



Robustness of European climate projections from dynamical downscaling

Jens Hesselbjerg Christensen^{1,2,3} · Morten A. D. Larsen⁴ · Ole B. Christensen³ · Martin Drews⁴ · Martin Stendel³

Received: 23 October 2018 / Accepted: 24 May 2019 / Published online: 1 June 2019
© The Author(s) 2019

Abstract

How climate change will unfold in the years to come is a central topic in today's environmental debate, in particular at the regional level. While projections using large ensembles of global climate models consistently indicate a future decrease in summer precipitation over southern Europe and an increase over northern Europe, individual models substantially modulate these distinct signals of change in precipitation. So far model improvements and higher resolution from regional downscaling have not been seen as able to resolve these disagreements. In this paper we assess whether 2 decades of investments in large ensembles of downscaling experiments with regional climate model simulations for Europe have contributed to a more robust model assessment of the future climate at a range of geographical scales. We study climate change projections of European seasonal temperature and precipitation using an ensemble-suite comprised by all readily available pan-European regional model projections for the twenty-first-century, representing increasing model resolution from ~ 50 to ~ 12 km grid distance, as well as lateral boundary and sea surface temperature conditions from a variety of global model simulations. Employing a simple scaling with global mean temperature change we identify emerging robust signals of future seasonal temperature and precipitation changes also found to resemble current observed trends, where these are judged to be statistically significant.

Keywords Regional climate model ensembles · PRUDENCE · ENSEMBLES · CORDEX · Pattern scaling · Climate projections

1 Introduction

For more than 2 decades, coordinated efforts of applying regional climate models (RCMs) to downscale global climate model (GCM) simulations for Europe have been pursued by an ever-increasing group of scientists (Rummukainen et al. 2015; Rummukainen 2016). This endeavour showed its first results during framework projects supported by the European Union and sister projects in North America all initiated in the 1990s (Christensen et al. 1997; Machenhauer et al. 1998; Takle et al. 1999; Hagemann et al. 2001). Here, the foundation for today's advanced World Climate Research Programme (WCRP) initiative COordinated Regional Downscaling EXperiment (CORDEX) (Giorgi and Gutowski 2015; Gutowski et al. 2016) was laid out, as the first ensembles of coordinated RCM simulations aiming to assess future regional climate change emerged. It was already realized, however, at this early stage that systematic model biases in GCMs as well as RCMs made this task very challenging (Christensen et al. 1997, 2001; Pan

✉ Jens Hesselbjerg Christensen
hesselbjerg@nbi.ku.dk

Morten A. D. Larsen
madla@dtu.dk

Ole B. Christensen
obc@dmi.dk

Martin Drews
mard@dtu.dk

Martin Stendel
mas@dmi.dk

¹ University of Copenhagen, Juliane Maries Vej 30, 2100 Copenhagen Ø, Denmark

² NORCE, Norwegian Research Centre AS, Bergen, Norway

³ Danish Meteorological Institute, Lyngbyvej 100, 2100 Copenhagen Ø, Denmark

⁴ Technical University of Denmark, Produktionstorvet, Building 424, 2800 Kgs. Lyngby, Denmark

et al. 2001; Schiermeier 2004). As an immediate outcome, the idea was therefore conceived to undertake even more concerted efforts by constructing even more well-defined and structured sets of common simulations; this led to the PRUDENCE project (Christensen et al. 2002; Christensen and Christensen 2007) (2001–2004). Additional coordinated efforts involving an increased number of GCMs and RCMs then followed in the ENSEMBLES project (Van der Linden and Mitchell 2009; Christensen et al. 2010) (2004–2009) and continue in the ongoing Euro- and Med-CORDEX initiatives (Jacob et al. 2014; Ruti et al. 2015) (2011–present). Meanwhile, along with the overall coordination, model resolutions have increased from a grid point distance of about 50 km (PRUDENCE) to 12 km (Euro-CORDEX11) and from time slice simulations covering 30 years (PRUDENCE) to transient experiments representing the time span of 1951–2100 (ENSEMBLES and CORDEX); from two, but one dominating, driving GCMs and the SRES (Nakicenovic et al. 2000) A2 and B2 emission scenarios (PRUDENCE) to several GCMs (Euro-CORDEX) and multiple RCP (Meinshausen et al. 2011) scenarios. So far, this wealth of simulations has mainly been used to provide a measure of baseline change according to a particular emission scenario, or relating to the passage of global mean temperature thresholds, e.g., 2 °C (Vautard et al. 2014). This has typically been defined by a multi-model mean change with associated uncertainties estimated from model spread within the specific ensemble (Christensen and Christensen 2007; Deque et al. 2007; Van der Linden and Mitchell 2009; Jacob et al. 2014) and has resulted in largely incomparable projections that only leave room for relatively simple statements about the future climate conditions in Europe. Only in a few cases a comparison between some parts of the modelling suites have been attempted (Vautard et al. 2014; Rajczak and Schär 2017; Fernández et al. 2019), but never across the entire suite.

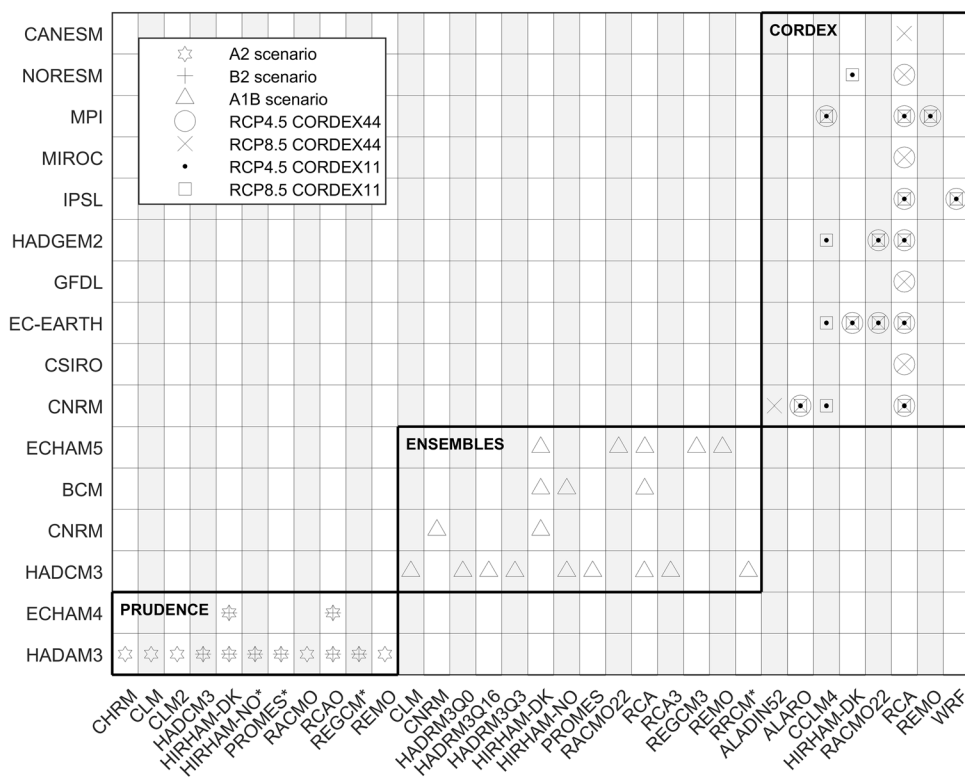
Climate projections such as those derived from the above-mentioned efforts are widely used to explore projected *impacts* of future climate change. Specifically, for use in risk-based impact analyses there is an increasing demand for high-resolution probabilistic climate change information at the regional scale based on such multi-model approaches (e.g., Jacob et al. 2014). The interest in achieving robust projections is shared by different scientific communities as well as by practitioners and other stakeholders striving to identify means and measures to construct robust estimates of future changes, their impacts and consequences at the relevant scales, where measures to adapt to these changes need to be taken.

The ability to simulate regional climate realistically is a formidable challenge and while improvements to do so have been steady they have also been very slow (Collins et al. 2013; Christensen et al. 2013). As a result, the full plausible range of climate change for any given scenario cannot

in practice be assessed using the presently available model information (Hawkins and Sutton 2009), such as that defined by CORDEX, nor even when including also the extensive coarser-resolution and larger-scale Coupled Model Inter-comparison Project Phase 5 (CMIP5) (Taylor et al. 2011) database. Consequently, pattern scaling (i.e., simple scaling of model-mean changes of temperature and precipitation patterns with global mean temperature change) has long been used to generalize climate change information beyond the information available from individual climate models (Santer et al. 1990; Tebaldi and Arblaster 2014; Matte et al. 2019). This includes the Fifth Assessment Report from the Intergovernmental Panel on Climate Change (IPCC AR5) (Collins et al. 2013; IPCC 2013), where this approach was used to directly compare CMIP3 (Meehl et al. 2007) and CMIP5 annual multi-model mean temperature and precipitation change signals. Here it was demonstrated that the many fundamental differences between the model set-ups in CMIP3 and CMIP5 made the ensembles different in a statistical sense, even though visually the annual multi-model mean changes are largely indistinguishable in the two model compilations (Collins et al. 2013).

In this paper we investigate the projected regional climate change scenarios for Europe across PRUDENCE, ENSEMBLES and Euro-CORDEX, the latter at two different model resolutions (P–E–C, hereafter). Using a methodology based on the abovementioned pattern scaling to circumvent the many differences that exist between the three generations of multi-model ensemble approaches within the P–E–C experimental sequence (see Fig. 1), we systematically reassess their results across different emission scenarios, GCM/RCM combinations, model resolution, project and model versions and improvements. The added value of regional climate models in terms of increased resolution over coarser GCMs, e.g., in terms of an improved representation of relevant climate processes at the regional to local scales, is well documented. Here we assess to which degree model improvements and higher resolution from regional downscaling over time have been able to resolve apparent inconsistencies in climate change projections for Europe, e.g., the opposite trends in precipitation change inferred from individual RCMs for certain areas and seasons. Specifically, we consider the projected temperature and (relative) precipitation changes on a seasonal scale, i.e., model mean regional climate change signals for summer (June–July–August; JJA) and winter (December–January–February; DJF) for all the available P–E–C simulations using the original model resolution, followed by a scaling in relation to the global average warming of the driving global model. The scaled projections are also compared with observed trends from the middle of the last century to the present, a period over which the global mean temperature rose close to 1 °C. The overarching objective is to identify components of the model projections that

Fig. 1 GCM/RCM combinations in the PRUDENCE, ENSEMBLES, CORDEX11 and CORDEX44 experimental sequence. RCMs are sorted alphabetically within each model family. Models omitted in the EOF analysis (see text) are marked with an asterisk (*). The total number of individual simulations studied are 15 (PRUDENCE), 18 (ENSEMBLES), 34 (CORDEX44), and 32 (CORDEX11)



may possibly be considered robust and in line with already experienced change (e.g., Pal et al. 2004). This comparison is not an attempt to assess to what extent the observed changes are anthropogenic in origin, but to demonstrate whether or not the projected changes found using the P–E–C suite of models resemble current long-term observed trends when these are robust and statistically significant.

2 Methods

The three research projects in the P–E–C sequence span a period of nearly 20 years of European regional climate modelling developments. As a result, each of them is constrained in different ways (e.g., by computing resources). This reveals itself in the way the various experiments were conducted. PRUDENCE was initially set up to only use one specific set of boundary conditions for all RCMs, from the Hadley Centre HadAM3H (Buonomo et al. 2007) model. This atmosphere-only GCM was set up in comparatively high resolution at the time, 150 km grid distance, for two time slices: a present day time slice 1961–1990 with observed sea surface temperature (SST) and sea ice, and a future time slice 2071–2100 with forcing from SRES A2, while taking the SST and sea ice climate change signal from a HadCM3 (Johns et al. 2003) global coupled simulation and added to the fields of the observed present-day period. Corresponding simulations following the (weaker) SRES B2 scenario

were also performed. Regional simulations were set up with a common integration area of 50 km resolution. During the PRUDENCE project the HadAM3H-driven simulations were supplemented by a few simulations forced by the ECHAM4/OPYC (Roeckner et al. 1996) coupled global model, also following the SRES A2 and B2 scenario (see also Christensen et al. 2001). Several PRUDENCE simulations used somewhat different configurations resulting in different areas than the standard due to specific considerations at the respective institutions, who carried out the simulations.

The ENSEMBLES experiment was set up to include several sets of GCM boundary data and ended up using 9 different coupled GCM simulations for simulations reaching 2100, all following the SRES A1B scenario. Regional simulations were transient, covering the full integration period of 1951–2100, and used a common integration area with a 25 km grid distance. Further, in this project, the exact area configuration was fixed except for a few necessary exceptions due to specificities of the map projection used by certain regional climate models.

CORDEX is still ongoing, and new simulations are added continually. Two different resolutions are used for Europe, 0.44° and 0.11°, corresponding to around 50 km and 12 km grid separation, respectively, created by aggregating or disaggregating the ENSEMBLES domain. Scenarios are now based on the Representative Concentration Pathways (RCPs; van Vuuren et al. 2011), and many simulations exist with RCP2.6, RCP4.5 and RCP8.5. Only simulations from the

latter two are included in this study, in parts due to the relatively limited number of available simulations for RCP2.6 compared to the other two. Moreover, since the RCP2.6 simulations do not consistently project a global warming signal of more than 1 °C within the twenty-first-century, we want to avoid extrapolation when applying a pattern scaling approach (detailed in the following section).

2.1 Climate scenarios and pattern scaling

The pattern scaling approach we use to bring together different members of the PRUDENCE, ENSEMBLES and CORDEX ensembles is similar to the one adopted by IPCC AR5. It allows us to scale the geographical patterns of projected changes in temperature and precipitation to 1 °C of global warming for all simulations and thus to facilitate their detailed comparison. Specifically, we use the projected global mean temperature change in 2100 by the driving GCM for each of the individual RCMs and climate scenario (PRUDENCE: A2 and B2; ENSEMBLES: A1B; CORDEX: RCP4.5 and RCP8.5) as a general scaling parameter. The intensity of the simulated regional change is scaled with respect to the modelled annual mean global averaged temperature change (deduced from the driving GCM) at each grid point, which allows for identifying similarities and differences in the simulated spatial patterns across emission scenarios, GCM/RCM combinations, model resolution and model generation.

2.2 Uncertainty metrics and robustness definitions

To assert the estimation of temperature and precipitation change and the corresponding robustness for projected climate scenarios across the P–E–C suite, every available model simulation within these experiments should be used. Historical and projected runs in the suite include the period 1961–1990 and 2071–2100 for PRUDENCE whereas corresponding periods are 1981–2000 and 2081–2100 for ENSEMBLES and 1986–2005 and 2081–2100 for Euro-CORDEX. The climate change scenarios used include SRES A2/B2 (pre-CMIP3) for PRUDENCE, SRES A1B (CMIP3) for ENSEMBLES and RCP4.5/RCP8.5 (CMIP5) for Euro-CORDEX. Model resolutions are 50 km and 25 km for PRUDENCE and ENSEMBLES, respectively, whereas Euro-CORDEX simulations comes in both a 50 km (Euro-CORDEX44) and 12 km (Euro-CORDEX11) resolution, resulting in four different lines of model suites to be compared.

Within each model suite, the individual model mean difference between the periods mentioned above is calculated per season for DJF and JJA and subsequently scaled with the corresponding globally averaged annual mean temperature anomaly for the driving GCM and scenario in concern. For ENSEMBLES and CORDEX, these are available from CMIP3 and CMIP5, while the necessary data for PRUDENCE were collected in Christensen et al. (2001). After scaling, we infer the multi-model mean change(s). The colouring in Figs. 2 and 3 illustrates the result of this exercise,

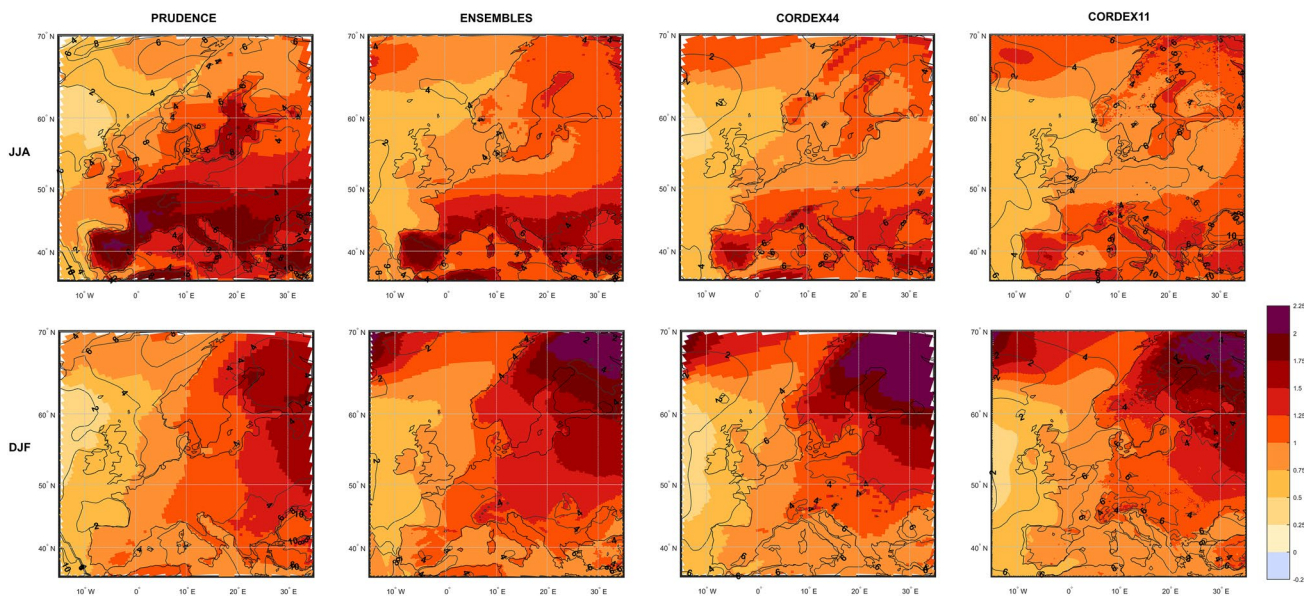


Fig. 2 Seasonal mean scaled temperature change [°C] for PRUDENCE, ENSEMBLES, Euro-CORDEX44 and Euro-CORDEX11 for the JJA and DJF seasons. The contour lines show the ratio between the multi-model mean and the multi-model standard deviation

(S/N; see Methods). Hatching indicates areas with $S/N < 1$. As indicated in the figure S/N is everywhere larger than one (for comparison see Fig. 3)

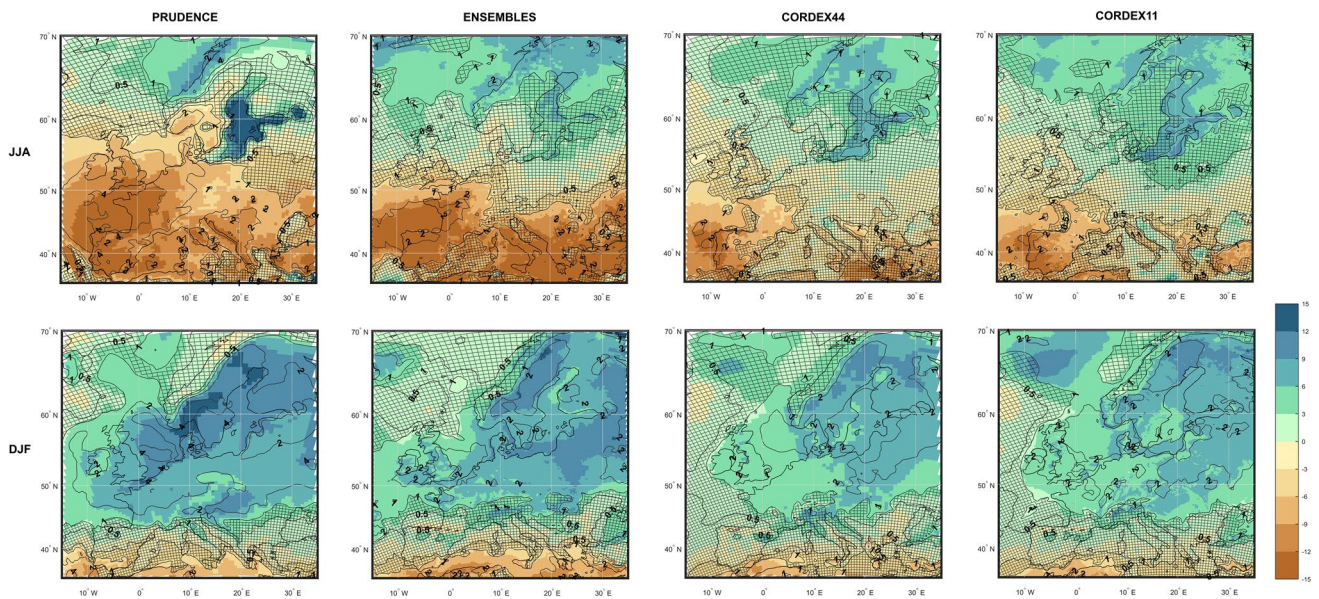


Fig. 3 Seasonal mean scaled relative precipitation change [%] for PRUDENCE, ENSEMBLES, Euro-CORDEX44 and Euro-CORDEX11 for the JJA and DJF seasons. As in Fig. 2 the contour lines show the S/N ratio, while hatching indicates areas where $S/N < 1$

adopting the same colour map and scale as used in IPCC AR5. Results are presented in the original model resolution corresponding to the particular part of the P–E–C model suite, e.g., 50 km for the PRUDENCE simulations. As a measure of robustness or model agreement, the inter-model standard deviation, which we hereafter refer to as ‘noise’ (N), within each model suite is calculated to overlay the plotted change ‘signal’ (S) with the corresponding S/N ratio plotted as contours.

For an ensemble of independent RCMs with a long-term, mean seasonal average temperature and precipitation of similar magnitude, the model-mean state, as here represented by S, has been demonstrated to be a representative measure of the projected future climate (Collins et al. 2013). Along the same lines, the model spread, here represented by N, is often used to represent the ensemble robustness in terms of specifying a range for the projected future climate. In Figs. 2, 3 and 6, areas with $S/N < 1$ are depicted with hatching, emphasizing areas of model disagreement. Conversely, regions without hatching thereby exhibit a certain level of robustness. This follows the approach by Laux et al. (2017) and Matte et al. (2019) and to some extent IPCC AR5 (IPCC 2013).

2.3 EOF analysis

To confirm potential changes in spatial patterns of change between the model suites, we use an Empirical Orthogonal Function (EOF) analysis (Lorenz 1956), based on the same models as are used to construct the mean climate change signal and for the S/N analysis. The analysis was carried

out using the implementation in MATLAB (MATLAB and Statistics Toolbox Release 2017a, The MathWorks, Inc., Natick, Massachusetts, United States). In order to perform this analysis, the results of all simulations in P–E–C were interpolated to the 50 km resolution adopted in PRUDENCE and Euro-CORDEX44. To ensure a suitable spatial extent of the EOF analysis, six PRUDENCE simulations (three models—scenario A2/B2) and one ENSEMBLES simulation however had to be omitted, since these unfortunately exclude larger parts of the European domain. In the following distributed plots of model EOFs, Principal Components (PCs) and variabilities will be used to support inter-comparison of the four lines of models discussed above.

2.4 Comparison to observed trends

The similarities between the P–E–C simulations has been compared to current observed trends. For demonstration we use the CRU TS 4.01 dataset (Harris et al. 2014), which covers all land areas except Antarctica with a resolution of 0.5° for the period 1901–2016. However, as pointed out in IPCC AR5 (Bindoff et al. 2013) only the latter half of the twentieth-century is clearly influenced by anthropogenic climate change at the regional scale. We thus restrict ourselves to the period 1950–2016. For these six-and-a-half decades, we calculate the linear trend scaled with the corresponding global mean temperature trend, its statistical significance together with the smallest and largest linear trends that consistently fit the observations within the error margins for each grid cell. This is done as follows: For the period 1950–2016, a least square fit has been applied for each land grid point (as the

CRU TS dataset is only defined over land). For temperature the best fit and the smallest and largest trends within the error margins are then calculated accordingly. The scaling with the global temperature trend is expressed as the quotient of the local and the globally averaged annual mean temperature trend for the same period. Thus, a value of 2 implies that the local trend is twice as large as the global one. For precipitation a similar approach is applied but it is here calculated as percent per 1 °C global warming: from the regression curve, we calculate the percentage change with respect to the first year; then the scaling with global temperature change is done as for temperature. To estimate the confidence interval for the slope of the regression line, the standard error (SE) of the distribution of the slope needs to be known. Assuming a normal distribution, SE can be expressed by the following equation

$$SE = \sqrt{\frac{(y_i - \hat{y}_i)^2}{n - 2}} / \sqrt{\sum(x_i - \bar{x})^2}$$

where y_i is the value of the dependent variable for observation i , \hat{y}_i is the estimated value of the dependent variable for observation i , x_i is the observed value of the independent variable for observation i , \bar{x} is the mean of the independent variable, n is the number of observations, and $n - 2$ is the number of the degrees of freedom. Significance levels are calculated with a p value of 0.05; for temperature and precipitation, we applied a one-sided, respectively, a two-sided Student t test.

In order to go beyond merely visual comparison of similarity, we calculate the ranking of the observational trend pattern among the simulation ensemble of climate change fields in terms of pattern correlation.

3 Results

3.1 Seasonal temperature and precipitation change

Figures 2 and 3 compare the projected seasonal multi-model mean JJA and DJF changes in temperature [°C] and precipitation [%] per degree of global annual mean temperature change across the P–E–C sequence. Broadly spoken, there is an overall agreement of the patterns of change in either variable and for both seasons. While the projected warming signal in the PRUDENCE simulations stands out as compared to the others and is considerably stronger during JJA, there is a better general agreement in the projected warming pattern in DJF. The excessive JJA warming signal in PRUDENCE has previously been attributed partly to systematic errors in the surface schemes in the older models (Christensen et al. 1997; Hirschi et al. 2011; Boberg and Christensen 2012),

but is also related to the strong dependence on the single dominating driving GCM. The projected change patterns in mean precipitation are quite similar in DJF, while there is a much stronger drying signal in PRUDENCE during JJA. This can be largely explained by the same systematic errors responsible for the excessive Mediterranean JJA warming (Seneviratne et al. 2006).

Figures 2 and 3 show that the amplitude of the multi-model mean climate change signal for both temperature and precipitation is mostly reduced along the P–E–C sequence. The temperature response over Europe in JJA as well as in DJF is everywhere large enough to ensure a signal-to-noise (S/N) ratio well above 1 (and generally higher in Euro-CORDEX11 than in PRUDENCE), as is documented by Fig. 2. Relative changes in precipitation are much more complicated to interpret, primarily due to the fact that wherever modest changes are projected in the multi-model mean, even a numerically small inter-model spread will result in a low S/N ratio. Roughly speaking, across the P–E–C suite of models, any projected change within $\pm 5\%$ has a S/N ratio below 1, while numerically larger changes appear more robust. The S/N ratio for precipitation in DJF is well above 1 except for a relatively narrow transition zone in southern Europe separating a Mediterranean drying from a central and northern European moistening. This signal is basically unaltered along the P–E–C sequence (Fig. 3). For JJA the reduced amplitude in the climate change signal along the P–E–C sequence results in an expansion of the transition zone between drying and moistening, suggesting that the only really robust climate change signal related to precipitation is the drying over the Iberian Peninsula and central parts of Turkey and a moistening in the northernmost parts of Scandinavia. This is also emphasized by the gradual reduction in amplitude of the climate change signal along the P–E–C sequence. We note that the relatively large transition zone in JJA with no statistically significant projected precipitation change may be seen as consistent with “no change”. We further note that the areas here identified with an $S/N < 1$ seem to closely resemble the regions identified in IPCC AR5 (IPCC 2013) to exhibit a natural variability exceeding the climate change signal.

3.2 EOF analyses

In order to test the robustness of the projected changes in temperature and precipitation, we perform an EOF analysis on the full set of scenarios from the P–E–C sequence. Figure 4 displays the loading on the first three principal components (PCs) for each of the RCM/GCM combinations shown in the same P–E–C sequence (in arbitrary units). Importantly, a few models can be identified as having a substantially anomalous loading on the PCs. Assuming that any identifiable outlier model could be responsible

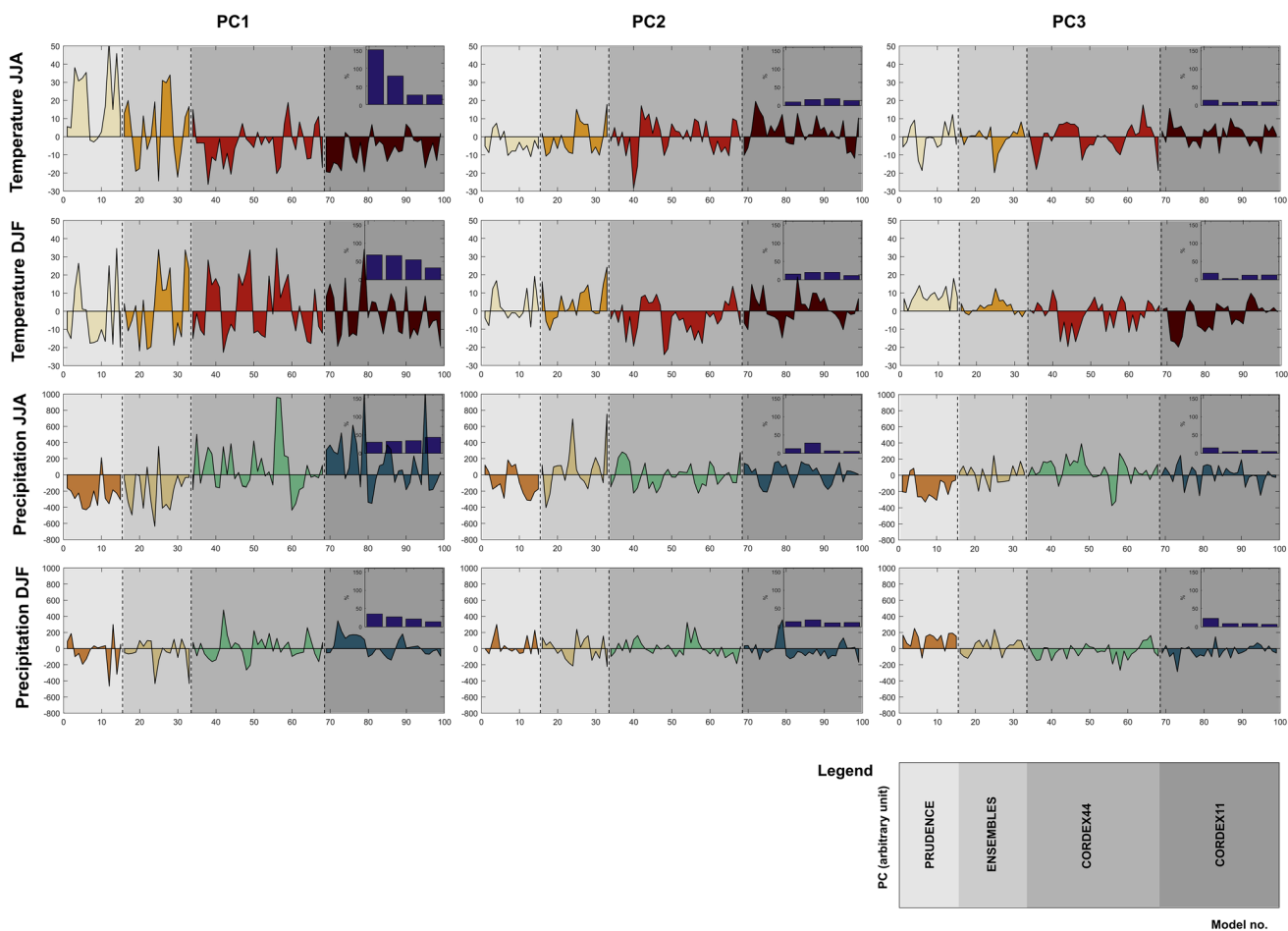


Fig. 4 Values (arbitrary units) of the first three principal components (PC1 to PC3) relating to individual models from the P–E–C multi-model ensembles for JJA temperature, DJF temperature, JJA precipitation and DJF precipitation. The grey shading (see the legend) dis-

tinguished the model sub-samples, whereas inserts show the variance within each sample normalized by the variance for the entire ensemble

for a poorer S/N ratio and thus significantly affect the perceived robustness of the regional climate change signals, we redid the EOF analysis leaving out those model runs that are identified as ‘outliers’ in Fig. 4. We found, however, the effect to be small with regards to both the multi-model mean and the S/N ratio for the full P–E–C sequence as well as for each of the four sub-samples (not shown). The small and in practical terms negligible effect of removing single outliers clearly reflects that the ensemble size proves to be large enough not to be strongly influenced by outliers. This suggests that the multi-model ensemble that is represented

by Euro-CORDEX11 yields a fairly robust estimate of the future European climate, i.e., scaled with global mean temperature change corresponding to RCP4.5 and possibly RCP8.5 but not beyond (i.e., any forcing stronger than RCP8.5 or beyond the twenty-first-century). This limitation is to avoid extrapolation rather than interpolation of the multi-model ensemble mean change.

We calculate the total variance for the investigated fields and seasons for each project individually (cf. Table 1). For each field a tendency towards higher inter-model agreement within each project along the P–E–C sequence is shown with

Table 1 Variance of PCs within each project (arbitrary units)

	PRUDENCE	ENSEMBLES	CORDEX44	CORDEX11
JJA Temperature	487	559	313	232
DJF Temperature	469	521	507	344
JJA Precipitation	164,565	266,718	260,717	278,264
DJF Precipitation	91,973	102,189	92,609	81,868

the exception of JJA precipitation; conversely, DJF precipitation shows very little change. Further, projection on individual EOFs (not shown) reveals that the observed increase in the variability of JJA precipitation rests solely on the first and dominating EOF, which is a generally homogeneous pattern across most of Europe. Thus, we find a large variability in the sensitivity of JJA precipitation in the climate projection for both CORDEX and ENSEMBLES as compared to PRUDENCE. This we ascribe to the already mentioned fact that PRUDENCE is dominated by one specific GCM simulation.

For each of the three dominating EOFs we have further investigated the corresponding PCs for each of the subensembles (PRUDENCE, ENSEMBLES, Euro-CORDEX44 and Euro-CORDEX11) separately. The ensemble variance can be split into the combined variance within each of the model ensembles and the variance among the individual model means, weighted by the number of simulations in each ensemble. The ratio of these two indicates the degree to which the ensembles are different with respect to the EOF in question. The values after proper normalization (explained in the caption of Table 2) can be seen in Table 2. Note that this estimation is highly approximate, as we are here treating the PCs of model results as independent, normally and identically distributed quantities (iid). For JJA temperature, PC1 is evidently the pattern with the largest difference between ensembles of all fields investigated.

Figure 5 shows EOFs for the fields referred to in Table 2. Signs of PCs and EOFs have been chosen to make the majority of grid points in EOF maps positive. Interpretations of PCs and EOFs from Figs. 4 and 5 for the individual fields/seasons follow below.

The first EOF pattern for JJA temperature shows a much increased warming in southern Central Europe, particularly southern France compared to the other ensembles. In Fig. 4 it is seen that the PC loadings for the PRUDENCE simulations is much more positive than the other simulations (cf. Table 1). The non-homogeneous distribution of PC values

among PRUDENCE simulations seen in Fig. 4 indicates that rather few simulations are responsible for this deviation.

For DJF temperature, only the third PC is seen to be very inhomogeneous across the different modelling projects in Table 2. This pattern shows a north–south gradient (Fig. 5 row 2). A sub-set of CORDEX at both resolutions shows a larger than average warming at the northern edge of the domain, whereas PRUDENCE shows lower values than the ensemble-mean (Figs. 2, 4). It should be noted, however, that the corresponding amplitude of this PC is not very high (Fig. 4); the various project signals shown in Fig. 2 do not look very different.

For JJA precipitation (Fig. 5 row 3) two inhomogeneous PCs are found: PC number 1, where most models of PRUDENCE as well as of ENSEMBLES (Fig. 4 row 3) generally show much larger drying over Europe than the CORDEX simulations, particularly over the Mediterranean region (see also Fig. 3). Also, PC number 3 indicates an anomalously high precipitation increase in the Baltic Sea and decrease in the western Mediterranean, which is higher for PRUDENCE than for the other ensembles. This is a feature most likely related to the dominating single GCM in this set of experiments (Christensen and Christensen 2007).

Finally, the last row of Fig. 5 indicates that the signal of DJF precipitation is quite independent of project across all the different ensembles, except for PC number 3, which indicates that PRUDENCE shows a significantly higher precipitation increase in Central and Northern Europe compared to the other projects (Fig. 3).

Overall, all of the abovementioned findings (Fig. 5) are consistent with the patterns shown in Figs. 2 and 3, and serve to document the robustness of these signals across the P–E–C sequence.

3.3 Observed trends

To test the results discussed above, we have compared the scaled regional climate model projections with observed trends from 1950 to 2016. Figure 6 depicts the observed linear trends in seasonal mean temperature and relative precipitation change, scaled by the observed global temperature change to allow for direct comparison with the scaled results from the three regional climate model ensembles (Figs. 2, 3). For observed temperatures, even the smallest trend within the estimated error margins can be found to be significantly different from zero ($p = 0.05$) almost everywhere, in particular in JJA (no hatching). Conversely, in the eastern Mediterranean, the observed trends are generally weak; and they are not significant in DJF. This corroborates findings, e.g., by Barkhordarian et al. (2016).

As expected, the observed seasonal temperature (Fig. 6) increases across European land masses are in general larger than the annual global mean, with the strongest increase

Table 2 Statistical F test quantity, basically the ratio between variance between ensemble means weighted by the number of simulations in each ensemble and the weighted mean intra-model variance

	PC1	PC2	PC3
Temperature JJA	15.6	2.2	0.5
Temperature DJF	0.6	4.0	11.0
Precipitation JJA	9.8	3.1	10.1
Precipitation DJF	1.5	0.6	13.1

Critical value for the hypothesis of ensemble-dependent signals is 2.7 at a 95% confidence level, assuming identically distributed (iid) quantities. Formally significant values (in bold) indicate that the PC values depend on the project

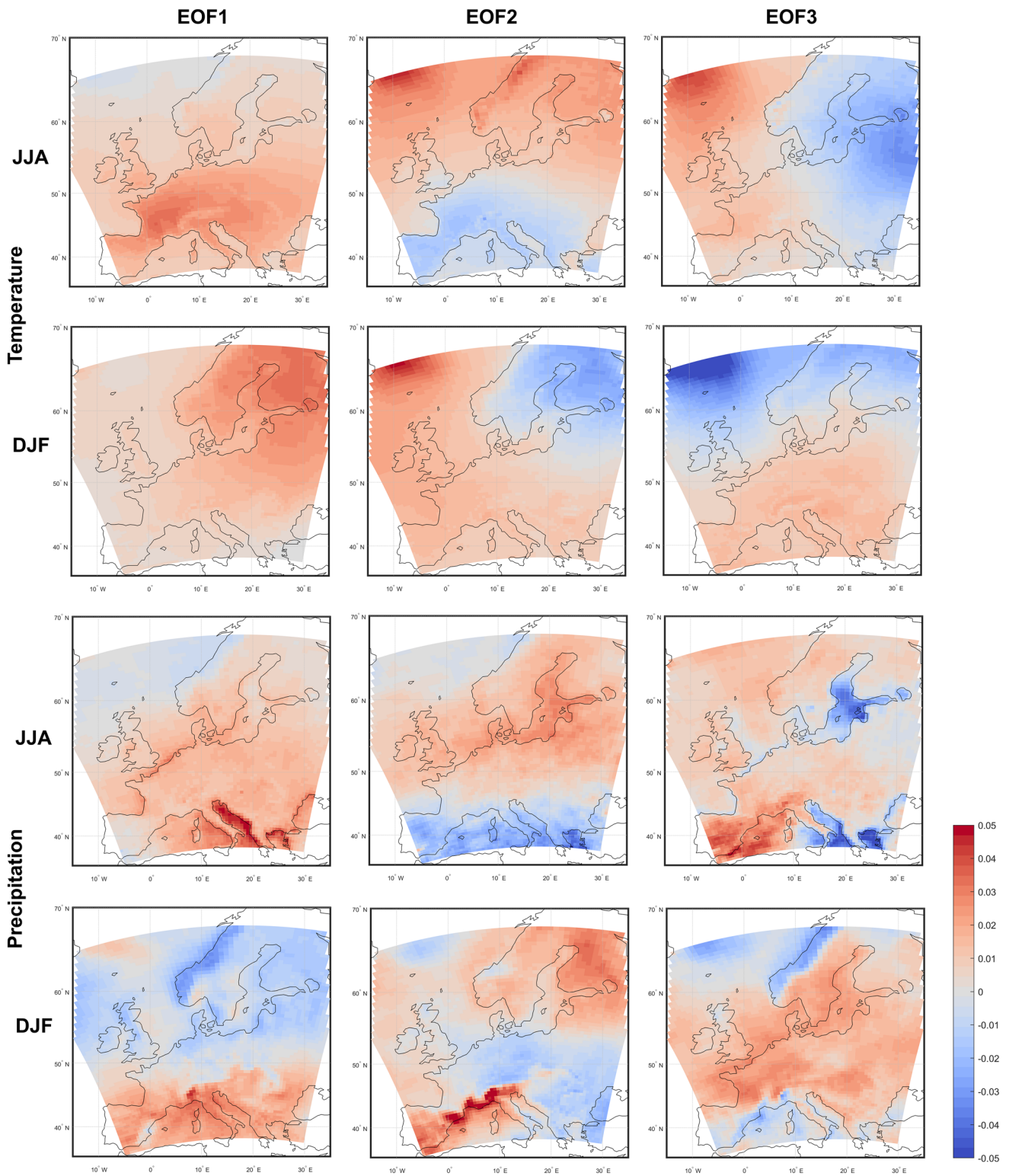


Fig. 5 Spatial patterns corresponding to the leading three PCs (EOF1–EOF3) for seasonal mean scaled temperature change, respectively, seasonal mean scaled relative precipitation change for PRUDENCE, ENSEMBLES, Euro-CORDEX44 and Euro-CORDEX11.

All patterns are normalized with the sum of squares being unity. First and third row indicate JJA, second and fourth row indicate DJF. The sign of the EOF plots is selected to reflect a positive sum across cells

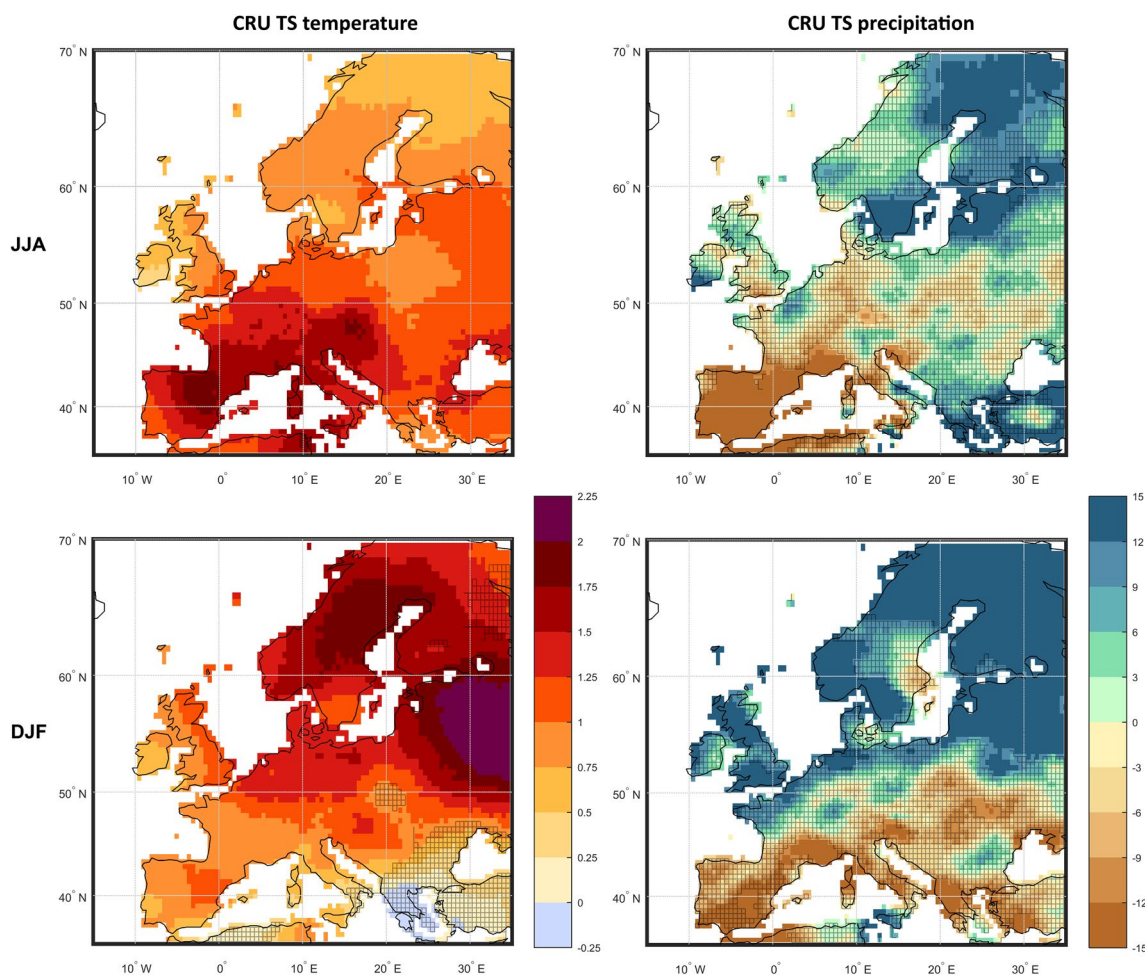


Fig. 6 Scaled observed linear trends in seasonal mean temperature [$^{\circ}\text{C}$] and relative precipitation change [%]. The hatching indicates where the pointwise trends are not considered significant with a p value of 0.05

depicted towards the northeast in DJF and towards the southwest in JJA. Considering precipitation on the other hand, the observed trends are far from significant everywhere: in DJF, we find significantly positive trends only over Russia, northern Scandinavia and Scotland. Isolated positive trends are also found further south, whereas significantly negative trends are found over the Iberian Peninsula. In JJA there is a significant increase in precipitation in the east (north-eastern Scandinavia and parts of Turkey). For precipitation, neither the smallest nor the largest trends within the error margins are significantly different from zero with the notable exceptions just mentioned (not shown).

In order to compare and quantify the robustness of the projected changes in temperature and precipitation with these patterns in observed trends, we examine the normalized patterns from observed trends and from all of the 99 P–E–C simulations. We calculate the root mean squared distance between each pair of patterns; this distance is obtained by subtracting the spatial average from each pattern and

normalizing with the spatial standard deviation over the grid points in question; then calculating the root of the squared sum of differences between these normalized patterns over the area. This is different from the EOF analysis, where the pointwise ensemble-mean pattern is subtracted, and no normalization is performed. We sum all distances from each pattern to all other patterns. The rank of each pattern among the 100-member set is calculated according to this sum. For the observational pattern, the lower this rank, the higher the agreement between the observational pattern and the ensemble of model patterns.

Examining the first line of Table 3, it seems that the pattern of observational trends has deviations from the set of model results, particularly for winter temperature, since the ranking of the observations is close to the bottom position of number 100. Considering only geographical points, which are not hatched in Fig. 6 (line 2 in Table 3; i.e., areas of significant observational change) the situation changes somewhat towards higher agreement between observation

Table 3 Ranking of the pattern of observations among the 100 patterns examined, according to the sum of distances to all other elements

Field	Temperature JJA	Temperature DJF	Precipitation JJA	Precipitation DJF
Rank for all 2903 observation grid points	78	97	65	88
Rank for only non-hatched grid points	78	87	39	96
Number of non-hatched grid points	2903	2429	378	711

As there are 100 elements in total, this corresponds to a percentile. Upper row: All grid points with observations included. Second row: Only non-hatched grid points, which have a significant observational trend, from Fig. 6 are included. The final row contains the number of non-hatched grid points used

and model ensemble, particularly for winter temperature and summer precipitation, but not for winter precipitation. The number of relevant points, which are land points in the observational lattice with a significant ($p=0.05$) observational trend inside the common model domain, are listed in line 3 of Table 3; for winter temperature all possible points have significant observational trend, whereas only around one eighth of the points have a significant summer precipitation trend (cf. Fig. 6).

4 Discussion and conclusion

In view of the known model- and GCM/RCM configuration deficits discussed above, particularly in the older part of the P–E–C sequence, the pattern-scaled projections of temperature and precipitation change across 20 years of model evolution appear quite similar. Likewise, the patterns of change of annual mean temperature and precipitation can be found to compare well with the results from global models (not shown; e.g., Collins et al. 2013). This apparent robustness of the P–E–C sequence represents an important contribution to credibility of the projections provided by the Euro-CORDEX data set representing the current state of the art.

A second argument towards the aim of credibility is the degree of similarity of the P–E–C climate change patterns with the scaled observed linear trends for both JJA and DJF temperature and precipitation where these trends are statistically significantly different from zero, as indicated by comparing Figs. 2 and 3 with Fig. 6 and specifically in Table 3. This only holds marginally for winter conditions, in particular for precipitation. In addition, it is not well established how to attribute the observed trends between decadal scale natural variations and a forced signal associated with global warming. The results regarding a lesser disagreement between models and observation when only considering points with a significant observational trend suggest that the uncertainty in values for observed trends caused by decadal climate fluctuations is, to a large degree, responsible for the disagreement between observations and models. There are, however, caveats to the conclusion: with fewer points, there will be a larger noise obscuring differences between model

ensemble and observations; also, many model simulations are very similar (e.g., members of both CORDEX11 and CORDEX44), and therefore the low rank of observations is probably too extreme compared to a case with more independent simulations.

A third objective would be the ability to demonstrate that internal decadal variability is small in comparison to the climate change signal. While this is beyond the scope of the present paper, we note that Matte et al. (2019) demonstrate this to hold for the Euro-CORDEX ensembles, whenever S/N is approximately larger than 1. Despite the higher degree of freedom following from a higher grid resolution of app. 12 km, the large-scale pattern and even local details in Euro-CORDEX11 are clearly in line with coarser resolution findings, suggesting that credible projections can be deduced from this data set. The usefulness of the multi-model mean values as a measure of the projected signal of change and the associated robustness is supported by recent work on the behaviour of multi-model ensemble data by (Christiansen 2017). That said, maps using multi-model statistics to represent the projected future climate change in Europe must in general be considered with caution, as no individual model will replicate the ensemble-mean patterns of change, i.e., the pattern representing the multi-model mean may not be plausible in a strictly physics based interpretation (Madsen et al. 2017).

In this study we consider the robustness of temperature and precipitation changes at the seasonal scale. However, there is supporting evidence that these findings could be extended to other time scales as well, such as the occurrence of heatwaves or extreme daily precipitation (Boberg et al. 2009, 2010; Fischer and Schär 2010; Christensen et al. 2015; Rajczak and Schär 2017). Likewise, we have here only considered the average climate conditions—a necessary prerequisite to moving on to climate parameters/conditions representing, e.g., extremes. Again, the apparent robustness of the multi-model mean climate change signal across the P–E–C model suite indicates that for the overall warming patterns, and for the drying and moistening tendencies, more complex climate variables may also be robust. This proposition clearly needs to be tested, and analyses such as comparing the results of regional climate simulations with observed

trends at the grid point level could be applied if the existence of observational data sets will allow for such a comparison. To this end, new data sets combining high resolution in both time and geography are critically needed. Yet, it still remains to be seen whether the implied matching trends are formally detectable or attributable to anthropogenic climate change (Allen and Ingram 2002; Hall and Qu 2006).

Acknowledgements The research leading to these results has received funding from the European Research Council (ERC) under the European Community's Seventh Framework Programme (FP7/2007–2013) ERC Grant agreement 610055 as part of the ice2ice Project, from NordForsk through Project No. 74456 “Statistical Analysis of Climate Projections” (eSACP) and from COWIfonden. We acknowledge the World Climate Research Programme (WCRP)'s Coordinated Regional Climate Downscaling Experiment (CORDEX) for its role in making available the WCRP CORDEX multi-model data set, and we thank the climate modelling groups for producing and making available their model outputs. We are grateful to an anonymous reviewer for pointing out original weaknesses in our comparison with observed trends and to Dr. Peter Thejll, DMI, for helpful discussions on the presentation of the approach taken to carry out the final analysis.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Allen M, Ingram W (2002) Constraints on future changes in climate and the hydrologic cycle. *Nature* 419:224–232
- Barkhordarian A, Storch H, Zorita E, Gómez-Navarro JJ (2016) An attempt to deconstruct recent climate change in the Baltic Sea basin. *J Geophys Res Atmos* 121:13207–13217
- Bindoff NL, Stott PA, AchutaRao KM, Allen MR, Gillett N, Gutzler D, Hansing K, Hegerl G, Hu Y, Jain S, Mokhov II, Overland J, Perlwitz J, Sebbani R, Zhang X (2013) Detection and attribution of climate change: from global to regional. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds) *Climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA
- Boberg F, Christensen JH (2012) Overestimation of Mediterranean summer temperature projections due to model deficiencies. *Nat Clim Change* 2:433–436
- Boberg F, Berg P, Thejll P, Gutowski WJ, Christensen JH (2009) Improved confidence in climate change projections of precipitation evaluated using daily statistics from the PRUDENCE ensemble. *Clim Dyn* 32:1097–1106
- Boberg F, Berg P, Thejll P, Gutowski WJ, Christensen JH (2010) Improved confidence in climate change projections of precipitation further evaluated using daily statistics from ENSEMBLES models. *Clim Dyn* 35:1509–1520
- Buonomo E, Jones R, Huntingford C, Hannaford J (2007) On the robustness of changes in extreme precipitation over Europe from two high resolution climate change simulations. *Q J R Meteorol Soc* 133:65–81
- Christensen JH, Christensen OB (2007) A summary of the PRUDENCE model projections of changes in European climate by the end of this century. *Clim Change* 81:7–30
- Christensen JH, Machenhauer B, Jones RG, Schar C, Ruti PM, Castro M, Visconti G (1997) Validation of present-day regional climate simulations over Europe: LAM simulations with observed boundary conditions. *Clim Dyn* 13:489–506
- Christensen JH, Raisanen J, Iversen T, Borge D, Christensen OB, Rummukainen M (2001) A synthesis of regional climate change simulations—a Scandinavian perspective. *Geophys Res Lett* 28:1003–1006
- Christensen JH, Carter TR, Giorgi F (2002) PRUDENCE employs new methods to assess European climate change. *EOS Trans Am Geophys Union* 83:147
- Christensen JH, Kjellstrom E, Giorgi F, Lenderink G, Rummukainen M (2010) Weight assignment in regional climate models. *Clim Res* 44:179–194
- Christensen JH, Krishna Kumar K, Aldrian E, An S-I, Cavalcanti IFA, de Castro M, Dong W, Goswami P, Hall A, Kanyanga JK, Kitoh A, Kossin J, Lau N-C, Renwick J, Stephenson DB, Xie S-P, Zhou T (2013) Climate phenomena and their relevance for future regional climate change. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds) *Climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA
- Christensen OB, Yang S, Boberg F, Maule CF, Thejll P, Olesen M, Drews M, Sorup HJD, Christensen JH (2015) Scalability of regional climate change in Europe for high-end scenarios. *Clim Res* 64:25–38
- Christiansen B (2017) Ensemble averaging and the curse of dimensionality. *J Clim* 31:1587–1596
- Collins M, Knutti R, Arblaster JM, Dufresne J-L, Fifelet T, Friedlingstein P, Gao X, Gutowski WJ, Johns T, Krinner G, Shongwe M, Tebaldi C, Weaver AJ, Wehner M (2013) Long-term climate change: projections, commitments and irreversibility. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds) *Climate change 2013: the physical science basis contribution of Working Group I to the Fifth Assessment Report of the intergovernmental panel on climate change*. Cambridge University Press, Cambridge
- Deque M, Rowell DP, Luthi D, Giorgi F, Christensen JH, Rockel B, Jacob D, Kjellstrom E, de Castro M, van den Hurk B (2007) An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections. *Clim Change* 81:53–70
- Fernández J, Frías MD, Cabos WD et al (2019) Consistency of climate change projections from multiple global and regional model intercomparison projects. *Clim Dyn* 52:1139. <https://doi.org/10.1007/s00382-018-4181-8>
- Fischer E, Schär C (2010) Consistent geographical patterns of changes in high-impact European heatwaves. *Nat Geosci* 3:398–403
- Giorgi F, Gutowski WJ (2015) Regional dynamical downscaling and the CORDEX initiative. *Annu Rev Environ Resour* 40:467–490
- Gutowski W, Giorgi F, Timbal B, Frigon A, Jacob D, Kang H, Raghavan K, Lee B, Lennard C, Nikulin G, O'Rourke E, Rixen M, Solman S, Stephenson T, Tangang F (2016) WCRP COordinated Regional Downscaling EXperiment (CORDEX): a diagnostic MIP for CMIP6. *Geosci Model Dev* 9:4087–4095
- Hagemann S, Botzet M, Machenhauer B (2001) The summer drying problem over south-eastern Europe: sensitivity of the limited area model HIRHAM4 to improvements in physical parameterization

- and resolution. *Phys Chem Earth Part B Hydrol Oceans Atmos* 26:391–396
- Hall A, Qu X (2006) Using the current seasonal cycle to constrain snow albedo feedback in future climate change. *Geophys Res Lett* 33:L03502
- Harris I, Jones PD, Osborn TJ, Lister DH (2014) Updated high-resolution grids of monthly climatic observations: the CRU TS3.10 Dataset. *Int J Climatol* 34:623–642
- Hawkins E, Sutton R (2009) The potential to narrow uncertainty in regional climate predictions. *Bull Am Meteorol Soc* 90:1095–1107
- Hirschi M, Seneviratne S, Alexandrov V, Boberg F, Boroneant C, Christensen O, Formayer H, Orlowsky B, Stepanek P (2011) Observational evidence for soil-moisture impact on hot extremes in southeastern Europe. *Nat Geosci* 4:17
- IPCC (2013) Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge
- Jacob D, Petersen J, Eggert B, Alias A, Christensen OB, Bouwer LM, 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, Samuelsen P, Somot S, Soussana J-F, Teichmann C, Valentini R, Vautard R, Weber B, Yiou P (2014) EURO-CORDEX: new high-resolution climate change projections for European impact research. *Reg Environ Change* 14:563–578
- Johns TC, Gregory JM, Ingram WJ, Johnson CE, Jones A, Lowe JA, Mitchell JFB, Roberts DL, Sexton DMH, Stevenson DS, Tett SFB, Woodage MJ (2003) Anthropogenic climate change for 1860 to 2100 simulated with the HadCM3 model under updated emissions scenarios. *Clim Dyn* 20:583–612
- Laux P, Nguyen PNB, Cullmann J, Van TP, Kunstmann H (2017) How many RCM ensemble members provide confidence in the impact of land-use land cover change? *Int J Climatol* 37:2080–2100
- Lorenz EN (1956) Empirical orthogonal functions and statistical weather prediction. Department of Meteorology, MIT Statistical Forecast Project Rep 1
- Machenhauer B, Windelband M, Botzet M, Christensen JH, Déqué M, Jones R, Ruti P, Visconti G (1998) Validation and analysis of regional present-day climate and climate change simulations over Europe. MPI report 275, Max-Planck-Institut für Meteorologie
- Madsen M, Langen P, Boberg F, Christensen J (2017) Inflated uncertainty in multimodel-based regional climate projections. *Geophys Res Lett* 44:11606–11613
- Matte D, Larsen MAD, Christensen OB, Christensen JH (2019) Robustness and scalability of regional climate projections over Europe. *Front Environ Sci* 6:163
- Meehl G, Covey C, Delworth T, Latif M, McAvaney B, Mitchell J, Stouffer R, Taylor K (2007) The WCRP CMIP3 multimodel dataset—a new era in climate change research. *Bull Am Meteorol Soc* 88:1383
- Meinshausen M, Smith S, Calvin K, Daniel J, Kainuma M, Lamarque JF, Matsumoto K, Montzka S, Raper S, Riahi K, Thomson A, Velders G, van Vuuren DP (2011) The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Clim Change* 109:213–241
- Nakicenovic N, Alcamo J, Grubler A, Riahi K, Roehrl RA, Rogner HH, Victor N (2000) Special report on emissions scenarios (SRES), a special report of Working Group III of the intergovernmental panel on climate change. Cambridge University Press, Cambridge
- Pal JS, Giorgi F, Bi X (2004) Consistency of recent European summer precipitation trends and extremes with future regional climate projections. *Geophys Res Lett* 31:L13202
- Pan Z, Christensen JH, Arritt RW, Gutowski WJ, Takle ES, Otieno F (2001) Evaluation of uncertainties in regional climate change simulations. *J Geophys Res Atmos* 106:17735–17751
- Rajczak J, Schär CCJD (2017) Projections of future precipitation extremes over Europe: a multimodel assessment of climate simulations. *J Geophys Res Atmos* 122:10773–710800
- Roeckner E, Arpe K, Bengtsson L, Christoph M, Claussen M, Dümenil L, Esch M, Giorgette M, Schlese U, Schulzweida U (1996) The atmospheric general circulation model ECHAM-4: model description and simulation of present-day climate. MPI report 218, Max-Planck-Institut für Meteorologie
- Rummukainen M (2016) Added value in regional climate modeling. *Wiley Interdiscip Rev Clim Change* 7:145–159
- Rummukainen M, Rockel B, Barring L, Christensen JH, Reckermann M (2015) Twenty-first-century challenges in regional climate modeling. *Bull Am Meteorol Soc* 96:ES135–ES138
- Ruti PM, Somot S, Giorgi F, Dubois C, Flaounas E, Obermann A, Dell’Aquila A, Pisacane G, Harzallah A, Lombardi E, Ahrens B, Akhtar N, Alias A, Arsouze T, Aznar R, Bastin S, Bartholy J, Béranger K, Beuvier J, Bouffies-Cloché S, Brauch J, Cabos W, Calmanti S, Calvet JC, Carillo A, Conte D, Coppola E, Djurdjevic V, Drobinski P, Elizalde-Arellano A, Gaertner M, Galàn P, Gallardo C, Gualdi S, Goncalves M, Jorba O, Jordà G, L’Heveder B, Lebeaupin-Brossier C, Li L, Liguori G, Lionello P, Maciàs D, Nabat P, Önl B, Raikovic B, Ramage K, Sevault F, Sannino G, Struglia MV, Sanna A, Torma C, Vervatis V (2015) Med-CORDEX initiative for Mediterranean climate studies. *Bull Am Meteorol Soc* 97:1187–1208
- Santer BD, Wigley TML, Schlesinger ME, Mitchell JFB (1990) Developing climate scenarios from equilibrium GCM results. Max-Planck-Institut für Meteorologie report no. 47, Max-Planck-Institut für Meteorologie, Hamburg, Germany
- Schiermeier Q (2004) Modellers deplore ‘short-termism’ on climate. *Nature* 428:593
- Seneviratne SI, Lüthi D, Litschi M, Schär C (2006) Land–atmosphere coupling and climate change in Europe. *Nature* 443:205–209
- Takle ES, Gutowski WJ, Arritt RW, Pan ZT, Anderson CJ, da Silva RR, Caya D, Chen SC, Giorgi F, Christensen JH, Hong SY, Juang MHM, Katzfey J, Lapenta WM, Laprise R, Liston GE, Lopez P, McGregor J, Pielke RA, Roads JO (1999) Project to intercompare regional climate simulations (PIRCS): description and initial results. *J Geophys Res Atmos* 104:19443–19461
- Taylor KE, Stouffer RJ, Meehl GA (2011) An overview of CMIP5 and the experiment design. *Bull Am Meteorol Soc* 93:485–498
- Tebaldi C, Arblaster J (2014) Pattern scaling: its strengths and limitations, and an update on the latest model simulations. *Clim Change* 122:459–471
- Van der Linden P, Mitchell JFBE (2009) ENSEMBLES: climate change and its impacts—summary of research and results from the ENSEMBLES project. Met Office Hadley Centre, Exeter
- van Vuuren DP, Edmonds J, Kainuma M, Riahi K, Thomson A, Hibbard K, Hurtt GC, Kram T, Krey V, Lamarque J-F, Masui T, Meinshausen M, Nakicenovic N, Smith SJ, Rose SK (2011) The representative concentration pathways: an overview. *Clim Change* 109:5–31. <https://doi.org/10.1007/s10584-011-0148-z>
- Vautard R, Gobiet A, Sobolowski S, Kjellström E, Stegehuis A, Watkiss P, Mendlik T, Landgren O, Nikulin G, Teichmann C, Jacob D (2014) The European climate under a 2 °C global warming. *Environ Res Lett* 9:034006