Climate Dynamics (2000) 16:169–182

F. Giorgi · R. Francisco

Uncertainties in regional climate change prediction: a regional analysis of ensemble simulations with the HADCM2 coupled AOGCM

Received: 12 January 1998 / Accepted: 7 July 1999

Abstract We analyze ensembles (four realizations) of historical and future climate transient experiments carried out with the coupled atmosphere-ocean general circulation model (AOGCM) of the Hadley Centre for Climate Prediction and Research, version HADCM2, with four scenarios of greenhouse gas (GHG) and sulfate forcing. The analysis focuses on the regional scale, and in particular on 21 regions covering all land areas in the World (except Antarctica). We examine seasonally averaged surface air temperature and precipitation for the historical period of 1961–1990 and the future climate period of 2046–2075. Compared to previous AOGCM simulations, the HADCM2 model shows a good performance in reproducing observed regional averages of summer and winter temperature and precipitation. The model, however, does not reproduce well observed interannual variability. We find that the uncertainty in regional climate change predictions associated with the spread of different realizations in an ensemble (i.e. the uncertainty related to the internal model variability) is relatively low for all scenarios and regions. In particular, this uncertainty is lower than the uncertainty due to inter-scenario variability and (by comparison with previous regional analyses of AOGCMs) with inter-model variability. The climate biases and sensitivities found for different realizations of the same ensemble were similar to the corresponding ensemble averages and the averages associated with individual realizations of the same ensemble did not differ from each other at the 5% confidence level in the vast majority of cases. These results indicate that a relatively small number of realizations (3 or 4) is sufficient to characterize an AOGCM transient climate change prediction at the regional scale.

F. Giorgi (✉) · R. Francisco
Abdus Salam International Centre for Theoretical Physics,
PO Box 586, 34100 Trieste, Italy

1 Introduction

The simulation of regional scale climatic changes due to anthropogenic forcings, such as increased emissions of greenhouse gases (GHGs) and sulfate aerosols, is critical for an assessment of regional impacts of climate modifications. Coupled atmosphere-ocean general circulation models (AOGCMs) are the most powerful tools today available for global climate simulation. Despite a marked increase in computer power, coupled AOGCMs are still run at relatively coarse horizontal spatial resolutions (~300 km). At this resolution, the effects of local forcings on regional climate are not captured, and a number of regionalization techniques have been developed to enhance the regional information provided by AOGCMs (Giorgi and Mearns 1991; IPCC 1996).

Any regionalization technique, however, relies on two factors: First, the ability of AOGCMs to provide accurate simulation of broad scale circulation and climate features over the region of interest; and, second, the ability of AOGCMs to produce reliable simulation, or "prediction", of the response of such broad scale climatic features to changes in external forcings. It is well known that the highly non-linear nature of atmospheric interactions poses limits to weather predictability of the order of 1–2 weeks. At longer time scales, such as monthly/seasonal to interannual/interdecadal, predictions can be based on the statistical properties of ensembles of model simulations given the knowledge of external forcings. For example, some success has been achieved in seasonal prediction via ensembles of global model simulations with varying initial conditions (e.g. Marx and Fennessy 1999). The use of ensemble simulations can thus be a powerful tool to increase the reliability of long term predictions of the climatic response to anthropogenic forcings.

Given these premises, the issue of uncertainty in climate change predictions at the regional scale (10^4–10^7 km^2) can be summarized in the following questions:

1. If one performs with the same model an ensemble of AOGCM climate simulations for a certain scenario of time dependent GHG, sulfate and other forcings (or "transient" experiment), say from the years 1860 to 2100, what is the range of simulated regional climate response for the given forcing scenario? In other terms, how representative is a single experiment of the model's climatology for the selected scenario over a given region?
2. If transient experiments for the same forcing scenario are conducted with different AOGCMs, what is the range of simulated regional climate responses?
3. What is the range of simulated regional climate responses to different forcing scenarios?
4. What is the effect of sub-GCM grid scale forcings on the simulated regional climate change signal?

We refer to these questions 1, 2 and 3 as reflecting the uncertainty associated with the internal AOGCM model variability, the inter-model variability and the inter-scenario variability, respectively. The compounded uncertainty implied by these three questions defines the uncertainty in the simulation of regional climate change by AOGCMs. We then have to compound this uncertainty with that implied by question 4, i.e. with the uncertainty due to sub-GCM grid scale forcings, which can be assessed with the regionalization techniques reviewed, for example, by Giorgi and Mearns (1991).

Past work on AOGCM analysis has mostly dealt with question 2, partly because only very recently ensemble transient runs for different scenarios have been completed. For example, Kittel et al. (1998) intercompared results from nine AOGCM transient experiments for present day conditions and for conditions at equivalent doubling of GHG concentration. They analyzed surface air temperature and precipitation over seven regions and drew the following conclusions:

a. For current climate the model simulations showed relatively large biases compared to observed climatologies. The biases were greater than the uncertainties in observations.
b. The biases varied substantially, both in magnitude and sign, among models and among regions.
c. The models generally showed greater agreement in their sensitivity to GHG doubling than their control (present-day) runs did with observations.

These conclusions of Kittel et al. (1998), whose results are also included in IPCC (1996), are generally in line with the present view of regional performance and climate change simulations by AOGCMs.

In this paper we turn our attention to our question 1, which is of foremost importance for a full understanding of the regional climate change prediction problem. To do so, we analyze data from a set of ensembles of transient experiments completed with the coupled AOGCM of the Hadley Centre for Climate Prediction and Research, version HADCM2 (e.g. Johns et al. 1997). Data were available from ensembles of 4 realizations of transient runs under different GHG and sulfate forcing scenarios. Therefore we can address question 1 also within the context of our question 3.

We divide all land areas of the globe (with the exception of Antarctica) into 21 regions. Surface air temperature and precipitation for each of the HADCM2 runs are regionally averaged, compared to observations for present day conditions in order to assess the model performance, and intercompared among different experiments and realizations for future climate conditions. Climatic averages and measures of interannual variability for the 30-year simulated periods of 1961–1990 (present day) and 2046–2075 (future climate) are analyzed.

The primary objectives of the analysis are to:

1. Assess to what extent an individual simulation is representative of the model climatology at the regional scale for present day and future conditions. This defines a measure of uncertainty related to internal regional climate variability in the model.
2. Compare the uncertainty due to internal model variability with that due to inter-model and inter-scenario variability.
3. Assess the dependency of the uncertainty in model prediction (in the sense indicated already) on season, regional setting, scenario of forcing, and variable under consideration (surface air temperature and precipitation).

The data used in this study were provided by the Data Distribution Centre (DDC), a data archive recently created under recommendation by the Task Group for Climate Impact Assessment (TGCIA) of the Intergovermental Panel for Climate Change (IPCC). The DDC is intended to provide a service to the climate impact assessment community through the maintenance and distribution of a variety of observed climate datasets and output from a range of model simulations. Information on the DDC can be found by accessing the website http://ipcc-ddc.cru.uea.ac.uk.

In Sect. 2 we first describe the model simulations, the measures of model performance and the regions utilized in this study. Section 3 presents the results of our analysis and Sect. 4 shows our conclusions.

2 Model, experiments and analysis

2.1 Model and experiments

The HADCM2 coupled AOGCM used to complete the simulations analyzed here is described in detail by Mitchell et al. (1995a, b) and Johns et al. (1997). Both the ocean and atmospheric components of the model have a spatial horizontal resolution of 3.75° longitude (~400 km) and 2.5° latitude (~300 km). The atmosphere and ocean models have 19 and 20 vertical levels, respectively. Clouds are explicitly calculated and an adjustment for horizontal heat and water fluxes is included in the ocean model. Land surface processes include vegetation, snow, a four layer soil model and surface runoff calculations. The surface albedo is a function of snow depth, vegetation type and snow/ice temperature.

The series of experiments analyzed in this study is as follows (see Table 1 and Mitchell et al. 1995a, b for further detail). A 1400-

Table 1 List of HADCM2 experiments analyzed in this work. For each experiment, ensembles of four realizations were available (see text and Fig. 1). See also Mitchell et al. (1995a, b) and Johns et al. (1997) for more detail

Experiment	Period of HADCM run	Period of analysis	GHG forcing	Sulfate forcing
HGHG	1860–1990	1961–1990	Historical	No
HSULF	1860–1990	1961–1990	Historical	Historical (IS92a)
F1GHG	1990–2100	2046–2075	1% yr^{-1} Comp. incr.	No
F1SULF	1990–2100	2046–2075	1% yr^{-1} Comp. incr.	IS92a
FdGHG	1990–2100	2046–2075	0.5% yr^{-1} Comp. incr.	No
FdSULF	1990–2100	2046–2075	0.5% yr^{-1} Comp. incr.	IS92a

year control simulation was first conducted with GHG concentration at pre-industrial levels and no aerosol effects. Note that this simulation did not show a perceptible drift in global mean surface temperature (Johns et al. 1997), a result common in AOGCMs employing ocean flux correction. Ensembles of 4 realizations of "historical" runs were then completed starting at 150 year intervals within the control simulation. These historical runs were of 130 years length, with GHG and sulfate forcing specified to represent the historical period 1860–1990. Because they start at different times within the control run, each individual realization has different initial conditions for the atmosphere, the ocean and the land surface (soil water content). Two sets of historical runs were available, one including only anthropogenic GHG forcing (HGHG/1-4) and one including both GHG and anthropogenic sulfate aerosol effects (HSULF/1-4). Sulfate aerosol direct radiative effects were described in the model via an increase in surface albedo; indirect aerosol effects due to interactions with clouds were not included; and the GHG forcing was expressed in terms of equivalent carbon dioxide concentration. Both the GHG and sulfate historical forcings are described in Mitchell et al. (1995a, b).

At the end of each simulated historical run (i.e. at year 1990) the following sets of future transient climate change simulations were available, each consisting of 4 realizations and each realization starting from the end of the corresponding historical run: (a) 1% $year^{-1}$ compounded GHG increase and no aerosol effects (F1GHG/1-4); (b) 1% $year^{-1}$ compounded GHG increase and inclusion of aerosol effects (F1SULF/1-4); (c) 0.5% $year^{-1}$ compounded GHG increase and no aerosol effects (FdGHG/1-4); and (d) 0.5% $year^{-1}$ compounded GHG increase and inclusion of aerosol effects (FdSULF/1-4). Each of these experiments was extended from year 1990 to year 2100 and the future aerosol forcing was taken from the IPCC IS92a scenario (IPCC 1992). We thus have (see Table 1) two ensembles of four realizations of 130-year historical climate runs, from 1860 to 1990 (experiments HGHG/1-4, HSULF/1-4); four ensembles of four realizations of 110-year long future climate experiments, from 1990 to 2100 (experiments F1GHG/1-4, F1SULF/1-4, FdGHG/1-4, FdSULF/1-4).

Out of these simulations we extracted for our analysis monthly data for two variables, surface air temperature (T) and precipitation (P). These variables are of interest for a number of reasons. First, they reflect the overall model behaviour over a given region as, for example, model errors in the simulation of large-scale circulations, storm tracks and water vapour fluxes strongly affect the simulation of precipitation. Second, most previous works on analysis of GCM output has focused on surface air temperature and precipitation, so that comparison of the present results with previous published literature is facilitated. Finally, both variables are of central interest for climate impact assessment research, as many impact studies make use of T and P fields provided by AOGCM simulations.

In this work we limit our analysis to regionally and seasonally averaged data and only to the extreme seasons (December–January–February, or DJF, and June–July–August, or JJA). The periods of analysis are 30 years in length, i.e. the period of 1961 to 1990 from the historical runs, and the period of 2046 to 2075 from the future climate runs. The 30-year period chosen for the future climate runs is centred around the time of GHG doubling compared to 1990 assuming a 1% $year^{-1}$ compounded increase (year 2060). This period, which can be assumed to be representative of climatic conditions in the second half of the twenty first century, was chosen in order to facilitate comparison with previous analyses and in particular with the work of Kittel et al. (1998). For the 0.5% $year^{-1}$ GHG increase scenario, the compounded increase of GHG concentration at the year 2060 compared to 1990 is a factor of about 1.4. A base length of 30 years for each simulated period was deemed sufficient to address issues of interannual variability without having to analyze the entire transient simulations. Also, a period of 30-year length reduces sampling problems associated with model decadal variability (Machenhauer et al. 1998). Hereafter, the experiment labels in Table 1 will refer to the corresponding 30-year periods in our analysis. Finally, in regard to the comparison with the results of Kittel et al. (1998), we should point out that they used 10-year long periods centred around GHG doubling and that they discussed only results from experiments not including aerosol effects.

For comparison of control and historical runs with observations we use the observed climatological dataset developed by the Climatic Research Unit (CRU) of the University of East Anglia (New et al. 1999a, b). This consists of a series of gridded monthly screen height temperature and precipitation over land areas at a 0.5° resolution for the period of 1901 to 1996 (New et al. submitted 1999b), out of which we extracted the period 1961–1990 for comparison with the historical runs of Table 1.

Some general remarks on the comparison of model and observed data are useful for an interpretation of our results. First, the observed dataset is characterized by a certain level of uncertainty, mostly due to the density of available observing stations. New et al. (1999a) provide estimates of this uncertainty for the 30 year climatology of 1961–1990 using both an internal cross-validation procedure and a comparison with other available observed climatologies. They conclude that uncertainties in observed climatic averages are of the order of 0.5 to 1.3 K for temperature and up to 10–25% for precipitation, and are largest over regions characterized by poor station coverage and high spatial variability, such as Greenland, the American Cordillera, the Himalaya-Tibetan region, Southeast Asia and regions of the African continent. It is also likely that uncertainties in the observations are greater for interannual variability than for climatological averages.

An additional source of uncertainty in model-observation comparison is related to elevation, which affects temperature and precipitation. In the CRU dataset, elevation effects are accounted for by including elevation as a predictor variable in the interpolation procedure from station location to the 0.5° grid (New et al. 1999a). On the other hand, uncertainty can originate from differences in the average elevation of model and observed data. To evaluate this effect we calculated the average elevation in the HADCM2 grid and in the elevation dataset used by New et al. (1999a) for the 21 regions examined in this work (see Sect. 2.2). The differences between average elevation in the HADCM2 and observations over these broad regions were small, less than 100 m, except for Southeast Asia. We therefore estimate that the effect of elevation differences between model and observations over the 21 selected regions is small for temperature, order of several tenths of a degree. For precipitation, this effect is difficult to estimate, as topography can affect precipitation in various ways for different seasons.

Finally, Mitchell et al. (1995a, b) and Johns et al. (1997) present an analysis of the global behaviour of the HADCM2 control and transient experiments (only one realization). The climate sensitivity

of the HADCM2 to equilibrium doubling of carbon dioxide concentration is about 2.5 K, i.e. towards the low end of comparable AOGCMs. Mitchell et al. (1995a, b) also show how inclusion of aerosol effects improves the agreement of simulated spatial patterns of surface climate with historical observations.

2.2 Regions

Because we are interested in the regional scale performance of the model, we divided all land areas in the World (except for Antarctica) into 21 regions, which are shown in Fig. 1 and described in Table 2. The regions of Fig. 1 were chosen based on the following criteria: (1) The size of the regions vary in the range of a few thousand to several thousand km in each direction, therefore each region contains at least several HADCM2 grid points and thus includes the smallest wavelength resolved by the HADCM2; (2) we wanted to approximately cover all land areas in the World with a manageable number of regions of simple shape; (3) the selection of specific regions was intended to represent different climatic regimes and physiographic settings (admittedly, this was a subjective choice, and different regional configurations could be devised); (4) some of the regions are close to those used by Kittel et al. (1998) (and also in IPCC 1996), such as Australia, Central North America, East Asia, Mediterranean Basin, and Southeast Asia.

Our calculation of regional averages of HADCM2 and observed data consists of the following steps:

1. Our 21 regions are defined using as base the CRU 0.5° grid. Hence, we first interpolate the model data from the coarse HADCM2 grid to this fine grid. For this purpose, we utilize a third order polynomial interpolation scheme in latitude and longitude (e.g. Phillips and Taylor 1996). Given the smoothness of the HADCM2 fields, this interpolation procedure produces only minimal modification of the original fields, both for temperature and precipitation (not shown). Only land points are used in the interpolation and, as a result, not all of the original 0.5° CRU grid points include interpolated HADCM2 values (e.g. when a CRU point is surrounded by ocean HADCM2 points). These grid points are not included in the regional averages.

2. After the HADCM2 data are interpolated onto the CRU grid, regional averages are produced using the interpolated 0.5° data weighted by the cosine of the latitude of the grid point loca-

Table 2 List of regions used in this study. Only land grid points are used in the analysis

Name	Acronym	Latitude (°)	Longitude (°)
Australia	AUS	45S–11S	110E–155E
Amazon Basin	AMZ	20S–12N	82W–34W
Southern South America	SSA	56S–20S	76W–40W
Central America	CAM	10N–30N	116W–83W
Western North America	WNA	30N–60N	130W–103W
Central North America	CNA	30N–50N	103W–85W
Eastern North America	ENA	25N–50N	85W–60W
Alaska	ALA	60N–72N	170W–103W
Greenland	GRL	50N–85N	103W–10W
Mediterranean Basin	MED	30N–48N	10W–40E
Northern Europe	NEU	48N–75N	10W–40E
Western Africa	WAF	12S–18N	20W–22E
Eastern Africa	EAF	12S–18N	22E–52E
Southern Africa	SAF	35S–12S	10W–52E
Sahara	SAH	18N–30N	20W–65E
Southeast Asia	SEA	11S–20N	95E–155E
East Asia	EAS	20N–50N	100E–145E
South Asia	SAS	5N–30N	65E–100E
Central Asia	CAS	30N–50N	40E–75E
Tibet	TIB	30N–50N	75E–100E
North Asia	NAS	50N–70N	40E–180E

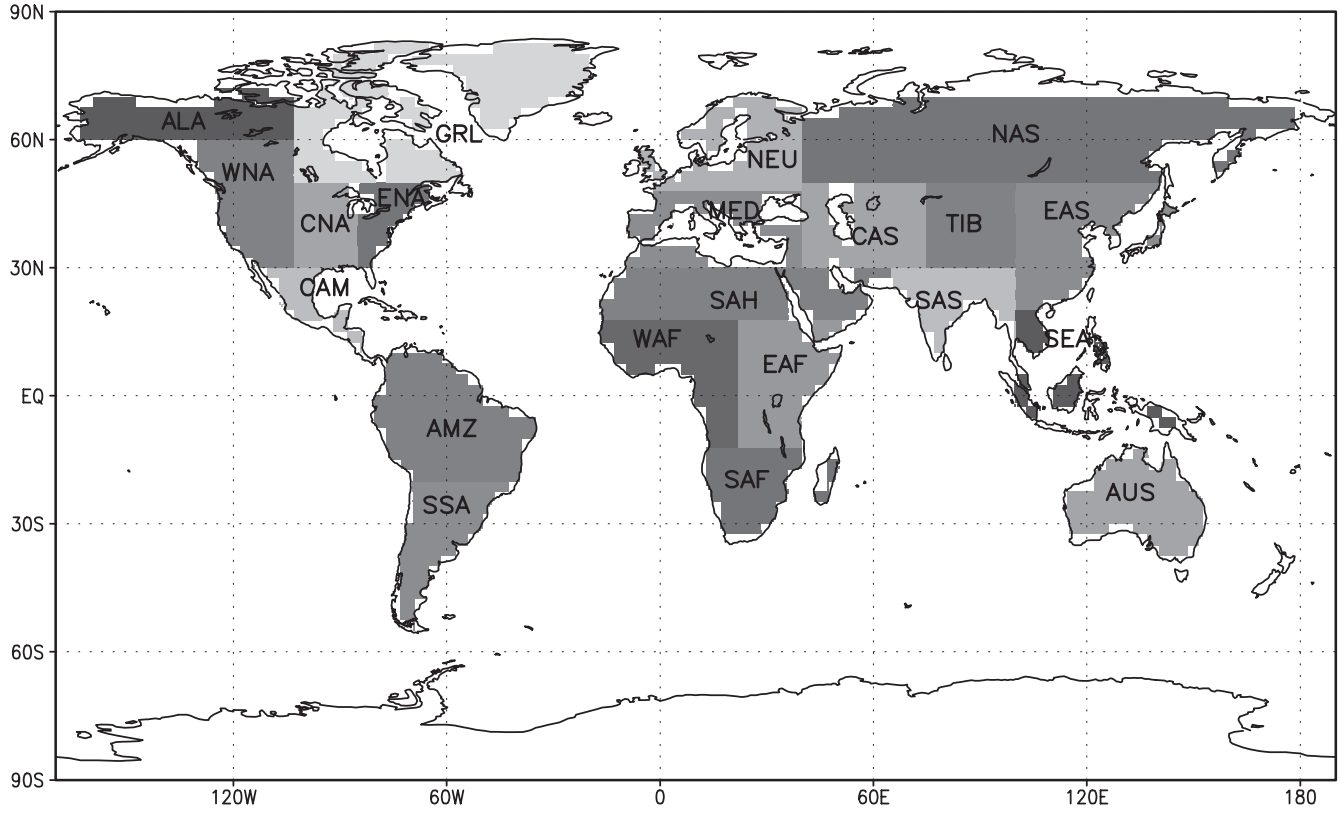

Fig. 1 Regions used in the analysis presented in this work (see Table 2 and text)

tion. This is done for each month of the 30 year periods to produce series of monthly regional values, out of which series of seasonal values as well as seasonal climatological averages are calculated.

3. For the CRU data, regional averages of monthly values are directly calculated as in point (2) above using only grid points for which HADCM2 values are available. i.e. the HADCM2 and CRU regional averages include only a common set of grid points.

We chose to use the 0.5° CRU grid as base for our study for simplicity in the calculations of spatial averages for the regions of Table 2, and because this grid can provide a common framework for analyzing output from different models characterized by different grid configurations. Although we recognize that some level of uncertainty can be related to different interpolation procedures and grid choice, we estimate that this effect is small when taking averages over broad regions such as in Fig. 1.

2.3 Measures of analysis

Our analysis is divided into two segments. In the first, we compare the historical experiments (1961–1990) with CRU observations to evaluate the model performance in representing present day climate, and in the second we analyze the differences between future climate conditions for the period 2046–2075 and the historical simulations.

In the analysis we use various quantities. In the following definitions, each variable (T, P) should be intended as already regionally and seasonally averaged. The 30-year mean for a variable (say T) and an individual realization is defined as

$$\bar{T} = \sum_{i=1}^{30} \frac{T_i}{30} \tag{1}$$

and the corresponding ensemble average for a set of (4) realizations is

$$\tilde{T} = \sum_{i=1}^{4} \frac{\bar{T}_i}{4} \tag{2}$$

The "spread" of the ensemble is defined as the difference between the maximum and minimum value attained by the different realizations, i.e.

$$T_{\text{spread}} = \bar{T}_{\max} - \bar{T}_{\min} \ . \tag{3}$$

The spread can be interpreted as a direct measure of the uncertainty (or error bar) associated with interpreting a single realization as representative of the entire ensemble. As a measure of the statistical significance of the difference between the averages of two 30-year long realizations we use a standard two-sided student t-test (e.g. Edwards 1984). To measure interannual variability, we use the standard deviation of the individual seasonal values during a 30-year period. For temperature, in the calculation of the standard deviations we first detrend the data by removing from each seasonal value the value implied by a least square fit linear approximation of the 30 year trend.

When comparing the historical runs with observations we refer to the term "bias" as the difference between the 30-year averaged value of an historical experiment (either for an individual realization or for an ensemble average) and the corresponding observed 30-year averaged value. We refer to the term "sensitivity" (e.g. ΔT) as the difference between the 30-year averaged value in a future climate simulation (2046–2075) and the corresponding value of the historical simulation (1961–1990).

3 Results

In this section we discuss the results from the HADCM2 simulations. In Sect. 3.1 we analyze the present day historical runs (i.e. HGHG and HSULF for the periods 1961–1990), focusing on the comparison between simulated and observed climatologies. In Sect. 3.2 we turn our attention to the differences, or sensitivities, of climatological variables between future climate conditions (i.e. F1GHG, F1SULF, FdGHG, and FdSULF for the periods 2046–2075) and present day historical conditions.

3.1 Present day simulations (1961–1990)

3.1.1 Surface air temperature

Figure 2a, b shows the seasonal surface air temperature bias for each 30-year (1961–1990) realization of the HGHG and HSULF experiments and each region of Fig. 1. The corresponding regional ensemble averages and spread values (in parentheses) of the bias are reported in Table 3.

Figure 2 shows that over most regions the temperature biases for present day conditions are in the range of ±2 K, with predominance of biases in the range of ±1 K

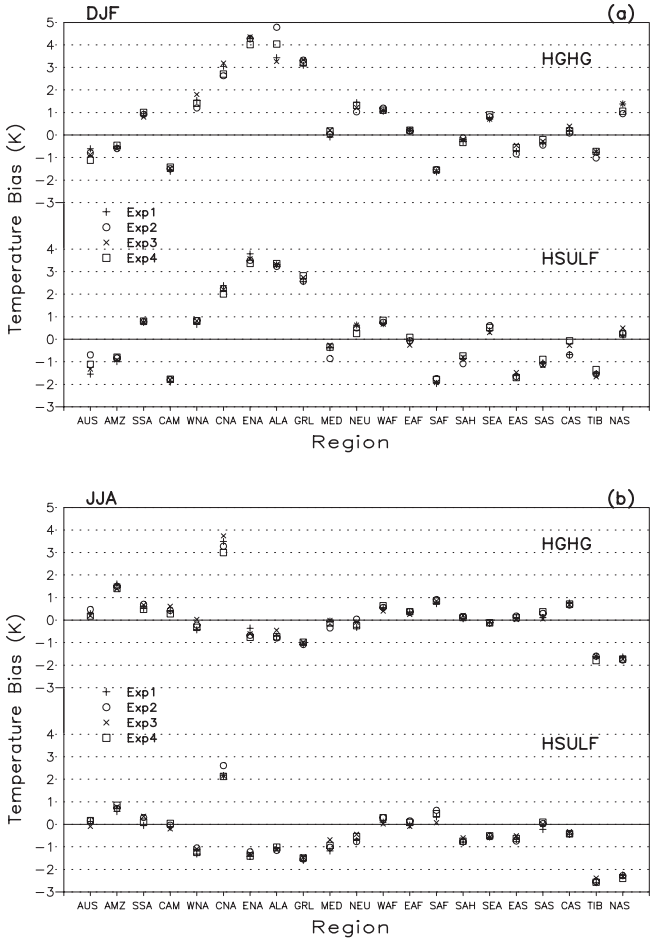

Fig. 2a, b Surface air temperature biases for different historical experiments (see Table 1) and all regions of Fig. 2. All realizations of the ensembles are included (EXP1-4). Units are K. **a** DJF; **b** JJA

Table 3 Ensemble average and spread of the regional surface air temperature biases with respect to the CRU observations for the two historical (1961–1990) experiments (see Table 1). The spreads of the biases are reported in parentheses. Units are K

Region	HGHG DJF	HSULF DJF	HGHG JJA	HSULF JJA
AUS	−0.85 (0.51)	−1.17 (0.86)	0.28 (0.28)	0.09 (0.26)
AMZ	−0.54 (0.13)	−0.88 (0.20)	1.47 (0.21)	0.72 (0.28)
SSA	0.91 (0.20)	0.77 (0.06)	0.59 (0.21)	0.18 (0.40)
CAM	−1.50 (0.18)	−1.82 (0.13)	0.43 (0.33)	−0.08 (0.23)
WNA	1.48 (0.60)	0.78 (0.16)	−0.24 (0.46)	−1.18 (0.26)
CNA	2.89 (0.56)	2.21 (0.38)	3.38 (0.74)	2.26 (0.47)
ENA	4.23 (0.34)	3.56 (0.43)	−0.61 (0.42)	−1.34 (0.20)
ALA	3.88 (1.52)	3.30 (0.13)	−0.68 (0.34)	−1.09 (0.15)
GRL	3.24 (0.26)	2.68 (0.26)	−1.03 (0.10)	−1.52 (0.15)
MED	0.08 (0.31)	−0.46 (0.60)	−0.14 (0.33)	−0.96 (0.49)
NEU	1.25 (0.41)	0.50 (0.36)	−0.15 (0.36)	−0.62 (0.33)
WAF	1.11 (0.13)	0.74 (0.17)	0.54 (0.25)	0.18 (0.28)
EAF	0.20 (0.03)	−0.08 (0.35)	0.34 (0.13)	0.03 (0.24)
SAF	−1.58 (0.10)	−1.85 (0.21)	0.83 (0.18)	0.37 (0.55)
SAH	−0.24 (0.20)	−0.87 (0.35)	0.13 (0.14)	−0.73 (0.19)
SEA	0.78 (0.17)	0.45 (0.30)	−0.11 (0.06)	−0.55 (0.07)
EAS	−0.65 (0.38)	−1.61 (0.22)	0.11 (0.14)	−0.64 (0.22)
SAS	−0.33 (0.26)	−1.05 (0.26)	0.21 (0.32)	−0.02 (0.33)
CAS	0.21 (0.30)	−0.43 (0.63)	0.71 (0.09)	−0.40 (0.11)
TIB	−0.81 (0.28)	−1.53 (0.32)	−1.66 (0.20)	−2.52 (0.22)
NAS	1.20 (0.45)	0.30 (0.31)	−1.72 (0.13)	−2.31 (0.17)

and with biases generally smaller in JJA than DJF. Notable exceptions in which the biases exceed 2 K are central and eastern North America, Alaska and Greenland during DJF and central North America, Tibet and north Asia in JJA. Comparison of HGHG and HSULF results shows that aerosol effects induce a surface cooling over all regions and seasons, in the range of a few tenths to 1.1 K, and generally greatest over the North America, European and Asian regions. While in the HGHG runs the regional biases are largely positive, in the HSULF runs they become more equally distributed between positive and negative, with prevalence of negative biases in the summer.

These biases are generally of the same order of, or greater than, the estimated uncertainties in observations by New et al. (1999a), and they are at the low end of those found by Kittel et al. (1998) for previous AOGCM experiments. This is an indication of a relatively good performance in surface air temperature simulation by the HADCM2 model which, recall, is likely influenced by the use of an ocean flux correction that tends to relax the model towards observed conditions.

Both Fig. 2a, b and Table 3 show that the spread of the biases is relatively small for all regions, less than 0.5 K, except for two instances in JJA and seven instances in DJF. Overall, the temperature spread values are greater in DJF than in JJA and are maximum in DJF over Alaska and Australia and in JJA over Central North America. Inclusion of aerosol effects does not influence in any systematic way the magnitude of the spread.

Also evident from Fig. 2a, b is that, overall, the spread of the biases is smaller than the ensemble average bias. This implies that different realizations are closer to each other than to observations. The *t*-tests revealed that the average regional temperatures simulated by individual realizations in an ensemble did not differ from each other at the 5% confidence level in 93% of regional cases (not shown). All these features indicate that, in terms of regionally and time averaged temperatures, different realizations give very close, and in fact statistically the same, simulations for present day conditions.

Figure 3 presents the ensemble averaged interannual standard deviations of surface air temperature for the observations and the historical runs. Compared to observations, the model overestimates the interannual standard deviation of temperature (and thus the temperature interannual variability) during JJA over all regions and by up to factors exceeding 2. During DJF the model underestimates temperature variability over the North American, northern Europe and Asian regions, while it overestimates it over the other regions. Overall the agreement between observed and simulated standard deviations is greater in DJF than JJA. The

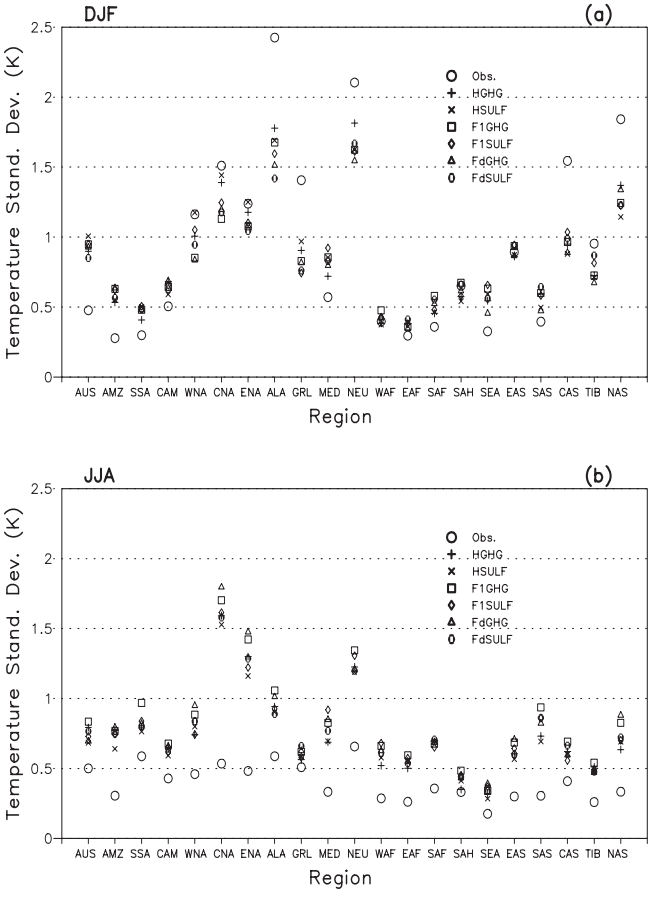

Fig. 3a, b Ensemble average of interannual standard deviation for surface air temperature over all regions of Fig. 2. Included are all historical and future climate transient experiments (see Table 1). Units are K. **a** DJF; **b** JJA

inclusion of aerosol effects mostly reduces variability in JJA and mostly increases it in DJF, although the effect is small. Similar conclusions were found for all individual realizations of the ensembles.

3.1.2 Precipitation

Figure 4a, b shows the DJF and JJA regional precipitation biases for the HGHG and HSULF experiments. The precipitation bias is expressed as percentage of observed CRU precipitation. Table 4 reports the corresponding ensemble average and spread of the bias.

From Fig. 4a, b we first note that the JJA regional precipitation biases are mostly in the range of ±20%. Exceptions are Alaska, Sahara and Tibet, where the bias is 35–60% and Southern Africa, where the bias is about −30%. During DJF the biases are higher, especially over Central America, Alaska, east Asia and Tibet, where they exceed 100%. Note that all these regions are characterized by complex topography, so that a higher uncertainty may be present in the observed data (New et al. 1999a) and the relatively coarse model resolution

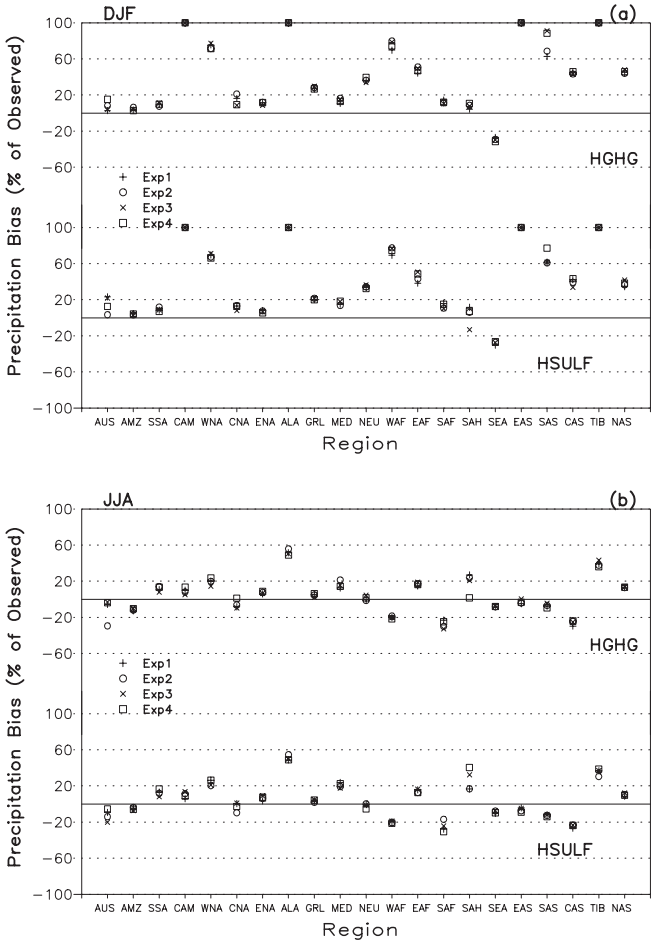

Fig. 4a, b Precipitation biases for different historical experiments (see Table 1) and all regions of Fig. 2. All realizations of the ensembles are included (EXP1-4). Units are percentage of observed precipitation. **a** DJF; **b** JJA

Table 4 Ensemble average and spread of the regional precipitations biases with respect to the CRU observations for the two historical (1961–1990) experiments (see Table 1). The spreads of the biases are reported in parentheses. Units are percentage of observed precipitation

Region	HGHG DJF	HSULF DJF	HGHG JJA	HSULF JJA
AUS	7 (12)	15 (19)	−10 (26)	−12 (14)
AMZ	4 (3)	3 (1)	−11 (2)	−5 (3)
SSA	9 (3)	9 (4)	11 (6)	12 (8)
CAM	117 (27)	113 (30)	9 (8)	9 (7)
WNA	73 (5)	68 (4)	19 (9)	23 (6)
CNA	13 (12)	11 (5)	−5 (10)	−2 (10)
ENA	10 (3)	7 (2)	7 (2)	6 (5)
ALA	157 (18)	154 (20)	51 (6)	50 (5)
GRL	27 (2)	20 (1)	4 (4)	3 (2)
MED	13 (5)	16 (4)	15 (8)	20 (6)
NEU	36 (5)	34 (4)	1 (5)	−1 (5)
WAF	75 (10)	74 (8)	−20 (3)	−20 (2)
EAF	47 (6)	45 (12)	16 (4)	14 (3))
SAF	12 (3)	12 (5)	−27 (9)	−24 (13)
SAH	7 (6)	2 (24)	18 (25)	26 (23)
SEA	−29 (4)	−27 (3)	−8 (0)	−8 (2)
EAS	158 (16)	139 (15)	−2 (5)	−6 (4)
SAS	77 (28)	65 (16)	−7 (5)	−12 (2)
CAS	44 (2)	39 (9)	−25 (5)	−24 (3)
TIB	321 (22)	301 (43)	39 (7)	35 (8)
NAS	45 (3)	37 (7)	13 (0)	10 (3)

likely plays a major role in generating the bias. Also during DJF, however, the bias is still of the order of 20% or less over many regions. Overall, it appears that the HADCM2 model tends to overpredict precipitation over land. In addition, both Fig. 4a, b and Table 4 show that the inclusion of aerosol effects does not strongly affect precipitation simulation. Comparison of the precipitation biases of Fig. 4a, b and Table 4 with those reported by Kittel et al. (1998) shows that the HADCM2 model exhibits biases that are at the low end of the range shown by previous AOGCMs. The model performance is somewhat better in JJA than DJF.

Especially for the JJA season, the spread in regionally averaged precipitation is small, all realizations highly resembling each other. Indeed, a *t*-test analysis indicated that in the vast majority of cases (over 90%), precipitation averages for pairs of realizations in the same ensemble did not differ from each other at the 5% confidence level. Note that in almost all individual realizations the bias for a given region is of the same sign as the ensemble average bias, and in most cases the spread is smaller than the ensemble averaged value. There also appears to be a direct relationship between the magnitude of the bias and the spread of the ensemble, greater spreads being associated with larger biases.

The model performance in simulating interannual variability of seasonal precipitation is poor. Figure 5 shows that the model tends to overestimate the interannual standard deviations (and thus interannual variability) of precipitation for both DJF and JJA and in each realization of the ensembles.

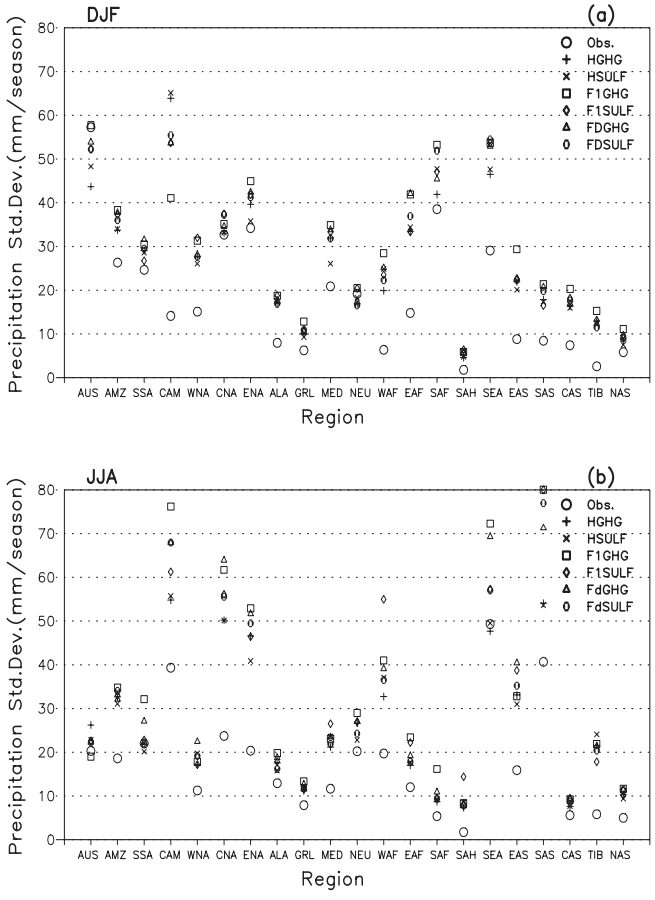

Fig. 5a, b Ensemble average of interannual standard deviation for precipitation over all regions of Fig. 2. Included are all historical and future climate transient experiments (see Table 1). Units are mm per season. **a** DJF; **b** JJA

3.1.3 Summary

Summarizing the results of this section, comparison with previous AOGCM experiments (Kittel et al. 1998) indicates that the HADCM2 simulations show a relatively good performance in simulating regional averages of temperature and precipitation for the 30-year historical period of 1961–1990. For the broad regions analyzed here, surface air temperature biases were mostly in the range of ± 1–2 K and precipitation was mostly within $\pm 20\%$ of observations. Inclusion of aerosol effects caused regional coolings of a few tenths to 1 K, but did not substantially affect precipitation. The model performance in reproducing interannual variability for the 30-year historical period was poor, with the model generally overestimating variability (as measured by the interannual standard deviation) both for temperature and precipitation.

Overall, the spread in seasonally and regionally averaged values between realizations was small, and in almost all cases the averages of different realizations were not statistically different from each other. In most cases the spread values for the biases were smaller than the ensemble mean biases, so that the bias in each realization had the same sign as, and close magnitude to, the ensemble mean bias.

3.2 Future climate simulations (2046–2075)

3.2.1 Surface air temperature

Figure 6a–d shows the surface air temperature sensitivities for DJF and JJA, all regions and all the realizations for the transient experiments of Table 1. The corresponding ensemble average and spread of the sensitivities are given in Tables 5 and 6.

As can be expected, the experiments yielding the most pronounced temperature sensitivity is F1GHG. The regional ensemble average sensitivities for this experiment are in the range of about 2.6 to 6 K in DJF and 2.2 to 3.9 K in JJA, with maximum sensitivity observed over Alaska, Greenland and north Asia. This trend of increased temperature sensitivities in colder climatic settings is consistent with previous AOGCM simulations, being largely related to the snow/ice albedo feedback mechanism.

Figure 7a, b shows the differences between the ensemble average temperature sensitivities of different scenarios. It can be seen that the inclusion of aerosol effects produces a systematic decrease in temperature sensitivity, which is variable in magnitude from region to region. For the F1SULF case, this aerosol-induced effect of reduced warming is in the range of 0.4 K (western North America) to 1.85 K (south Asia), generally greatest over the Asian, African and Central American regions. This response obviously depends on the sulfur emission scenario assumed in the experiments. For example, in the IPCC IS92a scenario used in the HADCM2 runs, the sulfur emission over the USA and Europe is assumed to rapidly stabilize, and in fact eventually decrease, after the 1990, while emissions over Asia are assumed to grow for several decades.

Reduction of rate of GHG increase to 0.5 % $year^{-1}$ in the absence of aerosol effects (FdGHG), induces a systematic reduction of the regional warmings by 0.75 K (Mediterranean) to 1.7 K (north Asia), and inclusion of aerosol effects within this scenario tends to further reduce the warming by up to several tenths of a degree (Fig. 7a, b). All the temperature sensitivities in Figs. 6 and 7 were statistically significant at the 5% confidence level.

It is important to note that both the inclusion of sulfate effects and the reduction of GHG increase rate produce changes in regional temperature sensitivities that are rather variable from region to region, from a few tenths of a degree to over 1.5 K. As can be expected, this inter-regional variability of response is more pronounced when including the sulfate forcing, which is more local in nature, but it is also present when only the GHG concentration is modified uniformly throughout the atmosphere.

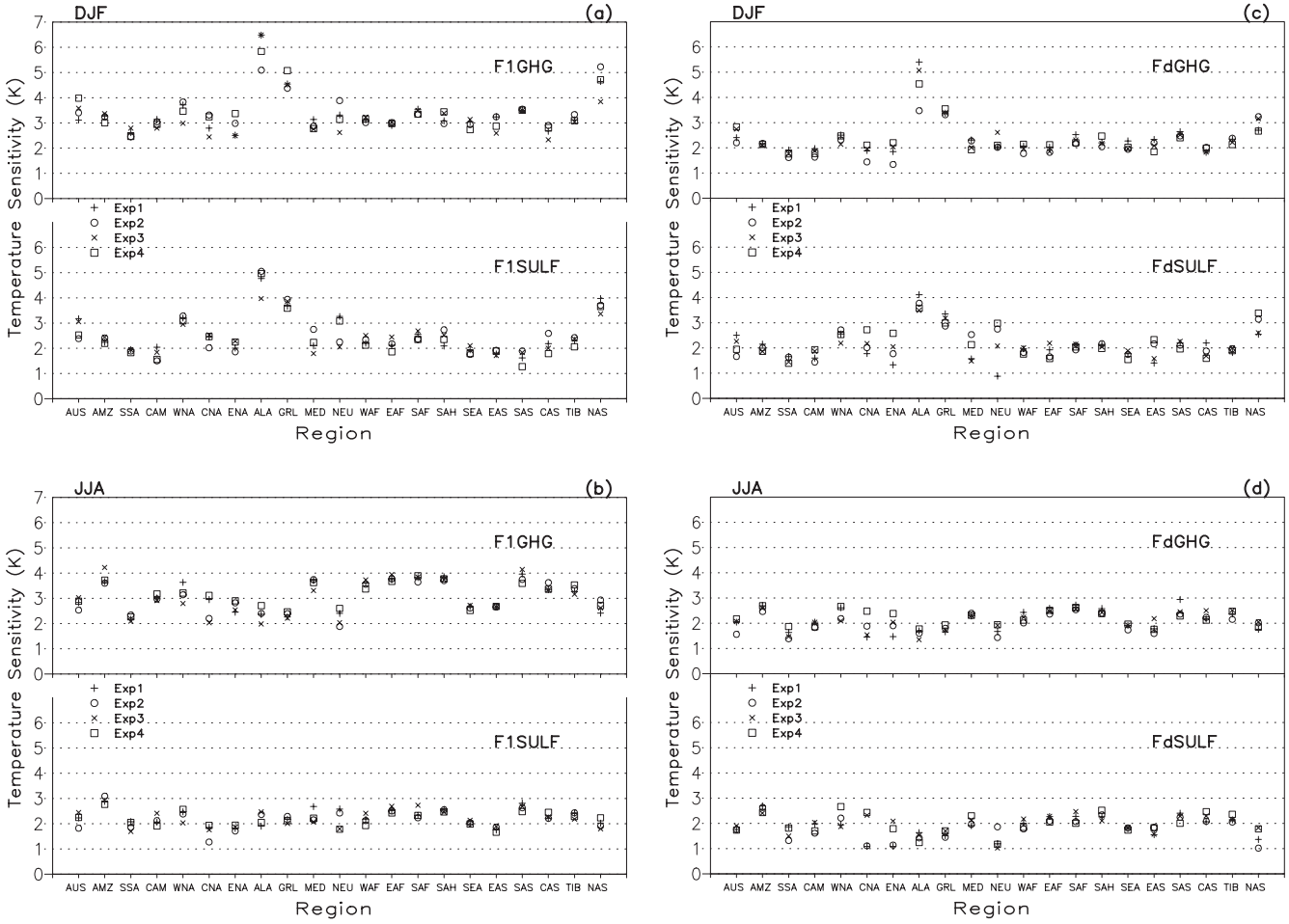

Fig. 6a–d Surface air temperature sensitivities for different future climate transient experiments (see Table 1) and all regions of Fig. 2. All realizations of the ensembles are included (EXP1-4). Units are K. **a** DJF, 1% year^{-1} increase in GHG concentration, with and without sulfate effects; **b** JJA, 1% year^{-1} increase in GHG concentration, with and without sulfate effects; **c** DJF, 0.5% year^{-1} increase in GHG concentration, with and without sulfate effects; **d** JJA, 0.5% year^{-1} increase in GHG concentration, with and without sulfate effects

This is evidently because of non-linear and feedback mechanisms. A strong (positive) feedback mechanism is the snow/ice albedo feedback, which plays a critical role in regulating climatic responses over cold-climate regions. As snow/ice melts due to surface warming, the surface albedo decreases, and thus the surface absorption of solar radiation increases leading to accelerated warming. Other possible feedback mechanisms are related to the surface energy and hydrological cycles. For example, radiational surface warming induces an increase in evaporation which, in the absence of increased precipitation, induces an overall surface drying that in turn enhances the warming.

This inter regional variability of response to changes in emission scenarios is relevant for the issue of using global results from simplified models to "scale" AOGCM-produced sensitivities from a given experiment in order to estimate regional sensitivities corresponding to different scenarios (e.g. Santer et al. 1990). Although systematic trends in the temperature response to sulfate and GHG forcing is evident from Fig. 7a, b, because of the regional dependence of the climatic responses to these forcings, caution should be used in employing a single global scaling factor for all regions.

Turning our attention to the spread of the ensembles, analysis of Fig. 6a, d and Tables 5, 6 suggests two general features. First, the spread in the sensitivities is generally greater than the spread in the corresponding biases. Over a few regions, the spread in temperature sensitivity is greater than 1 K, although in most instances it is less than 0.5 K. Second, there is a suggestion of a relationship between the magnitude of the sensitivity and the magnitude of the spread, with greater spreads being associated with greater sensitivities. This is especially true for cold climate regions, such as Alaska, Greenland and North Asia, where the snow/ice albedo feedback mechanism may play an important role in amplifying both the sensitivity and the spread.

Overall, the spread in sensitivities is still smaller than the the ensemble averages of the sensitivity, implying that each realization of a given experiment essentially gives a similar regional average temperature response. This was

Table 5 Ensemble average and spread of the regional surface air temperature sensitivities (2046–2075 minus 1961–1990) for the four future climate transient runs (see Table 1). Season is DJF. The spreads of the sensitivities are reported in parentheses. Units are K

Region	F1GHG	F1SULF	FdGHG	FdSULF
AUS	3.52 (0.86)	2.78 (0.79)	2.54 (0.63)	2.10 (0.65)
AMZ	3.22 (0.36)	2.32 (0.25)	2.14 (0.07)	1.97 (0.29)
SSA	2.58 (0.33)	1.90 (0.12)	1.76 (0.31)	1.53 (0.27)
CAM	2.98 (0.34)	1.74 (0.54)	1.81 (0.34)	1.71 (0.49)
WNA	3.50 (0.84)	3.13 (0.33)	2.35 (0.36)	2.50 (0.53)
CNA	2.94 (0.87)	2.37 (0.51)	1.85 (0.67)	2.17 (0.95)
ENA	2.84 (0.88)	2.10 (0.42)	1.86 (0.86)	1.93 (1.25)
ALA	5.97 (1.39)	4.69 (1.07)	4.62 (1.93)	3.74 (0.61)
GRL	4.62 (0.70)	3.77 (0.35)	3.40 (0.23)	3.10 (0.48)
MED	2.88 (0.38)	2.21 (0.96)	2.13 (0.39)	1.93 (1.05)
NEU	3.24 (1.26)	2.66 (1.21)	2.19 (0.59)	2.17 (2.10)
WAF	3.13 (0.21)	2.29 (0.39)	1.99 (0.38)	1.89 (0.25)
EAF	2.98 (0.14)	2.15 (0.58)	1.95 (0.31)	1.84 (0.61)
SAF	3.43 (0.20)	2.49 (0.34)	2.30 (0.34)	2.06 (0.23)
SAH	3.22 (0.47)	2.42 (0.64)	2.22 (0.42)	2.08 (0.19)
SEA	2.96 (0.40)	1.90 (0.34)	2.05 (0.33)	1.74 (0.34)
EAS	2.99 (0.65)	1.82 (0.21)	2.11 (0.49)	1.87 (0.94)
SAS	3.50 (0.05)	1.65 (0.62)	2.50 (0.24)	2.13 (0.30)
CAS	2.68 (0.57)	2.14 (0.80)	1.91 (0.19)	1.84 (0.61)
TIB	3.15 (0.30)	2.28 (0.36)	2.27 (0.25)	1.92 (0.18)
NAS	4.60 (1.37)	3.67 (0.61)	2.94 (0.57)	2.91 (0.84)

confirmed by a *t*-test analysis, which showed that in $\sim$ 84–99% of cases, and for all scenarios, each pair of realizations in an ensemble produced average temperatures for the period 2046–2075 that were not statistically different from each other at the 5% confidence level.

Table 6 Ensemble average and spread of the regional surface air temperature sensitivities (2046–2075 minus 1961–1990) for the four future climate transient runs (see Table 1). Season is JJA. The spreads of the sensitivities are reported in parentheses. Units are K

Region	F1GHG	F1SULF	FdGHG	FdSULF
AUS	2.81 (0.49)	2.19 (0.62)	1.96 (0.61)	1.78 (0.16)
AMZ	3.80 (0.62)	2.92 (0.33)	2.61 (0.23)	2.54 (0.29)
SSA	2.21 (0.25)	1.92 (0.40)	1.58 (0.49)	1.62 (0.53)
CAM	3.01 (0.26)	2.12 (0.51)	1.94 (0.21)	1.83 (0.43)
WNA	3.19 (0.84)	2.36 (0.54)	2.38 (0.57)	2.17 (0.79)
CNA	2.57 (1.08)	1.70 (0.66)	1.84 (1.03)	1.74 (1.34)
ENA	2.68 (0.45)	1.83 (0.23)	1.95 (0.91)	1.53 (1.00)
ALA	2.36 (0.73)	2.20 (0.58)	1.60 (0.42)	1.45 (0.39)
GRL	2.33 (0.25)	2.15 (0.28)	1.78 (0.26)	1.59 (0.25)
MED	3.60 (0.43)	2.28 (0.61)	2.33 (0.10)	2.08 (0.40)
NEU	2.23 (0.71)	2.14 (0.80)	1.74 (0.51)	1.31 (0.84)
WAF	3.56 (0.36)	2.16 (0.49)	2.20 (0.42)	1.95 (0.39)
EAF	3.78 (0.27)	2.53 (0.27)	2.48 (0.24)	2.17 (0.23)
SAF	3.80 (0.25)	2.41 (0.49)	2.61 (0.19)	2.21 (0.47)
SAH	3.76 (0.15)	2.51 (0.11)	2.44 (0.19)	2.31 (0.42)
SEA	2.63 (0.20)	2.04 (0.14)	1.88 (0.23)	1.79 (0.11)
EAS	2.67 (0.02)	1.79 (0.24)	1.83 (0.60)	1.70 (0.29)
SAS	3.86 (0.55)	2.65 (0.33)	2.51 (0.65)	2.23 (0.40)
CAS	3.42 (0.29)	2.28 (0.24)	2.26 (0.38)	2.23 (0.40)
TIB	3.33 (0.37)	2.33 (0.25)	2.37 (0.36)	1.49 (0.32)
NAS	2.67 (0.52)	1.99 (0.44)	1.92 (0.30)	2.17 (0.80)

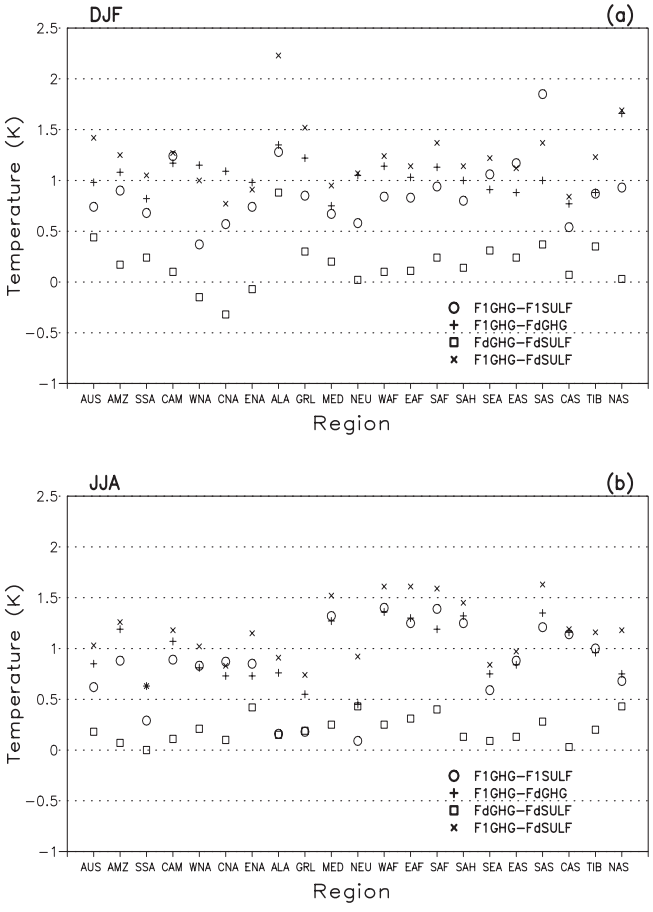

Fig. 7a, b Differences between ensemble average temperature sensitivites of different future climate transient experiments (see Table 1) for all regions of Fig. 2. Units are K. **a** DJF; **b** JJA

In relation to the discussion of Sect. 1, it is useful to compare the temperature sensitivity spreads with the maximum range of ensemble-average sensitivities associated with the four different scenarios (which is given by the F1GHG-FdSULF data in Fig. 7a, b). While the former is a measure of the uncertainty related to the internal model variability for a given scenario, the latter measures the uncertainty due to the use of different possible scenarios. Tables 5, 6 and Fig. 7a, b show that, while the spreads of the ensembles are mostly lower than 0.5–1 K, the maximum range of ensemble-averaged responses to the different emissions scenarios used in the HADCM2 runs mostly exceeds 1 K. This implies that, at least for the range of scenarios investigated in this work, the uncertainty due to the scenarios is greater than the uncertainty due to the internal model variability.

We can also compare the spreads of Tables 5, 6 with the range of sensitivites found by Kittel et al. (1998) for transient experiments at GHG doubling. They found an inter-model range of seasonal temperature sensitivities over 7 regions in the range of 3–5 K, therefore much larger than the spread in sensitivities of the present HADCM2 experiments. Although Kittel et al. (1998) only considered 10-year long periods, this still suggests

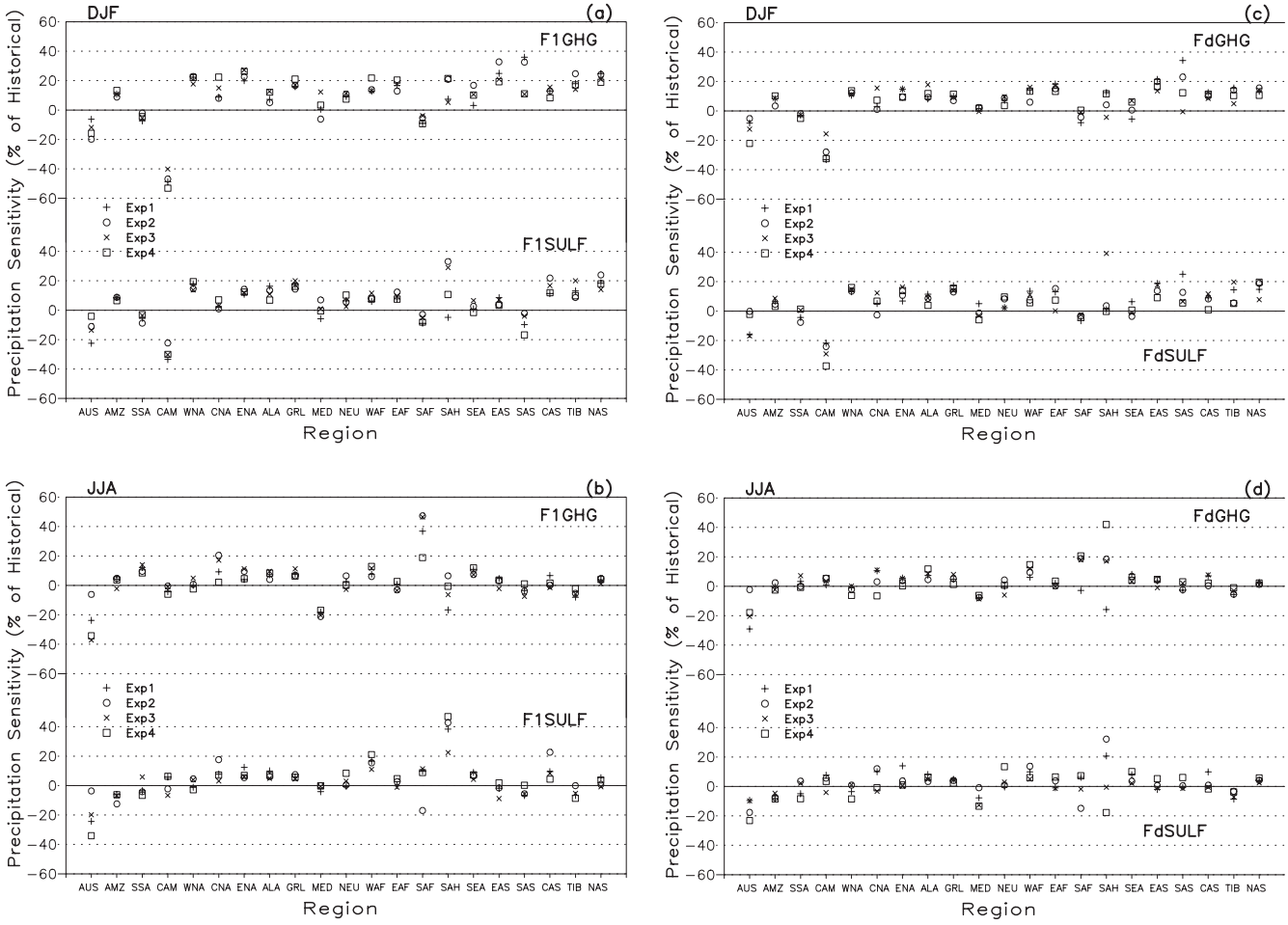

Fig. 8a, d Precipitation sensitivities for different future climate transient experiments (see Table 1) and all regions of Fig. 2. All realizations of the ensembles are included (EXP1-4). Units are percentage of historical (1961–1990) precipitation. **a** DJF, 1% year^{-1} increase in GHG concentration, with and without sulfate effects; **b** JJA, 1% year^{-1} increase in GHG concentration, with and without sulfate effects; **c** DJF, 0.5% year^{-1} increase in GHG concentration, with and without sulfate effects; **d** JJA, 0.5% year^{-1} increase in GHG concentration, with and without sulfate effects

that the uncertainty in internal model variability for temperature is much lower than the uncertainty in inter-model variability.

Figure 5 also shows the ensemble average interannual standard deviation of temperature for the future climate scenarios over all regions. It is difficult to draw definite conclusions, but the data in Fig. 5 would suggest a general, although small, increase in temperature variability during DJF. In JJA the signal is more mixed, as especially the North American regions show a general decrease in interannual variability in the future climate period. These relatively small changes in variability, however, cannot be taken with a high level of confidence, as they are much smaller than the errors in simulating variability by the model compared to observations.

3.2.2 Precipitation

Figure 8a–d shows the precipitation sensitivities for each realization and scenario, while Tables 7, 8 present the corresponding ensemble average sensitivities and spreads. In all cases, the sensitivities are expressed as a percentage of the 1961–1990 simulated values.

Over most regions, seasons and experiments the precipitation sensitivites are positive, i.e. precipitation tends to increase in the future climate change scenarios. This is in line with the notion, and previous modelling experience, of a general intensification of the atmospheric hydrologic budget under the warming induced by increased GHG concentration. This intensification is due to the higher evaporation rates implied by greater ocean and land temperatures and to the higher water holding capacity of a warmer atmosphere. It should be noted, however, that some of the regional sensitivites are negative. This is because regional precipitation changes are also associated with shifts in large-scale circulation patterns, such as jet stream location and storm tracks.

In general, the sensitivites in DJF are greater than in JJA. In DJF the greatest sensitivities are found over Australia, Central America, the North American and

Table 7 Ensemble average and spread of the regional precipitation sensitivities (2046–2075 minus 1961–1990) for the four future climate transient runs (see Table 1). Season is DJF. The spreads of the sensitivities are reported in parentheses. Units are percentage of historical (1961–1990) precipitation

Region	F1GHG	F1SULF	FdGHG	FdSULF
AUS	−13 (13)	−12 (18)	−12 (16)	−8 (16)
AMZ	10 (4)	8 (2)	7 (6)	5 (5)
SSA	−4 (5)	−5 (5)	−3 (3)	−2 (8)
CAM	−47 (12)	−29 (11)	−27 (17)	−28 (15)
WNA	21 (5)	16 (5)	11 (3)	14 (2)
CNA	13 (14)	3 (6)	6 (14)	5 (14)
ENA	23 (7)	12 (3)	12 (5)	11 (9)
ALA	9 (7)	11 (9)	11 (9)	8 (7)
GRL	17 (5)	16 (5)	9 (4)	14 (4)
MED	2 (18)	0 (12)	1 (2)	−1 (10)
NEU	9 (4)	6 (7)	7 (6)	5 (7)
WAF	15 (9)	8 (5)	12 (10)	9 (7)
EAF	16 (7)	9 (4)	15 (4)	8 (15)
SAF	−6 (5)	−6 (6)	−3 (8)	−4 (3)
SAH	13 (16)	16 (37)	5 (16)	10 (39)
SEA	10 (13)	2 (8)	1 (12)	0 (9)
EAS	24 (13)	5 (5)	17 (7)	14 (9)
SAS	22 (25)	−8 (14)	17 (34)	12 (19)
CAS	12 (7)	15 (10)	10 (4)	7 (10)
TIB	18 (10)	12 (11)	11 (10)	11 (14)
NAS	22 (5)	18 (9)	13 (4)	15 (11)

Table 8 Ensemble average and spread of the regional precipitation sensitivities (2046–2075 minus 1961–1990) for the four future climate transient runs (see Table 1). Season is JJA. The spreads of the sensitivities are reported in parentheses. Units are percentage of historical (1961–1990) precipitation

Region	F1GHG	F1SULF	FdGHG	FdSULF
AUS	−25 (31)	−20 (30)	−17 (26)	−15 (13)
AMZ	2 (6)	−7 (6)	0 (4)	−7 (3)
SSA	11 (5)	−2 (12)	2 (7)	−1 (12)
CAM	−2 (5)	0 (12)	3 (4)	3 (11)
WNA	0 (7)	1 (7)	−2 (6)	−2 (9)
CNA	12 (18)	8 (14)	4 (17)	4 (15)
ENA	7 (7)	7 (7)	3 (5)	4 (13)
ALA	7 (5)	6 (5)	8 (7)	5 (4)
GRL	7 (4)	5 (2)	4 (6)	3 (2)
MED	−19 (4)	−1 (4)	−7 (2)	−8 (12)
NEU	1 (9)	2 (9)	0 (10)	4 (13)
WAF	9 (6)	15 (10)	10 (8)	8 (8)
EAF	0 (6)	1 (5)	1 (3)	1 (7)
SAF	37 (28)	3 (28)	13 (23)	0 (22)
SAH	−4 (23)	37 (24)	15 (57)	8 (49)
SEA	9 (4)	6 (4)	5 (4)	5 (8)
EAS	2 (6)	−2 (10)	2 (5)	0 (7)
SAS	−3 (8)	−4 (7)	0 (5)	1 (7)
CAS	1 (8)	11 (18)	4 (7)	2 (11)
TIB	−5 (5)	−5 (8)	−3 (4)	−5 (5)
NAS	3 (3)	2 (6)	1 (1)	4 (3)

the Asian regions. In JJA, Australia, the Mediterranean and the West and Southern Africa regions have the highest sensitivites. In addition, the largest instances of precipitation sensitivity mostly occur for the F1GHG experiment, although a pronounced inter regional and inter-experiment variability is evident. The precipitation sensitivities were statistically significant at the 5% confidence level in 55–76% of the regional cases in DJF, and in 18–43% of the cases in JJA.

An interesting feature of the sensitivities in Tables 7, 8 is that, in almost all cases during DJF and in the majority of cases during JJA, the ensemble average regional precipitation sensitivities have the same sign in all four different forcing scenarios, be they positive or negative. This implies a coherency in the regional scale model response to the forcing of GHG and sulfate aerosol under different magnitudes of these forcings. Despite this coherency, a systematic trend in precipitation response among different scenarios, similar to that found for temperature, is not evident. This is because of the non-linear nature of precipitation, which depends on large scale circulations, local forcings and release of condensation heat. For this reason, the issue of regional "scaling" for different forcing scenarios is more difficult for precipitation than for temperature.

The spread for the precipitation sensitivities are more pronounced compared to those for the temperature and precipitation biases and the temperature sensitivities. In about 20–30% of regional cases the simulated sensitivities did not have the same sign in all realizations of the ensemble. In fact, there appears to be a direct relationship between the magnitude of the sensitivity and the spread, with greatest values for both quantitities in the F1GHG experiments.

Despite the generally greater importance of the spread for precipitation sensitivity than temperature sensitivity, Fig. 8a, d is still indicative of a prevailing agreement among the simulated responses by different realizations. Most of the spread values are less than 20%, and in many cases less than 10%, and in most of the cases in which the sign of the response does not agree among realizations the magnitude of the response is small. Indeed, regional precipitation averages between pairs of future climate realizations in the same ensemble did not differ from each other at the 5% confidence level in 84–98% of the cases, depending on the scenario.

An important difference between the precipitation and temperature results is that, for precipitation, the spread values are of the same order of magnitude as the range of sensitivities due to the different scenarios. This implies that, for precipitation, the uncertainty related to the internal model variability is of the same magnitude as the uncertainty associated with the different scenarios. On the other hand, the spreads of Tables 7, 8 are generally smaller than the range of sensitivites found for different models by Kittel et al. (1998), which was as high as 50–80% for some regions.

Finally, compared to the present day periods, in the future climate experiments the model showed a tendency to simulate an increase in interannual variability (Fig. 7a, b), especially during DJF over the Central

America, central North America, Southeast Asia and south Asia. Again, in view of the small simulated changes and the relatively large overprediction of observed variability, this simulated trend cannot be taken with a high level of confidence.

3.2.3 Summary

For the future climate period, the HADCM2 model simulated warming over all regions in the range of a few to over 5 K for the different scenarios and a prevailing, although not ubiquitous, increase in precipitation, mostly of magnitude within 20% of the present day value. The averaged regional sensitivities were all statistically significant at the 5% confidence level for temperature, while for precipitation they were significant in 55–76% of regional cases in DJF and 18–43% of cases in JJA.

The spread in the sensitivities was generally larger than the spread in the biases. However, the simulated sensitivities showed a high degree of coherency among different realizations of the same scenario, both for temperature and precipitation. The model showed a tendency to simulate larger spreads in correspondence of larger sensitivities, a behavior suggestive of the occurrence of positive feedback mechanisms. Both for surface air temperature and precipitation, most pairs of averages of future climate realizations in an ensemble were not statistically different from each other at the 5% confidence level.

Finally, the simulated interannual variability showed a tendency to increase in the future climate periods compared to the present day periods. However, this tendency was weak and much smaller than the model errors in simulating present day variability.

4 Discussion and conclusions

Our analysis of the HADCM2 ensemble transient runs has given a number of indications. First, compared with previous AOGCM experiments, the HADCM2 runs have shown a good performance in reproducing historical (1961–1990) regionally and seasonally averaged surface air temperature and precipitation. This result is likely influenced by the use of ocean flux correction in the model. It should be stressed that this does not imply that AOGCM climate data should be directly used in climate assessment work. Even though the averaged values of surface climate variables over broad regions may be accurately simulated by an AOGCM, the detailed spatial structure of the variables is determined by local forcings, such as topography and landuse, mesoscale circulation structure and local feedback processes. Indeed, regional climate model and statistical downscaling studies have shown that surface climate sensitivities are characterized by pronounced spatial variability at the sub-GCM grid scale (e.g. Wigley et al. 1990; Giorgi et al. 1994). Therefore it is always advisable to use a regionalization tool to enhance the information provided by AOGCMs. However, the relatively good performance of the HADCM2 is indicative of an improvement of the latest generation of AOGCMs in providing the broad scale information that influences regional climate change.

Despite the improvement shown by the HADCM2 in reproducing observed climatic averages, the model still showed relatively large errors in simulating observed interannual variability, although it is likely that the uncertainty in the observations is greater for interannual variability than for climatic averages. Since changes in variability at different temporal scales are often more important than changes in climatic averages for impact assessment studies, it is extremely important that variability in climate models be thoroughly investigated and possibly improved in future work.

The primary purpose of our work was to investigate the importance of what we referred to as internal model variability, i.e. variability associated with differences among different realizations of the same transient climate scenario. Our analysis showed that the spread among realizations was small compared to the model biases for the historical period and compared to the model sensitivites for the future climate periods. The climate averages for pairs of realizations in an ensemble were not statistically different from each other at the 5% confidence level in the vast majority of regional cases. Finally, with the exception of a few regions for precipitation, the spread of sensitivites among different realizations was smaller than the range of sensitivites found for the different GHG and sulfate scenarios and smaller than the range of sensitivities found for different models by Kittel et al. (1998). These conclusions are essentially valid for different scenarios, seasons, and regional settings, although positive feedback processes, such as the snow/ice albedo feedback mechanism, contributed to increase the internal model variability over specific regions. Both for temperature and precipitation, the regional sensitivities among different scenarios showed coherency but significant inter regional variability, a result which is relevant for the issue of regional scaling of climate change scenarios.

Going back to the discussion of Sect. 1, our results indicate that the uncertainty associated with the internal AOGCM model variability is lower than the uncertainty associated with inter-model or inter-scenario variability. Again, this result may depend on the use of ocean flux correction in the HADCM2. With this caveat, our analysis shows that, in terms of climatic averages, a small sample of realizations (3–4) is sufficient to characterize the regional climate change response of a given AOGCM to a given scenario.

Acknowledgements We thank the Hadley Centre for Climate Prediction and Research for making available the results from the HADCM2 simulations and the Climatic Research Unit of the University of East Anglia for making available the observation

datasets. We also thank two reviewers for useful comments on the manuscript.

References

Edwards AL (1984) An introduction to linear regression and correlation. WH Freeman, New York, pp 81–83

Giorgi F, Mearns LO (1991) Approaches to regional climate change simulation: a review. Rev Geophys 29: 191–216

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

IPCC (1992) Climate change 1992: the supplementary report to the IPCC scientific assessment. Prepared by IPCC Working Group I. Houghton JT, Callander BA and Varney SK (eds), and WMO/UNEP. Cambridge University Press, Cambridge, UK, 200 pp

IPCC (1996) Kattenberg A, Giorgi F, Grassl H, Meehl GA, Mitchell JFB, Stouffer RJ, Tokioka T, Weaver AJ, Wigley TML. Climate models-projections of future climate. In: JT Houghton, LG Meira Filho, BA Callander, N Harris, A Kattenberg, K Maskell (eds) Climate change 1995: the science of climate change, Contribution of Working Group I to the Second Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, New York, pp 285–357

Johns TC, Carnell RE, Crossley JF, Gregory JM, Mitchell JFB, Senior CA, Tett SFB, Wood RA (1997) The second Hadley Centre coupled ocean-atmosphere GCM: model description, spin up and validation. Clim Dyn 13: 103–134

Kittel TGF, Giorgi F, Meehl GA (1998) Intercomparison of regional biases and doubled CO_2 sensitivity of coupled atmosphere-ocean general circulation model experiments. Clim Dyn 14: 1–15

Machenhauer B, Windelband M, Botzet M, Christensen JH, Deque M, Jones RG, Ruti PM, Visconti G (1998) Validation and analysis of regional present-day climate and climate simulations over Europe. MPI Report 275, Max-Planck-Institute, Hamburg, 86 pp

Marx J, Fennessy M (1999) Seasonal atmospheric prediction. Mon Weather Rev (in press)

Mitchell JFB, Davis RA, Ingram WJ, Senior CA (1995a) On surface temperature, greenhouse gases and aerosols: models and observations. J Clim 8: 2364–2386

Mitchell JFB, Johns TC, Gregory JM, Tett SFB (1995b) Climate response to increasing levels of greenhouse gases and sulfate aerosols. Nature 376: 501–504

New MG, Hulme M, Jones PD (1999a) Representing twentieth-century space time climate variability. Part I: development of a 1961–1990 mean monthly terrestrial climatology. J Clim 12: 829–856

New MG, Hulme M, Jones PD (1999b) Representing twentieth-century space time climate fields. Part II: development of a 1901–1996 mean monthly terrestrial climatology. Submitted to J Clim

Phillips GM, Taylor PJ (1996) Theory and applications of numerical analysis. Academic press, London, pp 54–58

Santer BD, Wigley TML, Schlesinger ME, Mitchell JFB (1990) Developing climate scenarios from equilibrium GCM results. Max-Planck Institut fur Meteorologie, Report 47, Hamburg, 29 pp

Wigley TML, Jones PD, Briffa KR, Smith G (1990) Obtaining sub-grid-scale information from coarse resolution general circulation model output. J Geophys Res 95: 1943–19537