Model assessment of regional surface temperature trends (1949-1997)

T. R. Knutson, T. L. Delworth, K. W. Dixon, and R. J. Stouffer

Geophysical Fluid Dynamics Laboratory/NOAA, Princeton, New Jersey

Abstract. Analyses are conducted to assess whether simulated trends in SST and land surface air temperature from two versions of a coupled ocean-atmosphere model are consistent with the geographical distribution of observed trends over the period 1949-1997. The simulated trends are derived from model experiments with both constant and time-varying radiative forcing. The models analyzed are low-resolution (R15, ~4°) and medium-resolution (R30, ~2°) versions of the Geophysical Fluid Dynamics Laboratory (GFDL) coupled climate model. Internal climate variability is estimated from long control integrations of the models with no change of external forcing. The radiatively forced trends are based on ensembles of integrations using estimated past concentrations of greenhouse gases and direct effects of anthropogenic sulfate aerosols (G+S). For the regional assessment, the observed trends at each grid point with adequate temporal coverage during 1949-1997 are first compared with the R15 and R30 model unforced internal variability. Nearly 50% of the analyzed areas have observed warming trends exceeding the 95th percentile of trends from the control simulations. These results suggest that regional warming trends over much of the globe during 1949-1997 are very unlikely to have occurred due to internal climate variability alone and suggest a role for a sustained positive thermal forcing such as increasing greenhouse gases. The observed trends are then compared with the trend distributions obtained by combining the ensemble mean G+S forced trends with the internal variability "trend" distributions from the control runs. Better agreement is found between the ensemble mean G+S trends and the observed trends than between the model internal variability alone and the observed trends. However, the G+S trends are still significantly different from the observed trends over about 30% of the areas analyzed. Reasons for these regional inconsistencies between the simulated and the observed trends include possible deficiencies in (1) specified radiative forcings, (2) simulated responses to specified radiative forcings, (3) simulation of internal climate variability, or (4) observed temperature records.

1. Introduction

The interpretation of warming trends in climate records over the past century remains an area of uncertainty in climate assessments [*Santer et al.*, 1996b]. The problem has been framed in terms of (1) detection of a climate change (i.e., concluding that an observed change is outside the range of internal climate variability alone) and (2) attribution of a change to various external forcings (i.e., demonstrating that the change is consistent with an expected forced climate response in the presence of internal climate variability noise). The external forcing may be either natural (e.g., changes in solar insolation or volcanic aerosol loading) or anthropogenic (e.g., changes in greenhouse gas concentrations due to human activities). Detection/attribution conclusions in general depend upon estimates or assumptions about the nature of the internal (unforced) climate variability on various time and space scales.

One approach to the detection problem has been to use climate models to estimate internal variability, with the models ranging from energy balance models [e.g., *Wigley and Raper*, 1990; *North and Stevens*, 1998] to global coupled ocean atmosphere general circulation models [e.g., *Stouffer et al.*, 1994; *Santer et al.*, 1995; *Hegerl et al.*, 1996; *Tett et al.*, 1997; *North and Stevens*, 1998; *Knutson and Manabe*, 1998]. An alternative approach has been to fit statistical models to climate data and, using the fitted models, to

assess the statistical significance of secular trend components or recent anomalous climate behavior [e.g., *Ghil and Vautard*, 1991; *Bloomfield*, 1992; *Trenberth and Hoar*, 1996]. A third approach, paleoclimate reconstruction, offers an opportunity to gain longer-term perspectives on climate variability [e.g., *Mann et al.*, 1998].

The attribution of climate change to particular forcings has typically involved quantitative comparisons of the spatial patterns of temperature change from forced climate change experiments with observed patterns of change. For example, *Barnett and Schlesinger* [1987] and *Santer et al.* [1993] illustrated the use of pattern correspondence approaches in early attempts to detect a greenhouse gas-induced signal in observed temperature data. *Santer et al.* [1995] used centered and uncentered pattern similarity statistics and concluded that they were able to detect the spatial signature of combined CO_2 + anthropogenic sulfate aerosol signal (based on equilibrium climate change experiments) in the observed data but not a CO_2-only signal. Their results and the use of the pattern correlations in detection studies were recently further elucidated by *Wigley et al.* [1998] using analysis of synthetic data. *Mitchell et al.* [1995] first showed that in a greenhouse gas + aerosols coupled model experiment the spatial correlation of simulated surface temperature changes versus observations tended to increase in recent decades, although not with high statistical significance. The vertical structure of atmospheric temperature changes has been investigated as a possible fingerprint of anthropogenic climate change by a number of authors [e.g., *Karoly et al.*, 1994; *Tett et al.*, 1996; *Santer et al.*, 1996a; *Vinnikov et al.*, 1996; *Hansen et al.*, 1997].

Optimal signal detection theory approaches to the detection/ attribution problem have been proposed by several investigators [*Hasselmann*, 1979; *Bell*, 1982, 1986; *North et al.*, 1995]. These approaches are designed to filter the data in such a way as to enhance the possibility of detecting a predicted climate change signal if such a signal exists in the observed data. In a recent application of such a method, *Hegerl et al.* [1997] analyzed several global climate model control experiments (internal variability only) and forced climate change experiments. They concluded that the observed surface temperature changes in the climate record are consistent with a combination of greenhouse gas and aerosol forcing but inconsistent with either greenhouse gases alone or solar variability alone. *North and Stevens* [1998] analyzed the forced response of a two-dimensional energy balance model along with internal variability from several global climate models and from a noise-forced energy balance model. They found highly significant volcanic, greenhouse gas, and anthropogenic aerosol signals, but they were not able to detect a solar signal with high confidence. *Tett et al.* [1999] analyzed a series of coupled global climate model simulations of the 20th century climate forced by various combinations of greenhouse gases, anthropogenic aerosols, volcanic aerosols, and solar variability. They reaffirmed the *Hegerl et al.* [1997] conclusion that the warming in recent decades is attributable to changes in greenhouse gases and sulfate aerosols and, in addition, found that solar variability may have contributed to the early 20th century warming.

In other recent detection studies, *Leroy* [1998] examined the detection problem, including the optimal detection techniques, from a Bayesian statistics perspective. He noted the potential for rating models relatively according to their predictions using such methods, and the related need for estimates of uncertainties associated with model predictions. *Stott and Tett* [1998] examined the dependence of (nonoptimal) climate change detection on the time and space scale used by analyzing trends of different lengths and projecting trends onto spherical harmonics. In general, they noted that prospects for detection were poorer for shorter timescales (<50 years) and smaller spatial scales (<5000 km). *Barnett et al.* [1998] considered the effect of both model errors and modeled internal atmospheric variability on the detection problem.

The methodology in the present study is intended to complement the more sophisticated detection/attribution techniques cited above. The goal is to provide, within a relatively simple framework, a regional-scale assessment of where the Geophysical Fluid Dynamics Laboratory (GFDL) coupled climate model is or is not consistent with observations, taking into account the effects of internal climate variability. The model is considered inconsistent with the observations if the observed trend lies outside the range of trends simulated by the model under certain specified conditions. This assessment is divided into two parts. The first assessment (internal variability) compares observed trends in various regions to the distribution of trends simulated in an unforced version (time-invariant radiative forcing) of the climate model. The model versus observation comparison is done on a grid-point-by-grid-point basis, at least over those areas of the globe where the observed temporal coverage is sufficient during the period 1949-1997. The second assessment (*G+S* forcing) attempts to identify regions where the GFDL coupled model is or is not consistent with observations based on a radiative forcing scenario combining greenhouse gases and the direct effects of anthropogenic sulfate aerosols (*G+S*). The *G+S* forcing scenario used here contains uncertainties and neglects several potentially important radiative forcings such as changes in ozone, volcanic aerosols,

solar influences, and indirect effects of tropospheric aerosols [*Schimel et al.*, 1996].

The observed and simulated trends are compared for the period 1949-1997. During this period, SST records are more reliable and more generally available than in the first half century, giving us more confidence in the regional details of the observed trend patterns. In addition, the 49-year record length should be long enough to avoid some of the potential problems [*Barnett et al.*, 1998; *Stott and Tett*, 1998] with using relatively short record lengths (< 20 years) for climate change detection studies. A potential problem of the present approach, illustrated in this study, is that the time interval considered (49 years) is sufficiently brief that regional details of the trend patterns in the climate model's *G+S* integrations will differ substantially depending on changes in initial conditions. These differences between experiments with the same radiative forcing scenario (*G+S*) arise due to the amount of internal climate variability in the model (i.e., "noise") versus the size of the radiatively forced change over the period (i.e., "signal"). Therefore we use the ensemble mean of several *G+S* integrations to estimate the actual response of the model to the forcing over the period (signal), and we use variability from very long control integrations to estimate how much the model trends can be expected to vary from the ensemble mean estimate (or from zero for the unforced assessments) due to internal climate variability (noise).

2. Data Description

2.1. Model Description and Experiments

The coupled ocean-atmosphere models used are low- (R15) and medium- (R30) resolution versions of the model described by *Manabe et al.* [1991], *Manabe and Stouffer* [1996], and references therein. The model has interactive clouds and seasonally varying solar insolation. The atmospheric component is a global general circulation model (GCM), with 9 (R15) or 14 (R30) finite-difference (sigma) levels in the vertical. The horizontal distributions of variables in this model are represented in both spectral and grid point domains, with rhomboidal truncation at zonal wavenumber 30 and a 3.75° longitude x 2.2° latitude computational grid for the R30; the R15 has half the horizontal resolution of the R30. The R30 ocean component is a global 18-level grid point GCM (MOM1) [*Pacanowski et al.*, 1991] with a 1.875° longitude x 2.2° latitude resolution and a 40-m-thick top layer; the R15 ocean component has 12 vertical levels with a 50-m-thick top layer and again half the horizontal resolution of the R30.

The models use a flux adjustment technique for heat and salinity fluxes at the ocean surface to reduce model drift and insure that the CO_2 perturbations and internally generated variability occur relative to a reasonably realistic control run state [*Manabe et al.*, 1991]. Because of the flux adjustment, both R30 and R15 models are relatively free of substantial long-term drifts, although detectable adjustments of global temperatures on multicentury to multimillennial timescales occur in the control runs. For the R15 control run [*Manabe and Stouffer*, 1996], the effect of this long-term adjustment is reduced in this study by eliminating the first 2000 years of the 12,000-year control coupled integration from the analysis. For the R30 model, such a long integration is not available; we therefore use years 101-1000 of a 1000-year control integration, skipping the first 100 years due to the relatively large initial adjustments of the thermohaline circulation in the model during that period.

The R30 climate change experiments used in this report include an ensemble of five equivalent greenhouse gases plus direct sulfate aerosol forcing (G+S) integrations over the period 1865 to 2000. Each ensemble member was begun from a different ocean/atmosphere initial condition (derived from control run conditions at least 50 years apart), and each was forced using the IS92a scenario [Intergovernmental Panel on Climate Change (IPCC) 1992] of estimated past or projected (post-1990) G+S radiative forcing [Haywood et al., 1997]. The R30 coupled model used here is an updated version of the R30 coupled model examined by Knutson and Manabe [1998]. The newer version exhibits significantly less climate drift in high latitudes and stronger decadal sea surface temperature (SST) variability in the tropical Pacific, as will be discussed in this report. The principal difference between the model versions is increased diffusion in the ocean component in the present set of model experiments; a number of other presumably minor changes have also been incorporated.

The R15 climate change experiments in this report include a nine-member ensemble of G+S integrations. The individual ensemble members have different combinations of ocean/atmosphere initial conditions and start dates for the G+S scenario (i.e., 1750, 1865, or 1915). For calculations in this report requiring pre-1949 data, only six of the R15 integrations (1750 and 1865 start dates only) were used. The R15 experiments are described more fully by Dixon and Lanzante [1999].

The R15 and R30 models each have certain advantages and limitations for use in the study. The R15 model, being more computationally efficient, provides for very long integrations and larger ensembles, which better define the signal and noise elements in the problem. As we will show, the R30 model has more realistic levels of interannual SST variability in the tropical Pacific and thus, we believe, provides a better representation of the impact of El Niño on trend assessments than the R15 model.

2.2. Observed Data Sets

The primary observed surface temperature data set used for the study is a combination of the Jones [1994] land surface air temperatures (LSATs) and the Parker et al. [1995] SSTs, updated through 1997 and available on a 5° x 5° grid. The land and SST data sets have been merged as discussed by Parker et al. [1994]. In many cases, outlined in detail later, the climate model data are filtered by masking out locations and periods where observations are not available.

Some preliminary results are presented based on the Jones [1994] data extended back to 1900, but these analyses are limited by the relatively sparse data coverage early in the century. An alternative would be to use reconstructions of SSTs dating back to the late 19th century [e.g., Smith et al., 1996; Rayner et al., 1996; Kaplan et al., 1998], although we have chosen not to do so in this report, where the main focus is on the relatively well-sampled post-1950 period.

3. Evaluation of Simulated Internal Variability

In this section, the simulated internal variability in the R30 and R15 climate models is compared to observed variability. A precise evaluation of the model is not possible due to the limited length and other shortcomings of the observational record, as well as our inability to separate the observed record into internal and forced variability components. One approach, used in this section, is to remove the linear trends from both observations and the model

control runs and then to compare the residual variability in terms of its spatial pattern, magnitude, and spectral characteristics.

3.1. Global Time Series and Spectra

The (unfiltered) observed record of annual global temperature [Parker et al., 1995; Jones, 1994] is shown in Plate 1 (top, black solid curve). This record shows the familiar 20th century warming trend and variability on interannual to multidecadal timescales. The R30 (n=5) and R15 (n=6) G+S scenario curves in Plate 1 show a similar degree of 20th century warming to the observations, consistent with several previous studies [Mitchell et al., 1995; Haywood et al., 1997; Johns et al., 1997]. The "envelope" containing the R30 G+S curves is broader than that for the R15, reflecting larger internal decadal variability in the R30, as discussed below. One of the five R30 G+S integrations (see top diagram) reproduces the observed record particularly well, including the pronounced warming from about 1910 to 1945, followed by a slight cooling trend for several decades, and resumption of the pronounced warming trend in the last few decades of the century. This result indicates that for the R30 model the combination of G+S forcing and internal climate variability is sufficient to reproduce the main features of the 20th century global-mean surface temperature record. The other R30 ensemble members illustrate that the pronounced pre-1940 warming does not occur in all G+S integrations, suggesting a potentially important role for internal climate variability [e.g., Schlesinger and Raman-kutty 1994] in the interpretation of 20th century global temperature observations.

The control run curves for the R30 and R15 simulations (bottom two panels of Plate 1) show 135-year samples of unforced internal climate variability as simulated by the two models. The variability on the quasi-decadal scale is somewhat more pronounced in the R30 control series than in the R15 control. The R30 control shows a slight cooling tendency; the trend of global temperature over years 101-1000 from this run is -0.013 K 100 yr^{-1}.

In Figure 1 the variance spectra of global mean temperature from the control models are compared with that of the observed record (1898-1997). The spectra from the R15 control (Figure 1a) show substantially less variability than observed on the 3-to-5-year timescale characteristic of El Niño, consistent with previous studies showing an unrealistically low amplitude for El Niño variability in the R15 coupled model [Lau et al., 1992; Manabe and Stouffer, 1996; Knutson et al., 1997]. The R30 control model spectra (Figure 1b) indicate larger variability than observed on the quasi-decadal (~10 years) timescale. In Figures 1c and 1d are spectra for detrended global time series from the R15 and R30 G+S scenario experiments, as shown in Plate 1. The model G+S time series have been masked according to the available observations, which tends to enhance the variability at all frequencies. In addition, the response to the G+S forcing is not fully described by a linear trend (see Plate 1) which leads to an enhancement of the low-frequency variance in the spectra of the linearly detrended G+S data. Thus while the R15 and R30 control runs tend to have lower variability than observed at the lowest frequencies shown (i.e., >20-year periods), this difference appears relatively minor for the detrended and masked data from the G+S experiments (Figures 1c and 1d).

In short, the results in Figure 1 indicate that the R15 and R30 climate models simulate global multidecadal variability at a level comparable to that in the available observed temperature record. Although the true level of internal multidecadal climate variability
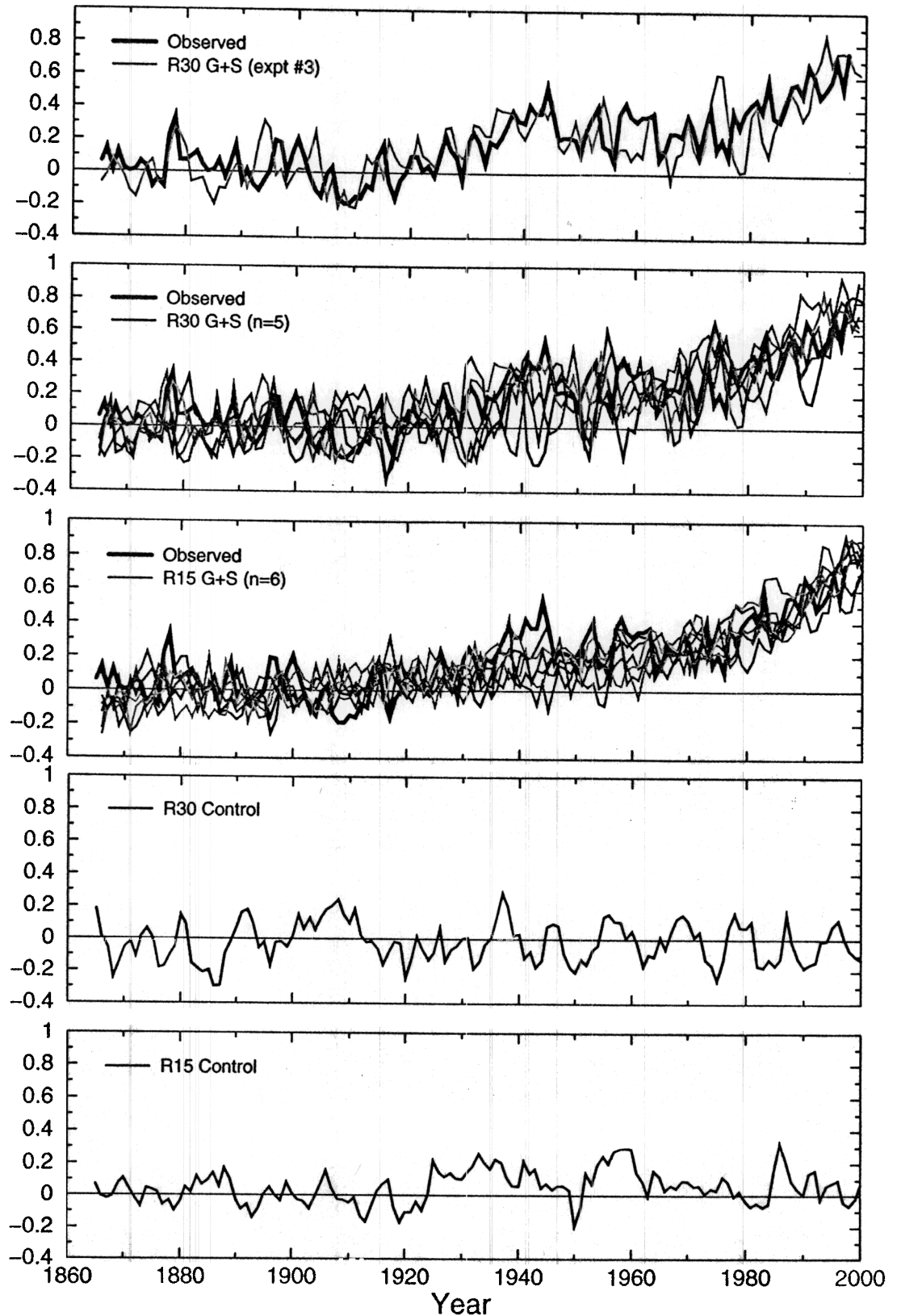
Global Temperature Indices

Plate 1. Annual global mean surface temperature anomalies (combined sea surface temperature (SST) over oceans, land surface air temperature (LSAT) over land) from 1865 to present based on observations (thick black curves, top three panels); R15 and R30 *G+S* experiments (thinner curves, various colors, top three panels); and sample 135-year segments from the R30 and R15 model control runs (bottom two panels). The global mean anomalies are relative to the 1880-1920 base period. The blue curve overlying the observed curve in the top panel is from R30 *G+S* experiment 3. R15 and R30 *G+S* scenario data have been sampled from the model SST and LSAT fields according to those times and locations where observations were available.
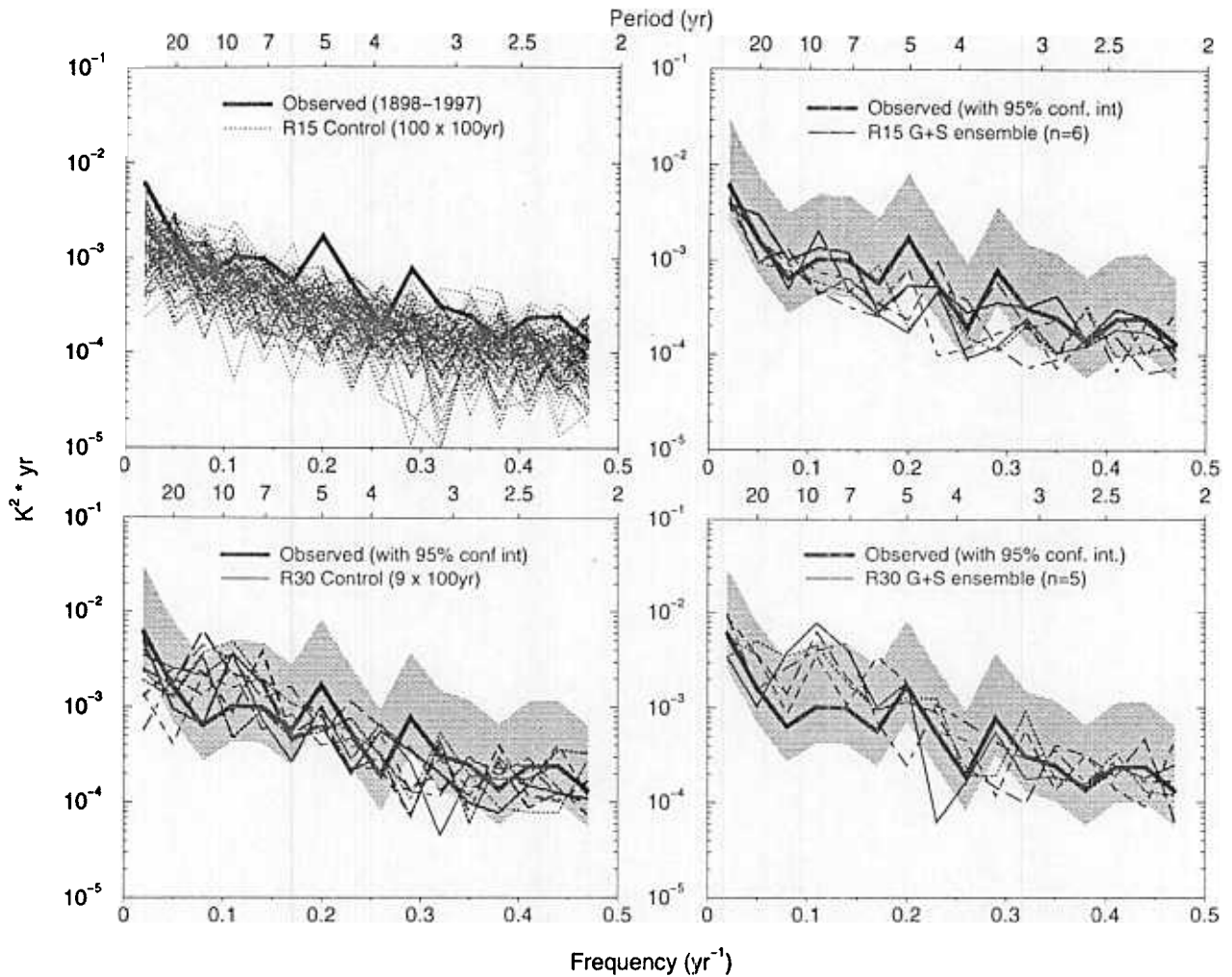
Figure 1. Variance spectra of global surface temperature series for observations (1898-1997) and for nonoverlapping 100-year time series segments from the various model experiments. The observed spectra are depicted by the thickest solid curve on each plot, with 95% confidence intervals shown by shading in Figures 1b-1d. The model spectra are depicted by the thinner lines, with various solid or dashed patterns representing different *G+S* experiments (1898-1997) or nonoverlapping control run segments. All 100-year annual mean series were detrended prior to computing the spectral coefficients. The raw spectra were smoothed using a nonoverlapping rectangular window of width 3 calculable frequencies. The models used are (a) R15 control run (10,000 year, or 100 x 100 year segments); (b) R30 control run (900 year, or 9 x 100 year segments); (c) R15 *G+S* ensemble experiments (*n*=6); and (d) R30 *G+S* ensemble experiments (*n*=5). (c, d) Model time series have been constructed using the observed data mask.

is not known, these results suggest that the models may provide a useful approximation of such variability.

3.2. Geographical Distribution

Maps of the interannual standard deviation of SST and land surface air temperature (LSAT) are shown in Plate 2. The observations are based on detrended data for the period 1949-1997. *Manabe and Stouffer* [1996] have presented a more comprehensive comparison with observations and a discussion of the mechanisms of internal surface temperature variability obtained from the GFDL R15 coupled model. See *Delworth* [1996] and *Knutson et al.* [1997] for further analysis of the GFDL climate model internal variability, or *Tett et al.* [1997] for an examination of the unforced variability in the Hadley Centre HADCM2 coupled model. The observations show regions of enhanced variability in middle to high latitudes over Asia and North America and in the equatorial Pacific associated with El Niño. Both R15 and R30 models show

enhanced variability over middle and high latitudes of Asia and North America in a similar pattern to the observations but with somewhat higher levels in high latitudes. The models have substantially greater variability than observed over subtropical land regions of both hemispheres (e.g., Australia, southern Africa, and the southern United States). Separate analyses [e.g., *Delworth and Manabe*, 1989] have demonstrated that a substantial fraction of this variability is related to land-surface processes. The relatively dry soil conditions in these regions during summer may contribute to the excessive variability of simulated surface air temperature. In terms of SST the R30 model has substantially higher interannual variability than the R15 model (or than observed) in two main regions: the central equatorial Pacific and the high-latitude North Atlantic. The North Atlantic feature is associated with pronounced variability of the R30 model thermohaline circulation to be reported on in a future study.

In the equatorial Pacific, neither the R15 nor the R30 model adequately simulates the geographical distribution of variance
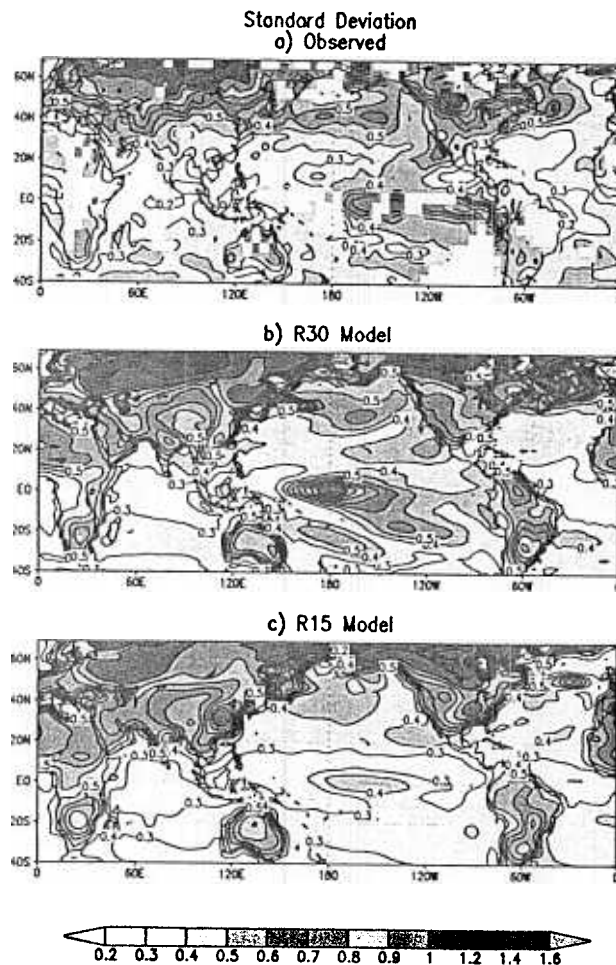
Standard Deviation
o) Observed

b) R30 Model

c) R15 Model

0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1 1.2 1.4 1.6

Plate 2. Standard deviation of annual mean SST (ocean regions) or surface air temperature (land regions) from (a) observations for 1949-1997; (b) the 900 year R30 control experiment; and (c) the 10,000 year R15 control experiment. The time series at each grid point were detrended prior to computing the standard deviations. Contour interval is 0.1 K for 0.1 to 1.0 K; and 0.2 K for ranges above 1.0 K.

along the equator apparent in the observations (i.e., a maximum extending from along the coast of South America westward to about 160°W). The local variability maximum in the R30 model is located substantially farther west than the observed maximum and somewhat west of the R15 simulated maximum. The maximum standard deviation values in the R30 model (near the dateline) are slightly higher than the observed maximum values (in the eastern equatorial Pacific). Regional spectral analysis and EOF analyses of the R30 model SST and subsurface temperature fields (not shown) indicate that the local maxima in SST variability in Plate 2b corresponds to an El Niño-like variability in the R30 model similar to that in previous studies of the GFDL coupled climate models [*Lau et al.*, 1992; *Knutson et al.*, 1997; *Knutson and Manabe*, 1998]. In the present R30 model, the phenomenon has a distinctly longer timescale (~8 to 10 years) than either the observed El Niño or the simulated El Niño in an earlier version of the R30 coupled model [*Knutson and Manabe*, 1998]. The longer timescale leads to a pronounced decadal peak in the spectra even for globally averaged temperature (Figures 1b and 1d). The unrealistically large decadal-scale SST variability in the equatorial Pacific in this version of the R30 model is a topic of continuing investigation.

4. Comparison of Observed and Simulated Trends (1949-1997)

4.1. Observed Trends

In this section, the coupled models are used to assess the possible roles for internal variability and radiative forcing in the observed trends over the period 1949-1997 and to assess the consistency between simulated and observed trends on a regional basis. Observed and simulated maps of SST and LSAT trends for 1949-1997 are shown in Plate 3 for annual mean and seasonal mean (JJA, DJF) data. For the observations (Plates 3a-3c), data coverage is assumed sufficient to compute a trend for a grid point if all decades have at least 50% of the years available, where a year is assumed available if at least 6 of 12 months (annual) or 2 of 3 months (seasonal) are available (nonmissing). The observed trend maps show more regions of warming than cooling, with the strongest warming occurring over middle- to high-latitude continental regions of Asia and western North America during boreal winter (DJF). These trend results are consistent with results of *Nicholls et al.* [1996]. The Pacific region trends are characterized by a broad triangular-shaped region of warming centered in the tropics [*Knutson and Manabe*, 1998], flanked by regions of cooling in the extratropics (central to western parts of the basin) in both hemispheres. Cooling trends are also indicated over much of the North Atlantic basin and over eastern North America, where they are most pronounced in the winter months (DJF). In the Southern Hemisphere, the warming appears fairly spatially uniform over most regions, with the exception of the cooling in the southwest Pacific to the east of Australia and over parts of South America.

4.2. Simulated Response to Radiative Forcing

Maps of the simulated trends for 1949-1997 from an ensemble (*n*=5) of R30 coupled model experiments are shown in Plates 3d-3f. Because of the presence of a cooling trend in the control integration as previously described, the results at each grid point are computed as a net trend (*G+S* experiment trend minus the control run trend over years 101-700 at that grid point) for each ensemble member. The simulation results show warming over almost the entire globe (and a greater fraction of the globe than in the observed trends). A few regions show a cooling, particularly in the high-latitude North Atlantic. A notable similarity between the model and the observed trends is the enhanced warming over middle- to high-latitude regions of Asia and western North America during DJF. In the model, this enhanced continental warming feature is largely confined to the DJF season and is much less evident during March-May and September-November (not shown). In the observations this enhanced continental warming feature is even more pronounced than in the model during DJF and is also fairly pronounced during the March-May season [*Nicholls et al.*, 1996]. Other notable regional features of the observations which are not apparent in the simulation are the cooling trends over the northern midlatitude Pacific, North Atlantic, and parts of the Southwest Pacific, Asia, and North America.

In assessing the regional discrepancies described above, it is important to consider the role of internal climate variability or "noise" in the model versus observation comparisons. The observed trend undoubtedly results from some combination of internal climate variability and radiative forcing, but it is not possible at present to distinguish these two contributions. The much smoother patterns in the model-simulated trend fields in Plate 3, compared to the observed 1949-1997 trends, are a consequence of using the ensemble mean trend for the model results,
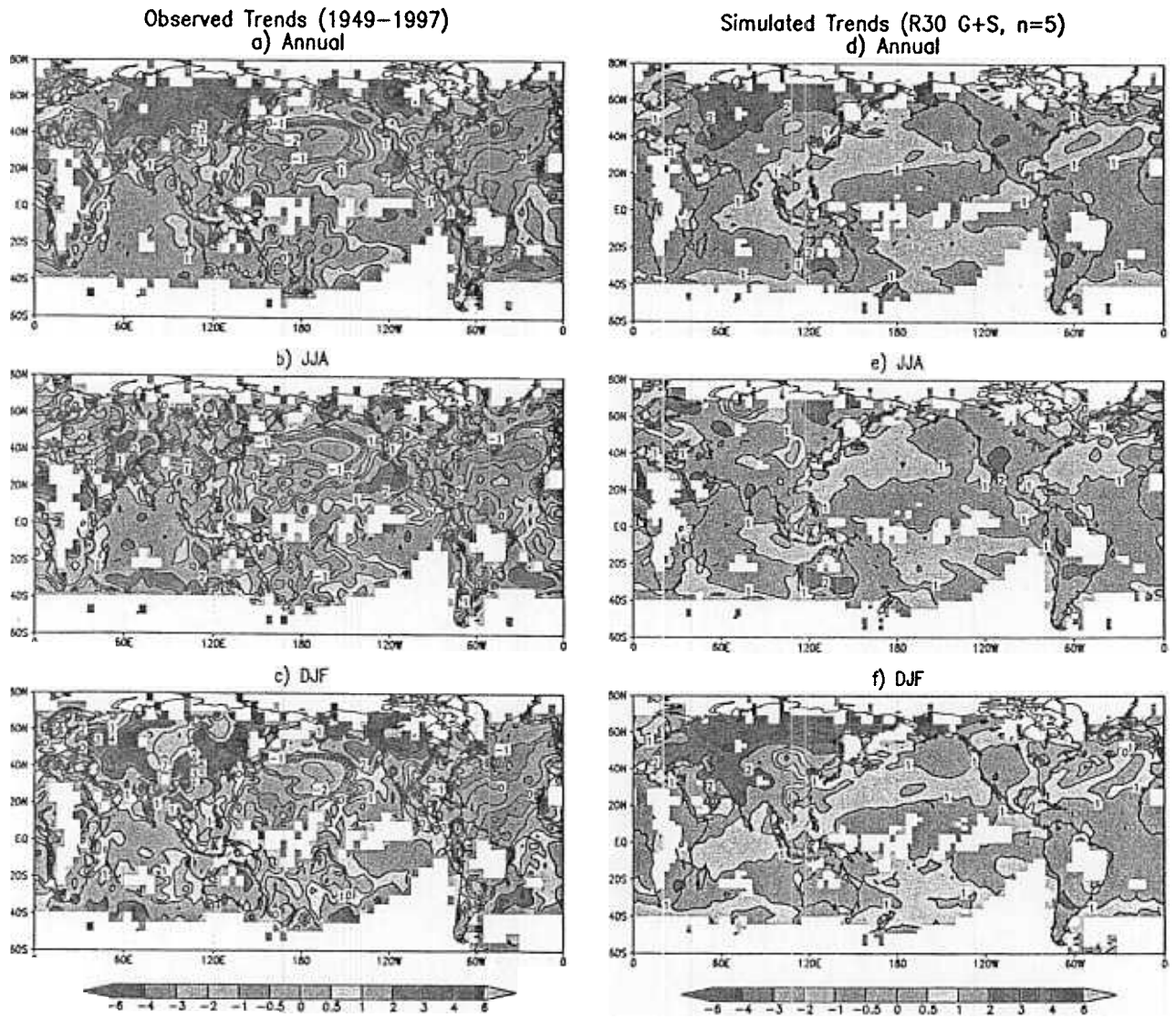
Observed Trends (1949−1997)
a) Annual

Simulated Trends (R30 G+S, n=5)
d) Annual

b) JJA

e) JJA

c) DJF

f) DJF

Plate 3. Linear trends in SST/LSAT for 1949-97 for observations (Plates 3a-3c) and for the ensemble mean of the R30 *G+S* set (*n*=5) of integrations (Plates 3d-3f). Results are for (a, d) annual, (b, e) June to August, or (c, f) December to February means, plotted with a contour interval of 1 K 100 yr^{-1} along with +/- 0.5 K 100 yr^{-1} contours. For the model trends the control run trend over the years 101-700 has been subtracted at each grid point. White regions denote grid boxes where the observed data coverage was found to be insufficient (see text for details). Approximately 73% of the globe was determined to have adequate coverage.
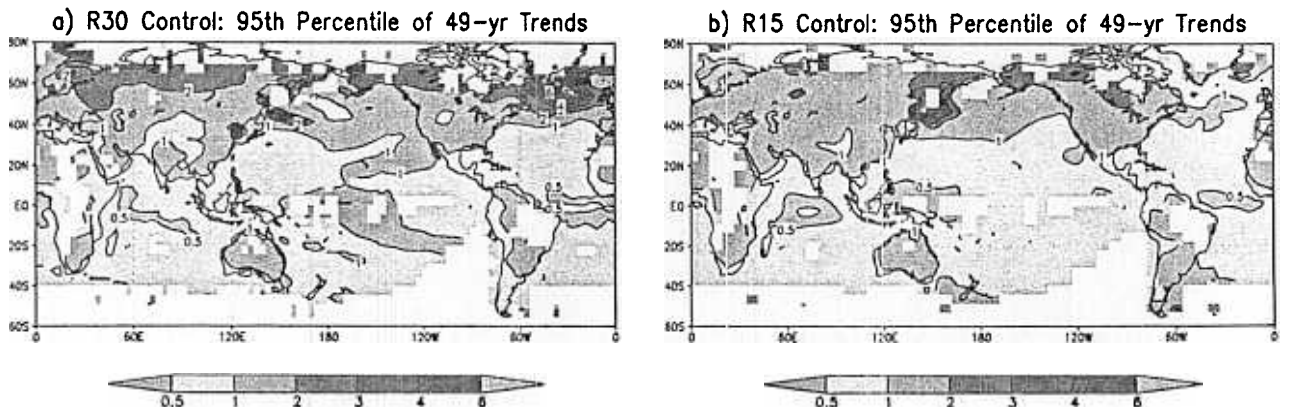
a) R30 Control: 95th Percentile of 49−yr Trends

b) R15 Control: 95th Percentile of 49−yr Trends

Plate 4. Geographic distribution of the 95th percentile of internally generated 49-year trends in annual SST/LSAT from the (a) R30 and (b) R15 control runs. All model time series were masked according to the availability of observations over the period 1949-97; white regions denote grid boxes where the observed data coverage was found to be insufficient (see text for details). Unit, K 100 yr^{-1}.

which strongly suppresses the spatial variation in the fields. This can be seen by examining the 1949-1997 trend maps from the individual G+S experiments in the Appendix (Plates 8 and 9). Each G+S experiment begins from a different initial condition, which leads to a different pattern of time-evolving noise superimposed on the radiatively forced signal. We have no reason to suspect, based on our previous experience with coupled climate change experiments, that the radiatively forced signal is strongly dependent upon the initial condition (i.e., upon which period from the control run is used as the initial condition for the climate change experiment).

Although in this study the ensemble mean of the trends obtained from several integrations is used as the model-estimated radiatively forced "signal," other approaches could be used. For example, *Jones and Hegerl* [1998] use the first empirical orthogonal function (EOF1), in combination with temporal filtering and a linear climate response model, to extract the dominant climate change response pattern and its time evolution from a coupled atmosphere/ocean GCM experiment. The EOF1 method assumes the climate change can be described by a single fixed pattern that explains the largest amount of variance in the time-evolving data. The linear trend method used here gives an alternative fixed pattern based on an assumed linear time dependence at each grid point.

4.3. Simulated Trends Due to Internal Variability

While the individual G+S simulations in the Appendix suggest a significant potential role for internal climate variability in the 1949-1997 trends, the small sample of independent realizations shown ($n=5$ for R30, or $n=9$ for R15) provides a rather limited sample of the probability distribution of simulated 1949-1997 trends. To better define this distribution of internally generated "trends," long control integrations without external forcing variations are used. The observed trends are then compared with the control run variability alone and with a combination of control run variability and the ensemble mean radiatively forced (G+S) signal.

The geographical distribution of internally generated 49-year trends in annual SST and LSAT from the climate models is illustrated by the 95th percentile of trends from 10,000 year (R15) or 900 year (R30) control runs (Plate 4). For these calculations all possible contiguous 49-year segments are included in the distributions (e.g., for the R30, trends for years 101-149, 102-150, 103-151, ..., and 852-900 are computed); thus the individual members of the distributions do not represent independent realizations. Note that the internally generated "trends" in Plate 4 were found to be sustained for 49 years in the model, although we report the trends in units of K 100 yr^{-1} in this study. The long-term underlying linear trends (900 years for R30; 10,000 years for R15) were removed prior to forming the 49-year segments. The data for each 49-year model segment was filtered using the observed data mask. The 5th percentile maps of the trend distributions (i.e., negative trends) are not shown but are very similar to the 95th percentile maps apart from a reversal of sign. As a measure of the similarity of the 95th and 5th percentile maps, the area mean of the absolute value of their sum (95th percentile + 5th percentile) at each grid point is only about 0.1 K 100 yr^{-1} (R30) and 0.03 K 100 yr^{-1} (R15).

According to Plate 4 the internally generated positive 49-year trends are generally of the order of 2 K 100 yr^{-1} or less over the analysis region. Values exceeding 2 K 100 yr^{-1} are simulated over much of northern Asia in the R30 model. Much larger trends are generated in a few smaller regions, including northeast of Japan in both R15 and R30 models and in the high-latitude North Atlantic in the R30 model. The latter feature is associated with fluctuations in the Atlantic Ocean overturning circulation on multidecadal

timescales in the R30 model. These fluctuations are particularly large during the first 300 years of the control simulation, associated with an initial adjustment of the thermohaline circulation, but weaken considerably after that. The large trend values northeast of Japan are a fairly small-scale feature and appear to be model dependent; these will not be investigated in this study. Elsewhere, the R15 and R30 models give fairly similar results with the notable exception of the central equatorial Pacific and some surrounding regions, where pronounced El Niño-like variability on 8-9 year timescales in the R30 control run leads to greater long-term trends in that model.

The high levels of simulated interannual variability in certain regions noted above and in section 3.2 (i.e., in the R30 model the decadal-scale equatorial Pacific variability and multidecadal North Atlantic variability, and in both the R15 and the R30 models the interannual variability over subtropical land regions) may imply that the control models generate unrealistically large multidecadal trends in those regions. However, such a conclusion is uncertain due to the relatively short observational records. An overestimate of multidecadal trends would be conservative for the purposes of the internal variability trend assessment (i.e., make it less likely that a significant climate change would be detected). On the other hand, for the G+S trend assessment an overestimate would inflate the uncertainty ranges for the simulated regional climate response, making the model less likely to fail a consistency test (versus observations) in those regions. The opposite biases would apply in regions where the simulated variability is too small, such as the eastern equatorial Pacific.

4.4. Trend Assessment: Model versus Observed

In this section, the distributions of internal climate variability from the model control runs (Plate 4) and the ensemble mean response to greenhouse gases plus sulfate aerosol (G+S) forcing (Plates 3 and 9) are used to assess where the observed trends over the period 1949-1997 (Plates 3a-3c) are or are not consistent with the models. The observed trends are here based on regridded data, where the observed 5°x5° data for each month has been regridded onto the R15 or R30 model grid using an area-weighted box-averaging technique requiring coverage of greater than 50% of the surrounding grid box for the grid point to be nonmissing for a given month.

Plate 5 shows the assessment of the observed trend in annual mean SST and LSAT (1949-1997) in comparison to simulated internal climate variability only. The regions with gray shading denote the grid boxes where the observed trend is within the range simulated in the R30 (Plates 5a and 5b) or R15 (Plates 5c and 5d) control run. The "range" referred to here is either the range from the 5th percentile to the 95th percentile of simulated trends for the grid point (i.e., the 5th/ 95th percentile range in Plates 5a and 5c), or the range from the minimum to the maximum simulated trend for the grid point (i.e., the Min/Max range in Plates 5b and 5d). The remaining color-shaded (nongray) regions and contours denote the observed trends for locations where the trends are outside the 5th/ 95th percentile range or the min/max range of the model-simulated internal climate variability. According to the model simulations the warming over large regions of the globe has been unusual (above the 95th percentile of the simulated trends). These regions include the Indian Ocean, much of Asia, the eastern tropical Pacific, the South Atlantic, and western Canada. The most unusual warming trends (i.e., above the simulated maximum) are over the Indian Ocean and parts of Asia, the eastern tropical Pacific, and the South Atlantic. Of these regions, the equatorial part of the eastern tropical

Table 1. Percent of Area Analyzed Where the Observed SST/LSAT Trend is Outside the Range of Trends Simulated by the Climate Model (< 5th Percentile or Absolute Minimum, or > 95th Percentile or Absolute Maximum)

	Annual					JJA					DJF				
	Min	5th Perc	95th Perc	Max	t Test	Min	5th Perc	95th Perc	Max	t Test	Min	5th Perc	95th Perc	Max	t Test
R30 (1949-1997)															
Internal variability	0.9	3.9	46.1	23.8			3.9	33.9	14.5		0.5	3.9	32.5	11.4	
G+S	12.7	29.2	7.6	1.5	30.7	8.6	24.7	6.9	1.6	25.3	7.5	21.8	7.2	2.1	23.5
R30 (1900-1997)															
Internal variability	0.3	0.7	78.3	60.8		0.3	1.2	63.8	42.5		0.2	0.4	61.8	37.8	
G+S	9.7	20.3	14.5	6.7	26.7	7.1	18.1	11.4	4.6	20.8	5.3	17.2	12.2	6.5	21.1
R30 (1910-1945)															
Internal variability	0.2	1.1	50.1	16.7		0.2	1.2	40.4	11.1		0.4	2.9	34.8	9.1	
G+S	0.7	3.0	30.1	6.0	27.2	0.4	3.2	25.0	4.9	23.1	1.1	6.1	25.0	4.5	24.6
R15 (1949-1997)															
Internal variability	0.1	4.0	48.4	15.3											
G+S	8.1	32.3	7.1	0.3	37.8										

Perc, percentile.

The columns labeled "t Test" show the percent of area analyzed where the G+S ensemble mean trends are significantly different from the observed trend at the 0.10 significance level according to a local two-sided t test. The internal variability results are based on control runs of length 900 years (R30) or 10,000 years (R15). The G+S results are based on ensemble-mean trends from R30 ($n=5$) or R15 ($n=9$) G+S experiments combined with the internal variability from the 900-year or 10,000-year control runs. The percent of globe analyzed (annual mean trends) was 73% for 1949-97, 33% for 1900-97, and 37% for 1910-45. Results are shown for annual means and the 3-month seasons of June-August and December-February.

Pacific in particular was noted in section 3.2 as having deficient (very low) levels of interannual variability.

The percent of area (nonmissing regions only) meeting various criteria is summarized in Table 1. The results indicate that the warming trends are above the simulated 95th percentile range in about half (46-48%) of the available areas according to the R30 and R15 control models, respectively, and above the maximum simulated value in 15% (R15) to 24% (R30) of the area analyzed. Note that the much longer control simulation for the R15 model allows for a greater possibility of simulating a few particularly extreme cases compared to the much more limited sample from the R30 model; this difference contributes to the above difference of 15% versus 24% for the maximum value results. In contrast, the cooling trends are below the 5th percentile range in only about 4% of the area and below the minimum simulated value over only about 0.1 to 0.9% of the area.

In Plate 6 is the assessment of the observed trends in comparison to trends from the G+S forcing experiments. For the percentile comparisons the "G+S forcing" trend distributions are obtained by linearly combining the internal variability distribution results from the control runs with the ensemble mean trends from the G+S experiments. Thus, for example, the 95th percentile of the trends at each grid point from the control simulation (Plate 4) is added to the ensemble mean net trend for that grid point from the G+S experiments (Plates 3 and 9) to obtain a 95th percentile value for the G+S trend distribution. ("Net" trend from the G+S experiment refers to the fact that the G+S experiment trends are adjusted by subtraction of the long-term underlying control run trends.) This construction of the G+S trend distribution assumes that the climate change signal and the internal variability are linearly additive. The assumption will be violated to the extent that the radiatively forced climate change analyzed here alters the model internal variability characteristics. However, there is little indication from our previous GCM studies of surface temperature variability [e.g., Knutson et al., 1997] and other sensitivity tests (not shown) that such an effect would be important for the present analysis.

The G+S forcing assessment in Plate 6 indicates that the observed trend is smaller or more negative than the lowest/coolest simulated trend in about 8% (R15) or 13% (R30) of the areas (Table 1). The main areas where this occurs tend to be oceanic

regions: over the northern midlatitude Pacific, parts of the North Atlantic basin, and in the Southwest Pacific near Australia and New Zealand. The fraction of analyzed areas with observed warming trends larger than any simulated in the models is comparatively minor (0.3% for R15 and 1.5% for the R30).

A caveat to the G+S trend assessment described above is that it does not account for the fact that both the observed trends and the G+S trends contain uncertainty, due to the small sample sizes ($n=5$ for R30; $n=9$ for R15; $n=1$ for the observations). To address this issue, a local two-sample t test was used to compare the ensemble mean G+S trend (sample size of 5 or 9) with the observed trend (sample size of 1) at each grid point. The trend distributions from the control simulations were used to provide the population variance for the t tests (M. Allen, personal communication, 1999). The dark contours in Plate 6 (see also Table 1) encompass regions where the null hypothesis for the test (H_0: that the G+S trends and observed trend have the same mean value) is rejected at the 0.10 or 0.05 level, according to a two-sided t test. In each case (Table 1) the percent of area where H_0 is rejected at the 0.10 level is similar to, though slightly smaller than, the percent of area outside the 5th to 95th percentile range (e.g., for R30 annual means, 30.7% versus 36.8%). For the R30 model, H_0 is rejected over a smaller percent of area than for R15 model (e.g., 30.7% versus 37.8%), as might be expected given the larger control run trends in some regions (Plate 4) and the smaller ensemble size ($n=5$) for the R30. The local t-test approach is not used here for the internal variability assessment since the underlying trend in the control run is known (~0) and is not being estimated from a relatively small sample as in the G+S case.

An issue in performing a series of local significance tests on fields which have spatial correlations (finite spatial degrees of freedom) is whether the percent of area rejecting the null hypothesis is significant from a global standpoint [e.g., Livezey and Chen 1983]. To confirm that the t-test results cited above are globally significant, we performed a series of Monte Carlo resampling experiments using the 900-year R30 control run data sets to derive synthetic "observed" trend maps and small ($n=5$) sample-sized "ensemble" trend maps, repeating the local t-test calculations described above on each synthetic sample derived from the control run. The resampling was repeated 10,000 times to generate a distri-

bution of percent area where H_0 is rejected on the basis of chance alone. The 95th percentile of this generated distribution was 22.7% for the R30 $G+S$ annual case, using a significance level of 0.10 for the local tests. Since the percent area rejecting H_0 at the 0.10 level using the actual observations and R30 $G+S$ ensemble was higher (30.7%), we conclude that according to these tests the differences between the $G+S$ ensemble trend and the observed trend exceed that expected due to chance alone at the 0.05 global significance level.

In summary, these assessment results indicate that the $G+S$ simulation trends are in better overall agreement with the observed trends than are the control-only simulated trends, although the $G+S$ simulations still have significant regional discrepancies versus observations.

The seasonal dependence of the above assessments was examined for the R30 simulations using 3-month mean (DJF and JJA) observations (Plate 3). The seasonal assessment results are shown for the 5th/95th percentile range of trends as simulated by the model with no external forcing (Plates 7a and 7b) or with $G+S$ forcing (Plates 7c and 7d); local t-test results are also presented in the case of the $G+S$ forcing. The results are generally consistent with the annual mean results. Over a large fraction of the area analyzed (33-36%) the observed regional warming exceeds the 95th percentile value simulated by the model with no external forcing (Plates 7a and 7b). With the $G+S$ radiative forcing, the fraction of area with disagreement between model and observation is smaller, although the model again fails to simulate some of the stronger observed regional cooling trends (Plates 7c and 7d). The fractional areas cited above are typically somewhat smaller in the seasonal analysis than in the annual mean analysis (Plate 6, Table

1). However, the differences in the seasonal analysis were still globally significant at the 0.05 level, according to the Monte Carlo resampling procedure described above. Similar results to those presented here were obtained for the transition seasons of March-May and September-November (not shown).

5. Dependence on Selected Initial Year

The assessments discussed above were done using the period 1949-1997. To explore the dependence of the results on the time period selected, the trend results are reexamined using different starting dates (all using 1997 as the end date). Figure 2 shows the global temperature linear trends to 1997 for start years varying from 1880 to 1978. Results are compared for the observed record and for the five R30 $G+S$ series from Plate 1. The R30-simulated trend values in Figure 2 have been adjusted by +0.020 K 100 yr^{-1} to account for the long-term global mean cooling trend of that magnitude in the R30 control experiment over years 101-700 (roughly coinciding with the model years used in the $G+S$ experiments).

For the 1949-1997 period each of the five R30 simulated global trends exceeds the observed trend (circle). There is a general tendency among the R30 ensemble members (Figure 2) for the simulated global trends to 1997 to exceed the observed global trends, particularly for starting years spanning the period from about 1915 to 1960. This apparently is related to the pronounced warm period in the observations centered near 1940 (Plate 1). The observed trends to 1997 for start years earlier (prior to 1910) or later (1960s and 1970s) in the record are in better agreement with the simulated distribution. The R30 $G+S$ experiment 3 is in closest

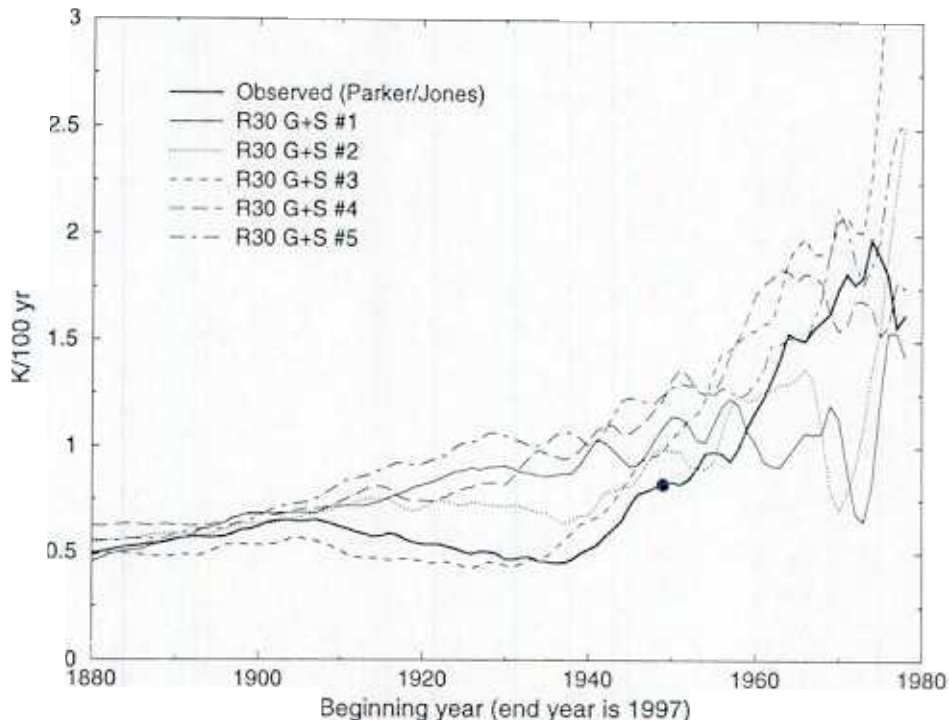## Linear Trends to 1997 of Global Temperature

Figure 2. Linear trends values (K 100 yr^{-1}) of the global temperature series for each member (n=5) of the R30 $G+S$ ensemble (Plate 1) as a function of start year, varying from 1880 to 1978, and using 1997 as the end year for all cases. The circle highlights the 1949-1997 value for the observations.

Model Assessment of Observed Trends − No External Forcing

Observed Trends vs. R30 Model (900 yr)
a) Outside 5th/95th percentile range

Observed Trends vs. R15 Model (10,000 yr)
c) Outside 5th/95th percentile range

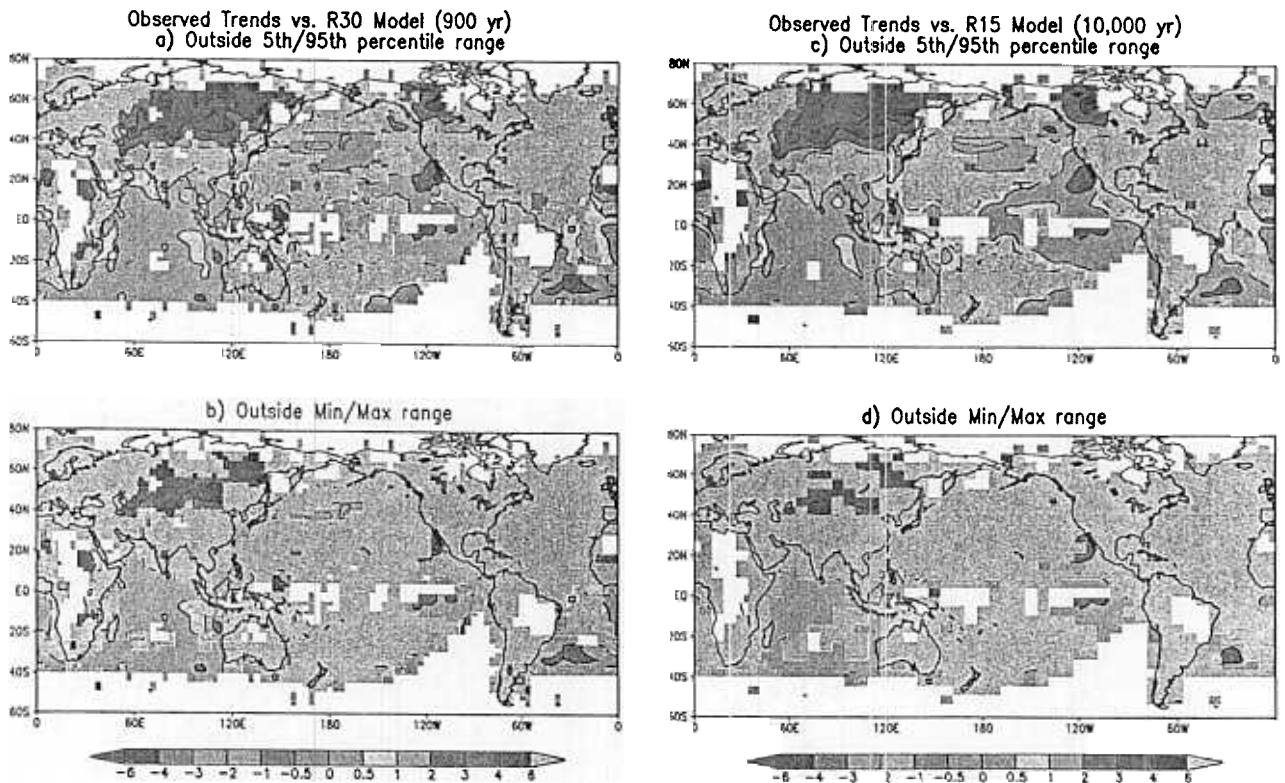b) Outside Min/Max range

d) Outside Min/Max range

Plate 5. Gray shading denotes grid boxes where the observed trend in annual mean SST/LSAT is within the (a, c) 5th to 95th percentile range or within the (b, d) minimum to maximum range as simulated in the (a, b) R30 or (c, d) R15 control run. Color-shaded (nongray) contoured regions denote grid boxes where the observed trend is outside of the simulated ranges, with the color and contouring indicating the magnitude of the observed trend (K 100 yr^{-1}) as in Plate 3. Contours and color shading conventions are as in Plate 3.

agreement with the observed trends to 1997 for start years early in the 20th century (1910-1940). As shown in Plate 1, this experiment simulates a pronounced warming trend from about 1910 to 1945 similar to the observations.

The regional trend assessment has been repeated (as in Plates 5-7) for both the 98-year period 1900-1997 and the 36-year period 1910-1945 using the R30 model simulations (figures not shown; see Table 1). These analyses cover about 33 and 37% of the globe, respectively, or about half the coverage of the 1949-97 analysis (73%) based on the same screening criteria. Coverage tended to be inadequate over the open regions of the Pacific, South Indian, and South Atlantic Oceans as well as remote land regions. The 1900-1997 assessment (Table 1) indicates that the 98-year warming trends since 1900 are above the maximum simulated value (internal variability) over about 60% of the area analyzed. In the G+S forced response assessment for 1900-1997 the percent of area with observed trends lower than the simulated minimum is roughly comparable to the percent area with observed trends greater than the simulated maximum; in that sense there appears to be less bias than in the 1949-1997 G+S assessment. In the 1910-1945 assessment (i.e., the period of strong early century global warming, Plate 1) there is some tendency in the observations for unusually strong regional warmings compared with the G+S experiments for that period (Table 1).

The midlatitude North Pacific cooling in the 1949-1997 trend assessment (e.g., Plate 3) was identified as one of the most pronounced inconsistencies between the model G+S experiments

and the observations. This region is notable for relatively strong interdecadal SST variability as shown or investigated in a number of previous observational and modeling studies [e.g., Trenberth, 1990; Trenberth and Hurrell, 1994; Graham, 1994; Miller et al., 1994; Deser and Blackmon, 1995; Deser et al., 1996; Latif and Barnett, 1996; Zhang et al., 1997; Zhang and Levitus, 1997; Nakamura et al., 1997; Kachi and Nitta, 1997; Knutson and Manabe, 1998; Meehl et al., 1998; White and Cayan, 1998; Barnett et al., 1999]. Furthermore, the pattern of associated North Pacific midlatitude cooling/tropical Pacific warming is reminiscent of a remote response (via an "atmospheric bridge") of North Pacific SSTs to warm interannual SST anomalies in the tropical Pacific [Alexander, 1990; Lau and Nath, 1996]. Here we are concerned with the nature of the cooling trend in the North Pacific over the period 1949-1997 and whether such a cooling trend has been present throughout the 20th century. An area-averaged SST index (30°-45°N, 155°E-145°W) was examined for this purpose. Because of the relatively sparse data coverage in this region prior to 1949, it was necessary to use indices derived from the LDEO (Kaplan et al. [1998] data through 1991) and GISST2.3b [Rayner et al., 1996] SST reconstructions. These indices (not shown) show the strong negative trend (1949-1997 or 1949-1991) but much smaller trends (closer to zero or even positive) using start years from earlier in the century. The period 1943-1956 appears as an unusually warm epoch in the midlatitude North Pacific, according to the reconstructions, leading to relatively large cooling trends over the period 1949-1997. Nonetheless, the underlying cause or
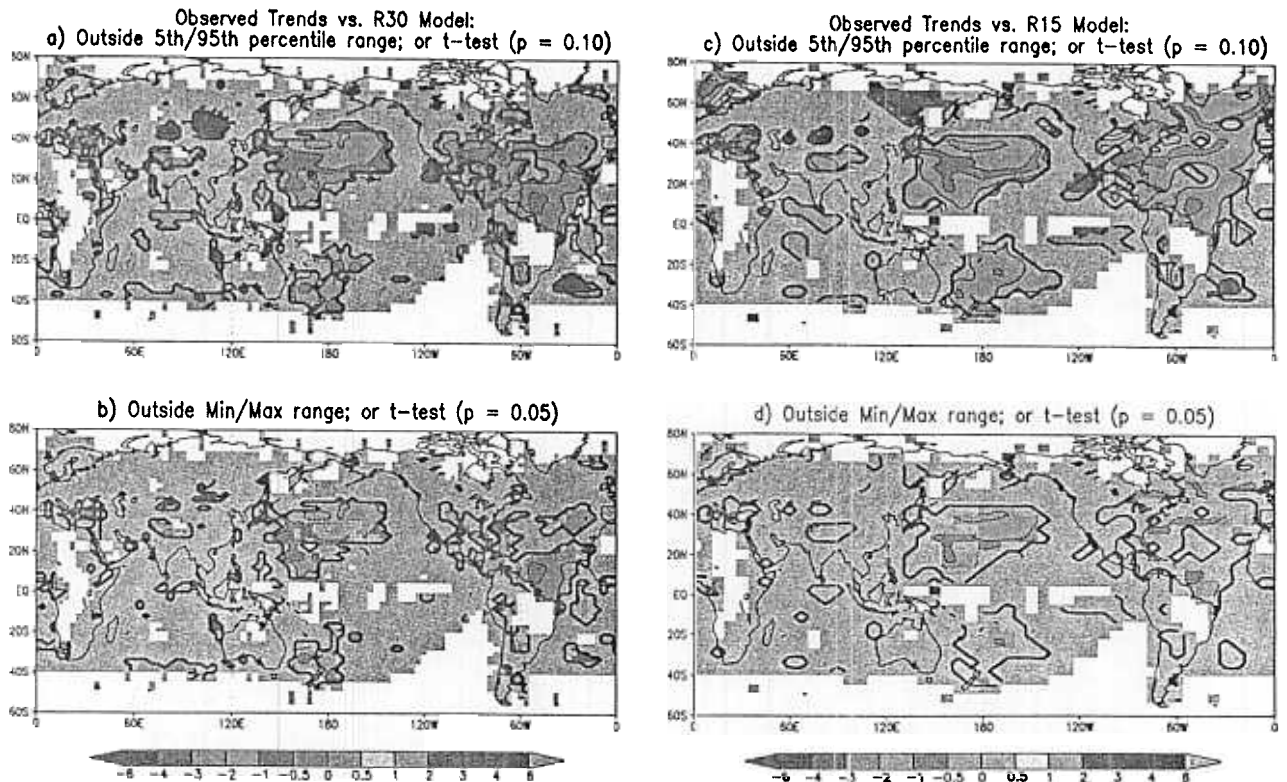
Observed Trends vs. R30 Model:
a) Outside 5th/95th percentile range; or t-test (p = 0.10)

Observed Trends vs. R15 Model:
c) Outside 5th/95th percentile range; or t-test (p = 0.10)

b) Outside Min/Max range; or t-test (p = 0.05)

d) Outside Min/Max range; or t-test (p = 0.05)

Plate 6. As in Plate 5 except that the observed trends are compared with the trend 5th/95th percentiles or max/min ranges from the model control runs (e.g., as illustrated by the 95th percentiles in Plate 4) to which the ensemble mean trends (1949-1997) from the *G+S* experiments (e.g., Plates 3d and 9) have been added. The additional heavy contours encompass regions where the observed trend and ensemble mean *G+S* trend are significantly different at the 0.10 (a, c) or 0.05 (b, d) significance level according to a local two-sided *t* test (see text).

R30 Model Assessment: Seasonal Observed Trends

No External Forcing:
a) JJA
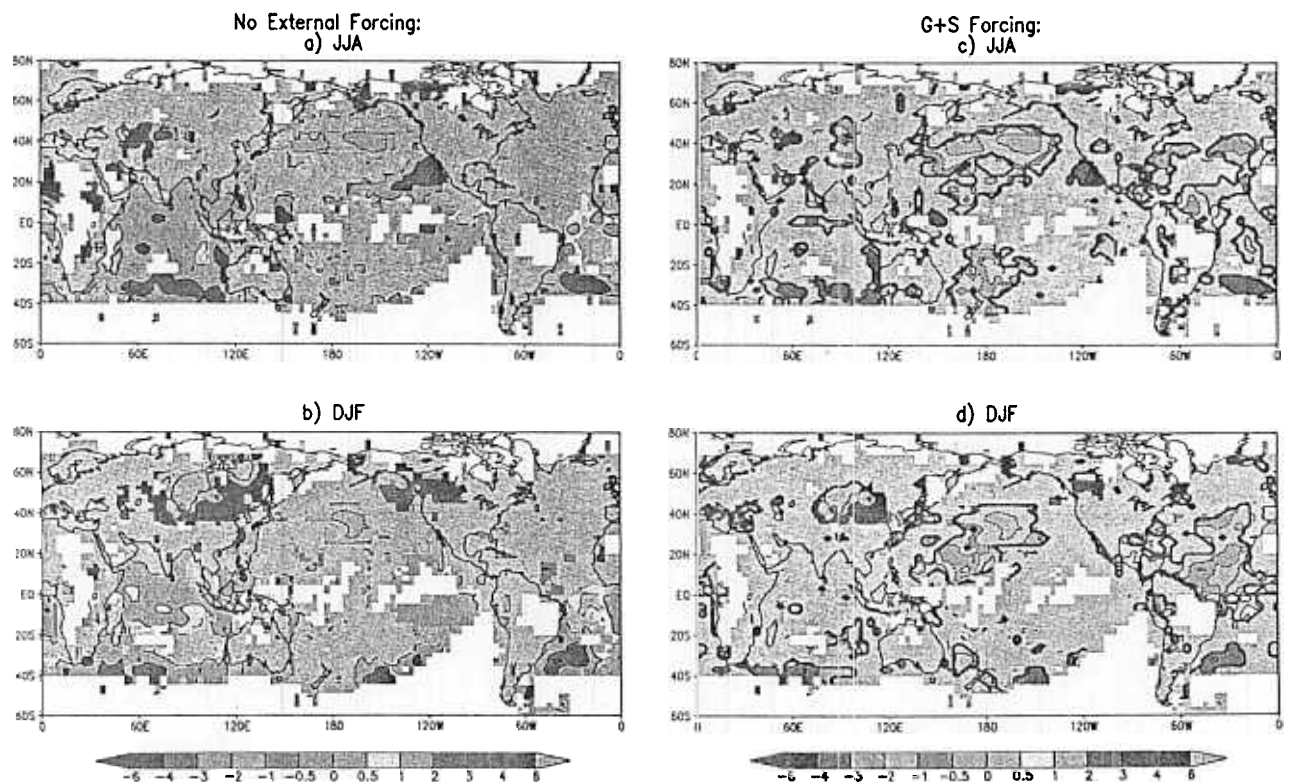
G+S Forcing:
c) JJA

b) DJF

d) DJF

Plate 7. As in Plates 5 and 6 except that seasonal mean trends (JJA for a, c and DJF for b, d) from the observations are assessed based on the R30 simulations. Gray shading and light contours denote grid boxes where the observed trends are within the 5th to 95th percentile range of the (a, b) R30 control run distribution, or (c, d) R30 control run distribution to which the ensemble mean trend from the R30 *G+S* integrations has been added. For Plates 7c and 7d the additional dark contours depict local *t* test results (0.10 significance level) as in Plate 6.

R30 G+S #1

R30 G+S #2

R30 G+S #3

R30 G+S #4

R30 G+S #5

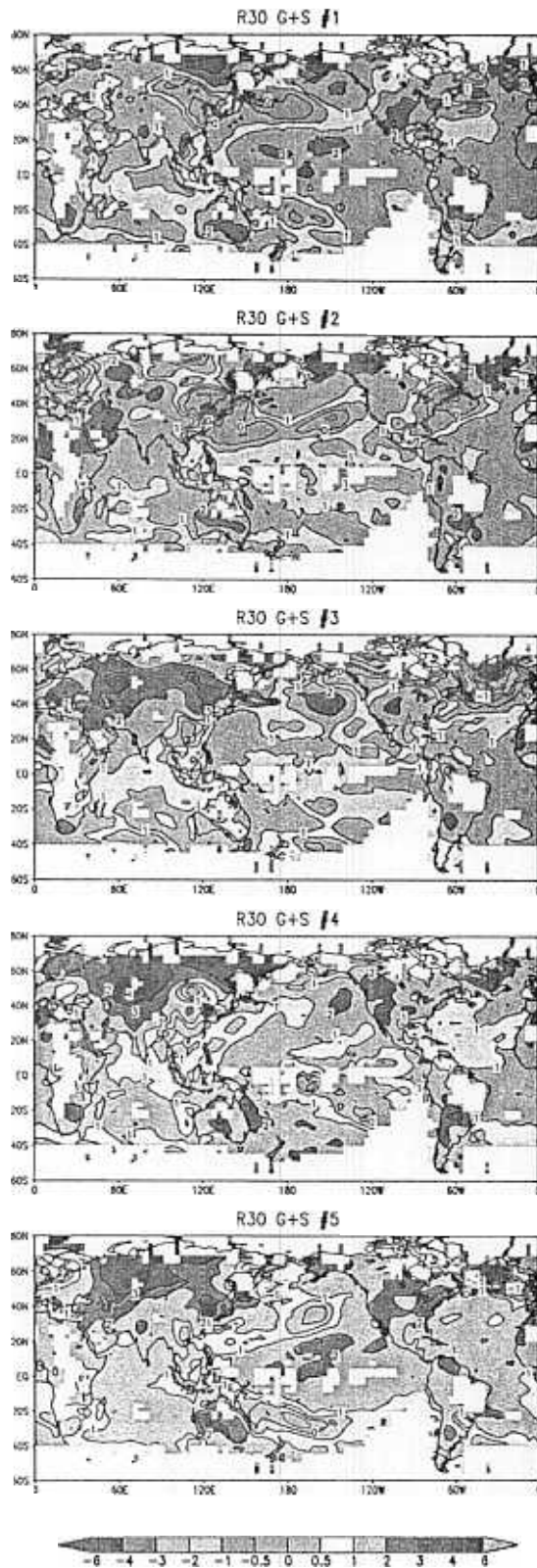-6 -4 -3 -2 -1 -0.5 0 0.5 1 2 3 4 6

Plate 8. Linear trends of annual mean SST/LSAT for the period 1949-1997 for each R30 model G+S ensemble member. The results at each grid point are computed as a net trend (G+S experiment trend minus the trend for years 101-700 of the control run) at that grid point. Contour interval of 1 K 100 yr^{-1} also with +/- 0.5 K 100 yr^{-1} contours shown.

causes of the inconsistency between the G+S experiments and the observed trends (1949-1997) in this region remains to be determined.

6. Summary and Conclusions

In terms of global mean temperature the R30 greenhouse gas plus aerosol (G+S) experiments in this report, like those previously presented for the R15 model [*Haywood et al.*, 1997; *Dixon and Lanzante*, 1999], simulate a 20th century warming trend broadly similar to that observed. One of the five R30 integrations tracks the multidecadal variability of the observed record particularly well, including the pronounced warming from about 1910 to 1945, the relatively stable temperatures from 1945 to 1980, and relatively rapid warming since about 1980.

The main focus of the present study is to document, from a regional perspective, where model-simulated trends in SST/LSAT are consistent with available observations and where they are not. The assessment is done for two sets of model results each from the R15 and R30 models: (1) control run (internal) variability and (2) control run variability plus the ensemble-mean-simulated response to greenhouse gases plus aerosols (G+S). The analysis focuses on the relatively data-rich period 1949-1997, although other periods have been considered.

As a preliminary step, the magnitude of the coupled model (detrended) internal variability is compared to observed (detrended) variability. The model global temperature spectra and local interannual variability maps resemble those of the observations, although shortcomings are apparent, in particular with the simulated El Niño variability. In the equatorial Pacific the R30 model has more realistic levels of interannual SST variability than the R15, but the R30 model variability is still deficient in the eastern equatorial Pacific, and its dominant timescale (~8-9 years) is about a factor of 2 longer than for the observed El Niño. The local magnitude of the simulated unforced interannual variability (Plate 2) appears excessive in the high-latitude North Atlantic (R30 model) and over subtropical land regions in both models.

In the regional trend assessment the observed trends (1949-1997) are first compared with the internally generated 49-year trends from the model control runs. The observed trends are unusually large (above the 95th percentile of simulated trends) for nearly half of the regions available for the comparison. According to the models the regions with the most unusual warming rates (i.e., above the simulated internal variability maximum) include much of the Indian Ocean and parts of Asia, the eastern tropical Pacific, and the South Atlantic. For the G+S forcing assessment, the observed trends are compared with the ensemble mean response to G+S forcing combined with the simulated internal variability from the control simulations. Better agreement is found between the ensemble mean G+S trends and the observed trends than between the model internal variability alone and the observed trends. However, the G+S trends are still significantly different from the observed trends (according to a local *t* test using the 0.10 significance level) over about 30% of the areas analyzed. In particular, the G+S experiments fail to simulate the large-scale cooling trends observed over the midlatitude North Pacific and parts of the Southwest Pacific and North Atlantic basins, even accounting for the effects of internal variability "noise" in the models.

The apparent "warm bias" in the G+S forcing assessment (1949-1997) is much reduced for 98-year trends beginning from 1900 and is even reversed ("cool bias") for trends over the period 1910-1945 (Table 1). This behavior apparently arises from the pronounced
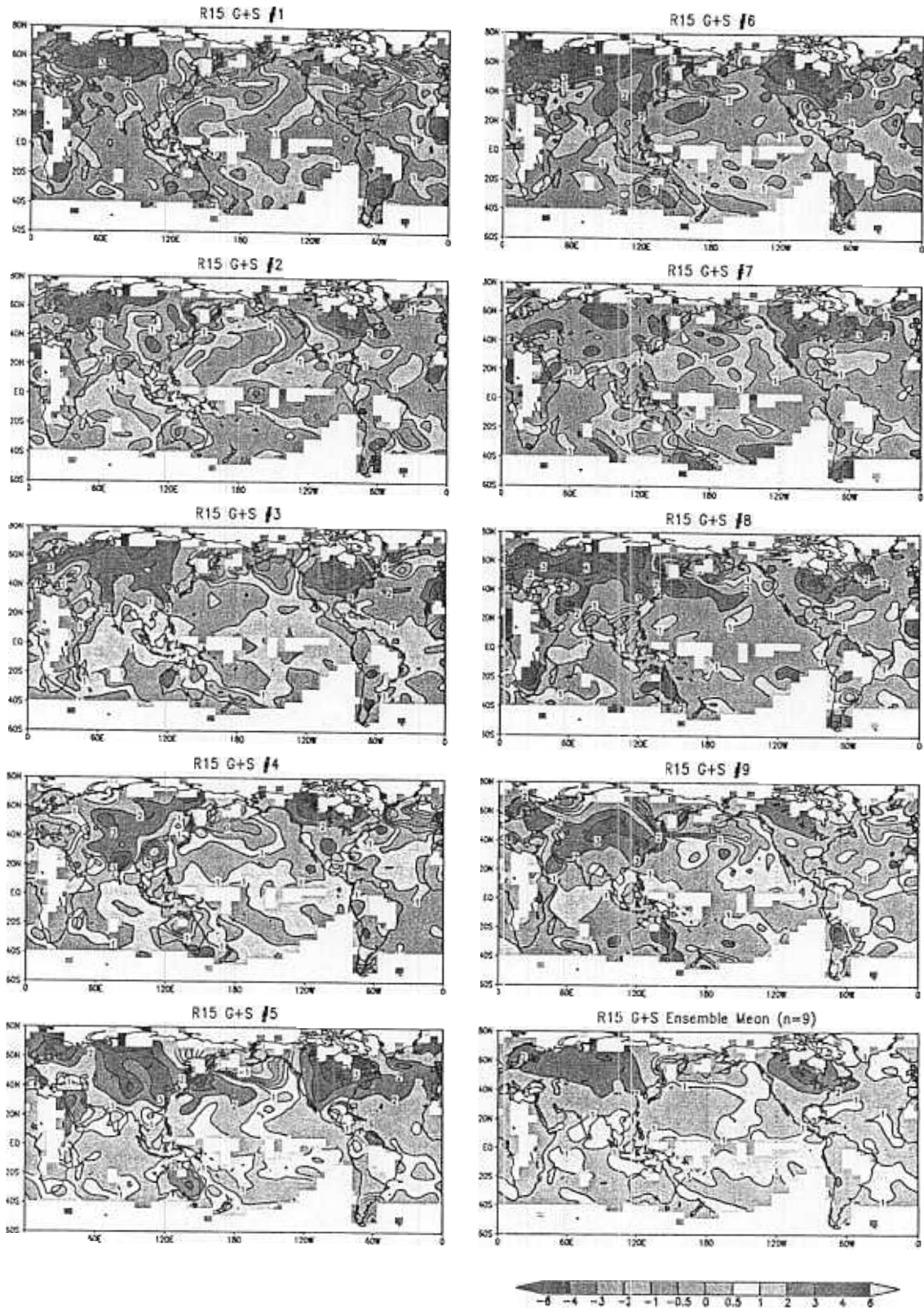
Plate 9. As in Plate 8 except for the R15 model. The results at each grid point are computed as a net trend (*G+S* experiment trend minus the 1500 year control run trend) at that grid point. The ensemble mean trend is shown in the bottom right panel of the diagram.

warming (of uncertain origin) in the observations during the early 20th century (1910-1945). Although the warm epoch around 1940 is well simulated, in terms of global temperature, in one of the R30 G+S ensemble members (experiment 3), it is not simulated in several of the other R30 G+S experiments (Plate 1) and thus not in the ensemble mean. The more limited data coverage in the first half century hinders attempts to resolve many regional details from these earlier periods.

For the G+S forcing assessment the differences between model and observation may be due to deficiencies in (1) the specified radiative forcings, (2) the simulated responses to specified radiative forcings, (3) the simulations of internal climate variability, or (4) the observed temperature records. As examples, the simulated internal multidecadal variability in certain regions (e.g., the North Atlantic or midlatitude North Pacific) could be too small, or the observed trends could have resulted from an unusual phasing of more than one internal variability mode. The simulated response to climate forcing could be in error in terms of the global mean or in terms of the degree of spatial structure in the response. The specified radiative forcing shortcomings include several neglected forcings (both natural and anthropogenic) as well as possible errors in the space/time history of the G+S forcing used. For example, potentially important neglected radiative forcings include indirect aerosol effects [Schimel et al., 1996], volcanic aerosols [North and Stevens, 1998], and solar variability [Tett et al., 1999]. At present, we cannot determine which of the above four potential deficiencies are most important for explaining the inconsistencies between the simulated and the observed trends.

The internal (unforced) variability trend assessment involves fewer assumptions than the G+S forcing assessment, as it does not depend upon factors (1) and (2) above. Thus assuming the model-simulated multidecadal internal variability (factor 3) and observed trend estimates (factor 4) are not too deficient, the trends in SST/LSAT analyzed here are very unlikely to be due to internal climate variability alone. This result is consistent with a number of previous studies suggesting that the observed global warming exceeds that expected from internal variability alone and strongly suggests a role for a sustained positive thermal forcing, such as increased greenhouse gases, in the 20th century warming.

Appendix: Geographical Distribution of Trends in the Individual G+S Experiments

A comparison of the 1949-97 trends from different G+S ensemble members provides an indication of the effect that different ocean/atmosphere initial conditions can have on regional features of the 1949-1997 simulated trend maps. The 1949-1997 trends for each of the R30 G+S experiments (annual data) are shown in Plate 8, while those for the R15 G+S experiments (n=9) are shown in Plate 9. The ensemble mean trend for the R15 experiments is shown in the bottom right panel of Plate 9; for the R30 the ensemble mean was shown in Plate 3d.

In the R30 simulations (Plate 8), three of the ensemble members (3, 4, and 5) show a pronounced warming over Asia; another (experiment 2) shows rather pronounced regions of cooling over eastern Asia, western Europe, eastern North America, and parts of the North Pacific; two (experiments 3,5) show particularly strong cooling in the high-latitude North Atlantic. None of the ensemble members show the degree of cooling seen in the observed record over the midlatitude North Pacific or the midlatitude and tropical North Atlantic (Plate 3a).

Similar regional variability of features is seen for the R15 ensemble members (Plate 9). None of the R15 ensemble members show cooling over the northern midlatitude Pacific or North Atlantic as pronounced as in the observed trends. The more uniform ensemble mean R15 trends in Plate 9 (bottom right panel) show warming over almost all regions and enhanced warming over northern Asia and North America.

The results illustrate how regional details of half-century trend patterns can vary substantially due to internal climate variability. A similar result has been shown previously by Kim and North [1995] using a simpler climate model framework.

References

Alexander, M. A., Simulation of the response of the North Pacific Ocean to the anomalous atmospheric circulation associated with El Niño, Clim. Dyn., 5, 53-65, 1990.

Barnett, T. P., G. C. Hegerl, B. Santer, and K. Taylor, The potential effect of GCM uncertainties and internal atmospheric variability on anthropogenic signal detection, J. Clim., 11, 659-675, 1998.

Barnett, T. P., D. W. Pierce, M. Latif, D. Dommenget, and R. Saravanan, Interdecadal interactions between the tropics and midlatitudes in the Pacific basin, Geophys. Res. Lett., 26, 615-618, 1999.

Barnett, T. P., and M. E. Schlesinger, Detecting changes in global climate induced by greenhouse gases, J. Geophys. Res., 92, 14,722-14,780, 1987.

Bell, T. L., Optimal weighting of data to detect climatic change: Application to the carbon dioxide problem, J. Geophys. Res., 87, 11,161-11,170, 1982.

Bell, T. L., Theory of optimal weighting of data to detect climatic change, J. Atmos. Sci., 43, 1694-1710, 1986.

Bloomfield, P., Trends in global temperature, Clim. Change, 21, 1-16, 1992.

Delworth, T. L., North Atlantic interannual variability in a coupled ocean-atmosphere model, J. Clim., 9, 2356-2375, 1996.

Delworth, T., and S. Manabe, The influence of soil wetness on near-surface atmospheric variability, J. Clim., 2, 1447-1462, 1989.

Deser, C., and M. L. Blackmon, On the relationship between Tropical and North Pacific sea surface temperature variations, J. Clim., 8, 1677-1680, 1995.

Deser, C., M. A. Alexander, and M. S. Timlin, Upper-ocean thermal variations in the North Pacific during 1970-1991, J. Clim., 9, 1840-1855, 1996.

Dixon, K. W., and J. R. Lanzante, Global mean surface air temperature and North Atlantic overturning in a suite of coupled GCM climate change experiments, Geophys. Res. Lett., 26, 1885-1888, 1999.

Ghil, M., and R. Vautard, Interdecadal oscillations and the warming trend in global temperature time series, Nature, 350, 324-327, 1991.

Graham, N. E., Decadal-scale climate variability in the tropical and North Pacific during the 1970s and 1980s: Observations and model results, Clim. Dyn., 10, 135-162, 1994.

Hansen, J., et al., Forcings and chaos in interannual to decadal climate change, J. Geophys. Res., 102, 25,679-25,720, 1997.

Hasselmann, K., On the signal-to-noise problem in atmospheric response studies, in Meteorology Over the Tropical Oceans, edited by D.B. Shaw, R. Meteorol. Soc., pp. 251-259, 1979.

Haywood, J. M., R. J. Stouffer, R. T. Wetherald, S. Manabe, and V. Ramaswamy, Transient response of a coupled model to estimated changes in greenhouse gas and sulfate concentrations, Geophys. Res. Lett., 24, 1335-1338, 1997.

Hegerl, G. C., K. Hasselmann, U. Cubasch, J. F. B. Mitchell, E. Roeckner, R. Voss, and J. Waszkewitz, Multi-fingerprint detection and attribution analysis of greenhouse gas, greenhouse gas-plus-aerosol and solar forced climate change, Clim. Dyn., 13, 613-634, 1997.

Intergovernmental Panel on Climate Change (IPCC), Climate Change 1992: The Supplementary Report to the IPCC Scientific Assessment,

edited by J. T. Houghton, B. A. Callender, and S. K. Varney, WMO/ UNEP, Cambridge Univ. Press, New York, 1992.

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, The second Hadley Centre coupled ocean-atmosphere GCM: Model description, spinup and validation, Clim. Dyn., 13, 103-134, 1997.

Jones, P. D., Hemispheric surface air temperature variations: A reanalysis and an update to 1993. J. Clim., 7, 1794-1802, 1994.

Jones, P. D., and G. C. Hegerl, Comparisons of two methods of removing anthropogenically related variability from the near-surface observational temperature field, J. Geophys. Res., 103, 13,777-13,786, 1998.

Kachi, M., and T. Nitta, Decadal variations of the global atmosphere-ocean system, J. Meteorol. Soc. Jpn., 75, 657-675, 1997.

Kaplan, A., M. A. Cane, Y. Kushnir, A. Clement, M. Blumenthal, and B. Rajagopalan, Analyses of global sea surface temperature 1856-1991, J. Geophys. Res., 103, 18,567-18,589, 1998.

Karoly, D. J., J. A. Cohen, G. A. Meehl, J. F. B. Mitchell, A. H. Oort, R. J. Stouffer, and R. T. Wetherald, An example of fingerprint detection of greenhouse climate change, Clim. Dyn., 10, 97-105, 1994.

Kim, K. Y., and G. R. North, Regional simulations of greenhouse warming including natural variability, Bull. Am. Meteorol. Soc., 76, 2171-2178, 1995.

Knutson, T. R., and S. Manabe, Model assessment of decadal variability and trends in the tropical Pacific Ocean, J. Clim., 11, 2273-2296, 1998.

Knutson, T. R., S. Manabe, and D. Gu, Simulated ENSO in a global coupled ocean-atmosphere model: Multidecadal amplitude modulation and CO$_2$ sensitivity, J. Clim., 10, 138-161, 1997.

Latif, M., and T. P. Barnett, Decadal climate variability over the North Pacific and North America: Dynamics and predictability, J. Clim., 9, 2407-2423, 1996.

Lau, N.-C., and M. J. Nath, The role of the "atmospheric bridge" in linking tropical Pacific ENSO events to extratropical SST anomalies, J. Clim., 9, 2036-2057, 1996.

Lau, N.-C., S. G. H. Philander, and M. J. Nath, Simulation of ENSO-like phenomena with a low-resolution coupled GCM of the global ocean and atmosphere, J. Clim., 5, 284-307, 1992.

Leroy, S. S., Detecting climate signals: Some Bayesian aspects, J. Clim., 11, 640-651, 1998.

Livezey, R. E., and W. Y. Chen, Statistical field significance and its determination by Monte Carlo techniques, Mon. Weather Rev., 111, 46-59, 1983.

Manabe, S., and R. J. Stouffer, 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, 1996.

Manabe, S., R. J. Stouffer, M. J. Spelman, and K. Bryan, Transient response of a coupled ocean-atmosphere model to gradual changes of atmospheric CO$_2$, part I, Annual mean response, J. Clim., 4, 785-818, 1991.

Mann, M. E., R. S. Bradley, and M. K. Hughes, Global-scale temperature patterns and climate forcing over the past six centuries, Nature, 392, 779-787, 1998.

Meehl, G. A., J. M. Arblaster, and W. G. Strand Jr., Global scale decadal climate variability, Geophys. Res. Lett., 25, 3983-3986, 1998.

Miller, A. J., D. R. Cayan, T. P. Barnett, N. E. Graham, and J. M. Oberhuber, Interdecadal variability of the Pacific Ocean: Model response to observed heat flux and wind stress anomalies, Clim. Dyn., 10, 135-162, 1994.

Mitchell, J. F. B., T. C. Johns, J. M. Gregory, and S. F. B. Tett, Climate response to increasing levels of greenhouse gases and sulfate aerosols, Nature, 376, 501-504, 1995.

Nakamura, H., G. Lin, and T. Yamagata, Decadal climate variability in the North Pacific during recent decades, Bull. Am. Meteorol. Soc., 78, 2215-2225, 1997.

Nicholls N., G. V. Gruza, J. Jouzel, T. R. Karl, L. A. Ogallo, and D. E. Parker, Observed climate variability and change, in Climate Change 1995: The Science of Climate Change, edited by J. T. Houghton, L. G. Meira Filho, B. A. Callander, N. Harris, A. Kattenberg, and K. Maskell, Cambridge Univ. Press, New York, 1996.

North, G. A., K.-Y. Kim, S. S. P. Shen, and J. W. Hardin, Detection of forced climate signals, part I, Filter theory, J. Clim., 8, 401-408, 1995.

North, G. A., and M. J. Stevens, Detecting climate signals in the surface temperature record, J. Clim., 11, 563-577, 1998.

Pacanowski, R., K. Dixon, and A. Rosati, The GFDL Modular Ocean Model Users Guide version 1, GFDL Ocean Group Tech. Rep. 2, NOAA/Geophys. Fluid Dyn. Lab., Princeton, N.J.,1991.

Parker, D. E., P. D. Jones, A. Bevan, and C. K. Folland, Interdecadal changes of surface temperature since the late 19th century, J.Geophys. Res., 99, 14,373-14,399, 1994.

Parker, D. E., C. K. Folland, and M. Jackson, Marine surface temperature: Observed variations and data requirements, Clim. Change, 31, 559-600, 1995.

Rayner, N. A., E. B. Horton, D. E. Parker, C. K. Folland, and R. B. Hackett, Version 2.2 of the global sea-ice and sea surface temperature data set, 1903-1994, Clim. Res. Tech. Note 74. Hadley Centre for Clim. Pred. and Res., Met. Office, Bracknell, U.K., 1996.

Santer, B. D., et al., A search for human influences on the thermal structure of the atmosphere, Nature, 382, 39-46, 1996a.

Santer, B. D., K. E. Taylor, T. M. L. Wigley, J. E. Penner, P. D. Jones, and U. Cubasch, Towards the detection and attribution of an anthropogenic effect on climate, Clim. Dyn., 12, 77-100, 1995.

Santer, B. D., T. M. L. Wigley, T. P. Barnett, and E. Anyamba, Detection of climate change and attribution of causes, in Climate Change 1995: The Science of Climate Change, edited by J. T. Houghton, L. G. Meira Filho, B. A. Callander, N. Harris, A. Kattenberg, and K. Maskell, Cambridge Univ. Press, New York, 1996b.

Santer, B. D., T. M. L. Wigley, and P. D. Jones, Correlation methods in fingerprint detection studies, Clim. Dyn., 8, 265-276, 1993.

Schimel, D., et al., Radiative forcing of climate change, in Climate Change 1995: The Science of Climate Change, edited by J. T. Houghton, L. G. Meira Filho, B. A. Callander, N. Harris, A. Kattenberg, and K. Maskell, Cambridge Univ. Press, New York, 1996.

Schlesinger, M. E., and N. Ramankutty, An oscillation in the global climate system of period 65-70 years, Nature, 367, 723-726, 1994.

Smith, T. M., R. W. Reynolds, R. E. Livezey, and D. C. Stokes, Reconstruction of historical sea surface temperatures using empirical orthogonal functions, J. Clim., 9, 1403-1420, 1996.

Stott, P. A., and S. F. B. Tett, Scale-dependent detection of climate change, J. Clim., 11, 3282-3294, 1998.

Stouffer, R. J., S. Manabe, and K. Y. Vinnikov, Model assessment of the role of natural variability in recent global warming, Nature, 367, 634-636, 1994.

Tett, S. F. B., J. F. B. Mitchell, D. E. Parker, and M. R. Allen, Human influence on the atmospheric vertical structure: Detection and observations, Science, 274, 1170-1173, 1996.

Tett, S. F. B., P. A. Stott, M. R. Allen, W. J. Ingram, and J. F. B. Mitchell, Causes of twentieth-century temperature changes near the Earth's surface, Nature, 399, 569-572, 1999.

Trenberth, K. E., Recent observed interdecadal climate changes in the Northern Hemisphere, Bull. Am. Meteorol. Soc., 71, 988-993, 1990.

Trenberth, K. E., and T. J. Hoar, The 1990-1995 El Niño-Southern Oscillation event: Longest on record, Geophys. Res. Lett., 23, 57-60, 1996.

Trenberth, K. E., and J. W. Hurrell, Decadal atmosphere-ocean variations in the Pacific, Clim. Dyn., 9, 303-319, 1994.

Vinnikov, K. Y., A. Robock, R. J. Stouffer, and S. Manabe, Vertical patterns of free and forced climate variations, Geophys. Res. Lett.; 23, 1801-1804, 1996.

White, W. B., and D. R. Cayan, Quasi-periodicity and global symmetries in interdecadal upper ocean temperature variability, J. Geophys. Res., 103, 21,335-21,354, 1998.

Wigley, T. M. L., and S. C. B. Raper, Natural variability of the climate system and detection of the greenhouse effect, Nature, 344, 324-327, 1990.

Wigley, T. M. L., P. J. Jaumann, B. D. Santer, and K. E. Taylor, Relative detectability of greenhouse-gas and aerosol climate change signals, Clim. Dyn., 14, 781-790, 1998.

Zhang, R.-H., and S. Levitus, Structure and cycle of decadal variability of upper-ocean temperature in the North Pacific, J. Clim., 10, 710-727, 1997.

Zhang, Y., J. M. Wallace, and D. S. Battistsi, ENSO-like interdecadal variability: 1900-93, J. Clim., 10, 1004-1020, 1997.

T. L. Delworth, K. W. Dixon, T. R. Knutson, and R. J. Stouffer, Geophysical Fluid Dynamics Laboratory/NOAA, P. O. Box 308, Forrestal Campus, Princeton, NJ 08542. (td@gfdl.gov; kd@gfdl.gov; tk@gfdl.gov;rjs@gfdl.gov.)