Calculation of Average, Uncertainty Range, and Reliability of Regional Climate Changes from AOGCM Simulations via the "Reliability Ensemble Averaging" (REA) Method

FILIPPO GIORGI

Abdus Salam International Centre for Theoretical Physics, Trieste, Italy

LINDA O. MEARNS

NCAR, Boulder, Colorado

ABSTRACT

The "reliability ensemble averaging" (REA) method for calculating average, uncertainty range, and a measure of reliability of simulated climate changes at the subcontinental scale from ensembles of different atmosphere–ocean general circulation model (AOGCM) simulations is introduced. The method takes into account two "reliability criteria": the performance of the model in reproducing present-day climate ("model performance" criterion) and the convergence of the simulated changes across models ("model convergence" criterion). The REA method is applied to mean seasonal temperature and precipitation changes for the late decades of the twenty-first century, over 22 land regions of the world, as simulated by a recent set of nine AOGCM experiments for two anthropogenic emission scenarios (the A2 and B2 scenarios of the Intergovernmental Panel for Climate Change). In the A2 scenario the REA average regional temperature changes vary between about 2 and 7 K across regions and they are all outside the estimated natural variability. The uncertainty range around the REA average change as measured by $\pm$ the REA root-mean-square difference (rmsd) varies between 1 and 4 K across regions and the reliability is mostly between 0.2 and 0.8 (on a scale from 0 to 1). For precipitation, about half of the regional REA average changes, both positive and negative, are outside the estimated natural variability and they vary between about -25% and $+30\%$ (in units of percent of present-day precipitation). The uncertainty range around these changes ($\pm$ rmsd) varies mostly between about 10% and 30% and the corresponding reliability varies widely across regions. The simulated changes for the B2 scenario show a high level of coherency with those for the A2 scenario. Compared to simpler approaches, the REA method allows a reduction of the uncertainty range in the simulated changes by minimizing the influence of "outlier" or poorly performing models. The method also produces a quantitative measure of reliability that shows that both criteria need to be met by the simulations in order to increase the overall reliability of the simulated changes.

1. Introduction

Projections of climatic changes for the twenty-first century at the broad regional, or subcontinental, spatial scale (10^6–10^7 km^2) are based on transient simulations with coupled atmosphere–ocean general circulation models (AOGCMs) including relevant anthropogenic forcings, for example, due to greenhouse gases (GHG) and atmospheric aerosols (e.g., Kattenberg et al. 1996). To date, such projections have been characterized by a low level of confidence and a high level of uncertainty deriving from different sources (Visser et al. 2000; Giorgi and Francisco 2000b): estimates of future anthropogenic forcings, the response of a climate model to a given forcing, the natural variability of the climate system. Quantifying uncertainties in the projection of future climate scenarios used for impact assessments has been identified as a critical research need both in the climate and impacts research communities (e.g., Carter et al. 1999; Mearns et al. 2001), and has inspired a recent flurry of research (e.g., Jones 2000a,b; New and Hulme 2000; Katz 2001).

One of the primary factors of uncertainty is that different AOGCMs can simulate quite different regional changes even under the same anthropogenic forcing scenario (e.g., Kittel et al. 1998; Giorgi and Francisco 2000b; Whetton et al. 1996) and it is very difficult to ascertain which of the different AOGCMs are most reliable. Therefore, a comprehensive assessment of regional change projections needs to be based on the collective information from the ensemble of AOGCM simulations. For example, Giorgi et al. (2001a) identified

Corresponding author address: Dr. Filippo Giorgi, Abdus Salam International Centre for Theoretical Physics, P.O. Box 586, Trieste 34100, Italy.
E-mail: giorgi@ictp.trieste.it

emerging patterns of the regional spatial structure of climatic changes by searching for consistent regional change signals, both in sign and magnitude, across a wide range of simulations with different AOGCMs. However, they did not quantify either the uncertainty or the reliability of the simulated changes.

Two general "reliability criteria" have been used to assess, mostly in a qualitative way, the reliability of regional climate change simulations (e.g., Kattenberg et al. 1996; Giorgi et al. 2001b). The first is based on the ability of AOGCMs to reproduce different aspects of present-day climate: the better a model performance in this regard, the higher the reliability of the climate change simulation. We refer to this as the "model performance" criterion. The second criterion is based on the convergence of simulations by different models for a given forcing scenario, greater convergence implying higher reliability of robust signals that are little sensitive to the differences among models. We refer to this as the "model convergence" criterion.

To date, these two reliability criteria have not been used together in a quantitative way to establish measures of uncertainty and reliability in regional climate change projections. Furthermore, procedures to estimate regional changes based on the collective information of different AOGCM simulations have been used only in limited regional contexts (Hulme and Carter 2000). Of relevance in this regard is the analysis of control climate simulations from 15 coupled AOGCMs presented by Lambert and Boer (2001) as part of the first phase of the Coupled Model Intercomparison Project (CMIP1). The study shows some evidence that the mean climatological fields averaged over the ensemble of models compare better with the observed climatology than the fields produced by any of the individual models. This suggests that the collective information from ensembles of model simulations may be more reliable than that of any individual model. Techniques for extracting information from ensembles of different model simulations have also been proposed by Krishnamurti et al. (1999, 2000) and Palmer et al. (2000) within the context of seasonal prediction.

In this paper, we present a quantitative procedure for calculating average, uncertainty range, and collective reliability of regional climate change projections from ensembles of different AOGCM simulations based on the model performance and model convergence criteria. We call this method "reliability ensemble averaging" (REA), and apply it to a recent set of AOGCM transient climate change experiments for the twenty-first century for two anthropogenic forcing scenarios. We then compare the results from this new method with those from a simpler averaging procedure. In the study we consider simulated changes in surface air temperature and precipitation for the late decades of the twenty-first century compared to present-day climate over 22 land regions covering most land areas of the world. Throughout this paper the term "ensemble" refers to simulations with different models and not to different realizations with the same model.

Note that, at present, in our method we do not follow a probabilistic approach. Recently, a number of articles have addressed the issue of estimating probability density functions (PDFs) for future climate variables (e.g., Jones 2000a,b; Schneider 2001; Wigley and Raper 2001; Andronova and Schlesinger 2001). For example, Wigley and Raper (2001) calculate the probabilities of globally averaged future climate conditions based on the emission scenarios developed by the Intergovernmental Panel on Climate Change (IPCC; Smart et al. 2000). However their method relies on a number of assumptions regarding the shape of the PDFs for the major uncertainty factors they consider (e.g., climate sensitivity, emission scenarios, radiative forcing). Jones (2000a,b) assume uniform PDFs for both regional and global temperature changes in a study for the south Australia region. While interpreting calculations of future climate in probabilistic terms is an important step, we prefer to employ a quantitative but nonprobabilistic method because we prefer not to make assumptions about the distribution of factors such as the climate model sensitivity, especially at the regional scale. In addition, we have a very limited sample size (nine AOGCMs for each scenario), which makes it very difficult to identify PDFs of specific regional climatic changes. It is more difficult at this stage of research to produce PDFs for regional climate change than for mean global conditions (the latter can be determined using simple climate models that can generate easily many simulations), since generating the large number of runs desirable for constructing PDFs at the regional scale is currently not feasible.

2. Experiments and methods

The set of experiments analyzed here includes 18 different AOGCM simulations (Table 1), 9 for each of 2 anthropogenic forcing scenarios, that is, the A2 and B2 marker scenarios developed by the IPCC (Smart et al. 2000). These scenarios, which are derived from different assumptions of future social and technological development, are considered as "high" (A2) and "medium low" (B2) in terms of cumulative GHG emissions (Smart et al. 2000).

Our analysis considers the difference in mean temperature and precipitation between the periods of 2071–2100 (future climate) and 1961–90 (present-day climate), a quantity we refer to as "change." Changes are calculated for December–January–February (DJF) and June–July–August (JJA). The data are interpolated onto a common 0.5° grid and are averaged over 22 regions covering nearly all land areas of the world and identified by Giorgi and Francisco (2000a,b; see Fig. 1). Only land areas are considered and the interpolation procedure is described by Giorgi and Francisco (2000a). For one of the models, three realizations were available, and

TABLE 1. List of AOGCM experiments and corresponding global temperature change, ΔT (K), defined as the difference on global annual surface air temperature between the periods 2071–2100 and 1961–90. Also given are references for the models and/or simulations analyzed.

Institution/model	Reference	ΔT-A2	ΔT-B2
CCSR-NIES/version 2	Nozawa et al. (2001)[a]	4.53	3.38
MRI/version 2	Noda et al. (1999)	1.25	0.92
CCC/CGCM2	Flato and Boer (2001)[b]	3.59[e]	2.49[e]
CSIRO/Mk2	Gordon and O'Farrell (1997)	3.50	2.71
NCAR/CSM	Dai et al. (2001)	2.29	1.71
DOE-NCAR/PCM	Washington et al. (2001)[c]	2.35	1.80
GFDL/R30-c	Knutson et al. (1999)	2.87	2.18
MPI-DMI/ECHAM4-OPYC	Stendel et al. (2000)	3.11	2.34
UKMO/HADCM3	Johns et al. (2001)[d]	3.38	2.42

[a] Manuscript submitted to *J. Meteor. Soc. Japan.*
[b] Manuscript submitted to *Geophys. Res. Lett.*
[c] Manuscript submitted to *Climate Dyn.*
[d] Personal communication.
[e] Average of three realizations.

the data analyzed here refers to their average. (Giorgi and Francisco 2000a,b show that 30-yr means do not vary substantially between different realizations of the same experiment). The global temperature changes for each simulation are reported in Table 1. It can be seen that the full range of simulated global temperature changes for a given scenario is about 3.5 K.

As a base for comparison of our REA method we use a simpler approach to the development of climate change estimates and associated uncertainty range. In this approach, taking as an illustrative example the temperature T, the estimated change is given by the ensemble average of all model simulations, that is,

$$\overline{\Delta T} = \frac{1}{N} \sum_{i=1,N} \Delta T_i, \tag{1}$$

where N is the total number of models, the overbar indicates the ensemble averaging and Δ indicates the model-simulated change.

In its generalized form, the uncertainty is measured by the corresponding root-mean-square difference (rmsd), or δ, defined by

$$\delta_{\Delta T} = \left[\frac{1}{N} \sum_{i=1,N} (\Delta T_i - \overline{\Delta T})^2 \right]^{1/2}. \tag{2}$$

The uncertainty range is then given by $\pm \delta_{\Delta T}$ and is centered around $\overline{\Delta T}$. To put this definition into perspective, if the changes followed a Gaussian PDF, the rmsd would be equivalent to the standard deviation and the $\pm \delta_{\Delta T}$ would approximately cover the 68.3% confidence interval. Note that the direct ensemble averaging does not explicitly take into account the reliability criteria and weighs equally all model simulations.
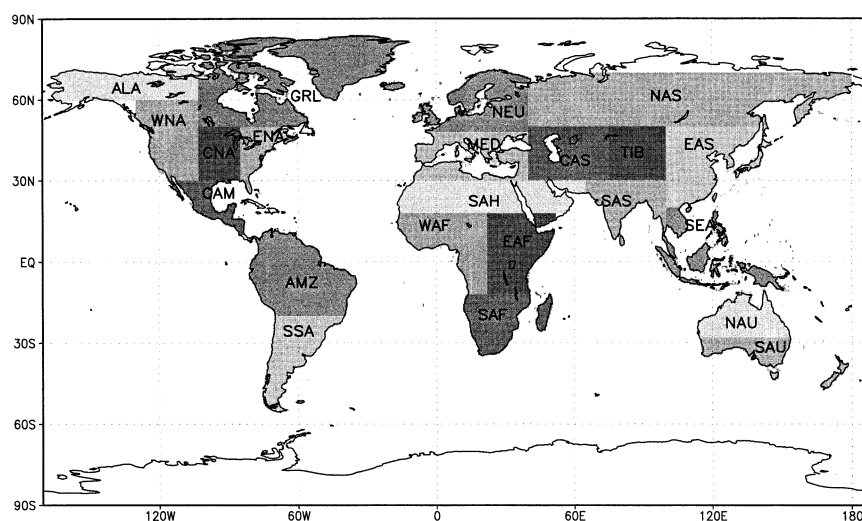

FIG. 1. Regions used in the analysis presented in this work. SAU, NAU, AMZ, SSA, CAM, WNA, CNA, ENA, ALA, GRL (and northern territories), MED, NEU, WAF, EAF, SAF, SAH, SEA, EAS, SAS, CAS, TIB, and NAS. Giorgi and Francisco (2000a) provide the regions' definition in terms of lat and lon.

In our REA method, the average change, $\widetilde{\Delta T}$, is given by a weighted average of the ensemble members, that is,

$$\widetilde{\Delta T} = \tilde{A}(\Delta T) = \frac{\sum_i R_i \Delta T_i}{\sum_i R_i}, \qquad (3)$$

where the operator $\tilde{A}$ denotes the REA averaging and R_i is a model reliability factor defined as

$$R_i = [(R_{B,i})^m \times (R_{D,i})^n]^{[1/(m \times n)]}$$

$$= \left\{ \left[\frac{\epsilon_T}{\mathrm{abs}(B_{T,i})} \right]^m \left[\frac{\epsilon_T}{\mathrm{abs}(D_{T,i})} \right]^n \right\}^{[1/(m \times n)]}. \qquad (4)$$

In Eq. (4), $R_{B,i}$ is a factor that measures the model reliability as a function of the model bias ($B_{T,i}$) in simulating present-day temperature, that is, the higher the bias the lower the model reliability. Here the bias is defined as the difference between simulated and observed mean temperature for the present-day period of 1961–90. Here $R_{D,i}$ is a factor that measures the model reliability in terms of the distance ($D_{T,i}$) of the change calculated by a given model from the REA average change, that is, the higher the distance the lower the model reliability. Therefore, the distance is a measure of the degree of convergence of a given model with the others. In other words, $R_{B,i}$ is a measure of the model performance criterion while $R_{D,i}$ is a measure of the model convergence criterion.

The choice of the particular function (4) is based on its simplicity and on the requirement that both criteria need to be met in order to yield a high reliability for a given model simulation. This is effectively achieved by using the product of the $R_{B,i}$ and $R_{D,i}$ factors. Note that in the experiments examined here there was little relationship between the biases and distances of individual simulations. We calculated the correlation between biases and distances across models for a given region and season, and found that the correlation was generally small (mostly less than 0.5 except for a few instances) and was statistically significant at the 95% confidence level only in a few regional cases. This suggests that, for most cases, a large individual model bias does not imply a corresponding large distance and vice versa, that is, that the main model outliers in the future climate simulation are not necessarily those that show the poorest performance in reproducing present-day climate. This is not surprising in view of the fact that often some model parameters are "tuned" to reproduce present-day climate but may be characterized by a pronounced sensitivity to strong climatic forcings.

The distance $D_{T,i}$ is calculated using an iterative procedure. A first guess of $D_{T,i}$ is the distance of each ΔT_i from the ensemble average change $\overline{\Delta T}$ of Eq. (1), that is, $[D_{T,i}]_1 = [\Delta T_i - \overline{\Delta T}]$. The first guess values are then used in Eqs. (3) and (4) to obtain a first-order REA average change $[\widetilde{\Delta T}]_1$, which is then used to recalculate

the distance of each individual model as $[D_{T,i}]_2 = [\Delta T_i - [\widetilde{\Delta T}]_1]$ and repeat the iteration. Typically, this procedure converges quickly after several iterations. Note that the distance from the REA average is only an estimated measure of the model convergence criterion given that future conditions are not known. This does not imply that the REA average represents the "true" climate response to a given forcing scenario but only that the REA average represents the best estimated response.

The parameters m and n in Eq. (2) can be used to weigh each criterion. For most calculations in this work, m and n are assumed to be equal to 1, which gives equal weight to both criteria. However, they could be different if different weight is given to the two criteria (see discussion in section 3e). Also, R_B and R_D are set to 1 when B and D are smaller than ϵ, respectively. Essentially, Eq. (4) states that a model projection is "reliable" when both its bias and distance from the ensemble average are within the natural variability, so that $R_B = R_D = R = 1$. As the bias and/or distance grow, the reliability of a given model simulation decreases. Note that, for R_B and R_D lower than 1, ϵ cancels out in the REA operator and the reliability factor effectively reduces to the reciprocal of the product of bias and distance.

The parameter ϵ in Eq. (2) is a measure of natural variability in 30-yr average regional temperature and precipitation. In order to calculate ϵ, we computed time series of observed regionally averaged temperature and precipitation for the twentieth century over our 22 regions from the dataset of New et al. (2000). We then computed 30-yr moving averages of the series after linearly detrending the data (to remove century-scale trends) and estimated ϵ as the difference between the maximum and minimum values of these 30-yr moving averages. Alternatively, ϵ could be defined as the difference between some upper and lower percentiles of the 30-yr moving averages, a definition that would be less dependent on the length of the observing record. However, given that the record is relatively short, and hence that ϵ is likely a lower estimate of variability, the use of the maximum and minimum values is more appropriate for this study.

In order to calculate the uncertainty range around the REA average change, we first calculate the REA rmsd of the changes, $\tilde{\delta}_{\Delta T}$, defined by

$$\tilde{\delta}_{\Delta T} = [\tilde{A}(\Delta T_i - \widetilde{\Delta T})^2]^{1/2} = \left[\frac{\sum_i R_i(\Delta T_i - \widetilde{\Delta T})^2}{\sum_i R_i} \right]^{1/2}. \qquad (5)$$

Equation (5) is the counterpart of Eq. (2) when the reliability criteria are accounted for. The upper and lower uncertainty limits are thus defined by

$$\Delta T_+ = \widetilde{\Delta T} + \tilde{\delta}_{\Delta T} \qquad (6a)$$

$$\Delta T_- = \widetilde{\Delta T} - \tilde{\delta}_{\Delta T}. \qquad (6b)$$

From this definition the total uncertainty range is giv-

en by $\Delta T_+ - \Delta T_- = \pm \tilde{\delta}_{\Delta T} = 2\tilde{\delta}_{\Delta T}$. The uncertainty limits defined by Eqs. (5), (6), and (2) encompass the range of changes defined by the rmsd with and without the explicit contribution of the reliability criteria. As mentioned, a rigorous probabilistic interpretation of these uncertainty limits is not possible because the PDFs of the changes are not known due to the small sample size. On the other hand an analogy may provide some guidance in this regard. Under the assumption that the changes are distributed following a Gaussian PDF, the rmsd is equivalent to the standard deviation, so that the $\pm \delta$ range would imply a 68.3% confidence interval. For a uniform PDF, that is, one in which each change has the same probability of occurrence, the $\pm \delta$ range implies a confidence interval of about 58%. In the REA method, the normalized reliability factors of Eq. (4) can be interpreted as the likelyhood of a model outcome, that is, the greater the factor, the greater the likelihood associated with the model simulation. As shown in section 3d, the distribution of the reliability factors is irregular and it changes from region to region. Therefore, assuming that the actual PDF of the changes is somewhere between a uniform and a Gaussian PDF, the $\pm \tilde{\delta}$ range can be interpreted as approximately representing a confidence interval of 60%–70%. Also note that the choice of an uncertainty range of $\pm \tilde{\delta}$ is only for illustrative purposes. A larger range, say $\pm 2\tilde{\delta}$, could be used to account for a greater confidence interval.

A quantitative measure of the collective model reliability ($\tilde{\rho}$) in the simulated changes can be obtained by applying the REA averaging operator to the reliability factor, that is,

$$\tilde{\rho} = \tilde{A}(R) = \frac{\sum_i R_i^2}{\sum_i R_i}. \tag{7}$$

In other terms, the collective reliability is given by the REA average of the individual model reliability factors. This definition of reliability is consistent with the fact that different model simulations are weighted differently in the calculation of the REA average.

Note that the reliability $\tilde{\rho}$ depends not only on the bias and distance but also on how these relate to the natural variability, which changes from region to region. In fact, while in Eqs. (3) and (5) ϵ cancels out under the condition of B and D greater than ϵ, in Eq. (7) ϵ does not cancel out. The underlying assumption is that more stringent conditions on B and D are required to increase the reliability over regions characterized by lower natural variability. The quantity $\tilde{\rho}$ can thus be interpreted as a reliability measure of the REA average and uncertainty range in relation to a certain level of natural variability. Given that different functions can be used to define the factors R_B and R_D, the relevance of $\tilde{\rho}$ is not in its absolute value, but as a tool to intercompare the level of reliability across regions.

Finally, we also define the two quantities:

$$\overline{R_B} = \frac{1}{N} \sum_{i=1,N} R_{B,i} \tag{8a}$$

$$\overline{R_D} = \frac{1}{N} \sum_{i=1,N} R_{D,i}, \tag{8b}$$

which provide a measure of the collective model reliability with respect to the two criteria separately.

3. Results

a. Model performance in reproducing present-day average climate

Prior to discussing the climate change results, it is useful to provide an overall view of the model performance in reproducing present-day climate. Figures 2 and 3 show, for each region, the ensemble average bias along with the largest positive and negative individual model biases for temperature and precipitation. The ensemble average regional temperature biases are mostly within ± 2 K (Figs. 2a,b). Noticeable exceptions are the Alaska (ALA) region in winter, where the ensemble average bias is about 6 K, and the Greenland (GRL) and Mediterranean (MED) regions in summer, where the average bias exceeds 3 K. The range of individual model biases (i.e., the difference between the largest positive and largest negative individual model biases) varies considerably across regions, and it is generally minimum in tropical and subtropical regions (order of 3–5 K) and maximum in mid- and high-latitude regions, where it can exceed 12 K. Except for the case of Alaska in DJF, where all models exhibit a warm bias, biases of both signs are found in the ensemble. The magnitude of the individual model biases varies from a few K to over 10 K (8 cases of bias in excess of 10 K are found).

Figures 3a,b show a wide range of precipitation biases across regions. The ensemble average biases are mostly within $\pm 50\%$ of observed precipitation, with a general predominance of positive biases, that is, an overestimate of precipitation, especially in the cold season. Precipitation is greatly overestimated over three regions, ALA and Tibet (TIB) in DJF and Sahara (SAH) in JJA, where the ensemble average bias is in excess of 200%. In the case of the Sahara region this large percentage overestimate is amplified by the very low observed precipitation amounts, while the problem over the Tibet region may be related to the generally coarse model resolution (300–500 km for the models considered), which would result in a poor representation of the Tibetan Plateau. The intermodel range of biases varies from about 30%–50% to over 200% across regions, and a number of instances of positive individual model bias greater than 100% can be observed. A few instances of very large individual model negative bias (precipitation underestimate) of up to 80%–90% also occur, in particular over the MED and Central Asia (CAS) in JJA and SAH in DJF. As with temperature, in most cases precipitation biases of both signs are found within the ensemble.
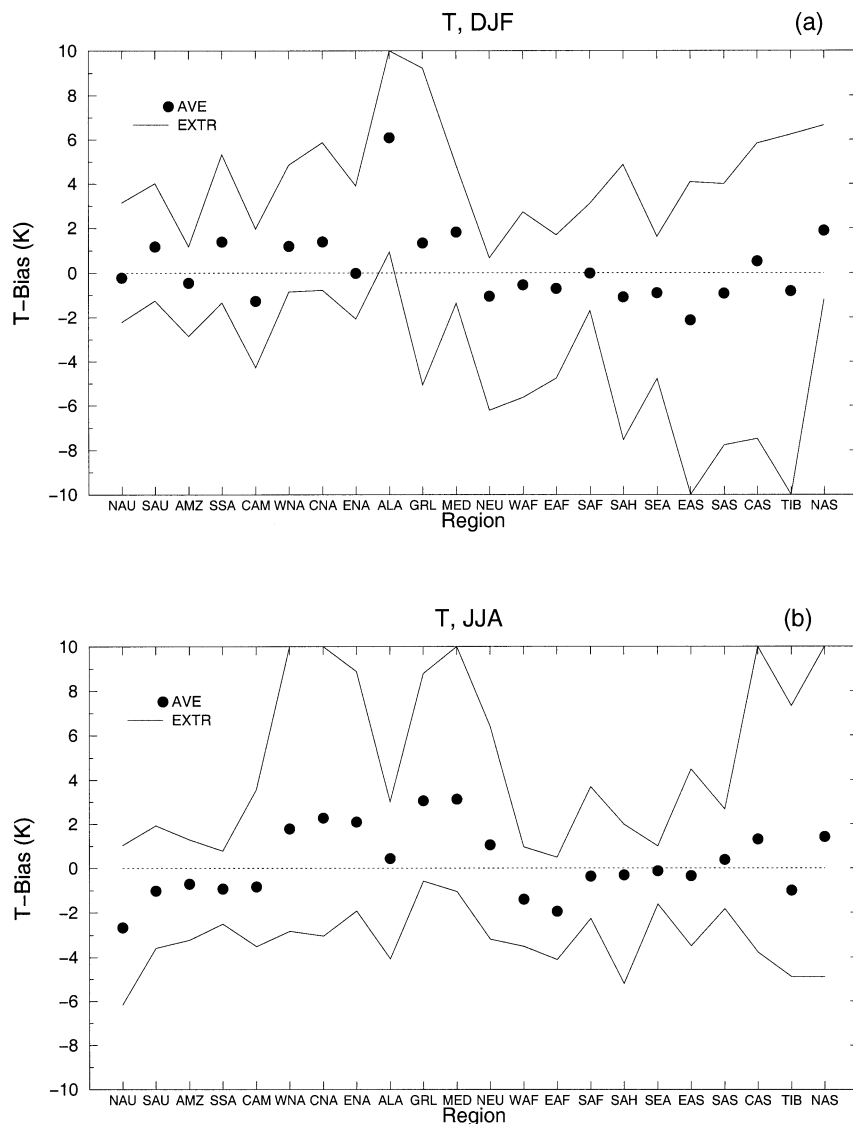
FIG. 2. Ensemble average temperature bias (AVE, dark circles) and largest positive or negative individual model temperature bias (EXTR, continuous lines) over the 22 regions of Fig. 1: (a) DJF, (b) JJA. Units are degrees K. Values with magnitude greater than 10 K have been set equal to 10 K for plotting purposes. The dotted line indicates the 0 level.

Figures 2 and 3 thus provide a picture of a wide range of model performance in reproducing present-day mean climate conditions, with occurrences of large errors, both by individual models and by the overall ensemble. Although individual model results are not shown, there was no model that performed best over all regions, and all models contributed at least one maximum positive or negative regional bias in Figs. 2 and 3. However, there were two to three models that exhibited an overall largest number of maximum bias within the ensemble. The marked interregional and intermodel variability of the biases illustrated by Figs. 2 and 3 and the occurrence of large biases clearly point to the need of including the model performance criterion in the evaluation of the simulated changes.

b. Estimates of change and uncertainty range

Figures 4 and 5 show the following variables for regional temperature and precipitation in the A2 scenario. The REA average change [Eq. (3)] along with the corresponding upper and lower uncertainty limits [Eqs. (6a) and (6b)] the ensemble average change plus/minus the corresponding rmsd and the natural variability estimates. Also shown are the highest and lowest simulated changes by individual models within the ensemble. The difference between these latter values can be considered as a measure of maximum uncertainty range that does not take into account the collective information of the ensemble of simulations. Note that in this paper the units
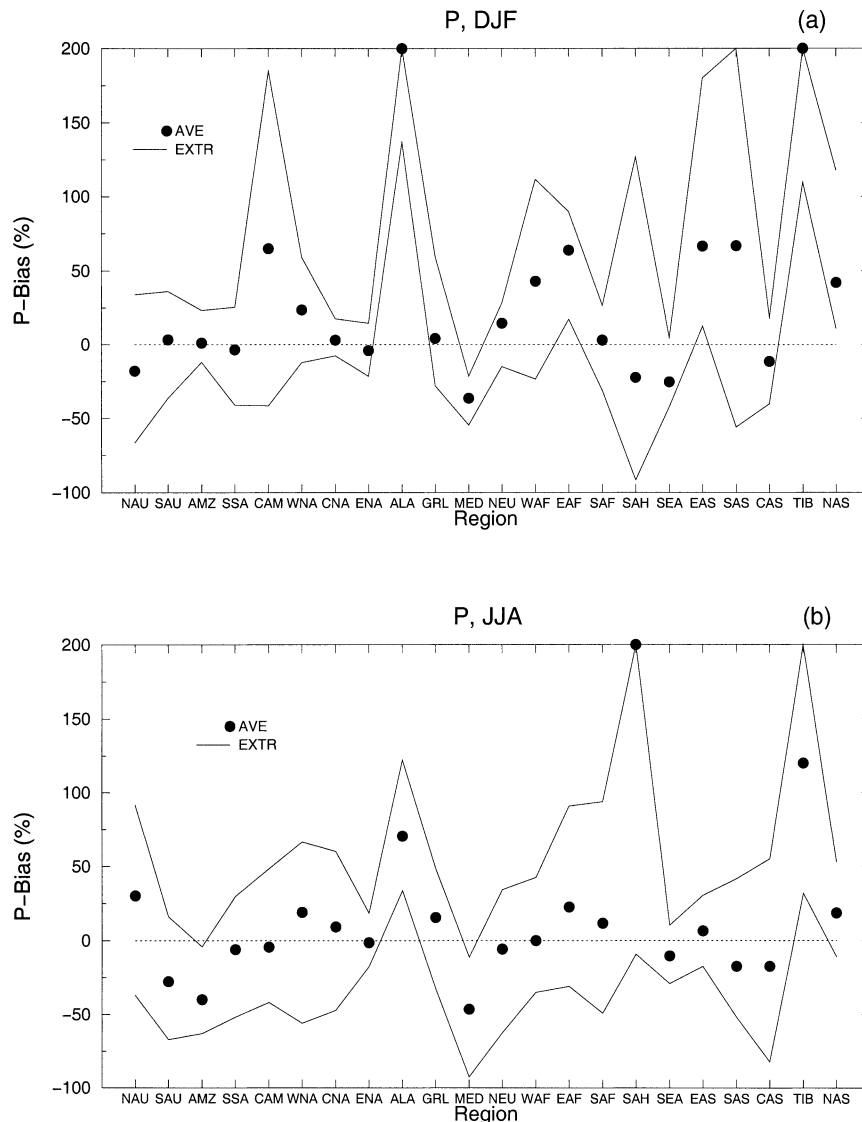
FIG. 3. Same as in Fig. 2 but for (a) DJF precipitation and (b) JJA precipitation. Units are percent of observed precipitation. Values greater than 200% have been set equal to 200% for plotting purposes.

used for the precipitation change is percentage of present-day precipitation.

For temperature, the difference between the REA average change and ensemble average change is of the order of a few tenths of K to about 1 K across regions. In DJF, maximum warming of 6.9–7.2 K is found over the Northern Hemisphere high-latitude regions of ALA, GRL, and northern Asia (NAS). The maximum northern high-latitude warming has been consistently found in previous AOGCM simulations (e.g., Giorgi and Francisco 2000a,b) and can be attributed in good part to the snow–ice albedo feedback mechanism, by which warming causes a decrease in snow and ice cover, and thus a decrease in the local albedo. This results in an increase of the absorption of solar radiation at the surface that

enhances the warming (e.g., Giorgi et al. 1997). Minimum DJF warming, of the order of 2.5–4 K is found over tropical and subtropical regions, along with the MED region.

In JJA the warming shows lower interregional variations than in DJF, mostly because of reduced warming over high-latitude northern regions. Maximum warming of 5–6 K occurs over the CAS, TIB, and NAS regions along with western North America (WNA) and central North America (CNA). Minimum warming in JJA of 2.2–2.7 K is calculated over southern South America (SSA), south Asia (SAS), and Southeast Asia (SEA).

The estimates of natural variability (ϵ_T) for DJF temperature are of the order of 0.25 to about 1.6 K, with maxima over northern high- and midlatitude continental

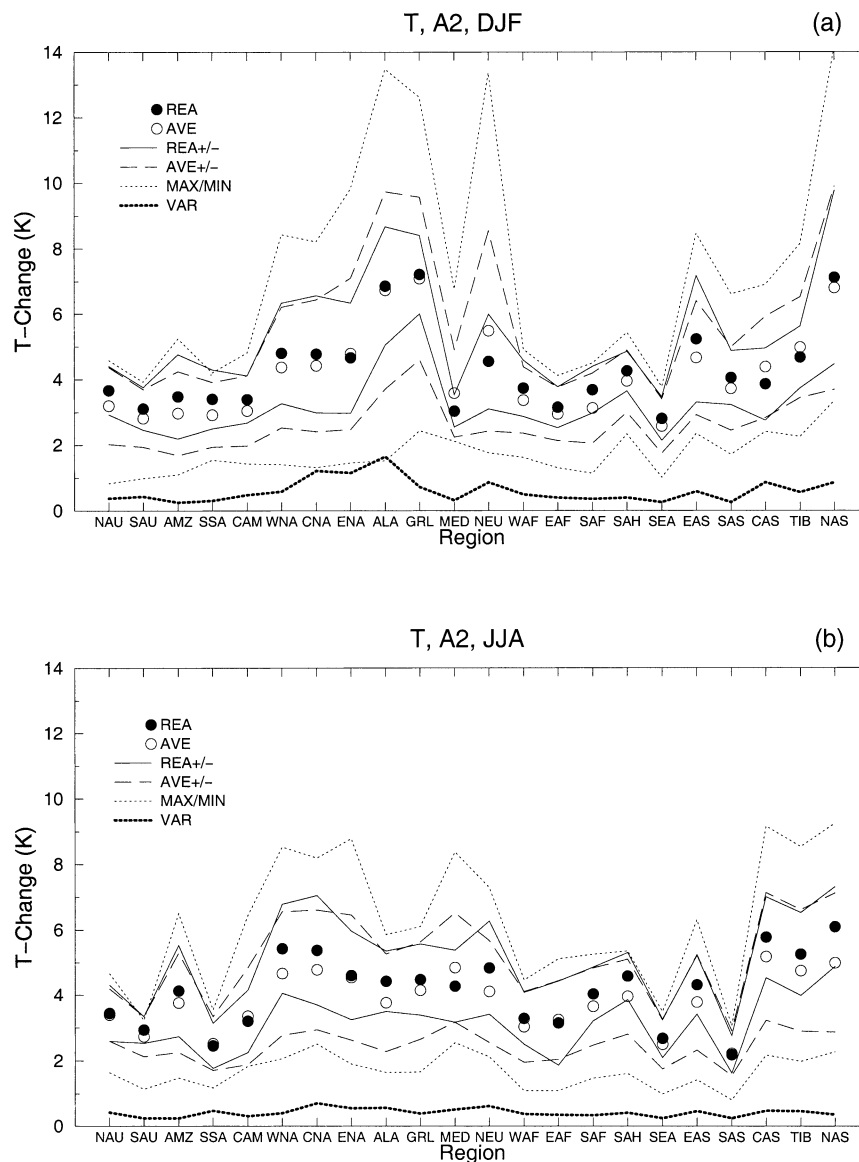FIG. 4. REA average change [REA, dark circles, see Eq. (3)] and corresponding upper and lower REA uncertainty limits [REA ±, continuous lines, see Eqs. (6a) and (6b)], ensemble average changes [AVE, open circles, see Eq. (1)], and corresponding ± rmsd [AVE±, dashed lines, see Eq. (2)], maximum and minimum changes simulated by individual models in the ensemble (MAX/ MIN, dotted lines), and estimated natural variability values (VAR, bold dotted line, see text) for the A2 scenario over the 22 regions of Fig. 1: (a) temperature, DJF; (b) temperature JJA. Units are degrees K.

regions. This is also likely due to the snow–ice albedo feedback mechanism, whereby relatively warm (or cold) periods are enhanced (or reduced) by the feedback process. Therefore, the largest temperature change signal occurs in regions characterized by the largest natural variability (e.g., Stott and Tett 1998; Fyfe and Flato 1999). In JJA there is less interregional variability of ϵ_T, which remains mostly in the range of about 0.25– 0.7 K. Note that the estimates of natural variability in Figs. 4 and 5 are generally consistent with analogous ones obtained by Giorgi and Francisco (2000a,b) from

multiple realizations of model experiments. All the simulated REA average and ensemble average regional warming values are well above the natural variability estimates.

The full range of individual model-simulated changes (dotted lines) is highly variable from region to region, mostly 3–12 K for DJF and 2–7 K for JJA. These ranges show pronounced maxima over mid- and high-latitude regions particularly in the cold season, that is, the intermodel range of warming is maximum in the regions where both the signal and the natural variability are
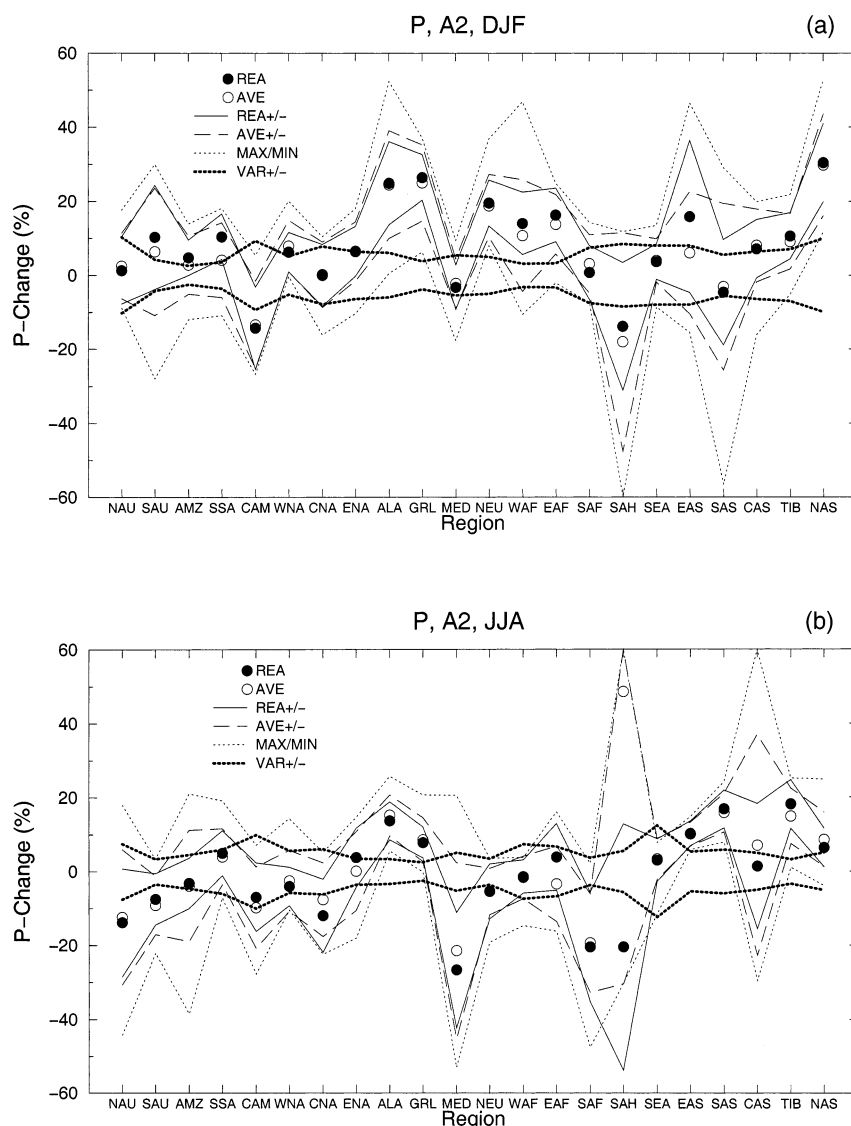
FIG. 5. Same as Fig. 4, but for (a) DJF precipitation and (b) JJA precipitation. Units are percent of present-day precipitation. Values with magnitude greater than 60% have been set equal to 60% for plotting purposes.

largest. This suggests on the one hand that the snow–ice albedo feedback gives an important contribution to the intermodel spread of results over these regions, and on the other hand that the treatment of snow and ice processes and related feedbacks is an important factor in determining the differences across model simulations.

We can compare the intermodel range of simulated changes of Figs. 4a,b with the corresponding range of model biases of Figs. 2a,b. This comparison shows that the range in bias is generally greater than the range in change, which is in agreement with the conclusions of an analysis by Kittel et al. (1998) of a previous generation set of AOGCM simulations. Especially during the cold season, a number of regions characterized by pronounced ranges in simulated change do not show

correspondingly large ranges in bias [most noticeably the ALA, northern Europe (NEU), and NAS regions]. This may be an indication that in the models some parameters within the ice physics representation are optimized for present-day conditions but show a different response to the GHG forcing. When comparing the ensemble average changes and corresponding average biases we can notice some correlation. The correlation between ensemble average changes and biases across regions is equal to 0.473 for DJF and 0.613 for JJA, in both cases statistically significant at the 95% confidence level. This indicates that, when considered collectively across models, the spatial structure of the simulated changes appears affected by the spatial structure of the biases.

The uncertainty range as defined by the rmsd [Eq. (2)], that is, $\pm\delta_{\Delta_T}$, is of the order of 2–5.5 K in DJF and 1.5–4 K in JJA (dashed lines in Fig. 4). A pronounced interregional variation of uncertainty range is found, with maxima in cold climate mid- and high-latitude regions. Use of the REA method (continuous lines) tends to overall reduce the uncertainty range, although not in all cases. This is because the contribution of model outliers and/or strongly biased models is effectively filtered out. With the REA method, the uncertainty estimates vary between 1 and 4 K in DJF and between 1 and 3.5 K in JJA across regions. The few cases in which the REA uncertainty range is greater than the corresponding rmsd-based uncertainty range generally occur when large biases are found in correspondence of small distances.

Moving to precipitation, Figs. 5a,b show that the differences between ensemble average and REA average values are generally less than 10% (in units of percent of present-day precipitation). In fact, in a number of cases the estimates of change with the two methods are quite close to each other. A noticeable exception is the SAH region in JJA, where a large difference is found between the ensemble average and the REA average. The main reason for this is that most of the model simulations exhibit a large precipitation bias over this region, in excess of 200%, with the exception of 3 simulations that have a bias of less than 100% (only 1 model has a bias lower than 10%). As a result, since the REA average is dominated by three simulations only, it can be substantially different from the ensemble average. In DJF we mostly find precipitation increases in the range of a few percent to about 30%, while decreases of about −15% occur over the Central America (CAM) and SAH regions. Smaller negative precipitation changes in DJF can be observed over the MED and SAS regions. The predominance of positive precipitation changes is consistent with an intensification of the hydrologic cycle in warmer conditions. In 11 out of 22 regions the DJF REA average change is well outside the estimated natural variability: positive over Southern Australia (SAU), SSA, ALA, GRL, NEU, NAS, and eastern Asia (EAS), Western and Eastern Africa (WAF and EAF); and negative over CAM and SAH.

In JJA the precipitation changes are more equally distributed between positive and negative except over the Asian regions, where the changes are mostly positive. Both negative and positive REA average changes are outside the natural variability. These include decreases over the Australian regions [northern Australia (NAU) and SAU], CNA, Southern Africa (SAF), SAH, and MED, where the REA average changes are in the range of −10% to −25%. Increases in JJA REA average precipitation outside the natural variability are found over ALA, GRL, TIB, EAS, and SAS. These increases are of the order of 10%–20%. Note that the models indicate an intensification of summer monsoon precipitation over Asia, a result not found in a previous set

of simulations (Giorgi and Francisco 2000b). This may be due to the decreased amounts of twenty-first century sulfate aerosols assumed in the A2 and B2 scenarios compared to the earlier scenarios used in the simulations analyzed by Giorgi and Francisco 2000b (see also Giorgi et al. 2001a).

The range of individual model-simulated changes can be large (dotted line in Figs. 5a,b). Over some regions this range can be as large as 60%–90%, for example, the SAH, EAS, and SAS regions in DJF and the NAU, MED, SAH, and CAS regions in JJA. This implies a pronounced spread of different model results over these regions. For the other regions the spread of model results is between 20% and 50%. The uncertainty range estimates based on the rmsd ($\pm\delta_{\Delta_P}$) is mostly in the range of 10%–40%, with some exceptions (SAS and SAH in DJF; MED, SAH, and CAS in JJA). The uncertainty range is reduced in most cases when using the REA method ($\pm\tilde{\delta}_{\Delta_P}$ mostly varying between 5% and 20%) because of the filtering of outliers.

It should be noted that only in a small number of cases do all models agree on the sign of the simulated precipitation change: The high-latitude Northern Hemisphere regions in DJF and ALA, GRL, EAS, SAS, and TIB in JJA. A greater number of cases is found in which the whole REA uncertainty range is of the same sign. Comparison of the changes in Fig. 5 and the biases in Fig. 3 clearly shows that the biases are markedly larger in magnitude. The correlation across regions of ensemble average changes and biases is small for DJF (0.329), but large for JJA (0.773), indicating that for the latter season the collective model biases somewhat influence the spatial structure of the collective simulation of changes.

For the B2 scenario the REA average and uncertainty analysis revealed a model behavior similar to that found for the A2 scenario. It is therefore more instructive to analyze the A2:B2 ratio of REA average changes to evaluate the extent to which the scenario affects the collective climate change simulation by the models. Figures 6a,b show the ratio of A2:B2 REA average regional changes of temperature and precipitation, respectively. For temperature, this ratio is quite uniform across regions, 1.30–1.52, which is also close to the ratio of the ensemble average global temperature changes for the two scenarios (about 1.34). This result indicates that the regional structure of simulated temperature change is a little sensitive to the details of the forcing scenarios, and implies that regional temperature changes at the subcontinental scale can be "scaled" by the global temperature change without addition of substantial uncertainty (Mitchell et al. 1999).

A higher interregional variability of the A2:B2 ratios of changes is found for precipitation, which is at least partially due to the smaller signal-to-noise ratio for the precipitation changes compared to the temperature changes. For the vast majority of cases, however, the ratio is greater than 1, indicating that the magnitude of
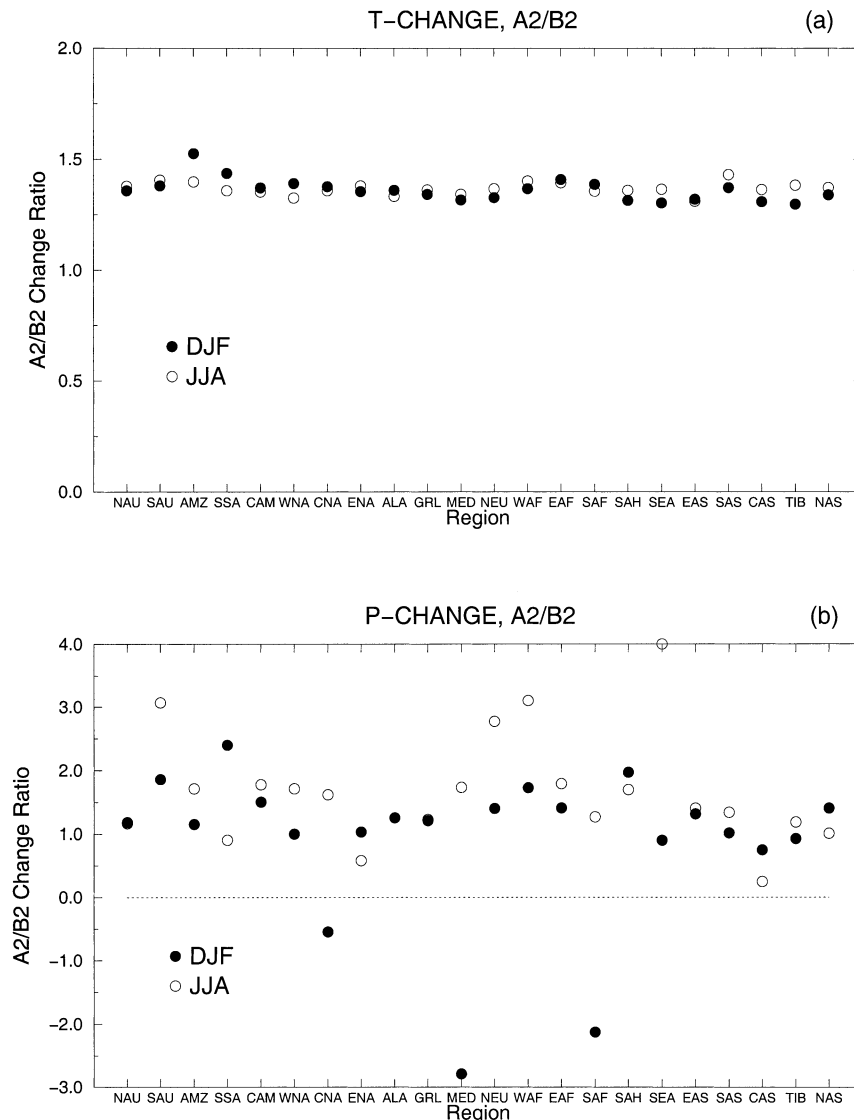
FIG. 6. Ratio of A2:B2 REA average changes over the 22 regions of Fig. 1 for DJF (dark circles) and JJA (open circles): (a) temperature and (b) precipitation. The dotted line indicates the 0 level.

the precipitation changes is amplified in the A2 experiments (i.e., the experiment with greater cumulative GHG forcing) compared to the B2 experiments regardless of the sign of the change. For most regions the A2: B2 ratio of precipitation changes is between 1 and 2. Only in three cases, CNA, MED, and SAF in DJF, this ratio is negative, that is, the REA average change simulated by the two scenarios is of opposite sign, and for all three cases the change is extremely small (see Figs. 5a,b). This result shows that there is a high level of consistency between the A2 and B2 scenarios in the sign of the simulated average precipitation changes. Noticeable cases of A2: B2 ratios of precipitation change greater than 2 include SAU in both seasons, SSA in DJF, and SAU, NEU, WAF, and SEA in JJA. The cor-

relation across regions between the A2 and B2 simulated REA average precipitation changes is high, about 0.95, indicating an overall high level of coherency between the changes in the two scenarios.

c. Reliability analysis

Figures 7 and 8 show the reliability $\tilde{\rho}$ [Eq. (7)] for the A2 scenario corresponding to the REA averages of Figs. 4 and 5 along with the collective reliability factors with respect to the model performance and convergence criteria [$\overline{R_B}$ and $\overline{R_D}$ in Eqs. (8a) and (8b), respectively]. It is evident that, both for temperature and precipitation, $\overline{R_D}$ is consistently greater than $\overline{R_B}$, which indicates that the model convergence in the simulation of changes is
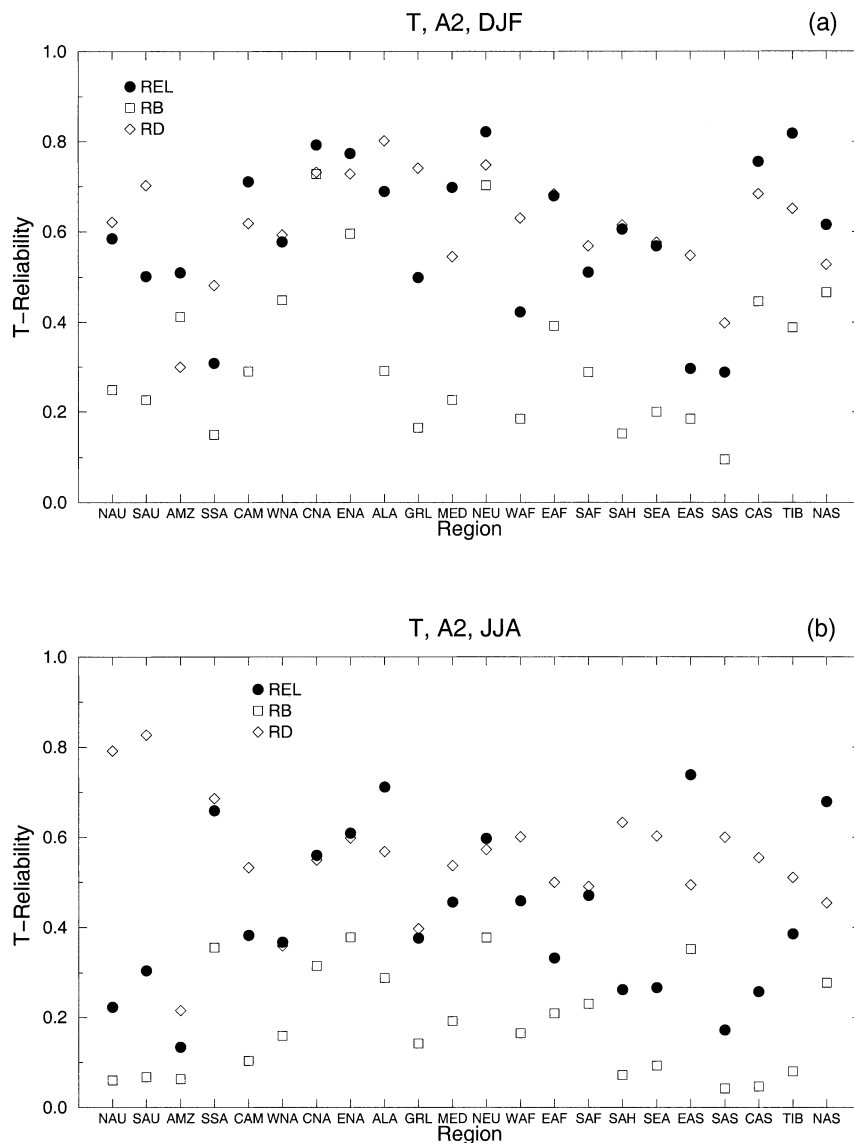
FIG. 7. Reliability [REL, dark circles, see Eq. (7)], and average reliability factors for the performance criterion [RB, open squares, see Eq. (8a)] and the convergence criterion [RD, open diamonds, see Eq. (8b)] for the A2 scenario over the 22 regions of Fig. 1: (a) DJF temperature and (b) JJA temperature.

greater than the model performance in reproducing present-day climate, a result evident from the comparison of Figs. 2, 3 and 4, 5. Therefore, for both variables the factor that most contributes to decreasing the reliability in the regional projections is the performance in reproducing present-day climate conditions. Note that in most cases the value of $\tilde{\rho}$ is in between those of R_B and R_D, which is intuitive from the definition of these quantities.

For temperature, the values of $\overline{R_D}$ are mostly in the range of 0.4–0.8 in DJF and somewhat lower in JJA. This implies distances from the ensemble average greater than the natural variability estimates by factors of up to 2.5. The values of $\overline{R_B}$ are generally lower, mostly 0.1–0.6 in DJF and even lower in JJA. This implies that

the overall magnitude of the model biases is greater than the natural variability by factors that can exceed 10. The overall reliability in the simulated temperature changes is in the range of 0.3–0.8 in DJF and 0.1–0.7 in JJA, which suggests that the winter-simulated changes are generally more reliable than the summer ones. In DJF the highest reliability values are found in some northern mid- and high-latitude regions: CNA and eastern North America (ENA), NEU, TIB, and CAS. The lowest reliability of DJF estimates occurs over SSA, EAS, and SAS. In JJA the highest reliability of the simulated temperature changes is found over SSA, EAS, and some mid- and high-latitude northern regions (CNA, ENA, ALA, NAS). The lowest reliability of JJA REA tem-
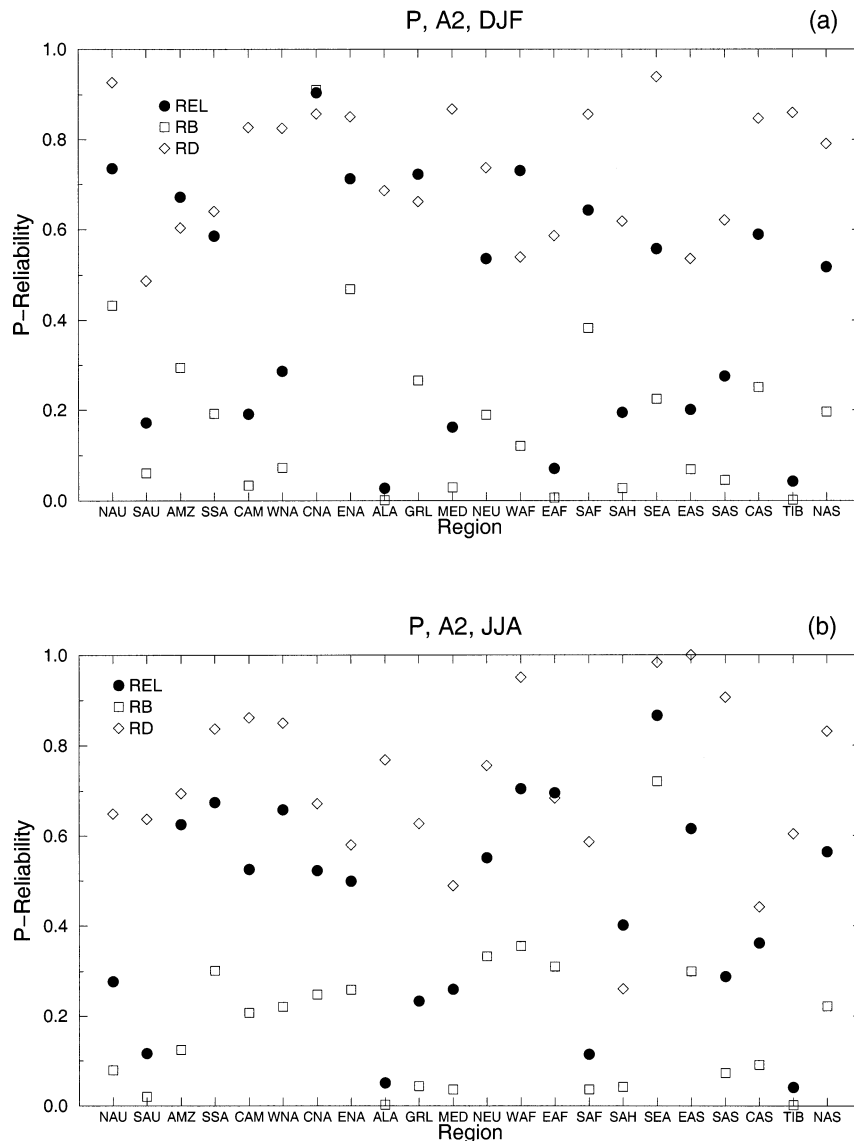
FIG. 8. Same as Fig. 7 but for (a) DJF precipitation and (b) JJA precipitation.

perature change is found over the Amazon (AMZ) and SAS regions. In evaluating the reliability values of Fig. 7, it should be recalled that the reliability measure $\tilde{\rho}$ depends on the natural variability estimate ϵ, lower values of ϵ implying more stringent reliability requirements.

The reliability of the simulated precipitation changes is highly variable across regions in both seasons, spanning values from less than 0.1 to 0.9. The values of $\overline{R_D}$ are consistently and substantially higher than those of $\overline{R_B}$. In fact $\overline{R_D}$ is mostly between 0.5 and 0.9, while $\overline{R_B}$ is mostly lower than 0.5. This is evidence of a generally poor model performance in reproducing present-day precipitation and of a magnitude of simulated changes not large compared to the natural variability estimates. Noticeable exceptions of larger $\overline{R_B}$ and thus

better model performance occur over CNA in DJF and SEA in JJA. Despite the poor collective model performance in simulating present-day precipitation amounts, it should be noted that given models might individually show better performance over individual regions and that the REA averaging maximizes the contribution of such models.

It is important to point out that, for precipitation, the reliability values should be assessed in relation to the magnitude of the changes. It is in fact possible that a high reliability simply implies that the change is within the estimated natural variability, that is, that there is high reliability in a simulation of no significant change. Indeed, many of the changes corresponding to reliability values in excess of 0.6 are small and within the natural variability estimates. Focusing on the changes that are

outside the natural variability, we can see that in DJF over GRL, NEU, WAF, and NAS the reliability is above 0.5, while over CAM, SAH, ALA, EAF, and EAS it is less than 0.4. In JJA, with the exception of ENA and EAS, the reliability for regions where the simulated changes are outside the natural variability is below 0.4, primarily as a result of high biases.

Overall, the reliability levels for the B2 experiments were similar to those for the A2 ones, so for brevity they are not discussed here.

d. Specific regional cases

In this section we discuss in more detail some specific regional cases to better illustrate the functioning of the REA method. We selected four cases, two for temperature and two for precipitation: Northern Europe DJF temperature (NEU-T), Amazon Basin JJA temperature (AMZ-T), Greenland DJF precipitation (GRL-P), and Mediterranean JJA precipitation (MED-P). In all these cases the REA changes are well outside the natural variability estimates (see Figs. 4 and 5). However, for the NEU-T and GRL-P cases the reliability is high (about 0.8 and 0.7, respectively), while for AMZ-T and MED-P it is low (0.15 and 0.25, respectively). Figures 9 and 10 show for each case the individual model changes and biases as a function of the normalized reliability factor R [i.e., the value of R_i of Eq. (4) divided by the sum of R_i over all models in the ensemble]. Also reported in the plots are the REA averages and the corresponding REA uncertainty limits.

The first feature to notice is the highly irregular nature of the distribution of the reliability factors. For example in the case of MED-P all the values of R are about 0.1 or less except for 1 instance of over 0.45. Similarly, for the GRL-P case only 2 values are greater than 0.1 (and both are equal to about 0.33). On the other hand, in the NEU-T case the R values are relatively well distributed between 0 and 0.2, while for the AMZ-T case for most models R is less than 0.08, for 2 models it is equal to about 0.17, and for 1 model it is about 0.25. The REA uncertainty range encompasses between 50% to 67% of the total number of the individual changes, with this percentage decreasing as the REA averaging is increasingly dominated by a smaller number of simulations.

In the NEU-T case, 7 out of 9 models show a small bias and 6 of them also show a small distance. The model with the largest bias also exhibits the largest distance. Evidently for this region both the convergence and performance criteria are met by most simulations. In the GRL-P case the averaging is dominated by the contribution of two simulations that show both small bias and distance, while again the model with the largest distance is also characterized by the largest bias. The GRL-P case thus illustrates a situation in which the high reliability is dominated by a relatively small number of high-performing models.

The MED-P case is an interesting one. As can be seen, most models show a negative precipitation change as well as a negative bias. In other words, most models are relatively dry over this region and become drier in the enhanced GHG forcing scenarios. The agreement in the sign of the change would imply a strong signal, however the fact that all models also show a negative bias indicates that perhaps the conditions of enhanced GHG forcing are amplifying a common model deficiency that leads to the negative bias. This feature has been noted also in the analysis of a previous set of simulations by Machenhauer et al. (1998). The addition of the performance criterion in the analysis thus indicates that the signal might not be the result of a physically realistic process but rather the result of a common model deficiency. As a consequence, although the change signal is strong, the reliability of this signal is very low.

Finally, the AMZ-T case illustrates the effect of the natural variability estimate on the overall reliability of the REA changes. In this case, both the biases and distances are not especially large compared to other regions. However, the estimated natural variability is small, about 0.25 K, therefore the reliability factor becomes small. Although admittedly the natural variability estimates are only approximate, and although this result depends to some extent on the function used to define the reliability factors, this example illustrates the point that different regions require different levels of "precision" because of the underlying regional characteristics of the natural variability.

e. Sensitivity to the weighting parameters m and n

In the previous sections the 2 parameters m and n of Eq. (4) were set equal to 1, implying the same weighting of the 2 reliability criteria. This is not necessarily the case if there are reasons to believe that one of the two criteria should have a greater weight. An example can illustrate this point. It might be argued that the model convergence toward a common answer regardless of the bias magnitude could be the indication of a robust signal. For example, Giorgi et al. (2001a) found a number of consistent patterns of change in regional temperature and precipitation structure across several model simulations. On the one hand, the consistency of these patterns might imply physical processes in action across models regardless of the model biases. On the other hand, the presence of large biases over a region might be a consequence of common model deficiences and thus a convergent response across models could result from the amplification of the common model biases. In the former case, one would set n greater than m in Eq. (4), giving more weight to the convergence criterion, while in the latter case one would set m greater than n, thereby giving more weight to the performance criterion. In order to test the sensitivity of our method to the values of m and n we repeated the A2 scenario calculations for the two cases $m = 1, n = 2$ (case CONV,
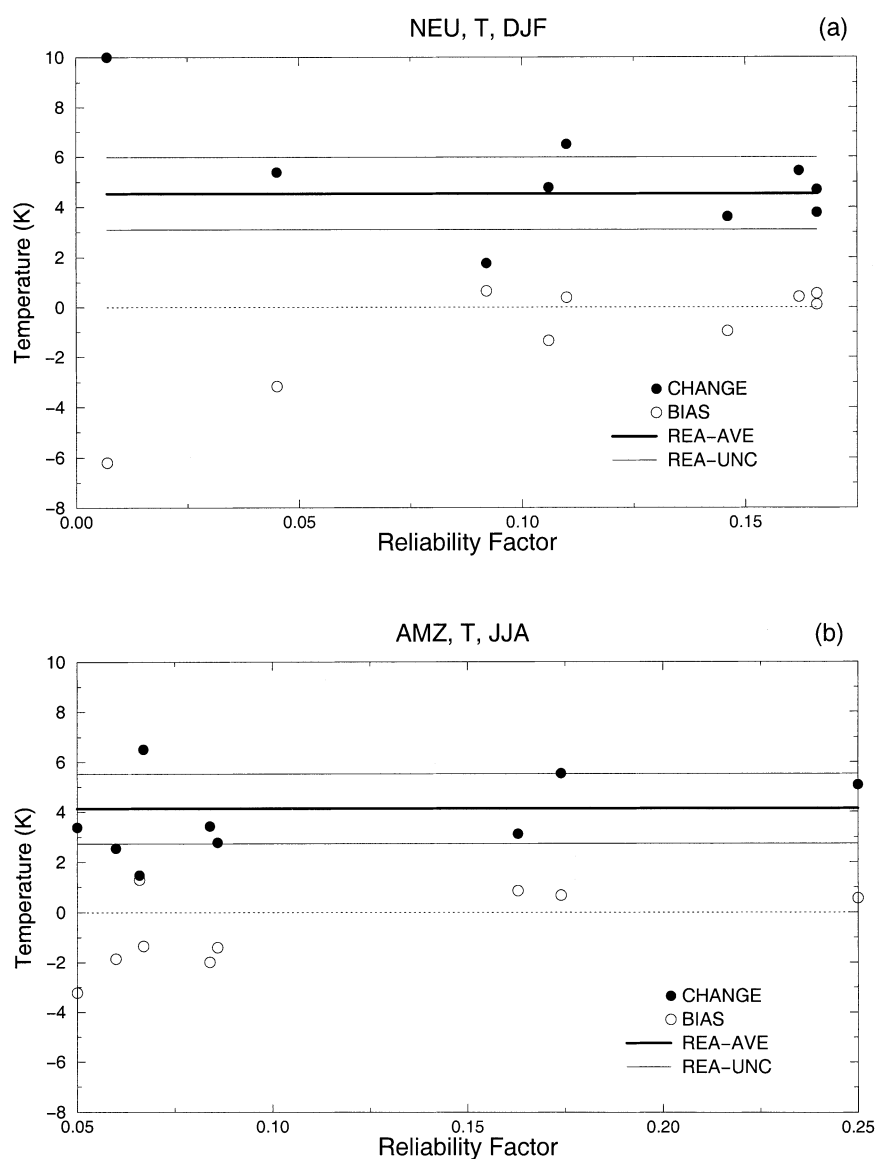
FIG. 9. Individual model change (dark circles) and bias (open circles) as a function of the normalized reliability factor [see Eqs. (3) and (4)]. Also shown are the REA average change (REA-AVE, bold continuous line) and corresponding REA uncertainty limits (REA-UNC, thin continuous lines). (a) NEU DJF temperature and (b) AMZ JJA temperature. Units are degrees K. The dotted line indicates the 0 level.

implying more weight to the convergence criterion) and $m = 2$, $n = 1$ (case PERF, implying more weight to the performance criterion).

For temperature, in all regions the REA changes [Eq. (3)] in the CONV and PERF cases were within 5% of those of the original case. A greater sensitivity was found in the calculation of upper and lower REA uncertainty ranges [Eqs. (6a) and (6b)], where in many regions the sensitivity to the changes in m and n exceeded 10%. Instances of particularly large sensitivity (higher than 20%) in the calculation of uncertainty range of temperature change occurred over the AMZ, ENA, WAF, EAF, and SAS regions in DJF and the NAU, AMZ,

CAM, WNA, CNA, and SAF regions in JJA for the CONV case; and the SAU, WNA, CNA, SAF, SAS, and NAS regions in DJF and the NAU, AMZ, WNA, and EAF regions in JJA for the PERF case. Concerning the calculation of the reliability [Eq. (7)], in the CONV case, a number of regions showed sensitivity of more than 10% compared to the original case (4 regions in DJF and 9 regions in JJA), while this number was lower in the PERF case (1 region in DJF and 5 regions in JJA).

For precipitation, in a few instances the REA change showed large sensitivity in both the CONV and PERF cases, although mostly in correspondence to instances of small estimated changes or small control run precip-

GRL, P, DJF   (a)
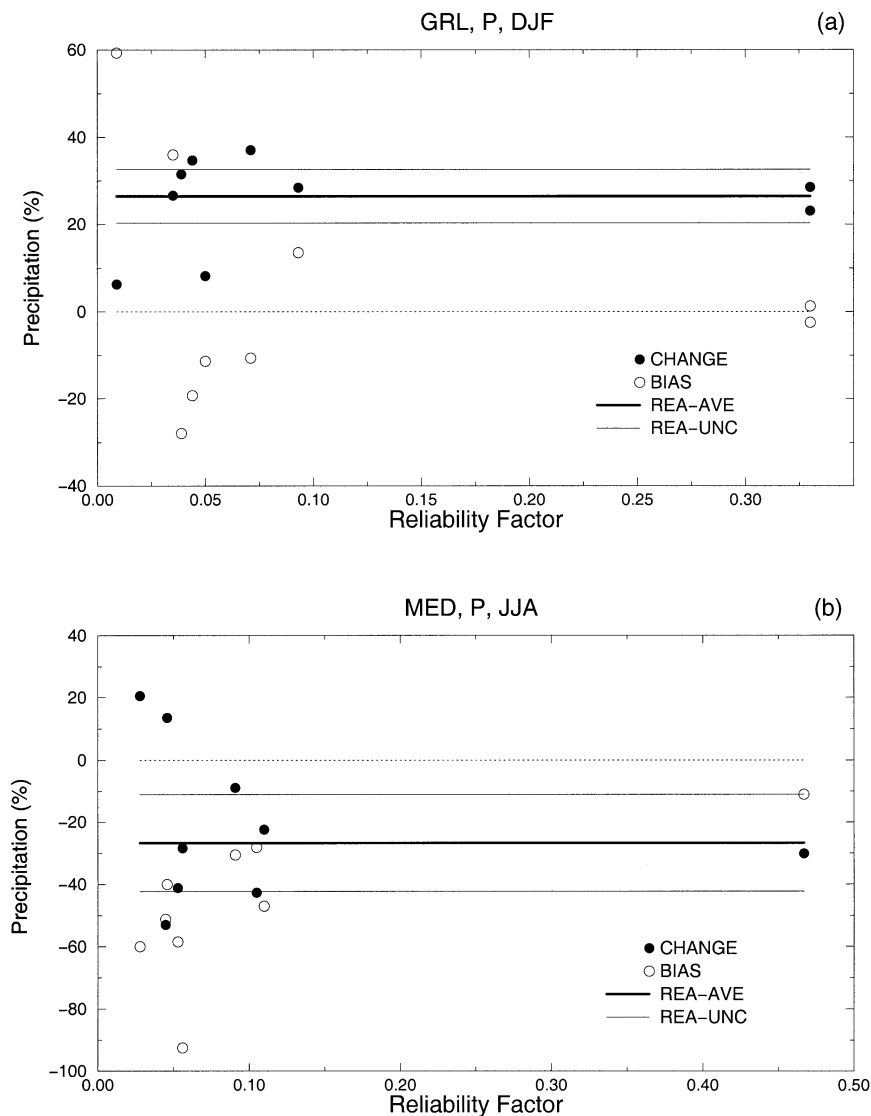

MED, P, JJA   (b)

FIG. 10. Same as in Fig. 9 but for (a) GRL DJF precipitation and (b) MED JJA precipitation. Units are percent of observed precipitation for the bias and percent of present-day precipitation for the change.

itation. The number of regions in which the uncertainty range showed sensitivity greater than 10% compared to the original case was lower than for temperature, indicating a somewhat lower intermodel spread of results. On the other hand, the calculation of reliability was more sensitive to the changes in m and n. Overall, both for temperature and precipitation, compared to the original results, the confidence was generally increased in the CONV case (and correspondingly decreased in the PERF case) because of the greater weight of the convergence criterion reliability factor, which has values generally higher than the performance criterion factor (see Figs. 7 and 8).

In summary, the REA method is flexible enough that the contribution of the two reliability criteria can be weighted differently by modifying the parameters m and n in Eq. (4). Some of the quantities we calculate, and in particular the reliability measure, appear sensitive to the choice of m and n, essentially because of the different contribution to the reliability by the convergence and performance criteria. However, it is our opinion that both criteria should be met in order to yield a high reliability in the projected changes.

4. Summary and discussion

In this paper we introduce the "reliability ensemble averaging" (REA), method for calculating average, uncertainty range and a reliability measure of regional climate changes from ensembles of experiments with dif-

ferent AOGCMs. The method explicitly and quantitatively takes into account what we have called reliability criteria, that is, the model performance and model convergence criteria. The philosophy underlying the REA approach is to minimize the contribution of simulations that either perform poorly in the representation of present-day climate over a region or provide outlier simulations with respect to the other models in the ensemble. Therefore, we extract only the most reliable information from each model. It is important to emphasize that the criteria are applied regionally and not globally, that is, that different models can be outliers or poor performers over different regions. In the set of experiments we analyzed, both for temperature and precipitation there was no model that ubiquitously performed better or worse than all the others, that is, that exhibited the lowest or highest bias over all regions. Concerning the simulated temperature changes, the same models were mostly (although not in all regions) the primary outliers since the magnitude of the regional changes is tied to the global model sensitivity. This was however not the case for precipitation, since precipitation changes are more tied to regional circulations and processes.

We applied our method to seasonal temperature and precipitation changes over 22 land regions of subcontinental scale as simulated by a recent set of transient climate change experiments with nine AOGCMs for two forcing scenarios. The REA average change differed from the ensemble average change by a few tenths of K to about 1 K for temperature and a few tenths of percent to about 10% (in units of percent of present-day values) for precipitation. The uncertainty range calculated using the REA method was generally lower than the corresponding range calculated using the rmsd, both for temperature and precipitation. This is because of the minimization of the contribution of outliers or poorly performing models in the estimation of the uncertainty.

The method also provides a quantitative measure of reliability in the simulated REA average changes based explicitly on the two reliability criteria. Both for temperature and precipitation, the overall reliability deriving from the model performance (as measured by the model bias) was consistently lower, and often quite markedly, than that deriving from the model convergence. In other words, the models showed biases substantially greater than the spread in the simulated changes. This conclusion is not obvious in view of the fact that often model parameters are "tuned" to reproduce present-day climate conditions. Because both criteria affect the overall reliability, the presence of large biases leads to a general reduction in the reliability of the simulated changes. This implies that the foremost requirement for a general improvement of the reliability of simulated regional climatic changes, at least as measured by the REA method, is the reduction of model biases (or more generally the improvement of model performance) in reproducing present-day regional climate conditions.

In the simulations for the A2 scenario the REA average regional temperature changes varied between about 2 and 7 K across regions and they were all outside the estimated natural variability. The uncertainty range around the REA average changes (as measured by $\pm \tilde{\delta}_{\Delta T}$) varied between 1 and 4 K across regions and the reliability level was mostly in the range of 0.2–0.8. For precipitation, about half of the REA average changes, both positive and negative, were outside the estimated natural variability, and they varied between about −25% and +30% (in units of percent of present-day precipitation). The uncertainty range around these changes ($\pm \tilde{\delta}_{\Delta T}$) mostly varied between about 10% and 30% and the corresponding reliability varied widely across regions. The ratio of A2:B2 REA changes was quite uniform across regions (1.30–1.52) for temperature, while for precipitation it showed larger variations but was nearly always positive (i.e., the average precipitation changes in the two scenarios were of the same sign).

The method can be expanded in various directions. For example, presently the calculation of the reliability factors is performed separately for each region and variable. More generally, consistent performance across variables and regions would be a further indicator of reliability. In addition, we employed relatively simple functions to describe the reliability factors, and more sophisticated ones could be utilized within the same conceptual framework. Finally, our method was applied in this study to averaged climate variables. However, it could be used to examine simulated changes in different climate statistics, for example variability measures such as the standard deviation. While the REA method was applied here to AOGCM simulations, it can also be used within the framework of different modeling tools used to produce regional climate information, such as regional climate models (e.g., Giorgi and Mearns 1991). We are currently exploring all these areas of development of the REA method to report in future work.

We will also investigate ways of further developing the method in probablistic terms. With the development of new research to calculate the probabilities of different climate model sensitivities (Andronova and Schlesinger 2001) and with the production of additional AOGCM model simulations for a larger number of emission scenarios, such extension may be possible without making numerous a priori distribution assumptions.

the Data Distribution Center (DDC) for making available some of the model output.

REFERENCES

Andronova, N. G., and M. E. Schlesinger, 2001: Objective estimation of the probability density function for climate sensitivity. *J. Geophys. Res.,* **106,** 22 605–22 611.

Carter, T. R., M. Hulme, and D. Viner, Eds.,1999: Representing uncertainty in climate change scenarios and impact studies. *Proc. ECLAT-2 Helsinki Workshop,* Norwich, United Kingdom, CRU, 128 pp.

Dai, A., T. M. L. Wigley, B. Boville, J. T. Kiehl, and L. Buja, 2001: Climates of the twentieth and twenty-first centuries simulated by the NCAR climate system model. *J. Climate,* **14,** 485–519.

Flato, G. M., and G. J. Boer, 2001: Warming asymmetry in climate change simulations. *Geophys. Res. Lett.,* **28,** 195–198.

Fyfe, J. C., and G. M. Flato, 1999: Enhanced climate change and its detection over the Rocky Mountains. *J. Climate,* **12,** 230–243.

Giorgi, F., and L. O. Mearns, 1991: Approaches to regional climate change simulation: A review. *Rev. Geophys.,* **29,** 191–216.

——, and R. Francisco, 2000a: Uncertainties in regional climate change predictions. A regional analysis of ensemble simulations with the HADCM2 GCM. *Climate Dyn.,* **16,** 169–182.

——, and ——, 2000b: Evaluating uncertainties in the prediction of regional climate change. *Geophys. Res. Lett.,* **27,** 1295–1298.

——, J. W. Hurrell, M. R. Marinucci, and M. Beniston, 1997: Elevation signal in surface climate change: A model study. *J. Climate,* **10,** 288–296.

——, and Coauthors, 2001a: Emerging patterns of simulated regional climatic changes for the 21st century due to anthropogenic forcings. *Geophys. Res. Lett.,* **28,** 3317–3320.

——, and Coauthors, 2001b: Regional climate information—Evaluation and projections. *Climate Change 2001: The Scientific Basis,* J. T. Houghton et al., Eds., Cambridge University Press, 583–638.

Gordon, H. B., and S. P. O'Farrell, 1997: Transient climate change in the CSIRO coupled model with dynamic sea ice. *Mon. Wea. Rev.,* **125,** 875–907.

Hulme, M., and T. R. Carter, 2000: The changing climate of Europe. *Assessment of the Potential Effects of Climate Change in Europe,* M. L. Parry, Ed., ACACIA Concerted Action Report, UEA, 350 pp.

Jones, R. N., 2000a: Managing uncertainty in climate change projections—Issues for impact assessment. *Climate Change,* **45,** 403–419.

——, 2000b: Analysing the risk of climate change using an irrigation demand model. *Climate Res.,* **14,** 89–100.

Kattenberg, G. A., and Coauthors, 1996: Climate models—Projections of future climate. *Climate Change 1995: The Science of Climate Change,* J. T. Houghton et al., Eds., Cambridge University Press, 285–358.

Katz, R. W., 2001: Techniques for estimating uncertainty in climate change scenarios and impact studies. *Climate Res.,* in press.

Kittel, T. G. F., F. Giorgi, and G. A. Meehl, 1998: Intercomparison of regional biases and doubled CO_2 sensitivity of coupled atmosphere–ocean general circulation model experiments. *Climate Dyn.,* **14,** 1–15.

Knutson, T. R., T. L. Delworth, K. W. Dixon, and R. J. Stouffer, 1999: Model assessment of regional surface temperature trends (1949–97). *J. Geophys. Res.,* **104,** 30 981–30 996.

Krishnamurti, T. N., C. M. Kishtawal, T. LaRow, D. Bachiochi, Z. Zhang, C. E. Williford, S. Gadgil, and S. Surendran, 1999: Improved weather and seasonal climate forecasts from multimodel superensemble. *Science,* **285,** 1548–1550.

——, ——, Z. Zhang, T. LaRow, D. Bachiochi, C. E. Williford, S. Gadgil, and S. Surendran, 2000: Multimodel ensemble forecasts for weather and seasonal climate. *J. Climate,* **13,** 4196–4216.

Lambert, S. J., and G. J. Boer, 2001: CMIP1 evaluation and intercomparison of coupled climate models. *Climate Dyn.,* **17,** 83–106.

Machenhauer, B., J. Windelband, M. Botzet, J. H. Christensen, M. Deque, R. G. Jones, P. M. Ruti, and G. Visconti, 1998: Validation and analysis of regional present-day climate and climate change simulations over Europe. Max-Planck-Institut für Meteorologie Rep. 275, 78 pp.

Mearns, L. O., M. Hulme, T. R. Carter, R. Leemans, M. Lal, and P. Whetton, 2001: Climate scenario development. *Climate Change 2001: The Scientific Basis,* J. T. Houghton et al., Eds., Cambridge University Press, 739–768.

Mitchell, J. F. B., T. C. Johns, M. Eagles, W. J. Ingram, and R. A. Davis, 1999: Towards the construction of climate change scenarios. *Climatic Change,* **41,** 547–581.

New, M., and M. Hulme, 2000: Representing uncertainty in climate change scenarios: A Monte-Carlo approach. *Integr. Assess.,* **1,** 203–214.

——, ——, and P. D. Jones, 2000: Representing twentieth-century space time climate fields. Part II: Development of a 1901–1996 mean monthly terrestrial climatology. *J. Climate,* **13,** 2217–2238.

Noda, A., K. Yoshimatsu, S. Yukimoto, K. Yamaguchi, and S. Yamaki, 1999: Relationship between natural variability and CO_2-induced warming pattern: MRI AOGCM experiment. Preprints, *10th Symp. on Global Change Studies,* Dallas, TX, Amer. Meteor. Soc., 359–362.

Palmer, T. N., C. Brankovic, and D. C. Richardson, 2000: A probability and decision model analysis of PROVOST seasonal multimodel ensemble integrations. *Quart. J. Roy. Meteor. Soc.,* **126,** 2013–2033.

Schneider, S. H., 2001: What is dangerous climate change? *Nature,* **411,** 17–19.

Smart, R., N. Nakicenovic, and S. Nakicenovic, Eds., 2000: *Emission Scenarios.* Cambridge University Press, 599 pp.

Stendel, M., T. Schmith, E. Roeckner, and U. Cubasch, 2000: The climate of the 21st century: Transient dimulations with a coupled atmosphere–ocean general circulation model. Danmarks Klimacenter Rep. 00-6.

Stott, P. A., and S. F. B. Tett, 1998: Scale-dependent detection of climate change. *J. Climate,* **11,** 3282–3294.

Visser, H., R. J. M. Folkert, J. Hoekstra, and J. J. deWolff, 2000: Identifying key sources of uncertainty in climate change projections. *Climatic Change,* **45,** 421–457.

Whetton, P. H., M. H. England, S. P. O'Farrell, I. G. Watterson, and A. B. Pittock, 1996: Global comparison of the regional rainfall results of enhanced greenhouse coupled and mixed layer ocean experiments: Implications for climate change scenario development. *Climatic Change,* **33,** 497–519.

Wigley, T. M. L., and S. C. B. Raper, 2001: Interpretation of high projections for global-mean warming. *Science,* **293,** 451–454.