How Well do Coupled Models Simulate Today's Climate?

Thomas Reichler and Junsu Kim

Department of Meteorology, University of Utah, Salt Lake City

Correspondence: Thomas Reichler (thomas.reichler@utah.edu)

Department of Meteorology, University of Utah
135 S 1460 E, Rm 819 (WBB)
Salt Lake City, UT 84112-0110
801-585-0040 Fax: 801-581-4362

Accepted Sept. 2007, Bulletin of the American Meteorological Society

Abstract

Information about climate and how it responds to increased greenhouse gas concentrations depends heavily on insight gained from numerical simulations by coupled climate models. The confidence placed in quantitative estimates of the rate and magnitude of future climate change is therefore strongly related to the quality of these models. In this study, we test the realism of several generations of coupled climate models, including those used for the 1995, 2001, and 2007 reports of the Intergovernmental Panel on Climate Change (IPCC). By validating against observations of present climate, we show that the coupled models have been steadily improving over time and that the best models are converging towards a level of accuracy that is similar to observation based analyses of the atmosphere.

Capsule Summary

This is the first systematic attempt to compare the performance of entire generations of climate models by exploring their ability to simulate present climate.

Introduction

Coupled climate models are sophisticated tools designed to simulate the earth climate system and the complex interactions between its components. Currently, more than a dozen centers around the world develop climate models to enhance our understanding of climate and climate change and to support the activities of the Intergovernmental Panel on Climate Change (IPCC). However, climate models are not perfect. Our theoretical understanding of climate is still incomplete, and certain simplifying assumptions are unavoidable when building these models. This introduces biases into their simulations, which sometimes are surprisingly difficult to correct (AchutaRao and Sperber 2006; Bony and Dufresne 2005; Covey et al. 2003; Mechoso et al. 1995; Oldenborgh et al. 2005; Sun et al. 2006). Model imperfections have attracted criticism, with some arguing that model based projections of climate are too unreliable to serve as a basis for public policy (Jones 2005; Lahsen 2005; Lindzen 2006; Singer 1999). In particular, early attempts at coupled modeling in the 1980s resulted in relatively crude representations of climate (Gates et al. 1993). Since then, however, we have refined our theoretical understanding of climate, improved the physical basis for climate modeling, increased the number and quality of observations, and multiplied our computational capabilities. Against the background of these developments, one may ask how much have climate models improved and how much can we trust the latest coupled model generation.

The goal of this study is to objectively quantify the agreement between model and observations using a single quantity derived from a broad group of variables, which is then applied to gauge several generations of coupled climate models. This approach is new, since previous model intercomparison studies either focused on specific processes (Lin et al. 2006; Oldenborgh et al. 2005; Stenchikov et al. 2002), avoided making quantitative performance statements (Bader 2004), or considered a rather narrow range of models.

Several important issues complicate the model validation process. First, identifying model errors is difficult because of the complex and sometimes poorly understood nature of climate itself, making it difficult to decide which of the many aspects of climate are important for a good simulation. Second, climate models must be compared against present (e.g., 1979-1999) or past climate, since verifying observations for future climate are unavailable. Present climate, however, is not an independent data set since it has already been used for the model development (Williamson 1995). On the other hand, information about past climate carries large inherit uncertainties, complicating the validation process of past climate simulations (e.g., Schmidt et al. 2004). Third, there is a lack of reliable and consistent observations for present climate, and some climate processes occur at temporal or spatial scales that are either unobservable or unresolvable. Finally, good model performance evaluated from the present climate does not necessarily guarantee reliable predictions of future climate (Murphy et al. 2004). Despite these difficulties and limitations, model agreement with observations of today's climate is the only way to assign model confidence, with the underlying assumption that a model that accurately describes present climate will make a better projection of the future.

Considering the above complications, it is clear that there is no single "ideal way" to characterize and compare model performances. Most previous model validation studies used conventional statistics to measure the similarity between observed and modeled data. For example, Taylor et al. (2001) and Boer and Lambert (2001) characterized model performance from correlation, root mean square (RMS) error, and variance ratio. Both studies found similar ways to combine these three statistics in a single diagram, resulting in nice graphical visualizations of model performance. This approach, however, is only practical for a small number of models and/or climate quantities. In addition, Taylor's widely used approach requires centered RMS errors with the mean bias removed. We, however, consider the mean bias as an important component of model error. Murphy et al. (2004) introduced a "Climate Prediction Index" (CPI), which measures the reliability of a model based on the composite mean square errors of a broad range of climate variables. More recently, Min and Hense (2006) introduced a Bayesian approach into model evaluation, where skill is measured in terms of a likelihood ratio of a model with respect to some reference.

Three generations of model data

This study includes model output from three different climate model intercomparison projects (CMIP): CMIP-1 (Meehl et al. 2000), the first project of its kind organized in the mid 90s; the follow-up project CMIP-2 (Covey et al. 2003; Meehl et al. 2005); and CMIP-3 (PCMDI 2007) (aka IPCC-AR4), representing today's state-of-the-art in climate modelling. The CMIP-3 data were taken from the "climate of the twentieth century" (20C3M) (hereafter simply 'present-day') and the "preindustrial control" (PICNTRL) (hereafter simply 'preindustrial') experiments. These simulations were driven by a rather realistic set of external forcings, which included the known or estimated history of a range of natural and anthropogenic sources, such as variations in solar output, volcanic activity, trace gases, and sulfate aerosols. The exact formulation of these forcings varied from model to model, with potential implications for model performance. In contrast, the CMIP-1 and CMIP-2 model output was derived from long "control runs", in which the forcings were held constant in time. These forcings were only approximately representative for present climate.

Measure of model performance

As outlined before, there are many different ways to measure and depict model performance. Given the extra challenge of this study to evaluate and depict a large number of models and climate variables, we decided to design our own measure. Our strategy was to calculate a single performance index, which can be easily depicted, and which consists of the aggregated errors in simulating the observed climatological mean states of many different climate variables. We focused on validating the time mean state of climate since this is the most fundamental and best observed aspect of climate, and because of restrictions imposed by available model data in calculating higher moments of climate (most CMIP-1 fields are archived as climatological means, prohibiting the derivation of temporal variability). This concept is somewhat similar to the CPI performance measure introduced by Murphy et al. (2004), but in contrast to the present study, Murphy et al. used a perfect model approach (real observations are replaced by model output) to calculate the CPI from a range of rather closely related models.

Our choice of climate variables, which is shown in Table 1, was dictated by the data available from the models. In most case, we were able to validate the model data against true observation based data, but for a few variables of the free atmosphere the usage of reanalyses as validation data was unavoidable. In terms of the specific uncertainties associated with each of those validating data sets, separate analysis (Reichler and Kim 2007) showed that the data can be considered as good approximations to the real state of present climate for the purpose of model validation.

[Table 1 about here]

We obtained the model performance index by first calculating multi-year annual mean climatologies from global gridded fields of models and validating data. The base period for the observations was 1979-1999, covering most of the well observed post-1979 satellite period. For some observations, fewer years were used if data over the entire period were not available. For the CMIP-1 models, long-term climatologies of the control run for Northern Hemisphere winter (December, January, February) and summer (June, July, August) conditions were downloaded from the archives and averaged to annual mean climatologies. The CMIP-2 climatologies were calculated by averaging the annual mean data of the control run over the years 61-80. The CMIP-3 'present-day' climatologies were formed using the same base period as for the observations, and the 'preindustrial' climatologies were taken from the last 20 simulation years of the corresponding control run. For any given model, only one member integration was included. In the rare case that a climate variable was not provided by a specific model, we replaced the unknown error by the mean error over the remaining models of the corresponding model generation. One model (BCC-CM1 from CMIP-3) was excluded because it only provided a small subset of variables needed for this study.

In determining the model performance index, we first calculated for each model and variable a normalized error variance e^2 by squaring the grid point differences between simulated (interpolated to the observational grid) and observed climate, normalizing on a grid point basis with the observed interannual variance, and averaging globally. In mathematical terms this can be written as

$$e_{vm}^{2}=\sum_{n}\left(w_{n}\left(\bar{s}_{vmn}-\bar{o}_{vn}\right)^{2}/\sigma_{vn}^{2}\right), \tag{1}$$

where $\bar{s}_{vmn}$ is the simulated climatology for climate variable (v), model (m), and grid point (n). $\bar{o}_{vn}$ is the corresponding observed climatology, w_n are proper weights needed for area and mass averaging, and σ_{vn}^{2} is the interannual variance from the validating observations. The normalization with the interannual variance helped to homogenize errors from different regions and variables. In order to ensure that different climate variables received similar weights when combining their errors, we next scaled e^2 by the average error found in a reference ensemble of models, i.e.,

$$I_{vm}^{2}=e_{vm}^{2}/\overline{e_{vm}^{2}}^{m=20C3M}, \tag{2}$$

where the overbar indicates averaging. The reference ensemble was the "present-day" CMIP-3 experiment. The final model performance index was formed by taking the mean over all climate variables (Table 1) and one model using equal weights,

$$I_{m}^{2}=\overline{I_{vm}^{2}}^{v}. \tag{3}$$

The final step combines the errors from different climate variables into one index. We justify this step from normalizing the individual error components prior to taking averages (Equ. 1 and 2). This guarantees that each component varies evenly around one and has roughly the same variance. In this sense, the individual I_{vm}^{2} values can be understood as rankings with respect to individual climate variables, and the final index is the mean over all ranks. Note that a very similar approach has been taken by Murphy et al. (2004).

Results

The outcome of the comparison of the 57 models in terms of the performance index I^2 is illustrated in the top three rows of Fig. 1. The I^2 index varies around one, with values greater than one for underperforming models and values less than one for more accurate models. Since I^2 is an indicator of model performance relative to the mean over the 'present-day' CMIP-3 ensemble, we used a logarithmic scale to display the index. The results indicate large differences from model to model in terms of their ability to match the observations of today's climate. Further, the results clearly demonstrate a continuous improvement in model performance from the early CMIP-1 to the latest CMIP-3 generation. To our knowledge, this is the first systematic attempt to compare the performance of entire generations of climate models by exploring their ability to simulate present climate. Fig. 1 also shows that the realism of the best models approaches that of atmospheric reanalysis (indicated by the green circle), but the models achieve this without being constrained by real observations.

We also obtained quantitative estimates of the robustness of the I^2 values by validating the models against a large synthetic ensemble of observational climatologies and by calculating the range of I^2 values encompassed the 5th and 95th percentiles. The synthetic ensemble was produced by selecting the years included in each climatology using bootstrapping (i.e., random selection with replacement). To the extent that the circles in Fig. 1 overlap, it is not possible to distinguish the performance of the corresponding models in a way that is statistically significant.

[Figure 1 about here]

Role of forcings

Given the more realistic forcing used for the ‘present-day’ CMIP-3 simulations, the superior outcome of the corresponding models is perhaps not too surprising. One might ask how important realistic forcing was in producing such good simulations. To this end, we included the ‘preindustrial’ CMIP-3 simulations into our comparison. Both the ‘present-day’ and the ‘preindustrial’ simulations were conducted with identical models. The only difference was the forcing used to drive the simulations, which was similar to preindustrial conditions for the ‘preindustrial’ and similar to present-day conditions for the ‘present-day’ experiments.

The outcome of validating the ‘preindustrial’ experiment against current climate is shown in the bottom row of Fig. 1. As expected, the I^2 values are now larger than for the ‘present-day’ simulations, indicating poorer performance. However, the mean difference between the two CMIP-3 simulations, which was due only to different forcings, is much smaller than that between CMIP-3 and the previous two model generations. The latter difference was due to different models and forcings combined. We conclude that the superior performance of the CMIP-3 models is mostly related to drastic model improvements, and that the forcings used to drive these models play a more subtle role.

Two developments, more realistic parameterizations and finer resolutions, are likely to be most responsible for the good performance seen in the latest model generation. For example, there has been a constant refinement over the years in how sub-grid scale processes are parameterized in models. Current models also tend to have higher vertical and horizontal resolution than their predecessors. Higher resolution reduces the dependency of models on parameterizations, eliminating problems since parameterizations are not always entirely physical. That increased resolution improves model performance has been shown in various previous studies (e.g., Mullen and Buizza 2002, Mo et al. 2005, Roeckner et al. 2006).

Sensitivity of the index

We now address the question of how sensitive our results are with respect to our particular choice of variables. We used bootstrapping to investigate how I^2 - averaged individually over the four model groups - varies with an increasing number v of variables. For any given v, we calculated I^2 many times, using every time different randomly chosen variable combinations taken from Table 1. As shown in Fig. 2, the spread of outcomes decreases with increasing number of variables. When six or more variables are used to calculate I^2, the average performances of the three model generations are well separated from each other - independent from the exact choice of variables. Only the two CMIP-3 experiments cannot be distinguished from each other, even for a very large number of variables. Also note that CMIP-3 performs always better than CMIP-1, and almost always better than CMIP-2, even when only one variable is included. These results indicate that I^2, when used to compare entire model generations, is robust with respect to the number and choice of selected variables.

[Figure 2 about here]

Value of the multi-model mean

We also investigated the performance of the multi-model means (black circles in Fig. 1), which are formed by averaging across the simulations of all models of one model generation and using equal weights. Notably, the multi-model mean usually outperforms any single model, and the CMIP-3 multi-model mean performs nearly as well as the reanalysis. Such performance improvement are consistent with earlier findings by Lambert and Boer (2001), Taylor et al. (2004), and Randall et al. (2007) regarding CMIP-1, AMIP-2, and CMIP-3 model output, respectively.

The use of multi-model ensembles is common practice in weather and short-term climate forecasting (Barnston et al. 2003; Krishnamurti et al. 2006; Palmer et al. 2004; Hagedorn et al. 2005), and it is starting to become important for long-term climate change predictions (Hewitt 2005; Murphy et al. 2004; Stainforth et al. 2005). For example, many climate change estimates of the recently released global warming report (IPCC 2007) of the Intergovernmental Panel on Climate Change are based on the multi-model simulations from the CMIP-3 ensemble. The report dealt with the problem of inconsistent predictions, resulting from the use of different models, by simply taking the average of all models as the best estimate for future climate change. Our results indicate that multi-model ensembles are a legitimate and effective means to improve the outcome of climate simulations. As yet, it is not exactly clear why the multi-model mean is better than any individual model. One possible explanation is that the model solutions scatter more or less evenly about the truth (unless the errors are systematic), and the errors behave like random noise that can be efficiently removed by averaging. Such noise arises from internal climate variability (Barnett et al. 1994), and probably to a much larger extent from uncertainties in the formulation of models (Murphy et al. 2004; Stainforth et al. 2005).

[Figure 3 about here]

Role of flux correction

When discussing coupled model performances, one must take into account that earlier models are generally flux corrected, whereas most modern models do not require such corrections (Fig. 3). Flux correction, or adding artificial terms of heat, momentum, and freshwater at the air-sea interface, prevents models from drifting to unrealistic climate states when integrating over long periods of time. The drift, which occurs even under unforced conditions, is the result of small flux imbalances between ocean and atmosphere. The effects of these imbalances accumulate over time and tend to modify the mean temperature and/or salinity structure of the ocean. The technique of flux correction attracts concern because of its inherently non-physical nature (McAvaney et al. 2001). The artificial corrections make simulations at the ocean surface more realistic, but only for artificial reasons. This is demonstrated by the increase in systematic biases (defined as the multi-model mean minus the observations) in sea surface temperatures from the mostly flux corrected CMIP-1 models to the generally uncorrected CMIP-3 models (Fig. 4a). Because sea surface temperatures exert an important control on the exchange of properties across the air-sea interface, corresponding errors readily propagate to other climate fields. This can be seen in Fig. 4b, which shows that biases in ocean temperatures tend to be accompanied by same-signed temperature biases in the free troposphere. On the other hand, the reduction of strong lower stratospheric cold biases in the CMIP-3 models indicates considerable model improvements. These cold biases are likely related to the low vertical and horizontal resolution of former model generations (Roeckner et al. 2006) and to the lack of parameterizations for small-scale gravity waves, which break, deposit momentum, and warm the middle atmosphere over the high latitudes. Modern models use appropriate parameterizations to replace the missing momentum deposition.

[Figure 4 about here]

Conclusion

Using a composite measure of model performance, we objectively determined the ability of three generations of models to simulate present-day mean climate. Current models are certainly not perfect, but we found that they are much more realistic than their predecessors. This is mostly related to the enormous progress in model development that took place over the last decade, which is partly due to more sophisticated model parameterizations, but also to the general increase in computational resources, which allows for more thorough model testing and higher model resolution. Most of the current models not only perform better, they are also no longer flux corrected. Both – improved performance and more physical formulation – suggest that an increasing level of confidence can be placed in model based predictions of climate. This, however, is only true to the extent that the performance of a model in simulating present mean climate is related to the ability to make reliable forecasts of long-term trends. It is to hope that these advancements will enhance the public credibility of model predictions and help to justify the development of even better models.

Given the many issues that complicate model validation, it is perhaps not too surprising that the present study has some limitations. First, we note the caveat that we were only concerned with the time mean state of climate. Higher moments of climate, such as temporal variability, are probably equally as important for model performance, but we were unable to investigate these. Another critical point is the calculation of the performance index. For example, it is unclear how important climate variability is compared to the mean climate, exactly which the optimum selection of climate variables is, and how accurate the used validation data are. Another complicating issue is that error information contained in the selected climate variables is partly redundant. Clearly, more work is required to answer the above questions, and it is to hope that the present study will stimulate further research in the design of more robust metrics. For example, a future improved version of the index should consider possible redundancies and assign appropriate weights to errors from different climate variables. However, we do not think that our specific choices in this study affect our overall conclusion that there has been a measurable and impressive improvement in climate model performance over the past decade.

Acknowledgements

We thank Anand Gnanadesikan, Karl Taylor, Peter Gleckler, Tim Garrett, and Jim Steenburgh for useful discussions and comments, Dan Tyndall for help with the figures, and Curt Covey and Steve Lambert for providing the CMIP-1 and CMIP-2 data. The comments of three anonymous reviewers, which helped to improve and clarify the paper, are also thanked. We acknowledge the modeling groups for providing the CMIP-3 data for analysis, the Program for Climate Model Diagnosis and Intercomparison for collecting and archiving the model output, and the JSC/CLIVAR Working Group on Coupled Modeling for organizing the model data analysis activity. The multi-model data archive is supported by the Office of Science, U.S. Department of Energy. This work was supported by NSF grant ATM0532280 and by NOAA grant NA06OAR4310148.

References

AchutaRao, K. and K. R. Sperber, 2006: ENSO simulation in coupled ocean-atmosphere models: are the current models better? *Clim. Dyn.*, **27,** 1-15.

Armstrong, R. L., M. J. Brodzik, K. Knowles, and M. Savoie, 2005: Global monthly EASE-Grid snow water equivalent climatology, Digital media, National Snow and Ice Data Center, Boulder, CO.

Bader, D., Ed., 2004: *An Appraisal of Coupled Climate Model Simulations*. Lawrence Livermore National Laboratory, 183 pp.

Barnett, T. P., L. Bengtsson, K. Arpe, M. Flugel, N. Graham, M. Latif, J. Ritchie, E. Roeckner, U. Schlese, U. Schulzweida, et al., 1994: Forecasting Global ENSO-Related Climate Anomalies. *Tellus-A*, **46,** 381-397.

Barnston, A. G., S. J. Mason, L. Goddard, D. G. Dewitt, and S. E. Zebiak, 2003: Multimodel Ensembling in Seasonal Climate Forecasting at IRI. *Bull. Amer. Meteor. Soc.*, **84,** 1783–1796.

Boer, G. J. and S. J. Lambert, 2001: Second order space-time climate difference statistics. *Clim. Dyn.*, **17,** 213- 218.

Bony, S. and J.-L. Dufresne, 2005: Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. *Geophys. Res. Lett.*, **32,** doi:10.1029/2005GL023851.

Covey, C., K. M. AchutaRao, U. Cubasch, P. Jones, S. J. Lambert, M. E. Mann, T. J. Phillips, and K. E. Taylor, 2003: An Overview of Results from the Coupled Model Intercomparison Project (CMIP). *Global Planet. Change*, **37,** 103-133.

Gates, W., U. Cubasch, G. Meehl, J. Mitchell, and R. Stouffer, 1993: An intercomparison of selected features of the control climates simulated by coupled ocean-atmosphere general circulation models, World Meteorological Organization, Geneva, World Climate Research Programme WCRP-82 WMO/TD No. 574, 46 pp.

Hagedorn, R., F. J. Doblas-Reyes, and T. N. Palmer, 2005: The rationale behind the success of multi-model ensembles in seasonal forecasting - I. Basic concept. *Tellus-A*, **57,** 219–233, doi:10.1111/j.1600-0870.2005.00103.x.

Hewitt, C. D., 2005: The ENSEMBLES Project: Providing ensemble-based predictions of climate changes and their impacts. *EGGS newsletter*, **13,** 22-25.

IPCC, 2007: Climate Change 2007: The physical science basis. Summary for policymakers**,** 21.

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

Jones, R., 2005: Senate Hearing Demonstrates Wide Disagreement About Climate Change. *FYI Number 142*, American Institute of Physics.

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, et al., 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77,** 437-471.

Krishnamurti, T. N., A. Chakraborty, R. Krishnamurti, W. K. Dewar, and C. A. Clayson, 2006: Seasonal Prediction of Sea Surface Temperature Anomalies Using a Suite of 13 Coupled Atmosphere-Ocean Models. *J. Clim.*, **19,** 6069-6088.

Lahsen, M., 2005: Seductive simulations? Uncertainty distribution around climate models. *Social Studies of Science*, 895-922.

Lambert, S. J. and G. J. Boer, 2001: CMIP1 evaluation and intercomparison of coupled climate models *Clim. Dyn.*, **17,** 83-116.

Levitus, S., T. P. Boyer, M. E. Conkright, J. A. T. O' Brien, L. S. C. Stephens, D. Johnson, and R. Gelfeld, 1998: NOAA Atlas NESDIS 18,World Ocean Database 1998: VOLUME 1: INTRODUCTION. *U.S. Government Printing Office***,** 346 pp.

Lin, J.-L., G. N. Kiladis, B. E. Mapes, K. M. Weickmann, K. R. Sperber, W. Lin, M. C. Wheeler, S. D. Schubert, A. Del Genio, L. J. Donner, et al., 2006: Tropical Intraseasonal Variability in 14 IPCC AR4 Climate Models. Part I: Convective Signals. *J. Clim.*, **19,** 2665-2690.

Lindzen, R., 2006: Climate of Fear. *Wall Street Journal*.

McAvaney, B. J., C. Covey, S. Joussaume, V. Kattsov, A. Kitoh, W. Ogana, A. J. Pitman, A. J. Weaver, R. A. Wood, and Z.-C. Zhao, 2001: The science of climate change Ch. 8: Model evaluation. *Climate Change 2001*, J. T. Houghton, Y. Ding, D. J. Griggs, M. Noguer, P. J. v. d. Linden, X. Dai, K. Maskell, and C. A. Johnson, Eds., Cambridge Univ. Press, 471-523.

Mechoso, C. R., A. W. Robertson, N. Barth, M. K. Davey, P. Delecluse, and S. I. P. R. Gene, B. Kirtman, M. Latif, H. Le Treut, T. Nagai, J.D. Neelin, S. G. H. Philander, J. Polcher, P. S. Schopf, T. Stockdale, M. J. Suarez, L. Terray, O. Thual, and J. J. Tribbia, 1995: The Seasonal Cycle over the Tropical Pacific in Coupled Ocean–Atmosphere General Circulation Models. *Mon. Wea. Rev.*, **123,** 2825–2838.

Meehl, G. A., G. J. Boer, C. Covey, M. Latif, and R. J. Stouffer, 2000: The Coupled Model Intercomparison Project (CMIP). *Bull. Amer. Meteor. Soc.*, **81,** 313-318.

Meehl, G. A., C. Covey, B. McAvaney, M. Latif, and R. J. Stouffer, 2005: Overview of the coupled model intercomparison project. *Bull. Amer. Meteor. Soc.*, **86,** 89-93.

Min, S.-K. and A. Hense, 2006: A Bayesian approach to climate model evaluation and multi-model averaging with an application to global mean surface temperatures from IPCC AR4 coupled climate models. *Geophys. Res. Lett.*, **33,** doi:10.1029/2006GL025779.

Mo, K. C., J.-K. Schemm, H. M. H. Juang, R. W. Higgins, and Y. Song, 2005: Impact of Model Resolution on the Prediction of Summer Precipitation over the United States and Mexico. *J. Climate*, **18,** 3910-3927.

Mullen, S. L. and R. Buizza, 2002: The Impact of Horizontal Resolution and Ensemble Size on Probabilistic Forecasts of Precipitation by the ECMWF Ensemble Prediction System. *Wea. Forecasting*, **17,** 173-191.

Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth, 2004: Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, **430,** 768-772.

Oldenborgh, G. J. v., S. Y. Philip, and M. Collins, 2005: El Nino in a changing climate: a multi-model study. *Ocean Sci.*, **1,** 81–95.

Palmer, T. N., A. Alessandri, U. Andersen, P. Cantelaube, M. Davey, P. Delecluse, Deque, M. , E. Diez, F. J. Doblas-Reyes, H. Feddersen, et al., 2004: Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER). *Bulletin of the American Meteorological Society*, **85,** 853-872.

Parker, D. E., C. K. Folland, A. Bevan, M. N. Ward, M. Jackson, and K. Maskell, 1995: Marine surface data for analysis of climate fluctuations on interannual to century timescale. *Natural Climate Variability on Decade-to-Century Time Scales*, National Academy Press, 241-250.

PCMDI, 2007: IPCC Model Ouput: http://www-pcmdi.llnl.gov/ipcc/about_ipcc.php.

Randall, D. A., R. A. Wood, S. Bony, R. Colman, T. Fichefet, J. Fyfe, V. Kattsov, A. Pitman, J. Shukla, J. Srinivasan, et al., 2007: Climate Models and Their Evaluation. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller, Ed., Cambridge University Press.

Reichler, T. and J. Kim, 2007: Uncertainties in the climatologies of reanalysis: Assessment using four global data sets. *J. Geophys. Res.***,** (submitted).

Roeckner, E., R. Brokopf, M. Esch, M. Giorgetta, S. Hagemann, L. Kornblueh, E. Manzini, U. Schlese, and U. Schulzweida, 2006: Sensitivity of Simulated Climate to Horizontal and Vertical Resolution in the ECHAM5 Atmosphere Model. *J. Climate*, **19,** 3771-3791.

Schmidt, G. A., D. T. Shindell, R. L. Miller, M. E. Mann, and D. Rind, 2004: General circulation modelling of Holocene climate variability. *Quaternary Sci. Rev.*, **23,** 2167-2181.

Simmons, A. J. and J. K. Gibson, 2000: The ERA-40 Project Plan. ERA-40 Project Rep., Series No. 1, 62 pp.

Singer, S. F., 1999: Human contribution to climate change remains questionable. *Eos Trans. AGU,* **80(16),** 183-183.

Stainforth, D. A., T. Aina, C. Christensen, M. Collins, N. Faull, D. J. Frame, J. A. Kettleborough, S. Knight, A. Martin, J. M. Murphy, et al., 2005: Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, **433,** 403-406.

Stenchikov, G., A. Robock, V. Ramaswamy, M. D. Schwarzkopf, K. Hamilton, and S. Ramachandran, 2002: Arctic Oscillation response to the 1991 Mount Pinatubo eruption: Effects of volcanic aerosols and ozone depletion. *J. Geophys. Res.*, **107,** doi:10.1029/2002JD002090.

Sun, D.-Z., T. Zhang, C. Covey, S. A. Klein, W. D. Collins, J. J. Hack, J. T. Kiehl, G. A. Meehl, I. M. Held, and M. Suarez, 2006: Radiative and Dynamical Feedbacks over the Equatorial Cold Tongue: Results from Nine Atmospheric GCMs. *J. Clim.*, **19,** 4059-4074.

Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, **106(D7),** 7183-7192, 10.1029/2000JD900719.

Taylor, K. E., P. J. Gleckler, and C. Doutriaux, 2004: Tracking Changes in the Performance of AMIP Models, LLNL, Livermore, 261 pp.

Williamson, D. L., 1995: Skill scores from the AMIP simulations. *First Int. AMIP Scientific Conf.*, Monterey, CA, World Meteorological Organization, 253–256.

Woodruff, S. D., R. J. Slutz, R. L. Jenne, and P. M. Steurer, 1987: A Comprehensive Ocean-Atmosphere Data Set. *Bull. Amer. Meteor. Soc.*, **68,** 1239-1250.

Xie, P. P. and P. A. Arkin, 1998: Global monthly precipitation estimates from satellite-observed outgoing longwave radiation. *J. Clim.*, **11,** 137-164.

Yu, L., R. A. Weller, and B. Sun, 2004: Improving latent and sensible heat flux estimates for the Atlantic Ocean (1988-1999) by a synthesis approach. *J. Clim.*, **17,** 373-393.

Zhang, Y., W. B. Rossow, A. A. Lacis, V. Oinas, and M. I. Mishchenko, 2004: Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global data sets: Refinements of the radiative transfer model and the input data. *J. Geophys. Res.*, **109,** D19105.

Captions

Fig. 1. Performance index I^2 for individual models (circles) and model generations (rows). Best performing models have low I^2 values and are located towards the left. Circles sizes indicate the length of the 95% confidence intervals. Letters and numbers identify individual models (see supplemental online material); flux corrected models are labeled in red. Grey circles show the average I^2 of all models within one model group. Black circles indicate the I^2 of the multi-model mean taken over one model group. The green circle (REA) corresponds to the I^2 of the NCEP/NCAR reanalyses (Kalnay et al. 1996). Last row (PICTRL) shows I^2 for the preindustrial control experiment of the CMIP-3 project.

Fig. 2. Spread of I^2 values (lowest to highest) for an increasing number of randomly chosen variables *v*. Shown are index values averaged individually over the four model groups (corresponding to the grey circles in Fig. 1). In order to avoid non-unity results for 20C3M, all values were normalized by the mean I^2 over all three model generations, and not by the mean of the 20C3M group alone (as in Fig. 1, see Equ. 2).

Fig. 3. Fraction of flux-adjusted models amongst the three model generations.

Fig. 4. Systematic biases for the three model generations. (a) Biases in annual mean climatological mean sea surface temperatures (in K). (b) Biases in zonal mean air temperatures (in K). Statistically significant biases that pass a Student's t-test at the 95%

level are shown in color; other values are suppressed and shown in white. Grey areas denote no or insufficient data.

Table 1. Climate variables and corresponding validation data. Variables listed as ‘zonal mean’ are latitude-height distributions of zonal averages on twelve atmospheric pressure levels between 1000 and 100 hPa. Those listed as ‘ocean’, ‘land’, or ‘global’ are single-level fields over the respective regions. The variable ‘net surface heat flux’ represents the sum of six quantities: Incoming and outgoing shortwave radiation; incoming and outgoing longwave radiation; and latent and sensible heat fluxes. Period indicates years used to calculate observational climatologies.

Table 1

Variable	Domain	Validation data	Period
sea level pressure	ocean	ICOADS (Woodruff et al. 1987)	1979-1999
air temperature	zonal mean	ERA-40 (Simmons and Gibson 2000)	1979-1999
zonal wind stress	ocean	ICOADS (Woodruff et al. 1987)	1979-1999
meridional wind stress	ocean	ICOADS (Woodruff et al. 1987)	1979-1999
2 m air temperature	global	CRU (Jones et al. 1999)	1979-1999
zonal wind	zonal mean	ERA-40 (Simmons and Gibson 2000)	1979-1999
meridional wind	zonal mean	ERA-40 (Simmons and Gibson 2000)	1979-1999
net surface heat flux	ocean	ISCCP (Zhang et al. 2004), OAFLUX (Yu et al. 2004)	1984 (1981) -1999
precipitation	global	CMAP (Xie and Arkin 1998)	1979-1999
specific humidity	zonal mean	ERA-40 (Simmons and Gibson 2000)	1979-1999
snow fraction	land	NSIDC (Armstrong et al. 2005)	1979-1999
sea surface temperature	ocean	GISST (Parker et al. 1995)	1979-1999
sea ice fraction	ocean	GISST (Parker et al. 1995)	1979-1999
sea surface salinity	ocean	NODC (Levitus et al. 1998)	variable

Figure 1

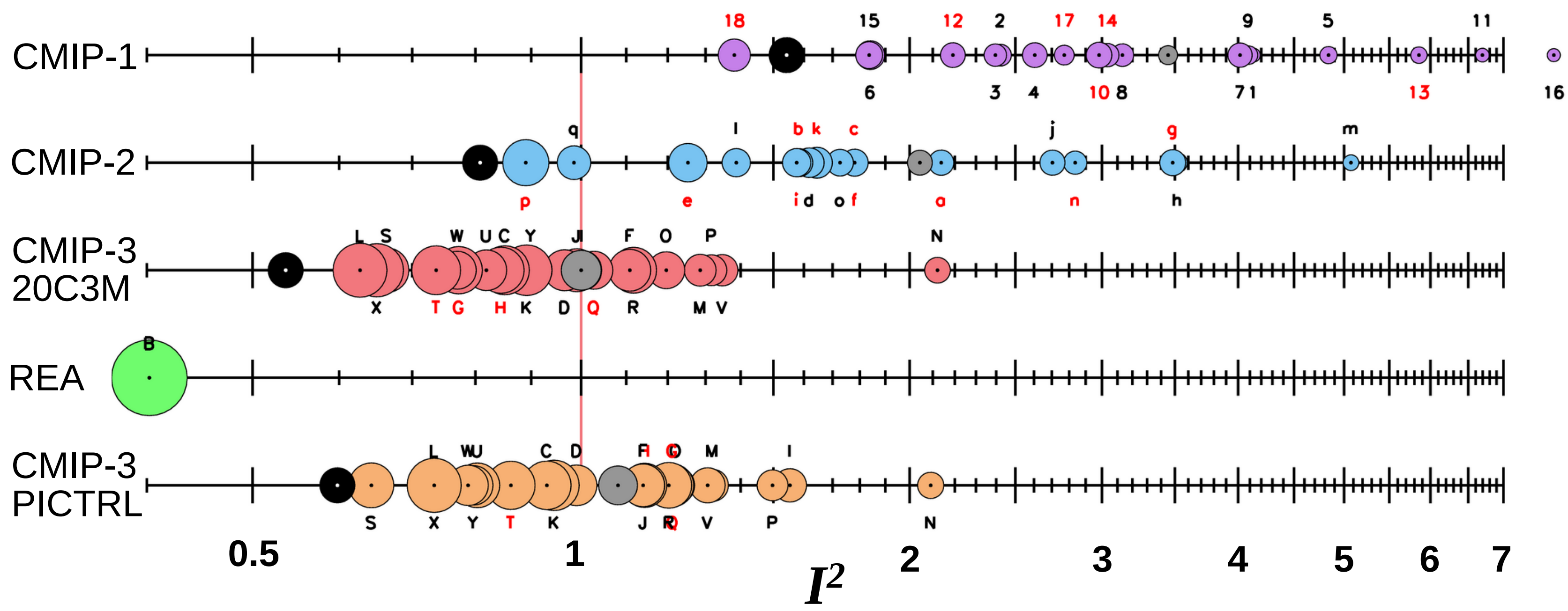
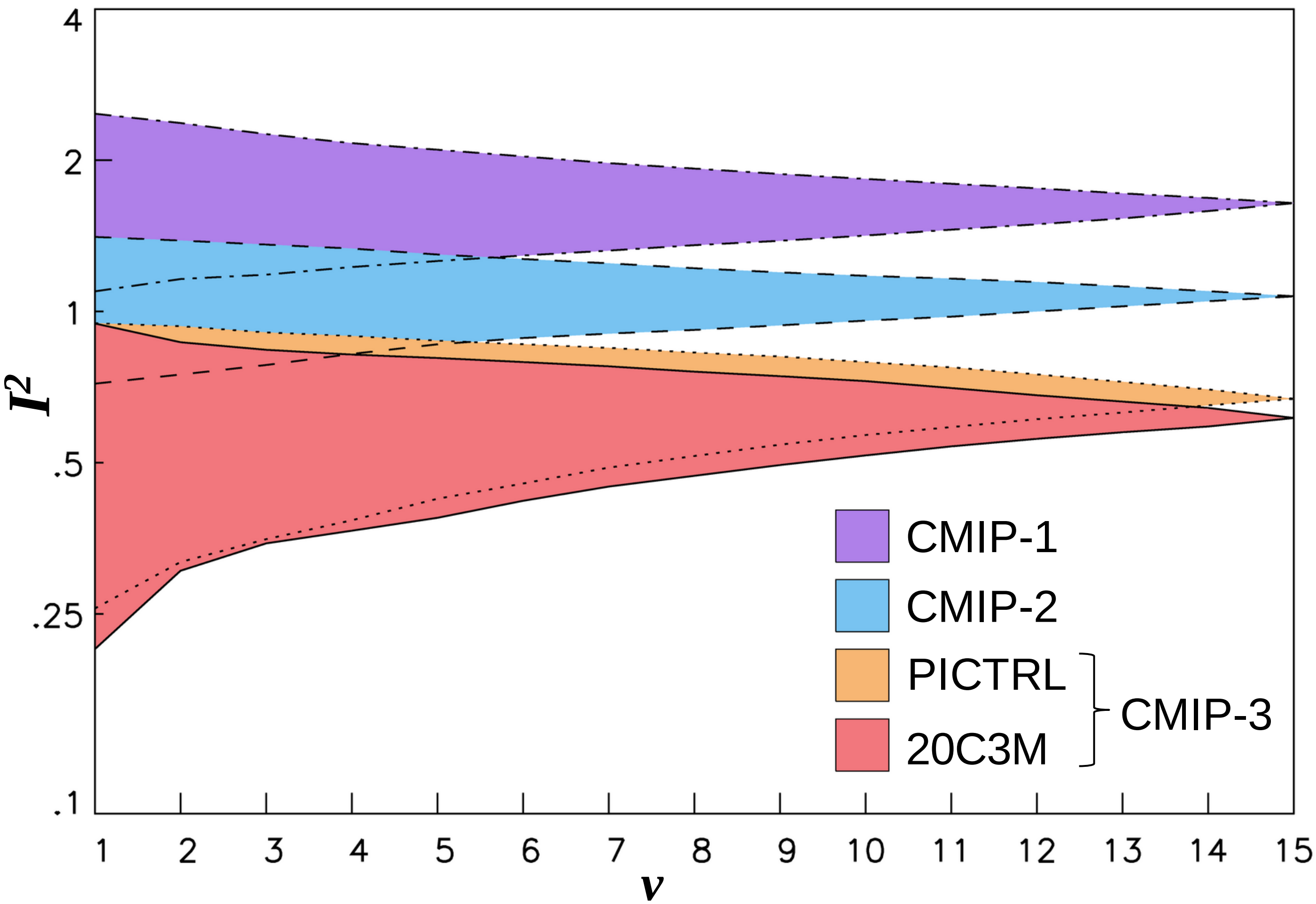
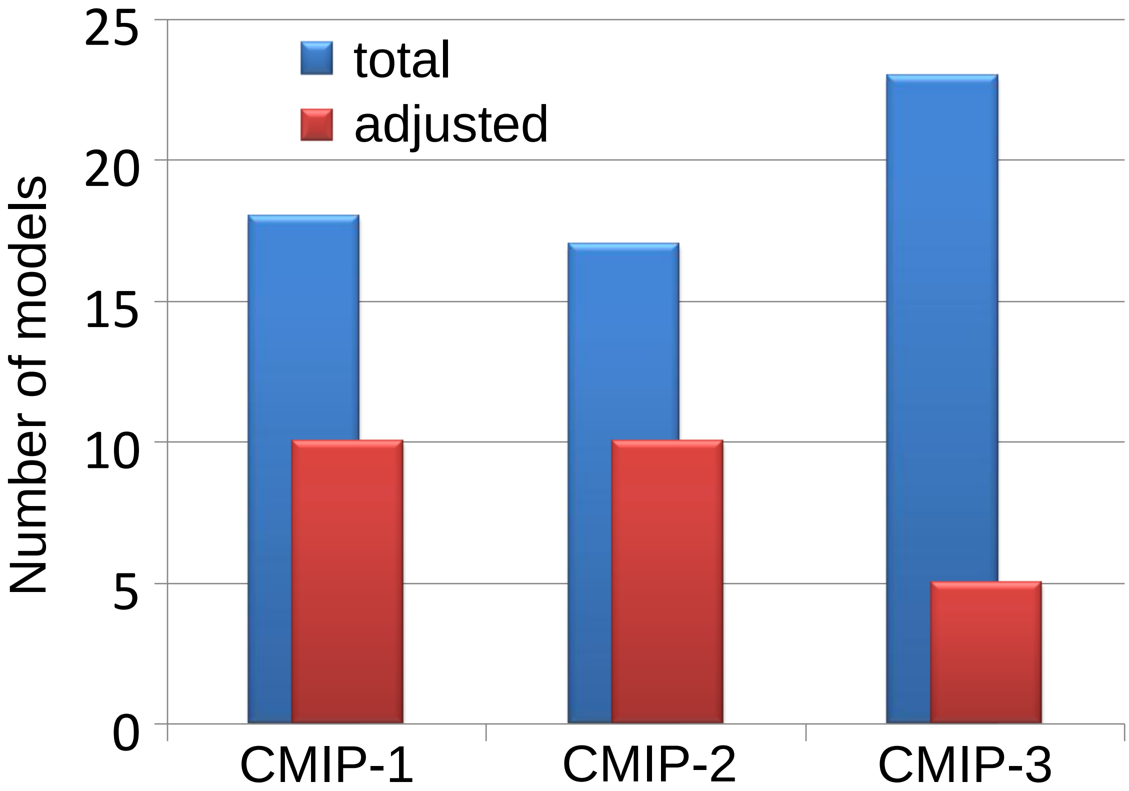


Figure 2



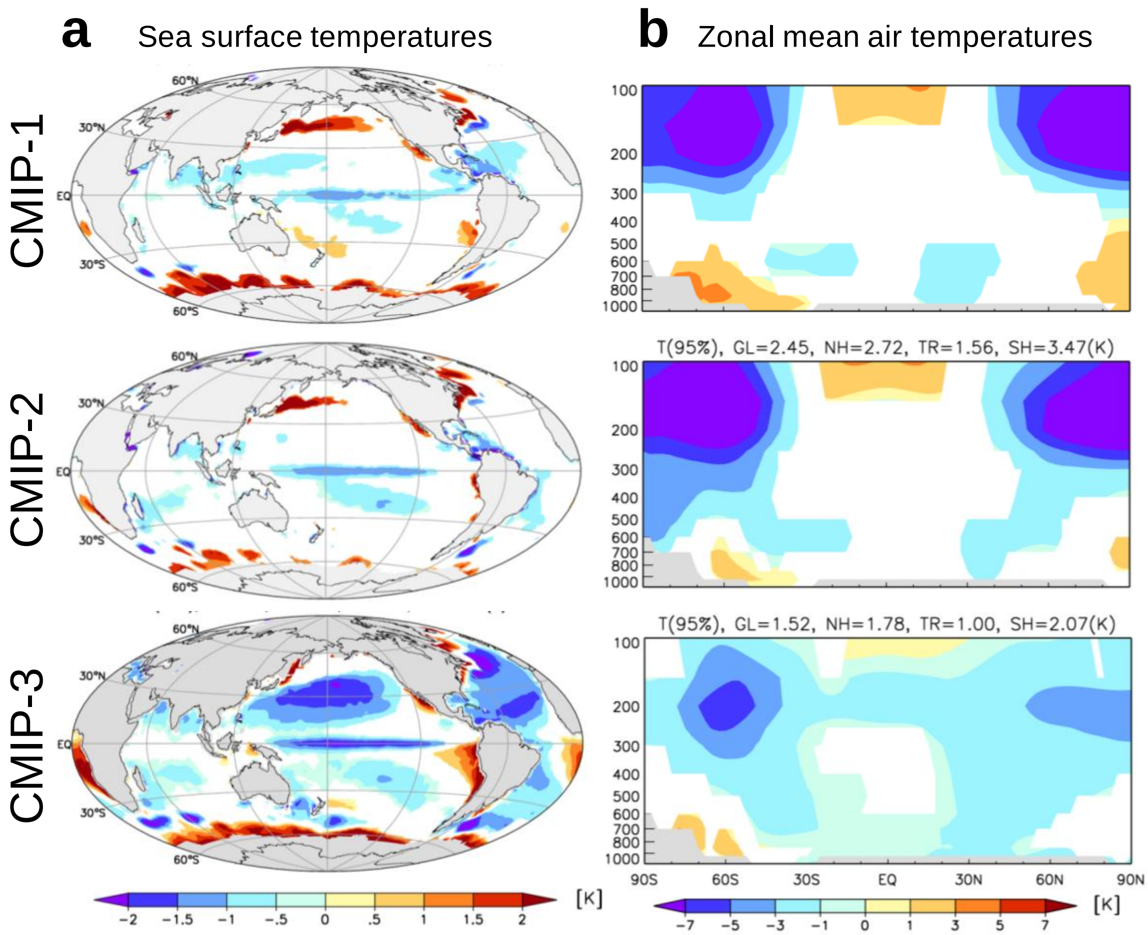
4
2
1
.5
.25
.1
I^2
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
v
CMIP-1
CMIP-2
PICTRL
20C3M
CMIP-3

Figure 3



Number of models
25
20
15
10
5
0
total
adjusted
CMIP-1
CMIP-2
CMIP-3

Figure 4



a Sea surface temperatures
b Zonal mean air temperatures
CMIP-1
CMIP-2
CMIP-3
T(95%), GL=2.45, NH=2.72, TR=1.56, SH=3.47(K)
T(95%), GL=1.52, NH=1.78, TR=1.00, SH=2.07(K)
90S 60S 30S EQ 30N 60N 90N
[K]

ONLINE SUPPLEMENT

Model identifiers and characteristics

Tables S1 to S3 list the names and identifiers of the different models investigated in this study along with some of their main characteristics.

Table S1: Identifiers and characteristics of the CMIP-1 models included in this study. Grid resolution: longitude x latitude. *L* denotes number of vertical layers. The column for flux adjustment uses the following notation: H: heat; M: momentum; W: water; X: none.

ID	Short Name	Model	Atmosphere	Ocean	Reference	Flux Adj.
01	BMRC	BMRC1, Australia	R21 (5.6x3.2), L9	5.6x3.2, L12	Power et al. 1993	X
02	CCCMA	CCCma1, Canada	T32 (3.8x3.8), L10	1.8x1.8, L29	Boer et al. 2000	H, W
03	CCSR	CCSR, Japan	T21 (5.6x5.6), L20	2.8x2.8, L17	Emori et al. 1999	H, W
04	CERFACS	CERFACS1, France	T21 (5.6x5.6), L30	2.0x2.0, L31	Guilyardi and Madec 1997	X
05	COLA	COLA1, U.S.A.	R15 (7.5x4.5), L9	1.5x1.5, L20	Schneider and Zhu 1998	X
06	CSIRO	CSIRO, Australia	R21 (5.6x3.2), L9	5.6x3.3, L21	Gordon and O'Farrell 1997	H, W, M
07	GFDL	GFDL_R15_a, U.S.A.	R15 (7.5x4.5), L9	3.7x4.5, L12	Manabe and Stouffer 1996	H, W
08	GISSM	GISS (Miller), U.S.A.	5.0x4.0, L9	5.0x4.0, L16	Miller and Jiang 1996	X
09	GISSR	GISS (Russell), U.S.A.	5.0x4.0, L9	5.0x4.0, L13	Russell et al. 1995	X
10	IAP	IAP/LASG1, China	R15 (7.5x4.5), L9	5.0x4.0, L20	Zhang et al. 2000	H, W, M
11	LMD	LMD/IPSL1, France	3.8x5.6, L15	2.0x2.0, L31	Braconnot et al. 1997	X
12	MPIE3	ECHAM3+LSG, Germany	T21 (5.6x5.6), L19	4.0x4.0, L11	Voss et al. 1998	H, W, M
13	MPIE4	ECHAM4+OPYC3	T42 (2.8x2.8), L19	2.8x2.8, L11	Roeckner et al. 1996	H, W, M
14	MRI	MRI1, Japan	5.0x4.0, L15	2.5x2.0, L21	Tokioka et al. 1996	H, W
15	NCARCSM	NCAR (CSM), U.S.A.	T42 (2.8x2.8), L18	2.4x2.0, L45	Boville and Gent 1998	X
16	NCARWM	NCAR (WM), U.S.A.	R15 (7.5x4.5), L9	1.0x1.0, L20	Washington and Coauthors 2000	X
17	NRL	NRL1, U.S.A.	T47 (2.5x2.5), L18	2.0x1.0, L25	Li and Hogan 1999	H, W
18	UKMO	UKMO (HadCM2), UK	3.75x2.5, L19	3.75x2.5, L20	Johns et al. 1997	H, W

Table S2: As Table S1 but for CMIP-2 models.

ID	Short Name	Model	Atmosphere	Ocean	Reference	Flux Adj.
a	BMRC	BMRC, Australia	R21 (5.6x3.2), L17	5.6x3.2, L12	Colman 2001	H, W, sfc SW rad.
b	CCCM	CCCma, CGCM1, Canada	T32 (3.8x3.8), L10	1.8x1.8, L29	Kim et al. 2003	H, W
c	CCSR	CCSR, Japan	T21 (5.6x5.6), L20	2.8x2.8, L17	Emori et al. 1999	H, W
d	CERF	CERFACS2 (ARPEGE/OPA2), France	T31 (3.9x3.9), L19	2.0x2.0, L31	Barthelet et al. 1998	X
e	CSIRO	CSIRO(Mk2), Australia	R21 (5.6x3.2), L9	5.6x3.2, L21	*Hirst et al. 2000*	H, W, M
f	MPIE3	ECHAM3+LSG, Germany	T21 (5.6x5.6), L19	4.0x4.0, L11	Voss et al. 1998	H, W, M
g	GFDL	GFDL_R15_a, U.S.A	R15 (7.5x4.5), L9	3.7x4.5, L12	Dixon et al. 2003	H, W
h	GISS	GISS (Russell), U.S.A	5.0x4.0, L9	5.0x4.0, L13	Russell and Rind 1999	X
i	IAP	IAP/LASG2, China	R15 (7.5x4.5), L9	5.0x4.0, L20	Zhang et al. 2000	H, W, M
j	LMD	LMD/IPSL2, France	5.6x3.8, L15	2.0x2.0, L31	Laurent et al. 1998	X
k	MRI	MRI2 (Tokioka), Japan	5.0x4.0, L15	2.5x2.0, L21	Tokioka et al. 1996	H, W
l	NCARC	NCAR(CSM), U.S.A	T42 (2.8x2.8), L26	1.0x(0.3-1.0),L40	Buja and Craig 2002	X
m	NCARW	NCAR-WM, U.S.A	R15 (7.5x4.5), L9	1.0x1.0, L20	Washington and Meehl 1996	X
n	NRL	NRL2, Monterey	T47 (2.5x2.5), L18	1.0x1.0, L25	Li and Hogan 1999	H, W
o	PCM	DOE-PCM, U.S.A	T42 (2.8x2.8), L18	0.67x0.67, L32	Washington and Coauthors 2000	X
p	UKMO	UKMO (HadCM2), UK	3.75x2.5, L19	3.75x2.5, L20	Johns et al. 1997	H, W
q	UKMO3	UKMO (HadCM3), UK	3.75x2.5, L19	1.25x1.25, L20	Gordon et al. 2000	X

Table S3: As Table S1 but for CMIP-3 models.

ID	Short name	Model	Atmosphere	Ocean	Reference	Flux Adj.
C	MIRCH	MIROC3.2 (hires), Japan	T106, L56	0.28x0.19, L47	K-1-model-developers 2004	X
D	MIRCM	MIROC3.2 (medres), Japan	T42, L20	1.4x(0.5-1.4) L43	K-1-model-developers 2004	X
F	BCCRC	BCCR-BCM2.0, Norway	T63, L31	1.5x0.5, L35	Furevik et al. 2003	X
G	C3T47	CGCM3.1 (T47), Canada	T47 (3.75x3.75), L31	1.85x1.85, L29	Kim et al. 2002	H,W
H	C3T63	CGCM3.1 (T63), Canada	T63 (2.8x2.8), L 31	1.4x0.94, L29	Flato and Boer 2001	H,W
I	CNRMC	CNRM-CM3, France	T63 (2.8x2.8), L45	1.875x(0.5-2), L31	Salas-Melia et al. 2005	X
J	CSIRO	CSIRO-Mk3.0, Australia	T63, L18	1.875x0.84, L31	Gordon et al. 2002	X
K	GFD20	GFDL-CM2.0, USA	2.5x2.0, L24	1.0x(1/3-1), L50	Delworth et al. 2006	X
L	GFD21	GFDL-CM2.1, USA	2.5x2.0, L24	1.0x(1/3-1), L50	Delworth et al. 2006	X
M	GISSA	GISS-AOM, USA	4x3, L12	4x3, L16	Lucarini and Russell 2002	X
N	GISSH	GISS-EH, USA	5x4, L20	5x4, L13	Schmidt et al. 2006	X
O	GISSR	GISS-ER, USA	5x4, L20	5x4, L13	Schmidt et al. 2006	X
P	IAPFG	IAP-FGOALS1-0-G, China	2.8x2.8, L26	1x1, L16	Yu et al. 2004	X
Q	INMCM	INM-CM3.0, Russia	5x4, L21	2.5x2, L33	Volodin and Diansky 2004	W
R	IPSLC	IPSL-CM4, France	2.5x3.75, L19	2x(1-2), L30	Marti et al. 2005	X
S	MPICM	ECHAM5/MPI-OM	T63, L32	1x1, L41	Min et al. 2005	X
T	MRICM	MRI-CGCM2-3-2A, Japan	T42, L30	2.5x(0.5-2.0)	Yukimoto and Noda 2002	H,M,W
U	NCARC	NCAR-CCSM3, USA	T85L26, 1.4x1.4	1x(0.27-1), L40	Collins et al. 2005	X
V	NCARP	NCAR-PCM, USA	T42 (2.8x2.8), L18	1x(0.27-1), L40	Kiehl and Gent 2004	X
W	UKMOC	UKMO-HadCM3, UK	3.75x2.5, L19	1.25x1.25, L20	Gordon et al. 2000	X
X	UKMOG	UKMO-HadGEM1, UK	1.875x1.25, L38	1.25x1.25, L20	Johns et al. 2004	X
Y	INGVE	INGV-SXG, Italy	T42, L19	2x(0.5-2), L31	Gualdi et al. 2003	X

Model References

Barthelet, P., S. Bony, P. Braconnot, A. Braun, D. Cariolle, E. Cohen-Solal, J.-L. Dufresne, P. Delecluse, M. Deque, L. Fairhead, et al., 1998: Simulations couplees globales de changements climatiques associes a une augmentation de la teneur atmospherique en CO2. *Comptes Rendus de l'Academie des Sciences, Sciences de la terre et des planetes*, **326,** 677-684.

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

Boville, B. A. and P. R. Gent, 1998: The NCAR Climate System Model, Version One. *J. Clim.*, **11,** 1115-1130.

Braconnot, P., O. Marti, and S. Joussaume, 1997: Adjustment and feedbacks in a global coupled ocean-atmosphere model. *Clim. Dyn.*, **13,** 507-519.

Buja, L. and T. Craig, 2002: CCSM2.0.1 User's Guide. *Accessible online at http://www.ccsm.ucar.edu/models/ccsm2.0.1/ccsm*.

Collins, N., G. Theurich, C. DeLuca, M. Suarez, A. Trayanov, V. Balaji, P. Li, W. Yang, C. Hill, and A. d. Silva, 2005: Design and implementation of components in the Earth System Modeling Famework. *International Journal of High Performance Computing Applications*, **19,** 341-350.

Colman, R. A., 2001: On the vertical extent of atmospheric feedbacks. *Clim. Dyn.*, **17,** 391-405.

Delworth, T. L., A. J. Broccoli, A. Rosati, R. J. Stouffer, V. Balaji, J. A. Beesley, W. F. Cooke, K. W. Dixon, J. Dunne, K. A. Dunne, et al., 2006: GFDL's CM2 Global Coupled Climate Models. Part I: Formulation and Simulation Characteristics. *J. Clim.*, **19,** 643-674.

Dixon, K. W., T. L. Delworth, T. R. Knutson, M. J. Spelman, and R. J. Stouffer, 2003: A comparison of climate change simulations produced by two GFDL coupled climate models. *Global Planet. Change*, **37,** 81-102.

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

Flato, G. M. and G. J. Boer, 2001: Warming asymmetry in climate change simulations. *Geophys. Res. Lett.*, **28,** 195-198.

Furevik, T., M. Bentsen, H. Drange, I. K. T. Kindem, N. G. Kvamstø, and A. Sorteberg, 2003: Description and validation of the Bergen Climate Model: ARPEGE coupled with MICOM. *Clim. Dyn.*, **21,** 27-51.

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

Gordon, H. B. and S. P. O'Farrell, 1997: Transient Climate Change in the CSIRO Coupled Model with Dynamic Sea Ice. *Mon. Wea. Rev.*, **125,** 875-908.

Gordon, H. B., L. D. Rotstayn, J. L. McGregor, M. R. Dix, E. A. Kowal-czyk, S. P. O’Farrell, L. J. Waterman, A. C. Hirst, S. G. Wil-son, M. A. Collier, et al., 2002: The CSIRO Mk3 climate system model, CSIRO Atmospheric Research, Aspendale.

Gualdi, S., A. Navarra, E. Guilyardi, and P. Delecluse, 2003: Assessment of the Tropical Indo-Pacific Climate in the SINTEX CGCM. *Annals of Geophysics*, **46,** 1-26.

Guilyardi, E. and G. Madec, 1997: Performance of the OPA/ARPEGE-T21 global ocean-atmosphere coupled model. *Clim. Dyn.*, **13,** 149-165.

Hirst, A. C., S. P. O'Farrell, and H. B. Gordon, 2000: Comparison of a coupled ocean-atmosphere model with and without oceanic eddy-induced advection. 1. Ocean spin-up and control integrations. *J. Clim.*, **13,** 139-163.

Johns, T., C. Durman, H. Banks, M. Roberts, A. McLaren, J. Rid-ley, C. Senior, K. Williams, A. Jones, A. Keen, et al., 2004: HadGEM1 model description and analysis of preliminary experiments for the IPCC Fourth Assessment Report, U.K. Met Office, Exeter, U.K.

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

K-1-model-developers, 2004: K-1 coupled model (MIROC) description, Center for Climate System Research, University of Tokyo.

Kiehl, J. T. and P. R. Gent, 2004: The Community Climate System Model, Version 2. *J. Clim.*, **17,** 3666-3682.

Kim, S. J., G. M. Flato, and G. J. Boer, 2003: A coupled climate model simulation of the Last Glacial Maximum, Part 2: Approach to equilibrium. *Clim. Dyn.*, **20,** 635-661.

Kim, S.-J., G. M. Flato, G. J. Boer, and N. A. McFarlane, 2002: A coupled climate model simulation of the Last Glacial Maximum, Part 1: Transient multi-decadal response. *Clim. Dyn.*, **19,** 515-537.

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

Li, T. and T. F. Hogan, 1999: The role of the annual mean climate on seasonal and interannual variability of the tropical Pacific in a coupled GCM. *J. Clim.*, **12,** 780-792.

Lucarini, V. and G. L. Russell, 2002: Comparison of mean climate trends in the Northern Hemisphere between National Centers for Environmental Prediction and two atmosphere-ocean model forced runs. *J. Geophys. Res.*, **107,** 4269.

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

Marti, O., B. P., J. Bellier, R. Benshila, S. Bony, P. Brock-mann, P. Cadulle, A. Caubel, S. Denvil, J. L. Dufresne, et al., 2005: The new IPSL climate system model: IPSL-CM4, IPSL, Paris, Case 101.

Miller, R. L. and X. Jiang, 1996: Surface energy fluxes and coupled variability in the Tropics of a coupled General Circulation Model. *J. Clim.*, **9,** 1599-1620.

Min, S.-K., S. Legutke, A. Hense, and W.-T. Kwon, 2005: Internal variability in a 1000-year control simulation with the coupled climate model ECHO-G. Part I: near surface temperature, precipitation,and mean sea level pressure. *Tellus*, **57A,** 605-621.

Power, S. B., R. A. Colman, B. J. McAvaney, R. R. Dahni, A. M. Moore, and N. R. Smith, 1993: The BMRC Coupled atmosphere/ocean/sea-ice model, Bureau of Meteorology Research Centre, Melbourne, 37, 55 pp.

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

Russell, G. L. and D. Rind, 1999: Response to CO2 transient increase in the GISS coupled model: Regional coolings in a warming climate. *J. Clim.*, **12,** 531-539.

Russell, G. L., J. R. Miller, and D. Rind, 1995: A Coupled Atmosphere-Ocean Model for Transient Climate Change Studies. *Atmosphere-Ocean*, **33,** 683-730.

Salas-Melia, D., F. Chauvin, M. Deque, H. Douville, J. Gueremy, P. Marquet, S. Planton, J. Royer, and S. Tyteca, 2005: Description and validation of the CNRM-CM3 global coupled model. *Clim. Dyn.***,** submitted.

Schmidt, G. A., R. Ruedy, J. E. Hansen, I. Aleinov, N. Bell, M. Bauer, S. Bauer, B. Cairns, V. Canuto, Y. Cheng, et al., 2006: Present-Day Atmospheric Simulations Using GISS ModelE: Comparison to In Situ, Satellite, and Reanalysis Data. *J. Clim.*, **19,** 153-192.

Schneider, E. K. and Z. Zhu, 1998: Sensitivity of the simulated annual cycle of sea surface temperature in the Equatorial Pacific to sunlight penetration. *J. Clim.*, **11,** 1932-1950.

Tokioka, T., A. Noda, A. Kitoh, Y. Nikaidou, S. Nakagawa, T. Motoi, S. Yukimoto, and K. Takata, 1996: A transient CO2 experiment with the MRI CGCM: Annual mean response, Center for Global Environmntal Research, National Institute for Environmental Studies, Environment Agency of Japan, Ibaraki, Japan, CGER-IO22-96, 86 pp.

Volodin, E. M. and N. A. Diansky, 2004: El-Nino reproduction in coupled general circulation model of atmosphere and ocean. *Russian meteorology and hydrology*, **12,** 5-14.

Voss, R., R. Sausen, and U. Cubasch, 1998: Periodically synchronously coupled integrations with the atmosphere-ocean general circulation model ECHAM3/LSG. *Clim. Dyn.*, **14,** 249-266.

Washington, W. M. and G. A. Meehl, 1996: High-latitude climate change in a global coupled ocean-atmosphere-sea ice model with increased atmospheric CO2. *J. Geophys. Res.*, **101,** 12795-12801.

Washington, W. M. and Coauthors, 2000: Parallel climate model (PCM) control and transient simulations. *Clim. Dyn.*, **16,** 755-774.

Yu, Y., X. Zhang, and Y. Guo, 2004: Global coupled ocean-atmosphere general circulation models in LASG/IAP. *Advanced Atmospheric Science*, **21,** 444-455.

Yukimoto, S. and A. Noda, 2002: Improvements of the Meteorological Research Institute Global Ocean-atmosphere Coupled GCM(MRI-CGCM2) and its climate sensitivity, NIES, Japan.

Zhang, X.-H., G.-Y. Shi, H. Liu, and Y.-Q. Yu(eds.), 2000: *IAP Global Atmosphere-Land System Model*. Science Press, 259 pp.