

C. Covey · A. Abe-Ouchi · G. J. Boer · B. A. Boville
 U. Cubasch · L. Fairhead · G. M. Flato · H. Gordon
 E. Guilyardi · X. Jiang · T. C. Johns · H. Le Treut
 G. Madec · G. A. Meehl · R. Miller · A. Noda
 S. B. Power · E. Roeckner · G. Russell · E. K. Schneider
 R. J. Stouffer · L. Terray · J.-S. von Storch

The seasonal cycle in coupled ocean-atmosphere general circulation models

Received: 28 February 1999 / Accepted: 29 February 2000

Abstract We examine the seasonal cycle of near-surface air temperature simulated by 17 coupled ocean-atmosphere general circulation models participating in the Coupled Model Intercomparison Project (CMIP). Nine of the models use ad hoc “flux adjustment” at the ocean surface to bring model simulations close to observations of the present-day climate. We group flux-adjusted and non-flux-adjusted models separately and examine the behavior of each class. When averaged over all of the flux-adjusted model simulations, near-surface air temperature falls within 2 K of observed values over the oceans. The corresponding average over non-flux-adjusted models shows errors up to ~6 K in extensive ocean areas. Flux adjustments are not directly applied over land, and near-surface land temperature errors are

substantial in the average over flux-adjusted models, which systematically underestimates (by ~5 K) temperature in areas of elevated terrain. The corresponding average over non-flux-adjusted models forms a similar error pattern (with somewhat increased amplitude) over land. We use the temperature difference between July and January to measure seasonal cycle amplitude. Zonal means of this quantity from the individual flux-adjusted models form a fairly tight cluster (all within ~30% of the mean) centered on the observed values. The non-flux-adjusted models perform nearly as well at most latitudes. In Southern Ocean mid-latitudes, however, the

C. Covey (✉)
 Program for Climate Model Diagnosis and Intercomparison,
 Mail Code L-264,
 Lawrence Livermore National Laboratory,
 Livermore, CA 94550, USA
 E-mail: coveyl@llnl.gov

A. Abe-Ouchi
 Center for Climate System Research 4-6-1,
 Komaba, Meguro, Tokyo, 153, Japan

G. J. Boer · G. M. Flato
 Canadian Center For Climate Modeling and Analysis,
 Victoria, B.C. V8W 2Y2, Canada

B. A. Boville · G. A. Meehl
 National Center for Atmospheric Research,
 Boulder, CO 80307, USA

U. Cubasch · E. Roeckner
 German Climate Computing Center and Max Planck
 Institute for Meteorology, Bundesstrasse 55,
 20146 Hamburg, Germany

L. Fairhead · H. Le Treut
 Laboratoire de Meteorologie Dynamique du CNRS,
 75252 Paris Cedex 05, France

H. Gordon
 CSIRO Division of Atmospheric Research, Mordialloc,
 Victoria 3195, Australia

E. Guilyardi · L. Terray
 CERFACS Climate Modeling and Global Change Team,
 31057 Toulouse Cedex, France

X. Jiang · R. Miller · G. Russell
 Goddard Institute for Space Studies, 2880 Broadway,
 New York, NY 10025, USA

T. C. Johns
 Hadley Center, UK Meteorological Office,
 London Road, Bracknell, Berkshire RG12 2SY,
 United Kingdom

G. Madec
 Laboratoire Dynamique d’Oceanographie
 et de Climatologie, 75252 Paris Cedex 05, France

A. Noda
 Meteorological Research Institute,
 Nagamine 1-1, Tsukuba, Ibaraki 305, Japan

S. B. Power
 Bureau of Meteorology Research Center,
 Melbourne, Victoria 3001, Australia

E. K. Schneider
 Center for Ocean-Land-Atmosphere Studies,
 4041 Powder Mill Road, Calverton, MD 20705, USA

R. J. Stouffer
 Geophysical Fluid Dynamics Laboratory,
 Princeton, NJ 08542, USA

J.-S. von Storch
 Institute of Meteorology, University of Hamburg,
 Bundesstrasse 55, 20146 Hamburg, Germany

non-flux-adjusted models overestimate the magnitude of January-minus-July temperature differences by ~ 5 K due to an overestimate of summer (January) near-surface temperature. This error is common to five of the eight non-flux-adjusted models. Also, over Northern Hemisphere mid-latitude land areas, zonal mean differences between July and January temperatures simulated by the non-flux-adjusted models show a greater spread (positive and negative) about observed values than results from the flux-adjusted models. Elsewhere, differences between the two classes of models are less obvious. At no latitude is the zonal mean difference between averages over the two classes of models greater than the standard deviation over models. The ability of coupled GCMs to simulate a reasonable seasonal cycle is a necessary condition for confidence in their prediction of long-term climatic changes (such as global warming), but it is not a sufficient condition unless the seasonal cycle and long-term changes involve similar climatic processes. To test this possible connection, we compare seasonal cycle amplitude with equilibrium warming under doubled atmospheric carbon dioxide for the models in our data base. A small but positive correlation exists between these two quantities. This result is predicted by a simple conceptual model of the climate system, and it is consistent with other modeling experience, which indicates that the seasonal cycle depends only weakly on climate sensitivity.

1 Introduction

The simulation of the seasonal cycle of near-surface air temperature (SAT) has long been considered a test of climate model performance. For example, Schneider and Londer (1984) point out that “every year the amount of incoming solar energy varies by a large fraction from winter to summer... (Climate system) feedback factors can be approximated, at least implicitly, by simply answering the following question: how many degrees of surface-temperature change accompanies how many watts-per-square-meter energy change as the sun moves its position in the sky from season to season?... By using nature’s ‘seasonal experiment,’ we can calibrate our climatic model’s internal response to external forcings.” Schneider and Londer (1984) cite work as early as 1971 using the seasonal cycle to verify the climate sensitivity of models (Lindzen et al. 1995 review the same subject from a different perspective). Seasonal changes are rapid enough that, as noted by Schneider and Londer (1984), “deep oceans and other slowly varying climatic subsystems do not participate significantly in the seasonal experiment, so a model that faithfully reproduces the seasonal climatic shifts is not verified for longer-term changes like CO_2 increases over decades.” Nevertheless, climate models including only the upper layers of the ocean have been remarkably successful in simulating equilibrium climate states (e.g., Manabe and Stouffer

1980; Washington and Meehl 1984). To first approximation these states may be independent of the deep ocean, in which case seasonal cycle amplitude and equilibrium climate sensitivity may be positively correlated.

In this study we assess both the accuracy of the climatological seasonal cycle and its connection with equilibrium climate sensitivity in 17 modern coupled ocean-atmosphere general circulation models. These models include the full ocean and atmosphere, as well as sea ice, and require only boundary conditions external to the physical climate system such as atmospheric CO_2 concentration and the amount of solar energy incident at the top of the atmosphere. We use model output compiled by the Coupled Model Inter-comparison Project (CMIP). Under the auspices of the World Climate Research Program’s Working Group on Coupled Models, CMIP has collected SAT and other output fields produced by 19 coupled GCMs during the later 1990s (Meehl et al. 1997). All output fields are available to “diagnostic subproject” investigators through the WGCM CMIP Panel (to propose a subproject see the CMIP Web pages at <http://www-pcmdi.llnl.gov/cmip>). We examined monthly mean SATs from 17 models in CMIP Phase 1 (Table 1). Two other CMIP1 models, not included in this study, directly constrain ocean surface salinity or sea ice to conform to observations.

External forcing is held constant in the model runs considered here, simulating the present-day (or, in some cases, the pre-industrial) global climate. This study compares the model-simulated seasonal cycle of SAT with observed data. Analysis of interannual and longer time scale variability in CMIP models is published separately (e.g., Barnett 1999; Bell et al. 2000; Stouffer et al. 2000).

Traditionally, coupled GCMs modify the exchange of heat, water and sometimes momentum at the ocean-atmosphere interface by “flux adjustment” (occasionally called “flux correction”). A mismatch between the surface fluxes calculated by the atmospheric sub-model, and the fluxes the ocean sub-model requires to maintain reasonable conditions, can cause the simulated climate to drift into unrealistic regimes. Flux adjustments correct the mismatch by adding ad hoc boundary terms to the partial differential equations (conservation laws for energy, mass and momentum) that describe the ocean and atmosphere. All but one of the CMIP models that employ flux adjustments use annually periodic adjustments; the exception is ECHAM4 + OPYC3, which uses seasonally constant adjustments. One might argue that for models using annually periodic flux adjustments, errors in the seasonal cycle of near-surface temperature are a statement of errors in the computation of flux adjustments and are not relevant to model validation. Furthermore, the possibility that flux adjustments distort the results of climate change experiments is a subject of vigorous debate (e.g., Sausen et al. 1988; Kerr 1994; Gregory and Mitchell 1997).

Of the 17 models considered here, however, eight are not flux-adjusted (Table 1). Thus the CMIP data base reveals that about half of modern coupled ocean-atmo-

sphere GCMs refrain from using flux adjustment. Several of these non-flux-adjusted models have run with useable results for ~ 100 simulated years or more. (In addition, two of the CMIP1 flux-adjusted models, GFDL and UKMO, have recently been run in versions without flux adjustment for >100 simulated years.) The presence of several non-flux-adjusted models in the CMIP data base allows us to quantify their seasonal cycle performance by asking how closely they agree with observations compared with the flux-adjusted models. In what follows we group flux-adjusted and non-flux-adjusted models separately and examine the behavior of each class.

Table 1 Model characteristics

Model ^a	Flux adjusted?	Run length [yr]	A [K] ^b	ΔT_{2x} [K] ^b
BMRC	No	105	9.03	2.1 ^c
CCCMA	Yes	150	9.16	3.5
CCSR	Yes	40	10.42	3.5
CERFACS	No	40	8.19	
COLA	No	50	8.78	2.4 ^c
CSIRO	Yes	100	9.19	4.3
ECHAM1 + LSG	Yes	960	8.66	2.6
ECHAM3 + LSG	Yes	1000	8.07	2.5 ^c
ECHAM4 + OPYC3	Yes	240	8.71	3.6 ^c
GFDL	Yes	1000	11.69	3.7
GISS (Miller)	No	89	9.78	3.6 ^c
GISS (Russell)	No	98	8.53	3.6 ^c
LMD + IPSL	No	24	7.87	3.4 ^c
MRI	Yes	100	9.79	4.8
NCAR CSM	No	300	9.95	2.1
NCAR (Washington & Meehl)	No	100	12.12	4.6
UKMO HadCM2	Yes	1085	8.81	4.1
Observed (Legates and Willmott)			9.46	

^a For identification and detailed model documentation, including model definitions of SAT, see the CMIP Web page <http://www.pcmdi.llnl.gov/modeldoc/cmip>

^b See text for definition

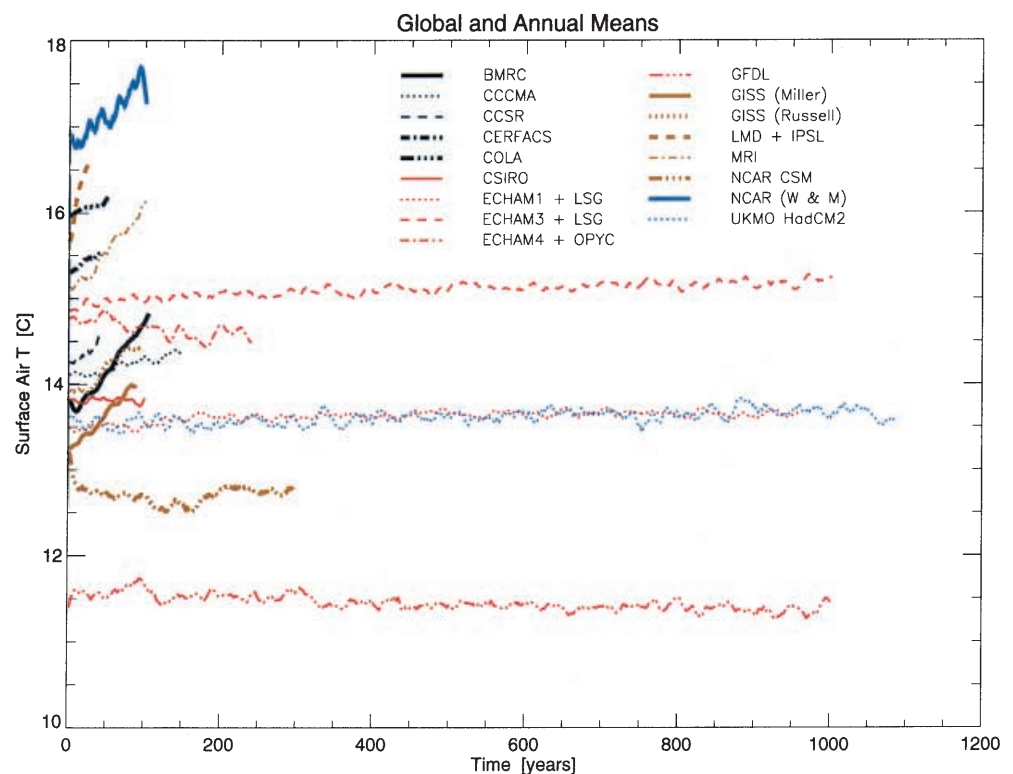
^c Similar but not identical atmospheric model versions used to measure A and ΔT_{2x}

2 Methods of analysis

Time periods used for averaging the model results to form a seasonal cycle climatology include the full range available from the CMIP1 data base, which ranges from 24 to 1085 years (Table 1). The only exception to this rule is the ECHAM3 + LSG model, from which we used just the first 400 of the total 1000 years. (The full 1000 years of data were not available when this study began.) As noted below, a sampling of results from the full 1000 years of ECHAM3 + LSG indicates that the time period we use makes little difference.

Our procedures make no correction for secular “climate drift,” which is substantial for some of the models. Figure 1 shows the globally and annually averaged SAT for each model as a function of time. Initial drifts up to several degrees K per century are apparent, mainly occurring in non-flux-adjusted models (thick lines). Averaging the magnitude of linear trend over the flux-adjusted and the non-flux-adjusted models gives 0.24 and 1.1 K/century respectively. The separation between the two populations is statistically significant (probability of such a large separation arising from sampling error <0.06 from a one-sided t -test). The figure also

Fig. 1 Time series of globally and annually averaged SAT from the 17 models used in this study. Curves are smoothed with an 11-point moving average. Flux-adjusted-model results are shown as *thin lines*; non-flux-adjusted model results are shown as *thick lines*. Individual models are distinguished from one another by varying line styles (*solid, dashed, etc.*) and colors as described in the figure



reveals a surprisingly large spread (>5 K) in the absolute value of globally and annually averaged SAT. This spread is due in part to differing definitions of SAT— e.g., four of the models simply take the temperature of the lowest model layer as SAT, while another 13 make an extrapolation from the lowest layer temperature, but other factors are responsible as well. For example, GFDL is too cold both because of the use of the lowest model level as a surrogate for SAT and because its boundary layer is too diffusive (global mean *sea surface* temperature is accurate but *near-surface air* temperature is too cold).

The present study involves seasonal cycle amplitude rather than the absolute value of annual mean SAT. The global mean seasonal cycle amplitude (defined later) changes by $<10\%$ in all models when the climatological average over all available years is replaced by results from just the first year. Global mean seasonal cycle amplitude changes by $<1\%$ when the number of years in the ECHAM3 + LSG model average increases from 400 to 1000. These results indicate that climate drifts do not create large problems for this study.

The difference between the climatological mean January and climatological mean July temperatures is a good approximation to seasonal cycle amplitude. Fourier analysis indicates that this approximation is quite accurate for both observed and model-simulated SAT (see Appendix 1). While a cyclical phenomenon is characterized by both amplitude and phase, outside the tropics the seasonal cycle is characterized by a rather uniform phase. The times of year in which Northern Hemisphere extratropical temperatures are warmest and coolest are July and January respectively, with the opposite effect applying to the Southern Hemisphere. We therefore focus our attention on amplitude rather than phase and measure amplitude by simply taking the difference between July and January values. In the tropics, the amplitude as defined is small, and a significant semiannual cycle of near-surface temperature is present. Accordingly, our method of analysis emphasizes the extratropics.

We use three observed-data sets to provide a sense of observational uncertainties in the annual cycle of SAT. We obtained the data of Legates and Wilmott (1990a, b) and the data of Shea (1986) from the NCAR Data Support Section (see Shea 1996). The third data set consists of several merged SAT observations (Gates et al. 1999, Appendix C). It uses data from Jones (1988), augmented by land surface temperatures from Schubert et al. (1992) and by observed (AMIP) sea surface temperatures (corrected to SAT with observed surface-air differences from da Silva et al. 1994). The Legates and Wilmott (1990a, b) data covers the period 1920–80, the Shea (1996) data 1950–79, and the augmented Jones (1988) data 1979–88. As shown later, these three observed data sets give nearly identical results when used to measure the seasonal cycle of SAT. Model-processed “reanalysis” of meteorological observations (e.g., Kalnay et al. 1996) give results that differ noticeably (though not drastically) from the observations used here. Because the reanalysis procedure infers near-surface temperatures indirectly, using model-dependent parametrizations of processes such as the hydrologic cycle, we assume the direct measurements are more reliable and include only them in our study.

3 Comparison of flux-adjusted and non-flux-adjusted models

January, July and July-minus-January climatological temperatures are shown for each individual model and observed data set on the CMIP Web site at <http://www-pcmdi.llnl.gov/cmip/scc.html>. Inspection of these 60 latitude-longitude maps shows that all models successfully capture the observed qualitative behavior of SAT. Here we summarize the results by first comparing the average over flux-adjusted models with the average over non-flux-adjusted models and with the observations,

then examining the zonally averaged seasonal cycle amplitude for the individual models and observations.

Figure 2 shows January and July SAT values averaged over both flux-adjusted and non-flux-adjusted models. In both seasons, the two averages over models generally agree with each other to within ~ 5 K (the color-scale interval in the graph). To compare with observed values, and to illustrate the detailed differences between the flux-adjusted and non-flux-adjusted models, Fig. 3 shows the difference between the fields shown in Fig. 2 and the observations of Jones. (We chose the Jones 1988 data for this figure because it seemed a reasonable compromise between the lack of coverage in Shea’s 1996 data, which excludes all of the middle and higher latitudes of the Southern Hemisphere, and the inclusion in Legates’ data of areas such as Antarctica in great detail despite sparse local observations. Within areas of common coverage, however, the Shea and Legates data sets lead to the same conclusions as that of Jones.) On average, both flux-adjusted and non-flux-adjusted models tend to underestimate SAT over land by several degrees K. This problem occurs especially at high surface elevation, where the model-derived SAT may not accurately reflect the sub-gridscale average over mountain areas. Since this error appears in both January and July, it tends to cancel when we take the difference between SATs for the two months. In ocean areas the flux-adjusted models are quite accurate, as expected, since flux adjustment techniques are applied only to ocean grid points of coupled models. The SAT averaged over flux-adjusted models agrees with observations to within ± 2 K at all ocean grid points. The average over non-flux-adjusted models, in contrast, exhibits SAT errors up to ~ 6 K over large ocean areas in both hemispheres and seasons, especially in the Southern Ocean.

Figures 4 and 5 show the difference between July and January SAT values for the two classes of models and for the Jones observations. Figure 4 (like Fig. 2) displays the gross situation with a contour interval of 5 K for the average over flux-adjusted models, the average over non-flux-adjusted models, and the observations. This figure shows a general qualitative agreement among the three data sets. The only obvious exception is the Southern Ocean at latitudes 50 – 60°S , where the average over models lacking flux adjustment (middle graph) overestimates the seasonal cycle amplitude by ~ 5 K. From Fig. 3 it appears that this error originates from January (austral summer) SAT being too warm in the non-flux-adjusted models at 50 – 60°S . Figure 5 confirms this error and also shows that the non-flux-adjusted models underestimate by >2 K the seasonal cycle amplitude in the latitude range 30 – 40°N in ocean subtropical gyre regions associated with the Gulf Stream and Kuroshio currents. (The green color code is the same as for the *overestimate* of seasonal cycle amplitude in the Southern Ocean, since the sign of the July-minus-January SAT difference changes between hemispheres.)

Surface Air Temperature [C]

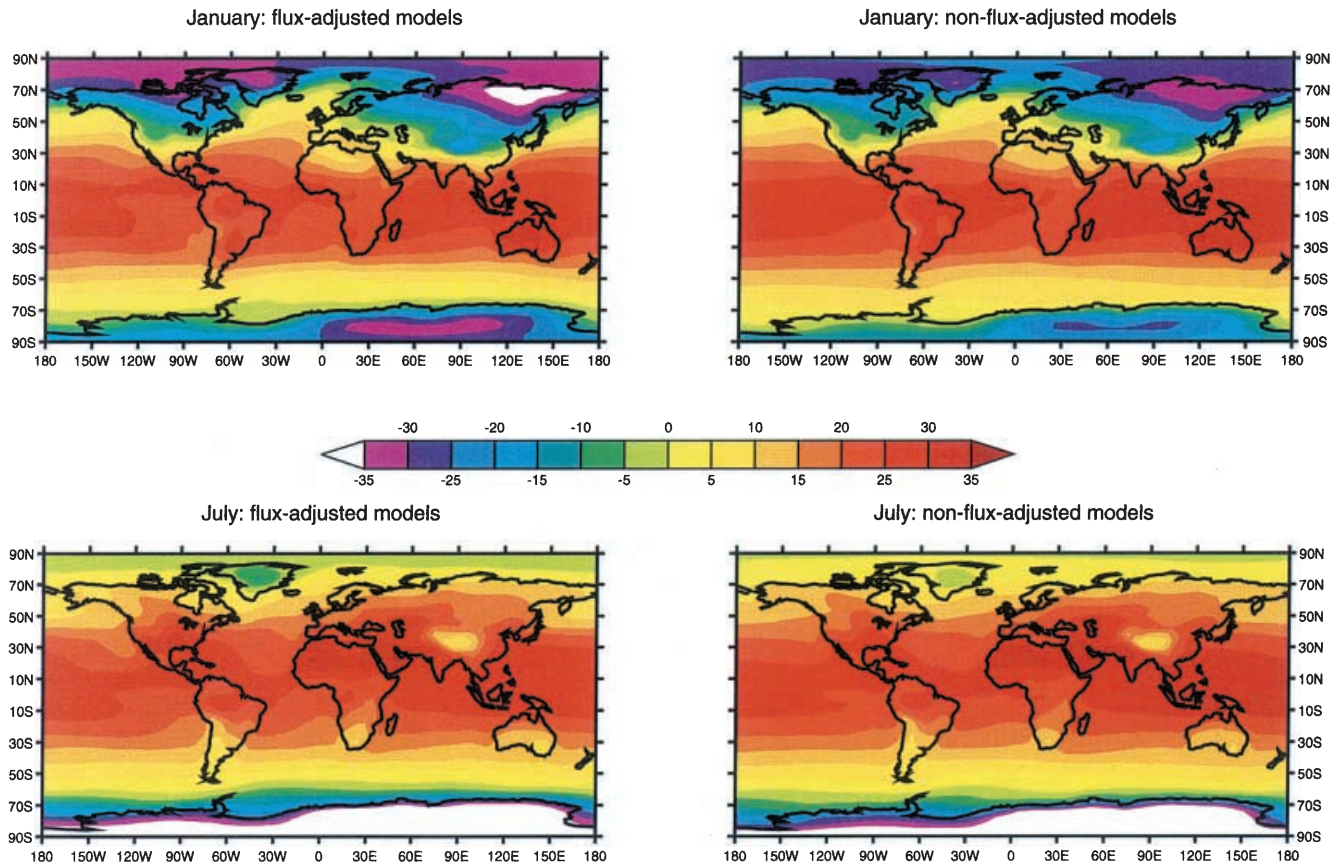


Fig. 2 Simulated January and July climatological mean SAT, averaged separately over flux-adjusted (*left*) and non-flux-adjusted (*right*) models

Over land, Fig. 5 shows substantial errors in July-minus-January SAT despite the partial cancellation of January and July SAT underestimates mentioned earlier. Averages over both flux-adjusted and non-flux-adjusted models show extensive land areas with errors > 2 K. On average, the flux-adjusted models perform only marginally better than the non-flux-adjusted models over land, as expected since flux adjustments are directly applied only to ocean areas.

To make a quantitative comparison of individual models and observations, Fig. 6 shows the zonal mean July-minus-January SAT for each of the 17 models and three observational data sets. These longitude averages are presented separately for land areas and ocean areas. The model results cluster fairly near the observations except in polar regions (where observations become uncertain, differences in model definitions of SAT are most problematic, and zonal means cover much smaller areas). Between about 60°S and 60°N , the differences among the models are greatest in northern mid-latitude land areas (Fig. 6a). Here the zonal mean July-minus-January SAT difference simulated by models is up to ~ 10 K above or below the observed values, or about $\pm 30\%$ maximum error relative to an average over

models of ~ 30 K SAT difference. In general the flux-adjusted models (thin lines) have smaller errors than the non-flux-adjusted models (thick lines) in northern mid-latitude land areas, even though flux adjustment is not applied directly over land. (Of course wind transport couples land with ocean areas to some extent.) Land areas between about 60°S and 30°N show considerably smaller errors with no obvious difference between flux-adjusted and non-flux-adjusted models. In Northern Hemisphere tropical land areas, nearly all models slightly overestimate seasonal cycle amplitude, consistent with Fig. 5.

In ocean areas (Fig. 6b) the model errors are smaller in absolute terms but about the same in relative terms, i.e., there is roughly a $\pm 30\%$ maximum spread of model-simulated July-minus-January SAT about the observed values. Here again the flux-adjusted models generally exhibit smaller errors than non-flux-adjusted models. This behavior is most apparent for ocean areas at $\sim 50^{\circ}\text{S}$, where (as discussed above) the non-flux-adjusted models tend to overestimate the amplitude of the seasonal cycle. At this latitude, Fig. 6b shows that four of the five models showing the greatest overestimate are not flux-adjusted (thick lines). Since the problem is common to

Surface Air Temperature: (Models - Observations) [C]

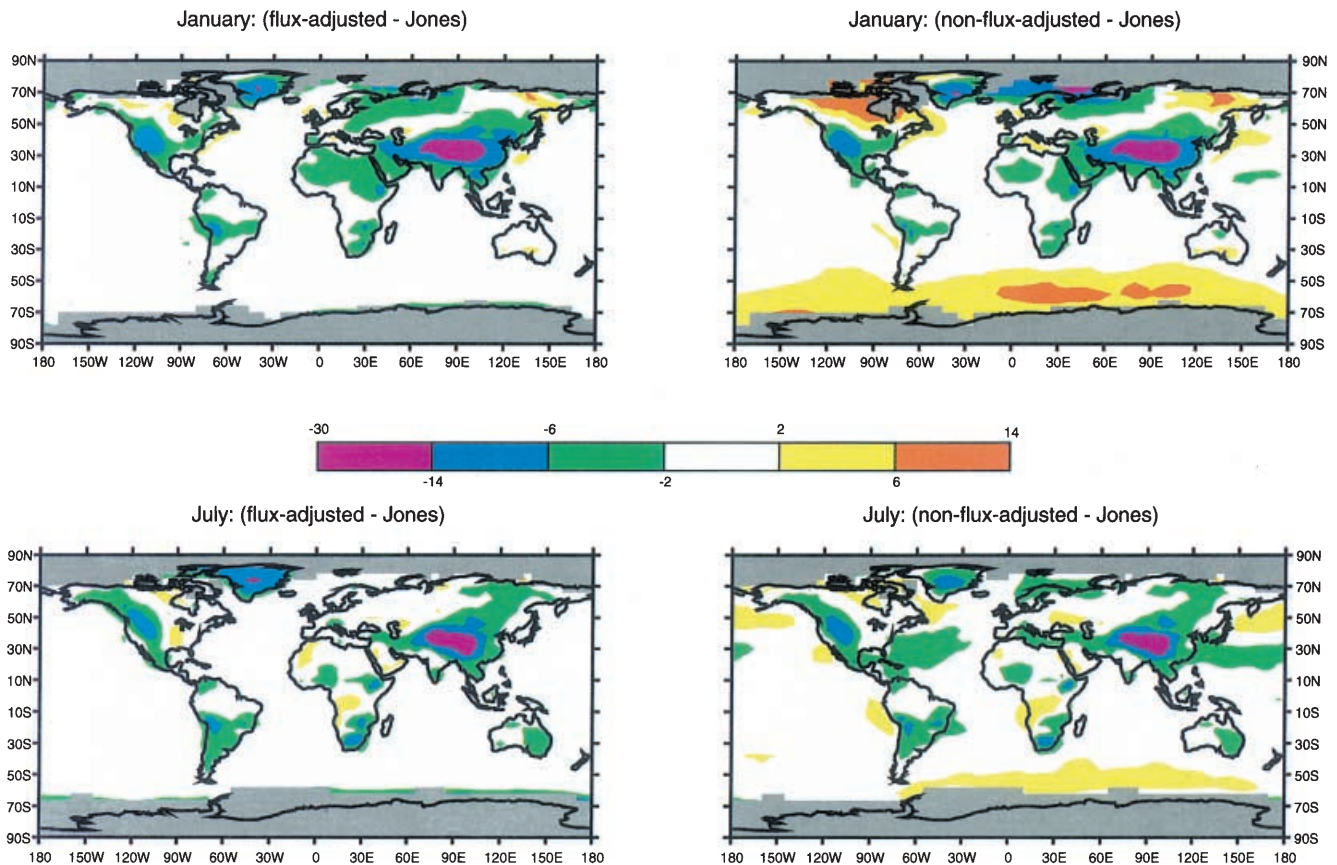


Fig. 3 Differences between the model-simulated SAT values shown in Fig. 2 and the corresponding observed values from the augmented Jones dataset. Note the nonlinear color scale in which error ranges

double with each color interval. Contour lines are ± 2 , ± 6 , ± 14 , ...; intervals between zero (not shown) and successive contour lines are 2, 4, 8, ...). Gray color indicates missing data

many of the non-flux-adjusted models, the error in the average over such models cannot be dismissed as an artifact of one or two egregiously “bad” cases.

To further quantify the difference between flux-adjusted and non-flux-adjusted models, we took the average and the standard deviation over models from Fig. 6a, b, for each class of model separately. The result (not shown) is that the average over flux-adjusted models and the average over non-flux-adjusted models are quite close to each other. Indeed, the difference between these two averages is always less than the standard deviation over either class of model. This result implies that there is little significant difference in zonal mean July-minus-January SAT between flux-adjusted models and non-flux-adjusted models. Statistically significant differences between the two classes of models appear only for ocean latitudes near 50°S (as discussed) and in the tropical oceans. In the tropical oceans, the flux-adjusted models show significantly smaller variance of results (i.e., inter-model scatter) than the non-flux-adjusted models. Elsewhere over ocean, and everywhere over land, the probability P that differences between the two classes of models arise by chance is > 0.1 , indicating

no significant difference, at nearly all latitudes ($P < 0.1$ at 10% or less of the latitudes, as would be expected by chance alone).

In summary, comparison of July-minus-January SAT between flux-adjusted and non-flux-adjusted models shows that the former group agrees better with observations, as expected. Errors distinguishing the non-flux-adjusted models include a systematic overestimate of seasonal cycle amplitude in the Southern Ocean and a greater range of disagreement with observations (compared with the flux-adjusted models) in Northern Hemisphere mid-latitude land areas. These errors are generally not extreme enough to make the non-flux-adjusted models distinguishable from flux-adjusted models, or from observations, after a zonal mean over land or ocean is taken and the two classes of models are treated as two statistical populations.

4 Implications for model evaluation

Is the extent of agreement between models and observations good enough to increase confidence in the

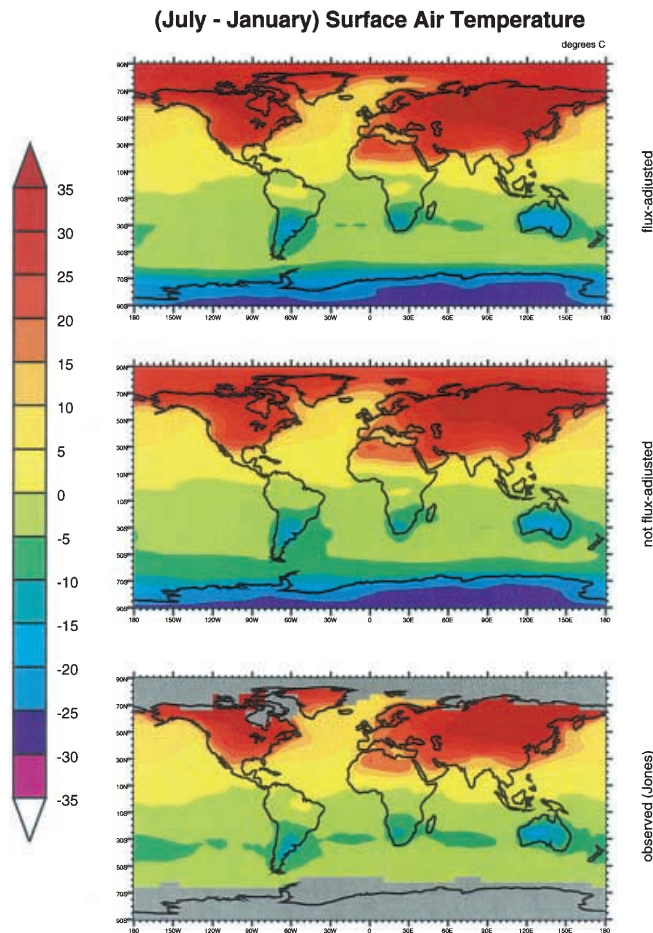


Fig. 4 Simulated difference between July and January climatological mean SAT averaged separately over flux-adjusted models (*top*) and non-flux-adjusted models (*middle*), and the corresponding observed values from the augmented Jones dataset (*bottom*)

models' climate-change simulations? The answer to this question obviously depends on the particular use made of a model. Surface temperature errors up to a few degrees K over large ocean areas could easily compromise a study of El Niño but might be acceptable in an assessment of global mean climate sensitivity. Climate sensitivity is usually measured by the equilibrium SAT response to a change in forcing (typically doubled atmospheric CO₂). With this application in mind, we consider the correlation between seasonal cycle amplitude and climate sensitivity. As discussed in the Introduction, one might expect the correlation to be positive. A more detailed analysis confirms this expectation but also implies that the correlation is weak (see Appendix 2).

To obtain a single number for seasonal cycle amplitude, we average the *peak-to-peak* amplitudes of the two hemispheres, i.e.,

$$A \equiv (\langle T_{\text{July}} - T_{\text{January}} \rangle_{\text{NH}} + \langle T_{\text{January}} - T_{\text{July}} \rangle_{\text{SH}}) / 2,$$

where the angle-brackets indicate an area average over the Northern Hemisphere (NH) or the Southern Hemisphere (SH). A is approximately the globally averaged absolute value of the quantity displayed in Fig. 6 (i.e.,

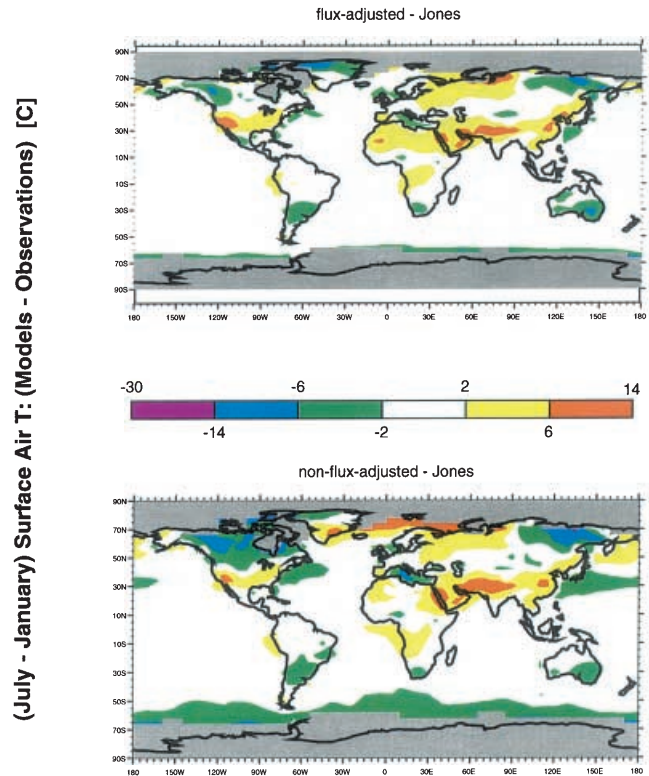
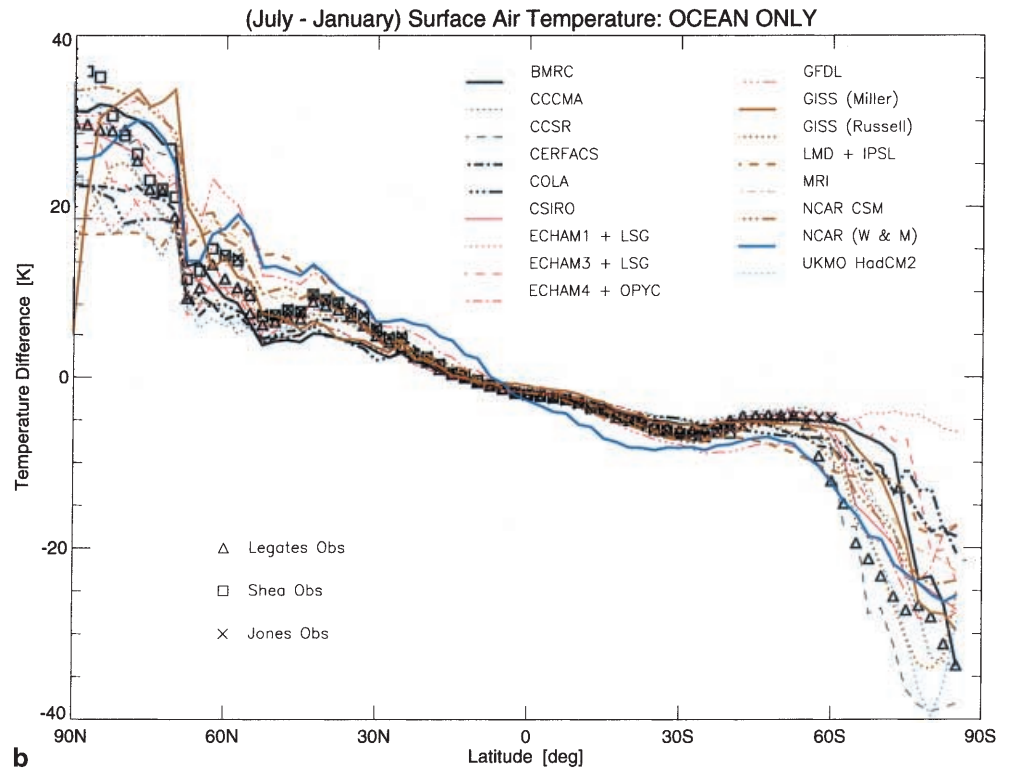
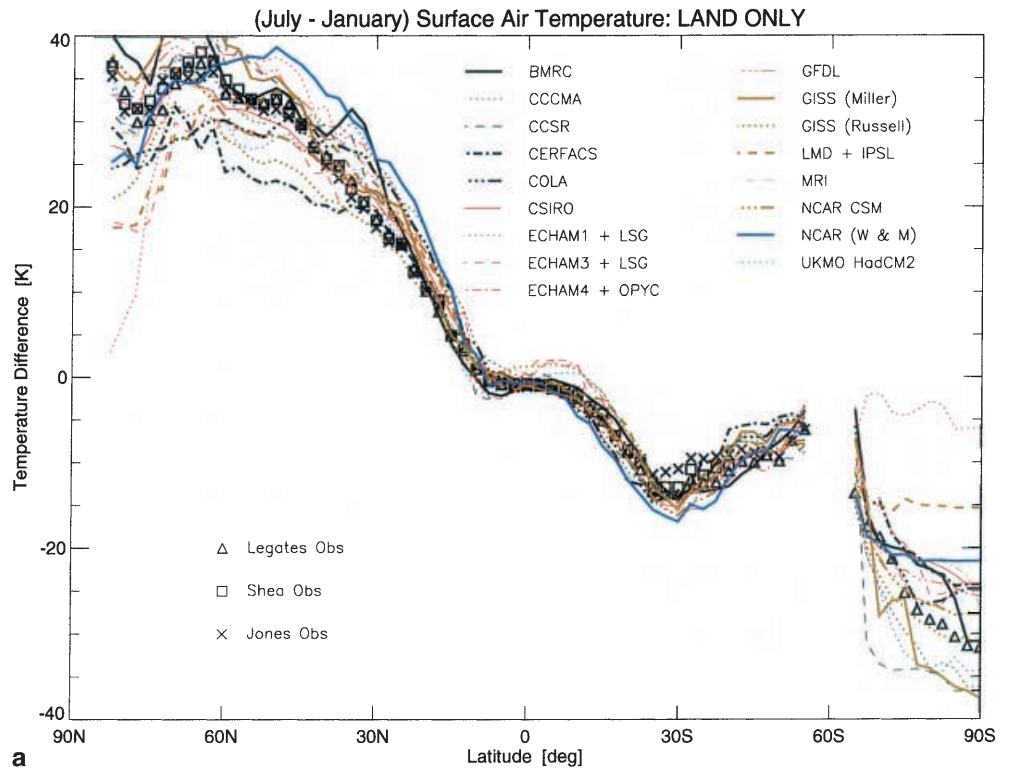


Fig. 5 Differences between the model-simulated and the observed values shown in Fig. 4, for flux-adjusted models (*top*) and non-flux-adjusted models (*bottom*). Note the nonlinear color scale as in Fig. 3

$A \approx \langle |T_{\text{July}} - T_{\text{January}}| \rangle_{\text{Global}}$). It also is approximately equal to the globally averaged peak-to-peak Fourier amplitude of the annual harmonic of SAT (see Appendix 1). The observed value of A from Legates and Willmott (1990a, b) is 9.46 K. (Legates and Willmott provide a globally complete data set. Our other two observational sets are missing data near the poles, which complicates the calculation of globally integrated quantities like A . The value from Legates and Willmott is consistent with $A \approx 9.3$ K, which may be obtained from Oort's 1983 hemispheric-averaged tables, keeping in mind that A is measured from peak to peak.) Values of A for each of the models in this study are given in Table 1. These vary by up to $\pm 30\%$ from the observed A , but the averages over flux-adjusted and non-flux adjusted models (9.48 K and 9.42 K respectively) are within 10% of each other and the observed value.

As a measure of climate sensitivity we take ΔT_{2x} , the equilibrium change in globally averaged SAT when the atmospheric concentration of CO₂ is doubled. ΔT_{2x} may be assessed by running a coupled ocean-atmosphere model to a statistical equilibrium state with doubled atmospheric CO₂. This procedure, however, consumes a great deal of computer time because the length of the run must exceed the adjustment time scale of the deep ocean, several hundred years. ΔT_{2x} is more typically assessed by running the atmosphere and sea-ice

Fig. 6a, b Zonally averaged difference between July and January climatological mean SAT from the 17 models and the 3 observational datasets used in this study, averaged separately over **a** land and **b** ocean areas. The line legend is the same as in Fig. 1, with flux-adjusted and non-flux-adjusted models distinguished by *thin* and *thick lines* respectively



components of a climate model together with a simple representation of the upper “mixed layer” of the ocean. The two procedures can give different results because the deep ocean includes feedback processes that can affect climate sensitivity (Stouffer and Manabe 1999). Here we

consider only ΔT_{2x} values assessed by an atmosphere/sea ice/mixed layer experiment. By thus omitting slow feedback processes in the deep ocean (which do not affect A) from assessment of ΔT_{2x} , we will presumably maximize the correlation between ΔT_{2x} and A .

An additional complication regarding ΔT_{2x} arises from the rapid pace of climate model development. In many cases the atmosphere and sea ice components of the models used to assess ΔT_{2x} are not identical to the corresponding parts of the coupled models in the CMIP1 data base (and used to assess A). These cases are indicated by the superscript c in Table 1. For one model (CERFACS) the difference seemed large enough to preclude meaningful comparison of ΔT_{2x} and A , in the judgment of the model's developers. ΔT_{2x} is omitted in Table 1 for this model.

ΔT_{2x} and A values from Table 1 are plotted in Fig. 7. Their correlation is weak but positive: $r = 0.4$. The probability of the null hypothesis (that such a correlation would arise from a chance sample of random pairs) is 0.13. Thus the association of ΔT_{2x} and A is statistically significant at nearly the 90% confidence level. Linear fits of ΔT_{2x} to A have positive slopes, both for the full set of models (solid line) and for the subset of models in which identical atmosphere and sea ice components were used to assess ΔT_{2x} and A (dashed line). Reversing the roles of ΔT_{2x} and A , so that A becomes the dependent variable, moves the lines considerably, but their slopes remain positive. These results indicate that the positive sign of the correlation is robust.

We conclude that equilibrium climate sensitivity is one of many factors influencing the amplitude of the seasonal cycle, accounting for perhaps $r^2 \approx 15\%$ of its variance among modern climate models.

5 Conclusions

Coupled ocean-atmosphere models produce reasonable seasonal cycles, and this success has been used to sup-

port their application to climate change simulations (Schneider and Londer 1984). One counter argument is that the models require arbitrary adjustments to simulate the seasonal cycle and other aspects of the present-day or pre-industrial climate (e.g., Singer 1997). This point of view may be unfair to flux-adjusted models, since they do not directly constrain surface temperature over land areas. But in any case the oft-heard claim (e.g., Nakamura et al. 1994) that coupled ocean-atmosphere models generally fail to simulate the current climate without ad hoc surface flux adjustments is now out of date. About half of modern coupled models do not use flux adjustments. For these models the rate of global mean surface temperature drift is greater than that of flux-adjusted models (Fig. 1), but in many cases it is small enough to make century time scale integrations useful, e.g., for assessing the effects of anthropogenic greenhouse gases (see Table 6.3 in Kattenberg et al. 1996). The zonal mean seasonal cycle amplitudes in non-flux-adjusted models agree with observations nearly as well as those of the flux-adjusted models, although local values exhibit greater errors (Sect. 3). The seasonal cycle phases are essentially indistinguishable from observations for both types of model (Appendix 1).

Taken together with more general studies of model output (Gates et al. 1996; Lambert and Boer submitted 2000), these results provide additional confidence in the reliability of climate models. The hard question is: how much more confidence? In this study we find a small but non-negligible positive correlation between the seasonal cycle amplitude and the equilibrium climate sensitivity of the models. The correlation arises because feedbacks involving fast-responding components of the climate system, the atmosphere, the ocean mixed layer and parts of the cryosphere, operate in response to seasonal cycle forcing as well as longer-term climate forcing. Our result is consistent with a simple conceptual model of the climate system (see Appendix 2). It is also consistent with our fairly limited experience correlating seasonal cycle amplitude with climate sensitivity in different versions of individual models. In particular, removal of water vapor feedback in the COLA model (Schneider et al. 1999) reduces equilibrium climate sensitivity by about a factor of 2 and produces a slight (<0.5 K) reduction of seasonal cycle amplitude at high latitudes (but no apparent change at lower latitudes).

The small magnitude of the correlation between the seasonal cycle and equilibrium climate sensitivity indicates that significant processes are not equally important for both phenomena. For example, the seasonal cycle is too fast to allow the ocean thermocline to respond, and water vapor and cloud processes are driven in large part by sea surface temperature changes, so these "atmospheric" feedbacks may differ in seasonal and long-term responses. Also, the climate responds to seasonal forcing to a large extent through storage and dynamical redistribution of heat in the upper mixed layer of the ocean (never reaching equilibrium). Global- and annual-mean climate, on the other hand, may not be sensitive to either

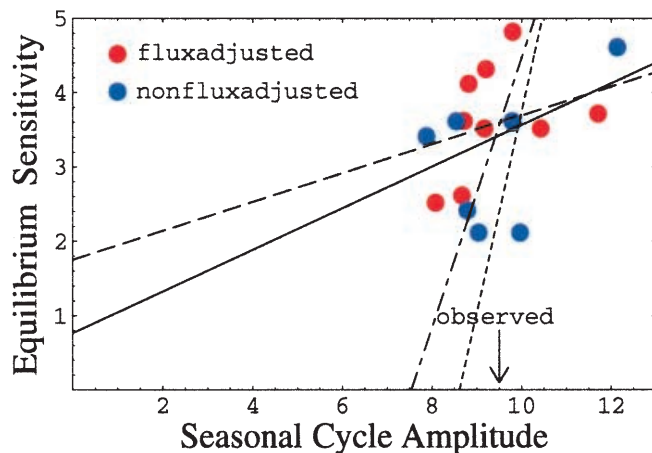


Fig. 7 Seasonal cycle amplitude (A) and equilibrium global mean SAT warming due to doubled atmospheric CO_2 (ΔT_{2x}) for the models. Red points represent flux-adjusted models and blue points represent non-flux-adjusted models. Least-squares linear fits are shown both for ΔT_{2x} as a function of A (solid line: all models; dashed line: only models for which identical versions were used to assess A and ΔT_{2x}) and for A as a function of T_{2x} (dash-dot line: all models; dotted line: only models for which identical versions were used to assess A and ΔT_{2x}). An arrow indicates the observed value of A .

of these processes, at least in the long term. In short, our results imply that an accurate seasonal cycle is important, but not by itself a sufficient indicator that a model contains the physical and dynamical processes appropriate for dependable global change simulations.

Acknowledgements We are grateful to Mike Fiorino for providing observed data and to the PCMDI Computations Staff, especially Susan Marlais and Dean Williams for visualization assistance and Anna McCravy for Web publication. We also thank Martin I. Hoffert and Stephen H. Schneider for extensive useful comments on the manuscript. This work was sponsored in part by the U.S. National Science Foundation (including grants ATM-9520579 and ATM-9907915 to EKS) and performed under auspices of the U.S. Department of Energy by University of California Lawrence Livermore National Laboratory under contract W-7405-Eng-48.

Appendix 1

(July–January) difference as an approximation to seasonal cycle amplitude

We demonstrate here that for both observations and model results, the difference between January and July climatological temperatures is an excellent approximation to the amplitude of the full seasonal cycle. First, we show results for the Jones observations described in the main text. (Nearly identical results are obtained from the Legates and Shea data sets.) Figure A1 displays both the amplitude and phase of the seasonal cycle determined from Fourier analysis. For each point, the length of the arrow gives the amplitude and the direction of the arrow gives the phase. An upward-pointing arrow indicates maximum SAT on January 1, a right-pointing arrow indicates maximum SAT on April 1, etc. Wherever seasonal cycle amplitudes are large, SAT reaches a maximum near the month of July for the Northern Hemisphere and near the month of January for the Southern Hemisphere. These phases suggest that simply taking the difference between July and January values would provide a reasonable approximation to the seasonal cycle amplitude.

To confirm this conclusion, we show in Fig. A2 both the amplitude of the seasonal cycle Fourier harmonic and the climatological (July–January) difference. Of course, the former quantity is positive everywhere while the latter carries a sign to first approximation positive in the Northern Hemisphere and negative in the Southern Hemisphere, as shown in Fig. 4 in the main text. We therefore take the absolute value of (July–January) SAT and divide by 2 to make this quantity directly comparable to the Fourier amplitude of SAT. The agreement between the two quantities is remarkable. The patterns are virtually identical, and the global mean of $|\text{July–January}|/2$ is more than 90% of the global mean of the Fourier amplitude. The global mean of $|\text{July–January}|$, in turn, is quite close to the peak-to-peak seasonal cycle amplitude A defined in the text above. For example, the two quantities differ by $<2\%$ for the average over either flux-adjusted or non-flux-adjusted models.

There remains the question of whether the (January–July) difference is a good measure of model-simulated as well as the observed seasonal cycle. As an indication of the models' behavior, Fig. A3 shows histograms of the annual mean harmonic components of *globally* averaged, model-simulated monthly SAT. Taking the global average of monthly mean SAT produces a time series with a strong residual seasonal cycle dominated by Northern Hemisphere land temperatures. (The Northern Hemisphere's seasonal cycle is roughly twice as strong as the Southern Hemisphere's, and the two are 180° out of phase. If the two hemispheres had equally strong seasonal cycles, they would cancel and there would be no seasonal cycle in the global mean.)

After taking global means, Fourier transformation gives a pair of numbers, amplitude and phase, for each model. These are used

to construct the histograms in Fig. A3. Although the amplitudes show considerable variation among models, there are essentially no differences in the phases of the first harmonic. The average over models is 7.14 and the standard deviation is 0.17 on a scale for which 7 indicates maximum global mean SAT in mid-July and 8 indicates maximum global mean SAT in mid-August. The observed value from Legates and Willmott (1990a, b) is 7.175. (As in our calculation of A , we chose Legates and Willmott's data because it has globally complete coverage.) These results support our decision to focus on the amplitude of the seasonal cycle of SAT and to represent it simply by the (July–January) difference.

Appendix 2

Expected relationship between seasonal cycle amplitude and climate sensitivity

We may investigate this subject via the following analysis by one of us (GJB) and M.I. Hoffert (personal communication). The deviation of hemispheric-average energy balance from its annual mean, neglecting inter-hemispheric exchange, is given by

$$C \frac{dT}{dt} = -\alpha T + \beta \cos \omega t \quad (1)$$

Here T is hemispheric mean temperature deviation and C is effective heat capacity per unit area (units of $\text{J m}^{-2} \text{K}^{-1}$), so the left-hand side of Eq. (1) gives the hemisphere's rate of heat storage in W m^{-2} . On the right-hand side of Eq. (1), αT is the net negative feedback of the system, assumed to be a linear function of the temperature deviation, β is the amplitude of seasonal cycle forcing, and $\omega = 2\pi/(1 \text{ year}) \approx 2 \times 10^{-7} \text{ s}^{-1}$.

The general solution of Eq. (1) is

$$T = \frac{\beta/C}{\omega^2 + (\alpha/C)^2} [\omega \sin \omega t + (\alpha/C) \cos \omega t] + B e^{-\alpha t/C} \quad (2)$$

where B is a constant determined by the initial conditions of the problem. Since we are interested in seasonal cycle equilibrium, we may assume that t is large enough to make the term involving B negligible. In that case we can rewrite Eq. (2) as

$$T = (A/2) \cos \omega(t - t_0) \quad (3)$$

where

$$\tan \omega t_0 = C\omega/\alpha \quad (4)$$

and

$$A = \frac{2\beta}{(C^2\omega^2 + \alpha^2)^{1/2}} \quad (5)$$

Note that A is the peak-to-peak seasonal cycle amplitude, as defined in the main text of this study.

Climate sensitivity is the equilibrium mean temperature change ΔT caused by a standard increment ΔQ in annual mean forcing. We use atmospheric CO_2 doubling ($\Delta T = \Delta T_{2x}$ and $\Delta Q = \Delta Q_{2x} \approx 4 \text{ W m}^{-2}$) as the standard. Assuming that the same feedback processes operate as in the seasonal cycle, we have by analogy with Eq. (1)

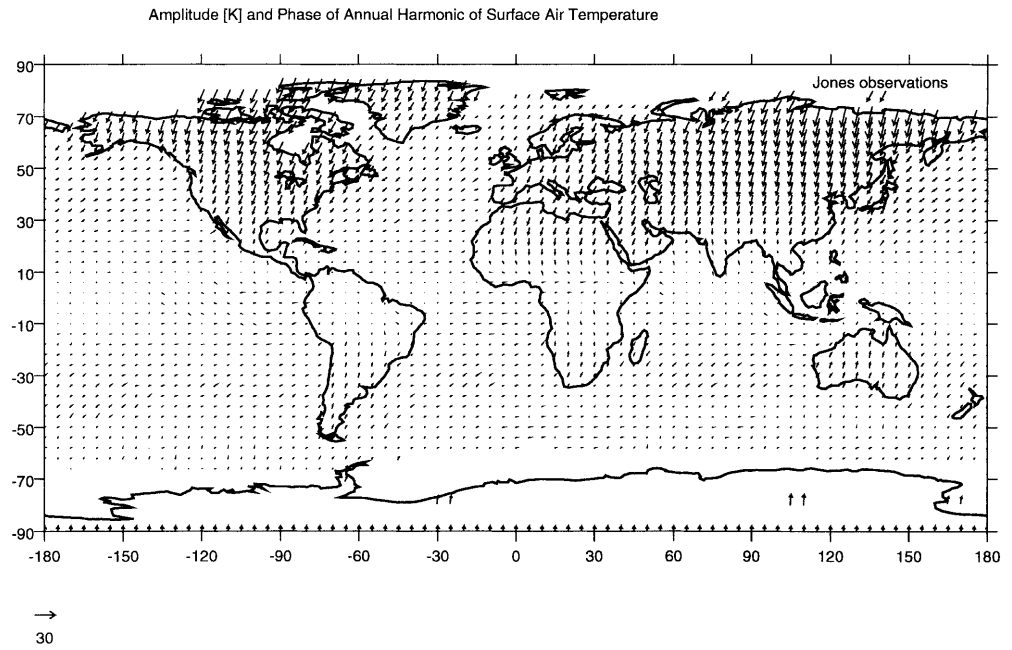
$$0 = -\alpha \Delta T_{2x} + \Delta Q_{2x} \quad (6)$$

showing that the feedback parameter α is inversely proportional to the equilibrium climate sensitivity. Eliminating α between Eqs. (5) and (6) gives

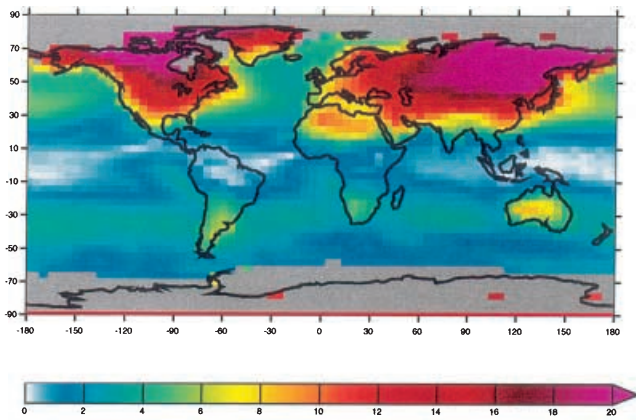
$$A = \frac{2\beta}{C\omega} \left[1 + \left(\frac{\Delta Q_{2x}}{\omega C \Delta T_{2x}} \right)^2 \right]^{-1/2} \quad (7)$$

which implies that A increases with increasing ΔT_{2x} . In particular, a plot of $1/A^2$ versus $1/(\Delta T_{2x})^2$ gives a straight line with positive slope.

Fig. A1 Amplitude and phase of the annual harmonic of SAT from the augmented Jones (observed) dataset. The arrow at lower left gives the amplitude scale in K. See text for phase information



Amplitude of First (Seasonal Cycle) Harmonic of T [K]



abs(July - January) / 2 [K]

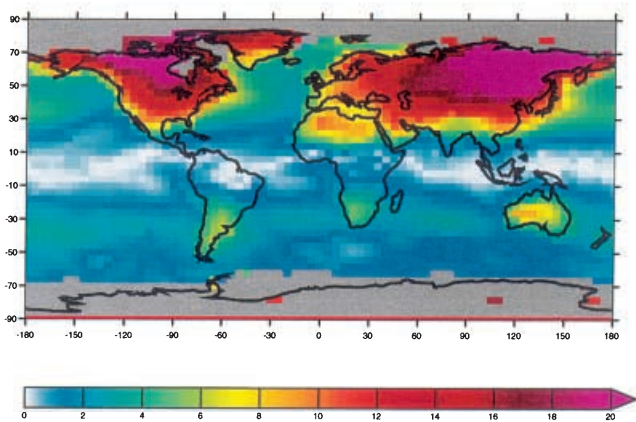


Fig. A2 Amplitude of the annual harmonic of SAT (*top*) and half the absolute value of July-minus-January SAT difference (*bottom*) from the augmented Jones (observed) dataset

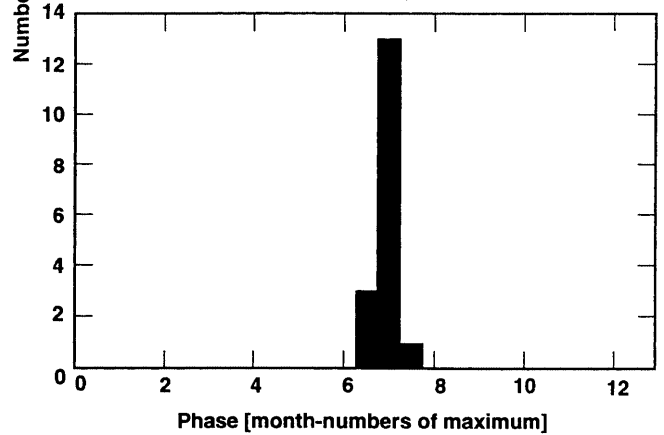
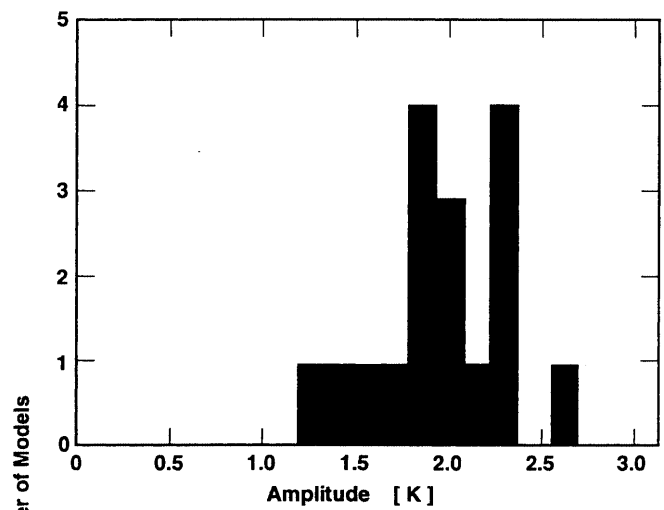


Fig. A3 Histograms of annual harmonic amplitude (*top*) and phase (*bottom*) for the globally averaged monthly mean SAT time series simulated by the 17 models used in this study

The strength of the interdependence between A and ΔT_{2x} depends on the ratio $\Delta Q_{2x}/\omega C \Delta T_{2x}$. This ratio is less than 1 if the climate's response time for CO_2 doubling is greater than a year, i.e., if $C \Delta T_{2x}/\Delta Q_{2x} > 1/\omega$. This inequality is easily satisfied by parameter choices consistent with conventional wisdom about climate sensitivity and response time. For example, Thompson and Schneider (1979, Appendix B) suggest that a reasonable value of C for the entire Earth (averaging over land and ocean areas) lies between about 4×10^7 and $2 \times 10^8 \text{ J m}^{-2} \text{ K}^{-1}$, so that $\Delta Q_{2x}/\omega C$ lies between about 0.1 and 0.5 K. Even for the largest of these values, the relation between A and ΔT_{2x} is weak: to first order

$$A \approx \frac{2\beta}{C\omega} \left[1 - \frac{1}{2} \left(\frac{0.5 \text{ K}}{\Delta T_{2x}} \right)^2 \right] \quad (8)$$

and the range $1.5 \text{ K} < \Delta T_{2x} < 4.5 \text{ K}$ implies a range in A of only about 5%. These simple considerations suggest that the correlation between A and ΔT_{2x} is both weak and nonlinear.

In this parameter regime A is near its large- C limit $2\beta/C\omega$, i.e., A is independent of ΔT_{2x} and inversely proportional to C . In a study applying Eq. (1) to the equatorial oceans, Schneider (1996) termed this case the "heat storage limit". The amplitude β of hemisphere-averaged seasonal cycle forcing is approximately $(S/4) \sin i = 136 \text{ W m}^{-2}$, where $S = 1360 \text{ W m}^{-2}$ is the solar energy flux at the top of the atmosphere (the "solar constant") and $i = 23.5^\circ$ is the tilt of Earth's axis (Milankovitch 1969). Substituting into $2\beta/C\omega$, one recovers the observed (9.5 K) value of A if $C \approx 1.4 \times 10^8 \text{ J m}^{-2} \text{ K}^{-1}$. This value of C lies within (but near the high end of) the "reasonable" range cited.

The phase as well as the amplitude of the seasonal cycle can be related to climate sensitivity. Eliminating α between Eqs. (4) and (6) gives

$$\tan \omega t_0 = \omega C \frac{\Delta T_{2x}}{\Delta Q_{2x}} \quad (9)$$

In the limit discussed ($C \Delta T_{2x}/\Delta Q_{2x} \gg 1/\omega$), the right-hand side of Eq. (9) becomes large and ωt_0 tends to $\pi/2$. This value indicates a phase lag of one-fourth the annual cycle, or three months, which would place the maximum and minimum hemispheric temperatures at the equinoxes. Such a prediction obviously disagrees with observations over land but may be roughly consistent with observations over oceans and sea ice. In any case, the phase lag implied by Eq. (9) is fairly robust because the inverse tangent function does not vary greatly at large values of its argument. For $1.5 \text{ K} < \Delta T_{2x} < 4.5 \text{ K}$ and values of C within the "reasonable" range cited ωt_0 differs from $\pi/2$ by at most 21%.

It is interesting to note the opposite, small- C limiting case of Eq. (7). In this case

$$A \approx 2\beta \frac{\Delta T_{2x}}{\Delta Q_{2x}} \approx (272 \text{ W m}^{-2}) \frac{\Delta T_{2x}}{\Delta Q_{2x}} \quad (10)$$

The seasonal cycle amplitude in this limit is independent of C and directly proportional to the equilibrium climate sensitivity. In effect the seasonal cycle becomes an equilibrium climate response (e.g., Eq. 9 shows that the phase lag $t_0 \rightarrow 0$ as $C \rightarrow 0$). If we substitute $\Delta Q_{2x} = 4 \text{ W m}^{-2}$ and $1.5 \text{ K} < \Delta T_{2x} < 4.5 \text{ K}$ Eq. (10) implies that A lies somewhere between 100 K and 300 K. Clearly this limit has little connection with Earth's climate, though with suitable changes of parameter values it may apply to an all-land planet such as Mars.

Our conclusions may be compared with earlier work using simple climate models. North and Coakley (1979) constructed a model of zonal mean surface temperature in which the seasonal cycle was simulated in agreement with observation. They found that albedo-temperature feedback, which increases the equilibrium climate sensitivity of their model, can be introduced without substantially altering the seasonal cycle. On the other hand, the model of Wigley and Raper (1991) confirms the positive sign of the correlation between A and ΔT_{2x} and additionally implies that their interdependence may be stronger than implied by Eq. (8) (T.M.L. Wigley personal communication). Wigley and Raper's (1991)

model resolves land and ocean areas into separate "boxes", unlike the North-Coakley (1979) type of model or the simple model used here. Aggregating land and ocean data to form zonal or hemispheric means may obscure a stronger connection between ΔT_{2x} and seasonal cycle amplitude over land, as suggested by the "all-land planet" limit discussed already. However, on recalculating A for land areas only, we find that the correlation shown in Fig. 7 weakens rather than strengthens.

References

- Barnett T (1999) Comparison of near surface air temperature variability in eleven coupled global climate models. *J Clim* 12: 511–518
- Bell J, Duffy P, Covey C, Sloan L, CMIP investigators (2000) Comparison of temperature variability in observations and sixteen climate model simulations. *Geophys Res Lett* 27: 261–264
- da Silva AM, Young CC, Levitus S (1994) Atlas of surface marine data 1994, vol. 1: Algorithms and procedures, NOAA Atlas NESDIS 6, US Department of Commerce, Washington, D.C.
- Gates WL, Henderson-Sellers A, Boer GJ, Folland CK, Kitoh A, McAvaney BJ, Semazzi F, Smith N, Weaver AJ, Zeng QC (1996) Climate models – evaluation. In: *Climate change 1995: the science of climate change*. Houghton JT et al. (eds) IPCC/Cambridge University Press, Cambridge, UK, pp 229–284
- Gates WL, Boyle JS, Covey C, Dease CG, Doutriaux CM, Drach RS, Fiorino M, Gleckler PJ, Hnilo JJ, Marlais SM, Phillips TJ, Potter GL, Santer BD, Sperber KR, Taylor KE, Williams DN (1999) An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). *Bull Am Meteorol Soc* 80: 29–55
- Gregory JM, Mitchell JFB (1997) The climate response to CO_2 of the Hadley Centre coupled AOGCM with and without flux adjustment. *Geophys Res Lett* 24: 1943–1946
- Jones PD (1988) Hemispheric surface air temperature variations: recent trends and an update to 1987. *J Clim* 1: 654–660
- Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Roelwsky C, Wang J, Leetma A, Reynolds R, Jenne R, Dennis J (1996) The NCEP/NCAR 40-year Reanalysis Project. *Bull Am Meteorol Soc* 77: 437–471
- Kattenberg A, Giorgi F, Grassl H, Meehl GA, Mitchell JFB, Stouffer RJ, Tokioka T, Weaver AJ, Wigley TML, Climate models – projections of future climate. In: *Climate change 1995: the science of climate change*. Houghton JT et al. (eds) IPCC/Cambridge University Press, Cambridge, UK, pp 285–357
- Kerr RA (1994) Climate modeling's fudge factor under fire. *Science* 265: 1528
- Lambert SJ, Boer GJ (2000) CMIP1 evaluation and intercomparison of coupled climate models. *J Clim*, submitted
- Legates DR, Willmott CJ (1990a) Mean seasonal and spatial variability in global surface air temperature. *Theoret Appl Clim* 41: 11–21
- Legates DR, Willmott CJ (1990b) Mean seasonal and spatial variability in gauge corrected, global precipitation. *Int J Clim* 10: 111–127
- Lindzen RS, Kirtman B, Kirk-Davidoff D, Schneider EK (1995) Seasonal surrogate for climate. *J Clim* 8: 1681–1684
- Manabe S, Stouffer RS (1980) Sensitivity of a global climate model to an increase of CO_2 concentration in the atmosphere. *J Geophys Res* 85: 5529–5554
- Meehl GA, Boer GJ, Covey C, Latif M, Stouffer RJ (1997) Intercomparison makes for a better climate model. *Eos* 78: 445–451
- Milankovitch M (1969) Canon of insolation and the Ice Age problem (p. 244), Royal Serbian Academy Special Publications, vol 132, Section on Mathematics and Natural Science, vol 33, Israel Program for Scientific Publications, Jerusalem

- Nakamura M, Stone PH, Marotzke J (1994) Destabilization of the thermohaline circulation by atmospheric eddy transports. *J Clim* 7: 1870–1882
- North GR, Coakley Jr JA (1979) Differences between seasonal and mean annual energy balance model calculations of climate and climate sensitivity. *J Atmos Sci* 36: 1190–1204
- Oort AH (1983) Global atmospheric circulation statistics 1958–1973 (Table 26, p. 95), NOAA Prof. Pap. 14, Geophysical Fluid Dynamics Laboratory, Princeton, NJ
- Sausen R, Barthel K, Hasselmann K (1988) Coupled ocean-atmosphere models with flux correction. *Clim Dyn* 2: 145–163
- Schnedier EK (1996) A note on the annual cycle of sea surface temperature at the equator. Report 36, Center for Ocean-Land-Atmosphere Studies, Calverton, MD, pp 11
- Schneider SH, Londer R (1984) The coevolution of climate and life (Ch. 6). Sierra Club Books, San Francisco, USA, pp 563
- Schnedier EK, Kirtman BP, Lindzen RS (1999) Tropospheric water vapor and climate sensitivity. *J Atmos Sci* 56: 1649–1658
- Schubert S, Wu CY, Zero J, Schemm JK, Park CK, Suarez M (1992) Monthly means of selected climate variables for 1985–1989. NASA Technical Mem, Goddard Space Flight Center, Greenbelt, MD, pp 376
- Shea DJ (1986) Climatological atlas: 1950–1979: Surface air temperature, precipitation, sea-level pressure, and sea-surface temperature (45S–90N). Technical Note TN-269 + STR, National Center for Atmospheric Research, Boulder, CO, 35 pp + 158 figures + microfiche
- Shea DJ (1996) An introduction to atmospheric and oceanographic datasets. Technical Note TN-404 + IA, National Center for Atmospheric Research, Boulder, CO, pp 134
- Singer SF (1997) Hot talk cold science: global warming's unfinished debate. The Independent Institute, Oakland, CA, pp 100
- Stouffer RJ, Manabe S (1999) Response of a coupled ocean-atmosphere model to increasing atmospheric carbon dioxide: sensitivity to the rate of increase. *J Clim* 12: 2224–2237
- Stouffer RJ, Hegerl G, Tett S (2000) A comparison of surface air temperature variability in three 1000-year coupled ocean-atmosphere model integrations. *J Clim* (in press)
- Thompson SL, Schneider SH (1979) A seasonal zonal energy balance model with an interactive lower layer. *J Geophys Res* 84: 2401–2414
- Washington WM, Meehl GA (1984) Seasonal cycle experiment on the climate sensitivity due to a doubling of CO₂ with an atmospheric general circulation model coupled to a simple mixed-layer ocean model. *J Geophys Res* 89: 9475–9503
- Wigley TML, Raper SCB (1991) Internally generated natural variability of global-mean temperatures. In: Greenhouse-gas-induced climatic change: a critical appraisal of simulations and observations. Elsevier Science, Amsterdam, pp 471–482