



Upgrades to the reliability ensemble averaging method for producing probabilistic climate-change projections

Ying Xu^{1,*}, Xuejie Gao^{1,2}, Filippo Giorgi²

¹National Climate Center, China Meteorological Administration, 100081 Beijing, China

²International Centre for Theoretical Physics (ICTP), PO Box 586, 34100 Trieste, Italy

ABSTRACT: We present an augmented version of the reliability ensemble averaging (REA) method designed to generate probabilistic climate-change information from ensembles of climate model simulations. Compared to the original version, the augmented method includes consideration of multiple variables and statistics in the calculation of performance-based weighting. In addition, the model convergence criterion previously employed has been removed. The method is applied to the calculation of changes in mean values and the variability of temperature and precipitation over different sub-regions of East Asia, based on the recently completed CMIP3 multi-model ensemble. Comparison of the new and old REA methods, along with the simple averaging procedure, and the use of different combinations of performance metrics shows that at fine sub-regional scales the choice of weighting is relevant. This is mostly because the models show a substantial spread in performance for the simulation of precipitation statistics, a result that supports the use of model weighting as a useful option to account for wide ranges in the quality of models. The REA method and, in particular, the upgraded method provide a simple and flexible framework for assessing the uncertainty related to the aggregation of results from ensembles of models in order to produce climate-change information at a regional scale.

KEY WORDS: Reliability ensemble averaging method · REA · Climate change · CMIP3

Resale or republication not permitted without written consent of the publisher

1. INTRODUCTION

Climate-change information at global, regional and local scales is required in order to develop adaptation and mitigation strategies at national and international levels. Such information is still affected by a high level of uncertainty both because of factors that are intrinsic to the nature of the climate-change prediction problem and the limits to our knowledge and modeling of the climate system (Giorgi 2005). It is thus necessary to characterize this uncertainty and to describe it in a way that can be used in impact assessment and adaptation/mitigation studies.

One way to do so is to approach the climate projection (or prediction) problem in a probabilistic way, in which a climate-change projection/prediction essentially consists of producing probability density func-

tions or cumulative distribution functions (PDFs or CDFs) of the changes in climatic variables of interest. The spread of the PDF is a measure of the uncertainty in the projection, and the PDF itself can be used in risk-based impact assessment and adaptation studies.

In the last several years a number of techniques have been proposed to produce PDFs or CDFs of climate-change variables at both global and regional scales (Wigley & Raper 2001, Giorgi & Mearns 2002, 2003, Murphy et al. 2004, Tebaldi et al. 2004). These are based on the statistical processing of larger or more moderate ensembles of climate-change simulations performed with climate models of varying complexity.

In this study, we were specifically interested in the approach proposed by Giorgi & Mearns (2002, 2003), named the reliability ensemble averaging (or REA) method. In this method, given an ensemble of different

*Email: xuying@cma.gov.cn

model projections, a weight is assigned to each model based on a set of reliability criteria, where higher weighting is assigned to more reliable models. Given these weights, the REA method allows us to calculate the mean change, the uncertainty limits around the mean change estimate, the CDF and PDF of the changes and the overall reliability of the estimated change. Because of the various types of information provided and the simplicity of its application, the REA method is considered a flexible tool that can be applied at both global and regional scales. For example, Moise & Hudson (2008) applied it to the calculation of climate-change PDFs over Australia and southern Africa based on an ensemble of global model simulations.

A 2-fold difficulty in the original formulation of the REA method presented itself in the calculation of weights, which was based on single variables and included criteria of both model performance in reproducing present-day climate and model convergence in future climate projections. (1) Using single variables might provide only a relatively weak measure of model performance, while (2) the model convergence criterion, although broadly used by the climate modeling community in various forms (e.g. Mearns et al. 2001), may cause an artificial narrowing of the PDF of change, which might then mask out important information on outcomes at the tails of the PDFs.

In the present paper we supply an updated formulation of the REA method that addresses these 2 points and that provides a more general framework to calculate model weights. As an illustrative example, we apply the updated REA method to calculate CDFs of temperature and precipitation change over different regions of East Asia based on the recently completed CMIP3 20th and 21st century climate simulation ensemble (Meehl et al. 2007). Within this general REA framework, we also provide an uncertainty analysis of the choice of criteria used for model weighting.

In Section 2, the paper presents a description of the updates to the REA method and the dataset used in the analysis. This is followed by a description and discussion of the results (Sections 3 and 4) and by our main conclusions (Section 5).

2. DATA AND METHODOLOGY

Here we apply the REA method to the calculation of temperature and precipitation changes based on ensembles of model simulations. The dataset employed is the ensemble of coupled atmosphere–ocean general circulation model (AOGCM) simulations conducted in Phase 3 of the coupled model intercomparison project (CMIP3) in support of the 4th assessment report of the Intergovernmental Panel on Climate

Change (IPCC AR4). It consists of simulations of 20th and 21st century climate by 23 models from laboratories worldwide, spanning a horizontal grid spacing in the range of about 1 to 4 degrees (~120 to 500 km) and performing all, or part, of the following simulations: 20th century climate using observed greenhouse gas (GHG) and aerosol forcing (referred to as C20C experiments) and 21st century climate using GHG and aerosol forcing from the A1B, A2 and B1 emission scenarios described in IPCC (2000). The CO₂ concentration associated with the B1, A1B and A2 scenarios is expected to reach a maximum, by 2100, of 550, 720 and 850 ppm, respectively.

Since our paper is mostly illustrative of the behavior of REA updates, we limit our analysis to the mid-level A1B scenario. Out of the 23 models we thus selected 18 that completed at least 1 realization of 21st century climate for the A1B scenario, a mid-level scenario used to be illustrative in the IPCC AR4. The models are: CCMA-3-T47, CNRM-CM3, CSIRO-MK3, GFDL-CM2.0, GFDL_CM2_1, GISS_AOM, GISS_EH, GISS-ER, INMCM3, IPSL-CM4, MIROC3-2H, MIROC3-2M, MIUB-ECHO-G, MPI-ECHAM5, MRI-CGCM2, NCAR-CCSM3, NCAR-PCM1 and UKMO-HADCM3. Some models include multiple realizations for the same scenario simulation, in which case only 1 realization was used in the calculations, so that each model was treated in the same way. Parallel calculations based on ensemble averages of different realizations for each model showed only a weak sensitivity of the results to this factor, confirming previous work that identified the inter-realization spread as a secondary contribution to uncertainty compared to the inter-model spread (e.g. Giorgi & Francisco 2000). We also stress that in our analysis all models are treated as being independent of each other. However, this may not be entirely correct, since some models originated from the same laboratory and thus might share some common model components. The importance of this assumption is difficult to assess, but, in order to procure a larger sample size, we decided to use all available models. More details on the models and simulations can be found at <http://www.pcmdi.llnl.gov> and in Table 8.1 of Randall et al. (2007).

Monthly data for the CMIP3 experiments are obtained from the PCMDI web site and are interpolated onto a common 1-degree grid, as described by Giorgi & Bi (2005a). Only results over land are analyzed. In addition, a common 1-degree land mask grid is defined (see Fig. 1), based on the half-degree grid of the observed dataset from the Climatic Research Unit (CRU) of the University of East Anglia (New et al. 2000). The CRU observation datasets of surface air temperature and precipitation are employed to validate model performance and calculate model weights.

Here we focus on the end-of-the-century change to maximize the change signal; therefore, the term 'change' refers to the difference between 20 yr means in the scenario and C20C simulations. The future time period for the scenario run is 2081–2100, while the reference period for the C20C simulations is 1961–1980, following Giorgi & Mearns (2003). The period 1961–1980 is the most recent (thus presumably with better observations) before the climate shift of the late 1970s in which the GHG-induced climate-change signal became most visible. Since this work is illustrative in nature, the choice of reference period is not important. All changes thus refer to differences between 2081–2100 and 1961–1980. The study area is shown in Fig. 1. It encompasses China, Mongolia and the surrounding regions. Five sub-regions within East Asia are further identified: North and Northeast China (NNE), Southeast China (SEC), eastern part of Northwest China (NWE), Southwest China (SWC) and Northwest China and Tibetan Plateau (NWT).

The general framework of the REA method is the same as that described by Giorgi & Mearns (2002, 2003). The original REA method calculates reliability weights for different models based on the following 2

criteria: (1) the model's ability to reproduce the observed 20th century climatology ('performance criterion'); this component of reliability is calculated from the model bias, and is referred to as R_B ; and (2) the model convergence with other models in terms of simulated climate-change signal ('model convergence' criterion); this component is calculated from the distance of a model's climate-change signal from the corresponding ensemble mean change, and is referred to as R_D .

The assumption underlying the model performance criterion is that a model's reliability is deemed higher and thus the model weight greater, if the model shows a good performance in reproducing present-day climate. Although this can be expected intuitively, it is also true that, since model parameters are generally optimized by comparison with present-day observations, good model performance does not strictly imply a reliable simulation of future climate. In other words, the model performance criterion is a necessary, but not sufficient, condition of model reliability for future climate projection.

The model convergence criterion stems from the notion that increased reliability can be expected of an ensemble of model projections if there is good agree-

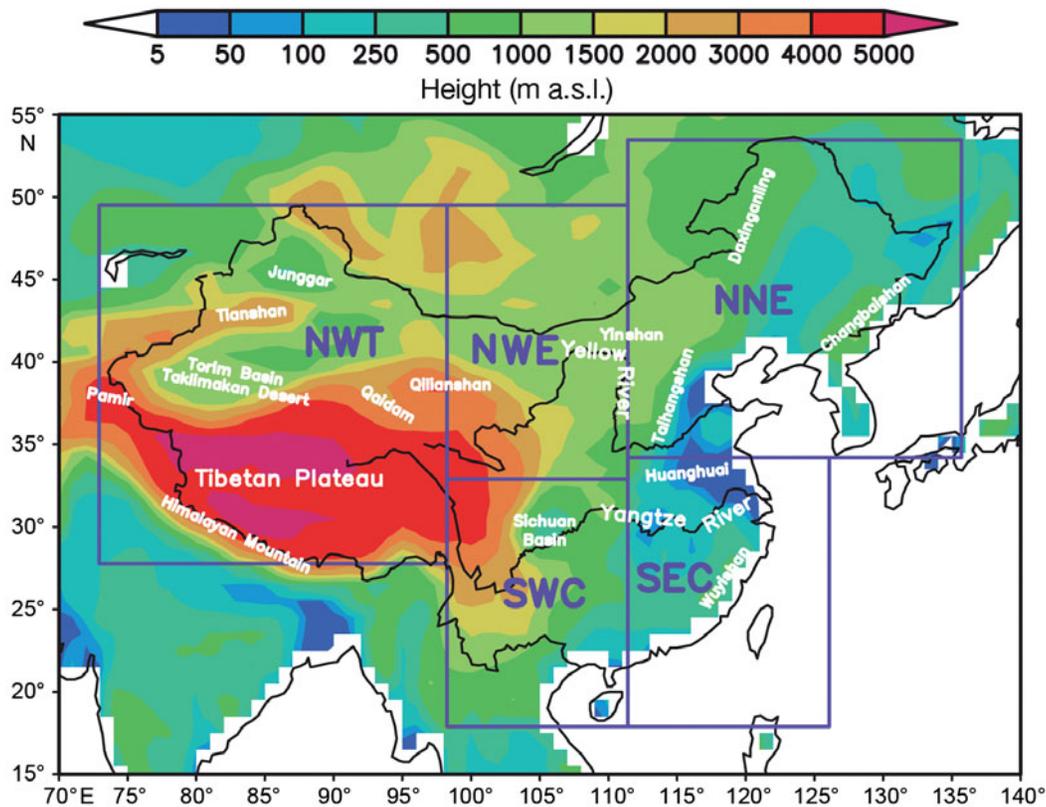


Fig. 1. Study area, sub-regions (North and Northeast China [NNE], Southeast China [SEC], eastern part of Northwest China [NWE], Southwest China [SWC] and Northwest China and Tibetan Plateau [NWT]), and the corresponding 1° resolution topography (m above sea level, a.s.l.)

ment across models in terms of the simulated change signal. By this notion, in the original REA method, the weight of a model is decreased if the change simulated by that model is far from the ensemble average change, i.e. if the model is an 'outlier'.

Given a model i , the model weight is thus expressed as the product of 2 reliability factors, one based on the model performance as measured by the bias ($R_{B,i}$) and the other on the model convergence as measured by the distance of the i -model projected change from the ensemble average change ($R_{D,i}$). In Giorgi & Mearns (2002) these reliability factors are essentially given by the ratio of the model bias, or distance, and a measure of the observed natural variability (ϵ) of the field of interest (see below, this section). If the bias, $B_{T,i}$, of a model or the distance of the simulated change from the ensemble mean, $D_{T,i}$, is within the bounds of the observed natural variability, the corresponding reliability is set to 1. The reliability factor thus ranges from 0 (unreliable) to 1 (reliable), and a model is deemed reliable (i.e. $R_i = 1$) when both the bias and distance are within natural variability. The uncertainty range around the REA mean is given by the REA-weighted root mean square distance of the individual models from the REA mean (or standard deviation between the models). Probabilities of regional climate change are then calculated using the REA method by assuming each model's reliability parameter as a measure of likelihood as described in Giorgi & Mearns (2003).

From a formal point of view, in the original REA method, the REA average change and associated root mean square error are therefore given by:

$$\Delta T = \tilde{A}(\Delta T) = \frac{\sum_i R_i \Delta T_i}{\sum_i R_i} \quad (1)$$

$$\tilde{\delta}_{\Delta T} = [\tilde{A}(\Delta T_i - \Delta T)^2]^{1/2} = \left[\frac{\sum_i R_i (\Delta T_i - \Delta T)^2}{\sum_i R_i} \right]^{1/2} \quad (2)$$

where T denotes temperature (in this paper we take the temperature as an illustrative example), ΔT denotes average change and is given by a weighted average of the ensemble numbers, the operator \tilde{A} denotes the REA averaging and R_i is a model reliability equivalent to the model weight, defined as:

$$R_i = [(R_{B,i})^m \times (R_{D,i})^n] = \left\{ \left[\frac{\epsilon_T}{\text{abs}(B_{T,i})} \right]^m \left[\frac{\epsilon_T}{\text{abs}(D_{T,i})} \right]^n \right\} \quad (3)$$

The parameters m and n in Eq. (3) can be used to weight each criterion. Following Giorgi & Mearns (2003), ϵ_T is a measure of natural variability defined as the difference between the maximum and minimum values of the 20 yr moving averages in the CRU observations for the 20th century after linearly detrending

the data to remove century-scale trends and abs denotes absolute value.

The uncertainty range is then given by:

$$\Delta T_+ = \Delta T + \tilde{\delta}_{\Delta T}; \Delta T_- = \Delta T - \tilde{\delta}_{\Delta T} \quad (4)$$

Once the reliability weight for each model is calculated, the likelihood of a model's outcome is defined by Giorgi & Mearns (2003) such that:

$$P_i = \frac{R_i}{\sum_{j=1}^N R_j} \quad (5)$$

i.e. P_i is the contribution from model m_i to the overall likelihood of a change. Threshold probabilities can be derived by summing over all P_i exceeding a given threshold of climate change. For example, the probability of a temperature change exceeding a certain threshold ΔT_{th} is given by:

$$P^{\Delta T > \Delta T_{th}} = \sum_i P_i, \Delta T_i > \Delta T_{th} \quad (6)$$

We upgraded the original REA method in 2 respects. First, the root mean square error (Eq. 2) used in defining the uncertainty range (Eq. 4) was replaced by:

$$\bar{\delta}_{\Delta T} = \sqrt{\frac{N_{\text{eff}}}{N_{\text{eff}} - 1}} \times \tilde{\delta}_{\Delta T} \quad (7)$$

where N_{eff} is the 'effective number' of models defined as:

$$N_{\text{eff}} = 1 / \sum_{i=1}^N P_i^2 \quad (8)$$

For uniform weighting ($P_i = 1/N$ for all models) N_{eff} equals N ; for non-uniform weights $N_{\text{eff}} < N$. This revision makes the uncertainty ranges obtained from different weighting options more directly comparable with each other. If we assume that the model simulations come from a population in which the climate changes and model weights are uncorrelated, then the square of Eq. (7) will give an unbiased estimate of the variance within this population. However, Eq. (2) gives a downward biased estimate, with the bias increasing with decreasing N_{eff} so that 'aggressive' weighting with very non-uniform weights for the models tends to provide an artificially narrowed uncertainty range.

The second major upgrade concerns the model weight. Eq. (3) shows that in the original REA method the model weight is given by the multiplication of 2 reliability factors that are measures of the model performance and convergence criteria. As mentioned, this equation has limitations. First, the model convergence criterion has the potential of artificially narrowing the PDF of changes by weighting the model outliers less. Second, the weight for a variable, say T , depends only on the model performance in simulating the mean value of that variable, i.e. it depends on a single variable and a single statistic.

Both these limitations are addressed here in the revised version of the REA method. First, we abandon the use of the convergence criterion by eliminating the convergence-based reliability factor from the definition of the overall weight. Second, we include multiple variables and multiple statistics in the definition of the reliability weight. In the revised REA version the weight is thus given by:

$$R_i = [f_1(\bar{T})]^{m_1} \times [f_2(T_{\text{var}})]^{m_2} \times [f_3(\bar{P})]^{m_3} \times [f_4(P_{\text{var}})]^{m_4} \times [f_5(\text{SLP}_{\text{corr}})]^{m_5} \quad (9)$$

where

$$f_1(\bar{T}) = \frac{\epsilon_T}{\text{abs}(\text{Bias})}; f_2(T_{\text{var}}) = \frac{\epsilon_{\text{STD}}}{\text{abs}(\text{STD}_{\text{model}} - \text{STD}_{\text{obs}})}$$

$$f_3(\bar{P}) = \frac{\epsilon_P}{\text{abs}(\text{Bias})}; f_4(P_{\text{var}}) = \frac{\epsilon_{\text{CV}}}{\text{abs}(\text{CV}_{\text{model}} - \text{CV}_{\text{obs}})}$$

$$f_5(\text{SLP}_{\text{corr}}) = \text{corr}(\text{SLP}_{\text{model}}, \text{SLP}_{\text{obs}})$$

Eq. (9) states that a model *weight* in its general form is given by the multiplication of 5 functional factors. The functions f_1 and f_3 measure the model ability to reproduce mean temperature and precipitation, respectively, and are the same as in the original REA version. The functions f_2 and f_4 measure the model performance in reproducing observed interannual variability (T_{var} or P_{var}). This variability is measured by the interannual standard deviation (STD) for temperature and by the interannual coefficient of variation (CV, standard deviation divided by the mean) for precipitation (Räsänen 2002), where the STD and CV are calculated for each 20 yr period of analysis. The function f_5 is a measure of the spatial correlation ($_{\text{corr}}$) between simulated and observed sea level pressure (SLP) patterns over a broad area encompassing the regions of interest (including both land and the surrounding ocean points), i.e. a measure of the model's ability to reproduce the basic circulation climatology affecting the region. Note that the functions f_1 to f_4 can be calculated down to the grid-point level, while the function f_5 is based on a spatial correlation and thus needs to be calculated over a broad region. In our case, we selected a region encompassing China and its adjacent areas,

extending from 70 to 140° E and from 15 to 60° N, i.e. including the primary circulation patterns of relevance for East Asia, such as the Western Pacific Subtropical High, the motion of which determines the evolution of the East Asia monsoon.

It is evident from Eq. (9) that the model weight accounts for multiple variables (local temperature and precipitation, and the broad spatial distribution of sea level pressure) and multiple statistics (mean and inter-annual variability). The multiplication of the 5 functions provides a stringent test of model performance, in that, for a model to have a relatively high weight, all 5 functions need to have high values, i.e. the model needs to provide a relatively good performance in all metrics considered. In addition, the parameters m_1 to m_5 , which can be any real numbers, can be used to weight each function differently. Setting 1 of these parameters to 0 is equivalent to removing the corresponding performance function.

Although Eq. (9) provides a general framework for the definition of the weight, a large number of different combinations of the 5 functions could be used in different contexts. As a sensitivity exercise we consider here 6 different formulations of the REA weights, corresponding to 6 combinations of m_1 to m_5 (Table 1).

In REA1 all parameters m_1 to m_5 are equal to 1, i.e. all 5 functions are considered and are equally weighted. This is our reference case. In REA2 m_3 , m_4 and m_5 are set to 0 for the temperature weights and m_1 , m_2 and m_5 are set to 0 for the precipitation weights. This corresponds to using only temperature-based functions for temperature and precipitation-based functions for precipitation, which, in turn, corresponds to removing the condition of using multiple variables. In REA3 and REA4, m_1 to m_5 is set to different values (see Table 1). In REA3, this corresponds to removing the condition of multiple statistics and giving a higher weight to the function including the mean of the variable considered, and, in REA4, the SLP is removed. Finally, REA-ORIG is the original formulation of Giorgi & Mearns (2002), and AVE is the case of simple averaging, i.e. all models are equally weighted.

Table 1. Different formulations of the reliability ensemble averaging (REA) weights in Eq. (9). AVE: the case of simple averaging, i.e. all models equally weighted

Methods	Temperature weight	Precipitation weight
REA1	$m_1 = 1, m_2 = 1, m_3 = 1, m_4 = 1, m_5 = 1$	$m_1 = 1, m_2 = 1, m_3 = 1, m_4 = 1, m_5 = 1$
REA2	$m_1 = 1, m_2 = 1, m_3 = 0, m_4 = 0, m_5 = 0$	$m_1 = 0, m_2 = 0, m_3 = 1, m_4 = 1, m_5 = 0$
REA3	$m_1 = 2, m_2 = 0, m_3 = 1, m_4 = 0, m_5 = 0$	$m_1 = 1, m_2 = 0, m_3 = 2, m_4 = 0, m_5 = 0$
REA4	$m_1 = 1, m_2 = 1, m_3 = 1, m_4 = 1, m_5 = 0$	$m_1 = 1, m_2 = 1, m_3 = 1, m_4 = 1, m_5 = 0$
AVE	m_1 to $m_5 = 0$	m_1 to $m_5 = 0$
REA-ORIG	Original formulation of Giorgi & Mearns (2002)	Original formulation of Giorgi & Mearns (2002)

3. RESULTS

3.1. Validation of the model simulations in present-day climate

Since the new version of the REA method is essentially based on the model performance criterion, we first present a brief evaluation of the ability of the model ensemble in simulating the observed climatology over the region of interest. We present results for the mean and interannual variability, since both are used in the calculation of model weights.

Fig. 2 presents for December-January-February (DJF) and June-July-August (JJA) the observed mean temperature from the CRU dataset, the unweighted model average (AVE) and the mean model bias (AVE-CRU) for the period 1961–1980. In DJF, the CMIP3 ensemble shows a predominant cold bias of up to several degrees over East Asia. The cold bias reaches -6°C in the Sichuan Basin and a maximum value of about -8°C over the Tibetan Plateau. Conversely, areas of warm bias are found in the Qilianshan and Tianshan Mountains of northwestern China. The largest biases are found in areas of complex topography, where, on the one hand, the model topography is excessively smooth and, on the other hand, the observed data are relatively sparse. The ensemble average temperature in JJA shows features similar to those found in DJF, except that the cold bias is smaller (usually -1 to -4°C).

Fig. 3 shows observed and ensemble average DJF and JJA precipitation for the period 1961–1980. In DJF, the prevailing northwesterly winds bring cold and dry air from the inner continent to East Asia. Less than 5 mm mo^{-1} of precipitation is observed over the broad areas of northern and northwestern China (Fig. 3a). Precipitation increases towards the southeast, and a maximum of 75 mm mo^{-1} is observed over the coastal regions of eastern China. The CMIP3 ensemble captures the observed precipitation pattern, with an increasing precipitation gradient from the north to the southeast; however, it overestimates the precipitation by up to 100% over most areas of China (Fig. 3c,e). This overestimate may be at least partially due to the smooth representation of the Tibetan Plateau in the CMIP3 AOGCMs. We also note that precipitation observations over much of the plateau may be affected by large uncertainties due to the relatively sparse station distribution.

JJA is the monsoon season over East Asia, with maximum precipitation over the southern coastal areas of China and decreasing precipitation towards the north and west. Monsoon precipitation in the CMIP3 ensemble (Fig. 3d) shows similar features to those in earlier global models, with a simulated monsoon precipitation center located too far north and west over central

China (e.g. Gao et al. 2001, Yu et al. 2001). As discussed by Gao et al. (2006), this is primarily caused by the coarse resolution of the global models. As a result of the excessively northward and westward extension of the monsoon front, the CMIP3 ensemble also overestimates precipitation over northwestern China, which is a region characterized by a prevalence of arid and semi-arid areas.

Interannual variability in observed and simulated temperatures (as measured by the interannual STD) for DJF and JJA are shown in Fig. 4. Interannual variability in the observed temperature is generally greater in DJF than JJA (Fig. 4a,b), and in both seasons it tends to increase towards the northern regions. In the cold season this is at least partially due to the effect of the snow-albedo feedback mechanism, which tends to enhance the surface response to temperature anomalies and, thus, the temperature variability. The CMIP3 ensemble captures both these features of interannual variability in temperature. However, the variability is overestimated throughout most of East Asia in JJA and over the central regions of China in DJF, while it is underestimated over the northern regions of the domain considered.

Concerning the interannual variability in precipitation (as measured by the CV; Fig. 5), in the observations this is found to be greater in DJF than in JJA, and the maximum is found in both seasons over northern and northwestern China. The CMIP3 ensemble captures these features, although it tends to underestimate interannual variability in both seasons, in particular over the Tibetan Plateau in DJF, where the relatively low density of observation stations adds an important element of uncertainty.

In summary, the validation analysis presented in this section shows that the CMIP3 ensemble captures the broad spatial patterns of the temperature and precipitation climatology of East Asia, both for the mean and for interannual variability. Some systematic biases, however, still persist in this generation of models, such as a general cold bias over the region, an excessively northwestward extension of the monsoon precipitation and a prevailing underestimate of interannual variability in DJF.

3.2. Future changes

In this section we focus on the future changes of the mean values and interannual variability in temperature and precipitation, as obtained with the revised REA method (case REA1) and compare them with the results obtained with the original REA method by Giorgi & Mearns (2002, 2003) (REA-ORIG) and the unweighted ensemble mean (AVE). We also explore

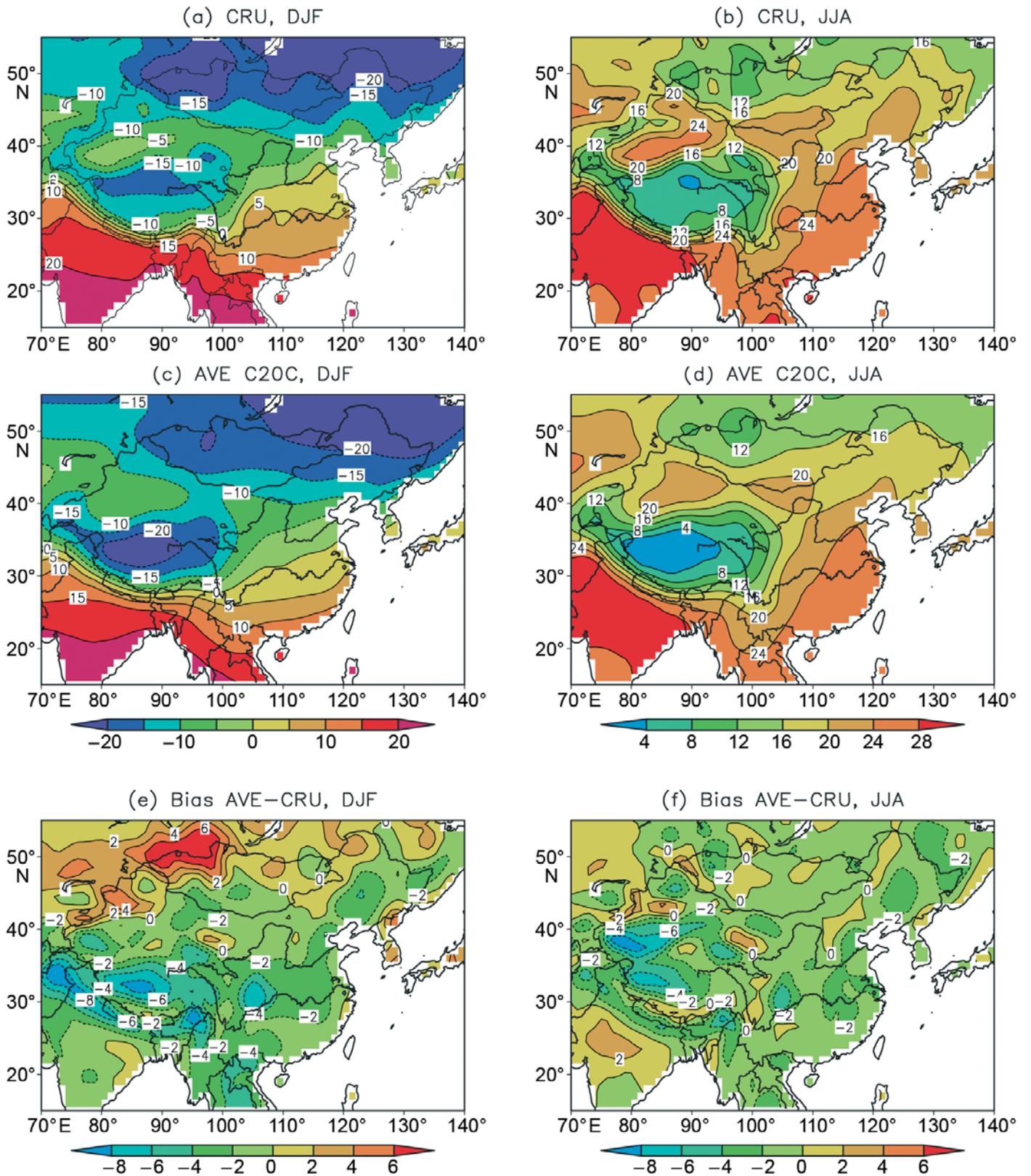


Fig. 2. (a,b) Observed (climate research unit; CRU), and (c,d) ensemble mean (unweighted model average; AVE) temperature ($^{\circ}\text{C}$) and (e,f) temperature bias (Bias, $^{\circ}\text{C}$) over East Asia for the period from 1961 to 1980. Left panels: DJF (December–February, winter); right panels: JJA (June–August, summer)

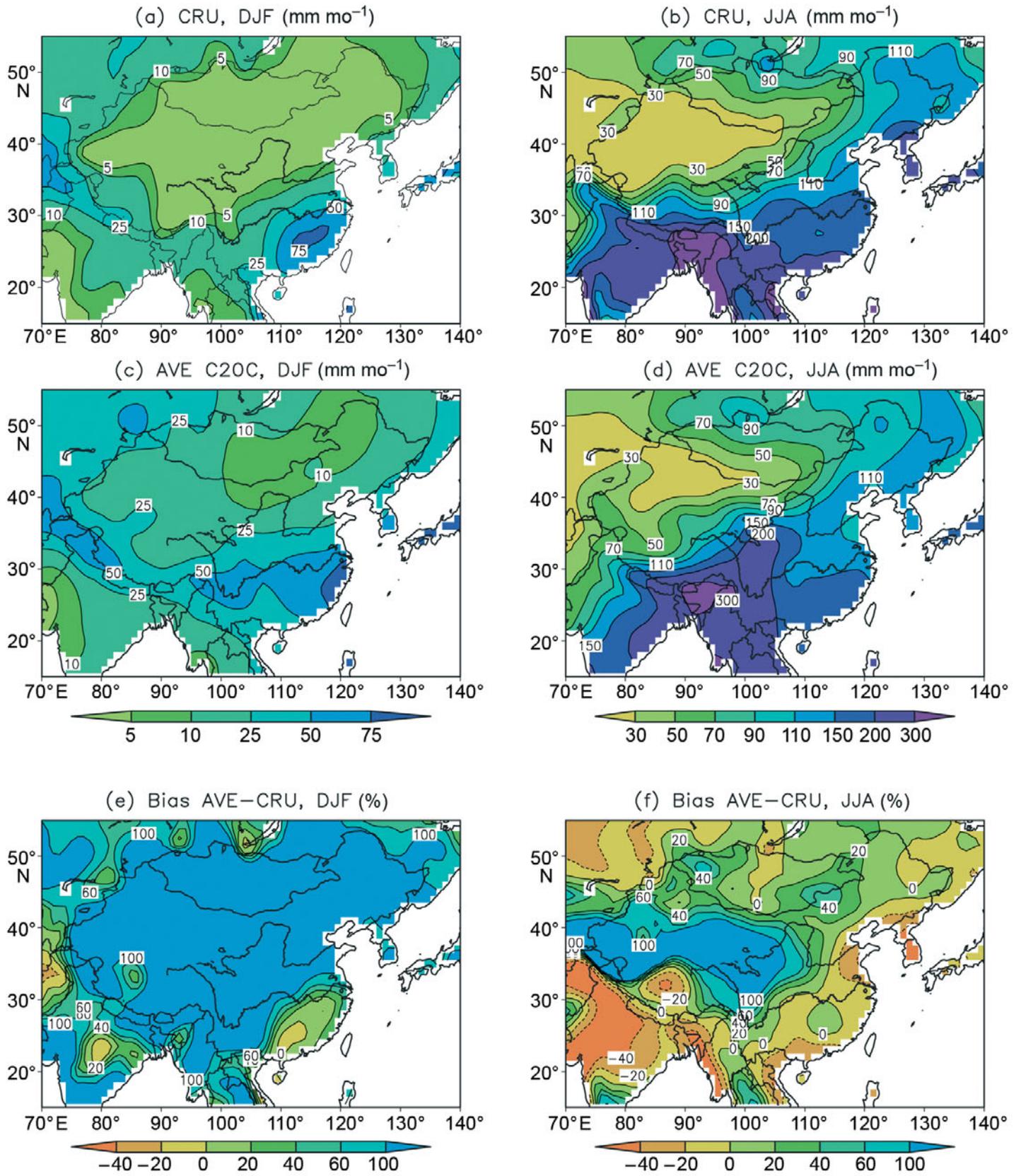


Fig. 3. As in Fig. 2, but for mean precipitation

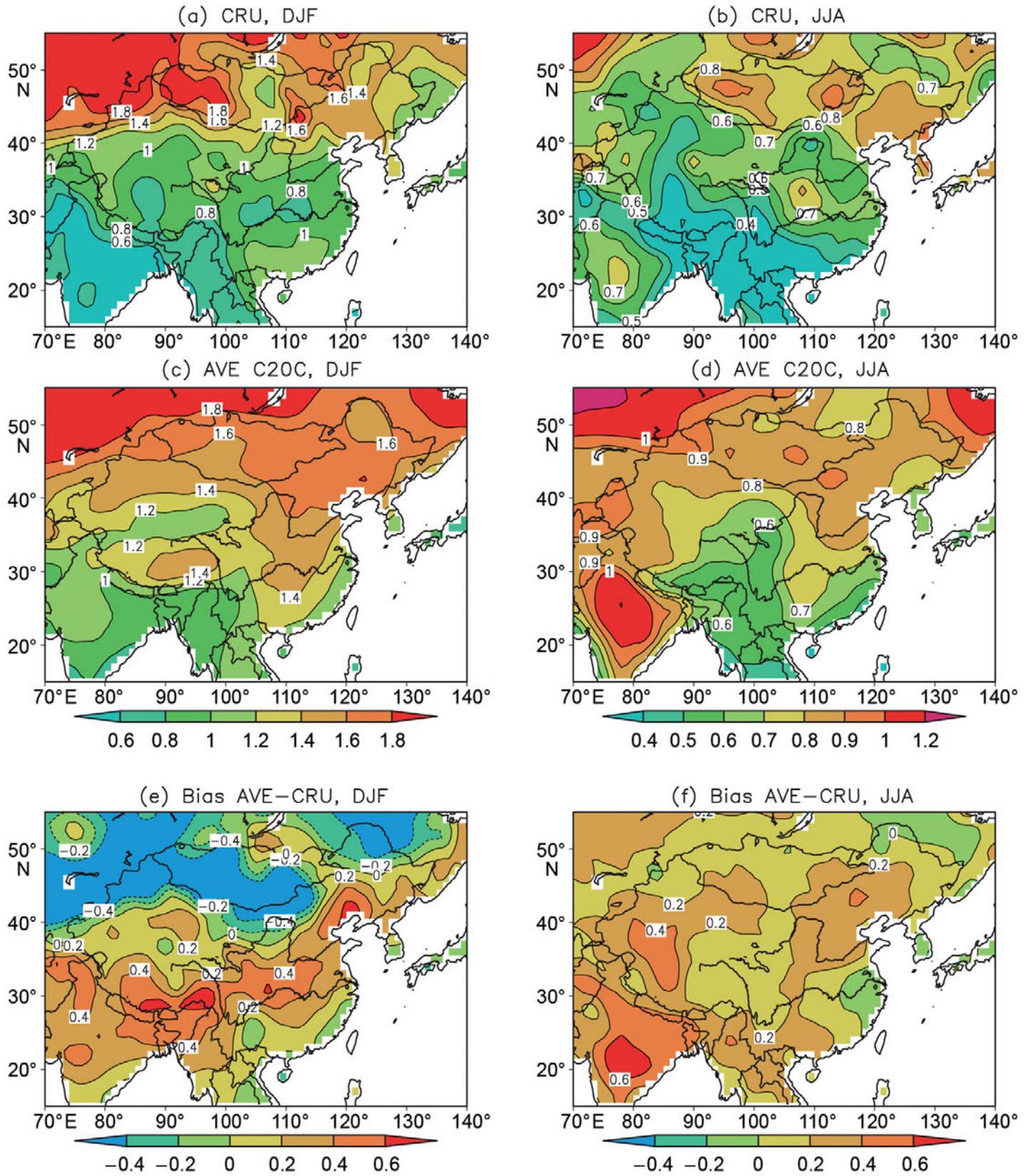


Fig. 4. Interannual variability in temperature ($^{\circ}\text{C}$), as measured by the interannual standard deviation for the period from 1961 to 1980. Abbreviations as in Fig. 2

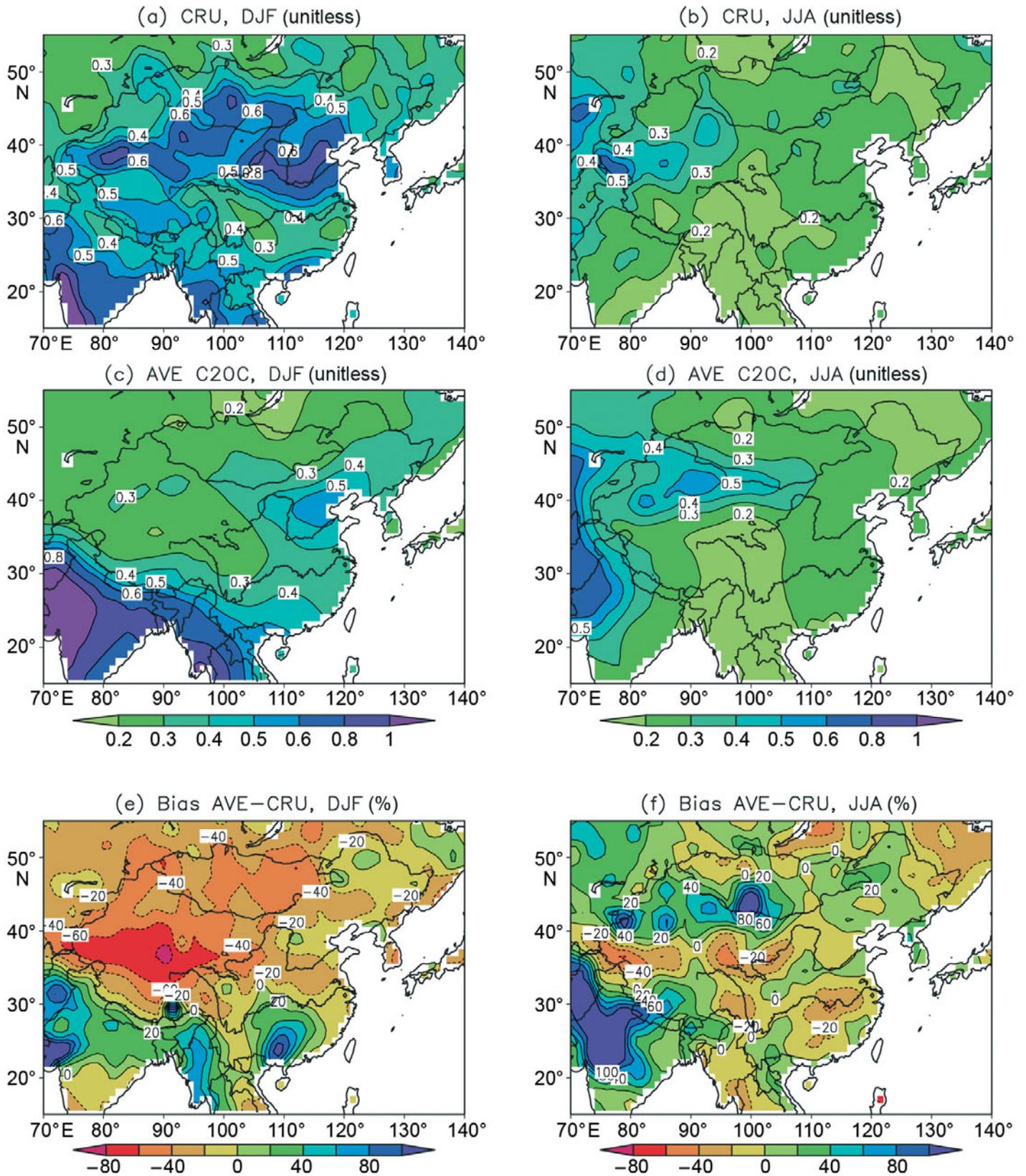


Fig. 5. As in Fig. 4, but for the interannual variability in precipitation as measured by the interannual coefficient of variation. Abbreviations as in Fig. 2

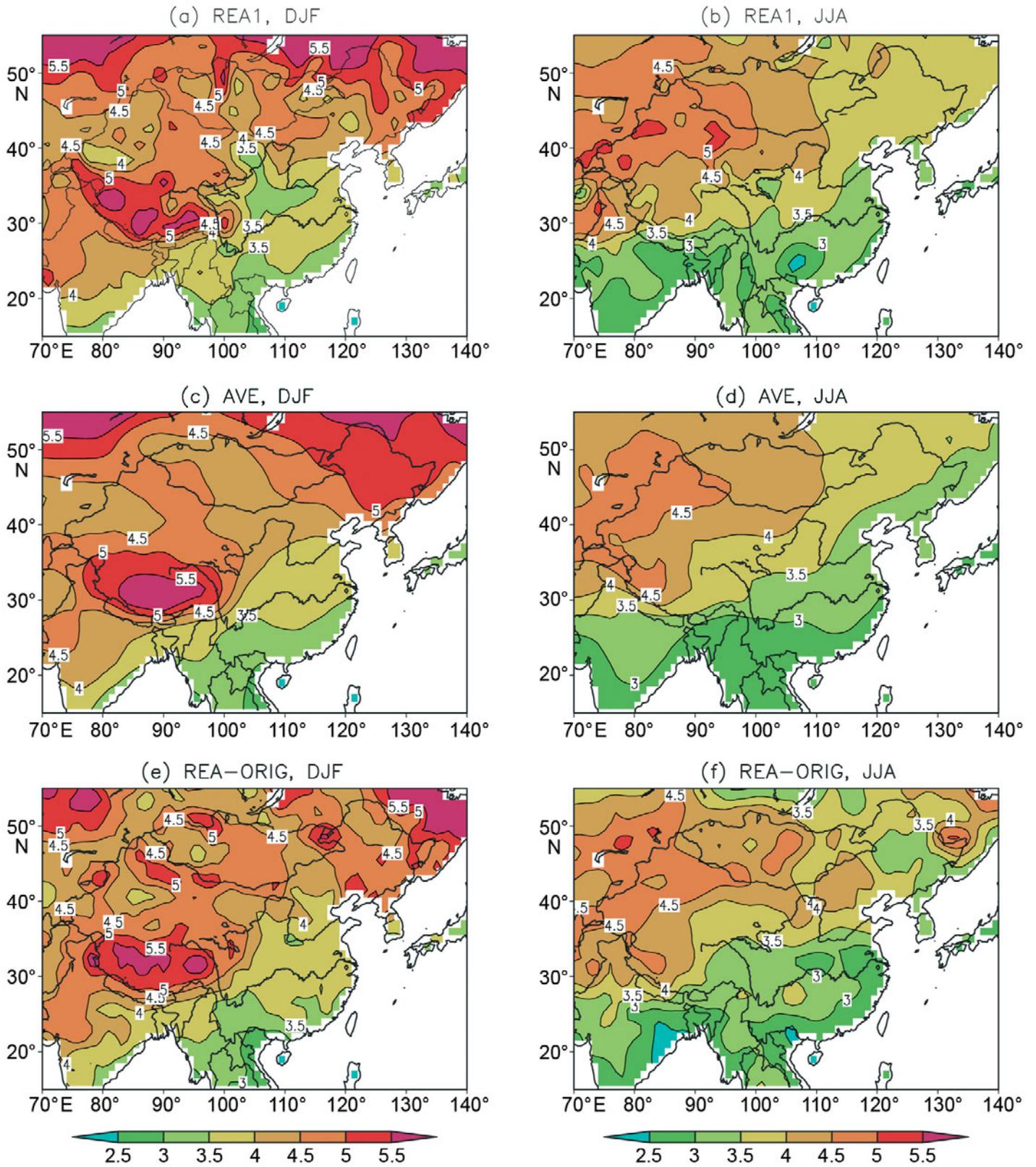


Fig. 6. DJF and JJA temperature change (°C) according to the A1B–C20C scenario (2081 to 2100 minus 1961 to 1980) calculated by reliability ensemble averaging (REA1) or simple averaging (AVE). REA-ORIG: original REA method. Further abbreviations as in Fig. 2

the sensitivity of the results to different combinations of the functions f_1 to f_5 (cases REA2 to REA4 in Table 1).

3.2.1. Changes in mean temperature and precipitation

Fig. 6 compares the temperature change signal (2081–2100 minus 1961–1980) for DJF and JJA as obtained with the REA1, AVE and REA-ORIG approaches. The AVE case shows a warming over the region in both seasons, generally greater in DJF than in JJA. In DJF, the warming is in the range of from 3.0 to 5.0°C, and its maximum value is over the northern regions of the analysis domain and over the Tibetan Plateau. This is consistent with the large-scaled distribution of warming found in most climate-change experiments, with maximum warming over high latitudes and high-elevation northern hemisphere regions due to the melting of snow and sea ice and the associated albedo feedback (Giorgi et al. 1997, Meehl et al. 2007). The ensemble average (AVE) warming is somewhat less pronounced in JJA, ranging between <3.0°C to >4.5°C, and it is maximum over the northwestern areas of the analysis domain.

Comparison of the REA1 and REA-ORIG results with AVE indicates that the basic patterns of warming remain unchanged, but that significant differences are found at sub-regional and local scales. Both REA schemes generally show finer scale details than AVE, and areas of substantial differences in warming. For example, both REA cases produce greater warming over northwestern China in JJA than AVE. In general, local differences in warming between the REA and AVE cases can be up to 1°C (Fig. 6). The fine-scale structure of the REA change signal is clearly associated to the use of locally calculated weights, so that the weighting might add some noise to the change fields, especially for aggressive weighting resulting in small N_{eff} values.

The mean DJF and JJA precipitation changes according to REA1, AVE and REA-ORIG are shown in Fig. 7. In DJF we find a dipolar structure of the precipitation change signal, with an increase over the northern regions of the domain and a decrease in the southern regions. This is consistent with a poleward shift of the mid-latitude storm track observed in climate-change simulations, which is associated with greater mid-tropospheric warming in the tropics than at high latitudes (Meehl et al. 2007). Increased northward water vapor transport resulting from the increased moisture content of the warmer atmosphere also contributes to these changes (Christensen et al. 2007). Thus, DJF precipitation increases over central and northern China, Mongolia and eastern Russia and decreases over southern China, India and the Indochina peninsula. All 3

approaches yield the same north–south dipolar structure; however, local differences of 10 to 20% between the REA and AVE cases are found, for example, over the Tibetan Plateau, Mongolia and northeastern China, as well as in areas of India and southeastern China. Again, the REA averaging tends to produce a more fine-scale structure of the change signal than simple averaging, because of the contribution of a smaller number of ‘effective’ models.

In JJA the AVE precipitation change is positive throughout our East Asia domain, as found in previous analyses of AOGCM projections (e.g. Giorgi et al. 2001, Hu et al. 2003, Giorgi & Bi 2005a, Xu et al. 2007) except for a region west of the Tibetan Plateau (Fig. 7d). The maximum increase in precipitation is over northeastern and southwestern China and the minimum is over a northwest–southeast-oriented region of central China. The REA-ORIG case shows a pattern generally in line with the AVE case; however, larger differences are found for the REA1 case. In fact, in REA1, the precipitation change over a large portion of central and southeastern China actually becomes negative. The more pronounced spatial variability of the change in REA1 is attributed to the fact that the weights vary more across the models (i.e. N_{eff} is smaller) in the new REA1 formulation than in REA-ORIG and obviously more than in AVE (see Fig. 11).

Changes of interannual variability in temperature for the REA1 and AVE cases are shown in Fig. 8a–d (note that the REA-ORIG approach cannot be directly applied to interannual variability, since the weights are only based on mean biases). In general, these changes exhibit a large amount of spatial variability and broad-scale consistency for the 2 approaches. The interannual variability mostly increases in JJA (consistent with the results of Räisänen 2002 and Giorgi & Bi 2005b) and shows areas of increase and decrease in DJF. Again, in both seasons, the spatial variability of the change is much more pronounced with the REA1 averaging approach, with areas of different signs of change between the 2 averaging methods. Similar conclusions are found for the interannual variability in precipitation (Fig. 8e–h).

Summarizing the results of this section, we can draw 2 basic conclusions. On the one hand, the broad-scale patterns of change are similar in the different averaging methods, especially for temperature and for mean changes. On the other hand, when REA weighting is used, the changes show greatly increased spatial variability, resulting in more detail at sub-regional to local scales and even in areas where REA changes have the opposite sign compared to the simple average. This is due to the use of locally calculated weights that locally reduce the number of effective models contributing to the calculation of change (see also the following subsection).

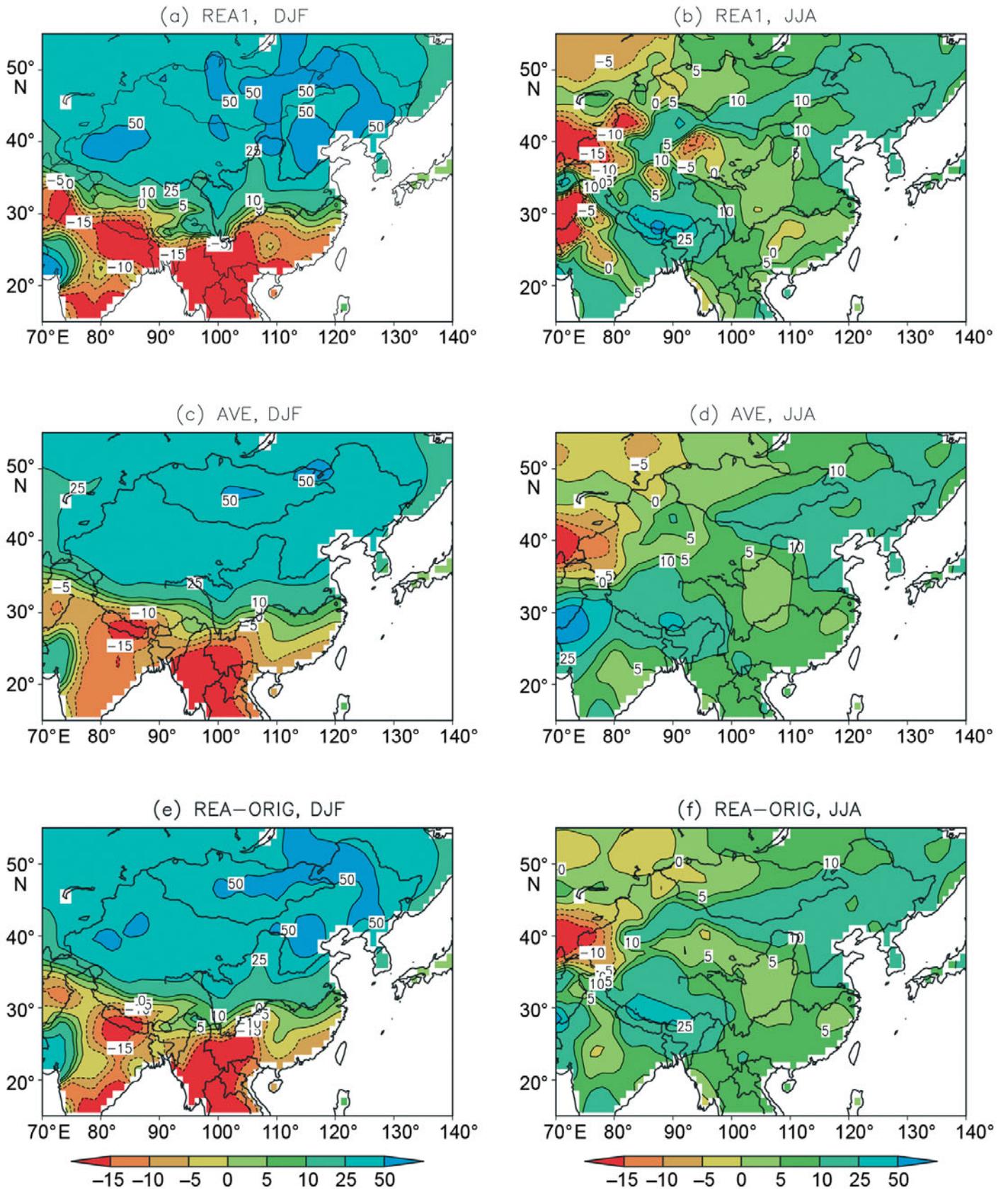


Fig. 7. As in Fig. 6, but for precipitation (percent of 1961 to 1980 values)

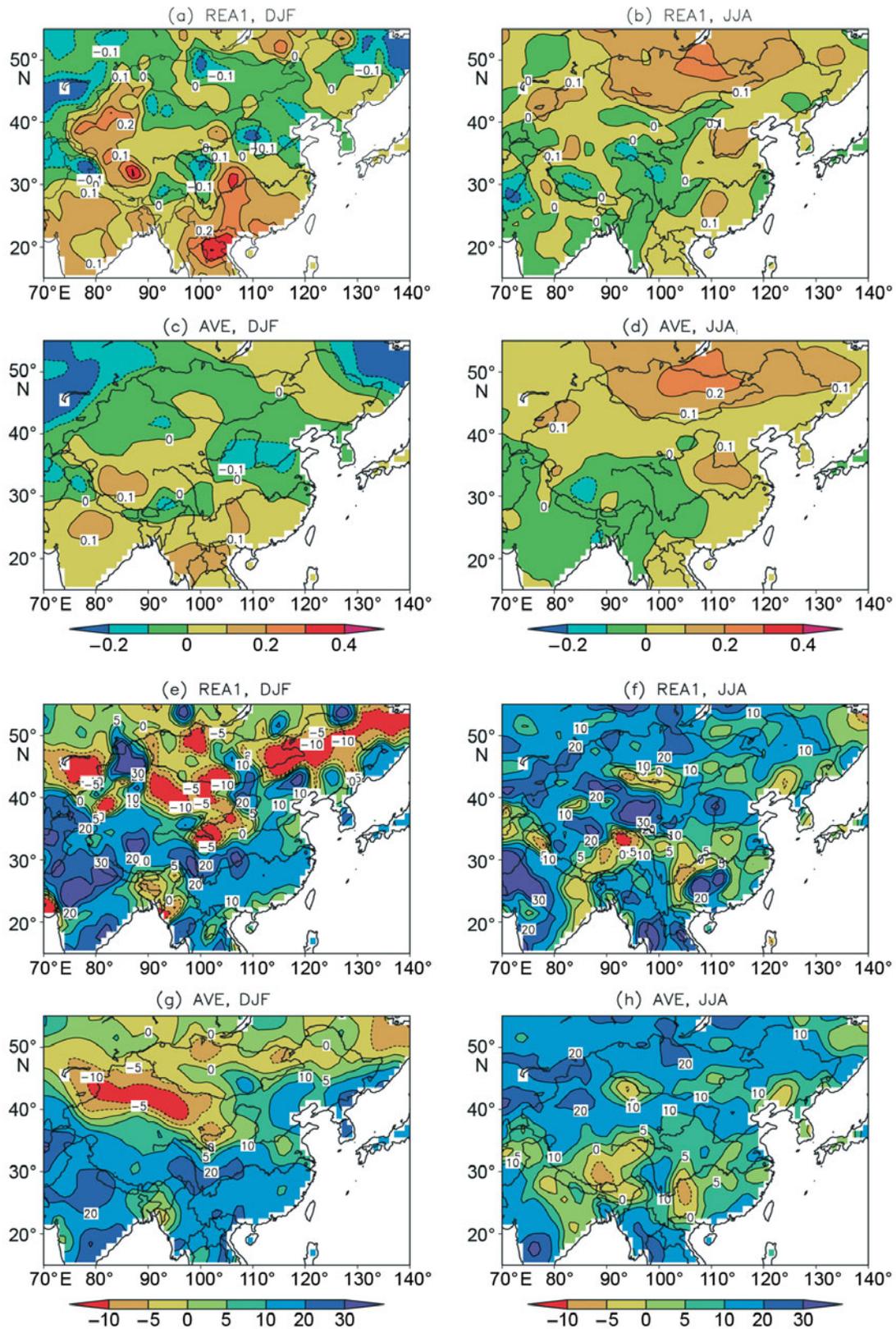


Fig. 8. (a–d) Change of interannual variability in temperature ($^{\circ}\text{C}$, 1960–1980 value; as measured by the interannual standard deviation) and (e–h) in precipitation (% , 1961 to 1980 value; as measured by the interannual coefficient of variation) from different averaging procedures; A1B–C20C scenario

3.2.2. Sensitivity to weighting scheme and contributions of the different functions entering the weights

We have seen in the previous section that the different averaging procedures can provide significantly different results at fine scales. Because the choice of weighting has a subjective component, it is useful to assess the sensitivity of the results to different weighting assumptions. For this purpose, we investigated the sensitivity of the results to the 6 averaging schemes given in Table 1.

Fig. 9a–d shows the mean REA change (Eq. 1) along with the uncertainty range for DJF and JJA temperature and precipitation for the 6 weighting schemes given in Table 1, the 5 Chinese sub-regions (Fig. 1) and the whole East Asia region. Fig. 9e–h shows the same data for the change of interannual variability in temperature and precipitation (except for the original REA method, which is not applicable to interannual variability).

For mean temperature change, the sensitivity to the weighting scheme is between a few tenths of a degree to slightly $<1^{\circ}\text{C}$, while, for precipitation, it is generally $<10\%$, except for the North China region in DJF, when it exceeds 20% (Fig. 9a–d). The change in interannual variability has a somewhat greater sensitivity to the weighting scheme, with instances in which the sub-regional scale REA mean change actually changes sign when different weighting schemes are used. This happens, however, only when the change is small. Overall, we find a general consistency across weighting schemes at the sub-regionally averaged scale.

Fig. 10 shows the contributions of the 5 functions in Eq. (9) to the overall weight for the 5 Chinese sub-regions along with the whole East Asia value. Values are reported for DJF and JJA and include the ensemble average, as well as the highest and lowest individual model values for each function. Finally, the bottom right panels in both figure parts show the ensemble average, maximum and minimum weights for the REA1 case (Eq. 8). The ensemble average weight can be considered a measure of the reliability of the projection by the ensemble of models, since it is a measure of the overall performance of the ensemble.

Looking first at the temperature-based functions f_1 and f_2 , we observe relatively high values in both DJF and JJA, with means in the range of 0.4 to 0.9, maximum values of 1 (bias lower than the estimated natural variability) and minimum values of 0.1 to 0.2. This indicates that the CMIP3 ensemble of models performs relatively well in simulating both mean values and interannual variability in temperature. The performance is weaker for mean values and variability in precipitation (functions f_3 and f_4), especially in

DJF. In this case, there is a wide inter-model spread, with values ranging from close to 1 (bias less than the natural variability) to about 0. In addition, we also find strong variability in the weighting functions across regions, especially in DJF, with model mean values ranging from 0.1 to 0.9. Note that all models show a f_3 value of 1 in JJA over northeastern China, implying a bias lower than the natural variability for all models.

Finally, the function f_5 measures the model's ability to reproduce the observed mean SLP patterns on a broad scale, and thus it is the same for all sub-regions. This function, which is the correlation of observed and simulated SLP patterns, shows high values, between 0.4 and 0.94 (mean of 0.8) in DJF, but lower values of 0.1 to 0.8 (mean of ~ 0.45) in JJA. This implies that the ensemble of CMIP3 general circulation models is more successful in representing winter circulation patterns than summer patterns. Some of the models, with correlations close to 1, show very good performance in this regard.

Coming to the total weight, i.e. the product of the 5 functions (Fig. 10, bottom panels), we find that, on average, this is relatively low, <0.2 in DJF. The minimum model weights are close to 0, mostly as a result of the precipitation metrics. This is a result of the well-known general difficulty of coarse resolution models to simulate precipitation over East Asia (Gao et al. 2006). On the other hand, in JJA, the top model weights are relatively high, up to 0.75, which indicates a large spread across models and, in particular, that a few models do perform well over the region. This suggests that weighting may be important over this area so that the contribution of the low-performing models is filtered out.

A useful quantity for the evaluation of the REA scheme is the effective number of models (N_{eff}), which measures the spread of the weights and thus the relative contribution of the different models in the ensemble. Fig. 11 shows the values of N_{eff} for the REA1 and REA-ORIG cases over the Chinese sub-regions shown in Fig. 1. It can be seen that N_{eff} is high, >14 , in the original case over all regions, implying that most models are similarly weighted. Evidently the REA-ORIG configuration results in only weak weighting across models, and thus most models contribute to the ensemble. The N_{eff} values decrease considerably in the REA1 case, showing a pronounced variability across regions. N_{eff} values are in the range of slightly <2 to about 12. In the former case, this implies that only a minority of the models contribute to the REA ensemble averaging. Addition of multiple variables and multiple statistics in the calculation of the weights thus leads to a much greater performance-based selection across models, i.e. in a more 'aggressive' weighting.

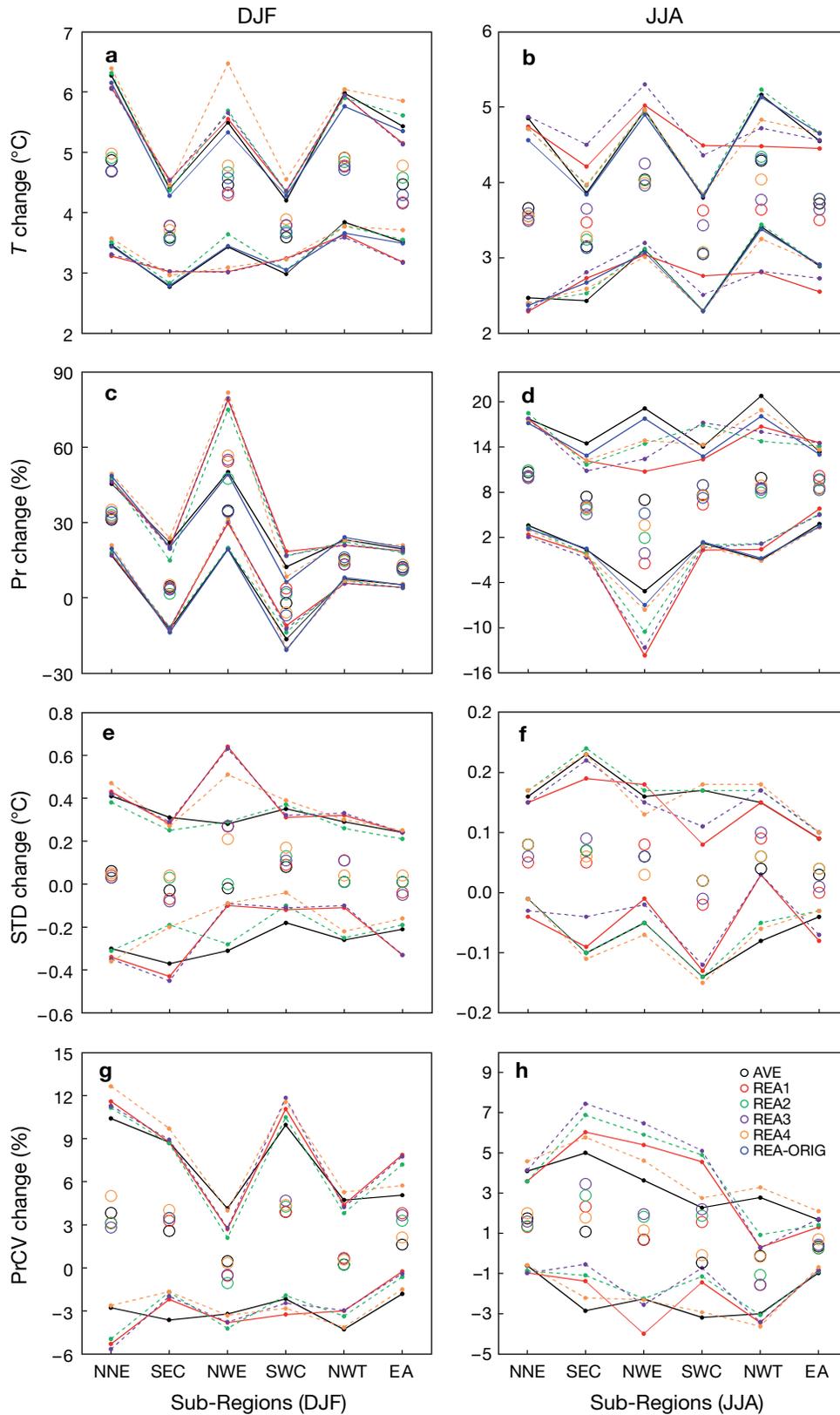


Fig. 9. Different formulations of REA and simple average change in (a,b) temperature (T); (c,d) precipitation (Pr); (e,f) standard deviation of temperature (STD); and (g,h) coefficient of variation of precipitation (PrCV) for winter (DJF) and summer (JJA). Corresponding upper and lower uncertainty limits are also given. Abbreviations for sub-regions as in Fig. 1

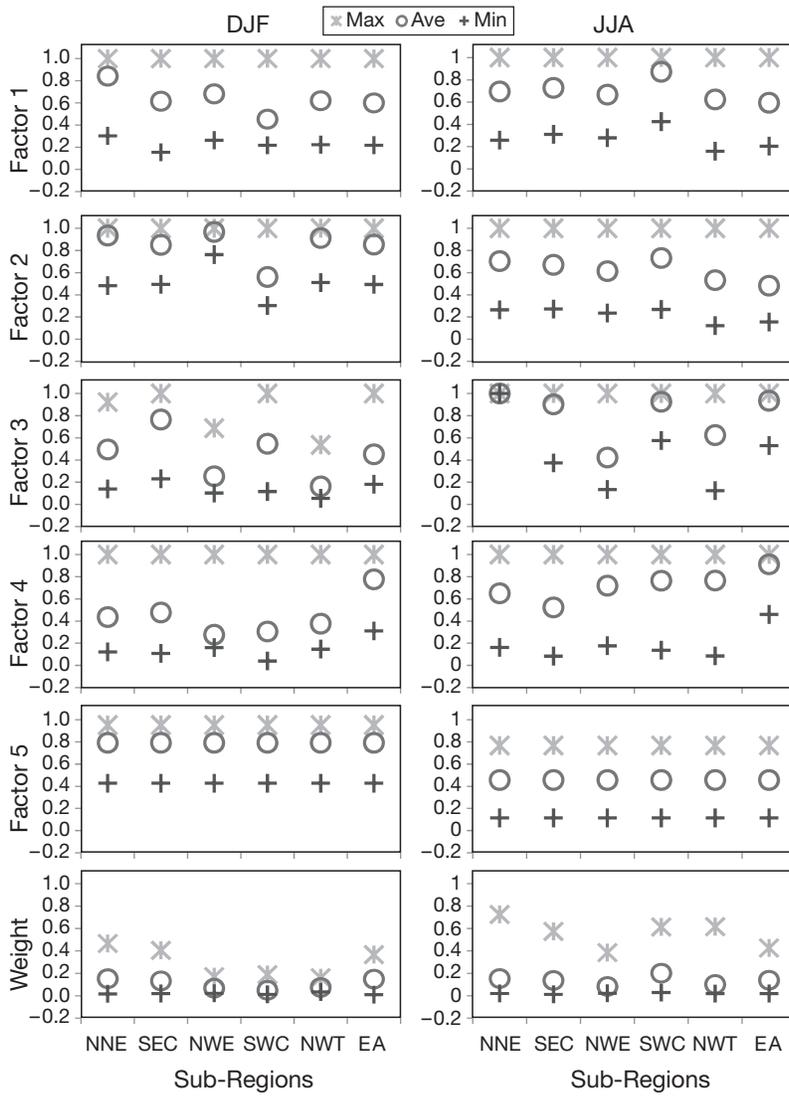


Fig. 10. Contribution of the 5 functions of Eq. (9) to the overall weight, along with the weight itself, averaged over the 5 sub-regions shown in Fig. 1 and the whole East Asia region (EA) for (a) winter (DJF) and (b) summer (JJA). The average, maximum and minimum values across models are indicated

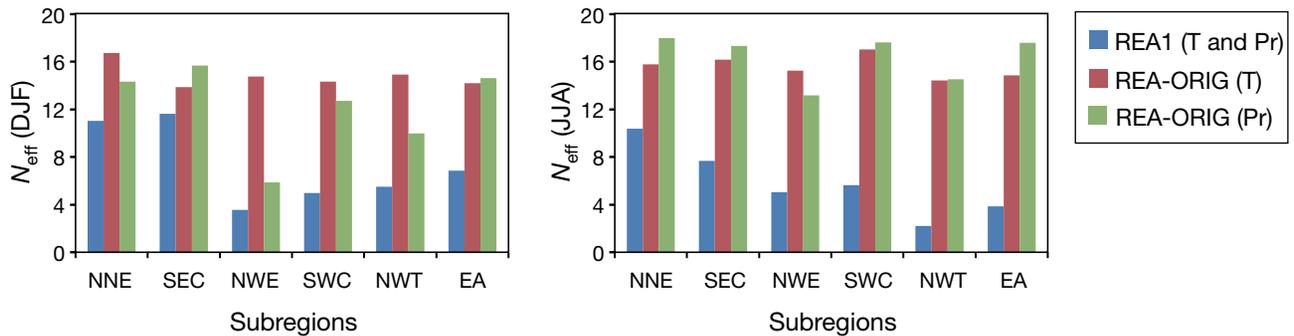
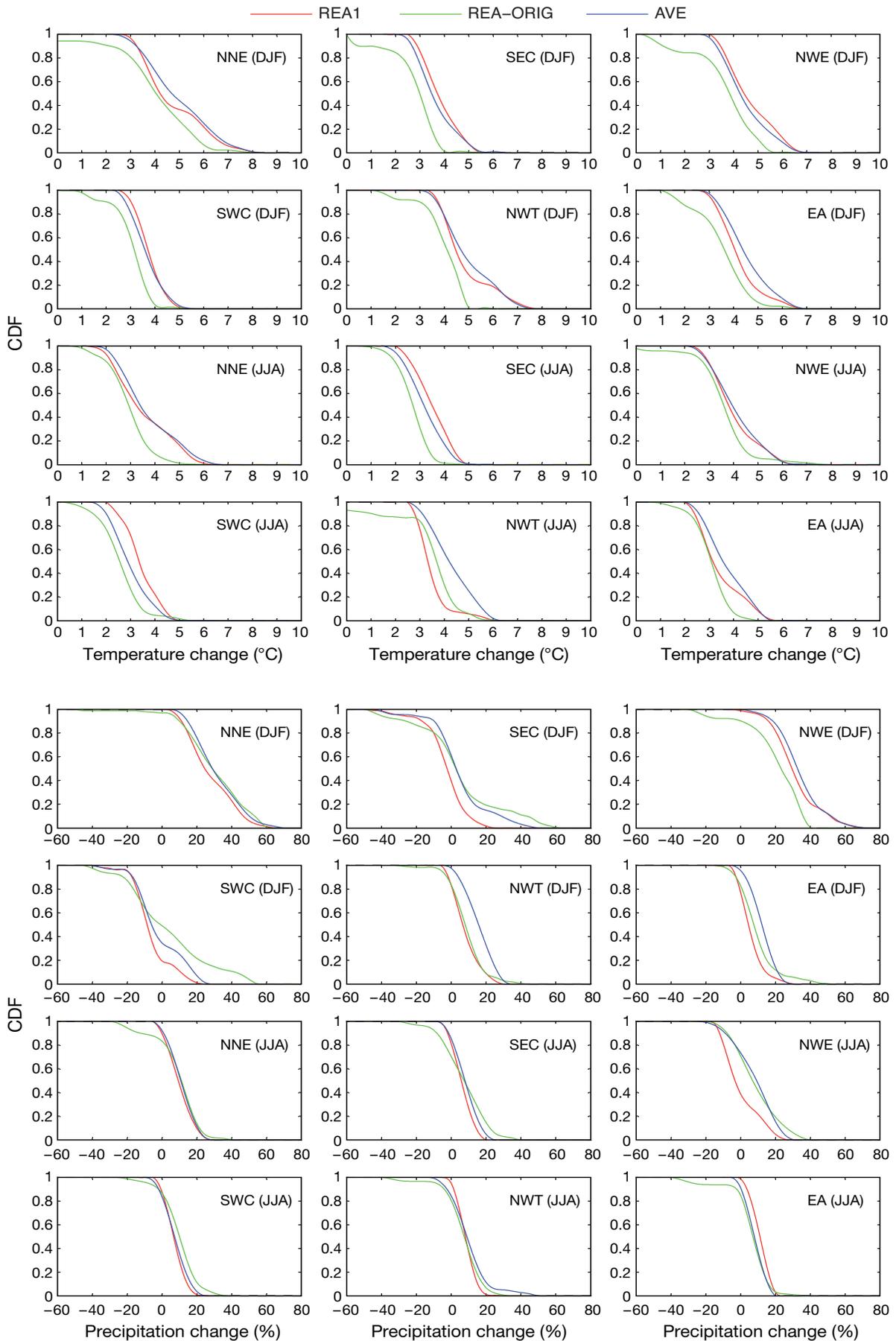


Fig. 11. Effective number (N_{eff}) of model values for the REA1 and REA-ORIG methods. T : temperature; Pr : precipitation; DJF: winter; JJA: summer; subregions as in Fig. 1

3.2.3. CDFs of temperature and precipitation change

As mentioned in Section 2, one of the applications of the REA method consists of providing CDFs of climate change using Eq. (6) and PDFs by performing a derivative of the CDFs. Due to the relatively small sample size of our ensemble, we limit our analysis to the CDFs. Fig. 12 shows the CDFs obtained from Eq. (6) for the REA1, REA-ORIG and AVE (corresponding to a weight of 1 for each model) methods, DJF and JJA temperature and precipitation change (2081–2100 minus 1961–1980), the 5 sub-regions shown in Fig. 1 and the whole East Asia region.

First, we note that the CDFs in Fig. 12 exhibit both unimodal and multi-modal distributions, the latter evidenced by multiple changes in the steepness of the CDF. Multi-modal distributions imply the aggregation of models in different clusters within the whole ensemble. We also find varying widths and shapes of the distributions over the different regions. Intercomparison of the CDFs in Fig. 12 shows that the 3 methods produce generally consistent CDFs of change; however, some significant differences across methods are found. For example, markedly different CDFs between the different approaches are found for the JJA temperature change over the NWT region. The differences across methods are not systematic, but depend on the region, season and variable. However, it can be noted that, in many cases, the AVE and REA1 methods yield results that are closer to each other than the REA-ORIG method. This is be-



cause the latter employs the ‘model convergence’ criterion in the calculation of weights, which is based on the distance of each model-simulated change from the REA-averaged change. As mentioned, it has been argued that use of this criterion might artificially narrow the distribution of future changes; therefore, this criterion was removed in the present upgrade of the REA method. Instances of marked differences between the REA1 and AVE methods are also observed, due to the relatively low value of the effective number of contributing models in the REA1 case.

4. DISCUSSION

The issue of model weighting is still much debated. Given the uncertainty related to different climate sensitivities and regional responses of climate models to external forcings, climate-change projection clearly needs to be based on ensembles of different model simulations, eventually leading to CDFs or PDFs of change. The issue is thus whether in this process all models should be treated equally, the assumption most used to date, or whether they should be treated differently according to some pre-defined criterion. The latter option would involve some type of model ‘weighting’ intended in the broad sense. While, for example, past IPCC reports have mostly treated all model information in the same way (no weighting), it has been increasingly recognized that some models are more ‘credible’ than others, and thus their information should be given more weight (e.g. Murphy et al. 2004).

But, what determines the credibility of a model? It is virtually impossible to choose a universal credibility metric (CM), since climate models have a very large number of variables and parameters. This implies that a subjective element in the choice of CM is unavoidable. In previous studies, different choices of CM have been proposed, from the simplest one of the original REA method to the climate prediction index suggested by Murphy et al. (2004), which included 32 variables. Obviously, comprehensiveness in the choice of CM is at the expense of simplicity, and this may require a compromise choice. In addition, if a ‘steep’ weighting procedure is used, leading to small N_{eff} values, the presumed improved quality of the retained models could be more than compensated by the increase in sampling errors.

Many questions can be raised concerning the choice of CM, for example:

1. Does the model performance in reproducing present-day climate conditions increase the credibility of the model’s future climate projection? The answer to this question is certainly yes. However, since some model parameters are often tuned to reproduce present-day conditions, there is no guarantee that they provide a correct response to different climatic states. In other words, this is a necessary, but not a sufficient condition.

2. Should the choice of CM be tied to specific applications or should it be of a general nature? Certainly the specific application matters, but the overall model performance for a range of statistics may be a better indication of the model’s credibility.

3. Should the CM be local/regional or global? The local/regional performance is certainly a prerequisite, but the global performance also matters.

4. How to combine different CMs into a single weight? It is important that the model does well for the specific metric of interest, but it adds value for it to do well for all metrics. For example, the multiplicative approach taken in the REA method for the functions f_1 to f_5 (Eq. 9) is a stringent test of model performance, since it assumes that a model gets a high weight only if it performs well for all CMs.

5. If there are multiple options to the choice of given CMs along with their combinations, how do we know which one is right?

It is evident from the rather general answers to the first 4 questions and the lack of an answer to the fifth that there is high uncertainty underlying the issue of model weighting. This can be approached by either not doing any weighting (therefore avoiding the problem) or by considering the weighting itself as a source of uncertainty to be assessed. Here we argue for the latter approach, and, in fact, the REA method was designed with this viewpoint in mind. It is a simple and flexible approach, which, at the moment, includes multiple variables (T, P, SLP), multiple statistics (mean, interannual variability), local (f_1 to f_4) as well as non-local (f_5) CMs, and flexibility on how to combine the CMs (parameters m_1 to m_5 in Eq. 9). It is by no means fully comprehensive: more CMs and statistics could be added. For example, CMs based on extremes could be particularly targeted in impact assessment studies, or trend-based CMs and past-climate simulations substantially different from present could be used to better test the method in a climate-change context. All this can be easily incorporated into the general REA framework. The bottom line is that, since it is difficult to unambiguously and objectively find the ‘best way’ of

Fig. 12. Cumulative distribution function (CDF) of temperature (upper panels) and precipitation (lower panels) change over the 5 subregions shown in Fig. 1 and the whole East Asia region (EA) obtained using the REA1, REA-ORIG and AVE models for winter (DJF) and summer (JJA)

doing the weighting, the uncertainty associated with it should be explored through sensitivity tests to parameter and functional choices in the weighting schemes.

5. SUMMARY AND CONCLUSIONS

In the present paper we implemented a number of upgrades to the REA method originally introduced by Giorgi & Mearns (2002, 2003) in order to address some limitations of the original approach. These upgrades essentially modify the way the model weights are calculated. More specifically, compared to the old method, the new one removes the model convergence criterion and bases the calculation of weights on multiple variables (temperature, precipitation and circulation) and multiple statistics (mean and standard deviation). In addition, the use of an 'effective' number of models is introduced (as was suggested by one of the anonymous reviewers). The framework for calculating the weights is designed so that the contribution of different variables and statistics can be easily added or removed from the overall weight. This makes the REA framework quite simple to use and flexible to apply in different contexts.

The upgraded REA method was applied to the calculation of CDFs of changes in temperature and precipitation over different East Asia sub-regions based on the CMIP3 ensemble of model projections for the 21st century under the A1B emission scenario (IPCC 2000). The new and original methods were compared with each other and with simple ensemble averaging. In addition, the sensitivity to specific choices of variables and statistics in the weighting procedure was assessed.

We found that the 2 methods are consistent in terms of the broad-scale distribution of changes, as might be expected. However, significant differences can be found at the finer sub-regional scales; in particular, the updated method tends to produce finer scale structure than both simple ensemble averaging and the original method. This is because the model performance, and thus the weight, varies more across models as we move to finer spatial scales. At these scales, it may thus be important to weight the better performing models more strongly, which results in only a relatively small number of models (small N_{eff}) contributing effectively to the calculations. It should be recognized, however, that such aggressive weighting could pose problems of sampling error. In addition, the climate-change CDFs can show significant differences across the various weighting choices.

Most of the model sensitivity to the choice of weights is related to precipitation performance metrics. This is an indication that there is a wider spread across models in the simulation of precipitation statistics than of

temperature and SLP statistics. It also implies that it is important to include hydrologic cycle metrics in performance-based weights, as these are more stringent indicators of model performance. In particular, while the current set of CMIP3 AOGCMs tends to show a poor overall performance in the simulation of precipitation over East Asia, some individual models can perform quite well, and thus their results are more heavily weighted.

This observation was confirmed by the calculation of the effective number of models contributing to the REA calculations (N_{eff}), which, for the REA1 case, was as low as ~ 2 over some Chinese sub-regions. This implies that a minority of the models contributed to the REA calculations, i.e. that a much more stringent test of model performance was applied. The parameter N_{eff} can, in fact, provide valuable information on the spread of performance across models.

Being based on a broader set of model metrics, the updated REA method is clearly an improvement over the original one; however, as mentioned, the choice of specific metrics and ways of combining them into an individual weight still has a subjective component. In particular, additional, more-specifically targeted metrics could be used (e.g. trends and extremes) and alternatives to the multiplication of metrics could be employed to formulate the weights. Users of the REA method are thus strongly advised to always carry out an analysis of the method sensitivity to the choice and combination of different metrics and of the added value of using the weighting compared to simple averaging. In other words, we argue that method of combining the information for an ensemble of models based on model weighting represents an important aspect of the uncertainty underlying climate-change projections that should be explored. The REA method provides a simple and flexible framework to carry out such analysis.

Acknowledgements. This research was jointly supported by the European Union project ENSEMBLES (Ensemble-Based Predictions of Climate Change and their Impacts), the R&D Special Fund for Public Welfare Industry (meteorology) (GYHY200806010) and the National Basic Research Program of China (2009CB421407, 2007BAC03A01, 2006CB403707). The authors acknowledge the modeling groups associated with the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP's Working Group on Coupled Modeling (WGCM) for their roles in making available the WCRP CMIP3 multi-model dataset. Support for this dataset is provided by the Office of Science, US Department of Energy. The useful comments and feedback from 3 anonymous reviewers helped to substantially improve the quality of the paper. In particular, we thank one of the reviewers for his/her many useful and constructive comments, for introducing us to the use of the N_{eff} concept and for suggesting its addition (along with some specific revisions of the text) to the upgraded REA method.

LITERATURE CITED

- Christensen JH, Hewitson B, Busuioc A, Chen A and others (2007) Regional climate projections. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) *Climate change 2007: the physical science basis. Contribution of Working Group I to the 4th Assessment Report of the IPCC*. Cambridge University Press, Cambridge, p 848–940
- Gao XJ, Zhao ZC, Ding YH, Huang RH, Giorgi F (2001) Climate change due to greenhouse effects in China as simulated by a regional climate model. *Adv Atmos Sci* 18: 1224–1230
- Gao XJ, Xu Y, Zhao ZC, Pal JS, Giorgi F (2006) On the role of resolution and topography in the simulation of East Asia precipitation. *Theor Appl Climatol* 86:173–185
- Giorgi F (2005) Climate change prediction. *Clim Change* 73: 239–265
- Giorgi F, Bi X (2005a) Updated regional precipitation and temperature changes for the 21st century from ensembles of recent AOGCM simulations. *Geophys Res Lett* 32: L21715. doi: 10.1029/2005GL024288
- Giorgi F, Bi X (2005b) Regional changes in surface climate interannual variability for the 21st century from ensembles of global model simulations. *Geophys Res Lett* 32: L1370. doi: 10.1029/2005GL023002
- Giorgi F, Francisco R (2000) Evaluating uncertainties in the prediction of regional climate change. *Geophys Res Lett* 27:1295–1298
- 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 based on the Reliability Ensemble Averaging (REA) method. *Geophys Res Lett* 30:1629. doi: 10.1029/2003GL017130
- Giorgi F, Hurrell JW, Marinucci MR, Beniston M (1997) Elevation signal in surface climate change: a model study. *J Clim* 10:288–296
- Giorgi F, Whetton PH, Jones RG, Christensen JH and others (2001) Emerging patterns of simulated regional climatic changes for the 21st century due to anthropogenic forcings. *Geophys Res Lett* 28:3317–3320
- Hu ZZ, Yang S, Wu R (2003) Long-term climate variations in China and global warming signals. *J Geophys Res* 108(D19):4614. doi: 10.1029/2003JD003651
- IPCC (Intergovernmental Panel on Climate Change) (2000) *Special report on emissions scenarios*. Cambridge University Press, Cambridge
- Mearns LO, Hulme M, Carter TR, Leemans R, Lal M, Whetton PH (2001) Climate scenario development. In: Houghton JT, Ding Y, Griggs DJ, Noguer M and others (eds) *Climate change 2001: the scientific basis*. Cambridge University Press, Cambridge, p 739–768
- Meehl GA, Stocker TF, Collins WD, Friedlingstein P and others (2007) Global climate projections. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) *Climate change 2007: the physical science basis. Contribution of Working Group I to the 4th assessment Report of the IPCC*. Cambridge University Press, Cambridge, p 747–846
- Moise AF, Hudson DA (2008) Probabilistic predictions of climate change for Australia and southern Africa using the reliability ensemble average of IPCC CMIP3 model simulations. *J Geophys Res* 113:D15113. doi: 10.1029/2007JD009250
- Murphy JM, Sexton DMH, Barnett DN, Jones GS, Webb MJ, Collins M, Stainforth DA (2004) Quantification of modeling uncertainties in a large ensemble of climate change simulations. *Nature* 430:768–772
- New MG, Hulme M, Jones PD (2000) Representing twentieth century space time climate fields. II. Development of a 1901–1996 mean monthly terrestrial climatology. *J Clim* 13:2217–2238
- Räisänen J (2002) CO₂-induced changes in interannual temperature and precipitation variability in 19 CMIP2 experiments. *J Clim* 15:2395–2411
- Randall DA, Wood RA, Bony S, Colman R and others (2007) Climate models and their evaluation. In: Solomon S, Qin D, Manning M, Chen Y, Marquis M, Averyt KB, Tignor M, Miller HL (eds) *Climate change 2007: the physical science basis. Contribution of Working Group I to the 4th Assessment Report of the IPCC*. Cambridge University Press, Cambridge, p 590–662
- Tebaldi C, Nychka D, Mearns LO (2004) From global mean responses to regional signals of climate change: simple pattern scaling, its limitations (or lack of) and the uncertainty in its results. In: *Proc18th Conf on Probability and Statistics in the Atmospheric Sciences*. AMS Annu Meet, Seattle, WA, www.image.ucar.edu/~tebaldi/papers/ams_2004-paper.pdf
- Wigley TM, Raper SC (2001) Interpretation of high projections for global-mean warming. *Science* 293:451–454
- Xu CH, Shen XY, Xu Y (2007) An analysis of climate change in East Asia by using the IPCC AR4 simulations. *Ad Clim Change Res* 3:287–292 (in Chinese with English abstract)
- Yu RC, Zhang MH, Yu YQ, Liu YM (2001) Summer monsoon rainfalls over mid-eastern China lagged correlated with global SSTs. *Adv Atmos Sci* 18:179–196

Editorial responsibility: Bryson Bates, Wembley, Western Australia, Australia

*Submitted: December 18, 2008; Accepted: November 10, 2009
Proofs received from author(s): February 5, 2010*