch between ERA-Interim and ARPEGE physics is imposed on the model. CNRM’s cold bias over high mountains is to some extent related to the model’s snow scheme and a too persistent snow cover (Vautard et al., 2013b).

The availability of different configurations of WRF allows comparison of the bias spread obtained for this particular model to the spread across different models. The fact that the temperature bias range of the three WRF-11 experiments often corresponds to the bias range of the entire ensemble illustrates the uncertainty introduced by the choice of parameterizations and parameter settings (e.g., Bellprat et al., 2012a; Mooney et al., 2013). This, however, is not apparent for precipitation biases where the different WRF setups approximately agree on sign and magnitude of their bias. In wintertime, the wet bias of WRF seems to be closely related to the distinct negative bias of mean sea-level pressure (compare Figures 3 and 4), indicating a

too high intensity of low pressure systems passing the continent. Circulation types and storm tracks, however, have not been analyzed in detail in the present study and possible relations between precipitation and mean sea-level pressure biases remain speculative.

Mostly independent of the season and the sub-domain under consideration, the relative ranking of models with respect to seasonal mean temperature is stable, with CNRM, KNMI and SMHI showing coldest temperatures as opposed to warmer conditions in CLMCOM and CSC. For seasonally and regionally averaged precipitation sums the relative ranking is less fixed, although the high resolution versions of SMHI, CRP-GL and IPSL-INERIS are often found at the wet end while CNRM typically belongs to the driest models.

For sub-domain mean values at seasonal resolution, no apparent benefit of a finer grid resolution is identified. For temperature and depending on sub-domain and season, the 0.11° experiments can be warmer or colder than their 0.44° counterparts and no systematic bias reduction in the high resolution experiments is found. This also holds for the 95th percentile of absolute temperature biases (95%-P). In case of precipitation, seasonal mean biases are typically larger in the EUR-11 ensemble as precipitation sums are generally overestimated by both ensembles and the increase of resolution is mostly associated with a further increase of precipitation. The latter might be related to stronger orographic gradients in the high resolution experiments due to a better resolved topography. Our analysis also highlights the potential of error compensation when restricting the analysis to mean values for relatively large sub-domains. Especially for precipitation a metric such as 95%-P can provide further insight into model biases on grid cell level in addition to the metrics PACO and RSV which measure the accuracy of horizontal distribution and spatial variation over a selected sub-domain.

The absence of obvious benefits of a finer grid resolution in our analysis does not rule out such an added value in general. The 0.22° resolution of the gridded observations, coarser than that of the 0.11° RCM simulations, allows conclusions concerning a lack of large-scale bias improvements by the 0.11° experiments, but hinders identification of benefits at smaller scale. In orographically structured terrain, we expect an added value of an increased spatial resolution for parameters such as meso-scale circulations, the precipitation intensity distribution at daily resolution or snow cover dynamics. These aspects have not been addressed in the current work, partly since this would require observational reference data with a better reliability than E-OBS at high temporal and spatial scales. Such data are

currently not available at a European level but only for smaller sub-regions (mostly individual countries), such as the REGNIE or HYRAS precipitation data for Germany (Rauthe et al., 2013) or the SAFRAN re-analysis over France (Quintana-Segui et al., 2008). A detailed investigation of the added value of high resolution experiments based on such data will be the subject of upcoming studies, possibly applying dedicated added value metrics (e.g., Kanamitsu and DeHaan, 2011). Indeed, recent studies by Bauer et al. (2011), Prein et al. (2013a and 2013b) and Warrach-Sagi et al. (2013) indicate that an increase of RCM resolution (in their case to convection-permitting scales) bears added value, but this added value can cancel out by spatial and temporal averaging.

Further cautionary notes concern the influence of (1) internal model variability, (2) uncertainties in the observational reference data, and (3) deficiencies of the driving re-analysis on the computed skill metrics. (1) can influence the simulated mean climatology even in decadal and multi-decadal RCM experiments that are subject to an identical boundary forcing (e.g., Bellprat et al., 2012a; Lucas-Picher et al., 2008; Roesch et al., 2008) in particular over large model domains as in our simulations. As the EUR-11 and EUR-44 ensembles consist of only one experiment for each setup, a quantification of the effect of internal variability on the model evaluation is not possible. Instead, slight nuances of bias characteristics should not be over-interpreted as they could, to some degree, result from internal random variability. A similar reasoning is true for uncertainties in the E-OBS observational reference. Finally, model evaluation has been carried out in a perfect boundary context and basically assumes a bias-free representation of the lateral atmospheric boundary forcing and of sea surface temperatures by the driving ERA-Interim re-analysis. Although the recent studies by Brands et al. (2012 and 2013) suggest a negligible re-analysis uncertainty for the Northern Hemispheric extratropics, a certain influence of a biased boundary forcing on the evaluation results cannot be ruled out.

5.2 RCM deficiencies and capabilities

One of the most prominent deficiencies across members of both the EUR-11 and EUR-44 ensembles is the predominant cold bias in most seasons and for most sub-domains. The spatially averaged bias often reaches -1 to -2 °C but can be larger in individual cases. For some regions such as Norway and the Alpine ridge, this cold bias might partly be related to the pronounced topography of the respective region, associated with large elevation differences between the individual RCMs and the E-OBS reference at grid point level. This,

in turn, potentially amplifies inaccuracies of the assumption of a spatially and temporally uniform lapse rate for elevation correction (see Section 3.3). Exceptions to the general picture of a predominant cold model bias are the CSC model that mostly shows too high temperatures as well as the summer season in southern and south-eastern Europe where most models have a tendency to overestimate temperatures. This result is consistent with previous findings (e.g., Hagemann et al., 2004 for PRUDENCE, and Christensen et al., 2008 for ENSEMBLES) and is probably related to an underestimation of summertime precipitation (compare Figures 3 and 4) and soil moisture-temperature coupling: In soil moisture-controlled evaporative regimes, low soil moisture contents (e.g., resulting from preceding precipitation deficits) limit the amount of energy used for the latent heat flux and increase the sensible heat flux, ultimately leading to an increase of air temperature (e.g., Seneviratne et al. 2010). This feedback is sensitive to all processes that interfere with the regional balances of water and energy, and this includes land-surface, boundary layer, convective and radiative processes. Related to this is the overestimation of inter-annual temperature variability in the summer season by both ensembles (RIAVs larger than 1). This widespread and systematic model bias has previously also been reported for the PRUDENCE and ENSEMBLES experiments (e.g., Fischer et al., 2012; Lenderink et al., 2007; Vidale et al., 2007). The warm summer biases do not coincide with pronounced positive mean sea-level pressure biases (compare Figures 2 and 4) which indicates the dominant role of regional-scale land surface-atmosphere interactions and only a minor contribution of large-scale circulation biases (e.g., too persistent blocking regimes). The former were also identified as driving factors for the correct representation of summer heat waves in the EURO-CORDEX ensemble (Vautard et al., 2013b).

Regarding regionally averaged precipitation biases, the most striking feature is a pronounced wet bias of both ensembles over most sub-domains and for most seasons (except CNRM and except the dry biases in southern and south-eastern Europe). As a consequence of a general tendency to higher precipitation sums with increased model resolution, this wet bias is typically more pronounced in the 0.11° experiments. Based on the restricted detail of our analysis, a full explanation of this bias is not possible at this point. Note that the E-OBS reference has not been corrected for the systematic undercatch of rain gauges (cf. Section 2.2). If one assumes a mean systematic undercatch of 20% of true precipitation, wet model biases can in some cases be explained by this shortcoming of the observational reference.

Another important deficiency of simulated precipitation are the low rank correlations (CRCO metric) of simulated and observed climatological monthly means in sub-domains FR, ME, AL and EA. Here, CRCOs are typically lower than 0.7 and partly close to zero or even negative, indicating a reversal of the observed annual cycle by the RCMs. When analyzing CRCO it has to be noted, though, that low or negative correlations are more likely in regions with a weak annual cycle of precipitation. A similar reasoning is true for model biases of the mean annual amplitude of precipitation (ROYA metric). As the numbers in the left panel of Figure 14 (CRCO) indicate, the standard deviation of the mean annual cycle is smallest - only 13 to 18% of the annual mean monthly precipitation - in sub-domains FR, ME and AL. The difference between maximal and minimal mean monthly precipitation is also smallest for these three sub-domains (right panel of Fig. 14, ROYA). It amounts to only 42 to 53% of the annual mean monthly precipitation. For sub-domains IP and MD this normalized difference is more than twice as large (131 and 113%, respectively), indicating a pronounced annual variation of precipitation in these regions. This is confirmed by the high values of the normalized standard deviation (44 and 31%; Figure 14 left panel) in these sub-domains. Hence, the bad model performance with respect to CRCO in FR, ME and AL and the considerable overestimation of ROYA in FR does not necessarily indicate a severe model bias but rather shows that the respective model cannot reproduce small monthly deviations from a rather uniform annual distribution of precipitation. This is, however, not the case for the partly weak mean annual correlation in EA. In particular CCLM and CNRM seem to have serious problems to correctly reproduce the annual cycle of precipitation in this eastern part of the model domain.

Concerning the general overestimation of spatial temperature and precipitation variability within sub-domains (RSV metric), this deficiency does very probably not only reveal true model biases but also deficiencies of the E-OBS reference relating to the spatial smoothing and an effective resolution lower than 0.22° and 0.44°, respectively, in regions of a low network density (see Section 2.2). This effect would lead to an apparent overestimation of RSV by the model experiments, although the true spatial variability might actually be well represented. Sub-domains like ME with little orographic variability and, furthermore, a rather dense station network (cf. Haylock et al., 2009) would be less affected by this artifact and, indeed, show a better model performance with respect to RSV (Fig. 9). Unfortunately, not a single data set currently exists that provides homogenized climate data for the entire European continent with an effective spatial resolution equal or higher than the actual resolution of

modern RCMs used for long-term climate simulations. Hence, more detailed investigations of small-scale climatological features can be carried out only for specific sub-regions where appropriate high resolution reference data exist.

When analyzing the temporal correlation between the simulated and observed seasonal mean values over the 20-year long evaluation period (metric TCOIAV), an obvious feature is the much better correlation for winter (mostly larger than 0.9 for both temperature and precipitation) compared to summer (often smaller than 0.6). The better performance for winter reflects the fact that European summer climate is much more controlled by local- to regional-scale processes, giving the RCMs a higher degree of freedom to alter the conditions imposed by the boundary forcing (e.g., Déqué et al., 2005). In contrast, winter climate in mid-latitudes is more affected by the synoptic-scale transport of warm or cold and moist or dry air masses which couples the internal solution of the RCMs closer to the temporal evolution of the lateral boundary values. Comparing TCOIAV for temperature and precipitation, smaller correlations are typically obtained for precipitation, reflecting a weaker control of the large-scale boundary conditions on sub-domain mean precipitation compared to sub-domain mean temperature.

Despite the mentioned shortcomings in the representation of specific climatic features over the European continent, the evaluation indicates a considerable skill of the EUR-11 and EUR-44 ensembles to reproduce larger-scale horizontal variability of climatological seasonal mean values (expressed, for instance, as differences of mean values between the individual sub-domains). In most sub-domains, especially for temperature, also the shape and the amplitude of regionally averaged mean annual cycles are reproduced to a large extent (ROYA and CRCO metrics). The climatological fields of mean sea-level pressure as represented by the driving ERA-Interim re-analysis are mostly captured well and are only slightly distorted in some cases.

For temperature the spatial variability within the individual sub-domains is fairly well captured (PACO mostly > 0.9). This good performance is, however, to some extent a simple result of the systematic elevation dependency of air temperature. As continental-scale gradients and biases thereof are not sampled by the sub-domains, high-elevation regions will typically have lower temperatures than their low-elevation counterparts in a given sub-domain, both in the observations and in the models. As grid-scale topography can be assumed to be realistically represented by the models and as, additionally, an elevation correction is

carried out for temperature this will lead to high values of PACO. This effect will generally be less pronounced in sub-domains without strong orographic gradients (such as ME). In the case of precipitation, the spatial variability within sub-domains is simulated less accurately (PACO typically between 0.4 and 0.9). This partly reflects the fact that seasonal precipitation sums are also affected by topography, but on regional scales far less systematic than temperature. Instead, RCMs can suffer from considerable systematic biases of the spatial precipitation field in orographic terrain such as the windward-lee effect (overestimation of precipitation on the windward side, underestimation on the lee side; e.g., Warrach-Sagi et al., 2013).

6 Conclusions and outlook

The present work evaluates the ERA-Interim driven RCM ensembles of the EURO-CORDEX initiative on a European scale. Our analysis mainly considers the standard parameters 2m temperature and precipitation and is based on monthly and seasonal mean values. Several simple and reproducible metrics covering a range of aspects of model performance are used to compare simulation results to the E-OBS observational reference. This enables a quantitative assessment of the newest generation of RCMs to simulate European climate conditions and a direct comparison with results of the previous ENSEMBLES simulations. The validation exercise serves as a quality standard for further simulations and future model developments. The added value of the high-resolution experiments (EUR-11) compared to their coarser resolution counterparts (EUR-44) is not specifically addressed in this study.

The model evaluation highlights the general ability of today's regional climate models to represent the basic spatio-temporal patterns of the European climate, but also indicates considerable deficiencies for selected metrics, regions and seasons. Some of these deficiencies, such as a predominant cold and wet bias in most seasons and over most of Europe, are found in the majority of experiments and reflect common model biases. Furthermore, many experiments are subject to a warm and dry summer bias over southern and south-eastern Europe. The latter had previously been identified for the ENSEMBLES experiments, but for this specific case the bias appears to be reduced in the EURO-CORDEX ensembles. However, neglecting the influence of slightly incompatible setups (different driving re-analysis, different simulation and, hence, evaluation period), no general improvements of the EURO-CORDEX simulations with respect to ENSEMBLES could be

identified for the temporal and spatial scales considered in the present work. In addition to common model deficiencies found across the range of different RCMs, a number of model-dependent biases could be identified. Except for a few consistent outliers, these biases typically depend on the region and season under consideration.

Identifying possible reasons for both common and model-specific bias characteristics and formulating specific recommendations for model development will require a deeper and dedicated analysis, including additional metrics and variables and explicitly taking into account uncertainties in the observational reference and the effect of RCM-internal climate variability. These aspects will be the subject of upcoming studies within the EURO-CORDEX community. The same is true for studies explicitly addressing the added value of an increased grid resolution. In terms of regionally and seasonally averaged quantities the present work could not identify such an added value. This does, however, not rule out benefits of an increased resolution, and we would expect such benefits for quantities such as daily precipitation intensities, small-scale spatial climate variability in topographically structured terrain or snow cover dynamics. These aspects still need to be investigated in more detail. Further analyses will consider (1) the relation between present-day model biases and simulated climate change signals, (2) the question of whether model biases are temporally stable and bias correction methods are feasible and can be reliably applied, (3) inter-comparisons of the performance of different types of downscaling methodologies, as well as (4) the assessment of trends of simulated climatic parameters within the observed period. For the latter aspect the current 20-year long EURO-CORDEX evaluation experiments are not well suited, but extended simulations covering the full ERA-Interim period (1979 to present) are already under way and will be available for such analyses. Furthermore, applying the same quantitative metrics used in the present study to the EURO-CORDEX GCM-driven experiments would allow separating the contribution of the driving global climate model from the intrinsic RCM contribution to the overall bias structure.

Acknowledgements

Calculations for WRF IPSL-INERIS were made using the TGCC super computers under the GENCI time allocation GEN6877. The contribution from CRP-GL was funded by the Luxembourg National Research Fund (FNR) through Grant FNR C09/SR/16 (CLIMPACT). The KNMI-RACMO2 simulations were supported by the Dutch Ministry of Infrastructure

and the Environment. The CCLM and REMO simulations were supported by the Federal Ministry of Education and Research (BMBF) and performed under the ‘‘Konsortial’’ share at the German Climate Computing Centre (DKRZ). The CCLM simulations were furthermore supported by the Swiss National Supercomputing Centre (CSCS) under project ID s78. The contribution from UHOH was funded by the German Science Foundation (DFG) through project FOR 1695. Part of the SMHI contribution was carried out in the Swedish Mistra-SWECIA programme founded by Mistra (the Foundation for Strategic Environmental Research). We acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (http://ensembles-eu.metoffice.com) and the data providers in the ECA&D project (http://eca.knmi.nl).

References

Awan, N. K., Truhet, H., and Gobiet, A.: Parameterization-Induced Error Characteristics of MM5 and WRF Operated in Climate Mode over the Alpine Region: An Ensemble-Based Analysis, J. Clim., 24, 3107–3123, 2011.

Baldauf, M. and Schulz, J. P.: Prognostic precipitation in the Lokal-Modell (LM) of DWD, COSMO Newslett., 4, 177–180, 2004.

Balsamo, G. P., Viterbo, A., Beljaars, B. J. J. M., van den Hurk, B., Hirschi, M., Betts, A., and Scipal, K.: A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the Integrated Forecast System, J. Hydrometeor., 10, 623-643, 2009.

Bauer, H.-S., Weusthoff, T., Dorninger, M., Wulfmeyer, V., Schwitalla, T., Gorgas, T., Arpagaus, M., Warrach-Sagi, K.: Predictive skill of a subset of the D-PHASE multi-model ensemble in the COPS region, Q. J. R. Meteorol. Soc., 137, 287–305, 2011

Bellprat, O., Kotlarski, S., Lüthi, D., and Schär, C.: Exploring Perturbed Physics Ensembles in a Regional Climate Model, J. Clim., 25, 4582-4599, 2012a.

Bellprat, O., Kotlarski, S., Lüthi, D., and Schär, C.: Objective calibration of regional climate models, J. Geophys, Res., 117, D23115, 2012b.

Bellprat, O., Kotlarski, S., Lüthi, D., and Schär, C.: Physical constraints for temperature biases in climate models, Geophys. Res. Lett., 40, 4042-4047, 2013.

Benestad, R. E., Hanssen-Bauer, I., and Chen, D.: Empirical-statistical Downscaling, World Scientific Publishing, Singapore, 2008.

Bergant, K., Belda, M., and Halenka, T.: Systematic errors in the simulation of European climate (1961–2000) with RegCM3 driven by NCEP/NCAR reanalysis, Int. J. Clim., 27, 455-472, 2007.

Boberg, F. and Christensen, J. H.: Overestimation of Mediterranean summer temperature projections due to model deficiencies, Nature Clim. Change, 2, 433-436, 2012.

Böhm, U., Keuler, K., Österle, H., Kücken, M., and Hauffe, D.: Quality of a climate reconstruction for the CADSES regions, Meteorologische Zeitschrift, 17(4), 477-485, 2008.

Bosshard, T., Carambia, M., Goergen, K., Kotlarski, S., Krahe, P., Zappa, M., and Schär, C.: Quantifying uncertainty sources in an ensemble of hydrological climate-impact projections, Water Resour. Res., 49,1-14, 2013.

Bougeault, P.: A simple parameterization of the large-scale effects of cumulus convection, Mon. Weather Rev., 113, 2108–2121, 1985.

Brands, S., Gutiérrez, J. M., Herrera, S., and Cofiño, A. S.: On the Use of Reanalysis Data for Downscaling, J. Clim., 25, 2517-2526, 2012.

Brands, S., Herrera, S., Fernández, J., and Gutiérrez, J. M.: How well do CMIP5 Earth System Models simulate present climate conditions in Europe and Africa? A performance comparison for the downscaling community, Clim. Dyn., 41, 803-817, 2013.

Buser, C. M., Künsch, H. R., Lüthi, D., Wild, M., and Schär, C.: Bayesian multi-model projection of climate: bias assumptions and interannual variability, Clim. Dyn., 33, 849-868, 2009.

CH2011: Swiss Climate Change Scenarios CH2011, Published by C2SM, MeteoSwiss, ETH, NCCR Climate, and OcCC. Zurich, Switzerland (available on http://www.ch2011.ch), 2011.

Champeaux, J. I., Masson, V., and Chauvin, F.: ECOCLIMAP: a global database of land surface parameters at 1 km resolution, Meteorol. Appl., 12, 29–32, 2003.

Christensen, J. H., Carter, T. R., Rummukainen, M., and Amanatidis, G.: Evaluating the performance and utility of regional climate models: the PRUDENCE project, Clim. Change, 81, 1-6, 2007.

Christensen, J. H., Boberg, F., Christensen, O. B., Lucas-Picher, P.: On the need for bias correction of regional climate change projections of temperature and precipitation, Geophys. Res. Lett., 35, L20709, 2008.

Christensen, J. H., Kjellström, E., Giorgi, F., Lenderink, G., Rummukainen, M.: Weight assignment in regional climate models, Clim. Res., 44, 179-194, 2010.

Claussen, M., Lohmann, U., Roeckner, E., and Schulzweida, U.: A global data set of landsurface parameters. MPI for Meteorology, Hamburg, Report No. 135, Hamburg, 1994.

Collins, W.D., Rasch, P. J., Boville, B. A., Hack, J. J., McCaa, J. R., Williamson, D. L., Kiehl, J. T., Briegleb, B., Bitz, C., Lin, S.-J., Zhang, M., and Dai, Y.: Description of the

NCAR Community Atmosphere Model (CAM 3.0), NCAR technical note, NCAR/TN-464+STR, 2004.

Coppola, E., Giorgi, F., Rauscher, S. A., Piani, C.: Model weighting based on mesoscale structures in precipitation and temperature in an ensemble of regional climate models, Clim. Res., 44, 121-134, 2010.

Cuxart, J., Bougeault, P., and Redelsperger, J.-L.: A turbulence scheme allowing for mesoscale and large-eddy simulations, Q. J. R. Meteorol. Soc., 126, 1–30, 2000.

Dee, D. P., Uppala, S. M., Simmons, A. J. et al.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. R. Meteorol. Soc., 137, 553-597, 2011.

Déqué, M., Jones, R. G., Wild, M., Giorgi, F., Christensen, J. H., Hassell, D. C., Vidale, P. L., Rockel, B., Jacob, D., Kjellström, E., de Castro, M., Kucharski, F., and van den Hurk, B.: Global high resolution versus Limited Area Model climate change projections over Europe: quantifying confidence level from PRUDENCE results, Clim. Dyn., 25, 653-670, 2005.

Déqué, M. and Somot, S.: Analysis of heavy precipitation for France using high resolution ALADIN RCM simulations, Idöjárás – Q. J. Hung. Meteorol. Serv., 112(3-4), 179-190, 2008.

Déqué, M., and Somot, S.: Weighted frequency distributions express modeling uncertainties in the ENSEMBLES regional climate experiments, Clim. Res., 44, 195-209, 2010.

Dickinson, R. E., Errico, R. M., Giorgi, F., and Bates, G. T.: A regional climate model for the Western United States, Clim. Change, 15, 383-422, 1989.

Doms, G., Förstner, J., Heise, E., Herzog, H.-J., Raschendorfer, M., Schrodin, R., Reinhardt, T., and Vogel, G.: A description of the nonhydrostatic regional model LM. Part II: physical parameterization. Available online at http://www.cosmomodel.org/content/model/documentation/core/cosmoPhysParamtr.pdf, 2007.

Donat, M. G., Leckebusch, G. C., Wild, S., and Ulbrich, U.: Benefits and limitations of regional multi-model ensembles for storm loss estimations, Clim. Res., 44, 211-225, 2010.

Douville, H., Planton, S., Royer, J. F., Stephenson, D. B., Tyteca, S., Kergoat, L., Lafont, S., and Betts, R. A.: The importance of vegetation feedbacks in doubled-CO_2 time-slice experiments, J. Geophys. Res., 105,14841–14861, 2000.

ECWMF-IFS: IFS documentation-Cy31r1. PART IV: Physical Processes, available from http://www.ecmwf.int/research/ifsdocs/CY31r1/PHYSICS/IFSPart4.pdf, 2007.

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., and Liebert, J.: HESS Opinions - Should we apply bias correction to global and regional climate model data? Hydrol. Earth Syst. Sci., 16, 3391-3404, 2012.

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., Tarpley, J. D.: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model, J. Geophys. Res., 108(D22), 8851, 2003.

de Elía, R., Caya, D., Côté, H., Frigon, A., Biner, S., Giguère, M, Paquin, D., Harvey, R., and Plummer, D.: Evaluation of uncertainties in the CRCM-simulated North American climate, Clim. Dyn., 30, 113-132, 2008.

Evans, J. P. and McCabe, M. F.: Regional climate simulation over Australia's Murray-Darling basin: A multitemporal assessment, J. Geophys. Res., 115, D14114, 2010.

Evans, J. P., Ekström, M., and Ji, F.: Evaluating the performance of a WRF physics ensemble over South-East Australia, Clim. Dyn., 39, 1241-1258, 2012.

Fischer, E. M., Seneviratne, S. I., Lüthi, D., and Schär, C.: Contribution of land-atmosphere coupling to recent European summer heat waves, Geophys. Res. Lett., 34, L06707, 2007.

Fischer, E.M. and Schär, C.: Future changes in daily summer temperature variability: driving processes and role for temperature extremes, Clim. Dyn., 33, 917-935, 2009.

Fischer, E. M., Rajczak, J., and Schär, C.: Changes in European summer temperature variability revisited, Geophys. Res. Lett., 39, L19702, 2012.

Fouquart, Y. and Bonnel, B.: Computations of solar heating of the earth's atmosphere : A new parameterization, Beitr. Phys. Atmos., 53, 35-62, 1980.

Fowler, H.J., Blekinsop, S., and Tebaldi, C.: Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modeling, Int. J. Clim., 27, 1547-1578, 2007.

Frei, C., Christensen, J. H., Déqué, M., Jacob, D., Jones, R.G., and Vidale, P. L.: Daily precipitation statistics in regional climate models: Evaluation and intercomparison for the European Alps, J. Geophys. Res., 108 (D3), 4124, 2003.

Frei, C., Schöll, R., Fukutome, S., Schmidli, J., and Vidale, P. L.: Future change of precipitation extremes in Europe: Intercomparison of scenarios from regional climate models, J. Geophys. Res., 111, D06105, 2006.

Giorgetta, M. and Wild, M.: The water vapor continuum and its representation in ECHAM4. MPI for Meteorology, Hamburg, Report No. 162, 1995.

Giorgi, F.: Simulation of Regional Climate Using a Limited Area Model Nested in a General Circulation Model, J. Clim., 3, 941-963, 1990.

Giorgi, F.: Regional climate modeling: Status and perspectives, J. Phys. IV France, 139, 101-118, 2006.

Giorgi, F., Jones, C., and Asrar, G.R.: Addressing climate information needs at the regional level: the CORDEX framework, WMO Bulletin, 58(3), 175-183, 2009.

Gobiet, A., Suklitsch, M., Leuprecht, A., Peßenteiner, S., Mendlik, T., and Truhetz, H.: Klimaszenarien für die Steiermark bis 2050 (STMK12),110 pp., available from http://www.technik.steiermark.at/cms/ziel/95576483/DE (in German), 2012.

Grell, G. A. and Devenyi, D.: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques, Geophys. Res. Lett., 29(14), doi:10.1029/2002GL015311, 2002.

Greve, P., Warrach-Sagi, K., and Wulfmeyer, V.: Evaluating Soil Water Content in a WRF-Noah Downscaling Experiment, J. Appl. Meteor. Climatol., 52, 2312–2327, 2013.

Hagemann, S.: An improved land surface parameter dataset for global and regional climate models. MPI for Meteorology, Hamburg, Report No. 336, 2002.

Hagemann, S., Machenhauer, B., Jones, R., Christensen, O. B., Déqué, M., Jacob, D., and Vidale, P. L.: Evaluation of water and energy budgets in regional climate models applied over Europe, Clim. Dyn., 23, 547-567, 2004.

Haslinger, K., Anders, I., Hofstätter, M.: Regional climate modelling over complex terrain: an evaluation study of COSMO-CLM hindcast model runs for the Greater Alpine Region, Clim. Dyn., 40, 511-529, 2013.

Hanel, M. and Buishand, A.: On the value of hourly precipitation extremes in regional climate model simulations, J. Hydrol., 393, 265-273, 2010.

Hanel, M. and Buishand, A.: Analysis of precipitation extremes in an ensemble of transient regional climate model simulations for the Rhine basin, Clim. Dyn., 36, 1135-1153, 2011.

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., and New, M.: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, J. Geophys. Res., 113, D20119, 2008.

Herrera, S., Fita, L., Fernández, J., and Gutiérrez, J. M.: Evaluation of the mean and extreme precipitation regimes from the ENSEMBLES regional climate multimodel simulations over Spain, J. Geophys. Res., 115, D21117, 2010.

Herrera, S., Gutiérrez, J. M., Ancell, R., Pons, M. R., Frías, M. D., and Fernández, J.: Development and analysis of a 50-year high-resolution daily gridded precipitation dataset over Spain (Spain02), Int. J. Clim., 32, 74-85, 2012.

Herrmann, M., Somot, S., Calmanti, S., Dubois, C., and Sevault, F.: Representation of spatial and temporal variability of daily wind speed and of intense wind events over the Mediterranean Sea using dynamical downscaling: impact of the regional climate model configuration, Nat. Hazards Earth Syst. Sci., 11, 1983-2001, 2011.

Hewitson, B. C. and Crane, R. G.: Climate downscaling: techniques and applications, Clim. Res., 7, 85-95, 1996.

Hirschi, M., Seneviratne, S. I., Hagemann, S., and Schär, C.: Analysis of seasonal terrestrial water storage variations in regional climate simulations over Europe, J. Geophys. Res., 112, D22109, 2007.

Hofstra, N., Haylock, M., New, M., and Jones, P. D.: Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature, J. Geophys. Res., 114, D21101, 2009.

Hofstra, N., New, M., and McSweeney, C.: The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data, Clim. Dyn., 35, 841-858, 2010.

Hohenegger, C., Brockhaus, P., Bretherton, C. S., and Schär, C.: The soil moisture-precipitation feedback in simulations with explicit and parameterized convection, J. Clim., 22(19), 5003–5020, 2009.

Hollweg, H.-D., Böhm, U., Fast, I., Hennemuth, B., Keuler, K., Keup-Thiel, E., Lautenschlager, M., Legutke, S., Radtke, K., Rockel, B., Schubert, M., Will, A., Woldt, M., and Wunram, C: Ensemble simulations over Europe with the regional climate model CLM forced with IPCC AR4 global scenarios, M&D Technical Report No 3, Hamburg, Germany, ISSN 1619-2257, 2008.

Holtanova, E., Miksovský, J., Kalvová, J., Pisoft, P., and Motl, M.: Performance of ENSEMBLES regional climate models over Central Europe using various metrics, Theor. Appl. Climatol., 108, 463-470, 2012.

Hong, S.-Y., Dudhia, J., and Chen, S.-H.: A revised approach to microphysical processes for the bulk parameterization of cloud and precipitation, Mon. Weather Rev., 132, 103–120, 2004.

Hong, S.-Y. and Lim, J-O. J.: The WRF single-moment 6-class microphysics scheme (WSM6), J. Korean. Meteorol. Soc., 42, 129–151, 2006.

Hong, S.-Y., Noh, Y., and Dudhia, J.: A new vertical diffusion package with an explicit treatment of entrainment processes, Mon. Weather. Rev., 134, 2318–2341, 2006.

Jacob, D., Bärring, L., Christensen, O. B., Christensen, J. H., de Castro, M., Déqué, M., Giorgi, F., Hagemann, S., Hirschi, M., Jones, R., Kjellström, E., Lenderink, G., Rockel, B., Sánchez, E., Schär, C., Seneviratne, S. I., Somot, S., van Ulden, A., and van den Hurk, B.: An inter-comparison of regional climate models for Europe: model performance in present-day climate, Clim. Change, 81, 31-52, 2007.

Jacob, D., Göttel, H., Kotlarski, S., Lorenz, P., and Sieck, K.: Klimaauswirkungen und Anpassung in Deutschland – Phase 1: Erstellung regionaler Klimaszenarien für Deutschland, Climate Change 11/08, German Federal Environment Agency, 154 pp, available from http://www.umweltbundesamt.de/publikationen/klimaauswirkungen-anpassung-in-deutschland, 2008.

Jacob, D., Elizalde, A., Haensler, A., Hagemann, S., Kumar, P., Podzun, R., Rechid, D., Reca Remedio, A., Saeed, F., Sieck, K., Teichmann, C., and Wilhelm, C.: Assessing the Transferability of the Regional Climate Model REMO to Different COordinated Regional Climate Downscaling EXperiment (CORDEX) Regions, Atmosphere, 3, 181-199, 2012.

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O. B., Bouwer, L. M., Braun, A., Colette, A., Déqué, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Menut, L., Nikulin,

G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Kröner, N., Kotlarski, S., Kriegsmann, A., Martin, E., van Meijgaard, E., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J.-F., Teichmann, C., Valentini, R., Vautard, R., Weber, B., and Yiou, P.: EURO-CORDEX: new high-resolution climate change projections for European impact research, Reg. Environ. Change,1-16, 2013.

Jaeger, E. B., Anders, I., Lüthi, D., Rockel, B., Schär, C., and Seneviratne, S. I.: Analysis of ERA40-driven CLM simulations for Europe, Meteorologische Zeitschrift, 17(4), 1-19, 2008.

Joint Research Centre: Global land cover 2000 database. European Commission, Joint Research Centre. http://bioval.jrc.ec.europa.eu/products/glc2000/glc2000.php, 2003.

Kain, J. S.: The Kain–Fritsch convective parameterization: an update, J. Appl. Meteorol., 43, 170–181, 2004.

Kain, J. S. and Fritsch, J. M.: A one-dimensional entraining/detraining plume model and its application in convective parameterization, J. Atmos. Sci., 47, 2784–2802, 1990.

Kain, J. S. and Fritsch, J. M.: Convective parameterization for mesoscale models: the Kain-Fritsch scheme. The representation of cumulus convection in numerical models, Meteorol. Monogr., 24, 165–170, 1993.

Kanamitsu, M., DeHaan, L.: The Added Value Index: A new metric to quantify the added value of regional models, J. Geophys. Res., 116, D11106, 2011.

Kim, J., Waliser, D. E., Mattmann, C. A., Mearns, L. O., Goodale, C. E., Hart, A. F., Crichton, D. J., McGinnis, S., Lee, H., Loikith, P. C., and Boustani, M.: Evaluation of the Surface Climatology over the Conterminous United States in the North American Regional Climate Change Assessment Program Hindcast Experiment Using a Regional Climate Model Evaluation System, J. Clim., 26, 5698–5715, 2013.

Kjellström, E., Boberg, F., de Castro, M., Christensen, J. H., Nikulin, G, and Sánchez, E.: Daily and monthly temperature and precipitation statistics as performance indicators for regional climate models, Clim. Res., 44, 135-150, 2010.

Kotlarski, S., Block, A., Böhm, U., Jacob, D., Keuler, K., Knoche, R., Rechid, D., and Walter, A.: Regional climate model simulations for hydrological applications: Evaluation of uncertainties, Adv. Geosciences, 5, 119-125, 2005.

Kotlarski, S., Paul, F., and Jacob, D.: Forcing a Distributed Glacier Mass Balance Model with the Regional Climate Model REMO. Part I: Climate Model Evaluation, J. Clim., 23, 1589-1606, 2010.

Kotlarski, S., Bosshard, T., Lüthi, D., Pall, P., and Schär, C.: Elevation gradients of European climate change in the regional climate model COSMO-CLM, Clim. Change, 112, 189-215, 2012.

Kunz, M., Mohr, S., Rauthe, M., Lux, R., and Kottmeier, C.: Assessment of extreme wind speeds from Regional Climate Models - Part 1: Estimation of return values and their evaluation, Nat. Hazards Earth Syst. Sci., 10, 907-922, 2010.

Kyselý, J. and Plavcová, E.: A critical remark on the applicability of E-OBS European gridded temperature data set for validating control climate simulations, J. Geophys. Res., 111, D23118, 2010.

Lacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., and Collins, W. D.: Radiative forcing by long-lived greenhouse gases: calculations with the AER radiative transfer models, J. Geophys. Res., 113, D13103, 2008.

Langhans, W., Schmidli, J., Fuhrer, O., Bieri, S., and Schär, C.: Long-term simulations of thermally-driven flows and orographic convection at convection-parameterizing and cloud-resolving resolutions, J. Appl. Met. Clim., in press, 2013.

Laprise, R.: Regional climate modeling, J. Comput. Phys., 227, 3641-3666, 2008.

Lemond, J., Dandin, P., Planton, S., Vautard, R., Pagé, C., Déqué, M., Franchistéguy, L., Geindre, S., Kerdoncuff, M., Li, L., Moisselin, J. M., Noël, T., and Y. M. Tourre: DRIAS – A step toward french climate services, Adv. Sci. Res., 6, 179-186, 2011.

Lenderink, G.: Exploring metrics of extreme daily precipitation in a large ensemble of regional climate model simulations, Clim. Res., 44, 151-166, 2010.

Lenderink, G., and Holtslag, A. A. M.: An updated length-scale formulation for turbulent mixing in clear and cloudy boundary layers, Q. J. R. Meteorol. Soc., 130, 3405–3427, 2004.

Lenderink, G., van Ulden, A., van den Hurk, B., and van Meijgaard, E.: Summertime inter-annual temperature variability in an ensemble of regional model simulations: analysis of the surface energy budget, Clim. Change., 81, 233-247, 2007.

Lohmann, U. and Roeckner, E.: Design and performance of a new cloud microphysics scheme developed for the ECHAM general circulation model, Clim. Dyn., 12, 557–572, 1996.

Lorenz, P. and Jacob, D.: Validation of temperature trends in the ENSEMBLES regional climate model runs driven by ERA40, Clim. Res., 44, 167-177, 2010.

Louis, J.-F.: A parametric model of vertical eddy fluxes in the atmosphere, Bound. Layer Meteorol., 17, 187–202, 1979.

Lucas-Picher, P., Caya, D., de Elía, R., and Laprise, R.: Investigation of regional climate models' internal variability with a ten-member ensemble of 10-year simulations over a large domain, Clim. Dyn., 31, 927-940, 2008.

Lucas-Picher, P., Somot, S., Déqué, M., Decharme, B., Alias, A.: Evaluation of the regional climate model ALADIN to simulate the climate over North America in the CORDEX framework, Clim. Dyn., 41, 1117-1137, 2013.

Lucas-Picher, P., Somot, S., Déqué, M., Decharme, B., Alias, A.: Evaluation of the regional climate model ALADIN to simulate the climate over North America in the CORDEX framework, Clim. Dyn., 41, 1117-1137, 2013.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themessl, M., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C., Vrac, M., and Thiele-Eich, I.: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, Rev. Geophys., 48, RG3003, 2010.

Maraun, D.: Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums, Geophys. Res. Lett., 39, L06706, 2012.

Maraun, D., Osborn, T. J., and Rust, H. W.: The influence of synoptic airflow on UK daily precipitation extremes. Part II: regional climate model and E-OBS data validation, Clim. Dyn., 39, 287-301, 2012.

Markovic, M., Jones, C., Vaillancourt, P. A., Paquin, D., Winger, K., and Paquin-Ricard, D.: An evaluation of the surface radiation budget over North America for a suite of regional climate models against surface station observations, Clim. Dyn., 31, 779-794, 2008.

Masson, V., Champeaux, J. L., Chauvin, F., M'eriguet, C., and Lacaze, R.: A global database of land surface parameters at 1 km resolution for use in meteorological and climate models, J. Clim., 16, 1261–1282, 2003.

McGregor, J. L.: Regional Climate Modelling, Meteorol. Atmos. Phys., 63, 105-117, 1997.

Mearns, L. O., Gutowski, W. J., Jones, R., Leung, L.-Y., McGinnis, S., Nunes, A. M. B., and Qian, Y.: A regional climate change assessment program for North America, EOS, 90(36), 311-312, 2009.

Menut, L., Tripathi, O., Colette, A., Vautard, R., Flaounas, E., and Bessagnet, B.: Evaluation of regional climate simulations for air quality modelling purposes, Clim. Dyn., 40, 2515-2533, 2013.

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., and Clough, S. A.: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave, J. Geophys. Res., 102D, 16663-16682, 1997.

Mooney, P. A., Mulligan, F. J., and Fealy, R.: Evaluation of the Sensitivity of the Weather Research and Forecasting Model to Parameterization Schemes for Regional Climates of Europe over the Period 1990-95, J. Clim., 26, 1002-1017, 2013.

Morcrette, J. J.: Impact of changes to the radiation transfer parameterizations plus cloud optical properties in the ECMWF model, Mon. Weather. Rev. 118, 847–873, 1990.

Morcrette, J. J., Smith, L., and Fouquart, Y.: Pressure and temperature dependence of the absorption in longwave radiation parameterizations, Beiträge zur Physik der Atmosphäre, 59(4), 455–469, 1986.

Morrison, H., Thompson, G., and Tatarskii, V.: Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: comparison of one- and two-moment schemes, Mon. Weather Rev., 137, 991–1007, 2009.

Neggers, R. A. J., Koehler, M., and Beljaars, A. C. M.: A dual mass flux framework for boundary layer convection. Part I: Transport, J. Atmos. Sci., 66, 1465-1487, 2009.

Neggers, R. A. J.: A dual mass flux framework for boundary layer convection. Part II: Clouds, J. Atmos. Sci., 66, 1489-1506, 2009.

Nikulin, G., Jones, C., Giorgi, F., Asrar, G., Büchner, M., Cerezo-Mota, R., Christensen, O. B., Déqué, M., Fernandez, J., Hänsler, A., van Meijgaard, E., Samuelsson, P., Bamba Sylla,

M., and Sushama, L.: Precipitation Climatology in an Ensemble of CORDEX-Africa Regional Climate Simulations, J. Clim., 25, 6057-6078, 2012.

Nordeng, T. E.: Extended versions of the convection parametrization scheme at ECMWF and their impact upon the mean climate and transient activity of the model in the tropics, ECMWF Tech. Memo. No. 206, 1994.

Paeth, H., Born, K., Podzun, R., and Jacob, D.: Regional dynamical downscaling over West Africa: model evaluation and comparison of wet and dry years, Meteorologische Zeitschrift, 14(3), 349-367, 2005.

Paeth, H.: Postprocessing of simulated precipitation for impact research in West Africa. Part I: model output statistics for monthly data, Clim. Dyn., 36, 1321-1336, 2011.

Pfeifer, S.: Modeling cold cloud processes with the regional climate model REMO, MPI for Meteorology, Hamburg, Reports on Earth System Science No. 23, 2006.

Prein, A. F., Gobiet, A., Suklitsch, M., Truhetz, H., Awan, N. K., Keuler, K., and Georgievski, G.: Added value of convection permitting seasonal simulations, Clim. Dyn., 41, 2655-2677, 2013a.

Prein, A., Holland, G. A., Rasmussen, R. M., Done, J., Ikeda, K., Clark, M. P., and Liu, C. H.: Importance of Regional Climate Model Grid Spacing for the Simulation of Heavy Precipitation in the Colorado Headwaters, J. Clim., 26, 4848-4857, 2013b.

Prömmel, K., Geyer, B., Jones, J. M., and Widmann, M.: Evaluation of the skill and added value of a reanalysis-driven regional simulation for Alpine temperature, Int. J. Clim., 30, 760-773, 2010.

Quintana-Seguı´, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchisteguy, L., and Morel, S.: Analysis of near-surface atmospheric variables: validation of the SAFRAN analysis over France, J. Appl. Meteorol. Climatol., 47, 92–107, 2008.

Rajczak, J., Pall, P., and Schär, C.: Projections of extreme precipitation events in regional climate simulations for Europe and the Alpine Region, J. Geophys. Res., 118, 3610-3626, 2013.

Räisänen, J. and Eklund, J.: 21st century changes in snow climate in Northern Europe: a high-resolution view from ENSEMBLES regional climate models, Clim. Dyn., 38, 2575-2591, 2012.

Rasch, P. J. and Kristjánsson, J. E.: A comparison of the CCM3 model climate using diagnosed and predicted condensate parameterizations, J. Clim., 11, 1587–1614, 1998.

Rauscher, S. A., Coppola, E., Piani, C., and Giorgi, F.: Resolution effects on regional climate model simulations of seasonal precipitation over Europe, Clim. Dyn., 35, 685-711, 2010.

Rauthe, M., Steiner, H., Riediger, U., Mazurkiewicz, A., and Gratzki, A.: A Central European precipitation climatology - Part I: Generation and validation of a high-resolution gridded daily data set (HYRAS), Meteorologische Zeitschrift, 22(3), 235-256, 2013.

Rechid, D., Raddatz, T., and Jacob, D.: Parameterization of snow-free land surface albedo as a function of vegetation phenology based on MODIS data and applied in climate modelling Theor. Appl. Climatol., 95, 245-255, 2009.

Ricard, J. L. and Royer, J. F.: A statistical cloud scheme for use in an AGCM, Ann. Geophys., 11, 1095–1115, 1993.

Ritter, B. and Geleyn, J.-F.: A comprehensive radiation scheme of numerical weather prediction with potential application to climate simulations, Mon. Weather Rev., 120, 303-325, 1992.

Rockel, B. and Geyer, B.: The performance of the regional climate model CLM in different climate regions, based on the example of precipitation, Meteorologische Zeitschrift, 17(4), 487-498, 2008.

Roesch, A., Jaeger, E. B., Lüthi, D., and Seneviratne, S. I.: Analysis of CCLM model biases in relation to intra-ensemble model variability, Meteorologische Zeitschrift, 17(4), 369-382, 2008.

Rubel, F. and Hantel, M.: BALTEX 1/6-degree daily precipitation climatology 1996–1998, Meteorol. Atmos. Phys., 77, 155-166, 2001.

Salzmann, N. and Mearns, L. O.: Assessing the Performance of Multiple Regional Climate Model Simulations for Seasonal Mountain Snow in the Upper Colorado River Basin, J. Hydromet., 13, 539-556, 2012.

Samuelsson, P., Gollvik, S., and Ullerstig, A.: The land-surface scheme of the Rossby Centre regional atmospheric climate model (RCA3). SMHI Rep Met 122, 25, 2006.

Sass, B. H., Rontu, L., Savijaärvi, H., and Räisänen, P.: HIRLAM-2 radiation scheme: documentation and tests. SMHI HIRLAM Technical Report No. 16, 1994.

Savijärvi, H.: A fast radiation scheme for mesoscale model and short-range forecast models, J. Appl. Meteorol., 29, 437–447, 1990.

Seneviratne, S.I., Lüthi, D., Litschi, M., and Schär, C.: Land–atmosphere coupling and climate change in Europe, Nature, 443, 205-209, 2006.

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., Teuling, A. J.: Investigating soil moisture–climate interactions in a changing climate: A review, Earth-Science Rev., 99, 125-161, 2010.

Sevruk, B.: Correction of precipitation measurements summary report. In: Sevruk, B. (ed) Correction of precipitation measurements, Züricher Geographische Schriften 23, 13-23, 1986.

Siebesma, A.P., Soares, P. M. M., and Teixeira, J.: A Combined Eddy-Diffusivity Mass-Flux Approach for the Convective Boundary Layer, J. Atmos. Sci., 64, 1230-1248, 2007.

Smiatek, G., Kunstmann, H., Knoche, R., and Marx, A.: Precipitation and temperature statistics in high-resolution regional climate models: Evaluation for the European Alps, J. Geophys. Res., 114, D19107, 2009.

Steger, C., Kotlarski, S., Jonas, T., and Schär, C.: Alpine snow cover in a changing climate: A regional climate model perspective, Clim. Dyn., 41, 735-754, 2013.

Suklitsch, M., Gobiet, A., Leuprecht, A., and Frei, C.: High Resolution Sensitivity Studies with the Regional Climate Model CCLM in the Alpine Region, Meteorologische Zeitschrift, 17(4), 467-476, 2008.

Suklitsch, M., Gobiet, A., Truhetz, H., Awan, N. K., Göttel, H., and Jacob, D.: Error characteristics of high resolution regional climate models over the Alpine area, Clim. Dyn., 37, 377-390, 2011.

Themeßl, M. J., Gobiet, A., and Heinrich, G.: Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal, Clim. Change, 112, 449-468, 2012.

Tiedtke, M.: A comprehensive mass flux scheme for cumulus parameterization in large-scale models, Mon. Weather Rev., 117, 1779–1799, 1989.

Tiedtke, M.: Representation of clouds in large-scale models, Mon. Weather. Rev., 121, 3040-3061, 1993.

Tompkins, A. M., Gierens, K., and Rädel, G.: Ice supersaturation in the ECMWF Integrated Forecast System, Q. J. R. Meteorol. Soc., 133, 53-63, 2007.

Uppala, S. M., Kallberg, P. W., Simmons, A. J. et al.: The ERA-40 re-analysis, Q. J. Roy. Meteor. Soc., 131, 2961–3012, 2005.

van den Hurk, B. J. J. M., Viterbo, P., Beljaars, A. C. M., and Betts, A. K.: Offline validation of the ERA40 surface scheme, ECMWF Tech. Report No. 75, ECMWF, 2000.

van den Hurk, B. J. J. M., Klein Tank, A., Lenderink, G., van Ulden, A., van Oldenborgh, G. J., Katsman, C., van den Brink, H., Keller, F., Bessembinder, J., Burgers, G., Komen, G., Hazeleger, W., and Drijfhout, S.: New climate change scenarios for the Netherlands, Water Sci. Technol., 56(4), 27–33, 2007.

van der Linden, P. and Mitchell, J. F. B.: ENSEMBLES: Climate change and its Impacts: Summary of research and results from the ENSEMBLES project, Met Office Hadley Centre, Exeter, UK, 2009.

Vautard, R., Noël, T., Li, L., Vrac, M., Martin, E., Dandin, P., Cattiaux, J., and Joussaume, S.: Climate variability and trends in downscaled high-resolution simulations and projections over metropolitan France, Clim. Dyn., 41, 1419-1437, 2013a.

Vautard, R., Gobiet, A., Jacob, D., Belda, M., Colette, A., Déqué, M., Fernández, J., García-Díez, M., Goergen, K., Güttler, I., Halenka, T., Karakostas, T., Katragkou, E., Keuler, K., Kotlarski, S., Mayer, S., van Meijgaard, E., Nikulin, G., Patarčić, M., Scinocca, J., Sobolowski, S., Suklitsch, M., Teichmann, C., Warrach-Sagi, K., Wulfmeyer, V., and Yiou, P.: The simulation of European heat waves from an ensemble of regional climate models within the EURO-CORDEX project, Clim. Dyn., 41(9-10), 2555-2575, 2013b.

Vidale, P. L., Lüthi, D., Wegmann, R., and Schär, C.: European summer climate variability in a heterogeneous multi-model ensemble, Clim. Change, 81, 209-232, 2007.

Wang, Y., Leung, L. R., McGregor, J. L., Lee, D.-K., Wang, W.-C., Ding, Y., and Kimura, F.: Regional Climate Modeling: Progress, Challenges, and Prospects, J. Met. Soc. Jpn., 82(6), 1599-1628, 2004.

Warrach-Sagi, K., Schwitalla, T., Wulfmeyer, V., and Bauer, H.-S.: Evaluation of a climate simulation in Europe based on the WRF-NOAH model system: precipitation in Germany, Clim. Dyn., 41, 755-774, 2013.

Wehner, M. F.: Very extreme seasonal precipitation in the NARCCAP ensemble: model performance and projections, Clim. Dyn., 40, 59-80, 2013.

Widmann, M., Bretherton, C. S., and Salathé Jr., E. P.: Statistical Precipitation Downscaling over the Northwestern United States Using Numerically Simulated Precipitation as a Predictor, J. Climate, 16, 799-816, 2003.

Wilby, R.L. and Fowler, H. J.: Regional Climate Downscaling, In: Fung F, Lopez A, New M (ed) Modelling the Impact of Climate Change on Water Resources, Blackwell Publishing Ltd., Chapter 3, pp 34-85, 2011.

Table 1. Overview on the models analyzed and their main characteristics.

Model	Institution (abbreviation)	Boundary forcing	Horizontal resolution	No. of vertical levels	Radiation scheme	Convection scheme	Microphysics scheme	Land-surface scheme	Land use	Planetary boundary layer scheme	Soil initialization and spin-up
ARPEGE 5.1	Centre National de Recherches Météorologiques **(CNRM)**	nudged towards ERA- Interim outside the EURO-CORDEX domain	0.11°, 0.44°	31	Morcrette (1990)	Bougeault (1985)	Ricard and Royer (1993)	Douville et al. (2000)	ECOCLIMAP (Champeaux et al., 2003; Masson et al., 2003)	Ricard and Royer (1993)	Year 1989 simulated twice after model cold start (i.e., one year spin-up)
CCLM 4.8.17	CLM Community **(CLMCOM)**	ERA- Interim directly	0.11°, 0.44°	40	Ritter and Geleyn (1992)	Tiedtke (1989)	Doms et al.(2007), Baldauf and Schulz (2004)	TERRA-ML: Doms et al.(2007)	GLC2000 (Joint Research Centre, 2003)	Louis (1979)	0.11°: Idealized profile for soil temperature and climatology for soil moisture 0.44°: 10-year spin-up with standard initialization (cold start) in 1979)
HIRHAM 5	Danish Meteorological Institute (**DMI**)	ERA-Interim directly	0.11°, 0.44°	31	Morcrette et al. (1986), Giorgetta and Wild (1995)	Tiedtke (1989)	Lohmann and Roeckner (1996)	Hagemann (2002)	Claussen et al. (1994)	Louis (1979)	Soil moisture initialized full, soil temperature climatology. One year spin-up 1988
REMO 2009	Climate Service Center **(CSC)**	ERA- Interim directly	0.11°, 0.44°	27	Morcrette et al. (1986), Giorgetta and Wild (1995)	Tiedtke (1989), Nordeng (1994), Pfeifer (2006)	Lohmann and Roeckner (1996)	Hagemann (2002), Rechid et al. (2009)	USGS (Hagemann, 2002)	Louis (1979)	Soil temperature and moisture are initialized using ERA-Interim (cold start)
RACMO 2.2	Royal Netherlands Meteorological Institute **(KNMI)**	ERA- Interim directly	0.11°, 0.44°	40	Fouquart and Bonnel (1980), Mlawer et al. (1997)	Tiedtke (1989), Nordeng (1994), Neggers et al. (2009)	Tiedtke (1993), Tompkins et al. (2007), ECMWF-IFS (2007), Neggers (2009)	Van den Hurk et al. (2000), Balsamo et al. (2009)	ECOCLIMAP (Champeaux et al., 2003; Masson et al., 2003)	Lenderink and Holtslag (2004), Siebesma et al. (2007)	Soil temperature and soil moisture initialized from ERA-Interim on 1st January 1979
RCA 4	Swedish Meteorological and Hydrological Institute **(SMHI)**	ERA- Interim directly	0.11°, 0.44°	40	Savijärvi (1990), Sass et al. (1994)	Kain and Fritsch (1990), Kain and Fritsch (1993)	Rasch and Kristjánsson (1998)	Samuelsson et al. (2006)	ECOCLIMAP (Champeaux et al., 2003; Masson et al., 2003)	Cuxart et al. (2000)	Soil temperature and soil moisture initialized with ERA-Interim in January 1979 (10-year spin-up)
WRF 3.3.1	University of Hohenheim **(UHOH)**	ERA- Interim directly	0.11°	50	CAM 3.0: Collins et al. (2004)	Modified Kain-Fritsch: Kain (2004)	Morrison 2-moment scheme : Morrison et al. (2009)	NOAH: Ek et al. (2003)	IGBP-MODIS 30''	YSU : Hong et al. (2006)	Soil temperatures and soil moisture initialized on 1st January 1987 with ERA-Interim, 2-year spin-up
WRF 3.3.1	Institut Pierre Simon Laplace / Institut National de l'Environnement Industriel et des Risques **(IPSL-INERIS)**	ERA- Interim directly	0.11°, 0.44°	32	RRTMG : Lacono et al. (2008)	Grell and Devenyi (2002)	Hong et al. (2004)	NOAH: Ek et al. (2003)	USGS Land Use	YSU : Hong et al. (2006)	Soil moisture and soil temperature initialized with ERA-Interim (cold start), no spin-up
WRF 3.3.1	Centre de Recherche Public Gabriel Lippmann **(CRP-GL)**	ERA- Interim, CRP-GL-11 nested into CRP-GL-44 (one-way double nesting)	0.11°, 0.44°	50	CAM 3.0: Collins et al. (2004)	Modified Kain-Fritsch: Kain (2004)	WSM 6-class: Hong and Lim (2006)	NOAH: Ek et al. (2003)	IGBP-MODIS 30''	YSU : Hong et al. (2006)	Soil moisture and soil temperature initialized with ERA-Interim (cold start), no spin-up

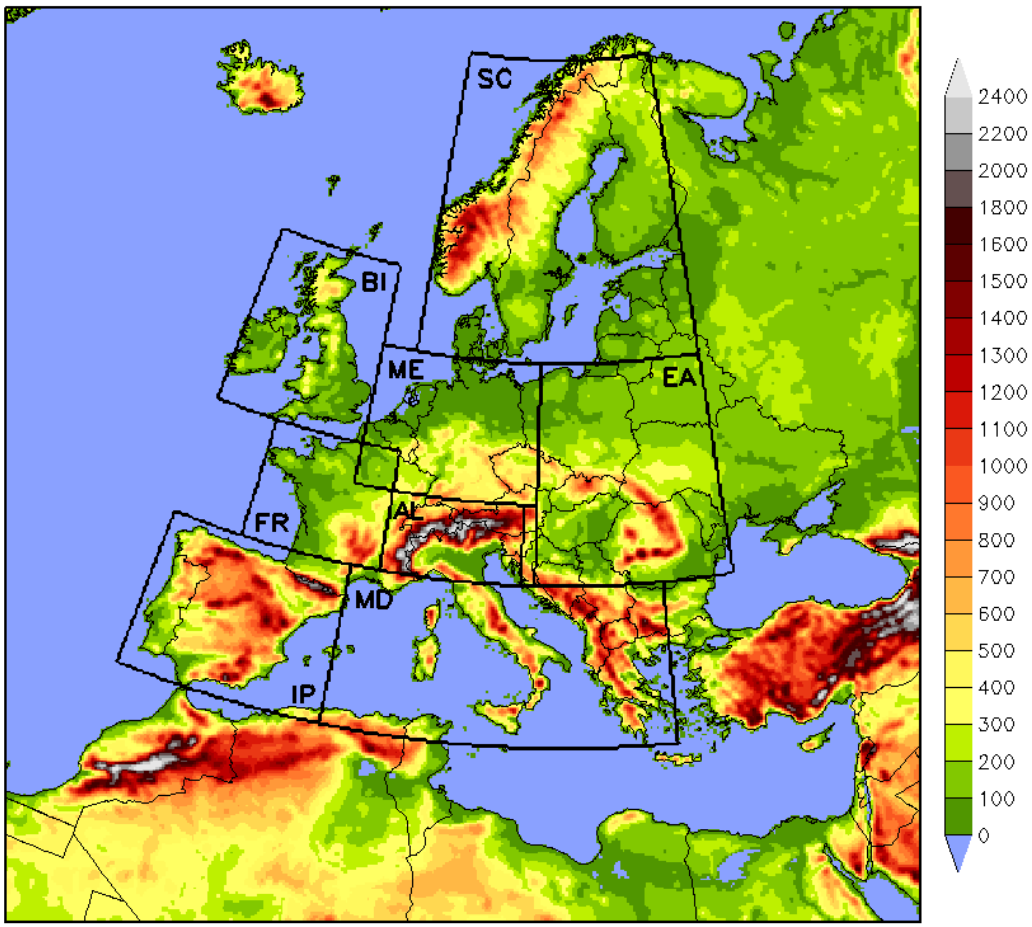


Figure 1. The common EURO-CORDEX analysis domain and location of the eight sub-domains used for model evaluation. The color represents the orography of the CLMCOM-11 setup [m].

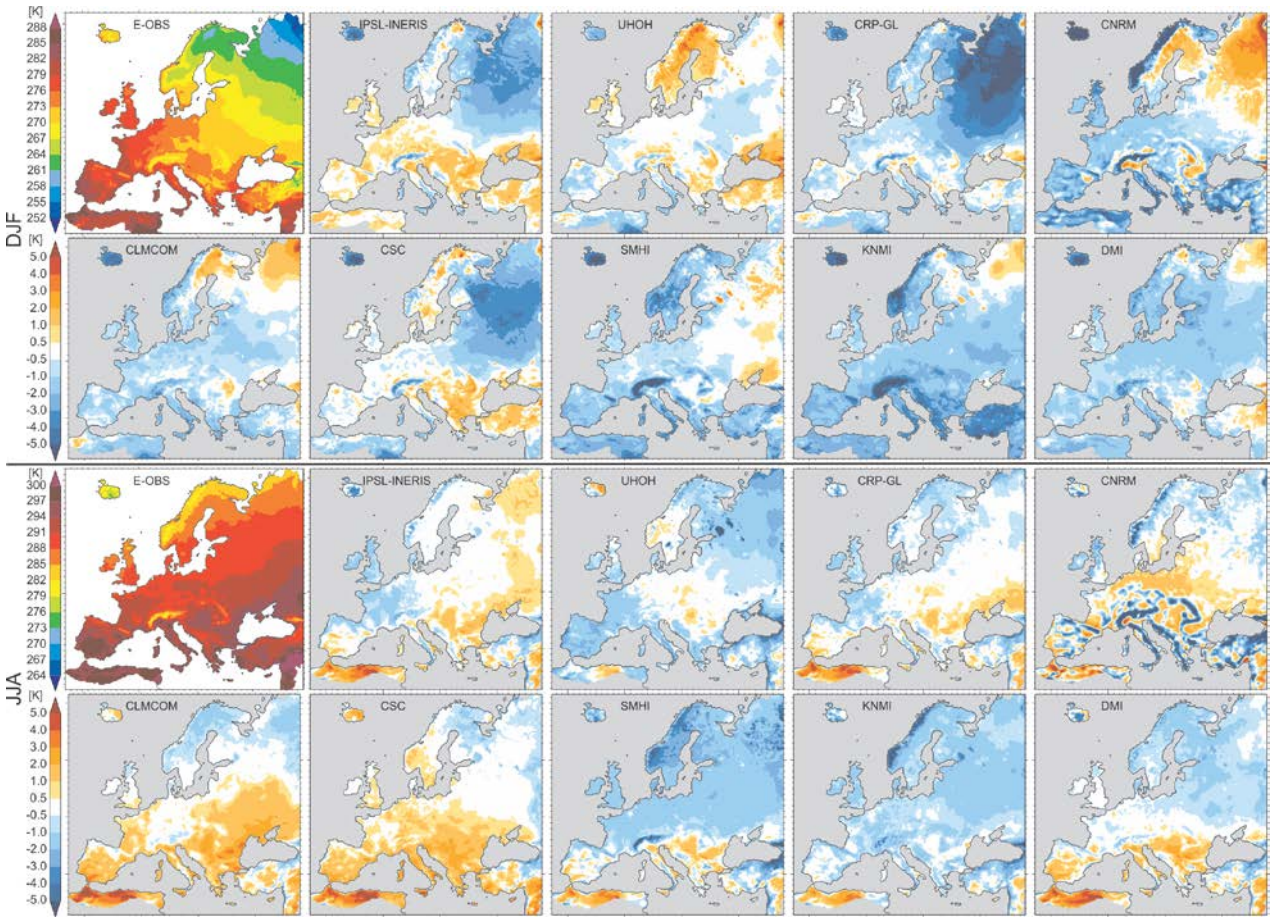


Figure 2. Mean seasonal temperature bias [K] for all experiments of the EUR-11 ensemble and the period 1989-2008. Upper rows: winter (DJF), lower rows: summer (JJA). The upper left panel of each section shows the horizontal pattern of mean seasonal temperature as provided by the E-OBS reference [K].

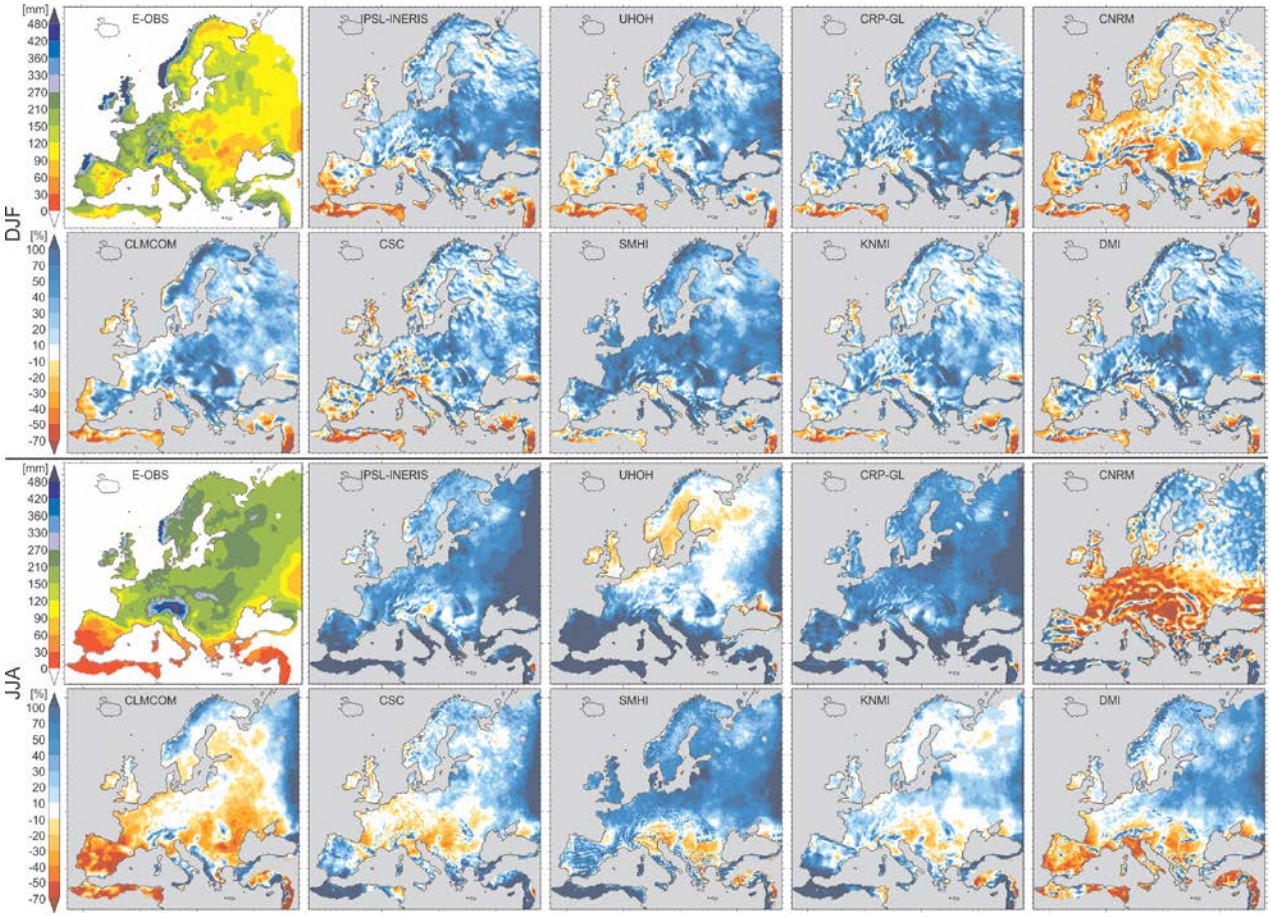


Figure 3. As Figure 2 but for the mean relative seasonal precipitation bias [%]. The upper left panel of each section shows the horizontal pattern of mean seasonal precipitation as provided by the E-OBS reference [mm month^{-1}].

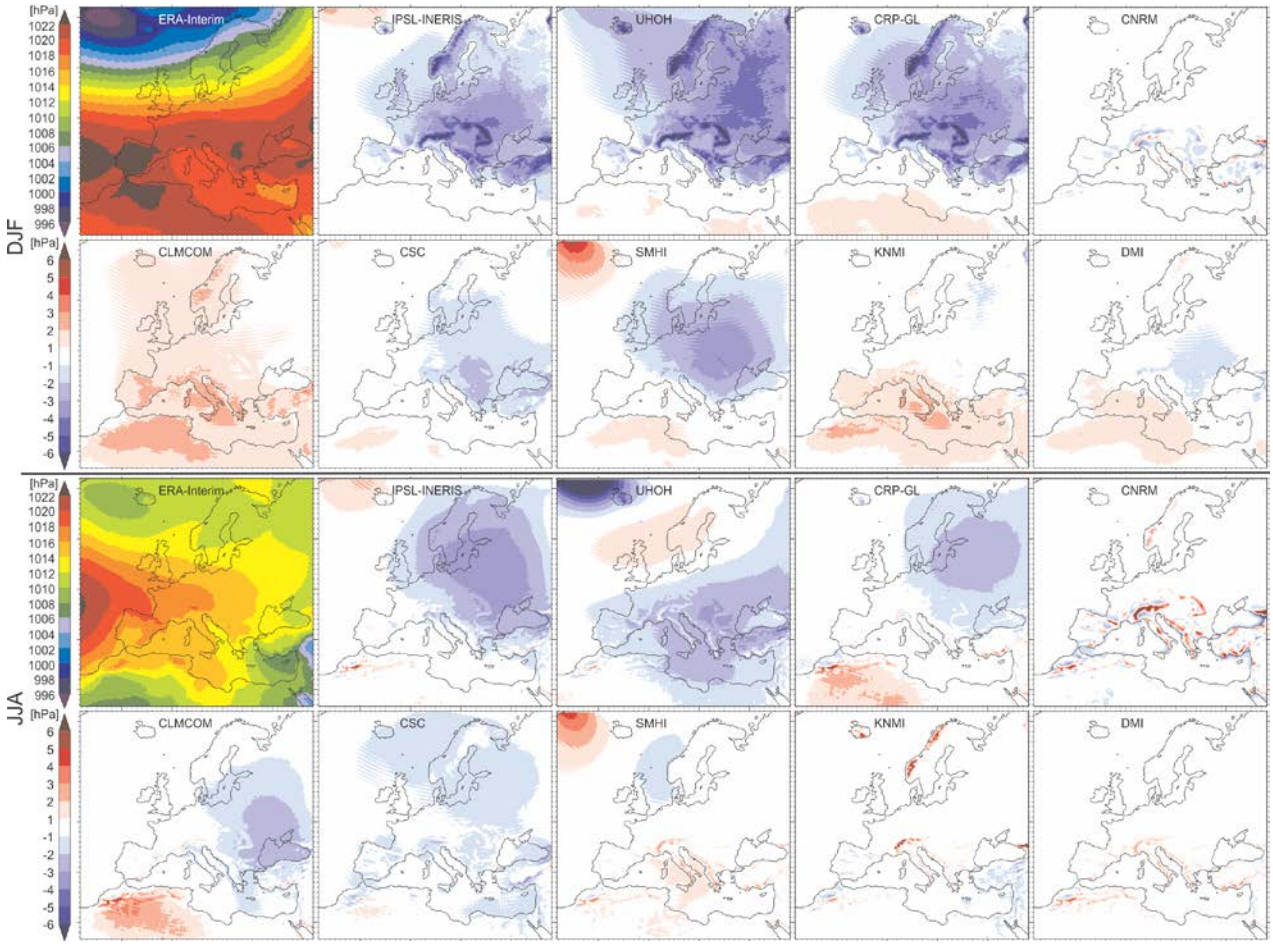
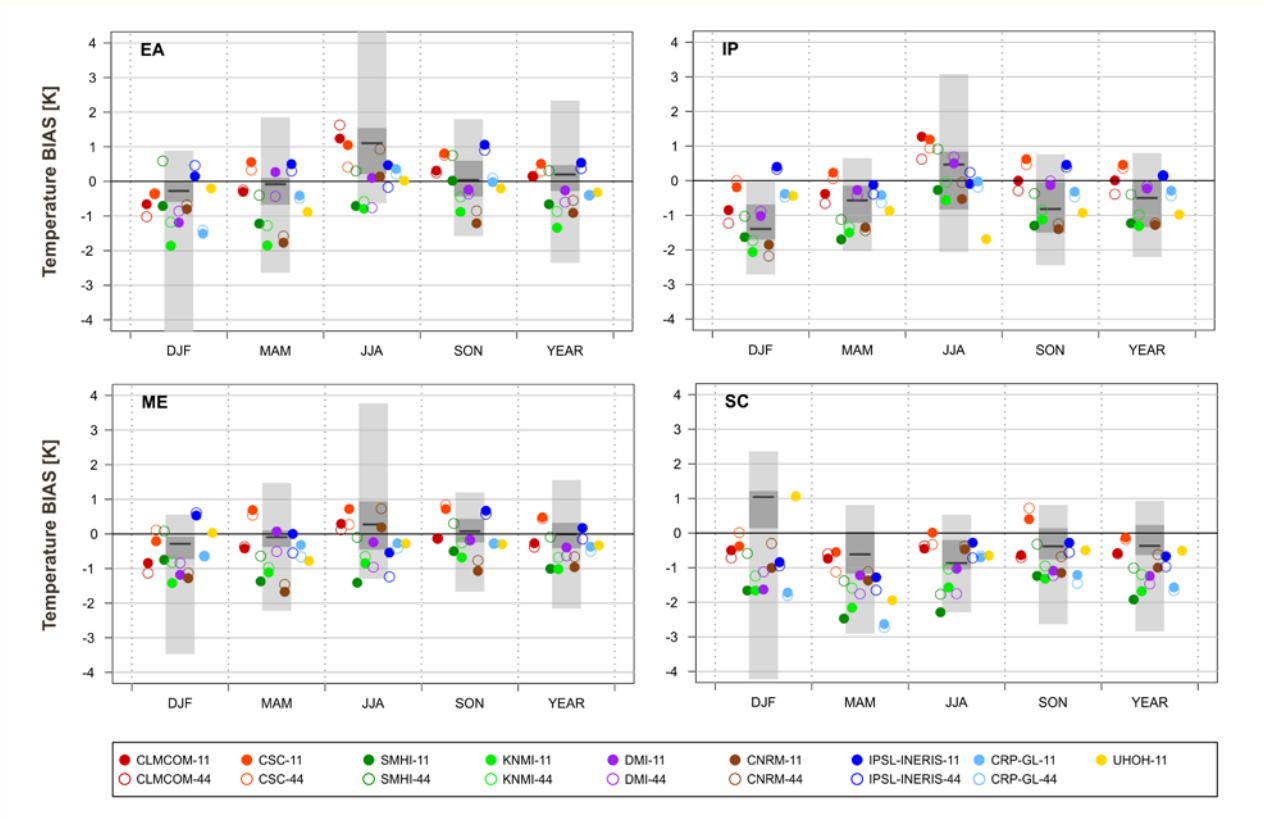


Figure 4. As Figure 2 but for the mean seasonal mean sea-level pressure bias [hPa]. The upper left panel of each section shows the horizontal pattern of mean seasonal precipitation as provided by the ERA-Interim reference [hPa].



1

2 Figure 5. Mean seasonal and annual temperature bias (BIAS) [K] for the EUR-11 (filled circles)
3 and the EUR-44 ensemble (open circles) and for sub-domains EA, IP, ME and SC (see Figure
4 B1 for sub-domains AL, BI, FR and MD). The gray bars denote the BIAS range of the ENS-22
5 ensemble: entire range in light gray, inter-quartile range (corresponding to eight models) in dark
6 gray and median as solid line.

7

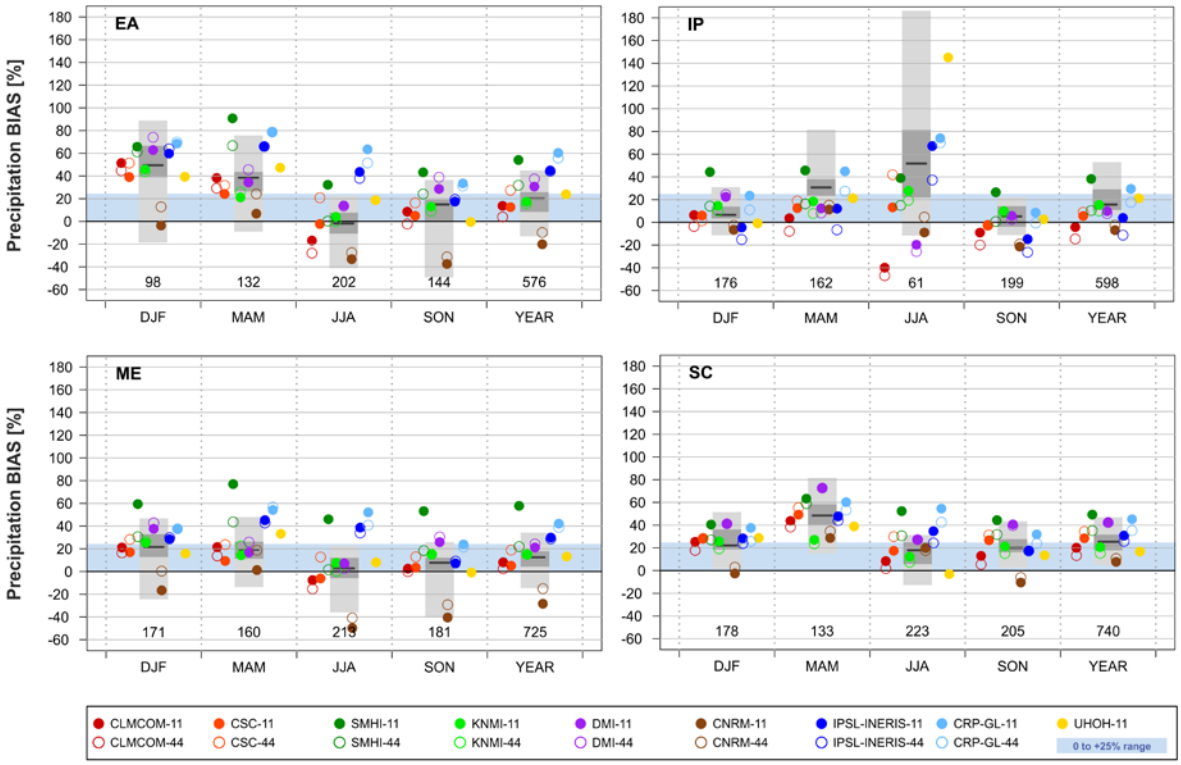


Figure 6. As Figure 5 but for the mean seasonal and annual relative precipitation bias (BIAS) [%]. The numbers along the x-axis indicate mean seasonal [mm/season] and mean annual [mm/year] precipitation sums for the period 1989-2008 in the E-OBS reference. The blue shading indicates a bias range between 0 and +25%, corresponding to acceptable model biases in case of a systematic rain gauge undercatch of up to 20% of true precipitation. See Figure B2 for sub-domains AL, BI, FR and MD.

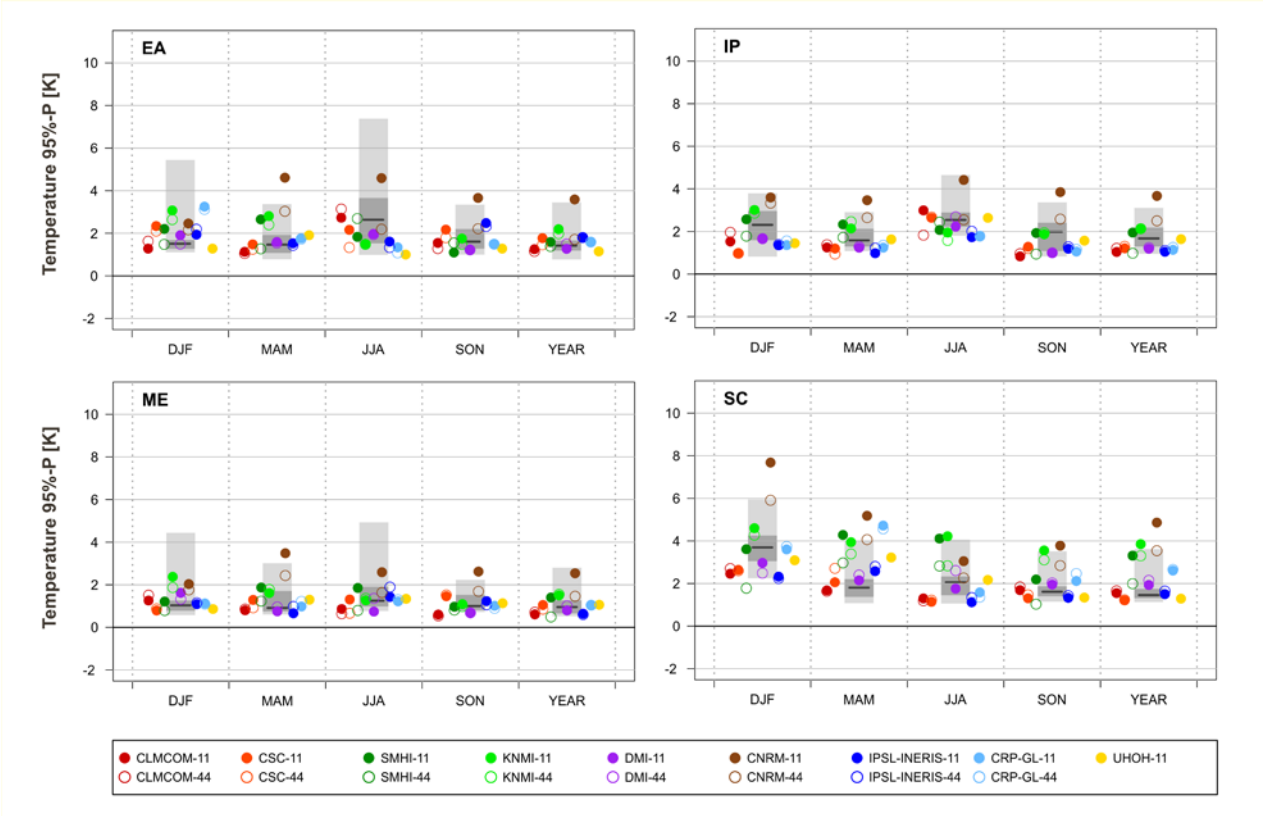


Figure 7. 95th percentile of mean seasonal and annual absolute temperature biases (95%-P) [K] for the EUR-11 (filled circles) and the EUR-44 ensemble (open circles) and for sub-domains EA, IP, ME and SC (see Figure B3 for sub-domains AL, BI, FR and MD). The gray bars denote the 95%-P range of the ENS-22 ensemble: entire range in light gray, inter-quartile range (corresponding to eight models) in dark gray and median as solid line.

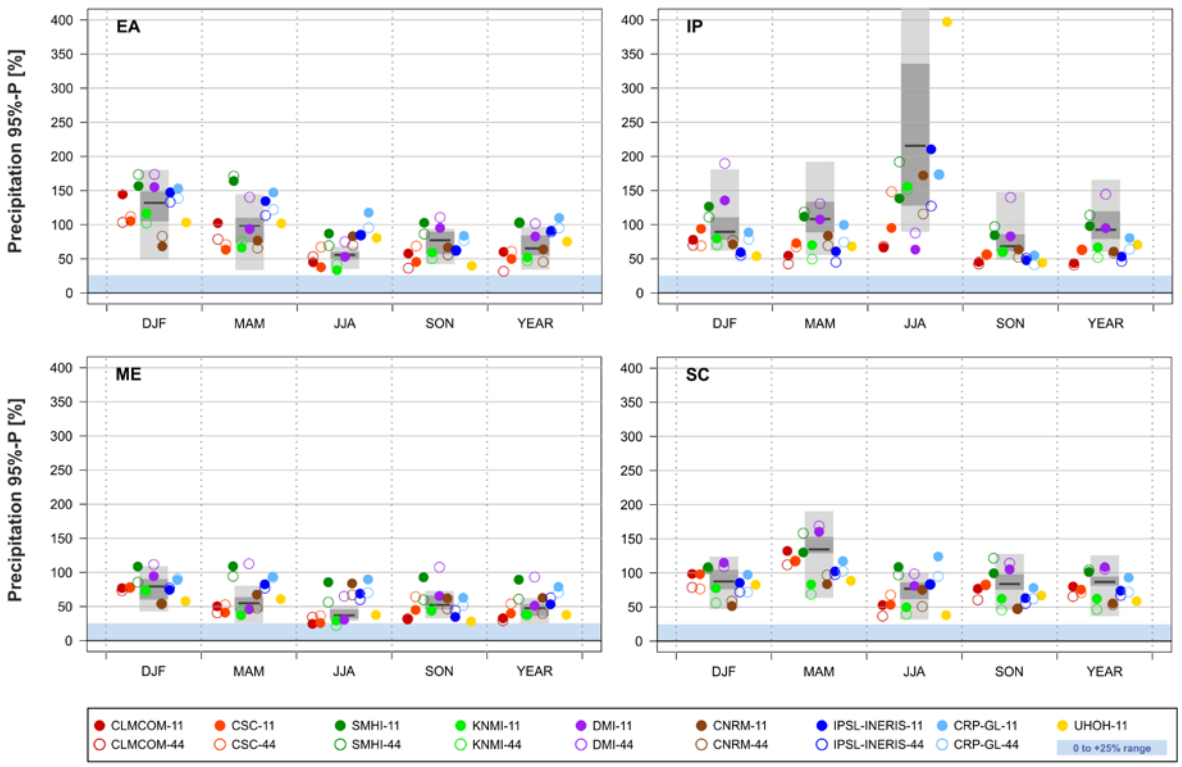


Figure 8. As Figure 7 but for the 95th percentile of mean seasonal and annual relative precipitation biases (absolute values; 95%-P) [%]. The blue shading indicates a bias range between 0 and +25%, corresponding to acceptable wet model biases in case of a systematic rain gauge undercatch of up to 20% of true precipitation. See Figure B4 for sub-domains AL, BI, FR and MD.

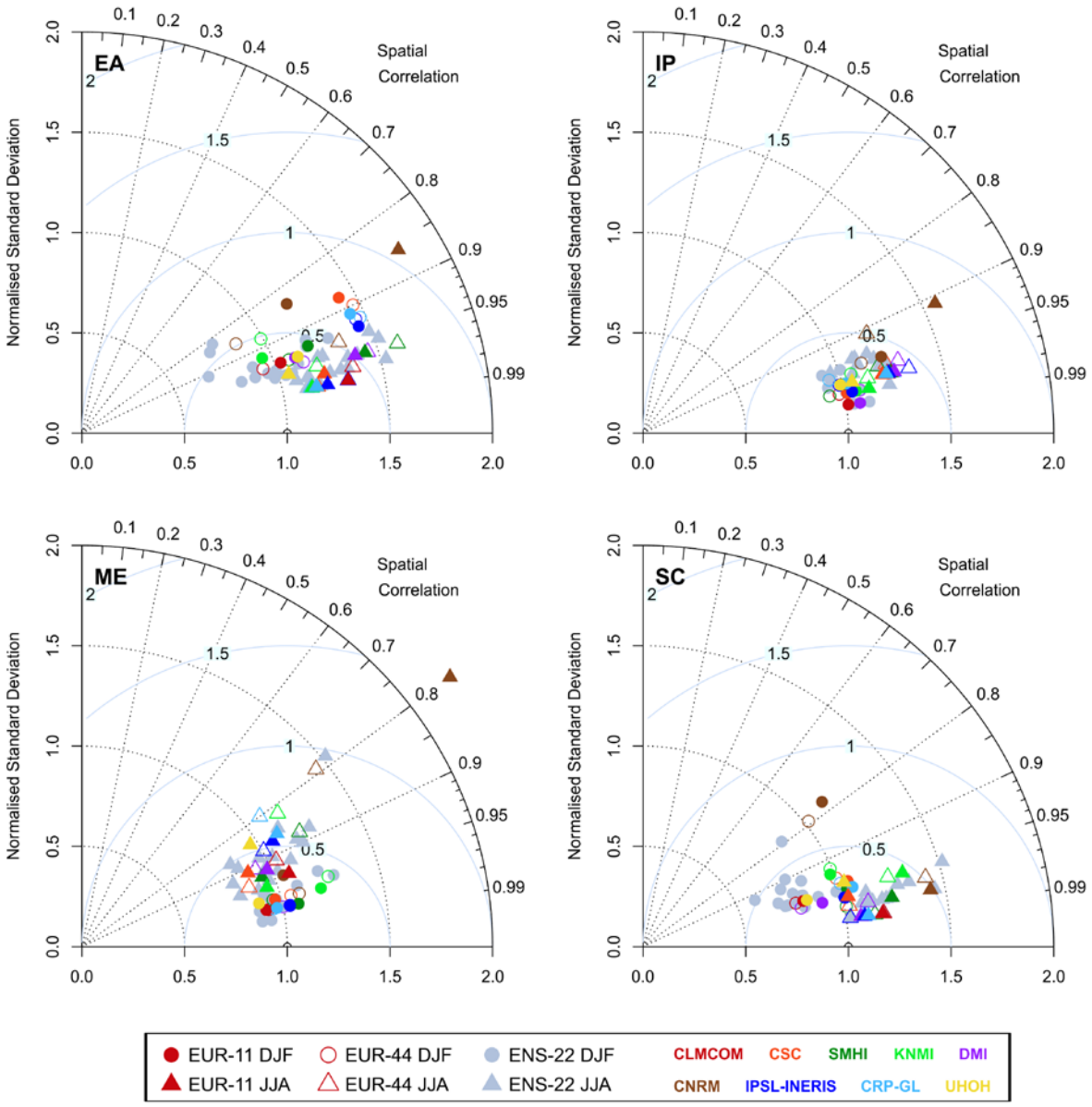


Figure 9. Spatial Taylor diagrams exploring the model performance with respect to the spatial variability of mean winter (circles) and mean summer (triangles) temperature within sub-domains EA, IP, ME and SC (see Figure B5 for sub-domains AL, BI, FR and MD). Filled markers: EUR-11 ensemble, non-filled markers: EUR-44 ensemble, gray markers: ENS-22 ensemble. The diagrams combine the spatial pattern correlation (PACO, cos(azimuth angle)) and the ratio of spatial variability (RSV, radius). The distance from the 1-1 location corresponds to the normalized and centered root-mean-square difference (which does not take into account the mean model bias), expressed as multiples of the observed standard deviation. Note the different number of underlying grid cells per sub-domain in the individual ensembles.

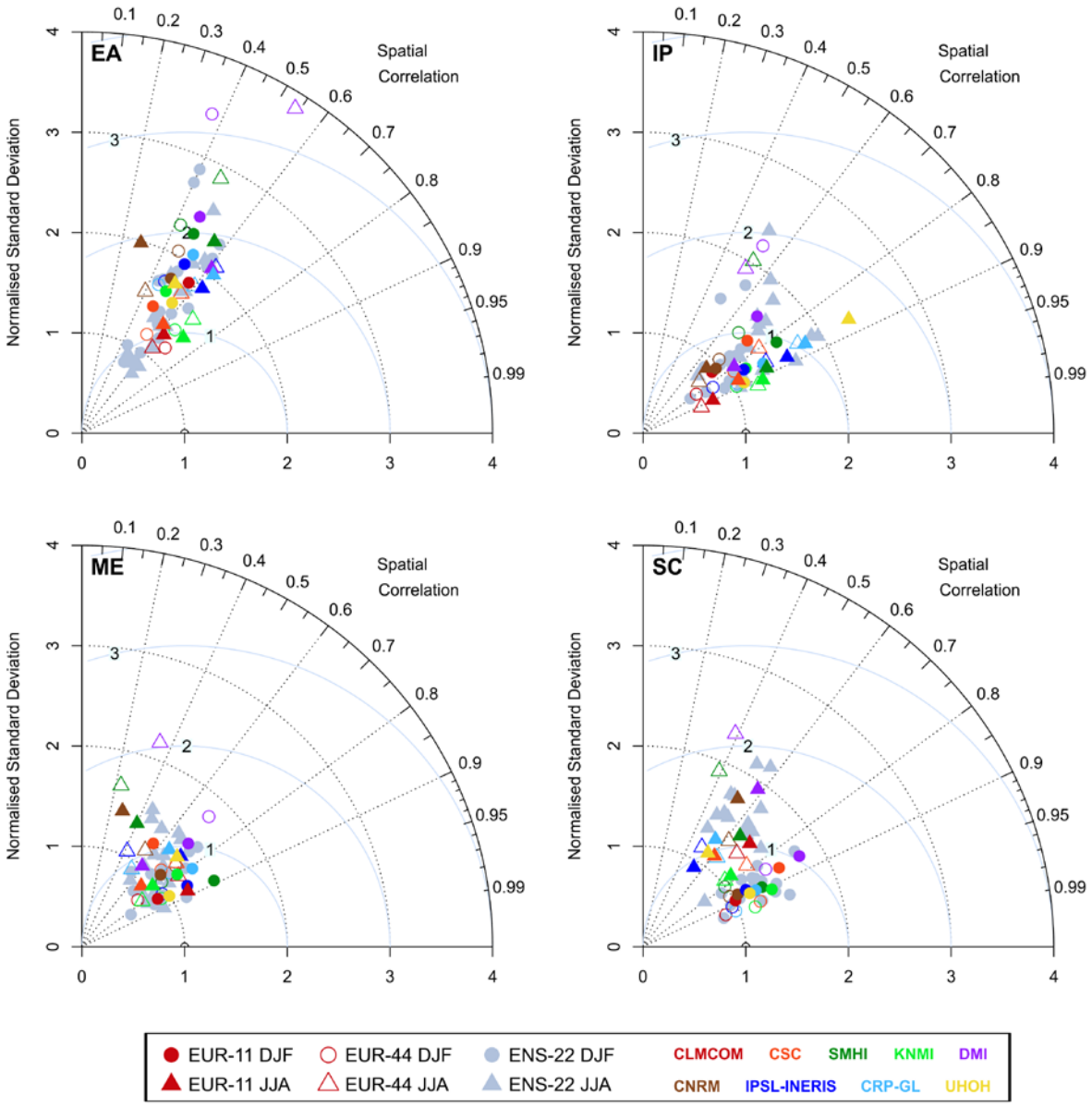


Figure 10. As Figure 9 but for mean winter (circles) and mean summer (triangles) precipitation. See Figure B6 for sub-domains AL, BI, FR and MD.

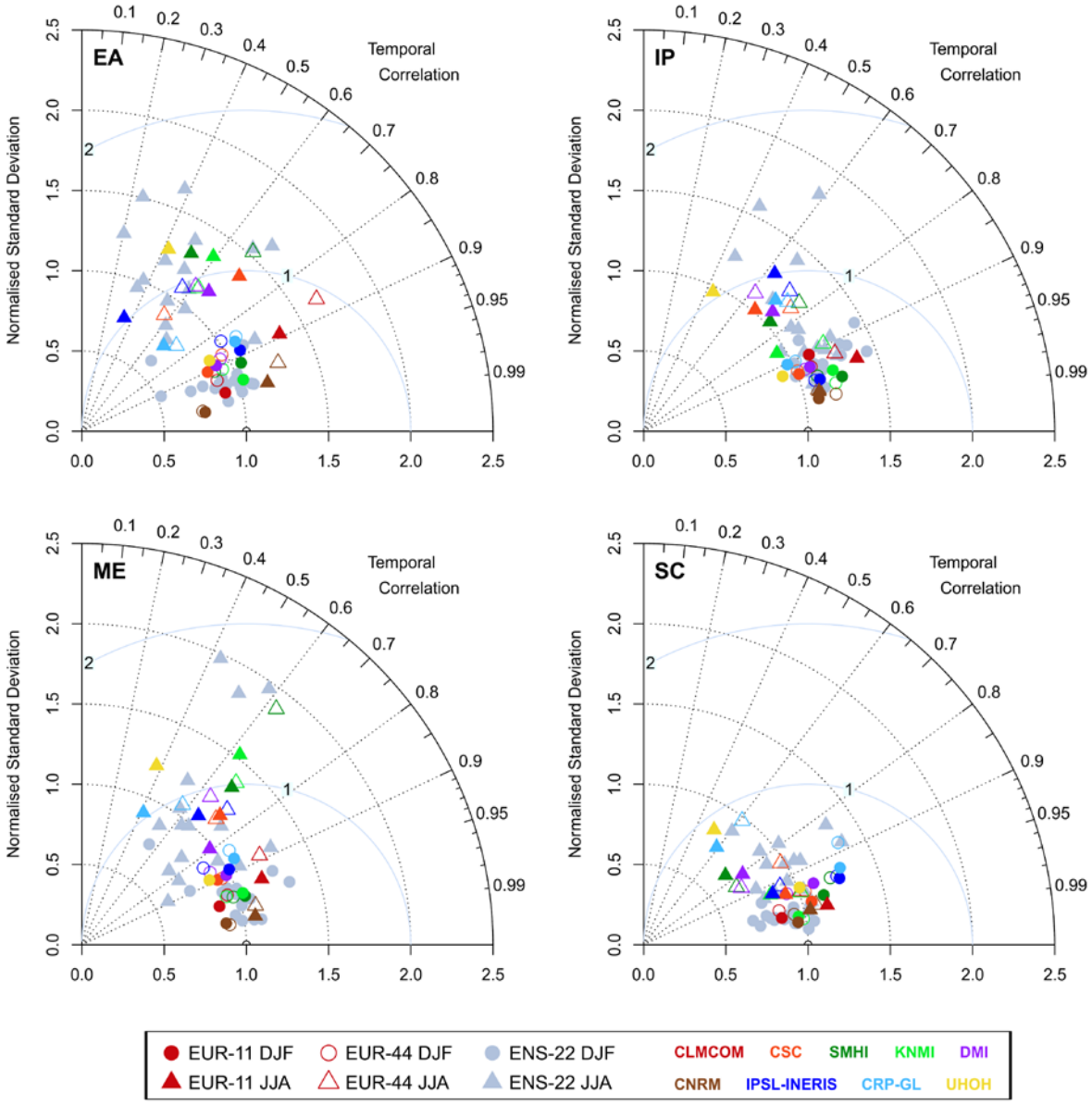


Figure 11. Temporal Taylor diagrams exploring the model performance with respect to the inter-annual temporal variability of mean winter (circles) and mean summer (triangles) temperature as averages over sub-domains EA, IP, ME and SC (see Figure B7 for sub-domains AL, BI, FR and MD). Filled markers: EUR-11 ensemble, non-filled markers: EUR-44 ensemble, gray markers: ENS-22 ensemble. The diagrams combine the temporal correlation of interannual variability (TCOIAV, cos(azimuth angle)) and ratio of interannual variability (RIAV, radius). The distance from the 1-1 location corresponds to the normalized and centered root-mean-square difference (which does not take into account the mean model bias), expressed as multiples of the observed standard deviation.

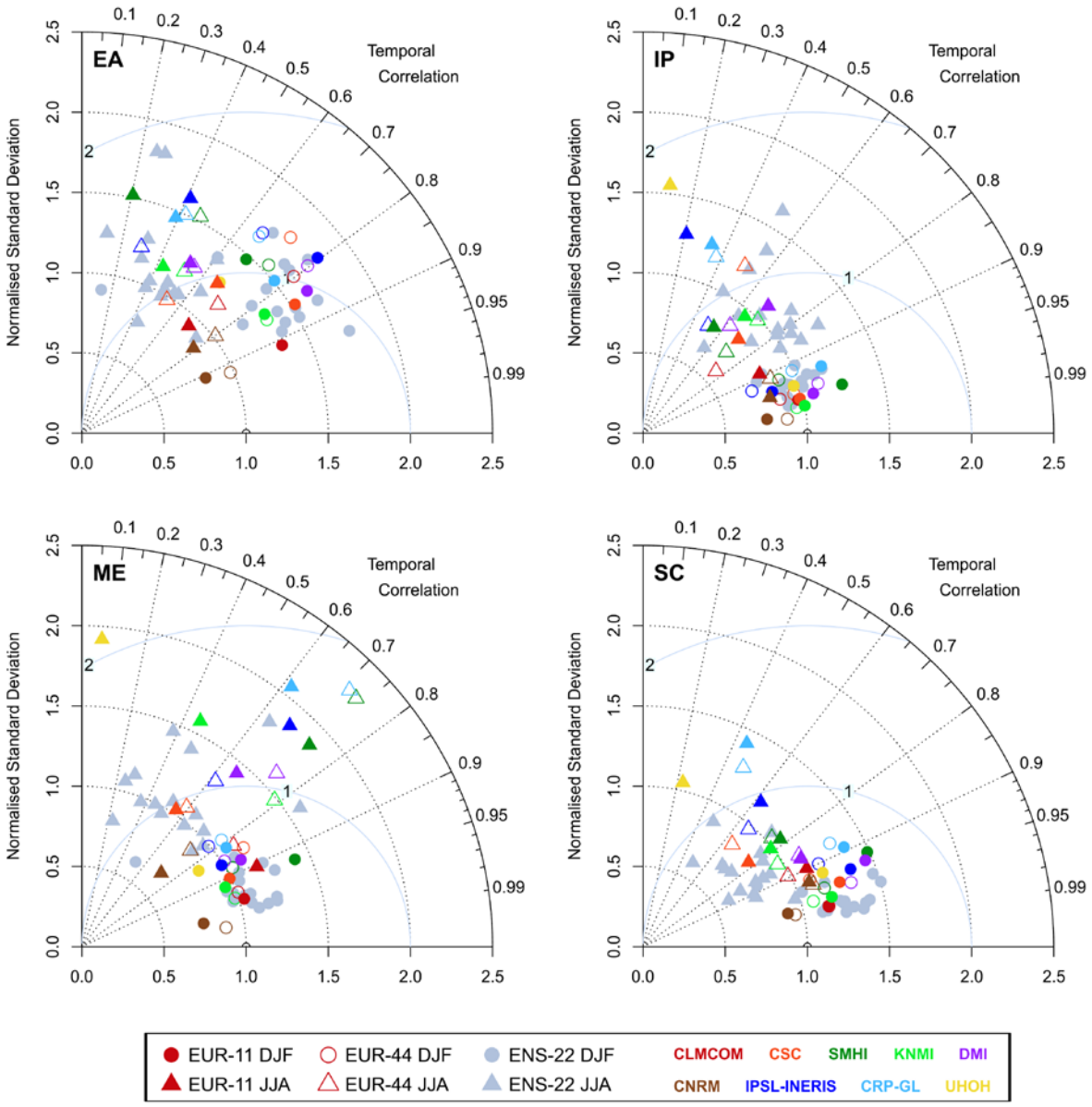


Figure 12. As Figure 11 but for mean winter (circles) and mean summer (triangles) precipitation. See Figure B8 for sub-domains AL, BI, FR and MD.

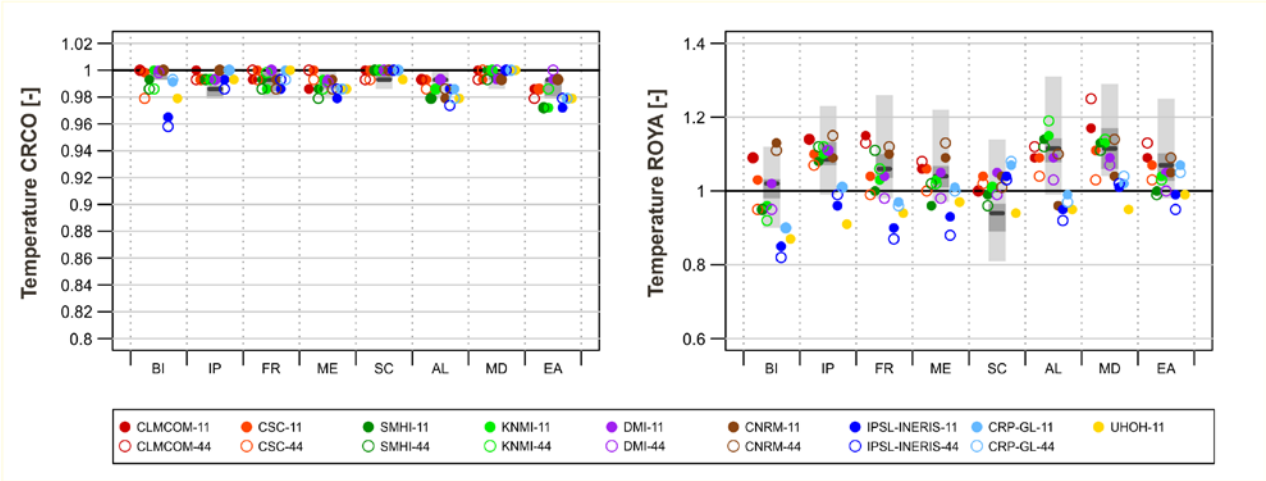


Figure 13. Model performance with respect to the mean annual cycle of temperature over each sub-domain as expressed by the climatological rank correlation (CRCO, left panel) and the ratio of yearly amplitudes (ROYA , right panel). The gray bars denote the CRCO and ROYA ranges of the ENS-22 ensemble: entire range in light gray, inter-quartile range (corresponding to eight models) in dark gray and median as solid line.

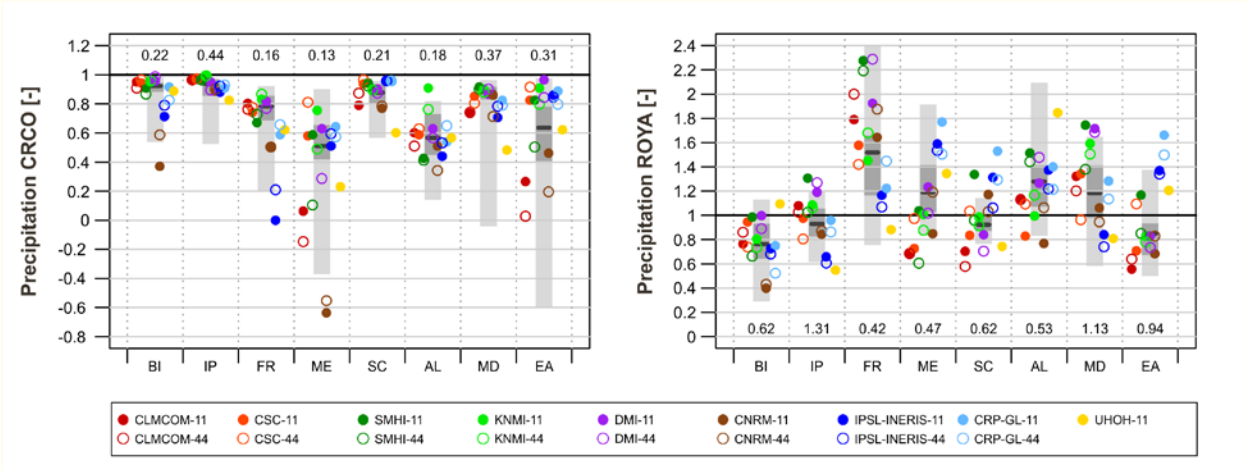


Figure 14. As Figure 13 but for the mean annual cycle of precipitation. The numbers along the x-axes indicate the standard deviation of the mean annual cycle (left panel) and the maximum difference between the climatological monthly means (right panel), both normalized by the annual mean monthly precipitation and based on average precipitation sums over the respective sub-domain in the E-OBS reference.

Appendix A: Definition of evaluation metrics

Let M_{nki} and R_{nki} be the annual, seasonal or monthly mean value of any variable of the model simulation (*M*) and the reference data (*R*) of year *i* at grid point *n* with

$n = 1,\ldots,N$ *N* = number of grid points of sub-region *SR*

$k = 1,\ldots,K$ *K*= number of analyzed periods per year:

K=12 for monthly, *K* =4 for seasonal, *K* =1 for yearly values

$i = 1,\ldots,I$ *I* = number of years (20 in this case)

The simulated spatial mean of period *k* and year *i* across a sub-region *SR* is defined as:

$$\hat{M}_{ki} = \frac{1}{N}\sum_{n \in SR} M_{nki} \qquad \text{(A1)}$$

The climatological mean of period *k* at grid point *n* is defined as:

$$\overline{M}_{nk} = \frac{1}{I}\sum_{i=1}^{I} M_{nki} \qquad \text{(A2)}$$

The climatological mean of period *k* averaged across a sub-region *SR* is then computed as:

$$\hat{\overline{M}}_{k} = \frac{1}{N}\sum_{n \in SR} \overline{M}_{nk} = \frac{1}{I}\sum_{i=1}^{I} \hat{M}_{ki} = \overline{\hat{M}}_{k} \qquad \text{(A3)}$$

The corresponding means for the reference data *R* are defined accordingly. Annual means are calculated by an unweighted average over 12 monthly means beginning with January. Seasonal means of year *i* are calculated by an unweighted average over three consecutive monthly means beginning with December of year *i*-1 for the winter season (DJF) and ending with November of year *i* for the fall season (SON).

Using these definitions, the applied evaluation metrics are calculated as follows.

Mean bias (*BIAS*)

for climatological annual (*k=1*), seasonal (*k=1,…,4*) and monthly (k=1,…12) mean values averaged across a sub-region:

$$BIAS_k = \hat{\overline{M}}_{k} - \hat{\overline{R}}_{k} \qquad \text{(A4)}$$

For precipitation, the relative difference with respect to the reference data is used.

95% percentile of the absolute value of grid point differences (*95%-P*)

$$95\%-P_k = \max_{n\in X}|\overline{M}_{nk} - \overline{R}_{nk}| \qquad X = \{n \in SR \mid Rank\ |\overline{M}_{nk} - \overline{R}_{nk}| \le 0.95N\} \tag{A5}$$

For precipitation, relative differences with respect to the reference data are used.

Pattern correlation (*PACO*)

$$PACO_k = \frac{1}{N-1}\sum_{n\in SR}\frac{(\overline{M}_{nk} - \hat{\overline{M}}_k)\cdot(\overline{R}_{nk} - \hat{\overline{R}}_k)}{\sigma_{SM_k}\cdot\sigma_{SR_k}} \tag{A6}$$

with the spatial variances

$$\sigma^2_{SM_k} = \frac{1}{N-1}\sum_{n\in SR}(\overline{M}_{nk} - \hat{\overline{M}}_k)^2 \quad \text{and} \quad \sigma^2_{SR_k} = \frac{1}{N-1}\sum_{n\in SR}(\overline{R}_{nk} - \hat{\overline{R}}_k)^2 \tag{A7}$$

Ratio of spatial variability (*RSV*)

$$RSV_k = \frac{\sigma_{SM_k}}{\sigma_{SR_k}} \tag{A8}$$

Temporal correlation of interannual variability (*TCOIAV*)

$$TCOIAV_k = \frac{1}{I-1}\sum_{i=1}^{I}\frac{(\hat{M}_{ki} - \hat{\overline{M}}_k)\cdot(\hat{R}_{ki} - \hat{\overline{R}}_k)}{\sigma_{TM_k}\cdot\sigma_{TRk}} \tag{A9}$$

with the temporal variances

$$\sigma^2_{TM_k} = \frac{1}{I-1}\sum_{i=1}^{I}(\hat{M}_{ki} - \hat{\overline{M}}_k)^2 \quad \text{and} \quad \sigma^2_{TR_k} = \frac{1}{I-1}\sum_{i=1}^{I}(\hat{R}_{ki} - \hat{\overline{R}}_k)^2 \tag{A10}$$

Ratio of interannual variability (*RIAV*)

$$RIAV_k = \frac{\sigma_{TM_k}}{\sigma_{TR_k}} \tag{A11}$$

Climatological rank correlation (*CRCO*)

$$CRCO = 1 - \frac{6}{12 \cdot (12^2 - 1)} \sum_{k=1}^{12} (Rank\ \hat{\overline{M}}_k - Rank\ \hat{\overline{R}}_k)^2 \qquad \text{(A12)}$$

Ratio of yearly amplitudes (*ROYA*)

$$ROYA = \frac{\max(\hat{\overline{M}}_k) - \min(\hat{\overline{M}}_k)}{\max(\hat{\overline{R}}_k) - \min(\hat{\overline{R}}_k)} \quad for\ \ k = 1,...,12 \qquad \text{(A13)}$$

Appendix B: Evaluation for sub-domains AL, BI, FR and MD

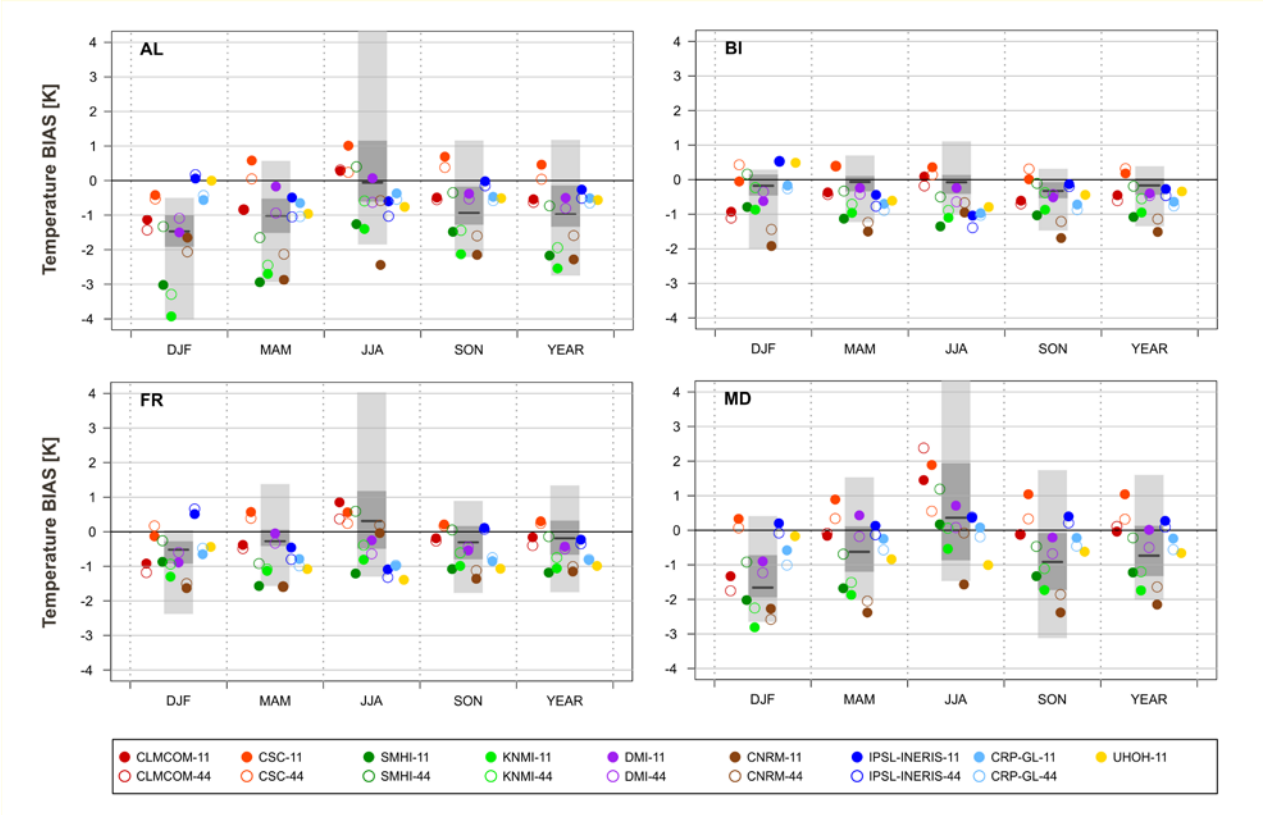


Figure B1. Mean seasonal and annual temperature bias (BIAS) [K] for the EUR-11 (filled circles) and the EUR-44 ensemble (open circles) and for sub-domains AL, BI, FR and MD (see Figure 5 for sub-domains EA, IP, ME and SC). The gray bars denote the BIAS range of the ENS-22 ensemble: entire range in light gray, inter-quartile range (corresponding to eight models) in dark gray and median as solid line.

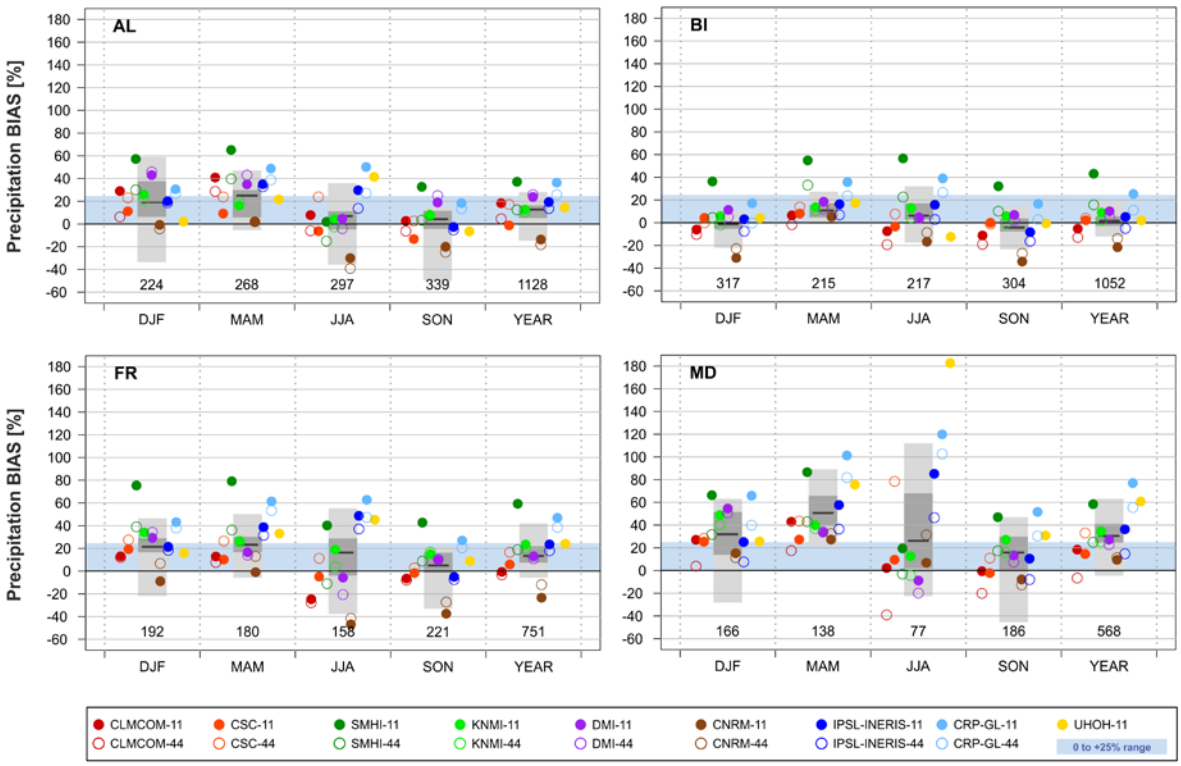
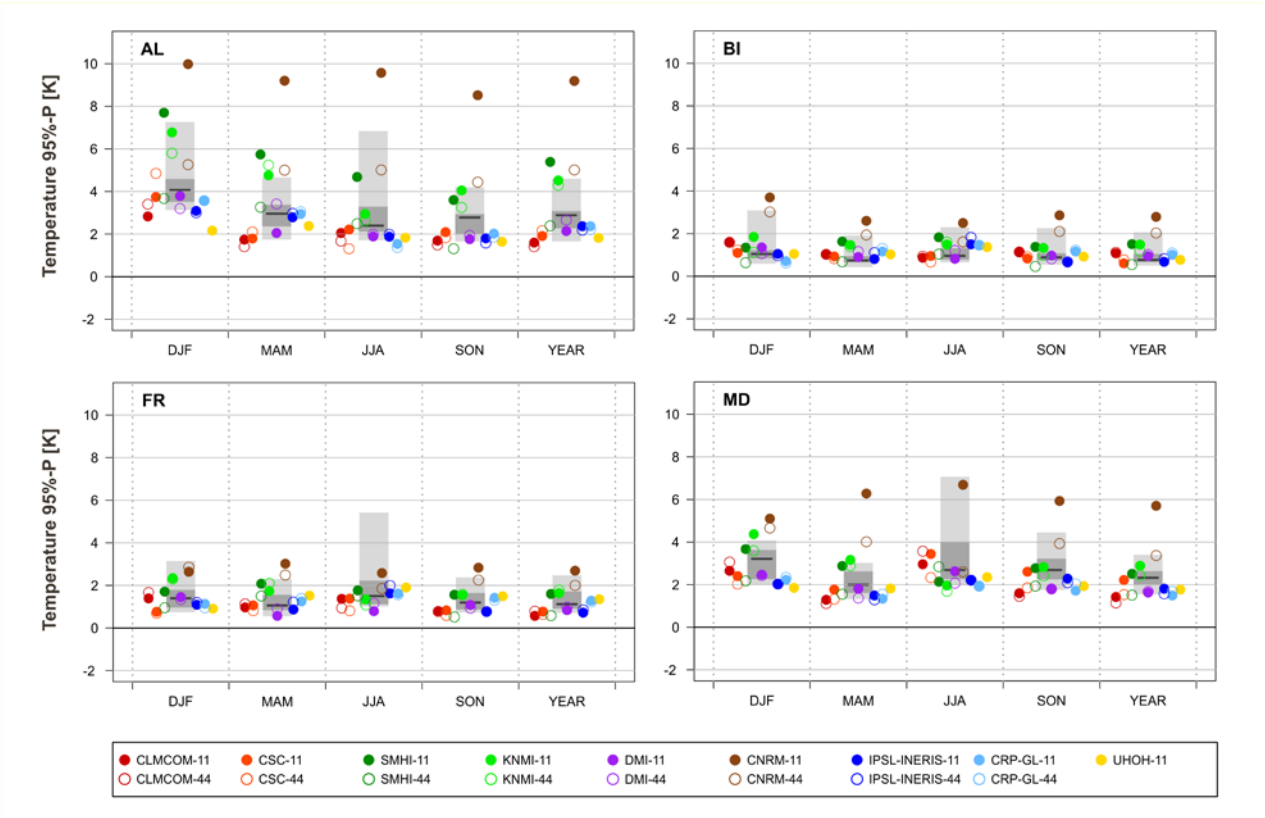


Figure B2. As Figure B1 but for the mean seasonal and annual relative precipitation bias (BIAS) [%]. The numbers along the x-axis indicate mean seasonal [mm/season] and mean annual [mm/year] precipitation sums for the period 1989-2008 in the E-OBS reference. The blue shading indicates a bias range between 0 and +25%, corresponding to acceptable model biases in case of a systematic rain gauge undercatch of up to 20% of true precipitation. See Figure 6 for sub-domains EA, IP, ME and SC.



1

2 Figure B3. 95th percentile of mean seasonal and annual absolute temperature biases (95%-P) [K]
3 for the EUR-11 (filled circles) and the EUR-44 ensemble (open circles) and for sub-domains AL,
4 BI, FR and MD (see Figure 7 for sub-domains EA, IP, ME and SC). The gray bars denote the
5 95%-P range of the ENS-22 ensemble: entire range in light gray, inter-quartile range
6 (corresponding to eight models) in dark gray and median as solid line.
7

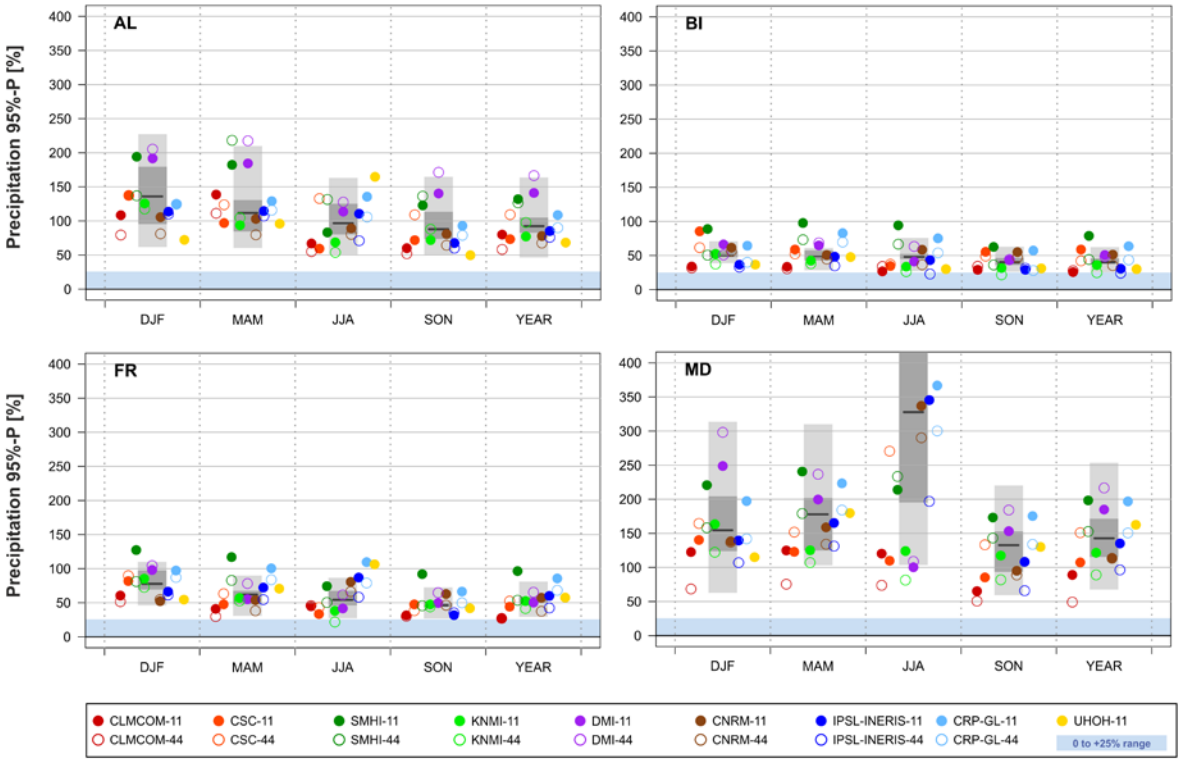


Figure B4. As Figure B3 but for the 95th percentile of mean seasonal and annual relative precipitation biases (absolute values; 95%-P) [%]. The blue shading indicates a bias range between 0 and +25%, corresponding to acceptable wet model biases in case of a systematic rain gauge undercatch of up to 20% of true precipitation. See Figure 8 for sub-domains EA, IP, ME and SC.

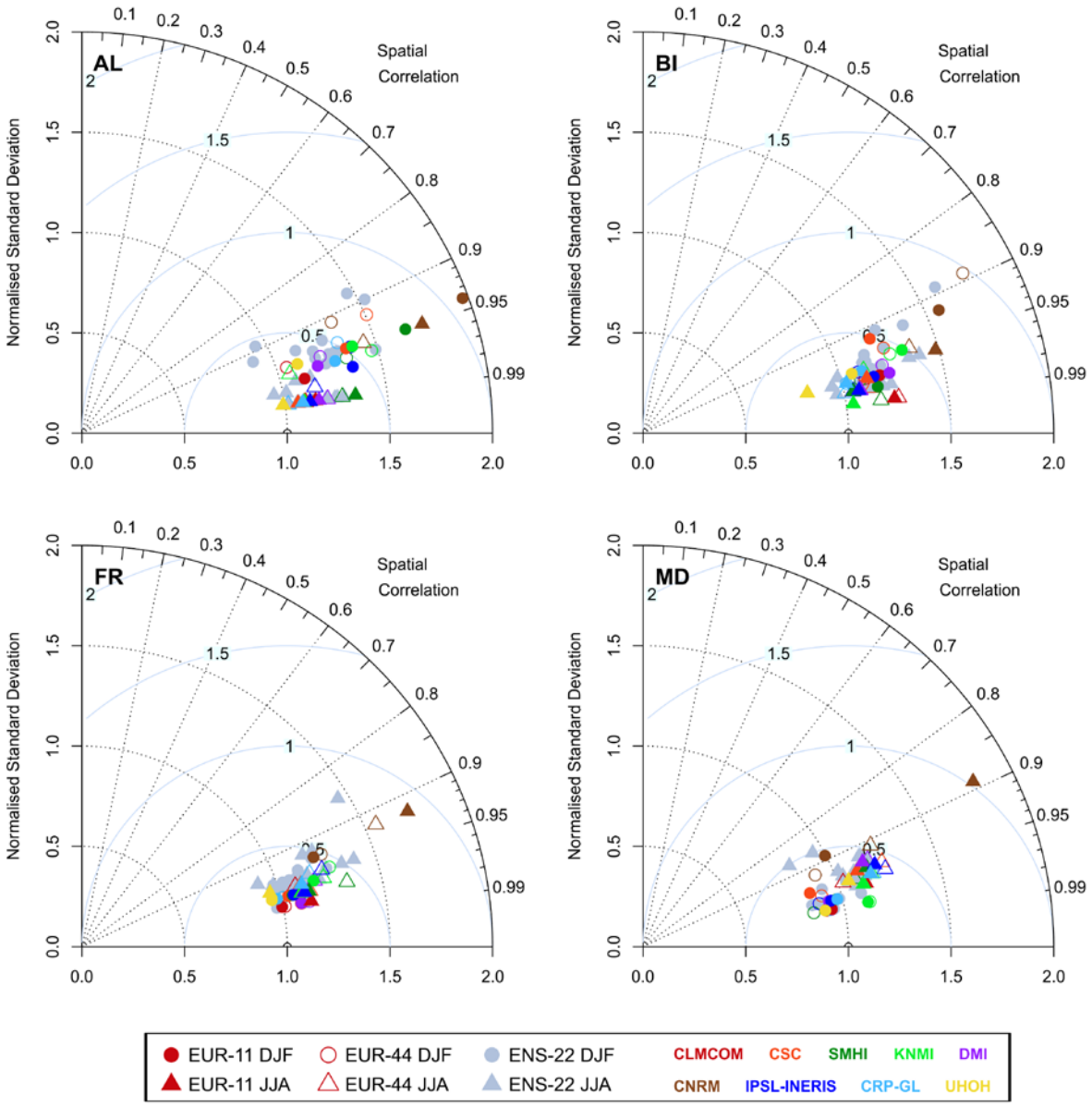


Figure B5. Spatial Taylor diagrams exploring the model performance with respect to the spatial variability of mean winter (circles) and mean summer (triangles) temperature within sub-domains AL, BI, FR and MD (see Figure 9 for sub-domains EA, IP, ME and SC). Filled markers: EUR-11 ensemble, non-filled markers: EUR-44 ensemble, gray markers: ENS-22 ensemble. The diagrams combine the spatial pattern correlation (PACO, cos(azimuth angle)) and the ratio of spatial variability (RSV, radius). The distance from the 1-1 location corresponds to the normalized and centered root-mean-square difference (which does not take into account the mean model bias), expressed as multiples of the observed standard deviation. Note the different number of underlying grid cells per sub-domain in the individual ensembles.

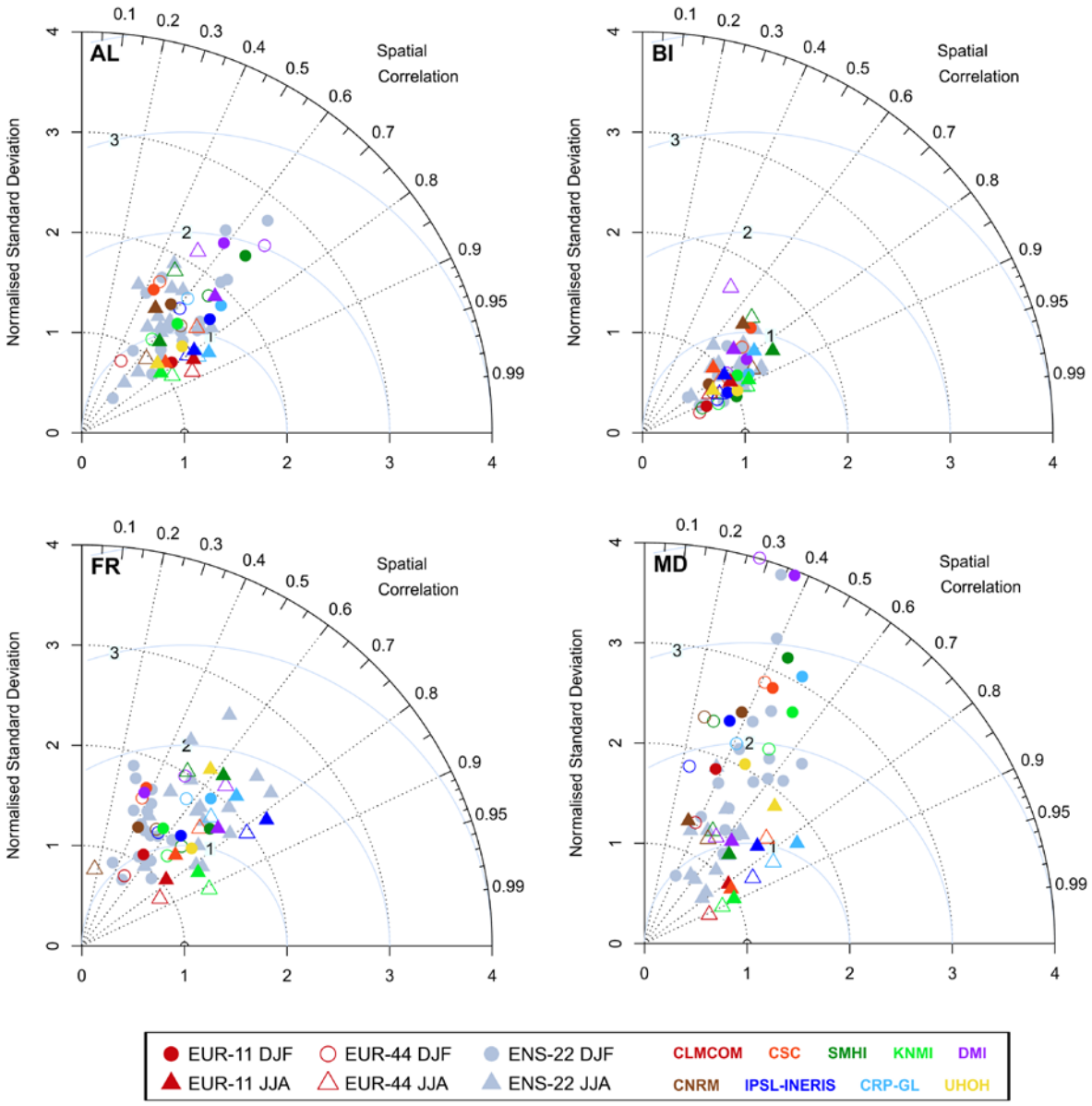


Figure B6. As Figure B5 but for mean winter (circles) and mean summer (triangles) precipitation. See Figure 10 for sub-domains EA, IP, ME and SC.

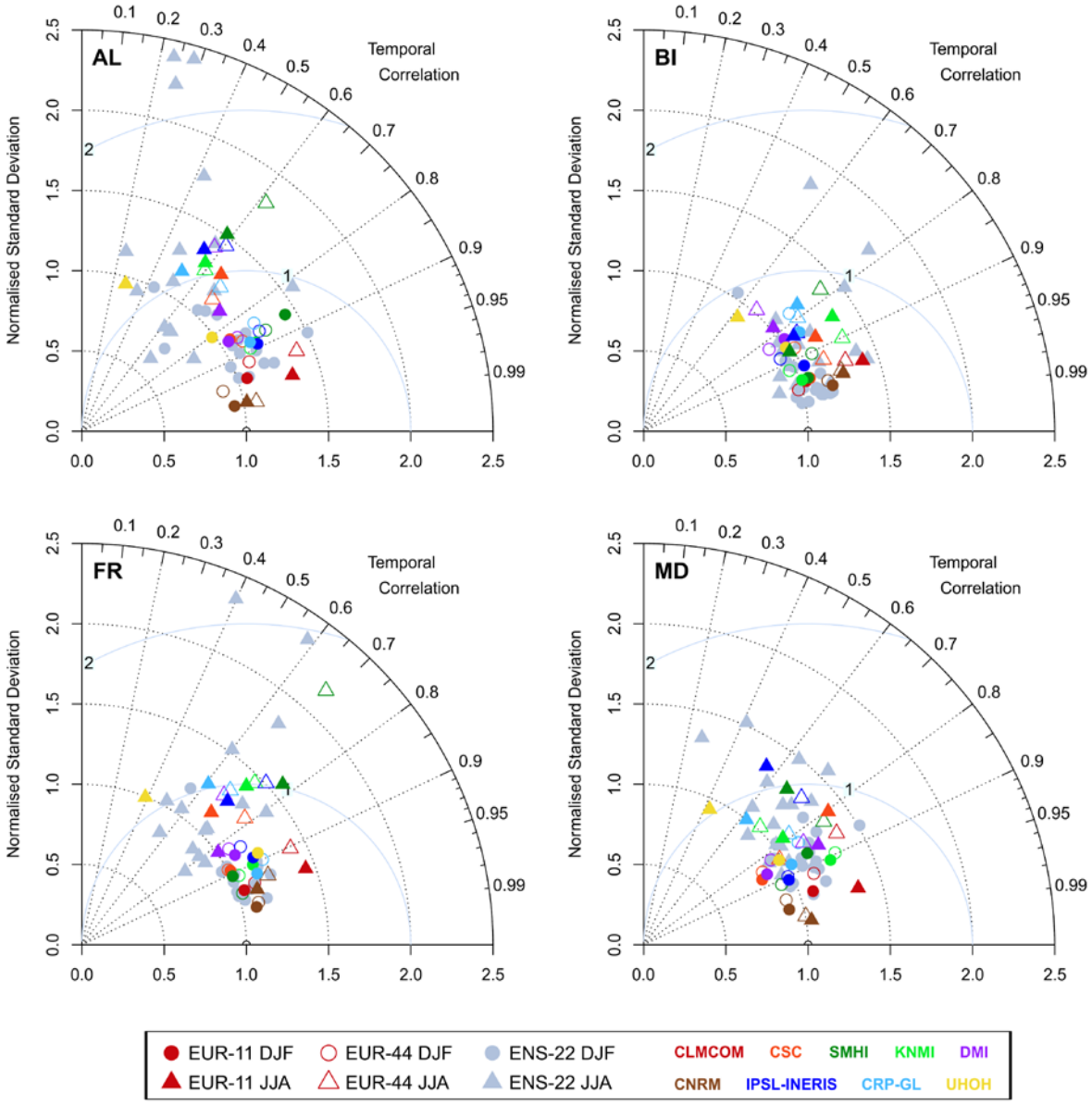


Figure B7. Temporal Taylor diagrams exploring the model performance with respect to the inter-annual temporal variability of mean winter (circles) and mean summer (triangles) temperature as averages over sub-domains AL, BI, FR and MD (see Figure 11 for sub-domains EA, IP, ME and SC). Filled markers: EUR-11 ensemble, non-filled markers: EUR-44 ensemble, gray markers: ENS-22 ensemble. The diagrams combine the temporal correlation of interannual variability (TCOIAV, cos(azimuth angle)) and ratio of interannual variability (RIAV, radius). The distance from the 1-1 location corresponds to the normalized and centered root-mean-square difference (which does not take into account the mean model bias), expressed as multiples of the observed standard deviation.

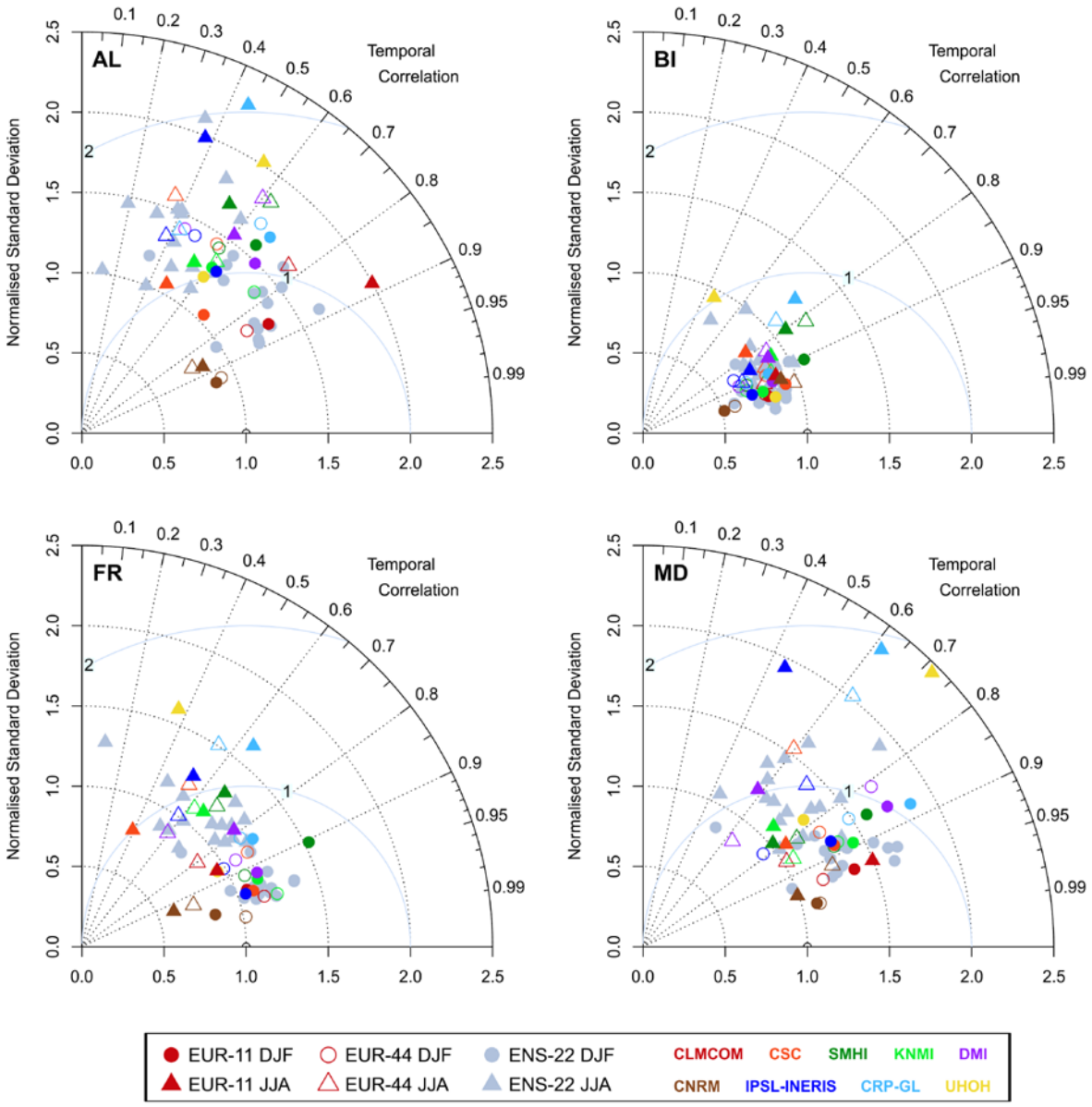


Figure B8. As Figure B7 but for mean winter (circles) and mean summer (triangles) precipitation. See Figure 12 for sub-domains EA, IP, ME and SC.