Assessment of Twentieth-Century Regional Surface Temperature Trends Using the GFDL CM2 Coupled Models

T. R. KNUTSON, T. L. DELWORTH, K. W. DIXON, I. M. HELD, J. LU,* V. RAMASWAMY, AND M. D. SCHWARZKOPF

Geophysical Fluid Dynamics Laboratory/NOAA, Princeton, New Jersey

G. STENCHIKOV

Rutgers–The State University of New Jersey, New Brunswick, New Jersey

R. J. STOUFFER

Geophysical Fluid Dynamics Laboratory/NOAA, Princeton, New Jersey

(Manuscript received 3 May 2005, in final form 27 September 2005)

ABSTRACT

Historical climate simulations of the period 1861–2000 using two new Geophysical Fluid Dynamics Laboratory (GFDL) global climate models (CM2.0 and CM2.1) are compared with observed surface temperatures. All-forcing runs include the effects of changes in well-mixed greenhouse gases, ozone, sulfates, black and organic carbon, volcanic aerosols, solar flux, and land cover. Indirect effects of tropospheric aerosols on clouds and precipitation processes are not included. Ensembles of size 3 (CM2.0) and 5 (CM2.1) with all forcings are analyzed, along with smaller ensembles of natural-only and anthropogenic-only forcing, and multicentury control runs with no external forcing.

Observed warming trends on the global scale and in many regions are simulated more realistically in the all-forcing and anthropogenic-only forcing runs than in experiments using natural-only forcing or no external forcing. In the all-forcing and anthropogenic-only forcing runs, the model shows some tendency for too much twentieth-century warming in lower latitudes and too little warming in higher latitudes. Differences in Arctic Oscillation behavior between models and observations contribute substantially to an underprediction of the observed warming over northern Asia. In the all-forcing and natural-only forcing runs, a temporary global cooling in the models during the 1880s not evident in the observed temperature records is volcanically forced. El Niño interactions complicate comparisons of observed and simulated temperature records for the El Chichón and Mt. Pinatubo eruptions during the early 1980s and early 1990s.

The simulations support previous findings that twentieth-century global warming has resulted from a combination of natural and anthropogenic forcing, with anthropogenic forcing being the dominant cause of the pronounced late-twentieth-century warming. The regional results provide evidence for an emergent anthropogenic warming signal over many, if not most, regions of the globe. The warming signal has emerged rather monotonically in the Indian Ocean/western Pacific warm pool during the past half-century. The tropical and subtropical North Atlantic and the tropical eastern Pacific are examples of regions where the anthropogenic warming signal now appears to be emerging from a background of more substantial multidecadal variability.

* UCAR Visiting Scientist.

Corresponding author address: Thomas R. Knutson, GFDL/NOAA, Forrestal Campus, U.S. Route 1, Princeton, NJ 08542.
E-mail: Tom.Knutson@noaa.gov

1. Introduction

The earth's mean surface temperature has warmed by ~0.6°C over the past century according to historical temperature records, supported by other climate observations (Folland et al. 2001). The purpose of the present study is to compare observed surface temperature changes to those produced by two versions of a new global climate model forced by estimated historical changes in a number of climate forcing agents. Such comparisons assess our ability to interpret past climate variations in terms of known climate forcing agents based on our physical understanding of the earth's climate system as embodied in the coupled climate models.

In this study, we use the Geophysical Fluid Dynamics Laboratory's (GFDL's) new CM2 coupled climate models (CM2.0 and CM2.1; Delworth et al. 2006) to simulate surface temperature variations over the period 1861–2000. The climate forcing agents included in the simulations include changes in well-mixed greenhouse gases, ozone, anthropogenic aerosols (direct effect only), solar irradiance, land cover type, and volcanic aerosols. Ensembles of all-forcing experiments are run for both CM2.0 ($n = 3$) and CM2.1 ($n = 5$) coupled models, along with additional natural-only and anthropogenic-only forcing experiments ($n = 1$ for CM2.0, $n = 3$ for CM2.1). Multicentury control integrations without external forcing are used to assess internal climate variability and reduce impacts of any model drifts unrelated to external forcing changes.

Global climate model historical experiments using similar sets of anthropogenic and natural forcings (~1860–2000) have been performed previously (e.g., Tett et al. 1999; Stott et al. 2000; Broccoli et al. 2003; Meehl et al. 2004; Hansen et al. 2005). These studies provide model-based evidence that both natural and anthropogenic forcings made significant contributions to early twentieth-century surface temperature changes, with anthropogenic forcings (e.g., greenhouse gases) being the dominant cause of the warming in the second half of the twentieth century. Delworth and Knutson (2000) found that internal climate variability also could have played a substantial role in the observed early twentieth-century warming.

The present study represents a substantial step beyond previous work at GFDL on this topic (e.g., Knutson et al. 1999; Delworth and Knutson 2000; Broccoli et al. 2003) in two important respects. First, the experiments in this paper include a more comprehensive and physically based set of climate forcing agents than our previous studies. Second, we use two versions of a completely updated global coupled climate model developed over a period of several years at GFDL (Delworth et al. 2006). The new models are substantially improved over previous GFDL coupled climate models in that they provide much-improved simulations of El Niño variability (Wittenberg et al. 2006), and they do not use flux adjustments. The new models include more realistic treatments of sea ice, land surface processes, and other physical processes in the atmosphere and oceans (Delworth et al. 2006 and references therein) than the previous models.

For comparing model-simulated and observed temperature variations, one has the choice of using multivariate "fingerprint" detection and attribution techniques (e.g., Hegerl et al. 1997) or simpler univariate methods based on local trends (e.g., Knutson et al. 1999; Boer et al. 2000; Karoly and Wu 2005) or climate indices (e.g., Karoly et al. 2003). A review of various methodologies and conclusions obtained through their application to previous coupled models is provided by Mitchell et al. (2001) in chapter 12 of the Intergovernmental Panel on Climate Change (IPCC) Working Group 1 Third Assessment Report (TAR; see also International Ad Hoc Detection and Attribution Group 2005). Multivariate methods enhance one's chances of detecting and attributing climate change, while a univariate analysis of the significance of local trends retains the advantage of being relatively easy to understand and communicate to nonspecialists. In addition to the presentation of trend maps and their significance, as in Knutson et al. (1999), we have also included an extensive set of figures showing time series case studies for a number of regions of interest. This approach avoids the dependence of the trends on the selection of starting/ending dates. Concerning regional-scale analysis, recent studies have applied formal multivariate detection/attribution methods at decreasing spatial scales and report evidence for detectable anthropogenic warming signals down to the continental scale (Zwiers and Zhang 2003; Stott 2003). Finally, while we focus on surface temperature changes using a single pair of models in this report, a number of studies (not reviewed here) have begun to assess twentieth-century climate changes using multimodel ensembles and climate variables other than surface temperature.

The paper is structured as follows. In sections 2 and 3, we present a brief overview of the main characteristics of the models and the climate forcing agents used in the historical simulations. In section 4, characteristics of the model control runs (without changes in climate forcings) are examined. In section 5, we examine changes in global mean temperature in the historical forcing runs. In section 6, we compare the simulated and observed surface temperature changes on a re-

gional basis. Arctic Oscillation influences are investigated in section 7. Section 8 contains our summary and conclusions.

2. Model description

The two coupled models used for the present study (CM2.0 and CM2.1) are described in detail in Delworth et al. (2006) and references therein. Further information is available online at http://nomads.gfdl.noaa.gov/CM2.X/references/, and model output data for a large portion of the experiments are available at http://nomads.gfdl.noaa.gov/. These coupled models are composed of four component models: atmosphere, land, sea ice, and ocean. The coupling between the component models (V. Balaji 2005, personal communication) occurs at 1- and 2-h intervals in CM2.0 and CM2.1, respectively, which couples the diurnal cycles of the atmosphere and ocean components.

The climate sensitivity of the models to a doubling of CO_2 is 2.9°C for CM2.0 and 3.4°C for CM2.1 (Stouffer et al. 2006). This was determined from the global mean surface air temperature difference between control and $2 \times CO_2$ experiments in which the oceanic components of the coupled models were replaced with a static "slab" ocean model. In these experiments, the original atmosphere, land, and sea ice components from CM2.0 and CM2.1 were retained. Heat flux adjustments were specified at the air–sea interface such that a realistic SST/sea ice climatology was maintained in the coupled slab model control runs. Identical adjustments were used in the $2 \times CO_2$ experiments. Further details will be reported elsewhere. The transient climate response (TCR) of the coupled models (with full ocean dynamics) was determined from the change in global mean surface air temperature around the time of CO_2 doubling in $+1\%\ yr^{-1}$ compounded CO_2 increase experiments. The TCR was about 1.6°C for both models (Stouffer et al. 2006).

The atmospheric model has a grid spacing of 2.5° longitude by 2° latitude and 24 vertical levels. The model contains a completely updated suite of model physics compared to the previous GFDL climate model, including new cloud prediction and boundary layer schemes, and diurnally varying solar insolation. The radiation code allows for explicit treatment of numerous radiatively important trace gases (including tropospheric and stratospheric ozone, halocarbons, etc.), a variety of natural and anthropogenic aerosols (including black carbon, organic carbon, tropospheric sulfate aerosols, and volcanic aerosols), and dust particles. Aerosols in the model do not interact with the cloud scheme, so that indirect aerosol effects on climate are not considered. A full description of the atmospheric model is contained in GFDL Global Atmospheric Model Development Team (2004) with updates as described in Delworth et al. (2006). CM2.1 and CM2.0 have essentially the same horizontal resolution, but differ in several aspects including different dynamical cores [finite volume (Lin 2004) in CM2.1 versus a B-grid finite difference dynamical core in CM2.0], further tuning of the cloud scheme for CM2.1, and a modified formulation of evaporation from land grid points with frozen soil (see Delworth et al. 2006).

The land model used in both CM2.0 and CM2.1 is the Land Dynamics model (LaD) as described in Milly and Shmakin (2002). Surface water is routed to ocean destination points on the basis of specified drainage basins. The land cover type in the model uses a classification scheme with 10 different land cover types.

The ocean model (Gnanadesikan et al. 2006; Griffies et al. 2005) has a nominal grid spacing of 1° in latitude and longitude, with meridional grid spacing decreasing in the Tropics to 1/3° near the equator, and uses a tripolar grid to avoid polar filtering over the Arctic. The model has 50 vertical levels, including 22 levels with 10-m thickness each in the top 220 m. A novel aspect is the use of a true freshwater flux boundary condition. The ocean components of CM2.0 and CM2.1 differ in parameter settings for some subgrid-scale physics and time stepping (Delworth et al. 2006).

The sea ice model, identical in the two models, is a dynamical model with three vertical layers and five ice thickness categories. The model uses the elastic viscous plastic rheology to calculate ice internal stresses, and a modified Semtner three-layer scheme for thermodynamics (Winton 2000).

For comparison to observed surface temperatures, the HadCRUT2v dataset (available online at http://www.cru.uea.ac.uk/cru/data/temperature/) is used. This dataset (1870–2004) combines the land surface air temperature data of Jones and Moberg (2003) with the HadSST1 sea surface temperature (SST) data of Parker et al. (1995) and Rayner et al. (2003), where variance adjustments have been applied to both land and ocean data (Jones et al. 2001). Trends in this study refer to simple linear trends computed using least squares regression.

3. Climate forcings for historical runs

For the historical forcing simulations described in this report, the models were integrated for a period of 140 yr using a time-varying specification of various climate forcing agents representative of conditions from 1861–2000. An ensemble of such integrations was created for each model using initial conditions obtained from widely separated (40 yr +) points in multicentury

control integrations with fixed 1860 radiative forcing and land surface conditions.

The time-varying forcing agents included CO_2, CH_4, N_2O, halocarbons, tropospheric and stratospheric O_3, anthropogenic tropospheric sulfates, black and organic carbon, volcanic aerosols, solar irradiance, and the distribution of land cover types. The time variations were based on a combination of observations and reconstructions for the late nineteenth and twentieth centuries. The change in net radiative forcing at the tropopause from 1860 to 2000 in the all-forcing experiments is 2.8 W m^{-2}. Atmospheric dust and sea salt concentrations are specified as a function of season but do not change from year to year. A brief description of the land cover, solar, and volcanic aerosol datasets is given below, as well as some comments on the (omitted) indirect aerosol effects. A comprehensive discussion of the data sources for the time variations and other details of all the forcings and their uncertainties are provided in V. Ramaswamy et al. (unpublished manuscript; see also http://nomads.gfdl.noaa.gov/CM2.X/faq/question_13html and http://nomads.gfdl.noaa.gov/CM2.X/references/).

The land cover–type change history is based on the Hurtt et al. (2006) global land use reconstruction history. Changes in land cover type affect the model's surface albedo, surface roughness, stomatal resistance, and effective water capacity. Solar variations over the period 1882–2000 are implemented as a function of wavelength, using data provided by J. Lean (Lean et al. 1995; J. Lean 2003, personal communication; see also Houghton et al. 2001).

The volcanic aerosol dataset was developed based on volcanic aerosol optical depth for visible band compiled from different sources by Sato et al. (1993). This dataset was subsequently improved by Hansen et al. (2002) and the column-averaged zonal mean effective radius was provided for the entire period. During the satellite era the aerosol characteristics are mostly based on Stratospheric Aerosol and Gas Experiment (SAGE) observations. Using these data, the aerosol extinction, single scattering albedo, and asymmetry parameter were calculated for the entire spectrum following Stenchikov et al. (1998).

As mentioned previously, the model runs do not include the indirect effects of aerosols [see Lohmann and Feichter (2005) for definitions and detailed discussion]. While the first indirect effect (cloud albedo) likely leads to a large negative forcing value at the tropopause or top of atmosphere (-1 W m^{-2}), the full indirect effect (i.e., first + semidirect + second effects together) leads to a flux change that could be substantially different from -1 W m^{-2}. Therefore, it is difficult to have confidence in speculations on the response to the omitted indirect aerosol effects, particularly since cloud feedbacks are also involved in the indirect effect. However, there is evidence that the forcing is likely to be negative in Northern Hemisphere midlatitudes (Ming et al. 2005). Of equal or even greater importance than the tropopause/top-of-atmosphere forcing could be the surface forcing due to the indirect effect, but this too has great uncertainty associated with it.

4. Control simulations

Multicentury control runs of both the CM2.0 and CM2.1 models, with constant radiative forcing agents and land cover type appropriate to 1860 conditions, were performed as a preliminary step. For these integrations, the model ocean was initialized from observed (Levitus) conditions and integrated for 300 (CM2.0) or 220 (CM2.1) years using 1860 radiative forcing and land surface conditions. The end of this spinup was used as the initial condition for the control runs described here. Further details of the initialization of the runs are given in Stouffer et al. (2006) and Delworth et al. (2006), and are similar to that described in Stouffer et al. (2004). The control runs provide an estimate of the climate model's internal variability (variability in the absence of external forcing changes) and an estimate of any long-term drift in the model that results from the coupled model's climate not being fully equilibrated with the 1860 forcing. The long control runs also provide widely separated initial conditions (particularly three-dimensional ocean initial conditions) to use to initiate independent ensemble members for the historical forcing scenarios. This statistical sampling of control run ocean initial conditions is necessary because the true three-dimensional state of the ocean from 1860 is poorly constrained by observations. A separate pair of control runs with constant 1990 climate forcings was also completed, and is used in this paper only for the maps comparing observed and simulated local interannual variability. The 1990 control runs were used for this comparison since the observations are from this general time period.

a. Global mean temperature variability

Figure 1 summarizes a number of aspects of the CM2.0 and CM2.1 control runs used in the study in terms of global mean reference atmospheric temperature at 2 m above the surface (T_ref). The CM2.0 control run (Fig. 1a) consists of 500 yr of simulation, which was preceded by a 300-yr spinup period (Delworth et al. 2006), of which the last 90 yr is shown in Fig. 1a. The time series of T_ref shows that the CM2.0 model continues to warm at a rate of about 0.2°C (100 yr)$^{-1}$ for at

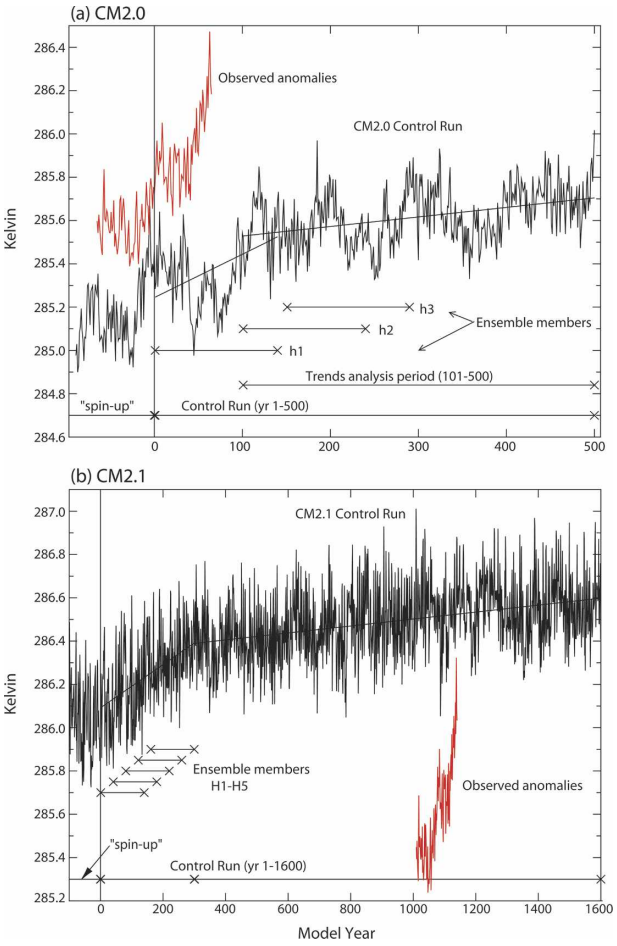


FIG. 1. Time series of global mean reference temperature (T_ref) from (a) CM2.0 control run and (b) CM2.1 control run (black curves). The red curves in each diagram are the observed surface temperature anomalies from HadCRUT2v (see text). The observations are offset by an arbitrary constant and are presented only as a reference trend comparison for the model results. The straight line segments connecting "x" marks depict different segments of the control run, including the segments corresponding to the historical forcing runs. The line segments superimposed on the time series illustrate the periods used for detrending of the post spinup segments of the control run. The spinup segments correspond to preliminary parts of the integration that are not analyzed. Note that the two diagrams have time axes of different lengths.

least 100 yr past the initial spinup period. After about year 100, the model continues to warm but at a much slower rate. This latter period (years 101–500) of fairly constant, more moderate drift is used as the trend analysis period for the CM2.0 runs, except as described in the text.

Figure 1a shows the three 140-yr periods of the CM2.0 control run that correspond to the three 140-yr all-forcing historical runs for that model (h1–h3). Note that the first CM2.0 historical run (h1) was initiated from a point in the control run (year 1) in which the model was continuing to drift warm at a substantial rate. In retrospect, a better choice of initial condition for h1 would have been after year 100 of the control run. However, the need to have several completed historical scenario runs for an IPCC model intercomparison project necessitated the choices made. To adjust for the effects of control run drift, the 140-yr time series from historical run h1 are adjusted by subtracting the trend of the control run over the same 140-yr time interval. Historical runs h2 and h3 are adjusted by subtracting the trend of the control run over the longer subsequent period (years 101–500).

The evolution of global mean T_ref in the CM2.1 control run is shown in Fig. 1b. Following a precontrol run spinup period of 220 yr (of which only the last 100 yr are shown), the model exhibits a moderate warming trajectory of about 0.1°C $(100\ \text{yr})^{-1}$ for years 1–300 of the control run, followed by a more gradual warming trend during years 301–1600. For our trends analysis in this study, we consider these two epochs (years 1–300 and 301–1600) separately, and remove the long-term drift from them separately. Note that all five all-forcing historical runs for CM2.1 (H1–H5) were run during the period in which the control run had a moderate but relatively stable degree of drift. These historical runs are adjusted for control run drift by subtracting the trend of the control run over years 1–300.

The red curves in Figs. 1a,b are the observed global mean annual mean temperature anomaly series based on the HadCRUT2v dataset. An arbitrary vertical offset has been added to the anomaly series to display it for reference alongside each control run annual mean temperature series. The pronounced warming in the observed global mean series clearly exceeds the internal variability of either the CM2.0 or the CM2.1 control runs. This is evident in Fig. 1 after accounting for the long-term adjustments mentioned above, by comparing observations with the later (relatively low drift) centuries of the control runs. This finding is consistent with numerous previous studies (e.g., Stouffer et al. 2000), which find that the observed global mean surface temperature warming exceeds changes due to internal climate variability as simulated by long control runs of several global coupled climate models.

b. Geographical distribution of interannual variability

A preliminary assessment of a climate model's internal variability can be made by comparing maps of the local interannual standard deviation of surface temperature between model and observations. The standard deviation of annual means includes variance on

time scales from two years out to the length of the record, and thus includes substantial contributions from time scales that are much shorter than the multidecadal trends that are the main focus of our study. Nonetheless, the comparison of interannual standard deviations provides a useful benchmark before focusing on longer time scales.

Figure 2 shows the interannual standard deviations for observations (Fig. 2a), CM2.0 (Fig. 2b), and CM2.1 (Fig. 2c). The model fields are based on SST (ocean regions) and T_ref (land regions) for years 101–200 for versions of the CM2.0 and CM2.1 control runs with constant 1990 climate forcings. Thus, the observed variability maps will include some contributions from time-varying climate forcing agents that are not incorporated in the control simulations. All observed and modeled series in Fig. 2 have been detrended using linear least squares regression. The observed field is based on years 1949–2003 from the combined SST–land surface temperature HadCRUT2v dataset. While there is a similar overall pattern of the standard deviation fields for the models and observations, with enhanced variability over continental regions relative to the surrounding oceans, and locally enhanced SST variability in the tropical Pacific associated with El Niño, a clear deficiency of the models is the greater than observed simulated interannual variability over many land regions.

The simulated El Niño region SST variability is excessive in magnitude, particularly in CM2.1, and in both models the most pronounced El Niño variability is displaced westward from the observed maximum location near the South American coast. However, as detailed in Wittenberg et al. (2006), CM2.0 and CM2.1 provide much-improved simulations of El Niño variability compared with previous GFDL coupled models in terms of the spatial structure and time scale of the SST variability. For example, the simulated ENSO-like SST variability in the GFDL R15 coupled model was much weaker and displaced west compared with the observed (Knutson et al. 1997) while the GFDL R30 coupled model's interannual SST variability was localized near the date line in the equatorial Pacific and had a substantially longer time scale (8–9 yr) than the observed El Niño (Knutson et al. 1999). CM2.0 and CM2.1 also have more realistic simulations of the equatorial Pacific mean subsurface thermocline structure (e.g., Wittenberg et al. 2006) than the earlier GFDL coupled models.

The enhanced variability over land regions in high latitudes does not appear to be attributable to excessive El Niño variability, as seen by comparing the maps for CM2.0 (Fig. 2b) and CM2.1 (Fig. 2c). Despite less excessive El Niño variability compared to CM2.1, CM2.0 shows clearly excessive variability over continents in high latitudes. Since CM2.0 and CM2.1 also have a number of parameter differences, which could potentially affect high-latitude continental variability, we also show Fig. 2d, which is a version of CM2.1 with

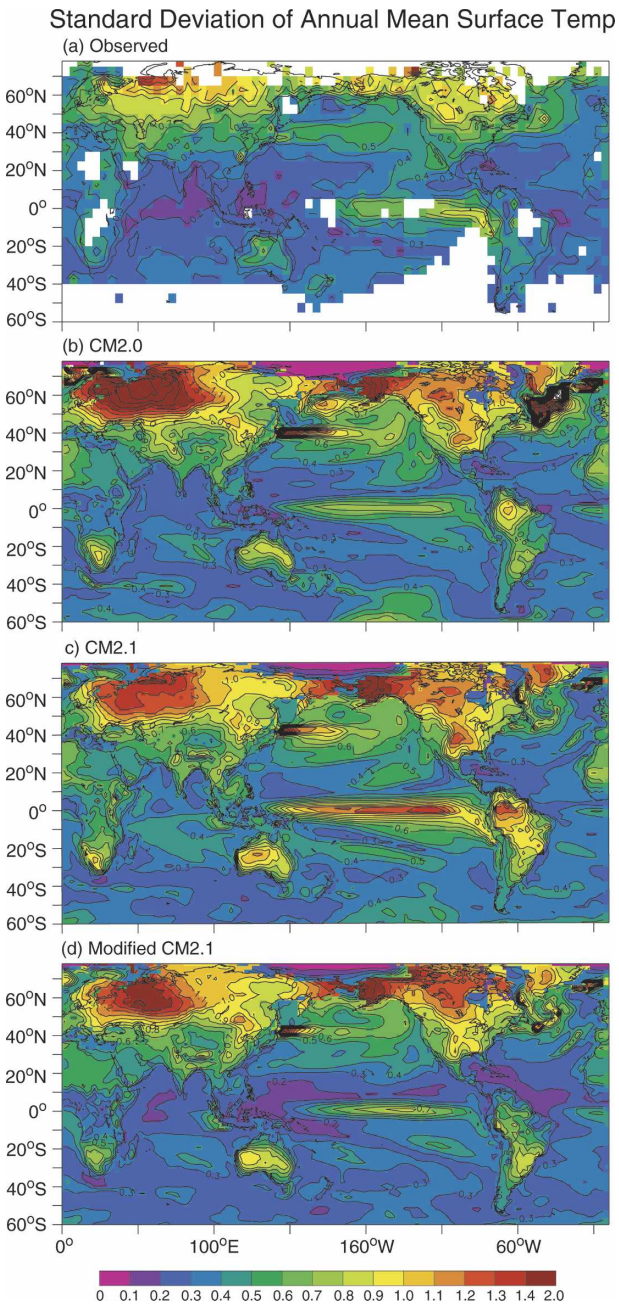


FIG. 2. Standard deviations of annual mean surface temperature (SST over ocean, surface air temperature over land) for (a) observations from HadCRUT2v (1949–2003), (b) CM2.0 control run, (c) CM2.1 control run, and (d) modified CM2.1 control run (see text). Model segments are 100 yr in length. Contour interval: 0.1°C. All time series were detrended prior to computing the standard deviations.

reduced El Niño region variability. This sensitivity experiment (Fig. 2d) was based on previous modeling experience (e.g., Wittenberg et al. 2006) that the model's El Niño variability is sensitive to the cumulus momentum transport in the atmospheric model. For the sensitivity experiment shown in Fig. 2d, cumulus momentum transport was turned off, producing a version of CM2.1 with weaker El Niño variability. The modified model shows a reduction, compared to CM2.1 (Fig. 2c), in the excessive interannual variability biases in a number of tropical and subtropical land regions, including northern South America and the southern United States. However, substantial biases remain in the northern extratropics, Australia, and southern South America, confirming that the extratropical biases are not primarily due to excessive El Niño variability in the model.

Bell et al. (2000) have noted a tendency for atmospheric models using simplified "bucket" land surface schemes to produce unrealistically high surface temperature variability over land. CM2.0 and CM2.1 use a "modified bucket" scheme, in the nomenclature of Bell et al., with spatially varying specified moisture holding capacities and stomatal conductance. In their study, such models typically exhibited greater temperature variability than observed over land, and greater variability than more physical models that included explicit representations of vegetation canopy and other features. In any case, the cause of the excessive simulated variability over land and in the tropical Pacific in the CM2 models is a topic of continuing investigation.

An important implication of the results in Fig. 2, in the context of the present study, is that there is no evidence that the CM2.0 and CM2.1 models substantially underestimate local internal climate variability, at least as aggregated over all resolvable time scales in annual mean data. Rather, the models show a tendency to overestimate this variability. This suggests that the models may provide a conservatively high estimate of internal climate variability when assessing whether observed multidecadal trends are within the range of expected internal climate variability. We will return to this issue in a later section, when we assess the variance spectra and standard deviation of observed and simulated global mean temperature in the historical simulations.

5. Historical simulations: Global mean temperature

In this section and section 6, we analyze the CM2.0 and CM2.1 historical simulations, beginning in this section with an analysis and comparison of global mean surface temperature time series and their variance spectra.

a. Simulated versus observed global mean temperature time series

Global mean surface temperature anomaly time series for various historical forcing runs are compared with observed temperature anomalies in Fig. 3. For these series, the model data have been masked out for periods and locations where data are missing according to the HadCRUT2v observed dataset. Sea surface temperature is used for model grid points with more than 50% ocean coverage, while T_ref is used for all other model grid points. Prior to computing the global means, the model time series at each grid point have been adjusted to account for long-term drifts in the control simulation as described in section 4. The data are displayed as annual mean anomalies relative to the 40-yr means for 1881–1920. The thick black curve in the panels denotes observations (HadCRUT2v) while the thick red lines are the ensemble means of various historical runs. Thin dashed green lines are the individual ensemble members.

The time series for both the CM2.0 (Fig. 3a) and CM2.1 (Fig. 3b) all-forcing experiments are in good agreement with the observed series in terms of the overall warming through the twentieth century. There is a clear tendency for the observed anomalies to fall within the range of the model ensemble members, with some notable exceptions as discussed below. In general agreement with observations, both all-forcing ensembles show a tendency for periods of fairly rapid warming early in the twentieth century, followed by a few decades of relatively little warming near the mid–twentieth century, with resumed rapid warming from about the mid-1960s (mid-1970s in the observations) through the end of the experiments in 2000.

The agreement between the CM2.1 anthropogenic-only series (Fig. 3e) and observations is also quite good, with the largest discrepancy shown being the strong temporary warming period around 1940 in the observations, which is not reproduced in the model runs. In contrast, the CM2.1 natural-only forcing experiments (Fig. 3d) clearly fail to reproduce the strong warming in the late twentieth century seen in the observations, although natural forcings appear to contribute significantly to the early twentieth-century warming in these experiments.

Other notable features of the global mean temperature curves for the all-forcings and natural-only forcing runs are the pronounced, short-lived cooling periods clearly evident in the model results in the 1880s, 1900s,

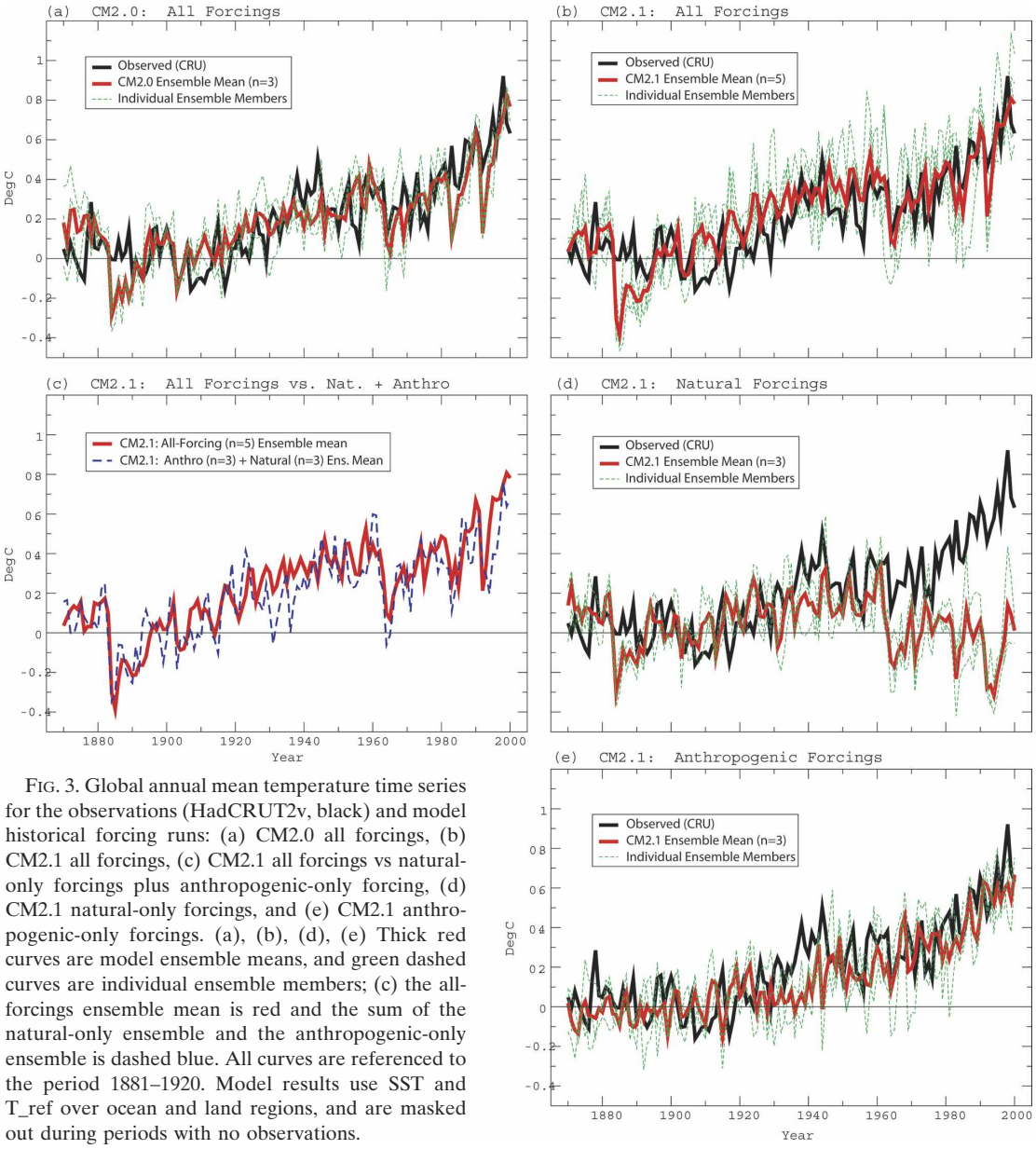


FIG. 3. Global annual mean temperature time series for the observations (HadCRUT2v, black) and model historical forcing runs: (a) CM2.0 all forcings, (b) CM2.1 all forcings, (c) CM2.1 all forcings vs natural-only forcings plus anthropogenic-only forcing, (d) CM2.1 natural-only forcings, and (e) CM2.1 anthropogenic-only forcings. (a), (b), (d), (e) Thick red curves are model ensemble means, and green dashed curves are individual ensemble members; (c) the all-forcings ensemble mean is red and the sum of the natural-only ensemble and the anthropogenic-only ensemble is dashed blue. All curves are referenced to the period 1881–1920. Model results use SST and T_ref over ocean and land regions, and are masked out during periods with no observations.

1960s, 1980s, and 1990s. These correspond to periods following strong volcanic aerosol forcing associated with the Krakatau (1883), Santa Maria (1902), Agung (1963), El Chichón (1982), and Pinatubo (1991) eruptions, respectively. In the all-forcing runs (Figs. 3a,b), the period of resumed rapid global warming in the late twentieth century begins following the cooling associated with Agung in the mid-1960s.

There is an impression from the all-forcing results (Figs. 3a,b) that the model's cooling response to several of the volcanic eruptions is larger than observed, particularly for Krakatau, El Chichón, and Pinatubo. However, this initial impression requires further investigation, since the timing of short-lived climate fluctuations associated with El Niño in the model and in the real world can complicate such interpretations (Santer et al. 2001; Soden et al. 2002). For example, with regard to Pinatubo, Fig. 4 of Soden et al. indicates that without the warming influence of El Niño, lower-tropospheric temperatures during 1992 would have been roughly 0.1°C cooler on average than what actually occurred. This compares with the model error (model versus observations) of about 0.25°C for 1992 in CM2.1 (Fig. 3b). In addition, ensemble mean tropical Pacific SSTs in our model simulations were below normal in 1992 for both CM2.0 and CM2.1 (not shown), presumably by coinci-

dence.[1] Similarly, for the El Chichón eruption, the occurrence in the same year of the strong 1982–83 El Niño event probably counteracted some of the global cooling from the eruption, while in the CM2.0 and CM2.1 ensemble means, equatorial Pacific temperatures were unusually cool, not warm, during this period. Thus, the model–observational discrepancies for Pinatubo and El Chichón appear to be partially attributable to El Niño influences, although further investigation (outside the scope of this paper) is needed.

The large cooling discrepancy in the 1880s is fairly prolonged in both models, covering the latter two-thirds of the decade (Figs. 3a,b). In contrast to the late twentieth century, these discrepancies do not appear to be exacerbated by strong El Niño influences. A likely strong contributor to the relatively prolonged cooling behavior is the occurrence of multiple significant volcanic eruptions during the period (Krakatau in 1883, Tarawera in 1886, and Bandai in 1888). The discrepancy between the models and observations during this period is unresolved, but could be due to problems in the observations (eruption or surface temperature data), radiative forcing specification, or the model's sensitivity to short-time-scale volcanic forcing. Tree-ring data, used as a proxy for Northern Hemisphere summer temperatures, provide some independent support for the observed temperature records indicating little global cooling during the period (Jones et al. 1995). Further investigation of individual forcing factors and internal variability will be needed for more definitive conclusions.

The low simulated temperatures during the 1880s in CM2.0 and CM2.1 natural-only and all-forcing runs could significantly affect trends that begin in the late 1800s. Later in this report, we present some long-term trend analyses using 1901 as the starting year, which reduces the influence of these Krakatau-era discrepancies on the trend statistics.

In a study with the GFDL R30 coupled model, Broccoli et al. (2003) concluded that the simulated response of that model to volcanic eruptions since the late 1800s was larger than observed, based on a superposed epoch analysis. We note that their simulations used a different prescribed volcanic forcing derived from another global model by Andronova et al. (1999) and was a more idealized implementation of volcanic forcing than used here.

[1] Adams et al. (2003) present empirical evidence, based on volcanic and paleoclimate proxies, for a link between explosive volcanic eruptions and the probability of El Niño occurrence the following winter. The limited ensembles in our experiments do not provide model-based evidence for such behavior.

The rapid warming in the early twentieth century appears likely due to a combination of anthropogenic and natural forcings (Fig. 3b) including increased greenhouse gases, reduced negative volcanic forcing following Krakatau and other eruptions just discussed, and a positive contribution from solar variations. This finding is in agreement with earlier studies (e.g., Stott et al. 2000). The results in Figs. 3d,e also suggest that either natural-only forcing or anthropogenic-only forcing, in combination with unusually strong temporary warming from internal climate variability, can reproduce an early twentieth-century warming in the model that is fairly close to the observations. For example, one anthropogenic-only ensemble member shows a temporary warming peak in the mid-1940s, similar to the observed peak a few years earlier. An even closer "anthropogenic forcing plus internal variability" analog for the early-twentieth-century global warming was shown for an earlier model by Delworth and Knutson (2000). However, regarding the late-twentieth-century warming, the results in Fig. 3 have little ambiguity: the rapid warming during the late twentieth century in CM2.1 is a consequence of anthropogenic forcing, since the runs with natural-only forcing fail to reproduce this strong observed late-twentieth-century warming. Solar variability contributes little positive radiative forcing during the late twentieth century while volcanic eruptions produce strong temporary negative forcings late in the century—particularly from Pinatubo (V. Ramaswamy et al., unpublished manuscript).

Figure 3c examines the linearity of the model's global response to the forcings by comparing the anomaly ensemble mean series from the CM2.1 all-forcings ensemble to the sum of the ensemble mean anomalies from the CM2.1 natural-only and anthropogenic-only forcing runs. The curves are very similar, indicating that the model's global response to these forcings is approximately linear. A similar result has been found in previous studies (e.g., Ramaswamy and Chen 1997; Haywood et al. 1997; Gillett et al. 2004; Meehl et al. 2004; see also Hansen et al. 1997 and Ramaswamy et al. 2001).

b. Variance spectra of global mean temperature

Figure 4 shows variance spectra of the observed global mean temperature series (1871–2000, dark red curve in each panel) in comparison to variance spectra of 130-yr segments (1871–2000) from the CM2.0 and CM2.1 (Figs. 4a,b, respectively) all-forcing historical runs. The ensemble spectra from the all-forcing runs (medium black lines) indicate that the model's variance spectrum is fairly similar to observations in terms of its general shape. However, there is a clear tendency to-

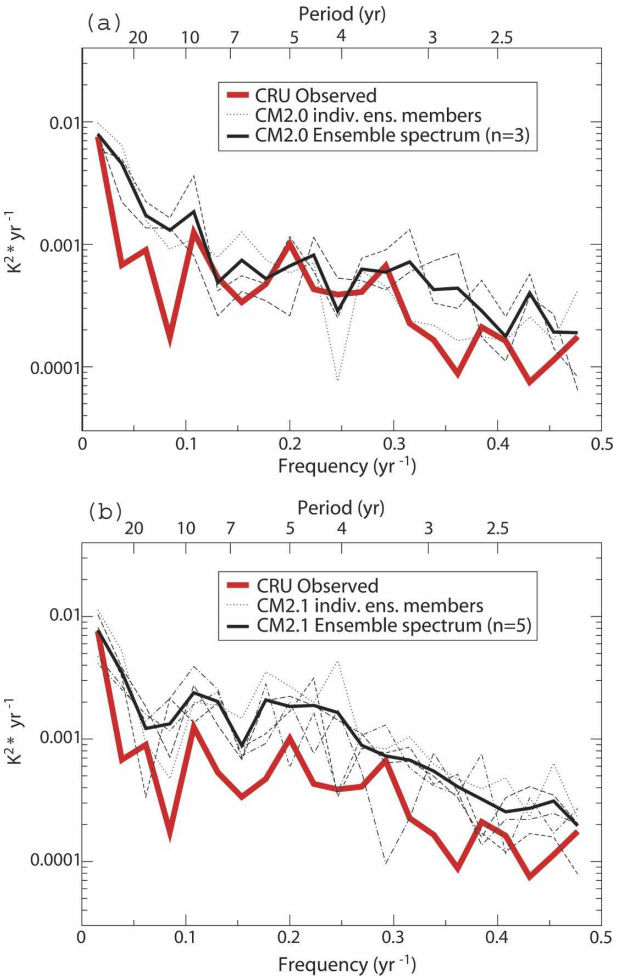


FIG. 4. Variance spectra of global mean surface temperature series for observations (1871–2000; red curves) and for 130-yr segments from various all-forcing historical experiments for (a) CM2.0 and (b) CM2.1. The models' ensemble mean spectra are depicted by medium solid black lines, and are computed as the mean of the spectra of the individual ensemble members (thin lines). The raw spectra were smoothed using a nonoverlapping boxcar window of width three calculable frequencies. All series were detrended prior to spectral calculations. Model time series have been constructed using the observed data mask.

ward excessive magnitude of variability, compared to observations, on almost all time scales, especially in CM2.1. In fact, the CM2.1 ensemble spectrum lies above the observed spectrum at all frequencies shown except at the lowest frequency (~65 yr). The CM2.0 ensemble spectrum also tends to lie above the observed spectrum, although not as consistently as for CM2.1.

The standard deviations of the raw and detrended annual mean and the detrended 10-yr running mean global mean time series for observations and the all-forcing historical runs are presented in Table 1. The annual statistics aggregate variability across all time

TABLE 1. Standard deviations of annual mean or 10-yr running mean global mean surface temperature time series (1871–2000) for observations and all-forcing model historical runs (unit: °C). For the detrended columns, a least squares linear trend has been removed from the data prior to computing the standard deviations. The ratio is obtained by dividing the average model standard deviation by the observed standard deviation.

	Unfiltered annual means	Detrended annual means	Detrended 10-yr running means
Observed	0.221	0.128	0.082
CM2.0 h1	0.218	0.166	0.095
CM2.0 h2	0.214	0.164	0.090
CM2.0 h3	0.211	0.147	0.078
CM2.0 average	0.214	0.159	0.088
Ratio: CM2.0 to Obs	0.970	1.242	1.069
CM2.1 H1	0.284	0.204	0.105
CM2.1 H2	0.247	0.169	0.087
CM2.1 H3	0.283	0.183	0.111
CM2.1 H4	0.203	0.159	0.071
CM2.1 H5	0.255	0.163	0.075
CM2.1 average	0.254	0.176	0.090
Ratio: CM2.1 to Obs	1.151	1.372	1.095

scales resolvable by annual means in the 130-yr records, as opposed to a spectral decomposition, which aggregates variance into different frequency bands. For detrended annual data, the average standard deviation of the CM2.1 all-forcing runs exceeds the observed standard deviation by 37%. The CM2.0 all-forcing scenario standard deviations also substantially exceed the observed value (by 24%). The fact that all eight independent CM2.0 or CM2.1 all-forcing runs have larger detrended standard deviations than the detrended observations indicates that model's global annual mean variability is very likely larger than observed variability. For example, if the modeled and observed standard deviation samples came from the same population, the probability that the observed standard deviation is lower than any of the eight model standard deviations, as in Table 1, would only be about 11%. Standard deviations of 10-yr running mean (detrended) data, which focus on low-frequency variations, show a smaller positive bias of variability in the models compared with observations (+7% and +10% for CM2.0 and CM2.1, respectively).

As previously discussed in the context of the control simulations, the excessive global mean surface temperature variability in the models, likely related to the excessive El Niño variability in CM2.1 as well as the excessive interannual variability over continental land surfaces, has the consequence that the models are likely to be providing a high (conservative) estimate of the

potential contribution of internal climate variability to observed long-term trends, as assessed in the following section. On the other hand, the excessive variability in the models will increase the spread among ensemble members, making it more likely that observations will fall within the spread of the individual ensemble members.

6. Assessment of regional surface temperature trends

In this section, the surface temperature trends in the models and observations are compared for different regions. We focus first on the latter half of the twentieth century (1949–2000), a period during which global data coverage was substantially more complete than during the late nineteenth and early twentieth centuries. Later in the section we present a trend analyses on the longer (century) scale for regions with at least moderate data coverage. Finally, the temporal character of low-frequency temperature variations in various case study regions are examined using area-averaged time series.

a. 1949–2000 trends

Regional surface temperature trends in the observations and models are assessed for the period 1949–2000 by comparing the observed pattern of trends with the internally generated 52-yr trends in the control runs and with the trend patterns from the 1861–2000 historical simulations over the same 52-yr period. The control run sample segments and model segments from the historical runs are masked with the time-varying observed missing data mask prior to computing the trend samples. For these analyses, simulation results for CM2.0 and CM2.1 have been combined into a single superensemble. Although not presented here, we have performed assessments for the CM2.0 and CM2.1 ensembles separately and obtained broadly similar results for the two sets of ensembles. This provides some justification for our combining the results from the two models into a single combined assessment.

The assessment of the annual mean trend maps is shown in Fig. 5. The observed trend map (Fig. 4a) shows pronounced warming [2°–4°C (100 yr)$^{-1}$] over many northern middle- and high-latitude continental regions with a broad region of more gradual warming [1°–2°C (100 yr)$^{-1}$] over much of the Indian Ocean/western Pacific warm pool region, Australia, the eastern tropical Pacific, and South Atlantic (north of 40°S). Regions of cooling, smaller in overall extent than the warming regions, are seen in the North Pacific, much of the North Atlantic basin, the southeastern United States, New Zealand and vicinity, and a few other small regions.

The observed trend assessment, based on the model control runs (Fig. 5b, color shaded regions), indicates that the observed warming trends over much of the globe are unusual in comparison to the trends found in the control run. About 68% of the global area examined has observed trends that are outside the 5th–95th percentile range of the trends for that region from the combined control run samples (37 nonoverlapping 52-yr segments). The areas of significant trends are dominated by warming trends, with only small regions of the globe identified as having significant cooling trends.

The ensemble mean trend map for the eight CM2.0 ($n = 3$) and CM2.1 ($n = 5$) all-forcing runs in Fig. 5c shows a broad-scale warming pattern over much of the globe with enhanced warming over many continental regions. A large area of cooling is simulated in the North Pacific similar to the observations although extending further north and east in the basin and not as far equatorward as the observed cooling trend region. Smaller secondary regions of cooling appear in extreme eastern Asia, central Asia, and the North Atlantic. The all-forcing ensemble mean does not show the particularly strong warming trends in middle- and high-latitude land regions of the Northern Hemisphere that are evident in the observations.

Figure 5d shows an assessment of where the ensemble mean simulated trends and observed trends are significantly different from one another. The field shown is a map of model-minus-observed trend differences (i.e., Fig. 5c minus Fig. 5a), but with a modified local Student's t test applied such that areas are blacked out where the null hypothesis that the simulated and observed trends are equal cannot be rejected at the 0.10 level using a two-sided test. By this definition, the remaining color-shaded regions are areas where the simulations are significantly different from the observations. The modified Student's t test used (following Knutson et al. 1999) is analogous to a standard two-sided two-sample Student's t test except that the long control runs are used to estimate the variance of the trends (assumed to be the same for both the model and observations). The n1 and n2 parameters for the Student's t test are n1 = 1 for the observation and n2 = 8 for the eight-member CM2.0/CM2.1 ensemble. The modified Student's t test statistic assumes 37 degrees of freedom based on the total number of nonoverlapping 52-yr segments ($n = 37$) from the control runs (assumed independent) that are used to estimate the variance of 52-yr trends.

The Student's t test results for the all-forcing runs in Fig. 5d indicate that these runs have significantly less warming than observed during 1949–2000 in a number

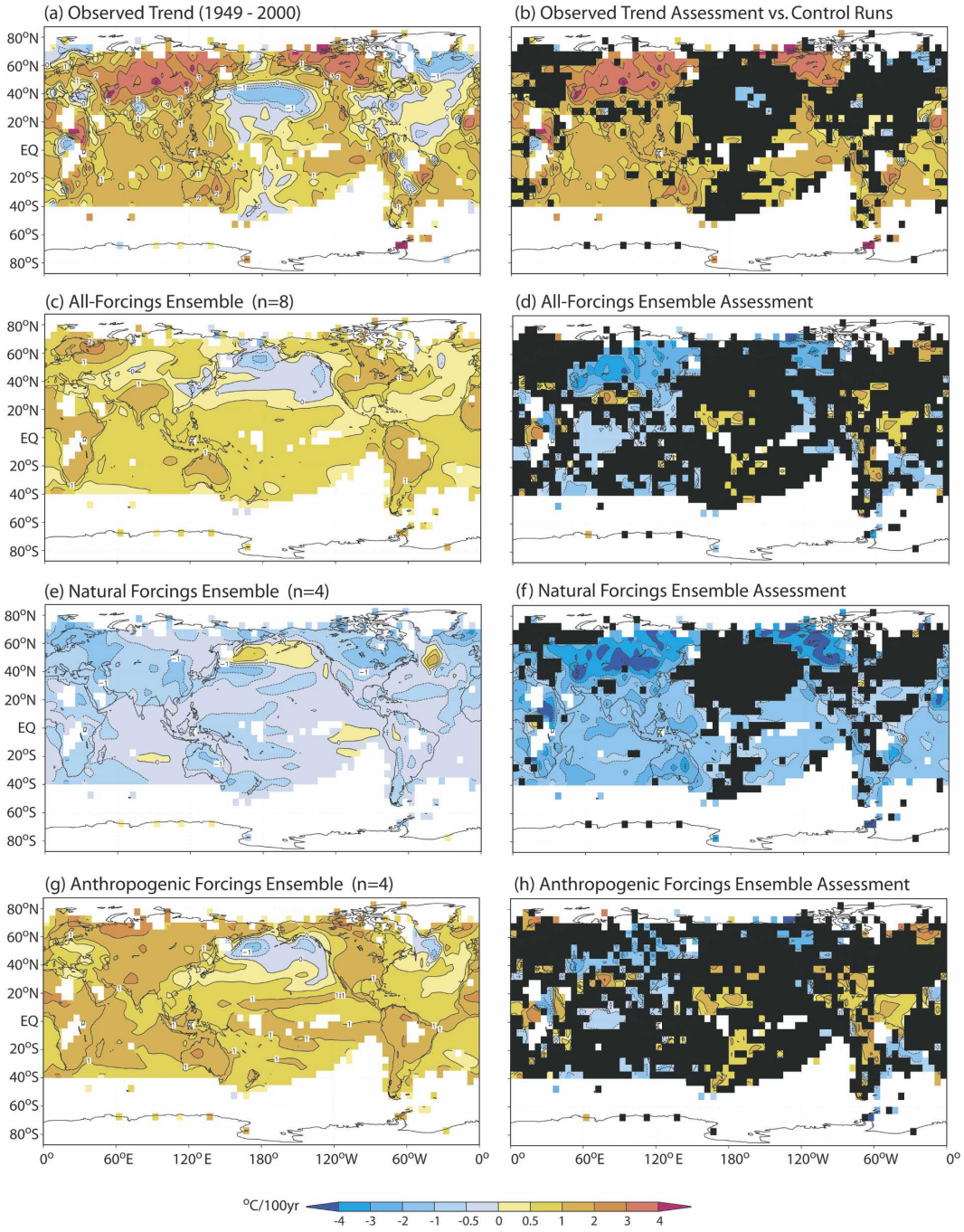


FIG. 5. (a) Observed surface temperature trends (1949–2000) in °C per 100 yr. (b) Black shading indicates regions where the observed trend is within the 5th–95th percentile range of the simulated trends from 37 nonoverlapping 52-yr segments of the control run. Color-shaded regions thus show where observed trends are unusual in comparison to model internal variability according to this measure. (c), (e), (g) Simulated surface temperature trends for 1949–2000 based on the ensemble mean ($n = 8$) of the CM2.0 and CM2.1 all-forcing (c, $n = 8$), natural-only forcing (e, $n = 4$), or anthropogenic-only forcing (g, $n = 4$) historical runs. (d), (f), (h) Model minus observed trend differences with black regions depicting areas of agreement between model and observations for the all-forcing (d), natural-only forcing (f), or anthropogenic-only forcing (h) runs. See text for details.

of regions, including much of northern Asia, and parts of Canada, the Indian Ocean/warm pool region, the southern Indian Ocean, and the South Atlantic. Most of the regions of inconsistency are areas where the all-forcing runs do not warm rapidly enough. Some smaller regions where the model warms too rapidly include parts of the central tropical Pacific, northern tropical Atlantic, South America, and equatorial Africa. The tests indicate that the all-forcing ensemble is not significantly different from the observations (black regions) in the North Pacific cooling region and most of the North Atlantic cooling regions in Fig. 5a. The percent of global area tested where the ensemble simulation has significantly different trends from observations is 31%, as compared with 68% for the control run comparison in Fig. 5b.

The ensemble mean simulated trend for the ($n = 4$) natural-only forcing runs (Fig. 5e) shows a cooling in most regions during 1949–2000, and thus the disagreement with observed trends is quite large. About 70% of the global area examined has significantly different simulated trends than the observed for this period according to the modified Student's *t* test (Fig. 5f). This is similar to the degree of inconsistency in the control run (no external forcing) comparison (Fig. 5b).

The ensemble mean results for the anthropogenic-only runs (Fig. 5g) show greater warming rates in many regions than the all-forcing runs (Fig. 5c). In terms of cooling regions, the ensemble mean map shows similar regional features to the all-forcing runs (Fig. 5c) in the extratropical North Pacific and somewhat more cooling in the high-latitude North Atlantic. The Student's *t* test assessment (Fig. 5h) shows a slightly smaller percent area with significant differences (27%) than the all-forcing ensemble (31%), with more of a balance between areas with negative and positive trend differences.

We have not performed Monte Carlo simulations to assess whether the locally significant differences in Fig. 5 represent significant changes from a field significance or global significance perspective (e.g., Livezey and Chen 1983). However, we found in a similar earlier study with the GFDL R30 coupled model, that a percent area exceeding about 23% represented a globally significant (95%) result (Knutson et al. 1999) for the above type of test, indicating that even the cases mentioned above with 27% and 31% of areas having locally significant differences probably represent globally significant findings.

Seasonal versions of the combined CM2.0/CM2.1 assessments for the observed trend (1949–2000) and the all-forcing ensemble are shown in Fig. 6. Broadly similar results were obtained for the CM2.0 and CM2.1 ensembles separately (not shown). Many of the characteristic features of the annual mean results just discussed also appear in the seasonal mean assessment results. A distinctive feature of the observed seasonal trend maps is the very pronounced seasonality of the warming in extratropical Northern Hemisphere land regions, with maximum warming in December–February (DJF) and March–May (MAM; e.g., upper left panels in Figs. 6a,b). This seasonally modulated character of the warming is largely absent in the simulated ensemble mean trend maps. In addition, the area with significant differences between model and observations is much larger in the winter and spring than in summer and fall in those regions. The northern Pacific cooling feature appears in all seasons in the observations and the model ensembles (both in CM2.1 and CM2.0 individually as well, though not shown). The high-latitude North Atlantic cooling feature appears in all seasons in the observations, but much more weakly in the model ensemble. However, the significance test results indicate that the observed cooling trends over the North Pacific and North Atlantic are mostly not significantly different from the all-forcing model ensemble trends.

b. 1901–2000 trends

Trend maps and their model-based assessment for the entire twentieth century (1901–2000) are shown for the observations and model historical runs in Fig. 7. The figure is constructed similarly to that for 1949–2000 (Fig. 5), except that a less stringent data availability requirement has been used for the 100-yr trends, owing to the more sparse data coverage in the first half of the twentieth century. Specifically, we have required 20% coverage for each of the five 20-yr periods (1901–20, 1921–40, etc.) as a screening criterion to determine whether to plot results for a grid point. Effectively, this screen applies to the first half of the twentieth century, since data coverage generally has been much greater since around 1950. The control run statistics are based on 20 nonoverlapping 100-yr segments from CM2.0 and CM2.1 combined.

A notable feature of the observed trend map for the past century (Fig. 7a) is the pervasiveness of the warming on the regional scale: almost all areas of the globe analyzed appeared to have warmed over the twentieth century. A few relatively small areas of cooling are seen, including a region south of Greenland and another covering the southeastern United States. The cooling trends in these regions generally do not appear to be statistically significant according to comparison with the control run 100-yr trends (Fig. 7b). On the other hand, the warming trends over much of the globe are statistically significant (compared to internal climate variability) according to these tests. The tropical

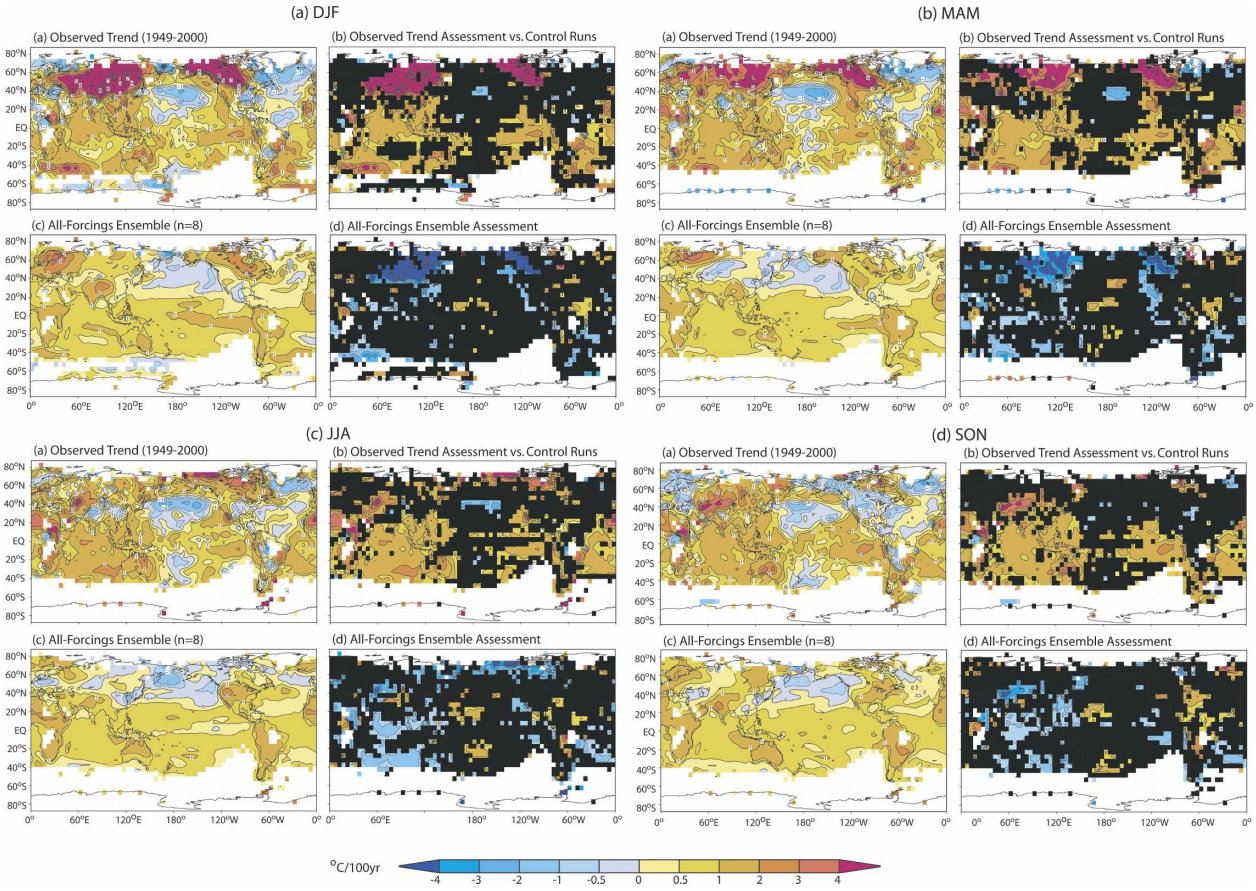


FIG. 6. As in Fig. 5 but for separate 3-month seasons: (a) DJF, (b) MAM, (c) June–August (JJA), and (d) September–November (SON). For each season, the same four-panel display format as for Figs. 5a–d is repeated.

North Atlantic is a notable region of significant warming that was not identified as having significant warming in the 52-yr trend analysis (Fig. 5). Overall, the results in Fig. 7b indicate that 89% of the areas analyzed have trends that lie outside of the 5th–95th percentile range of the model-simulated trends from the control run, with almost all of these being positive trends. This finding strongly suggests that over the vast majority of the global area analyzed, the observed twentieth-century warming trends on a regional scale are unlikely to be due to internal climate variability alone.

Comparing the observed trends with the all-forcing simulations (Fig. 7a versus Figs. 7c,d) there is some broad-scale similarity, with the model ensemble trend map also showing a preponderance of warming trends. Some broad-scale differences are also apparent. For example, the difference field (Fig. 7d) shows a tendency for too little warming (blue colors) in the extratropics and too much warming (yellow colors) in the Tropics and subtropics in the all-forcing runs. Another discrepancy between the model and observations is the presence of a large cooling region in the extratropical North Pacific in the all-forcing simulations—a feature that is absent in the observations. This discrepancy appears to be a statistically significant difference (Fig. 7d). There are substantial areas of consistency in 100-yr trends between observations and the all-forcing runs (i.e., black shading in Fig. 7d) include much of the North Atlantic Ocean, Indian Ocean, western Pacific, Europe, and North America. According to the Student's t tests, about 42% of the areas analyzed have significant differences between all-forcing run trends and the observations, indicating that trends over about 58% of the regions analyzed are consistent between the all forcing runs and observations.

For the anthropogenic-only forcing runs (Figs. 7g,h), significant differences with observations are found over a somewhat smaller fraction of the globe (33%) than for the all-forcing runs, indicating a somewhat greater degree of consistency in the anthropogenic-only forcing runs. Various regional features are broadly similar to those found for the all-forcing runs, and are not discussed in detail here.

The natural-only forcing runs 100-yr trend maps

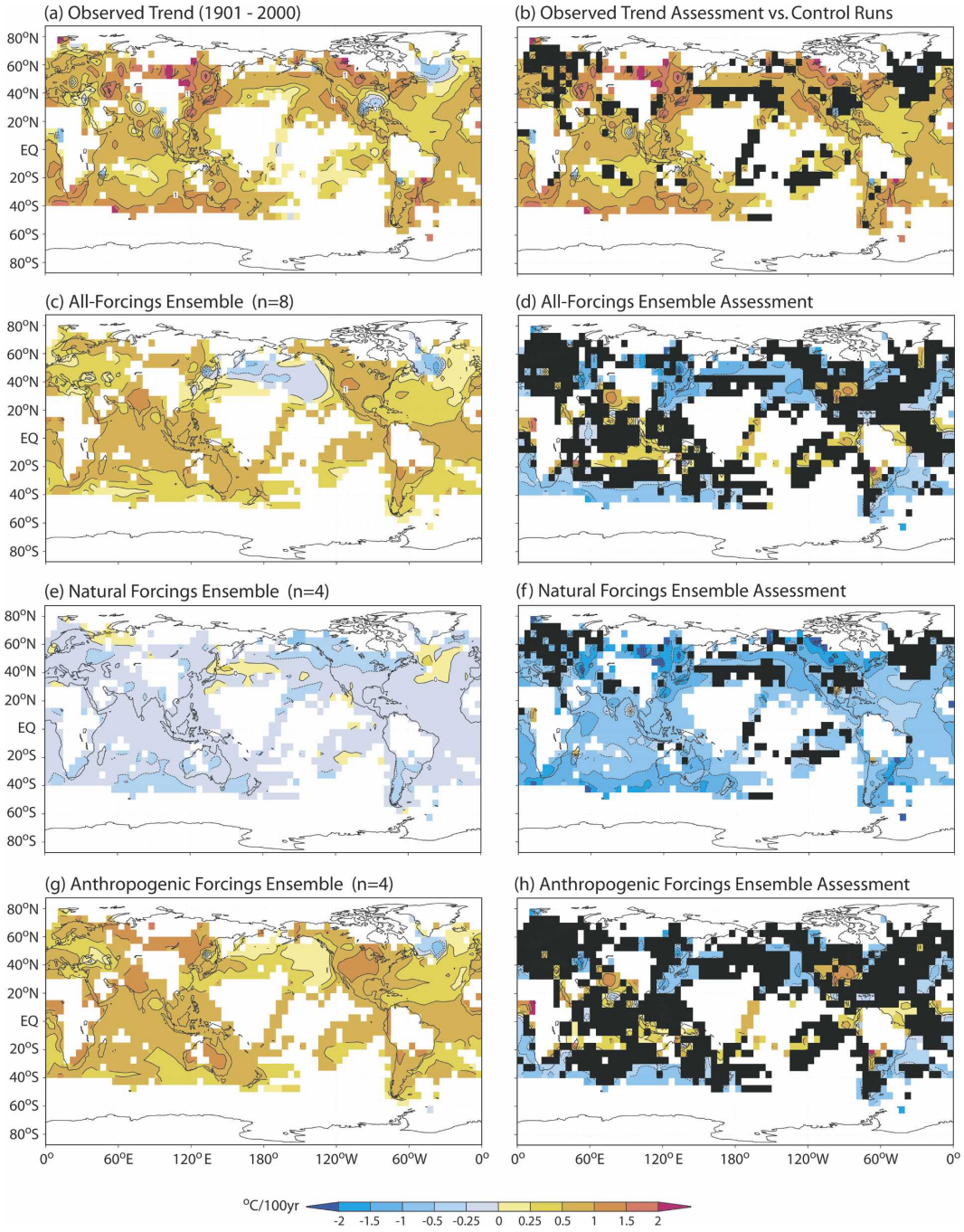


FIG. 7. As in Fig. 5, but for 100-yr trends over the period 1901–2000. Results are plotted only for grid points where the observations have at least 20% data availability of monthly mean anomalies for each of five 20-yr periods (1901–20, 1921–40, . . . , 1981–2000). The model fields are masked out during periods when observed data are missing. White regions denote grid points with insufficient observed data coverage.

(Figs. 7e,f) show mostly cooling trends, in poor agreement with observations. The differences between the simulated and observed trends are statistically significant (Fig. 7f) in most regions. About 84% of the analyzed areas have significant differences between model and observed trends. This result, compared to those for the all-forcing and anthropogenic forcing runs, constitutes substantial model-based evidence that the pervasive twentieth-century warming trends in the observations are not solely due to natural variability, but that

anthropogenic forcing has played an important role in producing these warming trends.

A caveat to the results in Fig. 7 is the relatively sparse data coverage and other data issues as one extends such analyses further back in time. The data issues include uncertainties due to adjustments to SST observations prior to the early 1940s (e.g., Folland et al. 2001). These imply a greater uncertainty in the observed trend details for the 100-yr trends (Fig. 7) than for the most recent half-century (Fig. 5). Although not used in the present study, an alternative approach would be to use currently available SST historical reconstructions (e.g., Kaplan et al. 1998; Rayner et al. 2003; Smith and Reynolds 2003), which attempt to address at least the missing data shortcoming by filling in data gaps in the SST records based on space–time statistical methods.

c. *Regional case studies*

The trend maps in the previous discussions are useful for revealing coherent spatial patterns and other features in the trend fields, but they require the rather arbitrary selection of starting and end dates, to which linear trend results can be sensitive. An alternative to trend maps are direct comparisons of observed and modeled time series. Because of space considerations, only a limited number of case study regions are chosen for focused examination in this study. Each of the time series has been adjusted for control run drift, screened with the observed data coverage for the region, and low-pass filtered (10-yr running mean) to emphasize low-frequency variations, which are most relevant to the present study. The time series are anomalies with respect to years 1881–1920 for these figures. The areas used to construct the regional time series are identified in Fig. 8. We also attempt to assess the possible effect of missing observations on the observed regional time series by constructing error ranges based on masked versus unmasked versions of the model time series.

1) GLOBAL, EXTRATROPICAL, AND TROPICAL MEANS

Figures 9a,b show the observed global mean temperature versus the all-forcing historical run ensemble members for CM2.0 and CM2.1, respectively. Figure 9c shows the ensemble means from the all-forcing ($n = 5$), natural-only forcing ($n = 3$), and anthropogenic-only forcing ($n = 3$) CM2.1 runs. The simulated and observed global mean curves (Figs. 9a–c) are in fairly good agreement for the all-forcings and anthropogenic-only forcing runs, but not for the natural-only forcing ensemble (Fig. 9c, blue dashed), as discussed previously for Fig. 3. The global mean results are provided here mainly as a reference comparison for the other regional case study time series.

Figure 9c also shows the estimated uncertainty due to missing observations for the observed global mean temperature series (10-yr means). The shaded region in (Fig. 9c) is the ± 2 standard error range about the observations, which was estimated by differencing model 10-yr running mean series obtained with and without use of the observational mask. The error estimate is based on the sample of eight available all-forcing scenario runs (three for CM2.0 and five for CM2.1) and eight natural-only (four) or anthropogenic-only (four) forcing runs. The error range decreases over time from about 0.12°C in the late 1800s to about 0.06°C in the late twentieth century. Note that this is only a partial error estimate, as it does not include the uncertainty due to errors in the available data or the uncertainty range due to internal climate variability (i.e., the range of results in a hypothetical world with an ensemble of observations available). For example, Fig. 2.8 of Folland et al. (2001) includes uncertainties due to bias corrections in SSTs prior to the early 1940s and due to urbanization, both of which appear to have a substantial impact on the uncertainty estimates. However, since it is straightforward for us to estimate at least the missing data component of the uncertainty for the global mean series as well as the regional series, these partial error estimates are included in our case study figures.

For the northern extratropics (Figs. 9d–f) the all-forcing simulations show a similar time evolution to the observed in some respects, including relatively rapid warming early in the twentieth century, relatively little warming or even some cooling from the 1930s (1940s in the observations) to the 1970s, and resumed strong warming from the 1970s onward. The southern extratropics (Figs. 9g–i), shows a more monotonic warming behavior in the all-forcing simulations, as do the observations. There is some tendency for the all-forcing simulations to warm less than the observations in the extratropics, particularly for the southern extratropics in CM2.0 (Fig. 9g). In the Tropics (Figs. 9j–l), the all-forcing runs tend to warm slightly more than the observations, with the observed curve typically falling on the lower edge or below the range of the model ensemble members. The natural-only ensemble again shows a clear deficiency in not simulating the strong late-twentieth-century warming in any of the regions shown in Fig. 9. The strong cooling in the 1880s discussed earlier appears most pronounced in the model in the northern extratropics (Figs. 9d–f) with a lesser expression in the Tropics (Figs. 9j–l) and only a minor

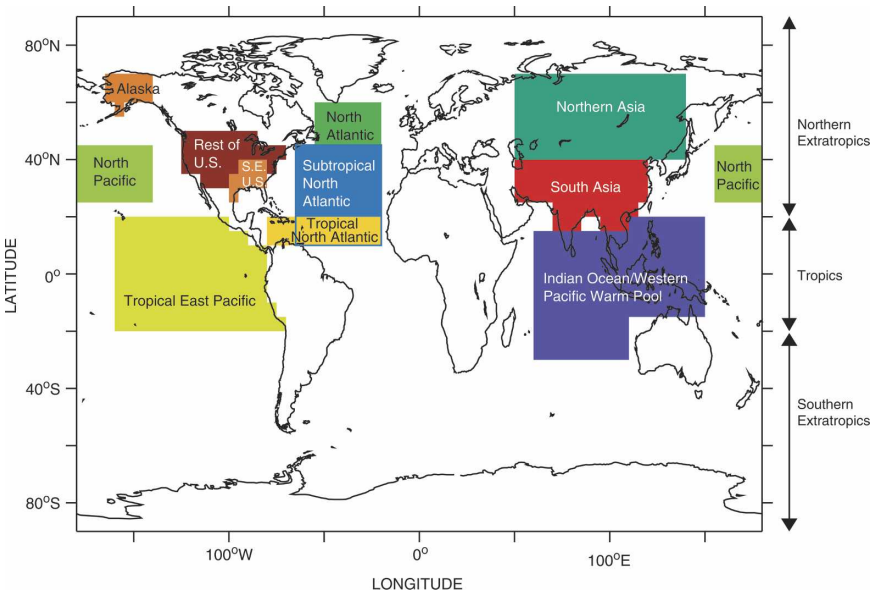


FIG. 8. Map showing the areas used to create various area-averaged time series for the case studies in Figs. 9–11. The percent of global area covered for each regions: Alaska: 0.3%, southeast United States: 0.4%, rest of continental United States (excluding southeast United States): 1.3%, Indian Ocean/western Pacific warm pool: 8.6%, subtropical North Atlantic: 3.3%, northern Asia: 3.7%, South Asia: 3.0%, North Atlantic: 0.8%, North Pacific: 2.6%, tropical east Pacific: 7.4%, tropical North Atlantic: 1.4%, Tropics: 34.2%, northern extratropics: 32.9%, southern extratropics: 32.9%.

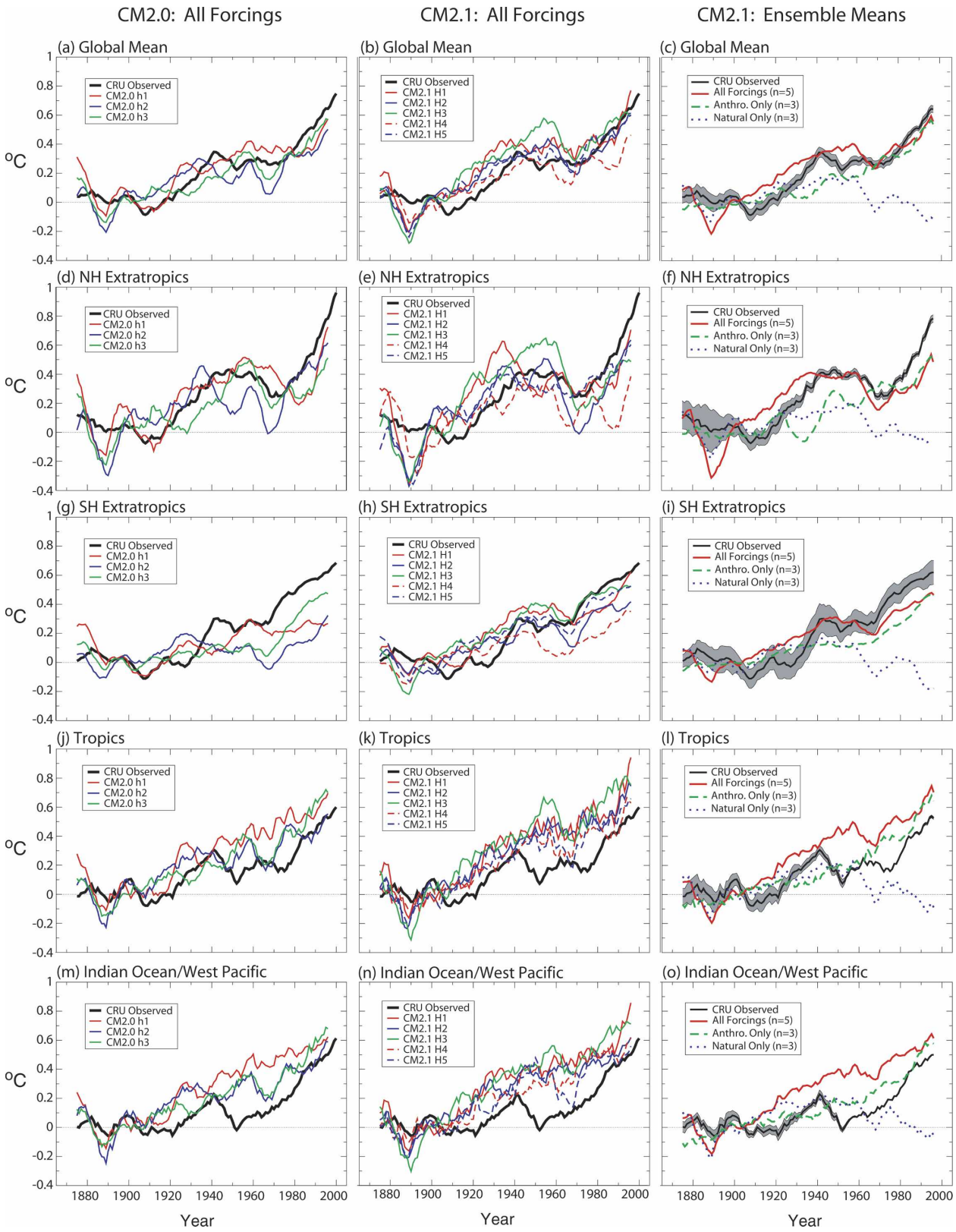
expression of the cooling in the southern extratropics (Figs. 9g–i).

The finding that the all-forcing model tends to warm too rapidly in the Tropics and not rapidly enough in the extratropics implies that the model simulates less poleward amplification of warming than observed during the twentieth century, as was shown in Fig. 7. To investigate the poleward amplification issue in more detail, we have examined time versus latitude plots (not shown) of zonal mean temperature anomalies for the observations, historical runs, and a future climate change scenario [CM2.1 IPCC Special Report on Emission Scenarios (SRES) Scenario A1B to 2100]. While the historical run ensemble mean shows less poleward amplification of warming than the observations, the A1B future scenario shows the Northern Hemisphere extratropics warming more than the Tropics in the model, while the deep Southern Hemisphere warms very little. A similar pattern of response to the A1B scenario is seen in +1% yr^{-1} CO2 transient experiments with CM2.0 and CM2.1 (Stouffer et al. 2006). Doubled CO_2 equilibrium experiments, with the atmospheric components of CM2.0 and CM2.1 coupled to a slab ocean model, show significant poleward amplification of warming in both hemispheres (Stouffer et al. 2006). These climate change experiments, showing marked polar amplification of warming in response to a large positive radiative forcing, suggest that internal climate variability could be playing a role in the apparent discrepancies in the Northern Hemisphere in the historical runs. Further investigation, outside the scope of this paper, will be needed to clarify this issue.

The impact of missing data on the observed curves in Fig. 9 is estimated to be greatest for the Southern

→

FIG. 9. Ten-yr running-mean area-averaged time series of surface temperature anomalies (°C) relative to 1881–1920 for observations and models for various regions: (a)–(c) global mean, (d)–(f) Northern Hemisphere extratropics (20°–90°N), (g)–(i) Southern Hemisphere extratropics (20°–90°S), (j)–(l) Tropics (20°N–20°S), and (m)–(o) Indian Ocean/western Pacific warm pool. The left column and middle columns are based on all-forcing historical runs 1871–2000 and observations 1871–2004 for CM2.0 (n = 3) and CM2.1 (n = 5), respectively. The right column is based on observed and model data through 2000, with ±2 standard error ranges (shading) obtained by sampling several model runs according to observed missing data. The red, blue, and green curves in the right-hand-column diagrams are ensemble mean results for the CM2.1 all-forcing (n = 5), natural-only (n = 3), and anthropogenic-only (n = 3) forcing historical runs. Model data were masked according to observed data coverage.


CM2.0: All Forcings
CM2.1: All Forcings
CM2.1: Ensemble Means
(a) Global Mean
(b) Global Mean
(c) Global Mean
(d) NH Extratropics
(e) NH Extratropics
(f) NH Extratropics
(g) SH Extratropics
(h) SH Extratropics
(i) SH Extratropics
(j) Tropics
(k) Tropics
(l) Tropics
(m) Indian Ocean/West Pacific
(n) Indian Ocean/West Pacific
(o) Indian Ocean/West Pacific
CRU Observed
CM2.0 h1
CM2.0 h2
CM2.0 h3
CM2.1 H1
CM2.1 H2
CM2.1 H3
CM2.1 H4
CM2.1 H5
All Forcings (n=5)
Anthro. Only (n=3)
Natural Only (n=3)
°C
Year
1880
1900
1920
1940
1960
1980
2000

Hemisphere extratropics (Fig. 9i). The error range is also relatively large (about ±0.15°C) in the 1880s in the northern extratropics.

Finally, we note that much of the low-frequency (multidecadal) fluctuations in the all-forcing runs appear to have a broadly similar timing or phasing across the different ensemble members, as well as for the observations. For example, in the northern extratropics (Figs. 9d,e), the temporary period of slight cooling from around the 1950s to around 1970, followed by strong warming, is seen in the envelope of the ensemble members. However, the ensemble mean for CM2.1 all forcing (Fig. 9f) also suggests a longer break between major warming periods than seen in the observations. A second example is the pronounced cooling episode seen in the 1880s in all panels. This common behavior among ensemble members indicate that these features are being forced in the model, and perhaps in observations, by volcanic activity and other specified climate forcing agents, as opposed to being randomly timed internal climate variations.

2) TROPICAL REGIONAL INDICES

Various tropical and subtropical regional indices are shown at the bottom of Figs. 9m–o and in Fig. 10. The Indian Ocean/western Pacific warm pool index (Figs. 9m–o) shows a relatively pronounced warming trend in the observations. The observed curve shows fairly rapid warming during the 1920s and 1930s, followed by temporary cooling during the 1940s, and then a relatively monotonic warming trend from around 1950 through the end of the twentieth century. The strong monotonic warming since 1950 in this region produced a broad-scale statistically significant warming signal in an earlier model-based assessment (Knutson et al. 1999; Hoerling et al. 2004)—a finding that was reconfirmed in the present study as well (Fig. 5b). A similar, though slightly noisier, behavior is evident for the Tropics as a whole (Figs. 9j–l). The simulated warming trend over the twentieth century is more monotonic for the ensemble mean of the CM2.1 all-forcing runs (Fig. 9o) than the observations, although individual ensemble members (Figs. 9m,n) show a few multidecadal variability episodes similar in amplitude to the observed fluctuation from 1920–50. The timing of the model-generated events does not coincide with the observed one. The models tend to warm more rapidly than observed prior to 1950, and less rapidly than observed from 1950 to 2000. Thus the observed warming nearly catches up with the simulated warming by the end of the period, as the observations reach the lower edge of the ensembles of the all-forcing runs for both CM2.0 and CM2.1.

In the tropical east Pacific (Figs. 10a–c), the simulated and observed time series are characterized by a much greater degree of noise relative to the warming signal, due to the influence of El Niño and perhaps Pacific decadal variability. Despite the high noise level in this region, the observed 10-yr running mean temperatures since about 1980 are the warmest in the record. During a temporary warming around 1900, temperatures reached levels within a few tenths of a degree of those of the most recent decades. The all-forcing ensemble members, particularly for CM2.1, tend to show a more monotonic warming evolution in this region, which results in a period of several decades, from about 1940 to 1980, when the model appears substantially warmer than the observations. However, from about 1980 on, the strong late-century warming in the observations has considerably narrowed the discrepancy with the simulations such that the recent observations are again within the envelope of the ensembles. This is similar to the behavior for the Indian Ocean/western Pacific warm pool (Figs. 9m–o) discussed above.

Although the estimated errors in the observations due to missing data are larger as one goes further back in the records, the error for the tropical east Pacific is estimated to be smaller than for the Southern Hemisphere extratropics, or the global mean, as seen by comparing the shaded regions in Fig. 10c with those in Fig. 9.

The tropical North Atlantic region from 10° to 20°N was selected for particular focus, as this region is known as the main development region for Atlantic hurricanes that evolve from tropical easterly waves emanating from Africa (e.g., Goldenberg et al. 2001). The discrepancies noted above for the Indian Ocean and eastern tropical Pacific simulations for the 1940s through 1980 are not evident in the tropical North Atlantic index (Figs. 10d–f). In fact, the model ensembles appear to partially capture certain aspects of the temporal evolution of the observed index for this region, including a tendency for more warming in the first half of the twentieth century than during the second half of the twentieth century. The model tends to be too cool during the 1880s, similar to the behavior discussed for the global mean and other regions, and slightly too warm at times during the early twentieth century.

The extension of the observed tropical North Atlantic index through 2004 in the left and middle columns (Figs. 10d,e) reveals a continued warming in the region that has recently brought the 10-yr running mean index to unprecedented warm levels, a finding that also appears consistent with the upward trend in each of the all-forcing model ensemble members during the 1990s.

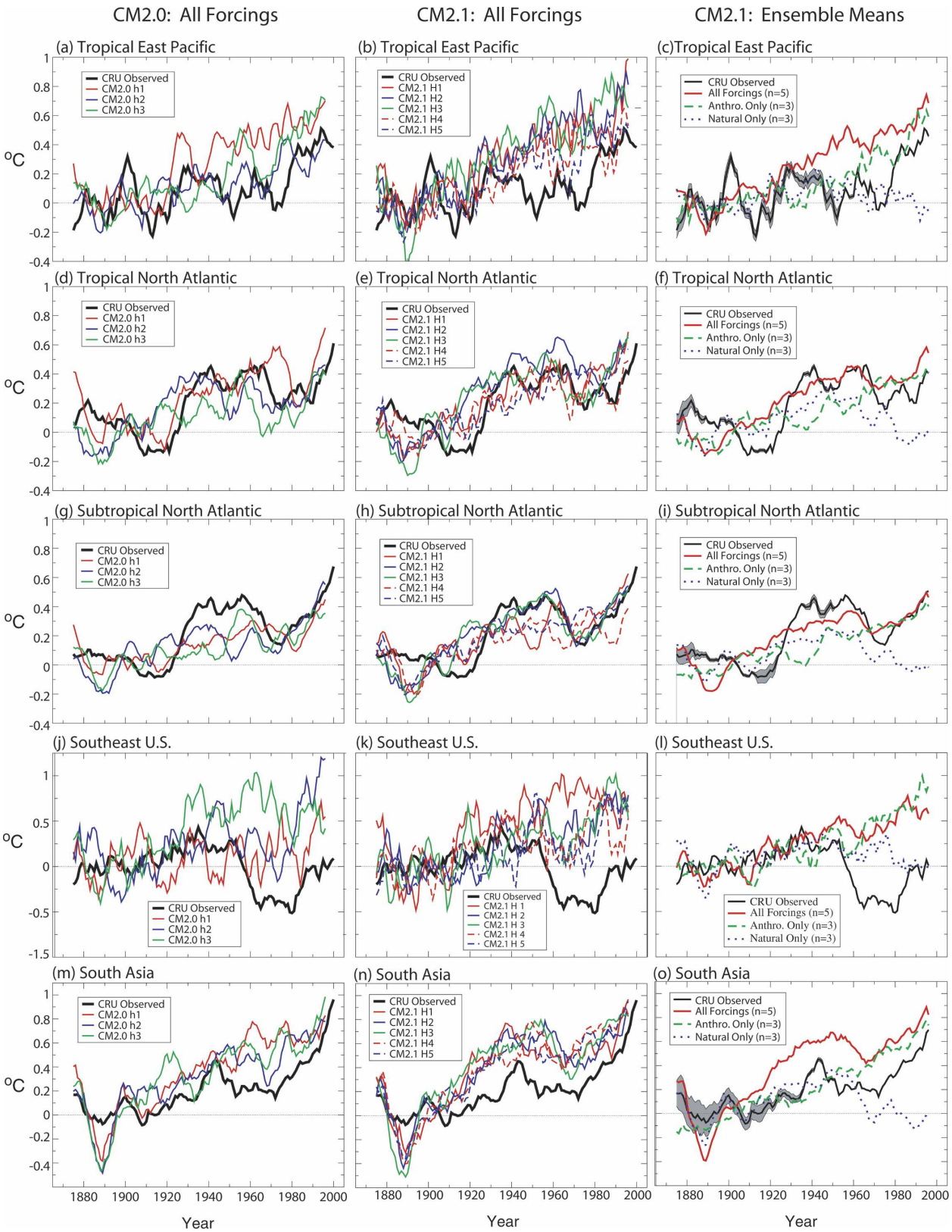


FIG. 10. As in Fig. 9, but for the following regions: (a)–(c) tropical eastern Pacific, (d)–(f) tropical North Atlantic, (g)–(i) subtropical North Atlantic, (j)–(l) southeast United States, and (m)–(o) South Asia. See Fig. 8 for illustrated location of specific regions.

This region is not one identified as having significant warming over the past half-century in Fig. 5, although the 100-yr warming trends evaluated in Fig. 7 appeared to be statistically significant. The time series results in Figs. 10d–f indicate that this result is due to strong multidecadal variability (either forced or internal) obscuring the century-scale warming trend signal over the past half-century. If this interpretation is correct, the warming late in the twentieth century in this region represents the emergence of a long-term anthropogenically forced warming signal from the background of substantial multidecadal variability.

The apparent anthropogenic warming trends in various tropical ocean basins (e.g., Figs. 7, 9, 10) may have important implications for tropical climate. For example, hurricane intensity theories (Emanuel 1987; Holland 1997; Tonkin et al. 1997) and hurricane model simulations (Knutson and Tuleya 2004) indicate a link between tropical SST warming and increased hurricane intensities and precipitation, with tropospheric lapse rate behavior also being an important factor (Shen et al. 2000). Recent observational studies (Emanuel 2005; Webster et al. 2005) suggest that a hurricane intensification signal may already be emerging in the observational data. Emanuel (2005) finds a strong correlation since the 1950s between variability/trends in tropical storm basin SSTs and a measure of the power dissipation of tropical cyclones in those basins. Landsea (2005) finds no evidence for an upward trend in Atlantic hurricane power dissipation measures. Potential long-term tendencies toward drought conditions in Africa are another area of concern, with CM2 simulations indicating that continued anthropogenic warming of tropical SSTs may lead to further drying of the Sahel (Held et al. 2005).

Figures 10g–i show temperature anomalies for a large region of the Atlantic that extends from 10° to 45°N, thus encompassing most of the subtropical North Atlantic. Note that this region, extending equatorward to 10°N, overlaps the tropical North Atlantic region just discussed. Relatively good agreement between all-forcing simulations and observations is also seen for this larger region, particularly for CM2.1. As was the case for the tropical North Atlantic, this region has been characterized by relatively little net warming over at least the last half-century, but may well be entering a period of pronounced warming. The underlying warming trend is likely driven by anthropogenic forcing (as inferred from the CM2.0 and CM2.1 simulations) as natural-only forcing runs fail to simulate the strong late-twentieth-century warming. The multidecadal variations are conceivably related to fluctuations of the North Atlantic thermohaline circulation as inferred from the North Atlantic Oscillation (NAO)-related heat and wind stress forcing for the twentieth century (e.g., Fig. 5 in Eden and Jung 2001). The timing of fluctuations in these Atlantic regions appears to be similar to that of the northern extratropics as a whole (Figs. 9d–f), and a muted form of these fluctuations is even discernible in the global mean (Figs. 9a–c).

The remaining subtropical time series in Fig. 10 include the southeast United States (Figs. 10j–l) and South Asia (Figs. 10m–o). The South Asia results appear similar in many respects to those for the nearby Indian Ocean/western Pacific warm pool (Figs. 9m–o). On the other hand, the southeast United States time series is an example of a striking disagreement between the models and observations (Figs. 10j–l). The observations there show a very gradual warming trend of about 0.25°C from the late 1800s to the mid–twentieth century, following by a pronounced cooling into the 1960s and 1970s, followed by a partial warming recovery since around 1980. In contrast, the CM2.1 all-forcing ensemble and its individual ensemble members show a consistent gradual warming through the entire 130-yr period. CM2.0 shows some tendency for long-term warming in two of three all-forcing ensemble members, but the third shows little warming over the century. Thus, any simulated warming appears to be small relative to internal climate variability in CM2.0. A caveat is that this region is one in which the model's simulated interannual surface temperature variability appears excessive (Fig. 2), which suggests the possibility of excessive simulated multidecadal climate variability in the region. The time series for the observations in the southeastern United States shows relatively small levels of decadal variability on shorter decadally resolved time scales, but contains a distinct very low frequency fluctuation during the second half of the twentieth century. The unusual behavior in the observed temperatures in this region and its contrast to the model simulations is a topic worthy of further investigation. Missing data effects appear to be minimal (Fig. 10l).

3) EXTRATROPICAL REGIONAL INDICES

In Fig. 11 are time series comparisons for several extratropical regional indices. The "rest of the United States" series (Figs. 11a–c) is shown as a contrast to the behavior for the southeast United States just discussed. This region includes the continental United States, but excludes Alaska, Hawaii, and the southeast United States. The observed series resembles that for the Northern Hemisphere extratropics (Fig. 9c) with strong warming from about 1910 to 1940, followed by a few decades of temporary cooling, followed by resumed strong warming after about 1970. The observed warming is within the range of the ensemble members, al-

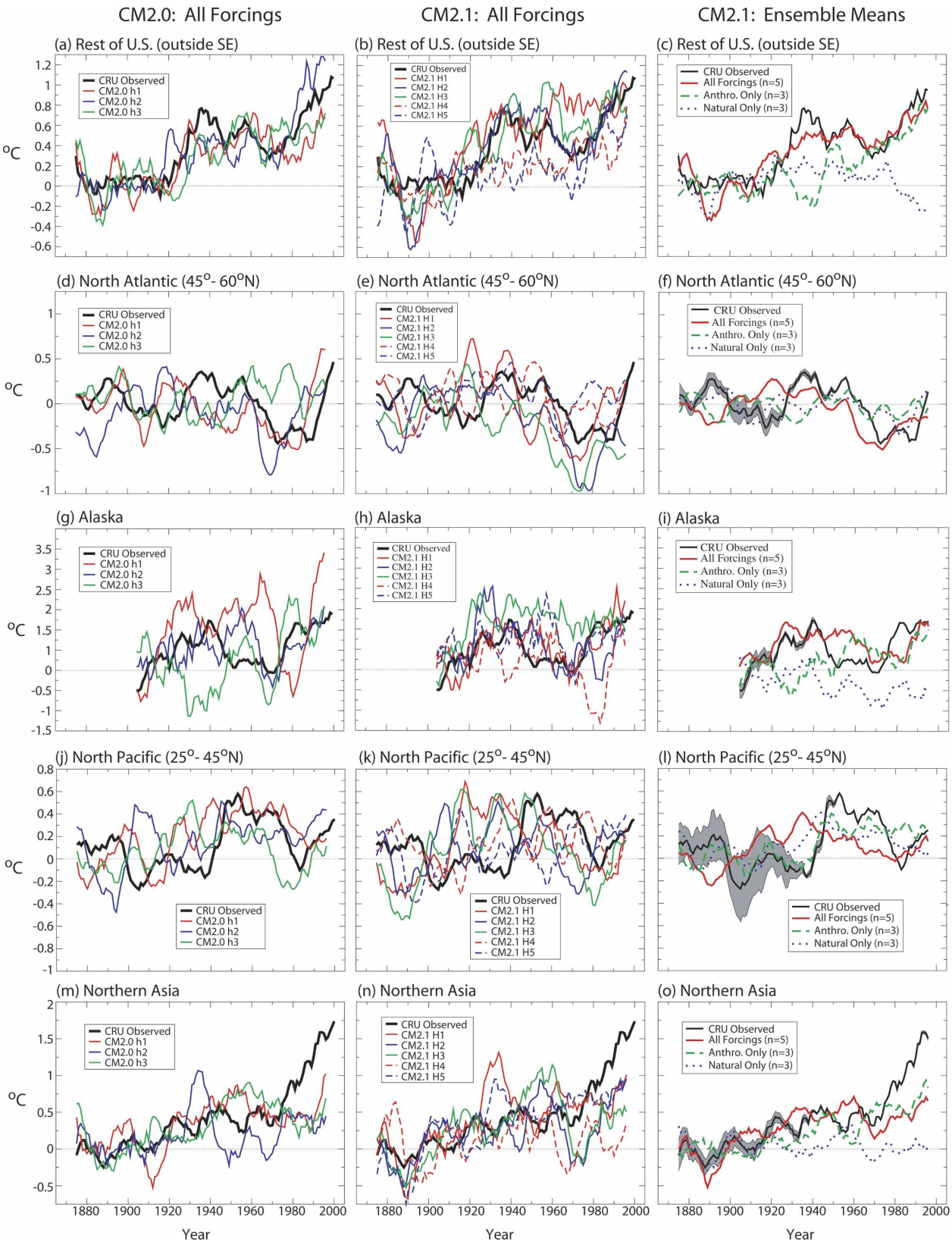


FIG. 11. As in Fig. 9, but for the following regions: (a)–(c) rest of continental United States (excluding southeast United States), (d)–(f) North Atlantic (45°–60°N), (g)–(i) Alaska, (j)–(l) North Pacific (25°–45°N); (m)–(o) northern Asia. See Fig. 8 for illustrated location of specific regions.

though CM2.1 shows a fairly large spread in behavior between different ensemble members. The all-forcing and anthropogenic-only forcing scenarios (Fig. 11c) depict a much more realistic late-twentieth-century warming than the natural-only forcing ensemble.

The high-latitude North Atlantic region (45°–65°N) was chosen for a case study (Figs. 11d–f) because of the intriguing cooling trends shown there in many of the modeled and observed trend maps (Figs. 5, 7) for both the last half-century and the full twentieth century. The observed 10-yr running mean time series for this region (Figs. 11d–f) shows pronounced multidecadal variability with a particularly strong cooling period from the 1950s to the 1970s and rapid warming in the 1990s. The observed curve typically falls within the range of the all-forcing ensemble members, particularly for the larger ($n = 5$) CM2.1 set of ensembles. During the second half of the twentieth century, several of the CM2.0 and CM2.1 individual all-forcing simulations (and the CM2.1 all-forcing ensemble mean) tend to follow a similar evolution to the observations in the late twentieth century. This agreement may be coincidental, as a few ensemble members show a quite different temporal evolution.

A notable characteristic of the model time series for the high-latitude North Atlantic region is the large scatter between ensemble members (note the increase in scale of the vertical axis compared with some of the previous figures). This enhanced internal climate variability leads to lower signal-to-noise ratios for radiative forcing responses in this region. The impact of missing observations is estimated to be moderate (Fig. 11f) during the period prior to 1920, but is smaller than the scatter of the different ensemble members.

Time series for Alaska (Figs. 11g–i) also show evidence of pronounced internal climate variability in the model, indicated by the scatter among ensemble members. A pronounced warming occurred beginning in the 1970s, although a warm period around 1940 appears nearly as warm as the 1990s. Alaska did not exhibit large-scale statistically significant warming trends in the annual trend map analyses (Figs. 5, 7), although parts of Alaska and large regions of Canada have apparently significant warming trends, particularly during winter and spring (Fig. 6).

The extratropical North Pacific index (Figs. 11j–l), like the high-latitude North Atlantic and Alaska indices, has large multidecadal variations in both the observations and model simulations, and a relatively large degree of scatter between individual all-forcing ensemble members. The observed cooling trend in this region during the last half of the twentieth century, a pronounced feature of Fig. 5a, appears to be part of a multidecadal variation, since Fig. 11l shows a comparable period of rapid warming prior to 1950. The observations typically fall within the range of the model ensemble members, although the CM2.1 all-forcing ensemble mean time series is not that similar to the observed series, apparently reflecting the impact of the large internal climate variability in the region. The potential impact of missing data appears to be substantial in this region, particularly prior to 1920. The error range is of the same order of magnitude as the model's internal variability (scatter between ensemble members).

Northern Asia (40°–70°N) was chosen for a case study (Figs. 11m–o) because the models did not warm as strongly as observed in this region according to the trend map analyses (e.g., Figs. 5–7). The observed time series for this region shows fairly monotonic warming, with a relatively moderate warming trend from the late 1800s through about the 1970s, and a much more rapid warming in recent decades. The total warming since the late 1800s in this region appears to exceed 1.5°C, or about twice the global mean warming. Such a strong warming is not apparent in any of the all-forcing model ensembles, although some individual ensemble members show some evidence for large internal climate variability in this region. In that regard, a caveat is the tendency in the model for excessive simulated interannual variability in this region (Fig. 2), which suggests that the internal climate variability may be exaggerated in the northern Asian region even on decadal time scales. In fact, the observed time series gives the appearance of having only modest decadal-scale variability, with the dominant feature being either very low-frequency (i.e., multicentury) variability or a long-term monotonic warming trend. We will return to this region for additional analysis in the next section.

7. Analysis of Arctic Oscillation impacts

A notable regional discrepancy in the CM2 historical simulations versus observations is the lack of pronounced warming over northern Asia in winter and spring late in the twentieth century (e.g., Figs. 6, 11m–o). Surface temperatures, including a substantial fraction of recent winter warming trends, in this region are known to be associated with (i.e., linearly congruent with) the NAO (Hurrell 1995) and Arctic Oscillation (AO; e.g., Thompson and Wallace 2000; Thompson et al. 2000). Therefore, a question arises as to whether the temperature discrepancies can be explained in terms of discrepancies in the observed and simulated AO behavior. To investigate this issue, we have performed additional analyses on northern Asia time series from the CM2.1 all-forcing runs, where we estimate the potential

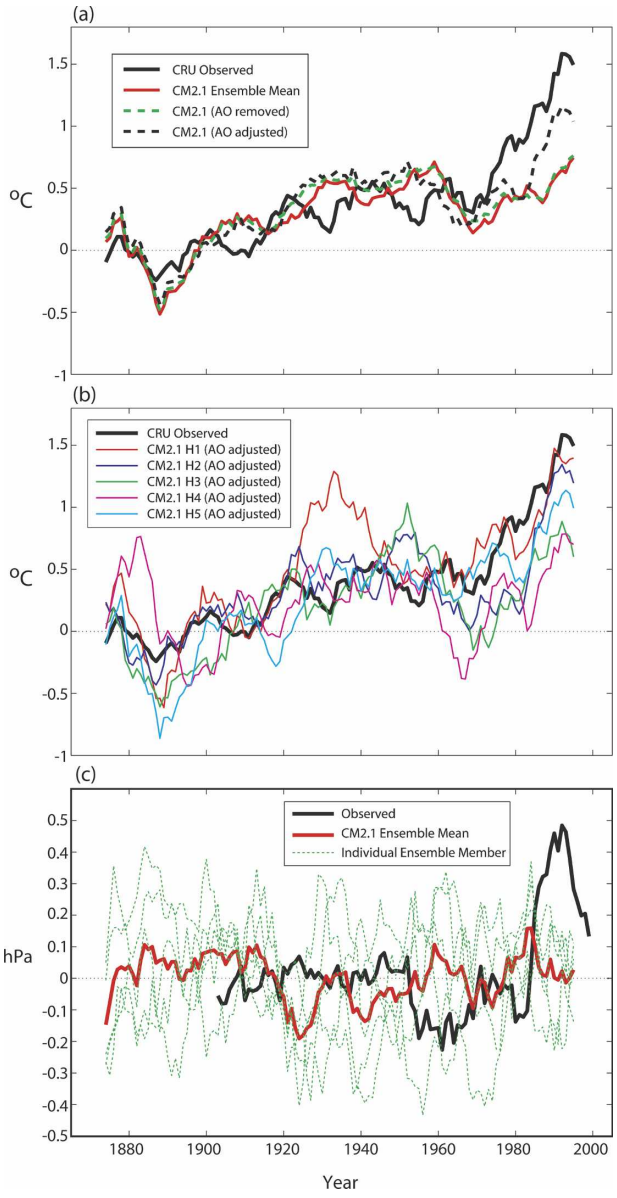


FIG. 12. (a), (b) Time series of 10-yr running mean northern Asia temperature index (Fig. 8) for observations (solid black) and the CM2.1 historical runs in °C. (a) Model ensemble mean (red), model ensemble mean with model AO contribution removed (green dashed) and model ensemble mean with model AO contribution replaced with an adjustment according to observed AO variability (black dashed). (b) Observed northern Asia index (thick black) and AO-adjusted indices for individual CM2.1 ensemble members (thin colored); (c) AO indices (hPa) for observations (black), model ensemble mean (red) and individual ensemble members (green). See text for details.

influence of recent observed positive AO anomalies on simulated northern Asia temperatures.

Figure 12c shows the observed AO index through 2004 (10-yr running means). The AO index was computed following Thompson and Wallace (2000) by projecting sea level pressure (SLP) fields onto the leading empirical orthogonal function (EOF) of SLP poleward of 20°N based on all months from January 1958 to April 1997. The monthly SLP data were obtained from the Trenberth SLP dataset available from the National Centers for Atmospheric Research (NCAR) Data Distribution Center (information online at http://dss.ucar.edu/datasets/ds010.1/) with corrections as described in Trenberth and Paolino (1980). The index shows little trend over most of the twentieth century, followed by a large shift toward positive (high index) values from around 1980 to the early 1990s, and a return toward values more typical of the early twentieth century in recent years.

For the model, data were projected onto the leading EOF of the five (combined) CM2.1 all-forcing scenario runs to form individual model AO indices (green dashed curves in Fig. 12c). Although not shown here, CM2.1 has a highly realistic simulation of the observed AO pattern (Delworth et al. 2006). None of the simulated AO indices has anomalies (from the 1901–30 reference period) as large as observed for the 1980s and 1990s although there is substantial variability in the model AO series. The ensemble mean of the model AO indices (red) shows little evidence for either a trend or for strong positive anomalies late in the twentieth century. Gillett et al. (2000; 2003) similarly found that model-simulated trends in the AO or NAO indices, using greenhouse gas and tropospheric sulfate aerosol forcing, tended to be significantly underestimated compared to the observed trends. However, Gillett (2000) also reported that the detection of a global response to these forcings was robust to exclusion of the AO-related warming, a result generally supported by Zwiers and Zhang's (2003) regional-scale analysis. Selten et al. (2004) also found little greenhouse gas–induced trend in the related NAO index in a large (n = 62) ensemble of 1940–2080 simulations, and that the late-twentieth-century NAO trend may have resulted largely from internal variability.

The impact on surface temperatures of our model's failure to reproduce the observed AO behavior was examined by using the observed AO index to adjust the model's northern Asia temperature series. Linear regressions were first computed between the AO index and the northern Asia time series for each model ensemble member, using annual mean values. The AO influence on the northern Asia temperature series was then removed from each model series using the model's linear regression relationship. The modified 10-yr running-mean ensemble mean northern Asia temperature index (dashed green curve in Fig. 12a) is quite similar to the original ensemble mean index (red curve). This in-

dicates that the AO in the model has minimal impact on multidecadal variability of the model's northern Asia temperature index. However, the AO does have a significant influence on northern Asia temperatures in the model in general, since if one considers all interannual time scales (i.e., by using unfiltered annual data), the average correlation between the model's AO and northern Asia temperature indices is about 0.43.

An adjusted northern Asia temperature index was obtained by adding to the model's AO-removed time series the product of the model's ensemble mean regression coefficients and the observed AO index. The resulting curves are labeled AO-adjusted in Figs. 12a,b. The ensemble AO-adjusted index (black dashed in Fig. 12a) shows a stronger warming beyond the mid-1980s than does the original unadjusted index (red curve in Fig. 12a). Thus, the strong positive observed AO anomalies late in the twentieth century, which are mostly absent in the model simulations, appear to account for about half of the ensemble mean warming discrepancy between the model and observations during this period for the northern Asia temperature index.

Figure 12b shows the individual AO-adjusted northern Asia temperature indices from the five CM2.1 all-forcing runs. None of the individual adjusted members have northern Asia warm anomalies as strong as observed in the 1990s according to this comparison, although one of the ensemble members comes fairly close to matching the observed warm anomalies.

The analysis shows that the presence of a positive AO trend in the observations, but not in the model, contributes to the differences in high-latitude temperature trends between the model and observations in the late twentieth century. In fact, one could argue that if the model had reproduced the observed warming, but without the strong AO anomalies, the agreement in surface temperatures for this region would have been for the wrong physical reasons.

8. Summary and conclusions

Climate models contain our hypotheses about the physical climate system, including how different components of the climate system interact and about the physics required to incorporate important climate forcing agents into climate change experiments. In the experiments described in this paper, two new climate models are forced with our current best estimates of the natural and anthropogenic forcings over the period 1861–2000. We compare model and observed surface temperature variations to check for consistency (or lack of consistency) over various periods. In addition, we use the model as a tool with which to probe and attempt to interpret aspects of the observed surface temperature record. The main findings are as follows:

1) The climate model simulations provide additional evidence for the hypothesis that the observed surface temperature warming during the twentieth century is too large to have been caused by internal climate variability or natural climate forcings alone. The observed trends (1901–2000 and 1949–2000) exceed model-generated internal variability and are significantly inconsistent with trends from natural-only forcing experiments over a wide majority of the global regions tested. There is little evidence that the model simulates too little internal climate variability; in fact, the model appears to simulate excessive internal climate variability on interannual time scales in many regions and at the global scale.
2) The historical simulations, which include forcing by both natural and anthropogenic agents, or by anthropogenic forcings alone, simulate the observed global mean temperature evolution during 1871–2000 reasonably well. These runs also simulate regional trends more realistically than runs forced by natural forcings alone. This provides further support for the view that the observed warming is a response of the climate system to the climate forcing agents used in the simulations, with anthropogenic forcings being of particular importance to the late-twentieth-century warming.
3) An exception to the agreement in the second finding is that the model shows possible excessive short-term cooling episodes that coincide with periods of strong volcanic activity. However, the discrepancies for Pinatubo and El Chichón in the late twentieth century may be due in a large part to El Niño influences during the posteruption periods. The causes for the simulated versus observed discrepancies during the volcanically active 1880s require further investigation.
4) Considering the twentieth century as a whole (1901–2000), one shortcoming of the all-forcing and anthropogenic-only forcing model simulations on a regional scale is the tendency for too much warming in the Tropics, with too little warming in higher latitudes. The causes of this apparent discrepancy will be the subject of further investigation, but may be partly related to internal variability. For example, in future warming scenarios with large positive radiative forcings, the model shows a much greater amplification of warming in the northern extratropics relative to the Tropics than is simulated in the historical scenario ensemble means.
5) Time series case studies for a number of regions

highlight a number of notable features of the historical simulations and observations. The model all-forcing runs agree with the observations reasonably well for the tropical and subtropical Atlantic regions. The results suggest that a century time-scale anthropogenic warming signal is emerging from a background of strong multidecadal variations in these regions as well as in the eastern tropical Pacific. The Indian Ocean/western Pacific warm pool region has a nearly monotonic warming trend since the 1950s. Rising tropical SSTs in various basins have possible links to increasing tropical cyclone intensities and African drought according to some recent studies. The extratropical North Pacific and high-latitude North Atlantic are regions characterized by very strong multidecadal variations in both models and observations, with long-term trends being difficult to discern. Northern Asia and the southeastern United States emerge as challenging regions to simulate, with the modeled warming being too weak in the former and too strong in the latter. The seasonality of the modeled warming over northern Asia is also not very realistic. The discrepancies in this region appear due in large part to the lack of a positive trend in the model's AO.

Several factors can contribute to the broad-scale regional discrepancies found in the study, such as too little warming over northern Asia, Canada, and the southern Indian Ocean, and too much warming over the southeastern United States. These include 1) errors or omissions in the specified forcing agents, 2) errors in the simulated response to forcing agents, 3) errors in the simulation of internal climate variability, and 4) errors in observed temperature data. Of these, the omission of indirect aerosol forcing from the climate forcing agents is particularly notable, as it may well be a substantial negative forcing (section 3). Inclusion of such a large negative forcing could degrade, rather than improve, the global mean agreement between model and observations presented here. With a smaller net positive forcing, a combination of higher climate sensitivity or reduced rate of heat uptake by the oceans may be required to attain a similar level of agreement on the global scale. Regionally, indirect aerosol forcing could potentially help reduce discrepancies between observed and simulated trends in areas such as the southeast United States or South Asia, and perhaps lead to greater discrepancies in other regions such as northern Asia, although our confidence in such speculations on regional responses is low.

All four of the potential sources of error mentioned above, which affect assessments of historical regional and global climate change in general, will require further investigation as the models, specified climate forcing agents, and observational datasets are further refined. In the context of the present model, improved simulations of interannual variability of surface temperature over extratropical land regions and evaluation of the aerosol forcing, including incorporation of indirect aerosol forcing, are important model development goals.

In summary, anthropogenic forcing agents, as opposed to natural forcing agents or internal climate variability, are simulated to be the dominant cause of the pronounced warming in the late twentieth century. This agrees with a number of previous global modeling studies of natural versus anthropogenic forcing (e.g., Tett et al. 1999; Stott et al. 2000; Broccoli et al. 2003; Meehl et al. 2004). Establishing the relative contributions of individual forcing agents in CM2 will require further specialized experiments and will be the topic of a future study.

Acknowledgments. We thank Andrew Wittenberg and Kirsten Findell and two anonymous reviewers for helpful comments on our manuscript, and Phil Jones and other developers of the HadCRUT2v dataset.

REFERENCES

Adams, J. B., M. E. Mann, and C. M. Ammann, 2003: Proxy evidence for an El Niño-like response to volcanic forcing. *Nature,* **426,** 274–278.

Andronova, N. G., E. V. Rozanov, F. Yang, M. E. Schlesinger, and G. L. Stenchikov, 1999: Radiative forcing by aerosols from 1850 to 1994. *J. Geophys. Res.,* **104,** 16 807–16 826.

Bell, J., P. Duffy, C. Covey, and L. Sloan, and the CMIP investigators, 2000: Comparison of temperature variability in observations and sixteen climate model simulations. *Geophys. Res. Lett.,* **27,** 261–264.

Boer, G. J., G. Flato, M. C. Reader, and D. Ramsden, 2000: A transient climate change simulation with greenhouse gas and aerosol forcing: Experimental design and comparison with the instrumental record for the twentieth century. *Climate Dyn.,* **16,** 405–425.

Broccoli, A. J., K. W. Dixon, T. D. Delworth, T. R. Knutson, R. J. Stouffer, and F. Zeng, 2003: Twentieth-century temperature and precipitation trends in ensemble climate simulations including natural and anthropogenic forcing. *J. Geophys. Res.,* **108,** 4798, doi:10.1029/2003JD003812.

Delworth, T. L., and T. R. Knutson, 2000: Simulation of early 20th century global warming. *Science,* **287,** 2246–2250.

——, and Coauthors, 2006: GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *J. Climate,* **19,** 643–674.

Eden, C., and T. Jung, 2001: North Atlantic interdecadal variability: Oceanic response to the North Atlantic Oscillation (1865–1997). *J. Climate,* **14,** 676–691.

Emanuel, K. A., 1987: The dependence of hurricane intensity on climate. *Nature,* **326,** 483–485.

——, 2005: Increasing destructiveness of tropical cyclones over the past 30 years. *Nature,* **436,** 686–688.

Folland, C. K., and Coauthors, 2001: Observed climate variability and change. *Climate Change 2001: The Scientific Basis: Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, J. T. Houghton et al., Eds., Cambridge University Press, 99–181.

GFDL Global Atmospheric Model Development Team, 2004: The new GFDL global atmosphere and land model AM2/LM2: Evaluation with prescribed SST simulations. *J. Climate,* **17,** 4641–4673.

Gillett, N. P., G. C. Hegerl, M. R. Allen, and P. A. Stott, 2000: Implications of changes in the Northern Hemisphere circulation for the detection of anthropogenic climate change. *Geophys. Res. Lett.,* **27,** 993–996.

——, F. W. Zwiers, A. J. Weaver, and P. A. Stott, 2003: Detection of human influence on sea-level pressure. *Nature,* **422,** 292–294.

——, M. F. Wehner, S. F. B. Tett, and A. J. Weaver, 2004: Testing the linearity of the response to combined greenhouse gas and sulfate aerosol forcing. *Geophys. Res. Lett.,* **31,** L14201, doi:10.1029/2004GL020111.

Gnanadesikan, A., and Coauthors, 2006: GFDL's CM2 global coupled climate models. Part II: The baseline ocean simulation. *J. Climate,* **19,** 675–697.

Goldenberg, S. B., C. W. Landsea, A. M. Mesta-Nuñez, and W. M. Gray, 2001: The recent increase in Atlantic hurricane activity: Causes and implications. *Science,* **293,** 474–479.

Griffies, S. M., and Coauthors, 2005: Formulation of an ocean model for global climate simulations. *Ocean Sci.,* **1,** 45–79.

Hansen, J. A., M. Sato, and R. Ruedy, 1997: Radiative forcing and climate response. *J. Geophys. Res.,* **102,** 6831–6864.

Hansen, J. E., and Coauthors, 2002: Climate forcing in Goddard Institute for Space Studies SI2000 simulations. *J. Geophys. Res.,* **107,** 4347, doi:10.1029/2001JD001143.

——, and Coauthors, 2005: Earth's energy imbalance: Confirmation and implications. *Science,* **303,** 1431–1435, doi:10.1126/science.1110252.

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

Held, I. M., T. L. Delworth, J. Lu, K. L. Findell, and T. R. Knutson, 2005: Simulation of Sahel drought in the 20th and 21st centuries. *Proc. Natl. Acad. Sci. U.S.A.,* **102,** 17 891–17 896.

Hoerling, M. P., J. W. Hurrell, T. Xu, G. T. Bates, and A. S. Phillips, 2004: Twentieth century North Atlantic climate change. Part II: Understanding the effect of Indian Ocean warming. *Climate Dyn.,* **23,** 391–405.

Holland, G. J., 1997: The maximum potential intensity of tropical cyclones. *J. Atmos. Sci.,* **54,** 2519–2541.

Houghton, J. T., and Coauthors, 2001: *Climate Change 2001: The Scientific Basis: Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 881 pp.

Hurrell, J. W., 1995: Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science,* **269,** 676–679.

Hurtt, G. C., S. Frolking, M. G. Fearon, B. Moore III, E. Shevliakova, S. Malyshev, S. W. Pacala, and R. A. Houghton, 2006: The underpinnings of land-use history: Three centuries of global gridded land-use transitions, wood harvest activity, and resulting secondary lands. *Global Change Biol*., in press.

International Ad Hoc Detection and Attribution Group, 2005: Detecting and attributing external influences on the climate system: A review of recent advances. *J. Climate,* **18,** 1291–1314.

Jones, P. D., and A. Moberg, 2003: Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001. *J. Climate,* **16,** 206–223.

——, K. R. Briffa, and F. H. Schweingruber, 1995: Tree-ring evidence of the widespread effects of explosive volcanic eruptions. *Geophys. Res. Lett.,* **22,** 1333–1336.

——, T. J. Osborn, K. R. Briffa, C. K. Folland, E. B. Horton, L. V. Alexander, D. E. Parker, and N. A. Rayner, 2001: Adjusting for sampling density in grid box land and ocean surface temperature time series. *J. Geophys. Res.,* **106,** 3371–3380.

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

Karoly, D. J., and Q. Wu, 2005: Detection of regional surface temperature trends. *J. Climate,* **18,** 4337–4343.

——, K. Braganza, P. A. Stott, J. M. Arblaster, G. A. Meehl, A. J. Broccoli, and K. W. Dixon, 2003: Detection of a human influence on North American climate. *Science,* **302,** 1200–1203.

Knutson, T. R., and R. E. Tuleya, 2004: Impact of CO_2-induced warming on simulated hurricane intensity and precipitation: Sensitivity to the choice of climate model and convective parameterization. *J. Climate,* **17,** 3477–3495.

——, 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.

——, T. L. Delworth, K. W. Dixon, and R. J. Stouffer, 1999: Model assessment of regional surface temperature trends (1949–1997). *J. Geophys. Res.,* **104,** 30 981–30 996.

Landsea, C. W., 2005: Hurricanes and global warming. *Nature,* **438,** E11–E12.

Lean, J., J. Beer, and R. S. Bradley, 1995: Reconstruction of solar irradiance since 1610: Implications for climate change. *Geophys. Res. Lett.,* **22,** 3195–3198.

Lin, S.-J., 2004: A "vertically Lagrangian" finite-volume dynamical core for global models. *Mon. Wea. Rev.,* **132,** 2293–2307.

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

Lohmann, U., and J. Feichter, 2005: Global indirect aerosol effects: A review. *Atmos. Chem. Phys.,* **5,** 715–737.

Meehl, G. M., W. M. Washingon, C. M. Amman, J. M. Arblaster, T. M. L. Wigley, and C. Tebaldi, 2004: Combinations of natural and anthropogenic forcings in twentieth-century climate. *J. Climate,* **17,** 3721–3727.

Milly, P. C. D., and A. B. Shmakin, 2002: Global modeling of land water and energy balances. Part I: The land dynamics (LaD) model. *J. Hydrometeor.,* **3,** 283–299.

Ming, Y., V. Ramaswamy, P. A. Ginoux, L.H. Horowitz, and L.M. Russell, 2005: GFDL GCM investigation of the indirect radiative effects of anthropogenic sulfate aerosol. *J. Geophys. Res.,* **110,** D22206, doi:10.1029/2005JD006161.

Mitchell, J. F. B., D. J. Karoly, G. C. Hegerl, F. W. Zwiers, M. R. Allen, and J. Marengo, 2001: Detection of climate change and

attribution of causes. *Climate Change 2001: The Scientific Basis: Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change,* J. T. Houghton et al., Eds., Cambridge University Press, 695–738.

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

Ramaswamy, V., and C.-T. Chen, 1997: Linear additivity of climate response for combined albedo and greenhouse perturbations. *Geophys. Res. Lett.,* **24,** 567–570.

——, and Coauthors, 2001: Radiative forcing of climate change. *Climate Change 2001: The Scientific Basis: Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change,* J. T. Houghton et al., Eds., Cambridge University Press, 349–416.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.,* **108,** 4407, doi:10.1029/2002JD002670.

Santer, B. D., and Coauthors, 2001: Accounting for the effects of volcanoes and ENSO in comparisons of modeled and observed temperature trends. *J. Geophys. Res.,* **106,** 28 033–28 059.

Sato, M., J. Hansen, M. P. McCormick, and J. Pollack, 1993: Stratospheric aerosol optical depth, 1850–1990. *J. Geophys. Res.,* **98,** 22 987–22 994.

Selten, F. M., G. W. Branstator, H. A. Dijkstra, and M. Kliphuis, 2004: Tropical origins for recent and future Northern Hemisphere climate change. *Geophys. Res. Lett.,* **31,** L21205, doi:10.1029/2004GL020739.

Shen, W., R. E. Tuleya, and I. Ginis, 2000: A sensitivity study of the thermodynamic environment on GFDL model hurricane intensity: Implications for global warming. *J. Climate,* **13,** 109–121.

Smith, T. M., and R. W. Reynolds, 2003: Extended reconstruction of global sea surface temperatures based on COADS data (1854–1997). *J. Climate,* **16,** 1495–1510.

Soden, B. J., R. T. Wetherald, G. L. Stenchikov, and A. Robock, 2002: Global cooling after the eruption of Mount Pinatubo: A test of climate feedback by water vapor. *Science,* **296,** 727–730.

Stenchikov, G., I. Kirchner, A. Robock, H. Graf, J. Antuna, R. G. Grainger, A. Lambert, and L. Thomason, 1998: Radiative forcing from the 1991 Mount Pinatubo volcanic eruption. *J. Geophys. Res.,* **103,** 13 837–13 857.

Stott, P. A., 2003: Attribution of regional-scale temperature changes to anthropogenic and natural causes. *Geophys. Res. Lett.,* **30,** 1728, doi:10.1029/2003GL017324.

——, S. F. B. Tett, G. S. Jones, M. R. Allen, J. F. B. Mitchell, and G. J. Jenkins, 2000: External control of 20th century temperature by natural and anthropogenic forcings. *Science,* **290,** 2133–2137.

Stouffer, R. J., G. Hegerl, and S. Tett, 2000: A comparison of surface air temperature variability in three 1000-yr coupled ocean–atmosphere model integrations. *J. Climate,* **13,** 513–537.

——, A. J. Weaver, and M. Eby, 2004: A method for obtaining pre-twentieth century initial conditions for use in climate change studies. *Climate Dyn.,* **23,** 327–339.

——, and Coauthors, 2006: GFDL's CM2 Global Coupled Climate Models. Part IV: Idealized climate change. *J. Climate,* **19,** 723–740.

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

Thompson, D. W. J., and J. M. Wallace, 2000: Annular modes in the extratropical circulation. Part I: Month-to-month variability. *J. Climate,* **13,** 1000–1016.

——, ——, and G. C. Hegerl, 2000: Annular modes in the extratropical circulation. Part II: Trends. *J. Climate,* **13,** 1018–1036.

Tonkin, H., G. Holland, C. Landsea, and S. Li, 1997: Tropical cyclones and climate change: A preliminary assessment. *Assessing Climate Change: Results from the Model Evaluation and Consortium for Climate Assessments,* W. Howe and A. Henderson-Sellers, Eds., Taylor and Francis Group, 327–360.

Trenberth, K. E., and D. A. Paolino Jr., 1980: The Northern Hemisphere sea-level pressure data set: Trends, errors, and discontinuities. *Mon. Wea. Rev.,* **108,** 855–872.

Webster, P. J., G. J. Holland, J. A. Curry, and H.-R. Chang, 2005: Changes in tropical cyclone number, duration, and intensity in a warming environment. *Science,* **309,** 1844–1846.

Winton, M., 2000: A reformulated three-layer sea ice model. *J. Atmos. Oceanic Technol.,* **17,** 525–531.

Wittenberg, A. T., A. Rosati, N.-C. Lau, and J. J. Ploshay, 2006: GFDL's CM2 global coupled climate models. Part III: Tropical Pacific climate and ENSO. *J. Climate*, **19,** 698–722.

Zwiers, F. W., and X. Zhang, 2003: Toward regional-scale climate change detection. *J. Climate,* **16,** 793–797.