PCMDI
Program for Climate Model Diagnosis and Intercomparison

UCRL-ID-109810

PCMDI Report No. 1

THE VALIDATION OF ATMOSPHERIC MODELS

by

W. Lawrence Gates

PCMDI, Lawrence Livermore National Laboratory

March 1992

PROGRAM FOR CLIMATE MODEL DIAGNOSIS AND INTERCOMPARISON
UNIVERSITY OF CALIFORNIA, LAWRENCE LIVERMORE NATIONAL LABORATORY
LIVERMORE, CA 94550

THE VALIDATION OF ATMOSPHERIC MODELS

W. Lawrence Gates

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

March 1992

This paper was presented as a portion of the proceedings of the First Demetra Conference on the Dilemmas of Global Warming, held in Chianciano Terme, Italy, 28 October - 1 November 1991.

ABSTRACT

The validation of atmospheric models is a key part of the modelling enterprise, but one to which increased attention needs to be given if systematic progress is to be made in the development of predictive climate models. The validation of current AGCMs in terms of the mean seasonal distribution of primary variables such as pressure, temperature, and wind shows a reasonable ability to simulate the observed large-scale features, while at the same time identifying a number of systematic errors. More recent validations have included the simulation of variability, which reveals a modest level of skill but with further systematic errors. Recent results from mesoscale models nested within AGCMs, however, have shown substantial skill in the simulation of regional climate. In addition to conventional data sources of various resolutions, current model validation is enriched by the use of satellite observations and other special data sets, as well as by the analyses from operational models.

A comprehensive atmospheric model validation program includes examination of not only the mean and variance, but of the complete frequency distribution. Moreover, in addition to the primary dynamical and physical variables, the various derived quantities associated with fluxes and processes and the occurrence of specific events should also be evaluated. A complete validation would also include evaluation of a model's ability to simulate more than just the present climate and/or its ability to simulate observed climate change (the latter aspect necessarily including the oceans). This effort will require the acquisition, calibration and processing of global observational data sets specifically for the purpose of model validation. Useful approaches toward such a program include the reanalysis of recent decades with modern AGCMs and advanced data assimilation systems, and the extension of the various atmospheric model intercomparison projects now underway to include a wider variety of diagnostics.

1. Introduction

The validation (defined as the determination of the degree of correctness or validity) of atmospheric general circulation models is a necessary step in the orderly development and use of models for both weather prediction and climate simulation. In the latter application in particular, it is essential to know a model's accuracy in simulating the balances that characterize the general atmospheric circulation as well as the nature of regional systematic errors. Since the atmosphere is the central and most variable component of the climate system, knowledge of an atmospheric model's errors is basic to an understanding of the performance of coupled atmosphere-ocean models and of models of the coupled climate system in general, even though the errors of the ocean and other non-atmospheric model components may be even less well-known than those of the atmosphere itself.

Atmospheric model validation has always been limited by the available data, which has included direct observations of the atmosphere's three-dimensional structure and circulation for only approximately the last fifty years. Even then the observations necessary to adequately characterize the large-scale features of the atmosphere have been largely confined to the more populated land areas of the Northern Hemisphere. Global observations of the cloudiness have been available from satellites for only about twenty years, while space-based measurements of the atmospheric radiation balance, atmospheric water vapor content, precipitation rate and winds have been made for only about a decade. While atmospheric modelers have, of course, used portions of these data to perform at least a preliminary validation of their models, in general there has not been an organized and comprehensive validation effort. After examining their model's simulation of the mean seasonal distributions of temperature, pressure, wind and precipitation, most modeling groups have not made a corresponding effort validating the model's simulation of variances of these and other fields or of the simulation of regional or extreme events, even though these features may be of more importance in applications of the model than the climatic means themselves.

The purposes of the present paper are to briefly review the present status of atmospheric model validation, and to suggest ways in which future validation may be made more systematic and useful in the light of necessarily limited observational and computational resources.

2. Current status of atmospheric GCM validation

The present state of atmospheric model validation is basically as characterized by the IPCC assessment in 1990 (Gates et al., 1990). Here the principal findings of that report will be summarized insofar as they apply to atmospheric models, followed by a brief review of more recent work.

Comparison of the observed large-scale distribution of the seasonally-averaged climate with that simulated by current atmospheric GCMs (in which the sea-surface temperature and sea ice have been taken from climatology) shows that there is a considerable degree of skill in the models. This is illustrated in Fig.1 in the case of precipitation for three current atmospheric GCMs of moderately high resolution. It is characteristic that while one model may resemble the observed precipitation most closely in a particular region, generally another model's simulation is superior in another region, i.e., no single GCM is universally the best. We may also note that the models show a number of features in common that are not prominent in the observed data, such as the simulation of DJF precipitation maxima in the North Atlantic and North Pacific oceans and the lack of a precipitation maximum in the Southern ocean near Antarctica. (In this case, however, the possible unreliability of the observed data themselves must be kept in mind.) A generally similar level of skill is present in other seasons, and for the corresponding large-scale distributions of pressure, temperature and circulation.

A useful overview of the accuracy of atmospheric models is given by zonally-averaged statistics. This validation measure is shown in Fig. 2 for five representative atmospheric GCMs. Here we note that the models tend to overestimate the observed precipitation, especially in the Northern Hemisphere during winter, and that there is considerable scatter among the models' results in the tropics (where a large fraction of the precipitation is associated with parameterized subgrid-scale convective processes). These features are similar to those seen in earlier model simulations (see, for example, Gates, 1987), and are representative of the current state of the art. Overall, the intermodel precipitation differences are about 1-2 mm/day, and the apparent systematic model error is of the same order.

Similar remarks may be made for the zonally-averaged mean sea-level pressure shown in Fig. 3 where the several GCMs' results display a scatter of abou 10-20 hPa while for the most part being clustered about the observed distribution. An exception is in high southern latitudes, where the models show substantial dis-

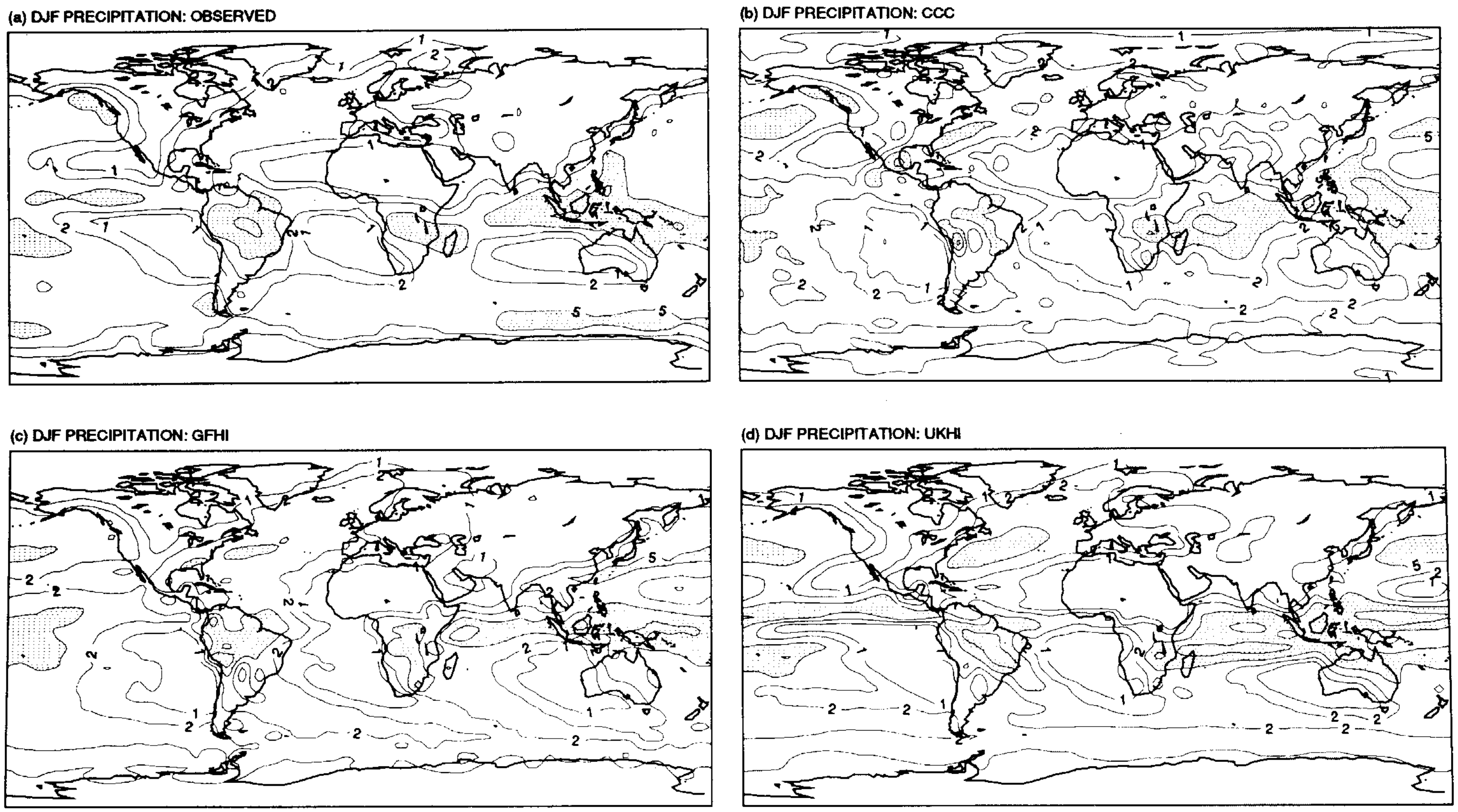


Figure 1. The global distribution of average precipitation (mm/day) for December— January — February as observed (a) (Jaeger, 1976) and as simulated by the CCC (b), the GFDL (c) and the UKMO (d) GCMs. (From Gates et al., 1990)

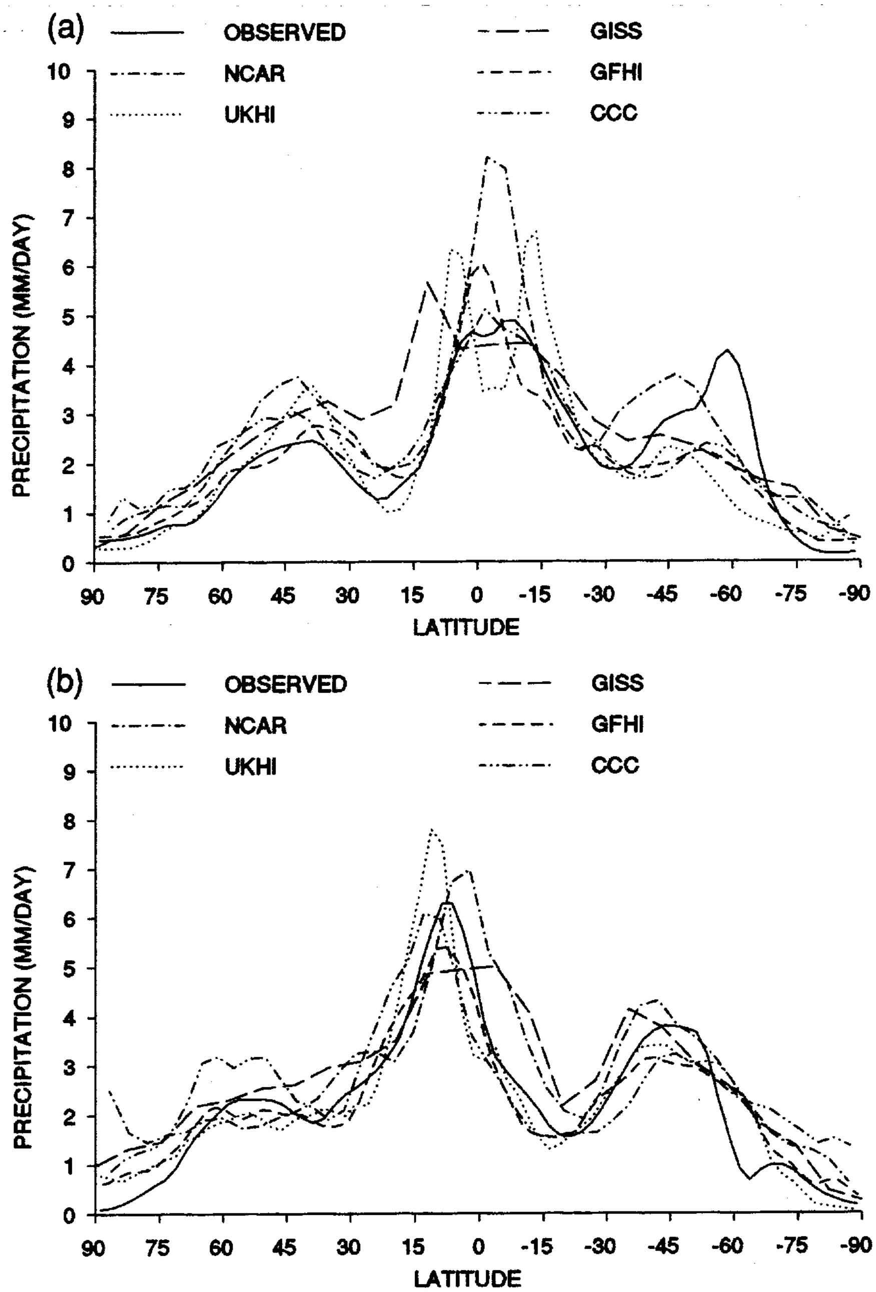
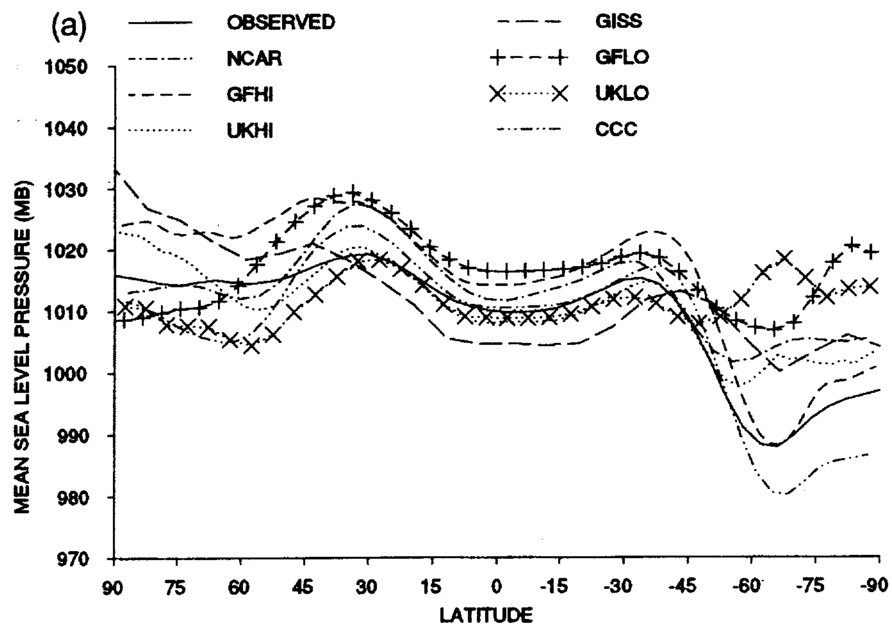


Figure 2. The zonally-averaged precipitation as observed (Jaeger, 1976) for December — January — February (a) and June — July — August (b), and as simulated by the NCAR, UKMO, GISS, GFDL and CCC atmospheric models. (From Gates et al., 1990)

agreement and large apparent errors. As was the case for precipitation, these features have characterized atmospheric models' zonally-averaged sea-level pressure for some time (Gates, 1987) and are representative of the current intermodel differences to be expected. The data in Fig. 3 also provide evidence of a systematic improvement of the simulations with increased (horizontal) model resolution, i.e., the change between GFLO and GFHI and between UKLO and UKHI, although such error reduction is not seen in all climate variables.

A further example of the current state of atmospheric model validation is given by the zonally-averaged zonal wind shown in Fig. 4 for the same models used in Figs. 2 and 3. Here we see a reasonably good simulation of the overall structure and seasonal shift of the tropospheric zonal circulation, which is a reflection of the models' success in simulating the mean meridional tropospheric temperature distribution. Here there is an intermodel variability of about $10ms^{-1}$ with a typical systematic model error of about half this amount. It is evident that the models with higher resolution simulate the strength of the mid-latitude westerlies and tropical easterlies with more accuracy.

Atmospheric models' simulations of other important climate variables such as snow cover, clouds and radiation have been validated in recent years by satellite observations. The gross features of the seasonal cycle of snow cover as given by the limited observational data available are successfully portrayed by GCMs, although there are large intermodel differences on regional scales. The ability of models to simulate the radiation balance of the atmosphere depends critically upon their simulation of the temperature, moisture, cloudiness and snow cover. In general, model-simulated total cloudiness is in broad agreement with that given by ISCCP data, although the vertical distribution of cloudiness is more difficult to validate. The outgoing long-wave radiation at the top of the atmosphere (OLR) is realistically simulated by most GCMs in the tropics, where it is strongly correlated with precipitation, while the somewhat lower OLR in the higher latitudes in winter are generally underestimated. Apart from the polar regions, the zonally-averaged OLR is within $20Wm^{-2}$ of that given by Nimbus 7 satellite observations. The corresponding model simulations of planetary albedo, on the other hand, exhibit intermodel differences of about 0.1 - 0.2, with a tendency for models to overestimate the albedo in lower latitudes. (A comparison of the observed zonally-averaged OLR and planetary albedo with those simulated by the five GCMs used in Figs. 2 and 4



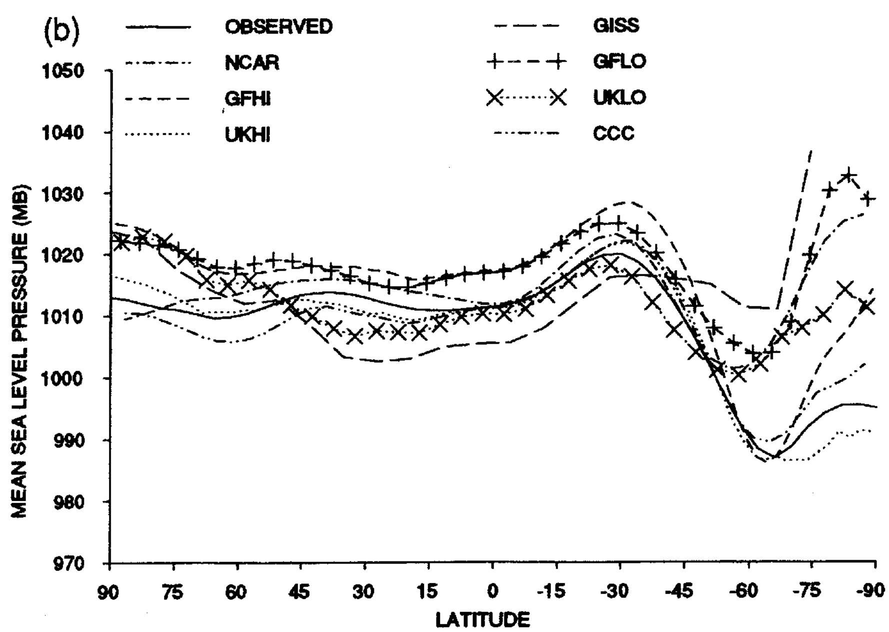


Figure 3. The zonally-averaged mean sea-level pressure for December — January — February (a) and June — July — August (b) as observed (Schutz and Gates, 1971, 1972) and as simulated by the NCAR, GFDL, UKMO, GISS and CCC atmospheric models. The labels HI and LO refer to high- and low- resolution model versions; specifically GFHI = R30, GFLO - R15, UKHI = 2.5° x 3.75 ° and UKLO = 5° x 7.5°. (From Gates et al., 1990)

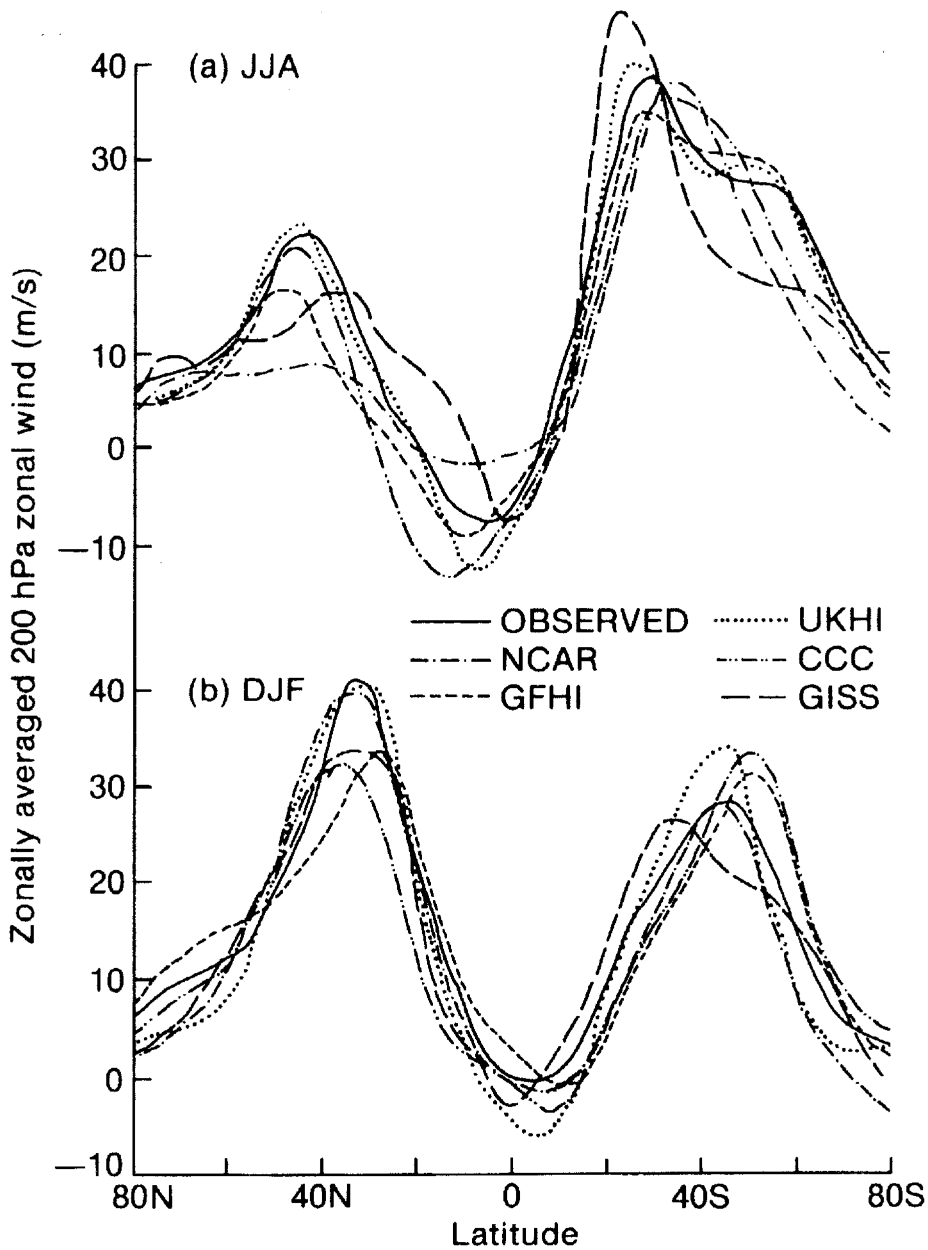


Figure 4. The zonally-averaged zonal wind at 200 hPa for June — July — August (a) and December — January — February (b) as observed (K. Trenberth, personal communication) and as simulated by the NCAR, GFDL, UKMO, CCC and GISS atmospheric models. (From Gates et al., 1990).

is given in Gates et al., 1990.)

More recent model validation studies provide evidence that atmospheric GCMs are capable of simulating aspects of the variability observed over a relatively wide range of time scales, in addition to the mean seasonal climate shifts discussed above. The diurnal variation of surface air temperature has been shown to be reasonably well simulated by modern atmospheric GCMs (see, for example, Cao et al., 1991), although the daily variability is commonly overestimated, especially in the higher northern latitudes during winter. Atmospheric models have also been shown to be capable of simulating much of the observed atmospheric interannual variability in the tropics and the associated ENSO-like phenomena when observed sea-surface temperatures are used (Palmer et al., 1991), although there is a tendency for overestimation of the interannual variability of surface air temperature in high latitudes in winter (Mearns et al., 1991). When coupled to an ocean model, atmospheric GCMs also simulate a realistic level of decadal variability in area-averaged surface air temperature as illustrated in Fig. 5 (although this is not a true validation, since specific dates cannot be assigned to such experiments).

In general the validation of atmospheric GCMs on regional scales has not shown as much skill as on the larger scales, as might be expected in view of the models' relatively coarse horizontal resolution and the uncertainty of many of the models' parameterizations of subgrid-scale processes. Some progress has recently been made, however, in simulating regional climate by nesting a mesoscale model within a GCM and using the GCM's time-dependent solution to determine the boundary conditions necessary to run the nested model. The surface air temperature from such a simulation is shown in Fig. 6, where a mesoscale model of about 50 km resolution over Europe was nested within a GCM with an approximate 500 km grid size. Compared to the GCM's simulation alone (not shown), the one-way nested model results show a significant amount of detail on regional scales that agrees well with observations. It should be noted, however, that the success of this technique depends on the quality of the large-scale GCM simulation, and would be expected to be most effective in regions with considerable topographic variations.

In addition to conventional climatological data on both large and regional scales and a growing amount of observations from satellites, confidence in the use of atmospheric models for the simulation of the (present) climate has been strengthened by the ongoing verification of the models used for operational

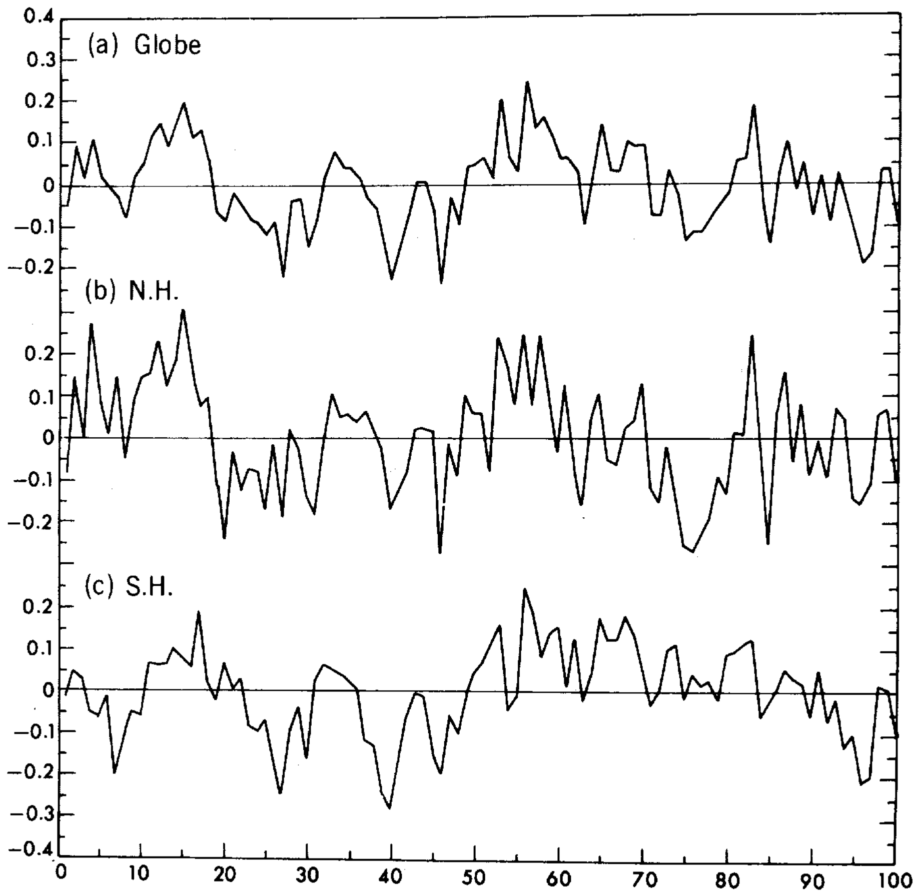


Figure 5. The variation of the area-averaged deviation of annual mean surface air temperature (°C) from the the corresponding 100-yr average in a control integration of the GFDL coupled atmosphere-ocean GCM for the globe (a), the Northern Hemisphere (b) and the Southern Hemisphere (c). (From Manabe et al., 1991)

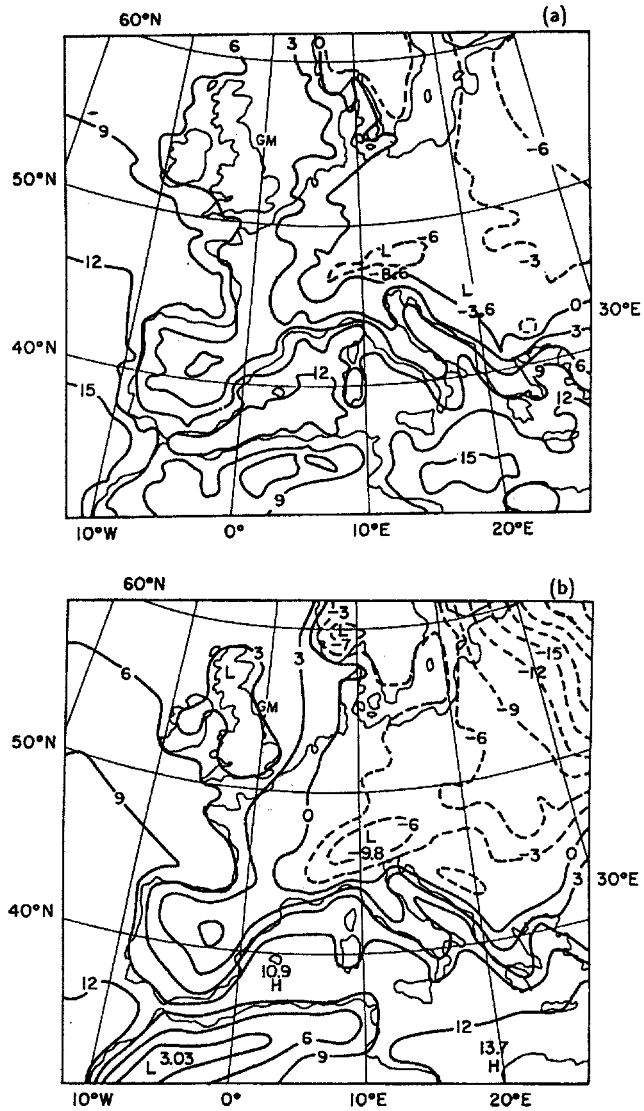


Figure 6. The average January surface air temperature (°C) over Europe as given by observations on a local scale (a) and as simulated by a mesoscale model nested within an atmospheric GCM (b). (From Giorgi et al., 1990)

weather prediction and by the results of paleoclimate simulations, although these model uses rely on quite different validation data bases. In the case of numerical weather prediction, atmospheric models with the highest possible resolution are used and the model's daily forecasts are routinely compared with a relatively dense network of weather observations; this has resulted in the identification of model deficiencies and the subsequent introduction of model improvements many of which have also been incorporated into models used for climate simulation. In paleoclimate modeling, on the other hand, the paucity of reliable and representative proxy data has prevented a detailed model validation, although such simulations provide unique tests of atmospheric models under conditions significantly different than today's.

3. Toward more systematic model validation

While model validation is a natural part of the work of every atmospheric modeling group, in most cases validation extends only to the average values of variables that are of particular interest and for which observed data are at hand. A somewhat more systematic approach to model validation is needed if we are to identify and progressively reduce model errors and to more effectively use atmospheric models in climate research.

It is useful to first recall some of the characteristics of climate and of (atmospheric) climate models. As is well known, the climate always appears to be changing, albeit at a relatively slow rate, and is characterized by fluctuations on a wide variety of time scales. Focussing on periods that may be documented in the observed atmospheric record, it is useful to introduce a more precise definition of climate for the purposes of model validation. To this end the concept of a climatic (or in our case an atmospheric) *state* was introduced (National Academy of Sciences, 1975), whose definition is the complete statistical description of the atmosphere over a specified timeperiod in a particular domain. We may thus, for example, speak of specific monthly, seasonal, yearly or decadal climate states of the entire atmosphere or any portion of it. The problem of validation thus becomes one of assessing the differences in modeled and observed climate states of the same kind. Since a climate state includes all statistical properties of the atmosphere, this definition formally ensures that all variables' mean and variance, for example, are sampled over the same time period. In many cases, however, it is not possible to follow this procedure due to the lack of appropriate observational data, and approximate or mixed climate states are often compared. (Although this definition does not address the problems raised by the possible existence of multiple equilibrium states, it is ideally suited to the analysis of climate experiments.)

With these thoughts in mind, a comprehensive validation of atmospheric models would include all of the elements of the "validation matrix" shown in Fig. 7. Here, for a given climate state such as the decadal statistics over the Northern Hemisphere, the mean, variability and complete frequency distribution (and corresponding error estimates) for the full suite of simulated variables and the associated fluxes, processes and phenomena would be determined from both modeled and observed data insofar as possible. For many of the matrix elements a reliable identification of model errors may require innovative techniques for time series analysis and pattern recognition (see, for example, Santer and Wigley, 1990). If all the elements of such a validation strategy were examined for at least the principal version of each modeling group's atmospheric model, then we would have a much more complete view of current simulation errors than we now possess (and observed data might be more effectively assembled for model validation purposes). As may be recognized from the discussion in section 2, most atmospheric model validation has been concerned with elements in the upper left portion of the matrix in Fig. 7 although there are scattered studies of other elements. In general, however, the climate state being considered has not been clearly defined in validation studies and error estimates have not been made.

Even with the assembly of the complete validation matrix, however, our knowledge of the ability of atmospheric models to simulate climate and climate change would be incomplete. Since a certain amount of tuning has occurred in the parameterization of physical processes in atmospheric GCMs in order to bring the models' overall performance in accord with the present climate (interpreted as, say, the most recent 30-yr climate state), we would have no *a priori* reason to believe that the models were capable of simulating a climate state that was distinctly different from the present. The only way this can be achieved is by using the available paleo-climatic conditions. Even though the data for the best-reconstructed paleoclimate—the last ice age—falls far short of that required for a comprehensive validation, the evidence suggests that atmospheric GCMs are indeed capable of reproducing the general features of at least some observed paleoclimates (Crowley

CLIMATE STATISTIC

VARIABLE, PROCESS OR PHENOMENON	Mean	Variance	Frequency distribution
Variables (temperature, wind, ...)			
Processes (fluxes, feedbacks, ...)		Simulated ± error Observed ± error	
Phenomena (cyclones, monsoons, ...)			

Figure 7. An atmospheric model validation matrix whose elements correspond to the climate statistics for model variables and processes that should be examined for each selected climate state with both simulated and observed data.

and North, 1991).

Finally, even though an atmospheric model might successfully simulate both the present and one (or more) past climate states, we would still not know whether it could correctly simulate the time-dependent transition between climate states. This question is of obvious importance to the use of GCMs for predicting the future course of climate in response, say, to increasing CO_2. Since paleoclimatic data are inadequate for this purpose, we are left with only the slowly accumulating record of modern observations. Simulating the historical climate record, however, may prove difficult since it may be largely the result of natural or internal (and therefore unpredictable) variability. Under these conditions, modelers usually make the implicit assumption that atmospheric GCMs are capable of simulating climate changes in response to at least relatively small changes in external forcing.

4. Limiting factors and future efforts

As might be deduced from the preceeding discussion, there are two general factors that will serve to limit (or at least to shape) future atmospheric model validation. The first of these is the need for reliable observations of the appropriate variables and processes. This has been a longstanding need in the development of atmospheric models (and is even more acute in the case of ocean models). Hopes for alleviating the data problem rest partly on the development of improved satellite sensors for a wider variety of atmospheric variables, and partly on the increased use of the daily observations that support operational weather prediction through the process known as reanalysis (Bengtsson and Shukla, 1988). While satellite programs are slowly advancing, the implementation of reanalysis (whereby modern data assimilation and initialization techniques are used with all available observations to reconstruct the daily state of the atmosphere over the past several decades or so) promises to provide a wealth of data that are especially suited for model validation. The second critical factor for future atmospheric model validation is of course the models themselves. Although we are familiar with the overall performance of the standard or control versions of many atmospheric GCMs (since these are usually the only versions whose results are published), there is a need for increased documentation and/or validation of model versions that employ alternative parameterizations of key physical processes such as convection, radiation, cloud physics and surface interactions. The climatic effects of alternative numerical algorithms also need to be more carefully examined, of which the effects of increased model resolution are perhaps the most important. Recent atmospheric model studies in which GCMs are run under standard conditions expressly for the purposes of intercomparison are revealing a marked dependence of the simulated feedbacks on model formulation (Cess et al., 1990).

Further such intercomparisons are now underway as part of a comprehensive international atmospheric model intercomparison project (AMIP) being carried out by the WCRP Working Group on Numerical Experimentation (WGNE), in which virtually all of the world's atmospheric GCMs will simulate a common decade. When combined with comprehensive validation and diagnosis (and possibly reanalysis), this effort is expected to provide valuable support to systematic model improvement.

Acknowledgement. This work was performed under the auspices of the U.S. Department of Energy Environmental Sciences Division at the Lawrence Livermore National Laboratory under contract W-7405-ENG-48.

REFERENCES

Bengtsson, L., and J. Shukla, 1988: Integration of space and *in situ* observations to study global climate change. *Bull. Amer. Meteor. Soc.*, 69, 1130-1143.

Cao, H.-X., J.F.B. Mitchell and J.R. Lavery, 1991: Simulated diurnal range and variability of surface temperature in a global climate model for present and doubled CO_2. *J. Climate. (in press)*

Cess, R.D., G.L. Potter, J.P. Blanchet, G.J. Boer, A.D. Del Genio, M. Déqué, V. Dymnikov, V. Galin, W.L. Gates, S.J. Ghan, J.T. Kiehl, A.A. Lacis, H. Le Treut, Z.-X. Li, X.-Z. Liang, B.J. McAvaney, V.P. Meleshko, J.F.B. Mitchell, J.-J. Morcrette, D.A. Randall, L. Rikus, E. Roeckner, J.-F. Royer, U. Schelse, D.A. Sheinin, A. Slingo, A.P. Sokolov, K.E. Taylor, W.M. Washington, R.T. Wetherald, I. Yagai and M.-H. Zhang, 1990: Intercomparison and interpretation of climate feedback processes in nineteen atmospheric general circulation models. *J. Geophys. Res.*, 95, 16601-16615.

Crowley, T.J., and G.R. North, 1991: *Paleoclimatology.* Oxford University Press, New York, 339 pp.

Gates, W.L., 1987: Problems and prospects in climate modeling. In *Toward Understanding Climate Change* (U. Radok, ed.), Westview Press, Boulder, pp. 5-33.

Gates, W.L., P.R. Rowntree and Q.-C. Zeng, 1990: Validation of climate models. In *Climate Change, The IPCC Scientific Assessment* (J.T. Houghton, G.J. Jenkins and J.J. Ephraums, eds.), Intergovernmental Panel on Climate Change, Cambridge, UK, pp. 93-130.

Giorgi, F., M.R. Marinucci and G. Visconti, 1990: Use of a limited-area model nested in a general circulation model for regional climate simulation over Europe. *J. Geophys. Res.*, 95, 18413-18431.

Jaeger, L., 1976: Monatskarten des Niederschlags für die ganze Erde. *Berichte Deutschen Wetterdienstes*, 18, Nr. 139, 38 pp.

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

National Academy of Sciences, 1975: *Understanding Climatic Change.* Report of U.S. GARP Committee's Panel on Climatic Variation (W.L. Gates and Y. Mintz, Co-chairmen), Washington, DC, 239 pp.

Palmer, T.N., C. Brankovic, P. Viterbo and M.J. Miller, 1991: Modelling interannual variations of summer monsoons. *J. Climate.* (in press)

Santer, B.D., and T.M.L. Wigley, 1990: Regional validation of means, variances and spatial patterns in general circulation model control runs. *J. Geophys. Res.*, 95, 829-850.

Schutz, C., and W.L. Gates, 1971: Global climatic data for surface, 800 mb and 400 mb: January. Report R-915-ARPA, Rand Corporation, Santa Monica, 173 pp.

Schutz, C., and W.L. Gates, 1972: Global climatic data for surface, 800 mb and 400 mb: July. Report R-1029-ARPA, Rand Corporation, Santa Monica, 180 pp.