CHAPTER 5

How well can the observed vertical temperature changes be reconciled with our understanding of the causes of these changes?

Convening Lead Author: Benjamin D. Santer, DOE LLNL
Lead Authors: J.E. Penner, Univ. of MI; P.W. Thorne, U.K. Met. Office
Contributing Authors: W. Collins, NSF NCAR; K. Dixon, NOAA; T.L. Delworth, NOAA; C. Doutriaux, DOE LLNL; C.K. Folland, U.K. Met. Office; C.E. Forest, MIT; I.M. Held, NOAA; J.R. Lanzante, NOAA; G.A. Meehl, NSF NCAR; V. Ramaswamy, NOAA; D.J. Seidel, NOAA; M.F. Wehner, DOE LBNL; T.M.L. Wigley, NSF NCAR

KEY FINDINGS

Fingerprint Pattern Studies

Fingerprint studies use rigorous statistical methods to compare spatial and temporal patterns of climate change in computer models and observations.

1. Both human and natural factors have affected Earth's climate. Computer models are the only tools we have for estimating the likely climate response patterns ("fingerprints") associated with different forcing mechanisms.

To date, most formal fingerprint studies have focused on a relatively small number of climate forcings. Our best scientific understanding is that:

- Increases in well-mixed greenhouse gases (which are primarily due to fossil fuel burning) result in large-scale warming of the Earth's surface and troposphere, and cooling of the stratosphere.
- Human-induced changes in the atmospheric burdens of sulfate aerosol particles cause regional cooling of the surface and troposphere.
- Depletion of stratospheric ozone cools the lower stratosphere and upper troposphere.
- Large volcanic eruptions cool the surface and troposphere (for 3 to 5 years) and warm the stratosphere (for 1 to 2 years).
- Increases in solar irradiance warm globally throughout the atmospheric column (from the surface to the stratosphere).

2. Results from many different fingerprint studies provide consistent evidence of a human influence on the three-dimensional structure of atmospheric temperature over the second half of the 20th century.

Robust results are:

- Detection of greenhouse-gas and sulfate aerosol signals in observed surface temperature records.
- Detection of an ozone depletion signal in stratospheric temperatures.
- Detection of the combined effects of greenhouse gases, sulfate aerosols, and ozone in the vertical structure of atmospheric temperature changes (from the surface to the stratosphere).

3. Natural factors have influenced surface and atmospheric temperatures, but cannot fully explain their changes over the past 50 years.

- The multi-decadal climatic effects of volcanic eruptions and solar irradiance changes are identifiable in some fingerprint studies, but results are sensitive to analysis details.

Trend Comparisons

Linear trend comparisons are less powerful than "fingerprinting" for studying cause-effect relationships, but when treated with caution can highlight important differences (and similarities) between models and observations.

4. When run with natural and human-caused forcings, model global-mean temperature trends for individual atmospheric layers are consistent with observations.

5. Comparing trend differences between the surface and the troposphere exposes potential discrepancies between models and observations in the tropics.

- Differencing surface and tropospheric temperature time series (a simple measure of the temperature lapse rate) removes much of the common variability between these layers. This makes it easier to identify discrepancies between modeled and observed lapse-rate changes.
- For globally averaged temperatures, model-predicted trends in tropospheric lapse rates are consistent with observed results.
- In the tropics, most observational data sets show more warming at the surface than in the troposphere, while most model runs have larger warming aloft than at the surface.

Amplification of Surface Warming in the Troposphere

6. In the tropics, surface temperature changes are amplified in the free troposphere. Models and observations show similar amplification behavior for monthly and interannual temperature variations, but not for decadal temperature changes.

- Tropospheric amplification of surface temperature anomalies is due to the release of latent heat by moist, rising air in regions experiencing convection.
- Despite large inter-model differences in variability and forcings, the size of this amplification effect is remarkably similar in the models considered here, even across a range of timescales (from monthly to decadal).
- On monthly and annual timescales, amplification is also a ubiquitous feature of observations, and is very similar to values obtained from models and basic theory.
- For longer-timescale temperature changes over 1979 to 1999, only one of four observed upper-air data sets has larger tropical warming aloft than in the surface records. All model runs with surface warming over this period show amplified warming aloft.
- These results could arise due to errors common to all models; to significant non-climatic influences remaining within some or all of the observational data sets, leading to biased long-term trend estimates; or a combination of these factors. The new evidence in this Report (model-to-model consistency of amplification results, the large uncertainties in observed tropospheric temperature trends, and independent physical evidence supporting substantial tropospheric warming) favors the second explanation.
- A full resolution of this issue will require reducing the large observational uncertainties that currently exist. These uncertainties make it difficult to determine whether models still have common, fundamental errors in their representation of the vertical structure of atmospheric temperature change.

Other Findings

7. It is important to account for both model and observational uncertainty in comparisons between modeled and observed temperature changes.

- There are large "construction uncertainties" in the process of generating climate data records from raw observations. These uncertainties can critically influence the outcome of consistency tests between models and observations.

8. Inclusion of spatially variable forcings in the most recent climate models does not fundamentally alter simulated lapse-rate changes at the largest spatial scales.

- Changes in black carbon aerosols and land use/land cover (LULC) may have had significant influences on regional temperatures, but these influences have not been quantified in formal fingerprint studies.
- These forcings were included for the first time in about half the global model simulations considered here. Their incorporation did not significantly affect simulations of lapse-rate changes at very large spatial scales (global and tropical averages).

Chapter 5: Recommendations

1. Separate the uncertainties in climate forcings from uncertainties in the climate response to forcings.

The simulations of 20th century (20CEN) climate analyzed here show climate responses that differ because of differences in:

- Model physics and resolution;
- The forcings incorporated in the 20CEN experiment;
- The chosen forcing history, and the manner in which a specific forcing was applied.
- Model initial conditions.

We consider it a priority to partition the uncertainties in climate forcings and model responses, and thus improve our ability to interpret differences between models and observations. This could be achieved by better coordination of experimental design, particularly for the 20CEN simulations that are most relevant for direct comparison with observations.

2. Quantify the contributions of changes in black carbon aerosols and land use/land cover to recent large-scale temperature changes.

We currently lack experiments in which the effects of black carbon aerosols and LULC are varied individually (while holding other forcings constant). Such "single forcing" runs will help to quantify the contributions of these forcings to global-scale changes in lapse rates.

3. Explicitly consider model and observational uncertainty.

Efforts to evaluate model performance or identify human-induced climate change should always account for uncertainties in both observations and in model simulations of historical and future climate. This is particularly important for comparisons involving long-term changes in upper-air temperatures. It is here that current observational uncertainties are largest and require better quantification.

4. Perform the "next generation" of detection and attribution studies.

Formal detection and attribution studies utilizing the new generation of model and observational data sets detailed herein should be undertaken as a matter of priority.

1. INTRODUCTION

A key scientific question addressed in this report is whether the Earth's surface has warmed more rapidly than the troposphere over the past 2-3 decades (NRC, 2000). Chapter 1 noted that there are good physical reasons why we do not expect surface and tropospheric temperatures to evolve in unison at all places and on all timescales. Chapters 2, 3, and 4 summarized our current understanding of observed changes in surface and atmospheric temperatures. These chapters identified important differences between surface and tropospheric temperatures, some of which may be due to remaining problems with the observational data, and some of which are likely to be real.

Climate models can be used to simulate the response to changes in a single forcing or a combination of forcings, and thus have real advantages for studying cause-effect relationships.

In Chapter 5, we seek to explain and reconcile the apparently disparate estimates of observed changes in surface and tropospheric temperatures. We make extensive use of computer models of the climate system. In the real world, multiple "climate forcings" vary simultaneously, and it is difficult to identify and separate the climate effects of individual factors. Furthermore, the experiment that we are performing with the Earth's climate system lacks a suitable control – we do not have a convenient "parallel Earth" on which there are no human-induced changes in greenhouse gases, aerosols, or other climate forcings. Climate models can be used to perform such controlled experiments, or to simulate the response to changes in a single forcing or combination of forcings, and thus have real advantages for studying cause-effect relationships. However, models also have systematic errors that can diminish their usefulness as a tool for interpretation of observations (Gates *et al.,* 1999; McAvaney *et al.,* 2001).

BOX 5.1: Climate Models

Climate models provide us with estimates of how the real world's climate system behaves and is likely to respond to changing natural and human-caused forcings. Because of limitations in our physical understanding and computational capabilities, models are simplified and idealized representations of a very complex reality. The most sophisticated climate models are direct descendants of the computer models used for weather forecasting. While weather forecast models seek to predict the specific timing of weather events over a period of days to several weeks, climate models attempt to simulate future changes in the *average distribution* of weather events.

Because the climate system is chaotic, fully coupled models of the atmosphere and ocean cannot simulate exactly the same sequence of individual weather events that occurred in the real world (see Section 2). Such models can, however, capture many of the statistical characteristics of observed weather and climate variability, on timescales of days to decades. Many models have demonstrated skill in their portrayal of major modes of observed climate variability, such as the North Atlantic Oscillation (Hurrell *et al.,* 2003), the El Niño/Southern Oscillation (ENSO; AchutaRao and Sperber, 2006) or the Atlantic Multidecadal Oscillation (Knight *et al.,* 2005). This variability contributes to the background "noise" against which any signal of human effects on climate must be detected[a.] (Box 5.5).

Simulations of 21st century climate are typically based on "scenarios" of future emissions of GHGs, aerosols and aerosol precursors, which in turn derive from scenarios of population changes, economic growth, energy usage, developments in energy production technology, *etc.* Climate models are also used to "hindcast" the climate changes that we have observed over the 20th century. When run in "hindcast" mode, a climate model is not constrained by actual weather observations from satellites or radiosondes. Instead, it is driven by our best estimates of changes in some (but probably not all) of the major forcings, such as GHG concentrations, the Sun's energy output, and the amount of volcanic dust in the atmosphere. In hindcast experiments, a climate model is free to simulate the full four-dimensional (latitude, longitude, height/depth and time) distributions of temperature, moisture, *etc.* Comparing the results of such an experiment with long observational records constitutes a valuable test of model performance.

A more complete assessment of climate models and their ability to represent many different aspects of the climate system will be covered in CCSP Synthesis and Assessment Product 3.1: "Climate Models: An Assessment of Strengths and Limitations for User Applications."

a. There is some evidence that human-induced climate change may modulate the statistical behavior of existing modes of climate variability (Hasselmann, 1999).

We evaluate published research that has made rigorous quantitative comparisons of modeled and observed temperature changes, primarily over the satellite and radiosonde eras. Some new model experiments (performed in support of the IPCC Fourth Assessment Report) involve simultaneous changes in a wide range of natural and human-induced climate forcings. These experiments are highly relevant for direct comparison with satellite-, radiosonde-, and surface-based temperature observations. We review their key results here.

2. MODEL SIMULATIONS OF RECENT TEMPERATURE CHANGE

Many different types of computer model are used for studying climate change issues (Meehl, 1984; Trenberth, 1992; see Box 5.1). Models span a large range of complexity, from the one- or two-dimensional energy-balance models (EBMs) through Earth system Models of Intermediate Complexity (EMICs) to full three-dimensional atmospheric General Circulation Models (AGCMs) and coupled atmosphere-ocean GCMs (CGCMs). Each type has advantages and disadvantages for specific applications. The more complex AGCMs and CGCMs are most appropriate for understanding problems related to the atmosphere's vertical temperature structure, since they explicitly resolve that structure, and incorporate many of the physical processes (*e.g.,* convection, interactions between clouds and radiation) thought to be important in maintaining atmospheric temperature profiles. They are also capable of representing the horizontal and vertical structure of unevenly distributed climate forcings that may contribute to differential warming of the surface and troposphere. Examples include volcanic aerosols (Robock, 2000) or the sulfate and soot aerosols arising from fossil fuel or biomass burning (Penner *et al.,* 2001; Ramaswamy *et al.,* 2001a,b).

AGCM experiments typically rely on an atmospheric model driven by observed time-varying changes in sea-surface temperatures (SSTs) and sea-ice coverage. This is a standard reference experiment that many AGCMs have performed as part of the Atmospheric Model Intercomparison Project ("AMIP"; Gates *et al.,* 1999). The AMIP-style experiments discussed here also include specified changes in a variety of natural and human-caused forcing factors (Hansen *et al.,* 1997, 2002; Folland *et al.,* 1998; Tett and Thorne, 2004).

In both observations and climate models, variations in the El Niño-Southern Oscillation (ENSO) have pronounced effects on surface and tropospheric temperatures (Yulaeva and Wallace, 1994; Wigley, 2000; Santer *et al.,* 2001; Hegerl and Wallace, 2002; Hurrell *et al.,* 2003). When run in an AMIP configuration, an atmospheric model "sees" the same changes in ocean surface temperature that the real world's atmosphere experienced. The time evolution of ENSO effects on atmospheric temperature is therefore very similar in the model and observations. This facilitates the direct comparison of modeled and observed temperature changes[1]. Furthermore, AMIP experiments reduce climate noise by focusing on the random variability arising from the atmosphere rather than on the variability of the coupled atmosphere-ocean system (which is larger in amplitude). This "noise reduction" aspect of AMIP runs has been exploited in efforts to identify human effects on year-to-year changes in atmospheric temperatures (Folland *et al.,* 1998; Sexton *et al.,* 2001) and volcanic influences on surface air temperature (Mao and Robock, 1998).

Climate models are also used to "hindcast" the climate changes that we have observed over the 20th century. Comparing the results of such an experiment with long observational records constitutes a valuable test of model performance.

One disadvantage of the AMIP experimental set-up is that significant errors in one or more of the applied forcing factors (or omission of key forcings) are not "felt" by the prescribed SSTs. Such errors are more obvious in a CGCM experiment, where the ocean surface is free to respond to imposed forcings. The lack of an ocean response, combined with the masking effects of natural variability, make it difficult to use an AMIP-style experiment to estimate the slow response of the climate system to an imposed forcing change[2]. CGCM experiments

[1] This does not mean, however, that the atmospheric model will necessarily capture the correct amplitude and horizontal and vertical structure of the tropospheric temperature response to the specified SST and sea-ice changes. Even with the specification of observed ocean boundary conditions, the time evolution of modes of variability that are forced by both the ocean and the atmosphere (such as the North Atlantic Oscillation; see Rodwell *et al.,* 1999) will not be the same in the model and in the real world (except by chance).

[2] Volcanic forcing provides an example of the signal

are more useful for this specific purpose (see Chapter 1, Figure 1.3).

The CGCM experiments of interest here involve a model that has been "spun-up" until it reaches some quasi-steady climate state[3]. The CGCM is then run with estimates of how a variety of natural and human-caused climate forcings have changed over the 20th century. We refer to these subsequently as "20CEN" experiments. Since the true state of the climate system is never fully known, the same forcing changes are applied *n* times,[4] each time starting from a slightly different initial climate state. This procedure yields *n* different realizations of climate change. All of these realizations contain some underlying "signal" (the climate response to the imposed forcing changes) upon which are superimposed *n* different manifestations of "noise" (natural internal climate variability). Taking averages over these *n* realizations yields less noisy estimates of the signal (Wigley *et al.,* 2005a).

AMIP-style experiments and CGCM runs are both useful tools for exploring the possible causes of differential warming.

In a CGCM, ocean temperatures are fully predicted rather than prescribed. This means that even a (hypothetical) CGCM which perfectly captured all important aspects of ENSO physics would not have the same timing of El Niño and La Niña events as the real world (except by chance). The fact that ENSO variability – and its effects on surface and atmospheric temperatures – does not "line up in time" in observations and CGCM experiments hampers direct comparisons between the two[5]. This problem can be ameliorated by statistical removal of ENSO effects (Santer *et al.,* 2001; Hegerl and Wallace, 2002; Wigley *et al.,* 2005a)[6].

The bottom line is that AMIP-style experiments and CGCM runs are both useful tools for exploring the possible causes of differential warming. We note that even if these two experimental configurations employ the same atmospheric model and the same climate forcings, they can yield noticeably different simulations of changes in atmospheric temperature profiles. These differences arise for a variety of reasons, such as AGCM-versus-CGCM differences in sea-ice coverage, SST distributions, and cloud feedbacks, and hence in climate sensitivity (Sun and Hansen, 2003)[7].

Most models undergo some adjustment of poorly-known parameters which directly affect key physical processes, such as convection and rainfall. Parameters are varied within plausible ranges, which are generally derived from direct observations. The aim of this procedure is to reduce the size of systematic model errors and improve simulations of present-day climate. Adjustment of uncertain model parameters is *not* performed over the course of a 20CEN experiment.

Several groups are now beginning to explore model parameter space, and are investigating the possible impact of parameter uncertainties on simulations of mean present-day climate and future climate change by running "perturbed physics" ensembles (Allen, 1999; Forest *et al.,* 2002; Murphy *et al.,* 2004; Stainforth *et al.,* 2005). Such work will help to quantify one component of model uncertainty. Another component of model uncertainty arises from differences in the basic structure of models[8].

estimation problem. The aerosols injected into the stratosphere during a massive volcanic eruption are typically removed within 2-3 years (Sato *et al.,* 1993; Hansen *et al.,* 2002; Ammann *et al.,* 2003). Because the large thermal inertia of the oceans causes a lag in response to this forcing, the cooling effect of the aerosols on the troposphere and surface persists for much longer than 2-3 years (Santer *et al.,* 2001; Free and Angell, 2002; Wigley *et al.,* 2005a). In the real world and in "AMIP-style" experiments, this slow, volcanically induced cooling of the troposphere and surface is sometimes masked by the warming effects of El Niño events (Christy and McNider, 1994; Wigley, 2000; Santer *et al.,* 2001), thus hampering volcanic signal estimation.

3 There are a variety of different spin-up strategies.

4 In most of the experiments reported on here, *n* is between 3 and 5.

5 If *n* is large enough to adequately sample the (simulated) effects of natural variability on surface and tropospheric temperatures, it is not necessarily a disadvantage that the simulated and observed variability does not line up in time. In fact, this type of experimental set-up allows one to determine whether the single realization of the observations is contained within the "envelope" of possible climate solutions that the CGCM simulates.

6 Residual effects of these modes of variability will remain in the data.

7 See, for example, the Ocean A and Ocean E results in Figure 3 of Sun and Hansen (2003).

8 The computer models constructed by different research groups can have quite different "structures" in terms of their horizontal and vertical resolution, atmospheric dynamics (so-called "dynamical cores"), numerical implementation (*e.g.,* spectral versus grid-point), and physical parameterizations. They do, however, share many common assumptions.

Box 5.2: Uncertainties in Simulated Temperature Changes

In discussing the major sources of uncertainty in observational estimates of temperature change, Chapter 2 partitioned uncertainties into three distinct categories: "structural," "parametric," and "statistical." Uncertainties in simulated temperature changes fall into similar categories. In the modeling context, "structural" uncertainties can be thought of as the uncertainties resulting from the choice of a particular climate model, model configuration (Section 2), or forcing data set (Section 3).

Within a given model, there are small-scale physical processes (such as convection, cloud formation, precipitation, *etc.*) that cannot be simulated explicitly. Instead, so-called "parameterizations" represent the large-scale effects of these unresolved processes. Each of these processes has uncertainties in the values of one or more key parameters.[a] Varying these parameters within plausible ranges introduces "parametric" uncertainty in climate change simulations (Allen, 1999; Forest *et al.,* 2002; Murphy *et al.,* 2004). Finally (analogous to the observational case), there is statistical uncertainty that arises from the unpredictable "noise" of internal climate variability, from the choice of a particular statistical metric to describe climate change, or from the application of a selected metric to noisy data.

a. Note that some of these parameters influence not only the climate response, but also the portrayal of the forcing itself. Examples include parameters related to the size of sulfate aerosols, and how aerosol particles scatter incoming sunlight.

Section 5 considers results from a range of state-of-the-art CGCMs, and thus samples some of the "structural uncertainty" in model simulations of 20th century climate change (Table 5.1). A further component of the spread in simulations of 20th century climate is introduced by uncertainties in the climate forcings with which models are run (Table 5.2). These are discussed in the following Section.

3. FORCINGS IN SIMULATIONS OF RECENT CLIMATE CHANGE

In an ideal world, there would be reliable quantitative estimates of all climate forcings – both natural and human-induced – that have made significant contributions to surface and tropospheric temperature changes. We would have detailed knowledge of how these forcings had changed over space and time. Finally, we would have used standard sets of forcings to perform climate-change experiments with a whole suite of numerical models, thus isolating uncertainties arising from structural differences in the models themselves (see Box 5.2).

Unfortunately, this ideal situation does not exist. As part of the IPCC Third Assessment Report, Ramaswamy *et al.* (2001b) assigned subjective confidence levels to our current "level of scientific understanding" (LOSU) of the changes in a dozen different climate forcings. Only in the case of well-mixed greenhouse gases ("GHGs"; carbon dioxide [CO_2], methane, nitrous oxide, and halocarbons) was the LOSU characterized as "high." The LOSU of changes in stratospheric and tropospheric ozone was judged to be "medium." For all other forcings (various aerosols, mineral dust, land use-induced albedo changes, solar, *etc.*), the LOSU was estimated to be "low" or "very low" (see Chapter 1, Table 1.1 and Section 1.2)[9].

In selecting the forcings for simulating the climate of the 20th century, there are at least three strategies that modeling groups can adopt. The first strategy is to incorporate only those forcings whose changes and effects are thought to be better understood, and for which time- and space-resolved data sets suitable for performing 20CEN experiments are readily available. The second strategy is to include a large number of different forcings, even those for which the LOSU is "very low." A third strategy is to vary the size of poorly known 20CEN forcings. This yields a range of simulated climate responses, which are then used to estimate the levels of the forcings that are consistent with observations (*e.g.,* Forest *et al.,* 2002).

The pragmatic focus of Chapter 5 is on climate forcings that have been incorporated in many CGCM simulations of 20th century climate. The primary forcings that we consider are changes in well-mixed GHGs, the direct effects of sulfate aerosol particles, tropospheric and stratospheric ozone, volcanic aerosols, and solar

[9] We note that there is no direct relationship between the LOSU of a given forcing and the contribution of that forcing to 20th century climate change. Forcings with "low" or "very low" LOSU may have had significant climatic impacts at regional and even global scales.

Box 5.3: Example of a Spatially-Heterogeneous Forcing: Black Carbon Aerosols

Carbon-containing aerosols (also known as "carbonaceous" aerosols) exist in a variety of chemical forms (Penner *et al.,* 2001). Two main classes of carbonaceous aerosol are generally distinguished: "black carbon" (BC) and "organic carbon" (OC). Both types of aerosol are emitted during fossil fuel and biomass burning. Most previous modeling work has focused on BC aerosols rather than OC aerosols. Some of the new model experiments described in Section 5 have now incorporated both types of aerosol in CGCM simulations of 20th century climate changes (see Tables 5.2 and 5.3).

Black carbon aerosols absorb sunlight and augment the GHG-induced warming of the troposphere (Hansen *et al.,* 2000; Satheesh and Ramanathan, 2000; Penner *et al.,* 2001; Hansen, 2002; Penner *et al.,* 2003)[a]. Their effects on atmospheric temperature profiles are complex, and depend on such factors as the chemical composition, particle size, and height distribution of the aerosols (*e.g.,* Penner *et al.,* 2003).

Menon *et al.* (2002) showed that the inclusion of fossil fuel and biomass aerosols over China and India[b] directly affected simulated vertical temperature profiles by heating the lower troposphere and cooling the surface. In turn, this change in atmospheric heating influenced regional circulation patterns and the hydrological cycle. Krishnan and Ramanathan (2002) found that an increase in black carbon aerosols has reduced the surface solar insolation (exposure to sunlight) over the Indian subcontinent. Model experiments performed by Penner *et al.* (2003) suggest that the net effect of carbonaceous aerosols on global-scale surface temperature changes depends critically on how aerosols affect the vertical distribution of clouds. On regional scales, the surface temperature effects of these aerosols are complex, and vary in sign (Penner *et al.,* 2006).

[a] Note that soot particles are sometimes transported long distances by winds, and can also have a "far field" effect on climate by reducing the reflectivity of snow in areas remote from pollution sources (Hansen and Nazarenko, 2003; Jacobson, 2004).

[b] During winter and spring, black carbon aerosols contribute to a persistent haze over large areas of Southern Asian and the Northern Indian Ocean (Ramanathan *et al.,* 2001).

irradiance. These are forcings whose effects on surface and atmospheric temperatures have been quantified in rigorous fingerprint studies (see Section 4.4). This does not diminish the importance of other climate forcings, whose global-scale contributions to "differential warming" have not been reliably quantified to date.

Examples of these "other forcings" include carbon-containing aerosols produced during fossil fuel or biomass combustion, human-induced changes in land surface properties, and the indirect effects of tropospheric aerosols on cloud properties. There is emerging scientific evidence that such spatially variable forcings may have had important impacts on regional and even on global climate (NRC, 2005). Some of this evidence is summarized in Box 5.3 and Box 5.4 for the specific cases of carbonaceous aerosols and land use change. These and other previously neglected forcings have been included in many of the new CGCM simulations of 20th century climate described in Section 5 (see Tables 5.1 and 5.2).

Clearly, we will never have complete and reliable information on all forcings that are thought to have influenced climate over the late 20th century. A key question is whether those forcings most important for understanding the differential warming problem are reliably represented. This is currently difficult to answer. What we can say, with some certainty, is that the expected atmospheric temperature signal due to forcing by well-mixed GHGs alone is distinctly different from the signal due to the combined effects of multiple natural and human forcing factors (Chapter 1; Santer *et al.,* 1996a; Tett *et al.,* 1996; Hansen *et al.,* 1997, 2002; Bengtsson *et al.,* 1999; Santer *et al.,* 2003a).

This is illustrated by the 20CEN and "single forcing" experiments performed with the Parallel Climate Model (PCM; Washington *et al.,* 2000). In PCM, changes in the vertical profile of atmospheric temperature over 1979 to 1999 are primarily forced by changes in well-mixed GHGs, ozone, and volcanic aerosols (Figure 5.1). Changes in solar irradiance and the scattering effects of sulfate aerosols are of secondary importance over this period. Even without performing formal statistical tests, it is visually obvious from Figure 5.1 that radiosonde-based estimates of observed stratospheric and tropospheric temperature changes are in better agreement with the PCM 20CEN experiment than with the PCM "GHG only" run.

This illustrates the need for caution in comparisons of modeled and observed atmospheric temperature change. The differences evident in such comparisons have multiple interpretations. They may be due to real errors in the

Box 5.4: Example of a Spatially-Heterogeneous Forcing: Land Use Change

Humans have transformed the surface of the planet through such activities as conversion of forest to cropland, urbanization, irrigation, and large water diversion projects (see Chapter 4). These changes can affect a variety of physical properties of the land surface, such as the albedo (reflectivity), the release of water by plants (transpiration), the moisture-holding capacity of soil, and the surface "roughness." Alterations in these physical properties may in turn affect runoff, heat and moisture exchanges between the land surface and atmospheric boundary layer, wind patterns, and even rainfall *(e.g.,* Pitman *et al.,* 2004). Depending on the nature of the change, either warming or cooling of the land surface may occur (Myhre and Myhre, 2003).

At the regional level, modeling studies of the Florida peninsula (Marshall *et al.,* 2004) and southwest Western Australia (Pitman *et al.,* 2004) have linked regional-scale changes in atmospheric circulation and rainfall to human transformation of the natural vegetation. Modeling work focusing on North America suggests that the conversion of natural forest and grassland to agricultural production has led to a cooling in summertime (Oleson *et al.,* 2004). The global-scale signal of land use/land cover (LULC) changes from pre-industrial times to the present is estimated to be a small net cooling of surface temperature (Matthews *et al.,* 2003, 2004; Brovkin *et al.,* 2004; Hansen *et al.,* 2005a; Feddema *et al.,* 2005). Larger regional trends of either sign are likely to be evident *(e.g.,* Hansen *et al.,* 2005a)[a].

a. Larger regional trends do not necessarily translate to enhanced detectability. Although the signals of LULC and other spatially-heterogeneous forcings are likely to be larger regionally than globally, the "noise" of natural climate variability is also larger at smaller spatial scales. It is not obvious *a priori,* therefore, how signal-to-noise relationships (and detectability of a given forcing's climate effects) behave as one moves from global to continental to regional scales.

models,[10] errors in the forcings used to drive the models, the neglect of important forcings, and residual inhomogeneities in the observations themselves. They may also be due to different manifestations of natural variability noise in the observations and a given CGCM realization. All of these factors can be important in model evaluation work.

4. PUBLISHED COMPARISONS OF MODELED AND OBSERVED TEMPERATURE CHANGES

A number of observational and modeling studies have attempted to shed light on the possible causes of "differential warming"[11]. We have attempted to organize the discussion of results so that investigations with similar analysis methods are grouped together[12]. Our discussion proceeds from simple to more complex and statistically rigorous analyses.

4.1 Regression Studies Using Observed Global-mean Temperature Data

One class of study that has attempted to address the causes of recent tropospheric temperature change relies on global-mean observational data only (Jones, 1994; Christy and McNider, 1994; Michaels and Knappenberger, 2000; Douglass and Clader, 2002). Such work uses a multiple regression model to quantify the statistical relationships between various "predictor variables" (typically time series of ENSO variability,

[10] These may lie in the physics, parameterizations, inadequate horizontal or vertical resolution, *etc.*

[11] We do not discuss studies which provide empirical estimates of "equilibrium climate sensitivity" – the steady-state warming of the Earth's surface that would eventually be reached after the climate system equilibrated to a doubling of pre-industrial CO_2 levels. This is often referred to as ΔT_{2xCO_2}. Estimates of ΔT_{2xCO_2} have been obtained by studying Earth's temperature response to "fast," "intermediate," and "slow" forcing of the climate system. Examples include the "fast" (<10-year) response of surface and tropospheric temperatures to massive volcanic eruptions (Hansen *et al.,* 1993; Lindzen and Giannitsis, 1998; Douglass and Knox, 2005; Wigley *et al.,* 2005a,b; Robock, 2005); the "intermediate" (100- to 150-year) response of surface temperatures to natural and human-caused forcing changes over the 19th and 20th centuries (Andronova and Schlesinger, 2001; Forest *et al.,* 2002; Gregory *et al.,* 2002; Harvey and Kaufmann, 2002) or to solar and volcanic forcing changes over the past 1-2 millennia (Crowley, 2000), and the "slow" (100,000-year) response of Earth's temperature to orbital changes between glacial and interglacial conditions (Hoffert and Covey, 1992; Hansen *et al.,* 1993). These investigations are not directly relevant to elucidation of the causes of changes in the vertical structure of atmospheric temperatures, which is the focus of this Chapter.

[12] It is useful to mention one technical issue relevant to model-data comparisons. As noted in Chapter 2, the satellite-based Microwave Sounding Unit (MSU) monitors the temperature of very broad atmospheric layers. To facilitate comparisons with observed MSU data sets, many of the studies reported on here calculate "synthetic" MSU temperatures from climate model experiments. Technical aspects of these calculations are discussed in Chapter 2, Box 2.1.

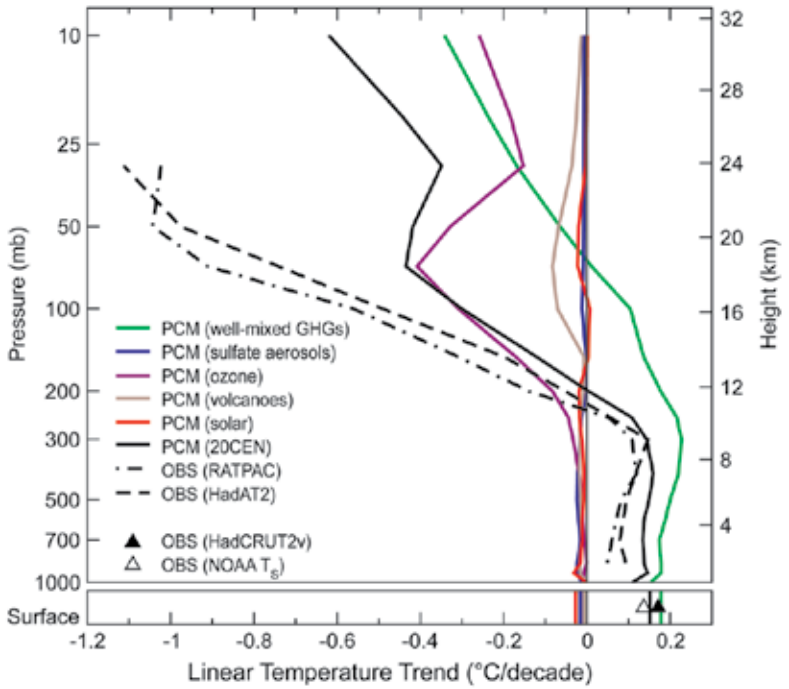

Figure 5.1: Vertical profiles of global-mean atmospheric temperature change over 1979 to 1999. Surface temperature changes are also shown. Results are from two different radiosonde data sets (HadAT2 and RATPAC; see Chapter 3) and from single forcing and combined forcing experiments performed with the Parallel Climate Model (PCM; Washington *et al.,* 2000). PCM results for each forcing experiment are averages over four different realizations of that experiment. All trends were calculated with monthly mean anomaly data.

volcanic aerosol loadings, and solar irradiance) and a single "predictand" (typically T_{2LT}). The aim is to remove the effects of the selected predictors on tropospheric temperature, and to estimate the residual trend that may arise from human-induced forcings. The quoted values for this residual trend in T_{2LT} range from 0.04 to 0.09ºC/decade[13].

These studies often make the unrealistic assumption that the uncertainties inherent in such statistical signal separation exercises are very small. They do not explore the sensitivity of regression results to uncertainties in the predictor variables or the observational record, and generally use solar and volcanic forcings as predictors rather than the climate responses to those forcings. Distinctions between forcing and response are important (Wigley *et al.,* 2005a). Accounting for uncertainties in predictor variables (and use of responses rather than forcings as predictors) expands the range of uncertainties in estimates of residual T_{2LT} trends (Santer *et al.,* 2001)[14].

Regression methods have also been used to estimate the net effects of ENSO and volcanoes on trends in global-mean surface and tropospheric temperatures. For T_{2LT}, both Jones (1994) and Christy and McNider (1994) found that ENSO effects induced a small net warming of 0.03 to 0.05ºC/decade over 1979 to 1993, while volcanoes caused a cooling of 0.18ºC/decade over the same period. Michaels and Knappenberger (2000) also reported a relatively small ENSO influence on T_{2LT} trends[15]. Santer *et al.* (2001) noted that over 1979 to 1997, volcanoes had likely cooled the troposphere by more than the surface. Removing the combined volcano and ENSO effects from surface and UAH T_{2LT} data helped to explain some of the observed differential warming: the "raw" T_S-minus-T_{2LT} trend over 1979 to 1997 decreased from roughly 0.15ºC/decade to 0.05-0.13ºC/decade.[16] Removal of volcano and ENSO influences also brought observed lapse rate trends closer to model results, but could not fully reconcile modeled and observed lapse rate trends[17].

[13] The studies by Jones (1994) and Christy and McNider (1994) remove volcano and ENSO effects from T_{2LT}, and estimate residual trends of 0.093 and 0.090ºC/decade over 1979 to 1993. A similar investigation by Michaels and Knappenberger (2000) obtained a residual trend of 0.041ºC/decade over 1979 to 1999. The error bars on these residual trend estimates are either not given, or claimed to be very small (*e.g.,* ± 0.005ºC/decade in Christy and McNider). A fourth study removed combined ENSO, volcano, and solar effects from T_{2LT}, and estimated a residual trend of 0.065 ± 0.012ºC/decade over 1979 to 2000 (Douglass and Clader, 2002).

[14] Santer *et al.* (2001) obtain residual T_{2LT} trends ranging from 0.06 to 0.16ºC/decade over 1979 to 1999. Their regression model is iterative, and involves removal of ENSO and volcano effects only.

[15] The ENSO components of their T_{2LT} trends were 0.04ºC/decade over 1979 to 1998 and 0.01ºC/decade over 1979 to 1999. This difference in the net ENSO influence on T_{2LT} (with the addition of only a single year of record) arises from the El Niño event in 1997/98, and illustrates the sensitivity of this kind of analysis to so-called "end effects."

[16] The latter results were obtained with the HadCRUTv surface data (Jones *et al.,* 2001) and version d03 of the UAH T_{2LT} data. The range of residual lapse-rate trends arises from parametric uncertainty, *i.e.,* from the different choices of ENSO predictor variables and volcano parameters.

[17] Santer *et al.* (2001) analyzed model experiments performed with the ECHAM4/OPYC model developed at the Max-Planck Institute for Meteorology in Hamburg (Roeckner *et al.,* 1999). The experiments included forcing by well-mixed greenhouse gases, direct and indirect sulfate aerosol effects, tropospheric and stratospheric ozone, and volcanic aerosols (Pinatubo only).

4.2 Regression Studies Using Spatially Resolved Temperature Data

Other regression studies have attempted to remove natural variability influences using spatially resolved temperature data. Regression is performed "locally" at individual grid-points and/or atmospheric levels. To obtain a clearer picture of volcanic effects on atmospheric temperatures, Free and Angell (2002) removed the effects of variability in ENSO and the Quasi-Biennial Oscillation (QBO) from Hadley Centre radiosonde data[18]. Their work clearly shows that the cooling effect of massive volcanic eruptions has been larger in the upper troposphere than in the lower troposphere. The implication is that volcanic effects probably contribute to slow changes in observed lapse rates.

Hegerl and Wallace (2002) used regression methods to identify and remove different components of natural climate variability from gridded fields of surface temperature data, UAH T_{2LT}, and "synthetic" T_{2LT} calculated from radiosonde data. They focused on the variability associated with ENSO and the so-called "cold ocean warm land" (COWL) pattern (Wallace *et al.,* 1995). While ENSO and COWL variability made significant contributions to the month-to-month and year-to-year variability of temperature differences between the surface and T_{2LT}, their analysis indicated that it had very little impact on decadal fluctuations in lapse rate. The authors concluded that natural variability alone was unlikely to explain these slow lapse-rate changes. However, the removal of ENSO and COWL effects more clearly revealed a volcanic contribution, consistent with the findings of Santer *et al.* (2001) and Free and Angell (2002). A climate model control run (with no changes in forcings) and a 20CEN experiment were unable to adequately reproduce the observed decadal changes in lapse rate[19].

[18] The HadRT2.1 data set of Parker *et al.* (1997). Like Santer *et al.* (2001), Free and Angell (2002) also found some sensitivity of the estimated volcanic signals to "parametric" uncertainty.

[19] The model was the ECHAM4/OPYC CGCM used by Bengtsson *et al.* (1999). The 20CEN experiment analyzed by Hegerl and Wallace (2002) involved combined changes in well-mixed greenhouse gases, the direct and indirect effects of sulfate aerosols, and tropospheric ozone. Forcing by volcanoes and stratospheric ozone depletion was not included.

4.3 Other Studies of Global and Tropical Lapse-rate Trends

Several studies have investigated lapse-rate trends without attempting to remove volcano effects or natural climate noise. Brown *et al.* (2000) used surface, radiosonde, and satellite data to identify slow, tropic-wide changes in the lower tropospheric lapse rate[20]. In their analysis, the surface warmed relative to the troposphere between the early 1960s and mid-1970s and after the early 1990s. Between these two periods, the tropical troposphere warmed relative to the surface. The spatial coherence of these variations (and independent evidence of concurrent variations in the tropical general circulation) led Brown *et al.* (2000) to conclude that tropical lapse rate changes were unlikely to be an artifact of residual errors in the observations.

The cooling effect of massive volcanic eruptions has been larger in the upper troposphere than in the lower troposphere.

Very similar decadal changes in lower tropospheric lapse rate were reported by Gaffen *et al.* (2000)[21]. Their study analyzed radiosonde-derived temperature and lapse rate changes over two periods: 1960 to 1997 and 1979 to 1997. Tropical lapse rates decreased over the longer period[22] and increased over the satellite era[23]. To evaluate whether natural climate variability could explain these slow variations, Gaffen *et al.* (2000) computed lapse rates from the control runs performed with three different CGCMs. Each control run was 300 years in length. These long runs provided estimates of the "sampling variability" of modeled lapse rate changes on timescales relevant to the two observational periods (38 and 19 years)[24]. Model-based esti-

[20] The Brown *et al.* (2000) study employed UKMO surface data (HadCRUT), version d of the UAH T_{2LT}, and an early version of the Hadley Centre radiosonde data set (HadRT2.0) that was uncorrected for instrumental biases.

[21] Gaffen *et al.* (2000) used a different radiosonde data set from that employed by Brown *et al.* (2000). The two groups also analyzed different surface temperature data sets.

[22] Corresponding to a tendency towards a more stable atmosphere.

[23] These lapse-rate changes were accompanied by increases and decreases in tropical freezing heights (which were inferred from the same radiosonde data).

[24] Each control run was used to generate distributions of 38-year and 19-year lapse rate trends. For example, a 300-year control run can be split up into 15 different "segments" that are each of length 19 years (assuming there is no overlap between segments). From these segments, one obtains 15 different estimates of how the lapse rate might vary in the absence of any forcing

mates of natural climate variability could not adequately explain the observed tropical lapse rate changes over 1979 to 1997. Similar conclusions were reached by Hansen *et al.* (1995) and Santer *et al.* (2000). Including natural and anthropogenic forcings in the latter study narrowed the gap between modeled and observed estimates of recent lapse-rate changes, although a significant discrepancy between the two still remained.

Different climate forcings have different characteristic patterns of temperature response ("fingerprints").

It should be emphasized that all of the studies reported on to date in Section 4 relied on satellite data from one group only (UAH), on early versions of the radiosonde data[25], and on experiments performed with earlier model "vintages." It is likely, therefore, that this work may have underestimated the structural uncertainties in observed and simulated estimates of lapse rate changes. We will consider in Section 5 whether modeled and observed lapse rate changes can be better reconciled by the availability of more recent 20CEN runs and more comprehensive estimates of structural uncertainties in observations.

4.4 Pattern-based "Fingerprint" Detection Studies

Fingerprint detection studies rely on patterns of temperature change (Box 5.5). The patterns are typically either latitude-longitude "maps" (*e.g.,* for T_4, T_2, T_S, *etc.*) or latitude-height cross-sections through the atmosphere[26]. The basic premise in fingerprinting is that different climate forcings have different characteristic patterns of temperature response ("fingerprints"), particularly in the free atmosphere (Chapter 1, Figure 1.3; Hansen *et al.,* 1997, 2002, 2005a; Bengtsson *et al.,* 1999; Santer *et al.,* 1996a; Tett *et al.,* 1996).

Most analysts rely on a climate model to provide physically based estimates of each fingerprint's structure, size, and evolution. The model fingerprints are searched for in observational climate records, using rigorous statistical methods to quantify the degree of correspondence with observed patterns of climate change[27]. Fingerprints are also compared with patterns of climate change in model control runs. This helps to determine whether the correspondence between the fingerprint and observations is truly significant, or could arise through internal variability alone (Box 5.5). Model errors in internal variability[28] can bias detection results, although most detection work tries to guard against this possibility by performing "consistency checks" on modeled and observed variability (Allen and Tett, 1999), and by using variability estimates from multiple models (Hegerl *et al.,* 1997; Santer *et al.,* 2003a,b).

The application of fingerprint methods involves a variety of decisions, which introduce uncertainty in detection results (Box 5.5). Our confidence in fingerprint detection results is increased if they are shown to be consistent across a range of plausible choices of statistical methods, processing options, and model and observational data sets.

Surface Temperature Changes

Most fingerprint detection studies have focused on surface temperature changes. The common denominator in this work is that the model fingerprints resulting from forcing by well-mixed GHGs and sulfate aerosols[29] are statistically identifiable in observed surface temperature records (Hegerl *et al.,* 1996, 1997; North and Stevens, 1998; Tett *et al.,* 1999, 2002; Stott *et*

changes. The observed lapse rate change over 1979 to 1997 is then compared with the model trend distribution to determine whether the observed result could be explained by natural variability alone.

25 These radiosonde data sets were either unadjusted for inhomogeneities, or had not been subjected to the rigorous adjustment procedures used in more recent work (Lanzante *et al.,* 2003; Thorne *et al.,* 2005).

26 In constructing these cross-sections, the temperature changes are generally averaged along individual bands of latitude. Zonal averages are then displayed at individual pressure levels, starting at the lowest model or radiosonde level and ending at the top of the model atmosphere or highest reported radiosonde level (see, *e.g.,* Chapter 1, Figure 3).

27 The fingerprint can be either the response to an individual forcing or a combination of forcings. One strategy, for example, is to search for the climate fingerprint in response to combined changes in a suite of different human-caused forcings.

28 For example, current CGCMs fail to simulate the stratospheric temperature variability associated with the QBO or with solar-induced changes in stratospheric ozone (Haigh, 1994). Such errors may help to explain why one particular CGCM underestimated observed temperature variability in the equatorial stratosphere (Gillett *et al.,* 2000). In the same model, however, the variability of temperatures and lapse rates in the tropical troposphere was in reasonable agreement with observations.

29 Most of this work considers only the direct scattering effects of sulfate aerosols on incoming sunlight, and not indirect aerosol effects on clouds.

Box 5.5: Fingerprint Studies

Detection and attribution ("D&A") studies attempt to represent an observed climate data set as a linear combination of the climate signals ("fingerprints") arising from different forcing factors and the noise of natural internal climate variability (Section 4.4). A number of different fingerprint methods have been applied to the problem of identifying human-induced climate change. Initial studies used relatively simple pattern correlation methods (Barnett and Schlesinger, 1987; Santer *et al.,* 1996a,b; Tett *et al.,* 1996). Later work involved variants of the "optimal detection" approach suggested by Hasselmann (1979, 1993, 1997)[a]. These are essentially regression-based techniques that seek to estimate the strength of a given fingerprint pattern in observational data *(i.e.*, how much a given fingerprint pattern has to be scaled up or down in order to best match observations). For example, if the regression coefficient for a GHG-induced T_S fingerprint is significantly different from zero, GHG effects are deemed to be "detected" in observed surface temperature records. Attribution tests address the question of whether these regression coefficients are also consistent with unity – in other words, whether the size of the model fingerprint is consistent with its amplitude in observations (*e.g.,* Allen and Tett, 1999; Mitchell *et al.,* 2001).

There are two broad classes of regression-based D&A methods (Mitchell *et al.,* 2001). One class assumes that although the fingerprint's amplitude changes over time, its spatial pattern does not (Hegerl *et al.,* 1996, 1997; Santer *et al.,* 2003a,b, 2004). The second class explicitly considers both the spatial structure and time evolution of the fingerprint (Allen and Tett, 1999; Allen *et al.,* 2006; Stott and Tett, 1998; Stott *et al.,* 2000; Tett *et al.,* 1999, 2002; Barnett *et al.,* 2001, 2005). This is particularly useful if the time evolution of the fingerprint contains specific information (such as a periodic 11-year solar cycle) that may help to distinguish it from natural internal climate variability (North *et al.,* 1995; North and Stevens, 1998).

A number of choices must be made in applying D&A methods to real-world problems. One of the most important decisions relates to "reduction of dimensionality". D&A methods require some knowledge of the correlation structure of natural climate variability[b]. This structure is difficult to estimate reliably, even from long model control runs, because the number of time samples available to estimate correlation behavior is typically much smaller than the number of spatial points in the field. In practice, the total amount of spatial information (the "dimensionality") must be reduced. This is often done by using a mathematical tool (Empirical Orthogonal Functions) to reduce a complex space-time data set to a very small number of spatial patterns ("EOFs") that capture most of the information content of the data set[c]. Different analysts use different procedures to determine the number of patterns to retain. Further decisions relate to the choice of data used for estimating fingerprint and noise, the number of fingerprints considered, the selection of observational data, the treatment of missing data, *etc.*[d].

D&A methods have some limitations. They do not work well if fingerprints are highly uncertain, or if the fingerprints arising from two different forcings are similar[e]. They make at least two important assumptions: that model-based estimates of natural climate variability are a reliable representation of "real-world" variability, and that the sum of climate responses to individual forcing mechanisms is equivalent to the response obtained when these factors are varied in concert. Testing the validity of both assumptions remains an important research activity (Allen and Tett, 1999; Santer *et al.,* 2003a; Gillett *et al.,* 2004a).

a. Hasselmann (1979) noted that the engineering field had extensive familiarity with the problem of identifying coherent signals embedded in noisy data, and that many of the techniques routinely used in signal processing were transferable to the problem of detecting a human-induced climate change signal.
b. The relationship between variability at different points in a spatial field.
c. The number of patterns retained is often referred to as the "truncation dimension." How the truncation dimension should be determined is a key decision in optimal detection studies (Hegerl *et al.,* 1996; Allen and Tett, 1999).
d. Another important choice determines whether global-mean changes are included or removed from the detection analysis. Removal of global means focuses attention on smaller-scale features of modeled and observed climate-change patterns, and provides a more stringent test of model performance.
e. This problem is known as "degeneracy." Formal tests of fingerprint degeneracy are sometimes applied *(e.g.,* Tett *et al.,* 2002).

al., 2000). These results are robust to a wide range of uncertainties (Allen *et al.,* 2006)[30]. In summarizing this body of work, the IPCC concluded that "There is new and stronger evidence that most of the warming observed over the last 50 years is attributable to human activities" (Houghton *et al.,* 2001, page 4). The causes of surface temperature change over the first half of the 20th century are more ambiguous (IDAG, 2005).

[30] For example, to uncertainties in the applied greenhouse-gas and sulfate aerosol forcings, the model responses to those forcings, and model-based estimates of natural internal climate variability.

Most of the early fingerprint detection work dealt with global-scale patterns of surface temperature change. The positive detection results obtained for "GHG-only" fingerprints were driven by model-data pattern similarities at very large spatial scales (*e.g.,* at the scale of individual hemispheres, or land-versus-ocean behavior). Fingerprint detection of GHG effects becomes more challenging at continental or sub-continental scales[31]. It is at these smaller scales that spatially heterogeneous forcings, such as those arising from changes in aerosol loadings and land use patterns, may have large impacts on regional climate (see Box 5.3 and 5.4). This is illustrated by the work of Stott and Tett (1998), who found that a combined GHG and sulfate aerosol signal was identifiable at smaller spatial scales than a "GHG-only" signal.

Preliminary investigations raise the intriguing possibility of formal detection of anthropogenic effects at regional scales that are of direct relevance to policymakers.

Recently, Stott (2003) and Zwiers and Zhang (2003) have reported positive identification of the continental- or even sub-continental features of combined GHG and sulfate aerosol fingerprints in observed surface temperature records.[32] Using a variant of "classical" fingerprint methods,[33] Min *et al.* (2005) identified a GHG signal in observed records of surface temperature change over East Asia. Karoly and Wu (2005) suggest that GHG and sulfate aerosol effects are identifiable at even smaller spatial scales ("of order 500 km in many regions of the globe"). These preliminary investigations raise the intriguing possibility of formal detection of anthropogenic effects at regional scales that are of direct relevance to policymakers.

Changes in Latitude/Longitude Patterns of Atmospheric Temperature or Lapse Rate

Fingerprint methods have also been applied to spatial "maps" of changes in layer-averaged atmospheric temperatures (Santer *et al.,* 2003b; Thorne *et al.,* 2003) and lapse rate (Thorne *et al.,* 2003). The study by Santer *et al.* (2003b) compared modeled and observed changes in T_2 and T_4. Model fingerprints were estimated from 20CEN experiments performed with PCM (see Table 5.1), while observations were taken from two different satellite data sets (UAH and RSS; see Christy *et al.,* 2003, and Mears *et al.,* 2003). The aim of this work was to assess the sensitivity of detection results to structural uncertainties in observed MSU data.

For the T_4 layer, the model fingerprint of combined human and natural effects was consistently detectable in both satellite data sets. In contrast, PCM's T_2 fingerprint was identifiable in RSS data (which show net warming over the satellite era), but not in UAH data (which show little overall change in T_2; see Chapter 3). Encouragingly, once the global-mean differences between RSS and UAH data were removed, the PCM T_2 fingerprint was detectable in both observed data sets. This suggests that the structural uncertainties in RSS and UAH T_2 data are most prominent at the global-mean level, and that this global-mean difference masks underlying similarities in smaller-scale pattern structure (Chapter 4; Santer *et al.,* 2004).

Thorne *et al.* (2003) applied a "space-time" fingerprint method to six individual climate variables. These variables contained information on patterns[34] of temperature change at the surface, in broad atmospheric layers (the upper and lower troposphere), and in the lapse rates between these layers[35]. Thorne *et al.* explicitly considered uncertainties in the searched-for fingerprints, the observed radiosonde data[36], and in various data processing/fingerprinting options. They also assessed the detectability of fingerprints arising from multiple forcings[37].

[31] This is partly due to the fact that natural climate noise is larger (and models are less skillful) on smaller spatial scales.

[32] Another relevant "sub-global" detection study is that by Karoly *et al.* (2003). This showed that observed trends in a variety of area-averaged "indices" of North American climate (*e.g.,* surface temperature, daily temperature range, and the amplitude of the seasonal cycle) were consistent with model-predicted trends in response to anthropogenic forcing, but were inconsistent with model estimates of natural climate variability.

[33] Involving Bayesian statistics.

[34] The "patterns" are in the form of temperature averages calculated over large areas rather than temperatures on a regular latitude/longitude grid.

[35] Thorne *et al.* (2003) calculated the lapse rate changes between the surface and lower troposphere, the surface and upper troposphere, and the lower and upper troposphere.

[36] The model fingerprint was estimated from 20CEN runs performed with two different versions of the Hadley Centre CGCM (HadCM2 and HadCM3). Observational data were taken from two early compilations of the Hadley Centre radiosonde data (HadRT2.1 and HadRT2.1s).

[37] Well-mixed greenhouse gases, the direct effects of

The "bottom-line" conclusion of Thorne *et al.* is that two human-caused fingerprints – one arising from changes in well-mixed GHGs alone, and the other due to combined GHG and sulfate aerosol effects – were robustly identifiable in the observed surface, lower tropospheric, and upper tropospheric temperatures. Evidence for the existence of a detectable volcanic signal was more equivocal. Volcanic and human-caused fingerprints were not consistently identifiable in observed patterns of lapse rate change[38].

Changes in Latitude/Height Profiles of Atmospheric Temperature

Initial detection work with zonal-mean profiles of atmospheric temperature change used pattern correlations to compare model fingerprints with radiosonde data (Karoly *et al.,* 1994; Santer *et al.,* 1996a; Tett *et al.,* 1996; Folland *et al.,* 1998; Sexton *et al.,* 2001). These early investigations found that model fingerprints of the stratospheric cooling and tropospheric warming in response to increases in atmospheric CO_2 were identifiable in observations (Chapter 1, Figure 1.3A). The pattern similarity between modeled and observed changes generally increased over the period of the radiosonde record.

The inclusion of other human-induced forcings in 20CEN experiments – particularly the effects of stratospheric ozone depletion and sulfate aerosols – tended to improve agreement with observations (Santer *et al.,* 1996a; Tett *et al.,* 1996; Sexton *et al.,* 2001). The addition of ozone depletion cooled the lower stratosphere and upper troposphere. This brought the height of the "transition level" between stratospheric cooling and tropospheric warming lower down in the atmosphere, and in better accord with observations (Chapter 1, Figure 1.3F). It also improved the agreement between simulated and observed patterns of T_4 (Ramaswamy *et al.,* 1996), and decreased the size of the "warming maximum" in the upper tropical troposphere, a prominent feature of CO_2-only experiments (compare Figures 1.3A and 1.3F in Chapter 1).

Early work on the direct scattering effects of sulfate aerosols suggested that this forcing was generally stronger in the Northern Hemisphere (NH) than in the Southern Hemisphere (SH), due to the larger emissions of sulfur dioxide in industrialized regions of the NH. This asymmetry in the distribution of anthropogenic sulfur dioxide sources should yield greater aerosol-induced tropospheric cooling in the NH (Santer *et al.,* 1996a,b). Other forcings can lead to different hemispheric temperature responses. Increases in atmospheric CO_2, for example, tend to warm land more rapidly than ocean (Chapter 1). Since there is more land in the NH than in the SH, the expected signal due to CO_2 increases is greater warming in the NH than in the SH. Because the relative importance of CO_2 and sulfate aerosol forcings evolves in a complex way over time (Tett *et al.,* 2002; Hansen *et al.,* 2002),[39] the "imprints" of these two forcings on NH and SH temperatures must also vary with time (Santer *et al.,* 1996b; Stott *et al.,* 2006).

Initial attempts to detect sulfate aerosol effects on atmospheric temperatures did not account for such slow changes in the hemispheric-scale features of the aerosol fingerprint. They searched for a time-invariant fingerprint pattern in observed radiosonde data (Santer *et al.,* 1996a). This yielded periods of agreement and periods of disagreement between the (fixed) aerosol fingerprint and the time-varying effect of aerosols on atmospheric temperatures. Some have interpreted the periods of disagreement as "evidence of absence" of a sulfate aerosol signal (Michaels and Knappenberger, 1996). However, subsequent studies (see below) illustrate that such behavior is expected if one uses a fixed sulfate aerosol fingerprint, and that it is important for detection studies to account for large temporal changes in the fingerprint.

sulfate aerosols, combined greenhouse-gas and sulfate aerosol effects, volcanic aerosols, and solar irradiance changes.

[38] The failure to detect volcanic signals is probably due to the coarse time resolution of the input data (five-year averages) and the masking effects of ENSO variability in the radiosonde observations. Note that the two models employed in this work yielded different estimates of the size of the natural and human-caused fingerprints.

[39] See, for example, Figure 1a in Tett *et al.* (2002) and Figure 8b in Hansen *et al.* (2002).

Table 5.1: Acronyms of climate models referenced in this Chapter. All 19 models performed simulations of 20th century climate change ("20CEN") in support of the IPCC Fourth Assessment Report. The ensemble size "ES" is the number of independent realizations of the 20CEN experiment that were analyzed here.

	Model Acronym	Country	Institution	ES
1	CCCma-CGCM3.1(T47)	Canada	Canadian Centre for Climate Modelling and Analysis	1
2	CCSM3	United States	National Center for Atmospheric Research	5
3	CNRM-CM3	France	Météo-France/Centre National de Recherches Météorologiques	1
4	CSIRO-Mk3.0	Australia	CSIRO[a.] Marine and Atmospheric Research	1
5	ECHAM5/MPI-OM	Germany	Max-Planck Institute for Meteorology	3
6	FGOALS-g1.0	China	Institute for Atmospheric Physics	3
7	GFDL-CM2.0	United States	Geophysical Fluid Dynamics Laboratory	3
8	GFDL-CM2.1	United States	Geophysical Fluid Dynamics Laboratory	3
9	GISS-AOM	United States	Goddard Institute for Space Studies	2
10	GISS-EH	United States	Goddard Institute for Space Studies	5
11	GISS-ER	United States	Goddard Institute for Space Studies	5
12	INM-CM3.0	Russia	Institute for Numerical Mathematics	1
13	IPSL-CM4	France	Institute Pierre Simon Laplace	1
14	MIROC3.2(medres)	Japan	Center for Climate System Research / NIES[b.] / JAMSTEC[c.]	3
15	MIROC3.2(hires)	Japan	Center for Climate System Research / NIES[b.] / JAMSTEC[c.]	1
16	MRI-CGCM2.3.2	Japan	Meteorological Research Institute	5
17	PCM	United States	National Center for Atmospheric Research	4
18	UKMO-HadCM3	United Kingdom	Hadley Centre for Climate Prediction and Research	1
19	UKMO-HadGEM1	United Kingdom	Hadley Centre for Climate Prediction and Research	1

[a.] CSIRO is the Commonwealth Scientific and Industrial Research Organization.
[b.] NIES is the National Institute for Environmental Studies.
[c.] JAMSTEC is the Frontier Research Center for Global Change in Japan.

Fingerprint detection studies provide consistent evidence that human-induced changes in greenhouse gases and sulfate aerosols are identifiable in radiosonde records of free atmospheric temperature change.

"Space-time" optimal detection schemes explicitly account for time variations in the signal pattern and in observational data (Box 5.5). Results from recent space-time detection studies support previous claims of an identifiable sulfate aerosol effect on surface temperature (Stott *et al.,* 2006) and on zonal-mean profiles of atmospheric temperature (Allen and Tett, 1999; Forest *et al.,* 2001, 2002; Thorne *et al.,* 2002; Tett *et al.,* 2002; Jones *et al.,* 2003). This work also illustrates that the identification of human effects on atmospheric temperatures can be achieved using tropospheric temperatures alone (Thorne *et al.,* 2002). Positive detection results are not solely driven by the inclusion of strong stratospheric cooling in the vertical pattern of temperature change (as has been claimed by Weber, 1996).

In summary, fingerprint detection studies provide consistent evidence that human-induced changes in greenhouse gases and sulfate aerosols are identifiable in radiosonde records of free atmospheric temperature change. The fingerprint evidence is much more equivocal in the case of solar and volcanic signals in the troposphere. These natural signals have been detected in some studies (Jones *et al.,* 2003) but not in others (Tett *et al.,* 2002), and their identification appears to be more sensitive to specific processing choices that are made in applying fingerprint methods (Leroy, 1998; Thorne *et al.,* 2002, 2003).

5. NEW COMPARISONS OF MODELED AND OBSERVED TEMPERATURE CHANGES

In this section, we evaluate selected results from recently completed CGCM 20CEN ex-

periments that have been performed in support of the IPCC Fourth Assessment Report (AR4). The runs analyzed here were performed with 19 different models, and involve modeling groups in nine different countries (Table 5.1). They use new model versions, and incorporate historical changes in many (but not all) of the natural and human forcings that are thought to have influenced atmospheric temperatures over the past 50 years[40] (Table 5.2). These new experiments provide our current best estimates of the expected climate change due to combined human and natural effects.

The new 20CEN runs constitute an "ensemble of opportunity" (Allen and Stainforth, 2002). The selection and application of natural and anthropogenic forcings was not coordinated across modeling groups.[41] For example, only seven of the 19 models were run with time-varying changes in LULC (Table 5.2). Modeling groups that included LULC effects did not always use the same observational data set for specifying this forcing, or apply it in the same way (Table 5.3). Only six models included some representation of the indirect effects of anthropogenic aerosols, which are thought to have had a net cooling influence on surface temperatures through their effects on cloud properties (Ramaswamy *et al.,* 2001b).

One important implication of Tables 5.2 and 5.3 is that model-to-model differences in the applied forcings are intertwined with model-to-model differences in the climate responses to those forcings. This makes it more difficult to isolate systematic errors that are common to a number of models, or to identify problems with a specific forcing data set. Note, however, that the lack of a coordinated experimental design is also an advantage, since the "ensemble of opportunity" spans a wide range of uncertainty in current estimates of climate forcings.

In addition to model forcing and response uncertainty, the 20CEN ensemble also encompasses uncertainties arising from inherently unpredictable climate variability (Boxes 5.1, 5.2). Roughly half of the modeling groups that submitted 20CEN data performed multiple realizations of their historical forcing experiment (see Section 2 and Table 5.1). For example, the five-member ensemble of CCSM3.0 20CEN runs contains an underlying signal (which one might define as the ensemble-average climate response to the forcings varied in CCSM3.0) plus five different sequences of climate noise. Such multi-member ensembles provide valuable information on the relative sizes of signal and noise. In all, a total of 49 20CEN realizations were examined here[42].

The following Section presents preliminary results from analyses of these 20CEN runs and the new observational data sets described in Chapters 2-4. Our primary focus is on the tropics, since previous work by Gaffen *et al.* (2000) and Hegerl and Wallace (2002) suggests that this is where any differences between observations and models are most critical. We also discuss comparisons of global-mean changes in atmospheric temperatures and lapse rates. We do not discount the importance of comparing modeled and observed lapse-rate changes at much smaller scales (particularly in view of the incorporation of regional-scale forcing changes in many of the runs analyzed here), but no comprehensive regional-scale comparisons were available for us to assess.

In order to facilitate "like with like" comparisons between modeled and observed atmospheric temperature changes, we calculate synthetic MSU T_4, T_2, and T_{2LT} from the model 20CEN results (see Chapter 2, Box 1). Both observed and synthetic MSU T_2 data include a contribution from the cooling stratosphere (Fu *et al.,* 2004a,b), and hence complicate the interpretation of slow changes in T_2. To provide a less ambiguous measure of "bulk" tropospheric

[40] This was not the case in previous model intercomparison exercises, such as AMIP (Gates *et al.,* 1999) and CMIP2 (Meehl *et al.,* 2000).

[41] In practice, experimental coordination is very difficult across a range of models of varying complexity and sophistication. Aerosols are a case in point. Some modeling groups that contributed 20CEN simulations to the IPCC AR4 do not have the technical capability to explicitly include aerosols, and instead attempt to represent their net radiative effects by adjusting the surface albedo.

[42] 49 individual realizations of the IPCC 20CEN run were available at the time this Chapter was written. An analysis of lapse-rate changes in these realizations has been published (Santer *et al.,* 2005). At present, the IPCC database contains 82 realizations of the 20CEN experiment. Relevant analyses of these additional 33 realizations are currently unpublished and unreviewed, and have not been included here.

Table 5.2: Forcings used in IPCC simulations of 20th century climate change. This Table was compiled using information provided by the participating modeling centers (see http://www-pcmdi.llnl.gov/ipcc/model.documentation/ipcc_model_documentation.php). Eleven different forcings are listed: well-mixed greenhouse gases (G), tropospheric and stratospheric ozone (O), sulfate aerosol direct (SD) and indirect effects (SI), black carbon (BC) and organic carbon aerosols (OC), mineral dust (MD), sea salt (SS), land use/land cover (LU), solar irradiance (SO), and volcanic aerosols (V). Shading denotes inclusion of a specific forcing. As used here, "inclusion" means specification of a time-varying forcing, with changes on interannual and longer timescales. Forcings that were varied over the seasonal cycle only are not shaded.

	MODEL	G	O	SD	SI	BC	OC	MD	SS	LU	SO	V
1	CCCma-CGCM3.1(T47)											
2	CCSM3											
3	CNRM-CM3											
4	CSIRO-Mk3.0											
5	ECHAM5/MPI-OM											
6	FGOALS-g1.0											
7	GFDL-CM2.0											
8	GFDL-CM2.1											
9	GISS-AOM											
10	GISS-EH											
11	GISS-ER											
12	INM-CM3.0											
13	IPSL-CM4											
14	MIROC3.2(medres)											
15	MIROC3.2(hires)											
16	MRI-CGCM2.3.2											
17	PCM											
18	UKMO-HadCM3											
19	UKMO-HadGEM1											

temperature changes, we use the statistical approach of Fu *et al.* (2004a, 2005) to remove stratospheric influences, thereby obtaining T^*_G and T^*_T in addition to T_{2LT}[43]. As a simple measure of lapse-rate changes, we consider temperature differences between the surface and three different atmospheric layers (T_{2LT}, T^*_G, and T^*_T). Each of these layers samples slightly different portions of the troposphere (Chapter 2, Figure 2.2).

The trend comparisons shown in Sections 5.1 and 5.2 do not involve any formal statistical significance tests (see Appendix A). While such tests are entirely appropriate for comparisons of individual model and observational trends,[44] they are less relevant here, where we compare a 49-member ensemble of model trends with a relatively small number of observationally based estimates. The model ensemble encapsulates uncertainties in climate forcings and model responses, as well as the effects of climate noise on trends. The observational range characterizes current structural uncertainties in historical changes. We simply assess whether the observations are contained within the simu-

[43] There is still some debate over the reliability of T^*_G trends estimated with the Fu *et al.* (2004a) statistical approach (Tett and Thorne, 2004, Gillett *et al.,* 2004; Kiehl *et al.,* 2005; Fu *et al.,* 2004b; Chapter 4). T^*_T is derived mathematically (from the overlap between the T_4 and T_2 weighting functions) rather than statistically, and is now generally accepted as a reasonable measure of temperature change in the tropical troposphere.

[44] For example, such tests have been performed by Santer *et al.* (2003b) in comparisons between observed MSU trends (in RSS and UAH) and synthetic MSU trends in four PCM 20CEN realizations.

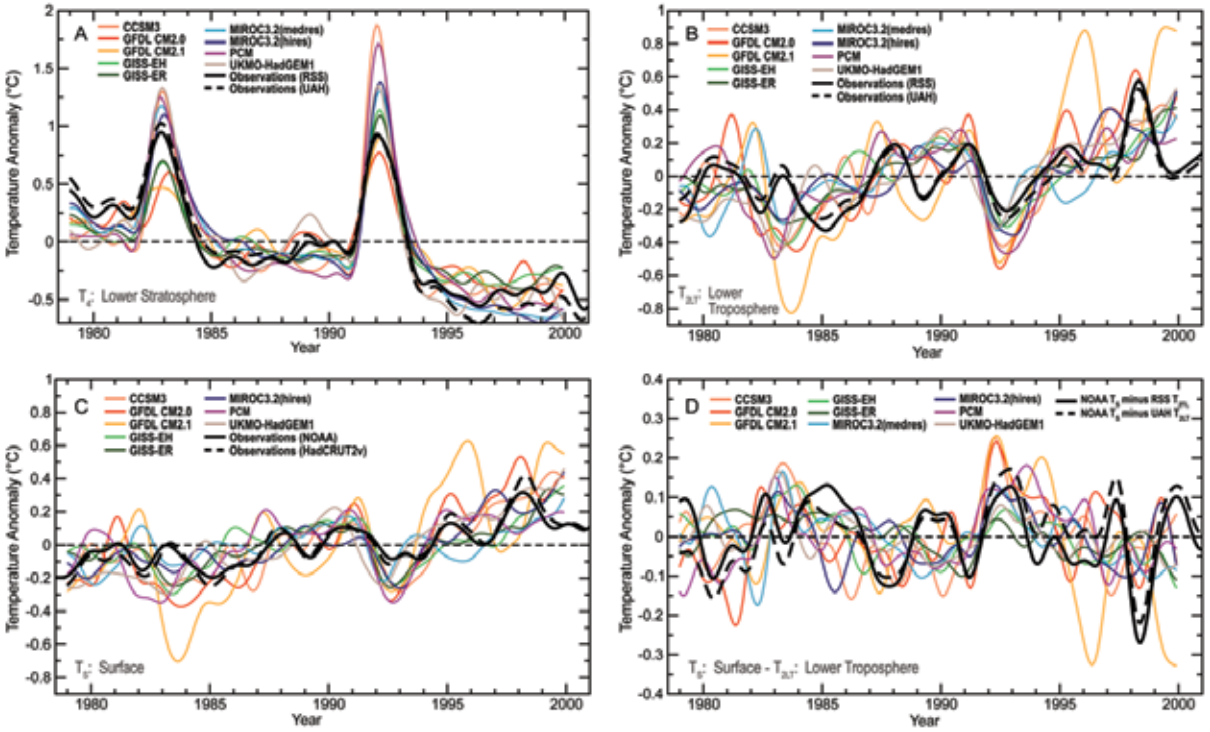

Figure 5.2: Modeled and observed changes in global-mean monthly-mean temperature of the lower stratosphere (T_4; A), the lower troposphere (T_{2LT}; B), the surface (T_S; C), and the surface minus the lower troposphere ($T_S - T_{2LT}$; D). A simple weighting function approach (Box 2.1, Chapter 2) was used to calculate "synthetic" T_4 and T_{2LT} temperatures (equivalent to the MSU T_4 and T_{2LT} monitored by satellites) from model temperature data. Simulated surface and atmospheric temperatures are from 20CEN experiments performed with nine different models (see Table 5.1). These models were chosen because they satisfy certain minimum requirements in terms of the forcings applied in the 20CEN run: all nine were driven by changes in well-mixed GHGs, sulfate aerosol direct effects, tropospheric and stratospheric ozone, volcanic aerosols, and solar irradiance (in addition to other forcings; see Table 5.2). Observed satellite-based estimates of T_4 and T_{2LT} changes were obtained from both RSS and UAH (see Chapter 3). Observed T_S results in C are from NOAA and HadCRUT2v, while observed $T_S - T_{2LT}$ differences in D use both observed T_{2LT} datasets, but NOAA T_S data only. All anomalies are expressed as departures from a 1979 to 1999 reference period average, and were smoothed with the same filter. To make it easier to compare temperature variability in models with different ensemble sizes (see Table 5.1), only the first 20CEN realization is plotted from each model. This also facilitates comparisons of modeled and observed variability.

lated trend distributions.[45] Our goal here is to determine where model results are qualitatively consistent with observations, and where serious inconsistencies are likely to exist. This does not obviate the need for the more rigorous statistical comparisons described in Box 5.5, which should be a high priority (see Recommendations).

5.1 Global-Mean Temperature and Lapse-Rate Trends

In all but two of the 49 20CEN realizations, the global-mean temperature of the lower stratosphere experiences a net cooling over 1979 to 1999 (Figures 5.2A, 5.3A)[46]. The model average T_4 trend is –0.25ºC/decade (Table 5.4A). Most of this cooling is due to the combined effects of stratospheric ozone depletion and increases in well-mixed GHGs (Ramaswamy *et al.,* 2001a,b), with the former the dominant influence on T_4 changes over the satellite era (Ramaswamy *et al.,* 1996; Santer *et al.,* 2003a). The

[45] The 49 20CEN realizations analyzed here are a very small sample from the large population of results that could have been generated by accounting for existing uncertainties in physics parameterizations and historical forcings (*e.g.,* Allen, 1999; Stainforth *et al.,* 2005). Likewise, the observational datasets that we consider in this report probably only capture part of the true "construction uncertainty" inherent in the development of homogeneous climate records from raw temperature measurements. We do not know *a priori* whether temperature changes inferred from these small samples are representative of the true temperature changes that would be estimated from the much larger (but unknown) populations of model and observational results. This is another reason why we are cautious about making formal assessments of the statistical significance of differences between modeled and observed temperature trends. We do, however, attempt to characterize some basic statistical properties of the model results (see Tables 5.4A,B).

[46] In the following, all inter-model and model-data comparisons are over January 1979 to December 1999. This is the longest period of overlap (at least during the satellite era) between the model experiments (which generally end in 1999) and the satellite data (which start in 1979).

Table 5.3: Forcings used in 20CEN experiments performed with the PCM, CCSM3.0, GFDL CM2.1, and GISS-EH models. Shading indicates a forcing that was not incorporated or that did not vary over the course of the experiment.

	PCM	CCSM3.0	GFDL CM2.1	GISS-EH
Well-mixed greenhouse gases	IPCC Third Assessment Report.	IPCC Third Assessment Report.	IPCC Third Assessment Report and World Meteorological Organization (2003).	CH_4, N_2O and CFC spatial distributions are fit to Minschwaner *et al.* (1998).
Sulfate aerosols (direct effects)	Spatial patterns of sulfur dioxide [SO_2] emissions prescribed over seasonal cycle. Year-to-year changes scaled by estimates of historical changes in SO_2 emissions.[a]	Sulfur cycle model using time and space-varying SO_2 emissions (Smith *et al.*, 2001, 2005).[b]	Computed from an atmospheric chemistry transport model.[c]	Based on simulations of Koch *et al.* (1999) and Koch (2001).[d]
Sulfate aerosols (indirect effects)	Not included.	Not included.	Not included.	Parameterization of aerosol indirect effects on cloud albedo and cloud cover.[d]
Stratospheric ozone	Assumed to be constant up to 1970. After 1970 prescribed from a NOAA dataset.[a]	Assumed to be constant up to 1970. After 1970 prescribed from a NOAA dataset.[b]	Specified using data from Randel and Wu (1999).	Specified using data from Randel and Wu (1999).[d]
Tropospheric ozone	Computed from an atmospheric chemistry transport model. Held constant after 1990.[a]	Computed from an atmospheric chemistry transport model. Held constant after 1990.[b]	Computed from an atmospheric chemistry transport model.[c]	Computed from an atmospheric chemistry transport model (Shindell *et al.*, 2003).[d]
Black carbon aerosols	Not included.	Present-day estimate of distribution and amount of black carbon, scaled by population changes over 20th century.[b]	Computed from an atmospheric chemistry transport model.[c]	Based on simulations of Koch *et al.* (1999) and Koch (2001).[d]
Organic aerosols	Not included.	Not included.	Computed from an atmospheric chemistry transport model.[c]	Based on simulations of Koch *et al.* (1999) and Koch (2001).[d]
Sea salt	Not included.	Distributions held fixed in 20th century at year 2000 values.[b]	Distributions held fixed at 1990 values.	Decadally varying.
Dust	Not included.	Distributions held fixed in 20th century at year 2000 values.[b]	Distributions held fixed at 1990 values.	Decadally varying.
Land use change	Distributions held fixed at present-day values.	Distributions held fixed at present-day values.	Knutson *et al.* (2006) global land use reconstruction history. Includes effect on surface albedo, surface roughness, stomatal resistance, and effective water capacity.	Uses Ramankutty and Foley (1999) and Klein Goldewijk (2001) time-dependent datasets. Effects on albedo and evapotranspiration included, but no irrigation effects.[d]
Volcanic stratospheric aerosols	Ammann *et al.* (2003).	Ammann *et al.* (2003).	"Blend" between Sato *et al.* (1993) and Ramachandran *et al.* (2000).	Update of Sato *et al.* (1993).
Solar irradiance	Hoyt and Schatten (1993).	Lean *et al.* (1995).	Lean *et al.* (1995).	Uses solar spectral changes of Lean (2000).

[a] See Dai *et al.* (2001) for further details.
[b] See Meehl *et al.*(2005) for further details.
[c] The chemistry transport model (MOZART; see Horowitz *et al.*, 2003; Tie *et al.*, 2005) was driven by meteorology from the Middle Atmosphere version of the Community Climate Model ("MACCM"; version 3). "1990" weather from MACCM3 was used for all years between 1860 and 2000.
[d] See Hansen *et al.* (2005a) for further details.

model average cooling is larger (–0.35ºC/decade) and closer to the satellite-based estimates if it is calculated from the subset of 20CEN realizations that include forcing by ozone depletion. The range of model T_4 trends encompasses the trends derived from satellites, but not the larger trends estimated from radiosondes. The most likely explanation for this discrepancy is a residual cooling trend in the radiosonde data (Chapter 4)[47]. The neglect of stratospheric water vapor increases in most of the 20CEN runs considered here (Shine *et al.,* 2003) may be another contributory factor[48].

Table 5.4A: Summary statistics for global-mean temperature trends calculated from 49 different realizations of 20CEN experiments performed with 19 different coupled models. Results are for four different atmospheric layers (T_4, T_2, T^*_G, and T_{2LT}), the surface (T_S), and differences between the surface and the troposphere (T_S minus T^*_G and T_S minus T_{2LT}). All trends were calculated over the 252-month period from January 1979 to December 1999 using global-mean monthly-mean anomaly data. Results are in °C/decade. The values in the "Mean" column correspond to the locations of the red lines in the seven panels of Figure 5.3. For each layer, means, medians and standard deviations were calculated from a sample size of n = 19, *i.e.,* from ensemble means (if available) and individual realizations (if ensembles were not performed). This avoids placing too much weight on results from a single model with a large number of realizations. Maximum and minimum values were calculated from all available realizations (*i.e.,* from a sample size of n = 49).

Layer	Mean	Median	Std. Dev. (1σ)	Minimum	Maximum
T_4	-0.25	-0.28	0.19	-0.70	0.08
T_2	0.14	0.12	0.08	0.02	0.35
T^*_G	0.18	0.17	0.08	0.05	0.38
T_{2LT}	0.20	0.19	0.07	0.06	0.39
T_S	0.16	0.16	0.06	0.05	0.33
$T_S - T^*_G$	-0.02	-0.02	0.05	-0.11	0.08
$T_S - T_{2LT}$	-0.03	-0.03	0.03	-0.10	0.05

Superimposed on the overall cooling of T_4 are the large stratospheric warming signals in response to the eruptions of El Chichón (in April 1982) and Mt. Pinatubo (in June 1991)[49]. Nine of the 19 IPCC models explicitly included volcanic aerosols (Figure 5.2A and Table 5.2)[50]. Seven of these nine models overestimate the observed stratospheric warming after Pinatubo. GFDL CM2.1 simulates the Pinatubo response reasonably well, but underestimates the response to El Chichón. Differences in the magnitude of the applied volcanic aerosol forcings must account for some of the inter-model differences in the T_4 warming signals (Table 5.3)[51].

Over 1979 to 1999, the global-mean troposphere warms in all 49 20CEN simulations considered here (Figures 5.2B, 5.3B-D). The shorter-term cooling signals of the El Chichón and Mt. Pinatubo eruptions are superimposed on this gradual warming[52]. Because of the influence of stratospheric cooling on T_2, the model average

[47] Recent work suggests that this residual trend is largest in the lower stratosphere and upper troposphere, and is primarily related to temporal changes in the solar heating of the temperature sensors carried by radiosondes (and failure to properly correct for this effect; see Sherwood *et al.,* 2005; Randel and Wu, 2006).

[48] Recent stratospheric water vapor increases are thought to be partly due to the oxidation of methane, and are expected to have a net cooling effect on T_4. To our knowledge, CH_4-induced stratospheric water vapor increases were explicitly incorporated in only two of the 19 models considered here (GISS-EH and GISS-ER; Hansen *et al.,* 2005a).

[49] These warming signals occur because volcanic aerosols absorb both incoming solar radiation and outgoing thermal radiation (Ramaswamy *et al.,* 2001a).

[50] The MRI-CGCM2.3.2 model incorporated volcanic effects indirectly rather than explicitly, using estimated volcanic forcing data from Sato *et al.* (1993) to adjust the solar irradiance at the top of the model atmosphere. This procedure would not yield volcanically-induced stratospheric warming signals.

[51] More subtle details of the forcing are also relevant to interpretation of inter-model T_4 differences, such as different assumptions regarding the aerosol size distribution, the vertical distribution of the volcanic aerosol relative to the model tropopause, *etc.* Note that observed T_4 changes over the satellite era are not well-described by a simple linear trend, and show evidence of a step-like decline in stratospheric temperatures after the El Chichón and Mt. Pinatubo eruptions (Pawson *et al.,* 1998; Seidel and Lanzante, 2004). Inter-model differences in the applied ozone forcings and solar forcings may help to explain why the GFDL, GISS, and HadGEM1 models appear to reproduce some of this step-like behavior, particularly after El Chichón, while T_4 decreases in PCM are much more linear (Dameris *et al.,* 2005; Ramaswamy *et al.,* 2006).

[52] Because of differences in the timing of modeled and observed ENSO events (Section 5.2), the tropospheric and surface cooling caused by El Chichón is more noticeable in all models than in observations (where it was partially masked by the large 1982/83 El Niño; Figures 5.2B,C).

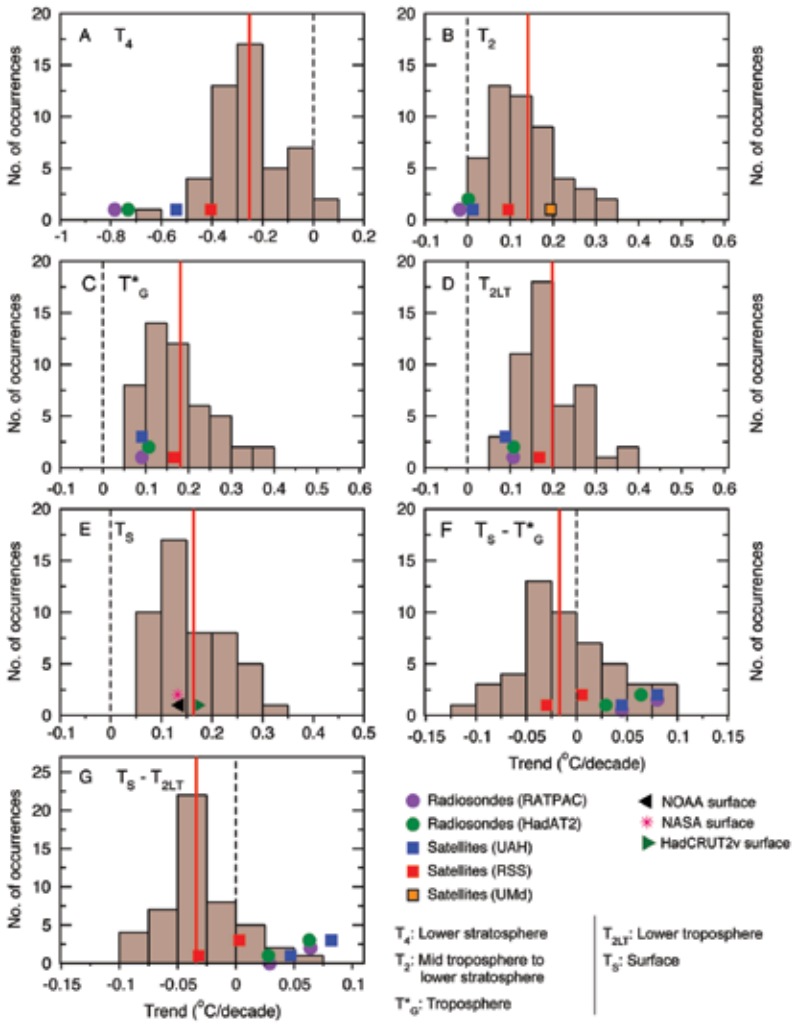

Figure 5.3: Modeled and observed trends in time series of global-mean T_4 (panel A), T_2 (panel B), T^*_G (panel C), T_{2LT} (panel D), T_S (panel E), T_S minus T^*_G (panel F), and T_S minus T_{2LT} (panel G). All trends were calculated using monthly-mean anomaly data. The analysis period is 1979 to 1999. Model results are displayed in the form of histograms. Each histogram is based on results from 49 individual realizations of the 20CEN experiment, performed with 19 different models (Table 5.1). The applied forcings are listed in Table 5.2. The vertical red line in each panel is the mean of the model trends, calculated with a sample size of n = 19 (see Table 5.4A). Observed trends are estimated from two radiosonde and three satellite datasets (T_2), two radiosonde and two satellite datasets (T_4, T^*_G and T_{2LT}), and three different surface datasets (T_S) (see Chapter 3). The bottom "rows" of the observed difference trends in panels F and G were calculated with NOAA T_S data. The top "rows" of observed results in panels F and G were computed with HadCRUT2v T_S data. The vertical offsetting of observed results in these panels (and also in panels B-E) is purely for the purpose of simplifying the visual display – observed trends bear no relation to the y-axis scale. To simplify the display, the Figure does not show the statistical uncertainties arising from the fitting of linear trends to noisy data. GISS observed T_S trends (not shown) are very close to those estimated with NOAA T_S data (see Chapter 3).

trend is smaller for this layer than for either T_{2LT} or T^*_G, which are more representative of temperature changes in the bulk of the troposphere (Table 5.4A)[53]. All of the satellite- and radiosonde-based trends in T_{2LT} and T^*_G are contained within the spread of model results. This illustrates that there is no fundamental discrepancy between modeled and observed trends in global-mean tropospheric temperature.

In contrast, the T_2 trends in both radiosonde data sets are either slightly negative or close to zero, and are smaller than all of the model results. This difference is most likely due to contamination from residual stratospheric and upper-tropospheric cooling biases in the radiosonde data (Chapter 4; Sherwood *et al.,* 2005; Randel and Wu, 2006). The satellite-based T_2 trends are either close to the model average (RSS and VG) or just within the model range (UAH; Fig. 5.3B). Even without formal statistical tests, it is clear that observational uncertainty is an important factor in assessing the consistency between modeled and observed changes in mid- to upper tropospheric temperature (Santer *et al.,* 2003b).

Observed T_S trends closely bracket the model average (Figure 5.3E). There is no inconsistency between modeled and observed surface temperature changes. Structural uncertainties in observed T_S trends are much smaller than for trends in T_4 or tropospheric layer-average temperatures (see Chapter 4).

The model-simulated ranges of lapse-rate trends also encompass virtually all observational results (Figures 5.3F,G)[54]. Closer inspection reveals that the model-average trends in tropospheric lapse rate are slightly negative,[55] indicating larger warming aloft than at the surface. Most combinations of observed T_S, T^*_G, and T_{2LT} data sets yield the converse result, and show smaller warming aloft than at the surface. As in the case of global-mean T^*_G and T_{2LT} trends, RSS-based lapse-rate trends are invariably closest to the model average results. Both models and observations show a tendency towards positive values of T_S minus T_{2LT} for several years after the El Chichón and Mt. Pinatubo eruptions, indicative of larger

[53] Due to ozone-induced cooling of the lower stratosphere, the model-average T_2 trend is slightly smaller (0.12°C/decade) and closer to the RSS result if it is estimated from the subset of 20CEN runs that include stratospheric ozone depletion. Subsetting in this way has little impact on the model-average T_{2LT} and T^*_G trends.

[54] Note that the subtraction of temperature variability common to surface and troposphere decreases (by about a factor of two) the large range of model trends in T_S, T^*_G, and T_{2LT} (Table 5.4A).

[55] Values are –0.02°C/decade in the case of T_S minus T^*_G and –0.03°C/decade for T_S minus T_{2LT}.

cooling aloft than at the surface (Figure 5.2D; Section 5.4).

5.2 Tropical Temperature and Lapse-Rate Trends

The previous section examined whether simulated global-mean temperature trends were contained within current estimates of structural uncertainty in observations. Since ENSO is primarily a tropical phenomenon, its influence on surface and tropospheric temperature is more pronounced in the tropics than in global averages. Observations contain only one specific sequence of ENSO fluctuations from 1979 to present, and only one sequence of ENSO effects on tropical temperatures. The model 20CEN runs examined here provide many different sequences of ENSO variability. We therefore expect – and find – that these runs yield a wide range of trends in tropical surface and tropospheric temperature (Figure 5.4)[56]. It is of interest whether this large model range encompasses the observed trends.

At the surface, results from the multi-model ensemble include all observational estimates of tropical temperature trends (Figure 5.4E; Table 5.4B). Observed results are close to the model average T_S trend of +0.16ºC/decade. There is no evidence that the models significantly over- or underestimate the observed surface warming. In the troposphere, all observational results are still within the range of possible model solutions, but the majority of model results show tropospheric warming that is larger than observed (Figures 5.4B-D). As in the case of the global-mean T_4 trends, the cooling of the tropical stratosphere in both radiosonde data sets is larger than in any of the satellite data sets or model results (Figure 5.4A)[57]. The UAH and RSS T_4 trends are close to the model average[58].

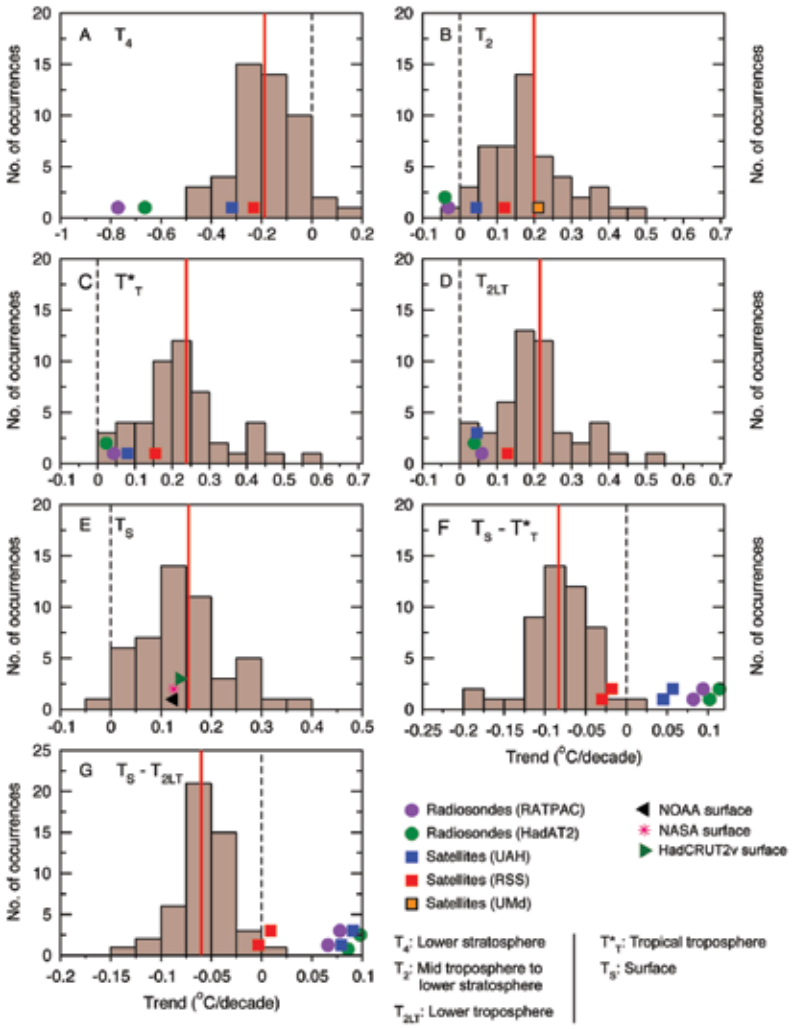

Figure 5.4: As for Figure 5.3, but for trends in the tropics (20°N-20°S).

In the model results, trends in the two measures of tropical lapse rate (T_S minus T_{2LT} and T_S minus T^*_T) are almost invariably negative, indicating larger warming aloft than at the surface (Figure 5.4F,G). Similar behavior is evident in only one of the four upper-air data sets examined here (RSS)[59]. The RSS trends are just within the range of model solutions[60].

[56] This would be true even for a hypothetical "perfect" climate model run with "perfect" forcings. The large model range of tropical temperature trends is not solely due to the effects of ENSO and other modes of internal variability. It also arises from uncertainties in the models and forcings (see Boxes 5.1 and 5.2 and Table 5.2).

[57] This supports recent findings of a residual cooling bias in tropical radiosonde data (Sherwood *et al.,* 2005; Randel and Wu, 2006).

[58] The model average is –0.27ºC/decade when estimated from the subset of 20CEN runs that include stratospheric ozone depletion.

[59] The UMd group does not provide either a stratospheric or lower-tropospheric temperature retrieval, and so could not be included in the comparison of modeled and observed trends in T_S minus T^*_T or T_S minus T_{2LT}. Assuming that the relationships between the UMd T_2, T_{2LT} and T^*_T trends were similar to those for the UAH and RSS data, the UMd data would yield T_{2LT} and T^*_T trends that were larger than in RSS. This would expand the range of observational uncertainty shown in Figures 5.4F,G.

[60] Three of the four RSS-based results in Figures 5.4 F and G are within two standard deviations of the model average values (see Table 5.4B). Note also that for their tropical T_{2LT} trend, RSS claims a 2σ uncertainty of ± 0.09°C/decade (Mears and Wentz 2005; Mears personal communication). This uncertainty is not included here.

Table 5.4B: As for Table 5.4A, but for tropical temperature trends (calculated from spatial averages over 20°N-20°S).

Layer	Mean	Median	Std. Dev. (1σ)	Minimum	Maximum
T_4	-0.19	-0.19	0.15	-0.49	0.13
T_2	0.20	0.19	0.10	-0.01	0.48
T^*_T	0.24	0.21	0.11	0.01	0.56
T_{2LT}	0.22	0.19	0.09	0.01	0.51
T_S	0.16	0.14	0.07	-0.02	0.37
$T_S - T^*_T$	-0.08	-0.08	0.04	-0.19	0.02
$T_S - T_{2LT}$	-0.06	-0.05	0.03	-0.15	0.01

The model results that overlap with the RSS-derived tropical lapse-rate trends exhibit less surface warming than the observations. This analysis is revisited in Section 5.4 using a metric that more directly addresses the relationship between surface and tropospheric temperature changes. Tropical lapse-rate trends in both radiosonde data sets and in the UAH satellite data are always positive (larger warming at the surface than aloft), and lie outside the range of model results.

This comparison suggests that discrepancies between our current best estimates of simulated and observed lapse-rate changes may be larger and more serious in the tropics than in globally averaged data. Large structural uncertainties in the observations (even in the sign of the trend in tropical lapse-rate changes) make it difficult to reach more definitive conclusions regarding the significance and importance of model-data discrepancies (see Section 5.4).

5.3 Spatial Patterns of Lapse-Rate Trends

Maps of the trends in lower tropospheric lapse rate help to identify geographical regions where the model-data discrepancies in Figures 5.4F and 5.4G are most pronounced. We focus on four U.S. models run with the most complete set of forcings: CCSM3.0, PCM, GFDL CM2.1, and GISS-EH (Table 5.3). These show qualitatively similar patterns of trends in T_S minus T_{2LT} (Figures 5.5A-D). Over most of the tropical ocean, the simulated warming is larger in the troposphere than at the surface. All models have some tropical land areas where the surface warms relative to the troposphere. The largest relative warming of the surface occurs at high latitudes in both hemispheres.

To illustrate structural uncertainties in the observed data, we show two different patterns of trends in T_S minus T_{2LT}. Both rely on the same NOAA surface data, but use either UAH (Figure 5.5E) or RSS (Figure 5.5F) as their source of T_{2LT} results. The "NOAA minus UAH" combination provides a picture that is very different from the model results, with coherent warming of the surface relative to the troposphere over much of the world's tropical oceans. While "NOAA minus RSS" also has relative warming of the surface in the Western and tropical Pacific, it shows relative warming of the troposphere in the eastern tropical Pacific and Atlantic Oceans. This helps to clarify why simulated lapse-rate trends in Figures 5.4F and 5.4G are closer to NOAA minus RSS results than to NOAA minus UAH results.

As pointed out by Santer *et al.* (2003b) and Christy and Spencer (2003), we cannot use such model-data comparisons alone to determine whether the UAH or RSS T_{2LT} data set is closer to (an unknown) "reality." As the next section will show, however, models and basic theory can be used to identify aspects of observational behavior that require further investigation, and may help to constrain observational uncertainty.

5.4 Tropospheric Amplification of Tropical Surface Temperature Changes

When surface and lower tropospheric temperature changes are spatially averaged over the deep tropics, and when day-to-day tropical temperature changes are averaged over months, seasons, or years, it is evident that temperature changes aloft are larger than at the surface. This "amplification" behavior has been described in many observational and modeling studies, and is a consequence of the release of latent heat by moist convecting air (*e.g.,* Manabe and Stouffer, 1980; Horel and Wallace, 1981; Pan and

Oort, 1983; Yulaeva and Wallace, 1994; Hurrell and Trenberth, 1998; Soden, 2000; Wentz and Schabel, 2000; Hegerl and Wallace, 2002; Knutson and Tuleya, 2004)[61].

A recent study by Santer *et al.* (2005) examined this amplification behavior in the same 20CEN runs and observational data sets considered in the present report. The sole difference (relative to the data used here) was that Santer *et al.* analyzed a version of the UAH T_{2LT} data that had not yet been adjusted for a recently discovered error (Mears and Wentz, 2005)[62]. The amplification of tropical surface temperature changes was assessed on different timescales (monthly, annual, and multi-decadal) and in different atmospheric layers (T^*_T and T_{2LT}).

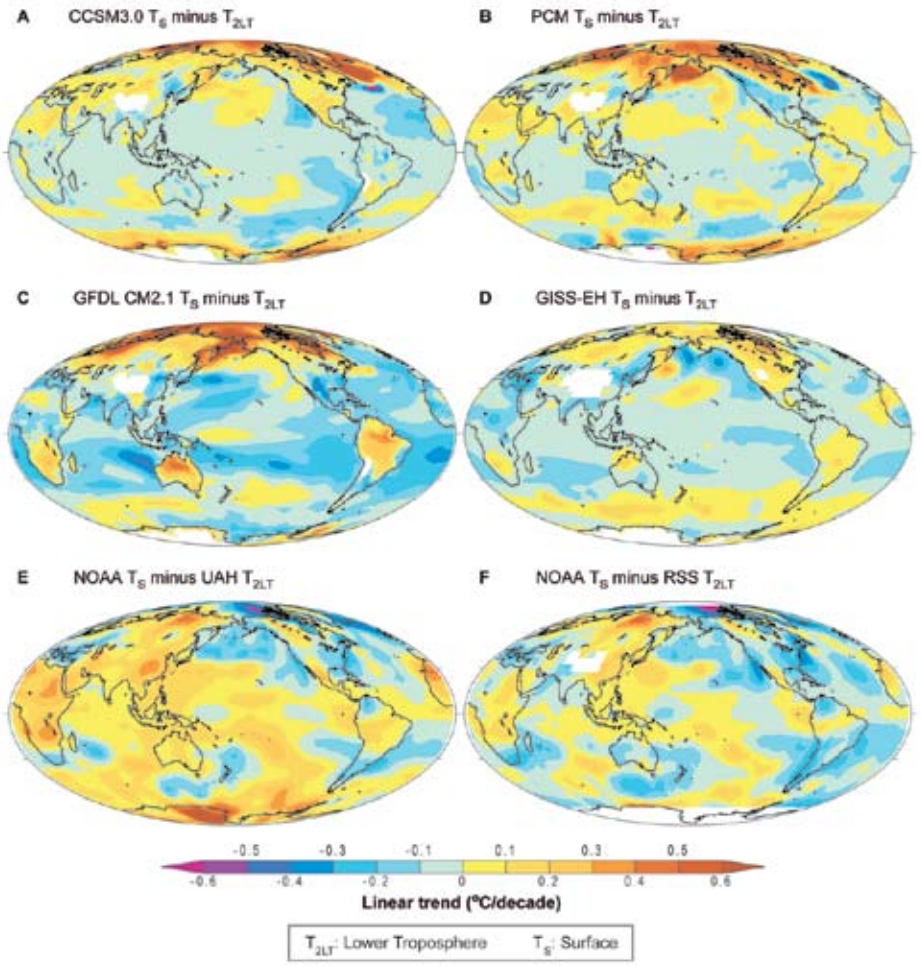

Figure 5.5: Modeled and observed maps of the differences between trends in T_S and T_{2LT}. All trends in T_S and T_{2LT} were calculated over the 252-month period from January 1979 to December 1999. Model results are ensemble means from 20CEN experiments performed with CCSM3.0 (panel A), PCM (panel B), GFDL CM2.1 (panel C), and GISS-EH (panel D). Observed results rely on NOAA T_S trends and on two different satellite estimates of trends in T_{2LT}, obtained from UAH (panel E) and RSS (panel F). White denotes high elevation areas where it is not meaningful to calculate synthetic T_{2LT} (panels A-D). Note that RSS mask T_{2LT} values in such regions, while UAH do not (compare panels E, F).

On short timescales (month-to-month and year-to-year variations in temperature), the estimated tropospheric amplification of surface temperature changes was in good agreement in all model and observational data sets considered, and was in accord with basic theory. This is illustrated in Figure 5.6, which shows the standard deviations of monthly-mean T_S anomalies plotted against the standard deviations of monthly-mean anomalies of T_{2LT} (panel A) and T^*_T (panel B). All model and observational results lie above the black line indicating equal temperature variability aloft and at the surface. All have similar "amplification factors" between their surface and tropospheric variability[63]. In the models, these similarities occur despite differences in physics, resolution, and forcings, and despite a large range (roughly a factor of 5) in the size of simulated temperature variability. In observations, the scaling ratios estimated from monthly temperature variability are relatively unaffected by the structural uncertainties discussed in Chapter 4.

[61] The essence of tropical atmospheric dynamics is that the tropics cannot support large temperature gradients, so waves (Kelvin, Rossby, gravity) even out the temperature field between convecting and non-convective regions. The temperature field throughout the tropical troposphere is more or less on the moist adiabatic lapse rate set by convection over the warmest waters. This is why there is a trade wind inversion where this profile finds itself inconsistent with boundary layer temperatures in the colder regions.

[62] The error was related to the UAH group's treatment of systematic drifts in the time of day at which satellites sample Earth's diurnal temperature cycle (see Chapter 4).

[63] Note that the slope of the red regression lines that has been fitted to the model results is slightly steeper for T^*_T than for T_{2LT} (c.f. panels 5.6A and 5.6B). This is because T^*_T samples more of the mid-troposphere than T_{2LT} (see Prospectus). Amplification is expected to be larger in the mid-troposphere than in the lower troposphere.

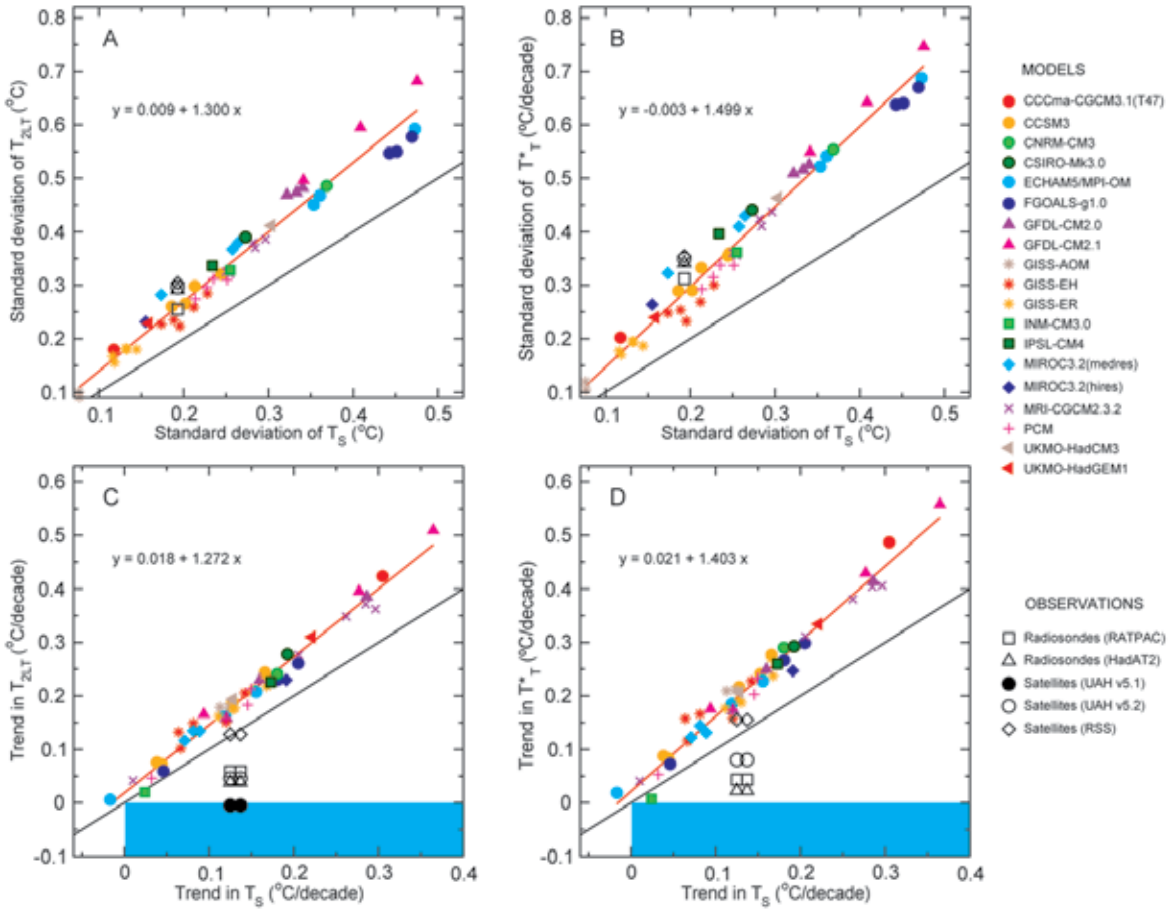

Figure 5.6: Scatter plots showing the relationships between tropical temperature changes at Earth's surface and in two different layers of the troposphere. All results rely on temperature data that have been spatially-averaged over the deep tropics (20°N-20°S). Model data are from 49 realizations of 20CEN runs performed with 19 different models (Table 5.1). Observational results were taken from four different upper-air datasets (two from satellites, and two from radiosondes) and two different surface temperature datasets (see Chapter 3). The two upper panels provide information on the month-to-month variability in T_S and T_{2LT} (panel A) and in T_S and T^*_T (panel B). The two bottom panels consider temperature changes on multi-decadal timescales, and show the trends (over 1979 to 1999) in T_S and T_{2LT} (panel C) and in T_S and T^*_T (panel D). The red line in each panel is the regression line through the model points. Its slope provides information on the amplification of surface temperature variability and trends in the free troposphere. The black line in each panel is given for reference purposes, and has a slope of 1. Values above (below) the black lines indicate tropospheric amplification (damping) of surface temperature changes. There are two columns of observational results in C and D. These are based on the NOAA and HadCRUT2v T_S trends (0.12 and 0.14°C/decade, respectively). Note that panel C show results from published and recently-revised versions of the UAH T_{2LT} data (versions 5.1 and 5.2). Since the standard deviations calculated from NOAA and HadCRUT2v monthly T_S anomalies are very similar, observed results in A and B use NOAA standard deviations only. The blue shading in the bottom two panels defines the region of simultaneous surface warming and tropospheric cooling.

A different picture emerges if amplification behavior is estimated from decadal changes in tropical temperatures. Figures 5.6C and 5.6D show multi-decadal trends in T_S plotted against trends in T_{2LT} and T^*_T. The 20CEN runs exhibit amplification factors that are consistent with those estimated from month-to-month and year-to-year temperature variability[64]. Only one observational upper-air data set (RSS) shows amplified warming aloft, and similar amplification relationships on short and on

[64] As in the case of amplification factors inferred from short-timescale variability, the factors estimated from multi-decadal temperature changes are relatively insensitive to inter-model differences in physics and the applied forcings (see Table 5.3). At first glance, this appears to be a somewhat surprising result in view of the large spatial and temporal heterogeneity of certain forcings (see Section 3). Black carbon aerosols, for example, are thought to cause localized heating of the troposphere relative to the surface (Box 5.3), a potential mechanism for altering amplification behavior. The fact that amplification factors are similar in experiments that include and exclude black carbon aerosols suggests that aerosol-induced tropospheric heating is not destroying the connection of large areas of the tropical ocean to a moist adiabatic lapse rate. Single-forcing experiments (see Recommendations) will be required to improve our understanding of the physical effects of black carbon aerosols and other spatially-heterogeneous forcings on tropical temperature-change profiles.

long timescales. The other observational data sets have scaling ratios less than 1, indicating tropospheric damping of surface warming (Fu and Johanson, 2005; Santer *et al.,* 2005)[65].

These analyses shed further light on the differences between modeled and observed changes in tropical lapse rates described in Section 5.2. They illustrate the usefulness of comparing models and data on different timescales. On short timescales, it is evident that models successfully capture the basic physics that controls "real world" amplification behavior. On long timescales, model-data consistency is sensitive to structural uncertainties in the observations. One possible interpretation of these results is that in the real world, different physical mechanisms govern amplification processes on short and on long timescales, and models have some common deficiency in simulating such behavior. If so, these "different physical mechanisms" need to be identified and understood.

Another interpretation is that the same physical mechanisms control short- and long-term amplification behavior. Under this interpretation, residual errors in one or more of the observed data sets must affect their representation of long-term trends, and must lead to different scaling ratios on short and long timescales. This explanation appears to be the more likely one in view of the large structural uncertainties in observed upper-air data sets (Chapter 4) and the complementary physical evidence supporting recent tropospheric warming (see Section 6).

"Model error" and "observational error" are not mutually exclusive explanations for the amplification results shown in Figures 5.6C and D. Although a definitive resolution of this issue has not yet been achieved, the path towards such resolution is now more obvious. We have learned that models show considerable consistency in terms of what they tell us about tropospheric amplification of surface warming. This consistency holds on a range of different timescales. Observations display consistent amplification behavior on short timescales, but radically different behavior on long timescales. Clearly, not all of the observed lapse-rate trends can be equally probable. Intelligent use of "complementary evidence" – from the behavior of other climate variables, from remote sensing systems other than MSU, and from more systematic exploration of the impacts of different data adjustment choices – should ultimately help us to constrain observational uncertainty, and reach more definitive conclusions regarding the true significance of modeled and observed lapse-rate differences.

Models show considerable consistency in terms of what they tell us about tropospheric amplification of surface warming.

5.5 Vertical Profiles of Atmospheric Temperature Change

Although formal fingerprint studies have not yet been completed with atmospheric temperature-change patterns estimated from the new 20CEN runs, it is instructive to make a brief qualitative comparison of these patterns. This helps to address the question of whether the inclusion of previously neglected forcings (like carbonaceous aerosols and land use/land cover changes; see Section 2) has fundamentally modified the "fingerprint" of human-induced atmospheric temperature changes searched for in previous detection studies.

We examine the zonal-mean profiles of atmospheric temperature change in 20CEN runs performed with four U.S. models (CCSM3, PCM, GFDL CM2.1, and GISS-EH). All four show a common large-scale fingerprint of stratospheric cooling and tropospheric warming over 1979 to 1999 (Figures 5.7A-D). The pattern of temperature change estimated from HadAT2 radiosonde data is broadly similar, although the transition height between stratospheric cooling and tropospheric warming is noticeably lower than in the model simulations (Figure 5.7E). Another noticeable difference is that the HadAT2 data show a relative lack of warming in the tropical troposphere,[66] where all four models simulate maximum warming. This particular aspect of the observed temperature-change pattern is very sensitive to data adjustments

[65] The previous version of the UAH T_{2LT} data yielded a negative amplification factor for multi-decadal changes in tropical temperatures. The UMd data set, which exhibits greatest warming in T_2, has not to date produced a T_{2LT} or T^*_T product, and so could not be included in Figure 5.6. However, assuming an internally consistent set of channel records, the UMd data would show larger T_{2LT} and T^*_T trends than RSS, and would therefore have amplification factors consistent with or greater than those inferred from the models.

[66] Despite the "end point" effect of the large El Niño event in 1997-1998 (see Chapter 3).

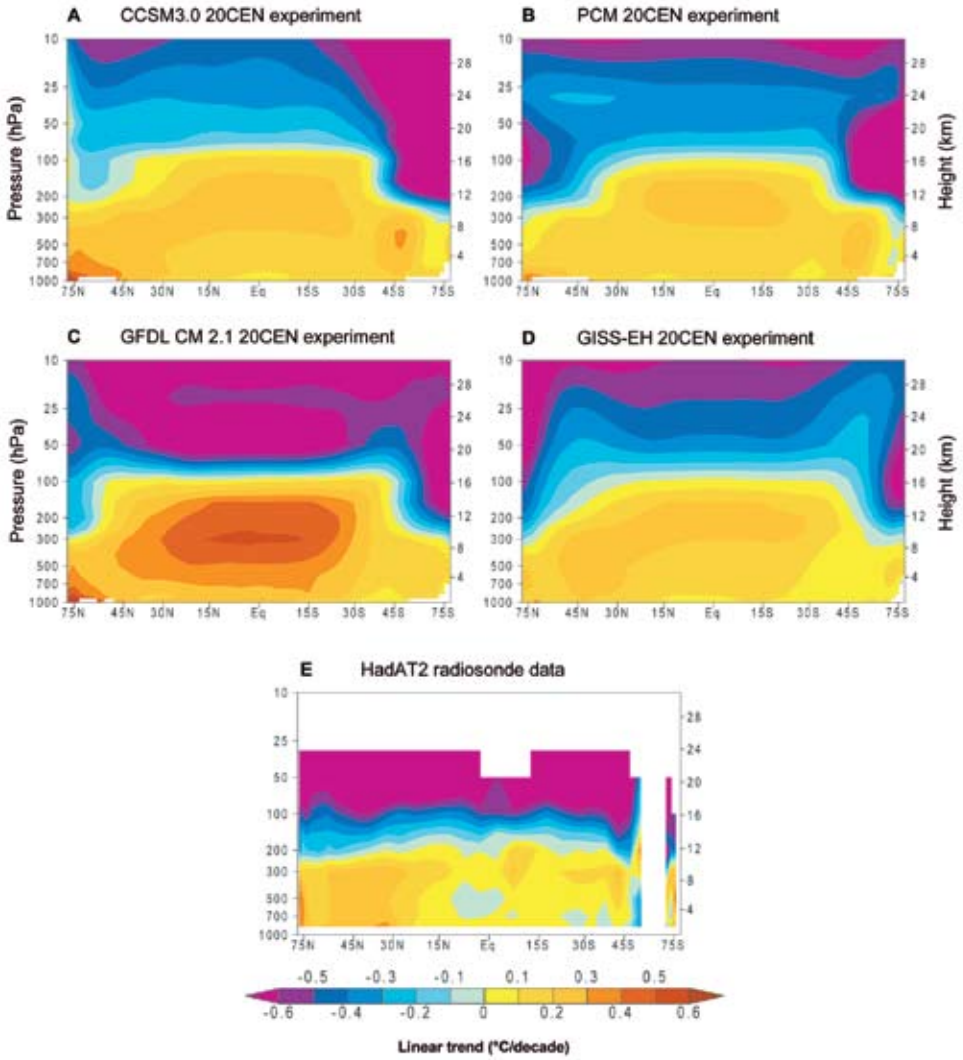

Figure 5.7: Zonal-mean patterns of atmospheric temperature change in "20CEN" experiments performed with four different climate models and in observational radiosonde data. Model results are for CCSM3.0 (panel A), PCM (panel B), GFDL CM 2.1 (panel C), and GISS-EH (panel D). The model experiments are ensemble means. There are differences between the sets of climate forcings that the four models used in their 20CEN runs (Table 5.3). Observed changes (panel E) were estimated with HadAT2 radiosonde data (Thorne *et al.,* 2005, and Chapter 3). The HadAT2 temperature data do not extend above 30 hPa, and have inadequate coverage at high latitudes in the Southern Hemisphere. All temperature changes were calculated from monthly-mean data and are expressed as linear trends (in °C/decade) over 1979 to 1999.

(Sherwood *et al.,* 2005; Randel and Wu, 2006). Tropospheric warming in the observations is most obvious in the NH extra-tropics, where our confidence in the reliability of radiosonde records is greatest.

Note that some of the details of the model fingerprint pattern are quite different. For example, GFDL's cooling maximum immediately above the tropical tropopause is not evident in any of the other models. Its maximum warming in the upper tropical troposphere is noticeably larger than in CCSM3.0, PCM, or GISS-EH. While CCSM and GFDL CM2.1 have pronounced hemispheric asymmetry in their stratospheric cooling patterns, with largest cooling at high latitudes in the SH,[67] this asymmetry is less apparent in PCM and GISS-EH.

Future work should consider whether the conclusions of detection studies are robust to such fingerprint differences. This preliminary analysis suggests that the large-scale "fingerprint" of stratospheric cooling and tropospheric warming over the satellite era – a robust feature of previous detection work – has not been fundamentally altered by the inclusion of hitherto-neglected forcings like carbonaceous aerosols and LULC changes (see Table 5.3). This does not diminish the need to quantify the individual contributions of these forcings in appropriate "single forcing" experiments.

6. CHANGES IN "COMPLEMENTARY" CLIMATE VARIABLES

Body temperature is a simple metric of our physical well-being. A temperature of 40ºC (104ºF) is indicative of an illness, but does not by itself identify the cause of the illness. In medicine, investigation of causality typically requires the analysis of many different lines of evidence. Similarly, analyses of temperature alone provide incomplete information on the causes of climate change. For example, there is evidence that major volcanic eruptions affect not only the Earth's radiation budget (Wielicki *et al.,* 2002; Soden *et al.,* 2002) and atmospheric temperatures (Hansen *et al.,* 1997, 2002; Free and Angell, 2002; Wigley *et al.,* 2005a), but also water vapor (Soden *et al.,* 2002), precipitation (Gillett *et al.,* 2004c), atmospheric circulation patterns

[67] This may be related to an asymmetry in the pattern of stratospheric ozone depletion: the largest ozone decreases over the past two to three decades have occurred at high latitudes in the SH.

(see, *e.g.,* Robock, 2000, and Ramaswamy *et al.,* 2001a; Robock and Oppenheimer, 2003), ocean heat content and sea level (Church *et al.,* 2005), and even global-mean surface pressure (Trenberth and Smith, 2005). These responses are physically interpretable and internally consistent[68]. The combined evidence from changes in all of these variables makes a stronger case for an identifiable volcanic effect on climate than evidence from a single variable only.

A "multi-variable" perspective may also be beneficial in understanding the possible causes of differential warming. The value of "complementary" climate data sets for studying this specific problem has been recognized by Wentz and Schabel (2000) and by Pielke (2004). The former found internally consistent increases in SST, T_{2LT}, and marine total column water vapor over the 12-year period from 1987 to 1998[69]. Multi-decadal increases in surface and lower tropospheric water vapor were also reported in the IPCC Second Assessment Report (Folland *et al.,* 2001).[70] More recently, Trenberth *et al.* (2005) found significant increases in total column water vapor over the global ocean[71]. At constant relative humidity, water vapor increases non-linearly with increasing temperature (Hess, 1959). Slow increases in tropospheric water vapor therefore provide circumstantial evidence in support of tropospheric warming. However, water vapor measurements are affected by many of the same data quality and temporal homogeneity problems that influence temperature measurements (Elliott, 1995; Trenberth *et al.,* 2005), so the strength of this circumstantial evidence is still questionable[72].

A "multi-variable" perspective may also be beneficial in understanding the possible causes of differential warming. At constant relative humidity, water vapor increases non-linearly with increasing temperature.

Other climate variables also corroborate the warming of Earth's surface over the second half of the 20th century. Examples include increases in ocean heat content (Levitus *et al.,* 2000, 2005; Willis *et al.,* 2004), sea-level rise (Cabanes *et al.,* 2001), thinning of major ice sheets and ice shelves (Krabill *et al.,* 1999; Rignot and Thomas, 2002; Domack *et al.,* 2005), and widespread glacial retreat, with accelerated rates of glacial retreat over the last several decades (Arendt *et al.,* 2002; Paul *et al.,* 2004)[73].

Changes in some of these "complementary" variables have been used in detection and attribution studies. Much of this work has focused on ocean heat content. When driven

[68] The physical consistency between the temperature and water vapor changes after the Pinatubo eruption has been clearly demonstrated by Soden *et al.* (2002). The surface and tropospheric cooling induced by Pinatubo caused a global-scale reduction in total column water vapor. Since water vapor is a strong GHG, the reduction in water vapor led to less trapping of outgoing thermal radiation by Earth's atmosphere, thus amplifying the volcanic cooling. This is referred to as a "positive feedback." Soden *et al.* "disabled" this feedback in a climate model experiment, and found that the "no water vapor feedback" model was incapable of simulating the observed tropospheric cooling after Pinatubo. Inclusion of the water vapor feedback yielded close agreement between the simulated and observed T_{2LT} responses to Pinatubo. This suggests that the model used by Soden *et al.* captures important aspects of the physics linking the real world's temperature and moisture changes.

[69] The Wentz and Schabel study used NOAA optimally interpolated SST data, a version of the UAH T_{2LT} data that had been corrected for orbital decay effects, and information on total column water vapor from the satellite-based Special Sensor Microwave Imager (SSM/I).

[70] More specifically, Folland *et al.* (2001) concluded, "Changes in water vapor mixing ratio have been analyzed for selected regions using *in situ* surface observations as well as lower-tropospheric measurements based on satellites and weather balloons. A pattern of overall surface and lower-tropospheric water vapor mixing ratio increases over the past few decades is emerging, although there are likely to be some time-dependent biases in these data and regional variations in trends. The more reliable data sets show that it is likely that total atmospheric water vapor has increased several percent per decade over many regions of the Northern Hemisphere since the early 1970s. Changes over the Southern Hemisphere cannot yet be assessed."

[71] Trenberth *et al.* (2005) reported an increase in total column water vapor over 1988 to 2001 of "1.3 ± 0.3% per decade for the ocean as a whole, where the error bars are 95% confidence intervals." This estimate was obtained with an updated version of the SSM/I data set analyzed by Wentz and Schabel (2000).

[72] Note, however, that SSM/I-derived water vapor measurements may have some advantages relative to temperature measurements obtained from MSU. Wentz and Schabel (2000) point out that (under a constant relative humidity assumption), the 22 GHz water vapor radiance observed by SSM/I is three times more sensitive to changes in air temperature than the MSU T_2 54 GHz radiance. Furthermore, while drift in sampling the diurnal cycle influences MSU-derived tropospheric temperatures (Chapter 4), it has a much smaller impact on SSM/I water vapor measurements.

[73] Folland *et al.* (2001) note that "Long-term monitoring of glacier extent provides abundant evidence that tropical glaciers are receding at an increasing rate in all tropical mountain areas". Accelerated retreat of high-elevation tropical glaciers is occurring within the tropical lower tropospheric layer that is a primary focus of this report, and provides circumstantial support for warming of this layer over the satellite era.

by anthropogenic forcing, a number of different CGCMs capture the overall increase in observed ocean heat content estimated by Levitus *et al.* (2000; 2005), but not the large decadal variability in heat content (Barnett *et al.,* 2001; Levitus *et al.,* 2001; Reichert *et al.,* 2002; Sun and Hansen, 2003; Pielke, 2003; Gregory *et al.,* 2004; Hansen *et al.,* 2005b)[74]. It is still unclear whether this discrepancy between simulated and observed variability is primarily due to model deficiencies or is an artifact of how Levitus *et al.* (2000; 2005) "infilled" data-sparse ocean regions (Gregory *et al.,* 2004; AchutaRao *et al.,* 2006).

In summary, the behavior of complementary variables enhances our confidence in the reality of large-scale warming of the Earth's surface, and tells us that the signature of this warming is manifest in many different aspects of the climate system. Pattern-based fingerprint detection work performed with ocean heat content (Barnett *et al.,* 2001; Reichert *et al.,* 2002; Barnett *et al.,* 2005; Pierce *et al.,* 2006), sea-level pressure (Gillett *et al.,* 2003), and tropopause height (Santer *et al.,* 2003a, 2004)[75] suggests that anthropogenic forcing is necessary in order to explain observed changes in these variables. This supports the findings of the surface- and atmospheric temperature studies described in Section 4.4. To date, however, investigations of complementary variables have not enabled us to narrow uncertainties in satellite- and radiosonde-based estimates of tropospheric temperature change over the past 2-3 decades. Formal detection and attribution studies involving water vapor changes may be helpful in this regard, since observations suggest a recent moistening of the troposphere, consistent with tropospheric warming.

7. SUMMARY

This chapter has evaluated a wide range of scientific literature dealing with the possible causes of recent temperature changes, both at the Earth's surface and in the free atmosphere. It shows that many factors – both natural and human-related – have probably contributed to these changes. Quantifying the relative importance of these different climate forcings is a difficult task. Analyses of observations alone cannot provide us with definitive answers. This is because there are important uncertainties in the observations and in the climate forcings that have affected them. Although computer models of the climate system are useful in studying cause-effect relationships, they, too, have limitations. Advancing our understanding of the causes of recent lapse-rate changes will best be achieved by comprehensive comparisons of observations, models, and theory – it is unlikely to arise from analysis of a single model or observational data set.

[74] Model control runs cannot generate such large multi-decadal increases in the heat content of the global ocean.

[75] The tropopause is the transition zone between the turbulently-mixed troposphere, where most weather occurs, and the more stably-stratified stratosphere (see Preface and Chapter 1). Increases in tropopause height over the past 3-4 decades represent an integrated response to temperature changes above and below the tropopause (Highwood *et al.,* 2000; Santer *et al.,* 2004), and are evident in both radiosonde data (Highwood *et al.,* 2000; Seidel *et al.,* 2001) and reanalyses (Randel *et al.,* 2000). In model 20CEN simulations, recent increases in tropopause height are driven by the combined effects of GHG-induced tropospheric warming and ozone-induced stratospheric cooling (Santer *et al.,* 2003a). Available reanalysis products do not provide a consistent picture of the relative contributions of stratospheric and tropospheric temperature changes to recent tropopause height increases (Pielke and Chase, 2004; Santer *et al.,* 2004).