Climate Dynamics (2001) 17:83–106

S. J. Lambert · G. J. Boer

CMIP1 evaluation and intercomparison of coupled climate models

Received: 26 January 2000 / Accepted: 9 June 2000

Abstract The climates simulated by 15 coupled atmosphere/ocean climate models participating in the first phase of the Coupled Model Intercomparison Project (CMIP1) are intercompared and evaluated. Results for global means, zonal averages, and geographical distributions of basic climate variables are assembled and compared with observations. The current generation of climate models reproduce the major features of the observed distribution of the basic climate parameters, but there is, nevertheless, a considerable scatter among model results and between simulated and observed values. This is particularly true for oceanic variables. Flux adjusted models generally produce simulated climates which are in better accord with observations than do non-flux adjusted models; however, some non-flux adjusted model results are closer to observations than some flux adjusted model results. Other model differences, such as resolution, do not appear to provide a clear distinction among model results in this generation of models. Many of the systematic differences (those differences common to most models), evident in previous intercomparison studies are exhibited also by the CMIP1 group of models although often with reduced magnitudes. As is characteristic of intercomparison results, different climate variables are simulated with different levels of success by different models and no one model is ''best'' for all variables. There is some evidence that the ''mean model'' result, obtained by averaging over the ensemble of models, provides an overall best comparison to observations for climatological mean fields. The model deficiencies identified here do not suggest immediate remedies and the overall success of the models in simulating the behaviour of the complex non-linear climate system apparently depends on the slow improvement in the balance of approximations that characterize a coupled climate model. Of course, the results of this and similar studies provide only an indication, at a particular time, of the current state and the moderate but steady evolution and improvement of coupled climate models.

S. J. Lambert (✉) · G. J. Boer
Canadian Centre for Climate Modelling and Analysis,
Meteorological Service of Canada, University of Victoria,
PO Box 1700, Victoria, B.C. V8W 2Y2, Canada
E-mail: Steven.Lambert@ec.gc.ca

1 Introduction

The potential climatic consequences of increasing concentrations of atmospheric greenhouse gases (GHGs) and of other perturbations to climate forcing are primarily studied using mathematical/physical models of the climate system. The most sophisticated of such models, the CGCMs (coupled global climate models), simulate the hour-by-hour evolution of the three-dimensional state of the atmosphere, ocean, cryosphere, and land surface. The simulated data produced by these models are treated in the same way as observations of the real system, although the global coverage and range of variables from the models is far more complete than that available from the observational network. The model climate is obtained as the statistics of these data. Although attention is given mainly to mean quantities, the three-dimensional distributions of other statistical moments are potentially available and increasing attention is given to second order moments and to extremes.

The CGCM is the best tool currently available for studying the complex climate system, its variability, and its response to changing forcing. In practice, however, CGCMs represent a ''balance of approximations'' that is necessary in order to represent the processes that govern the working of the climate system at the finite resolutions allowed by current computer capability and the limitations of available knowledge of the physical processes involved. The steps taken to pass from the physical principles governing the system to a working global model are described, for instance, in Boer et al. (1992). At each step, some errors and approximations are introduced into the model and it is necessary to quantify

this error, to understand its origins, and to reduce it in so far as possible. This is the task of model verification which can be approached in several ways including model intercomparison.

Several model intercomparison projects (MIPs) are currently underway. These include the Atmospheric Model Intercomparison Project (AMIP, Gates et al. 1999), the Paleoclimate Model Intercomparison Project (PMIP), the Seasonal Forecast Model Intercomparison Project (SMIP) and the Coupled Model Intercomparison Project (CMIP). These projects are community analysis/verification activities which seek: (1) to document the ability of models to simulate the current or past climate; (2) to identify common deficiencies in model results; (3) to formulate hypotheses concerning the causes of model deficiencies; (4) to perform experiments to clarify causes of climate model deficiencies; and (5) to document model evolution. The use of models to simulate potential climate change is a compelling impetus for model analysis. The Intergovernmental Panel on Climate Change (IPCC) has included chapters on model evaluation in each of its reports (IPCC 1990; IPCC 1995).

While model intercomparison has a fairly lengthy history and is certainly an area of considerable current activity, the analysis and intercomparison of sophisticated coupled climate models, which include a three-dimensional representation of both the atmosphere and the ocean, are comparatively recent (Gates et al. 1993; IPCC 1995). Here we evaluate and intercompare the control or unperturbed climates of some 15 coupled atmosphere/ocean models. This is part of the Coupled Model Intercomparison Project (CMIP) sponsored by the Working Group on Coupled Models (WGCM) of the World Climate Research Program (WCRP) which is described briefly in Meehl et al. (1997). Infrastructure support for this and other MIPs is provided by the Program for Climate Model Diagnosis and Intercomparison (PCMDI). Intercomparison information and links to other sources of information may be found at http://www-Pcmdi.llnl.gov/PCMDI.html. (Covey 1998).

2 Coupled models

Tables 1 and 2, based on Phillips (1998), give some basic information on the simulations which are part of this intercomparison. Results from 15 models from 12 institutions are analyzed. Simulations from three other modelling groups submitted to CMIP are not included since in one case surface salinity is controlled by relaxation to observed climate values, in the second sea-ice is prescribed, and in the third no ocean data are available.

2.1 The atmospheric component

As indicated in Table 1, the atmospheric components (AGCMs) of 10 of the coupled models employ spectral numerics in the horizontal. The resolutions range from a triangular truncation at 21 waves (transform grid size 5.6°) to a triangular truncation at 42 waves (transform grid 2.8°). The remaining five grid point models employ resolutions ranging from $4 \times 5°$ to $1.6 \times 3.8°$. The number of points in the horizontal, and hence the number of points at the surface (Pts) has a median value of 3.2×10^3 with a factor of 4 between the highest and lowest horizontal resolutions. In the vertical, almost half of the AGCMs have 9 or 10 levels, and the rest between 15 and 20 with one model having 30 levels in the vertical. While it can be misleading, the product of the number of grid points in the horizontal and the number of vertical levels gives

Table 1 The resolutions of the atmosphere, the ocean, and the coupled models

Model	Atmosphere resolution				Ocean resolution				Coupled model			
	Horizontal			Ver	Horizontal			Ver	A	O	A + O	A/O
	Res	Grid	Pts 10^3	Lev	Res	Grid	Pts 10^3	Lev	APts 10^3	OPts 10^3	AOpt 10^3	
COLA	R15	48×40	1.9	9	1.5×1.5	240×120	28.8	20	17.3	322.6	340	0.05
GFDL	R15	48×40	1.9	9	4.5×3.8	96×40	3.8	12	17.3	22.6	40	0.77
NCAR_WM	R15	48×40	1.9	9	1.0×1.0	360×180	64.8	20	17.3	661.0	678	0.03
CCSR	T21	64×32	2.0	20	2.8×2.8	128×64	8.2	17	41.0	64.1	105	0.64
CERFACS	T21	64×32	2.0	30	2.0×2.0	180×90	16.2	31	61.4	271.2	333	0.23
MPI_E3/L	T21	64×32	2.0	19	4.0×4.0	90×45	4.1	11	38.9	23.6	62	1.65
GISS_M	5×4	72×45	3.2	9	5.0×4.0	72×45	3.2	16	29.2	28.0	57	1.04
GISS_R	5×4	72×45	3.2	9	5.0×4.0	72×45	3.2	13	29.2	24.0	53	1.22
MRI	5×4	72×45	3.2	15	2.5×2.0	144×90	13.0	21	48.6	152.4	201	0.32
CSIRO	R21	64×56	3.6	9	5.6×3.3	64×56	3.6	21	32.3	39.4	72	0.82
LMD	1.6×3.8	96×48	4.6	15	2.0×2.0	180×90	16.2	31	69.1	275.0	344	0.25
CCCma	T32	96×48	4.6	10	1.8×1.8	192×96	18.4	29	46.1	256.6	303	0.18
UKMO	3.8×2.5	96×72	6.9	19	2.5×3.8	96×72	6.9	20	131.3	73.2	204	1.79
MPI_E4/O	T42	128×64	8.2	19	2.8×2.8	128×64	8.2	11	155.6	43.3	199	3.67
NCAR_CSM	T42	128×64	8.2	18	2.4×2.0	150×90	13.5	45	147.5	291.6	439	0.51
Median			3.2	15			8.2	20	41.0	73.2	201	0.64
Range			1.9–8.2	9–30			4–65	11–45	17–155	22–323	40–678	0.03–3.7

For the atmosphere and ocean components, Pts is the number of points in the horizontal and hence at the surface. For the coupled model, APts is the product of the number of levels and the number of points in the atmospheric grid and OPts is the product of the number of levels and the number of non-land points in the ocean grid and their sum, AOPts is the total number of points specifying the value of a variable like temperature at a particular time in the coupled system. The ratio of APts and OPts is denoted "A/O" in the rightmost column. Shading indicates models which do not employ flux adjustments

Table 2 Some basic information on the participating models

Model	Reference	OGCM			Ice	Adjust	Land		IC
CCCma	Flato et al. (2000); Boer et al. (2000)	P	R	Z	T	H W	BB	R	E
CCSR	Abe-Ouchi et al. (1996)	P	R	Z	T	H W	B	R	E
CERFACS	Guilyardi and Madec (1997)	P	R	Z	DIAG	None	MC	N	I
COLA	Schneider et al. (1997); Schneider and Zhu (1998)	P	R	Z	T	None	MC	N	I
CSIRO	Gordon and O'Farrell (1997); Hirst et al. (2000)	P	R	Z	T DYN	H W M	MC	R	E
GFDL	Manabe et al. (1991); Manabe and Stouffer (1996)	P	R	Z	T F	H W	B	R	E
GISS_M	Miller and Jiang (1996)	P	R	Z	T	None	MC	N	E
GISS_R	Russell et al. (1995)	P	F	M	T	None	M	R	I
LMD	Braconnot et al. (1997); Fichefet (1997)	P	R	Z	DIAG	None	MC	N	I
MPI_E3/L	Voss et al. (1998)	Q	F	Z	T	H W M	MC	R	E
MPI_E4/O	Roeckner et al. (1996)	P	F	D	T DYN	Ha Wa	MC	R	E
MRI	Tokioka et al. (1996)	P	R	Z	T F	H W	M	R	E
NCAR_CSM	Boville and Gent (1998)	P	R	Z	T DYN	None	MC	N	E
NCAR_WM	Meehl and Washington (1995); Washington and Meehi (1996)	P	R	Z	T DYN	None	B	N	I
UKMO	Johns (1996); Johns et al. (1997)	P	R	Z	T F	H W	BB MC	R	IE

The entries for the "OGCM" column are for dynamics (P-primitive equations; Q-quasi-geostrophic), upper boundary condition (R-rigid lid; F-free surface), and vertical coordinate (Z-height; M-mass-weighted; D-density/isopycnal. Features of the ice models are (T-thermodynamic; DIAG-diagnostic based on surface temperature; DYN-ice dynamics with rheology; F-ocean-forced drift. The "Adjust" column lists the variables which are flux adjusted (H-heat; W-water; M-momentum). Adjustments are monthly except Ha and Wa which are annual. Shading indicates models which do not employ flux adjustments. Characteristics of the land surface scheme are given in the "Land" column (B-single level bucket; BB-modified single layer scheme with variable layer depth and evapotranspiration; M-multi-level scheme; MC-modelled canopy; R-runoff to oceans; N-no runoff routing. The right-most column refers to initialization (E-substantially equilibrated ocean and atmosphere; I-observation-based; IE-partial upper ocean equilibration

some indication of the model resolution in terms of the number of values that specify the temperature or other variables at each time step. This number is listed in the APts column under Coupled Model in Table 1. There is almost a factor of 10 between the highest and lowest number of points which serve to represent the atmospheric distribution of variables.

A suite of subgrid scale physical parametrizations (radiation, convection, precipitation and cloudiness, boundary layer, etc.) are basic in atmospheric models. They are strong determinants of the model climate but differences cannot be characterized simply as can be done for the numerical methods and resolutions. Table 2 and Phillips (1998) give basic information on the parametrizations employed in the models.

2.2 The ocean component

The ocean components of the coupled models (OGCMs) have a stronger family resemblance than do most atmospheric models for two reasons. Firstly, all of the ocean models are grid-point models and many of the OGCMs are some version of the "modular ocean model" (MOM) (Pacanowski et al. 1995) or its predecessor. Secondly, current ocean models have fewer subgrid scale physical parametrizations than do atmospheric models so provide less opportunity for differences in this aspect. As noted in Phillips (1998) and references therein, there are, nevertheless, a variety of ways in which subgrid mixing processes, for instance, differ among ocean models.

As was the case for the atmosphere, ocean model resolutions vary considerably from a high of $1 \times 1°$ to a low of $4 \times 5°$ in the horizontal and from 11 to 45 levels in the vertical. The number of grid points that specify a variable at a particular time (OPts) differs by a factor of 15 among the models as shown in Table 1. The resolution of ocean models is typically higher than that of atmospheric models but they do not dominate the computational load of the coupled model because of the computational expense of the atmospheric model's physical parametrizations.

Table 2 gives a very broad characterization of the ocean component and the use or not of flux adjustment in the coupling between atmosphere and ocean. References describing the various models and model components are also given.

2.3 Sea-ice

Climate and climate change depend on a variety of feedback processes including the albedo-temperature feedback that involves the amount and distribution of snow on land and sea-ice. All models have an interactive representation of snow cover on land. Sea-ice is occasionally treated diagnostically where the ice cover is simply specified based on ocean surface temperature, but more commonly and more physically, it is treated thermodynamically whereby its thickness and distribution are determined interactively by an ice model. Some models include ice movement by allowing the ice to drift with the ocean currents while others attempt to include the more complete physical effects of wind forcing on ice motion including the effects of ice rheology. The various ice treatments used in the models are indicated in Table 2.

2.4 Land surface

The representation of the land surface in models varies considerably in complexity. The simplest representation is the "bucket" scheme which treats the land surface as a single 15 cm deep layer. Single layer schemes can be modified to allow variation in depth, soil type, and vegetation cover. The most sophisticated land surface schemes are multi-layer and can include representation of a canopy. Some indication of the land-surface treatment is indicated in Table 2 and a more complete description is included in Phillips (1998).

2.5 Initial conditions

Coupled models may be initialized from available observations and this is done for five of the 15 coupled models. Global atmospheric analyses are readily available but ocean observations are sparse both in space and in time. The ocean component is initialized with three-dimensional climatological values, typically those of Levitus and Boyer (1994) and Levitus et al. (1994). The resulting state is not in equilibrium and will necessarily adjust or "drift" for some period of time. Typically there is a comparatively short time scale in the upper ocean and a much longer time scale in the deep ocean.

The remaining 10 models attempt to initialize the system near equilibrium so that drift is reduced. Both the atmospheric and oceanic components are spun-up separately to equilibrium and then coupled, possibly using flux adjustment.

2.6 Flux adjustment

Atmospheric and oceanic components typically exchange information once per simulated day. The atmospheric model simulates surface fluxes of heat, fresh water, and momentum H_a, W_a, $\boldsymbol{M}_a$, given the sea-surface temperature T_o. These fluxes are used to force the ocean model which produces new values of surface temperature and so on. In a *flux adjusted* (FA) coupled model, the ocean receives the interactively calculated fluxes from the atmospheric model modified by *flux adjustments*, so that $(H_o, W_o, \boldsymbol{M}_o) = (H_a, W_a, \boldsymbol{M}_a) + (\delta H, \delta W, \delta \boldsymbol{M})$. The flux adjustments, $(\delta H, \delta W, \delta \boldsymbol{M})$ are not themselves interactive and are obtained *once and for all* and do not subsequently change.

The purpose of flux adjustment, or equivalently of "anomaly coupling", is to ensure that the control climate is reasonable, without large trends and drifts, and that climatic feedback processes are operating in their normal range. In order to determine the climate response to a change in GHG or other forcing in models with large drifts, it is necessary to subtract the control simulation from the climate change simulation. Such a procedure can introduce errors in the determination of the climate change response. The argument against flux adjustment is that it is not physical and may itself distort the climate response to changes in forcing. The flux adjustments are typically proportional to the differences between simulated and observed values of temperature and salinity so that there is a compromise between the size of the flux adjustment and the size of the error accepted in the coupled models. Models without flux adjustment typically exhibit larger control climate differences from observations than do flux-adjusted models. Of course, as models improve, flux adjustments will become smaller, as will differences between observed and simulated values in non-flux-adjusted models. In what follows we stratify some results depending on whether the coupled models are flux adjusted or not.

3 Data

Model intercomparison, as the name implies, analyzes model results as a group but the comparison with observations is typically part of any MIP.

Table 3 The atmospheric fields available in the CMIP1data base

CMIP1 atmosphere model data
Global distributions (DJF and JJA means)
–Precipitation
–Mean sea level pressure
–Surface wind stress
–Surface heat flux
–Surface fresh water flux
–Snow
–Sea-ice
Zonal cross sections (DJF and JJA means)
–Zonal wind
–Meridional wind
–Temperature
–Specific humidity
Standard deviations
–Annual zonal mean temperature
–Decadal zonal mean temperature
Time series of individual monthly mean
–Surface air temperature

Table 4 The oceanic fields available in the CMIP1data base

CMIP1 ocean model data
Global distributions (DJF and JJA means)
–Sea surface temperature
–Sea surface salinity
–Flux adjustments
Zonal cross sections (DJF and JJA means)
–Atlantic Basin temperature
–Indian Basin temperature
–Pacific Basin temperature
–Global temperature
–Atlantic Basin salinity
–Indian Basin salinity
–Pacific Basin salinity
–Global salinity
Global distribution (annual mean)
–Barotropic stream function
Zonal cross sections (annual mean)
–Atlantic Basin meridional overturning stream function
–Indian Basin meridional overturning stream function
–Pacific Basin meridional overturning stream function
–Global meridional overturning stream function
Integrated oceanic transport
–Heat

3.1 Model data

The data available for CMIP1 analysis are listed in Covey (1998) and summarized in Tables 3 and 4. The model data are available on a variety of grids, atmospheric pressure levels, and ocean depths. To facilitate intercomparison, all the global data are interpolated to 192×96 Gaussian grids. Atmospheric cross sections are interpolated to 96 Gaussian latitudes and the standard pressure levels: 50, 70, 100, 150, 200, 250, 300, 400, 500, 700, 850, 925, and 1000 mb. The oceanic cross sections are also interpolated to 96 Gaussian latitudes and depths: 0, 50, 100, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, and 5500 m.

3.2 Observation-based data

The comparison of simulated results with the observation-based climate record is difficult because the global distributions of a number of basic climatic quantities, such as precipitation, are not well known. The observational data used for comparison are given in Table 5. Da Silva et al. (1994) corrected their net heat flux values to ensure energy balance. The observed data are available on a variety of grids and are interpolated to the common Gaussian grid to aid in the comparisons with the simulated data.

4 Analysis

Basic analysis of coupled model behaviour compares model results with observation-based climatologies. Surface air temperature (SAT) is a measure of the surface energy budget and is a robust climate parameter; robust in the sense that it is comparatively large-scale, exhibits modest variability, and is comparatively well observed. Precipitation is a measure of the hydrological cycle but it is much less robust than temperature. Precipitation exhibits comparatively small scales, is highly variable and not well observed, especially over oceans, but is a primary meteorological variable of considerable practical importance. The amount and distribution of snow and sea-ice are measures of cryospheric processes and are analyzed in a separate intercomparison. Mean sea-level pressure is

Table 5 Sources of observational data

Field		Source	Reference
Surface heat flux	H	UWM COADS	Da Silva et al. (1994)
MSL pressure	PMSL	ECMWF reanalyses	Gibson et al. (1997)
Precipitation	P	Xie-Arkin	Xie and Arkin (1997)
Surface air temperature	SAT	NCAR	Jenne (1975)
SAT inter-annual variance		NCEP reanalyses	Kalnay et al. (1996)
Sea surface temperature	SST	GISST 2.2	Rayner et al. (1996)
Ocean temperature	T	Levitus 1994	Levitus and Boyer (1994)
Salinity	S	Levitus 1994	Levitus et al. (1994)
Ocean heat transport	F_o	Trenberth-Solomon	Trenberth and Solomon (1994)
North Atlantic stream function			Macdonald and Wunsch (1996) Schmitz (1995)
ACC transport			Whitworth et al. (1982) Whitworth and Petersen (1985)

a measure of both thermodynamical and dynamical balances at the surface. It is more robust than precipitation but less so than temperature.

Ocean variables, excepting sea-surface temperature (SST), are comparatively poorly known and are often given less attention in climate simulations. The oceanic temperature and salinity, the thermohaline overturning circulation, and the barotropic stream function are very basic measures of the energy and fresh water balances and of oceanic flow.

4.1 Means

A variable from a control simulation that is in statistical equilibrium under constant external forcing is represented as $Y = Y_m + Y'$, the sum of a long term time-independent climatological mean on which is superimposed random natural variability which averages to zero over sufficiently long time. Quantities are functions of latitude, longitude and height in the atmosphere or depth in the ocean. The mean is estimated as the long-term time average $\bar{Y} = Y_m + \bar{Y}' \approx Y_m$. The averaging periods used for the CMIP data and for the observed data are long enough to provide good estimates of the climate means.

The basic intercomparison is between long-term means of observed and modelled quantities since these are the only statistics available for most variables in the CMIP1 archive. A sequence of individual monthly means is available for SAT which permits the analysis of other statistics.

4.2 SAT trend and variability

The sequence of monthly means for SAT allows the analysis of the trend and natural variability of this quantity. If there is drift in either the observed or modelled climate, it may be partially removed by fitting to some function. For instance, a linear trend $X(t) = \bar{X} + bt + X' = \bar{X} + X_t + X'$ may be fitted by least squares. The space-time variance $\sigma^2 = \langle (X - \bar{X})^2 \rangle = \langle X_t^2 \rangle + \langle X'^2 \rangle = \sigma_t^2 + \sigma_n^2$, where $\langle\ \rangle$ denotes a global average, has a contribution from the linear trend, if it is present, and from the variability about the trend.

Trends in observed temperature may be associated with global warming due to greenhouse gas increases in the atmosphere (IPCC 1995). In the CMIP control simulations, however, there is no change in the external forcing and a non-zero trend over the length of the record indicates that the modelled system is not in equilibrium with the imposed constant forcing or, possibly, the presence of unforced long time scale variability in the simulated result.

4.3 Second order difference statistics

We deal here primarily with the global spatial patterns of mean climate variables. Observation-based quantities are denoted by X and simulated values by Y. These quantities may be further decomposed as $X = \langle X \rangle + [X]^+ + X^*$. At any level in the atmosphere or depth in the ocean, the global mean $\langle X \rangle$ is a single number; the zonal average, $[X]$ and $[X]^+ = [X] - \langle X \rangle$, the north-south meridional structure about the global mean, are functions of latitude; and $X^* = X - [X]$, the geographical pattern about the zonal mean, is a function of latitude and longitude. The decomposition takes into account that the strong north-south or meridional structure is a dominant feature of many climate variables while the remaining geographic pattern, although of great practical importance, accounts for less of the total variance. The decomposition allows the two components to be analyzed separately.

Differences between simulated and observed quantities may also be decomposed into components as $d = Y - X = \langle d \rangle + d^+ = \langle d \rangle + [d]^+ + d^*$ .The global average mean square difference (*msd*) is $\langle d^2 \rangle = \langle d \rangle^2 + \langle [d]^{+2} \rangle + \langle d^{*2} \rangle$. The total *msd* or its components may be represented by the root mean square difference $\sqrt{\langle d^2 \rangle}$ which has the units of the original variable.

The *relative* mean square difference (*rmsd*) measures the *msd* against the variance of the observed field as $\langle d^{+2} \rangle / \langle X^{+2} \rangle$ and/or $\sqrt{\langle d^{+2} \rangle} / (\sqrt{\langle X^{+2} \rangle})$. The relative *msd* of the components, $\langle [d]^{+2} \rangle / \langle [X]^{+2} \rangle$ and $\langle d^{*2} \rangle / \langle X^{*2} \rangle$, is also of interest. Finally, the spatial correlation coefficient may be calculated for the total field and separately for the components. Information on these second order measures of agreement between simulated and observed fields, namely the relative *msd*, correlation, and the ratio of simulated to observed variance are presented on a single diagram.

4.4 Model ensemble statistics

There are 15 models in the intercomparison giving 15 values of each climatological quantity, one from each model. In order to condense the amount of data and to perform simple statistical calculations, we follow Lambert and Boer (1989) and average the information across models. More particularly, we adopt a "multi-model ensemble" approach to the analysis of simulated climate. The ensemble view assumes that each model result represents a plausible solution to the governing equations and hence an independent realization of the climate.

The multi-model ensemble-average estimate of the climate mean is the average across the individual model values Y_α, namely, $\{Y\} = \frac{1}{N} \sum_{\alpha=1}^{N} Y_\alpha$. This is the *mean model* result. We may compare the *mean model* result with observations just as we do individual model results. The ensemble mean difference from the observations and the intermodel variance are $\{d\} = \{Y\} - X$ and $\sigma_I^2 = \frac{1}{N} \sum_\alpha (Y_\alpha - \{Y\})^2$. The ensemble mean difference indicates how successful the models are in simulating the observed climate "on average" while the intermodel standard deviation measures the scatter among model results and indicates how consistent the models are. As models improve, the ensemble mean will converge toward the observed value and both the mean difference and intermodel scatter will decrease to a value governed by sampling error.

To the extent that model differences are random and independent, they will cancel on averaging and $\{d\}$ will approach zero. The

mean model climate, $\{Y\} = X + \{d\} \approx X$, may then be a better estimate, on average, of the actual climate than the result from a particular model. The part of the difference $\{d\}$ that survives averaging over all model values indicates a systematic deficiency in models. A small value of σ_I indicates agreement among models and supports the assumption that they are capturing the processes that govern that variable and hence its climate. A large value of intermodel scatter, on the other hand, indicates disagreement and unreliability.

We will be interested in $\{d\}$ and σ_I as useful general measures of model difference from observations. If the observations are assumed to have errors much smaller than the intermodel scatter, then it may also be of some interest to calculate the *t*-statistic $t = \{d\}/\sigma_I$ to test if the model ensemble mean differs in a statistically significant way from the observed value. If it does so, then there is evidence of *systematic error* in the models. Certain systematic errors are a common and persistent feature of model results (e.g. Boer et al. 1992; Gates et al. 1999). Their causes are often difficult to identify and to remedy. The identification and amelioration of systematic error is a major goal of model intercomparison.

5 Presentation

The various statistics may also be calculated from simulated data stratified by model type, for instance flux-adjusted (FA) models and non-flux-adjusted (NFA) models, models with higher and lower resolution, models with different methods of initialization, and so forth. The available statistics, although they represent only a subset of model output, are nevertheless voluminous. The data are condensed for consideration and display in several ways under various forms of averaging. In particular, results are displayed in the form of:

1. Global means, $\langle Y \rangle$, which are single numbers
2. Zonal averages, $[Y]$, which are functions of latitude
3. Geographical distributions, Y, which are functions of latitude and longitude

which are available for each model. In addition as discussed already, the results may be further condensed by calculating ensemble means and intermodel standard deviations.

For SAT, the analysis can be expanded in several ways since a time-series of monthly means is available. Trends and interannual variability is examined here and

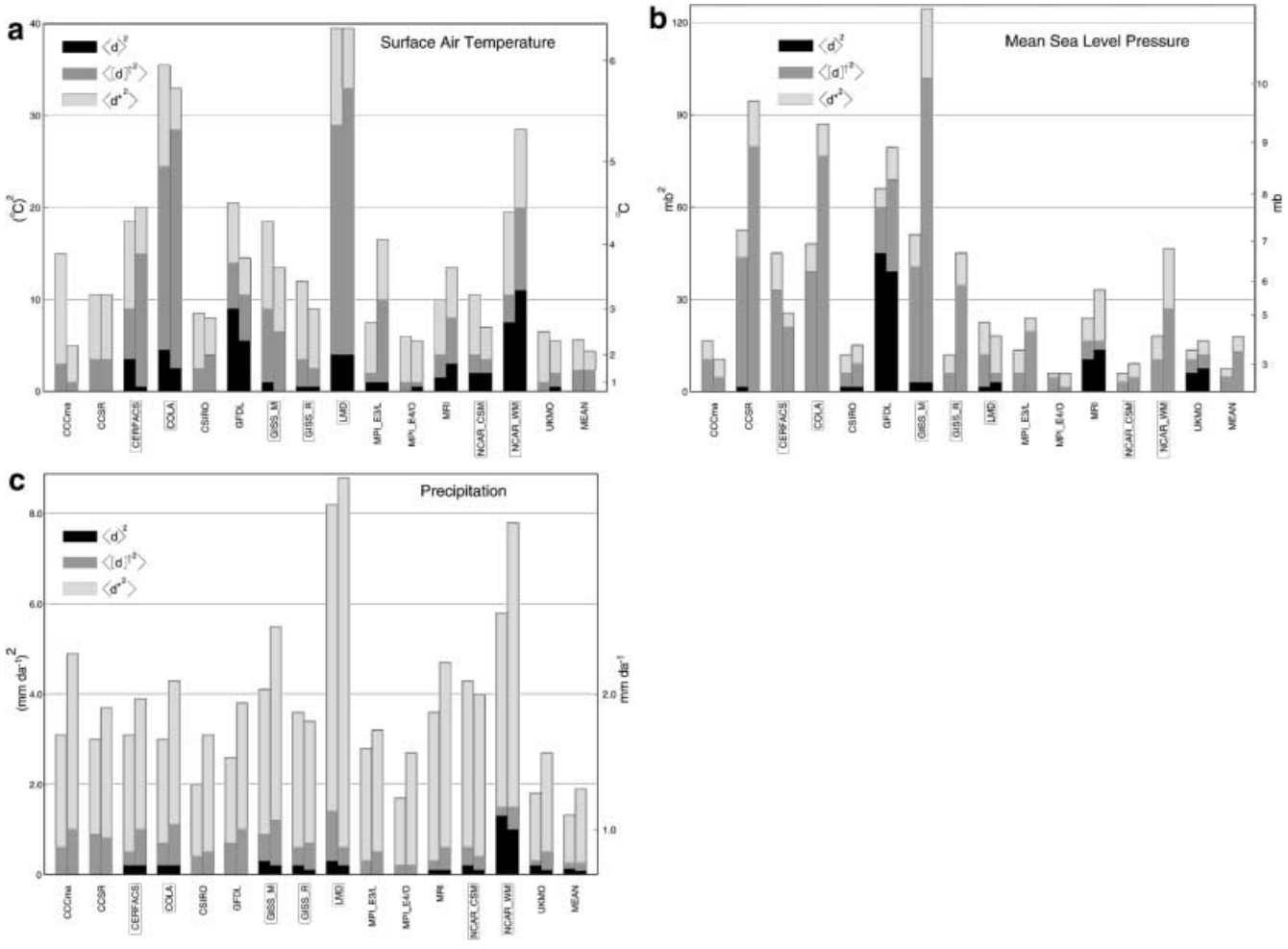


Fig. 1a–c Mean square differences between the global distributions of simulated and observed climate means of surface air temperature, precipitation and mean sea-level pressure. Contributions are from the difference in global means, meridional structures, and the remaining geographical pattern as discussed in the text. The *first column* is the result for the December–February and the *second* for the June–August seasons. Non-flux-adjusted models have their identifiers *enclosed in boxes*. The *mean model* result is the *rightmost entry*

annual cycle aspects of the SAT are reported in Covey et al. (1999).

5.1 Global statistics

Many fields are available from the models participating in CMIP and only a small subset can reasonably be used for intercomparison. In selecting this subset, the importance of the variables must be balanced against the reliability of the available observations. We concentrate mainly on surface air temperature (SAT), precipitation (P), and mean sea level pressure (MSLP). These geographically distributed quantities are of considerable practical importance and are also associated with aspects of the energy, fresh water, and momentum fluxes between atmosphere and ocean. Some information is also supplied on sea-surface temperature (SST), sea-surface salinity (SSS) and heat flux (H) between atmosphere and ocean.

5.2 Atmospheric surface variables

Figure 1 displays the globally averaged mean square difference (*msd*) $\langle d^2 \rangle$ between simulated and observed quantities and the components $\langle d \rangle^2$, $\langle [d]^{+2} \rangle$, $\langle d^{*2} \rangle$ indicating differences in the global mean, the meridional structure, and the geographical pattern. Results for SAT, P, and MSLP, for the DJF and JJA seasons are shown with NFA model results indicated by enclosing their identifiers in boxes. The ensemble *mean model* result is the rightmost entry. The scale on the right-hand-side of the diagram measures the root mean square difference $\sqrt{\langle d^2 \rangle}$ and may be applied to the components individually to indicate the associated *rmsd* values.

Figure 1 shows how differences in the mean, zonal, and geographic patterns contribute to the total *msd*. An alternative, and in some ways more revealing, view is given in Fig. 2 in which second order statistics comparing observed and modelled fields are shown. Figure 2 is an extension of a diagram due to Taylor (Gates et al. 1999) as discussed in Boer and Lambert (2000). The diagram displays relative mean square differences, pattern correlations and the ratio of variances for simulated and observed quantities. The statistics for the zonal structure, $[d^+]$ and the geographical pattern d^{*2} are shown separately by dots and triangles respectively. These are forced components and the diagrams provide information on the model's ability to reproduce the climatological spatial structure of the fields. Figures 1 and 2 provide complimentary views of the agreement of simulated and observed quantities.

We note only a few of the many features of these diagrams. Figure 1 indicates that different variables have different *msd* characteristics both for the total and the components. Mean square differences in meridional structure are comparatively large for SAT and MSLP while the geographic pattern is the largest component for precipitation. Differences in global means are comparatively small. There is a small but generally consistent seasonal difference in *msd* values for the variables.

In terms of relative *msd*, however, Fig. 2 shows that, with the exception of MSLP, the meridional structures are comparatively better represented than the geographic patterns and this is particularly true for SAT. The large variance associated with the pole to equator SAT structure is relatively well captured by the models although differences are still appreciable in an absolute sense. The zonal structures of the precipitation and especially the MSLP fields are not as well captured as that of the SAT in this relative sense.

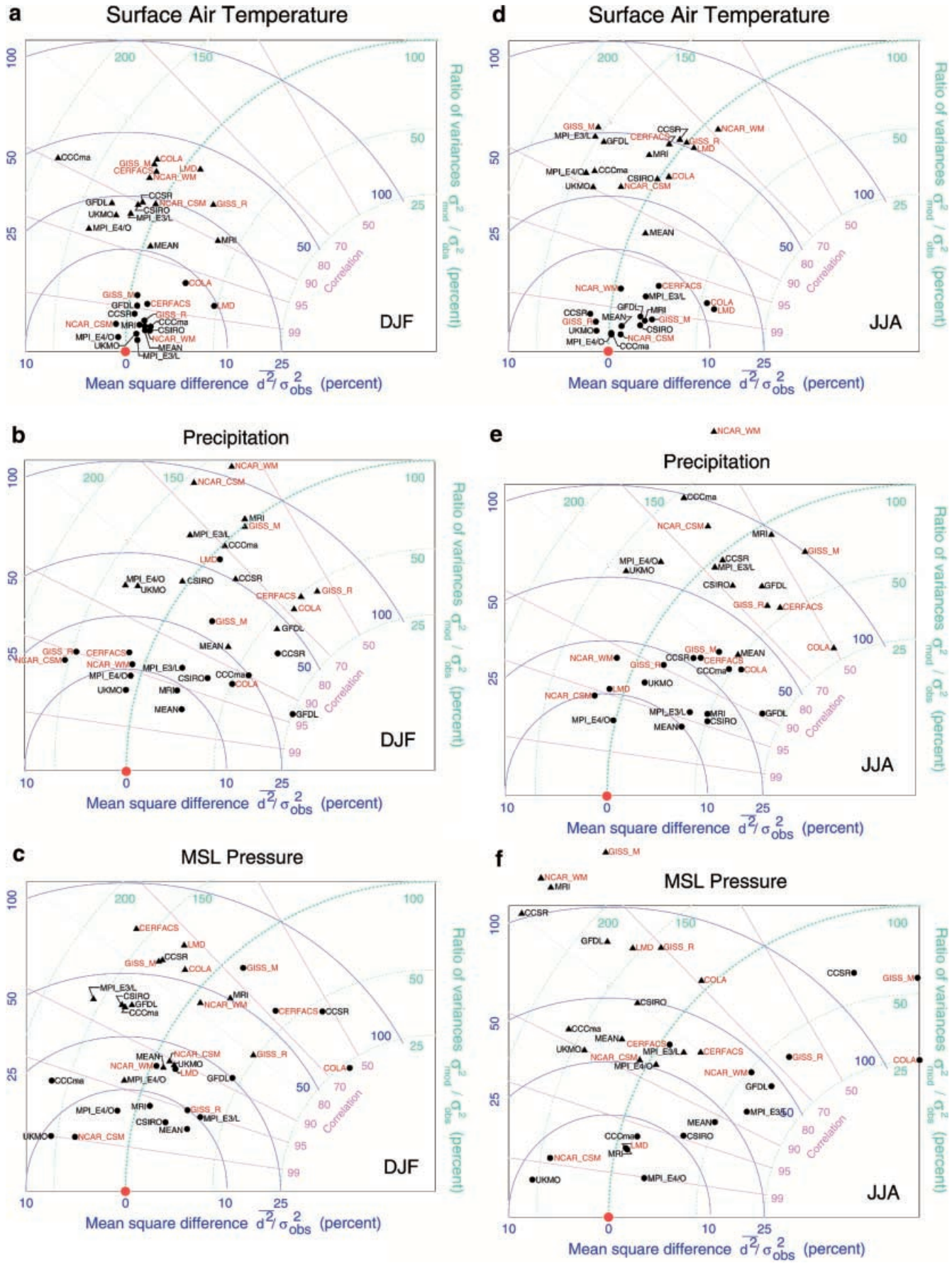
The geographic patterns Y^* remaining after subtracting out the mean and zonal averages of the fields typically have less variance than the meridional structures and are a sensitive test of a model's ability to simulate the details of the climatological distributions of the fields. These relative *msd*s are also shown in Fig. 2. Except for MSLP, they are typically larger, and the correlations smaller than for meridional structures.

Figure 2 indicates that the structure of SAT is generally best represented followed by MSLP and P and that the DJF season is generally better simulated than JJA. Both figures indicate that FA model results are generally closer to the observations, at least for this generation of models, than are NFA model results although some individual NFA model results are closer to observations than some FA model results. The figures also show that different models have different errors in different variables and that no model is best for all variables.

According to Fig. 1, the *msd* for the *mean model* is seldom larger than that for any particular model. In other words, the *mean model* is generally the best model measured in this way. This is also the case for the second order statistics of Fig. 2 and particularly so for the geographic pattern (indicated by the triangles) where generally the *mean model* result is visually separate from the cloud of individual model results. The *mean model* value is obtained by averaging over the set of model values and this averaging might be expected to result in smoother patterns and smaller variance than for the observed field but this is by no means always the case. The *mean model* results generally support the discussion of Sect. 4.5 and the idea that the model ensemble mean may provide a better overall simulated climate (and by implication climate change) than the results of any particular model.

The atmospheric surface variables SAT, P, and MSLP are global in extent and their climatologies are comparatively well known. The climatologies of the ocean surface variables SST, SSS, H tend to be somewhat less well known and, of course, are not global in extent but confined to ocean areas. We consider the agreement between simulated and modelled values in terms of spatial correlation coefficients in Table 6.

The north-south structures $[Y]^+$ of these quantities are generally well captured in model results as measured


a
Surface Air Temperature
DJF
b
Precipitation
DJF
c
MSL Pressure
DJF
d
Surface Air Temperature
JJA
e
Precipitation
JJA
f
MSL Pressure
JJA
Ratio of variances $\sigma^2_{mod} / \sigma^2_{obs}$ (percent)
Correlation
Mean square difference $\overline{d^2} / \sigma^2_{obs}$ (percent)

◀

Fig. 2a–f Second order difference statistics. The three interlinked quantities, the relative mean square difference, the ratio of the simulated to observed variance, and the correlation between simulated and observed components are shown for the meridional structure (*dots*) and the geographical pattern (*triangles*) as discussed in the text. The *mean model* result is also shown

by the spatial correlation coefficient. This is particularly true for SST and H but less the case for SSS. The geographical pattern component, however, is much less well captured. In Table 6, the median of the 15 model values is given together with the range of the correlation values for FA and NFA model results, and the result for the *mean model* is also given. Since flux adjustment is designed to reproduce the observed SST and SSS patterns, it is not surprising that the correlation is generally higher for FA models than for NFA models. In one case the separation is so great that the ranges do not overlap, so that the highest NFA correlation is lower than the lowest FA correlations. Correlations are comparatively high for SST but low for SSS and H. The *mean model* correlation value is always higher than the median and near the maximum in the range of individual model values.

5.3 SAT trend and variability

The "stability" of the control simulations is of interest when assessing the results of climate change experiments. One measure of stability is the lack of trends in these control simulations. Table 7 gives the linear trend in the global mean SAT and also the contribution to the overall SAT variance from the local linear trend and from the variability about the linear trend as described in Sect. 4.2. For models with over 100 years of available data, only the first 100 years are used and for those models with less than 100 years, the trend is computed using the available data. If there is significant long-period variability in the simulations, then computing the trend over 100 or fewer years may not be entirely representative of a model's drift.

As expected, the FA models generally exhibit smaller trends than do the NFA models. The latter tend to have short time series indicating the difficulty in producing stable control climates without flux adjustment in this generation of models. The trends in global mean SAT indicate a systematic tendency for warming in the control runs of current models. Although the drift in the control simulation may be subtracted out of the results of a climate change experiment, this assumes linearity of the response. It is preferred that the drifts are small compared to those expected in the climate change experiment. These are of the order of 0.5 °C/100 years for the forcing change during the twentieth century and are expected to be in the range 1.7–4.0 °C for the next 100 years (IPCC 1995).

In the mean, the FA models exhibit higher variability about the linear trend than do the NFA models; however, the difference between the two groups is not statistically significant.

Table 7 Linear trend for globally averaged surface air temperature (C/100 years) and the standard deviation associated with the local trend and interannual variability (°C). Shaded entries indicate NFA models. Length is the number of years used to compute the trend

Model	Trend in $\langle T \rangle$	σ_t	σ_n	Length
CCCma	0.240	0.144	5.85	100
CCSR	0.735	0.156	6.32	40
CERFACS	0.553	0.120	4.71	40
COLA	0.352	0.117	5.44	50
CSIRO	−0.009	0.058	5.31	100
GFDL	0.197	0.170	6.51	100
GISS_MILL	0.929	0.445	6.14	89
GISS_RUSS	0.702	0.542	5.02	98
LMD	4.206	0.321	5.47	24
MPI_E3/L	0.188	0.105	4.92	50
MPI_E4/O	−0.124	0.165	5.60	100
MRI	1.057	0.341	5.66	100
NCAR_CSM	−0.211	0.195	5.62	100
NCAR_WM	0.892	0.520	6.78	100
UKMO	0.030	0.080	5.62	100

6 Meridional and geographical structures

Geographical displays provide a spatial picture of climate variables but the number of possible diagrams is unmanageable. We use this type of display to present results averaged over groups of models rather for individual models. Zonal averages of a climate parameter may be plotted on a single diagram. The resulting "spaghetti" diagram provides an immediate indication of the scatter of the individual model results and their departures from the observation-based meridional structure.

Table 6 Range of pattern correlations between components of the observed and simulated patterns of SST, SSS, and H for the meridional structure, $[Y]^+$, and the geographical, Y^*, components

	Sea surface temperature		Sea surface salinity		Heat flux	
	DJF	JJA	DJF	JJA	DJF	JJA
$[Y]^+$ correlation range	0.94–1.00	0.95–1.00	0.56–0.97	0.47–0.93	0.94–0.98	0.94–0.98
Y^* correlation median value	0.77	0.78	0.49	0.53	0.61	0.29
FA range	0.73–0.92	0.78–0.93	0.47–0.76	0.38–0.69	0.56–0.79	0.17–0.52
NFA range	0.50–0.77	0.42–0.72	0.37–0.49	0.32–0.53	0.41–0.66	0.21–0.38
Mean model value	0.91	0.89	0.71	0.68	0.75	0.45

6.1 Surface air temperature

6.1.1 Means

Figures 3 and 4 display SAT results for DJF and JJA. The model mean and the intermodel standard deviation are displayed in the upper left panel. The intermodel scatter is largest in the polar regions and over high terrain where the height of the terrain varies from model to model. The zonally averaged values (upper right) show that the large north-south variation in temperature is well captured by the models in a relative sense (as reflected also in Fig. 2). The diagram visually indicates that, as expected, NFA model results (the dashed lines) tend to be farther from the observations than do FA model results. The geographic patterns of SAT for FA and NFA models are shown separately in the two bottom panels in the figure. Differences from observations are generally smallest over the oceans (as expected for FA models) and tend to be large over regions of high terrain where there is some ambiguity in the representation of SAT in both model and observations. The spatial pattern of differences between simulated and observed SAT is similar for FA and NFA models although NFA differences are larger. This is especially evident in the southern oceans. Noticeable differences also occur off the west coasts of Africa, North America and South America and this is likely the result of a lack of simulated low-level cloudiness. The FA models can alleviate this systematic error by adjusting the fluxes in these areas.

6.1.2 Interannual temperature variability

Results for the Northern Hemisphere extra-tropical inter-annual variance of SAT for January are given in Fig. 5. The left panel displays the observed field computed from the NCEP reanalyses. The largest variance is found in the North Atlantic and is associated with the sea-ice edge. The right panel gives the ensemble mean of the SAT variance (contours) which is, in effect, the *mean model* result for this second order quantity. The ratio of the ensemble mean variance to the observed variance is indicated by shading. The shading pattern was chosen to highlight the extrema in the departure field and it should be noted that it is not symmetric about the value one. The model mean does not exhibit the large variance near the ice edge seen in the observations. The model mean variance exhibits the broad aspects of the observed pattern such as the continental maxima over North America and Asia, but the shading shows that differences of a factor of two or more exist.

6.2 Ocean temperatures

Sea surface temperatures are very close to the SATs of Figs. 3 and 4 and are not plotted separately here.

Fig. 3 Intercomparison results for December–February for surface air temperature (K). The *upper left panel* displays the ensemble model mean (contoured) and the intermodel standard deviation (*shaded*). The *upper right panel* displays zonal averages for each model and for the observations with non-flux-adjusted models as *dashed lines*. The *lower left panel* gives the ensemble mean and the difference from observations for flux-corrected models and the *lower right panel* the same quantities for non-flux corrected models

Latitude-depth cross sections of ocean temperature are displayed in Fig. 6 for DJF. The annual cycle of temperature damps quickly with depth in the ocean so results for other seasons are similar and this also allows the comparison with the observed January-February-March (JFM) seasonal climatology. The model mean and intermodel standard deviation are given in the upper left panel. The models are generally too cold in the first few hundred metres and too warm in the deep ocean. The pattern of excessive warmth is similar for both FA (lower left) and NFA (lower right) models although the difference is larger for NFA models.

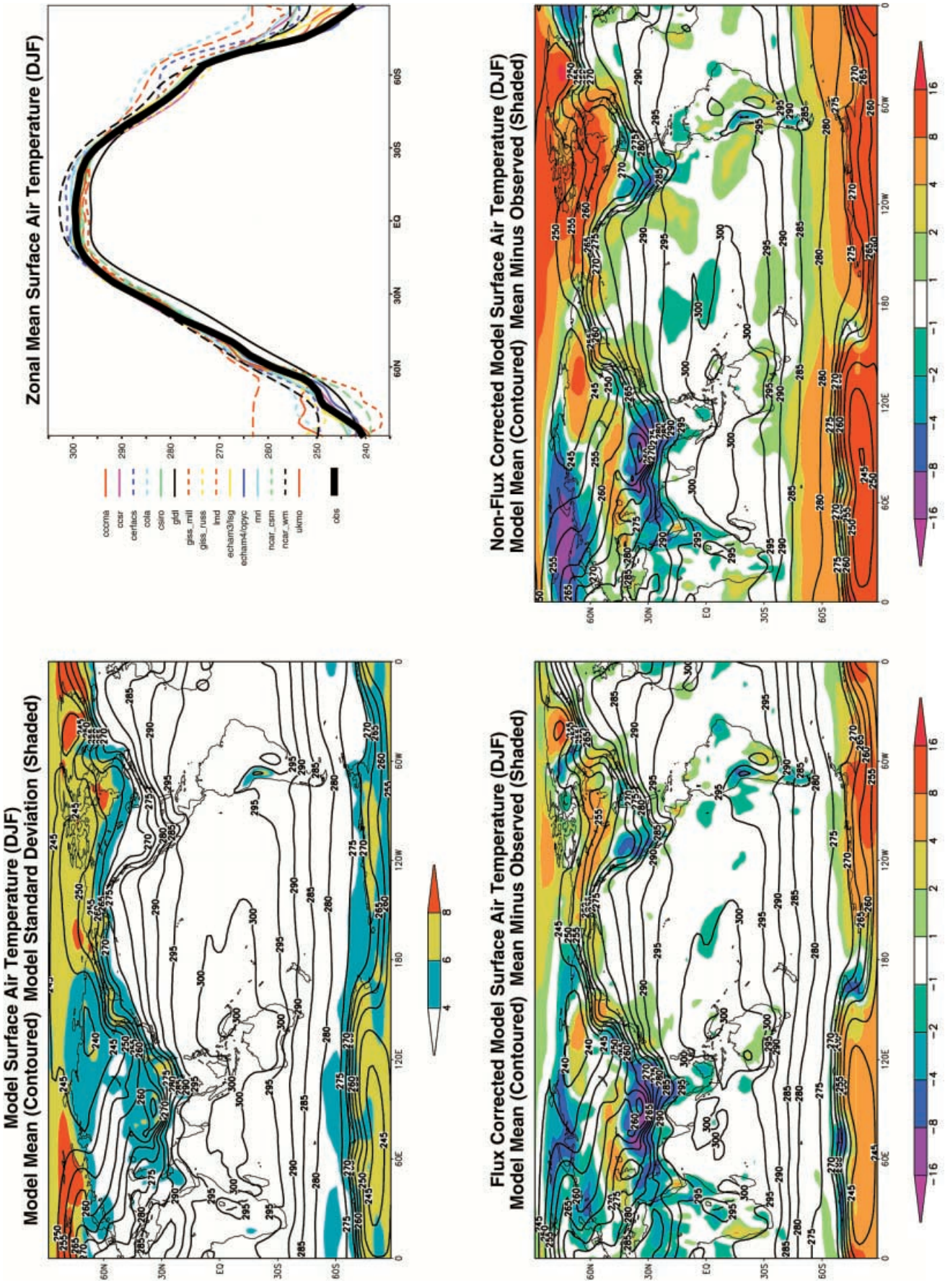
The rather large differences and the scatter among model results is also illustrated in the upper right plot of temperature profiles at 15°S (results are similar over a wide range of latitudes). Although the majority of ocean models share the same structure and numerics, the basic distribution of temperature, nevertheless, varies considerably among models due to differences in surface forcing, differences in internal sub-grid scale mixing, and whether the model is in equilibrium with its forcing.

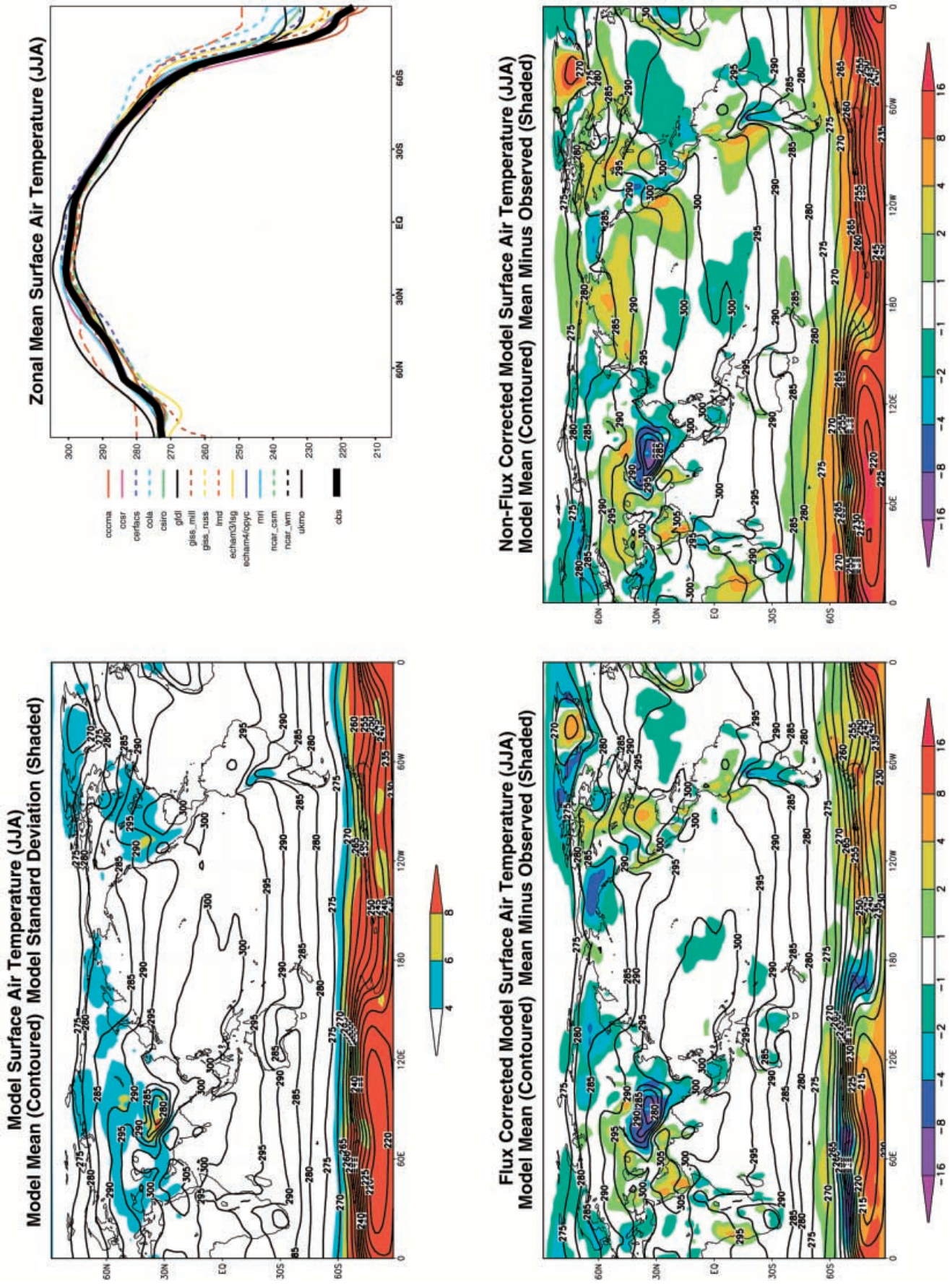
6.3 Precipitation

Precipitation is a climatological variable of immense practical importance and is a critical component of the freshwater flux at the ocean surface. Observational estimates of freshwater flux, however, combine the two poorly observed fields of evaporation and precipitation so are questionable for use in model intercomparison. We concentrate here on precipitation which is displayed in Fig. 7.

Zonal averages of precipitation show considerable variation in the position and the magnitude of the tropical maximum. The middle latitude precipitation maxima tend to be overestimated by most models and this is especially evident in the Northern Hemisphere where there is a region where all the model values are greater than the observational estimate. Previous intercomparison studies of zonally averaged precipitation have shown persistent and characteristic systematic differences similar to those seen here, although the magnitude of the differences have been decreasing over time (Boer 2000).

Large differences in modelled and observed precipitation are seen in equatorial regions where models underestimate precipitation in the South Pacific Convergence Zone and any shift in magnitude and position of the ITCZ has a large influence on the result. The


Zonal Mean Surface Air Temperature (DJF)
ccsm
ccsr
cerfacs
cola
csiro
gfdl
giss_mll
giss_russ
lmd
echam3/lsg
echam4/opyc
mri
ncar_csm
ncar_wm
ukmo
obs
Model Surface Air Temperature (DJF)
Model Mean (Contoured) Model Standard Deviation (Shaded)
Non-Flux Corrected Model Surface Air Temperature (DJF)
Model Mean (Contoured) Mean Minus Observed (Shaded)
Flux Corrected Model Surface Air Temperature (DJF)
Model Mean (Contoured) Mean Minus Observed (Shaded)


Zonal Mean Surface Air Temperature (JJA)
ccma
ccsr
cerfacs
cola
csiro
gfdl
giss_mll
giss_russ
lmd
echam3/lsg
echam4/opyc
mri
ncar_csm
ncar_wm
ukmo
obs
Non-Flux Corrected Model Surface Air Temperature (JJA)
Model Mean (Contoured) Mean Minus Observed (Shaded)
Flux Corrected Model Surface Air Temperature (JJA)
Model Mean (Contoured) Mean Minus Observed (Shaded)
Model Surface Air Temperature (JJA)
Model Mean (Contoured) Model Standard Deviation (Shaded)

◀

Fig. 4 As Fig. 3 except for the June–August season

Asian monsoon region is an area of large differences in JJA. Spatial patterns of precipitation differences for FA and NFA models (not shown) are similar although differences are larger for NFA model results.

6.4 Salinity

The freshwater fluxes at the surface produce effective sources and sinks of salt for the ocean. The SSS for DJF is plotted in Fig. 8 (very similar results for JJA are not shown). Freshwater flux-adjustment directly affects SSS and FA model results are much closer to observed values than are NFA model results. Low salinities occur in the Arctic Ocean north of Siberia which may be associated with river run-off in this area.

Latitude-depth cross sections of salinity for the Atlantic basin are displayed in Fig. 9. The model mean and intermodel standard deviation are given in the upper left and the differences from observations for the FA and the NFA models are given in the lower panels. Overall the models are too salty with large differences seen for NFA models between 500 m and 1000 m. The zonal average at 1000 m (upper right) shows the large differences of individual, especially NFA, model results from observations. Both at the surface and at 1000 m, the models tend to be too salty but this tendency reverses south of about 45°S.

Both SSS and P tend to be high in NFA models indicating that either evaporation is excessive or the models are not in equilibrium and that salt is being exchanged with the deeper ocean. The simulation of salinity with depth remains a challenge for this generation of NFA coupled models.

6.5 Mean sea-level pressure

Results for MSLP are displayed in Fig. 10. Extrapolation of pressure over high terrain to mean sea-level compromises the comparison between models and observations in those regions. The zonal mean plots show that many of the model results lie close to the observed values in tropical and middle attitudes. Differences are larger at higher latitudes and this is especially noticeable in the Antarctic trough and over Antarctica. The intercomparison history of zonal MSLP shows that these persistent and characteristic differences remain, although slow progress has been made (Boer 2000).

The geographic plots of MSLP show that the semi-permanent features of the observed climatology such as the Aleutian Low, the Icelandic Low, and the Siberian High are simulated with roughly the correct intensities and positions. The models are less successful in simulating the depth and position of the Antarctic trough. There is an apparent northward displacement of the

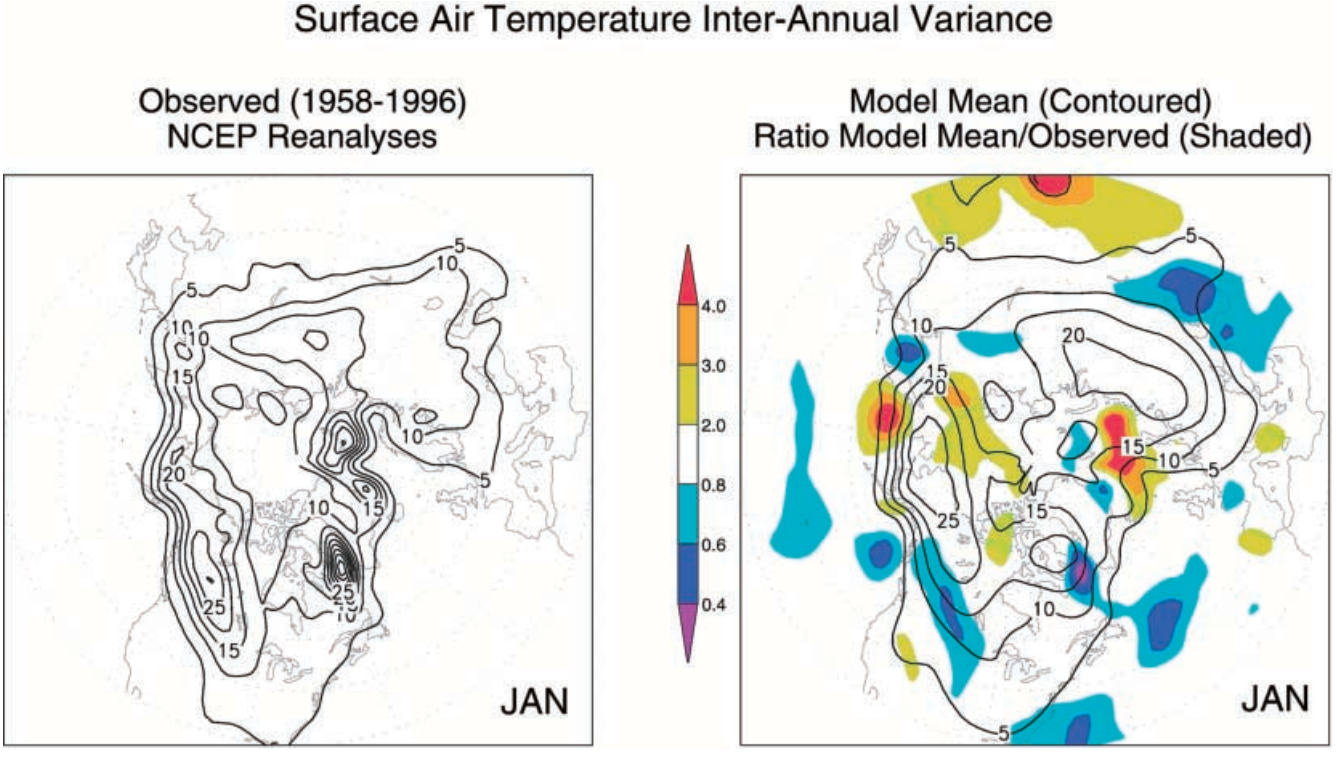
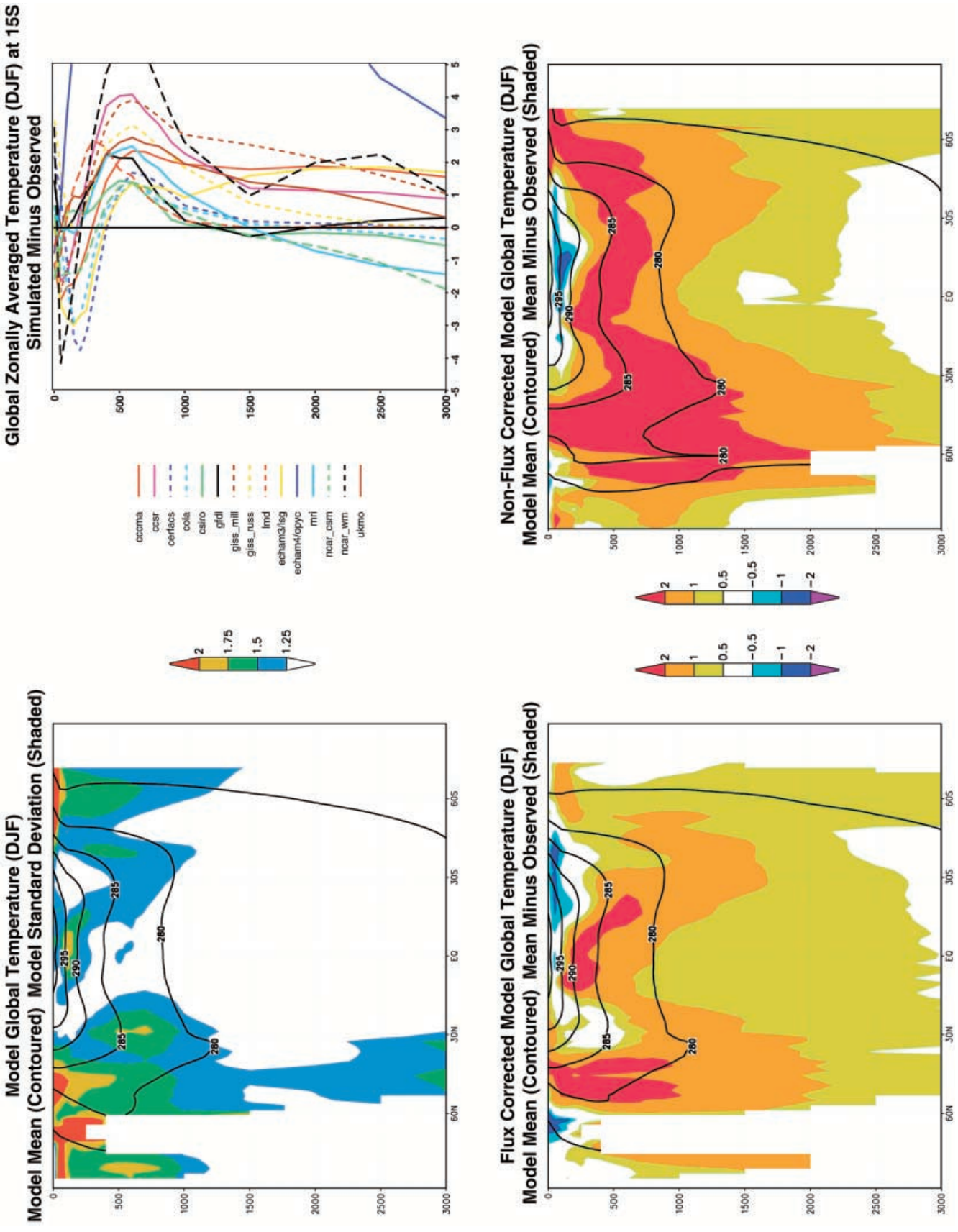


Fig. 5 January inter-annual surface air temperature variance. The *left panel* gives the observed field (C^2) and the *right panel* displays the model ensemble variance (contoured) and the ratio with the observations (*shaded*)


Model Global Temperature (DJF)
Model Mean (Contoured) Model Standard Deviation (Shaded)
Global Zonally Averaged Temperature (DJF) at 15S
Simulated Minus Observed
Non-Flux Corrected Model Global Temperature (DJF)
Model Mean (Contoured) Mean Minus Observed (Shaded)
Flux Corrected Model Global Temperature (DJF)
Model Mean (Contoured) Mean Minus Observed (Shaded)
cccma
ccsr
cerfacs
cola
csiro
gfdl
giss_mill
giss_russ
lmd
echam3/lsg
echam4/opyc
mri
ncar_csm
ncar_wm
ukmo
60N
30N
EQ
30S
60S

◄

Fig. 6 Latitude-depth cross-sections of ocean temperature (K) for December–February. The *upper left panel* displays the ensemble model mean (contoured) and the intermodel standard deviation (*shaded*). The *upper right panel* displays profiles of model values minus observed values at 15°S. The *lower left panel* gives the mean and difference from observed for flux-adjusted models and the *lower right panel* that for non-flux-adjusted models

Pacific storm track and an extension into the continent. The storm track over northern Scandinavia is displaced to the south. These results are consistent with SAT results which are too warm in the northern Pacific, Alaska, western Canada, and Europe and too cold in northern Scandinavia. Differences for JJA are typically of opposite sign in the Northern Hemisphere, but show little seasonal change in the south. Although not shown, means of FA and NFA results show that FA results are generally closer to observations and this is especially true for the position and depth of the Antarctic trough.

The MSLP is linked through geostrophy to surface winds and thence to the surface momentum flux between atmosphere and ocean. We do not compare model wind stress values with observation-based estimates but a sense of the differences can be inferred from the MSLP distributions.

6.6 Ocean circulation

Comprehensive observations of the ocean circulation are not readily available rendering evaluation of the modelled ocean circulations difficult. Two broad measures of the ocean circulation are tabulated in Table 8. These are the maximum of the overturning stream function in the North Atlantic and the magnitude of the barotropic stream function in Drake Passage. Observation-based estimates are also included. The overturning stream function is linked to poleward heat transport in the Atlantic and to the ocean "conveyer belt" which transports heat and salt throughout the ocean. The connection between changes in this circulation and climate change is stressed by many authors (IPCC 1995). Model results range from basically no overturning at all to a maximum of nearly 40 Sv.

The latitude-depth *mean model* Atlantic baroclinic overturning stream function is shown in Fig. 11 together with individual model and model mean profiles. The profiles show large differences among both FA and NFA results. These dramatic differences for the Atlantic overturning stream function indicate considerable uncertainty in the simulation of this quantity. Individual cross sections of the global stream function (not shown) also display large differences in the Southern Hemisphere.

The *mean model* barotropic stream function and the intermodel standard deviation are plotted in Fig. 12. Global observational estimates are not available for this quantity. The broad features of the flow in the Northern Hemisphere are more or less as expected, although perhaps somewhat weak in the Atlantic. The much larger values in the Southern Hemisphere are accompanied by large intermodel standard deviations, which are of the order of the mean itself, over much of the Indian Ocean and the adjacent Pacific and particularly the high latitude southern ocean. The values of the stream function for Drake Passage given in Table 8 indicate that there is considerable variability in the model simulations of this quantity; ranging from near zero to approximately twice the observational estimate. It is clear that results differ importantly in this generation of models.

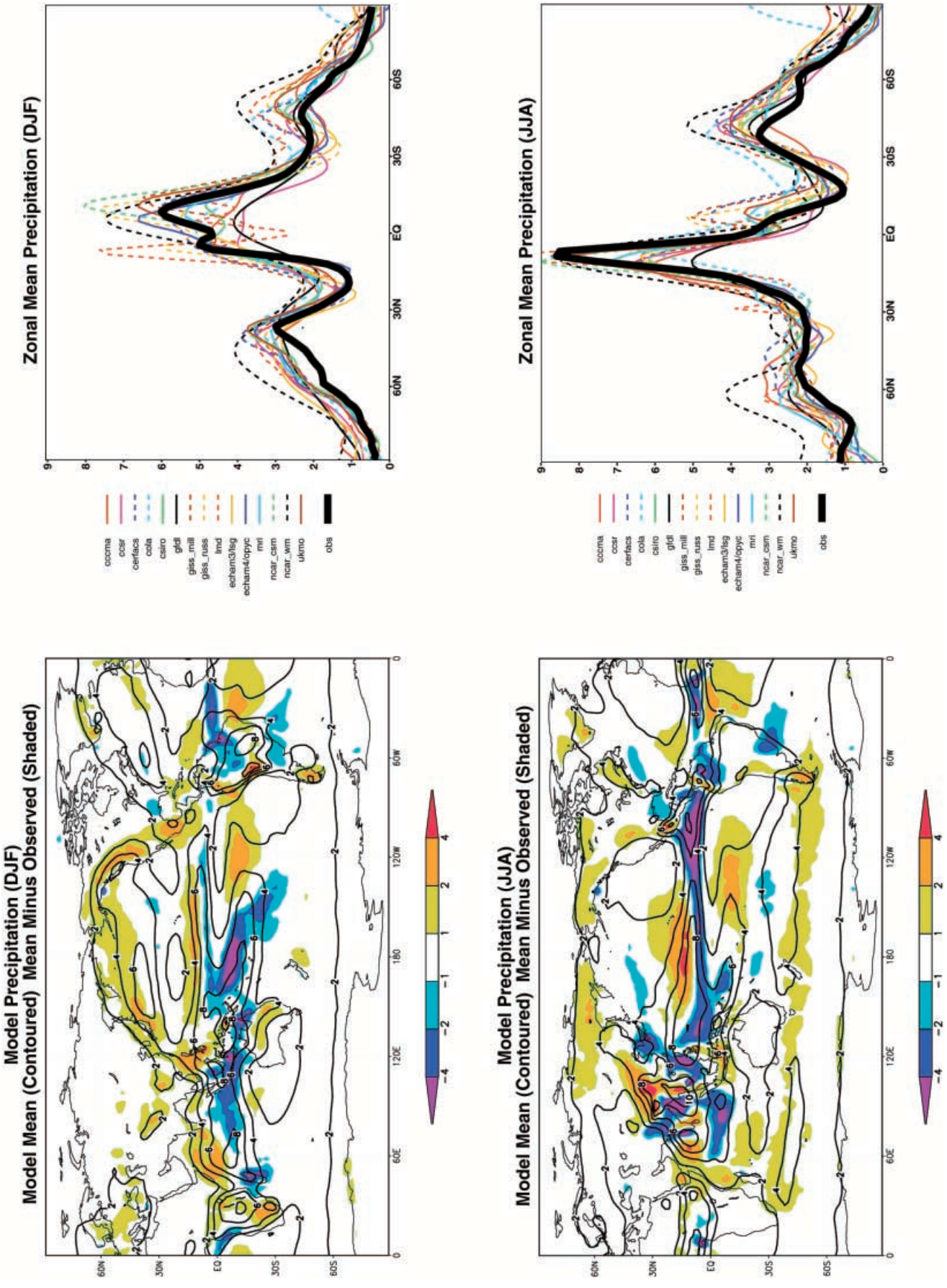
6.7 Ocean heat flux and heat transport

The energetic connection between the atmosphere and ocean is by means of the heat flux plotted in Fig. 13. (These do not include the flux adjustment terms.) The first order heat balance in the ocean, the gain of heat in the summer and loss of heat in the winter, is clearly captured by the models. Fluxes are large, however, and differences between modelled and observed estimates are also large, exceeding 100 Wm^{-2} in some regions. The patterns of surface heat flux differences from observation-based estimates tend to change sign with season. Values are large along the western coast of the Americas where a lack of stratus clouds is a well-known systematic error of models. The difficulty in modelling the transport and structure of the Gulf Stream also has a signature in the surface heat flux.

Surface heat flux is adjusted in FA models in order to maintain observed temperatures, rather than to reproduce estimated heat flux values, and so adjustment may degrade rather than enhance the agreement. While the meridional structure of the heat flux is relatively well captured by the models, the remaining geographical pattern is not, as previously noted in conjunction with Table 6.

The north-south transport of heat by the ocean responds to and balances the annual average flux across the surface which results after averaging over the very large seasonal cycle of gain and loss at the surface. The annual average is much smaller than the summer and winter values of Fig. 13 and correspondingly more difficult to simulate correctly. The annual average gain of heat in the tropics is transported poleward in both hemispheres where it lost from the ocean at high latitudes.

The oceanic heat transport is shown in Fig. 14 for FA and NFA models. In the Northern Hemisphere, the FA models simulate only about one-half the maximum observed transport with relatively little scatter among results. The NFA models simulate about three-quarters of the observed maximum transport on average, with a larger inter-model variability. This situation is essentially reversed in the Southern Hemisphere. Ocean transports which differ from the actual values imply some combination of erroneous surface fluxes and/or trans-


Zonal Mean Precipitation (DJF)
Zonal Mean Precipitation (JJA)
cccma
ccsr
cerfacs
cola
csiro
gfdl
giss_mill
giss_russ
lmd
echam3/lsg
echam4/opyc
mri
ncar_csm
ncar_wm
ukmo
obs
60N
30N
EQ
30S
60S
Model Precipitation (DJF)
Model Mean (Contoured) Mean Minus Observed (Shaded)
Model Precipitation (JJA)
Model Mean (Contoured) Mean Minus Observed (Shaded)
60E
120E
180
120W
60W
0
-4
-2
-1
1
2
4

◄ ►

Fig. 7 *Left panels* give the model ensemble mean (contoured) precipitation rate (mm da^{-1}) and the difference from observed (*shaded*) for the December–February and June–August seasons. *Right panels* give the zonal average precipitation rates for individual models and for the observations. Flux adjusted model results are *dashed*

Fig. 8 As Fig. 3 but for sea surface salinity (psu)

Fig. 9 As Fig. 6 but for the salinity of the Atlantic ocean (psu). The *upper right panel* is the zonal average at 1000 m

Fig. 10 As Fig. 7 except for mean sea-level pressure (hPa)

port in the ocean model and may be associated with large errors in surface temperatures and other parameters.

The correct simulation of the distribution of surface heat flux and the associated poleward transport of heat in the ocean is another major challenge to coupled modelers. Higher resolution in ocean models may help produce more vigorous ocean transport but is not in itself sufficient since some of the largest differences in basic climatological quantities are found for models with high ocean resolution. As is so often the case in modelling a coupled system like climate, balanced improvement of all aspects of atmospheric and oceanic models is required to attain modest evolutionary improvement.

7 Summary

The evaluation of global models used for day-to-day weather forecasting is a relatively simple matter of comparing model predictions to observations. The evaluation of coupled global climate models, used for the prediction of climate change, cannot be undertaken in the same way. Consequently, the reliability of climate models is assessed by examining their ability to simulate the current climate, by intercomparing the future climates simulated by groups of models, and by the simulation of the climate of the past century for which there is an observational record. The first of these approaches is used here to evaluate a group of CGCMs as part of CMIP1, the first phase of the coupled model intercomparison project.

The global distributions of a range of basic climate variables are compared among models and with observation-based estimates of these quantities. Second order difference statistics are calculated and displayed for global means, meridional structures and the remaining geographical patterns. Zonal averages, zonal cross sections, and global maps of these quantities and the difference between modelled and observed values are displayed.

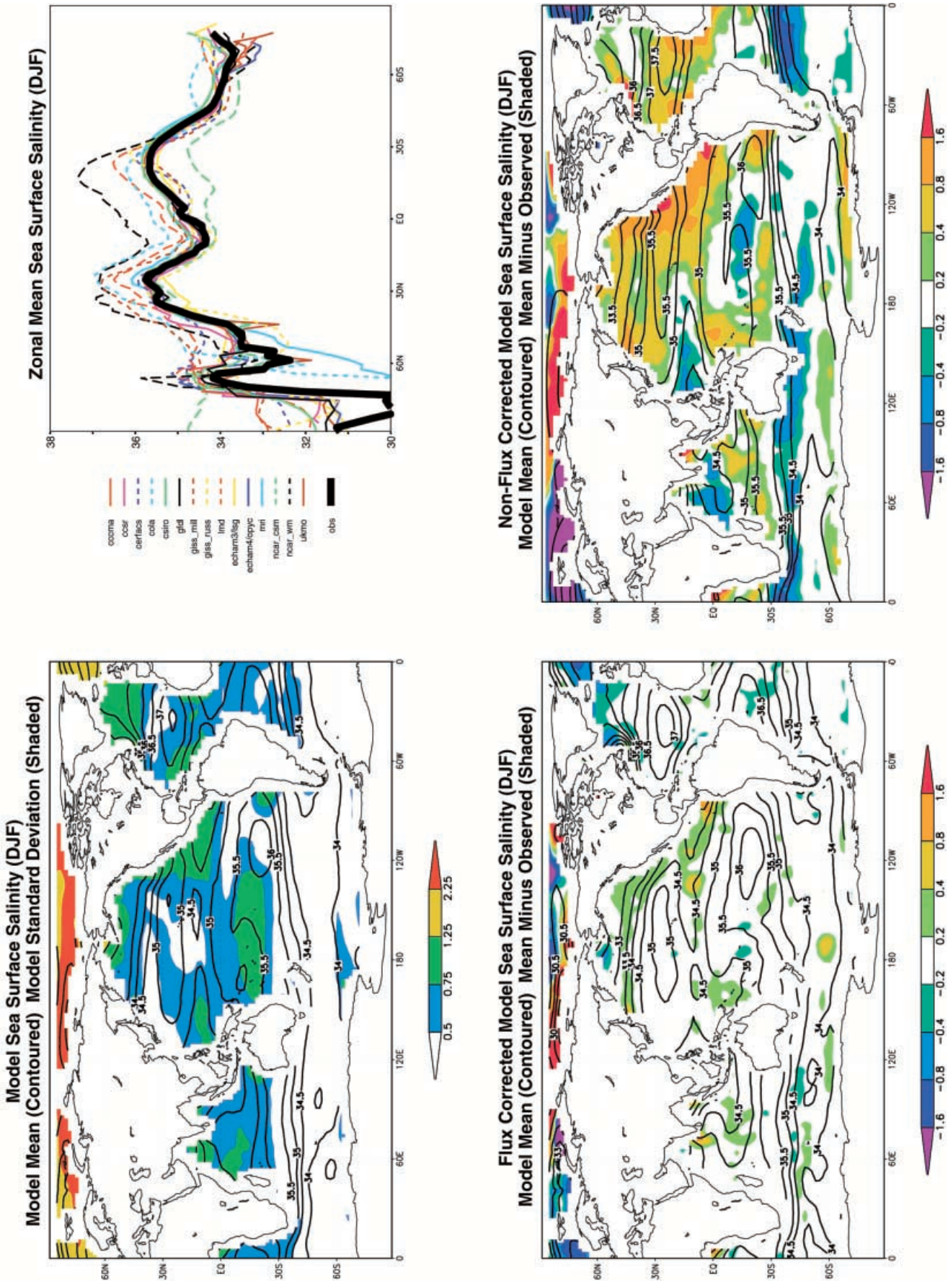
Results may be summarized briefly as follows: (1) the current generation of climate models, on average, reproduce the major features of the observed distribution of the basic climate parameters; (2) there is, nevertheless, a considerable scatter between model results and between simulated and observed values; (3) this is particularly true of oceanic variables; (4) flux-adjusted models generally produce simulated climates which are in better accord with observations than do non-flux-adjusted models (as, of course, is the intent of flux adjustment); (5) some non-flux-adjusted model results are, nevertheless, closer to the observations than some flux-adjusted model results; (6) other model differences, such as resolution, do not appear to provide a clear distinction among model results in this generation of models; (7) systematic differences; i.e., differences that are common to most models, are apparent; (8) some of these systematic differences, i.e., for precipitation and mean sea-level pressure, have been known for some time and show a slow improvement with model evolution; (9) as is characteristic of intercomparison results, different climate variables are simulated with different levels of success by different models and no one model is "best" for all variables; and (10) there is evidence that the *mean model* result, obtained by averaging over the ensemble of model results, provides an overall "best" comparison to observations for climatological mean fields.

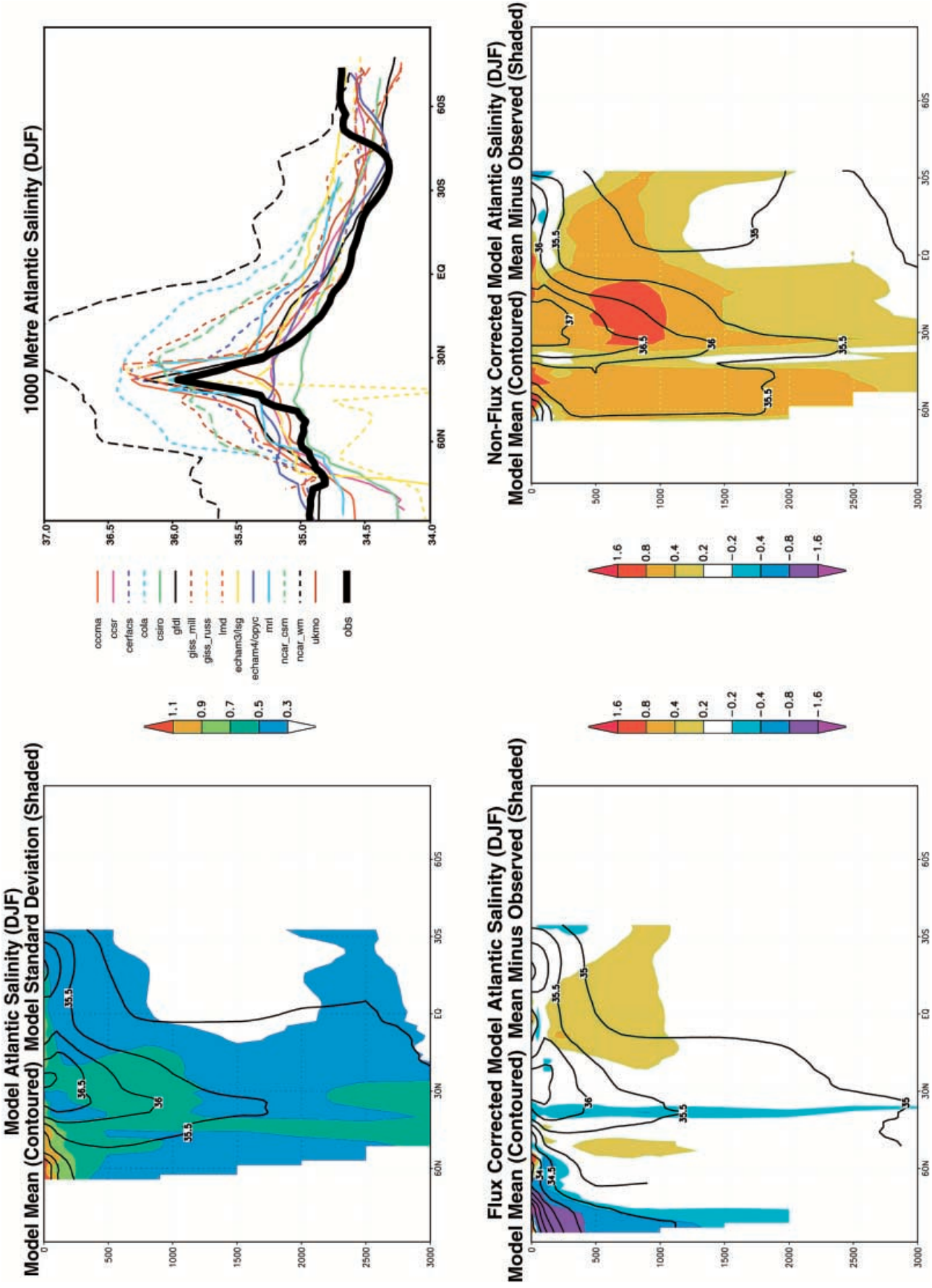
Although flux adjusted models generally display better agreement with observations, the overall progress achieved in modelling the climate system from first principles is, perhaps, best reflected by non-flux-adjusted results which do not benefit by having their deficiencies masked by the adjustment terms. The new generation of non-flux-adjusted models under development will, however, likely initially accept somewhat larger errors in surface and deep ocean quantities than the current flux-adjusted versions of the models.

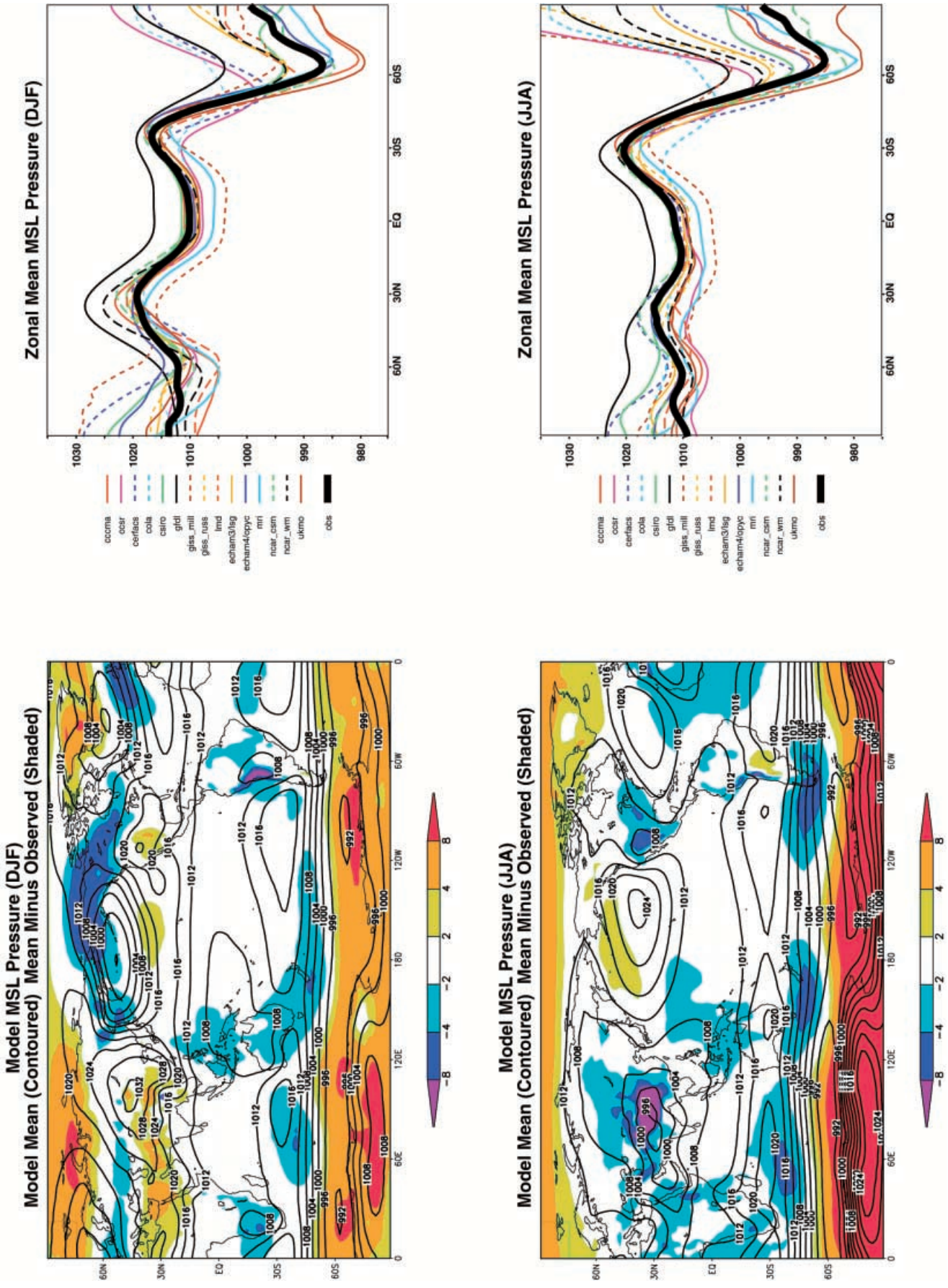
The model deficiencies identified here do not suggest immediate remedies. The overall success of the models in simulating the behaviour of the complex non-linear climate system apparently depends on the slow improve-

Table 8 The maximum North Atlantic baroclinic overturning stream function (Sv) and the value of the barotropic stream function in Drake Passage (60S 56W) for the individual models and the *mean model*. The average and intermodel standard deviation and estimates from observations are also given

Model	Stream function	
	Baroclinic	Barotropic
CCCma	24	61.3
CCSR	28	186.4
CERFACS	24	56.6
COLA	18	8.4
CSIRO	16	91.9
GFDL	17	61.9
GISS_M	21	67.2
GISS_R	10	
LMD	15	60.5
MPI_E3/L	30	64.6
MPI_E4/O	24	86.0
MRI	~0	34.2
NCAR_CSM	32	229.1
NCAR_WM	39	83.0
UKMO	19	195.7
Mean model	17	80
Average (standard deviation)	21.1 (9.1)	91.9 (62.3)
Observation-estimate	18–27	130–140


Zonal Mean Sea Surface Salinity (DJF)
38
36
34
32
30
60N
30N
EQ
30S
60S
cccma
ccsr
cerfacs
cola
csiro
gfdl
gias_mll
gias_russ
lmd
echam3/lag
echam4/opyc
mri
ncar_ccm
ncar_wm
ukmo
obs
Model Sea Surface Salinity (DJF)
Model Mean (Contoured) Model Standard Deviation (Shaded)
0.5
0.75
1.25
2.25
Non-Flux Corrected Model Sea Surface Salinity (DJF)
Model Mean (Contoured) Mean Minus Observed (Shaded)
Flux Corrected Model Sea Surface Salinity (DJF)
Model Mean (Contoured) Mean Minus Observed (Shaded)
−1.6
−0.8
−0.4
−0.2
0.2
0.4
0.8
1.6
0
60E
120E
180
120W
60W


1000 Metre Atlantic Salinity (DJF)
37.0
36.5
36.0
35.5
35.0
34.5
34.0
60N
30N
EQ
30S
60S
occma
ocsr
cerfacs
coda
csiro
gfdl
giss_mill
giss_russ
lmd
echam3/lsg
echam4/opyc
mri
ncar_csm
ncar_wm
ukmo
obs
Model Atlantic Salinity (DJF)
Model Mean (Contoured) Model Standard Deviation (Shaded)
Non-Flux Corrected Model Atlantic Salinity (DJF)
Model Mean (Contoured) Mean Minus Observed (Shaded)
Flux Corrected Model Atlantic Salinity (DJF)
Model Mean (Contoured) Mean Minus Observed (Shaded)
1.6
0.8
0.4
0.2
−0.2
−0.4
−0.8
−1.6
1.1
0.9
0.7
0.5
0.3
0
500
1000
1500
2000
2500
3000


Zonal Mean MSL Pressure (DJF)
Zonal Mean MSL Pressure (JJA)
Model MSL Pressure (DJF)
Model Mean (Contoured) Mean Minus Observed (Shaded)
Model MSL Pressure (JJA)
Model Mean (Contoured) Mean Minus Observed (Shaded)
ccma
ccsr
cerfacs
cola
csiro
gfdl
giss_mll
giss_russ
lmd
echam3/lsg
echam4/oppyc
mri
ncar_csm
ncar_wm
ukmo
obs
60N
30N
EQ
30S
60S
1030
1020
1010
1000
990
980
-8
-4
-2
2
4
8

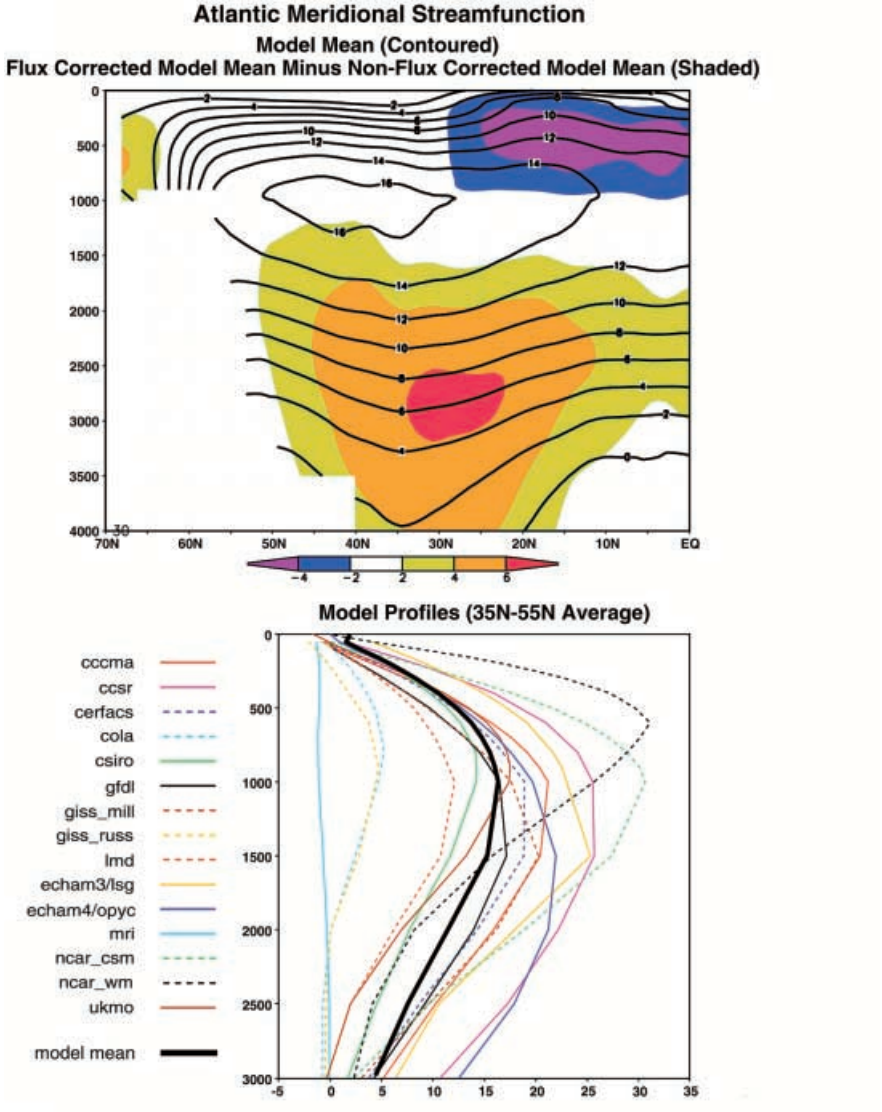


Fig. 11 The ensemble mean Atlantic meridional overturning stream function (Sv) and the difference between flux-adjusted and non-flux-adjusted model results (*shaded*) and individual model profiles for 35–55°N

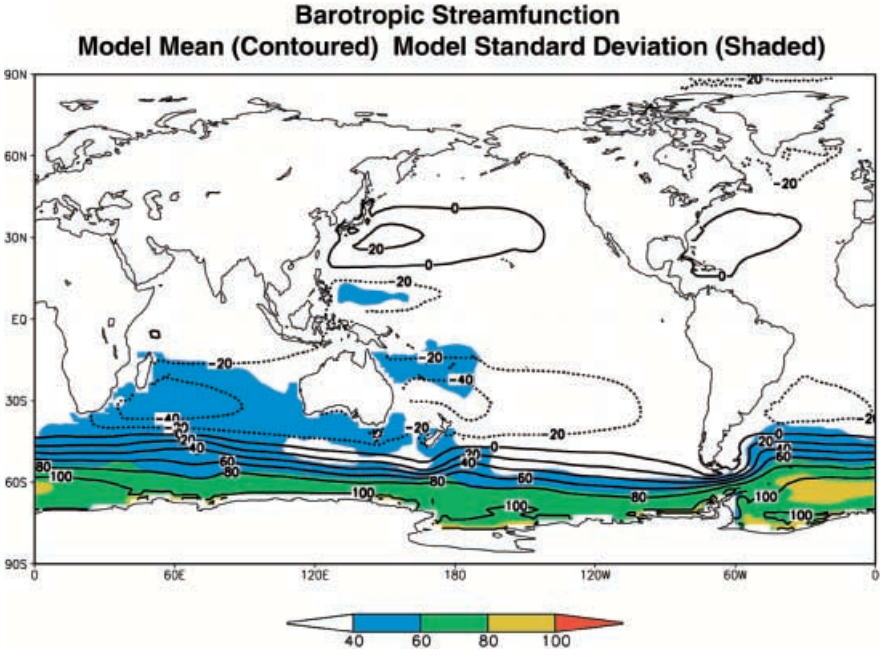
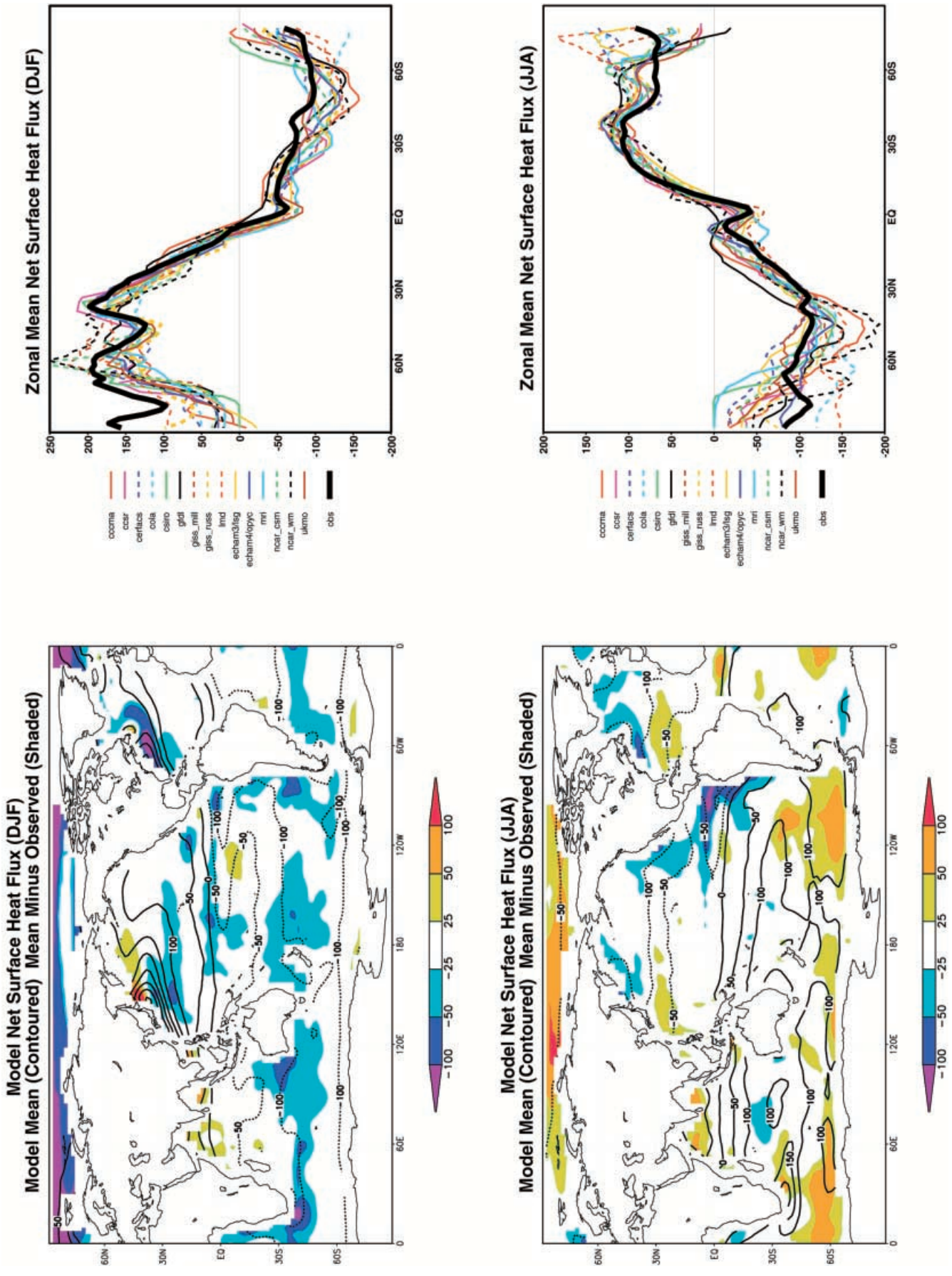


Fig. 12 Ensemble mean barotropic stream function (Sv) and intermodel standard deviation (*shaded*)

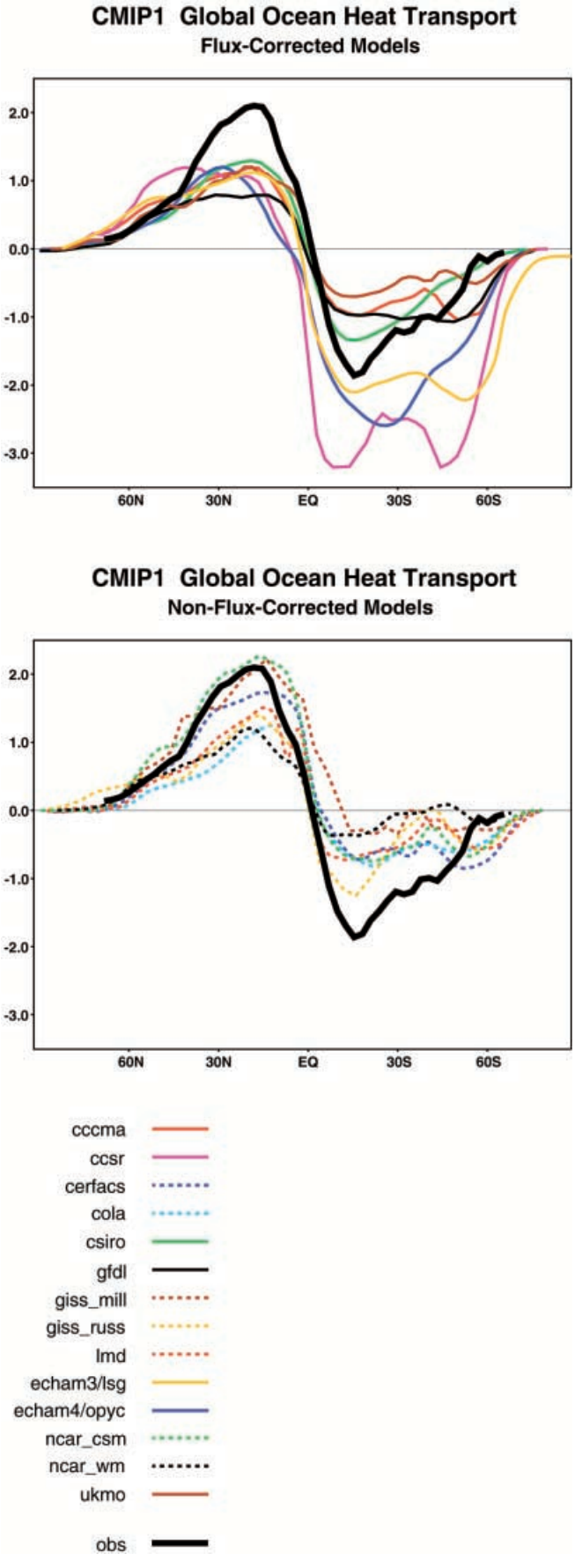

Zonal Mean Net Surface Heat Flux (DJF)
Zonal Mean Net Surface Heat Flux (JJA)
Model Net Surface Heat Flux (DJF)
Model Mean (Contoured) Mean Minus Observed (Shaded)
Model Net Surface Heat Flux (JJA)
Model Mean (Contoured) Mean Minus Observed (Shaded)
cccma
ccsr
cerfacs
cola
csiro
gfdl
giss_mll
giss_russ
lmd
echam3/lsg
echam4/opyc
mri
ncar_csm
ncar_wm
ukmo
obs
60N
30N
EQ
30S
60S
60E
120E
180
120W
60W
0

◀

Fig. 13 As Fig. 7 but for surface net heat flux (W m^{-2})

▶

Fig. 14 Observed and simulated annual ocean heat transport (PW). The *upper panel* gives the results for the flux-corrected models and observations and the *lower panel* gives the results for the non-flux corrected models and observations



ment in the balance of approximations that characterize a coupled climate model. Of course, the results of this and similar studies provide only an indication, at a particular time, of the current state and the moderate but steady evolution and improvement of coupled climate models.

References

Abe-Ouchi A, Yamanaka Y, Kimoto M (1996) Outline of the CCSR coupled atmosphere and ocean model and experiment. Internal Report, Centre for Climate System Research, University of Tokyo

Boer GJ (2000) Climate model intercomparison. In: Numerical modelling of the global atmosphere in the climate system. Kluwer, Dordrecht, The Netherlands

Boer GJ, Lambert SJ (2000) Second order space-time climate difference statistics. Climate Dym 17: 213–218

Boer GJ, Arpe K, Blackburn M, Deque M, Gates WL, Hart TL, Le Treut H, Roeckner E, Sheinin DA, Simmonds I, Smith RNB, Tokioka T, Wetherald RT, Williamson D (1992) Some results from an intercomparison of the climates simulated by 14 atmospheric general circulation models. J Geophys Res 97: 12 771–12 786

Boer GJ, Flato G, Ramsden D (2000) A transient climate change simulation with greenhouse gas and aerosol forcing: experimental design and comparison with the instrumental record for the 20th century. Clim Dyn 16: 427–450

Boville BA, Gent PR (1998) The NCAR Climate System Model, Version One. J Clim 11: 1115–1130

Branconnet P, Marti O, Joussaume S (1997) Adjustments and feedbacks in a global coupled ocean-atmosphere model. Clim Dyn 13: 507–519

Covey C (1998) CMIP1 model output. At http://www-pcmdi.llnl.gov/cmip/diagsub.html#CMIP1 model output. Program for Climate Model Diagnosis and Intercomparison (PCMDI), the Lawrence Livermore National Laboratory, Livermore, California

Covey CC, Abe-Ouchi A, Boer GJ, Flato GM, Boville BA, Meehl GA, Cubasch U, Roeckner E, Gordon H, Guilyardi E, Terray L, Jiang X, Miller R, Russell G, Johns TC, Le Treut H, Fairhead L, Madec G, Noda A, Power SB, Schneider EK, Stouffer RJ, von Storch J-S (1999) The seasonal cycle in coupled ocean-atmosphere general circulation models. Program for Climate Model Diagnosis and Intercomparison Report Series, Report 51, 41 pp., Livermore, USA

Da Silva AM, Young CC, Levitus S (1994) Atlas of surface marine data 1994, volume 1: algorithms and procedures. NOAA Atlas NESDIS 6, US Department of Commerce, Washington, DC

Fichefet T (1997) European coupled climate models: IPSL. Model documentation compiled for Euroclivar. Accessible on the World Wide Web at http://www. knmi.nl/euroclivar/ipsl.html

Flato GM, Boer GJ, Lee WG, McFarlane NA, Ramsden D, Reader MC, Weaver AJ (2000) The Canadian Centre for Climate Modelling and Analysis Global Coupled Model and its climate, Clim Dyn 16: 451–467

Gates WL, Cubash U, Meehl GA, Mitchell JFB, Stouffer RJ (1993) An intercomparison of selected features of the control climates simulated by coupled ocean-atmosphere general circulation

models. World Climate Research Programme, WCRP-82. WMO/TD No. 574, WMO, Geneva
Gates WL, Boyle JS, Covey C, Dease CG, Doutriaux CM, Drach RS, Fiorino M, Gleckler PJ, Hnilo JJ, Marlais SM, Phillips TJ, Potter GL, Santer BD, Sperber KR, Taylor KE, Williams DN (1999) An overview of the results of the Atmospheric Model Intercomparison Project (AMIP). Bull Am Meteorol Soc 80: 29–56
Gibson JK, Kallberg P, Uppala S, Hernandez A, Nomura A, Serrano E (1997) ECMWF Re-Analysis Project Report No. 1. ERA Description. ECMWF, Shinfield Park, Reading, UK
Gordon HB, O'Farrell SP (1997) Transient climate change in the CSIRO coupled model with dynamic sea ice. Mon Weather Rev 125: 875–907
Guilyardi E, Madec G (1997) Performance of the OPA/ARPEGE-T21 global ocean-atmosphere coupled model. Clim Dyn 13: 149–165
Hirst AC, O'Farrell SP, Gordon HB (2000) Comparison of a coupled ocean-atmosphere model with and without oceanic eddy-induced advection. I. Ocean spin-up and control integrations. J Clim 13: 139–163
IPCC (1990) Climate change: the IPCC scientific assessment. Houghton JT, Jenkins GJ, Ephraums JJ (eds) Cambridge University Press, Cambridge, UK
IPCC (1995) Climate change 1995: the science of climate change. Houghton JT, Meira Filho LG, Callender BA, Harris N, Kattenberg A, Maskell K (eds) Cambridge University Press, Cambridge, UK
Jenne RL (1975) Data sets for meteorological research. NCAR Technical Note, NCAR-TN/1A, National Center for Atmospheric Research, Boulder, USA
Johns TC (1996) A description of the Second Hadley Centre Coupled Model (HadCM2). Climate Research Technical Note 71, Hadley Centre, United Kingdom Meteorological Office, Bracknell Berkshire RG12 2SY, UK
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
Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Ropelewski C, Wang J, Leetmaa A, Reynolds R, Jenne R, Joseph D (1996) The NCEP/NCAR 40-year reanalysis project. Bull Am Meteorol Soc 77: 437–471
Lambert SJ, Boer GJ (1989) Atmosphere-ocean fluxes and stresses in general circulation models. Atmosphere-Ocean 27: 692–715
Levitus S, Boyer TP (1994) World Ocean Atlas 1994 vol 4: temperature. NOAA Atlas NESDIS 4
Levitus S, Burgett R, Boyer TP (1994) World Ocean Atlas 1994 vol 3: salinity. NOAA Atlas NESDIS 3
Macdonald AM, Wunsch C (1996) An estimate of global ocean circulation and heat fluxes. Nature 382: 436–439
Manabe S, Stouffer RJ (1996) Low-frequency variability of surface air temperature in a 1000-year integration of a coupled atmosphere-ocean-land surface model. J Clim 9: 376–393
Manabe S, Stouffer RJ, Spelman MJ, Bryan K (1991) Transient responses of a coupled ocean-atmosphere model to gradual changes of atmospheric CO_2. Part I: annual mean response. J Clim 4: 785–818
Meehl GA, Washington WM (1995) Cloud albedo feedback and the super greenhouse effect in a global coupled GCM. Clim Dyn 11: 299–411
Meehl GA, Boer GJ, Covey C, Latif M, Stouffer RJ (1997) Intercomparison makes for a better climate model. Eos, Trans Am Geophys Union 78: 445–446, 451
Miller RL, Jiang X (1996) Surface energy fluxes and coupled variability in the tropics of a coupled general circulation model. J Clim 9: 1599–1620
Pacanowski RC (1995) MOM 2 documentation: users guide and reference manual, version 1.0. GFDL Ocean Group Technical Report 3, Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey
Phillips T (1998) Summary documentation: CMIP I model features and experimental implementation (version 1.0). At http://www-pcmdi.llnl.gov/modeldoc/cmip/index.html. Program for Climate Model Diagnosis and Intercomparison (PCMDI), Lawrence Livermore National Laboratory, Livermore, California
Rayner NA, Horton EB, Parker DE, Folland CK, Hackett RB (1996) Version 2.2 of the global sea-ice and sea surface temperature dataset, 1903–1994. CRTN 74, Hadley Centre, Meteorological Office, Bracknell, United Kingdom
Roeckner E, Oberhuber JM, Bacher A, Christoph M, Kirchner I (1996) ENSO variability and atmospheric response in a global coupled atmosphere-ocean GCM. Clim Dyn 12: 737–754
Russell GL, Miller JR, Rind D (1995) A coupled atmosphere-ocean model for transient climate change studies. Atmos-Ocean 33: 683–730
Schneider EK, Zhu Z (1998) Sensitivity of the simulated annual cycle of sea surface temperature in the equatorial Pacific to sunlight penetration. J Clim 11: 1932–1950
Schneider EK, Zhu Z, Giese BS, Huang B, Kirtman BP, Shukla J, Carton JA (1997) Annual cycle and ENSO in a coupled ocean-atmosphere general circulation model. Mon Weather Rev 125: 680–702
Schmitz WJ (1995) On the interbasin-scale thermohaline circulation. Rev Geophys 33: 151–173
Tokioka T, Noda A, Kitoh A, Nikaidou Y, Nakagawa S, Motoi T, Yukimoto S, Takata K (1996) A transient CO_2 experiment with the MRI CGCM: annual mean response. CGER's Supercomputer Monograph Report vol 2, CGER-IO22-96, ISSN 1341–4356, Centre for Global Environmental Research, National Institute for Environmental Studies, Environment Agency of Japan, Ibaraki, Japan
Trenberth KE, Solomon A (1994) The global heat balance: heat transport in the atmosphere and ocean. Clim Dyn 10: 107–134
Voss R, Sausen R, Cubasch U (1998) Periodically synchronously coupled integrations with the atmosphere-ocean general circulation model ECHAM3/LSG. Clim Dyn 14: 249–266
Washington WM, Meehl GA (1996) High-latitude climate change in a global coupled ocean-atmosphere-sea ice model with increased atmospheric CO_2. J Geophys Res 101(D8): 12 795–12 801
Whitworth T, Petersen RG (1985) Volume transport of the Antarctic circumpolar current from bottom pressure measurements. J Phys Oceanogr 15: 810–816
Whitworth T, Nowlin WD, Worley SJ (1982) The net transport of the Antarctic circumpolar current through Drake passage. J Phys Oceanogr 12: 960–971
Xie P, Arkin A (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