



Validation of temperature trends in the ENSEMBLES regional climate model runs driven by ERA40

Philip Lorenz^{1,2,*}, Daniela Jacob¹

¹Max Planck Institute for Meteorology, Bundesstrasse 53, 20146 Hamburg, Germany

²Present address: Institute of Meteorology, Freie Universität Berlin, Carl-Heinrich-Becker-Weg 6–10, 12165 Berlin, Germany

ABSTRACT: In the EU-Project ENSEMBLES, a major objective was the development of a weighting system for regional climate models (RCMs) based on the ability of the RCMs to reproduce observed characteristics of important atmospheric variables. Within this context, a suite of state-of-the-art RCMs was employed for the creation of probabilistic future regional climate change scenarios for Europe. Among others, one measure to determine the individual RCM-weights is their ability to reproduce the observed trends in 2 m temperature in the reanalysis (ERA40)-driven experiments for the period 1960–2000. As the reference, we used the new ENSEMBLES observational gridded dataset for Europe (E-OBS). As additional datasets for comparisons, we also used the near-surface temperature datasets from the CRU observations and from the ERA40 and NCEP/NCAR reanalysis. The analysis was performed for the land fraction of 8 different European regions, the so-called PRUDENCE regions defined within the PRUDENCE project (<http://prudence.dmi.dk>). Annual and seasonal linear trends in near-surface temperature were computed for each ENSEMBLES RCM, the E-OBS dataset, and for the additional datasets mentioned above. In all regions, the computed linear temperature trends based on annual mean temperatures showed smaller values for the RCMs and the NCEP/NCAR reanalysis than for both observational datasets, and in most regions also smaller values than for the ERA40 reanalysis dataset. Depending on the magnitude of the difference in linear trends between the individual RCMs and the E-OBS dataset, skill scores were assigned to each RCM. The resulting skill scores were of similar magnitude (0.7–0.9) for the different models and regions (except for Scandinavia, which had lower skill scores around 0.6–0.8). Spatially aggregated for all of Europe, and combined from annual and seasonal into one value for each RCM, these skill scores were included in the general ENSEMBLES RCM weighting system

KEY WORDS: Temperature trends · Trend analysis · Climate sensitivity · Regional Climate Model · Europe · ENSEMBLES · ERA40 · PRUDENCE

Resale or republication not permitted without written consent of the publisher

1. INTRODUCTION

The results of different global and regional climate models (GCMs and RCMs, respectively) continue to show a large spread, even when the models are applied under identical forcing conditions (Solomon et al. 2007). One way to judge the quality of model results is by an in-depth comparison against observations, which does not immediately imply that good results for the past are an indicator for good quality future climate

projections. Such a comparison includes different hydro-meteorological quantities and their related pattern on different temporal and spatial scales (e.g. Jacob et al. 2007, Christensen et al. 2008, Feldmann et al. 2008, Früh et al. 2010). The overall assessment of the different comparisons can be interpreted as the model quality. Within the EU-Project ENSEMBLES (van der Linden & Mitchell 2009), various aspects of model quality have been determined against observational data, and a weighting system for an ensemble of

*Email: Philip.Lorenz@fu-berlin.de

RCMs has been developed (Christensen et al. 2010, this Special).

The sensitivity of a climate model to external forcing (like greenhouse gas concentrations) may be described by the trends in temperatures or other climate parameters. However, it is difficult to validate trends in climate simulations: the observed change in 2 m temperature since preindustrial conditions until the end of the 20th century is less than 1 K for Europe (Jones & Moberg 2003), and about 1.5 K for large parts of Europe until the end of 2007 (van Oldenborgh et al. 2009). State-of-the-art climate models show systematic absolute temperature biases of a similar or even larger magnitude (Solomon et al. 2007), although these kinds of biases do not necessarily influence the temperature trend. Especially in the case of control experiments, internal multi-decadal variability masks long-term trends (Giorgi 2002). Furthermore, trends can be computed by different statistical methods, leading to different estimations of trends. Despite of these difficulties, it is important for a climate model to simulate a reasonable temperature trend, especially when applying it to future climate simulations; consequently the measure of temperature trends is part of the ENSEMBLES RCM weighting system.

Here we used the ENSEMBLES RCM experiments for Europe forced by the global reanalysis ERA40 (Uppala et al. 2005) for the period 1961–2000. As this is a rather short period for calculations of long-term trends, all trends computed within this analysis were simple linear trends covering the whole period 1961–2000. For all RCMs and for the new ENSEMBLES observational gridded dataset as a reference, we computed linear yearly and seasonal trends for the land fractions of 8 regions for this time period. The difference between the calculated trend for each RCM

and the trend in the reference dataset was then used to compute skill scores for each model, season, and region. For aggregation of the skill scores for the 8 regions into a general skill score for Europe, 2 different methods were applied. Our final skill score was used in the ENSEMBLES RCM weighting system as described by Christensen et al. (2010).

2. MODELS AND DATA

Validation of temperature trends was performed for the ENSEMBLES RCM simulations for Europe on a horizontal resolution of about 25 km; these simulations were initialized and driven at their lateral boundaries by the reanalysis data set ERA40 (Uppala et al. 2005) from the European Centre for Medium Range Weather Forecasts (ECMWF). The respective institutions, model names, and acronyms for referencing the individual simulations within this paper are listed in Table 1. For all RCM simulations listed in Table 1, we used monthly mean near-surface (2 m) temperature downloaded from the ENSEMBLES RCM data base at DMI (<http://ensembles-eu.metoffice.com>).

The gridded land-only ENSEMBLES data set E-OBS (v. 2) based on station measurements (Haylock et al. 2008, Klok & Klein Tank 2009) was used as the reference. This dataset is available for Europe in several grid resolutions; we used the version on a regular 0.25° grid.

Three additional global near-surface temperature datasets were incorporated for comparison. First, we used the observational dataset from the Climate Research Unit (CRU) v. TS3, which is a gridded land-only dataset based on station measurements (Mitchell & Jones 2005). Both the E-OBS and the CRU datasets

Table 1. Institutions, model names, and acronyms used in this paper for the ENSEMBLES RCM simulations. The Hadley Centre performed 3 ERA40-driven regional climate model (RCM) simulations with different versions of the HadRm: HC-Q0 is the version with 'normal' climate sensitivity, HC-Q3 with 'low' climate sensitivity, and HC-Q16 with 'high' climate sensitivity

Institution	RCM	Acronym
Met Eireann, Community Climate Change Consortium for Ireland	RCA3.0	C4I
Czech Hydrometeorological Institute	ALADIN	CHMI
Météo-France CNRM	RM4.5	CNRM
Danish Meteorological Institute	HIRHAM5	DMI
Swiss Institute of Technology	CLM	ETHZ
UK Met Office, Hadley Centre for Climate Prediction and Research	HadRM3	HC-Q0, HC-Q3, HC-Q16
The Abdus Salam International Centre for Theoretical Physics	RegCM3	ICTP
The Royal Netherlands Meteorological Institute	RACMO2	KNMI
The Norwegian Meteorological Institute	HIRHAM	METNO
Max Planck Institute for Meteorology	REMO	MPI-M
OURANOS Consortium on Regional Climatology and Adaptation to Climate Change	MRCC4.2.3	OURANOS
Swedish Meteorological and Hydrological Institute	RCA3.0	SMHI
Universidad de Castilla La Mancha	PROMES	UCLM

make use of partly the same underlying station data, which makes these observational datasets not completely independent from each other. Furthermore, near-surface temperatures from ERA40 and from the NCEP/NCAR reanalysis (hereafter NCEP) provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their web site at www.esrl.noaa.gov/psd/ (Kalnay et al. 1996) were investigated. All 3 additional datasets were remapped from their original resolutions (CRU: 0.5°, ERA40: about 1.125°, NCEP: about 1.9°) to the regular 0.25° grid using only land points before performing the analysis.

3. METHODS

We first describe the computation of linear trends on an annual and seasonal basis for several European sub-regions. Second, the definition of skill scores depending on the difference of the computed linear trends from the respective RCM against the computed linear trend from the E-OBS dataset is given. Third, the aggregation from the annual and seasonal skill scores computed for each European sub-region into annual and seasonal skill scores valid for all of Europe is defined. Last, the combination of the annual and seasonal skill scores into a general skill score resulting in 1 general weight per RCM is introduced.

3.1. Linear trend computation

To analyze the temperature trend, we used linear regression. In this method, a regular time series can be described by Eq. (1):

$$y_i = \alpha + \beta t_i + r_i \quad (1)$$

where y_i is the value of any variable at time step i , α and β are the regression coefficients, t_i is the year at time step i , and r_i is the residual value between the temperature value and the value given by the regression line at time step i . In our case (estimation of trends), we were interested in the regression parameter β , which gives the magnitude and direction of the slope of the regression line.

For the computation of the linear regression coefficients α and β , the sum Q of the squares of the residuals for each time step (n is the number of time steps) has to be minimized (method of least squares):

$$Q = \sum_{i=1}^n r_i^2 \quad (2)$$

For minimized Q , the resulting regression parameter β can be computed using Eq. (3):

$$\beta = \frac{\sum_{i=1}^n (t_i - \bar{t})(y_i - \bar{y})}{\sum_{i=1}^n (t_i - \bar{t})^2} \quad (3)$$

with overbars indicating the mean value of the time series.

With this method, linear temperature trends were computed for all datasets mentioned in the previous section. The trends were computed for the 8 European sub-regions shown in Fig. 1, the so called 'PRUDENCE regions' defined within the EU-project PRUDENCE. In a first step, monthly area mean values of 2 m temperature were computed for the land fraction of each sub-region. Out of these monthly values, annual and seasonal mean values were computed, which built the basis for the computation of the linear trends.

For the annual and seasonal means, the slope parameter β (referred to hereafter as 'trend') and the sum of the squared residuals Q (Eq. 2) were computed for the period 1961–2000 for each region. The resulting values for β were multiplied by 10; therefore, these values represent the decadal linear trend (in K decade⁻¹).

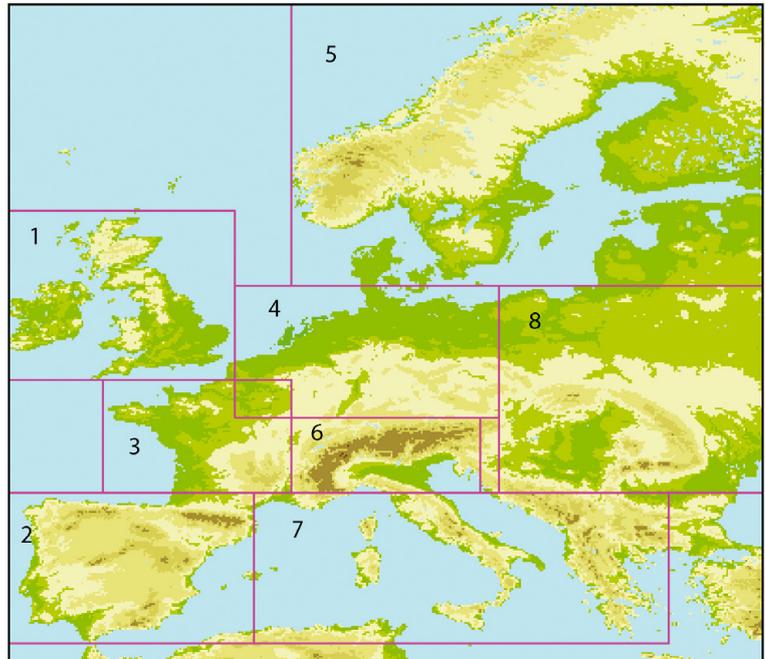


Fig. 1. Eight European sub-regions defined within the PRUDENCE project after Rockel (figure originally appeared in Christensen & Christensen 2007). 1: British Isles (BI), 2: Iberian Peninsula (IP), 3: France (Fr), 4: Mid-Europe (ME), 5: Scandinavia (Sc), 6: Alps (Al), 7: Mediterranean (Md), 8: Eastern Europe (EE)

3.2. Skill score computation

Skill scores (S) were computed for each RCM, each season, and each region from the trends by:

$$S = 1 - \frac{|\beta - \beta_{\text{REF}}|}{\zeta + |\beta - \beta_{\text{REF}}|} \quad (4)$$

S by definition is between 0 and 1, where 0 indicates a large absolute difference between the estimated trend in the RCM (β) and in the reference data set (β_{REF}) and 1 indicates no difference. ζ is a scaling parameter for the resulting skill score S : for very small absolute trend differences between the trends in the RCM (β) and in the reference data set (β_{REF}), S reaches the value of 1; for differences as large as ζ , S gets a value of 0.5; and for differences much larger than ζ , S reaches the value 0. The choice of ζ influences the magnitude of the skill score spread of the RCM suite, but it does not influence the skill score order of the RCMs. Tests with different values of ζ were carried out (0.1, 0.5, and 2.0 K decade⁻¹); in the ENSEMBLES weighting system, we assigned ζ a value of 0.5 K decade⁻¹.

3.3. Spatial aggregation of skill scores

For combining the individual skill scores for the 8 PRUDENCE regions into 1 aggregated skill score for all of Europe, 2 different methods were applied:

First, a simple arithmetic mean of the 8 skill scores of the regions was calculated for each model, resulting in equal weights for each region. This representation of an unweighted aggregated skill is denoted in the figures and tables as EUR-U.

Second, the skill scores of the individual PRUDENCE regions were weighted depending on the magnitude of the sum of the squared residuals Q (Eq. 2) of the reference E-OBS dataset for the individual regions. For smaller (larger) values of Q for a specific region, the time series is more (less) correlated with the regression line, and therefore the estimated trend corresponds more (less) to a linear trend function and should be given a higher (lower) weight. Consequently, the weights W_r for each region (r) were computed for each season (Eq. 5) based on the Q values from the reference E-OBS dataset:

$$W_r = 1 - \frac{Q_r}{\xi + Q_r}; \quad r = 1, \dots, 8 \quad (5)$$

Similar to ζ in Eq. (4), ξ is a scaling parameter: for very small values of Q , W reaches 1; if Q is equal to ξ , W gets the value 0.5; and for values of Q much larger than ξ , W reaches 0. Tests with different values of ξ were carried out (2, 10, and 50 K). For the ENSEMBLES weighting system, a value of 10 K was used for ξ .

The aggregated skill score for Europe, S_{EUR} , was computed by Eq. (6):

$$S_{\text{EUR}} = \frac{\sum_{r=1}^8 W_r S_r}{\sum_{r=1}^8 W_r} \quad (6)$$

This weighted aggregated skill is denoted in the figures and tables as EUR-W.

3.4. Combination of annual and seasonal skill scores

Finally, the skill scores based on annual and seasonal mean values were combined into general skill scores. For the ENSEMBLES weighting system, the ability of the models to realistically simulate the general trend can be considered more important than their ability to simulate the individual seasonal trends. The computed seasonal trends could be influenced by a shift in the phase of the annual cycle of 2 m temperature. As the analysis of the annual cycle is part of another metric of the ENSEMBLES RCM weighting system (Christensen et al. 2010), here we decided to give the skill scores based on the annual mean values a higher weight than the skill score based on seasonal mean values, namely a weight of 0.5 for the skill score based on the annual mean, and 0.125 for each of the skill scores based on the 4 seasonal means, leading to Eq. (7):

$$S_{\text{combined}} = 0.5 \cdot S_{\text{Year}} + 0.125 \cdot (S_{\text{DJF}} + S_{\text{JJA}} + S_{\text{MAM}} + S_{\text{SON}}) \quad (7)$$

4. RESULTS

Computationally derived linear trends are statistical quantities derived from time series; therefore, we first show examples of the temporal evolutions of annual mean near-surface temperatures in the observational datasets and in the ENSEMBLES RCMs. Fig. 2 shows the time series for the PRUDENCE regions British Isles (BI), Iberian Peninsula (IP), France (Fr), Mid Europe (ME), Scandinavia (Sc), Alps (Al), Mediterranean (Md), and Eastern Europe (EE). Both observational datasets E-OBS and CRU showed very similar characteristics. For some regions, the 2 m temperature from the ERA40 reanalysis was very similar to E-OBS and CRU (e.g. Fr), while there were significant differences in other regions, e.g. Sc. These systematic biases between ERA40 and CRU, which were strongest for the beginning of the analysis period, have also been identified for Northern Europe by Keup-Thiel et al. (2006). This can partly be explained by gaps in the availability of synoptic surface data in ERA40 prior to 1967 (Simmons & Jones 2004). The first major result from this analysis

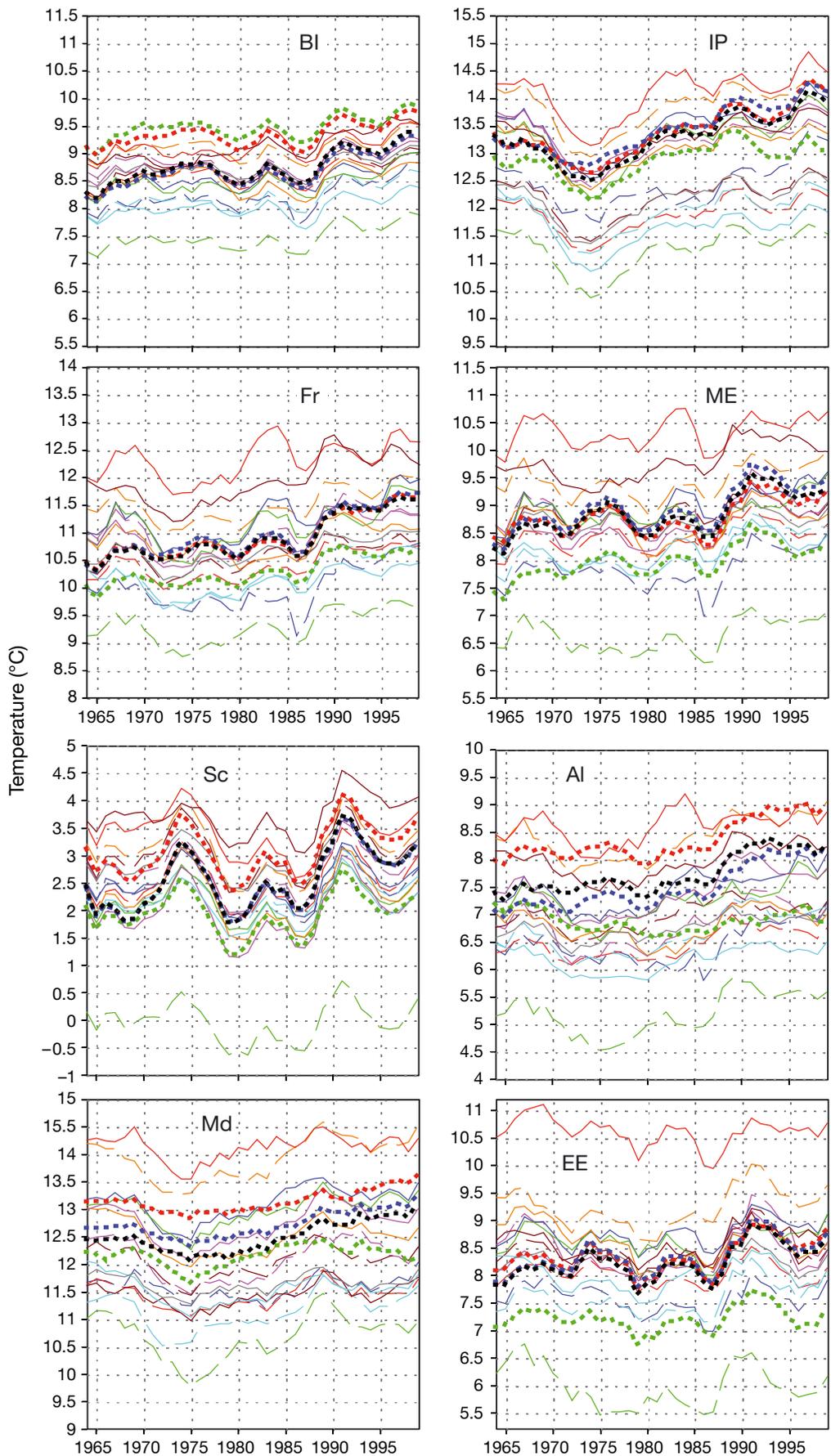
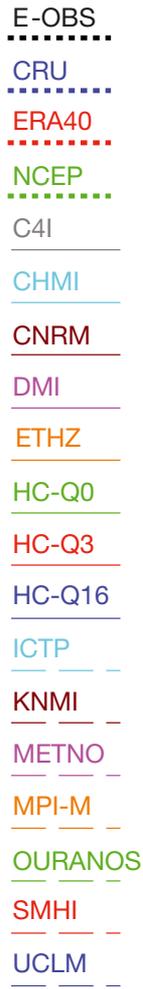


Fig. 2. Annual mean 2 m temperatures ($^{\circ}\text{C}$) (running 5 yr mean) in the observational datasets and in the ENSEMBLES RCMs for the PRUDENCE regions British Isles (BI), Iberian Peninsula (IP), France (Fr) and mid-Europe (ME), Scandinavia (Sc), Alps (Al), Mediterranean (Md), and Eastern Europe (EE)

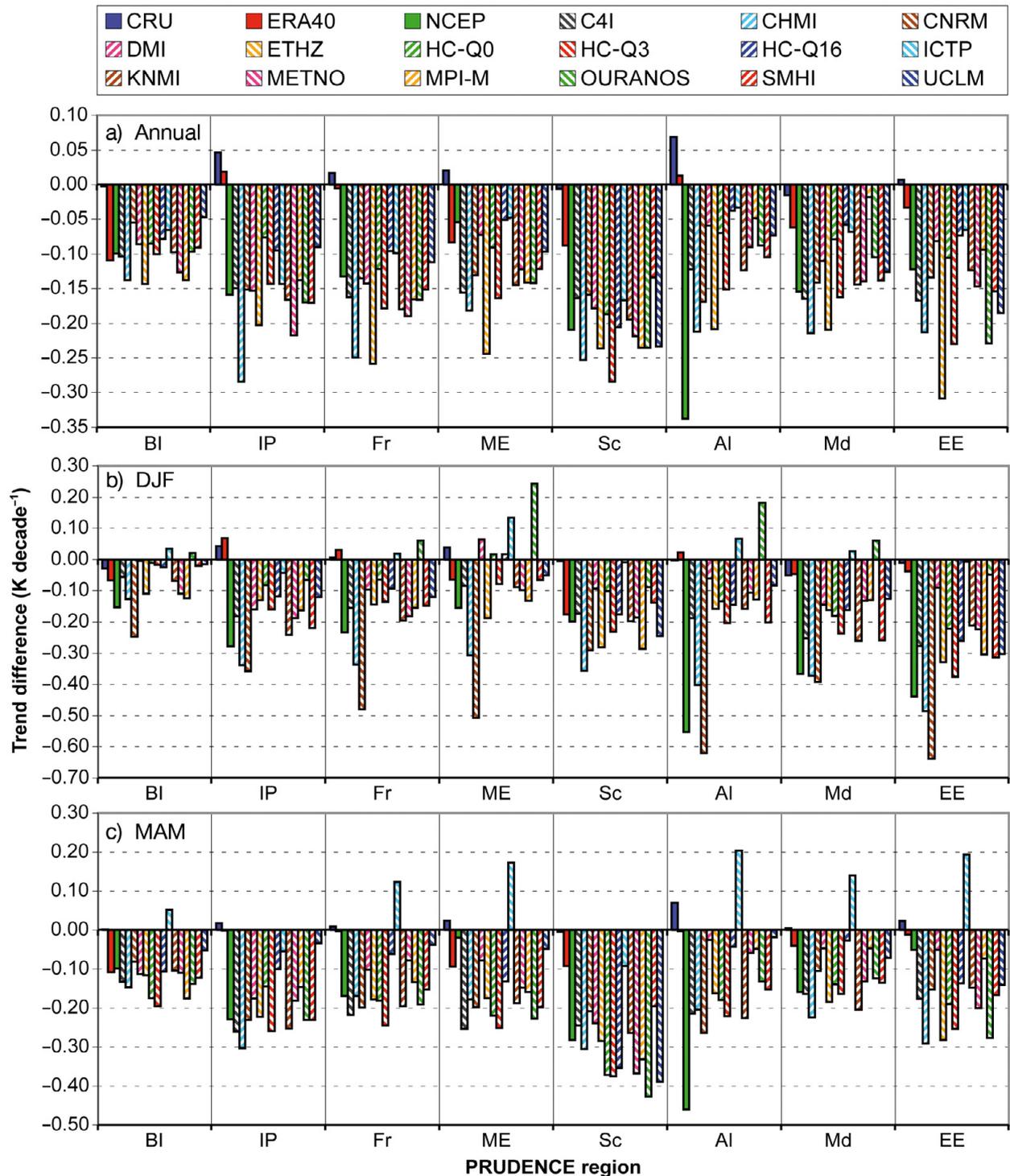


Fig. 3 (this and the next page). Differences of linear trends in 2 m temperature (K decade^{-1}) with respect to E-OBS based on (a) annual, (b) winter (DJF), (c) spring (MAM), (d) summer (JJA), and (e) autumn (SON) mean values for CRU (solid purple bars), ERA40 (solid red bars), NCEP (solid green bars), and all ENSEMBLES RCMs (hatched bars) for all PRUDENCE regions. The sorting order of the bars for each region corresponds to the sorting order within the key

is that the 2 m temperatures in ERA40 are not similar in all regions to the those from the 2 observational datasets, which were very close to each other. For the

NCEP reanalysis dataset, in most regions larger differences to the observational datasets were found than in ERA40. In contrast to ERA40, this reanalysis did not

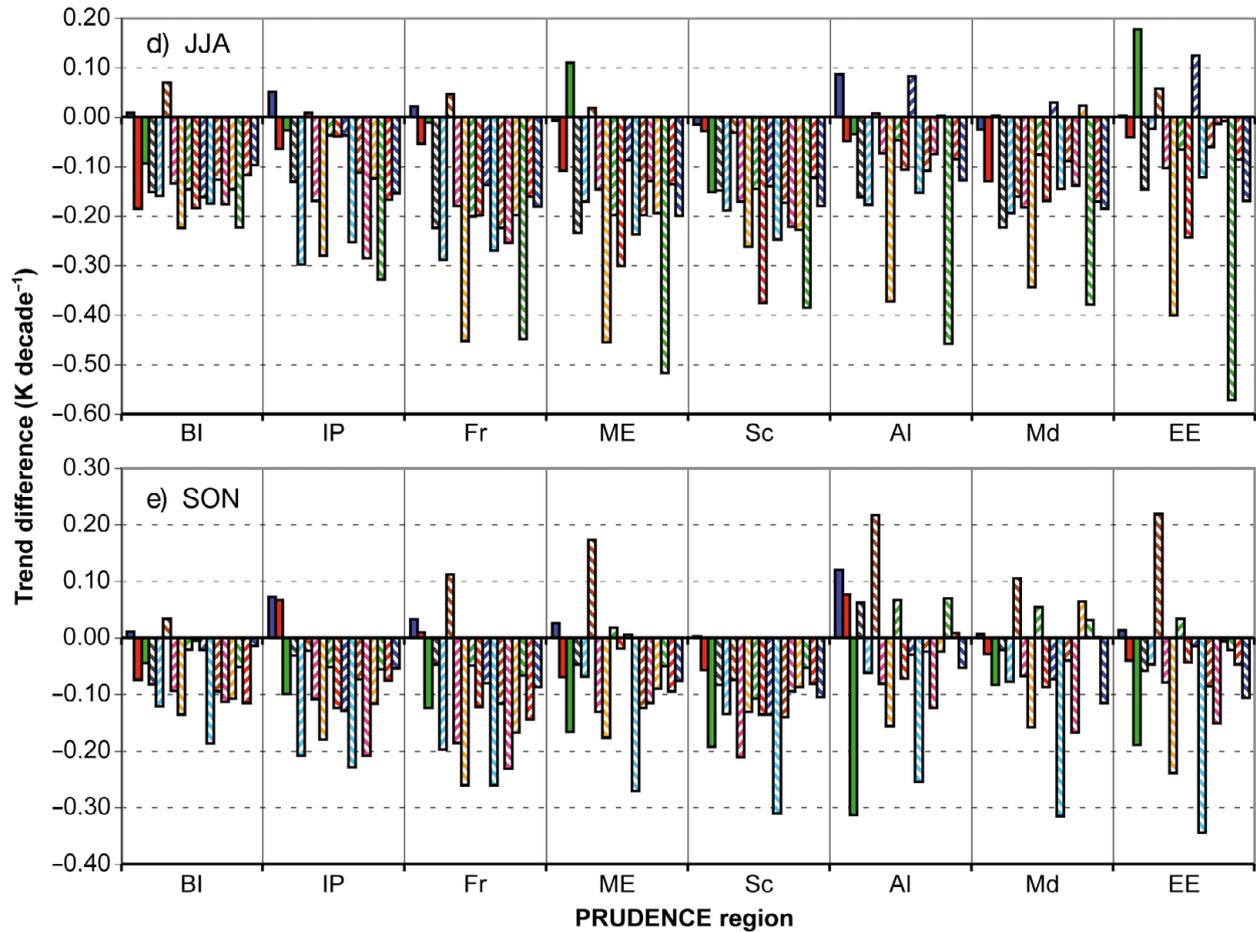


Fig. 3 (continued)

incorporate surface temperature observations, and hence gave an estimate for what the temperature trends would be in a GCM driven by sea surface temperatures (SSTs) and forced to have the observed circulation.

In the ENSEMBLES RCMs, systematic biases in the absolute values of seasonal means are obvious with respect to observations for near-surface temperature (Christensen et al. 2010). However, these kinds of biases are not the subject of the present analysis, which concentrated on trends only. For E-OBS and CRU, no clear trends were visible in Fr within the first half of the analysis period (1960–1980; Fig. 2), while there were clear rising trends for the second half (1980–2000), which could be associated with aerosol decreases in this period (e.g. Ruckstuhl et al. 2008, van Oldenborgh et al. 2009). For the RCMs, very small downward temperature trends were visible for 1960–1980, and like in the observations, a clear positive trend for 1980–2000. Although there were variations, the previously described findings for Fr in principle also hold for all other regions except Sc, where it

was difficult to see a clear trend in the time series plot (Fig. 2). In most regions, a positive trend was visible in the second half of the analysis period (1980–2000) in the observations and in the simulated data.

Linear trends were computed using Eq. (3) for all observational datasets and RCMs, and for all PRUDENCE regions based on the seasonal means as well as on the annual mean values. Fig. 3a shows the computed linear trend differences with respect to the E-OBS reference dataset based on annual mean values. It is remarkable that for all regions, the estimated trends from the individual RCM simulations were weaker than the trends computed for E-OBS, visible in the negative sign of the trend differences to E-OBS (for the different regions, the calculated absolute trend values for E-OBS ranged between 0.20 and 0.34 K decade⁻¹). The computed trends for CRU were stronger in some regions than the trends for E-OBS, and in the other regions they only showed small differences to E-OBS. Furthermore, in all but 1 region (Md for the MPI-M RCM), they were larger than the trends computed for each individual RCM simulation. The trends com-

puted for ERA40 were smaller in most regions than for E-OBS or CRU, but still larger than these for most of the RCMs. The trend differences for NCEP were generally larger than those for ERA40 and of similar magnitude to those computed for the RCMs, which could be due to missing processes in both the RCMs and the NCEP/NCAR reanalysis.

Concerning the 3 realizations done by the Hadley Centre (HC-Q0, HC-Q3, HC-Q16), for all regions the high sensitivity run (HC-Q16) showed larger temperature trends than the normal sensitivity run (HC-Q0), while the latter showed larger trends than the low sensitivity run (HC-Q3). This set of experiments shows that the sensitivity of the model influences the calculated trends.

The trends for the RCMs were similar for some regions, such as BI, while they were quite different—sometimes even with opposite sign of the absolute trend value (not shown)—for other regions, such as EE. All analyzed RCM simulations used the same driving dataset (ERA40) at their lateral boundaries. The climate in Europe is dominated by pressure systems travelling from west to east like in most regions of the mid-latitudes, leading to the strongest influence of the lateral boundary forcing data on the western boundary of the RCM simulations. Furthermore, SSTs from the ERA40 reanalysis were prescribed for all RCMs. Consequently, region BI, as it is close to the western boundary and is surrounded by water, was more strongly influenced by the forcing data than the EE region, which is far away from the western boundary and has less sea areas (see Fig. 1) and a more continental climate, leading to a higher spread between the individual RCM simulations in EE than in BI. This can be summarized as follows: for a continental climate, the ensemble of RCM simulations showed a larger spread in 2 m temperature than for a maritime climate. This was also reported in the PRUDENCE project (Roeckner & Jacob 2008), which was based on an ensemble of GCM-driven simulations.

With regard to the seasonal temperature trends, for winter (DJF; Fig. 3b), the values for linear trends of most RCMs agreed better in general with the trends for E-OBS (visible by smaller magnitudes of the trend differences), CRU, and ERA40. The trends in winter are caused to a large extent by the shift towards a more predominately westerly circulation (e.g. van Oldenborgh et al. 2009). As this circulation is prescribed at the boundaries, the relatively good agreement is not very surprising. However, in winter, the RCMs from CHMI and CNRM produced significantly larger negative trend differences to the observational datasets. The reason for this is still unclear.

For spring (MAM; Fig. 3c), most RCM results again showed significantly smaller trend values than E-OBS

and CRU for all regions (and also smaller than ERA40 for most regions). Here the exception was the ICTP RCM, which showed the largest trends of all RCMs in all regions and exceeded the values for E-OBS and CRU in most regions.

For summer (JJA; Fig. 3d), the trend differences to E-OBS calculated for most RCMs were again negative in many regions. Furthermore, the trend differences for the RCMs were again larger than those for the CRU data, but of similar magnitude to the trend differences for ERA40 data. For many regions, the CNRM RCM results showed only small differences to E-OBS and CRU data. Results from the OURANOS RCM showed rather large negative trend difference values in some regions (the data from the ETHZ RCM showed the same behavior in weaker form). Interestingly, for the PRUDENCE region of Sc, NCEP and all RCMs except the one from CNRM calculated negative absolute trend values, while E-OBS, CRU, and ERA40 data showed positive absolute trends (not shown).

For autumn (SON), E-OBS, CRU, and ERA40 data had small positive absolute trend values for the western regions BI, IP, and Fr, while they had almost no absolute trend for all other regions, except EE, where they had a clear negative absolute trend (not shown). Most of the RCMs in general simulated small negative trend differences compared to E-OBS and CRU data (Fig. 3e). However, in all regions, the CNRM RCM simulated the highest trends of all RCMs, even exceeding the trends from E-OBS and CRU data in all regions except IP. Furthermore, for all regions, the ICTP RCM data included the largest negative trend differences.

The skill scores computed by Eq. (4) depict the absolute linear temperature trend difference of the respective RCM results (or CRU, ERA40, and NCEP data) with respect to the reference dataset E-OBS: the smaller the absolute difference was, the larger was the assigned value for S . As an example, the skill scores based on the annual mean values (Fig. 4a) in general showed the highest values for comparison with CRU data (above 0.9) followed by ERA40 data. For most regions, the RCMs simulated similar values of around 0.7–0.9, with significant lower values of around 0.6–0.8 for Sc for the chosen value of $0.5 \text{ K decade}^{-1}$ for ζ in Eq. (4). The other tested values of $0.1 \text{ K decade}^{-1}$ ($1.0 \text{ K decade}^{-1}$) for ζ did not change the skill score order of the RCMs, but changed the range to about 0.3–0.7 (0.85–0.95) for most regions (not shown).

For the aggregation of the individual skill scores for the 8 PRUDENCE regions, weighting factors were computed by Eq. (5) with a value of 10 K for ξ . They showed moderate variations for the different regions and seasons (Fig. 5). Tests with values of 2 K (50 K) for ξ in Eq. (5) did not change the order of the weighting factors of the different regions, but made the spread

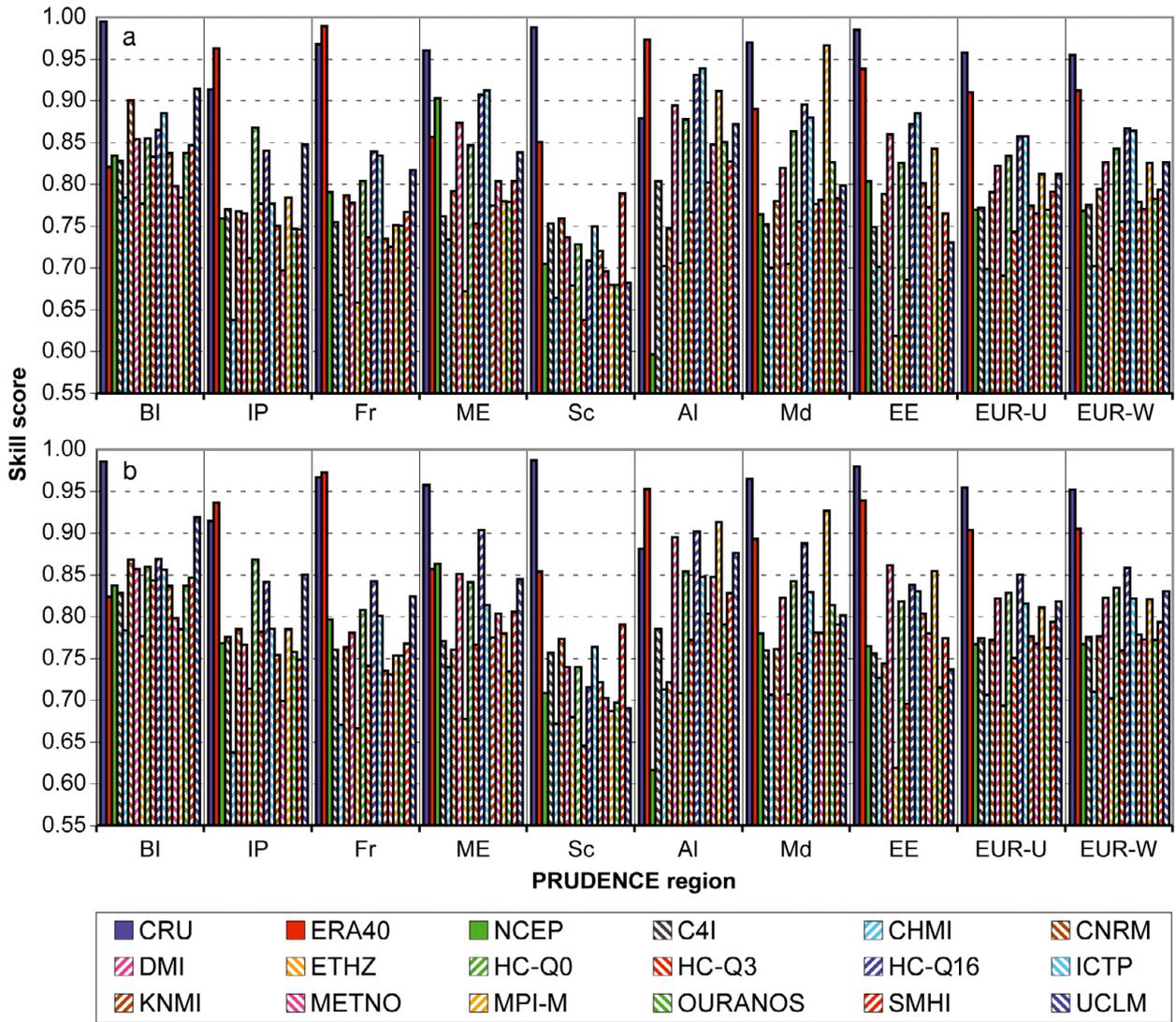


Fig. 4. Skill scores based on (a) annual mean values and (b) combined skill scores for CRU, ERA40, and the ENSEMBLES RCMs for all PRUDENCE regions and aggregated for Europe (EUR-U, unweighted and EUR-W, weighted). The sorting order of the bars for each region corresponds to the sorting order within the key

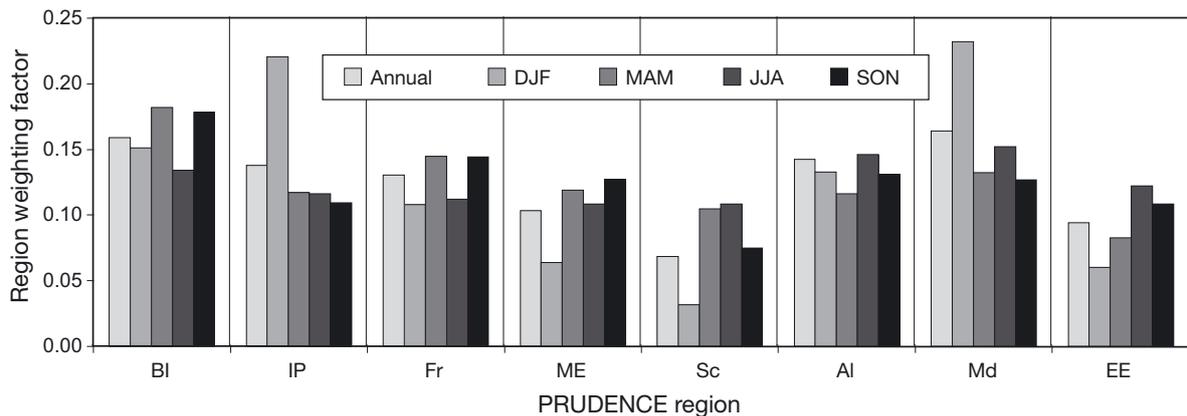


Fig. 5. Region weighting factors of the PRUDENCE regions for the annual mean and all 4 seasons

larger (smaller). The highest regional weighting factors were found for the regions BI, IP, and Md, indicating that for these regions the temperature time series was slightly better represented by a linear trend than for regions like Sc and EE with comparable small area weighting factors. As a consequence of the described relatively small variations in the weighting factors for the different regions, the unweighted European mean skill scores showed very similar values as the weighted European mean skill scores (see columns EUR-U and EUR-W in Fig. 4a). This finding also holds when the region weighting factors were computed with values for ξ (Eq. 5) of 2 K and 50 K instead of 10 K.

As a third method, European skill scores were computed considering the whole of Europe as 1 separate region (not shown). Again, the resulting skill scores were very similar to the simple unweighted (EUR-U) and weighted (EUR-W) European skill scores.

Finally, for all regions and aggregated for Europe, the skill scores based on the annual and on seasonal mean values were combined into overall skill scores by Eq. (6). Given that in this formula, the skill score based on the annual temperatures had a much larger influence than the skill scores based on the seasonal temperatures, it is understandable that the combined skill scores (Fig. 4b) were similar to the skill scores based on the annual values (Fig. 4a). The combined and weighted European skill scores (EUR-W) shown in Fig. 4b were included in the general ENSEMBLES RCM weighting system (Christensen et al. 2010).

5. CONCLUSIONS

The linear trend values based on annual mean near-surface temperatures for 1961–2000 in all analyzed regions for all RCMs (and also the NCEP/NCAR reanalysis) were weaker than in the observational datasets E-OBS and CRU. This is surprising, since all RCMs were driven by quasi-observed lateral boundary conditions and SSTs from the global reanalysis dataset ERA40. The fact that all RCMs underestimated the trends seen by the observational E-OBS dataset implies that no combination of the RCMs with any set of positive weights can replicate the observed trend.

For ERA40 data, the linear trends in 2 m temperatures in most regions showed smaller values than the ones in the E-OBS or CRU data, but in general they were not as small as for the RCMs. The investigation of the reasons for the systematic underestimation of the linear trends in the RCMs is beyond the scope of the present analysis. Nevertheless, having the largest trends in the observational dataset, and the smallest in the RCM results, points to deficiencies in their modeling systems, and even in the reanalysis systems of

NCEP and ERA40. Possible causes include missing processes such as a realistic description of dust aerosols or regional feedback processes. Especially in the summer season, such regional feedback processes include soil moisture exhaustion or decreasing cloud cover, which can impact the temperature developments. Also, potential shortcomings in the observational datasets may play a role and need further investigation.

In the analysis of linear trends based on seasonal mean temperatures, some models showed rather large (or small) trends for all analysis regions for individual seasons, mostly associated with opposite behavior in other season(s). A possible reason for this behavior may be shifts in the phase of the annual cycle of near-surface temperatures in the RCM simulations.

In this study, we used a simple linear trend analysis for 8 large European regions to evaluate the ability of the ENSEMBLES RCMs to realistically simulate the observed trends of near-surface temperatures. This task could also be tackled by other more elaborate methods to investigate the robustness of the results. Furthermore, it would be interesting to analyze trends for smaller sub-regions, which has not been done and which would require reliable observational data sets with a high density in the regions under investigation.

The computed skill scores have been used for inclusion in the ENSEMBLES RCM weighting system (Christensen et al. 2010). Skill scores were of similar magnitude for all analysis regions except for Sc, where most RCMs showed a lower skill score. Nevertheless, the magnitude of the spread of the computed skill scores for different RCMs and different regions is influenced by the choice of the scaling parameters ζ (Eq. 4) and ξ (Eq. 5). Furthermore, the spatially aggregated (temporally combined) skill scores were determined by the formulation of Eq. (5) (Eq. 7), which again has a subjective component. Therefore, these subjective parts in the definition of the presented weighting metric influence the overall model weights of the ENSEMBLES RCM weighting system and should be further investigated.

Acknowledgements. This work was supported by the European Commission Programme FP6 under the contract GOCE-CT-2003-505539 (ENSEMBLES). We thank the global and regional modelers of the project, as well as the developers of the gridded observational dataset (E-OBS). We also thank 3 anonymous reviewers for their valuable comments, which significantly helped to improve this paper.

LITERATURE CITED

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 (Suppl 1):7–30

- Christensen JH, Boberg F, Christensen OB, Lucas-Picher P (2008) On the need for bias correction of regional climate change projections of temperature and precipitation. *Geophys Res Lett* 35:L20709 doi:10.1029/2008GL035694
- Christensen JH, Kjellström E, Giorgi F, Lenderink G, Rummukainen M (2010) Weight assignment in Regional Climate Models. *Clim Res* 44:179–194
- Feldmann H, Früh B, Schädler G, Panitz HJ, Keuler K, Jacob D, Lorenz P (2008) Evaluation of regional climate model precipitation climatology for southwestern Germany. *Meteorol Z* 17:455–465
- Früh B, Feldmann H, Panitz HJ, Schädler G, Jacob D, Lorenz P, Keuler K (2010) Determination of precipitation return values in complex terrain and their evaluation. *J Clim* 23: 2257–2274
- Giorgi F (2002) Variability and trends of sub-continental scale surface climate in the twentieth century. I. observations. *Clim Dyn* 18:675–691
- Haylock MR, Hofstra N, Klein Tank AMG, Klok EJ, Jones PD, New M (2008) A European daily high-resolution gridded dataset of surface temperature and precipitation for 1950–2006. *J Geophys Res* 113:D20119 doi:10.1029/2008JD010201
- Jacob D, Bärring L, Christensen OB, Christensen JH and others (2007) An intercomparison of regional climate models for Europe: design of the experiments and model performance. *Clim Change* 81(Suppl 1):31–52
- Jones PD, Moberg A (2003) Hemispheric and large-scale surface air temperature variations: an extensive revision and an update to 2001. *J Clim* 16:206–223
- Kalnay E, Kanamitsu M, Kistler R, Collins W and others (1996) The NCEP/NCAR 40-year reanalysis project. *Bull Am Meteorol Soc* 77:437–470
- Keup-Thiel E, Göttel H, Jacob D (2006) Regional climate simulations for the Barents Sea region. *Boreal Environ Res* 11:329–339
- Klok EJ, Klein Tank AMG (2009) Updated and extended European dataset of daily climate observations. *Int J Climatol* 29:1182–1191
- Mitchell TD, Jones PD (2005) An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int J Climatol* 25: 693–712
- Roeckner E, Jacob D (2008) Der Klimawandel ist voll im Gange. In: Lozan J, Grassl H, Jendritzky G, Karbe L, Reise K (eds) Warnsignal Klima: Gesundheitsrisiken; Gefahren für Pflanzen, Tiere und Menschen. GEO Wissenschaftliche Auswertungen, Hamburg, p 19–34
- Ruckstuhl C, Philipona R, Behrens K, Coen MC and others (2008) Aerosol and cloud effects on solar brightening and the recent rapid warming. *Geophys Res Lett* 35:L12708 doi:10.1029/2008GL034228
- Simmons AJ, Jones PD (2004) Comparison of trends and low-frequency variability in CRU, ERA-40, and NCEP/NCAR analyses of surface air temperature. *J Geophys Res* 109: D24115 doi:10.1029/2004JD005306
- Solomon S, Qin D, Manning M, Chen Z, and others (eds) (2007) *Climate change 2007: the physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge
- Uppala SM, Kållberg PW, Simmons AJ, Andrae U and others (2005) The ERA-40 re-analysis. *QJR Meteorol Soc* 131: 2961–3012
- van der Linden P, Mitchell JFB (eds) (2009) *ENSEMBLES: Climate change and its impacts: summary of research and results from the ENSEMBLES project*. Met Office Hadley Centre, Exeter
- van Oldenborgh GJ, Drijfhout SS, van Ulden A, Haarsma R and others (2009) Western Europe is warming much faster than expected. *Clim Past Discuss* 4:897–928

Submitted: January 11, 2010; Accepted: November 8, 2010

Proofs received from author(s): November 30, 2010