Vol. 44: 121–134, 2010
doi: 10.3354/cr00940

CLIMATE RESEARCH
Clim Res

Published December 10

Contribution to CR Special 23, 'Regional Climate Model evaluation and weighting'

Model weighting based on mesoscale structures in precipitation and temperature in an ensemble of regional climate models

E. Coppola*, F. Giorgi, S. A. Rauscher, C. Piani

Earth System Physics Section, International Centre for Theoretical Physics, Trieste, Italy

ABSTRACT: We present a weighting scheme specifically designed for regional climate models (RCMs) in that it is based on the model performance in simulating the sub-global climate model (GCM) mesoscale climate signal. The functional form of the weights is based on multiple variables (temperature and precipitation) and metrics (correlation and root mean square error). The weighting scheme is applied to an ensemble of RCM simulations for the European region recently completed as part of the ENSEMBLES project. As a test of the successful implementation of the scheme, the weighting leads to an overall improvement of the performance of the ensemble when measured with the same metrics used in the weighting. The improvement is particularly pronounced over topographically complex regions (e.g. the Alps) in which a larger inter-model range of performance, and thus a more aggressive weighting, is found. When applied to the generation of probabilistic climate change projections, this scheme is designed to be used in conjunction with other RCM weighting metrics developed in the ENSEMBLES project and corresponding weighting schemes for the GCMs driving the regional models.

KEY WORDS: Weighting mesoscale signal · Regional climate model ensemble · European climate

1. INTRODUCTION

The availability of relatively large global and regional multi-model ensemble projections of climate change (e.g. Meehl et al. 2007, Déqué et al. 2005, 2007) requires the development of techniques to compound this information towards the production of regional climate change scenarios (Giorgi 2005). Within this context, there is an ongoing debate regarding whether all models should be treated equally or whether the models should be weighted by some measure of model reliability. Model weighting essentially implies that the information of more reliable models provides a stronger contribution when compounding the full model ensemble. Following this consideration, in the last few years several techniques have been proposed to weight the models according to different performance metrics and methodological approaches (e.g. Giorgi & Mearns 2002, 2003, Murphy et al. 2004, Tebaldi et al. 2005, Piani et al. 2007, Tebaldi & Knutti 2007).

Giorgi & Mearns (2002) first introduced the reliability ensemble averaging (REA) method, in which climate model ensemble members were weighted based on their bias and distance from the ensemble mean. In this first important step towards ensemble weighting, the weight applied to single ensemble members was dependent on a single variable (temperature or precipitation). More recently, Xu et al. (2010) upgraded the REA method to account for multiple metrics, variables and statistics in the definition of the model weights. In Murphy et al. (2004) and Piani et al. (2007), global climate models (GCMs) were weighted using a broad range of normalized variables and the associated root mean square error (RMSE) relative to available observations. In particular, Murphy et al. (2004) produced a probabilistic forecast for global climate sensitivity

*Email: coppolae@ictp.it

whereas Piani et al. (2007) produced probabilistic forecasts of regional values of temperature and precipitation changes. As a final example of weighting approach, methods using a Bayesian approach were proposed by Tebaldi et al. (2005) and Tebaldi & Knutti (2007).

Most of the previous model-weighting methods have been applied to ensembles of GCM simulations. However, as part of the EU project ENSEMBLES (http://ensembles-eu.metoffice.com/) an ensemble of regional climate model (RCM) projections were completed for the European region (see Section 3). A component of the project consisted of the development of weighting schemes specifically designed for RCMs that could be used to produce RCM-based probabilistic climate change forecasts. Different groups produced different techniques — described separately in the papers of this special issue — and then the different weighting metrics were compounded to obtain a final weight, as described in Christensen et al. (2010, this Special), and to produce probabilistic forecasts, as described in Déqué & Somot (2010, this Special).

Within this context, in the present study we describe a weighting approach specifically aimed at the use of ensembles of RCMs, where the performance metrics used to calculate the weights are based on variables for which the RCMs are expected to provide added value compared to GCMs. The weighting procedure is then implemented within the ENSEMBLES RCM set of models. More specifically, we used the ensemble of simulations driven by ERA40 reanalysis (Uppala et al. 2005) fields to determine the performance metrics and the associated weights. We then assessed to what extent the weighting procedure influences the ensemble performance of the models. This was done in order to test the implementation of the scheme, as it should be expected that a weighted model performance should improve, particularly for those metrics used to generate the weights. Further, we performed a number of case studies in order to explore the sensitivity of the results to the method of deriving the weights. Our weights are compounded with others by Christensen et al. (2010) and used in a climate projection context by Déqué & Somot (2010).

2. THE WEIGHTING METHOD

The method developed here is based on metrics for which the use of RCMs is expected to provide significant added value with respect to coarse-scale GCMs. Following this objective, the metrics are based on sub-GCM grid-scale processes that can be resolved by RCMs. In order to define such metrics, we thus first decomposed the RCM signal into a large-scale component and a mesoscale signal as done, for example, in Giorgi et al. (1994).

The large-scale component was identified by carrying out a 9×9 grid point running spatial average of the RCM fields. Because the RCM grid spacing is ~25 km, this yields a signal roughly at a scale of 200 to 250 km, which is typical of present-day global coupled climate models (GCMs). The mesoscale signal was then obtained by simply subtracting the calculated large-scale component from the full RCM fields. This generates an anomaly field in which the large-scale component is filtered out. Performance metrics based on the mesoscale signal are thus essentially measures of the 'added value' due to the increased resolution of RCMs compared with present-day GCMs in response, for example, to sub-GCM grid-scale topography and coastlines. In order to calculate the performance metrics, the mesoscale signal was calculated for both the RCMs and suitable observations.

We identified the following 5 functions that measure the model mesoscale performance for temperature and precipitation:

$$g_1 = \mathrm{R}(p_{\mathrm{mod}}, p_{\mathrm{obs}}) \qquad (1)$$

$$g_2 = \mathrm{R}(T_{\mathrm{mod}}, T_{\mathrm{obs}}) \qquad (2)$$

$$g_3 = \sigma(p)_{\mathrm{obs}} / \mathrm{RSME}(p) \qquad (3)$$

$$g_4 = \sigma(T)_{\mathrm{obs}} / \mathrm{RSME}(T) \qquad (4)$$

$$g_5 = [1 - (|\mathrm{R}(p_{\mathrm{obs}}, T_{\mathrm{obs}}) - \mathrm{R}(p_{\mathrm{mod}}, T_{\mathrm{mod}})|/2] \qquad (5)$$

Each function was calculated separately for the 4 seasons: December–January–February (DJF), March–April–May (MAM), June–July–August (JJA) and September–October–November (SON). In Eqs. (1–5), R is the spatial correlation coefficient between the observed and simulated mean mesoscale signal calculated for the 4 seasons over a preselected region of interest (e.g. Europe) and for a given time period (e.g. the ERA-40 period, 1961–2000). The symbols p and T represent seasonally averaged mesoscale components of precipitation and temperature, respectively; RMSE is the root mean square error between simulated and observed mean mesoscale signals calculated in space over the preselected region after averaging the fields over the selected time period; and σ is a measure of the interannual variability of the observed mesoscale signal. To obtain σ, the mean mesoscale signal was first calculated at each grid point for every season of each year of the selected time period. This time series of seasonal values was then used to compute the interannual standard deviation at each grid point, which was then averaged over all the grid points in the domain of interest to yield σ. In Eqs. (3) & (4), if the RMSE is lower than the observed interannual variability of the same variable (which is taken as a measure of observation uncertainty), then the values of g_3 and g_4 were set equal to 1 (i.e. the model is perfect in that metric). In other words, σ is used as a scaling metric to yield a non-dimensional value. The last function (Eq. 5) measures the ability of the models to simulate the spatial correlation between precipitation and temperature mesoscale signals, with R having the same meaning as in g_1 and g_2 except that it now measures the spatial correlation between temperature and precipitation mesoscale signals. The specific functional form of g_5 is such that it yields values between 0 and 1.

Therefore, the set of functional metrics include multiple spatial statistics (spatial correlation, RMSE) and multiple variables (temperature and precipitation). The rationale for the choice of the 5 specific functions is the following: g_1 and g_2 (for p and T, respectively) measure the model's ability to reproduce the observed spatial patterns of the mesoscale signal (e.g. as affected by complex topography), which is a key aspect of the RCM performance; g_3 and g_4 (for p and T, respectively) measure the overall model performance (as measured by the RMSE) in quantitatively reproducing the magnitude and sign of the signal; and g_5 measures the model's ability to reproduce the (spatial) interconnections between the temperature and precipitation mesoscale signals.

Note that the functions g_1–g_5 are all normalized to yield non-dimensional values between 0 and 1 over a preselected domain, in our case the common ENSEMBLES RCM domain.

Given the functional metrics g_1–g_5, in its general form, the weight for a given model i is given by:

$$w_i = g_1^j \times g_2^j \times g_3^j \times g_4^j \times g_5^j \quad (6)$$

where the exponent j can be used to give more weight to one metric than the others. From these weights, the mean simulated value of a variable $\overline{X}$ (temperature or precipitation) obtained from the ensemble of models is given by:

$$\overline{X} = \sum_i w_i X_i \Big/ \sum_i w_i \quad (7)$$

where X_i is the value of the variable for model i. A few points should be stressed. In Eq. (6), any of the functions g_1–g_5 can be set to 1 to remove a particular metric from the weight. Also, from Eq. (7), it is clear that the weighting metrics can all be scaled by a common multiplicative factor without changing the weighting process. Also note that, following Giorgi & Mearns (2002), the weight is given by the product of all performance functional metrics, which implies a stringent test on model performance, i.e. in order to have a high weight, a model needs to perform well in all metrics. Finally, the weights are not calculated locally, by construction, but over a pre-selected region.

In order to test the weighting procedure, we calculated the values of the functions g_1–g_5 as well as the weights for 15 ENSEMBLES RCMs by comparing the mesoscale signals in the ERA40-forced RCM simulations with corresponding observed values. All quantities were calculated for the seasonal averages for DJF, MAM, JJA and SON. In addition, to explore the sensitivity of the approach to different combinations of weighting functions, we tested 5 cases of different combinations of g_1–g_5 as described in Table 1. In Case 1, all the g_1–g_5 functions are used; in Cases 2a and 2b, only the precipitation-based functions or temperature-based functions are included; in Case 3, the function g_5 is removed; and in Case 4, functions g_1 and g_2 provide a stronger contribution to the weight than the others.

Finally, we calculated the weights for the European common region across the ENSEMBLES models and observations and for a subregion including only the Alpine area, where the mesoscale signal is expected to be particularly strong.

3. MODEL AND EXPERIMENTS

In the present study we used results from 15 RCMs participating in the ENSEMBLES project. The list of models and originating research centers is given in Table 2. A more detailed specification of the RCMs, including references, can be found in Christensen et al. (2010). The simulations cover the period 1961–2000 and the meteorological initial and lateral boundary conditions were obtained from the ERA40 reanalysis product. The domain covers the entire European region with a resolution of approximately 25 km.

In order to calculate the functions g_1–g_5, observations are needed at a resolution similar to that of the models. For this purpose we used the observed monthly data set for Europe at 10 minute resolution (0.16°) developed by the Climate Research Unit (CRU) at the University of East Anglia (Mitchell et al. 2004) (data set version CRU TS1.2). The observed

Table 1. The different function combinations used to obtain the weights

Case	Function combination
1	$g_1 \times g_2 \times g_3 \times g_4 \times g_5$
2a	$g_1 \times g_3$
2b	$g_2 \times g_4$
3	$g_1 \times g_2 \times g_3 \times g_4$
4	$(g_1)^2 \times (g_2)^2 \times g_3 \times g_4$

Table 2. The 13 institutes and the 15 regional climate models used

Institute	Model
Météo-France (CNRM)	ALADIN
Swiss Institute of Technology (ETHZ)	CLM
The Abdus Salam Intl. Centre for Theoretical Physics (ICTP)	RegCM3
The Royal Netherlands Meteorological Institute (KNMI)	RACMO2
UK Met Office, Hadley Centre for Climate Prediction and Research (HC)	HadRM3Q0
UK Met Office, Hadley Centre for Climate Prediction and Research (HC)	HadRM3Q3
UK Met Office, Hadley Centre for Climate Prediction and Research (HC)	HadRM3Q16
OURANOS	CRCM
The Community Climate Change Consortium for Ireland (C4I)	RCA3
Max-Planck-Institute for Meteorology (MPI)	REMO
Swedish Meteorological and Hydrological Institute (SMHI)	RCA
Universidad de Castilla La Mancha (UCLM)	PROMES
Danish Meteorological Institute (DMI)	HIRHAM5
The Norwegian Meteorological Institute (METNO)	HIRHAM
Czech Hydrometeorological Institute (CHMI)	ALADIN

data set was interpolated onto the 25 km common ENSEMBLES model grid. We note that an additional observation data set (called E-OBS) was produced as part of ENSEMBLES, although it was not yet available when we carried out this work. However, a careful comparison of the CRU and E-OBS data sets is given in Rauscher et al. (2010) who concluded that, although the 2 data sets differ in horizontal extent and station density, their climatologies are very close to each other and would thus lead to similar conclusions. The authors also stress that the methodology presented here can be applied using any set of relevant observations.

4. RESULTS

Figs. 1 & 2 compare the RCM ensemble mean and observed mesoscale signals over the European region for DJF and JJA surface air temperature and precipitation. In both cases, the mesoscale signal is essentially tied to the main European topographical features. For temperature, it is negative in correspondence with the mountain peaks and positive in the surrounding valley areas as a consequence of the relatively smooth orography used by the models at the present horizontal resolution. Therefore, for example, it attains large negative values over the Alpine chain and positive values over the Po valley of northern Italy. For both the European and the Alpine regions, the model ensemble captures the observed spatial pattern of the mesoscale signal ($R \approx 0.92$ to 0.99) (Table 3). This can be expected in view of the fact that the mesoscale temperature signal, on average, is a linear function of elevation that roughly follows the atmospheric temperature lapse rate.

Topography is also a key factor in modulating the precipitation mesoscale signal (Fig. 2). For example, positive values are found in the mountain areas in southern Norway and Scotland and the upwind (westward) slope of the Alps, Carpathians and Balkan massifs in DJF and, in general, over all mountainous areas in JJA. The model ensemble generally reproduces this pattern, although the correlation with the observed mesoscale signal is lower, 0.68 to 0.75 for the whole European region and 0.58 to 0.78 for the Alpine region, with minimum values in MAM and maximum values in JJA (Table 4). Overall, the RCM ensemble reproduces reasonably well the observed spatial patterns of the mesoscale signal for seasonal temperature and precipitation (Figs. 1 & 2, Tables 3 & 4). In addition, biases and RMSE of the mesoscale signals are generally small, for both temperature and precipitation (Tables 3 & 4).

Fig. 3 shows the values of the full weight (Eq. 6) for Case 1, for all models and seasons. The distribution of weights across models shows relatively uniform values except for 2 models (3 and 4) that exhibit weights that are much larger than all the others. This indicates that these 2 models overall have a substantially better performance in simulating the mesoscale signal as measured by the metrics of Eqs. (1) to (5). If the full weight is used, these models would be the greatest contributors to the ensemble. Intermediate weights are found for Models 1, 2, 8, 10, 12 and 15, whereas 6 models have relatively low weights (5, 6, 7, 9, 13 and 14). The inter-model distribution of weights thus appears substantially skewed. The weights also show substantial variations across seasons, with the largest values (and thus better model performance) generally found in summer.

A useful diagnostic when analyzing model weighting is the effective number (N_{eff}) of models (Xu et al. 2010), defined as:

$$N_{eff} = 1 \Big/ \sum_{i=1}^{N} P_i^2 \tag{8}$$

where P_i is defined as:

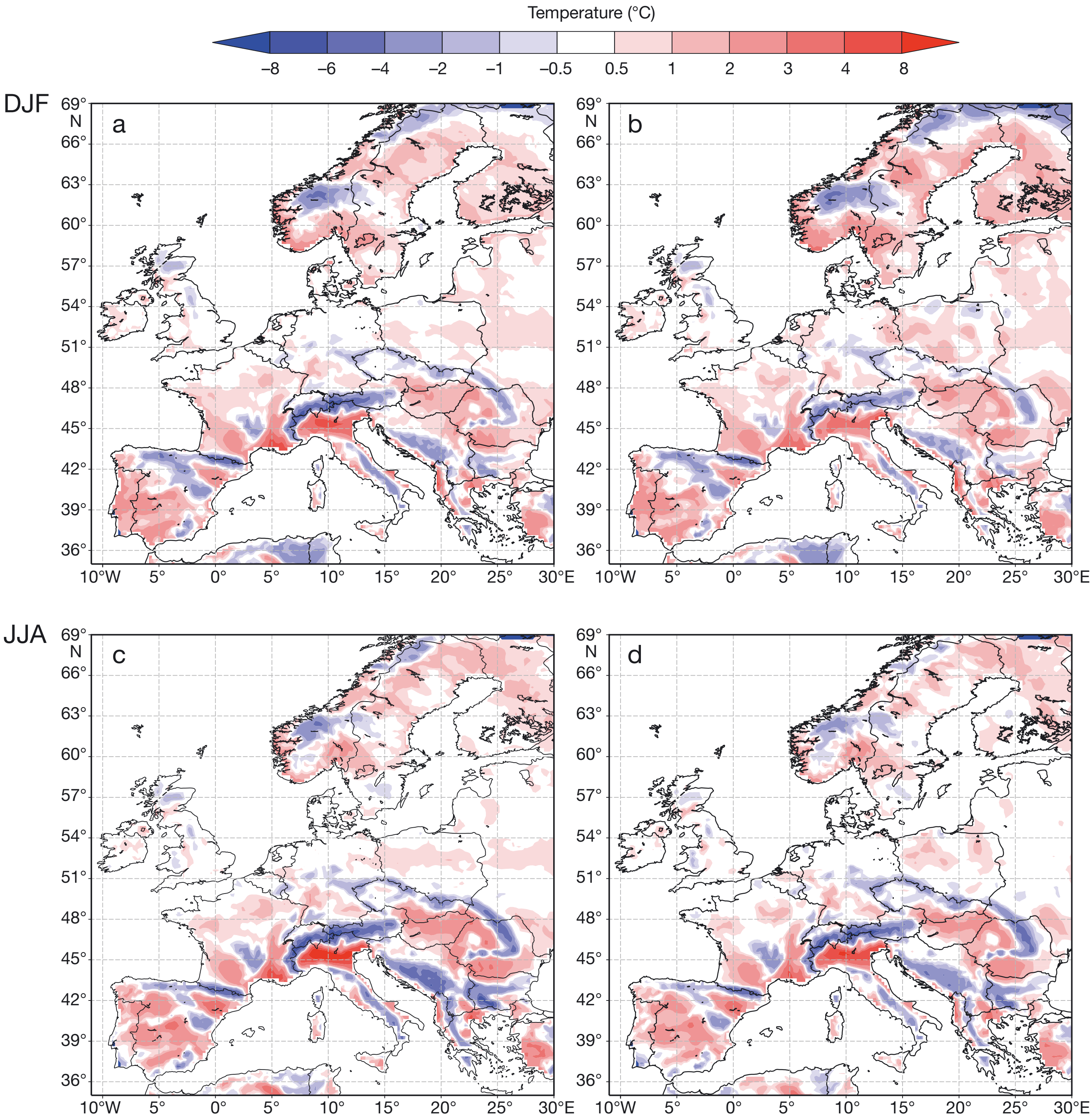

Fig. 1. (a,c) 15 model means and (b,d) CRU temperature mesoscale signals for DJF and JJA, respectively

$$P_i = \frac{w_i}{\sum_{j=1}^{N} w_j} \tag{9}$$

where j indicates the model.

For uniform weighting ($P_i = 1/N$ for all models, where N is the number of models), $N_{eff} = N$; for non-uniform weights, $N_{eff} < N$ and the lower the value of N_{eff}, the more aggressive the weighting. Also note that if one model has a zero weight and the remaining models have equal weights, then $N_{eff} = N - 1$.

The values of N_{eff} are reported in Tables 3 & 4. In the tables, the best score for each case and for each of the statistical indices is highlighted in bold. For all seasons

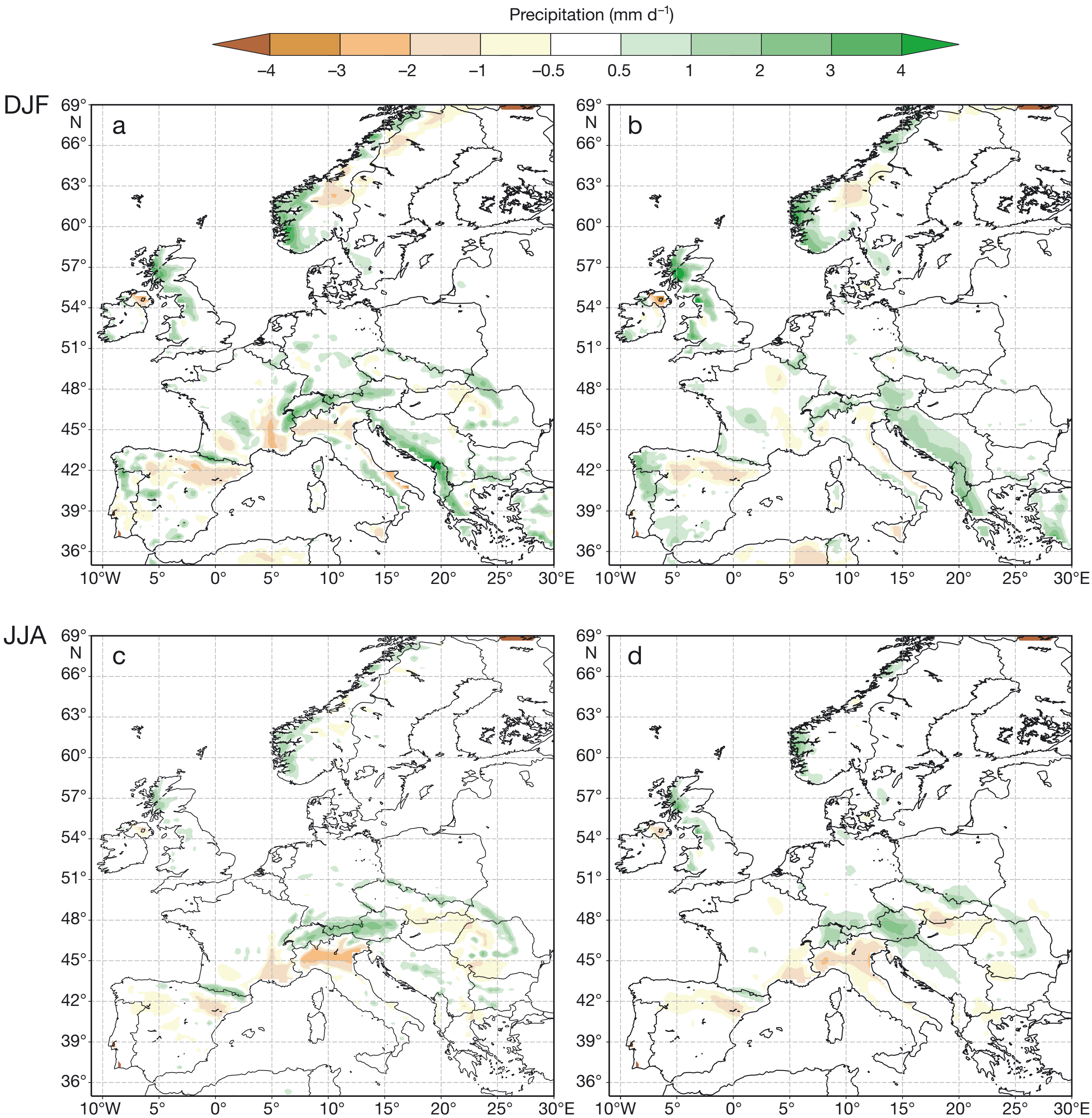

Fig. 2. (a,c) 15 model means and (b,d) CRU precipitation mesoscale signals for DJF and JJA, respectively

and for both domains the best scores are usually associated with a lower value of N_{eff}. This is because for lower N_{eff} values, the better models provide stronger contributions to the ensemble.

In order to illustrate how the 5 functions g_1–g_5 contribute to the intermodel spread of weights, the values of each individual function for each model in DJF and JJA were plotted (Fig. 4). The functions g_2 and g_5 show the largest values in both seasons, especially in JJA, and the values of g_2 are relatively uniform across models. The lowest values are found for the functions g_3 and g_4, which are both measures of the model perfor-

Table 3. Bias, spatial root mean square error (RMSE) (mm d^{-1}) and spatial correlation (R) of the 15-model and 15-model weighted mean mesoscale temperature (Cases 1–4) compared with CRU observed temperature mesoscale signals. All indices are computed for the whole European region and for the small Alpine region. The effective model number (N_{eff}) is indicated for both regions. **Bold**: best results

Season	Europe				Alps			
	Bias	RMSE	R	N_{eff}	Bias	RMSE	R	N_{eff}
Mean no. weights								
DJF	**–0.107**	0.611	0.922	15	–0.279	1.153	0.961	15
MAM	**–0.010**	0.388	0.975	15	–0.061	0.787	0.989	15
JJA	0.059	0.493	0.966	15	0.029	0.741	0.990	15
SON	**–0.023**	0.382	0.968	15	**–0.205**	0.714	0.986	15
Case 1								
DJF	–0.110	**0.605**	**0.923**	13.6	–0.301	1.013	**0.964**	9.8
MAM	–0.010	0.368	**0.976**	12.2	**–0.042**	0.639	**0.991**	10.5
JJA	0.052	0.476	**0.968**	12.6	0.006	0.630	0.991	11.9
SON	–0.030	0.374	**0.969**	12.7	–0.215	0.628	**0.988**	11.5
Case 2a								
DJF	–0.110	0.609	0.921	13.9	–0.307	1.103	0.961	12.1
MAM	–0.010	0.381	0.974	13.5	–0.050	0.742	0.988	13.1
JJA	0.059	0.492	0.966	13.7	0.025	0.768	0.988	13.2
SON	–0.025	0.376	0.968	13.7	–0.210	0.664	0.987	13.2
Case 2b								
DJF	–0.108	0.609	0.922	14.8	**–0.267**	1.058	**0.964**	14.0
MAM	–0.014	0.377	**0.976**	14.4	–0.055	0.691	**0.991**	13.6
JJA	0.052	0.480	**0.968**	14.4	0.016	**0.624**	**0.992**	13.4
SON	–0.027	0.380	**0.969**	14.6	–0.212	0.670	**0.988**	14.4
Case 3								
DJF	–0.112	0.607	0.922	13.7	–0.296	**1.005**	**0.964**	9.9
MAM	–0.015	0.369	**0.976**	12.5	–0.045	0.644	**0.991**	10.8
JJA	0.052	0.475	**0.968**	12.6	0.006	0.627	0.991	11.7
SON	–0.030	0.374	**0.969**	12.8	–0.216	0.628	**0.988**	11.6
Case 4								
DJF	–0.115	0.606	0.922	12.8	–0.309	**1.005**	**0.964**	8.7
MAM	–0.017	**0.367**	**0.976**	11.2	–0.048	**0.638**	**0.991**	9.9
JJA	**0.051**	**0.474**	**0.968**	11.3	**0.003**	0.632	0.991	10.8
SON	–0.032	**0.373**	**0.969**	11.4	–0.220	**0.625**	**0.988**	10.6

mance in reproducing magnitude and sign of the signal. It is worth noting that no model shows the best performance in all metrics. The relatively high weights of Models 3 and 4 are mostly based on relatively higher values of the functions g_3 and g_4, along with relatively high values of g_1 (precipitation mesoscale signal). Thus, in general the weights appear to be more sensitive to the precipitation than the temperature performance metrics.

In order to allow an assessment of the effect of the weighting, Figs. 5 & 6 compare the temperature and precipitation mesoscale signals as obtained from the standard unweighted average and the weighted average using the weights of Case 1. As expected from the fact that the mesoscale signal is primarily induced by topography, the differences occur mainly in correspondence with the main mountainous systems, particularly the Alps. They are on the order of a few tenths of a degree for temperature and a few tenths of a millimeter per day for precipitation and do not show any systematic behavior, i.e. they are both positive and negative.

Given that the region where the mesoscale-based weighting appears most relevant is the Alps, we repeated our calculations only for the Alpine region (e.g. as shown in Fig. 8). Because, as mentioned above, the weights are calculated over a preselected region rather than locally, they need to be recalculated for the Alps. Fig. 7 shows the weights (Case 1) for the different models and seasons. As can be seen, a greater inter-model variability in the calculated weights is found compared to the whole European region case. Models 3 and 4 still have relatively high weights, but relatively high weights were also found for Models 1, 8 and 15 (for 2 seasons). Conversely, the lowest weights were found for Models 2, 5, 6 and 14. Overall, the magnitude of the weights varies by a factor of ~5. Fig. 8, which shows the contributions of the different functional metrics for the Alpine region, indicates that again functions g_2 and g_5 are relatively uniform across models, whereas most of the inter-model variability is due to functions g_2, g_3 and g_4.

The number of effective models was also calculated for this region (Tables 3 & 4). Again, we observed a correspondence of the best scores with the lowest value of N_{eff} but, for the same case and the same season, the value of N_{eff} over the Alpine domain is always lower than that obtained for the whole European region. This is because of the greater inter-model performance over the Alps than the whole European region, i.e. because of the more aggressive weighting that resulted over the Alps.

Figs. 9 & 10 compare the mesoscale signal over the Alpine region as obtained from the unweighted and weighted (Case 1) ensemble average for temperature and precipitation in DJF and JJA. As noticed earlier, the effect of the weighting is stronger over the Alps than the rest of the domain, with differences between the weighted and unweighted means reaching almost 1°C for temperature and 1 mm d^{-1} for precipitation.

Some systematic differences can be observed over the Alpine region between the 2 averaging processes.

Table 4. Bias, spatial root mean square error (RMSE) (mm d^{-1}) and spatial correlation (R) of the 15-model and 15-model weighted mean mesoscale precipitation (Cases 1–4) compared with CRU observed precipitation mesoscale signals. All indices are computed for the whole European region and for the small Alpine region. The effective model number (N_{eff}) is indicated for both regions. **Bold**: best results

Season	Europe				Alps			
	Bias	RMSE	R	N_{eff}	Bias	RMSE	R	N_{eff}
Mean no. weights								
DJF	−0.024	0.533	0.725	15	0.068	0.977	0.612	15
MAM	−0.022	0.401	0.683	15	0.138	0.965	0.585	15
JJA	−0.033	0.339	0.751	15	0.083	0.789	0.783	15
SON	−0.045	0.466	0.743	15	**−0.062**	0.820	0.676	15
Case 1								
DJF	−0.025	**0.499**	**0.741**	13.6	0.036	0.796	0.648	9.8
MAM	−0.022	**0.356**	0.701	12.2	0.121	**0.824**	0.594	10.5
JJA	−0.034	0.320	0.767	12.6	0.071	0.732	0.795	11.9
SON	−0.043	0.438	0.765	12.7	−0.089	0.721	0.696	11.5
Case 2a								
DJF	−0.026	0.500	0.738	13.9	0.039	0.827	0.644	12.1
MAM	−0.022	0.365	0.701	13.5	**0.118**	0.864	0.594	13.1
JJA	−0.034	0.322	0.766	13.7	**0.053**	0.741	0.796	13.2
SON	−0.044	0.443	0.760	13.7	−0.082	0.749	0.691	13.2
Case 2b								
DJF	**−0.023**	0.541	0.723	14.8	0.066	0.955	0.613	14.0
MAM	**−0.021**	0.403	0.682	14.4	0.140	0.928	0.586	13.6
JJA	**−0.032**	0.340	0.748	14.4	0.107	0.789	0.779	13.4
SON	−0.045	0.464	0.743	14.6	−0.069	0.787	0.681	14.4
Case 3								
DJF	−0.024	0.504	0.737	13.7	0.033	0.783	0.646	9.9
MAM	**−0.021**	0.362	0.704	12.5	0.121	0.828	0.595	10.8
JJA	−0.033	0.319	0.768	12.6	0.074	0.731	0.796	11.7
SON	−0.044	0.440	0.762	12.8	−0.089	0.719	0.697	11.6
Case 4								
DJF	**−0.023**	0.502	**0.741**	12.8	**0.031**	**0.771**	**0.656**	8.7
MAM	**−0.021**	0.359	**0.707**	11.2	0.121	0.826	**0.598**	9.9
JJA	−0.034	**0.316**	**0.772**	11.3	0.062	**0.726**	**0.798**	10.8
SON	**−0.042**	**0.435**	**0.768**	11.4	−0.088	**0.715**	**0.700**	10.6

In terms of temperature, the weighting produces higher temperatures over the mountain peaks and lower temperatures over the surrounding valleys, especially the Po valley. For precipitation, the weighting produces lower values at high elevations and higher values at low elevations for DJF, whereas a noisier pattern is found for JJA. These systematic differences are due to the fact that a smaller number of models (N_{eff} value) with common characteristics dominate the averaging process.

All our results are summarized in Tables 3 (temperature) & 4 (precipitation). One should expect that the bias and RMSE would be lower for the weighted than the unweighted case, whereas the correlation should be higher, i.e. the weighting should lead to an improved performance by the ensemble of models in simulating the mean mesoscale signal, given that the metrics used to measure the improvement are also used for the weighting. This is indeed true in the vast majority of cases for all seasons, variables and weighting schemes, showing that the weighting method has been properly implemented and provides the results expected. The improvement by the model weighting is especially evident in the Alpine region, particularly for the bias and RMSE of both temperature and precipitation.

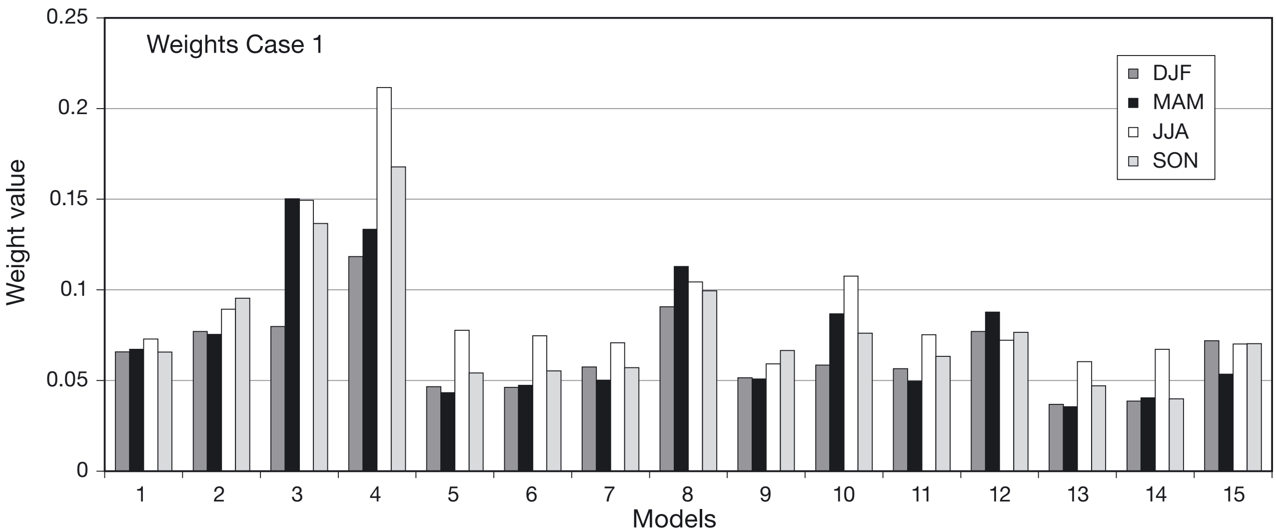

Fig. 3. Weight values for all 15 models and for all 4 seasons for Case 1

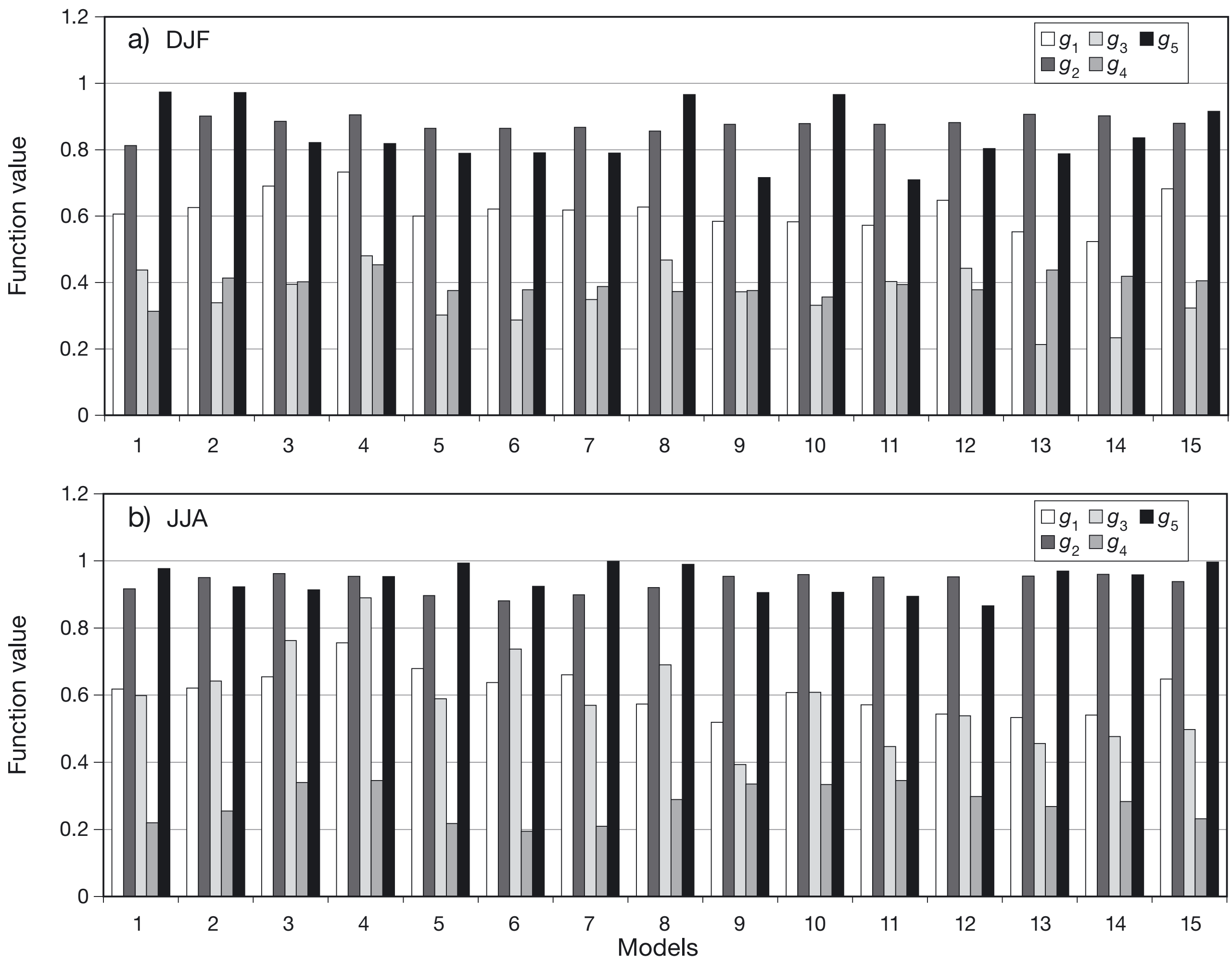

Fig. 4. Function values for all 15 models for (a) DJF and (b) JJA

5. CONCLUDING REMARKS

In the present study we presented a model-weighting method based on the performance of RCMs in simulating various statistics of the sub-GCM grid-scale mesoscale signal of temperature and precipitation. The method was applied to the ensemble of 15 RCMs participating in the ENSEMBLES EU project, using simulations for the period 1961–2000 and driven at the lateral boundaries by ERA40 reanalyses. The model weights are based on 5 functional metrics that measure the model's ability to simulate the spatial patterns of the mesoscale signal, its magnitude/sign and the covariance between temperature and precipitation mesoscale signals. Calculations were performed for both the entire European region and for the Alpine sub-region.

Results indicate that there is a substantial variability in weights, and thus performance, across models, especially for the Alpine region, mostly deriving from the magnitude/sign and precipitation-based functional metrics. This factor is measured by the effective number of models (N_{eff}), which is ~12 for Europe and smaller, ~10, for the Alps. As expected, the weighted mean leads to a general improvement of the simulation of the mesoscale signal compared with the unweighted mean, especially over the Alps, showing a good implementation and performance of the weighting procedure. Indeed, in general, this weighting method has its largest effects in areas characterized by complex topography, which exerts a strong forcing at the sub-GCM scale. We also tested the sensitivity of the weighting procedure to different combinations of the 5 functional metrics.

Our weighting scheme is specifically designed for use in ensembles of RCM simulations, because it is based on the model performance of those aspects that provide an added value with respect to the driving global model fields. This weighting scheme is compounded with others in ENSEMBLES by Christensen et al. (2010) and is used to produce probabilistic forecasts by Déqué

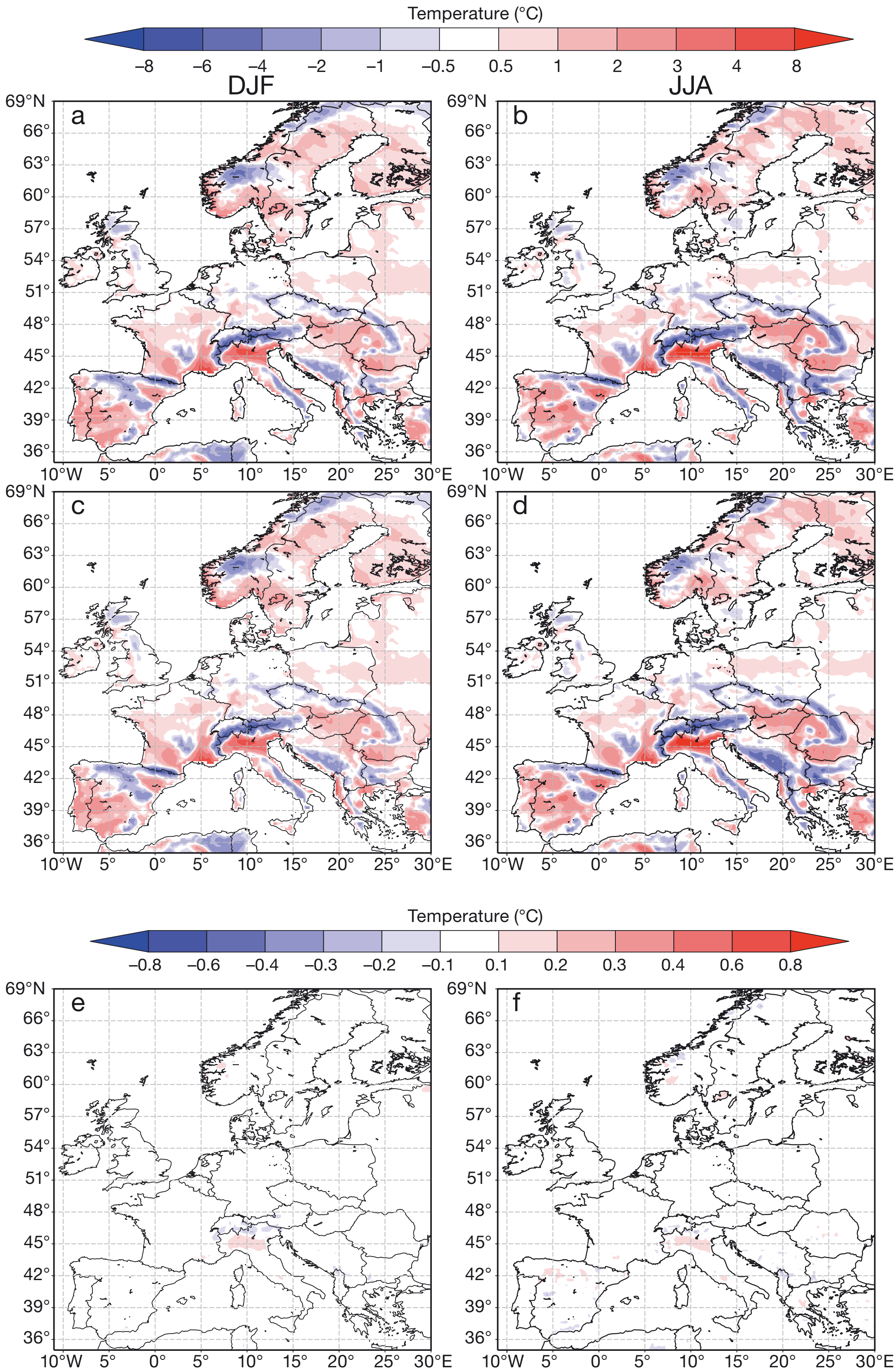

Fig. 5. Mesoscale temperature model mean and mesoscale temperature-weighted mean for (a,c) DJF and (b,d) JJA. The difference between the simple mean and the weighted mean for (e) DJF and (f) JJA is also shown

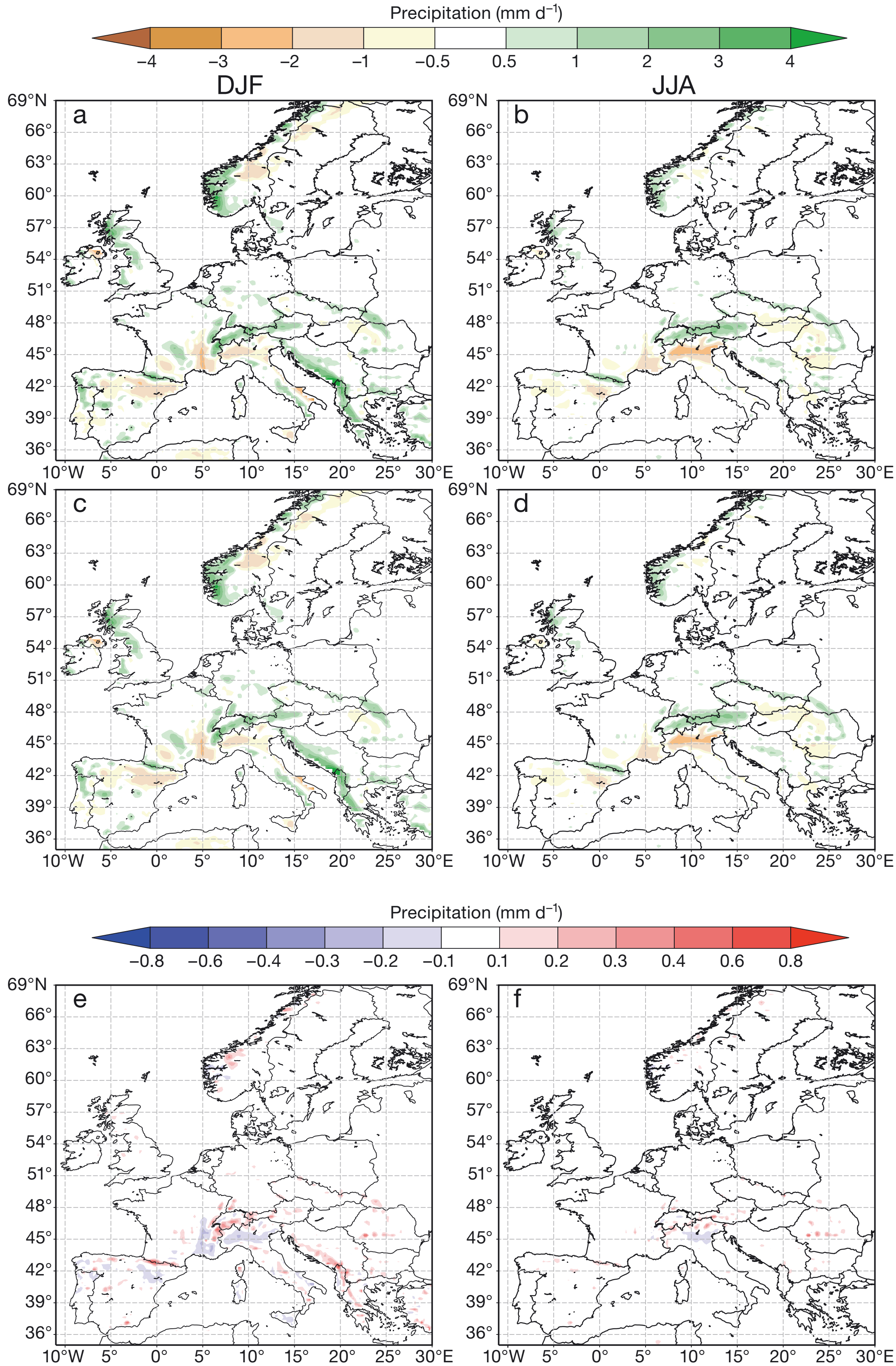

Fig. 6. Mesoscale precipitation model mean and mesoscale precipitation-weighted mean for (a,c) DJF and (b,d) JJA. The difference between the simple mean and the weighted mean for (e) DJF and (f) JJA is also shown

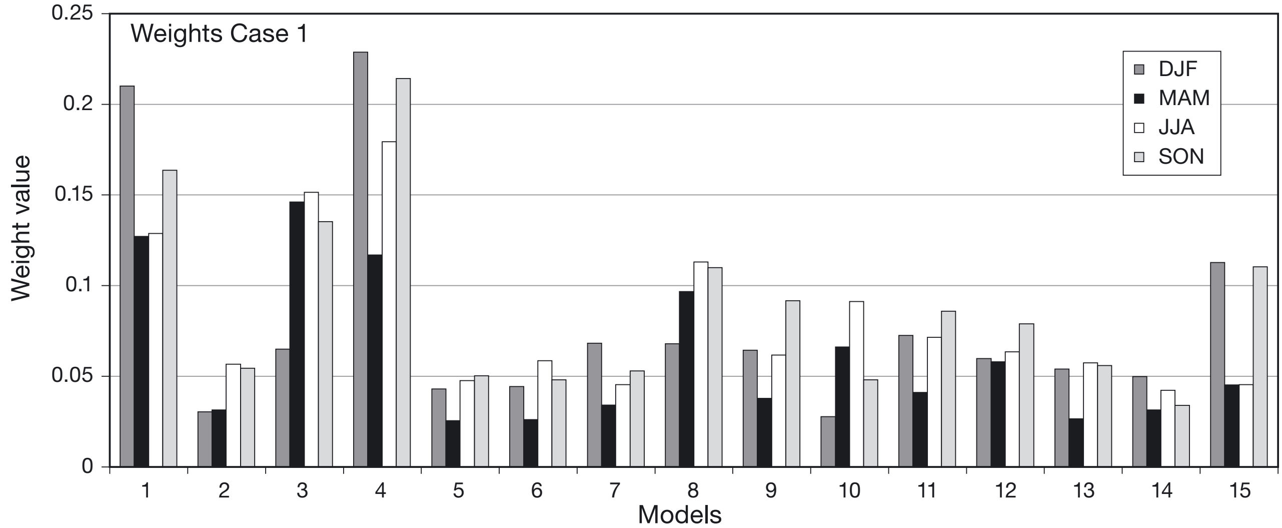

Fig. 7. Weight values for all 15 models and for all 4 seasons for Case 1 for the Alps region

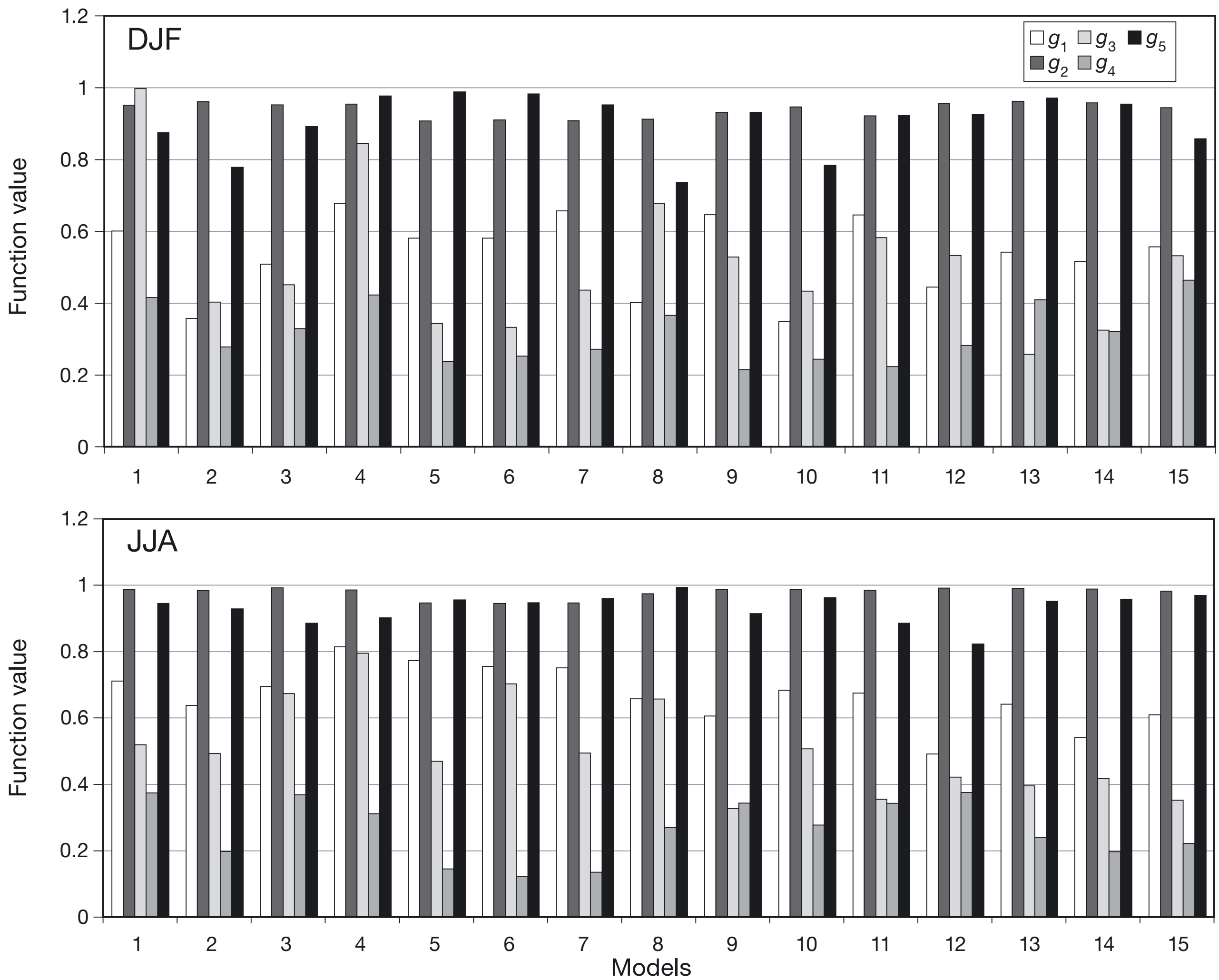

Fig. 8. Function values for all 15 models for (a) DJF and (b) JJA for the Alps region

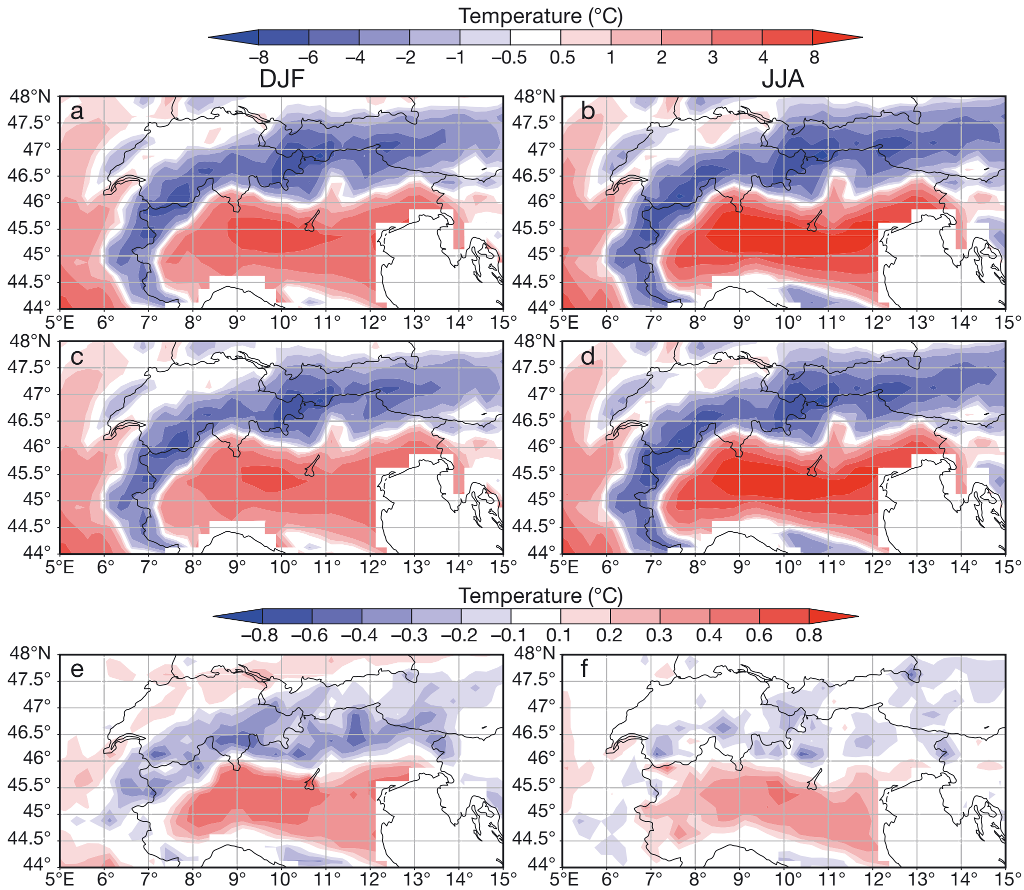

Fig. 9. Mesoscale temperature model mean and mesoscale temperature-weighted mean for the Alps region for (a,c) DJF and (b,d) JJA. The difference between the simple mean and the weighted mean for (e) DJF and (f) JJA is also shown

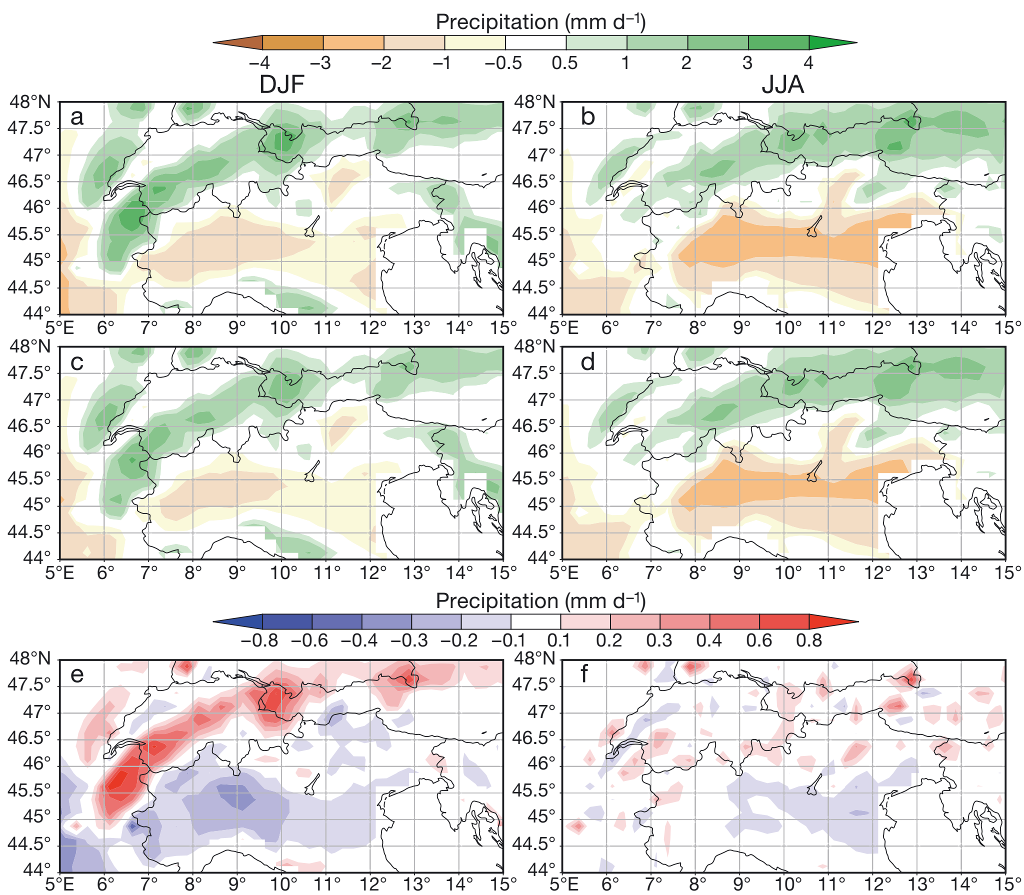

Fig. 10. Mesoscale precipitation model mean and mesoscale precipitation-weighted mean for the Alps region for (a,c) DJF and (b,d) JJA. The difference between the simple mean and the weighted mean for (e) DJF and (f) JJA is also shown

& Somot (2010). We note that, when ensembles of GCM-driven simulations are available, the method should be used in conjunction with a weighting scheme for the driving GCMs. We also stress that, as in any other weighting scheme, ours includes a subjective component in terms of the choice of variables, metrics and combination of functions to produce weights. This subjectiveness is unavoidable, as it is virtually impossible to construct a universally valid performance metric. In addition, the fact that the weighting improves the overall performance of the ensemble in reproducing present climate mesoscale signals does not necessarily imply that its use enhances the robustness of the climate change signals produced by the ensemble. As a result, we suggest that the weighting itself represents a source of uncertainty in the generation of RCM-based regional projections, which should be assessed using suitable sensitivity experiments.

Acknowledgements. This paper was complete as part of the EU project ENSEMBLES (contract no. GOCE-CT-2003-505539). We thank the ENSEMBLES RCM modeling groups for making available their simulation results.

LITERATURE CITED

Christensen JH, Kjellström E, Giorgi F, Lenderink G, Rummukainen M (2010) Weight assigments regional climate models: exploring the concept. Clim Res 44:179–194

Déqué M, Jones RG, Wild M, Giogi F and others (2005) Global high resolution vs. regional climate model climate change scenarios over Europe: quantifying confidence level from PRUDENCE results. Clim Dyn 25:653–670

Déqué M, Rowell DP, Lüthi D, Giorgi F and others (2007) An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections. Clim Change 81:53–70

Déqué M, Somot S (2010) Weighted frequency distributions expressing modeling uncertainties in the ENSEMBLES regional climate experiments. Clim Res 44:195–209

Giorgi F (2005) Interdecadal variability of regional climate change: implications for the development of regional climate change scenarios. Meteorol Atmos Phys 89:1–15

Giorgi F, Mearns LO (2002) Calculation of average, uncertainty range and reliability of regional climate changes from AOGCM simulations via the 'reliability ensemble averaging (REA)' method. J Clim 15:1141–1158

Giorgi F, Mearns LO (2003) Probability of regional climate change calculated using the reliability ensemble averaging (REA) method. Geophys Res Lett 30:1629 doi:10.1029/2003GL017130

Giorgi F, Shields Brodeur C, Bates GT (1994) Regional climate change scenarios over the United States produced with a nested regional climate model. J Clim 7:375–399

Meehl GA, Stocker TF, Collins WD, Friedlingstein P and others (2007) Global climate projections. In: Solomon S, Qin D, Manning M, Chen Z and others (eds) 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, p 747–845

Mitchell TD, Carter TR, Jones PD, Hulme M, New M (2004) A comprehensive set of high resolution grids of monthly climate for Europe and the globe: the observed record (1901–2000) and 16 scenarios (2001–2100), Working Paper 55, Tyndall Centre, Norwich

Murphy JM, Sexton DMH, Barnett DN, Jones GS, Webb MJ, Collins M, Stainforth DA (2004) Quantification of modelling uncertainties in a large ensemble of climate change simulations. Nature 430:768–772

Piani C, Sanderson B, Giorgi F, Frame DJ, Christensen C, Allen MR (2007) Regional probabilistic climate forecasts from a multithousand, multimodel ensemble of simulations. J Geophys Res 112:D24108 doi:10.1029/2007JD008712

Rauscher SA, Coppola E, Piani C, Giorgi F (2010) Resolution effects on regional climate model simulations of seasonal precipitation over Europe. Climate Dynamics 35:685–711 doi:10.1007/s00382-009-0607-7

Tebaldi C, Knutti R (2007) The use of the multi-model ensemble in probabilistic climate projections. Philos Trans R Soc Lond A 365:2053–2075

Tebaldi C, Smith R, Nychka D, Mearns LO (2005) Quantifying uncertainties in projections of regional climate change: a Bayesian approach to the analysis of multi-model ensembles. J Clim 18:1524–1540

Uppala SM, KÅllberg PW, Simmons AJ, Andrae U and others (2005) The ERA-40 re-analysis. QJR Meteorol Soc 131:2961–3012

Xu Y, Gao X, Giorgi F (2010) Upgrades to the reliability ensemble averaging method for producing probabilistic climate-change projections. Clim Res 41:61–81

Submitted: December 22, 2009; Accepted: September 17, 2010

Proofs received from author(s): November 26, 2010