

**PCMDI Report No. 45**

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

W.L. Gates, J. S. Boyle, C. C. Covey, C. G. Dease, C. M. Doutriaux,  
R. S. Drach, M. Fiorino, P. J. Gleckler, J. J. Hnilo, S. M. Marlais,  
T. J. Phillips, G. L. Potter, B. D. Santer, K. R. Sperber,  
K. E. Taylor, and D. N. Williams

The Program for Climate Model Diagnosis and Intercomparison  
Lawrence Livermore National Laboratory  
Livermore, CA 94551

April 1998

## DISCLAIMER

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

This report has been reproduced  
directly from the best available copy.

Available to DOE and DOE contractors from the  
Office of Scientific and Technical Information  
P.O. Box 62, Oak Ridge, TN 37831  
Prices available from (615) 576-8401, FTS 626-8401

Available to the public from the  
National Technical Information Service  
U.S. Department of Commerce  
5285 Port Royal Rd.,  
Springfield, VA 22161

## 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 AMIP was, and continues to be, the comprehensive evaluation of the performance of atmospheric GCMs on climate and higher-frequency time-scales 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 atmospheric evolution subject to the observed sequence of monthly-averaged global sea-surface temperature and sea-ice distributions during the decade 1979-1988, along with standardized values of the solar constant and atmospheric CO<sub>2</sub> concentration. An agreed-upon standard output (see <http://www-pcmdi.llnl.gov/amip/output/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), we do not emphasize here the results of individual models nor 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 these 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 the Appendix.

## 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 consider selected variables for the illustrative season December-January-February (DJF) during the decade 1979-1988. The observed data used for model validation are from the ECMWF reanalysis for 1979-1988 whenever possible, although other sources are used for surface temperature and fluxes, precipitation, cloudiness and radiation.

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; 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  $2\ 1/2^\circ \times 2\ 1/2^\circ$  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 considerable scatter among the models' results in high latitudes, 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 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). 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 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), the ensemble standard deviation (Fig. 3c) and the ensemble mean error (Fig. 3d). Here (and in the other seasons) 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. As was the case for sea-level pressure, the ensemble distribution is superior to that of any individual model.

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. 4a, along with the observed distribution taken from the ECMWF reanalysis. The models' results reassuringly cluster around the observed data, although poleward of about 60 deg latitude there is a marked increase in the intermodel scatter. Much of this disagreement is likely due to the models' differences in orography and their procedures for reduction to sea level.

The zonally-averaged distributions of the models' DJF surface air temperature are shown in Fig. 4b, along with the observed distribution. 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 reflects 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. 4c, along with the distribution given by the ECMWF reanalysis for the AMIP decade. Except for a few apparent 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. 4d, along with the observational estimate given by the COADS data. 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 for a model's readiness to be coupled with an ocean GCM (Gleckler et al., 1995). Despite large observational uncertainties, examination of the surface net shortwave radiation and latent heat flux (the dominant components of surface heating) has revealed systematic errors in both components (Gleckler and Weare, 1997).

The zonally-averaged distributions of the outgoing longwave radiation (OLR) at the top of the atmosphere are shown in Fig. 5a, along with the observed distribution given by NCEP operational data for the AMIP decade. 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 lower latitudes near 30N and 30S, 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. 5b, along with the observed distribution given by ISCCP data during 1983-1990. 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. 5c, along with an estimate of the observed distribution from the merged NCEP database. 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 mid-latitudes of both hemispheres. A relatively large scatter among the models' results is also evident in Fig 5c, especially in the equatorial region, although there are apparent 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 30S, along with a systematic poleward displacement by about 10 deg latitude of the precipitation minima in the subtropics of both hemispheres.

Finally, the zonally-averaged distributions of the precipitation minus evaporation (P-E) simulated in DJF by the AMIP models is shown in Fig. 5d, together with an estimate of the observed distribution from COADS. In spite of the difficulties in simulating precipitation noted above, the bulk of the models successfully reproduce the observed structure 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 even more limited success of the simulation of the surface heat flux seen in Fig. 4d, indicates that considerable improvement is needed before the AMIP models could be successfully coupled to global ocean models without the use of flux corrections.

c. Meridional sections

A complement to the data shown in Figs. 1-5 given by the portrayal of selected simulated variables in the latitude-pressure meridional section. This is given for the DJF ensemble mean zonal wind in Fig. 6a, which is seen to closely resemble the observed distribution in Fig. 6b. The errors of the ensemble mean (Fig. 6d) 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. 6c) 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. 7a. The observed distribution shown in Fig. 7b is structurally similar to the models' portrayal of the Hadley circulation between about 30S and 50N, although the models systematically underestimate the observed strength. This is clearly shown in Fig. 7d where the dominant error pattern resembles the Hadley circulation itself. The variability among the models (Fig. 7c) 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, we note that the mean sea-level pressure rms error is generally larger in the Southern Hemisphere, with a maximum in JJA. The rms errors of the surface air temperature, on the other hand, are systematically smaller in the Southern Hemisphere, reflecting the predominance of the prescribed sea-surface temperatures. Interestingly enough, the rms errors of 200 hPa temperature are largest in the Southern Hemisphere in summer and autumn, while the rms error of the 200 hPa zonal wind is a minimum in the autumn in both hemispheres.

The rms error of the outgoing longwave radiation (OLR) shows relatively little seasonal or hemispheric variation, and is likely a result of 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 Hemisphere than in the Northern Hemisphere. The rms errors of precipitation and of precipitation minus evaporation are nearly the same in all seasons in both hemispheres, and represent a substantial fraction of the globally-averaged annual precipitation of  $2.7 \text{ mm day}^{-1}$  given by Xie and Arkin (1997). This reflects the continuing difficulty of both models and observations in accurately estimating this component of the hydrological cycle.

The AMIP performance errors given in Figs. 1-7 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, by the PCMDI staff, and by the modeling groups themselves. 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 (<http://www-pcmdi.llnl.gov/amipabs.html>). In particular, studies such as those by Gleckler et al. (1995), Srinivasan et al. (1995), Weare and Mokhov (1995), D'Andrea et al. (1996), Lau et al. (1996), Slingo et al. (1996), Tao et al. (1996), Weare et al. (1996), Sperber and Palmer (1996), Duvel et al. (1997), Ferranti et al. (1997), Gaffen et al. (1997), Gleckler and Weare (1997), Joubert (1997), Saji and Goswami (1997), Sperber et al. (1997), Zhang et al. (1997), Frei and Robinson (1998), Gadgil and Sajani (1998) Hide et al. (1998), and Mao and Robock (1998) have served to show the presence of important systematic errors in the AMIP models' simulations of a wide variety of processes and regional phenomena.

### **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.

#### **a. Seasonal variability**

Although there are many ways of portraying the seasons, a compact form which preserves geographical dependence is the amplitude of the mean seasonal cycle. This is shown in Fig. 8a for the average of the first annual harmonic of the models' simulation of mean sea-level pressure, and in Fig. 8b for the ECMWF reanalysis over the AMIP decade. 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. 8c is seen to closely resemble that from the reanalysis shown in Fig. 8d, 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 upwards of 90% of the total variability in the tropics and subtropics, a statistic which the models slightly overestimate. The models' average portrayal of the mean seasonal cycle of other variables (not shown) are 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 use here the mean sea-level pressure averaged over 25S-15S and between the longitudes 125E-135E and 135W-145W as suggested by Trenberth and Shea (1987). 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. 9a along with the corresponding observational result from the NCEP reanalysis. On the whole, the models simulate the broad aspects of the observed ENSO variations reasonably well (as would 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 major ENSO events.

As a measure of mid-latitude interannual variability, we have chosen the mean sea-level pressure averaged over the area 30N-65N, 160E-220E 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. 9b, along with the observed variability. (As in Fig. 9a, the mean annual cycle has been removed and variations less than 8 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 Taylor (unpublished) and illustrated in Fig. 10 for the AMIP model's simulation of the total space-time variability of the monthly-averaged mean sea-level pressure. Here the distance from the origin is equal to the standard deviation of the field (normalized by the observed standard deviation), while the distance from the "reference" point (which is here the ECMWF reanalysis) is equal to the rms pattern difference between the observed and modeled fields (again normalized by the observed standard deviation). The cosine of the polar angle is equal to the correlation between the simulated and observed monthly means. 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 rms error and high correlation. All statistics were computed over the full AMIP period (120 months) after interpolating modeled and observed data to a common 4° by 5° latitude-longitude grid and the removal of any global mean bias. It is clear from Fig. 10 that the AMIP models differ widely in their ability to simulate the total space-time variability of sea-level pressure, and that the model ensemble (which, if superior to any individual model, would lie near the dotted arc and closer to the ECMWF reanalysis

than the models shown in Fig. 10) still possesses considerable disagreement with observations in terms of the pattern of variability.

Fig. 10 may also be used to assess the practical limit of accuracy which we may expect a model to approach, given its inevitable errors and the observational uncertainties. This is estimated by the distance of the NCEP reanalysis relative to the reference ECMWF reanalysis, which is less than that of any AMIP model. This figure also shows that models with the same rms error (as given by the distance to the ECMWF reference) may differ substantially in the amplitude of their variability while having a similar mean correlation with observations (e.g., the LMD and GLA models). A few modeling groups have carried out several AMIP simulations with their models, in which each simulation was identical except for the initial conditions specified. The properties of these ensembles lie close to the location of the original single simulations shown in Fig. 10, from which we may conclude that the scatter of the AMIP models plotted in the figure cannot be explained as simply due to the differences that might be expected from a sensitive dependence on initial conditions. Considerable improvement is clearly possible in the models' simulations of the space-time variance.

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. 11. 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, as might be expected from the constraints of the AMIP experiments, while they 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 attributable to the extratropical cold bias, which enhances the meridional temperature gradient.

#### 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). Fig. 12 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. Here the abscissa SITES is a dimensionless measure of overall (squared) differences in the annual mean state, standardized by the combined temporal variance 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) measures whether modeled and observed spatial anomaly fields 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 pattern evolution.

Fig. 12 shows that there is a wide spread among the models' results; there is almost a factor of two 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 are larger than the current observational uncertainty (as given by the comparison between the ECMWF and NCEP reanalyses, which lie at SITES = 0, RBAR = 1, and SITES = 0.05, RBAR = 0.98, respectively). Model mean sea-level pressure errors 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 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. 8 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 good agreement with observation, 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 reanalysis for the ECMWF reanalysis used in Fig. 12. The primary conclusions were not modified, but it was noted that some changes occur in the relative location of some models. In Fig. 12 the MPI model has the second smallest error in the overall mean state (i.e., a small SITES value), 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 reanalysis is used as the reference observational dataset, the SITES score of the NMC model (which is a descendant of the model used in the NCEP reanalysis) is improved and that of the ECMWF model is degraded. This suggests that both reanalyses contain a non-trivial model "imprint", particularly over data-poor regions. 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 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 the Appendix . 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 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. 13 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. The corresponding distributions of precipitation and its error are shown in Fig. 14, in which it may be noted that in some areas the error of the ensemble mean has in fact increased with model revision.

The reduction of errors in the revised AMIP models' simulation of zonally-averaged cloudiness and precipitation in DJF are given in Fig. 15. 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 a measure of the progress that has been achieved in AMIP model revision, the root-mean-square error statistics shown in Table 1 for the complete original AMIP ensemble have been recalculated (for the Northern Hemisphere) for the original and revised AMIP model subsets. The results are shown in Table 2, and indicate that on a hemispheric mean seasonal basis, only the rms error of the total cloudiness has 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 models have generally higher outgoing longwave radiation and cloudiness errors than does the complete original AMIP ensemble. 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 thirty-one AMIP models was 1991 and that of the ten revised AMIP models was 1995.

## 6. CONCLUSIONS AND FUTURE PLANS

### a. Outstanding modeling problems

From the analysis 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. With the future incorporation of interactive chemical and biological processes into atmospheric models, and the routine extension of the models into the upper atmosphere, the representation of the direct and indirect effects of aerosols 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 be the most appropriate and accurate 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 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 (<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 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 fifty variables from virtually all atmospheric GCMs as of the early 1990's (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, including the Paleoclimate Modelling Intercomparison Project (PMIP) in coordination with PAGES, and the Coupled Model Intercomparison Project (CMIP), the Study of Tropical Oceans in Coupled Models (STOIC) and the El Nino Simulation Intercomparison Project (ENSIP) in coordination with CLIVAR. PCMDI has also cooperated with the Project for the Intercomparison of Land-surface Parameterization Schemes (PILPS) of GEWEX, and has assisted the GCM-Reality Intercomparison Project for the Stratosphere (GRIPS) 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 in the preparation of this report. Thanks are also due the World Climate Research Programme's AMIP Panel for their advice, and of course to the AMIP modeling groups for their participation and cooperation. We also thank the DOE Office of Biological and Environmental Research (BER) 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 No. W-7405-Eng-48.

## REFERENCES

- Boyle, J.S., 1996: Intercomparison of low-frequency variability of the global 200 hPa circulation for AMIP simulations. PCMDI Report No.32, Lawrence Livermore National Laboratory, Livermore, CA, 53 pp.
- D'Andrea, F., S. Tibaldi, M. Blackburn, G. Boer, M. Déqué, M.R. Dix, B. Dugas, L. Ferranti, T. Iwasaki, A. Kitoh, V. Pope, D. Randall, E. Roeckner, C. Hall, D. Straus, W. Stern, H. van den Dool, and D. Williamson, 1996: Northern Hemisphere atmospheric blocking as simulated by 15 atmospheric general circulation models in the period 1979-1988. WCRP-96, WMO/TD-No. 784, World Climate Research Programme/World Meteorological Organization, Geneva, 25 pp.
- 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.
- \_\_\_\_\_, 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.
- \_\_\_\_\_, 1994c: Atlas of Surface Marine Data, Vol. 4: Anomalies of Fresh Water Fluxes. NOAA Atlas NESDIS 9, U.S. Department of Commerce, Washington, DC.
- Duvel, J.P., S. Bony, H. Le Treut and Participating AMIP Groups, 1997: Clear-sky greenhouse effect sensitivity to sea surface temperature changes: An evaluation of AMIP simulations. *Climate Dyn.*, 13, 259-273.
- Ferranti, L., J.M. Slingo, T.N. Palmer and B.J. Hoskins, 1997: Relations between interannual and intraseasonal monsoon variability as diagnosed from AMIP integrations. *Quart. J. Roy. Meteor. Soc.*, 123, 1323-1357.
- Fiorino, M., 1998: The PCMDI Observational Data Set. PCMDI Report, Lawrence Livermore National Laboratory, Livermore, CA (in press).
- Frei, A., and D. Robinson, 1998: Evaluation of snow extent and its variability in the Atmospheric Model Intercomparison Project. *J. Geophys. Res.* (in press).
- Gadgil, S., and S. Sajani, 1998; Monsoon precipitation in the AMIP runs. *Climate Dyn.* (in press).
- Gaffen, D.J., R.D. Rosen, D.A. Salstein and J.S. Boyle, 1997: Evaluation of tropospheric water vapor simulations from the Atmospheric Model Intercomparison Project. *J. Climate*, 10, 1648-1661.
- Gates, W.L., 1992: AMIP: The Atmospheric Model Intercomparison Project. *Bull. Amer. Meteor. Soc.*, 73, 1962-1970.

- Gates, W.L., 1995: An overview of AMIP and preliminary results. In Proceedings First International AMIP Scientific Conference (15-19 May 1995, Monterey, CA), WCRP-92, WMO TD-No. 732, World Meteorological Organization, Geneva, pp. 1-8.
- Gates, W.L., A. Henderson-Sellers, G.J. Boer, C.K. Folland, A. Kitoh, B.J. McAvaney, F. Semazzi, N. Smith, A.J. Weaver and Q.-C. Zeng, 1996: Climate Models - evaluation. In *Climate Change 1995: The Science of Climate Change*. J.T. Houghton et al. eds., Cambridge University Press, pp. 229-284.
- Gibson, J.K., P. Kallberg, S. Uppala, A. Hernandez, A. Nomura and E. Serrano, 1997: ERA description. ECMWF Reanalysis Project Report Series No.1, European Centre for Medium-range Weather Forecasts, Reading, UK, 66 pp.
- Gleckler, P.J., D. Randall, G. Boer, R. Colman, M. Dix, V. Galin, M. Helfand, J. Kiehl, A. Kitoh, W. Lau, X.-Z. Liang, V. Lykossov, B. McAvaney, K. Miyakoda, S. Planton and W. Stern, 1995: Cloud-radiative effects on implied oceanic energy transports as simulated by atmospheric general circulation models. *Geophys. Res. Letters*, 22, 791-794.
- Gleckler, P.J. (editor), 1996: AMIP Newsletter No. 8, PCMDI, Lawrence Livermore National Laboratory, Livermore, CA (Available at <http://www-pcmdi.llnl.gov/amip/amipnews.html>.)
- Gleckler, P.J., and B.C. Weare, 1997: Uncertainties in global ocean surface heat flux climatologies derived from ship observations. *J. Climate*, 10, 2764-2781.
- Gruber, A., and A.F. Krueger, 1984: The status of the NOAA outgoing longwave radiation data set. *Bull. Amer. Meteor. Soc.*, 65, 958-962.
- Hide, R., J. Dickey, S. Marcus, R. Rosen and D. Salstein, 1998: Atmospheric angular momentum fluctuations in global circulation models during the period 1979-1988. *J. Geophys. Res.* (in press).
- Jones, P.D., 1988: Hemispheric surface air temperature variations: Recent trends and an update to 1987. *J. Clim.*, 1, 654-660.
- Joubert, A.M., 1997: Simulations by the Atmospheric Model Intercomparison Project of atmospheric circulation over southern Africa. *Internat. J. Climatology*, 17, 1129-1154.
- Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deavan, M. Iredell, S. Saha, G. White, J. Woolen, Y. Zhu, A. Leetmaa, R. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, H. Janowiak, K.C. Mo, C. Ropelewski, J. Wang, R. Jenne and D. Joseph, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, 77, 437-471.
- Lau, W.K.-M., J.H. Kim and Y. Sud, 1996: Intercomparison of hydrologic processes in AMIP GCMs. *Bull. Amer. Meteor. Soc.*, 77, 2209-2226.
- Mao, J., and A. Robock, 1998: Surface air temperature simulations by AMIP general circulation models: Volcanic and ENSO signals and systematic errors. *J. Climate* (in press).

- Phillips, T.J., 1994: A Summary Documentation of the AMIP Models. PCMDI Report No. 18, Lawrence Livermore National Laboratory, Livermore, CA, 343 pp.
- Phillips, T.J., 1996: Documentation of the AMIP models on the World Wide Web. *Bull. Amer. Meteor. Soc.*, 77, 1191-1196. (See also <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, Geneva, 76 pp.
- Saji, N.H., and B.N. Goswami, 1997: Intercomparison of the seasonal cycle of tropical surface stress in 17 AMIP atmospheric general circulation models. *Climate Dyn.*, 13, 561-585.
- Santer, B.D., K.E. Taylor and L.C. Corsetti, 1995: Statistical evaluation of AMIP model performance. In: Proceedings of the First International AMIP Scientific Conference, WCRP-92, WMO TD- No. 732, World Meteorological Organization, Geneva, pp.13-18.
- Slingo, J.M., K.R. Sperber, J.S. Boyle, J.-P. Ceron, M. Dix, B. Dugas, W. Ebisuzaki, J. Fyfe, D. Gregoro, J.-F. Gueremy, J. Hack, A. Harzallah, P. Inness, A. Kitoh, W.K.-M. Lau, B. McAvaney, R. Madden, A. Matthews, T.N. Palmer, C.-K. Park, D. Randall and N. Renno, 1996: Intraseasonal oscillations in 15 atmospheric general circulation models: Results from an AMIP diagnostic subproject. *Climate Dyn.*, 12, 325-357.
- Sperber, K.R., and T.N. Palmer, 1996: Interannual tropical rainfall variability in general circulation model simulations associated with the Atmospheric Model Intercomparison Project. *J. Climate*, 9, 2727-2750.
- Sperber, K.R., J.M. Slingo, P.M. Inness and W.K.-M. Lau, 1997: On the maintenance and initiation of the intraseasonal oscillation in the NCEP/NCAR reanalysis and in the GLA and UKMO AMIP simulations, *Climate Dyn.*, 13, 769-795.
- Srinivasan, G., M. Hulme and C. Jones, 1995: An evaluation of the spatial and interannual variability of tropical precipitation as simulated by GCMs. *Geophys. Res. Letters.*, 22, 1697-1700.
- Tao, X., J. Walsh and W. Chapman, 1996: An assessment of global climate model simulations of arctic air temperatures. *J. Climate*, 9, 1060-1076.
- Trenberth, K., and D. Shea, 1987: On the evolution of the Southern Oscillation, *Mon. Wea. Rev.*, 115, 3078-3096.
- Trenberth, K., and J. Hurrell, 1994: Decadal atmosphere-ocean variations in the Pacific. *Climate Dyn.*, 9, 303-319.

- Weare, B.C., and I. Mokhov, 1995: Evaluation of total cloudiness and its variability in the Atmospheric Model Intercomparison Project. *J. Climate*, 8, 2224-2238.
- Weare, B.C., and AMIP Modeling Groups, 1996: Evaluation of the vertical structure of zonally-averaged cloudiness and its variability in the Atmospheric Model Intercomparison Project. *J. Climate*, 9, 3419-3431.
- 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 Report No. 44, Lawrence Livermore National Laboratory, Livermore, CA, 29pp.
- 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.
- Zhang, Y., K.R. Sperber, J.S. Boyle, M. Dix, L. Ferranti, A. Kitoh, K.-M. Lau, K. Miyakoda, D. Randall, L. Takacs and R. Wetherald, 1997: East Asian winter monsoon: Results from eight AMIP models. *Climate Dyn.*, 13, 797-820.

## APPENDIX: 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 are given in parenthesis; 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 (2.3, 3.7, 3.7.1)
CCC	Canadian Climate Centre, Victoria (GCM II)
CCSR	Center for Climate System Research, Tokyo (CCSR/NIES AGCM)
CNRM	Centre National de Recherches Météorologiques, Toulouse (EMERAUDE, ARPEGE cy II)
COLA	Center for Ocean-Land-Atmosphere Studies, Calverton (1.1)
CSIRO	Commonwealth Scientific and Industrial Research Org., Melbourne (CSIRO9 Mark1)
CSU	Colorado State University, Ft. Collins (91)
DERF	Dynamic Extended Range Forecasting (GFDL), Princeton (SM392.2, SM195)
DNM	Department of Numerical Mathematics, Moscow (A5407.V1, A5407.V2)
ECMWF	European Centre for Medium-Range Weather Forecasts, Reading (Cy36)
GFDL	Geophysical Fluid Dynamics Laboratory, Princeton (CDG1)
GISS	Goddard Institute for Space Studies, New York (II Prime)
GLA	Goddard Laboratory for Atmospheres, Greenbelt (GCM-01.0 AMIP-01)
GSFC	Goddard Space Flight Center, Greenbelt (GEOS-1)
IAP	Institute of Atmospheric Physics, Beijing (IAP-2L)
JMA	Japan Meteorological Agency, Tokyo (GSM 8911)
LMD	Laboratoire de Météorologie Dynamique, Paris (LMD5, LMD6b, LMD6s)
MGO	Main Geophysical Observatory, St. Petersburg (AMIP92)
MPI	Max-Planck Institute for Meteorology, Hamburg (ECHAM3, ECHAM4)
MRI	Meteorological Research Institute, Tsukuba (GCM-II, GCM-IIb)
NCAR	National Center for Atmospheric Research, Boulder (CCM2)
NMC	National Meteorological Center, Washington (MRF)
NRL	Naval Research Laboratory, Monterey (NOGAPS3.2, NOGAPS 3.4)
RPN	Recherche en Prévision Numérique, Dorval (NWP-D40P29)
SUNGEN	State University of New York, Albany/NCAR, Boulder (GENESIS 1.5, GENESIS 1.5A)
SUNYA	State University of New York, Albany (CCM1-TG)
UCLA	University of California, Los Angeles (AGCM 6.4)
UGAMP	Universities' Global Atmospheric Modelling Project, Reading (UGCM 1.3)
UIUC	University of Illinois, Urbana-Champaign (MLAM-AMIP)
UKMO	United Kingdom Meteorological Office, Bracknell (HADAM 1)
YONU	Yonsei University, Seoul (Tr 5.1, Tr 7.1)

**Table 1.** Seasonal root-mean-square errors of the AMIP model ensemble mean for selected variables over the period 1979-1988.

Variable	Northern Hemisphere				Southern Hemisphere			
	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON
Mean sea-level pressure <sup>1</sup> (hPa)	2.2	1.8	2.4	1.6	2.8	4.2	5.0	4.2
Surface air temperature <sup>2</sup> (C)	3.8	4.0	3.0	3.5	1.6	2.7	2.5	1.6
Temperature <sup>1</sup> at 200 hPa (C)	3.2	4.7	4.2	3.8	6.9	6.0	4.3	5.0
Zonal wind <sup>1</sup> at 200 hPa (ms <sup>-1</sup> )	4.4	4.0	3.7	3.3	3.7	3.2	3.2	3.6
Outgoing longwave radiation <sup>3</sup> (Wm <sup>-2</sup> )	7.5	8.2	10.0	8.7	8.3	7.6	8.6	8.4
Cloudiness <sup>4</sup> (%)	16.5	17.4	16.0	16.0	23.2	20.5	23.6	23.6
Precipitation <sup>5</sup> (mm day <sup>-1</sup> )	1.1	1.3	1.7	1.3	1.1	1.0	0.9	1.0
Precipitation-evaporation <sup>6</sup> (mm day <sup>-1</sup> )	1.4	1.2	1.8	1.6	1.3	1.2	1.1	1.0

<sup>1</sup> Observations from ECMWF reanalysis (1979-1988) (Gibson et al., 1997)

<sup>2</sup> Observations from Jones (1988) and COADS merged data (1979-1988) (da Silva et al., 1994a); UGAMP not included

<sup>3</sup> Observations from NCEP operational data (1979-1990) (Gruber and Kreuger, 1984); RPN not included

<sup>4</sup> Observations from ISCCP C2 data (1983-1990) (Rossow et al., 1991); RPN not included

<sup>5</sup> Observations from NCEP merged data (1979-1988) (Xie and Arkin, 1997)

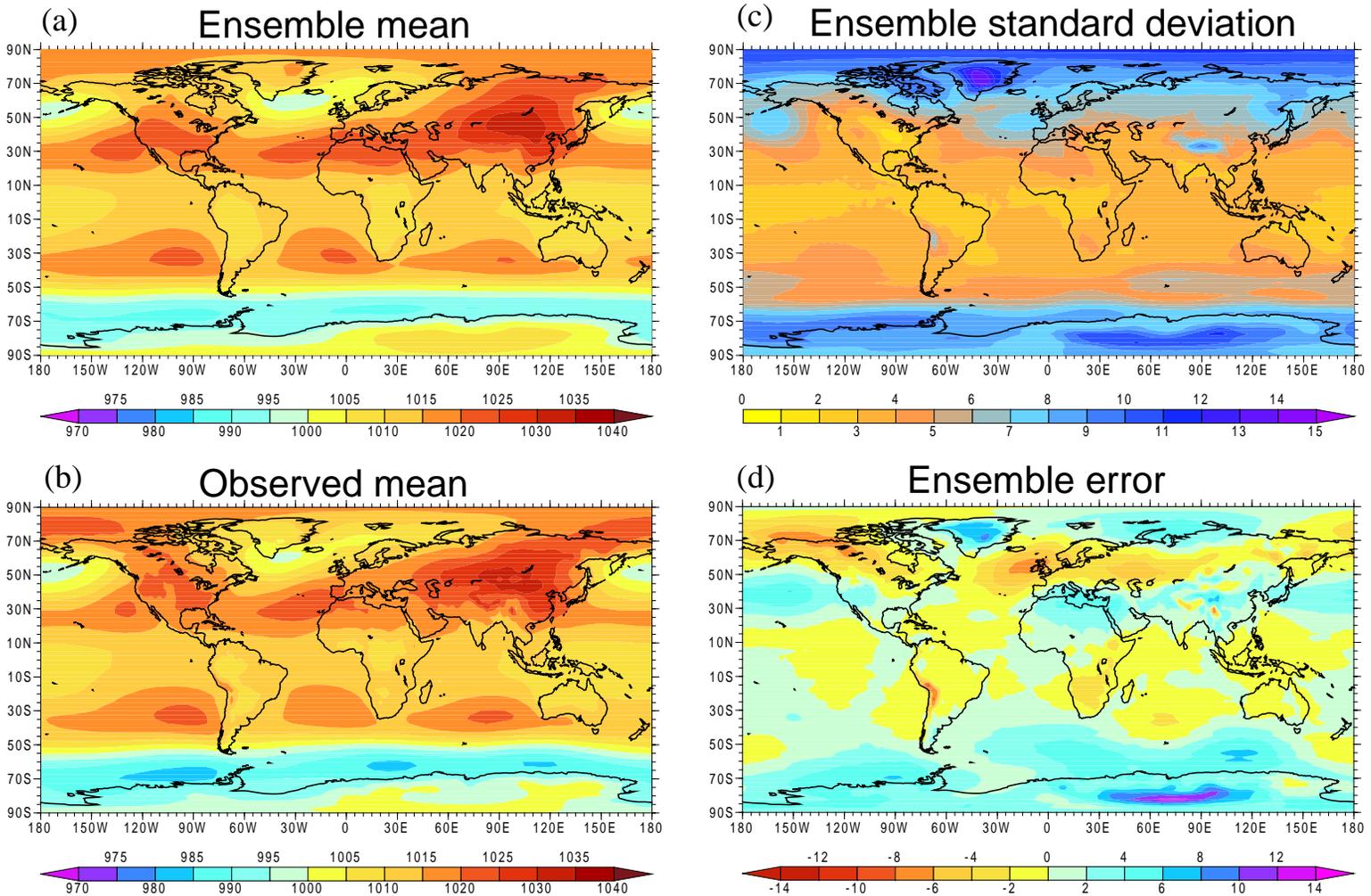
<sup>6</sup> Observations from COADS (1945-1989) (da Silva et al., 1994c); RPN not included

**Table 2.** Seasonal northern hemisphere root-mean-square errors of the ensemble mean of the subset of revised AMIP models for selected variables over the period 1979-1988<sup>7</sup>.

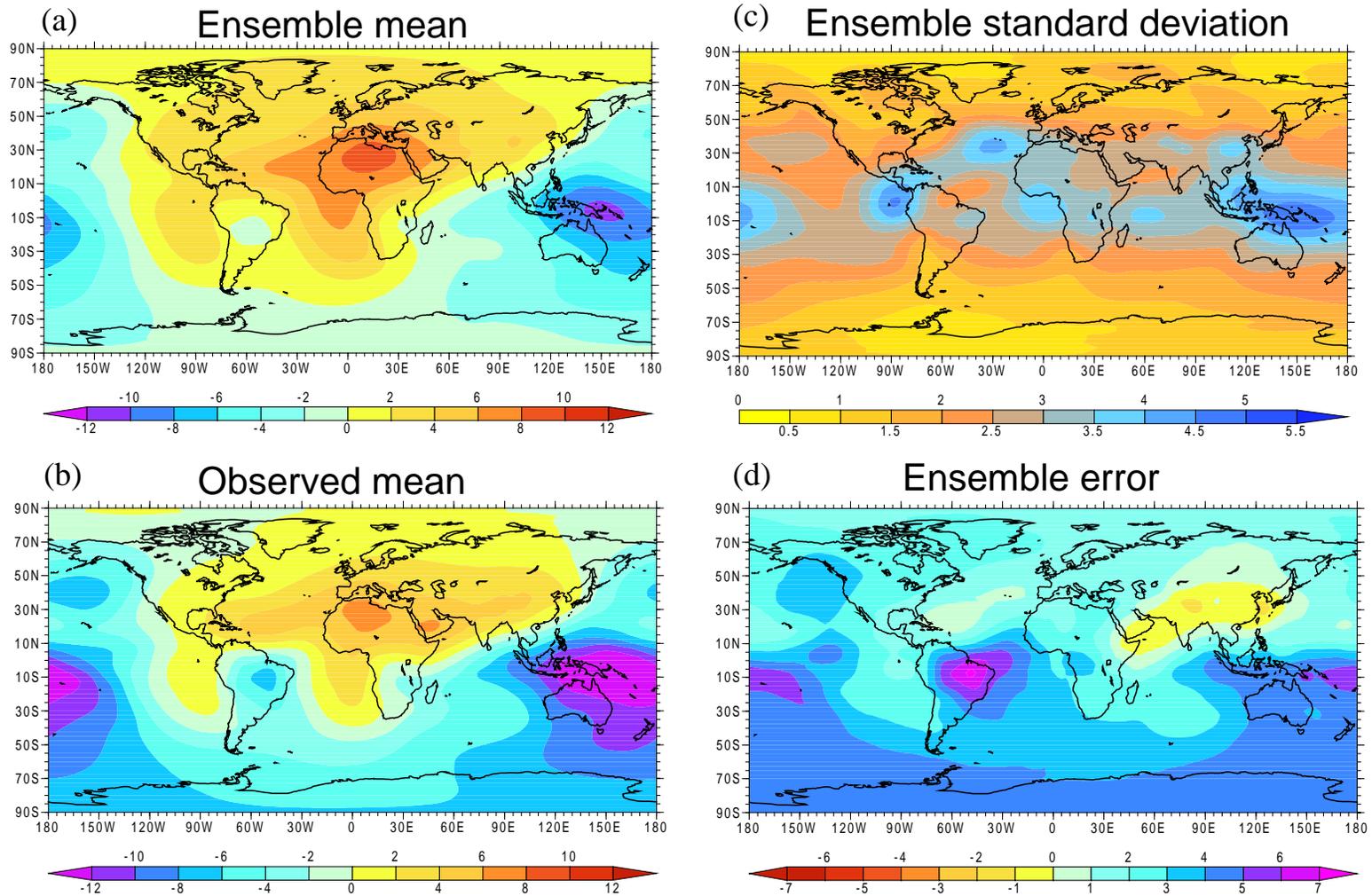
Variable	Original Models				Revised Models			
	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON
Mean sea-level pressure (hPa)	2.4	2.2	2.8	2.0	2.2	1.9	2.7	2.1
Surface air temperature (C)	4.0	3.9	3.0	3.5	4.0	4.0	3.1	3.6
Temperature at 200 hPa <sup>8</sup> (C)	3.9	5.2	4.4	4.3	4.2	5.6	4.5	4.3
Zonal wind at 200 hPa <sup>2</sup> (ms <sup>-1</sup> )	5.2	4.8	4.6	3.5	4.3	4.6	4.6	3.8
Outgoing longwave radiation (Wm <sup>-2</sup> )	9.7	11.7	14.9	13.0	9.1	10.2	13.3	11.6
Cloudiness (%)	19.8	20.6	19.3	20.8	17.3	16.8	15.5	15.8
Precipitation (mm day <sup>-1</sup> )	1.2	1.4	1.9	1.5	1.2	1.5	1.9	1.6
Precipitation-Evaporation <sup>2</sup> (mm day <sup>-1</sup> )	1.4	1.3	1.9	1.7	1.4	1.4	2.0	1.9

<sup>7</sup> The observed data are as in Table 1, and the revised AMIP models are those of the BMRC, CNRM, DERF, DNM, LMD, MPI, MRI, NRL, SUNGEN and YONU (see Appendix)

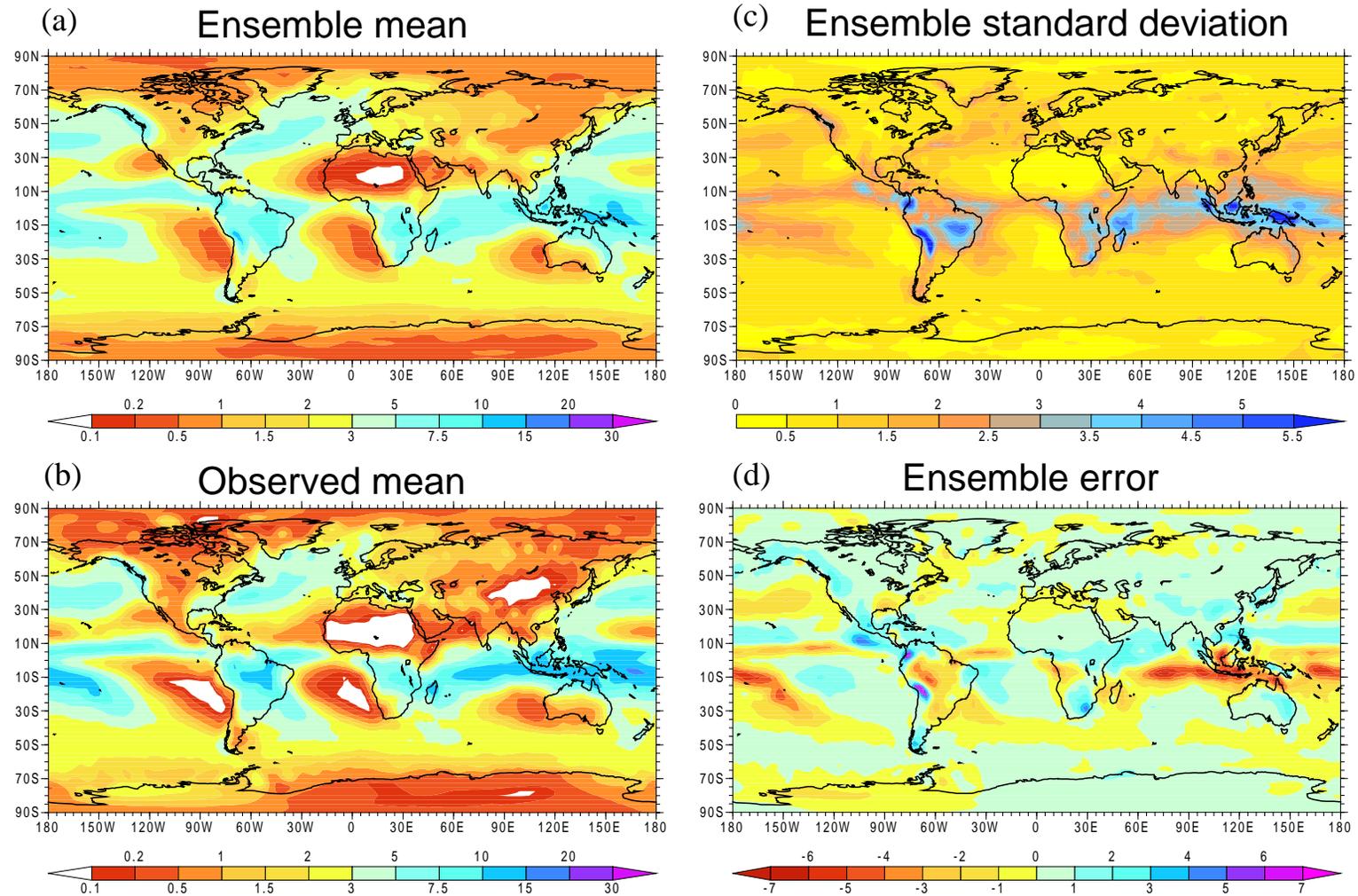
<sup>8</sup> DNM not included



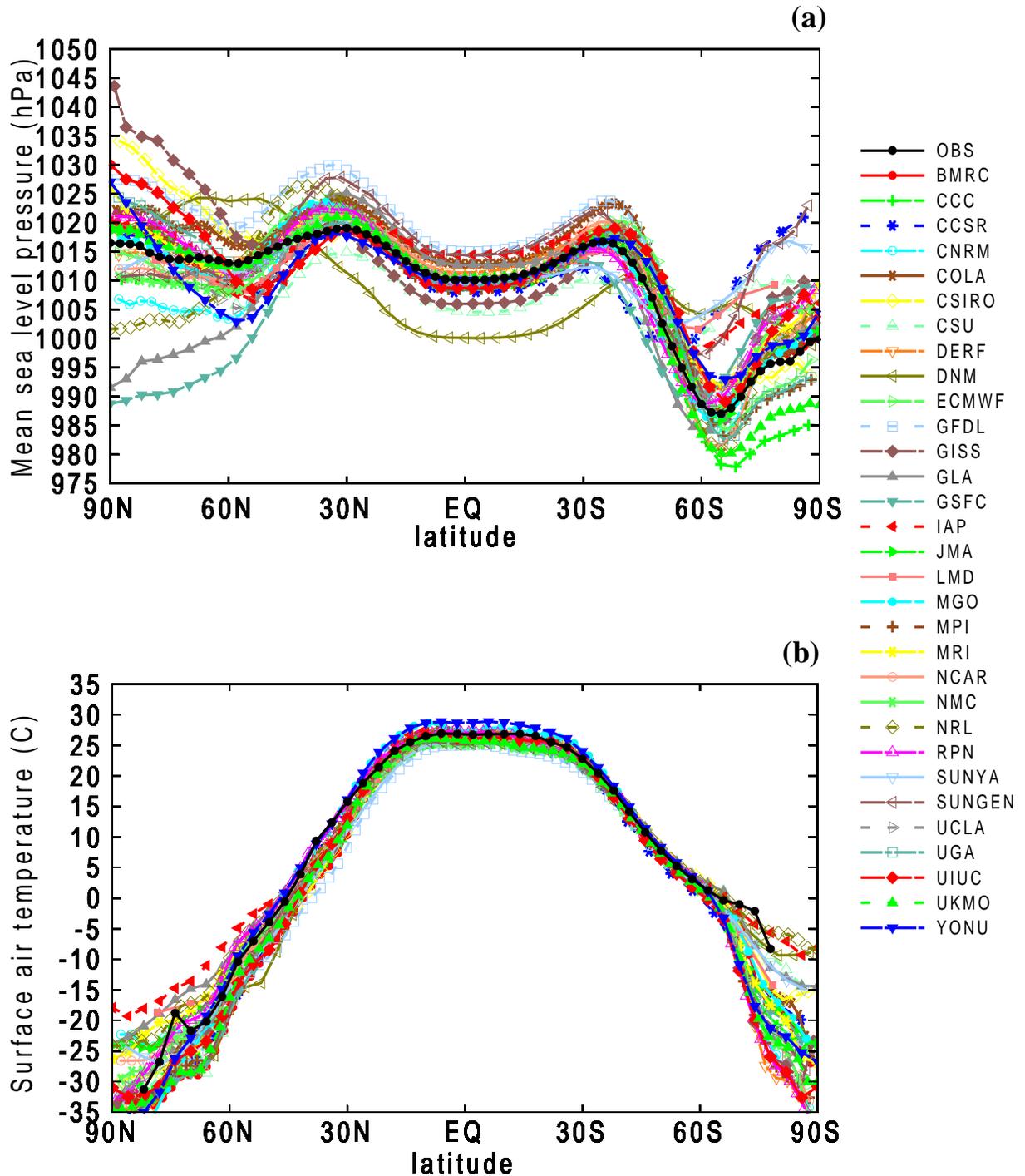
**Fig. 1.** The geographical distribution of mean sea-level pressure (hPa) in December-January-February of 1979-1988 given by the AMIP ensemble mean (a), and by data from the ECMWF reanalysis (Gibson et al., 1997) during 1979-1988 (b). The standard deviation (hPa) of the model ensemble is given in (c), and the error (ensemble mean minus observation; hPa) is given in (d).



**Fig. 2.** As in Fig. 1 except for the velocity potential at 200 hPa ( $106 \text{ m}^2\text{s}^{-1}$ ).



**Fig. 3.** As in Fig. 1 except for precipitation (mm day<sup>-1</sup>), with observations for 1979-1988 from the merged NCEP database (Xie and Arkin, 1997). Note the nonlinear scale in (a) and (b).



**Fig. 4** The zonally-averaged distribution of selected variables simulated by the AMIP models for December-January-February of 1979-1988 and that given by observations (solid black line). Panel (a) is the sea-level pressure, with observed data from the ECMWF reanalysis; panel (b) is the surface air temperature, with observed data from Jones (1988) and COADS (da Silva et al., 1994a); (c) is the zonal wind at 200 hPa, with observed data from the ECMWF reanalysis; (d) is the net ocean surface heat flux, with observational estimates from COADS (da Silva et al., 1994b). (See Appendix for model identification; UGAMP missing in (b), RPN missing in (d)).

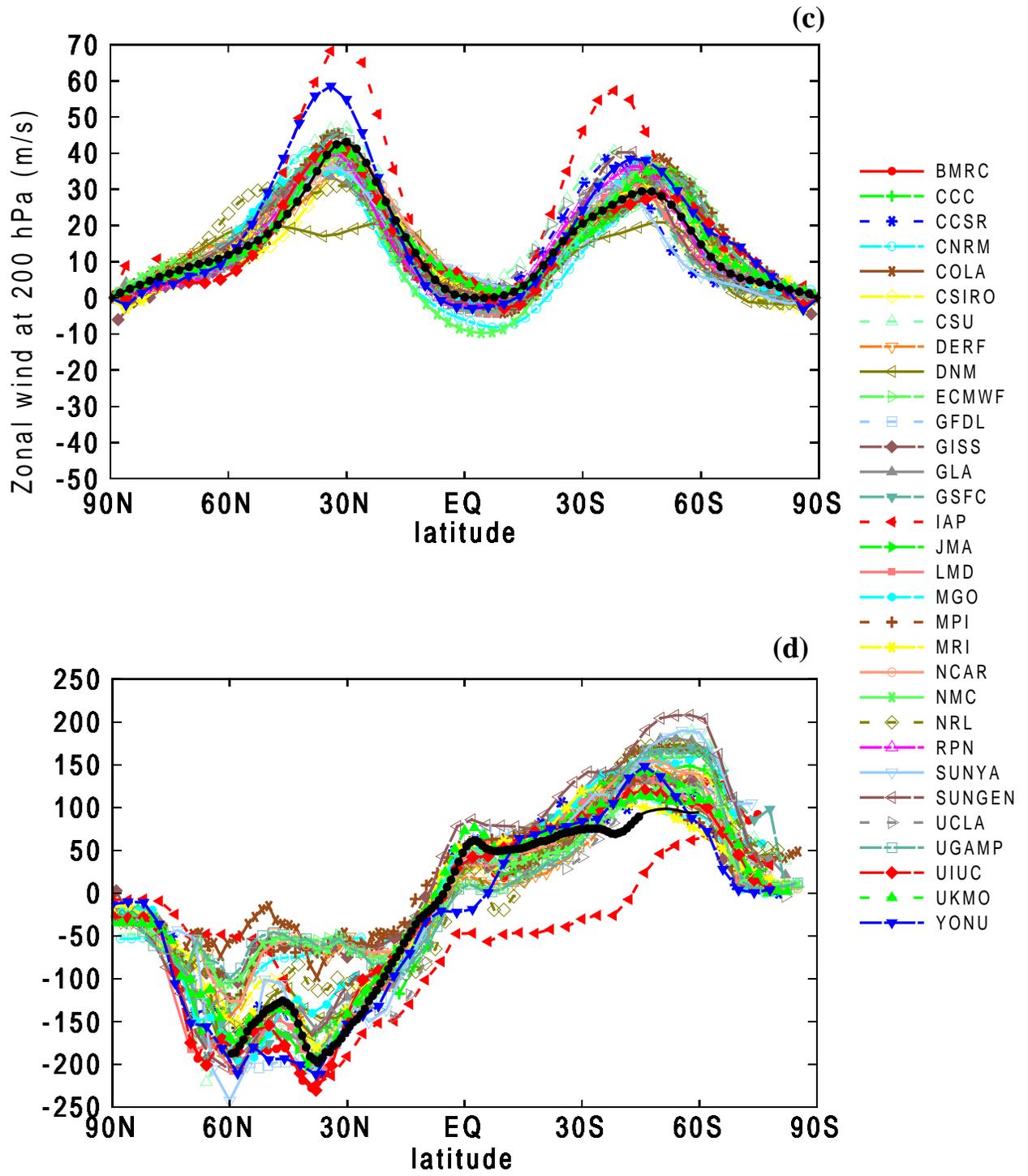
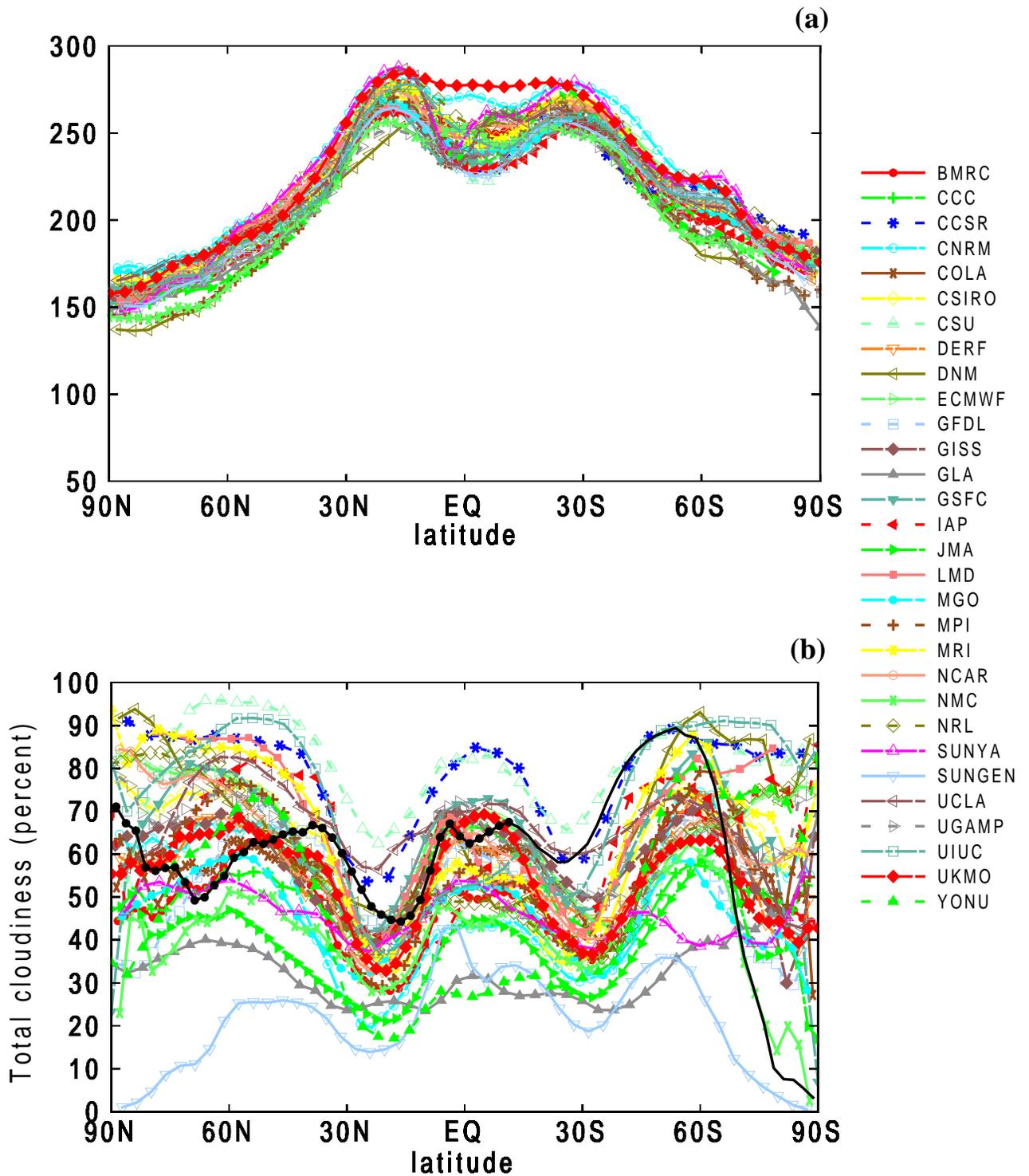


Fig. 4 - continued



**Fig. 5** As in Fig. 4 except for the outgoing long-wave radiation (a), with observations from the NCEP operational database (Gruber and Krueger, 1984); total cloudiness (b) with observations from ISCCP for 1983-1990 (Rossow et al., 1991); precipitation (c) with observations from the merged NCEP database (Xie and Arkin, 1997); precipitation minus evaporation over the ocean (d) with observations from COADS (da Silva et al., 1994c). (RPN missing in (a), (b) and (d)).

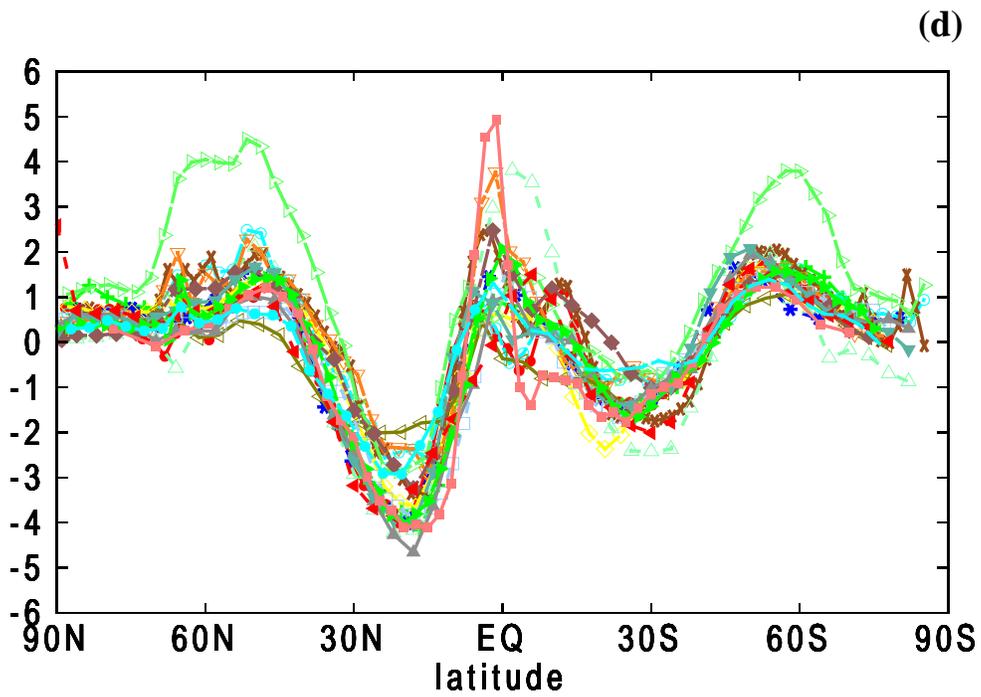
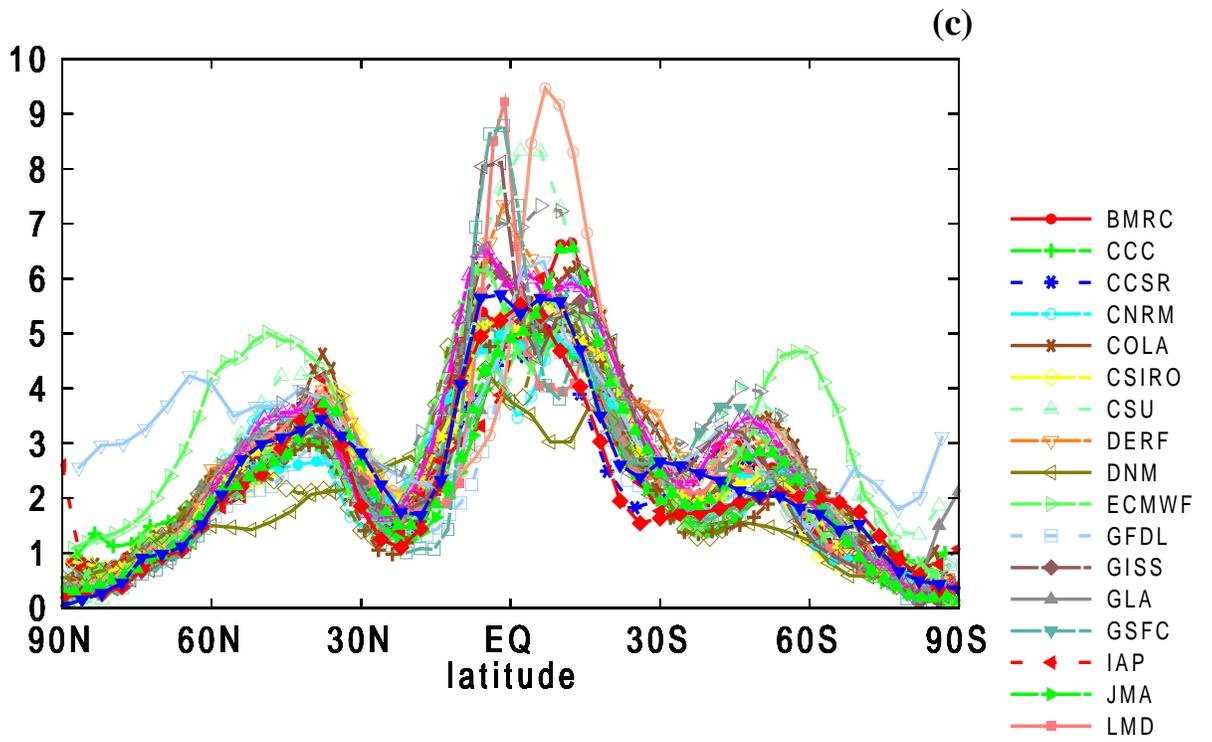
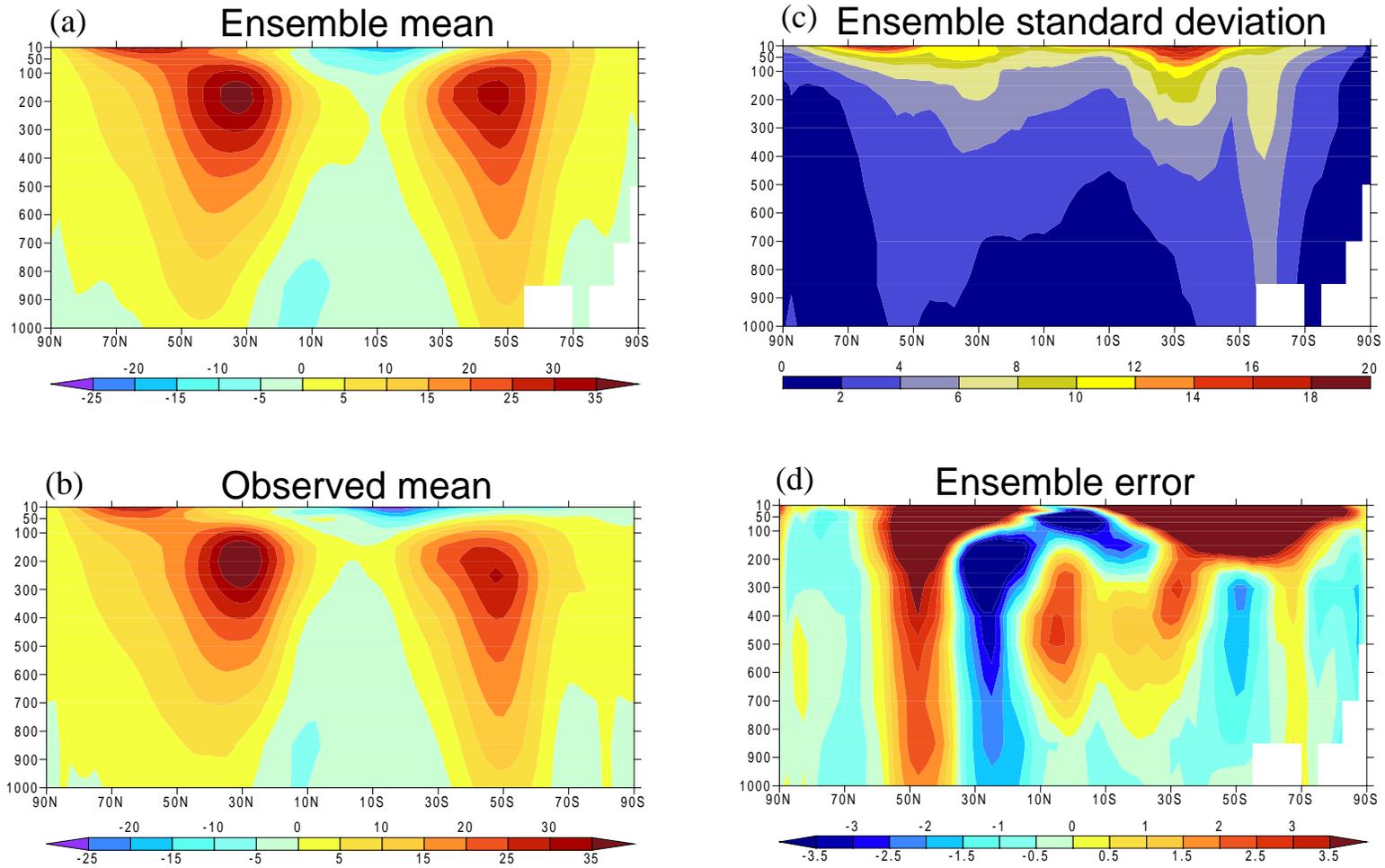
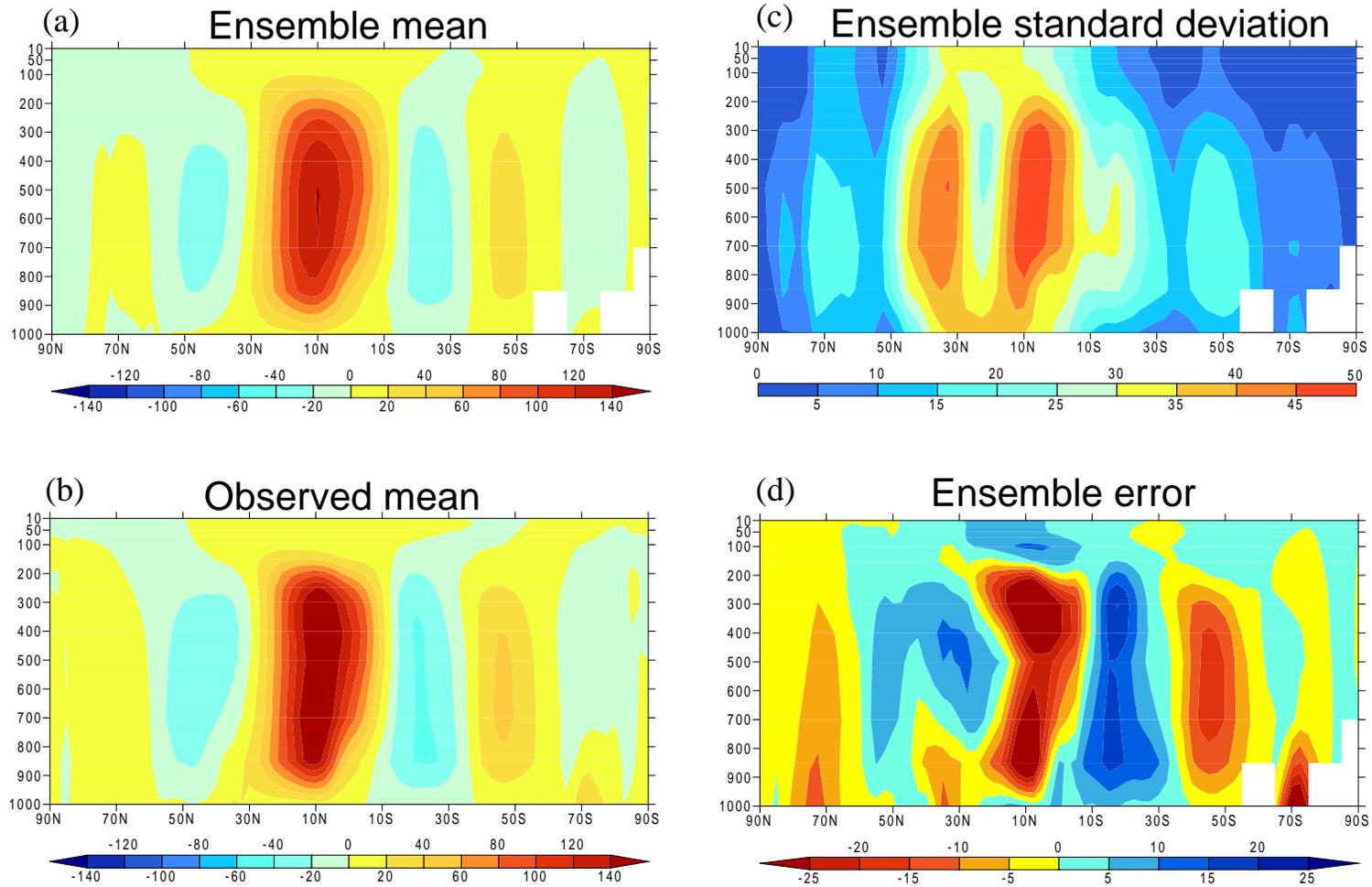


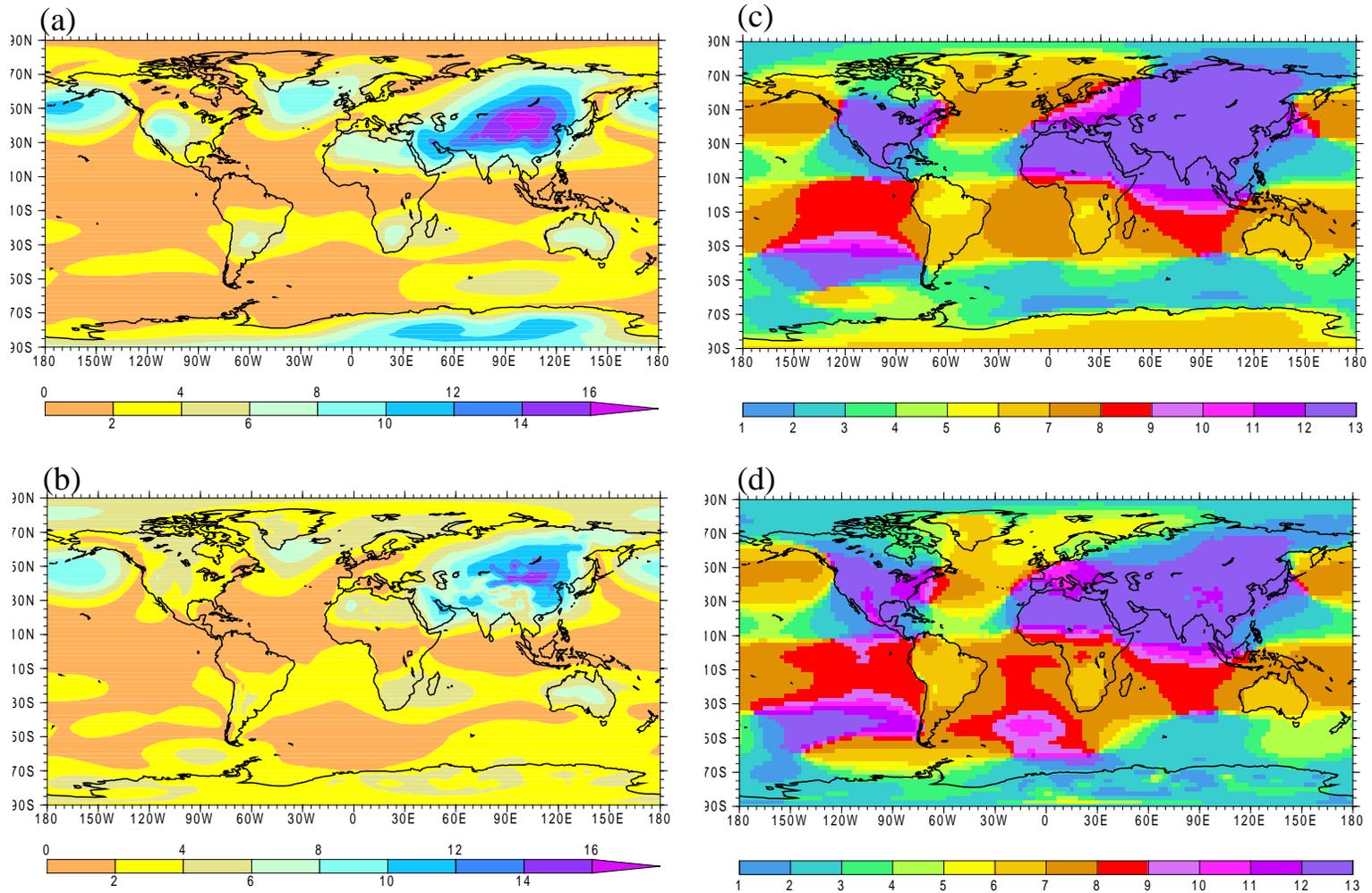
Fig. 5 - continued



**Fig. 6.** As in Fig. 1 except for the latitude-pressure meridional section of the zonal wind (ms<sup>-1</sup>) given by the AMIP ensemble mean.



**Fig. 7.** As in Fig 6 except for the streamfunction for the mean meridional circulation ( $109\text{kg s}^{-1}$ ).



**Fig. 8.** The mean seasonal cycle during 1979-1988 as simulated by the AMIP models in terms of the average amplitude of the annual harmonic of mean sea-level pressure (hPa) in (a), and that given by the ECMWF reanalysis (b). The average phase (month of maximum) of the simulated annual harmonic is given in (c) and as observed in (d).

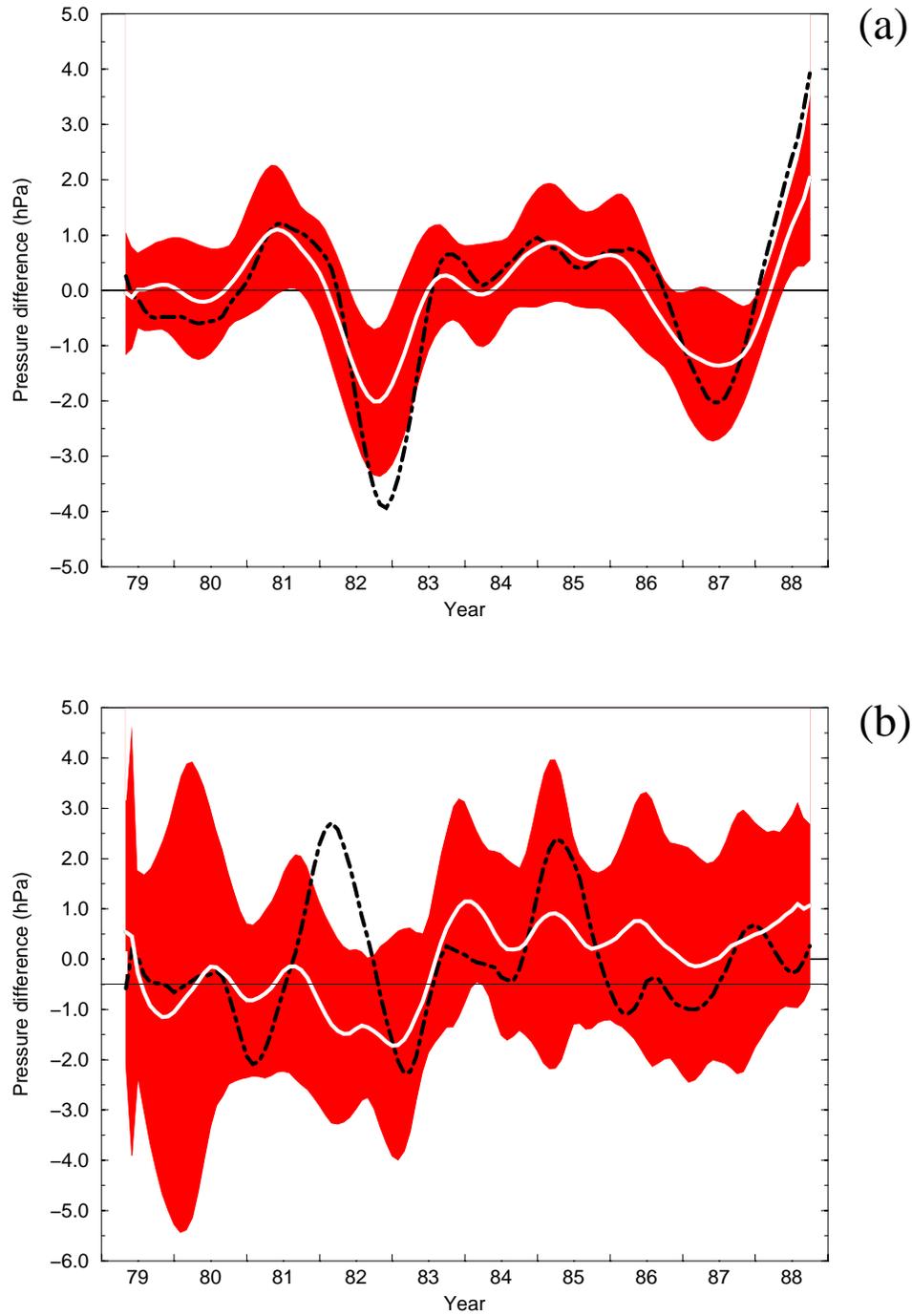
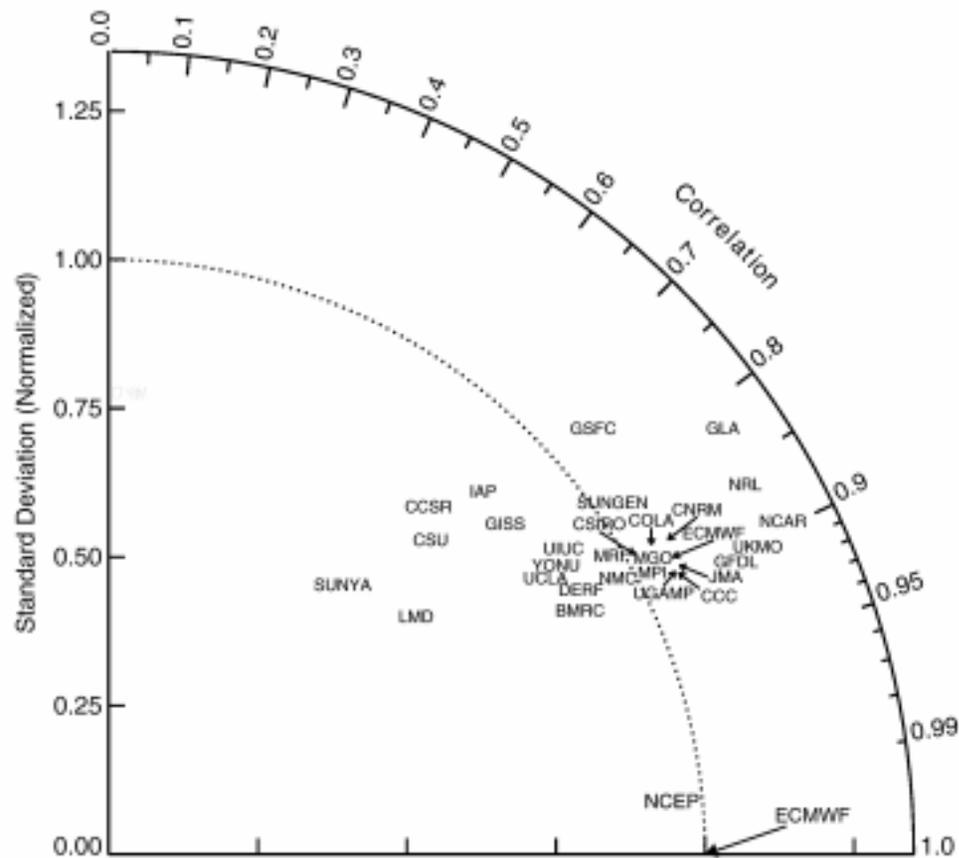
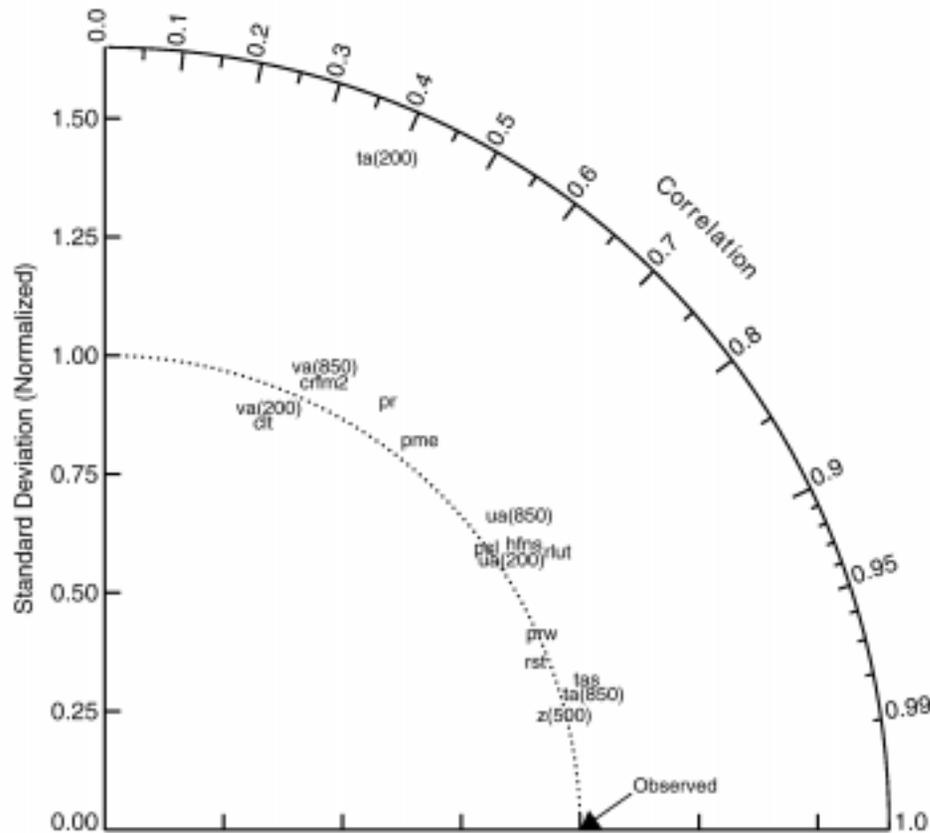


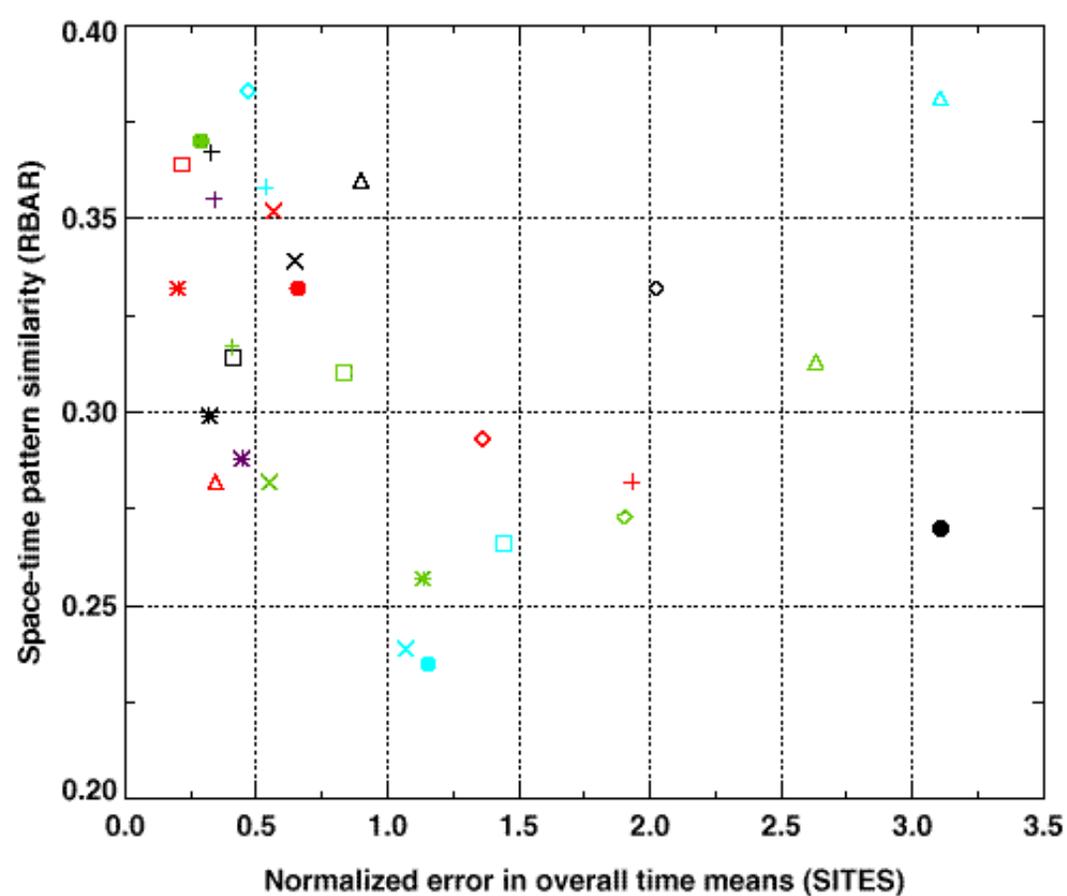
Fig. 9. The interannual variability during 1979-1988 as simulated by the AMIP ensemble ( full line) and as given by the NCEP reanalysis (dash-dotted line) for 1979-1988 (Kalnay et al.,1996). The thin dashed line shows the error of the ensemble mean (ensemble minus observation). Panel (a) shows the sea-level pressure difference averaged over the areas 25S-15S, 125E-135E and 25S-15S,135W-134W, and panel (b) shows the pressure anomaly (relative to the decadal mean) averaged over the area 30N-65N, 160E-220E. The shaded area surrounding the ensemble mean indicates the spread of two standard deviations about the mean.



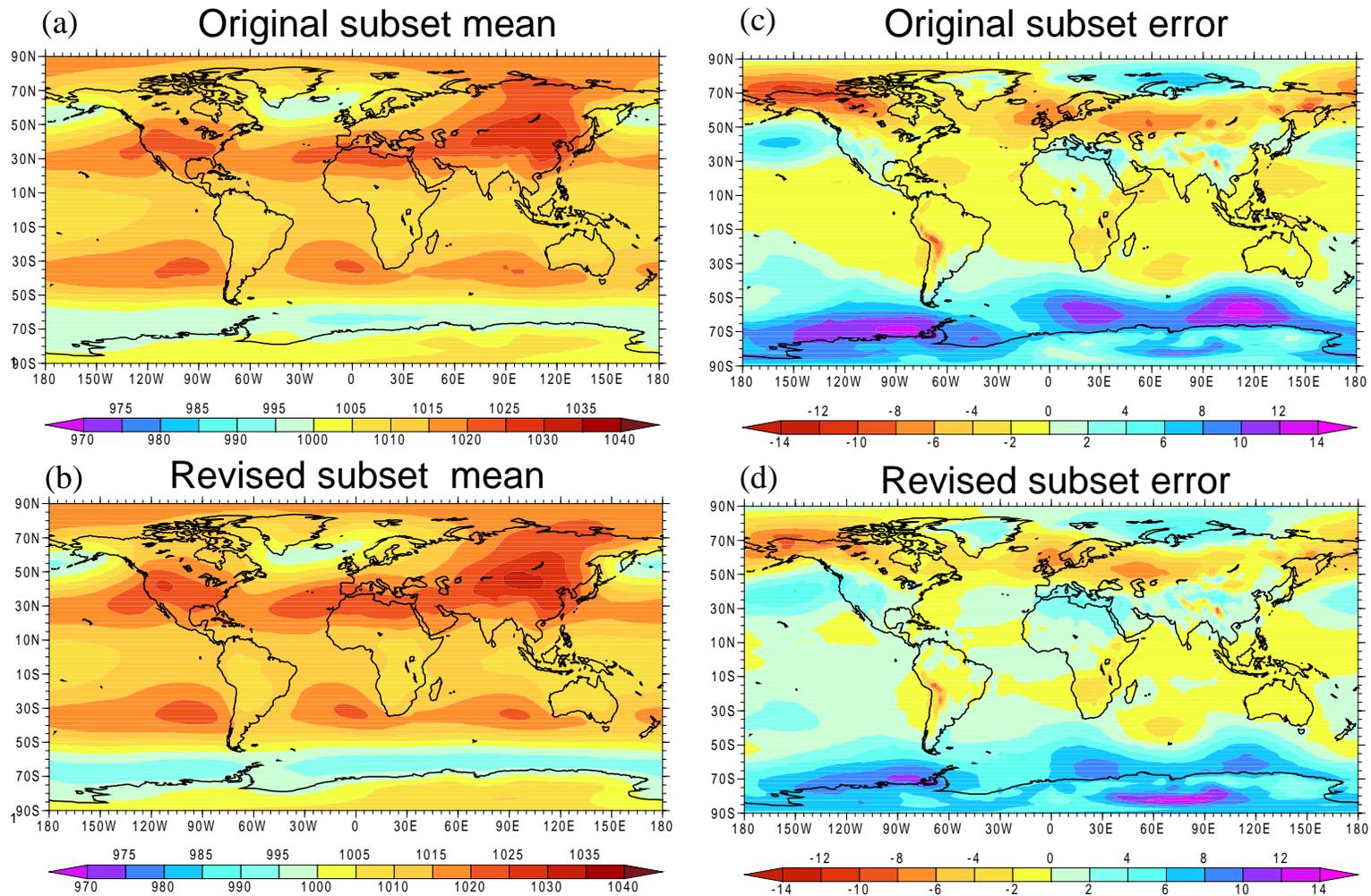
**Fig. 10.** A model performance diagram showing the total space-time pattern variability of the AMIP models' mean sea-level pressure, in terms of the standard deviation of the modeled monthly means (proportional to the distance from the origin), the rms difference between the simulated and observed monthly means (proportional to the distance from the ECMWF reference point), and the correlation between the simulated and observed monthly means over the period 1979-1988. The standard deviations and rms differences have been normalized by the observed standard deviation. The position of the NCEP reanalysis relative to the reference ECMWF reanalysis is also indicated. (See Appendix for model identification).



**Fig. 11.** As in Fig. 10 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 hPa and 850 hPa, crfm2 the cloud-radiative forcing, clt the total cloudiness, pr the precipitation, pme the precipitation minus evaporation, psl the sea-level pressure, (over ocean only), ua(200) and ua(850) the zonal wind at 200 hPa and 850 hPa, hfns the net surface heat flux, rlut the outgoing long-wave 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.



**Fig. 12.** A model performance diagram showing the AMIP models' simulation of the mean December-January-February sea-level pressure during 1979-1988 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.) (See text for fuller explanation).



**Fig. 13.** The geographical distribution of mean sea-level pressure (hPa) simulated in December-January-February of 1979-1988 by the subset of ten models that revisited AMIP with revised versions. Panel (a) shows the mean of the original models' simulation, and panel (b) shows the corresponding mean of the revised versions of the same set of models. Panels (c) and (d) show the errors of the original and revised model subset, respectively, relative to the ECMWF reanalysis.