ARTICLE IN PRESS

ELSEVIER

GLOBAL AND PLANETARY CHANGE

Global and Planetary Change 769 (2002) 1–31

www.elsevier.com/locate/gloplacha

An overview of results from the Coupled Model Intercomparison Project

Curt Covey[a,*], Krishna M. AchutaRao[a], Ulrich Cubasch[b], Phil Jones[c], Steven J. Lambert[d], Michael E. Mann[e], Thomas J. Phillips[a], Karl E. Taylor[a]

[a] *Program for Climate Model Diagnosis and Intercomparison (PCMDI), Lawrence Livermore National Laboratory, Mail Code L-264, Livermore, CA 94551, USA*

[b] *Max-Planck-Insitut fuer Meteorologie (MPI), Hamburg, Germany*

[c] *Climatic Research Unit (CRU), University of East Anglia, Norwich, UK*

[d] *Canadian Centre for Climate Modelling and Analysis (CCCma), Victoria, Canada*

[e] *Department of Environmental Sciences, University of Virginia, Charlottesville, VA, USA*

Received 1 March 2001; accepted 6 July 2002

Abstract

The Coupled Model Intercomparison Project (CMIP) collects output from global coupled ocean–atmosphere general circulation models (coupled GCMs). Among other uses, such models are employed both to detect anthropogenic effects in the climate record of the past century and to project future climatic changes due to human production of greenhouse gases and aerosols. CMIP has archived output from both constant forcing ("control run") and perturbed (1% per year increasing atmospheric carbon dioxide) simulations. This report summarizes results form 18 CMIP models. A third of the models refrain from employing ad hoc flux adjustments at the ocean–atmosphere interface. The new generation of non-flux-adjusted control runs are nearly as stable as—and agree with observations nearly as well as—the flux-adjusted models. Both flux-adjusted and non-flux-adjusted models simulate an overall level of natural internal climate variability that is within the bounds set by observations. These developments represent significant progress in the state of the art of climate modeling since the Second (1995) Scientific Assessment Report of the Intergovernmental Panel on Climate Change (IPCC; see Gates et al. [Gates, W.L., et al., 1996. Climate models—Evaluation. Climate Climate 1995: The Science of Climate Change, Houghton, J.T., et al. (Eds.), Cambridge Univ. Press, pp. 229–284]). In the increasing-CO_2 runs, differences between different models, while substantial, are not as great as one might expect from earlier assessments that relied on equilibrium climate sensitivity.

Keywords: CMIP; GCM; Climate

1. Introduction

Global coupled ocean–atmosphere general circulation models (coupled GCMs) that include interactive sea ice simulate the physical climate system, given only a small number of external boundary conditions

* Corresponding author. Tel.: +1-925-422-1828; fax: +1-925-422-7675.

E-mail address: covey1@llnl.gov (C. Covey).

0921-8181/02/$ - see front matter © 2002 Published by Elsevier Science B.V.
PII: S0921-8181(02)00193-5

such as the solar "constant" and atmospheric concentrations of radiatively active gases and aerosols. These models have been employed for decades in theoretical investigations of the mechanisms of climatic changes. In recent years, coupled GCMs have also been used to separate natural variability from anthropogenic effects in the climate record of the 20th century, and to estimate future anthropogenic climate changes including global warming. A number of coupled GCMs have been developed by different research groups. For some time it has been apparent that these models give somewhat contradictory answers to the same questions—e.g., a range from roughly 1.5 to 4.5 °C in the global mean surface air temperature increase due to a doubling of atmospheric carbon dioxide—due to subtle differences in their assumptions about clouds and other phenomena at scales smaller than the separation of model grid points (Cess et al., 1989; Mitchell et al., 1989).

In 1995, the JSC/CLIVAR Working Group on Coupled Models, part of the World Climate Research Program, established the Coupled Model Intercomparison Project (CMIP; see Meehl et al., 2000). The purpose of CMIP is to provide climate scientists with a database of coupled GCM simulations under standardized boundary conditions. CMIP investigators use the model output to attempt to discover why different models give different output in response to the same output, or (more typically) to simply identify aspects of the simulations in which "consensus" in model predictions or common problematic features exist. CMIP may be regarded as an analog of the Atmospheric Model Intercomparison Program (AMIP; see Gates et al., 1999). In the AMIP simulations, sea ice and sea surface temperature are prescribed to match recent observations, and the atmospheric response to these boundary conditions is studied; in CMIP, the complete physical climate system including the oceans and sea ice adjust to prescribed atmospheric concentrations of CO_2.

Details of the CMIP database, together with access information, may be found on the CMIP Web site at http://www-pcmdi.llnl.gov/cmip/diagsub.html. The first phase of CMIP, called CMIP1, collected output from coupled GCM control runs in which CO_2, solar brightness and other external climatic forcing is kept constant. (Different CMIP control runs use different values of solar "constant" and CO_2 concentration, ranging from 1354 to 1370 W m^{-2} and 290 to 345 ppm, respectively; for details see http://www-pcmdi.llnl.gov/cmip/Table.htm.). A subsequent phase, CMIP2, collected output from both model control runs and matching runs in which CO_2 increases at the rate of 1% per year. No other anthropogenic climate forcing factors, such as anthropogenic aerosols (which have a net cooling effect), are included. Neither the control runs nor the increasing-CO_2 runs in CMIP include natural variations in climate forcing, e.g., from volcanic eruptions or changing solar brightness.

CMIP thus facilitates the study of intrinsic model differences at the price of idealizing the forcing scenario. The rate of radiative forcing increase implied by 1% per year increasing CO_2 is nearly a factor of 2 greater than the actual anthropogenic forcing in recent decades, even if non-CO_2 greenhouse gases are added in as part of an "equivalent CO_2 forcing" and anthropogenic aerosols are ignored (see, e.g., Fig. 3 of Hansen et al., 1997). Thus, the CMIP2 increasing-CO_2 scenario cannot be considered as realistic for purposes of comparing model-predicted and observed climate changes during the past century. It is also not a good estimate of future anthropogenic climate forcing, except perhaps as an extreme case in which the world accelerates its consumption of fossil fuels while reducing its production of anthropogenic aerosols. Nevertheless, this idealized scenario generates an easily discernible response in all the CMIP models and thus provides the opportunity to compare and possibly explain different responses arising from different model formulations.

The purpose of this report is to give an overview of the CMIP simulations with emphasis on common model successes and failures in simulating the present-day climate, and on common features of the simulated changes due to increasing CO_2. We pay extra attention to the three fields that CMIP provides at monthly mean time resolution: surface air temperature, sea level pressure and precipitation. The other fields are described here in terms of annual mean quantities. Extensive analyses of seasonal variations in the CMIP1 control runs is given by Covey et al. (2000) and Lambert and Boer (2001), and a more complete "atlas" of CMIP2 output—from which much of this report is extracted—is available online at http://www-pcmdi.llnl.gov/pcmdi/pubs/pdf/report66. More specialized studies of the CMIP database are summarized by Meehl et al. (2000) and the CMIP Web site at http://www-pcmdi.llnl.gov/cmip/abstracts.html. Also, very brief extracts from this report are presented in the most recent Scientific Assessment Report of the Intergovernmental Panel on Climate Change (IPCC; see McAvaney et al., 2001).

In this report, we include 18 models from the CMIP database (see Table 1). For most of our analysis we use the latest (CMIP2) version of each model, but for long-term variability (Section 2.4) we use models from both CMIP1 and CMIP2 provided the control runs are more than 200 simulated years long. As indicated in table, three of the models we use to study variability did not provide enough data to appear in the other sections of this report or (in one case) provided data too late for full incorporation. We nevertheless decided to include these models in our variability study in order to consider the greater possible number of models with long control runs. Finally, we exclude two CMIP2 models that employed fixed sea ice boundary conditions and one whose control run was only 3 simulated years long. (These excluded models are not shown in the table.) Complete documentation of all CMIP models is available on the CMIP Web site at http://www-pcmdi.llnl.gov/cmip/Table.htm and links therein.

2. Present-day climate

In this section, we compare output from the model control run simulations with recent climate observations. It has become increasingly apparent that the detailed climate record of the past century (and indeed the past millenium) cannot be explained without considering changes in both natural and anthropogenic forcing (Tett et al., 1999; Santer et al., 2000;

Table 1
Models used for this study and sections in which they are used

	Model	Key references	Flux correction	Control run length (year)	Section
1	BMRC	Power et al., 1998	heat, water	80	2.1–2.3, 3
2	CCCMA	Flato et al., 2000; Boer et al., 2000; Flato and Boer, in press	heat, water	150	2.1–2.3, 3
3	CCSR	Emori et al., 1999	heat, water	200	2.1–2.3, 3
4	CERFACS	Barthelet et al., 1998a,b	NONE	80	2.1–2.3, 3
5	CSIRO	Gordon and O'Farrell, 1997	heat, water, momentum	100	2.1–2.3, 3
6	DOE PCM	Washington et al., 2000	NONE	300	2.1–2.4, 3
7	ECHAM1 + LSG	Cubasch et al., 1992; von Storch et al., 1997	heat, water, momentum	960	2.4
8	ECHAM3 + LSG	Cubasch et al., 1997; Voss et al., 1998	heat, water, momentum	1000	2.1–2.4, 3
9	ECHAM4 + OPYC3	Roeckner et al., 1996a,b	heat, water (ann. mean)	240	2.1, 2.3, 2.4, 3
10	GFDL	Manabe et al., 1991; Manabe and Stouffer, 1996	heat, water	1000	2.1–2.4, 3
11	GFDL R30	Delworth and Knutson, 2000	heat, water	300	2.4
12	GISS	Russell et al., 1995; Russell and Rind, 1999	NONE	98	2.1–2.3, 3
13	IAP/LASG	Wu et al., 1997; Zhang et al., 2000	heat, water, momentum	80	2.1–2.3, 3
14	LMD/IPSL	Laurent et al., 1998; Leclainche et al., submitted for publication	NONE	301	2.1–2.3, 3[a]
15	MRI	Tokioka et al., 1996	heat, water	80	2.1–2.3, 3
16	NCAR CSM	Boville and Gent, 1998	NONE	300	2.1–2.4, 3
17	UKMO HadCM2	Johns, 1996; Johns et al., 1997	heat, water	1085	2.1–2.4, 3
18	UKMO HadCM3	Gordon et al., 2000	NONE	400	2.1–2.4, 3

[a] The model used for variability study (Section 2.4) is a slight modification of the version used in other sections of this report (Dufresne et al., submitted for publication).

Crowley, 2000). Since the CMIP control run boundary conditions lack these forcing variations, we focus on means and other statistics that we judge to be largely unaffected by them. In the final part if this section we discuss the climate variability simulated by the CMIP control runs. This topic has also been addressed in more specialized studies (Barnett, 1999; Bell et al., 2000a, in press; Duffy et al., 2000).

For our observational data base we use the most recent and reliable sources we are aware of, including Jones et al. (1999) for surface air temperature, Xie and Arkin (1997) for precipitation, and reanalysis of numerical weather predictions initial conditions for sea level pressure. We sometimes use multiple sources to provide a sense of observational uncertainty, e.g., reanalysis from both the European Centre for Medium-Range Weather Forecasts (ERA15; Gibson et al., 1997) and the U.S. National Centers for Environmental Prediction (NCEP; Kalnay et al., 1996).

2.1. Global and annual means

Averaging over latitude and longitude to form global means reduces surface variable to one-dimensional time series. Additional averaging of monthly means to form annual means removes seasonal cycle variations (which can be substantial even for global means), providing a convenient entry point to three-dimensional model output. Fig. 1 shows the resulting time series for CMIP2 control run surface air temperature and precipitation.

The range among the models of global- and annual-mean surface air temperature is rather surprising. Jones et al. (1999 conclude that the average value for 1961–1990 was 14.0 °C and point out that this value differs from earlier estimates by only 0.1 °C. Taking into consideration all of the observational uncertainties, it appears that the actual value of surface air temperature was between 13.5 and 14.0 °C during the second half of the 20th Century and roughly 0.5 °C less in the late 19th Century. It therefore seems that several of the models (which simulate values from less than 12 °C to over 16 °C) are in significant disagreement with the observations of this fundamental quantity. Reasons for this situation are discussed briefly by Covey et al. (2000) in the context of the CMIP1 models. A natural question to ask is whether the spread in simulated temperatures is correlated with variations in planetary albedo among the models. Unfortunately, the CMIP1 and CMIP2 database does not include the energy balance at the top of the atmosphere. This information is being collected under an expanded version of the database (described in Section 4), and results to date are compared with observations in Table 2. While definite conclusions are not possible at this time, it is noteworthy that for the five models in hand the simulated values are close to each other and to the observations.

The CMIP2 models as a group also give a wide range of estimates for global- and annual-mean precipitation, compared with the best observed values from several sources (2.66–2.82 mm/day from Table 2 in Xie and Arkin, 1997). Precipitation, however, is notoriously difficult to measure globally, and the observational uncertainty of its global and annual mean may not be smaller than the range of model-simulated values in Fig. 1.

Perhaps the most striking aspect of Fig. 1 is the stability of model-simulated temperature and precipitation. The stability occurs despite the fact that 6 of the 16 CMIP2 models refrain from employing ad hoc flux adjustments at the air–sea interface. Until a few years ago, conventional wisdom held that in order to suppress unrealistic climate drift, coupled ocean–atmosphere general circulation models must add such unphysical flux "corrections" to their governing equations. The 1995 IPCC assessment (Gates et al., 1996) diplomatically expressed the concern that "[f]lux adjustments are relatively large in the models that use them, but their absence affects the realism of the control climate and the associated feedback processes". The CMIP1 experiments were conducted at about the same time as his assessment was written. Covey et al. (2000) note that averaging the magnitudes of linear trends of global- and annual-mean surface air temperature gives 0.24 and 1.1 °C/century, respectively, for flux-adjusted and non-flux-adjusted CMIP1 models. For the CMIP2 models shown in Fig. 1, however, the corresponding numbers for the average ± 1 standard deviation over each class of model are 0.13 ± 0.13 °C/century for the flux-adjusted models and 0.31 ± 0.31 °C/century for the non-flux-adjusted models. Nevertheless, it must be kept in mind that a small rate of global mean climate drift

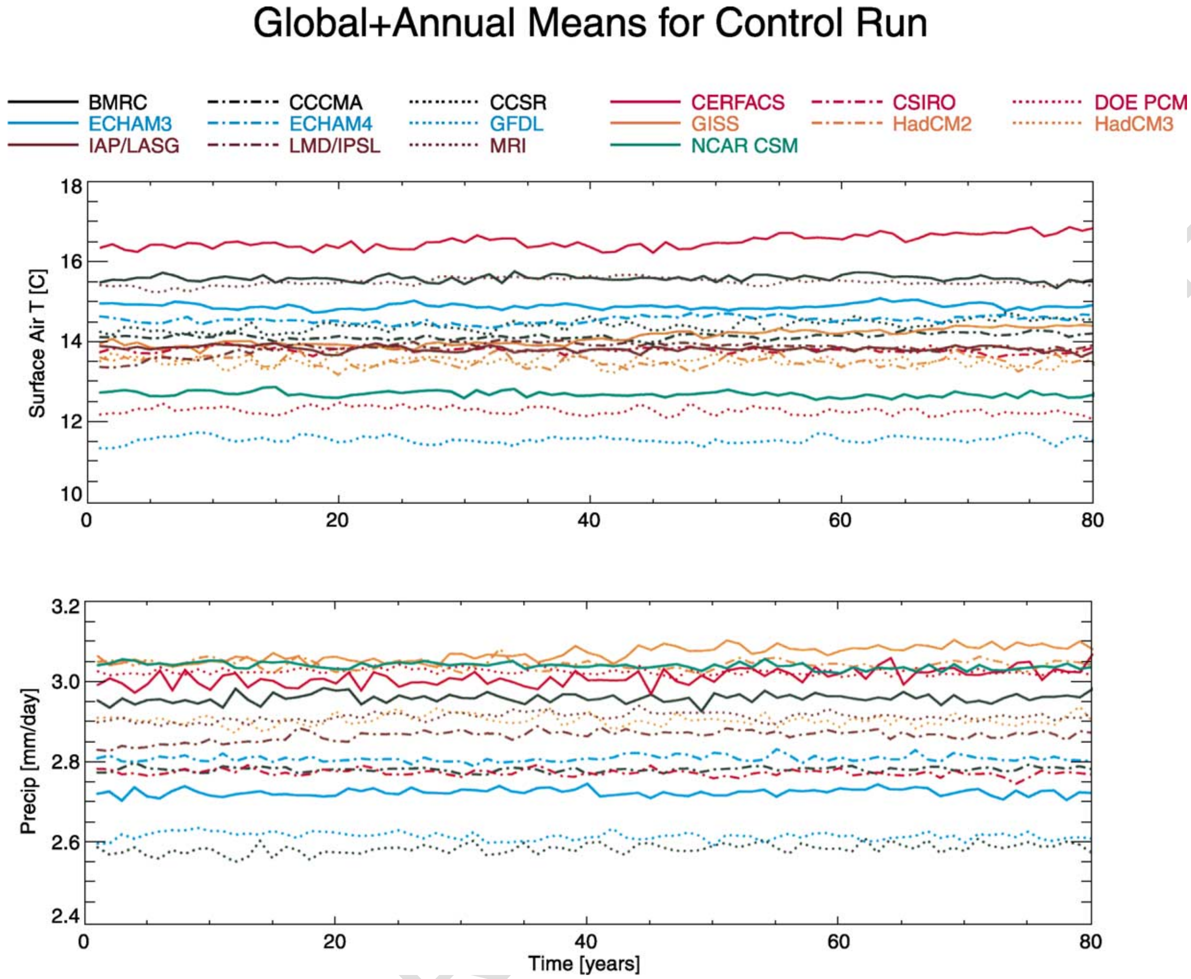


Fig. 1. Globally averaged annual mean surface air temperature (top) and precipitation (bottom) from the CMIP2 control runs.

does not preclude strong local drifts at the surface and problematic long-term drift in the deep ocean.

2.2. Long-term time means

As noted above, most of the CMIP2 output variables are present in the database as 20-years means that the average out of the seasonal cycle. In this subsection, we examine surface variables and the other two-dimensional quantities. To summarize the performance of the models in latitude–longitude space, we interpolate their output to the common Gaussian grid with 128 longitudes and 64 latitudes. We show both the model mean (the average over all the models)

Table 2
Global and annual mean top-of-atmosphere energy balance

	ERBE obs	CSM[a]	CSM[a]	GFDL R30	HadCM2	HadCM3	PCM
Outgoing long wave [W m^{-2}]	236.3	238.4	238.4	235.0	235.5	240.8	237.2
Absorbed solar [W m^{-2}]	241.1	238.3	238.5	235.3	235.0	240.6	237.2
Albedo	0.293	0.302	0.301	0.310	0.311	0.295	0.305

[a] The two CSM results are taken from two different non-overlapping segments of the same control run.

and the intermodel standard deviation (sd_m). Where possible, we compare the model means for the control simulation with observations. Lambert and Boer (2001) demonstrate that the model mean exhibits good agreement with observations, often better that any of the individual models. High values of sd_m indicate areas where the models have difficulty in reaching a consensus, implying reduced levels of confidence in the model result.

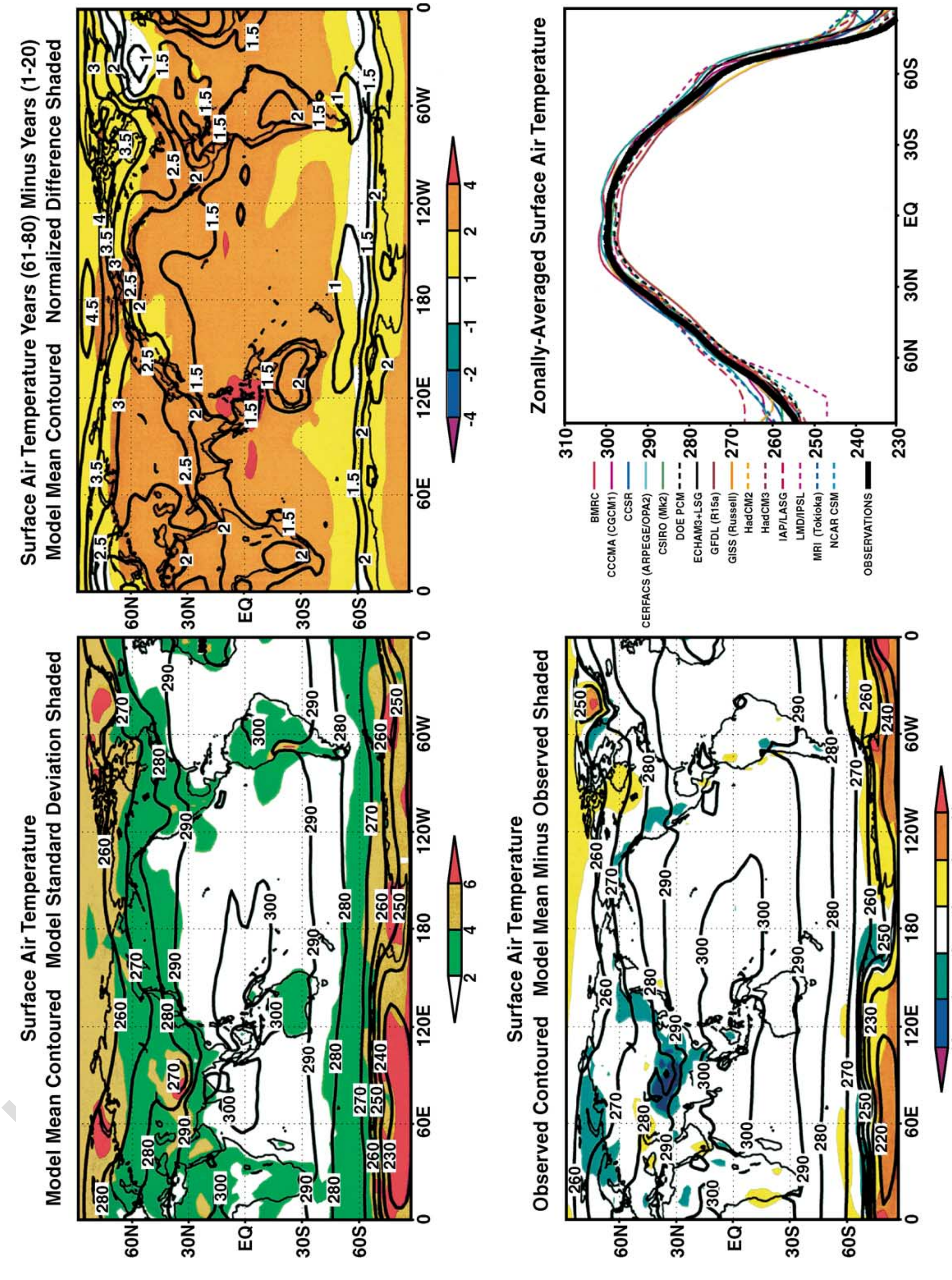
Results for which observations are available are presented as four-panel displays. The upper-left panel shows the model mean and sd_m, the lower-left panel shows the observed field and the departure of the model mean from this observed field, and the lower-right panel shows zonal averages for the individual models and the observations. These three panels contain only output from model control runs. The upper-right panel gives the differences between the model mean for years 60–80 and years 1–20 for the enhanced greenhouse warming simulations, together with these differences normalized by their standard deviation among the models. Result in the upper-right panel will be discussed in Section 3.

Fig. 2 displays result for annual mean surface air temperature (also known as screen temperature). Over most of the globe, the model mean differs from the Jones observations by less than two °C, although larger differences are evident in polar regions. These annual departures are much less that the winter and summer season errors reported by Lambert and Boer (2001). The zonally averaged results for the individual models show that all the quite successful in reproducing the observed structure, except in the polar regions. sd_m values show that the models tend to disagree in the polar regions and over high terrain but produce consistent simulations over ice-free oceans. This consistency may occur because the ocean components of coupled models tend to be more similar that their atmospheric components, or it may simply be due to the lack of terrain effect and strong horizontal gradients over open oceans.

Fig. 3 displays results for annual mean sea level pressure. As demonstrated by sd_m, the models are very consistent in their simulations. The largest variances occur in south polar regions and much of this results from extrapolation below ground. Comparison with the ECMWF/ERA reanalysis (Gibson et al., 1997) shows that the model mean is within 2 hPa of the observed field over most of the globe. The largest departures occur near Antarctica with lesser departures north of Scandinavia, Russia and western North America. The zonally averaged results demonstrate the agreement among the models. With the exception of one model and in the southern polar regions, the models agree with each other to within $\sim$ 5 hPa. Also evident from the zonally averaged results, however, is the difficulty that models have in simulating both the position and depth of the Antarctic trough. This difficulty implies (by geostropic balance) that most models have trouble correctly simulating wind stress in this region, an important factor in ocean–atmosphere coupling.

Fig. 4 displays result for annual mean precipitation. It is evident from the relatively large sd_m that the models have difficulty in producing consistent simulations. This result is expected because precipitation is a small-scale process. Likely contributors to inconsistency among models include differences in horizontal resolution and sub-gridscale parameterization schemes. Precipitation is a difficult field to observe and thus one must be somewhat cautious in using it for evaluation purposes. (Comparison of surface air temperature, sea level pressure and precipitation with alternate observational datasets is given Section 2.3.) Using the Xie and Arkin (1997) observations, we find that in general the models simulate $\sim$ 1 mm/day too much precipitation in mid-latitudes and somewhat too little in the tropics. The models correctly simulate the position of the annual mean ITCZ slightly north of the equator, but a disagreement with observations occur in the South Pacific. Here the model mean has a

Fig. 2. Summary of long-term time means for surface air temperature (K). The upper-left panel gives the control run 80-year mean averaged over all models (contours) and the intermodel standard deviation (color shading). The lower-left panel gives observed values (contours) and the difference between the control run model mean and the observations (color shading). The lower-right panel gives zonal averages for the individual model control runs and the observations. The upper-right panel gives the average over all models of the difference between the last 20-year mean and the first 20-year mean from the 80-year perturbation simulations, in which atmospheric carbon dioxide increases at a rate of 1% per year (contours), together with this difference normalized by the corresponding intermodel standard deviation (color shading).


Surface Air Temperature
Model Mean Contoured Model Standard Deviation Shaded
Surface Air Temperature Years (61-80) Minus Years (1-20)
Model Mean Contoured Normalized Difference Shaded
Surface Air Temperature
Observed Contoured Model Mean Minus Observed Shaded
Zonally-Averaged Surface Air Temperature
BMRC
CCCMA (CGCM1)
CCSR
CERFACS (ARPEGE/OPA2)
CSIRO (Mk2)
DOE PCM
ECHAM3+LSG
GFDL (R15a)
GISS (Russell)
HadCM2
HadCM3
IAP/LASG
LMD/IPSL
MRI (Tokioka)
NCAR CSM
OBSERVATIONS

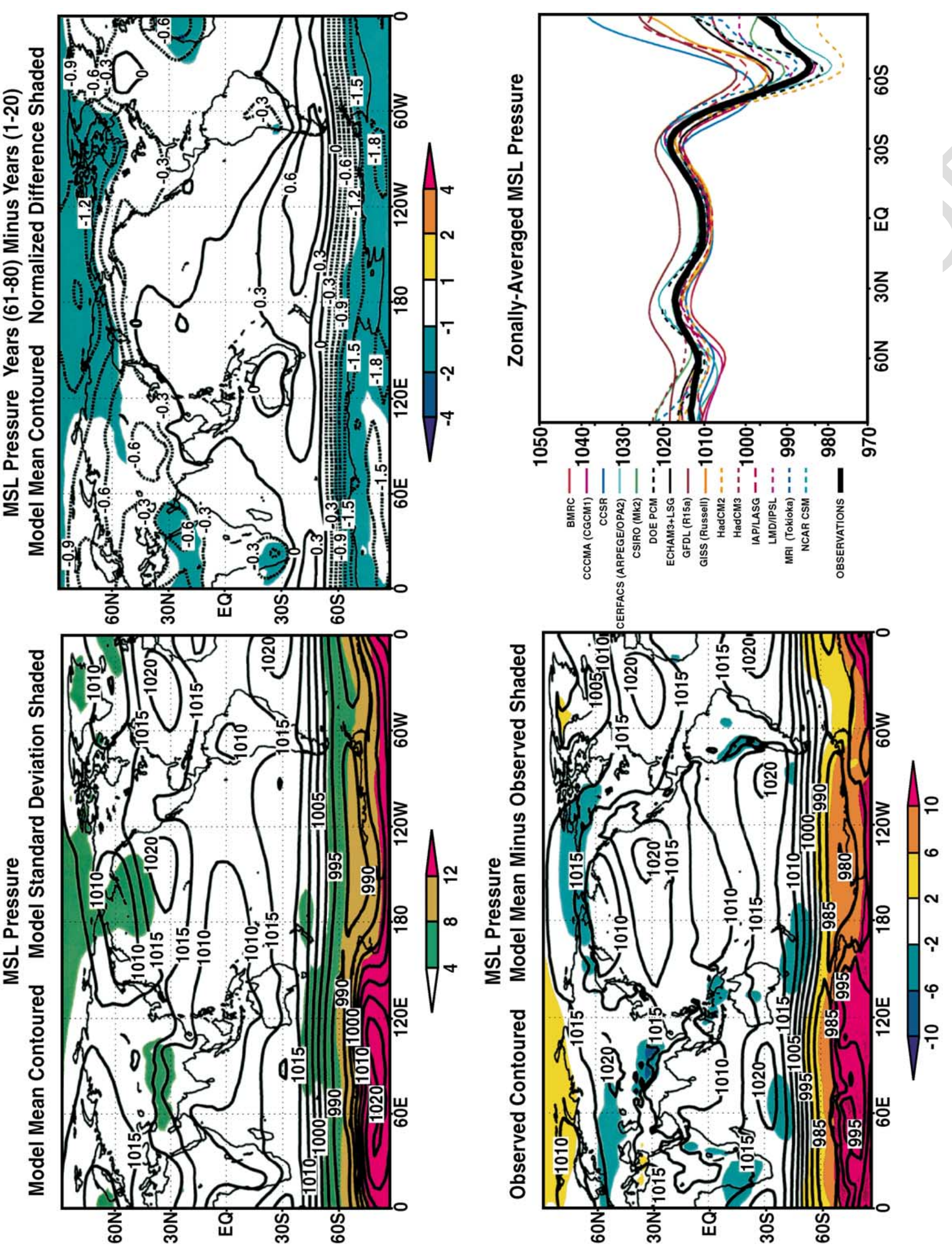


Fig. 3. Same as Fig. 2 for mean sea level pressure (hPa).

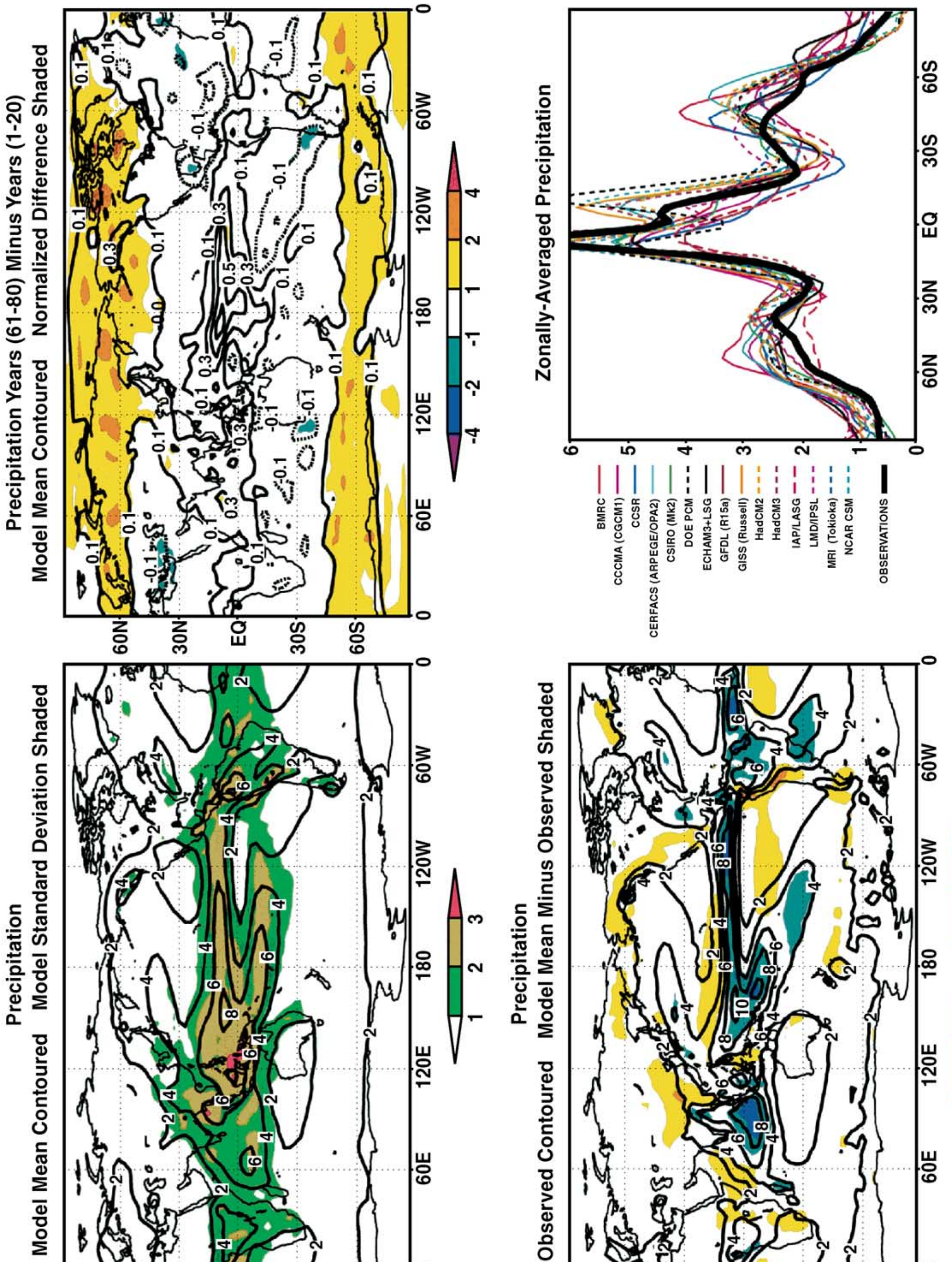


Fig. 4. Same as Fig. 2 for precipitation (mm/day).

second maximum band roughly parallel to the Equator, but the observations have a maximum with a northwest–southeast orientation north of New Zealand (the so-called South Pacific Convergence Zond or SPCZ). The zonally averaged results show that the "double ITCZ" problem is shared by several of the models.

We now turn to three-dimensional atmospheric quantities, presented here (after zonal averaging) as latitude-height sections. Fig. 5 shows zonal averaged annual mean air temperature. The pattern of the model mean isotherms is qualitatively close to observations, but compared with the ECMWF/ERA reanalysis, the model mean is generally too cold in the troposphere and polar stratosphere and too warm at lower latitudes in the stratosphere. The magnitude of these errors is comparable to sd_m, implying are common to most of the models. Results for the individual models at 925 hPa confirm this simulation for the cold bias at low levels, but they also show that near the surface the latitude gradient of temperature is accurately simulated outside the polar regions. The corresponding model-simulated mean zonal winds in the lower troposphere (not shown) agree to within ~ 2 m/s with each other and with the ECMWF/ERA reanalysis except in the vicinity of the Antarctic trough. Results for specific humidity (Fig. 6) display a fairly systematic underestimate in the low latitude troposphere, although the departure of the model mean from ECMWF/ERA reanalysis is rather small (~ 1 g/kg) and the pattern of the model mean in latitude-height space is again quite similar to observations.

Turning to ocean variables, we show (Fig. 7) the annual mean temperature at 1000 m depth. (Sea surface temperature is closely coupled to surface air temperature over the oceans and is not explicitly discussed in this report.) At this level the models are generally consistent in their simulation ($sd_m < 1$ °C) except in the North Atlantic, subtropical Pacific and Indian Oceans, and in the Arabian Sea. Available observations (Levitus and Boyer, 1994) indicate that the model mean is too warm over most of the ocean. The zonally averaged results show that outside the polar regions, all but one of the models simulate 1000 m temperatures that are at or above (by up to ~ 2 °C) the observations. An overly diffusive thermocline may be root of this problem. The corresponding results for salinity (not shown) exhibit relatively large sd_m values.

For the annual means of barotropic streamfunction (Fig. 8) and global overturning streamfunction (Fig. 9) we use three-panel displays because there are no complete observations of these quantities. Nevertheless, it is noteworthy that the model means for all three agree qualitatively with conventional wisdom among oceanographers. Quantitative disagreement among the models is most striking for the barotropic streamfunction in the Southern Hemisphere, where as noted earlier the near-surface temperature, pressure and wind stress simulations disagree significantly.

Poleward heat transport by the global ocean is given in Fig. 10. In the upper left-hand panel, the upper dashed line is the model mean plus one sd_m and the lower dashed line is the model mean minus one sd_m. The model mean, which is not plotted, is halfway between the two dashed lines. Observations of Trenberth and Solomon (1994) are shown as a bold solid in the both upper-left and bottom panels. From these observations, it appears that over most of the ocean the model-simulated transport is generally too weak.

The observation are uncertain, however. For example, an update (Trenberth, 1998) of the Trenberth and Solomon data reduces the peak ocean heat transport in the Southern Hemisphere by nearly a factor of 2.

Finally, control run sea ice thickness in the Arctic and Antarctic is given in the left-side panel of Fig. 11. Observations are not shown in the figure, but the limited data that exist on ice thickness (e.g., Rothrock et al., 1999) are in rough accord with CMIP model-mean values. This result is consistent with comparisons of observed sea ice extent and CMIP simulations (McAvaney et al., 2001, Table 8.3). However, inter-model standard deviations of sea ice thickness are comparable to the model-mean values, indicating significant disagreements among the models.

2.3. Global statistics

To begin to obtain a more quantitative picture of how well (or how poorly) the models agree with observations, we use a diagram developed by Taylor (submitted for publication). This technique, and others exhibited in this section, are part of the climate diagnostic software developed at the Program for

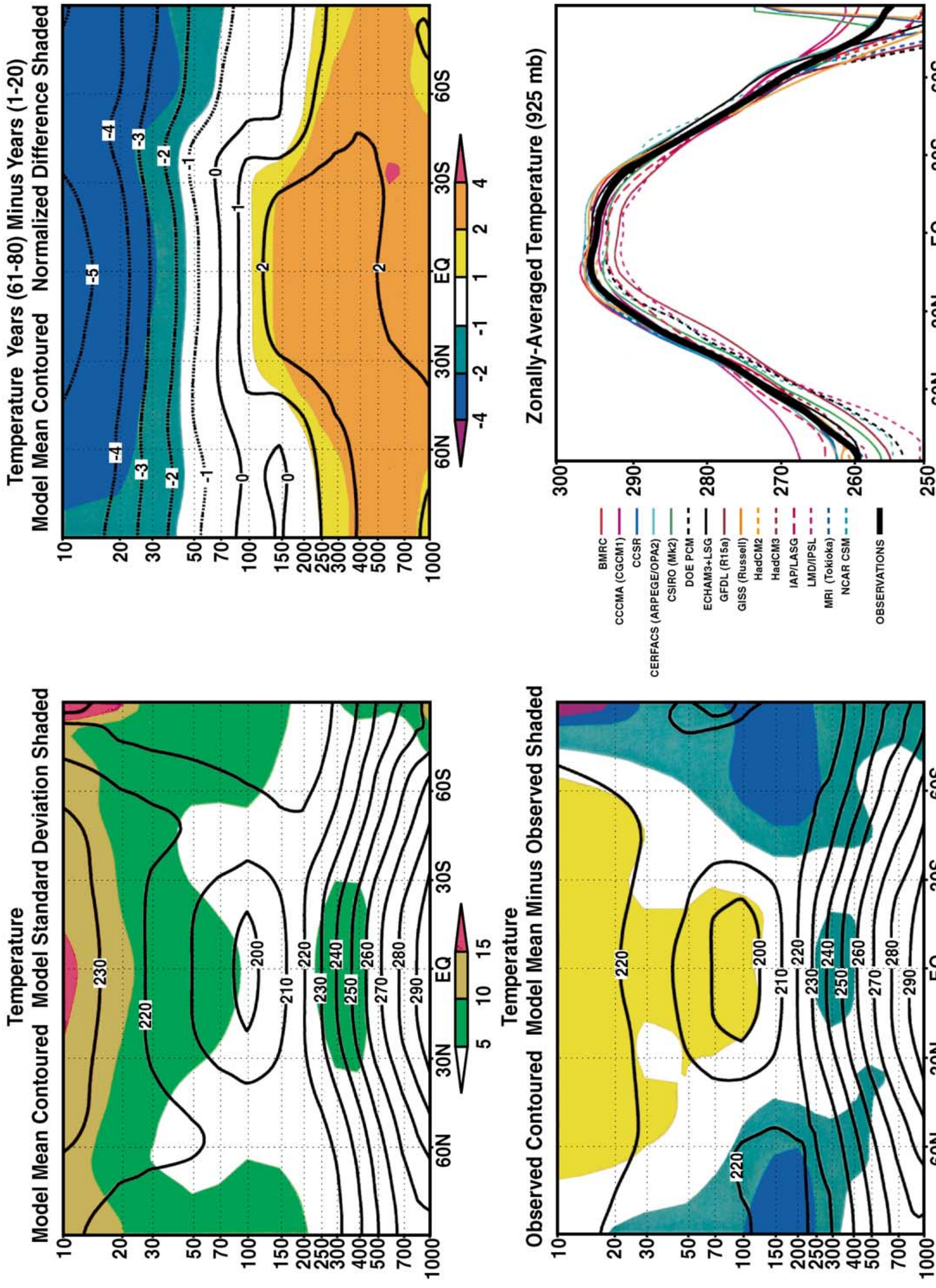


Fig. 5. Same as Fig. 2 for zonally averaged temperature (K).

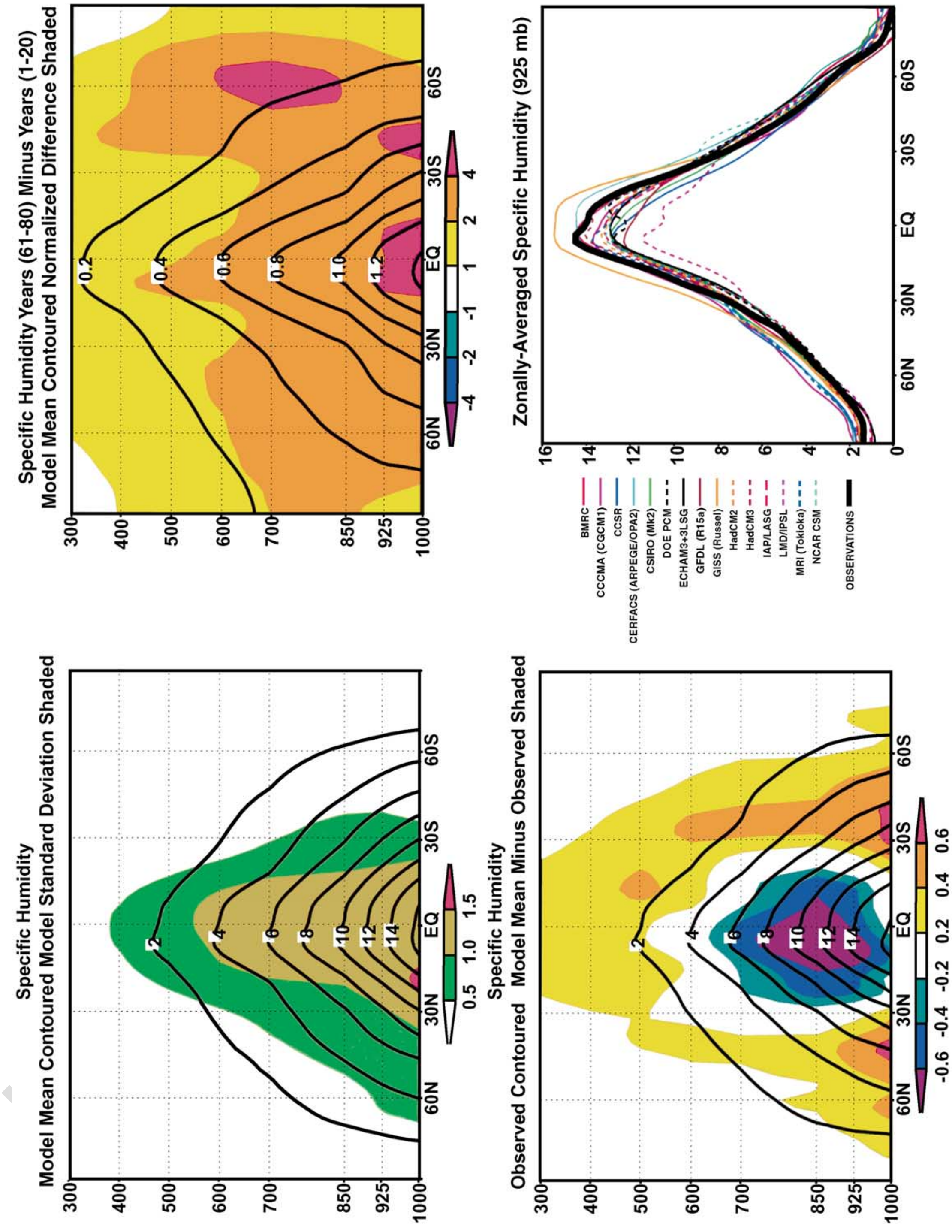


Fig. 6. Same as Fig. 2 for zonally averaged specific humidity (g/kg).

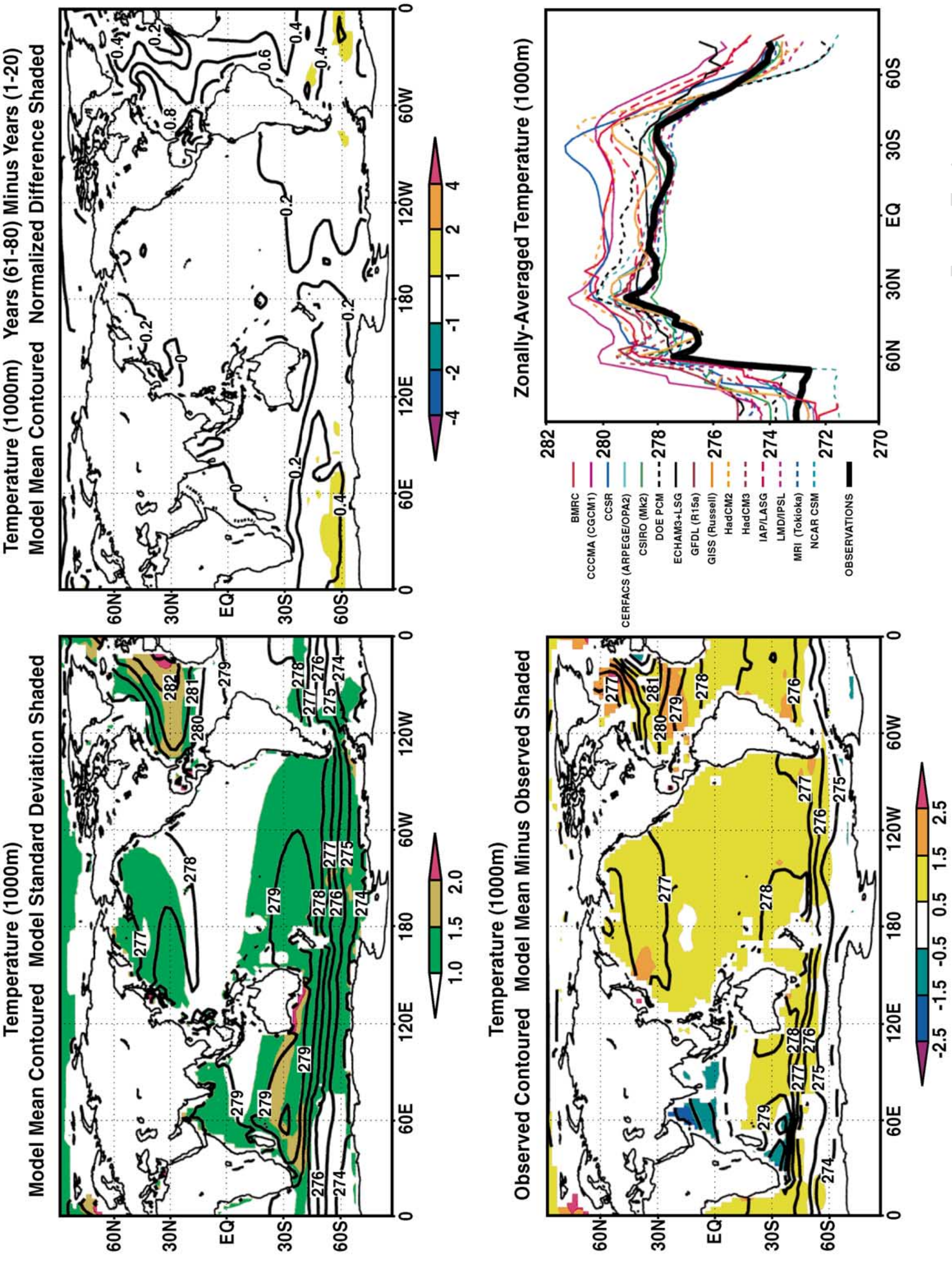


Fig. 7. Same as Fig. 2 for ocean temperature at 1000 m depth (K).

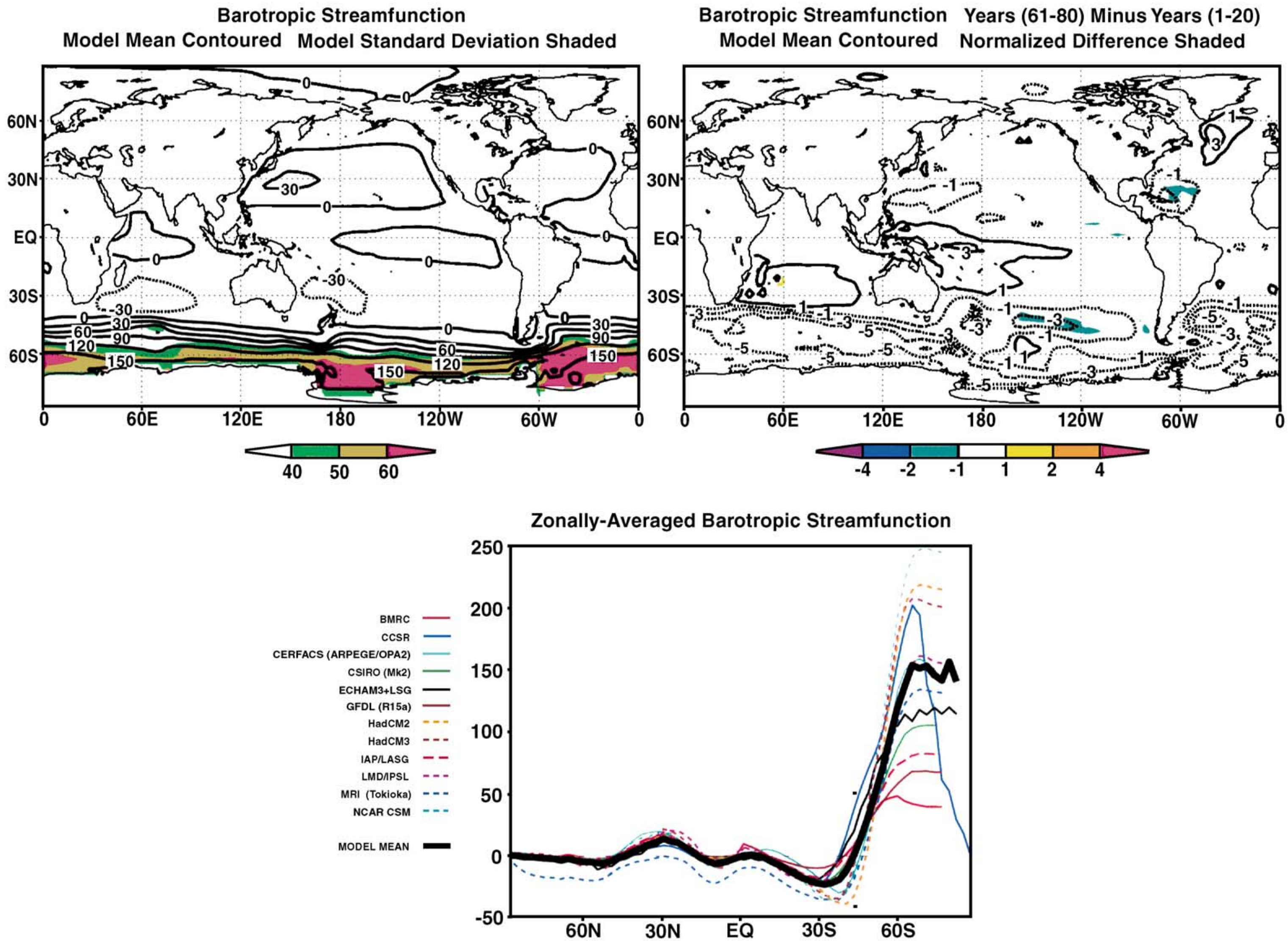


Fig. 8. Summary of long-term time means for the barotropic streamfunction (Sv). The upper-left panel gives the control run 80-year mean averaged over all models (contours) and the intermodel standard deviation (color shading). The bottom panel gives zonal averages for the individual model control runs and the model mean. The upper-right panel gives the average over all models of the difference between the last 20-year mean and the first 20-year mean from the 80-year perturbation simulations, in which atmospheric carbon dioxide increases at a rate of 1% per year (contours), and this difference normalized by the corresponding intermodel standard deviation (color shading).

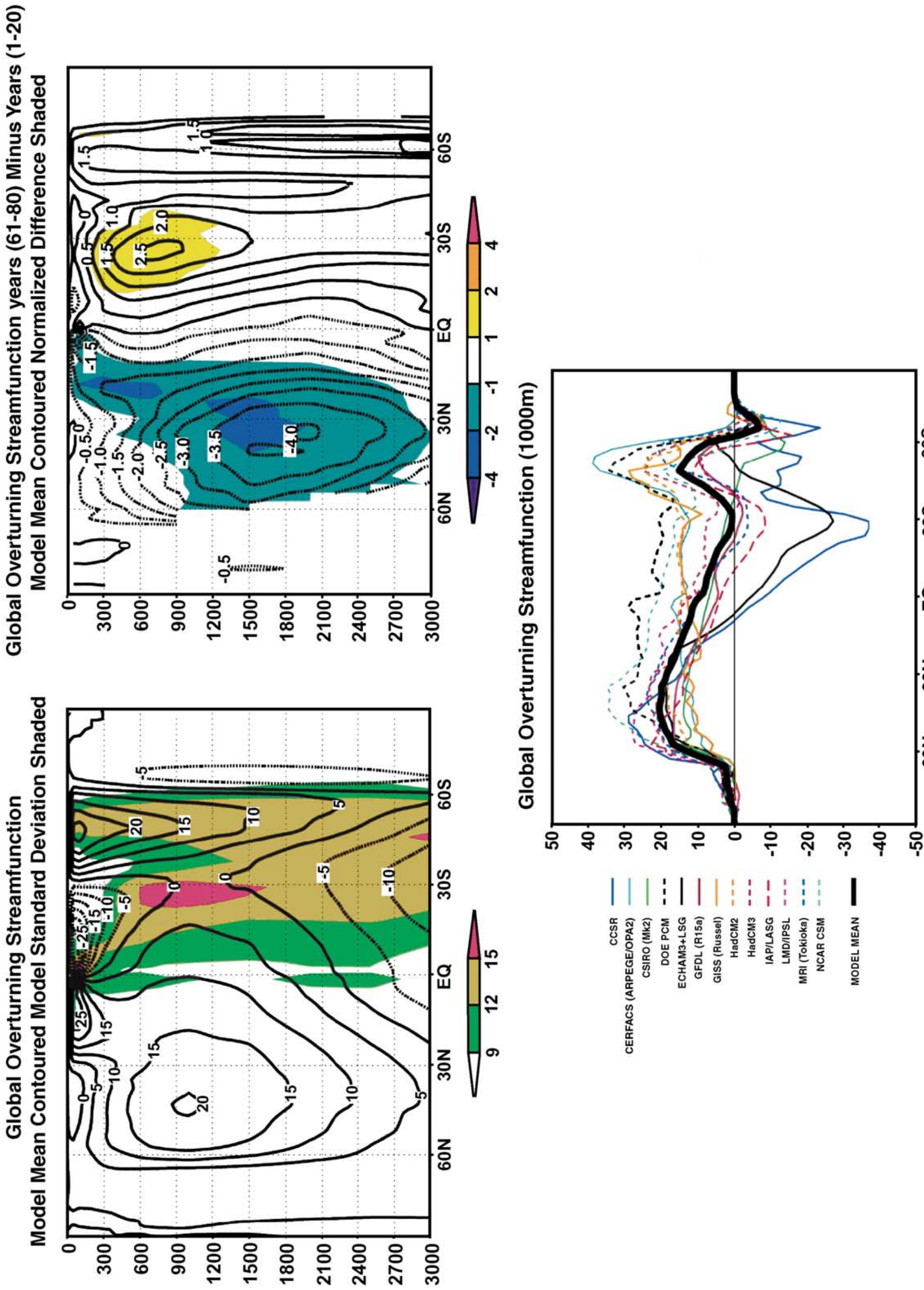
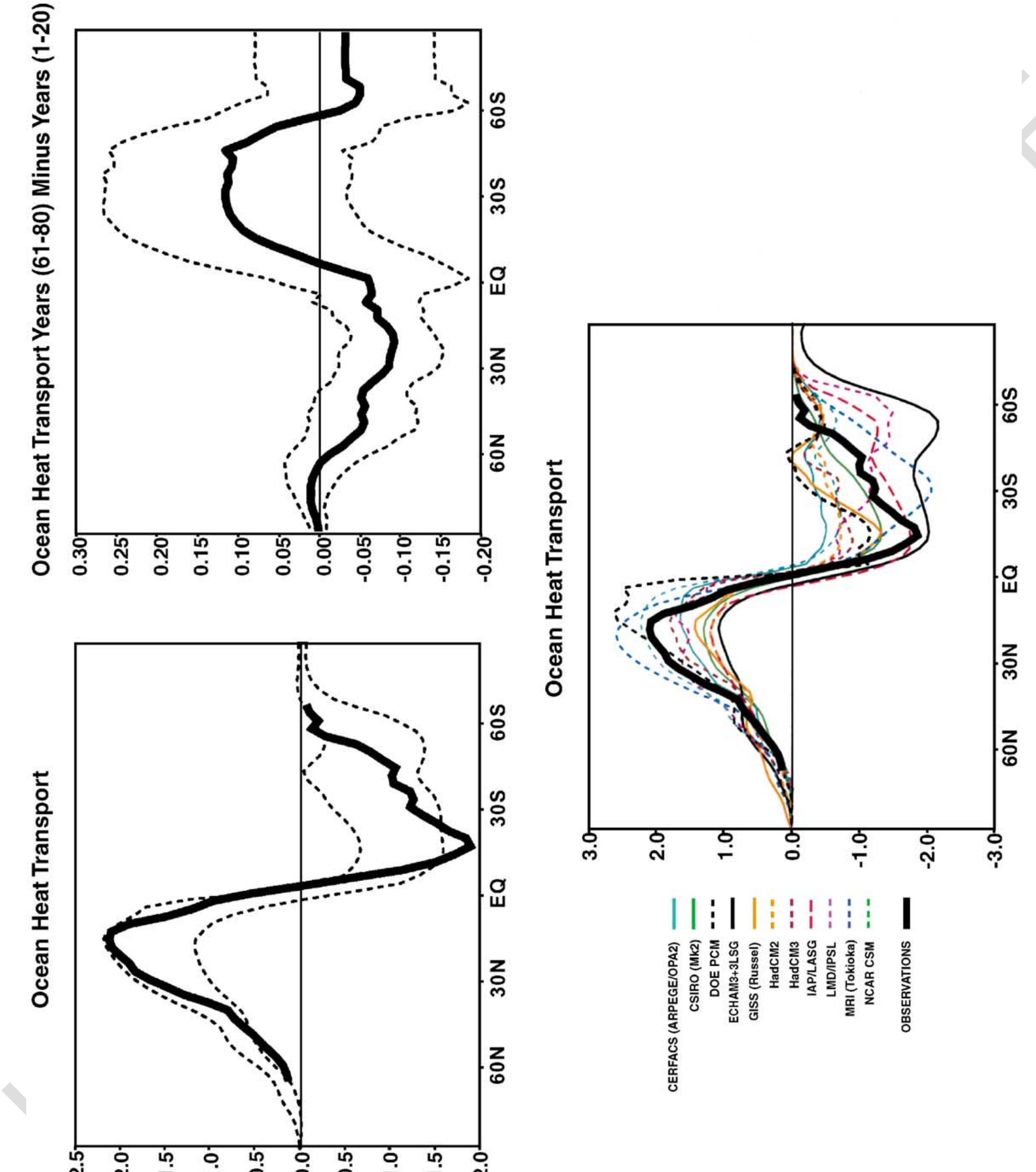


Fig. 9. Same as Fig. 8 for global overturning streamfunction (Sv).


Ocean Heat Transport
Ocean Heat Transport Years (61-80) Minus Years (1-20)
Ocean Heat Transport
60N
30N
EQ
30S
60S
CERFACS (ARPEGE/OPA2)
CSIRO (Mk2)
DOE PCM
ECHAM3+3LSG
GISS (Russel)
HadCM2
HadCM3
IAP/LASG
LMD/IPSL
MRI (Tokioka)
NCAR CSM
OBSERVATIONS

Climate Diagnosis and Intercomparison (PCMDI). Selected PCMDI software tools and their documentation can be downloaded from the Web site http://www-pcmdi.llnl.gov/software. We intend to make the software tools that produces Figs. 12, 14, etc., public via this Web site.

Fig. 12 is a Taylor diagram of the total spatial and temporal variability of three fields: surface air temperature, sea level pressure and precipitation. The variability shown in the figure includes the seasonal cycle but excludes the global mean. The radial coordinate is the ratio of the modeled to observed standard deviation. The cosine of the angle of the model point from the horizontal axis is the spatio-temporal correlation between model and observation. When plotted in these coordinates, the diagram also indicates the root-mean-square difference between model and observation: this differences is proportional to the linear distance between the model point and the "observed" point lying on the horizontal axis at unit distance from the origin. Thus, the diagram enables visualization of three quantities—standard deviation normalized by observation, correlation with observation, and r.m.s. difference from observation—in a two-dimensional space. This is possible because the three quantities are not independent of each other (Taylor, submitted for publication). Loosely speaking, the polar coordinate of the diagram gives the correlation between model and observation for space–time variations but contains no information about the amplitude of the variations, the radial coordinate compares the modeled and observed amplitude of the variations, and the distance between each point and the "observed" point gives the r.m.s. model error.

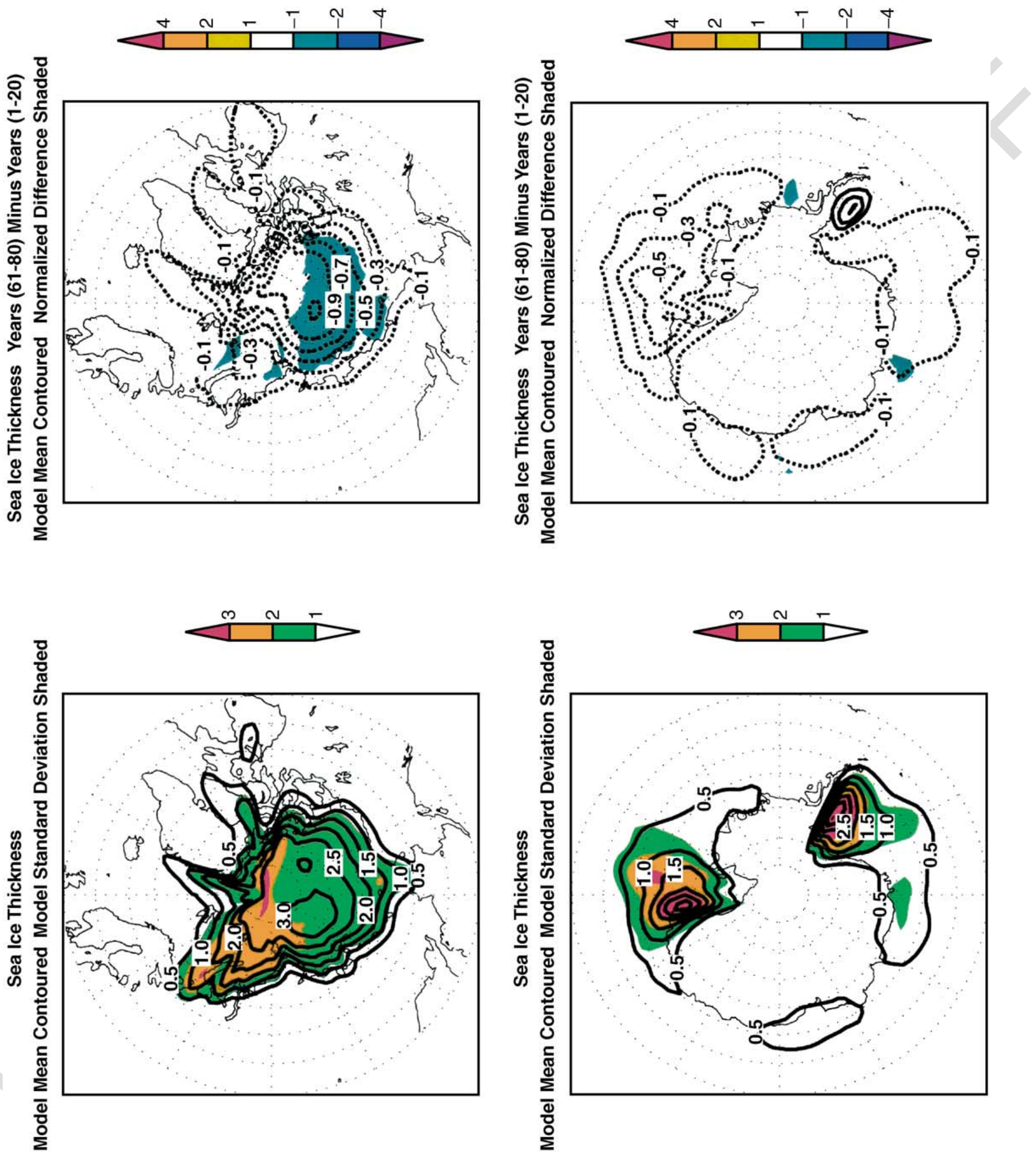
The most striking of the figure is the way it separates the three fields into separate groups. This separation agrees with the familiar qualitative statement that models simulate temperature best, sea level pressure less well, and precipitation worst (e.g., Gates et al., 1996). For surface air temperature, all models achieve a correlation with observation >0.93, and the standard deviation of space–time variations is within ± 15% of the observed value in nearly all models. (This achievement is especially noteworthy for the non-flux-adjusted models, which have no explicit constraints requiring surface temperatures to match observations.) For modeled sea level pressure, the correlation with observation falls mainly in the range 0.7–0.9; for modeled precipitation it falls in the range 0.4–0.7. The standard deviation of space–time variations is also modeled less well for precipitation and sea level pressure than it is for surface air temperature.

To provide a sense of observational uncertainty, we include two alternative observed data sets in Fig. 12: ECMWF/ERA reanalysis ("E") and NCEP reanalysis ("N"). These data sets are plotted as if they were model output. For all three fields, the alternate observed data sets fall closer to the baseline "observed" point than any model does—but not much closer than the closest model. For precipitation and surface air temperature, the r.m.s. difference between either of the reanalysis data sets and the baseline observations is more than half the smallest r.m.s. model error. Whether this result says something positive about the models or negative about reanalysis is unclear. More comparison between alternate sets of observations is provided in the following figures.

Fig. 12 displays the total space–time variance of the model runs. It is also useful to examine individual components of the variance. Fig. 13 shows how we divide a surface field (either model-simulated or observed) into components. Our procedure follows the usual practice space–time behavior:

1. the global and annual mean (not included in Fig. 12),
2. the zonal and annual mean, giving variations with latitude,
3. the annual mean deviations from the zonal mean, giving variations with longitude (mainly land–sea contrast),
4. the annual cycle of the zonal mean, giving seasonal variations as a function of latitude,
5. the annual cycle of deviations from the zonal mean, giving the remaining variance (apart from interannual variations, which are not considered here).

Fig. 10. Summary of long-term time means for northward global ocean heat transport (PW). The upper-left panel gives the observed values as a solid line; the dashed lines are the model mean plus and minus one intermodel standard deviation. The bottom panel gives zonal averages for the individual model control runs and the model mean. The upper-right panel gives the average over all models of the difference between the last 20-year mean and the first 20-year mean from the 80-year perturbation simulations, in which atmospheric carbon dioxide increases at a rate of 1% per year (solid line), and this difference plus and minus one corresponding intermodel standard deviation (dashed lines).


Sea Ice Thickness Years (61-80) Minus Years (1-20)
Model Mean Contoured Normalized Difference Shaded
Sea Ice Thickness Years (61-80) Minus Years (1-20)
Model Mean Contoured Normalized Difference Shaded
Sea Ice Thickness
Model Mean Contoured Model Standard Deviation Shaded
Sea Ice Thickness
Model Mean Contoured Model Standard Deviation Shaded

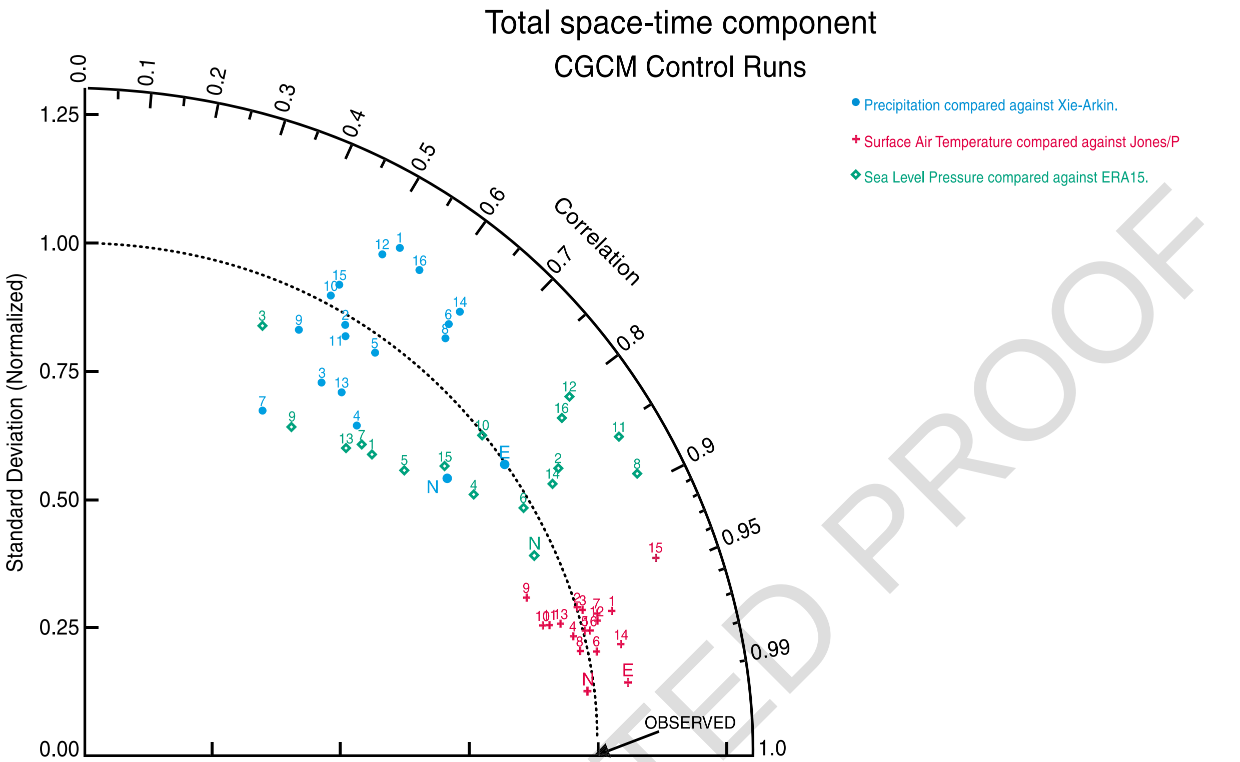


Fig. 12. Error statistics of surface air temperature, sea level pressure and precipitation. The radial coordinate gives the magnitude of total standard deviation, normalized by the observed value, and the angular coordinate gives the correlation with observations. It follows that the distance between the OBSERVED point and any model's point is proportional to the r.m.s. model error (Taylor, submitted for publication). Numbers indicate models counting from left to right in Figs. 14–16. Letters indicate alternate observational data sets compared with the baseline observations: E = 15-year ECMWF/ERA reanalysis ("ERA15"); N = NCEP reanalysis.

In Figs. 14–16, we divide the r.m.s. difference between each model and observation ("total error" of the model) into these components. The error component associated with the global and annual mean is called the bias, and the remaining error (the sum of components 2–5) is called the pattern error. The figures give—from top to bottom—the total error, the bias, the pattern error, and the remaining error components. For each component, errors are normalized by that component's observed standard deviation. The error amounts are color-coded so that blue indicates a small error compared with the observed standard deviation and red indicates a large error compared with the observed standard deviation.

Applying this metric to surface air temperature (Fig. 14), we find that nearly all error components in nearly all models are small, particularly the annual and zonal mean components. For three of the models—ECHAM + OPYC3, HadCM2 and HadCM3—all of the error components are about as small as for ERA and NCEP reanalyses when the latter are included as extra "models". Turning to sea level pressure (Fig. 15), we find that nearly all models have small errors for global and zonal means, but several of the models have large errors

Fig. 11. Summary of long-term time means for sea ice thickness (m), with North polar regions shown in top panels and South polar regions shown in bottom panels. The left-side panels give the control run 80-year mean averaged over all models (contours) and the intermodel standard deviation (color shading). The right-side panels give the average over all models of the difference between the last 20-year mean and the first 20-year mean from the 80-year perturbation simulations, in which atmospheric carbon dioxide increases at a rate of 1% per year (contours), together with this difference normalized by the corresponding intermodel standard deviation (color shading).

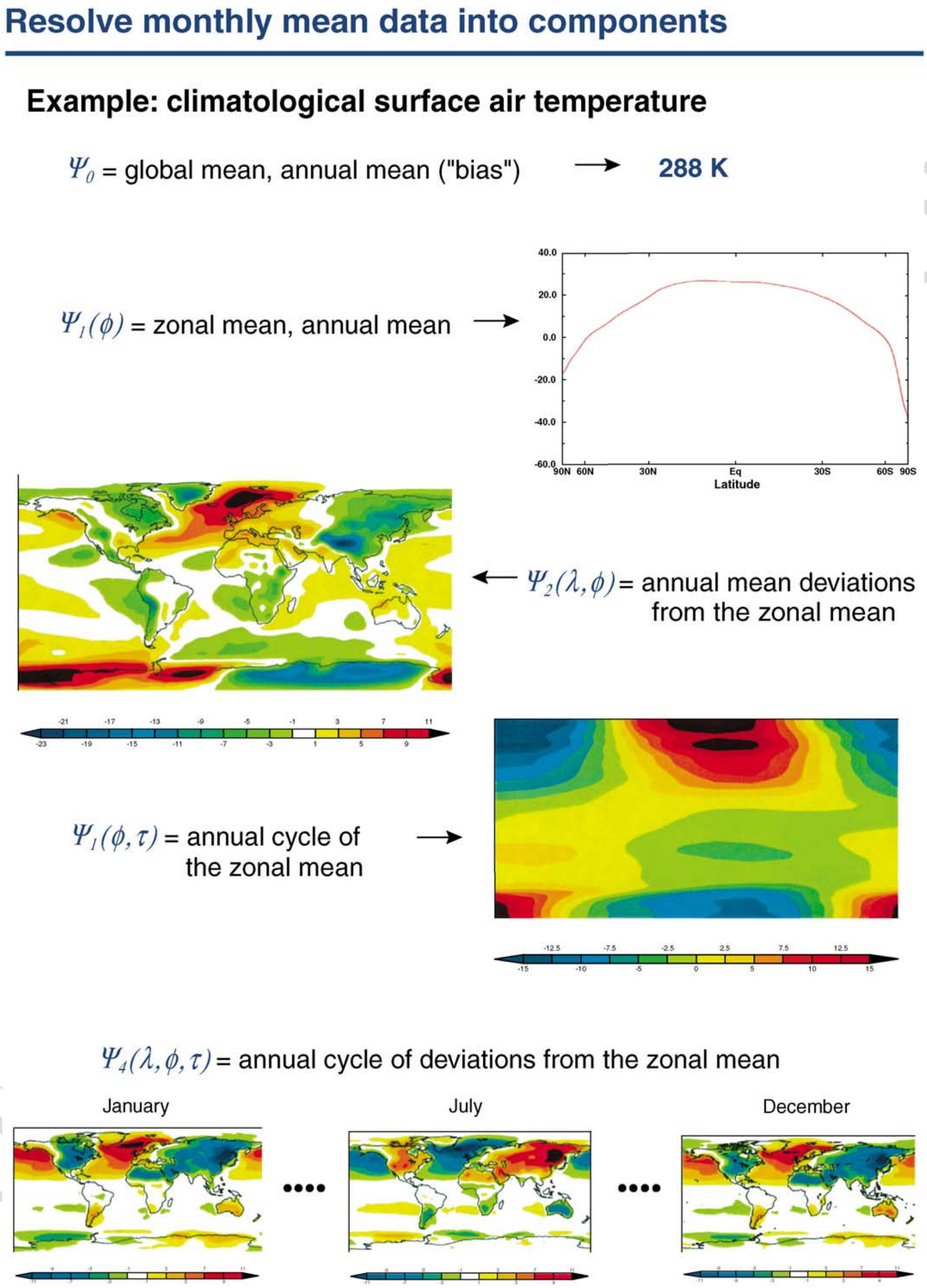


Fig. 13. Example showing division of a model output field into space and time components.

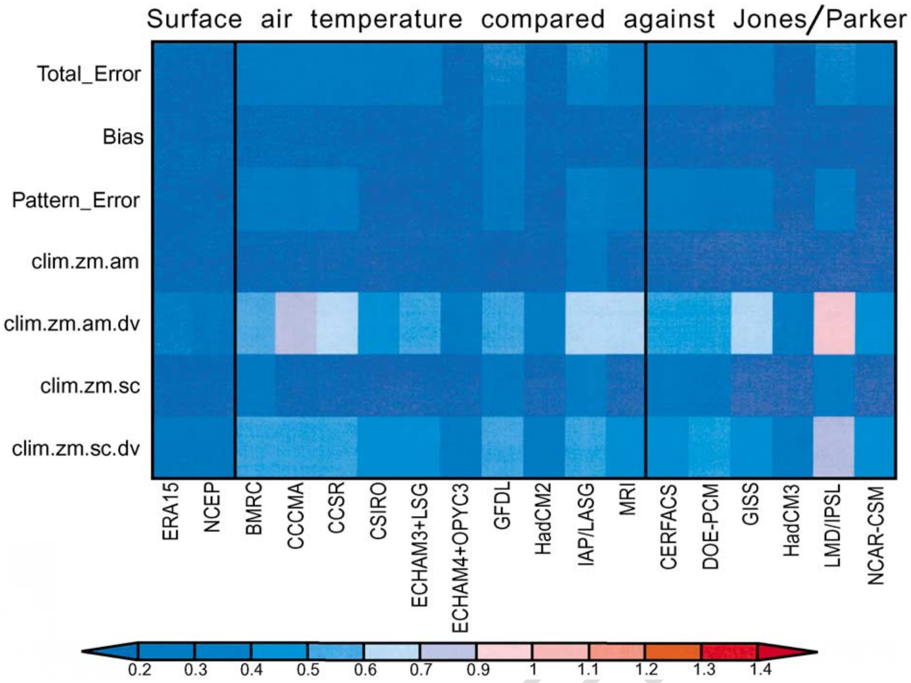


Fig. 14. Components of space–time errors in the climatological annual cycle of surface air temperature. Shown are the total error, the global and annual mean error ("bias"), the total r.m.s. ("pattern") error, and the following components (explained in Fig. 23): zonal and annual mean ("clim.zm.am") annual mean deviations from the zonal mean ("clim.zm.am.dv"), seasonal cycle of the zonal mean ("clim.zm.sc") and seasonal cycle of deviations from the zonal mean ("clim.zm.sc.dv"). For each component, errors are normalized by the component's observed standard deviation. The two left-most columns represent alternate observationally based data sets, ECMWF/ERA and NCEP reanalyses, compared with the baseline observations (Jones et al., 1999). Remaining columns give model results: the 10 models to the left of the second thick vertical line are flux adjusted and the six models to the right are not.

for more detailed space–time patterns. Surprisingly, even the NCEP reanalysis has a large "error" in one component (annual cycle of the zonal mean) when compared with the baseline observations from ERA. Turning to precipitation (Fig. 16), we find that model errors are concentrated in the annual cycle of deviations from the zonal means. Large errors in this component appear for all models except HadCM2 and the two reanalyses. These errors are unrelated to the "double ITCZ" problem discussed above, which would not appear in this component. Errors in the global and zonal means (including the seasonal cycle of the zonal mean) are small for all models. This situation is an improvement over earlier models in which even the global and annual mean precipitation value could be substantially erroneous, e.g., ~30% greater than observed in Version 1 of the NACAR Community Climate Model (Covey and Thompson, 1989, Table 1).

Figs. 14–16 can also be used to sort models into flux-adjusted and non-flux-adjusted classes, as explained in the figure captions. Differences between these two classes of models are not obvious from the figures. This result reinforces the inferences made above that in modern coupled GCMs the performance differences between flux-adjusted and non-flux-adjusted models are relatively small (see also Duffy et al., 2000). Evidently, for at least the century-time-scale integrations used to detect and predict anthropogenic climate change, several modeling groups now find it possible to dispense with flux adjustments. This development represents an improvement over the situation a decade ago, when most groups felt that coupled models could not satisfactorily reproduce the observed climate without including arbitrary (and often nonphysical) adjustment terms in their equations.

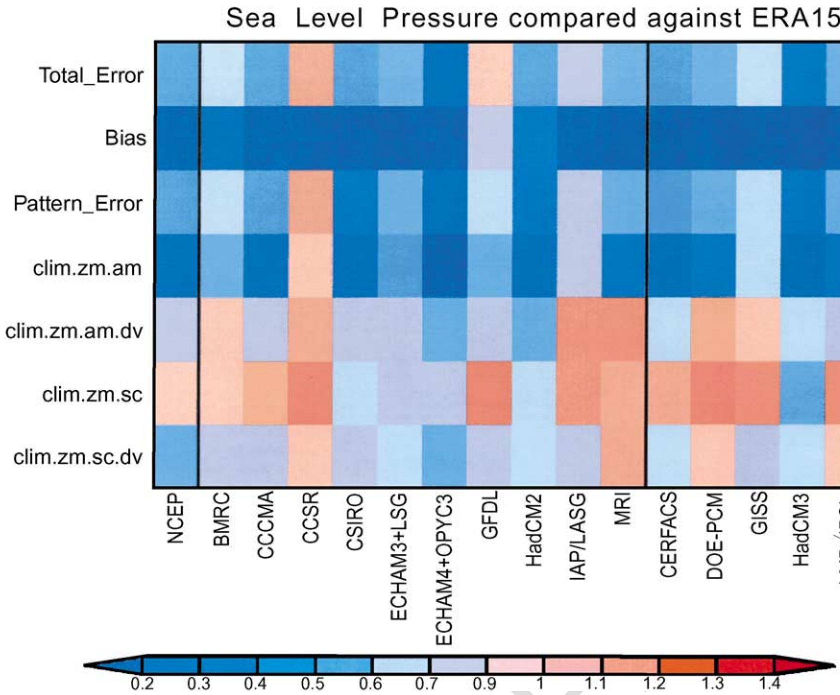


Fig. 15. Same as Fig. 14 for mean sea level pressure. Baseline observations are from ECMWF/ERA reanalysis.

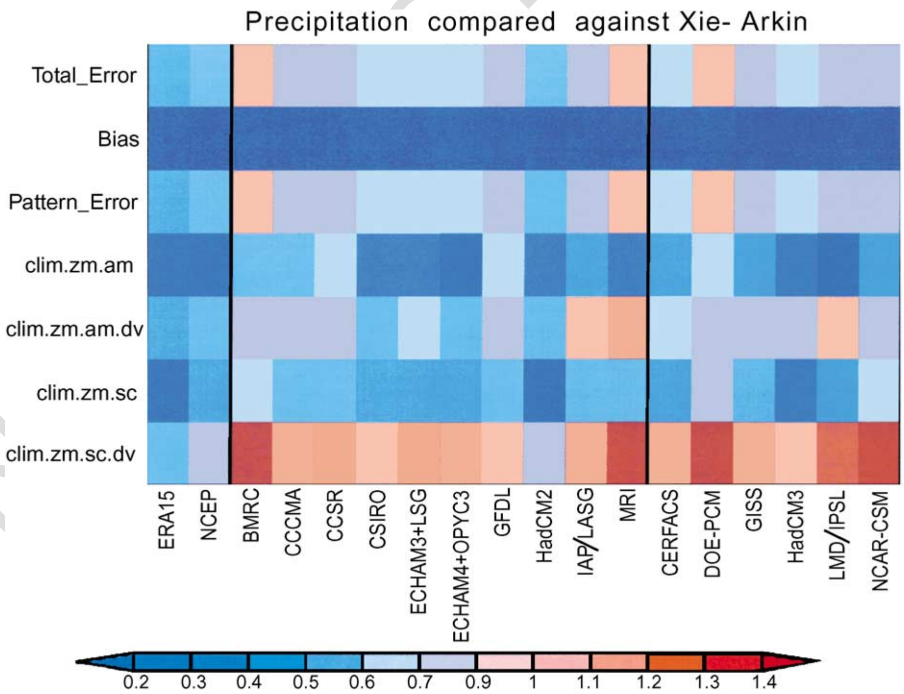


Fig. 16. Same as Fig. 14 for precipitation. Baseline observations are from Xie and Arkin (1997).

2.4. Climate variability

As noted in the Introduction, several detailed studies of climate variability have used the CMIP database. Here we confine discussion to the power spectra of globally or hemispherically averaged annual mean surface air temperature simulated by the CMIP control runs. We use the most complete set of model output available to CMIP and draw a few simple conclusions that were not emphasized in the detailed studies. Fig. 17 shows power spectra of detrended globally and annually average surface air temperature simulated by the 10 longest-running CMIP control runs. For comparison, we also show as "Observed" data the spectra obtained from the instrumental anomaly record of years 1861–1999 (Jones et al., 2001). All time series used for our spectra are available on the World-Wide Web at ftp://sprite.llnl.gov/pub/covey/Data. We detrended all time series before spectral analysis.

Our spectral analysis follows the algorithms described by Jenkins and Watts (1968), calculating the spectra from the autocovariance with lags up to 1/4 the length of each time series and using a Tukey window 1/10 the length of each time series. The same software was used to produce Fig. 8.1 in the IPCC's Second Scientific Assessment Report (Santer et al., 1996), which displayed power spectra from three coupled GCM's and an earlier version of Jones' observational dataset. In the earlier IPCC figure,

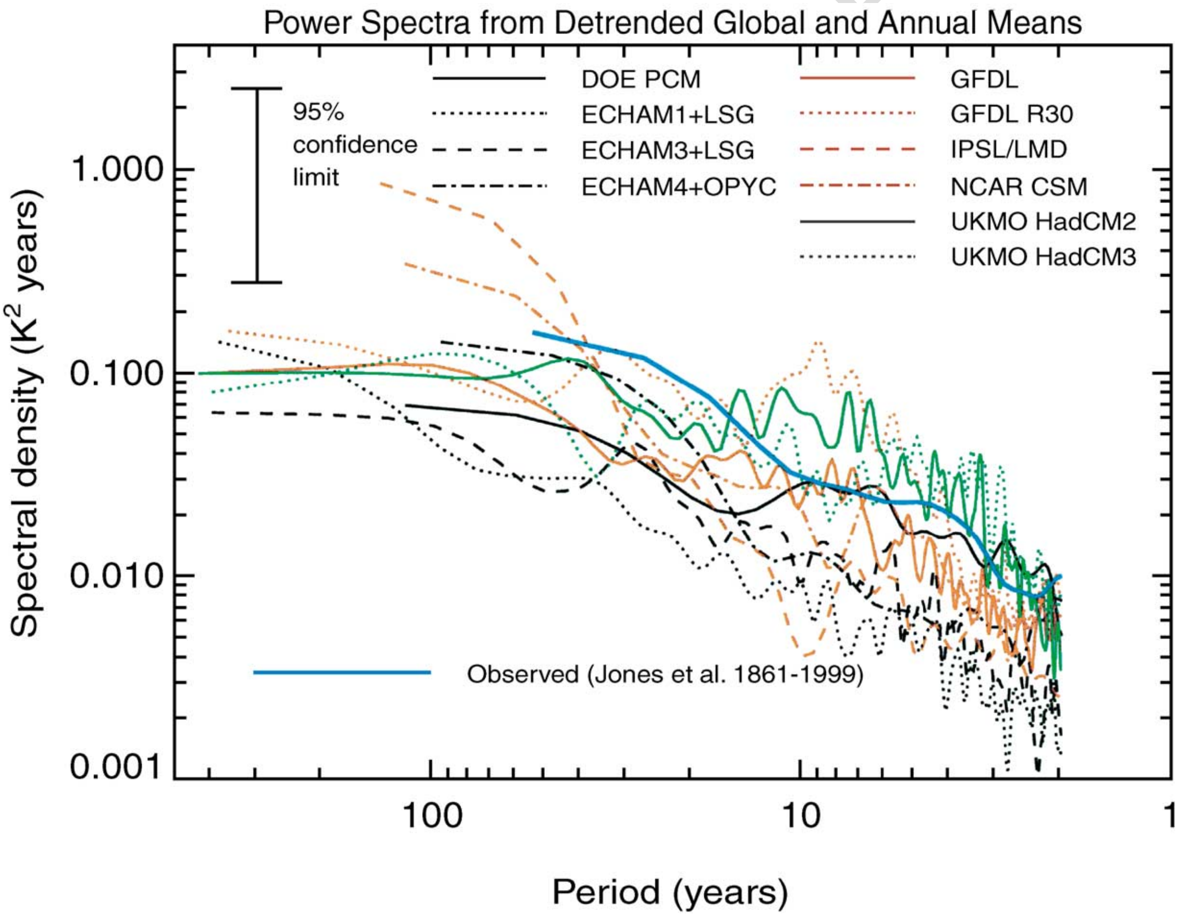


Fig. 17. Power spectra of detrended globally and annually averaged surface air temperature simulated by the 10 longest-running CMIP control runs and as observed by Jones et al. (2001).

however, the spectra were normalized so that the areas under all curves were identical. In our spectra, the areas under the curves (if the curves are plotted on linear scales) equal the total variances about the means of the detrended time series. The 95% confidence interval indicated by the vertical bar is based only on uncertainties due to finite sample size. This confidence interval is the same for all cases because the ratio (maximum lag)/(number of time points) is the same for all cases. Our spectra are quite similar to those shown in Fig. 13 of Stouffer et al. (2000) for a subset of the models considered in the present study, providing reassurance that the results are not sensitive to small changes in the analysis algorithm.

Most of the model-derived spectra fall below the observation-derived spectrum in Fig. 17. The instrumental record, however, may include an "anthropogenic overprint" that would not be included in model control runs. Thus, the instrumental data may overestimate natural variance at multidecadal time scales, because the nonlinear increase in global mean temperature during the 20th Century (temperature rising in the early and late parts of the century with a pause in between) leaves a residual long-term cycle after linear detrending. To address this issue, we present in Fig. 18 the spectra derived from the spectra derived from Northern Hemisphere area averages rather than global averages. This spatial averaging allows us to compare the model results with a proxy-based Northern Hemisphere surface air temperature reconstruction for the years 1000–1850 (Mann et al., 1998, 1999) as well as the instrumental data. The proxy time series actually extends to 1980, but we truncated it at 1850 to avoid an anthropogenic overprint.

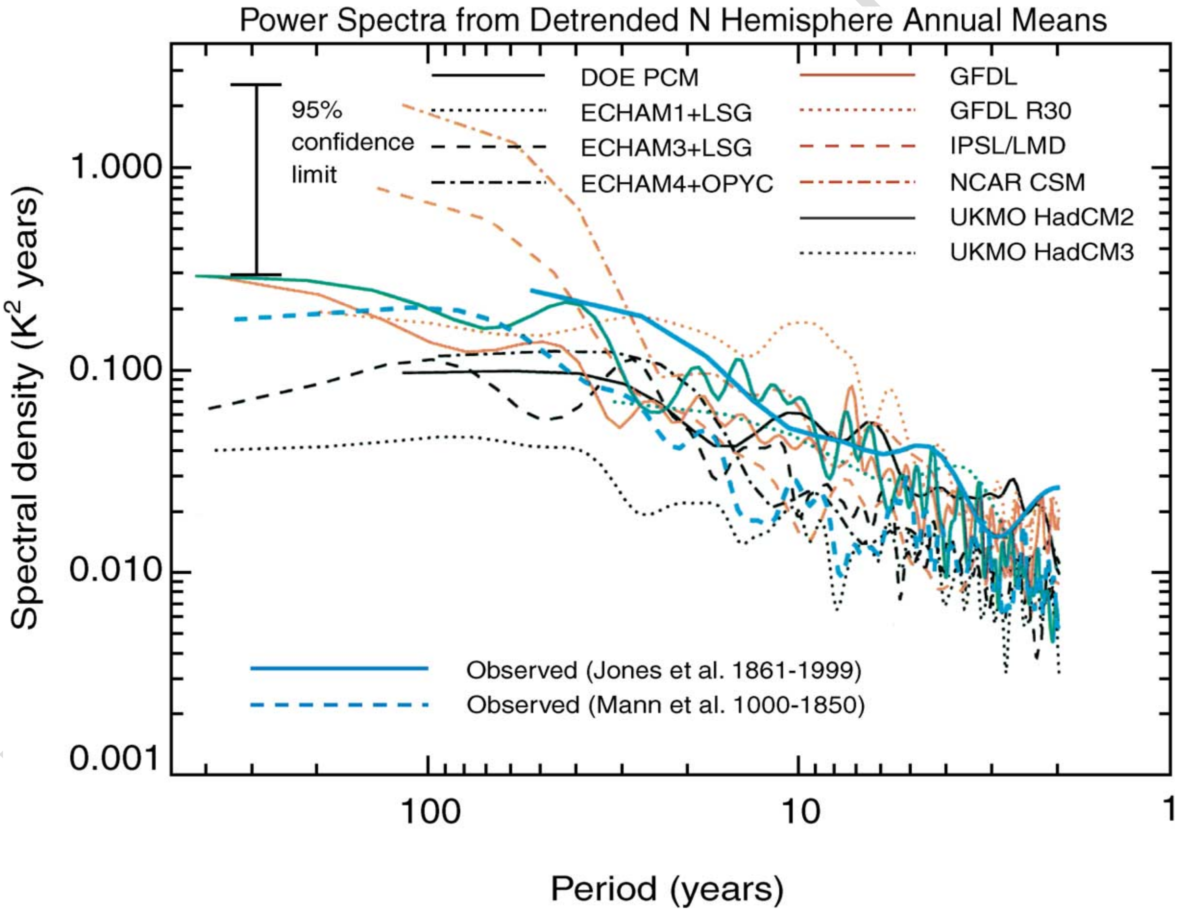


Fig. 18. Same as Fig. 17 for Northern Hemisphere average temperature; additional observed data are from Mann et al. (1999).

In addition to the error bar shown in the figures, a one-sided uncertainty arises in the proxy data from undercalibration of the true variance (as suggested in Fig. 18 by the nearly constant underestimate of the spectrum of the instrumental record by that of the proxy data over the two overlap). From Fig. 2 of Mann et al. (1999), this additional uncertainty may be estimated approximately 36% for periods of 2–50 years and about 100% for periods greater than 50 years. The proxy data, however, includes the combined influences of both naturally forced (e.g., solar and volcanic induced) and internal variability (Mann et al., 1998, Crowley and Kim, 1999; Crowley, 2000), while the CMIP simulations do not include naturally forced variability. The presence of a forced component of variability in the proxy data will thus lead to an overestimate of the spectrum of purely internal variability. Given the relevant estimates (Crowley, 2000), it can be argued that these two effects—undercalibration of true climatic variance and overestimate of the internal component of variability—largely cancel, and that a comparison of the spectrum of the proxy data with that of the CMIP control runs is in fact appropriate.

Incidentally, Fig. 18 shows indirectly that model control runs as well as the 20th Century observational record may contain long transient fluctuations. In the NCAR CSM 300 years run, the Northern Hemisphere mean temperature declines by about 1 °C over the first 150 years and then recovers over the next 50 years. After linear detrending and spectral analysis, this slow variation appears as high spectral power at the longest period for this model (~ 100 years). A similar though less severe effect appears in the IPSL/LMD model output. Of course the low-frequency "tail" of any power spectrum must be interpreted with caution.

In summary, the instrumental and proxy data provide plausible upper and lower limits, respectively, to the real world's natural climate variability, and it is gratifying to note that the CMP spectra generally fall in between these two limits. The assumption that model-simulated variability has realistic amplitudes at inter-annual to interdecadal time scales underlies many of the efforts to detect anthropogenic effects in the observational record, and Fig. 18 provides evidence supporting that assumption (see also Mann, 2000). However, more detailed comparison of the models and the observations—including seasonal as well as annual means—may uncover additional discrepancies (Bell et al., in press). Also, as noted above, one must keep in mind that the real world includes naturally forced climate variations that were not included in the CMIP boundary conditions. In Fig. 19, an example from one

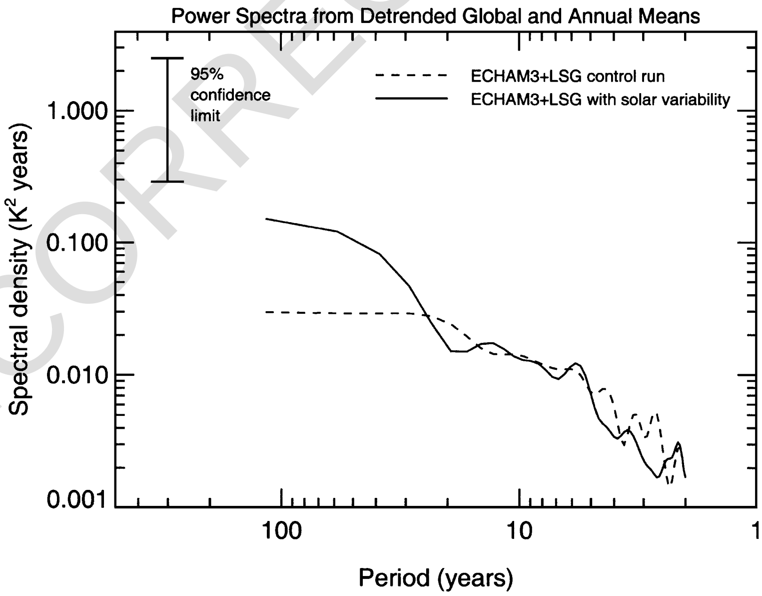


Fig. 19. Same as Fig. 17 for the ECHAM3 + LSG control run and for the same model run with an estimate of historical variations of solar energy output.

model (Experiment 2 from Cubasch et al., 1997) shows that inclusion of solar variations can boost low frequency spectral power by as much as a factor of 5 Similar results have been obtained by the UKMO Hadley Centre and by Crowley (2000).

3. Increasing-CO_2 climate

To begin our discussion of model responses to 1% per year increasing atmospheric CO_2, Fig. 20 shows global and annual mean changes in surface air temperature and precipitation under this scenario, i.e., differences between the increasing-CO_2 and control runs. The surface air temperature results are similar to those shown in the 1995 IPPC report (Kattenberg et al., 1996, Fig. 6.4). The models reach about 2 °C global mean surface warming by the time CO_2 doubles around year 70, and the range of model results stays within roughly $\pm 25\%$ of the average model result throughout the experiments. This rather narrow range contrasts with a greater spread of model output for experiments in which the models are allowed to reach equilibrium. The typical statement for the equilibrium results (from IPPC reports and similar sources) is that the surface warms by 3.0 ± 1.5 °C under doubled

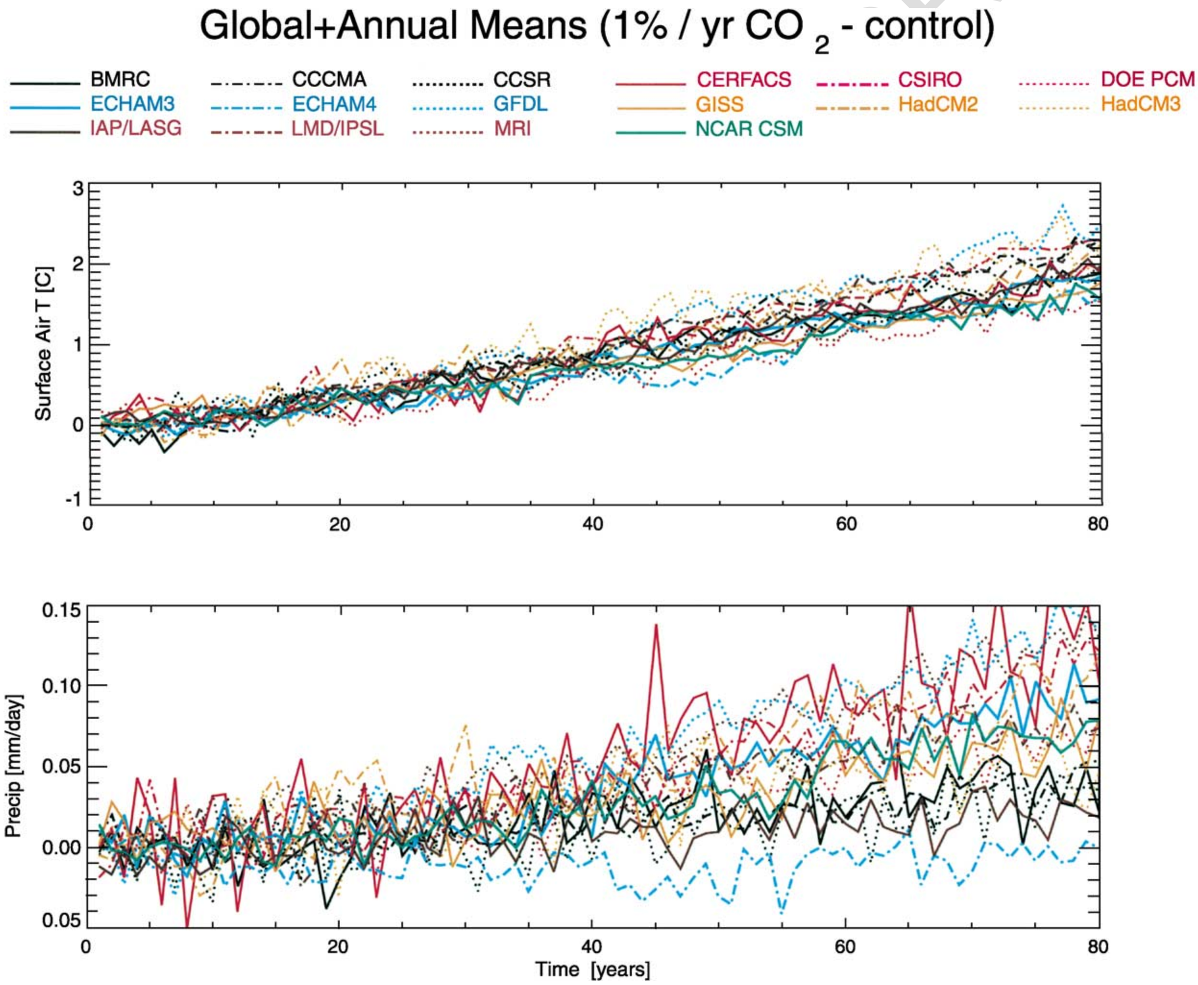


Fig. 20. Globally averaged difference between increasing-CO_2 and control run values of annual mean surface air temperature (top) and precipitation (bottom) for the CMIP2 models. Compare with Fig. 1, which gives control run values.

CO_2. While it is understandable that the ultimate equilibrium warming is greater than the warming at the moment that CO_2 reaches twice its initial value, it may seem surprising that the dispersion of results from different model—a factor of 3 in the equilibrium experiments—is reduced to $\pm$ 25% in the time-evolving (or "transient") experiments considered here.

The precipitation responses of the models span a much wider range than the temperature responses. As shown in Fig. 20, the increase in global and annual mean precipitation at the time of CO_2 doubling varies from essentially zero to $\sim$ 0.2 mm/day. With the exception of the ECHAM4 + OPYC3 model, global means of both surface air temperature and precipitation increase in all of the enhanced-CO_2 simulations; nevertheless the correlation between precipitation increases and temperature increases is weak (as is the correlation between precipitation increases and the control run temperatures shown in the top panel of Fig. 1). This lack of correlation is most obvious in the ECHAM4 + OPYC3 model, for which the global mean temperature increase at 80 years is 1.6 °C while the global mean precipitation increase is less than 0.02 mm/day. The reason for the small precipitation response in this model is the change in cloud radiative forcing in the global warming scenario (E. Roeckner, personal communication). Compared with other models, there is a large increase in the long wave component of cloud forcing, resulting in a positive feedback on the enhanced-CO_2 greenhouse effect, and at the same time a large increase in the short wave component of cloud forcing, resulting in negative feedback via increased reflection of sunlight back to space. These two cloud feedbacks largely cancel in the temperature response, but they act at different locations relevant to the precipitation response. The long wave cloud feedback heats the atmosphere while the short wave cloud feedback cools the surface. The cooler surface has less tendency to evaporate water even though the warmer atmosphere could potentially hold more water vapor; the net result is very little change in global mean evaporation and precipitation.

Turning to geographical and latitude-height distributions, we recall that the upper-right panels of Figs. 2–11 display changes simulated by the perturbation experiments. Contour lines give the model-mean difference between the first 20-year time mean and the last 20-year time mean of the 80-year simulation. This difference is the change over roughly 60 years during which time atmospheric CO_2 nearly doubles. The intermodel standard deviation (sd_m) of these 60-year differences is used to normalize the model mean differences. Absolute values of the normalized difference greater than one are shaded and indicate that the changes simulated by the models have a reasonable degree of consistency and therefore one might have increased confidence in the results.

For surface air temperature (Fig. 2), there is a globally averaged model mean increase of 1.73 °C. The largest changes occur in the polar regions and over land areas. The increases exceed sd_m by a factor of 2 over most of the globe. For mean sea level pressure (Fig. 3), the polar regions and land areas exhibit a decrease and the oceans tend to exhibit an increase, an indicator of monsoon-like circulations developing as a run results of land areas warming faster than ocean areas. The largest values of normalized sea level pressure difference are generally found in polar areas. Changes in precipitation (Fig. 4) show an increase over most of the globe. The globally averaged model mean increase is 0.07 mm/day. Only a few areas—generally in the sub-tropics—exhibit a decrease. The largest values of normalized difference occur in high mid-latitudes and probably have an association with storm tracks. Changes in net heat flux (not shown) are generally positive, showing a gain of heat by the oceans; the mean model change is generally less than sd_m, indicating that although the models all transport heat into the oceans in global warming scenarios, the locations at which they do so vary. The models also simulate changes in net fresh water flux (not shown) that are similar in sign to the control run results, indicating that dry areas will become drier and wet areas wetter. Changes in model mean zonally averaged temperature as a function of height (Fig. 5) show the expected pattern of warming in the troposphere and lower stratosphere and cooling in the remainder of the stratosphere. Changes in large areas of the troposphere and the stratosphere are more than twice sd_m. Model mean zonally averaged specific humidity (Fig. 6) increases everywhere and its changes are also large compared with sd_m consistent with the temperature changes.

Changes in model mean ocean temperature at 1000 m depth (Fig. 7) are generally small. The models do produce consistent simulations of slightly increased temperature (and salinity, not shown) off the coast of Antarctica. The model mean barotropic streamfunction (Fig. 8) decreases off Antarctica, indicating a slower Antarctic Circumpolar Current. As a result of the large scatter among models, however, the normalized differences are generally small. Model mean global overturning streamfunction (Fig. 9) decreases in magnitude, with a reasonable degree of agreement among the models. Results for ocean heat transport (Fig. 10) are displayed differently: the solid line represents the model mean difference and the dashed are one sd_m above and below the model mean. The enhanced greenhouse effect acts to reduce the ocean heat transport, consistent with the general slowdown in ocean circulation depicted in Figs. 8–10. Model-mean changes in sea ice thickness (Fig. 11) indicate thinning at essentially all locations. Only in portions of the Arctic, however, is the magnitude of the normalized difference greater than 1; elsewhere there is significant disagreement among the models.

4. Conclusions

Comparison of the CMIP2 control run output with observation of the present-day climate reveals improvements in coupled model performance since the IPCC's mid-1990s assessment (Gates et al., 1996). The most prominent of these is a diminishing need for arbitrary flux adjustments at the air–sea interface. About half of the newer generation of coupled models omit flux adjustments, yet the rates of "climate drift" they exhibit (Fig. 1) are within the bounds required for useful model simulations on time scales of a century or more. The flux-adjusted models exhibit less drift on average, however, and thus agree better with the limited information we possess on climate variations before the Industrial Revolution (e.g., Jones et al., 1998; Mann et al., 1999). Both flux-adjusted and non-flux-adjusted models produce a surprising variety of time-averaged global mean temperatures, from less than 12 °C to over 16 °C. Perhaps this quantity has not been the subject of as much attention as it deserves in model development and evaluation.

The spatial patterns of model control run output variables display numerous areas of agreement and disagreement with observations (Figs. 2–11). As always, it is difficult to determine whether or not the models are "good enough" to be trusted when used to study climate in the distant past or to make predictions of the future. The global statistics shown in Figs. 12–16 provide some encouragement. They indicate that the difference between a typical model simulation and a baseline set of observation is not much greater than the difference between sets of observation. To the extent that different sets of observations (including model-based reanalyses) are equally reliable, this result implies that coupled GCM control runs are nearly as accurate as observational uncertainty allows them to be—at least for the quantities highlighted by our global statistics.

The CMIP2 models do not yield the same simulation of climate change when they are all subjected to an identical scenario of 1% per year increasing CO_2. The range of model-simulated global mean warming, however, is less than the factor of 3 (1.5–4.5 °C) uncertainty commonly cited for equilibrium warming under doubled CO_2. Part of the explanation could involve the behavior of models not included in this report, which may give more extreme results than the CMIP2 models. An additional reason for the narrower range, however, is that the response time of the climate system increases with increasing climate sensitivity (Hansen et al., 1984, 1985; Wigley and Schlesinger, 1985). This introduces a partial cancellation of effects: models with larger sensitivity (greater equilibrium warming to doubled CO_2) are farther from equilibrium than less-sensitive models at any given time during the increasing-CO_2 scenario. Also, the CMIP2 models with larger equilibrium sensitivities have a greater efficiency of ocean heat uptake under increasing CO_2 than the models with smaller equilibrium sensitivities (Raper et al., submitted for publication). The enhanced ocean heat uptake further delays surface warming. Considering the narrowed range of surface temperature responses among the CMIP2 models, one might speculate that the uncertainty in model predictions of climate response to a given forcing is less than the uncertainty in future anthropogenic forcing itself (Hansen et al., 1997). On the other hand, simulated precipitation increases differ greatly among the CMIP2 models and appear to have no simple relationship with simulated temperatures.

Expansion of the CMIP model output set has begun under auspices of the JSC/CLIVAR Working Group on Coupled Models, and analysis of the existing database is continuing. (See the Web page http://www-pcmdi.llnl.gov.cmip/cmip2plusann.html for the most recent additions to the database.) We encourage all interested scientists to contribute to this ongoing effort.

Acknowledgements

We thank Benjamin D. Santer for providing spectral analysis software and for many helpful discussions, Clyde Dease and Anna McCravy of the PCMDI computations staff for assistance with data processing and Web publication, respectively, and of course the modelers whose contributions have made CMIP possible. CC also thanks his fellow IPCC Lead Authors for extensive discussions of climate model evaluation. This work was performed under the auspices of the U.S. Department of energy by University of California Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.

References

Barnett, T.P., 1999. Comparison of near-surface air temperature variability in 11 coupled global climate models. J. Climate 12, 511–518.

Barthelet, P., Bony, S., Braconnot, P., Braun, A., Cariolle, D., Cohen-Solal, E., Dufresne, J.-L., Delecluse, P., Deque, M., Fairhead, L., Filiberti, M.-A., Forichon, M., Grandpeix, J.-Y., Guilyardi, E., Houssais, M.-N., Imbard, M., LeTreut, H., Levy, C., Li, Z.X., Madec, G., Marquet, P., Marti, O., Planton, S., Terray, L., Thual, O., Valcke, S., 1998a. Simulations couplees globales de changements climatiques associes a une augmentation de la teneur atmospherique en CO2. C. R. Acad. Sci. Paris, Sci. Terre Planetes 326, 677–684 (in French with English summary).

Barthelet, P., Terray, L., Valcke, S., 1998b. Transient CO_2 experiment using the ARPEGE/OPAICE nonflux corrected coupled model. Geophys. Res. Lett. 25, 2277–2280.

Bell, J., Duffy, P.B., Covey, C., Sloan, L., and CMIP investigators, 2000a. Comparison of temperature variability in observations and sixteen climate model simulations. Geophys. Res. Lett. 27, 261–264.

Bell, J.F., Sloan, L.C., Covey, C., Duffy, P.B., Revenaugh, J., 2000b. Comparison of natural climate variability in multiple temperature reconstructions and global climate models. Eos (in press, abstract).

Boer, G.J., Flato, G., Ramsden, D., 2000. A transient climate change simulation with greenhouse gas and aerosol forcing: projected climate to the twenty-first century. Clim. Dyn. 16, 427–450.

Boville, B.A., Gent, P.R., 1998. The NCAR climate syatem model, version one. J. Climate 11, 1115–1130.

Cess, R.D., et al., 1989. Interpretation of cloud–climate feedback as produced by 14 atmospheric general circulation models. Science 245, 513–516.

Covey, C., Thompson, S.L., 1989. Testing the effects of ocean heat transport on climate. Glob. Planet. Change 75, 331–341.

Covey, C., et al., 2000. The seasonal cycle in coupled ocean–atmosphere general circulation models. Clim. Dyn. 16, 775–787.

Crowley, T.J., 2000. Causes of climate change over the past 1000 years. Science 289, 270–277.

Crowley, T.J., Kim, K.Y., 1999. Modeling the temperature response the forced climate change over the last six centuries. Geophys. Res. Lett. 26, 1901–1904.

Cubasch, U., Hasselmann, K., Höck, H., Maier-Reimer, E., Mikolajewicz, U., Santer, B.D., Sausen, R., 1992. Time-dependent greenhouse warning computations with a coupled ocean–atmosphere model. Clim. Dyn. 8, 55–69.

Cubasch, U., Voss, R., Hergel, G.C., Waszkewitz, J., Crowley, T.J., 1997. Simulation of the influence of solar radiation variations on the global climate with an ocean–atmosphere general circulation model. Clim. Dyn. 13, 757–767.

Delworth, T.L., Knutson, T.R., 2000. Simulation of early 20th century global warming. Science 287, 2246–2250.

Duffy, P.B., Bell, J., Covey, C., Sloan, L., and CMIP investigators, 2000. Effect of flux adjustments on temperature variability in climate models. Geophys. Res. Lett. 27, 763–766.

Dufresne, J.-L., Friedlingstein, P., Berthelot, M., Bopp, L., Ciais, P., Fairhead, L., Le Treut, H., Monfray, P., 2001. Direct and indirect effects of future climate change on land and ocean carbon uptake. Science (submitted for publication).

Emori, S., Nozawa, T., Abe-Ouchi, A., Numaguti, A., Kimoto, M., Nakajima, T., 1999. Coupled ocean–atmosphere model experiments of future climate change with an explicit representation of sulfate aerosol scattering. J. Meteorol. Soc. Jpn. 77, 1299–1307.

Flato, G.M., Boer, G.J., 2001. Warming asymmetry in climate change simulations. Geophys. Res. Lett. (in press).

Flato, G.M., Boer, G.J., Lee, W.G., MacFarlane, N.A., Ramsden, D., Reader, M.C., Weaver, A.J., 2000. The Canadian Centre for Climate Modelling and Analysis global coupled model and its climate. Clim. Dyn. 16, 451–467.

Gates, W.L., et al., 1996. Climate models—evaluation. In: Houghton, J.T., et al. (Eds.), Climate Climate 1995: the Science of Climate Change. Cambridge Univ. Press, pp. 229–284.

Gates, W.L., et al., 1999. An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). Bull. Am. Meteorol. Soc. 80, 29–55.

Gibson, J.K., Kallberg, P., Uppala, S., Hernandez, A., Nomura, A., Serrano, E., 1997. ERA Description, ECMWF Reanalysis Project Report Series No. 1. European Centre for Medium-range Weather Forecasts, Reading, UK. 66 pp.

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

Gordon, C., Cooper, C., Senior, C.A., Banks, H.T., Gregory, J.M., Johns, T.C., Mitchell, J.F.B., Wood, R.A., 2000. The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. Clim. Dyn. 16, 147–168.

Hansen, J., Lacis, A., Rind, D., Russell, G., Stone, P., Fung, I., Ruedy, R., Lerner, J., 1984. Climate sensitivity: analysis of feedback mechanisms. In: Hansen, J.E., Takahashi, T. (Eds.), Climate Processes and Climate Sensitivity. Geophysical Monograph Series, vol. 29. American Geophysical Union, Washington, DC, pp. 130–163.

Hansen, J., Russell, G., Lacis, A., Fung, I., Rind, D., Stone, P., 1985. Climate response times: dependence on climate sensitivity and ocean mixing. Science 229, 857–859.

Hansen, J., Sato, M., Lacis, A., Ruedy, R., 1997. The missing climate forcing. Philos. Trans. R. Soc. Lond., B 352, 231–240.

Jenkins, G.M., Watts, D.G., 1968. Spectral Analysis and its Applications. Holden-Day, pp. 310–311.

Johns, T.C., 1996. A Description of the Second Hadley Center Coupled Model (HadCM2). Climate Research Technical Note, vol. 71. Hadley Center, United Kingdom Meteorological Office, Bracknell Berkshire RG12 2SY, United Kingdom. 19 pp.

Johns, T.C., Carnell, R.E., Crossley, J.F., Gregory, J.M., Mitchell, J.F.B., Senior, C.A., Tett, S.F.B., Wood, R.A., 1997. The second Hadley Centre coupled ocean–atmosphere GCM: model description, spinup and validation. Clim. Dyn. 13, 103–134.

Jones, P.D., Briffa, K.R., Barnett, T.P., Tett, S.F.B., 1998. High-resolution palaeoclimatic records for the last millenium: interpretation, integration and comparison with general circulation model control-run temperatures. Holocene 8, 455–471.

Jones, P.D., New, M., Parker, D.E., Martin, S., Rigor, I.G., 1999. Surface air temperature and its changes over the past 150 years. Rev. Geophys. 37, 173–199.

Jones, P.D., Osborn, T.J., Briffa, K.R., Folland, C.K., Horton, E.B., Alexander, L.V., Parker, D.E., Rayner, N.A., 2001. Adjusting for sampling density in grid box land and ocean surface temperature time series. J. Geophys. Res. 106, 3371–3380.

Kalnay, E., et al., 1996. The NCEP/NCAR 40-year reanalysis project. Bull. Am. Meteorol. Soc. 77, 437–471.

Kattenberg, A., et al., 1996. Climate models—projections of future climate. In: Houghton, J.T., et al. (Eds.), Climate Climate 1995: the Science of Climate Change. Cambridge Univ. Press, pp. 285–357.

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

Laurent, C., LeTreut, H., Li, Z.X., Fairhead, L., Dufresne, J.L., 1998. The influence of resolution in simulating inter-annual and inter-decadal variability in a coupled ocean–atmostphere GCM with emphasis over the North Atlantic, IPSL Report N8.

Leclainche, Y., Braconnot, P., Marti, O., Joussaume, S., Dufresne, J.L., Filiberti, M.A., 2001. The role of the sea ice thermodynamics in the Norther Hemisphere climate as simulated by a global coupled ocean–atmosphere model. J. Climate (submitted for publication).

Levitus, S., Boyer, T.P., 1994. World Ocean Atlas 1994 Volume 4: Temperature. NOAA Atlas NESDIS, vol. 4. 117 pp.

Manabe, S., Stouffer, R.J., 1996. Low-frequency variability of surface air temperature in a 1000-year integration of a coupled atmosphere–ocean–land surface model. J. Climate 9, 376–393.

Manabe, S., Stouffer, R.J., Spelman, M.J., Bryan, K., 1991. Transient responses of a coupled ocean–atmosphere model to gradual changes of atmospheric CO_2: Part I. Annual mean response. J. Climate 4, 785–818.

Mann, M.E., 2000. Lessons for a new millennium. Science 289, 253–254.

Mann, M.E., Bradley, R.S., Hughes, M.K., 1998. Global-scale temperature patterns and climate forcing over the past six centuries. Nature 392, 779–787.

Mann, M.E., Bradley, R.S., Hughes, M.K., 1999. Northern Hemisphere temperature during the past millenium: interference, uncertainties, and limitations. Geophys. Res. Lett. 26, 759–762.

McAvaney, B.J., et al., 2001. Model evaluation. In: Houghton, J.T., et al. (Eds.), Climate Change 2001: the Scientific Basis. Cambridge Univ. Press, pp. 471–521.

Meehl, G.A., Boer, G.J., Covey, C., Latif, M., Stouffer, R.J., 2000. The Coupled Model Intercomparison Project (CMIP). Bull. Am. Meteorol. Soc. 81, 313–318.

Mitchell, J.F.B., Senior, C.A., Ingram, W.J., 1989. CO_2 and climate: a missing feedback? Nature 341, 132–134.

Power, S.B., Tseitkin, F., Colman, R.A., Sulaiman, A., 1998. A Coupled General Circulation Model for Seasonal Prediction and Climate Change Research, BMRC Research Report No. 66. Bureau of Meteorology, Australia.

Raper, S.C.B., Gregory, J.M., Stouffer, R.J., 2001. The role of climate sensitivity and ocean heat uptake on AOGCM transient temperature and thermal expansion response. J. Climate (submitted for publication).

Roeckner, E., Arpe, K., Bengtsson, L., Christoph, M., Claussen, M., Dümenil, L., Esch, M., Giorgetta, M., Schlese, U., Schulzweida, U., 1996a. The Atmospheric General Circulation Model ECHAM4: Model Description and Simulation of Present-Day Climate, MPI Report No. 218. Max-Planck-Institut für Meteorologie, Hamburg, Germany. 90 pp.

Roeckner, E., Oberhuber, J.M., Bacher, A., Christoph, M., Kirchner, I., 1996b. ENSO variability and atmospheric response in a global coupled atmospheric–ocean GCM. Clim. Dyn. 12, 737–754.

Rothrock, D.A., Yu, Y., Maykut, G.A., 1999. Thinning of the Arctic sea-ice cover. Geophys. Res. Lett. 26, 3469–3472.

Russell, G.L., Rind, D., 1999. Response to CO_2 transient increase in the GISS coupled model: regional coolings in a warming climate. J. Climate 12, 531–539.

Russell, G.L., Miller, J.R., Rind, D., 1995. A coupled atmosphere–ocean model for transient climate change studies. Atmos.-Ocean 33, 683–730.

Santer, B.D., et al., 1996. Detection of climate change and attribution of causes. In: Houghton, J.T., et al. (Eds.), Climate Climate 1995: the Science of Climate Change. Cambridge Univ. Press, pp. 411–443.

Santer, B.D., et al., 2000. Interpreting differential temperature trends at the surface and in the lower troposphere. Science 287, 1227–1232.

Stouffer, R.J., Hegerel, G., Tett, S., 2000. A comparison of surface air temperature variability in three 1000-year coupled ocean–atmosphere model integrations. J. Climate 13, 513–537.

Taylor, K.E., 2001. Summarizing in a single diagram multiple aspects of model performance. J. Geophys. Res. (submitted for publication [also available as PCMDI Report No. 55: http://www-pcmdi.llnl.gov/pcmdi/pubs/ab55.html]).

Tett, S.F.B., Stott, P.A., Allen, M.R., Ingram, W.J., Mitchell, J.F.B., 1999. Causes of twentieth-century temperature change near the Earth's surface. Nature 399, 569–572.

Tokioka, T., Noda, A., Kitoh, A., Nikaidou, Y., Nakagawa, S., Motoi, T., Yukimoto, S., Takata, K., 1996. A transient CO2 experiment with the MRI CGCM: annual mean response, CGER's Supercomputer Monograph Report Vol. 2, CGER-IO22-96, ISSN 1341-4356, Center for Global Environmental Research, National Institute for Environmental Studies, Environment Agency of Japan, Ibaraki, Japan. 86 pp.

Trenberth, K.E., 1998. The heat budget of the atmosphere and ocean. Proceedings of the First WCRP International Conference on Reanalysis, pp. 17–20. WCRP-104, WMO/TD-NO. 876.

Trenberth, K.E., Solomon, A., 1994. The global heat balance: heat transport in the atmosphere and ocean. Clim. Dyn. 10, 107–134.

von Storch, J.-S., Kharin, V.V., Cubash, U., Hegerl, G.C., Schriever, D., von Storch, H., Zorita, E., 1997. A description of a 1260-year control integration with the coupled ECHAM1/LSG general circulation model. J. Climate 10, 1525–1543.

Voss, R., Sausen, R., Cubash, U., 1998. Periodically synchronously coupled integrations with the atmosphere–ocean general circulation model ECHAM3/LSG. Clim. Dyn. 14, 249–266.

Washington, W.M., Weatherly, J.M., Meehl, G.A., Semtner Jr., A.J., Bettge, T.W., Craig, A.P., Strand, W.G., Arblaster, J., Wayland, V.B., James, R., Zhang, Y., 2000. Parallel Climate Model (PCM) control and transient simulations. Clim. Dyn. 16, 755–774.

Wigley, T.M.L., Schlesinger, M.E., 1985. Analytical solution for the effect of increasing CO_2 on global mean temperature. Nature 315, 649–652.

Wu, G.-X., Zhang, X.-H., Liu, H., Yu, Y.-Q., Jin, X.-Z., Gou, Y.-F., Sun, S.-F., Li, W.-P., 1997. Global ocean–atmosphere–land system model of LASG (GOALS/LASG) and its performance in simulation study. Q. J. Appl. Meteorol. 8, Supplement, 15–28 (in Chinese).

Xie, P., Arkin, P., 1997. Global precipitation: a 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. Bull. Am. Meteorol. Soc. 78, 2539–2558.

Zhang, X.-H., Shi, G.-Y., Liu, H., Yu, Y.-Q. (Eds.), 2000. IAP Global Atmosphere–Land System Model. Science Press, Beijing, China. 259 pp.