LETTERS

Detection of Regional Surface Temperature Trends

DAVID J. KAROLY AND QIGANG WU

School of Meteorology, University of Oklahoma, Norman, Oklahoma

(Manuscript received 29 March 2005, in final form 14 June 2005)

ABSTRACT

Trends in surface temperature over the last 100, 50, and 30 yr at individual grid boxes in a 5° latitude–longitude grid are compared with model estimates of the natural internal variability of these trends and with the model response to increasing greenhouse gases and sulfate aerosols. Three different climate models are used to provide estimates of the internal variability of trends, one of which appears to overestimate the observed variability of surface temperature at interannual and 5-yr time scales. Significant warming trends are found at a large fraction of the individual grid boxes over the globe, a much larger fraction than can be explained by internal climate variations. The observed warming trends over the last 50 and 30 yr are consistent with the modeled response to increasing greenhouse gases and sulfate aerosols in most of the models. However, in some regions, the observed century-scale trends are significantly larger than the modeled response to increasing greenhouse gases and sulfate aerosols in the atmosphere. Warming trends consistent with the response to anthropogenic forcing are detected at scales on the order of 500 km in many regions of the globe.

1. Introduction

The assessment of the possible causes of observed climate change in the Intergovernmental Panel on Climate Change (IPCC) Third Assessment Report concluded that "most of the observed warming over the last 50 years is likely to have been due to the increase in greenhouse gas concentrations" (Mitchell et al. 2001). This conclusion was based on many studies of global and very large scale climate variations. That assessment also concluded that "surface temperature changes are detectable only on scales greater than 5, 000 km." Since then, it has been shown that an anthropogenic climate change signal is detectable in continental-scale regions using surface temperature changes over the twentieth century (Karoly et al. 2003; Stott 2003; Zwiers and Zhang 2003; Karoly and Braganza 2005). For example, Stott (2003) used simulations with the third Hadley Centre Coupled Ocean–Atmosphere General Circulation Model (HadCM3) model to show that most of the observed warming over the last 50 yr in six separate regions of the globe, including North America, Eurasia, and Australia, was likely to be due to the increase in greenhouse gases in the atmosphere. These studies did not consider regions smaller than continental scale. Giorgi (2002) showed that there were significant observed warming trends at a subcontinental scale in most regions of the globe.

Mitchell et al. (2001) in the IPCC Third Assessment Report provide a definition of detection of climate change that we use here: "*Detection* is the process of demonstrating that an observed change is significantly different (in a statistical sense) than can be explained by natural internal variability." Almost no studies have considered the detection of surface temperature trends at a regional scale, such as at the scale of an individual model grid box. This is because detection of anthropogenic climate change is a signal-to-noise problem, and the noise associated with internal variations of surface temperatures at a regional scale is greater than at larger continental or global scales. However, Knutson et al. (1999) performed a model assessment of regional surface temperature trends using the Geophysical Fluid

---

Corresponding author address: Dr. David Karoly, School of Meteorology, University of Oklahoma, 100 E. Boyd Street, Norman, OK 73019.
E-mail: dkaroly@ou.edu

Dynamics Laboratory (GFDL) R30 model. They showed that the observed temperature trends over the 49-yr period 1949–97 at individual grid boxes were outside the range of trends that could be explained due to internal climate variations alone over a sizeable fraction of the model grid. They also showed that the observed surface temperature trends were consistent with the response to anthropogenic forcing over larger regions than for internal climate variations alone.

Here, the analysis of Knutson et al. (1999) is extended to assess the significance of observed surface temperature trends at individual grid boxes relative to model-based estimates of natural internal climate variability. The observed trends are considered over three different periods, 1903–2002, 1953–2002, and 1973–2002, to determine if the period affects the significance of any trend detection. Three different climate models are used to estimate the internal unforced variability of trends over these periods, and the largest estimate is used to provide a conservative assessment of the significance of the trends. A field significance test is applied to determine whether the fraction of grid boxes that show locally significant trends is greater than would be expected by chance. In addition, the observed trends are compared with the modeled response to increasing greenhouse gases and sulfate aerosols, to assess whether the observed trends are consistent with the response to changes in anthropogenic forcing.

2. Data

Observed monthly mean surface temperature data for 1881–2002 on a 5° latitude–longitude grid (HadCRUT2v; Jones and Moberg 2003) are used. These data were obtained from quality-controlled instrumental observations over land and sea surface temperature observations and have been used in virtually all detection studies considering surface temperature changes. Only regions with data available throughout most of the twentieth century are considered, with grid boxes having more than 66% of years available included in the trend analysis.

The observed climate variations are compared to simulations with three global coupled ocean–atmosphere climate models from

- Geophysical Fluid Dynamics Laboratory, United States (GFDL R30);
- Hadley Centre, United Kingdom (HadCM2);
- National Center for Atmospheric Research, United States [Parallel Climate Model (PCM)].

Details of these models, including their resolution in the ocean and atmosphere, and original references can be found in McAvaney et al. (2001). All the models include representations of important physical processes in the atmosphere and the ocean, as well as sea ice and land surface processes. Two of the models (GFDL R30 and HadCM2) include adjustments of heat fluxes at the surface to reduce climate drift in the coupled model simulations. The other model (PCM) has no flux adjustments and maintains a stable global-mean climate when external forcings are not varied. Constant external forcing simulations ("control" runs) represent the natural internal variability of the unforced climate system. We have data from control runs for 900 yr from GFDL R30, 1085 yr from HadCM2, and 1000 yr from PCM.

We also analyze simulations with changes in anthropogenic forcing, including changing atmospheric concentrations of greenhouse gases and a representation of the effects of changing sulfate aerosols, to represent the human influence on climate ("GS" runs). The effects of changing sulfate aerosols are represented in these models using changes in surface albedo based on sulfate emissions. We have data from GS runs for three different ensemble members from GFDL R30 (Delworth et al. 2002), four members from HadCM2 (Tett et al. 1999), and four members from PCM (Washington et al. 2000). The data from the model grids are at a slightly higher horizontal resolution than the observed data and have been interpolated onto the observed 5° grid for the analysis.

The observed variability of the detrended surface temperatures on interannual and longer time scales is compared with the variability in the control climate model simulations to evaluate the quality of the simulations of natural internal climate variability. Simple linear detrending is used to attempt to remove some of the possible anthropogenic signal in the observed temperatures. The results are insensitive to the order of the polynomial trend removed from the indices. The standard deviations of the detrended observed and control model temperatures are calculated at each of the grid boxes with sufficient observational data. The ratio of the modeled standard deviation divided by that observed was determined at each grid box and then zonally averaged to provide a simple measure of whether the amplitude of the modeled variability is generally larger or smaller than observed (Fig. 1). The HadCM2 model has more variability than the other two models in most latitude bands at both interannual and 5-yr time scales and has substantially larger variability than observed in the Tropics and subtropics and in the NH high latitudes, as noted by Stouffer et al. (2000). The variability in the GFDL and PCM models is more similar to that observed in most latitude bands, although the ob-

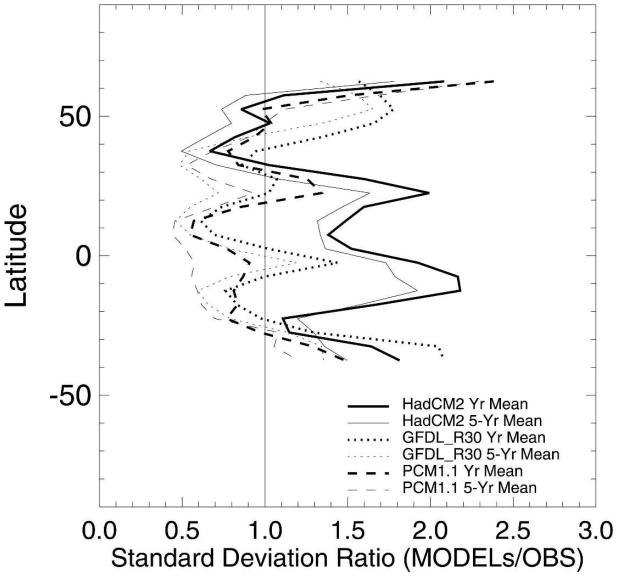


FIG. 1. Zonal average of the ratio of the standard deviation of surface temperature variations in individual grid boxes from control simulations divided by that for observed variations. Bold lines are for interannual variations, while thinner lines are for 5-yr average variations. The observed variations were linearly detrended to remove part of the possible anthropogenic influence.

served variability at longer time scales appears to be somewhat underestimated except at higher latitudes. Bell et al. (2000) noted that most climate models overestimate the variability of surface temperature over land in the NH middle and higher latitudes.

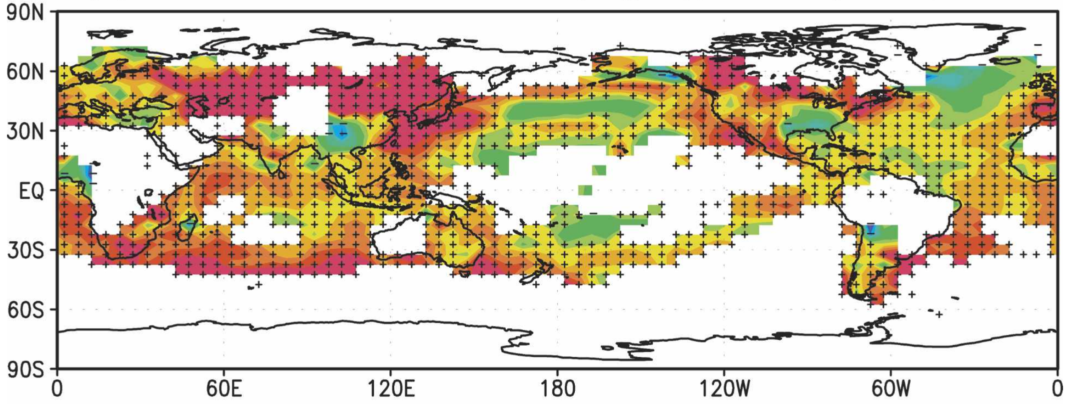
3. Assessment of observed trends

We compare the observed temperature trends over the last 100, 50, and 30 yr at individual grid boxes with the range of trends that might occur as a result of internal climate variability, as estimated from the long control simulations. At each grid box with sufficient observations, the observed linear trends over the periods 1903–2002, 1953–2002, and 1973–2002 are calculated. For each control simulation, the variability of 100-, 50-, and 30-yr linear trends is calculated for overlapping segments of the control simulations. The distribution of trends is approximately normally distributed, and we use the standard deviation of this distribution as the measure of the internal variability of linear trends. At each grid box, we identify warming trends that are significantly different from zero at the 95% level (using a one-sided test). Figure 2 shows the pattern of observed trends in color, with the significant trends identified using the internal variability estimated from the HadCM2 control run and marked with a plus symbol (+). About 80% of the individual grid boxes with sufficient observational data show significant warming trends over the period 1903–2002, 54% show significant warming over 1953–2002, and 45% over 1973–2002. The fraction of grid boxes with significant warming decreases for shorter trend periods primarily because the internal variability of trends increases for shorter periods, not because the magnitude of the trends decreases. The pattern of the warming is similar for the different periods, with more warming over land than ocean and more warming at higher latitudes in the Northern Hemisphere. However, the regions with significant warming occur roughly equally over land and ocean. Note that there is a very small number of grid boxes in Fig. 2 with locally significant cooling trends, marked by a minus symbol (−) but the fraction of such grid boxes with significant cooling trends is much less than 5%, and they are likely due to internal climate variations.

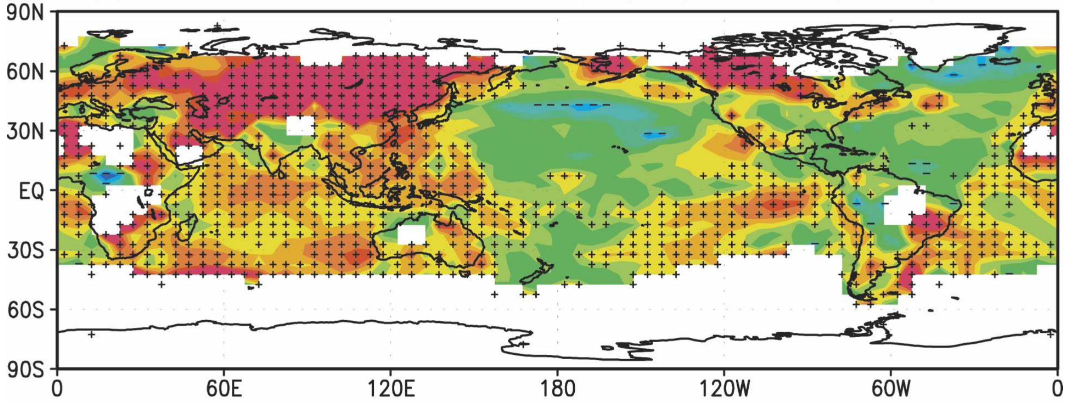
The analysis described above is a local test of the trend at each grid box. On average, in a stationary climate, we would expect 5% of the grid boxes to show warming trends significant at the 95% level due to random variability alone. As there is large spatial coherence of low-frequency variations of surface temperature, a much larger fraction of significant warming trends might occur by chance in a single test. Hence, we apply a field significance test based on the approach of Livezey and Chen (1983) to determine the range of fractions of grid boxes with significant trends that could occur due to internal variability, again estimated from the control simulations. We consider the linear trends at each grid box in a sample period from the control simulation and determine the fraction of grid boxes that show locally significant warming trends. This is repeated a large number of times by resampling different periods from the control simulation to determine the distribution of gridbox fractions with significant warming that could occur due to internal climate variations. Although the expected value is 5%, the 95th percentile for the distribution of fractions for significant 100-yr trends is 19%, 12% for 50-yr trends, and 17% for 30-yr trends. The fraction of grid boxes with significant observed warming trends is large for all the different trend periods and much greater than the range of trend fractions that could be expected due to internal climate variations, as simulated by the HadCM2 model.

The analysis has been repeated using the control simulations from the other two models to provide the estimate of internal climate variability of the trends, and the results for the fraction of grid boxes with locally significant observed warming trends are given in Table 1. This fraction is affected very little by the different model estimates of the internal variability of the trends,

(a) 1903–2002 OBS SAT Trend (deg. C), 80.4% (18.5%)



(b) 1953–2002 OBS SAT Trend (deg. C), 54.2% (11.6%)



(c) 1973–2002 OBS SAT Trend (deg. C), 45.4% (17.2%)

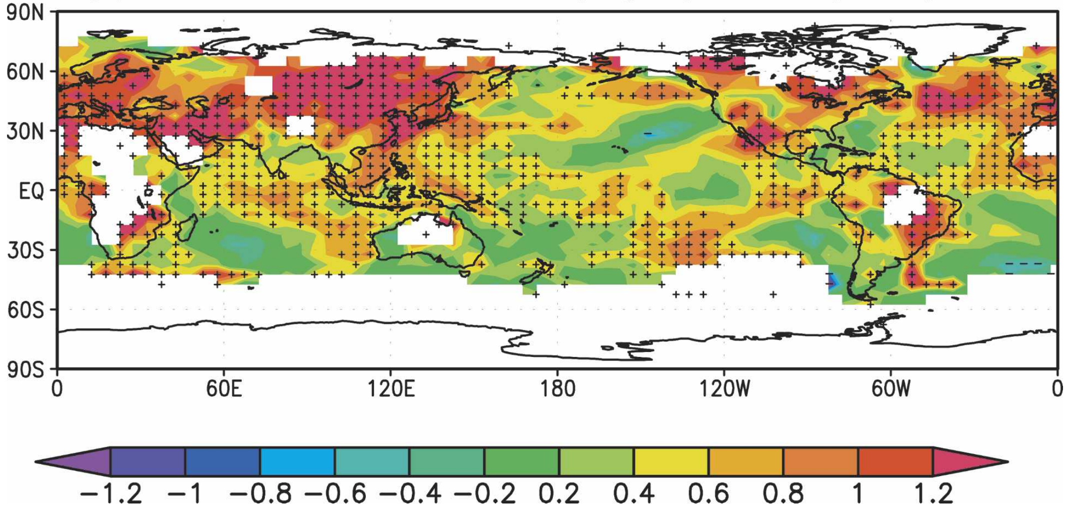


FIG. 2. Observed trends in surface temperature over the periods (top) 1903–2002, (middle) 1953–2002, and (bottom) 1973–2002. Plus (minus) symbols mark individual grid boxes where the observed trends are significantly larger (smaller) than zero at the 95% level using a one-sided test. Above each map is the fraction of grid boxes with significant warming trends and, in brackets, the possible range of fractions that could occur due to natural internal climate variability. The HadCM2 control simulation has been used to provide the estimate of the internal variability of the trends.

TABLE 1. (a) Fraction of grid boxes over the globe with observed linear warming trends locally significant at the 95% level over three different time intervals. Three different model estimates of the variability of trends due to natural internal climate variability are used. The number in brackets is the result of a field significance test to determine the largest value of this fraction due to natural climate variability. It represents at least the 95% significance level for the field significance of the fraction of grid boxes. (b) Same as in (a), but for the fraction of grid boxes where the observed trend is significantly different from the ensemble-mean model response to GS forcing, using a two-sided test at the 90% level.

(a) Significant observed trends?	1903–2002 warming	1953–2002 warming	1973–2002 warming
HadCM2	80% (19%)	54% (12%)	45% (17%)
GFDL R30	83% (12%)	64% (10%)	60% (11%)
PCM	80% (11%)	58% (12%)	51% (10%)
(b) Observed trend significantly different from GS response?	1903–2002 warming	1953–2002 warming	1973–2002 warming
HadCM2	34% (27%)	16% (22%)	10% (24%)
GFDL R30	37% (12%)	37% (11%)	19% (10%)
PCM	37% (12%)	20% (13%)	9% (9%)

with about 80% significant for 100-yr trends, 60% significant for 50-yr trends, and 50% significant for 30-yr trends. The main influence of the different model estimates of internal variability is in the field significance test, with the other two models showing smaller upper bounds for the possible fractions of grid boxes showing significant warming due to internal variability alone. This seems to be due to the larger and more spatially coherent internal variations in the HadCM2 simulations.

Next, we assess whether the observed warming trends at each of the grid boxes are locally consistent with the response to increasing greenhouse gases and sulfate aerosols, as estimated by the different models. We test whether the difference between the observed warming trend and the ensemble mean warming trend from the GS-forced simulations with each of the models is significantly different from zero in each grid box, this time using a two-sided test. Figure 3 shows the ensemble-mean warming trends over the three different periods from the HadCM2 model forced by increasing greenhouse gases and sulfate aerosols. Grid boxes where the model warming trend is significantly greater (smaller) than the observed trend are shown by a plus (minus) symbol. The fraction of the grid boxes where the observed trend is not consistent with the GS-forced HadCM2 model simulations is much smaller than in Fig. 2: 34% for the trend over 1903–2002, 16% for the trend over 1953–2002, and 10% for 1973–2002. Also shown in brackets in Fig. 3 is the 95th percentile of the distribution of fractions of grid boxes that give significantly different trends due to internal variability (estimated from the control simulation). The fraction of grid boxes where the observed trend over 1903–2002 is not consistent with the model GS response is outside the range that can be explained by internal variability. There are regions with model trends significantly smaller than observed, indicating that the HadCM2 model response to GS forcing alone cannot explain the observed trends over this period. There are many possible reasons for this, including that the HadCM2 response to GS forcing may be too small or that other radiative forcing factors may have contributed to the observed trends, such as changes in solar irradiance, volcanic aerosols, or land cover (Mitchell et al., 2001). The observed trends over the more recent periods are consistent with the HadCM2 response to GS forcing, as the fraction of grid boxes where the observed trends are not consistent with the model trends can be explained by internal climate variations.

The analysis has been repeated with the two other models' responses to GS forcing, and the results are given in Table 1. As for the HadCM2 model, the observed warming trends over 1903–2002 are not consistent with the PCM and GFDL model responses to GS forcing at a large fraction of grid boxes, and this cannot be explained by internal variability. The results for the more recent periods are similar to those using the HadCM2 model, with smaller regions where the observed trends are not consistent with the model response to GS forcing, and these can be explained by internal variations for the most recent period, 1973–2002, except for the GFDL model.

4. Discussion

We have shown that the observed warming trends over the last 100, 50, and 30 yr at individual grid boxes are significantly different from zero at large fractions of the grid boxes over the globe. These fractions are much too large to be explained as a chance occurrence due to internal climate variations. In addition, the observed trends over the most recent period (1973–2002) are consistent with the model response to increasing greenhouse gases and sulfate aerosols in most of the climate models but are significantly larger than the model trends over the last 100 yr in some regions.

It is possible that too many individual grid boxes could be identified with significant observed warming if the estimate of the internal variability of trends from the climate models is too small. However, at interannual and 5-yr time scales, the variability in the HadCM2 model may be too large, and using it to provide an estimate of internal variability may be conservative.

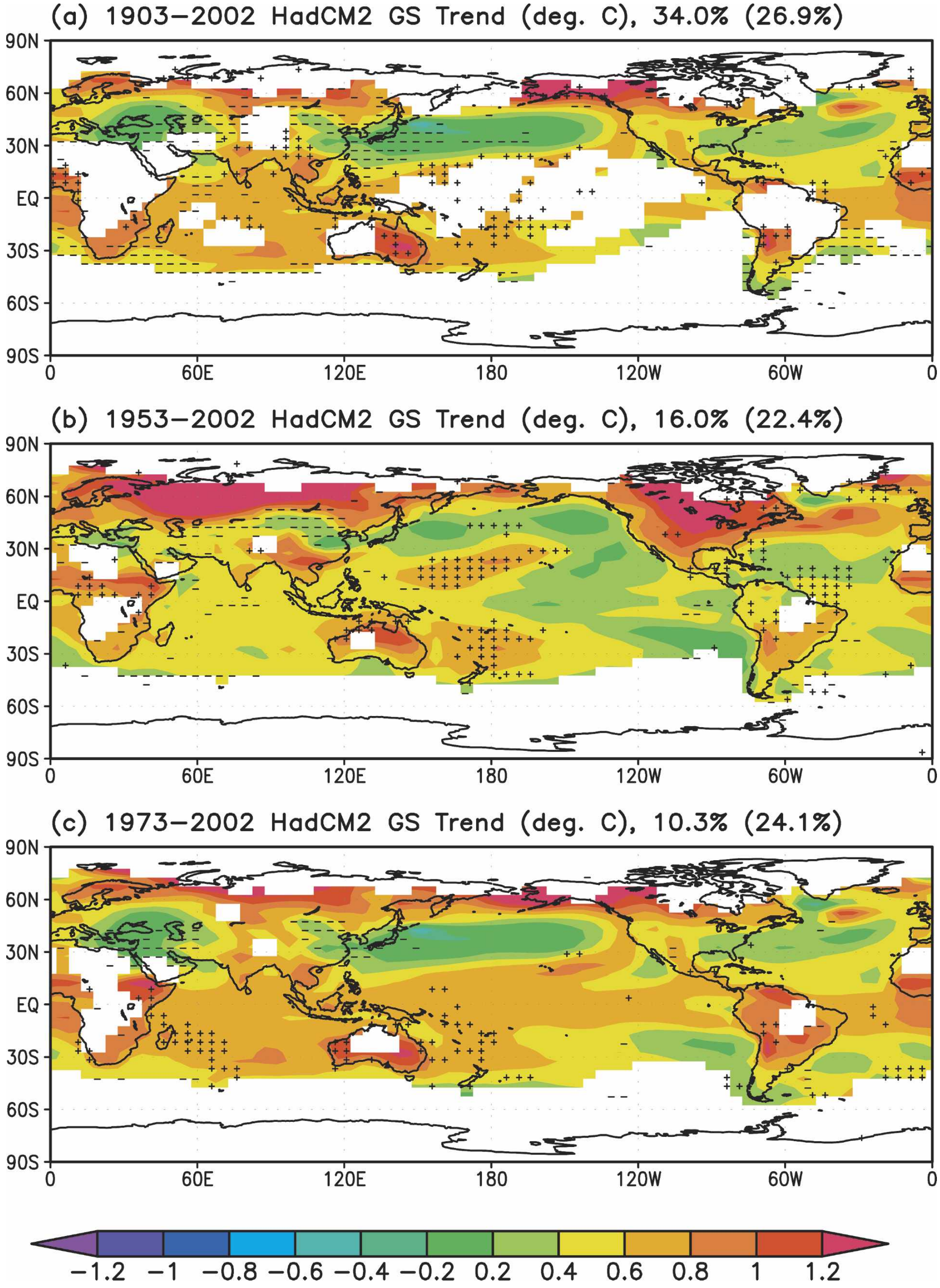


FIG. 3. Simulated trends in surface temperature in response to increasing greenhouse gases and sulfate aerosols over the periods (top) 1903–2002, (middle) 1953–2002, and (bottom) 1973–2002. The trends are the ensemble-mean values from the HadCM2 model. Plus (minus) symbols mark individual grid boxes where the model trends are significantly larger (smaller) than the observed trends at the 90% level using a two-sided test. Above each map is the fraction of grid boxes with significantly different trends than observed and, in brackets, the possible range of fractions that could occur due to internal variability. The HadCM2 control simulation has been used to provide the estimate of the internal variability of the trends.

The statement that the observed trends are consistent with the model response to GS forcing is not a strong attribution statement, as there are other climate forcings that may be more important at local rather than global scales. We have not considered the possible responses to land use or land cover changes, nor to increases in carbon black aerosols, any of which may be important contributors to the observed warming trends in some grid boxes.

We have shown that a significant warming trend can be detected in surface temperatures at scales on the order of 500 km in most regions of the globe. This result is primarily associated with the substantial global-mean warming that is larger than the variations in the spatial pattern of the warming. Note that we have not assessed the pattern of observed warming at 500-km scales and are not saying that variations in the warming at 500-km scales can be detected. In practice, we have shown that the observed warming trend is larger than can be explained by internal variability at 500-km scales. Hence, we believe that the statement in the IPCC Third Assessment Report on the scale of detectable temperature changes is no longer correct.

Our results are likely to be of considerable practical importance, as natural and human systems are more likely to be affected by regional temperature changes when these changes are outside the range normally experienced by the systems. They help to explain why a number of ecological systems appear to be changing in a way that is consistent with that expected as a response to regional and global warming trends (Parmesan and Yohe 2003; Root et al. 2003, 2005).

Acknowledgments. QW is supported by a Gary Comer Science and Education Foundation Postdoctoral Fellowship. We wish to acknowledge the efforts of the many scientists involved in developing, running, and archiving the model simulations that have been used in this study. Without their efforts, this analysis would not have been possible.

REFERENCES

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

Delworth, T. L., R. J. Stouffer, K. W. Dixon, M. J. Spelman, T. R. Knutson, A. J. Broccoli, P. J. Kushner, and R. T. Wetherald, 2002: Review of simulations of climate variability and change with the GFDL R30 coupled climate model. *Climate Dyn.,* **19,** 555–574.

Giorgi, F., 2002: Variability and trends of sub-continental scale surface climate in the twentieth century. Part I: Observations. *Climate Dyn.,* **18,** 675–691.

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

Karoly, D. J., and K. Braganza, 2005: Attribution of recent temperature changes in the Australian region. *J. Climate,* **18,** 457–464.

——, 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., T. L. Delworth, K. Dixon, and R. J. Stouffer, 1999: Model assessment of regional surface temperature trends (1949–97). *J. Geophys. Res.,* **104,** 30 981–30 996.

Livezey, R. E., and W. Y. Chen, 1983: Statistical field significance and its determination by Monte Carlo techniques. *Mon. Wea. Rev.,* **111,** 46–49.

McAvaney, B. J., and Coauthors, 2001: Model evaluation. *Climate Change 2001: The Scientific Basis*, J. T. Houghton et al., Eds., Cambridge University Press, 471–523.

Mitchell, J. F. B., D. J. Karoly, M. R. Allen, G. Hegerl, F. Zwiers, and J. Marengo, 2001: Detection of climate change and attribution of causes. *Climate Change 2001: The Scientific Basis*, J. T. Houghton et al., Eds., Cambridge University Press, 695–738.

Parmesan, C., and G. Yohe, 2003: A globally coherent fingerprint of climate change impacts across natural systems. *Nature,* **421,** 37–42.

Root, T. L., J. T. Price, K. R. Hall, S. H. Schneider, C. Roszenweig, and J. A. Pounds, 2003: Fingerprints of global warming on wild animals and plants. *Nature,* **421,** 57–60.

——, D. P. MacMynowski, M. D. Mastrandrea, and S. H. Schneider, 2005: Human modified temperatures induce species changes: Joint attribution. *Proc. Natl. Acad. Sci. USA,* **102,** 7465–7469.

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

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

Tett, S. F. B., P. A. 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.

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

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