A Comparison of Surface Air Temperature Variability in Three 1000-Yr Coupled Ocean–Atmosphere Model Integrations

Ronald J. Stouffer

NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey

Gabriele Hegerl

University of Washington, Joint Institute for the Study of the Atmosphere and Ocean, Seattle, Washington

Simon Tett

Hadley Centre for Climate Prediction and Research, Bracknell, United Kingdom

ABSTRACT

This study compares the variability of surface air temperature in three long coupled ocean–atmosphere general circulation model integrations. It is shown that the annual mean climatology of the surface air temperatures (SAT) in all three models is realistic and the linear trends over the 1000-yr integrations are small over most areas of the globe. Second, although there are notable differences among the models, the models' SAT variability is fairly realistic on annual to decadal timescales, both in terms of the geographical distribution and of the global mean values. A notable exception is the poor simulation of observed tropical Pacific variability. In the HadCM2 model, the tropical variability is overestimated, while in the GFDL and HAM3L models, it is underestimated. Also, the ENSO-related spectral peak in the globally averaged observed SAT differs from that in any of the models. The relatively low resolution required to integrate models for long time periods inhibits the successful simulation of the variability in this region. On timescales longer than a few decades, the largest variance in the models is generally located near sea ice margins in high latitudes, which are also regions of deep oceanic convection and variability related to variations in the thermohaline circulation. However, the exact geographical location of these maxima varies from model to model. The preferred patterns of interdecadal variability that are common to all three coupled models can be isolated by computing empirical orthogonal functions (EOFs) of all model data simultaneously using the common EOF technique. A comparison of the variance each model associated with these common EOF patterns shows that the models generally agree on the most prominent patterns of variability. However, the amplitudes of the dominant modes of variability differ to some extent between the models and between the models and observations. For example, two of the models have a mode with relatively large values of the same sign over most of the Northern Hemisphere midlatitudes. This mode has been shown to be relevant for the separation of the temperature response pattern due to sulfate aerosol forcing from the response to greenhouse gas forcing. This indicates that the results of the detection of climate change and its attribution to different external forcings may differ when unperturbed climate variability in surface air temperature is estimated using different coupled models. Assuming that the simulation of variability of the global mean SAT is as realistic on longer timescales as it is for the shorter timescales, then the observed warming of more than 0.5 K of the SAT in the last 110 yr is not likely to be due to internally generated variability of the coupled atmosphere–ocean–sea ice system. Instead, the warming is likely to be due to changes in the radiative forcing of the climate system, such as the forcing associated with increases in greenhouse gases.

1. Introduction

The study of the variability of surface air temperature (SAT) proves to be very important, especially in terms of detecting anthropogenic climate change (Santer et al.

1996). In order to detect anthropogenic climate change, one needs to compare the observed changes with the typical climate variations. Thus, a well-defined measure of the variability or ''noise'' in the climate system is needed. The record of surface temperature changes is presently the longest, most reliable observational record available. However, for the purposes of climate change detection, the observational record is too short to determine accurately the unperturbed variability of the climate system on timescales of several decades or longer. Furthermore, external natural climatic influences such

Corresponding author address: Mr. Ronald J. Stouffer, Geophysical Fluid Dynamics Laboratory/NOAA, Forrestal Campus, Forrestal Road/U.S. Route 1, P.O. Box 308, Princeton, NJ 08542.
E-mail: rjs@gfdl.gov

as volcanic eruptions or changes in solar output, as well the anthropogenic emission of significant amounts of greenhouse gases into the atmosphere, complicate the interpretation of the observational record. Therefore, the observational record contains both ''signals,'' that is, forced changes in the climate system, and noise, which is the unperturbed internal variability of the coupled ocean–atmosphere–land surface system.

To circumvent this problem, the variability found in unperturbed control integrations of coupled models of the ocean–atmosphere system has been used to estimate the noise in the observed climate system (e.g., Stouffer et al. 1994; Santer et al. 1995; Mitchell et al. 1996; Hegerl et al. 1996, 1997; Tett et al. 1996). Here, we compare the modeled variability obtained from GCMs to estimates of the observed variability on annual to decadal timescales. On these timescales, the long time-scale forcings and associated surface air temperature increases due to changes in the greenhouse gases should be small relative to the magnitude of the variability. However, even on the shorter timescales, climatic responses to forcings such as volcanoes and solar cycle variations are present in the observational record.

The primary purpose of this paper is to compare the SAT variability found in each of the three 1000-yr integrations to each other and to the observations. Some attempts have already been made to compare the surface air temperature variability from several sources, including models and observations. Kim et al. (1996) compared variability found in the Geophysical Fluid Dynamics Laboratory (GFDL) coupled model used here with an older version of the Max Planck Institute (MPI) model discussed here and the observations. Santer et al. (1996) and Hegerl et al. (1996, 1997) used variability found in several different coupled models, including some of those used here, for detection studies. All these authors found some differences in the magnitude and structure of the climate variability as simulated by the models. As an example of the utility of making estimates of variability from long coupled model integrations, we will use the simulated low frequency variability to re-assess the observed warming of this past century.

2. Model description and observed datasets

The three coupled models used here were developed at the Max Planck Institute in Germany (HAM3L), the Hadley Centre in the United Kingdom (HadCM2), and the National Oceanic and Atmospheric Administration's Geophysical Fluid Dynamics Laboratory in the United States (GFDL). All three institutions have a long history of climate research and these models represented the state of the art in the early 1990s (e.g., Gates et al. 1992). In order to integrate the model multiple centuries with existing computational resources, the resolution of the models used here is relatively low. All three institutions have obtained new integrations using higher-resolution coupled ocean–atmosphere models than those described here.

The atmospheric model components of the HAM3L and GFDL models use spectral dynamics and have approximately the same horizontal resolution. The HadCM2 model uses a gridpoint model for its atmospheric component and uses a higher resolution than the other two models. The atmospheric component of the coupled model is important because much of the variability on short timescales originates in the atmosphere, as will be discussed later. Both the HadCM2 and GFDL models use versions of the Bryan–Cox ocean model while the HAM3L model uses the large-scale geostrophic (LSG) ocean model. All three coupled models use flux adjustments for heat and freshwater; the HAM3L model also uses wind stress adjustments. These adjustments vary seasonally and geographically, but they have no interannual variations. Therefore they should not systematically act to damp or amplify surface anomalies. The subgrid-scale parameterizations used in the three models are quite different from each other and represent a good cross section of the parameterizations used in climate change research today. Below only a brief description of each model is given. For a more detailed description, please follow the references found in each section.

a. HAM3L

The coupled ocean–atmosphere model, HAM3L or ECHAM3/LSG (Voss et al. 1998) is an updated version of the MPI ocean–atmosphere GCM (Cubasch et al. 1992; Maier-Reimer et al. 1993; Roeckner et al. 1992), which uses the ECHAM3 atmospheric model. This atmospheric model has a horizontal resolution of T21 (with an associated 5.6° lat Gaussian grid) and 19 vertical levels. The LSG ocean model has 11 layers in the vertical and a horizontal resolution similar to the Gaussian grid of the atmospheric model. It includes a simple thermodynamic sea ice model. The atmospheric component's physical parameterizations include cloud liquid water as a prognostic variable and resolve the diurnal cycle (Roeckner et al. 1992). HAM3L has been used for a number of anthropogenic climate change simulations with greenhouse gas and direct sulfate aerosol forcing (Hasselmann et al. 1995; Cubasch et al. 1996), for several simulations with changes in solar irradiance forcing (Cubasch et al. 1997), and for paleoclimate studies (Schiller et al. 1997). The variability in a 1250-yr simulation using the older version of the coupled model (ECHAM1/LSG) is described in von Storch et al. (1997). The climate sensitivity to a doubling of CO_2 for the model used here is estimated to be on the order of 3.0 K Voss and Mikolajewicz (1999). Flux adjustments for heat, freshwater, and wind stress are included in the ocean–atmosphere coupling to reduce model drift.

We use data from the HAM3L simulation after the year 160, since the climate drift in the model during the
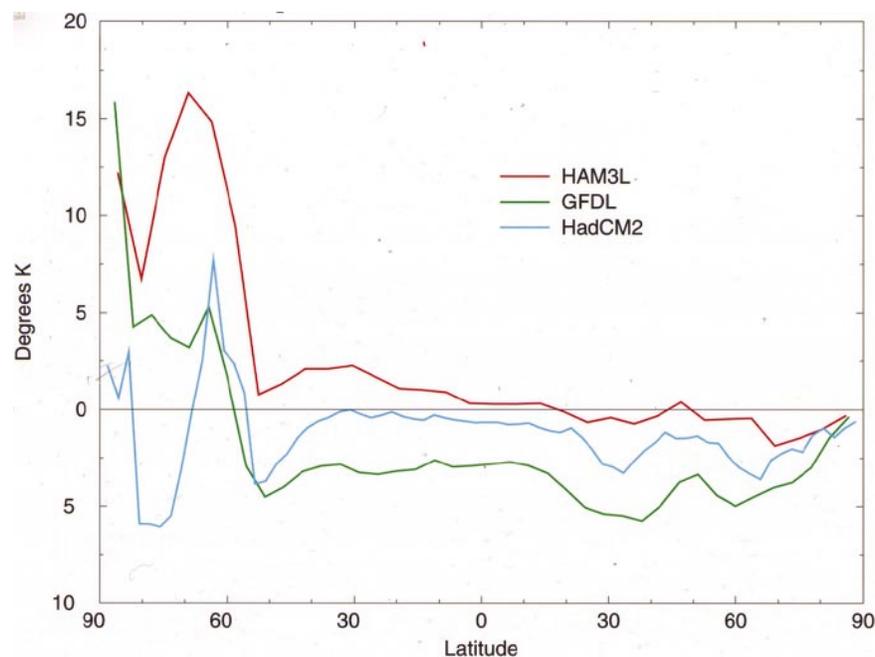
FIG. 1. Zonal averages of the 1000-yr time mean, annual mean surface air temperature difference (simulated minus observed) from HadCM2 (blue), GFDL (green), and HAM3L (red). The observations are obtained from Legates and Willmott (1990). The units are K.

first 160 yr of the simulation is quite large. The drift appears to originate near in the sea ice edge in the Southern Hemisphere and causes a global mean warming of about 0.5 K in the first 160 yr of the unperturbed control simulation. After that, the global mean SAT warms approximately 0.2 K in the next 1000 yr of integration, as is shown later in this paper.

b. GFDL

A 1000-yr simulation with the GFDL model (Manabe and Stouffer 1993, 1996) is used here. The GFDL coupled ocean–atmosphere GCM consists of a spectral (R15) atmospheric model using a 4.5° lat × 7.5° long Gaussian grid with nine vertical levels. The ocean component of the coupled model is a version of the Bryan–Cox model using a 4.5° lat × 3.75° long resolution with 12 vertical levels and contains a simple thermodynamic sea ice model that allows sea ice drift with the surface ocean currents. A simple land surface model includes interactive hydrology. The atmosphere and ocean models are coupled daily using flux adjustments for heat and freshwater. There is no diurnal variation of insolation in the model. The climate sensitivity of the model is 3.7 K as obtained from an atmosphere–mixed layer ocean model and 4.5 K when obtained using the coupled model (Stouffer and Manabe 1999). The coupled model and its variability are described in Stouffer et al. (1994), Manabe and Stouffer (1996, hereafter referred to as MS96), and Delworth et al. (1993, 1997). This model also has been used to study climate change due to in-

creasing CO_2 (Stouffer et al. 1989; Manabe et al. 1991; Manabe and Stouffer 1994; Stouffer and Manabe 1999; and others), CO_2 and sulfate aerosols (Haywood et al. 1997), carbon cycle studies (Sarmiento and Le Quéré 1996; Sarmiento et al. 1998), and various paleoclimate issues (Manabe and Stouffer 1988, 1995, 1997).

c. HadCM2

We use 1000 yr from a long control simulation (Tett et al. 1997) with HadCM2 (Johns et al. 1997). The HadCM2 model has a horizontal resolution of 2.5° lat × 3.75° long. The atmospheric model uses 19 vertical levels and the physical parameterizations include cloud water as a prognostic variable and a full diurnal cycle. The ocean model was developed from the Bryan–Cox model. For the tracer grid, it uses the same horizontal grid as the atmosphere model and it uses 20 vertical levels. A simple sea ice model is included that allows for leads and drift with the ocean surface currents. Flux adjustments of heat and freshwater are used to modify the fluxes that are passed from the atmosphere to the ocean daily. The climate sensitivity of this model is 4.1 K (Senior and Mitchell 1999, manuscript submitted to *J. Climate*). This model has been used for a number of anthropogenic climate change experiments (Mitchell and Johns 1997; Keen and Murphy 1997; Gregory and Mitchell 1997). The atmospheric component and the coupled model have been used for paleoclimate studies (Hewitt and Mitchell 1996, 1998).

d. Observations

The observed near-surface temperature data used have been compiled as anomalies with respect to the average of the years 1950–79 for monthly mean near surface temperature on a $5° \times 5°$ global grid by Jones and Briffa (1992) and Jones (1994a,b). The data range from the year 1854 to the present, with data coverage changing in time. Global mean near-surface temperature time series and the decadal mean pattern of temperature change based on these data are presented in Folland et al. (1992). Due to data coverage limitations in the earliest part of the observed record, only the years 1880–1990 are used for most comparisons with modeled variability. For the comparison of model variance fields with observations, annual means were computed from the monthly time series data. At least 10 months of a 12-month period were required to construct an annual mean. Otherwise, that year was ignored in our analysis. This procedure ensures that the higher variance associated with a time series composed of shorter averaging periods for the computation of the annual means does not lead to an overestimate of the observed variability. This criterion has been relaxed to three annual values (which consist of at least 1 month each) in 5 years for computing time series of the model common EOF patterns from the observations. Here, it was considered preferable to have as high as possible spatial coverage of EOF patterns, even if this possibly included poorly averaged annual mean values. However, we found that the common EOF results are not very sensitive to the criterion used for computing annual mean values.

As noted earlier, the observations contain not only internally generated climate variability, but also climate response to various natural or anthropogenic external radiative forcings such as volcanic eruptions, changes in solar radiation, ozone, aerosol forcing, and increasing greenhouse gas concentrations. By limiting the comparison of the simulated and observed variability to timescales from 1 to 5 yr (fluctuations of 2 to 10 yr), most of the climate changes due to increasing greenhouse gases are expected to be small relative to the internally generated variability. For the comparison of modeled versus observed variability on longer timescales, the observed data were locally detrended using a linear, least squares fit to determine the trend. Jones and Hegerl (1998) show that, at on annual timescales, the subtraction of an estimated greenhouse gas signal or greenhouse gas and sulfate aerosol signal changes the estimate of variability only marginally.

For computing the time evolution and therefore the variance of the common EOFs from the observations, locally subtracting a linear trend of varying length would distort the pattern variability. For that part of the comparison, the observations have been used unaltered. For comparison, we have also used an observational dataset where an estimate of the anthropogenic influence has been subtracted (Jones and Hegerl 1998). The sub-

tracted signal consists of a spatial pattern that is constant in time and that reflects the response of the HAM3L model to greenhouse gas and sulfate aerosol forcing. We show the time evolution of the EOF patterns from both the raw observations and those with the anthropogenic signal subtracted.

The study of the effect of the other external climate forcings on the observed temperature record is an area of active research. Given the presence of the external radiative forcings in the observations, exact agreement between observed and model variability cannot be expected. The integrations contain no changes in the radiative forcing except for the diurnal and seasonal cycles. If it is assumed that the effect of radiative forcings can be superimposed linearly on the model's variability (which should be the case for small forcings) and therefore increases the overall variability in the observations, we expect that the variability obtained from the simulations should be generally smaller than the observed variability.

It should also be noted that the definition of surface air temperature varies from model to model and between the models and the observations. In the HadCM2 and HAM3L results, SAT is a temperature above the surface at 1.5 and 2 m, respectively. It represents a value that is interpolated between the lowest model level and the surface temperature. In the GFDL results, SAT is the temperature of the lowest model level, which is about 80 m above the surface. In the observations, the values over land areas typically are measured about 2 m above the surface, while over the ocean they are obtained from sea surface temperature anomalies. Among the model results, the differences in the definition of SAT only slightly affect the results shown here since anomalies are mainly used in the analysis. In the comparison between the modeled results and the observations, the differences in using SST anomalies instead of SAT may be important over the oceans on short timescales. MS96 have shown that there are differences in the spectral estimates between SST and SAT over the oceans.

To analyze the temporal variations in the simplest way, the time series obtained from the three models are time averaged over nonoverlapping, consecutive time intervals of 1-, 5-, and 25-yr time series. This removes most of the variations shorter than 2, 10, and 50 yr, respectively. Obviously, it would be desirable to filter the data with a sharply defined bandpass filter. However, given the shortness and the number of gaps in the observed record such a filter was not used. For the variance analysis presented in the next section, the observations and the model results are linearly detrended.

3. Local variability of SAT

Before looking at the local (and later the global) variability of SAT, we will first examine the mean state or
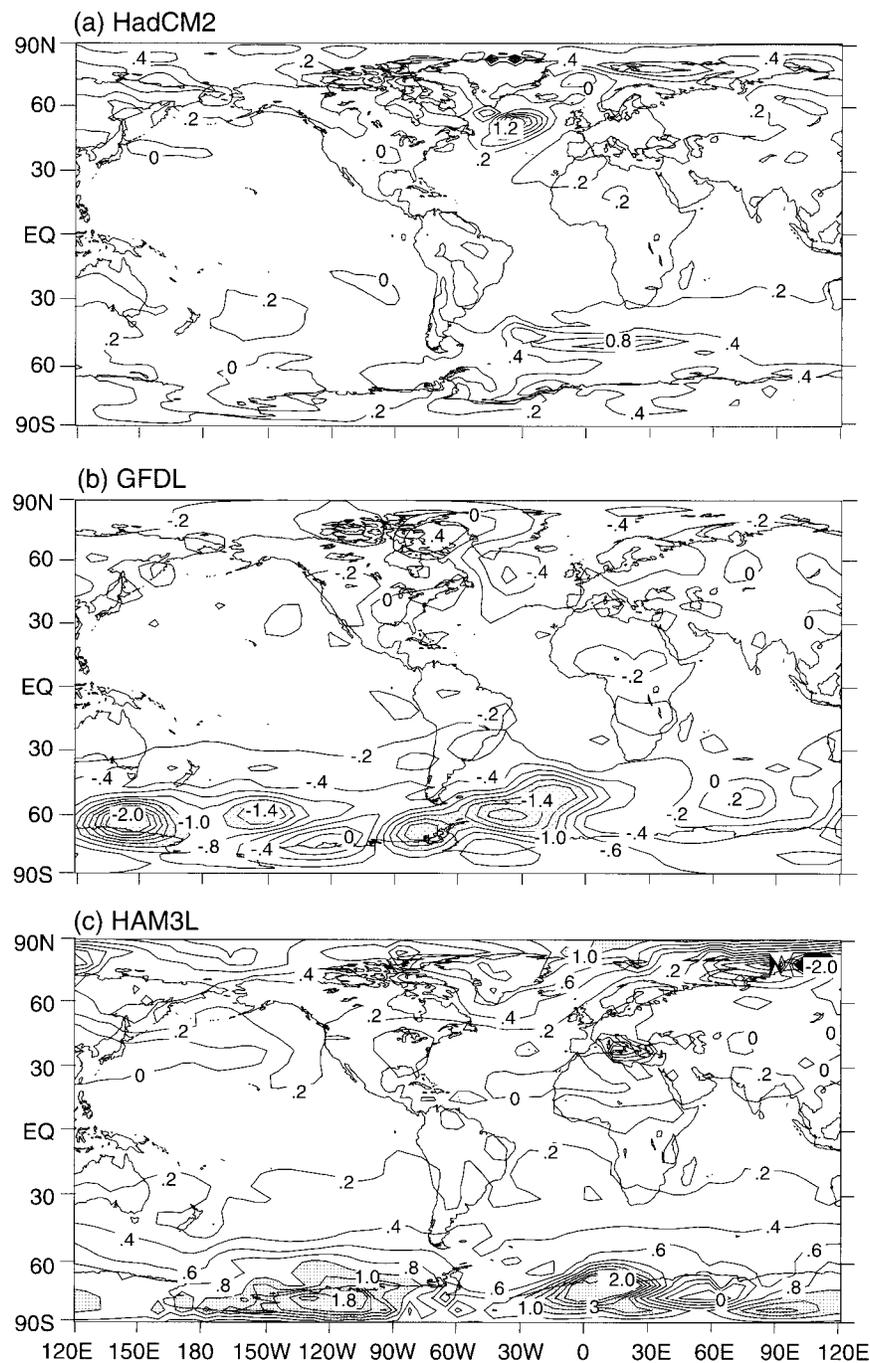
FIG. 2. Geographical distribution of the linear trend computed from the 1000-yr time mean, annual mean surface air temperature time series from (a) HadCM2, (b) GFDL, and (c) HAM3L. The units are K 1000 yr^{-1}. The shading indicates areas where the trend is greater than 1 K 1000 yr^{-1} (positive or negative).

climatology of the different models and the linear trends present in the datasets. Any trend present in the dataset, if it is large enough, will make the computation and interpretation of the variance difficult. In addition, the time mean state can have a large impact on the location and magnitude of the variability found in a simulation, since it can influence the nonlinear mechanisms that generate climate variability. For example, if the 0°C isotherm and the corresponding sea ice or snow cover boundary is misplaced, the location of the oceanic and atmospheric variability associated with the location of this boundary will also be misplaced.
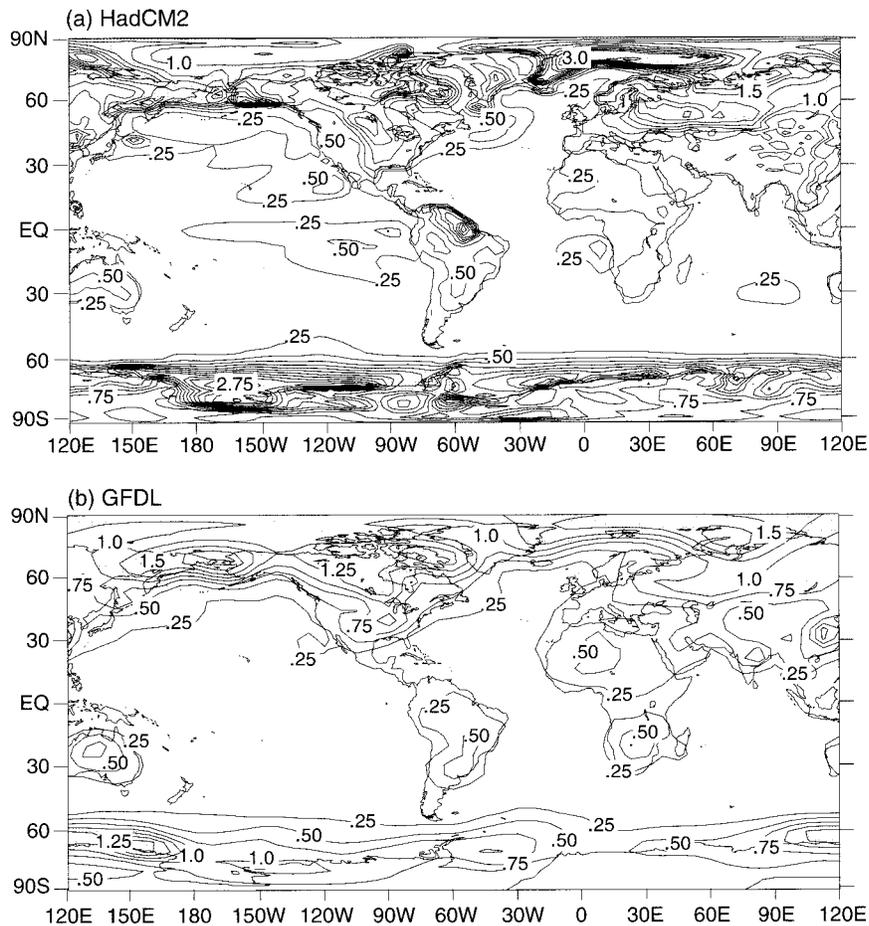
FIG. 3. Geographical distribution of the variance computed from the annual SAT anomalies obtained from (a) HadCM2, (b) GFDL, (c) HAM3L, and (d) observations (Jones and Briffa 1992). The units are K^2. The shading indicates areas where the variance is larger than 1.0 K^2.

a. Climatological mean state and drift

Therefore before examining the variance of SAT, we first show the zonally averaged, time mean SAT difference (simulated minus observed) computed over the 1000-yr time series (Fig. 1) and the linear trend (Fig. 2) computed at each grid location for all three models. Overall, the models do a fairly good job of simulating today's climate (Fig. 1). However, there are some large errors in the models' simulations. In the Southern Ocean of the HadCM2 and GFDL plots, there are areas where the zonal difference from the observed (Legates and Willmott 1990) is greater than 5 K. In the geographical distributions (not shown), there are points where the local SAT in the two models are more than 10 K warmer or colder than the observations. This region contains very few observations (Legates and Willmott 1990), so it is likely that the observed dataset has large uncertainties. In the HAM3L simulation (Fig. 1), the zonally averaged SAT is warmer than the observations in the Southern Ocean by more than 10 K. The unrealistically warm climatology of the high latitudes of the Southern

Hemisphere in the HAM3L model results from the strong warming trend during the first 160 yr of the simulation. (As described in the previous section, the data from this time period were not used in the analysis shown here.) This warming resulted in much less sea ice in the Southern Ocean in the HAM3L model when compared with the other two models or the observations. As will be shown later, this in turn effects the SAT variability as discussed above. In the Tropics and subtropics, the GFDL simulation is too cold by 2–4 K (Fig. 1). About 1 K of this error is due to the use of the lowest model level temperature for the GFDL SAT as discussed earlier. It is encouraging that in general, the simulated mean SAT is fairly realistic. Outside of the circumpolar region and mountainous areas, the mean SAT error is less than 4 K in most areas.

The linear trends found in all three models are quite small (Fig. 2). The largest trends are generally found over the ocean in the neighborhood of sea ice edges, which are also regions of deep to intermediate water formation in the ocean. It is likely that these trends result
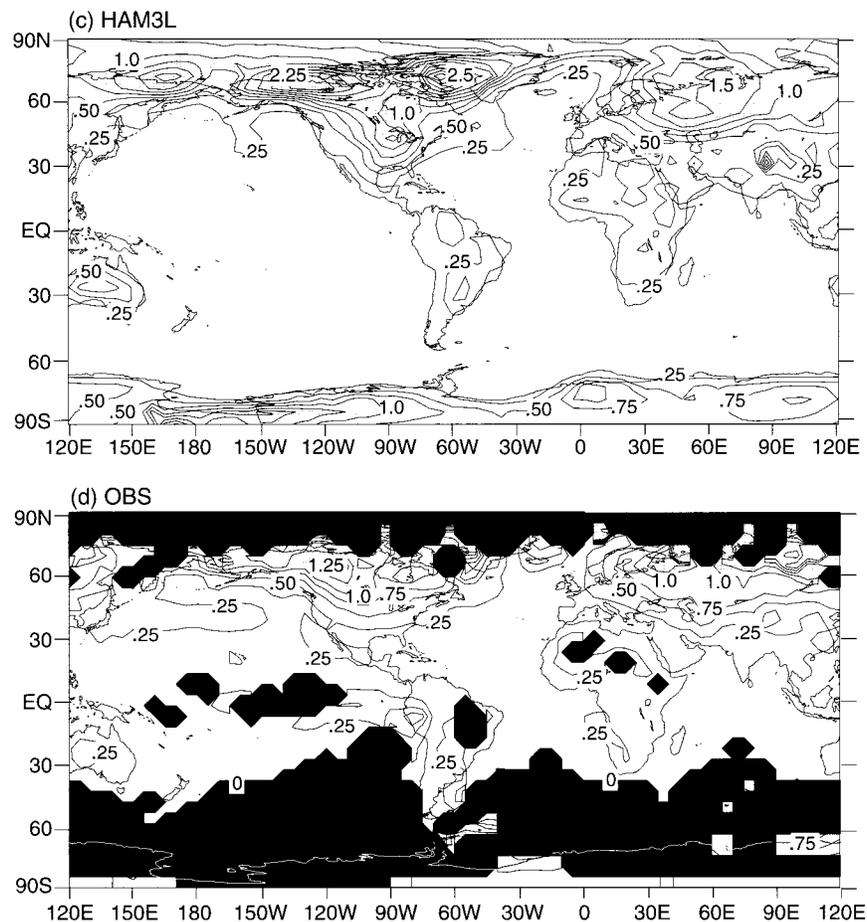
FIG. 3. (*Continued*)

from problems with the initialization of the oceanic component of the coupled model. Over land regions in all the models, the trends are very small and appear to be related to the global mean drift in each model. In the GFDL simulation, the largest local trends are about 1.5 K cooling per 1000 yr around Antarctica. In the HAM3L simulation, the largest local trends are also found around Antarctica and are about 2.0 K warming per 1000 yr. In the HadCM2 simulation, the largest local trend is found in the North Atlantic and is 1.2 K warming per 1000 yr. This local maximum of the SAT variability in the North Atlantic in the HadCM2 time series appears to be generated by a single event where the thermohaline circulation increases by 3–4 SV over a 50-yr period during the integration (Tett et al. 1997).

Despite the relatively large trends discussed above, at most locations in all three models, the linear trend is smaller than 0.2 K $(1000 \text{ yr})^{-1}$. The GFDL simulation has a general cooling trend, while HAM3L and HadCM2 have a general warming trend. The magnitude of the trends are much smaller than the variability at most locations on timescales shorter than 50 yr or so. Therefore, with the possible exception of the region around

Antarctica, the bias of the computed variance due to drift should be negligible.

In summary, along with the exceptions discussed above, the simulation of the time mean, globally distributed SAT by the models is fairly realistic and the local linear trends over the 1000-yr period are quite small.

b. Variance

The variability in each of the three long model control simulations has been previously published individually by authors from the respective institutions (MS96; Tett et al. 1997; Voss et al. 1998). Here, we compare the variability among the different models and the observations to assess the ability of the models to simulate local temperature variability. The variance maps computed from the annual mean SAT time series (Fig. 3) indicate that for both the models and the observations, the variance is generally larger in high latitudes than in low latitudes. At a given latitude, the variance is larger over the continents than over the surrounding oceans. The exception to these generalizations is found in the
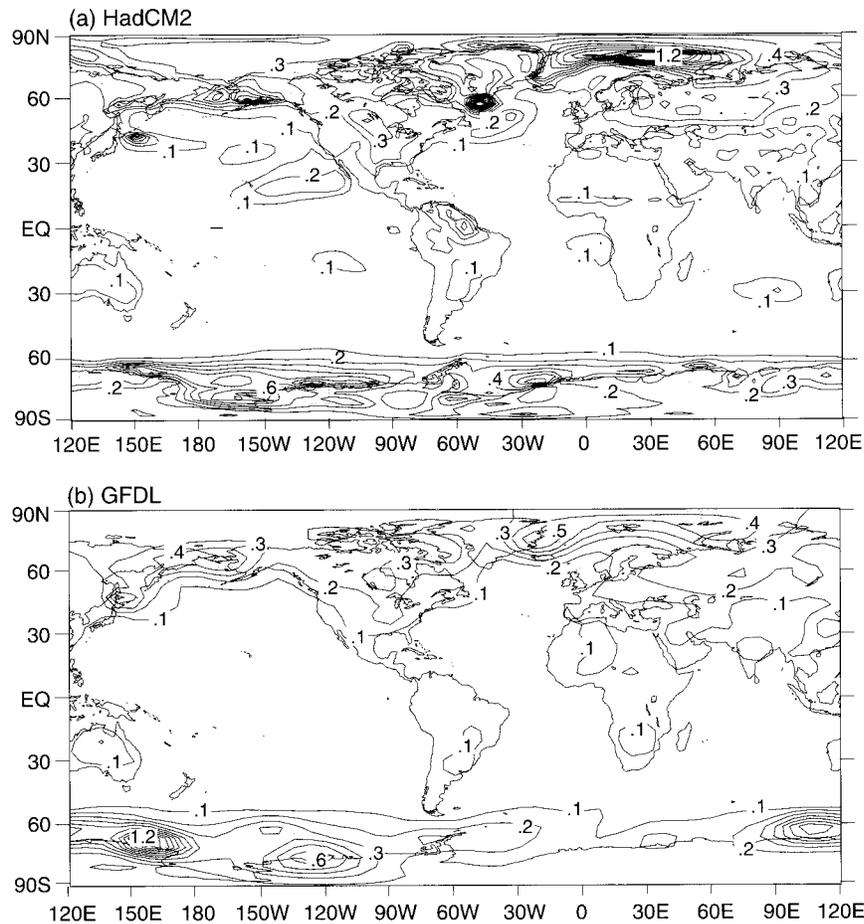
FIG. 4. Geographical distribution of the variance computed from the 5-yr mean SAT anomalies obtained from (a) HadCM2, (b) GFDL, (c) HAM3L, and (d) observations (Jones and Briffa 1992). The units are K². The shading indicates areas where the variance is larger than 0.3 K².

tropical Pacific, where the observations (Fig. 3d) show a local maximum in the variance due to the El Niño–Southern Oscillation (ENSO). Of the models, only HadCM2 (Fig. 3a) seems to simulate the correct magnitude of the SAT variability in the tropical Pacific region. All three models simulate an ENSO-like phenomenon to some degree. However, due to limitations in the horizontal resolution and physical parameterizations in the models, the magnitude of the variability is generally too small in GFDL and HAM3L. [For more information on the ENSO-like simulations from these three models see, for the GFDL model, Knutson et al. (1997); for HadCM2, Tett et al. (1997); and for HAM3L, Timmermann et al. (1999).]

Another area in the observations (Fig. 3d) where there is a local variance maximum is located in the central North Pacific Ocean. To some degree, this maximum is simulated by all three models with the HadCM2 model being the closest to the observations. However, it is not well simulated by any of the models. Over northeastern South America, the simulated variance is much too large in the HadCM2 results when compared with the obser-

vations. As noted earlier, the HAM3L is too warm in the Southern Ocean. This results in a much smaller sea ice coverage and associated SAT variance in this region when compared with the other two models. With the exceptions noted above, the observed variance computed from the annual mean anomalies of SAT is generally well simulated by the models.

The variance maps computed from the 5-yr mean time series for the models and the observations (Fig. 4) show a pattern similar to annual variance maps discussed above. Again in the tropical eastern Pacific, the observations and HadCM2 show a region where the variance is relatively high. As was the case for the variance computed from the annual mean timescales, the variance computed from the 5-yr mean time series is generally larger in high latitudes than in low latitudes and is generally larger over the continents when compared to the adjacent oceans. Both of these generalizations seem to explain more of the pattern in the 5-yr variance distribution (Fig. 4) than in the annual variance distribution (Fig. 3).

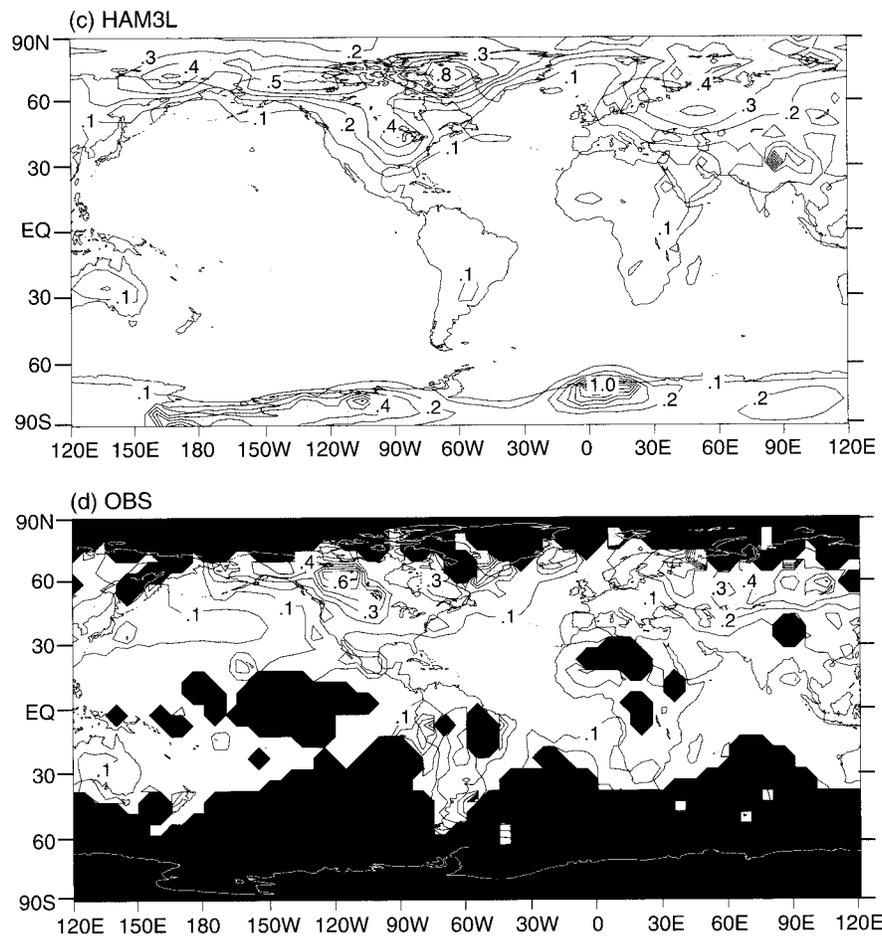By comparing Fig. 3 to Fig. 4, there is very close to

FIG. 4. (*Continued*)

a factor of 5 reduction in the variance for most areas. Later, we will discuss in more detail the fact that the local interannual variability is similar to white noise and therefore is nearly uncorrelated between consecutive years.

The locations where the models have poor simulations in the 5-yr variance distributions (Fig. 4) are similar to those found in the annual mean variance distributions (Fig. 3). However, it is important to note that our estimate of observed variability (Fig. 4d) based on 5-yr averages is rather poorly sampled and may be influenced by external climate forcings. In general, one can again conclude that overall, the observed variance computed from the 5-yr anomalies of SAT is well simulated by the models.

Due to the short record, variances are not computed from the observed 25-yr mean time series. Comparing just the models to each other (Fig. 5), one notes that the largest variances computed from the 25-yr mean time series are found in high latitudes. However, upon closer inspection, one notes that the exact location of the region of maximum variance is found at different locations in each model. In the HadCM2 results (Fig.

5a), the maxima are located near the southern tip of Greenland and in the Arctic Ocean north of Europe. In the GFDL results (Fig. 5b), there are three maxima: one in the northern North Atlantic, one north of Japan, and another in the Southern Ocean. In the HAM3L results (Fig. 5c), there are relatively large values found over land points of Antarctica near coastlines, unlike in the other two models. Also, there are maxima located near Baffin Island and over the Arctic Ocean north of Siberia. In HadCM2, a small variance maximum also occurs in the tropical Pacific.

It is also interesting that the magnitude of the variance in the local maxima described above is very similar to that found in the variances computed from the 5-yr mean time series, indicating that the variance is behaving like very red noise or shows pronounced very low-frequency modes in these areas. Deep and intermediate water formation regions in the ocean are typically located in high latitudes near the sea ice margin. In all the models, as will be discussed later, variations in the internal oceanic variability, deep to intermediate oceanic mixing, and/or the sea ice processes can lead to persistent anomalies
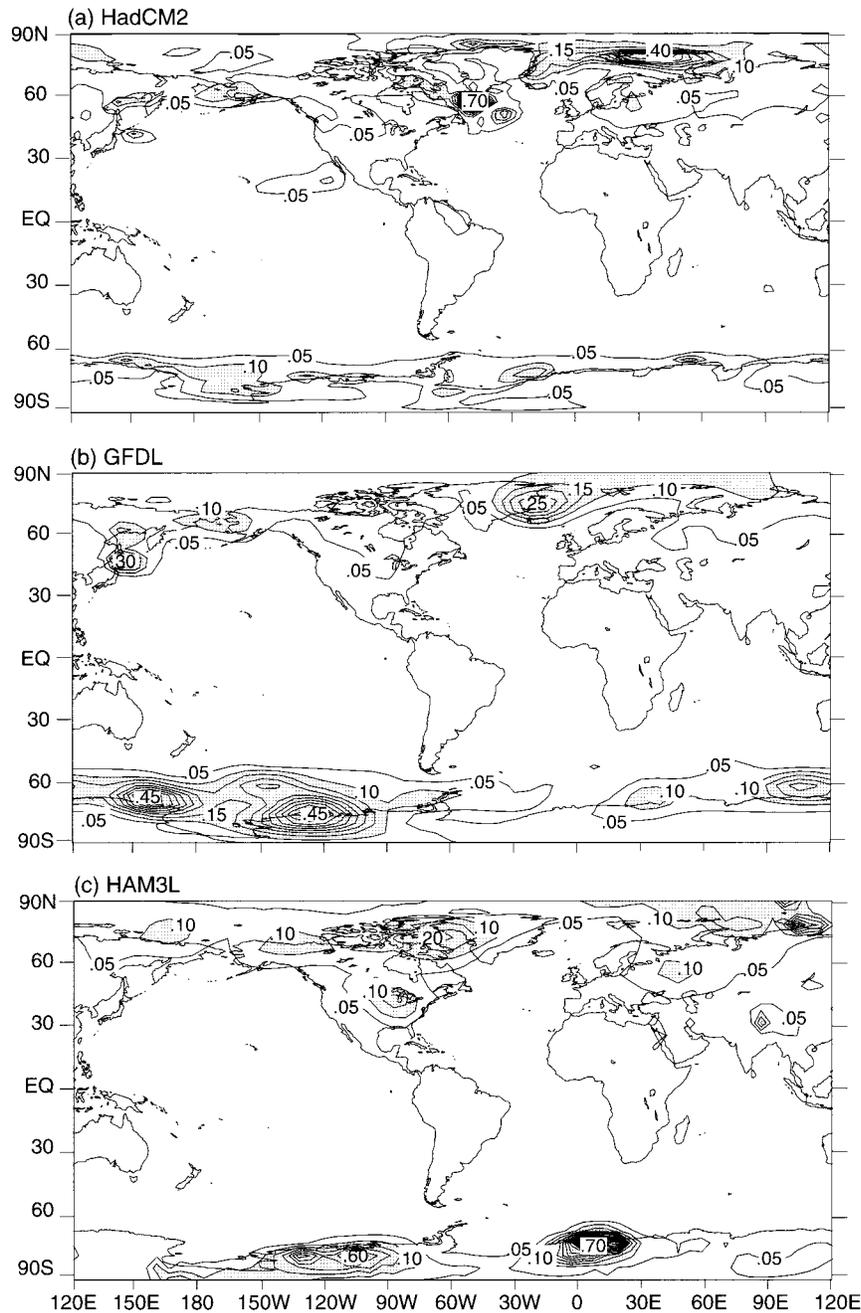
FIG. 5. Geographical distribution of the variance computed from the 25-yr mean SAT anomalies obtained from (a) HadCM2, (b) GFDL, and (c) HAM3L. The units are K^2. The shading indicates areas where the variance is larger than 0.1 K^2.

in these regions creating the large variances on long timescales.

c. Interpretation of modeled variability

In order to explain surface temperature variability, Hasselmann (1976) proposed a theory in which he envisioned an atmosphere that generates variability, which can be characterized as white noise on interannual and longer timescales. This white noise is integrated by the global ocean producing a red noise response. Note that this concept allows for the excitation of ocean variability in all frequency ranges by atmospheric noise, but not for an atmospheric response to the oceanic variability. MS96 applied a simplified, localized version of that theory to explain the local variability in their model. In

this theory, the local white noise forcing of the atmosphere produces a local red noise response in the sea surface temperature. In the simplified theory, the spectrum of local temperature can be explained by the thermal damping of the underlying surface, leading to red noise response up to a certain frequency. The redness of the spectrum depends on the magnitude of the thermal inertia of the underlying surface.

MS96 found that in most regions of the globe, the simplified version of Hasselmann's theory appears to provide a reasonable approximation of the air–sea interactions in their coupled model. This conclusion is supported by the analysis of Hall and Manabe (1997). By studying both sea surface temperature and sea surface salinity anomalies in the model and in observations outside of places of deep water formation or active eddy fields in the ocean, both the models and observations indicate that the simplified version of Hasselmann's theory provides a reasonable quantitative explanation of the origin of climate variability.

Understanding the relationship of SAT variability to the underlying surface is important in order to appreciate the results presented here. SAT is influenced both by the white noise of the free atmosphere and by the damping of the underlying surface (see MS96 for a more complete analysis). Over the continents, the damping is small due to the small heat capacity of the continental surface. Therefore the continental SAT spectrum is approximately white noise on annual and longer timescales. Over the oceans on short timescales, the damping is much larger due to the large heat capacity of the oceanic mixed layer. Since the SAT anomalies are closely linked to the sea surface temperature anomalies through the surface fluxes, the oceanic SAT spectra lie much closer to the red noise spectra of the sea surface temperatures.

Earlier it was noted that the annual mean SAT variance is very close to 5 times the variance obtained from the 5-yr time series in most continental areas; it appears the SAT time series is nearly white noise on these timescales. This suggests, according to Hasselmann's theory, that most of the surface temperature variance is dominated by atmospheric white noise. Over the oceans in these models, the dynamical ocean does not seem to add very much to the local variance except in the Tropics on 5-yr timescales and shorter and in high-latitude regions where the 25-yr mean temperature variance is much larger than one would expect from the 1- and 5-yr variance maps using Hasselmann's simplified theory (as shown in the preceding section). In these areas the local variability does not appear to be generated from local thermally integrated atmospheric white noise. The location of these maxima suggests that they are related to variations of the thermohaline circulation, areas of the deep to intermediate water formation processes, and/or interactions with sea ice. In these regions and on these timescales, the simplified version of Hasselmann's theory is not a good model of the local variability.

The red component of the variability can be better assessed by analyzing the persistence of anomalies of annual and decadal timescales. For purely white noise, the lag-1 autocorrelation should be zero (or very close to zero due to sampling uncertainty), indicating that the thermal inertia of the underlying surface is small. Over land areas, the lag-1 autocorrelation coefficient computed from the 1-yr mean time series is very close to zero (Fig. 6). Due to the very small heat capacity of the land surface, there is very little persistence of SAT anomalies from one year to the next.

Over the ocean, the magnitude of the correlation coefficients are much larger than over the land and shows that the spectral characteristics of the variability is significantly red in most regions, confirming the simplified version of Hasselmann's theory. However, the pattern from model to model is very different. In the HadCM2 results (Fig. 6a), there are high-latitude Northern Hemisphere and tropical maxima. In the GFDL results (Fig. 6b), the largest coefficients are found around Antarctica, the northwestern part of the Pacific Ocean, and the northern North Atlantic Ocean. In the HAM3L results (Fig. 6c), the largest lag-1 autocorrelation coefficients are found around Antarctica, the tropical Pacific Ocean, and some high-latitude Northern Hemisphere regions. Also in general, the lag-1 autocorrelation coefficients seem larger over the HadCM2 oceanic regions than in the HAM3L oceanic regions, which in turn are larger over most areas than those seen in the GFDL model. The latter may be caused by a different definition of SAT. In the GFDL results, as was mentioned earlier, SAT is defined as the temperatures of the lowest model level. This level is located in the boundary layer, about 80 m above the surface, allowing the free atmospheric noise to have a greater influence on the SAT variability in the GFDL results than in either the HadCM2 or HAM3L results, which are located much closer to the surface of the ocean. Generally, the maximum autocorrelation coefficients in high latitudes are found near the locations of the variability maxima computed from the 25-yr SAT times series in each model.

On longer timescales, the lag-1 autocorrelation coefficients computed from the 5-yr mean SAT time series are largest in high latitudes in all the models (Fig. 7). However, the correlation coefficients are large only in high latitudes of the Northern Hemisphere in HadCM2 and only in high latitudes of the Southern Hemisphere in HAM3L, while they are large in both hemispheres of the GFDL results. In HadCM2 results, negative values predominate in the Tropics. As will be seen later, there is a spectral peak in the HadCM2 global mean SAT between 5 and 10 yr.

Over land areas, the coefficients are randomly distributed and near zero, suggesting that on this timescale, variability is truly uncorrelated. In all three models, the relatively large coefficients found in the Tropics on shorter timescales are absent. This indicates that, in the models at least, there is little persistence of the SAT
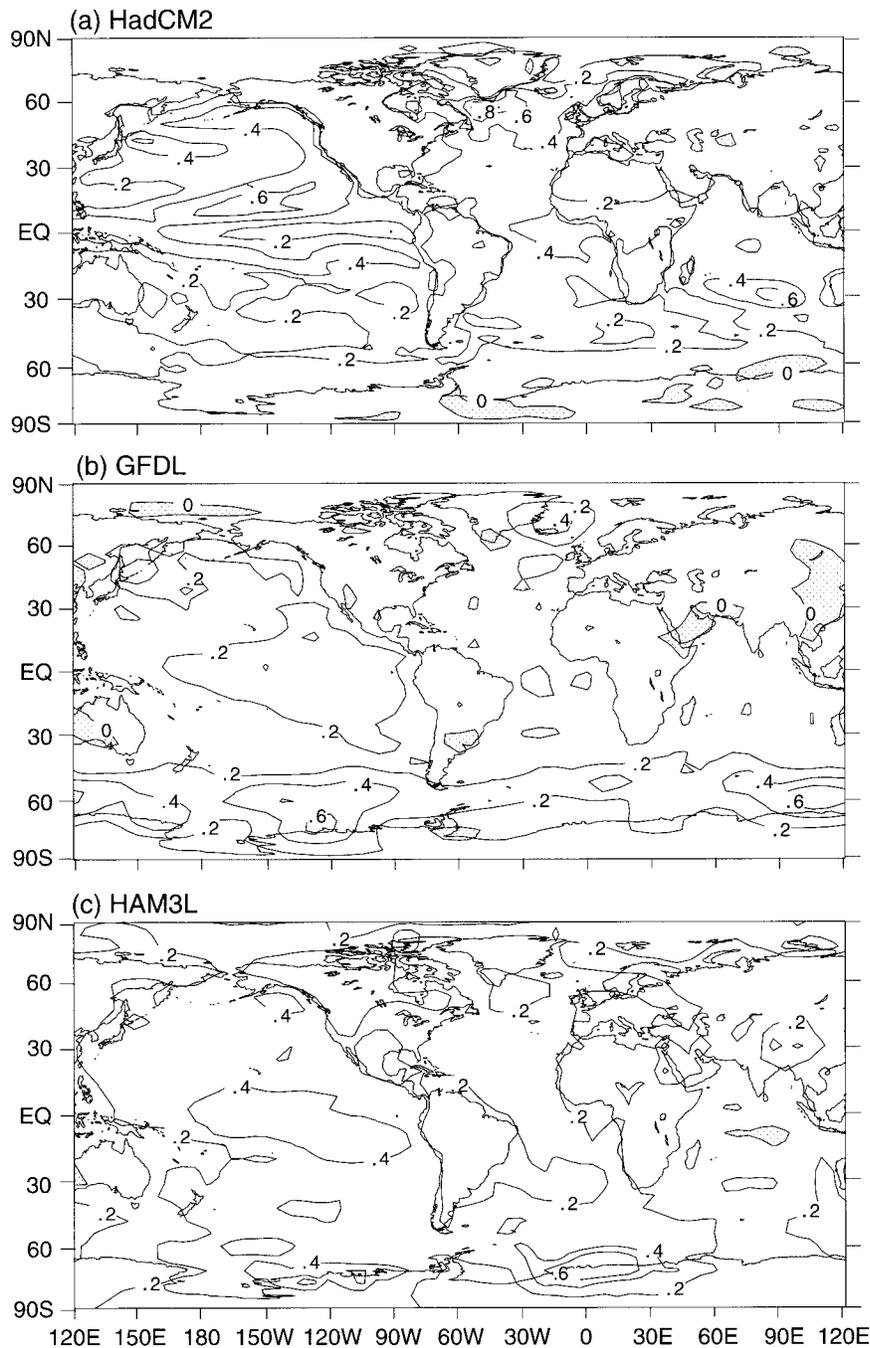
FIG. 6. Geographical distribution of the lag-1 autocorrelation coefficients computed from the annual mean SAT anomalies obtained from (a) HadCM2, (b) GFDL, and (c) HAM3L. The shading indicates areas of negative correlation.

anomalies from one 5-yr period to the next outside of high-latitude oceanic regions.

On multidecadal timescales (not shown), the pattern and magnitude of the correlation coefficients is reasonably similar among the models. In the HadCM2 and GFDL results, the autocorrelation coefficient has a maximum value of about 0.6 in the northern North Atlantic.

Around Antarctica, the maximum autocorrelation coefficient is also about 0.6 in the GFDL and HAM3L results, while it is smaller in the HadCM2 results. These regions correspond with the maxima in 25-yr averaged variance.

The long timescale SAT variability found in these regions results from a number of different processes. One is internal oceanic variability, as mentioned earlier.
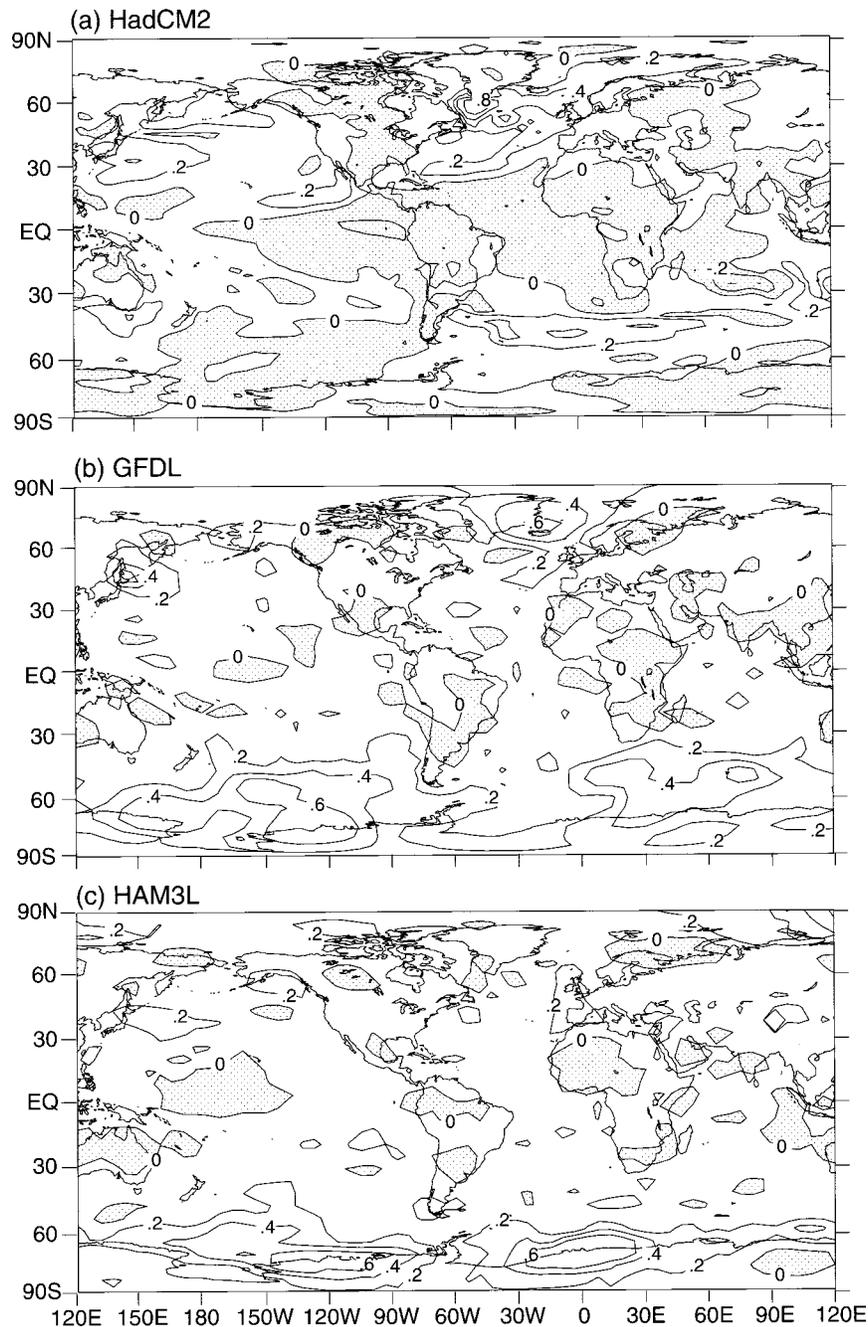
FIG. 7. Geographical distribution of the lag-1 autocorrelation coefficients computed from the 5-yr mean SAT anomalies obtained from (a) HadCM2, (b) GFDL, and (c) HAM3L. The shading indicates areas of negative correlation.

Internally generated oceanic variability can result from variations in the thermohaline circulation (see, e.g., Delworth et al. 1993). Futhermore, deep oceanic convection occurs in these regions and is influenced by complex interactions between atmospheric and sea ice processes as well as internal oceanic processes. Due to the large effective heat capacity of the ocean in regions of deep convection, these interactions also can lead to decadal or longer variability. Finally, sea ice processes interacting with the oceanic mixed layer also can lead to long timescale SAT variability through injections of salt and heat into the ocean, by modifying the surface fluxes into the atmosphere, and by albedo feedback. Further analysis is necessary to determine the exact role each plays in generating the long timescale SAT variability found in high latitudes.
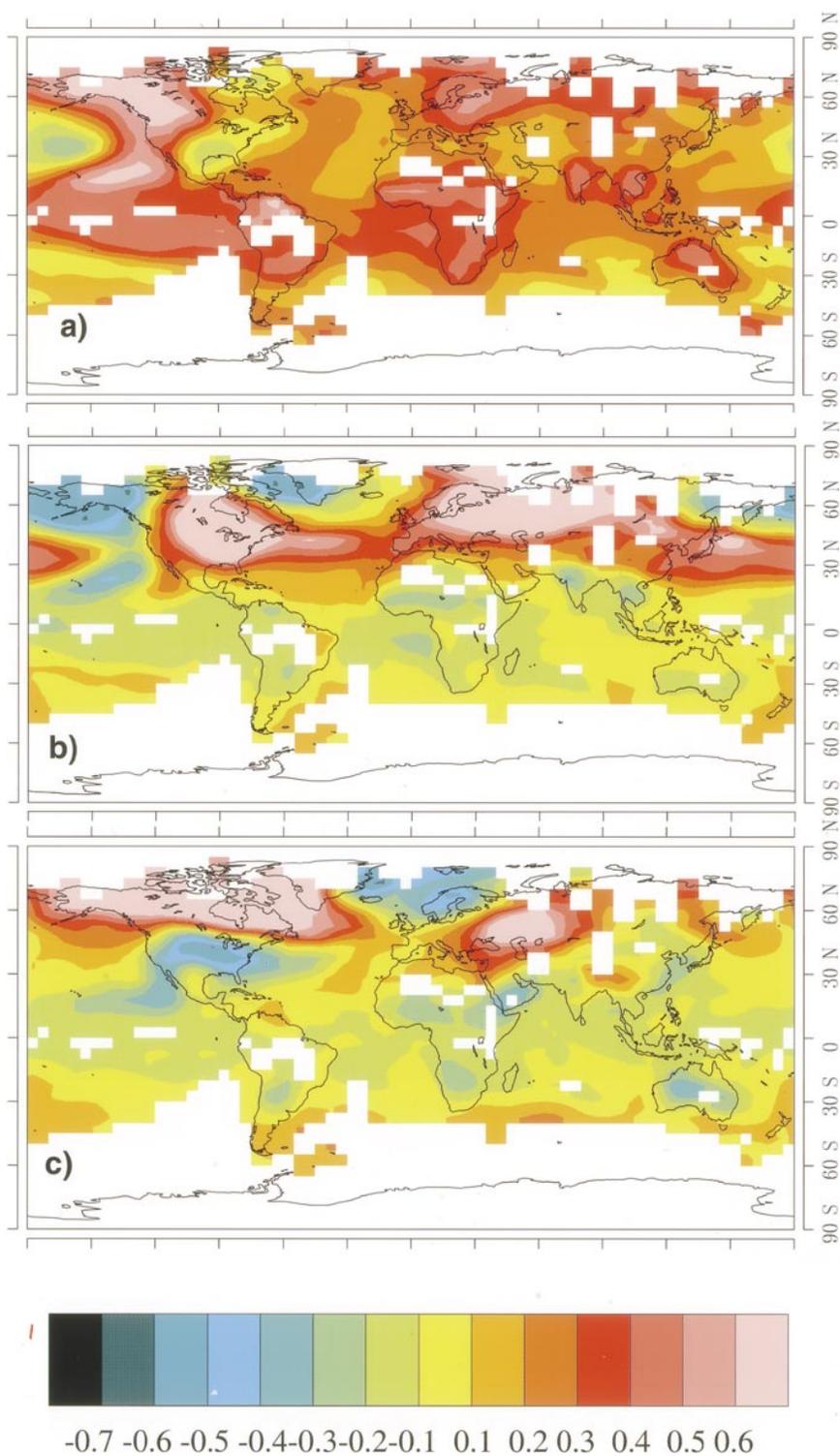
FIG. 8. Geographical distribution of the first three common EOFs for all three models. The EOFs have been computed from concatenated 5-yr mean SAT time series from the first 500 yr of the model integrations. The plotted values are unitless.
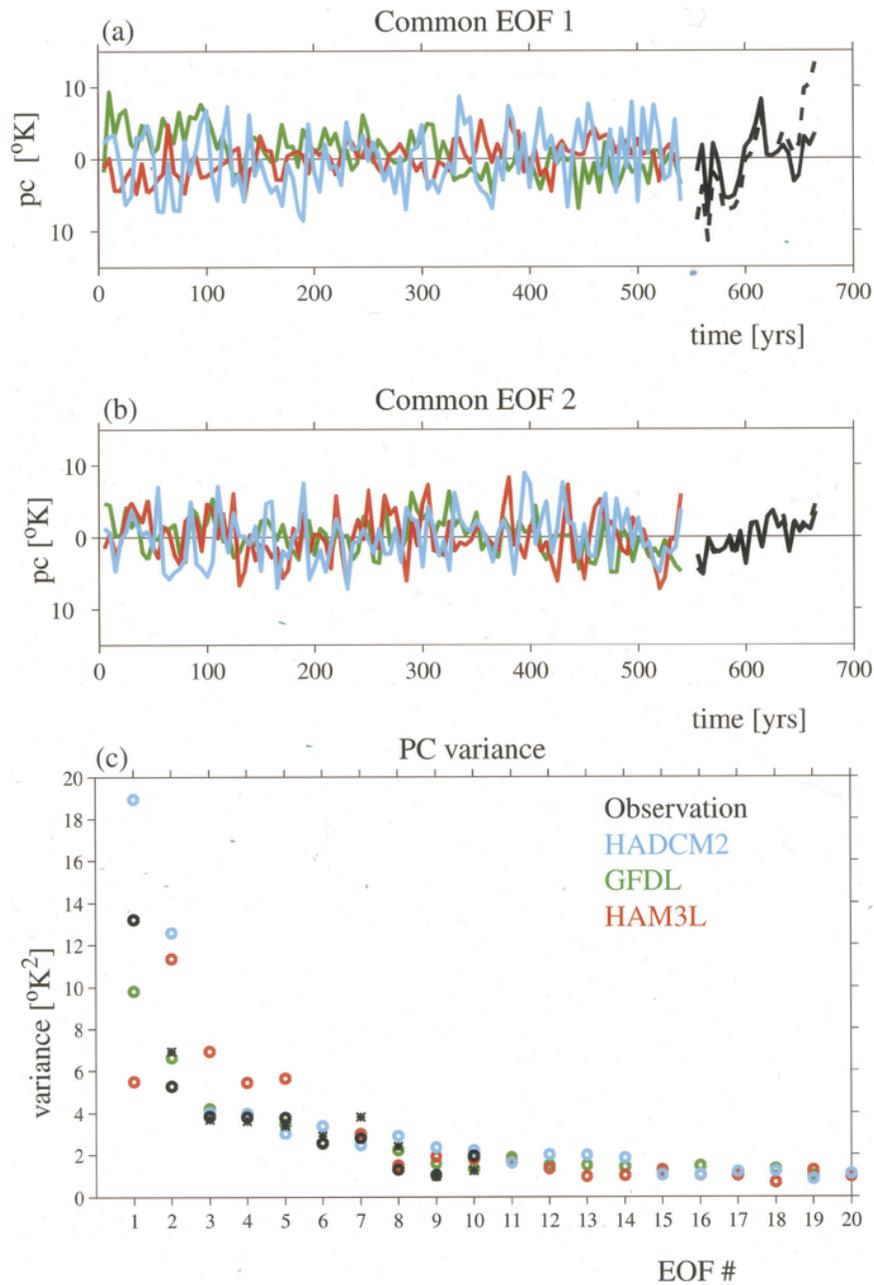
FIG. 9. (a) and (b) The time series of the common EOF coefficients associated with the first two EOF patterns in units of K. The black lines represent the observed data projected onto the eigenvectors computed from the model's datasets. The dashed line represents the raw observed data and the solid line shows the observed data after taking out an estimate of the human influence (Jones and Hegerl 1998). (c) The variance associated with each of the first 20 eigenvectors obtained from the common EOF analysis in units of K^2. The black stars represent the variance from the raw observations, the black ring from the anthropogenic signal-subtracted observations. Since the raw observations are strongly influenced by the trend, the variance of the first EOF of the raw data is very high (~ 35 K^2) and not plotted on the panel. For all the plots, the blue lines or markers represent data from HadCM2, green those from GFDL, red those from HAM3L, and black those from observations.

d. Comparison of structured variability

It is interesting to study not only the local magnitude of variance between the models, but also the covarying spatial patterns in which the different models prefer to arrange their variability. This can be studied by computing empirical orthogonal functions (EOFs), which are eigenvectors of the spatial noise covariance matrix. It has been shown that EOFs obtained from some of the individual model simulations used here differ (e.g., Kim et al. 1996). Part of the difference is likely the result of how the variance is distributed between modes in each model. For example, Santer et al. (1995) compare the spatial EOF patterns for the dominant modes of variability of the GFDL model with an older version of the Hamburg model. While the first EOF of the GFDL model shows a pattern that is associated with the whole globe warming or cooling, its second EOF shows some similarity with the Hamburg model's first EOF. However, the similarity cannot be assessed, since the GFDL model's second EOF is constrained to be orthogonal to the first. Also, if a model has several EOFs with similar eigenvalues, these EOFs can be mixed due to sampling uncertainty (North et al. 1982). For both these reasons, even models with similar teleconnections can yield rather different EOF patterns. This difference makes the interpretation of the results difficult and inhibits the comparison of variance associated with structures of covariability in the different models.

Therefore, we used the common EOF method [Verbeek (1997) and Frankignoul et al. (1989); also see Barnett (1999) for a similar analysis using results from 11 coupled model integrations] to compare the modes of variability found in the three models. The common EOF method isolates patterns of structured variability that are present in all three models. The variability associated with these patterns can be compared quantitatively between the models.

For the computation of the common EOFs, the data from all the models were interpolated onto the same 5° × 5° grid as the observations and area weighted by the square root of the cosine of the latitude. Then, all the model data were restricted to grid points that have been consistently covered by observations since the middle of this century. Other areas are indicated by the blank regions in the common EOF map [Fig. 8; this data space has been chosen in the same manner as Hegerl et al. (1996), (1997)]. This gridding and restriction of the areal coverage was done to enable the comparison with the observations below. To ensure that the covariance matrix was not dominated by models with larger variance, all model data were scaled so that the globally averaged grid point variance was identical for all three models. After that, the regridded and scaled data from the models were concatenated, and EOFs computed in the same way as for an individual model. After the EOFs were computed, the variance associated with each EOF pattern was computed from the unscaled data of each model.

The observations could not be included in the computation of EOFs, because the shortness and number of gaps in the observational record causes too many technical problems and would lead to a very high uncertainty in this analysis. However, the time evolution of the common EOF patterns obtained from the model results can be computed using the observations. Gaps in the observations disturb the orthogonality of the EOF patterns (if only the locations covered in a given year are used rather than the full pattern). Therefore a least square fit has been used to compute the time evolution of the first 10 observed EOF patterns (the results appear insensitive to varying the number of patterns).

The time evolution of the EOF patterns from the observations is influenced by changes in the earth's radiative forcing, which make a comparison of model results and observations problematic. Eliminating these influences by subtracting a linear trend from each grid point of the observations between 1880 and 1990 is difficult, since the data coverage in time varies from grid point to grid point. Thus, we have used both the raw observations and the observations with a smoothed estimate of the greenhouse gas plus aerosol signal removed (Jones and Hegerl 1998).

The first three common EOF patterns computed from 5-yr mean SAT time series are shown in Fig. 8. The pattern shown in Fig. 8a shows nearly uniform warming (cooling) over most of the globe with the exception of the tropical Pacific region. Here, the pattern resembles the decadal tropical Pacific oscillation of Zhang et al. (1997). This first EOF seems to be a compromise between the GFDL and HadCM2 SAT variability. The pattern shows some similarity with the GFDL model's first individual EOF on interannual timescales (Santer et al. 1995), but it shows a stronger expression of decadal tropical variability that is likely influenced by the HadCM2 results. As shown earlier, this model tends to produce stronger tropical variability than the other two models.

By looking at the EOF1 coefficient time series (Fig. 9a; for variance comparison see Fig. 9c), we note that the variability is largest in HadCM2 and smallest in HAM3L. In the GFDL results, and to a lesser extent in HAM3L, this mode also shows some trend in the component, which is associated with the global mean drift in these models. The variability associated with each EOF for the raw observed dataset and the observations after subtraction of the anthropogenic component (called "residual observations" in the following) is also shown in Fig. 9.

Because the sign of the first EOF pattern is consistent over most of the globe, EOF1 picks up much of the observed global mean warming trend in the raw data (Fig. 9a). Thus, its time evolution disagrees quite strongly with that of the residual observations, which had most of this trend removed. For both the raw and residual

observations, EOF2 shows a quite similar time evolution except at the end of the time series (Fig. 9b). There is some trend in the time evolution of the observed EOF2 component in both the raw and residual observations before 1950. It is likely that this trend results from the warming in the first half of this century, particularly over North America, which projects onto EOF2 (Fig. 8b). This warming is only partly removed by subtracting the estimated anthropogenic signal (Jones and Hegerl 1997) and is of uncertain origin. However, it is important to note that the changes in the earth's radiative forcing and the shortness of the record make the variance estimates from the observations uncertain.

Figure 9c, the variance spectrum (see Barnett 1999), shows how much variance is associated with the individual EOF patterns in the different models and in the raw and residual observations. In all models, the explained variance generally decreases with increasing number of the EOFs, which shows that the common EOF analysis captures the dominant modes of each model (note that this reduction is inherent in the concatenated variability data by construction but does not need to occur for each individual model's variance). An exception is HAM3L, where the first EOF shows smaller variability than the second, which is due to the small ENSO-like variability in that model. However, there are substantial differences between the amount of variance associated with each individual model, especially in the first few EOFs. Since EOF1 is influenced by tropical variability, it is not surprising that it is associated with far more variability in HadCM2 than in the other models, given that it has the largest variability in the tropical Pacific. [See Tett et al. (1997) for more discussion on this subject.]

In the HadCM2 and HAM3L results, EOF2 is associated with too much variance when compared to the observations. EOF2 has a hemispherically asymmetric structure with relatively large values over Northern Hemisphere midlatitude continents (Fig. 8b). This may make detection of a response to sulfate aerosol forcing more difficult when using these model simulations as a surrogate for the observed natural variability than when using the GFDL model data (see Santer et al. 1995; Mitchell et al. 1996; Hegerl et al. 1997). This is a result of EOF2 having a hemispherically asymmetric structure and relatively large values over the Northern Hemisphere midlatitude continents where the direct aerosol influence should be strongest (see Santer et al. 1995; Hegerl et al. 1997; Hegerl et al. 1999, manuscript submitted to *Climate Dyn.*).

Another difference found among the models is that the total percentage of variance explained by the first 20 eigenvectors varies between the models, from 66% (HadCM2) and 64% (HAM3L) to 54% (GFDL). This suggests that the variability structure present in the GFDL results is slightly different from that found in the other two models.

Common EOFs computed from SAT time series whose averaging period is slightly longer, such as decadally averaged data (not shown), yield consistent results with those presented above for the 5-yr mean variability. If the common EOFs are computed over the whole globe instead of over the observed data covered areas, the results are also quite similar to those discussed above. However, the patterns of variability tend to place more emphasis on the variance found in high latitudes. Also, the common EOFs shown here resemble the results of Barnett (1999), who computed common EOFs from 11 different coupled GCM control simulations, each 100 yr in length or longer. The results from the common EOF analysis suggest that while the overall model variability agrees fairly well with the observed variability, there is room for improvement. It is hoped that as the model resolution becomes higher and the physical parameterizations improve, the simulation of the observed variability features will improve.

4. Variability of global mean SAT

Before examining the global mean SAT time series, regression maps of the local SAT anomalies against the detrended global SAT are shown. The regression maps are used to demonstrate how the local anomalies discussed in the previous section are related to the global mean temperature anomalies. From the regression maps of the local anomalies against the global SAT computed using the 1-yr mean time series (Fig. 10), the HAM3L and GFDL patterns (Figs. 10b and 10c) are very similar, with relatively large values located over land and sea ice regions similar to the observed distribution (Fig. 10d). The values over the oceans are generally much smaller than those over the adjacent land areas. The regression also indicates that the HadCM2 1-yr SAT time series contains a very large ENSO-like signal (Fig. 10a) and is very different from the other two models. The ENSO signal in the HadCM2 model seems to involve the whole Tropics in addition to the more familiar Pacific–North American wavelike pattern. Both of the other models also show some influence of the tropical Pacific on global mean temperature but to a much smaller extent.

It is interesting to note that the pattern that is most clearly seen in the GFDL and HAM3L maps and to some extent HadCM2 (Fig. 10) is suggestive of the cold ocean–warm land, "COWL" pattern of Wallace et al. (1995). In the Wallace et al. analysis, the local anomalies about the global mean anomaly are analyzed. Here we use the regression obtained from the 1-yr mean time series of the local SAT anomalies against the global SAT anomalies to study this issue. The regression indicates that the largest local anomalies associated with a given global mean SAT anomaly are found over the continents. Therefore, if the global mean anomaly were removed as in the Wallace et al. analysis, the COWL pattern would be evident. Wallace et al. have shown that the COWL pattern is associated with short timescales
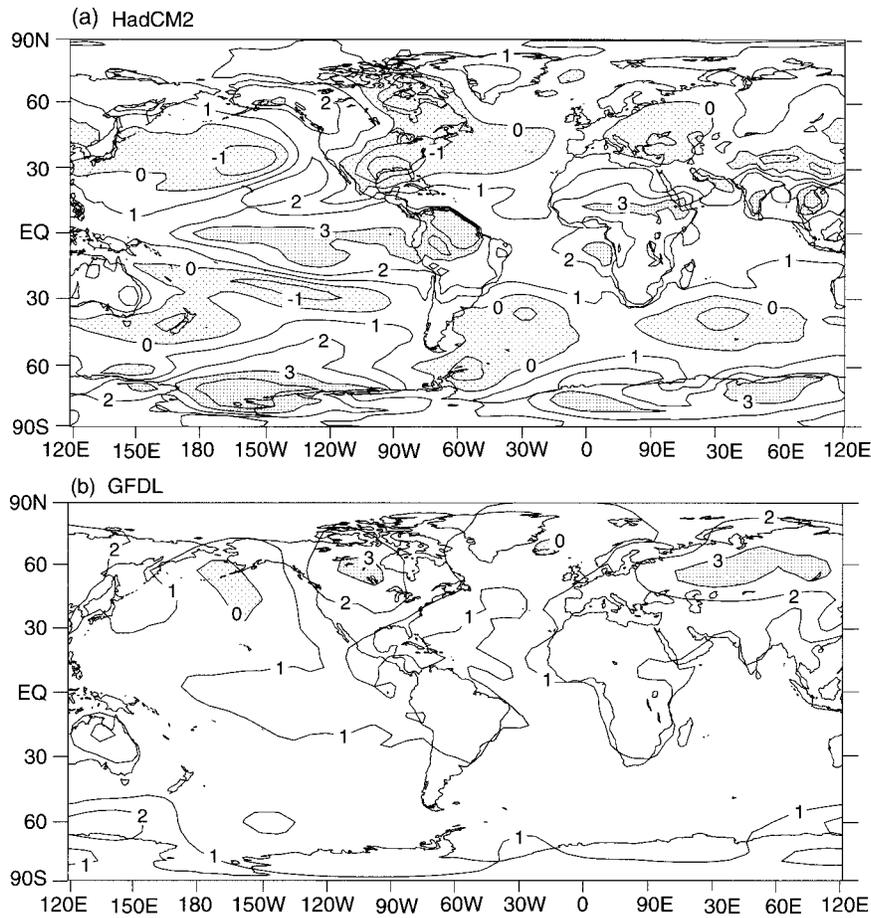
FIG. 10. Geographical distribution of the regression coefficient of the local 1-yr mean SAT anomaly against globally averaged, 1-yr mean SAT obtained from (a) HadCM2, (b) GFDL, (c) HAM3L, and (d) the observations (Jones and Briffa 1992). The regression coefficient obtained here indicates the slope of the line that is a least square fit of local vs globally averaged 1-yr mean SAT anomaly at each grid point. Shading indicates areas of negative regression coefficient.

and mainly results from wintertime dynamics in the atmosphere. They hypothesized that short-term wintertime dynamics tends to lead to a redistribution of warm and cold anomalies in the atmosphere. If the relatively warm anomalies happen to predominate over the continents and the relatively cold anomalies over the oceans, the difference in the heat capacity of the underlying surface causes larger surface temperature anomalies over the land than over the ocean, leading to a warm global mean anomaly. Broccoli et al. (1998) found that the GFDL coupled model used here simulates the observed COWL pattern very well when the SAT time series is analyzed in the same manner as Wallace et al. Furthermore, by using models of varying complexity, Broccoli et al. have confirmed that the contrast in the heat capacity between the land and the ocean is responsible for the COWL pattern.

Continuing to study how the global SAT anomalies are influenced by the local SAT anomalies, we show the regression of the local anomalies against the global

anomalies computed from the 25-yr mean time series (Fig. 11). With the exception of the tropical Pacific in the HadCM2 distribution, the values are largest in mid- and high latitudes. By comparing the regression maps obtained from the 1-yr mean (Fig. 10) and 25-yr mean (Fig. 11) SAT time series, it is seen that the tropical maxima are smaller as the timescales become longer in all the models. The high-latitude maxima generally become larger and the midlatitude maxima shift northward. Sea ice processes, internal ocean variability, and/ or changes in oceanic convection influence the SAT variability in the locations where the maxima on long timescales occur, suggesting that these processes become relatively more important in influencing the global mean SAT on the longer timescales. It is also noted that for most of the globe and especially over the land areas, the local values obtained from the regression are positive. This indicates that when the local anomalies are warmer (colder) than normal, the global mean anomalies tend to be slightly warmer (colder) than normal too.
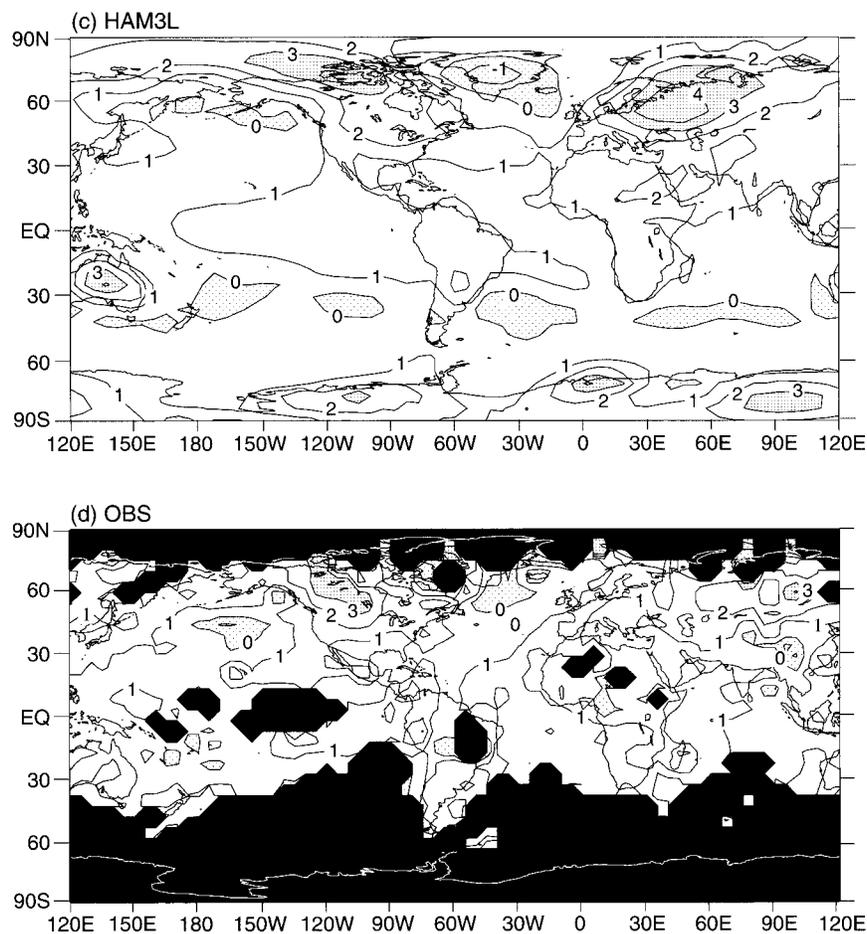
FIG. 10. (*Continued*)

The time series of global mean SAT obtained from the three models are shown in Fig. 12. The time series obtained from the HadCM2 and HAM3L models show a small warming trend over the 1000-yr time period, while the time series from the GFDL model contains a small cooling trend. One also notes that the variability appears largest in the time series obtained from the HadCM2 model. As shown earlier by the regression and common EOF patterns, this is partly the result of the whole Tropics oscillating in phase with the HadCM2 model's ENSO.

To more easily examine the variability of the global mean SAT time series on various timescales, the power spectrum has been computed from the models and compared to the observed estimates (Fig. 13). The linear trend has been removed both from the modeled and observed datasets. It is useful to keep in mind that all the modeled variability is only generated through interactions between the atmosphere, ocean, sea ice, and land surface. There are no changes in the radiative forcing on timescales longer than 1 yr in the models during these integrations. In the observations, this is not the case. Changes in the solar forcing, changes in strato-

spheric aerosols due to volcanic eruptions, and changes in land surface albedo and atmospheric chemistry due to human activities all can produce changes in the radiative forcing of the climate on various timescales. If one assumes a linear superposition of the climatic response to changes in radiative forcing in addition to the naturally occurring internally generated variability found in the observations, one would expect the observed spectral estimates to be larger than those obtained from the models. Because the observed radiative forcing is not well known, it is not possible at present to quantitatively estimate these effects. Furthermore, the observed temperature record contains many uncertainties due to various sampling problems. In addition, the removal of the trend from the observations potentially removes much of the variance on timescales longer than several decades, which are still present in the model's time series. Thus, the comparison of the simulated and observed spectral estimates must be viewed with these problems in mind.

Generally, in all three models and the observed globally averaged SAT time series, the spectral density increases as the period becomes longer (Fig. 13). This is
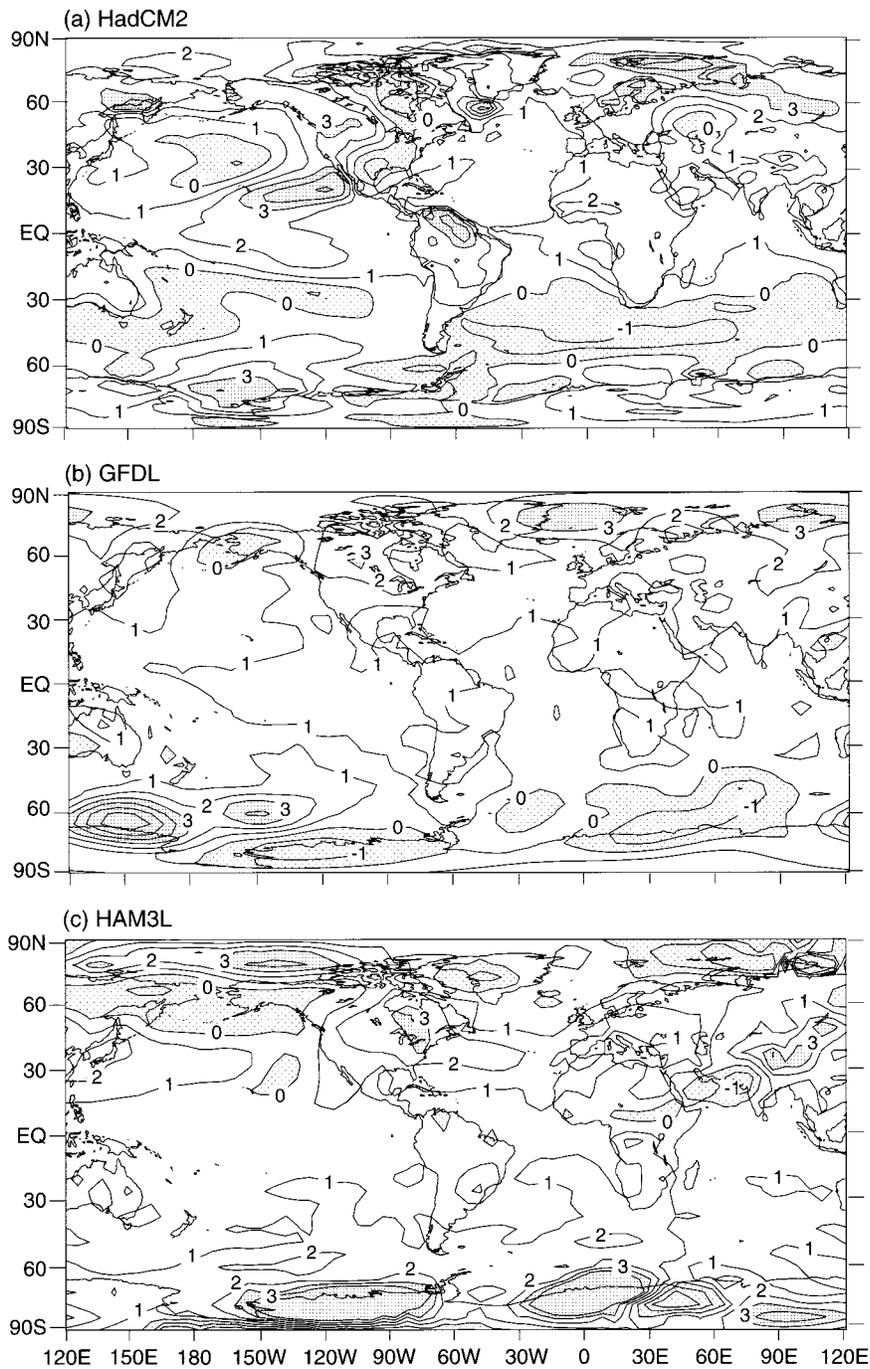
FIG. 11. Geographical distribution of the regression coefficient of the local 25-yr mean SAT anomaly on the globally averaged, 25-yr mean SAT obtained from (a) HadCM2, (b) GFDL, and (c) HAM3L. Shading indicates areas of negative regression coefficient.

consistent with the previously discussed concept that the global ocean thermally integrates atmospheric weather noise, reddening the global mean SAT spectrum. However, for periods longer than 50–100 yr, it appears that the spectral estimates obtained from the models remain approximately constant with increasing period.

There are a number of spectral peaks in the modeled and observed spectral estimates; however, none of these peaks are statistically significant. Some of the peaks have been shown to have physical significance. For example, in the observations, the peak in the observed spectral estimates around 3–5 yr has been associated with the ENSO phenomena in the tropical Pacific Ocean.
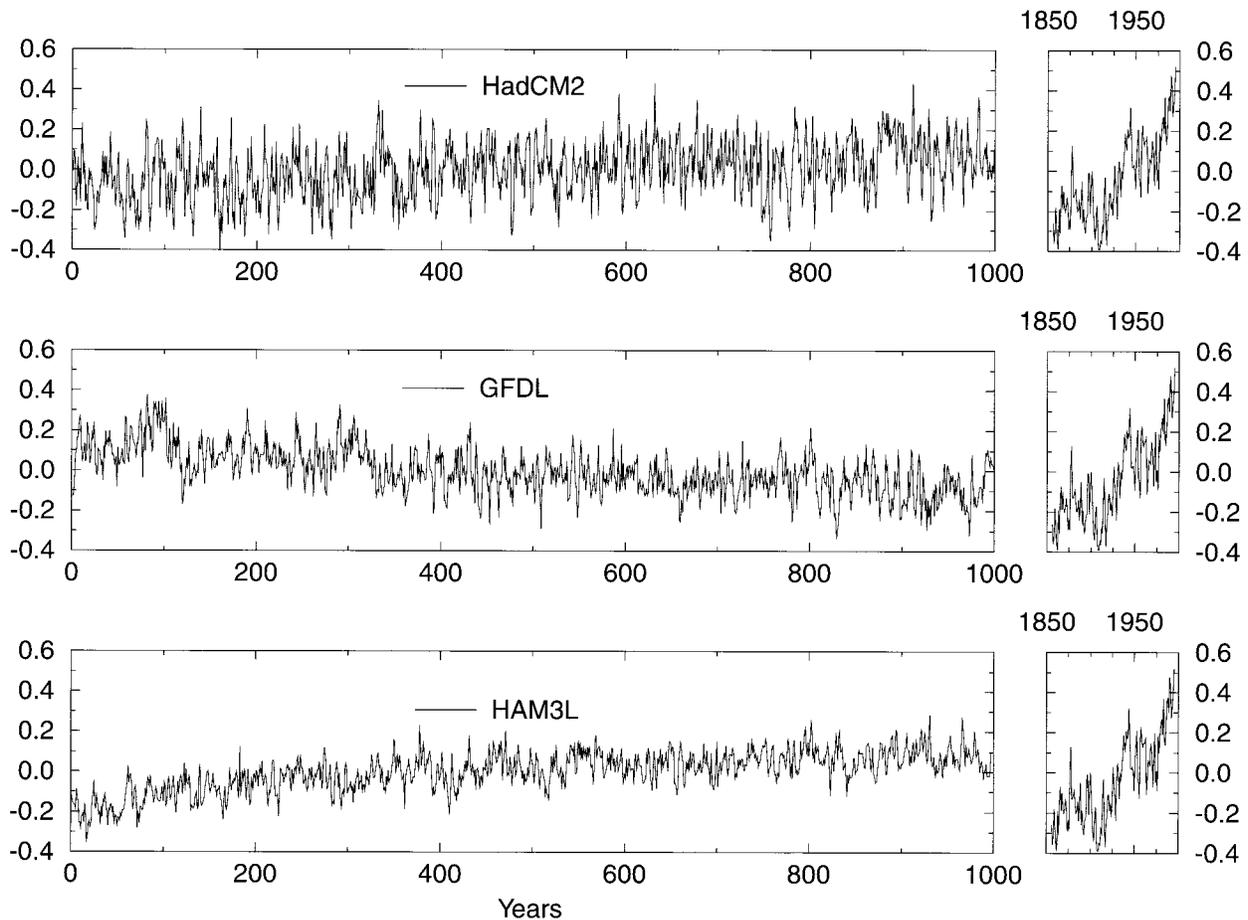
FIG. 12. Time series of globally averaged, annual mean SAT anomaly (K) from the three models (top left: HadCM2; middle left: GFDL; bottom left: HAM3L). The right plots are all identical and represent the globally averaged, SAT observations compiled by Jones and Briffa (1992).

The position and magnitude of the peak in the observed spectrum associated with ENSO is not well simulated by any of the models. The HadCM2 model, which has the most variance of the three models, has a weak spectral peak in the 6–8 yr time period. Of the three models, the power is generally largest in HadCM2 results and smallest in the HAM3L results, with the observed power intermediate between the HadCM2 and GFDL results.

The interpretation of the above results depends critically on the stationarity of the observed global mean SAT time series. From Figs. 12 and 13, it is clear that on long timescales, the model's global mean SAT is nearly stationary. If the observed time series is also stationary on these long timescales, then one could conclude that the HadCM2 model probably overestimates the global SAT variability. However, the assessment of the models' variability on timescales longer than a century awaits high quality paleotemperature and radiative forcing estimates.

5. A simple detection example

Following the example of Stouffer et al. (1994), we investigate the observed warming of the globally av-

eraged SAT. Over the past 110 yr, the observed globally averaged SAT has warmed by about 0.5 K. It is useful to assess whether or not this temperature rise could be due to internally generated variability of the climate system or if it is some radiatively forced climate change. We use the results obtained from the climate models as a surrogate for the observed long-term internal variability of the surface air temperature.

Since only one observed record is available that potentially includes both internally generated variability and radiatively forced climate change, it is very difficult to assess the model's simulation on centennial and longer long timescales as discussed above. Some preliminary results using paleodata for such an assessment indicate that secular variability in coupled climate models is smaller than in paleodata (Barnett et al. 1996). However, the use of paleodata introduces additional uncertainty and also may be influenced by external low-frequency forcing, such as changes in insolation (Cubasch et al. 1997).

For this simple detection example, we use the observed globally averaged surface temperature record.
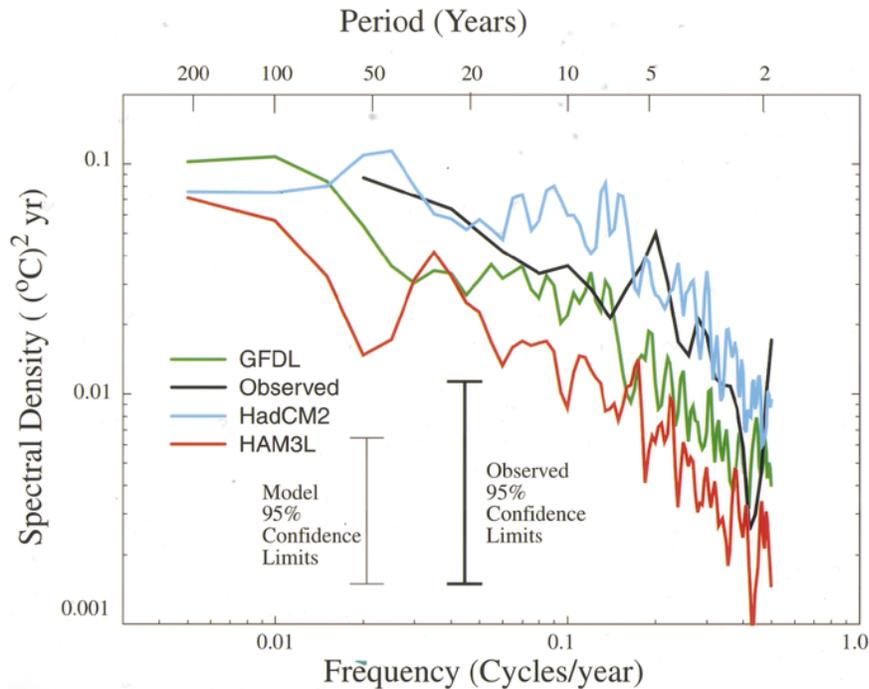
FIG. 13. Power spectra of the detrended globally averaged, annual mean SAT anomaly. The curves represent the estimates obtained from HadCM2 (blue), GFDL (green), and HAM3L (red). The observed (black line) is from the globally averaged, annual mean SAT anomalies compiled by Jones and Briffa (1992). The spectra are smoothed Fourier transforms of the autocovariance function using a Tukey window of 100 lags for the models and 30 lags for the observations. The two vertical lines represent the range of 95% confidence in the spectral estimates for the model and observations.

On the longer timescales, a temperature rise as large as that observed over the last 110 yr is not found in any of the model time series (Fig. 12). This suggests that if the model variability is realistic on these long time-scales, then the observed warming is radiatively forced and not the result of internally generated variability. This result is in agreement with findings of other detection studies (e.g., Stouffer et al. 1994) and detection/attribution studies that suggest that anthropogenic increases in greenhouse gases and sulfate aerosols are a primary explanation for the observed climate changes of the last century (Santer et al. 1995, 1996; Hegerl et al. 1996, 1997; Tett et al. 1996).

6. Summary and discussion

This paper examines the variability of SAT obtained from three long coupled atmosphere–ocean model integrations. In addition to their use in climate studies, these model integrations are the control integrations for various perturbation studies, such as those investigating the effect of increasing greenhouse gas concentrations in the atmosphere. In these control integrations, the radiative forcing repeats exactly year after year, so that any climate variations present in the integrations are due to interactions among the atmosphere, ocean, sea ice, and land surface components of the models. Since

the integrations are 1000 yr in length, variability shorter than the century timescale is fairly well sampled. By comparing the models to the observed SAT variability, it is found that the models' simulation of SAT variability compares favorably to the observed both globally and locally on timescales where the observational record allows such a comparison.

The geographical distribution of the variance computed from 1-yr mean and 5-yr mean SAT time series from the models and observations is largest over the extratropical continents and relatively small over the oceans. The tropical Pacific Ocean is the exception to this generalization. In this region, only the HadCM2 model adequately simulates the magnitude of the observed tropical SAT variability. However, the spatial extent and timescale of the HadCM2 tropical variability appear questionable. It is clear that higher-resolution ocean models and improved parameterizations are needed in order to properly simulate the variability in this region.

It should be noted that in high latitudes of the Northern Hemisphere and mid- and high latitudes of the Southern Hemisphere, there are large gaps in the observations so that the model performance cannot be evaluated. Elsewhere, the models generally do a good job of simulating the geographical distribution of the

observed variance on the 1- and 5-yr timescales with the exceptions noted above. Results from analyzing the 25-yr mean SAT time series show that on longer time-scales, the modeled variability tends to be largest over the oceans in high latitudes. These are regions near the sea ice edge and where deep to intermediate waters are formed. Both sea ice processes and/or changes in the deep oceanic mixing resulting from either coupled processes or internally generated oceanic variability may give rise to the long timescales found there. The exact location of the local SAT variability maxima are in different geographical positions from model to model.

The lag-1 autocorrelation of the SAT anomalies computed from the 1-yr mean SAT time series indicates that there is very little persistence from one year to the next over the continents. Over the oceans, relatively persistent SAT anomalies are found in the Tropics and high latitudes in the models. On longer timescales, as seen by analyzing the 5-yr mean SAT time series, persistent anomalies are mainly found in high latitudes. On timescales of 25 years or more, the differences in the exact placement of the most persistent anomalies varies from model to model but corresponds to the areas of maximum variance on the 25-yr timescale in each individual model.

A common EOF analysis of 5-yr mean data shows that the structured variability is broadly similar between the models. However, the amount of variance associated with different EOF patterns varies between the models and between the models and the observations. For example, HadCM2 and HAM3L seem to show too much variability of the same sign in Northern Hemisphere midlatitude regions. This variability structure makes the detection of a sulfate aerosol forced climate change more difficult. HAM3L and GFDL underestimate the decadal tropical variability, while HadCM2 overestimates it.

Both the common EOF analysis and the regression of the local SAT anomalies on the global mean temperature anomalies indicate that the large variability found over the continents contributes most to the variability of the global mean time series both in the models and in the observations. However, from point correlations, MS96 found that the variability over the two major land masses (Eurasia and North America) is not correlated in the GFDL model results, suggesting that the SAT anomalies over these two land masses are independent. When both land masses happen to be warm (cold), the global mean SAT anomaly is warm (cold).

Using a different analysis technique and applying it only to the observations, Wallace et al. (1995) have identified a cold land–warm ocean (COWL) pattern. This pattern is very similar to what is found when using the regression and EOF analysis. Given this similarity, it is not surprising that Broccoli et al. (1998) have found that the GFDL coupled used here simulates the observed COWL pattern very well. Their analysis demonstrates that the heat capacity of the underlying surface gives

rise to this pattern of variability. The lag-1 autocorrelation maps suggest that this pattern of relatively short timescale variability is also present in HAM3L and, to a lesser extent, in HadCM2.

The assessment of the quality of the model's variability simulation on longer than decadal timescales is very difficult as discussed earlier. The results presented here and by MS96 suggest that as the timescales become longer, the oceans play an increasingly important role in generating variability. The horizontal resolution of the oceanic component of the coupled models used here is coarse and does not resolve many oceanic processes such as oceanic eddies. The effects of the oceanic eddies are parameterized by subgrid-scale mixing schemes and the use of relatively large diffusion coefficients. The use of these subgrid-scale parameterizations may have an adverse effect on the simulation of oceanic variability. In addition, the observed climate sensitivity is not well known. The models used here lie in the 1.5–4.5 K range and above the Intergovermental Panel on Climate Change best guess value of 2.5 K for a doubling of the atmospheric CO_2 concentration (Kattenberg et al. 1996). Much of the uncertainty in this estimate is due to the radiative effects of clouds, which are controlled by subgrid processes in the atmospheric models. Therefore, it is important that any conclusions on the detection of climate change be reexamined with much higher resolution coupled models in the future when the influence of the subgrid-scale parameterizations should be smaller.

Overall, the quality of the models' simulation of the observed SAT variability on decadal timescales and shorter encouraged us to use these long model integrations to study the observed globally averaged SAT warming of about 0.5 K over the past century. In order to perform meaningful detection or attribution studies of human impact on the observed climate, measures of the internally generated variability on the long timescales are needed.

For this simple detection example, we use the models' simulated long-term SAT variability, which includes no changes in radiative forcing, as a surrogate for the observed variability. A temperature trend as large as the observed is not found in any of these model integrations. If the models' simulation of variability on long timescales is realistic, then the observed warming must be due to changes in the radiative forcing of the planet and not the result of internally generated variability.

REFERENCES

Barnett, T. P., 1999: Comparison of near-surface air temperature variability in eleven coupled global climate models. *J. Climate,* **12,** 511–518.

——, B. D. Santer, P. D. Jones, R. S. Bradley and K. R. Briffa, 1996: Estimates of low frequency natural variability in near-surface air temperature. *Holocene,* **6,** 255–263.

Broccoli, A. J., N.-C. Lau, and M. J. Nath, 1998: The cold ocean–warm land pattern in long integrations of a coupled atmosphere–ocean GCM. *J. Climate,* **11,** 2743–2763.

Cubasch, U., K. Hasselmann, H. Höck, E. Maier-Reimer, U. Mikolajewicz, B. D. Santer, and R. Sausen, 1992: Time-dependent greenhouse warming computations with a coupled ocean–atmosphere model. *Climate Dyn.,* **8,** 55–69.

——, G. C. Hegerl, and J. Waszkewitz, 1996: Prediction, detection and regional assessment of anthropogenic climate change. *Geophysica,* **32,** 77–96.

——, ——, R. Voss, J. Waszkewitz, and T. J. Crowley, 1997: Simulation of the influence of solar radiation variations on the global climate with an ocean–atmosphere general circulation model. *Climate Dyn.,* **13,** 757–767.

Delworth, T. L., S. Manabe, and R. J. Stouffer, 1993: Interdecadal variations of the thermohaline circulation in a coupled ocean–atmosphere model. *J. Climate,* **6,** 1993–2011.

——, ——, and ——, 1997: Multidecadal climate variability in the Greenland Sea and surrounding regions: A coupled model simulation. *Geophys. Res. Lett.,* **24,** 257–260.

Folland, C. K., T. R. Karl, N. Nicholls, B. S. Nyenzi, D. E. Parker, and K. Ya. Vinnikov, 1992: Observed climate variability and change. *Climate Change 1992. The Supplementary Report to the IPCC Scientific Assessment,* J. T. Houghton, B. A. Callander and S. K. Varney, Eds., Cambridge University Press, 135–170.

Frankignoul, C., C. Duchene, and M. Cane, 1989: A statistical approach to testing equatorial ocean models with observed data. *J. Phys. Oceanogr.,* **19,** 1191–1208.

Gates, W. L., J. F. B. Mitchell, G. J. Boer, U. Cubasch, and V. P. Meleshko, 1992: Climate modeling, climate prediction and model validation. *Climate Change 1992. The Supplementary Report to the IPCC Scientific Assessment.* J. T. Houghton, B. A. Callander and S. K. Varney, Eds., Cambridge University Press, 97–134.

Gregory, J. M., and J. F. B. Mitchell, 1997: The climate response to CO_2 of the Hadley Centre coupled AOGCM with and without flux adjustment. *Geophys. Res. Lett.,* **24,** 1943–1946.

Hall, A., and S. Manabe, 1997: Can local linear stochastic theory explain sea surface temperature and salinity variability? *Climate Dyn.,* **13,** 167–180.

Hasselmann, K., 1976: Stochastic climate models, Pt. 1, Theory. *Tellus,* **28,** 473–485.

——, and Coauthors, 1995: Detection of anthropogenic climate change using a fingerprint method. MPI Report No. 168, 20 pp. [Available from Max-Planck-Institut für Meteorologie, Bundesstrasse 55, 20147 Hamburg, Germany.]

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

Hegerl, G. C., H. von Storch, K. Hasselmann, B. D. Santer, U. Cubasch, and P. D. Jones, 1996: Detecting greenhouse gas induced climate change with an optimal fingerprint method. *J. Climate,* **9,** 2281–2306.

——, K. Hasselmann, U. Cubasch, J. F. B. Mitchell, E. Roeckner, R. Voss, and J. Waszkewitz, 1997: Multi-fingerprint detection and attribution of greenhouse gas, greenhouse gas-plus-aerosol, and solar forced climate change. *Climate Dyn.,* **13,** 613–634.

Hewitt, C. D., and J. F. B. Mitchell, 1996: GCM simulations of the climate of 6 kyr BP: Mean changes and interdecadal variability. *J. Climate,* **9,** 3505–3529.

——, and ——, 1998: A fully coupled GCM simulation of the climate of the mid-Holocene. *Geophys. Res. Lett.,* **25,** 361–364.

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

Jones, P. D., 1994a: Recent warming in global temperature series. *Geophys. Res. Lett.,* **21,** 1149–1152.

——, 1994b: Hemispheric surface air temperature variations: A reanalysis and an update to 1993. *J. Climate,* **7,** 1794–1802.

——, and K. R. Briffa, 1992: Global surface air temperature variations during the twentieth century: Part 1, spatial, temporal and seasonal details. *Holocene,* **2,** 165–179.

——, and G. C. Hegerl, 1998: Comparisons of two methods of removing ''anthropogenic-related'' variability from the near-surface observational temperature field. *J. Geophys. Res.,* **103,** 13 777–13 786.

Kattenberg, A., and Coauthors, 1996: *Climate Change 1995: The Science of Climate Models—Projections of Future Change.* J. T. Houghton et al., Eds., Cambridge University Press, 285–357.

Keen, A. B., and J. M. Murphy, 1997: Influence of natural variability and the cold start problem on the simulated transient response to increasing CO_2. *Climate Dyn.,* **13,** 847–864.

Kim, K.-Y., G. R. North, and G. C. Hegerl, 1996: Comparisons of the second-moment statistics of climate models. *J. Climate,* **9,** 2204–2221.

Knutson, T. R., S. Manabe, and D. Gu, 1997: Simulated ENSO in a global coupled ocean–atmosphere model: Multidecadal amplitude modulation and CO_2 sensitivity. *J. Climate,* **10,** 138–161.

Legates, D. R., and C. J. Willmott, 1990: Mean seasonal and spatial variability in global surface air temperature. *Theor. Appl. Climatol.,* **41,** 11–21.

Maier-Reimer, E., U. Mikolajewicz, and K. Hasselmann, 1993: Mean circulation of the Hamburg LSG model and its sensitivity to thermohaline surface forcing. *J. Phys. Oceanogr.,* **23,** 731–757.

Manabe, S., and R. J. Stouffer, 1988: Two stable equilibria of a coupled ocean–atmosphere model. *J. Climate,* **1,** 841–866.

——, and ——, 1993: Century-scale effects of increased atmospheric CO_2 on the ocean–atmosphere system. *Nature,* **364,** 215–218.

——, and ——, 1994: Multiple-century response of a coupled ocean–atmosphere model to an increase of atmospheric carbon dioxide. *J. Climate,* **7,** 5–23.

——, and ——, 1995: Simulation of abrupt climate change induced by freshwater input to the North Atlantic Ocean. *Nature,* **378,** 165–167.

——, and ——, 1996: Low-frequency variability of surface air temperature in a 1000-year integration of a coupled atmosphere–ocean–land surface model. *J. Climate,* **9,** 376–393.

——, and ——, 1997: Coupled ocean–atmosphere model response

to freshwater input: Comparison to Younger Dryas event. *Paleoceanography,* **12,** 321–336.

——, ——, M. J. Spelman, and K. Bryan, 1991: Transient responses of a coupled ocean–atmosphere model to gradual changes of atmospheric CO$_2$. Part I: Annual mean response. *J. Climate,* **4,** 785–818.

Mitchell, J. F. B., and T. C. Johns, 1997: On modification of global warming by sulfate aerosols. *J. Climate,* **10,** 245–267.

——, ——, J. M. Gregory, and S. F. B. Tett, 1996: Climate response to increasing levels of greenhouse gases and sulphate aerosols. *Nature,* **376,** 501–504.

North, G. R., T. L. Bell, R. F. Cahalan, and F. J. Moeng, 1982: Sampling errors in the estimation of empirical orthogonal functions. *Mon. Wea. Rev.,* **110,** 699–706.

Roeckner, E., and Coauthors, 1992: Simulation of the present-day climate with the ECHAM model: Impact of model physics and resolution. Report No. 93, 171 pp. [Available from Max-Planck-Institut für Meteorologie, Bundesstrasse 55, 20147 Hamburg, Germany.]

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

——, and Coauthors, 1996: A search for human influences on the thermal structure of the atmosphere. *Nature,* **382,** 39–46.

Sarmiento, J. L., and C. Le Quéré, 1996: Oceanic carbon dioxide uptake in a model of century-scale global warming. *Science,* **274,** 1346–1350.

——, T. M. C. Hughes, R. J. Stouffer, and S. Manabe, 1998: Response of the ocean carbon cycle to anthropogenic climate warming. *Nature,* **393,** 245–249.

Schiller, A., U. Mikolajewicz, and R. Voss, 1997: The stability of the North Atlantic thermohaline circulation in a coupled ocean–atmosphere general circulation model. *Climate Dyn.,* **13,** 325–347.

Stouffer, R. J., and S. Manabe, 1999: Response of a coupled ocean–atmosphere model to increasing atmospheric carbon dioxide: Sensitivity to the rate of increase. *J. Climate,* **12,** 2224–2237.

——, ——, and K. Bryan, 1989: Interhemispheric asymmetry in climate response to a gradual increase of atmospheric CO$_2$. *Nature,* **342,** 660–662.

——, ——, and K. Ya. Vinnikov, 1994: Model assessment of the role of natural variability in recent global warming. *Nature,* **367,** 634–636.

Tett, S. F. B., J. F. B. Mitchell, D. E. Parker, and M. R. Allen, 1996: Human influence on the atmospheric vertical temperature structure: Detection and observations. *Science,* **247,** 1170–1173.

——, T. C. Johns, and J. F. B. Mitchell, 1997: Global and regional variability in a coupled AOGCM. *Climate Dyn.,* **13,** 303–323.

Timmermann, A., M. Latif, R. Voss, and A. Groetzner, 1999: Modes of climate variability as simulated by the Coupled General Circulation Model. Part 1: ENSO-like variability and its low frequency variability. *Climate Dyn.,* **15,** 605–618.

Verbeek, J., 1997: Wind stress and SST variability in the North Atlantic area: Observations and five coupled GCMs in concert. *Mon. Wea. Rev.,* **125,** 942–957.

von Storch, J., V. Kharim, U. Cubasch, G. C. Hegerl, D. Schriever, H. von Storch, and E. Zorita, 1997: A description of a 1260-year control integration with the coupled ECHAM1/LSG general circulation model. *J. Climate,* **10,** 1525–1543.

Voss, R., and U. Mikolajewicz, 1999: Long-term climate changes due to increasing CO$_2$ concentration in the coupled atmosphere–ocean general circulation model. ECHAM3/LSG Rep. 298, 50 pp. [Available from Max-Planck Institut für Meteorologie, Bundesstrasse 55, 20147 Hamburg, Germany.]

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

Wallace, J. M., Y. Zhang, and J. A. Renwick, 1995: Dynamic contribution to hemispheric mean temperature trends. *Science,* **270,** 780–783.

Zhang, Y., J. M. Wallace, and D. S. Battisti, 1997: ENSO-like interdecadal variability: 1900–93. *J. Climate,* **10,** 1004–1020.