An Overview of the Results of the Atmospheric Model Intercomparison Project (AMIP I)

W. Lawrence Gates, James S. Boyle, Curt Covey, Clyde G. Dease, Charles M. Doutriaux, Robert S. Drach, Michael Fiorino, Peter J. Gleckler, Justin J. Hnilo, Susan M. Marlais, Thomas J. Phillips, Gerald L. Potter, Benjamin D. Santer, Kenneth R. Sperber, Karl E. Taylor, and Dean N. Williams
Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, Livermore, California

ABSTRACT

The Atmospheric Model Intercomparison Project (AMIP), initiated in 1989 under the auspices of the World Climate Research Programme, undertook the systematic validation, diagnosis, and intercomparison of the performance of atmospheric general circulation models. For this purpose all models were required to simulate the evolution of the climate during the decade 1979–88, subject to the observed monthly average temperature and sea ice and a common prescribed atmospheric CO_2 concentration and solar constant. By 1995, 31 modeling groups, representing virtually the entire international atmospheric modeling community, had contributed the required standard output of the monthly means of selected statistics. These data have been analyzed by the participating modeling groups, by the Program for Climate Model Diagnosis and Intercomparison, and by the more than two dozen AMIP diagnostic subprojects that have been established to examine specific aspects of the models' performance. Here the analysis and validation of the AMIP results as a whole are summarized in order to document the overall performance of atmospheric general circulation–climate models as of the early 1990s. The infrastructure and plans for continuation of the AMIP project are also reported on.

Although there are apparent model outliers in each simulated variable examined, validation of the AMIP models' ensemble mean shows that the average large-scale seasonal distributions of pressure, temperature, and circulation are reasonably close to what are believed to be the best observational estimates available. The large-scale structure of the ensemble mean precipitation and ocean surface heat flux also resemble the observed estimates but show particularly large intermodel differences in low latitudes. The total cloudiness, on the other hand, is rather poorly simulated, especially in the Southern Hemisphere. The models' simulation of the seasonal cycle (as represented by the amplitude and phase of the first annual harmonic of sea level pressure) closely resembles the observed variation in almost all regions. The ensemble's simulation of the interannual variability of sea level pressure in the tropical Pacific is reasonably close to that observed (except for its underestimate of the amplitude of major El Niños), while the interannual variability is less well simulated in midlatitudes. When analyzed in terms of the variability of the evolution of their combined space–time patterns in comparison to observations, the AMIP models are seen to exhibit a wide range of accuracy, with no single model performing best in all respects considered.

Analysis of the subset of the original AMIP models for which revised versions have subsequently been used to revisit the experiment shows a substantial reduction of the models' systematic errors in simulating cloudiness but only a slight reduction of the mean seasonal errors of most other variables. In order to understand better the nature of these errors and to accelerate the rate of model improvement, an expanded and continuing project (AMIP II) is being undertaken in which analysis and intercomparison will address a wider range of variables and processes, using an improved diagnostic and experimental infrastructure.

Corresponding author address: Dr. Lawrence Gates, Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, P.O. Box 808, Livermore, CA 94550.
E-mail: gates5@llnl.gov
In final form 18 September 1998.

1. Introduction

Since its establishment in 1989 by the World Climate Research Programme, the Atmospheric Model Intercomparison Project (AMIP) has become the most prominent international effort devoted to the

diagnosis, validation, and intercomparison of global atmospheric models' ability to simulate the climate. The participating modeling groups represent virtually every atmospheric and/or climate modeling center in the world, while analysis of the results involves much of the international climate diagnostics community. The primary purpose of the AMIP was, and continues to be, the comprehensive evaluation of the performance of atmospheric GCMs on climate and higher-frequency timescales and the documentation of their systematic errors in an effort to foster the models' improvement.

Under the guidance of the AMIP Panel of the International Working Group on Numerical Experimentation, support for the implementation of AMIP has been provided by the U.S. Department of Energy through the Program for Climate Model Diagnosis and Intercomparison (PCMDI) at the Lawrence Livermore National Laboratory. As described by Gates (1992), AMIP was designed to simulate the atmosphere's evolution subject to the observed sequence of monthly averaged global sea surface temperature and sea-ice distributions during the decade 1979–88, along with standardized values of the solar constant and atmospheric CO_2 concentration. An agreed-upon standard output (see http://www-pcmdi.llnl.gov/amip/OUTPUT/AMIP1/amip1so.html) consisting of the monthly averaged global distributions of selected atmospheric variables has been collected, quality controlled, archived, and distributed by PCMDI.

By virtue of its sustained support and near-universal participation, AMIP has become a de facto standard climate performance test of atmospheric GCMs. By documenting the models and archiving their results, AMIP has aided the process of model improvement through systematic diagnosis and experimentation. AMIP has also proven to be a useful reference for model sensitivity and predictability experiments, and has served as a prototype for similar projects on the validation and intercomparison of other models.

The purposes of this paper are to present an overview of the results of AMIP and to describe the planning for its continuation. In view of the current public availability of the original AMIP models' results and their widespread analysis by the AMIP diagnostic subprojects and other groups (Gates 1995; Gates et al. 1996), we do not emphasize here the results of individual models, nor do we consider the simulation of specific processes or events. Rather, we focus on the performance of the AMIP models as a whole and seek to summarize the systematic errors that were characteristic of atmospheric GCMs in the early 1990s. Against the background of the results, we are then able to document the improvement of a subset of the original AMIP models that have revisited the experiment with revised model versions. The AMIP models have been described by Phillips (1994, 1996) and are identified in appendix A; the AMIP diagnostic subprojects are listed in appendix B, and the observational data used for validation are described in appendix C.

2. Validation of AMIP ensemble mean

The AMIP results clearly demonstrate the degree to which atmospheric GCMs can simulate the observed mean seasonal climate when furnished with realistic boundary conditions. On the whole, the models provide a credible simulation of the large-scale distribution of the primary climate variables characterizing the atmospheric pressure, temperature, wind, hydrologic cycle, and radiation balance, although a number of common systematic model errors are apparent. Here we summarize the AMIP simulations in terms of the mean of the models as an ensemble and focus on selected variables for the illustrative season December–February (DJF) during the decade 1979–88. The observed data used for model validation are from the European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis for 1979–88 whenever possible, although other sources are used for surface temperature and fluxes, precipitation, cloudiness, and radiation; the specific observational data used are summarized in appendix C.

a. Geographical distributions

The DJF average of the AMIP models' ensemble mean of the global distribution of mean sea level pressure is shown in Fig. 1, along with the observed field taken from the ECMWF reanalysis for 1979–88 (Gibson et al. 1997). This figure also shows the ensemble standard deviation and the corresponding ensemble mean error relative to the observed mean. Here (and in subsequent figures) the ensemble mean has been constructed by interpolating each model's results to a common $2.5° \times 2.5°$ grid comparable to the resolution of the observed data. It is evident that the AMIP ensemble mean shown in Fig. 1a closely resembles the observed large-scale distribution shown in Fig. 1b in nearly every respect (and, in fact, resembles the observed field more closely than any individual model's result). As shown in Fig. 1c there is

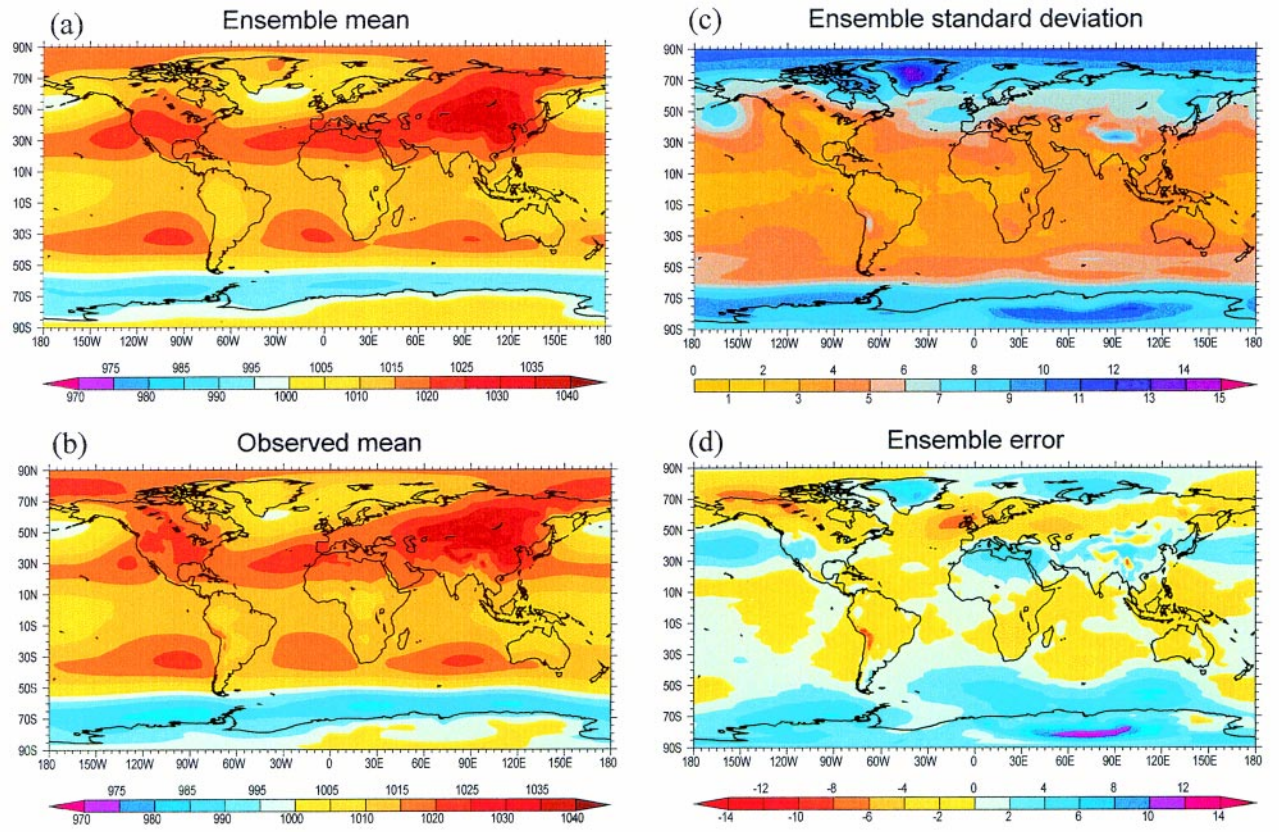

FIG. 1. The geographical distribution of mean sea level pressure (hPa) in DJF of 1979–88 given by (a) the AMIP ensemble mean, and (b) by data from the ECMWF reanalysis (Gibson et al. 1997) for 1979–88. (c) The standard deviation (hPa) of the model ensemble, and (d) the error (ensemble mean minus observation; hPa).

considerable scatter among the models' results in high latitudes, some of which is likely due to differences in the models' methods of reduction to sea level and differences in their orography. In addition to the maximum errors over Antarctica and Greenland, Fig. 1d shows the ensemble mean sea level pressure to be too high across the mid-Pacific and over the Mediterranean Sea; this error is likely to be related to a northward displacement of the westerlies. Similar results are found in the other seasons (not shown).

The DJF average of the AMIP models' ensemble mean of the global distribution of the 200-hPa velocity potential is shown in Fig. 2a, along with the observed distribution (Fig. 2b) taken from the ECMWF reanalysis for 1979–88 (Gibson et al. 1997). Although the model ensemble correctly positions the large-scale maxima and minima over North Africa, South America, and eastern Indonesia, it underestimates the strength of the latter two and hence underestimates the strength of the associated convergent flow in much of the Southern Hemisphere. The ensemble standard deviation (Fig. 2c) and the ensemble mean error (Fig. 2d) show that the principal intermodel disagreements occur in the Tropics and are likely to be a result of differing parameterizations of deep convection. Similar results are found in the other seasons (not shown).

The average DJF global distribution of the AMIP models' ensemble mean of precipitation is shown in Fig. 3a, together with the corresponding observed distribution (Fig. 3b) given by the National Centers for Environmental Prediction (NCEP) data of Xie and Arkin (1997). The ensemble standard deviation (Fig. 3c) and the ensemble mean error (Fig. 3d) are also shown. The models as a whole are seen to give a broadly realistic distribution of precipitation, although the models generally underestimate the observed DJF precipitation in the equatorial zones, which are also the regions of greatest disagreement among the models themselves. The models also generally underestimate the dryness in the subtropical dry zones, although this and other apparent errors are relative to the accuracy of the Xie and Arkin (1997) estimate of the observed precipitation (see appendix C). As was the case for sea level pressure, the ensemble precipitation distribution

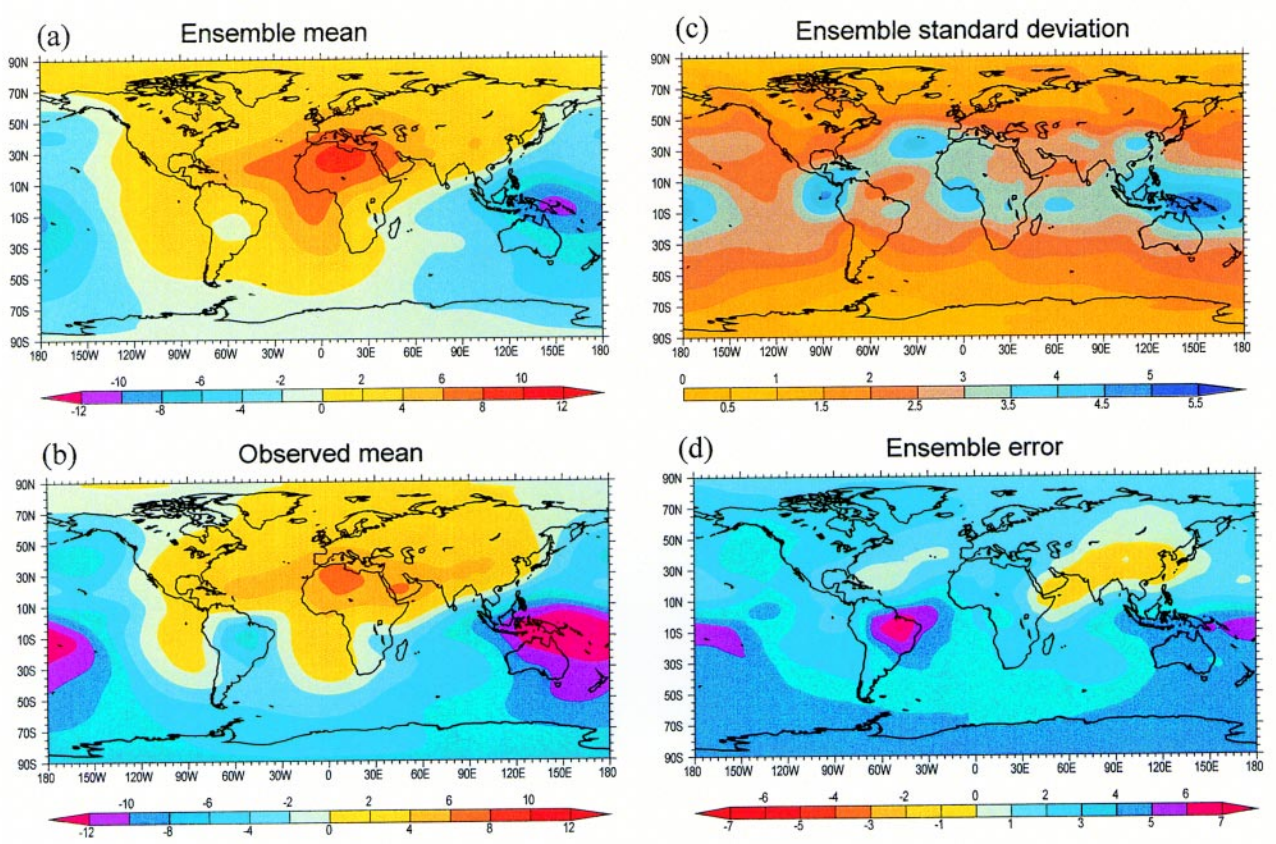

FIG. 2. As in Fig. 1 except for the velocity potential at 200 hPa (10^6 m^2 s^{-1}).

is superior to that of any individual model [a result first noted by Hulme et al. (1993)].

In order to illustrate the AMIP models' simulation of the seasonal change from winter to summer for at least one variable, we show the precipitation for June–August (JJA) in Fig. 4. From the ensemble mean given in Fig. 4a and the observational estimate from the NCEP database (Xie and Arkin 1997) given in Fig. 4b, we may see that the models on the whole reproduce the seasonal migration of the large-scale precipitation features reasonably well. We note, however, that both the ensemble standard deviation (Fig. 4c) and the ensemble error (Fig. 4d) are considerably larger in JJA than in DJF.

In order to show the models' seasonal precipitation in a relative sense, the ratio of the ensemble error to the observed mean is given for both DJF and JJA in Fig. 5. Here we clearly see that the models are generally too dry by more than 20% over large regions of the major precipitation zones. As noted earlier, the models are also too wet in the major dry zones in both DJF and JJA, although this effect is exaggerated by the low amounts of the observed precipitation.

b. Zonal averages

Although our emphasis is on the AMIP models' overall performance, it is instructive to show the individual models' results for a few selected variables. In doing so, however, it should be emphasized that the original AMIP simulations were made in the early 1990s and are not necessarily representative of the participating institutions' more recent models. With this understanding, the zonal averages of the DJF mean sea level pressure simulated by the individual AMIP models are shown in Fig. 6a, along with the observed distribution taken from the ECMWF reanalysis for 1979–88 (Gibson et al. 1997). The models' results reassuringly cluster around the observed data, although poleward of about 60° latitude there is a marked increase in the intermodel scatter. While some of this disagreement may be due to the models' differences in orography and their procedures for reduction to sea level, it suggests the presence of systematic model errors.

The zonally averaged distributions of the models' DJF surface air temperature are shown in Fig. 6b, along with the observed distribution as merged by Fiorino (1997) from data of da Silva et al. (1994), Jones (1988),

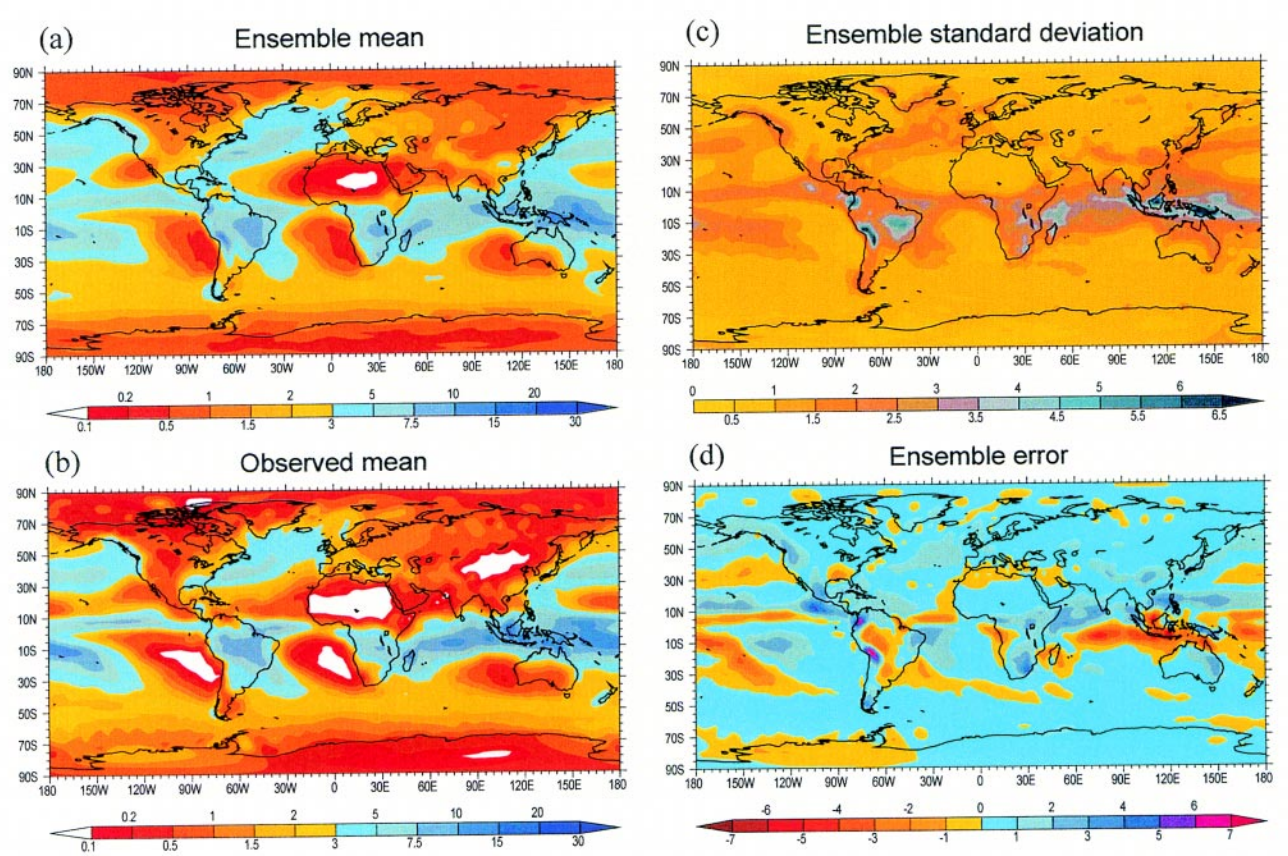

FIG. 3. As in Fig. 1 except for precipitation (mm day^{-1}), with observations for 1979–88 from the NCEP database (Xie and Arkin 1997). Note the nonlinear scale in (a) and (b).

and Schubert et al. (1992). As expected, the models' results closely follow the observed data in those latitudes where ocean predominates, doubtless due to the use of prescribed observed sea surface temperatures. The larger deviations among the models in higher latitudes reflect their strongly model-dependent simulation of the near-surface vertical temperature structure in the polar regions in both summer and winter. As was the case with sea level pressure, these data are also sensitive to the models' orography and to their definition of the surface air temperature.

The zonal averages of the simulated DJF distribution of zonal wind at 200 hPa are shown in Fig. 6c, along with the distribution given by the ECMWF reanalysis for the AMIP decade (Gibson et al. 1997). Except for a few outliers, the models closely follow the observations in this and the other seasons (not shown), in testimony to their generally realistic reproduction of the average tropospheric thermal structure. The most apparent systematic error is the tendency of many models to overestimate the strength of the westerly maximum in the Southern Hemisphere, and a slight northward displacement of the westerlies in the Northern Hemisphere.

The zonal averages of the net surface heat flux simulated over the ocean (only) for DJF are shown in Fig. 6d, along with the observational estimate given by the Comprehensive Ocean Atmosphere Data Set (COADS) (da Silva et al. 1994b). The models' results conform to the overall structure of the observed distribution in this and other seasons (not shown). However, the annual mean poleward gradient of net surface heating is clearly deficient in most models, which has important implications when the model is coupled to an ocean GCM. Although the observational uncertainties are large, the surface net shortwave radiation and the surface latent heat flux are the dominant components of surface heating.

The zonally averaged distributions of the outgoing longwave radiation (OLR) at the top of the atmosphere are shown in Fig. 7a, along with the observed distribution given by the National Environmental Satellite, Data and Information Service (NESDIS) data for the AMIP decade (Gruber and Krueger 1984). The models provide a generally realistic simulation in this and other seasons at all latitudes, although there is a suggestion of a systematic overestimate of the OLR in the

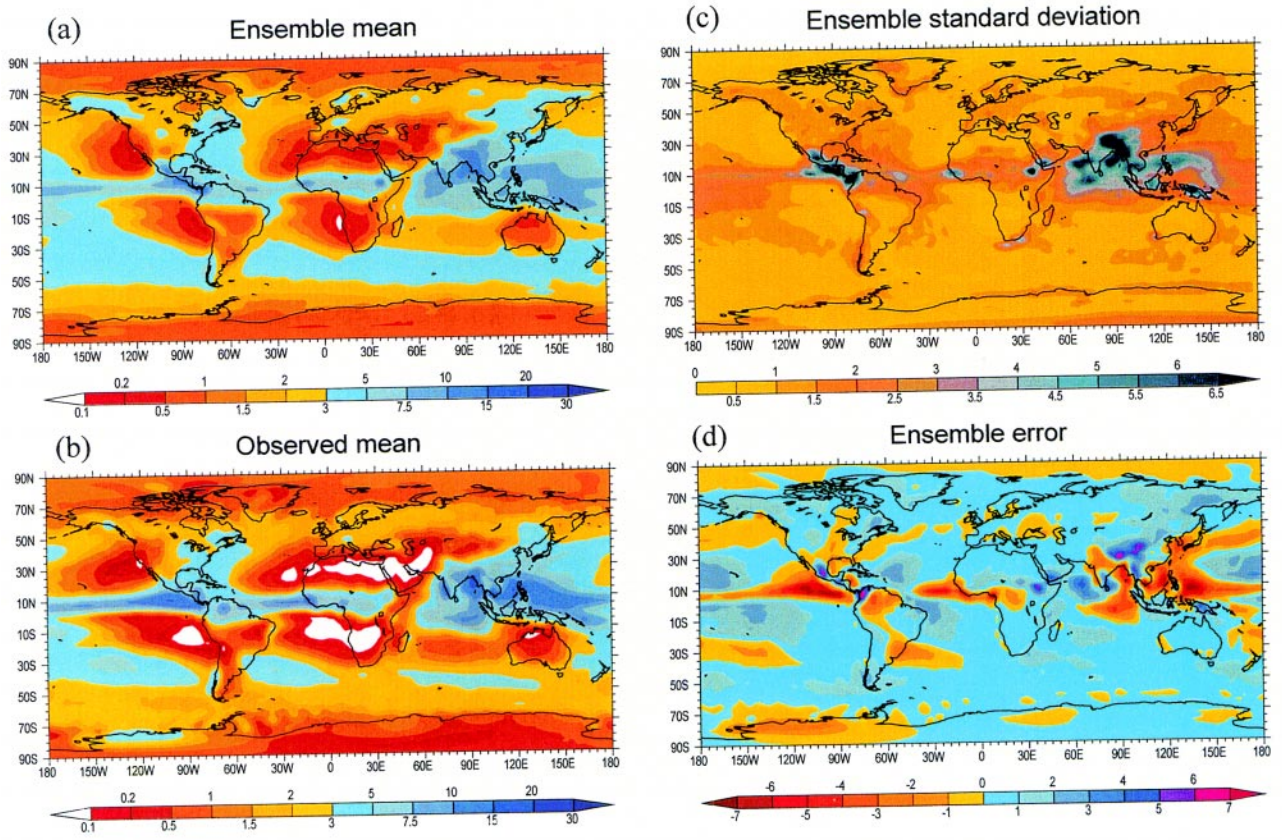

FIG. 4. As in Fig. 3 except for JJA.

lower latitudes near 30°N and 30°S, which is likely related to the models' underestimate of cloudiness at these latitudes (see below).

The zonally averaged distributions of the simulated DJF total cloudiness are shown in Fig. 7b, along with the observed distribution given by the International Satellite Cloud Climatology Project (ISCCP) data during 1983–90 (Rossow et al. 1991). Although the bulk of the models' results display the same general latitudinal variation as do the observations in this and the other seasons (not shown), it is apparent that there are large intermodel differences. Note, however, that part of the scatter in the models' results is due to varying definitions of total cloudiness among the models and between the models and ISCCP. Except in the high latitudes (where the observational estimates are especially uncertain), there is a tendency for most models to underestimate the total cloudiness.

The zonally averaged distributions of the total simulated precipitation are shown in Fig. 7c, along with an estimate of the observed distribution from the NCEP database (Xie and Arkin 1997). Overall, the models' results in DJF and the other seasons (not shown) display the same general latitudinal structure as the observational estimates, including the equatorial maximum and the secondary maxima in the midlatitudes of both hemispheres. A relatively large scatter among the models' results is also evident in Fig. 7c, especially in the equatorial region, although there are outliers in the higher latitudes as well. If the observational estimates are deemed reliable, these data indicate a model tendency to overestimate the precipitation at nearly all latitudes except south of about 30°S, along with a systematic poleward displacement by about 10° latitude of the precipitation minima in the subtropics of both hemispheres. The models' systematic underestimation of both the maximum and minimum regional precipitation seen in Fig. 5 is not evident here due to compensation in the zonal averages.

Finally, the zonally averaged distributions of the precipitation minus evaporation ($P–E$) simulated in DJF by the AMIP models is shown in Fig. 7d, together with an estimate of the observed distribution from the NCEP data of Xie and Arkin (1997) and the COADS data of da Silva et al. (1994c). In spite of the difficulties in simulating precipitation noted above, and in spite of the relatively large uncertainties in estimating the "observed" evaporation, the bulk of the models

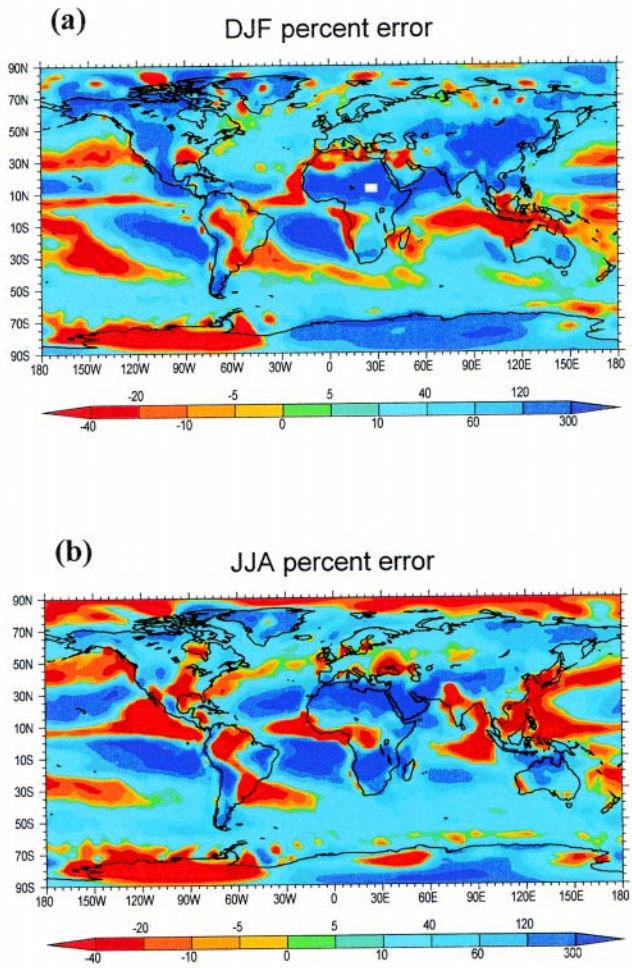

FIG. 5. The ratio of the AMIP ensemble mean precipitation to the observed precipitation (in %) for (a) DJF and (b) JJA. (The white dot in central Africa denotes a location of nearly zero observed precipitation in DJF.)

successfully reproduce the observed large-scale features of the observed *P–E*, although the standard deviation of the models' results is a large fraction of their average at most latitudes. This qualified success, along with the limited success of the simulation of the surface heat flux seen in Fig. 6d, indicates that considerable improvement is possible in atmospheric models' simulation of the surface fluxes that are critical to the global ocean circulation.

c. Meridional sections

A complement to the data shown in Figs. 1–7 is given by the portrayal of selected simulated variables in the latitude–pressure meridional section. This is given for the DJF ensemble mean temperature in Fig. 8a, and it closely resembles the observed distribution given by the ECMWF reanalysis (Gibson et al. 1997) in Fig. 8b. As found in earlier intercomparisons (Boer et al. 1992), the errors of the AMIP ensemble mean given in Fig. 8d show a marked systematic cold bias nearly everywhere, exceeding –11°C in the lower polar stratosphere. The cause of this error has remained elusive, although it may be related to errors in the models' advection schemes. The variability of the models' zonally averaged temperature about their ensemble mean (Fig. 8c) shows that the greatest model disagreement is in the lower tropical stratosphere.

The DJF ensemble mean zonal wind given in Fig. 9a is seen to closely resemble the observed distribution calculated from the ECMWF reanalysis (Gibson et al. 1997) given in Fig. 9b. The errors of the ensemble mean (Fig. 9d) are generally small, except in the core of the westerly jet in the Northern Hemisphere (which the model ensemble positions slightly too far north) and in the Southern Hemisphere stratosphere (where the model ensemble underestimates the easterlies). In the Tropics, however, the ensemble mean has a westerly bias relative to the observed average easterlies. The standard deviation of the simulations about their mean (Fig. 9c) is also seen to generally increase with altitude.

The corresponding structure of the AMIP models' ensemble mean of the streamfunction for the mean meridional circulation is given in Fig. 10a. The observed distribution calculated from the ECMWF reanalysis (Gibson et al. 1997) shown in Fig. 10b is structurally similar to the models' portrayal of the Hadley circulation between about 30°S and 50°N, although the models systematically underestimate the observed strength. This is likely due to the models' coarse resolution and is clearly shown in Fig. 10d, where the dominant error pattern resembles the Hadley circulation itself. The variability among the models (Fig. 10c) is largest in the portrayal of the tropical circulation.

d. Validation summary

A summary of the accuracy of the models' seasonal simulation is given in Table 1 in terms of the root-mean-square errors of the ensemble mean in comparison to the observations. Using the ECMWF reanalysis for verification, the rms errors of the ensemble mean sea level pressure and 200-hPa temperature are generally larger in the Southern Hemisphere, while the rmse of the ensemble mean surface air temperature are generally smaller, reflecting the predominance of the prescribed sea surface temperatures. Interestingly enough, the rmse's of 200-hPa temperature are largest in the Southern Hemisphere in summer, while the rmse of the 200-hPa zonal wind

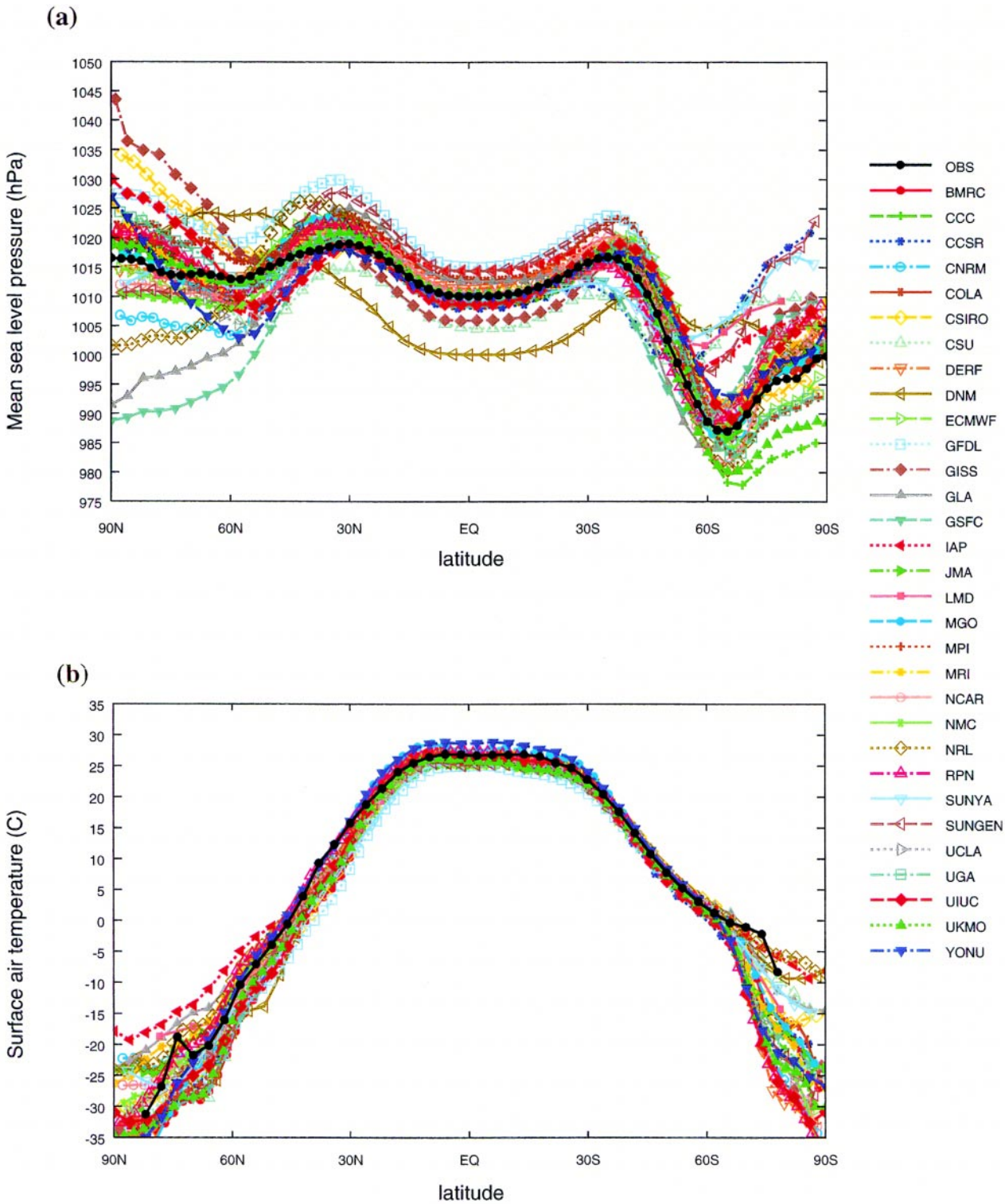

FIG. 6. The zonally averaged distribution of selected variables simulated by the AMIP models (see appendix A) for DJF of 1979–88 and that given by the ECMWF reanalysis for the same period (Gibson et al. 1997) (solid black line). (a) The sea level pressure, with observed data from the ECMWF reanalysis; (b) the surface air temperature, with observed data as merged by Fiorino (1997) from data of da Silva et al. (1994a), Jones (1988), and Schubert et al. (1992).

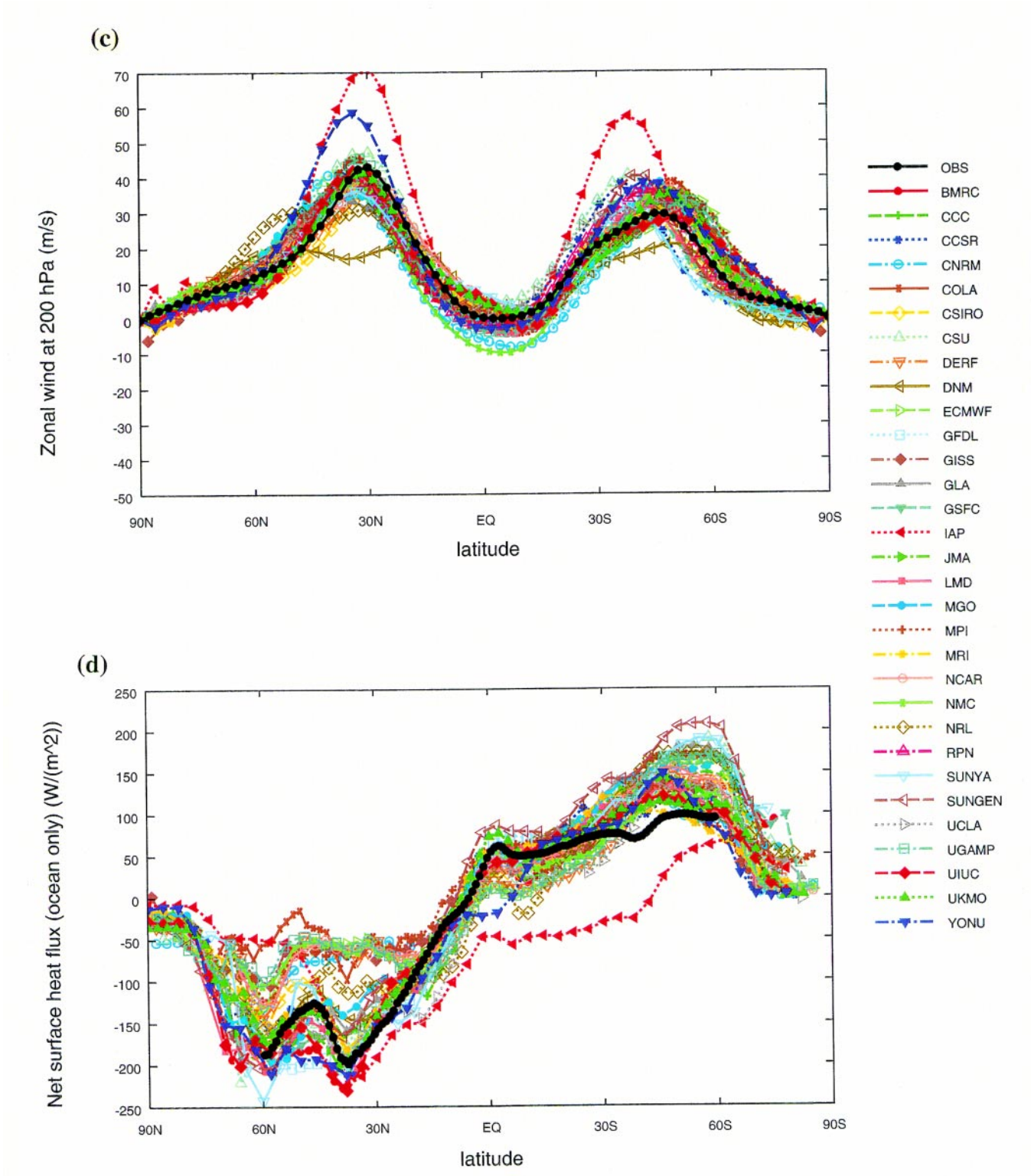

FIG. 6. *(Continued)* (c) The zonal wind at 200 hPa, with observed data from the ECMWF reanalysis; (d) the net ocean surface heat flux, with observational estimates from COADS (da Silva et al., 1994b). [See appendix A for model identification; UGAMP missing in (b), RPN missing in (d).]

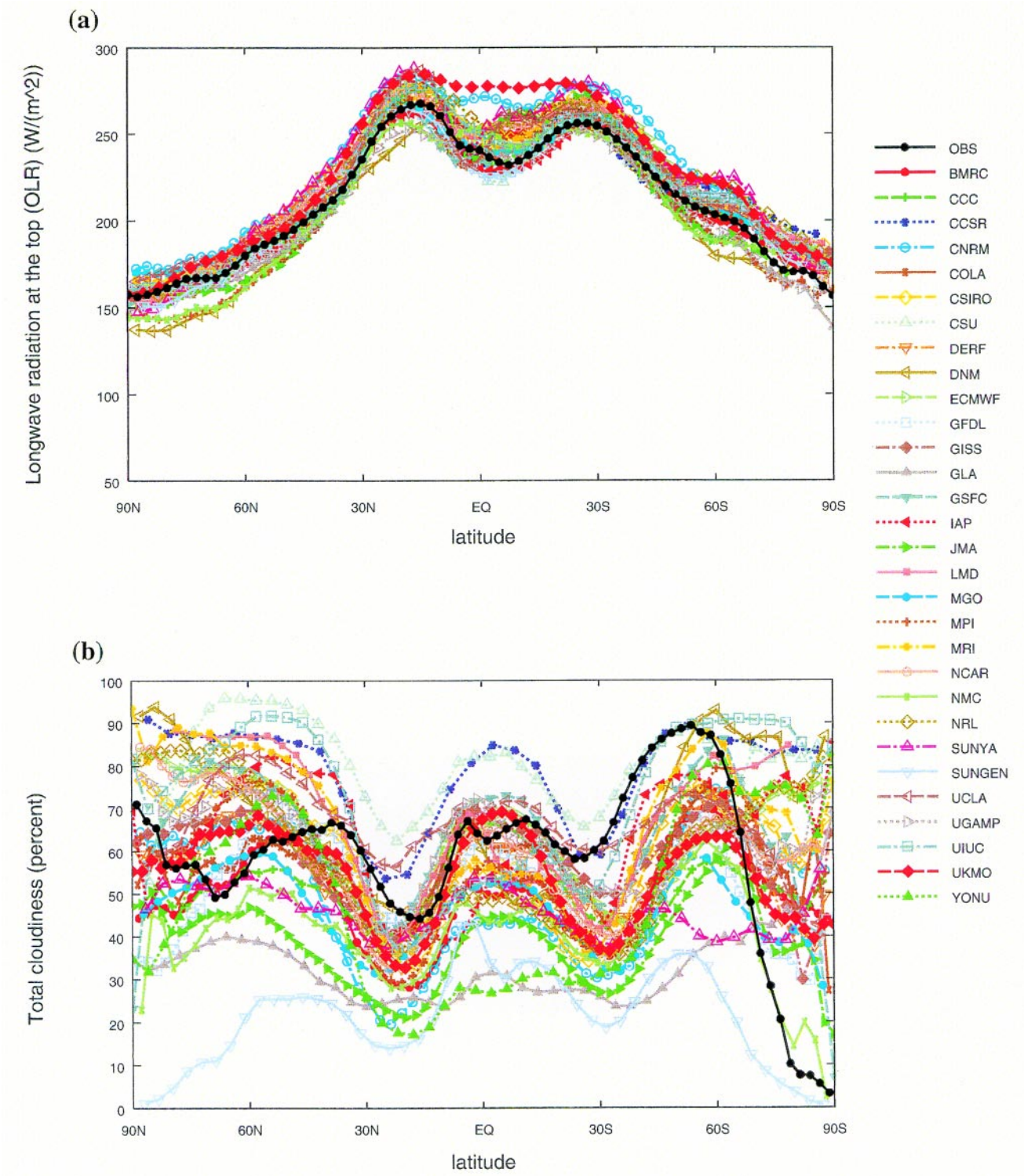

FIG. 7. As in Fig. 6 except for the (a) the outgoing longwave radiation, with observations from the NCEP database (Gruber and Krueger 1984); (b) total cloudiness with observations from ISCCP for 1983–90 (Rossow et al. 1991).

is a minimum in the autumn in both hemispheres. The rmse of the OLR, on the other hand, shows relatively little seasonal or hemispheric variation and may reflect the tuning of the models' cloud radiative properties. This is in contrast to the distribution of total cloudiness, which is notably larger in the Southern

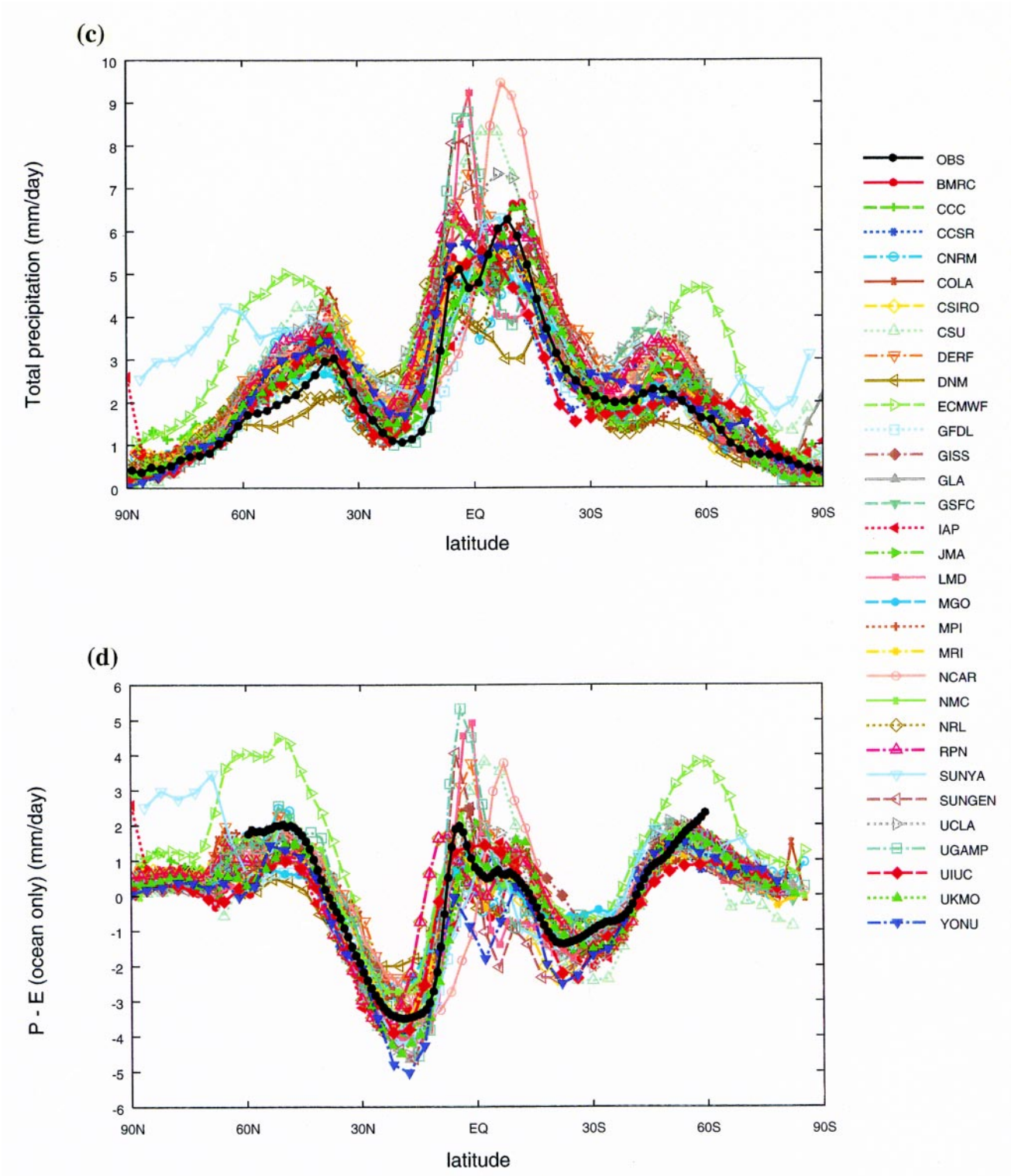

FIG. 7. *(Continued)* (c) Precipitation with observations from the NCEP database (Xie and Arkin 1997); (d) precipitation minus evaporation over the ocean with observations from the NCEP data of Xie and Arkin (1997), and the COADS data for da Silva et al. (1994c). [RPN missing in (a), (b), and (d).]

Hemisphere than in the Northern Hemisphere. The rmse's of precipitation and of precipitation minus evaporation are roughly the same in all seasons in both hemispheres and represent a substantial fraction of the globally averaged annual precipitation of 2.7 mm day^{-1} given by Xie and Arkin (1997). This

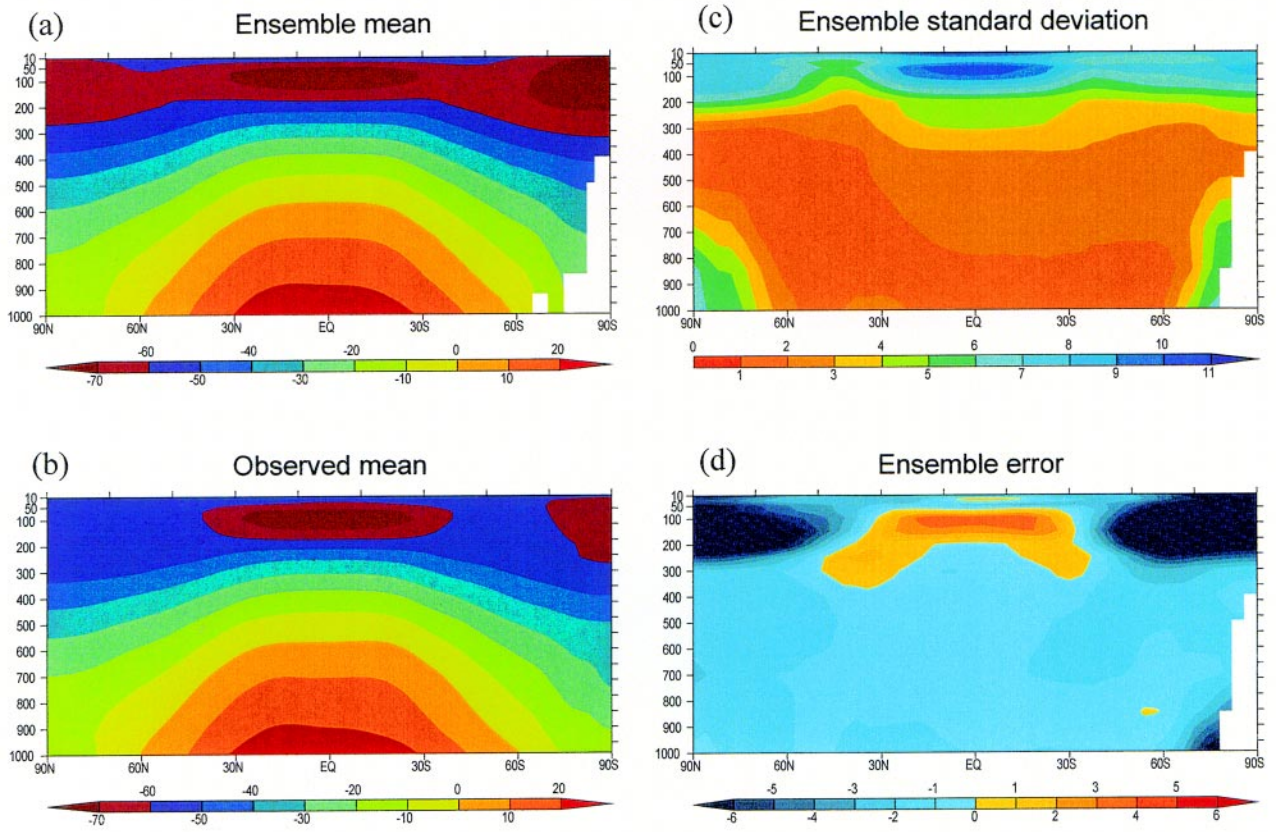

FIG. 8. (a) The latitude–pressure meridional section of the temperature (°C) given by the AMIP ensemble mean and (b) the observed data from the ECMWF reanalysis (Gibson et al. 1997). (c) The standard deviation of the ensemble mean. (d) The ensemble error. The pressure units are hPa.

reflects the continuing difficulty of both models and observations in accurately estimating this component of the hydrological cycle.

Also shown in Table 1 are the relative errors of the ensemble means, as given by the ratio of the rmse to the observed spatial standard deviations. By this measure the largest percent errors are found in the cloudiness and 200-hPa temperature, for which the relative seasonal errors can exceed 100%. The smallest percent errors, on the other hand, are found in the surface air temperature, as might have been anticipated.

The AMIP performance errors given in Figs. 1–10 and Table 1 replace the preliminary statistics given earlier (Gates 1995), which were incomplete in some respects and which did not use reanalysis for validation. It remains true, however, that the mean and rms errors of the ensemble mean are smaller than those for any individual model in the ensemble, in terms of latitudinal and seasonal averages.

Considerable additional diagnosis and validation of the AMIP models have been performed by the AMIP diagnostic subprojects (see appendix B), by the PCMDI staff, and by the modeling groups themselves. This work has served to show the presence of important systematic errors in the AMIP models' simulations of a wide variety of processes and regional phenomena. A summary of many of these studies is given in the *Proceedings of the First International AMIP Scientific Conference* (Gates 1995), and a comprehensive collection of abstracts of AMIP-related publications may be accessed on the Internet at http://www-pcmdi.llnl.gov/amip/ABSTRACTS.

3. Validation of AMIP ensemble variability

In addition to validation of the mean, attention should be given to validation of the variability about the time and/or space mean, since in some instances this is a more important and revealing aspect of model performance than the means themselves. To this end, we consider here the AMIP models' portrayal of both the seasonal cycle and interannual variability of selected variables.

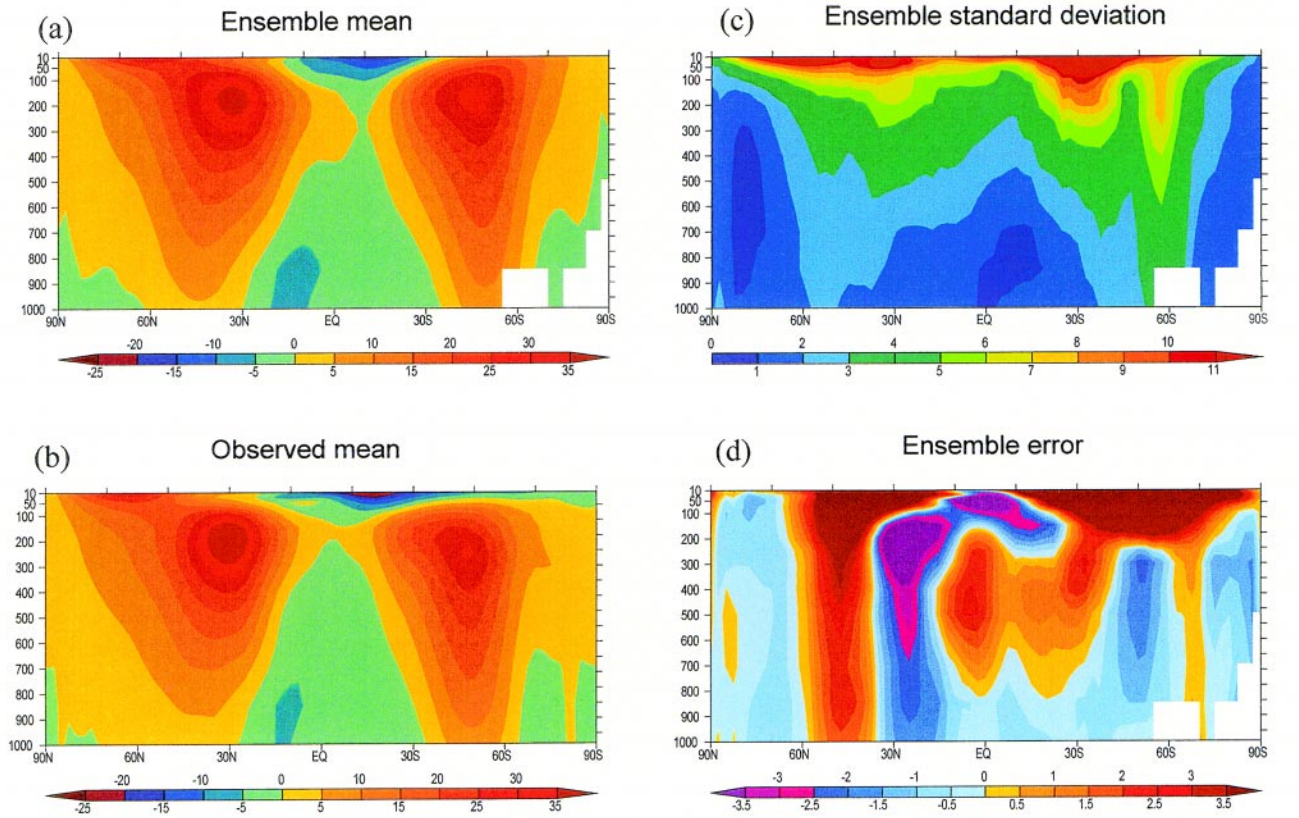

FIG. 9. As in Fig. 8 except for the zonal wind (m s^{-1}), with the observed estimate taken from the ECMWF reanalysis (Gibson et al. 1997).

a. Seasonal variability

Although there are many ways of portraying the seasons, a compact form that preserves geographical dependence is the amplitude of the mean seasonal cycle. This is shown in Fig. 11a for the average of the first annual harmonic of the models' simulation of mean sea level pressure and in Fig. 11b for the ECMWF reanalysis over the AMIP decade (Gibson et al. 1997). While the simulated and observed patterns are quite similar, the large amplitude of the seasonal variation observed over the Tibetan Plateau is overestimated by the models, as are the secondary maxima of seasonal sea level pressure variation near the Aleutian Islands, over western North America, and near Iceland. The average of the models' phasing of the first annual harmonic shown in Fig. 11c is seen to closely resemble that from the reanalysis shown in Fig. 11d, with the exception of high southern latitudes where differences in the models' calculation of the pressure reduction to sea level cause substantial disagreement.

In the ECMWF reanalysis the annual harmonic of sea level pressure explains upward of 90% of the total variability in the Tropics and subtropics, a statistic that the models slightly overestimate. The models' average portrayal of the mean seasonal cycle of other variables (not shown) is also in close agreement with observations, although there are large differences among some models.

b. Interannual variability

The AMIP decade is marked by two large ENSO events in 1982/83 and 1986/87, which provide an attractive opportunity to evaluate the models' portrayal of interannual variability. Instead of the commonly used Southern Oscillation index, however, we have used the mean sea level pressure averaged over 15°–25°S and between the longitudes 125°–135°E and 135°–145°W as suggested by Trenberth and Shea (1987) in order to accommodate the models' different horizontal resolutions as equitably as possible. After removing the mean annual cycle from each model's simulation and then using a filter to remove variations of less than 8 months, the results for the ensemble of AMIP models are shown in Fig. 12a along

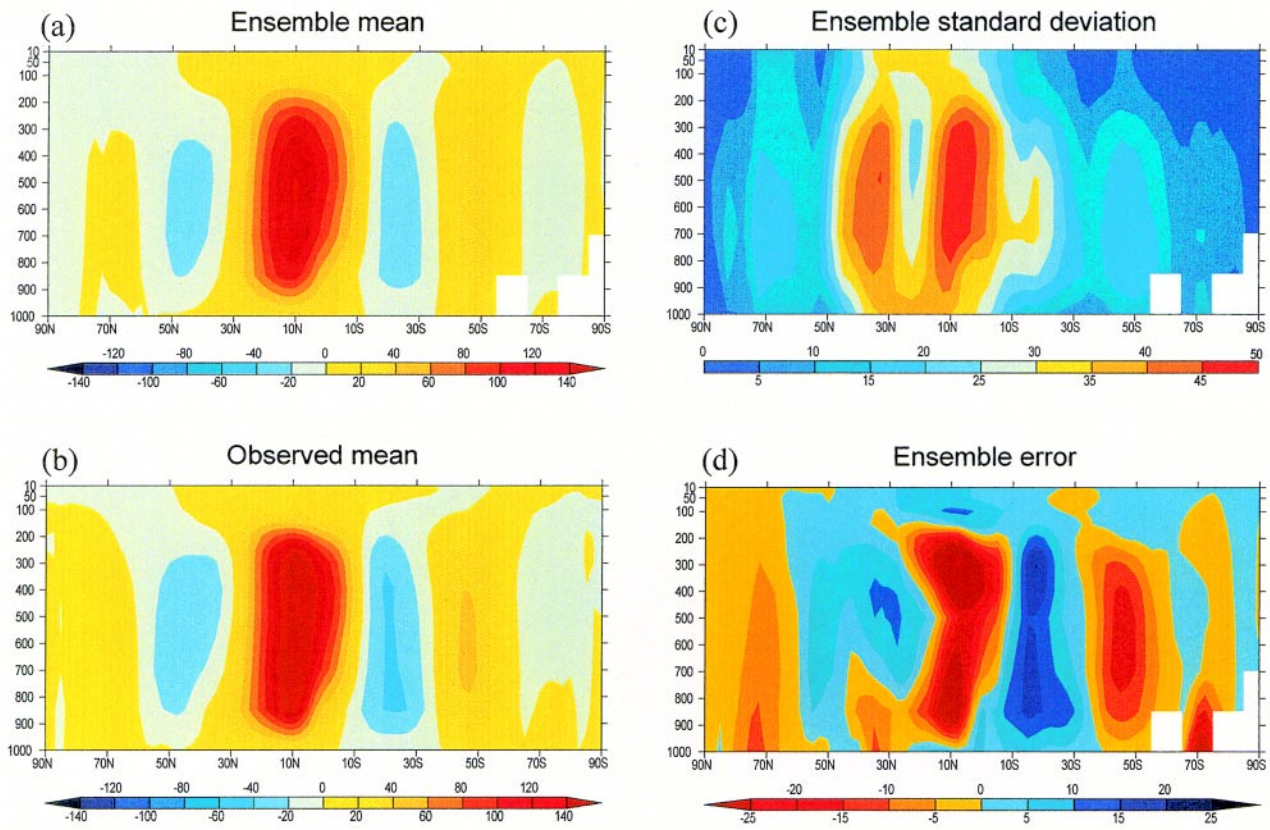

FIG. 10. As in Fig. 8 except for the streamfunction for the mean meridional circulation (10^9 kg s^{-1}).

with the corresponding observational result from the NCEP reanalysis (Kalnay et al. 1996). On the whole, the models simulate the broad aspects of the observed ENSO variations reasonably well (as might be expected since these are primarily driven by the prescribed SST), at least as portrayed by the ensemble mean and the surrounding two standard deviations. It may also be noted that the models generally underestimate the magnitude of the sea level pressure response in the major ENSO events in spite of the observed sea surface temperature anomalies having been prescribed as boundary conditions.

As a measure of midlatitude interannual variability, we have chosen the mean sea level pressure averaged over the area 30°–65°N, 160°E–140°W in the North Pacific, following Trenberth and Hurrell (1994). The results in terms of the AMIP ensemble mean and the associated standard deviations are shown in Fig. 12b, along with the observed variability. (As in Fig. 12a, the mean annual cycle has been removed and variations less than eight months have been filtered out.) There is poor correlation with the observations, and there is a relatively large spread among the models; approximately half of the models have variance that is greater than that in the reanalysis, in contrast to the Tropics where the models systematically underestimated the observed variability. Similar results are found when the interannual variance is examined over North America and the North Atlantic where the reanalysis is relatively robust, indicating the difficulty the models have in simulating extratropical variations that are not linked to the sea surface temperature.

c. Space–time variability

Further insight on the AMIP models' ability to simulate both the pattern and amplitude of the observed interannual variations is afforded by the diagram devised by K. E. Taylor (1998, unpublished manuscript) and illustrated in Fig. 13 for the AMIP model's simulation of the total space–time variability of the monthly averaged mean sea level pressure (excluding land areas). Here the distance from the origin is equal to the standard deviation of the field while the distance from the "reference" point (i.e., the ECMWF reanalysis) is equal to the rms difference between the observed and modeled fields (with the time-mean global average removed). Both the standard deviation and the rms differences have been normal-

TABLE 1. Seasonal rmse of the AMIP model ensemble mean and their percentage ratio to the observed spatial std dev (in parentheses) for selected variables over the decade 1979–88. The observational data used are identified in appendix C.

Variable	Northern Hemipshere				Southern Hemisphere			
	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON
Mean sea level pressure (hPa)	2.2 (37)	1.7 (44)	2.4 (52)	1.6 (40)	2.3 (25)	4.3 (41)	5.0 (44)	4.2 (35)
Surface air temperature* (°C)	3.9 (23)	4.1 (32)	3.0 (39)	3.6 (31)	1.6 (16)	2.9 (28)	3.2 (32)	2.0 (20)
Temperature at 200 hPa (°C)	3.1 (113)	4.6 (183)	4.1 (130)	3.7 (216)	6.8 (214)	5.9 (340)	4.2 (86)	4.9 (127)
Zonal wind at 200 hPa (m s^{-1})	4.1 (26)	3.9 (31)	3.6 (31)	3.2 (26)	3.0 (26)	3.0 (28)	3.1 (20)	3.6 (2.8)
Outgoing longwave radiation** (W m^{-2})	11.0 (29)	11.6 (38)	11.4 (41)	10.3 (35)	11.9 (46)	10.4 (32)	12.3 (29)	13.9 (40)
Total cloudiness (%)**	16.7 (86)	17.5 (101)	18.0 (99)	16.3 (91)	23.6 (118)	21.2 (117)	20.9 (102)	24.1 (122)
Precipitation (mm day^{-1})	1.1 (52)	1.3 (57)	1.7 (53)	1.3 (48)	1.2 (40)	1.0 (47)	0.9 (53)	1.0 (53)
Precipitation–evaporation** (mm day^{-1})	1.4 (55)	1.2 (54)	1.6 (66)	1.6 (66)	1.3 (58)	1.6 (65)	1.5 (61)	1.5 (63)

*UGAMP not included. **RPN not included.

ized by the observed standard deviation calculated from the ECMWF reanalysis. It can be shown that the relationship between the normalized variance, the normalized rmse, and the correlation coefficient implies that the cosine of the polar angle in Fig. 13 is equal to the correlation between the simulated and observed monthly mean fields. Thus, a model that is relatively accurate would lie near the dotted arc (indicating it had the correct variance) and close to the observed reference (indicating a small rmse and high correlation). All statistics here were computed over the full AMIP period (120 months) after interpolating modeled and observed data to a common grid and after removing any global mean bias. It is clear from Fig. 13 that the AMIP models differ widely in their ability to simulate the total space–time variability of sea level pressure and that models with the same rmse (given by the distance to the ECMWF reference) may differ substantially in the amplitude of the variability while having a similar correlation with observations [e.g., the Laboratoire de Météorologie Dynamique (LMD) and Goddard Laboratory for Atmospheres (GLA) models].

Observational uncertainty limits the extent to which any model can be expected to agree with observations or reanalyses. An estimate of this uncertainty is given in the Taylor diagram by the distance between alternative reanalyses in Fig. 13. Here the distance of the NCEP–National Center for Atmospheric Research (NCAR) Reanalysis (Kalnay et al. 1996) from the ECMWF reference is much less than the distance to any AMIP model, indicating that by this measure at least, observational uncertainty is smaller than the models' systematic errors.

A few modeling groups have carried out an ensemble of AMIP simulations, in which each simulation was identical except for the initial conditions used. The location of each ensemble member lies very close to the location of the original single simulation shown in Fig. 13, from which we may conclude that the scatter of the AMIP models plotted in the figure is not due to the differences

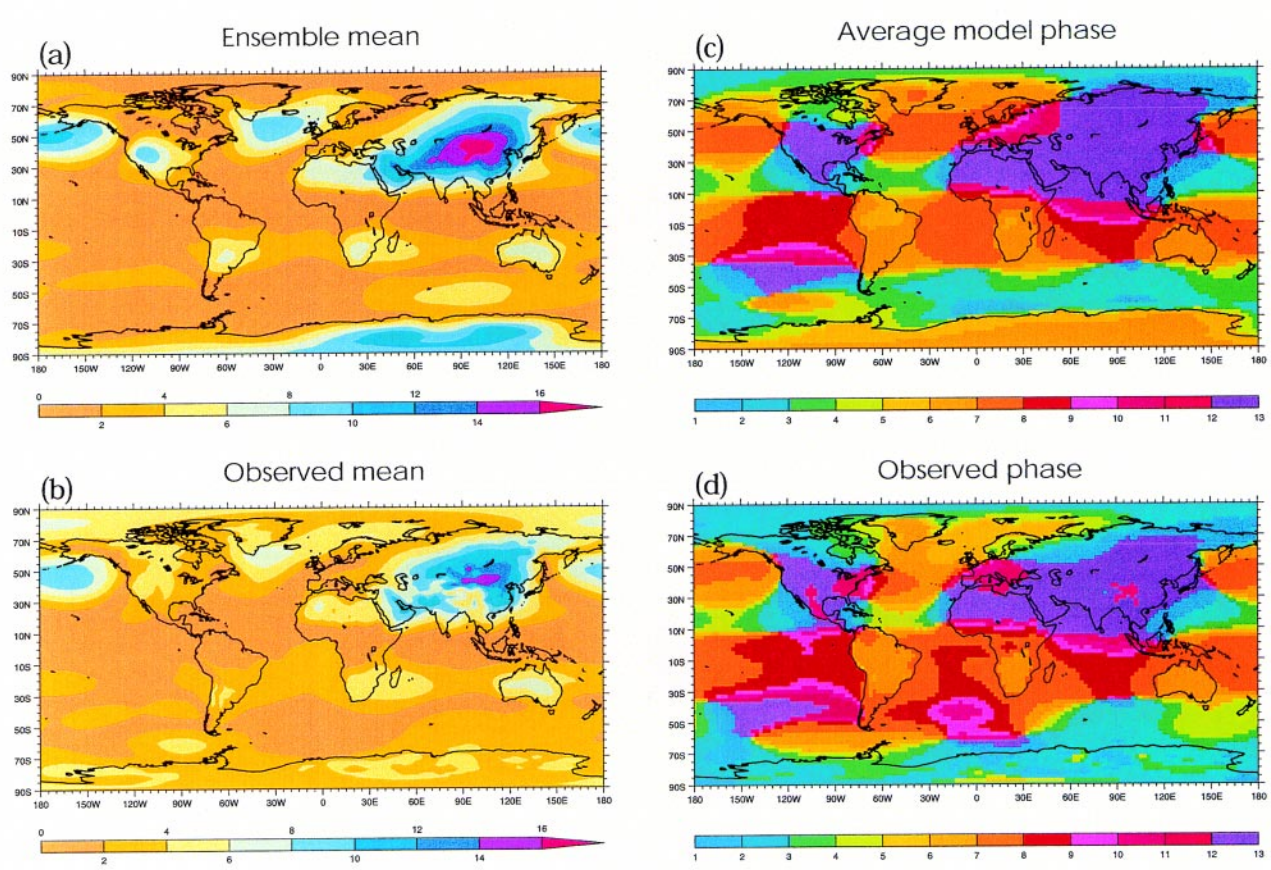

FIG. 11. The mean seasonal cycle during 1979–88 as (a) simulated by the AMIP models in terms of the average amplitude of the annual harmonic of mean sea level pressure (hPa), (b) and that given by the ECMWF reanalysis (Gibson et al. 1997). The average phase (month of maximum) of the (c) simulated annual harmonic and (d) as observed.

expected from their sensitive dependence on initial conditions. Thus, Fig. 13 indicates that differences in the AMIP models' performance are significant and that considerable improvement is possible in their simulation of the space–time pattern of variability.

The Taylor diagram may also be used to show the ability of the ensemble of the AMIP models to simulate the variability of selected variables, as in Fig. 14. For each variable the centroid of the collection of model results is plotted. It is evident that the models on the whole are relatively skillful in simulating the variability of the surface air and 850-hPa temperatures and the variability of the 500-hPa geopotential, as might be expected from the constraints of the AMIP experiments. On the other hand, the models simulate rather poorly the variability of the total cloud cover, the meridional wind, and precipitation. The models' general overestimate of the variance of the 200-hPa temperature is notable and is likely related to the systematic temperature bias seen earlier in Fig. 9. The observational uncertainty of the variables shown in Fig. 14 could be estimated by plotting the values given by the NCEP–NCAR Reanalysis relative to the ECMWF reference. These points (not shown) lie close to the dotted arc with correlations greater than 0.99.

4. Other model performance analyses

Experience in the diagnosis of AMIP model results has shown that it is useful to characterize model performance in terms of a variety of statistical measures. In an attempt to meet this need, Santer et al. (1995) computed a range of statistics following Wigley and Santer (1990). Figure 15 shows two selected statistics for the case of the mean sea level pressure simulated by the various AMIP models, using the ECMWF reanalysis as the observational reference. To avoid consideration of biases introduced by the use of different methods for reduction of pressure to sea level, the analysis was restricted to ocean areas only, after interpolation of both model and reanalysis results to a common equal-area grid (the grid of the LMD model). Additionally, the respective climatological global means were subtracted

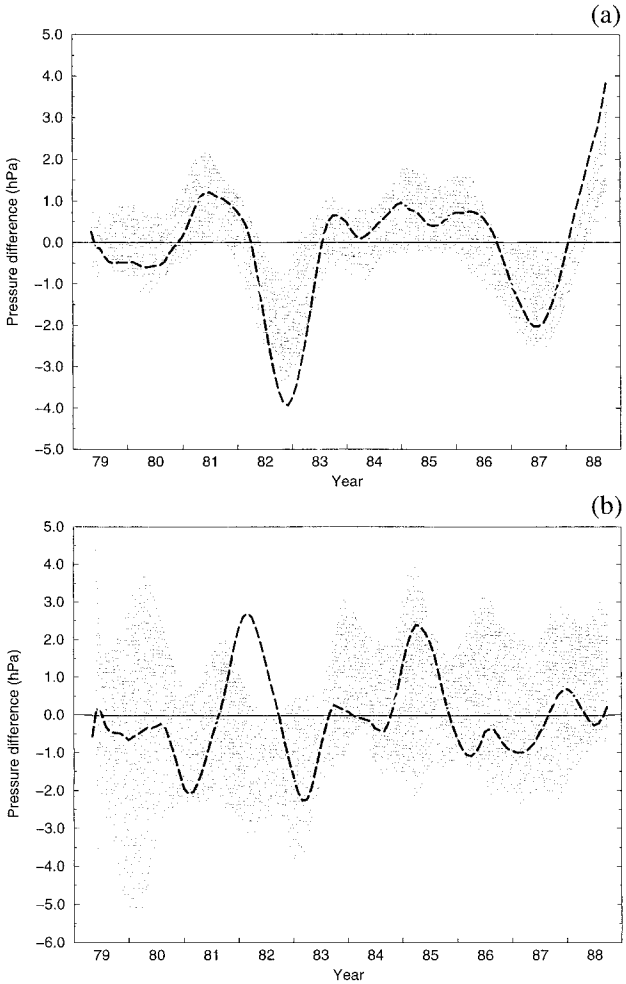

FIG. 12. The interannual variability during 1979–88 as simulated by the AMIP ensemble (full white line) and as given by the NCEP–NCAR Reanalysis (dashed line) for 1979–88 (Kalnay et al. 1996). (a) The sea level pressure difference averaged over the areas 15°–25°S, 125°–135°E, and 15°–25°S, 135°–145°W, and (b) the pressure anomaly (relative to the decadal mean) averaged over the area 30°–65°N, 160°E–140°W. The shaded area surrounding the ensemble mean indicates the spread of two std dev about the mean.

from all model and reanalysis pressure fields to compensate for overall differences in atmospheric mass.

In Fig. 15 the abscissa SITES is a dimensionless measure of overall (squared) differences in the annual mean state, standardized by the pooled temporal standard deviations of the model and observed datasets (Preisendorfer and Barnett 1983). Since climatological monthly means were not subtracted prior to analysis, the temporal variability used in the standardization of SITES has both seasonal and interannual components. Larger numerical values of SITES indicate larger overall errors in the simulation of the mean state. The ordinate RBAR (Wigley and Santer 1990) is a time-mean pattern correlation that measures whether modeled and observed spatial anomaly fields are similar and evolve in a similar way, in this case over both seasonal and interannual scales during the 120-month AMIP period; RBAR is bounded by +1 and −1, with larger values indicating greater similarity in the mean pattern and in pattern evolution.

Figure 15 shows that there is a wide spread among the models' results; there is almost a factor of 2 difference in the similarity of the models' space–time pattern evolution with observations (as measured by RBAR) and a wide variation in the models' error in time means (as measured by SITES), with no apparent correlation between the two performance measures. All of the models have errors in simulating the time-mean sea level pressure that is larger than the current differences between the two reanalysis products (as given by the comparison between the ECMWF and NCEP–NCAR reanalyses, which lies at SITES = 0.04, RBAR = 0.98). Model mean sea level pressure errors

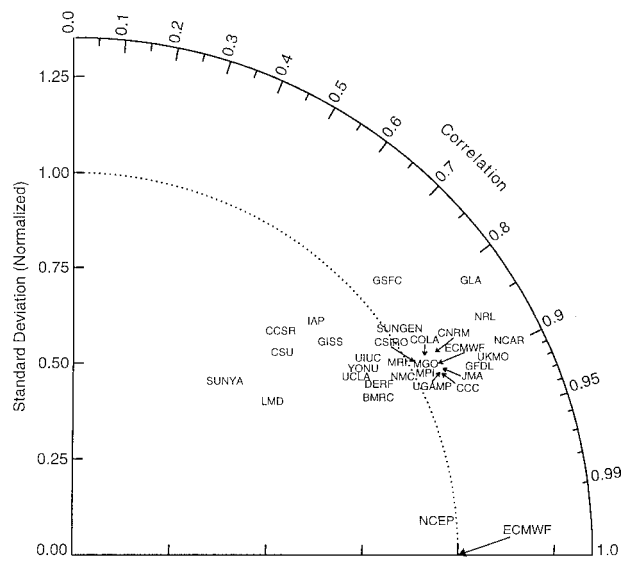

FIG. 13. A model performance diagram showing the total space–time pattern variability of the AMIP models' mean sea level pressure (excluding land areas), in terms of the std dev of the modeled monthly means (proportional to the distance from the origin), the rms difference between the simulated and observed monthly means (proportional to the distance from the ECMWF reference point), and the correlation between the simulated and observed monthly means over the period 1979–88. The std dev and rms differences have been normalized by the observed std dev. The position of the NCEP reanalysis relative to the reference ECMWF reanalysis is also indicated. Except where there are arrows, the center of each model acronym marks its position on the diagram. (See appendix A for model identification.)

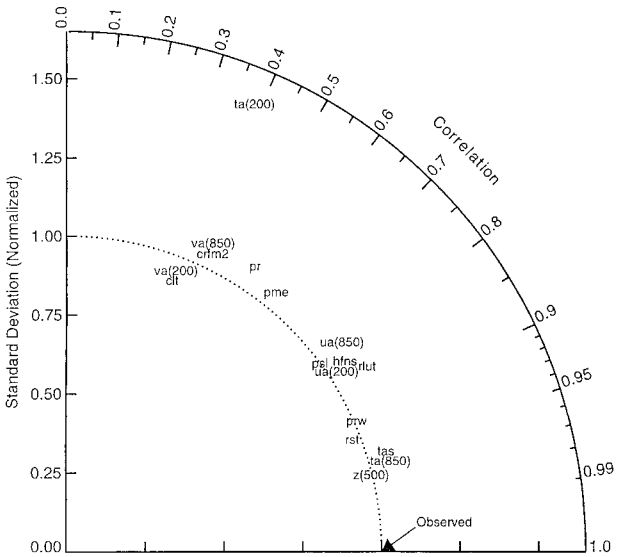

FIG. 14. As in Fig. 13 except for the AMIP ensemble's simulation of selected variables, relative to observations given by the ECMWF reanalysis. Here ta(200) is the temperature at 200 hPa, va(200) and va(850) are the meridional winds at 200 and 850 hPa, respectively, crfm2 the cloud-radiative forcing, clt the total cloudiness, pr the precipitation, pme the precipitation minus evaporation, psl the sea level pressure (excluding land areas), ua(200) and ua(850) the zonal wind at 200 and 850 hPa, hfns the net surface heat flux, rlut the outgoing longwave radiation, prw the precipitable water, rst the incoming solar radiation, tas the surface air temperature, ta(850) the temperature at 850 hPa, and z(500) the geopotential height at 500 hPa.

are also larger than the statistical differences expected from unpredictable atmospheric variability (as characterized by repeated AMIP ensemble calculations, not shown). The relatively high value of RBAR for the NCEP–NCAR Reanalysis in comparison with ECMWF indicates that the monthly mean sea level pressure anomaly fields evolve in a very similar way in the two reanalyses. All of the AMIP models have substantially less agreement with the observed anomaly pattern evolution (i.e., lower RBAR). Although Fig. 11 showed the models' simulation of the seasonal cycle of sea level pressure (as well as that of many other climate variables, not shown) to be in reasonable agreement with observations, there is considerable room for improvement in the models' simulation of interannual variability.

Do these results depend substantially on the choice of the observed dataset used for the data–model comparisons? We tested this possibility by repeating the tests, but now substituting the NCEP–NCAR Reanalysis for the ECMWF reanalysis used in Fig. 15. The primary conclusions were not modified, but it was noted that some changes occur in the relative location of some models. In Fig. 15 the MPI model has the second smallest error in the overall mean state (i.e., a small SITES values), and the ECMWF model the smallest error in space–time pattern similarity (i.e., the largest RBAR value); this may be related to the fact that both of these models and the model used in the reference ECMWF reanalysis evolved from a common progenitor. When the NCEP–NCAR Reanalysis is used as the reference observational dataset, the SITES scores of the National Meteorological Center (NMC, recently renamed NCEP) model (which is a descendant of the model used in the NCEP–NCAR Reanalysis) is improved and that of the ECMWF model is degraded. This suggests that both reanalyses contain a nontrivial model "imprint," particularly over data-poor regions, so that the differences between the ECMWF and NCEP–NCAR reanalyses are not due solely to "observational" uncertainties. For certain fields, therefore, it may be misleading to assess the performance of a model on the basis of a single reanalysis product. It is

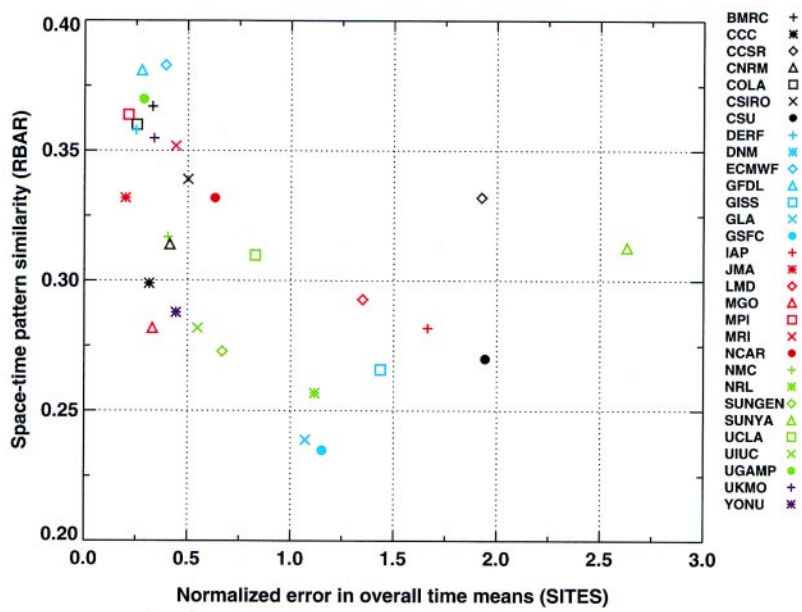

FIG. 15. A model performance diagram showing the AMIP models' simulation of the mean DJF sea level pressure during 1979–88 in terms of the normalized error of the time mean (SITES) and the evolution of the space–time pattern (RBAR), relative to the ECMWF reanalysis. (The DNM model's location at SITES = 9.35 and RBAR = 0.14 is not shown.)

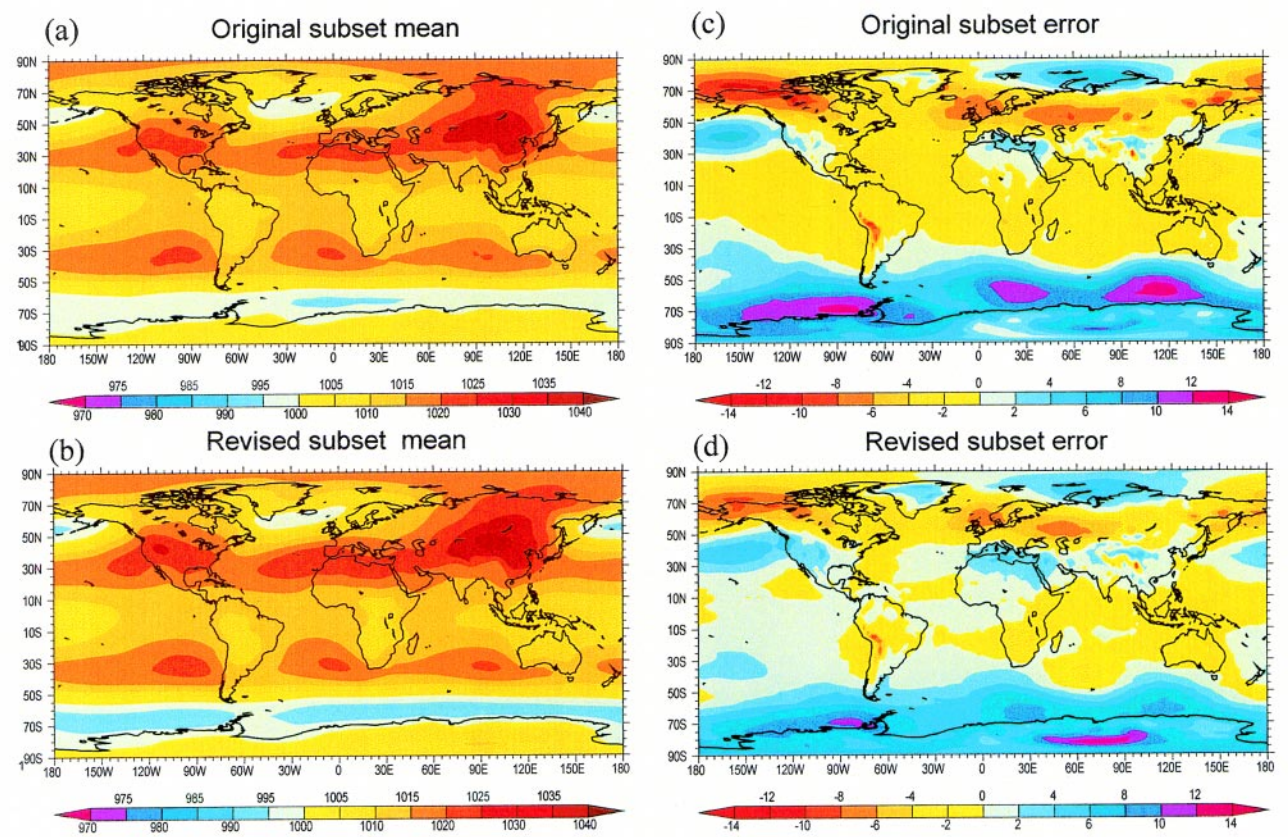

FIG. 16. The geographical distribution of mean sea level pressure (hPa) simulated in DJF of 1979–88 by the subset of 10 models that revisited AMIP with revised versions. (a) The mean of the original models' simulation and (b) the corresponding mean of the revised versions of the same set of models. (c) and (d) The errors of the original and revised model subset, respectively, relative to the ECMWF reanalysis.

clear that many statistical measures are required to adequately portray model performance and that a model's error budget is likely to be a unique complexion of interacting inadequacies.

5. Documenting model improvement

One of the purposes of AMIP was (and continues to be) promotion of the improvement of atmospheric GCMs. In this spirit approximately half of the participating modeling groups repeated the AMIP simulation with a revised version of their original AMIP model. The groups that completed such an AMIP "revisit" are identified in appendix A. Most of the revised models were intended to reduce specific systematic errors seen in the original AMIP versions and usually involved changes in the parameterization of cloudiness and/or convection. While these revisits have enabled the modeling groups to determine the extent to which their model revision has resulted in the anticipated improvement, here we focus on the revisits' improvement of the AMIP models as a whole. For this purpose we consider only the subset of the original AMIP models that have been revised and compare their ensemble mean performance with that of the original versions. In this way the influence of the unrevised models in the original AMIP ensemble is avoided.

In parallel with the analysis of the original AMIP ensemble mean given in section 2, we show in Fig. 16 the geographical distribution of sea level pressure (and its error relative to observations) given by the mean of the original model subset and by the mean of the corresponding revised model subset. On the whole, there is a small reduction of the models' error nearly everywhere, although the large-scale pattern of systematic error is unchanged. This result may be due to the fact that the models' resolution was not changed and that the revisions made may not have addressed the models' most important errors. The corresponding distributions of precipitation and its error are shown in Fig. 17, in which it may be noted that in some

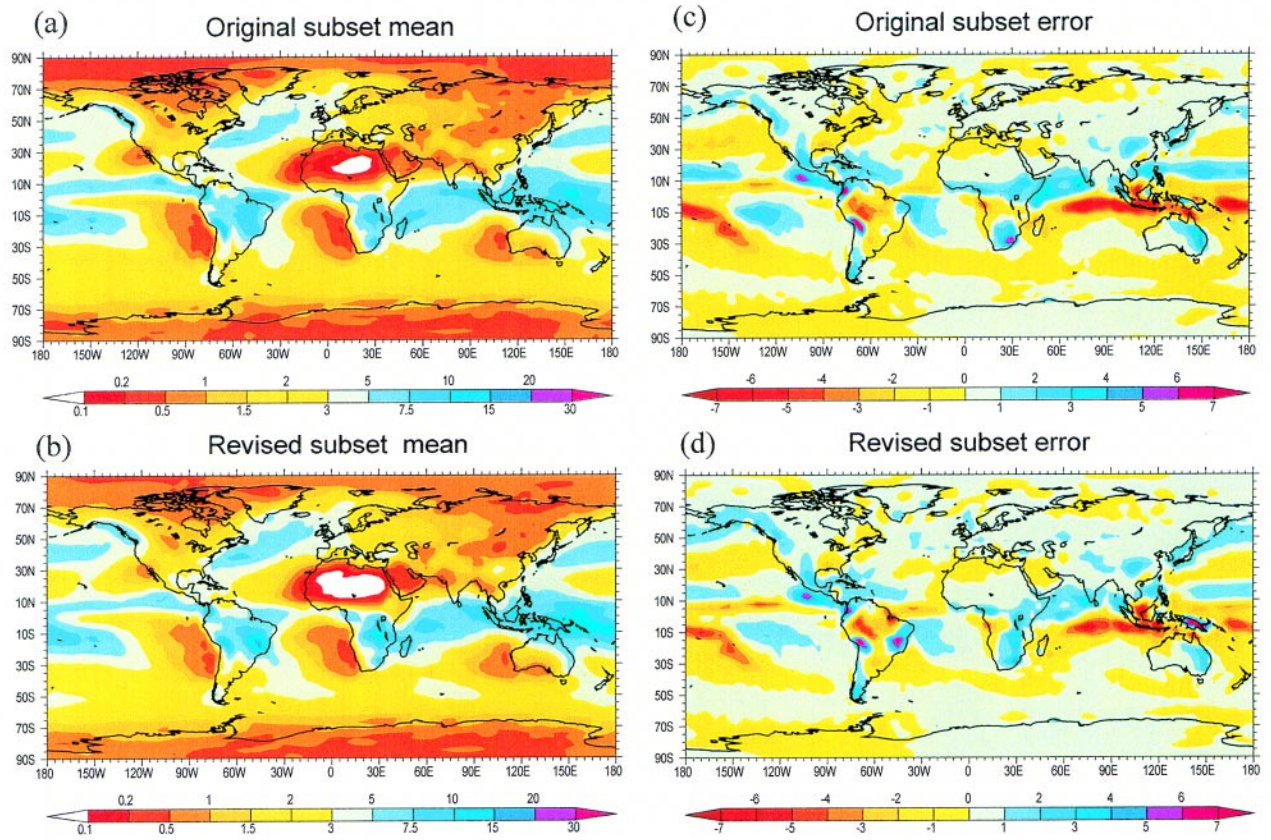

FIG. 17. As in Fig. 16 except for precipitation (mm day^{-1}), with the observed data from Xie and Arkin (1997).

areas the error of the ensemble mean has in fact increased with model revision. This result, however, may be partly due to the inadequacies of the observational estimate used.

The reduction of errors in the revised AMIP models' simulation of zonally averaged cloudiness and precipitation in DJF are given in Fig. 18. Here we see that the spread in simulated DJF cloudiness in the original model subset has been reduced, principally through the correction of several models whose original results were outliers. A similar reduction of model spread has, however, not occurred for DJF precipitation, and several models' results can be seen to have deteriorated. This same behavior is seen in other variables and other seasons (not shown), and indicates that in many cases model revision is only selectively effective in reducing systematic errors.

In order to provide an overall measure of the progress that has been achieved in AMIP model revision, the root-mean-square and percent error statistics shown in Table 1 for the complete original AMIP ensemble have been recalculated (for the Northern Hemisphere) for that subset of the original models that were revised and for the revised AMIP model themselves. The results are shown in Table 2 and indicate that on a hemispheric mean-seasonal basis, only the rms errors of the total cloudiness (and to a lesser extent the outgoing longwave radiation) have been significantly reduced in the revised subset of models compared to their original versions. By comparison with the Northern Hemisphere data in Table 1, however, we may note that the original versions of this subset of models generally had higher outgoing longwave radiation and cloudiness errors than did the complete original AMIP ensemble, and in this respect the revised model subset included some of the poorest models to begin with. In less than half of the cases shown in Table 2 have the models' errors been reduced in the revisions, and in most cases these changes are smaller than the observational uncertainties. Although some of the revised models were only slightly modified, the overall rate of model improvement may be judged by noting that the average "vintage" or year of production of the original AMIP models was 1991 and that of the revised AMIP models was 1995.

TABLE 2. As in Table 1 except for the subset of revised AMIP models in the Northern Hemisphere. The revised models are those of BMRC, CNRM, DERF, DNM, LMD, MPI, MRI, NRL, SUNGEN, and YONU (see appendix A).

Variable	Original models				Revised models			
	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON
Mean sea level pressure (hPa)	2.4 (39)	2.2 (56)	2.8 (60)	2.0 (49)	2.3 (38)	1.9 (50)	2.7 (57)	2.1 (51)
Surface air temprature (°C)	3.9 (23)	3.9 (31)	3.0 (39)	3.5 (31)	4.0 (24)	4.0 (32)	3.1 (40)	3.6 (32)
Temperature* at 200 hPa (°C)	3.9 (142)	5.2 (206)	4.4 (140)	4.3 (247)	4.2 (154)	5.6 (221)	4.5 (144)	4.3 (248)
Zonal wind at 200 hPa (m s^{-1})	5.2 (32)	5.2 (42)	4.6 (39)	3.5 (29)	4.3 (27)	5.0 (40)	4.8 (41)	3.9 (32)
Outgoing longwave radiation (W m^{-2})	10.2 (27)	11.4 (37)	14.5 (52)	13.2 (45)	9.8 (26)	10.6 (34)	13.6 (49)	12.3 (42)
Total cloudiness (%)	19.4 (100)	20.8 (120)	20.9 (115)	20.0 (112)	17.3 (89)	16.9 (98)	15.5 (86)	15.7 (88)
Precipitation (mm day^{-1})	1.2 (57)	1.4 (65)	1.9 (59)	1.5 (55)	1.2 (56)	1.5 (68)	2.0 (61)	1.5 (56)
Precipitation–evaporation (mm day^{-1})	1.5 (60)	1.3 (59)	1.6 (69)	1.7 (70)	1.4 (58)	1.3 (60)	1.8 (75)	1.7 (71)

*DNM not included.

6. Conclusions and future work

a. Outstanding modeling problems

From the analyses presented here and elsewhere, it is clear that much further work is needed to significantly reduce the errors of atmospheric GCMs. Continuing outstanding problems are the parameterization of clouds and their radiative interactions, the parameterization of convection and precipitation, and the portrayal of the interactions between the land surface and hydrologic processes. The increasing use of coupled atmosphere–ocean models for extended integrations has also emphasized the importance of an accurate portrayal of the surface fluxes in the marine boundary layer, although their effect on the sea surface temperature has been neglected in the case of AMIP. The future incorporation of interactive chemical and biological processes into atmospheric models, and the routine extension of the models into the upper atmosphere, will pose new challenges and opportunities for model improvement.

It should be recalled that a model's errors are defined with respect to observational data that are in many cases of limited quality and coverage, although the observed data used here are believed to have the broadest coverage available. Enhancements of the database through the development of new remote sensing capabilities and improvements in the retrieval and reanalysis of existing instrumental data are essential parts of a continuing model validation strategy.

b. AMIP's continuation

Following the discussion of the preliminary results of AMIP at the First International AMIP Scientific Conference (Gates 1995), PCMDI submitted a proposal for the continuation of the project. This initiative for an AMIP II was enthusiastically supported by the conference participants and was given widespread review and comment by the climate modeling and diagnostic communities during the following year (Gleckler 1996).

The principal planned enhancements of AMIP II relative to the original AMIP are improvement of the

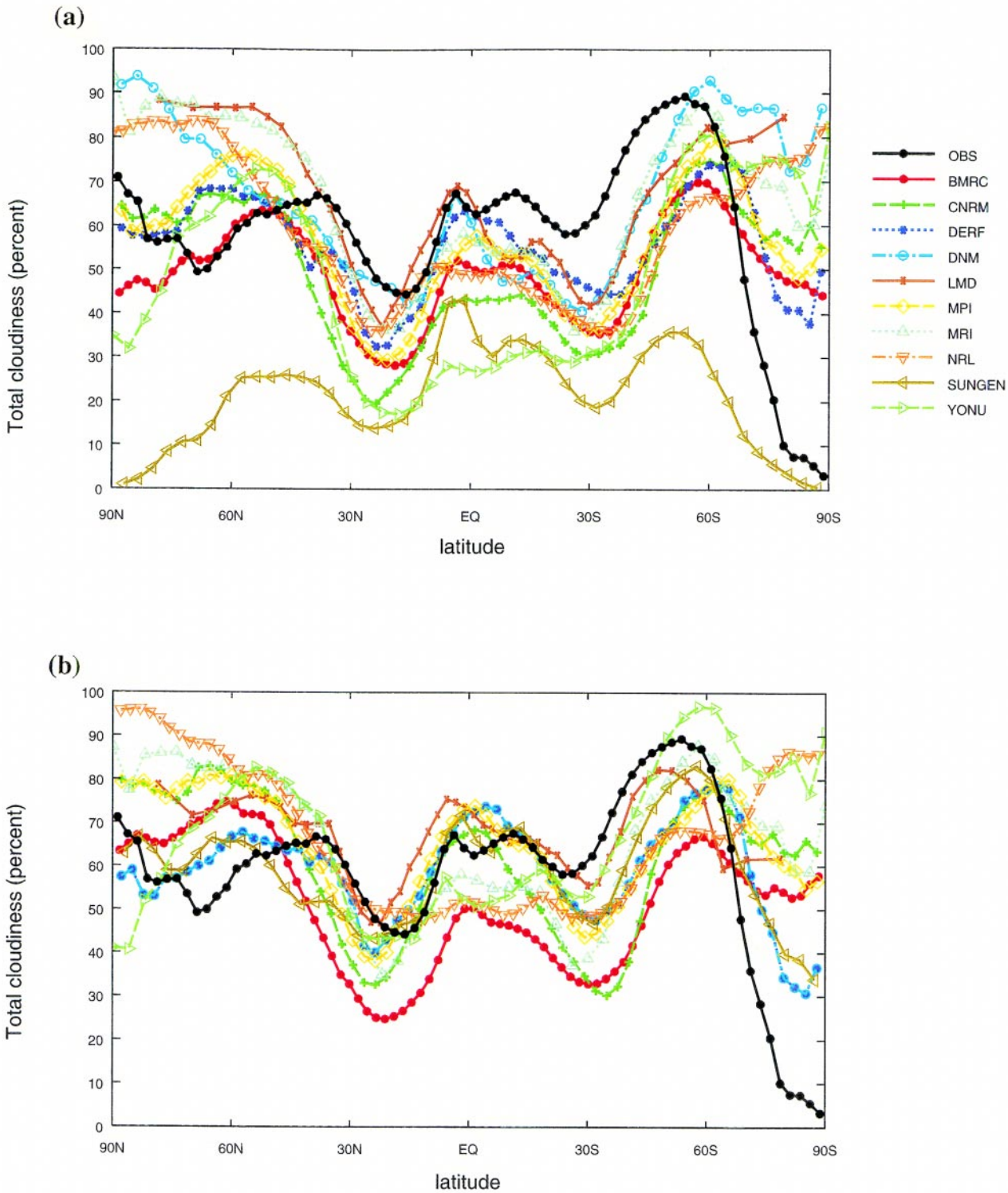

FIG. 18. The zonally averaged total cloudiness and precipitation simulated in DJF of 1979–88 by the subset of 10 models that revisited AMIP with revised versions. (a) and (b) The cloudiness from the original models and revised models, respectively, with the observed data from ISCCP for 1983–90 (Rossow et al. 1991) given by the solid black line.

experimental design; additional diagnosis of an expanded model output; the establishment of standards and software for data management, transmission, and analysis; the inclusion of numerical experimentation subprojects in addition to diagnostic subprojects; clarification of the participation protocol; and increased use of the Internet

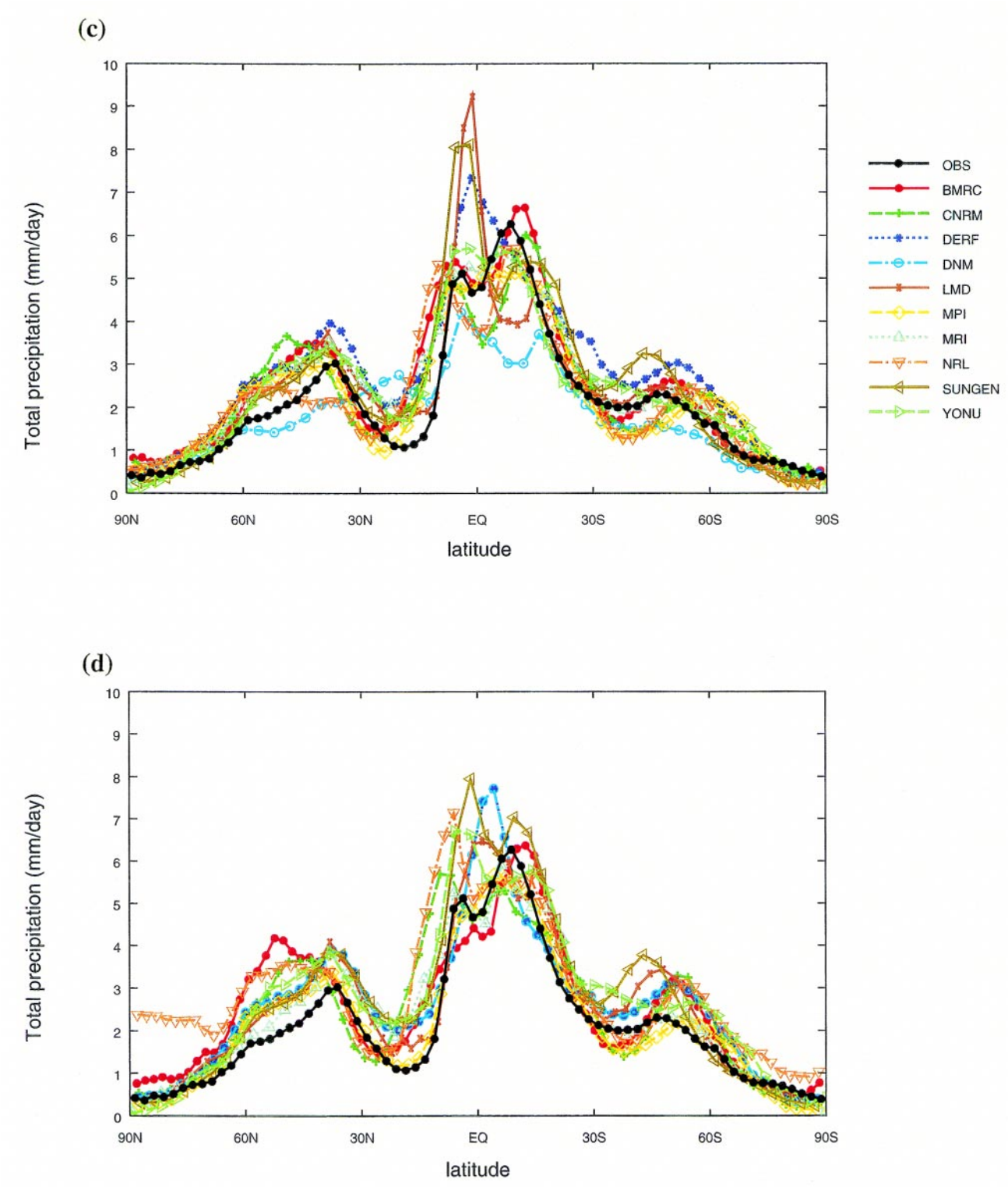

FIG. 18. *(Continued)* (c) and (d) The corresponding precipitation, with the observed data from the NCEP database (Xie and Arkin 1997).

(http://www-pcmdi.llnl.gov/amip) for project communication and coordination. It is expected that AMIP II will become an accepted community protocol for the continued diagnosis, validation, and improvement of the atmospheric GCMs, and will serve as a benchmark reference for the atmospheric component of coupled models.

c. AMIP's legacy

Beyond the ready availability of a decade of standardized and quality controlled output for some 50 variables from virtually all atmospheric GCMs as of the early 1990s (see http://www-pcmdi.llnl.gov/pcmdi/archives.html), the legacy of AMIP includes a suite of improved software for data storage, access, analysis, and visualization (Williams 1997), and documentation of the physics and numerics of the AMIP models in a common comprehensive format (Phillips 1994, 1996). Considerable effort has also been expended on the assembly and maintenance of an observational database to support model diagnosis and validation [Fiorino (1998); see http://www-pcmdi.llnl.gov/obs].

A further legacy of AMIP is the impetus it has provided for the international coordination of the diagnosis, validation, and intercomparison of climate models. Using its experience in supporting AMIP, PCMDI has actively supported other model intercomparison projects under the auspices of the World Climate Research Programme (WCRP), including the Paleoclimate Modelling Intercomparison Project in coordination with PAGES, and the Coupled Model Intercomparison Project, the Study of Tropical Oceans in Coupled Models, and the El Niño Simulation Intercomparison Project (ENSIP) in coordination with CLIVAR. PCMDI has also cooperated with the Project for the Intercomparison of Land-Surface Parameterization Schemes of Global Energy and Water Cycle Experiment, and has assisted the GCM-Reality Intercomparison Project for the Stratosphere in coordination with SPARC. AMIP has also served as a prototype for the intercomparison of sea-ice models and ocean carbon cycle models, and provides an approach that may be followed in the intercomparison of ocean models as well. Collectively, these projects are providing the framework for an international climate modeling and diagnostic infrastructure that should broaden, improve, and accelerate many aspects of climate research.

Acknowledgments. The authors gratefully acknowledge the technical assistance of Emmanuelle Cohen-Solal, Charles O'Connor, and Yi Zhang, and the editorial assistance of Harriet Moxley and Anna McCravy in the preparation of this report. Thanks are also due to the original AMIP panel members Lennart Bengtsson, George Boer, and David Burridge, who assisted the chairman (W. L. Gates) in the guidance of the project; to the AMIP modeling groups for their participation and cooperation without which this project would have been impossible; and to the anonymous reviewers for their assistance in improving the original manuscript. We also thank the DOE Office of Biological and Environmental Research for their vision and sustained support. This research was performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

Appendix A: Model identification

The groups that participated in the AMIP experiment are identified below by their institutional acronyms. The technical identification of the model(s) used by each group is given in parentheses; the first model listed is the group's original AMIP submission and the second model (where listed) is the group's revised AMIP model. (Both BMRC and LMD completed second revisits, whose performances are not considered here.) A comprehensive documentation of the AMIP models is given in Phillips (1994, 1996) and is available on the World Wide Web (see http://www-pcmdi.llnl.gov/modeldoc/amip).

BMRC	Bureau of Meteorology Research Centre, Melbourne, Australia (2.3, 3.7, 3.7.1)
CCC	Center for Climate Modelling and Analysis, Victoria, British Columbia, Canada (GCM II)
CCSR	Center for Climate System Research, Tokyo, Japan (CCSR/NIES AGCM)
CNRM	Centre National de Recherches Météorologiques, Toulouse, France (EMERAUDE, ARPEGE cy II)
COLA	Center for Ocean–Land–Atmosphere Studies, Calverton, Maryland (1.1)
CSIRO	Commonwealth Scientific and Industrial Research Organisation, Melbourne, Australia (CSIRO9 Mark1)
CSU	Colorado State University, Fort Collins, Colorado (91)
DERF	Dynamic Extended Range Forecasting, Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey (SM392.2, SM195)
DNM	Department of Numerical Mathematics, Moscow, Russia (A5407.V1, A5407.V2)
ECMWF	European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom (Cy36)
GFDL	Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey (CDG1)

GISS	Goddard Institute for Space Studies, New York, New York (II Prime)
GLA	Goddard Laboratory for Atmospheres, Greenbelt, Maryland (GCM-01.0 AMIP-01)
GSFC	Goddard Space Flight Center, Greenbelt, Maryland (GEOS-1)
IAP	Institute of Atmospheric Physics, Beijing, China (IAP-2L)
JMA	Japan Meteorological Agency, Tokyo, Japan (GSM 8911)
LMD	Laboratoire de Météorologie Dynamique, Paris, France (LMD5, LMD6b, LMD6s)
MGO	Main Geophysical Observatory, St. Petersburg, Russia (AMIP92)
MPI	Max Planck Institute for Meteorology, Hamburg, Germany (ECHAM3, ECHAM4)
MRI	Meteorological Research Institute, Tsukuba, Japan (GCM-II, GCM-IIb)
NCAR	National Center for Atmospheric Research, Boulder, Colorado (CCM2)
NMC	National Centers for Environmental Prediction, Washington, D.C. (MRF)
NRL	Naval Research Laboratory, Monterey, California (NOGAPS3.2, NOGAPS 3.4)
RPN	Recherche en Prévision Numérique, Dorval, Quebec, Canada (NWP-D40P29)
SUNGEN	The University at Albany, State University of New York, Albany, New York–NCAR, Boulder, Colorado (GENESIS 1.5, GENESIS 1.5A)
SUNYA	The University at Albany, State University of New York, Albany, New York (CCM1-TG)
UCLA	University of California, Los Angeles, Los Angeles, California (AGCM 6.4)
UGAMP	Universities' Global Atmospheric Modelling Project, Reading, United Kingdom (UGCM 1.3)
UIUC	University of Illinois, Urbana–Champaign, Urbana, Illinois (MLAM-AMIP)
UKMO	United Kingdom Meteorological Office, Bracknell, United Kingdom (HADAM 1)
YONU	Yonsei University, Seoul, South Korea (Tr 5.1, Tr 7.1)

Appendix B: Diagnostic subprojects

The AMIP I diagnostic subprojects that were established to examine model performance in terms of specific processes and regional phenomena are listed below. Further information on the subprojects is available on the Internet (http://www-pcmdi.llnl.gov/amip/AMIP1/AMIPsubprojects.html).

Subproject 1:	Tropical variability at synoptic to intraseasonal timescales
Subproject 2:	Low-frequency variability
Subproject 3:	Cyclone frequencies and extratropical intraseasonal variability
Subproject 4:	Clear-sky greenhouse sensitivity and water vapor distribution
Subproject 5:	Ocean surface boundary fluxes
Subproject 6:	Monsoons*
Subproject 7:	Hydrologic processes
Subproject 8:	Polar phenomena and sea ice*
Subproject 9:	High-latitude Southern Hemisphere circulation
Subproject 10:	Atmospheric blocking
Subproject 11:	Atmospheric humidity and soil moisture
Subproject 12:	Land surface processes*
Subproject 13:	Cloudiness variations
Subproject 14:	Cloud-radiative forcing
Subproject 15:	Atmospheric angular momentum fluctuations
Subproject 16:	Stratospheric circulation
Subproject 17:	Multiscale water/energy balances*
Subproject 18:	Extreme events
Subproject 19:	Temperature validation by microwave sounding unit data
Subproject 20:	Circulation features of southern Africa
Subproject 21:	Monthly/daily surface climatologies and regional anomalies
Subproject 22:	Atmospheric energetics in the wavenumber domain
Subproject 23:	Variations of the centers of action
Subproject 24:	Caspian Sea regional climates
Subproject 25:	East Asian climates
Subproject 26:	Monsoon precipitation

*Denotes coordination with other WCRP projects: No. 6 with MONEG and WGNE, No. 8 with ACSYS, No. 12 with PILPS and WGNE, No. 17 with GCIP/GEWEX.

Appendix C: Observational data

This appendix summarizes the monthly mean observational data used for validation of the AMIP simulations, together with a justification of their selection. These data were drawn from the PCMDI Observational Data Set (PODS) constructed by Fiorino (1998), which is described on the Internet (Table C1; see http://www-pcmdi.llnl.gov/obs/pods).

In these datasets there is a mixture of directly observed data (either in situ or remotely sensed) and that supplied by model-assisted assimilation and analysis. If the observational coverage is widespread, then those datasets' quality is generally higher than in those more heavily dependent on models. Except for the precipitation, cloudiness, fluxes, and surface temperature, the reanalyses were the datasets of choice.

TABLE C1. Monthly mean observational variables and data sources used for validation of the AMIP simulations with justification of their selection.

Variable	Data source	Selection justification
Temperature, wind, and sea level pressure	ECMWF (Gibson et al. 1997) NCEP (Kalnay et al. 1996)	Reanalyses
Surface air temperature	PCMDI (Fiorino 1997; da Silva et al. 1994c; Jones 1988; Schubert et al. 1992)	Merger of best available data
Total precipitation	NCEP (Xie and Arkin 1997)	Contains largest amount of land and satellite data
Net surface heat flux, surface evaporation	COADS (da Silva et al. 1994b)	Most observations with balance-corrected fluxes
Total cloud amount	ISCCP (Rossow et al. 1991)	Best coverage (although only for half AMIP decade)
Outgoing longwave radiation	NESDIS (Gruber and Krueger 1984)	Best coverage (operational product)

References

Boer, G. J., and Coauthors, 1992: Some results from an intercomparison of the climates simulated by 14 atmospheric general circulation models. *J. Geophys. Res.,* **97,** 12 771–12 786.

da Silva, A. M., C. C. Young, and S. Levitus, 1994a: *Atlas of Surface Marine Data, Vol. 2: Anomalies of Directly Observed Quantities.* NOAA Atlas NESDIS 7, U.S. Department of Commerce, Washington, DC, 416 pp. [Available from National Oceanographic Data Center, User Services Group, NOAA/NESDIS E/OC1, SSMC3, 4th Floor, 1315 East-West Highway, Silver Spring, MD 20910-3282.]

——, ——, and ——, 1994b: *Atlas of Surface Marine Data, Vol. 3: Anomalies of Heat and Momentum Fluxes.* NOAA Atlas NESDIS 8, U.S. Department of Commerce, Washington, DC, 411 pp. [Available from National Oceanographic Data Center, User Services Group, NOAA/NESDIS E/OC1, SSMC3, 4th Floor, 1315 East-West Highway, Silver Spring, MD 20910-3282.]

——, ——, and ——, 1994c: *Atlas of Surface Marine Data, Vol. 4: Anomalies of Fresh Water Fluxes.* NOAA Atlas NESDIS 9, U.S. Department of Commerce, Washington, DC, 308 pp. [Available from National Oceanographic Data Center, User Services Group, NOAA/NESDIS E/OC1, SSMC3, 4th Floor, 1315 East-West Highway, Silver Spring, MD 20910-3282.]

Fiorino, M., cited 1997: A merged surface air temperature data set for the validation of AMIP I. PCMDI Web Rep., Lawrence Livermore National Laboratory. [Available online at http://www-pcmdi.llnl.gov/obs/ipcc/.]

——, cited 1998: The PCMDI Observational Data Set. PCMDI Rep., Lawrence Livermore National Laboratory. [Available online at http://www-pcmdi.llnl.gov/obs/pods.]

Gates, W. L., 1992: AMIP: The Atmospheric Model Intercomparison Project. *Bull. Amer. Meteor. Soc.*, **73,** 1962–1970.

——, 1995: An overview of AMIP and preliminary results. *Proc. of the First Int. AMIP Scientific Conference,* WCRP-92, WMO TD-No. 732, Monterey, CA, World Meteorological Organization, 1–8.

——, and Coauthors, 1996: Climate models—Evaluation. *Climate Change 1995: The Science of Climate Change*, J. T. Houghton et al., Eds., Cambridge University Press, 229–284.

Gibson, J. K., P. Kallberg, S. Uppala, A. Hernandez, A. Nomura, and E. Serrano, 1997: ERA description. ECMWF Reanalysis Project Rep. Series 1, European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom, 66 pp.

Gleckler, P. J., Ed., 1996: AMIP Newsletter. No. 8, PCMDI, Lawrence Livermore National Laboratory, Livermore, CA. [Available online at http://www-pcmdi.llnl.gov/amip/NEWS/amipnl8.html.]

Gruber, A., and A. F. Krueger, 1984: The status of the NOAA outgoing longwave radiation data set. *Bull. Amer. Meteor. Soc.*, **65,** 958–962.

Hulme, M., K. R. Briffa, and P. D. Jones, 1993: General circulation model validation and climate change detection. Climatic Research Unit Rep., Norwich, United Kingdom, 24 pp. [Available from Climatic Research Unit, School of Environmental

Sciences, University of East Anglia, Norwich NR4 7TJ, United Kingdom.]

Jones, P. D., 1988: Hemispheric surface air temperature variations: Recent trends and an update to 1987. *J. Climate,* **1,** 654–660.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471.

Phillips, T. J., 1994: A summary documentation of the AMIP models. PCMDI Rep. 18, Lawrence Livermore National Laboratory, 343 pp. [Available from PCMDI, Lawrence Livermore National Laboratory, Livermore, CA 94550.]

——, 1996: Documentation of the AMIP models on the World Wide Web. *Bull. Amer. Meteor. Soc.,* **77,** 1191–1196. [Available online at http:/www-pcmdi.llnl.gov/modeldoc/amip.]

Preisendorfer, R. W., and T. P. Barnett, 1983: Numerical model-reality intercomparison tests using small-sample statistics. *J. Atmos. Sci.,* **40,** 1884–1896.

Rossow, W. B., L. C. Garder, P. J. Lu, and A. W. Walker, 1991: International Satellite Cloud Climatology Project (ISCCP): Documentation of cloud data. WMO/TD-No. 266, World Meteorological Organization, 76 pp.

Santer, B. D., K. E. Taylor, and L. C. Corsetti, 1995: Statistical evaluation of AMIP model performance. *Proc. of the First Int. AMIP Scientific Conference,* WCRP-92, WMO TD-732, Monterey, CA, World Meteorological Organization, 13–18.

Schubert, S., C.-Y. Wu, J. Zero, J.-K. Schemm, C.-K. Park, and M. Saurez, 1992: Monthly means of selected climate variables for 1985–1989. NASA Tech. Memo., Goddard Space Flight Center, 376 pp.

Trenberth, K., and D. Shea, 1987: On the evolution of the Southern Oscillation. *Mon. Wea. Rev.,* **115,** 3078–3096.

——, and J. Hurrell, 1994: Decadal atmosphere–ocean variations in the Pacific. *Climate Dyn.,* **9,** 303–319.

Wigley, T. M. L., and B. D. Santer, 1990: Statistical comparison of spatial fields in model validation, perturbation and predictability experiments. *J. Geophys. Res.,* **95,** 851–865.

Williams, D. N., 1997: The PCMDI software system: Status and future plans. PCMDI Rep. 44, Lawrence Livermore National Laboratory, 29 pp. [Available from PCMDI, Lawrence Livermore National Laboratory, Livermore, CA 94550.]

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