



LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Identification of Anthropogenic Climate Change Using a Second-Generation Reanalysis

Ben Santer, Tom Wiglet, Adrian Simmons, Per Kallberg, Graeme Kelly, Sakari Uppala, Caspar Ammann, James Boyle, Wolfgang Bruggemann, Charles Doutriaux, Mike Fiorino, Carl Mears, Gerald Meehl, Robert Sausen, Karl Taylor, Warren Washington, Michael Wehner, Frank Wentz

June 2, 2004

Journal of Geophysical Research (Atmospheres)

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

Identification of Anthropogenic Climate Change Using a Second-Generation Reanalysis

Benjamin D. Santer¹, Tom M. L. Wigley², Adrian J. Simmons³, Per W. Kållberg³,
Graeme A. Kelly³, Sakari M. Uppala³, Caspar Ammann², James S. Boyle¹,
Wolfgang Brüggemann⁴, Charles Doutriaux¹, Mike Fiorino¹, Carl Mears⁵, Gerald A.
Meehl², Robert Sausen⁶, Karl E. Taylor¹, Warren M. Washington²,
Michael F. Wehner⁷ & Frank J. Wentz⁵

¹ Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore
National Laboratory, Livermore, CA 94550, USA

² National Center for Atmospheric Research, Boulder, CO 80303, USA

³ European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading,
RG2 9AX, UK

⁴ University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

⁵ Remote Sensing Systems, Santa Rosa, CA 95401, USA

⁶ Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre,
Oberpfaffenhofen, D-82234 Wessling, Germany

⁷ Lawrence Berkeley National Laboratory Berkeley, CA 94720, USA

Submitted to *Journal of Geophysical Research (Atmospheres)*

May 31, 2004

Abstract

Changes in the height of the tropopause provide a sensitive indicator of human effects on climate. A previous attempt to identify human effects on tropopause height relied on information from ‘first-generation’ reanalyses of past weather observations. Climate data from these initial model-based reanalyses have well-documented deficiencies, raising concerns regarding the robustness of earlier detection work that employed these data. Here, we address these concerns using information from the new second-generation ERA-40 reanalysis. Over 1979 to 2001, tropopause height increases by nearly 200 meters in ERA-40, partly due to tropospheric warming. The spatial pattern of height increase is consistent with climate model predictions of the expected response to anthropogenic influences alone, significantly strengthening earlier detection results. Atmospheric temperature changes in two different satellite datasets are more highly correlated with changes in ERA-40 than with those in a first-generation reanalysis, thus illustrating the improved quality of temperature information in ERA-40. Our results provide support for claims that human activities have warmed the troposphere and cooled the lower stratosphere over the last several decades of the 20th century, and that both of these changes in atmospheric temperature have contributed to an overall increase in tropopause height.

1 Introduction

Reanalyses are synthesized atmospheric states, derived by reprocessing sequences of past weather observations using the data assimilation techniques developed to initiate numerical weather forecasts (Trenberth and Olson, 1988; Bengtsson and Shukla, 1988). *In situ* and satellite-based measurements of atmospheric properties for a particular analysis time are used in a statistically optimal way to correct a short-term forecast from the preceding analysis time. The numerical forecast model carries forward in time and spreads in space the information from earlier observations. Sets of observations that differ in accuracy and spatial and temporal coverage are thereby blended into a regular set of gridded products suitable for a wide range of applications.

The first reanalysis products were completed in the mid-1990s at the European Centre for Medium-Range Weather Forecasts (ECMWF) and the U.S. National Center for Environmental Prediction (NCEP). ECMWF reanalysed weather observations over the fifteen-year period from 1979 to early 1994 (Gibson et al., 1997). Output from this project is commonly referred to as ERA-15 (ECMWF Re-Analysis). NCEP and the National Center for Atmospheric Research (NCAR) jointly produced a second reanalysis product (NCEP-50) spanning the 50+ years from 1948 to present (Kalnay et al., 1996; Kistler et al., 2001). ERA-15 and NCEP-50 have been used for such diverse purposes as climate model evaluation (Gates et al., 1999), investigation of subseasonal and interannual variability (Sperber, 2003; Annamalai et al., 1999; AchutaRao and Sperber, 2002) analysis of changes in extreme events (Kharin and Zwiers, 2000), and climate change detection studies (Gillett et al., 2003; Santer et al., 2003a).

The present paper focusses on the use of reanalyses for identification of human effects on climate. Reanalysis products have a number of advantages and disadvantages for this specific purpose. Consider the advantages first:

- Reanalyses provide internally-consistent estimates of changes in climate: they are uncontaminated by the changes in model physics and resolution that typically affect the secular behavior of operational analyses (Trenberth and Olson, 1988; Basist and Chelliah, 1997).
- Reanalyses offer spatially-complete information for many different atmospheric variables. This facilitates the application of pattern-based “fingerprint” detection studies (Hasselmann, 1979; Barnett and Schlesinger, 1987; Santer et al., 1995; Hegerl et al., 1996; Allen and Tett, 1999; Stott et al., 2000)¹, particularly for poorly-observed variables.
- The existence of multiple reanalyses, generated with different numerical weather prediction models and data assimilation methods, enables assessment of the sensitivity of detection results to current uncertainties in reanalysis-based estimates of climate change (Santer et al., 2003a).

These advantages must be weighed against several deficiencies:

- Climate data from reanalyses, especially the “first generation” ERA-15 and NCEP-50 reanalyses, exhibit inhomogeneities related to temporal changes in

¹Sampling complex, time-evolving patterns of climate change with *in situ* observational networks that are spatially incomplete and vary over time can introduce non-trivial biases in global estimates of surface and atmospheric temperature changes (Duffy et al., 2001; Santer et al., 2000a).

the distribution, availability, and quality of assimilated satellite and radiosonde information (see, e.g., Kållberg et al., 1997; Basist and Chelliah, 1997; Pawson and Fiorino, 1998; Santer et al., 1999; Trenberth et al., 2001). Observing system changes introduce spurious non-climatic variability that is difficult to separate unambiguously from the true low-frequency climate changes that are of interest in detection work. One example is the transition to the widespread assimilation of satellite-based temperature retrievals in NCEP-50, which induces step-wise changes in such quantities as lower stratospheric temperature (Santer et al., 1999) and the variability of the tropical Quasi-Biennial Oscillation (QBO; Pawson and Fiorino, 1999). Inhomogeneities related to observing system changes are not restricted to the pre-satellite era (Trenberth et al., 2001; Santer et al., 2004).

- The climate data output from reanalyses are also sensitive to a number of specific technical choices. These are related to the physics and resolution of the selected numerical model, the properties and implementation² of the data assimilation system, and the procedures used to adjust for biases in the assimilated data (Harris and Kelly, 2001).

This list of advantages and disadvantages indicates that reanalyses should not be used uncritically in climate-change detection work; nor should they be entirely discounted for such studies. Our perspective is that reanalyses provide a valuable tool for

²This encompasses decisions on the nature of the observational data streams that are actually assimilated (e.g., whether temperature information is assimilated in the form of cloud-cleared radiances or retrievals), the relative weights assigned to different types of observations, and whether assimilated data are handled in a univariate or multivariate way (Dethof and Hólm, 2002).

exploring the robustness of “fingerprint” detection results to plausible uncertainties in current estimates of decadal-timescale climate change. However, some of the deficiencies noted above have raised concerns regarding the reliability of detection results that are based on first-generation reanalysis products. A case in point is the study by Santer et al. (2003a) (henceforth SAN03), which identified a model-predicted fingerprint of combined anthropogenic and natural influences in the tropopause height changes estimated from ERA-15 and NCEP-50. Criticism of this paper by Trenberth (personal communication) has suggested that first-generation reanalyses are of insufficient quality for identifying anthropogenically-induced increases in tropopause height. A related comment on SAN03 by Pielke and Chase (2004) contends that NCEP-50 provides highly-reliable estimates of tropospheric temperature change. Pielke and Chase note that the temperature changes driving tropopause height increases are quite different in NCEP-50 and in the climate model simulations analysed by SAN03³, and argue that this discrepancy invalidates the SAN03 detection results.

Here, we address these twin criticisms using data from the second-generation ERA-40 reanalysis (Simmons and Gibson, 2000). ERA-40 uses a refined numerical model with an advanced data assimilation system (Andersson et al., 1998; Simmons and Hollingsworth, 2002) to produce new estimates of atmospheric variability over the past 45 years. The ERA-40 project was completed by ECMWF in 2003, and spans the period from September 1957 to August 2002. We employ climate data from ERA-40 to address the following scientific questions:

1. Does the use of tropopause height changes inferred from ERA-40 confirm or negate detection results previously obtained by SAN03 with the “first-generation”

³This point was noted earlier and assessed by Santer et al. (2003b) and SAN03.

ERA-15 and NCEP-50 reanalyses?

2. Are the tropopause height increases in climate model and reanalysis data driven by similar layer-average changes in stratospheric and tropospheric temperature?
3. How do the layer-average atmospheric temperature changes in first- and second-generation reanalyses compare with temperature changes estimated from Microwave Sounding Units (MSUs) flown on polar orbiting satellites? Do such comparisons illustrate evolutionary improvements in reanalysis skill?
4. Are decadal-timescale changes in p_{LRT} , the pressure of the lapse-rate tropopause, sensitive to the vertical resolution of the input atmospheric temperature data used for calculating p_{LRT} ?

The structure of this paper is as follows. In Section 2, we provide a brief introduction to the reanalysis datasets employed in our detection study, and to the observational satellite datasets used in addressing questions 2 and 3 above. Section 2 also introduces the Parallel Climate Model (PCM; Washington et al., 2000) with which we define the expected spatial pattern of tropopause height change due to anthropogenic effects. Methods applied for calculation of the tropopause pressure and ‘synthetic’ MSU temperatures from reanalysis and climate model data are outlined in Section 3. Particular attention is devoted to the issue of the sensitivity of estimated tropopause pressure changes to the vertical resolution at which the calculation is performed. Changes in global means and spatial patterns of tropopause pressure are discussed in Section 4. Section 5 describes results from the revisit of the SAN03 tropopause height detection analysis with ERA-40 data. In Section 6, we use temporal and spatial correlations to compare the processed and synthetic MSU

temperatures in four different datasets. Conclusions and answers to the four scientific questions posed above are given in Section 7.

2 Reanalysis, satellite, and model data

2.1 Reanalysis data

Full descriptions of the ERA-15, NCEP-50, and ERA-40 reanalysis projects are given by Gibson et al. (1997), Kalnay et al. (1996), Kistler et al. (2001), Simmons and Gibson (2000), and the reanalysis websites.⁴ Our aim here is to highlight some of the principal differences between ERA-40 and earlier reanalyses. The first key difference is that ERA-40 has higher horizontal and vertical resolution. The operational numerical weather prediction (NWP) models in NCEP-50 and ERA-15 were run at T62 and T106 spectral truncation (respectively), while ERA-40 was run at T159 spectral truncation with a grid resolution of roughly 125 km. ERA-40 employs a hybrid sigma-pressure coordinate system with 60 vertical levels; the top model level is at 0.1 hPa (ca. 65 km). Both NCEP-50 (28 levels) and ERA-15 (31 levels) had substantially fewer levels in the vertical than ERA-40, lower top levels (ca. 3 hPa for NCEP-50 and 10 hPa for ERA-15), and a less extensive representation of the stratosphere. The enhanced vertical resolution of ERA-40 is useful for exploring the sensitivity of estimated p_{LRT} changes to the number of model levels used in calculating p_{LRT} (see Section 3.1).

⁴<http://wesley.wwb.noaa.gov> (NCEP-50), <http://www.ecmwf.int/research/era/ERA-15> (ERA-15), and <http://www.ecmwf.int/research/era> (ERA-40).

There are also important differences in the data assimilation systems. ERA-40 directly incorporates raw satellite radiances through a three-dimensional variational data assimilation system (3D-Var) (Andersson et al., 1998). The 3D-Var scheme is global, multivariate, and nonlinear. Implementation and subsequent refinement of the variational assimilation scheme in ECMWF operations has yielded pronounced increases in analysis accuracy and forecast skill (Andersson et al., 1998; Simmons and Hollingsworth, 2002). ERA-15 used a 1D-Var scheme to assimilate a less-extensive set of satellite radiances that had undergone many pre-processing steps.

Like ERA-40, NCEP-50 also used a 3D-Var scheme (Kalnay et al., 1996), but satellite data were assimilated through temperature retrievals rather than radiances. The retrievals were generated by the National Environmental Satellite Data and Information Service (NESDIS), and incorporate information from several different satellite-borne sensors.⁵ The NESDIS retrievals have documented biases in the temperature of the tropical stratosphere (Mo et al., 1995) and in the temperature and static stability of the troposphere (Kelly et al., 1991). These biases, together with changes in the retrieval algorithms themselves, can induce spurious temporal variability (Basist and Chelliah, 1997; Santer et al., 2004). The assimilation of satellite data is therefore fundamentally different in the three reanalyses, as are procedures for bias correction of satellite and radiosonde information (Gibson et al., 1997; Kalnay et al., 1996; Harris and Kelly, 2001; Andrae et al., 2004).

ERA-40 atmospheric temperature data required for p_{LRT} and equivalent MSU calculations were available on the model Gaussian grid at each of the 60 full model

⁵MSU, the Stratospheric Sounding Unit (SSU), and the High-Resolution Infrared Radiation Sounder (HIRS).

levels. ECMWF also interpolated model-level fields to 23 discrete pressure levels and archived temperature data at this reduced vertical resolution (Källberg et al., 2004). We made use of both the 60- and 23-level temperature datasets for computing p_{LRT} , but calculated synthetic MSU temperatures with 23-level data only. Most calculations employed monthly-mean data for the period January 1958 through December 2001, although some six-hourly data were used to test the sensitivity of p_{LRT} changes to the temporal resolution of the input temperature data (see Section 3.1).

NCEP-50 atmospheric temperature data were interpolated from the T62 Gaussian grid and 28 model levels to a regular $2.5^\circ \times 2.5^\circ$ latitude-longitude grid and 17 discrete pressure levels (spanning 1000-10 hPa). NCEP-50 data were available in the form of monthly means for the period January 1948 through December 2001. ERA-15 data were not used for the present study given the relatively short duration of this reanalysis.

2.2 Satellite data

To evaluate whether atmospheric temperature changes in ERA-40 are more reliable than those in NCEP-50 (and hence whether tropopause height detection times estimated from ERA-40 temperature data are more credible), we compare synthetic MSU temperatures calculated from both reanalyses with the MSU temperatures processed by Mears et al. (2003) of Remote Sensing Systems (RSS) and by Christy et al. (2003) at the University of Alabama in Huntsville (UAH). Our focus is on MSU channels 4 and 2, which provide information on layer-average stratospheric and tropospheric

temperatures.⁶ We refer to these temperatures as T4 and T2, respectively. We use the most recent versions of the RSS and UAH T4 and T2 data (versions 1.2 for RSS and 5.1 for UAH). Both datasets were available in the form of monthly means on a regular $2.5^\circ \times 2.5^\circ$ latitude-longitude grid, and span the period from January 1979 through December 2003.

RSS and UAH use different procedures to adjust the raw MSU radiances for intersatellite biases, uncertainties in instrument calibration coefficients, changes in instrument body temperature, and drift in sampling of the diurnal cycle (Mears et al., 2003; Christy et al., 2003). These processing differences lead to divergent estimates of T2 changes over 1979-2001: the troposphere warms by $0.09^\circ\text{C}/\text{decade}$ in RSS, while the T2 trend in UAH is close to zero (Table 1). Discrepancies between the RSS and UAH T2 changes strongly influence the detectability of model-predicted T2 fingerprints (Santer et al., 2003c).

2.3 Climate model data

Both our original SAN03 tropopause height detection study and the present work employ data from the Department of Energy Parallel Climate Model (PCM) developed by NCAR and Los Alamos National Laboratory (Washington et al., 2000). In addition to climate change detection work (Santer et al., 2003c), PCM has been used for a wide range of applications, including studies of forced changes in decadal variability (Meehl et al., 2000), factors affecting the amplitude of simulated ENSO variability (Meehl et al., 2001), the climate response to volcanic forcing (Ammann et al., 2003;

⁶The maxima of the weighting functions for MSU channels 4 and 2 are at roughly 74 and 595 hPa.

Wigley et al., 2004) and the differential responses to solar and greenhouse-gas forcing (Meehl et al., 2003).

We analyze two PCM experiments here. The first (ANTHRO) involves combined changes in three anthropogenic forcings (well-mixed greenhouse gases, the direct scattering effects of sulphate aerosols, and tropospheric and stratospheric ozone). The second (ALL) additionally includes the effects of changes in solar irradiance and volcanic aerosols. Our earlier study considered only the ALL experiment. Here, we use ANTHRO for detection purposes, while ALL is more relevant for direct visual comparison with observations. ALL commences in 1890, while ANTHRO starts in 1872. Both end in 1999. Four realizations of each experiment were performed. Further details of the model and imposed forcing changes are given in APPENDIX A.

Our fingerprint study requires model-based estimates of internally-generated climate noise for assessing statistical significance (APPENDIX B). These were obtained from two 300-year control integrations performed with PCM and the ECHAM4/OPYC model (“ECHAM”). Technical details of the ECHAM model are provided in Roeckner et al. (1999) and in APPENDIX A. Previous work with ECHAM has shown that tropopause height increases in anthropogenic climate-change experiments are large and readily identifiable relative to the unforced variability of p_{LRT} in the model control run (Sausen and Santer, 2003; Santer et al., 2003b).

3 Calculation of tropopause height and synthetic MSU temperatures

3.1 Tropopause height

We diagnose changes in p_{LRT} from reanalysis and model data by interpolation of the lapse rate in a p^κ coordinate system, where p denotes pressure, $\kappa = R/c_p$, and R and c_p are the gas constant for dry air and the specific heat capacity of dry air at constant pressure (Reichler et al., 2003). The algorithm identifies the threshold model level at which the lapse rate falls below $2^\circ\text{C}/\text{km}$, and then remains less than this critical value for a specified vertical distance (World Meteorological Organization, 1957). The exact pressure at which the lapse rate attains the critical value is determined by linear interpolation of lapse rates in the layers immediately above and below the threshold level. This definition of tropopause height is robust under most conditions. Exceptions include situations where the atmosphere is relatively isothermal or where multiple stable layers are present (Reichler et al., 2003).⁷ The algorithm is applied in a consistent way to monthly-mean profiles of atmospheric temperature in PCM, ECHAM, ERA-40, and NCEP-50.

Our previous work (Santer et al., 2003b) showed that calculations performed with the 28- and 17-level temperature data from NCEP-50 yielded similar decadal-timescale changes in global- and tropical-mean p_{LRT} . Concerns remain, however, regarding the reliability of p_{LRT} changes estimated from temperature data with coarse

⁷To avoid unrealistically high or low p_{LRT} values, search limits are restricted to pressure levels between roughly 600 and 75 hPa; the search proceeds upwards from 600 hPa.

vertical resolution (Ramaswamy et al., 2001). To address these concerns, p_{LRT} trends in ERA-40 were calculated from temperatures archived at the reduced set of 23 pressure levels and the full 60 model levels (Section 2.1).⁸ The latter dataset has higher vertical resolution in the vicinity of the tropopause (see Källberg et al., 2004).

The 60-level (‘L60’) and 23-level (‘L23’) calculations yield similar estimates of global-mean p_{LRT} changes in ERA-40, with linear trends of -2.36 hPa/decade and -2.66 hPa/decade (respectively) over 1979-2001 (Fig. 1a; Table 1).⁹ Both trends are significantly different from zero when temporal autocorrelation effects are properly accounted for (Santer et al., 2000b). The spatial fields of p_{LRT} trends over this 23-year period are also highly similar in the two calculations (Figs. 2a,b), with a pattern correlation of $r_{\text{L60:L23}} = 0.94$. This is an encouraging result, particularly for our pattern-based climate change detection work, since it illustrates that the large-scale pattern of recent tropopause height change is relatively insensitive to the vertical resolution of the temperature data used in p_{LRT} calculations (at least in ERA-40). The most pronounced pattern differences are in the tropics, where the L23 results have small but spatially-coherent decreases in p_{LRT} , while small p_{LRT} trends of both sign occur in the L60 case (Fig. 2). These slight differences are probably related to the improved representation of the tropical tropopause in the L60 case.

We note that the primary detection conclusions described in Section 5 are insensitive to our choice of L23 or L60 p_{LRT} data. For consistency with p_{LRT} calculations

⁸Model-level pressures near the tropopause depend only weakly on surface pressure; they were calculated using monthly-mean values of surface pressure and the vertical coordinate definition specified by Källberg et al. (2004).

⁹A decrease in p_{LRT} signifies an increase in tropopause height.

involving the (low vertical resolution) NCEP-50, PCM, and ECHAM data (see Table 1), all ERA-40 p_{LRT} results employed in our detection work are exclusively from L23 calculations.

Highwood et al. (2000) and Reichler et al. (2003) have pointed out that p_{LRT} is often difficult to define at high latitudes in the Southern Hemisphere (SH), particularly during SH winter. In the observation-sparse early years of ERA-40, this problem is compounded by the model’s wintertime stratospheric cold bias in polar regions of the SH, which can lead to an unrealistically high lapse-rate tropopause at individual Antarctic grid points (see, e.g., the August result in Fig. 3). The ‘high tropopause’ problem becomes less severe in the satellite era, when improved observational data constraints are introduced. The time-varying nature of the problem results in an annual cycle with spuriously large p_{LRT} anomalies in SH winter in the initial two decades of the reanalysis (Fig. 1a). For this reason, data poleward of 60°S were excluded from the tropopause height fingerprint analysis (Section 5) and from all subsequent calculations of spatially-averaged p_{LRT} changes. This removes the spurious annual cycle (c.f. Figs. 1a,b), thereby reducing the variance of the time series and the standard error of the p_{LRT} trend (Table 1). It also decreases differences between the L23 and L60 results early in the period. Averaging over 90°N - 60°S has the additional effect of decreasing the global-mean p_{LRT} trends themselves, since large p_{LRT} decreases occur poleward of 60°S (Fig. 2).

Finally, we examined the sensitivity of ERA-40’s decadal-timescale p_{LRT} trends to the temporal resolution of the input temperature data. As is the case with data from radiosondes (Highwood and Hoskins, 1998) and NCEP-50 (Santer et al., 2003b), ERA-40 p_{LRT} trends computed from monthly-mean temperature data are similar to

trends calculated from six-hourly data. This justifies our use of monthly-mean data for determining tropopause height changes.

3.2 Synthetic MSU temperatures

We use a static global-mean weighting function to compute synthetic MSU T4 and T2 temperatures from both climate model and reanalysis data (Santer et al., 1999). This procedure facilitates the ‘like with like’ comparison of synthetic MSU temperatures with the MSU data processed by RSS and UAH. The appropriate weighting function is applied to grid-point profiles of monthly-mean pressure-level temperatures in PCM, ERA-40, and NCEP-50. For global and hemispheric means, this approach yields results similar to those obtained with a complex radiative transfer code (Santer et al., 1999).

Locally, surface emissivity effects and large temporal changes in atmospheric moisture can yield differences between the equivalent T2 temperatures estimated with the static weighting function and full radiative transfer approaches. This is not a significant problem for our comparison of models and reanalyses, since we use a consistent approach for calculating synthetic MSU temperatures from these two types of data. However, comparisons of the synthetic and processed MSU T2 data should be made with caution over elevated terrain, particularly over the ice-covered surface of Antarctica.

4 Tropopause height changes in reanalyses and PCM

4.1 Global-mean changes

The height of the tropopause shows a sustained multidecadal increase since the early 1960s (Fig. 4). This overall increase is evident in ERA-40, NCEP-50, and the PCM ALL experiment.¹⁰ Superimposed on this increase are short-term height decreases in response to explosive volcanic eruptions. The p_{LRT} changes after the eruptions of Mt. Agung (1963), El Chichón (1982), and Pinatubo (1991) are invariably larger in ALL than in either reanalysis, primarily due to the excessive stratospheric warming responses in PCM (Fig. 5). This is a common deficiency in models with coarse vertical resolution in the stratosphere (see, e.g., Bengtsson et al., 1999).

Another relevant factor in the comparison of volcanic p_{LRT} responses is that ALL includes estimates of volcanic aerosol forcing and explicitly considers the aerosol’s radiative effects (Ammann et al., 2003), while NCEP-50 and ERA-40 do not incorporate observed estimates of volcanic aerosol properties. In both reanalyses, information on the climate signatures of volcanic eruptions is obtained indirectly through the assimilated satellite and *in situ* data. During eruptions in the pre-satellite era, such as that of Agung in 1963, the sparse coverage of available radiosonde data may bias reanalysis-based estimates of volcanically-induced climate signals, thus contributing to p_{LRT} differences between reanalyses and PCM.

¹⁰Note that ERA-40 diverges markedly from NCEP-50 prior to roughly 1975, during the period when observational coverage is relatively sparse and does not provide as strong a constraint on the reanalyses.

ERA-40 shows a pronounced lower stratospheric warming in 1975 (Fig. 5). Although a slight warming may have taken place in reality due to the eruption of Mt. Fuego in October 1974, the stratospheric warming in ERA-40 stems primarily from an error in the bias correction of radiances from the Vertical Temperature Profiler Radiometer (VTPR) on the NOAA-4 satellite, which were assimilated during 1975 and the first half of 1976. Bias correction coefficients computed for VTPR data from the NOAA-3 satellite were inadvertently applied in adjusting NOAA-4 VTPR data. The error had largest impact in the Southern Hemisphere stratosphere, but is also evident in the global mean of the synthetic T2 data for ERA-40 (see Fig. 9 below).

Figure 6 provides a simple conceptual model for interpreting the low- and high-frequency p_{LRT} changes shown in Fig. 5 (see also Highwood et al., 2000, who use a similar conceptual model). Calculations performed with radiative-convective models and more complex atmospheric GCMs illustrate that stratospheric cooling and tropospheric warming are robust signals of increases in atmospheric CO₂ (e.g., Hansen et al., 1984, 2002; Manabe and Wetherald, 1987; Ramaswamy et al., 1996, 2001). These temperature changes tend to increase tropopause height. Anthropogenically-induced depletion of stratospheric ozone also causes a net increase in tropopause height through strong cooling of the stratosphere.¹¹ Alternately, volcanic aerosols injected into the stratosphere absorb incoming solar radiation and outgoing longwave radiation, thus warming the stratosphere and cooling the troposphere. Both of these changes decrease tropopause height (Fig. 6).

¹¹Depletion of stratospheric ozone cools the stratosphere and the troposphere. These changes have effects of opposite sign on tropopause height. The stratospheric cooling influence predominates, so the net effect of ozone depletion is to raise tropopause height. In PCM, tropospheric ozone increases warm the troposphere and also contribute to tropopause height increases.

The actual temperature perturbations associated with these three forcings are more complex as a function of latitude and altitude than the idealized changes illustrated in Fig. 6 (e.g., Bengtsson et al., 1999; Hansen et al., 2002; Santer et al., 2003b). We note, however, that the global-scale p_{LRT} changes in PCM, NCEP-50, and ERA-40 are qualitatively consistent with this simple conceptual model.

Our quantitative comparisons of p_{LRT} changes in ERA-40 and NCEP-50 focus on 1979-2001. This period is characterized by a relatively stable observing system, and by higher quantity and quality of assimilated observations. PCM ALL results are given for the period 1979-1999.¹² In ERA-40, p_{LRT} decreases by 2.12 hPa/decade over 1979-2001 in the L23 calculation, corresponding to an overall increase in global-mean tropopause height of roughly 200 meters. The p_{LRT} decrease of 1.79 hPa/decade in NCEP-50 corresponds to a height increase of approximately 170 meters. Both trends are significantly different from zero at the 1% level (Table 1), and are consistent with height increases inferred directly from radiosondes (Highwood et al., 2000; Seidel et al., 2001). The PCM trend of -1.13 hPa/decade over 1979-1999 is smaller than in either reanalysis, in part due to the model's overestimated stratospheric warming in response to El Chichón and Pinatubo (see Figs. 5,6).

4.2 Spatial patterns

Despite their use of very different assimilation systems, input satellite data, and bias correction schemes, ERA-40 and NCEP-50 have striking similarities in their spatial patterns of tropopause height change (Figs. 7a-d). Both show increases in height

¹²Recall that the ALL experiment ends in 1999.

over most of the globe. These increases are small in the tropics, and are largest at high latitudes in the Southern Hemisphere. It is notable that pattern similarities are not restricted to 1979-2001 (Figs. 7c,d), but are also evident over the longer (and observationally less well-constrained) 1958-2001 period, particularly between 30°N-60°N, where radiosonde coverage is relatively dense (Figs. 7a,b). The largest differences between ERA-40 and NCEP-50 are poleward of 45°S, where radiosonde coverage is poor; height increases here are more coherent in ERA-40 than in NCEP-50 (*c.f.* Figs. 7a,b and 7c,d). There are also prominent differences off the coast of California (Figs. 7c,d).

The large-scale patterns of p_{LRT} change over 1979-1999 in the PCM ALL and ANTHRO experiments are qualitatively similar to those in the two reanalyses, with coherent height increases over most of the globe (*c.f.* Figs. 7e,f and 7c,d). As in ERA-40 and NCEP-50, increases are small and relatively unstructured in the tropics, and largest poleward of 45°S. In both model and reanalysis results, p_{LRT} changes tend to be noisy at the transition from the tropical to the extratropical tropopause. The similarity between the spatial fields of p_{LRT} change in the ALL and ANTHRO experiments (Figs. 7e,f) arises because both patterns are driven primarily by anthropogenic forcing, at least in PCM (see SAN03 and Section 5).

5 Fingerprint detection results

We next used the ensemble-mean p_{LRT} changes from the PCM ANTHRO experiment to define the expected signal in response to anthropogenic forcing. This is referred to here as the “fingerprint” pattern, \vec{f} . We applied a standard method to search for an

increasing expression of \vec{f} in ERA-40 and NCEP-50 p_{LRT} data, and to estimate the ‘detection time’ – the time at which \vec{f} becomes consistently identifiable at a stipulated 5% significance level (Hasselmann, 1979; Santer et al., 1995; SAN03). Details of the method are given in APPENDIX B. We consider the sensitivity of our detection results to different processing options. Detection times are a function of:

- Reanalysis dataset (ERA-40 or NCEP-50).
- Fingerprint pattern. We use either the ‘raw’ fingerprint, \vec{f} , or the optimized fingerprint, \vec{f}^* . The latter is rotated away from high-noise directions in an attempt to enhance signal-to-noise ratios and fingerprint detectability.
- The model control run (ECHAM or PCM) used for optimizing \vec{f} and assessing statistical significance.
- Treatment of spatial-mean p_{LRT} changes (spatial mean included or removed). Removal of the spatial mean ensures that positive detection results cannot be driven solely by large mean changes, and focusses attention on the correspondence between sub-global aspects of p_{LRT} changes in PCM and reanalyses (see, e.g., Hegerl et al., 1996; Santer et al., 2003c).

As noted previously, large, abrupt changes in the availability of satellite-based data on atmospheric temperature, moisture, and winds can introduce non-climatic variability in reanalysis products (Section 2.1). To minimize the impacts of such spurious variability on our detection study, we use only post-1978 reanalysis data. Additionally, restricting our attention to a 90°N-60°S spatial domain largely eliminates the twin problems of poorly-defined and unrealistically high p_{LRT} values over

the Antarctic continent (Section 3.1). Finally, we note that although SAN03 also used NCEP-50 p_{LRT} data, our detection analysis differs from theirs in its spatial domain and in its use of the PCM ANTHRO experiment to define the searched-for tropopause height fingerprint.¹³

Our detection results support the conclusions that SAN03 obtained with p_{LRT} data from first-generation reanalyses, and confirm that there has been an identifiable human influence on tropopause height over the past several decades. This finding is insensitive to statistical analysis details (Fig. 8). For each reanalysis dataset, there are 8 possible detection time estimates. The ANTHRO p_{LRT} fingerprint can be successfully identified in all 8 cases involving ERA-40 data, and in 7 of 8 cases that use NCEP-50 data. These results reflect similarities between the large-scale patterns of p_{LRT} change in ERA-40, NCEP-50, and PCM, such as their common spatial coherence and hemispheric asymmetry (c.f. Figs. 7c-f).

Positive detection of \vec{f} is not due solely to the large global-mean height increases in PCM and reanalyses (Fig. 4). This is evident when spatial mean p_{LRT} changes are removed, and the smaller-scale hemispheric asymmetry component of the fingerprint is emphasized. The ‘mean removed’ version of \vec{f} is identifiable 6-8 years earlier in ERA-40 than in NCEP-50, in part because the large height increases poleward of 45°S are more coherent in ERA-40 than in NCEP-50 (c.f. Figs. 7c,d), and are more similar to p_{LRT} changes in PCM. Removing the mean has little effect on detection times for ERA-40, but significantly degrades detection times for NCEP-50.

¹³The spatial domain in SAN03 was 85°N-85°S, and they relied on the PCM ALL experiment to define \vec{f} .

As noted above, our previous detection work (SAN03) employed the PCM ALL integration (rather than ANTHRO) to define \vec{f} . The p_{LRT} fingerprint is very similar in ANTHRO and ALL, reflecting the overall increase in tropopause height that is common to both experiments. In PCM, this increase is mainly driven by changes in well-mixed greenhouse gases and stratospheric ozone (included in both ANTHRO and ALL), and not by changes in solar irradiance and/or volcanic aerosols (included in ALL only). This is why use of the ANTHRO and ALL p_{LRT} fingerprints yields similar detection results. Positive identification of the ANTHRO fingerprint confirms that we are primarily identifying anthropogenic effects.

6 Analysis of synthetic MSU temperatures

6.1 Synthetic MSU temperatures in ERA-40 and PCM

As discussed above, both stratospheric cooling and tropospheric warming tend to increase tropopause height (Fig. 6; Section 4.1). In SAN03, we found that the height increase in NCEP-50 over 1979-2001 was driven by stratospheric cooling only, whereas recent height increases in PCM and ERA-15 were due to the combined effects of tropospheric warming and stratospheric cooling. We speculated that the tropopause height increase in NCEP-50 was partly the result of compensating errors, with excessive stratospheric cooling (related to the assimilation of biased temperature retrievals) offsetting the height decrease induced by a spurious cooling of NCEP's troposphere (Santer et al., 2004). It is important, therefore, to determine whether our positive identification of the PCM tropopause height fingerprint in ERA-40 arises from partly

compensating errors (as in the NCEP-50 case) or from real similarities in model and reanalysis profiles of atmospheric temperature change.

We address this issue by comparing synthetic MSU temperature trends in ERA-40 and PCM ALL. Over 1979-2001, the stratosphere cools by $0.30^{\circ}\text{C}/\text{decade}$ in ERA-40; its troposphere warms by $0.08^{\circ}\text{C}/\text{decade}$ (Table 1; Figs. 5,9). These results are similar to (and statistically consistent with) T4 and T2 trends in the PCM ALL experiment. This confirms that recent tropopause height increases in PCM and ERA-40 are being driven by similar global-scale atmospheric temperature changes both above and below the tropopause.

It is also instructive to compare the low-frequency variability of T2 changes in ERA-40 and the ALL experiment (Fig. 10). The four realizations of ALL represent four different manifestations of natural internal variability, each superimposed on the underlying climate response to combined anthropogenic and natural forcings. The ALL realizations define an ‘envelope’ of possible changes in tropospheric temperature. The low-frequency T2 changes in ERA-40 are generally contained within this envelope.¹⁴ Even aspects of the sub-decadal variability of T2 in ERA-40, such as the cooling responses to the Agung, El Chichón, and Pinatubo eruptions, are well-reproduced by the ensemble-mean T2 changes in ALL. This constitutes a more stringent test of PCM’s performance than comparison of trends alone.

¹⁴During times of major El Niño or La Niña events, high-frequency T2 changes in ERA-40 are often outside PCM’s envelope of inter-realization variability (Fig. 10). This is because the phasing of El Niño and La Niña events is not the same in the real world and in a coupled model experiment, except by chance. ERA-40 is also outside the PCM variability envelope for most of 1975, when it was affected by the VTPR bias correction error noted earlier (Section 4.1).

6.2 Comparison with processed MSU temperatures

To evaluate whether T4 and T2 changes in ERA-40 are more reliable than those in NCEP-50 (and hence whether tropopause height detection times estimated from ERA-40 are more credible), we compare synthetic MSU temperatures calculated from both reanalyses with the RSS and UAH MSU temperatures processed by Mears et al. (2003) and Christy et al. (2003) (Section 2.2). We analyze both temporal correlations between time series of global-mean temperature changes, and spatial correlations between patterns of temperature trends.

6.2.1 Temporal correlations

Despite fundamental differences in how Mears et al. (2003) and Christy et al. (2003) process raw MSU T4 and T2 radiances, global-mean atmospheric temperature changes in RSS and UAH are more highly correlated with each other than with the synthetic MSU changes in either reanalysis (Table 2). This conclusion holds for both T2 and T4 changes, and for correlations calculated with and without the overall linear trend (which emphasize low- and high-frequency components of the time series, respectively).

Another general result is that atmospheric temperature changes in the two observational MSU datasets correlate more highly with ERA-40 than with NCEP-50. In fact, ERA-40 invariably correlates better with RSS and UAH than it does with NCEP-50, and the lowest ‘between-dataset’ correlations always involve NCEP data (Table 2). This is probably due to spuriously large cooling in NCEP’s lower stratospheric temperatures (Fig. 5; Table 1), which is introduced both by the assimilation

of biased NESDIS temperature retrievals (Section 2.2) and by the transition from MSU to the Advanced MSU instrument in the late 1990s (Santer et al., 2004). Because of the non-negligible stratospheric contribution to T2 (Fu et al., 2004), NCEP’s excessive stratospheric cooling ‘leaks’ into its synthetic T2 temperature, contributing to the unrealistically large negative T2 trend in this dataset ($-0.11^{\circ}\text{C}/\text{decade}$; Fig. 9; Table 1).

6.2.2 Spatial correlations

Patterns of linear trends in T2 are qualitatively similar in ERA-40, RSS, and UAH (Figs. 11a-c). All show coherent warming over most of the Northern Hemisphere and cooling over the central Pacific and northern Siberia. Tropospheric temperature trends in these three datasets differ poleward of 45°S , where UAH cools markedly, RSS cools moderately, and ERA-40 has no net cooling. These differences are not fully understood, although differences in the treatment of surface emissivity effects over snow- and ice-covered surfaces are likely to be a contributory factor. The large-scale patterns of stratospheric cooling are similar in ERA-40, RSS, and UAH, with maximum cooling at high latitudes in the Southern Hemisphere, and cooling minima (or even slight warming) over the central Pacific, Alaska, and the South Indian Basin and Ross Sea (Figs. 12a-c). NCEP’s T2 and T4 changes (Figs. 11d, 12d) are distinctly different from those in ERA-40 and the two satellite datasets, with more coherent tropospheric cooling, and stronger cooling of the tropical and subtropical stratosphere. Possible reasons for this behavior were discussed in Section 6.2.1. PCM’s patterns of T2 and T4 trends (Figs. 11e, 12e) are more similar to those in ERA-40, RSS and UAH than to NCEP’s trend patterns.

Pattern correlations help to quantify these comparisons (Table 3). Correlations are calculated both with and without inclusion of the spatial means; these statistics are referred to as c and r , respectively (Barnett and Schlesinger, 1987). Removal of spatial means can reveal smaller-scale pattern similarities and/or differences that may be obscured by global-mean trend differences between two datasets. This is the case with T2 changes in the two reanalyses, for which $c_{\{\text{NCEP:ERA}\}} = -0.08$, while $r_{\{\text{NCEP:ERA}\}} = 0.74$. The lower value for c arises from large differences in the global-mean trends. In contrast, removal of spatial-mean T4 changes in NCEP-50 and ERA-40 degrades pattern similarity, pointing towards differences in the smaller-scale spatial structure of their stratospheric temperature trends ($c_{\{\text{NCEP:ERA}\}} = 0.86$, $r_{\{\text{NCEP:ERA}\}} = 0.41$; Figs. 12a,d).

For ‘observed’ (RSS and UAH) and synthetic MSU temperatures, correlations between the spatial patterns of trends yield three key results: 1) Despite fundamental differences in their satellite data adjustment procedures, atmospheric temperature changes in RSS and UAH are more similar to each other than to changes in either reanalysis dataset; 2) Temperature changes in the two reanalyses are more similar to changes in observed satellite data products than they are to each other; 3) Temperature changes in RSS and UAH are consistently more highly correlated with those in ERA-40 than with changes in NCEP-50. All three findings hold for both T2 and T4, and for ‘mean included’ and ‘mean removed’ correlations.

7 Conclusions

In Section 1, we posed four scientific questions. The first dealt with the robustness of the tropopause height detection results obtained by Santer et al. (2003; “SAN03”). SAN03 claimed that they could identify a model-predicted “fingerprint” of externally-forced tropopause height changes in the first-generation ERA-15 and NCEP-50 reanalyses. Deficiencies in these early reanalysis products prompted justifiable questions regarding the reliability of these detection claims (Trenberth, personal communication).

We addressed this criticism here by revisiting the SAN03 tropopause height detection study with the second-generation ERA-40 reanalysis (Simmons and Gibson, 2000), which differs from the earlier ERA-15 and NCEP-50 reanalyses in a number of important aspects (Section 2.1). The PCM fingerprint of anthropogenically-forced tropopause height changes was statistically identifiable in ERA-40 p_{LRT} data, confirming the conclusions of SAN03. The ERA-40 detection results were robust to a number of choices made in implementing and applying our fingerprint method (Section 5).

The second question focussed on the atmospheric temperature changes that influence tropopause height increases. Previous work showed that both stratospheric cooling and tropospheric warming can raise the height of the tropopause (Highwood et al., 2000; Santer et al., 2003b; Fig. 6). Although SAN03 identified the PCM ALL fingerprint in NCEP-50 p_{LRT} data, the temperature changes driving this positive result were very different: the troposphere warmed in ALL, but cooled markedly in NCEP-50 (Table 1). This discrepancy raised further concerns regarding the reliability of the SAN03 detection claims (Pielke and Chase, 2004).

Santer et al. (2004) speculated that the positive NCEP-50 detection results could be explained by error compensation, related to NCEP’s excessive stratospheric cooling (Section 6). ERA-40 provides some support for this interpretation. In ERA-40, as in PCM ALL, the troposphere warms and stratosphere cools over the last several decades, and both effects contribute to an increase in tropopause height. Unlike in the NCEP-50 case, overall tropopause height increases in PCM and ERA-40 are dictated by similar large-scale changes in atmospheric temperature.

The third question addressed the relative reliability of synthetic MSU temperatures in NCEP-50 and ERA-40, and hence the relative reliability of detection results based on these datasets. It considered whether processed satellite data, such as the RSS and UAH MSU products (Mears et al., 2003; Christy et al., 2003), can be used to evaluate the fidelity with which ERA-40 and NCEP-50 simulate changes in T4 and T2. One problem with such comparisons is that processed and synthetic MSU data are not strictly independent: both reanalyses assimilate MSU information, either in the form of radiances (ERA-40) or MSU-based temperature retrievals (NCEP-50).

Although independence is a valid concern (and one that is difficult to address without systematic observing system experiments), we note that the estimates of decadal-timescale T2 and T4 changes are generated in fundamentally different ways in reanalyses and processed satellite data. ERA-40 and NCEP-50 rely on bias correction procedures (Kalnay et al., 1996; Harris and Kelly 2001) and the assimilation system itself to correct for the satellite data problems that are identified and adjusted for by RSS and UAH. RSS and UAH make such adjustments in a univariate sense, using MSU radiance information only. In contrast, the multivariate assimilation procedures in ERA-40 and NCEP-50 seek to achieve physical consistency be-

tween different analysed variables, such as temperature and wind fields, and utilize multivariate observational information from radiosondes, aircraft, surface data, and a variety of satellite-based sensors. There are many factors, therefore, that can contribute to differences between the synthetic MSU temperatures in reanalyses and the MSU temperatures processed by RSS and UAH.

It is encouraging that current satellite-based estimates of T4 and T2 changes – despite the large uncertainties in these estimates (Mears et al., 2003; Christy et al., 2003) – invariably agree better with temperature changes in the second-generation ERA-40 reanalysis than with those in the earlier NCEP-50 reanalysis. This suggests that evolutionary improvements in reanalysis data assimilation systems have demonstrably improved the quality of estimated atmospheric temperature changes, thus answering our third question.

We note, however, that ERA-40 still manifests inhomogeneities, such as unrealistically large stratospheric warming in the mid-1970s related to an error in the bias correction of the NOAA VTPR radiances (Fig. 5). This does not affect our detection analysis, which uses post-1978 data. By restricting our attention to the post-1978 portion of ERA-40, we also reduce the impact of changes in the availability of *in situ* data.

Even the post-1978 period, however, is not completely devoid of data homogeneity problems. For example, comparisons with radiosonde data and with an AMIP simulation that employed the ERA-40 model indicate that ERA-40’s tropical temperatures at 100 hPa are biased cold in 1979 and the first half of the 1980s. This is due to difficulties in the assimilation of radiance data from the early TOVS (TIROS Operational Vertical Sounder) satellites, which probably resulted in an overestimation of

the warming trend in the upper tropical troposphere from 1979 to 2001, and a T4 cooling trend that is weaker than that of RSS and UAH in the tropics (Fig. 12). Minimizing the spurious climate signatures of inter-satellite biases and temporal changes in satellite data availability will remain a significant challenge, both for reanalyses and for groups that directly process satellite data (such as RSS and UAH).

The final question that we posed dealt with the sensitivity of estimated tropopause pressure changes to the vertical resolution of the temperature data used in calculating p_{LRT} . ERA-40 supplies an ideal test-bed for addressing this concern. Temperature data from ERA-40 were available at both high (L60) and low (L23) vertical resolution. The L60 and L23 calculations yielded similar global-mean p_{LRT} changes, and (more importantly for our fingerprint detection work) similar patterns of tropopause height increase (Figs. 1,2). Our detection conclusions are not sensitive to this source of uncertainty.¹⁵

In summary, our study has identified a model ‘fingerprint’ of anthropogenically-forced tropopause height changes in ERA-40 data, and indicates that p_{LRT} changes inferred from ERA-40 cannot be accounted for by natural variability alone. This confirms and improves upon an earlier result that SAN03 obtained with p_{LRT} changes inferred from NCEP-50, and shows that our tropopause height findings are robust to uncertainties in existing reanalysis products. Recent increases in tropopause height in ERA-40 and the PCM ALL experiment are occurring for the same reasons – large-scale stratospheric cooling and tropospheric warming. Our comparisons between ob-

¹⁵We also verified that the positive detection results obtained in SAN03 are not an artefact of the large p_{LRT} changes poleward of 60°S, where the lapse-rate tropopause is difficult to define and is influenced by model errors (Section 3.1).

served and reanalysis-derived estimates of atmospheric temperature change suggest that p_{LRT} detection results based on ERA-40 data are more reliable than those obtained with NCEP-50.

Acknowledgements

Work at Lawrence Livermore National Laboratory (LLNL) was performed under the auspices of the U.S. Dept. of Energy, Environmental Sciences Division, contract W-7405-ENG-48. The ERA-40 project was partially funded by the European Commission under contract EVK2-CT-1999-00027. ERA-40 received support from numerous sources: Fujitsu Ltd. (provision of computational resources), the many institutions that supplied observational datasets (NCAR and NCEP in particular), institutions that seconded staff to work at ECMWF on the project, and several partner organizations that took part in production monitoring and product validation. T.M.L. Wigley received support from the NOAA Office of Global Programs (“Climate Change Data and Detection”) grant no. NA87GP0105, and from the U.S. Dept. of Energy, grant no. DE-FG02-98ER62601. A portion of this study was supported by the U.S. Dept. of Energy, Office of Biological and Environmental Research, as part of its Climate Change Prediction Program. The MSU T2 and T4 data and static MSU weighting functions were provided by John Christy (Univ. of Alabama in Huntsville). Erich Roeckner (Max-Planck Institut für Meteorologie, Hamburg) supplied ECHAM control run data. We thank Myles Allen (Oxford University) for useful discussions and comments. Gary Strand (NCAR) provided assistance in running and processing the PCM ANTHRO simulations.

References

AchutaRao, K., and K. R. Sperber, 2002: Simulation of the El Niño Southern Oscillation: Results from the Coupled Model Intercomparison Project. *Clim. Dyn.*, **19**, 191-209.

Allen, M. R., and S. F. B. Tett, 1999: Checking for model consistency in optimal fingerprinting. *Clim. Dyn.*, **15**, 419-434.

Ammann, C., G. A. Meehl, W. M. Washington, and C. S. Zender, 2003: A monthly and latitudinally varying volcanic forcing dataset in simulations of 20th century climate. *Geophys. Res. Lett.*, **30**, doi:10.1029/2003/GL016875.

Andrae, U., N. Sokki, and K. Onogi, 2004: The radiosonde temperature bias corrections used in ERA-40. ERA-40 Project Report Series, Reading, U. K. (in preparation).

Annamalai, H., J. M. Slingo, K. R. Sperber, and K. Hodges, 1999: The mean evolution and variability of the Asian Summer Monsoon: Comparison of ECMWF and NCEP-NCAR reanalyses. *Mon. Weath. Rev.*, **127**, 1157-1186.

Andersson, E., and Coauthors, 1998: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). Part III: Experimental results. *Q. J. R. Meteorol. Soc.*, **124**, 1831-1860.

Barnett, T. P., and M. E. Schlesinger, 1987: Detecting changes in global climate induced by greenhouse gases. *J. Geophys. Res.*, **91**, 6659-6667.

- Basist, A. N., and M. Chelliah, 1997: Comparison of tropospheric temperatures derived from the NCEP/NCAR reanalysis, NCEP operational analysis, and the Microwave Sounding Unit. *Bull. Amer. Met. Soc.*, **7**, 1431-1447.
- Bengtsson, L., and J. Shukla, 1988: Integration of space and in situ observations to study global climate change. *Bull. Amer. Meteor. Soc.*, **69**, 1130-1143.
- Bengtsson, L., E. Roeckner, and M. Stendel, 1999: Why is the global warming proceeding much slower than expected? *J. Geophys. Res.*, **104**, 3865-3876.
- Christy, J. R., R. W. Spencer, W. B. Norris, and W. D. Braswell, 2003: Error estimates of version 5.0 of MSU-AMSU bulk atmospheric temperatures. *J. Atmos. Ocean. Tech.*, **20**, 613-629.
- Dai, A., T. M. L. Wigley, B. A. Boville, J. T. Kiehl, and L. E. Buja, 2001: Climates of the twentieth and twenty-first centuries simulated by the NCAR climate system model. *J. Clim.*, **14**, 486-519.
- Dethof, A., and E. Hólm, 2002: Ozone in ERA-40: 1991-1996. Technical Memorandum **377**, European Centre for Medium-Range Weather Forecasts, Reading, U.K., 37 pp.
- Duffy, P. B., C. Doutriaux, I. K. Fodor, and B. D. Santer, 2001: Effect of missing data on estimates of near-surface temperature change since 1900. *J. Clim.*, **14**, 2809-2814.
- Fu, Q., C. M. Johanson, S. G. Warren, and D. J. Seidel, 2004: Contribution of stratospheric cooling to satellite-inferred tropospheric temperature trends. *Nature*, **429**, 55-58.

- Gibson, J. K., and Coauthors, 1997: *ECMWF Re-Analysis Project Report Series. 1. ERA Description*. Reading, U. K., 66 pp.
- Gillett, N. P., F. W. Zwiers, A. J. Weaver, and P. A. Stott, 2003: Detection of human influence on sea-level pressure. *Nature* **422**, 292-294.
- Hansen, J., and Coauthors, 1984: *Climate Sensitivity: Analysis of feedback mechanisms*. In: *Climate Processes and Climate Sensitivity* (Eds. J. Hansen and T. Takahasi). Maurice Ewing Series 5, Geophysical Monograph 29, 130-163. American Geophysical Union, Washington D.C.
- Hansen, J., and Coauthors, 2002: Climate forcings in Goddard Institute for Space Studies SI2000 simulations. *J. Geophys. Res.*, **107**, doi:10.1029/2001JD001143.
- Harris, B. A., and G. A. Kelly, 2001: A satellite radiance bias correction scheme for data assimilation. *Q. J. R. Meteor. Soc.*, **127**, 1453-1468.
- Hasselmann, K., 1979: On the signal-to-noise problem in atmospheric response studies. *Meteorology of Tropical Oceans*, D. B. Shaw (Ed.), *Roy. Met. Soc. London*, 251-259.
- Hegerl, G. C., H. Storch, K. Hasselmann, B. D. Santer, U. Cubasch, and P. D. Jones, 1996: Detecting greenhouse-gas-induced climate change with an optimal fingerprint method. *J. Clim.*, **9**, 2281-2306.
- Highwood, E. J., and B. J. Hoskins, 1998: The tropical tropopause. *Q. J. R. Meteor. Soc.*, **124**, 1579-1604.

- Highwood, E. J., B. J. Hoskins, and P. Berrisford, 2000: Properties of the Arctic tropopause. *Q. J. R. Meteor. Soc.*, **126**, 1515-1532.
- Hoyt, D. V., and K. H. Schatten, 1993: A discussion of plausible solar irradiance variations, 1700-1992. *J. Geophys. Res.*, **98**, 18895-18906.
- Kållberg, P., 1997: *Aspects of the re-analysed climate, ECMWF Re-Analysis Project Report Series*, **2**, Reading, U. K., 89 pp.
- Kållberg, P., and Coauthors, 2004: *The ERA-40 Archive*. ERA-40 Project Report Series, Reading, U. K. (in preparation).
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437-471.
- Kelly, G., E. Andersson, A. Hollingsworth, P. Lönnberg, J. Pailleux, and Z. Zhang, 1991: Quality control of operational physical retrievals of satellite sounding data. *Mon. Weath. Rev.*, **119**, 1866-1880.
- Kharin, V. V., and F. W. Zwiers, 2000: Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere-ocean GCM. *J. Clim.*, **13**, 3760-3788.
- Kistler, E., and Coauthors, 2001: The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. *Bull. Amer. Meteor. Soc.*, **82**, 247-267.
- Kiehl, J. T., T. L. Schneider, R. W. Portmann, and S. Solomon, 1999: Climate forcing due to tropospheric and stratospheric ozone. *J. Geophys. Res.*, **104**, 31239-31254.

- Manabe, S., and R. T. Wetherald, 1987: Large scale changes of soil wetness induced by an increase in atmospheric carbon dioxide. *J. Atmos. Sci.*, **44**, 1211-1235.
- Mears, C. A., M. C. Schabel, and F. W. Wentz, 2003: A reanalysis of the MSU channel 2 tropospheric temperature record *J. Clim.*, **16**, 3650-3664.
- Meehl, G. A., W. M. Washington, J. M. Arblaster, T. W. Bettge, and W. G. Strand Jr., 2000: Anthropogenic forcing and decadal climate variability in sensitivity experiments of twentieth- and twenty-first century climate. *J. Clim.*, **13**, 3728-3744.
- Meehl, G. A., P. R. Gent, J. M. Arblaster, B. L. Otto-Bliesner, E. C. Brady, and A. Craig, 2001: Factors that affect the amplitude of El Niño in global coupled climate models. *Clim. Dyn.*, **17**, 515-526.
- Meehl, G. A., W. M. Washington, T. M. L. Wigley, J. M. Arblaster, and A. Dai, 2003: Solar and greenhouse gas forcing and climate response in the 20th century. *J. Clim.*, **16**, 426-444.
- Mo, K., X. L. Wang, R., Kistler, M., Kanamitsu, and E. Kalnay, 1995: Impact of satellite data on the CDAS-reanalysis system. *Mon. Weath. Rev.*, **123**, 124-139.
- Pawson, S., and M. Fiorino, 1998: A comparison of reanalyses in the tropical stratosphere. Part 1: Thermal structure and the annual cycle. *Clim. Dyn.*, **14**, 631-644.
- Pawson, S., and M. Fiorino, 1999: A comparison of reanalyses in the tropical stratosphere. Part 3: Inclusion of the pre-satellite data era. *Clim. Dyn.*, **15**, 241-250.
- Pielke, R. A. Sr., and T. N. Chase, 2004: Comment on “Contributions of anthro-

pogenic and natural forcing to recent tropopause height changes". *Science*, **303**, 1771b.

Ramaswamy, V., M. D. Schwarzkopf, and W. J. Randel, 1996: Fingerprint of ozone depletion in the spatial and temporal pattern of recent lower-stratospheric cooling. *Nature*, **382**, 616-618.

Ramaswamy, V., and Coauthors, 2001: Stratospheric temperature trends: observations and model simulations. *Rev. Geophys.*, **39**, 71-122.

Reichler, T., M. Dameris, and R. Sausen, 2003: Determining the tropopause height from gridded data. *Geophys. Res. Lett.*, **30**, doi: 10.1029/2003GL018240.

Roeckner, E., L. Bengtsson, J. Feichter, J. Lelieveld, and H. Rodhe, 1999: Transient climate change simulations with a coupled atmosphere-ocean GCM including the tropospheric sulfur cycle. *J. Clim.*, **12**, 3004-3032.

Santer, B. D., and Coauthors, 1995: Ocean variability and its influence on the detectability of greenhouse warming signals. *J. Geophys. Res.*, **100**, 10693-10725.

Santer, B. D., and Coauthors, 1999: Uncertainties in observationally based estimates of temperature change in the free atmosphere. *J. Geophys. Res.*, **104**, 6305-6333.

Santer, B. D., and Coauthors, 2000a: Interpreting differential temperature trends at the surface and in the lower troposphere. *Science*, **287**, 1227-1232.

Santer, B. D., and Coauthors, 2000b: Statistical significance of trends and trend differences in layer-average atmospheric temperature time series. *J. Geophys. Res.*,

105, 7337-7356.

Santer, B. D., and Coauthors, 2003a: Contributions of anthropogenic and natural forcing to recent tropopause height changes. *Science*, **301**, 479-483.

Santer, B. D., and Coauthors, 2003b: Behavior of tropopause height and atmospheric temperature in models, reanalyses, and observations: Decadal changes. *J. Geophys. Res.*, **108**, 4002, doi:10.1029/2002JD002258.

Santer, B. D., and Coauthors, 2003c: Influence of satellite data uncertainties on the detection of externally forced climate change. *Science*, **300**, 1280-1284.

Santer, B. D., and Coauthors, 2004: Response to Comment on “Contributions of anthropogenic and natural forcing to recent tropopause height changes”. *Science*, **303**, 1771c.

Sausen, R. & B. D. Santer, 2003: Use of changes in tropopause height to detect human influences on climate. *Meteorolog. Z.*, **12**, 131-136.

Seidel, D. J., R. J. Ross, J. K. Angell, and G. C. Reid, 2001: Climatological characteristics of the tropical tropopause as revealed by radiosondes. *J. Geophys. Res.*, **106**, 7857-7878.

Simmons, A. J., and J. K. Gibson, 2000: *The ERA-40 project plan, ERA-40 Project Report Series No. 1*, Reading, U. K., 62 pp.

Simmons, A. J., and A. Hollingsworth, 2002: Some aspects of the improvement in skill of numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **128**, 647-677.

- Smith, C. A., and P. Sardeshmukh, 2000: The effect of ENSO on the intraseasonal variance of surface temperature in winter. *Int. J. Clim.*, **20**, 1543-1557.
- Sperber, K. R., 2003: Propagation and vertical structure of the Madden-Julian Oscillation. *Mon. Weath. Rev.*, **131**, 3018-3037.
- Stott, P. A., and Coauthors, 2000: External control of 20th century temperature by natural and anthropogenic forcings. *Science*, **290**, 2133-2137.
- Stott, P. A., G. S. Jones, and J. F. B. Mitchell, 2003: Do models underestimate the solar contribution to recent climate change? *J. Clim.*, **16**, 4079-4093.
- Trenberth, K. E., D. P. Stepaniak, J. W., Hurrell, and M. Fiorino, 2001: Quality of reanalyses in the tropics. *J. Clim.*, **14**, 1499-1510.
- Trenberth, K. E., and J. G. & Olson, 1988: An evaluation and intercomparison of global analyses from NMC and ECMWF. *Bull. Amer. Meteor. Soc.*, **69**, 1047-1057.
- Washington, W. M., and Coauthors, 2000: Parallel Climate Model (PCM) control and transient simulations. *Clim. Dyn.*, **16**, 755-774.
- Wigley, T. M. L., P. J. Jaumann, B. D. Santer, and K. E. Taylor, 1998: Relative detectability of greenhouse-gas and aerosol climate change signals. *Clim. Dyn.*, **14**, 781-790.
- Wigley, T. M. L., 2000: ENSO, volcanoes, and record-breaking temperatures. *Geophys. Res. Lett.*, **27**, 4101-4104.

Wigley, T. M. L., C. M. Ammann, B. D. Santer, and S. C. B. Raper, 2004: The effect of climate sensitivity on the response to volcanic forcing. *J. Clim.* (submitted).

World Meteorological Organization, 1957: *Meteorology-A three-dimensional science: Second session of the commission for aerology WMO Bull.*, **IV(4)**, 134-138.

APPENDIX A

Forcings and model details

The PCM ALL and ANTHRO experiments provide estimates of expected changes in p_{LRT} . Full details of the historical forcings used in these integrations are given elsewhere (Dai et al., 2001; Washington et al., 2000; Kiehl et al., 1999). Here, it is sufficient to note that the anthropogenic forcings in ALL and ANTHRO are identical to those employed in experiments with the NCAR Climate System Model (CSM; Dai et al., 2001). Of relevance for detection studies is the neglect of indirect sulfate aerosol forcing (see, e.g., Stott et al., 2003), the absence of forcing by black and organic carbon (Hansen et al., 2002), and the (unrealistic) assumption that the spatial pattern of SO₂ emissions is time invariant (except over the seasonal cycle), and can be scaled by estimates of historical changes in global-mean SO₂ emissions.¹⁶

Natural external forcings were treated as follows. Total solar irradiance changes were prescribed according to Hoyt and Schatten (1993), updated as in Meehl et al. (2003), with no wavelength dependence of the forcing. Volcanic forcing was based on estimates of total sulfate loading and a simplified model of aerosol distribution and decay (Ammann et al., 2003).

Both PCM and the ECHAM model (the latter was used exclusively for estimating internally-generated climate noise) were run with T42 spectral truncation in their at-

¹⁶This assumption is likely to be more serious for detection work focusing on century-timescale changes than for our analysis, which relies on post-1979 p_{LRT} changes. This is because variations in the spatial pattern of SO₂ emissions are larger over the 20th century than over the past 25 years.

atmospheric model components, which is equivalent to a horizontal resolution of roughly 250-300 km in the tropics. PCM and ECHAM use 18 and 19 atmospheric levels respectively. PCM's ocean model component has relatively high spatial resolution, with 32 vertical layers and $2/3^\circ \times 2/3^\circ$ horizontal resolution, decreasing to 0.5° at the equator. The ECHAM ocean model has coarser vertical resolution (11 vertical layers) and coarser horizontal resolution poleward of 36° ($2.8^\circ \times 2.8^\circ$). Like PCM, ECHAM's ocean resolution decreases to 0.5° at the equator.

APPENDIX B

Fingerprint Detection Procedure

a. Definition of fingerprint

Let $\vec{s}(t)$ represent the time-evolving patterns of annual-mean p_{LRT} from a realization of the PCM ANTHRO experiment, expressed as anomalies relative to the smoothed ANTHRO initial state (1890-1909). The arrow denotes a vector in p -dimensional space, where p is the total number of model grid-points; t is time in years. The fingerprint \vec{f} is computed from the ensemble-mean $\vec{s}(t)$ data for the full period of the ANTHRO experiment (1890-1999), after first regridding to a 10° latitude \times 10° longitude grid and excluding data poleward of 60°S (see Fig. 1). We define \vec{f} as the first EOF, which explains a substantial fraction of the overall variance of $\vec{s}(t)$: 60% for the ‘mean included’ analysis, and 29% for the ‘mean removed’ case. Possible low-frequency changes in the signal pattern are not accounted for (Wigley et al., 1998), as they would be in an approach using space-time EOFs (Stott et al., 2000).

b. Estimation of detection time

We use a standard “fingerprinting” technique (Hasselmann 1979; Santer et al., 1995) to determine detection time – the time at which the fingerprint \vec{f} becomes consistently identifiable at some stipulated significance level. Our method relies on the defined fingerprint \vec{f} (see above), on annual-mean ‘observational’ data, $\vec{o}(t)$ (NCEP-50 or ERA-40), and on control integrations, $\vec{c}(t)$ and $\vec{c}_1(t)$ (PCM and ECHAM). Reanalysis

data are expressed as anomalies relative to 1979-2001; control anomalies are defined relative to the mean of the full 300-year integrations.

Two forms of detection time are computed: non-optimized ('raw') and optimized. To define raw detection times, $\vec{o}(t)$ and $\vec{c}_1(t)$ are projected onto the fingerprint \vec{f} , yielding (respectively) a test statistic time series $Z(t)$ and a 'signal free' time series $N(t)$. We fit least-squares linear trends of increasing length L to $Z(t)$, and then compare their slope parameters with the distribution of L -length trends in $N(t)$ until the trend exceeds and remains above the 5% significance level. The test is one-tailed and we assume a Gaussian distribution of trends in $N(t)$. Detection time is referenced to 1979, which marks the start date of more widespread satellite data assimilation in both reanalyses (Kalnay et al., 1996; Simmons and Gibson, 2000). We use a minimum trend length of 10 years, so the earliest possible detection time is in 1988.

Optimized detection times are determined similarly, but involve projection of $\vec{o}(t)$ and $\vec{c}_1(t)$ onto \vec{f}_m^* , a version of the fingerprint that has been rotated away from high noise directions. This rotation is performed in the subspace of the first m EOFs of $\vec{c}(t)$, where m is the 'truncation dimension'. We explore the sensitivity of optimized detection times by using three different values of m (5, 10, and 15). Our basic conclusions are insensitive to this choice, and Fig. 8 shows results for the $m = 15$ case only. Full details of the detection method are given in Santer et al., (1995).

Given the short observational record lengths, we use only the spatial properties of signal and noise in rotating \vec{f} . Other detection work involving longer datasets with more temporal structure has employed both spatial and temporal information for fingerprint optimization (Stott et al., 2000). In the eight cases shown in Fig. 8, optimization leaves detection times unchanged in four cases, improves detection time

in one case, and degrades detection time in three cases.

One possible explanation for the failure of optimization to consistently improve detection times is that important components of the fingerprint may be lost in projecting \vec{f} onto the subspace of only the first m control run EOFs. Significant differences between the noise used for optimizing \vec{f} and the noise used for calculating natural variability statistics can also reduce the effectiveness of optimization.

c. Analysis with mean removed

In the ‘mean removed’ case, time-varying spatial means of the ensemble-mean PCM ANTHRO anomalies are removed (from each grid-point, and at each time) prior to calculation of EOFs and \vec{f} . Time-varying spatial means are also subtracted from $\vec{o}(t)$, $\vec{c}(t)$, and $\vec{c}_1(t)$.

d. Sensitivity to significance level

Our detection times are not strongly sensitive to the choice of significance threshold. While we have used a nominal 5% significance threshold here, a more conservative test with a 1% significance threshold yields very similar results.

Table 1: Statistics for time series of changes in p_{LRT} , T4, and T2. Results for reanalyses, RSS and UAH were calculated using monthly-mean anomalies spanning the 276-month period from January 1979 through December 2001. Anomalies are relative to climatological monthly means for this period. PCM statistics were computed over January 1979 through December 1999, and are averages of results for the four ALL realizations. p_{LRT} anomaly data are either global means or spatial averages over 90°N-60°S (see Section 3.1). All T4 and T2 anomalies are global means. The vertical resolution of the atmospheric temperature data used for p_{LRT} , T4, and T2 calculations is indicated in the ‘Levels’ column. Trends and 1σ trend standard errors are in °C/decade (T4, T2) or in hPa/decade (p_{LRT}). The lag-1 autocorrelation of the time series (AR-1) was used to adjust standard errors for temporal autocorrelation effects (Santer et al., 2000). Due to their high AR-1 values and small effective sample sizes, adjusted standard errors could not be calculated reliably for PCM T4 data. Standard deviations are in °C (T4, T2) or hPa (p_{LRT}). One, two, or three asterisks denote trends significantly different from zero at the 10%, 5%, or 1% levels (respectively); tests are one-tailed. p_{LRT} data were available from reanalyses and PCM only.

Table 2: Correlations between time series of global-mean monthly-mean atmospheric temperature anomalies for four different datasets. Correlations were calculated over the 276-month period from January 1979 through December 2001, and were computed from both the raw anomaly data and linearly detrended data (underlined).

Table 3: Correlations between the spatial patterns of atmospheric temperature change in reanalysis and satellite datasets. Pattern correlations were calculated with the linear trend data in Figures 3 and 4 (for T2 and T4, respectively). All datasets were transformed to the RSS grid and masked with RSS coverage. Two forms of

correlation are given: with spatial means included (c) and spatial means subtracted (r ; Barnett and Schlesinger, 1987). The latter are underlined.

Figure 1: Effect of vertical resolution of the input atmospheric temperature data on estimates of global-scale p_{LRT} changes in ERA-40. p_{LRT} was calculated from ERA-40 temperatures archived at two different vertical resolutions: 23 pressure levels (L23) and 60 model levels (L60; see Section 3.1). Monthly-mean anomalies are either globally-averaged (**a**), or averaged over 90°N-60°S (**b**). Anomalies were defined relative to climatological monthly means over January 1979 to December 2001.

Figure 2: Effect of vertical resolution of the input temperature data on the estimated patterns of p_{LRT} change in ERA-40. p_{LRT} calculations were performed with both L23 and L60 atmospheric temperature data (**a,b**). Linear p_{LRT} trends over 1979-2001 were calculated using monthly-mean anomaly data, with anomalies defined as in Fig. 1. The L60 and L23 trend patterns are highly correlated ($r = 0.94$).

Figure 3: Atmospheric temperature profiles in ERA-40 at a selected Antarctic grid-point (84.67°S, 128.25°E). Temperatures are monthly-mean values for May, June, July, and August of 1958. Crosses denote the pressure of the lapse-rate tropopause estimated with the standard WMO criterion (WMO, 1957; Reichler et al.; 2003). Note the unrealistically high p_{LRT} value in August 1958 (Section 3.1).

Figure 4: Time series of monthly-mean p_{LRT} anomalies from the NCEP-50 and ERA-40 reanalyses and the ensemble mean of the PCM ALL experiment (Section 2). Results are spatial averages over 90°N-60°S. Bold lines denote data that were low-pass filtered to highlight changes on 5-10 year timescales; thin dotted lines are the raw monthly-mean anomalies. ERA-40 p_{LRT} values were calculated using the L23

temperature data (Section 3.1). Reanalysis p_{LRT} anomalies were defined relative to climatological monthly means computed over 1979-2001, while PCM anomalies were expressed relative to a 1979-1999 reference period.

Figure 5: Time series of global-mean, monthly-mean anomalies in lower stratospheric temperatures (MSU T4). Results are processed MSU T4 measurements (UAH, RSS) and synthetic T4 temperatures calculated from NCEP-50, ERA-40, and the ensemble mean of the PCM ALL experiment (Section 2.2). For definition of anomalies and explanation of bold and thin lines, refer to Fig. 4.

Figure 6: Conceptual model for the effect of three different forcings on tropopause height. The solid black lines are the baseline atmospheric temperature profiles. Forcing by stratospheric ozone depletion, increases in well-mixed greenhouse gases, and volcanic eruptions can perturb this base state. The effect of the first two forcings is to increase tropopause height (indicated by the upward-pointing arrows), while volcanic forcing causes height decreases.

Figure 7: Tropopause pressure changes in reanalyses and PCM. Least-squares linear trends in monthly-mean p_{LRT} data (in hPa/decade) were computed over 1958-2001 for ERA-40 and NCEP-50 (**a,b**). ERA-40 and NCEP-50 trends are also shown for the shorter period 1979-2001 (**c,d**). For PCM, p_{LRT} trends over 1979-1999 were calculated from the ensemble mean of the ALL and ANTHRO experiments (**e,f**).

Figure 8: Detection times for PCM tropopause height fingerprints in NCEP-50 and ERA-40 reanalyses. The detection analysis uses both the ‘mean included’ and ‘mean removed’ fingerprints calculated from the PCM ANTHRO experiment, with a 5% significance level as the detection threshold (Section 5). The longer the colored bar,

the earlier the detection time. If no bar is present, fingerprints could not be identified before the final year of the reanalyses (2001). ‘RAW’ denotes detection times for non-optimized fingerprints. Optimized detection times are given for a single choice of the truncation dimension m ($m = 15$; see APPENDIX B). To avoid the introduction of artificial skill, the model control run used for optimization was always different from the control run used for estimating natural variability statistics.

Figure 9: Time series of global-mean, monthly-mean anomalies in processed and synthetic tropospheric temperatures (MSU T2). For further details, refer to Fig. 5. The colored rectangles on the time axis use a composite index of SST and circulation changes (Smith and Sardeshmukh, 2000) to indicate the timing and duration of observed El Niño (in red) and La Niña (in blue) events, which influence the variability of T2 data in reanalyses, UAH, and RSS (Wigley 2000; Santer et al., 2003c).

Figure 10: Consistency between changes in synthetic T2 temperatures in ERA-40 and the PCM ALL experiment. Bold lines denote the low-pass filtered T2 data in ERA-40 (black) and in the ALL ensemble mean (blue). The yellow envelope defines the range between the highest and lowest T2 anomalies in the four realizations of ALL. The range was smoothed with the same low-pass filter that was applied to ERA-40 and the ALL ensemble mean. The thin dotted lines are the unfiltered T2 anomalies, defined as in Fig. 4. The colored rectangles provide information on observed ENSO events (see Fig. 9).

Figure 11: Tropospheric temperature changes in reanalyses and PCM. Least-squares linear trends over 1979-2001 in monthly-mean processed or synthetic MSU T2 temperatures from ERA-40 (a), RSS (b), UAH (c), and NCEP-50 (d). Also shown are T2 trends over 1979-1999 in the ensemble mean of the PCM ALL experiment (e).

Figure 12: Stratospheric temperature changes in reanalyses and PCM. Least-squares linear trends over 1979-2001 in monthly-mean processed or synthetic MSU T4 temperatures from ERA-40 (**a**), RSS (**b**), UAH (**c**), and NCEP-50 (**d**). Also shown are T4 trends over 1979-1999 in the ensemble mean of the PCM ALL experiment (**e**).

Table 1

		Area	Levels	Trend	Std. Error	Std. Dev.	AR-1
p_{LRT}	NCEP-50	S	17	-1.79 ***	± 0.37	1.95	0.73
	ERA-40	G	60	-2.36 ***	± 0.47	2.36	0.78
	ERA-40	G	23	-2.66 ***	± 0.48	2.46	0.80
	ERA-40	S	60	-1.78 ***	± 0.38	2.00	0.73
	ERA-40	S	23	-2.12 ***	± 0.38	2.12	0.75
	PCM ALL	S	18	-1.13	± 1.41	2.52	0.92
T4	UAH	G	1	-0.49 *	± 0.29	0.49	0.96
	RSS	G	1	-0.39 *	± 0.22	0.44	0.94
	NCEP-50	G	17	-0.82 ***	± 0.26	0.69	0.94
	ERA-40	G	23	-0.30	± 0.30	0.42	0.96
	PCM ALL	G	18	-0.35	—	0.58	0.98
T2	UAH	G	1	0.01	± 0.05	0.18	0.78
	RSS	G	1	0.09 **	± 0.05	0.19	0.78
	NCEP-50	G	17	-0.11 **	± 0.06	0.21	0.81
	ERA-40	G	23	0.08 **	± 0.05	0.20	0.76
	PCM ALL	G	18	0.07	± 0.06	0.20	0.80

G: Global-mean.

S: Spatial average over 90°N-60°S.

Table 2

T4 results	UAH	RSS	NCEP-50	ERA-40
UAH	1.00	0.99	0.95	0.96
RSS	<u>0.99</u>	1.00	0.91	0.97
NCEP-50	<u>0.92</u>	<u>0.90</u>	1.00	0.88
ERA-40	<u>0.98</u>	<u>0.97</u>	<u>0.93</u>	1.00

T2 results	UAH	RSS	NCEP-50	ERA-40
UAH	1.00	0.94	0.81	0.92
RSS	<u>0.98</u>	1.00	0.67	0.93
NCEP-50	<u>0.88</u>	<u>0.89</u>	1.00	0.68
ERA-40	<u>0.94</u>	<u>0.93</u>	<u>0.86</u>	1.00

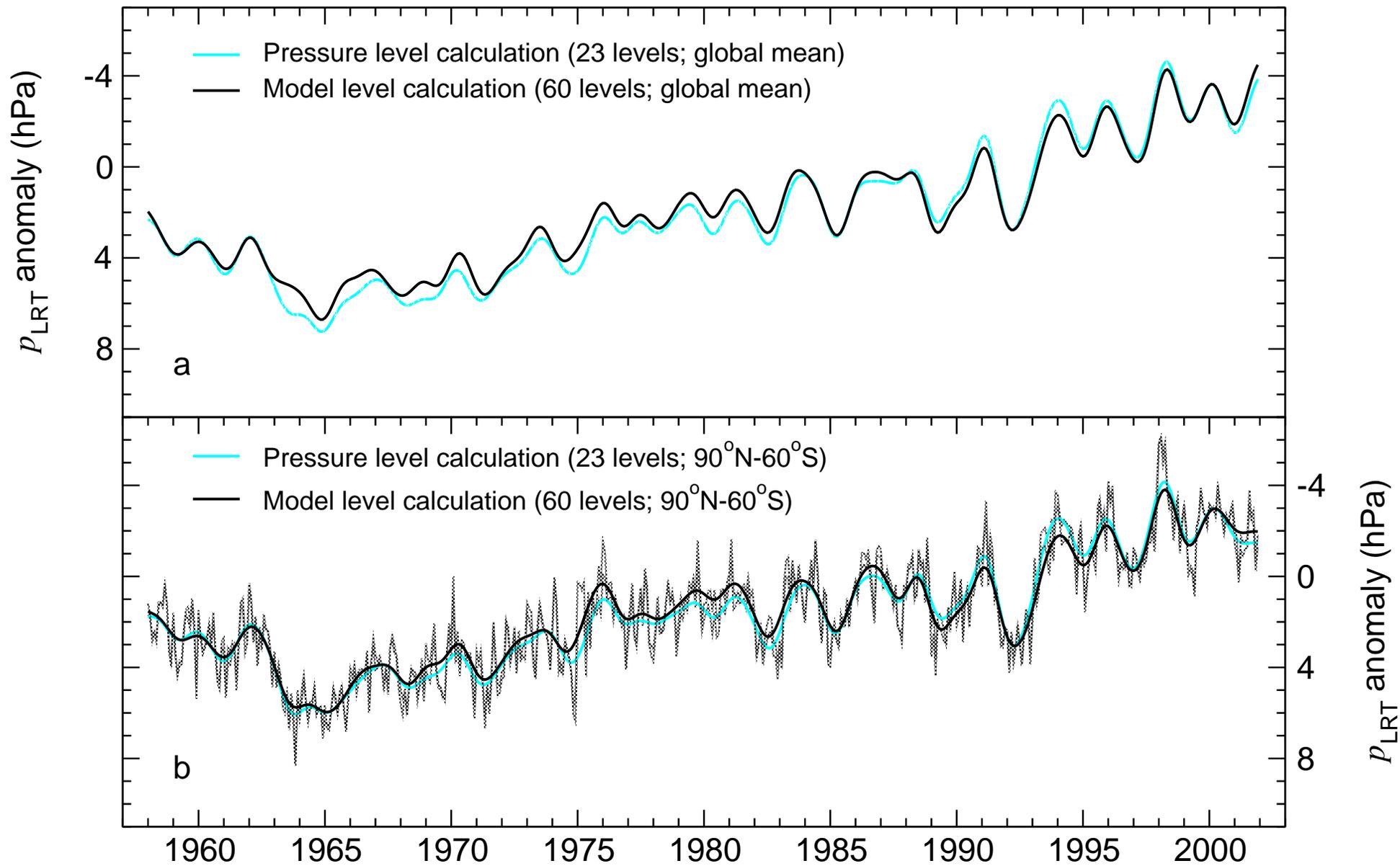
Table 3

T4 results	UAH	RSS	NCEP-50	ERA-40
UAH	1.00	0.99	0.91	0.98
RSS	<u>0.97</u>	1.00	0.91	0.98
NCEP-50	<u>0.30</u>	<u>0.47</u>	1.00	0.86
ERA-40	<u>0.97</u>	<u>0.96</u>	<u>0.41</u>	1.00

T2 results	UAH	RSS	NCEP-50	ERA-40
UAH	1.00	0.80	0.27	0.42
RSS	<u>0.95</u>	1.00	-0.15	0.73
NCEP-50	<u>0.46</u>	<u>0.51</u>	1.00	-0.08
ERA-40	<u>0.50</u>	<u>0.55</u>	<u>0.74</u>	1.00

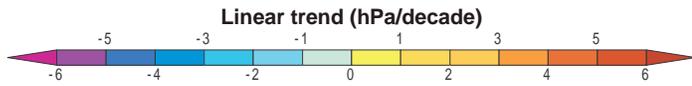
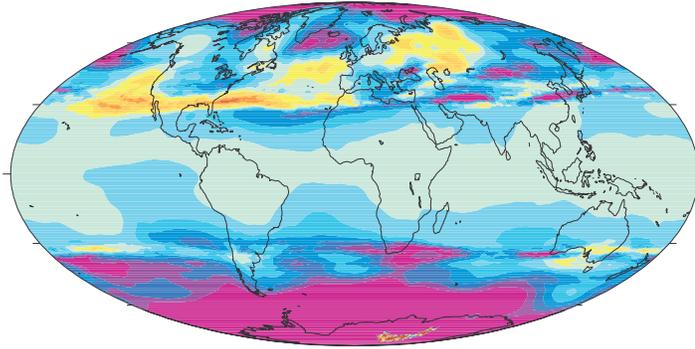
Lapse-Rate Tropopause Calculations in ERA-40

Anomalies w.r.t. 1979-2001. Bold: low-pass filtered data

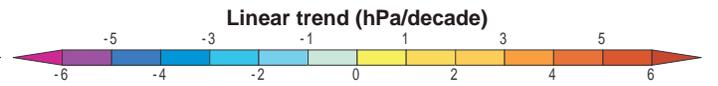
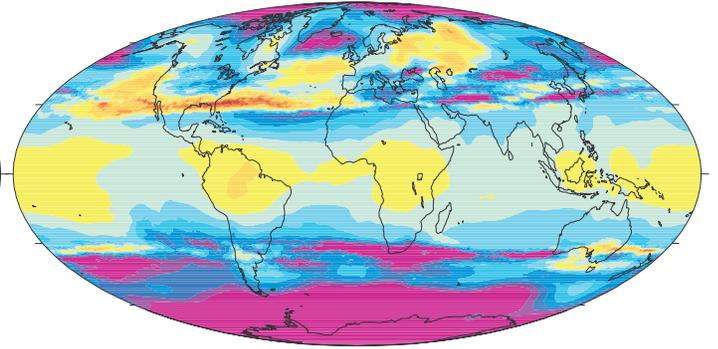


Tropopause Pressure Changes in ERA-40: Sensitivity to Vertical Resolution

a ERA-40 linear trend (23 levels; 1979–2001)



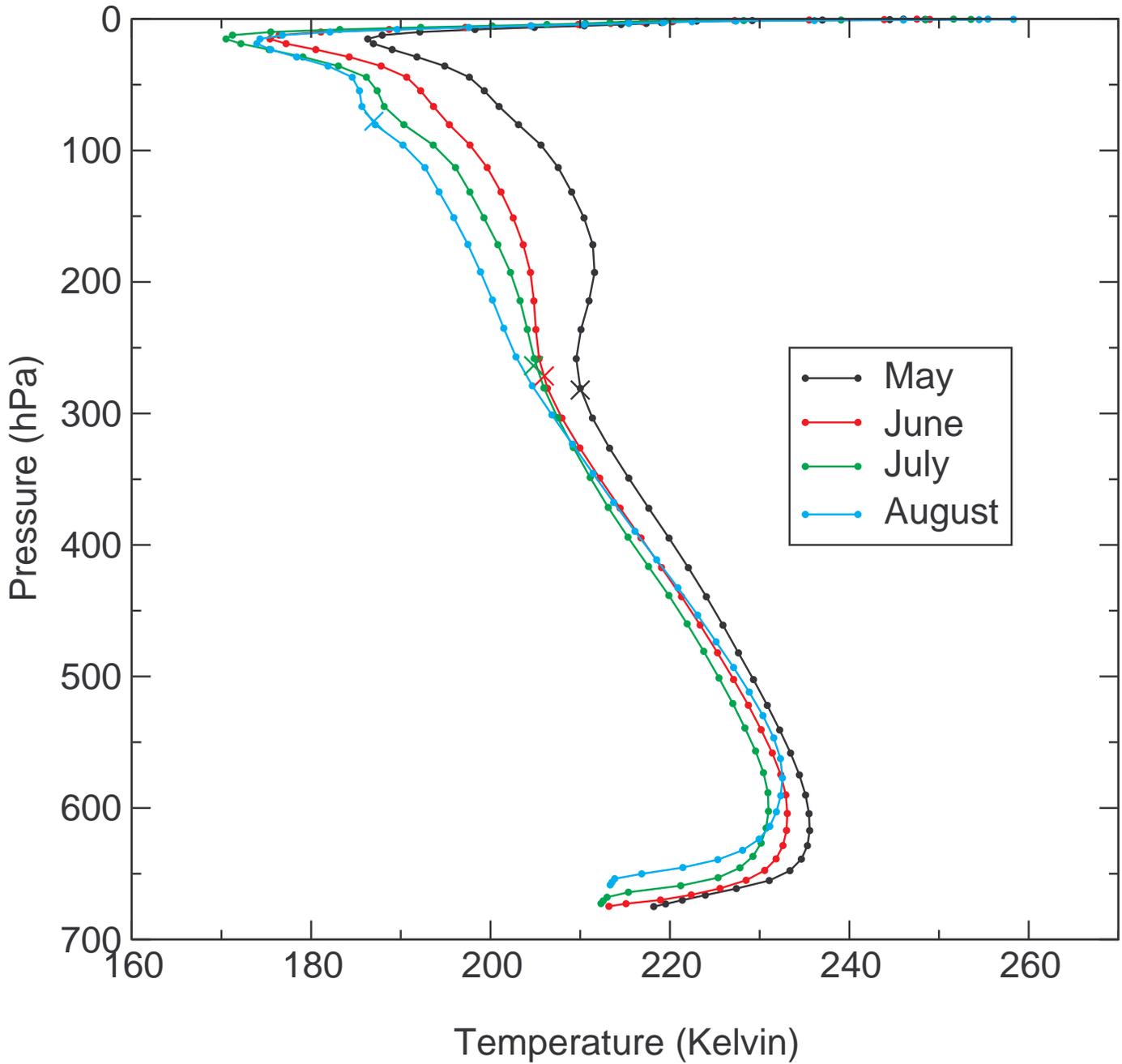
b ERA-40 linear trend (60 levels; 1979–2001)



Santer et al. Fig. 2

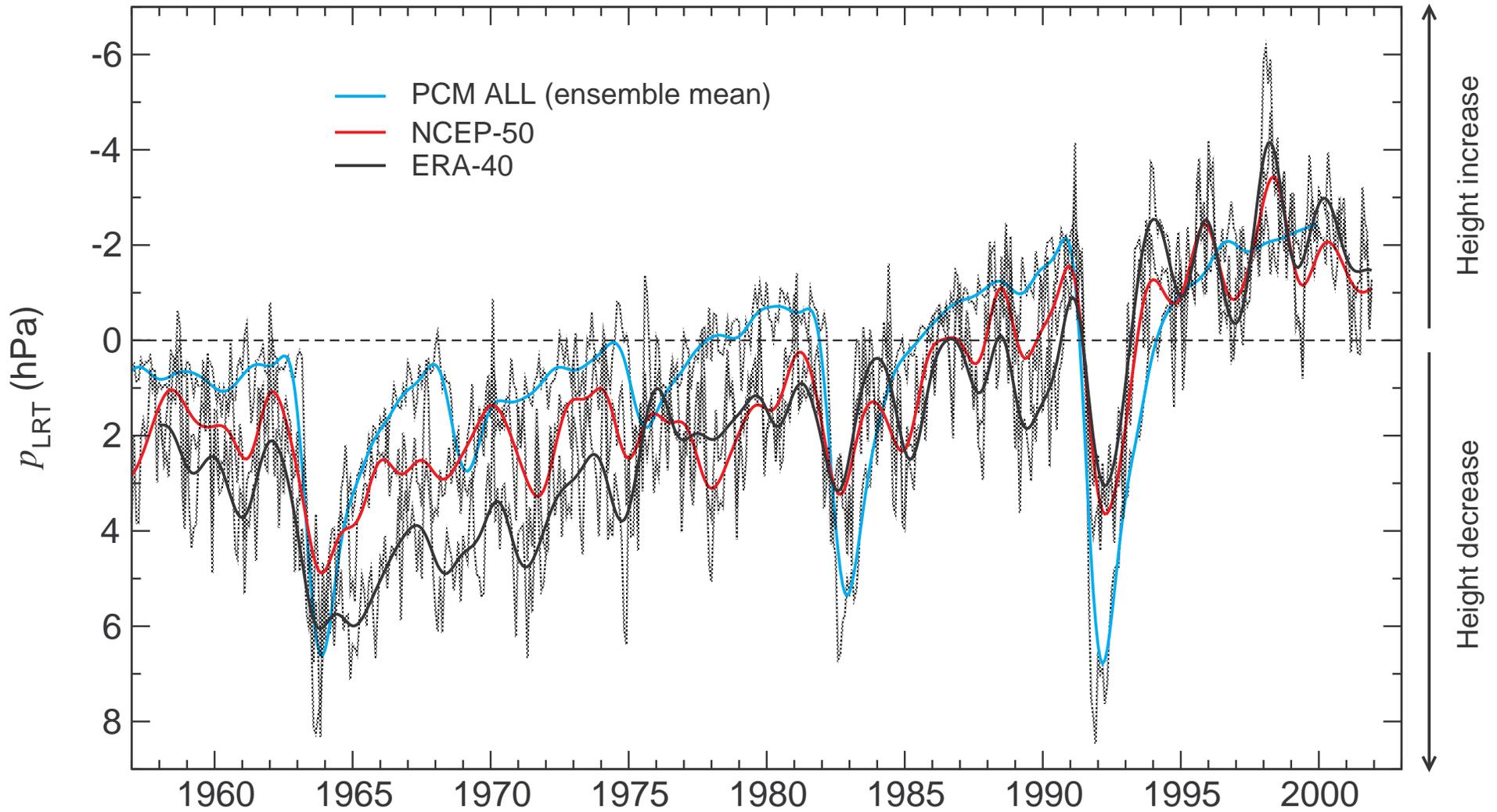
Estimated Lapse-Rate Tropopause in ERA-40

Longitude: 128.25°E. Latitude: 84.67°S. Time: 1958



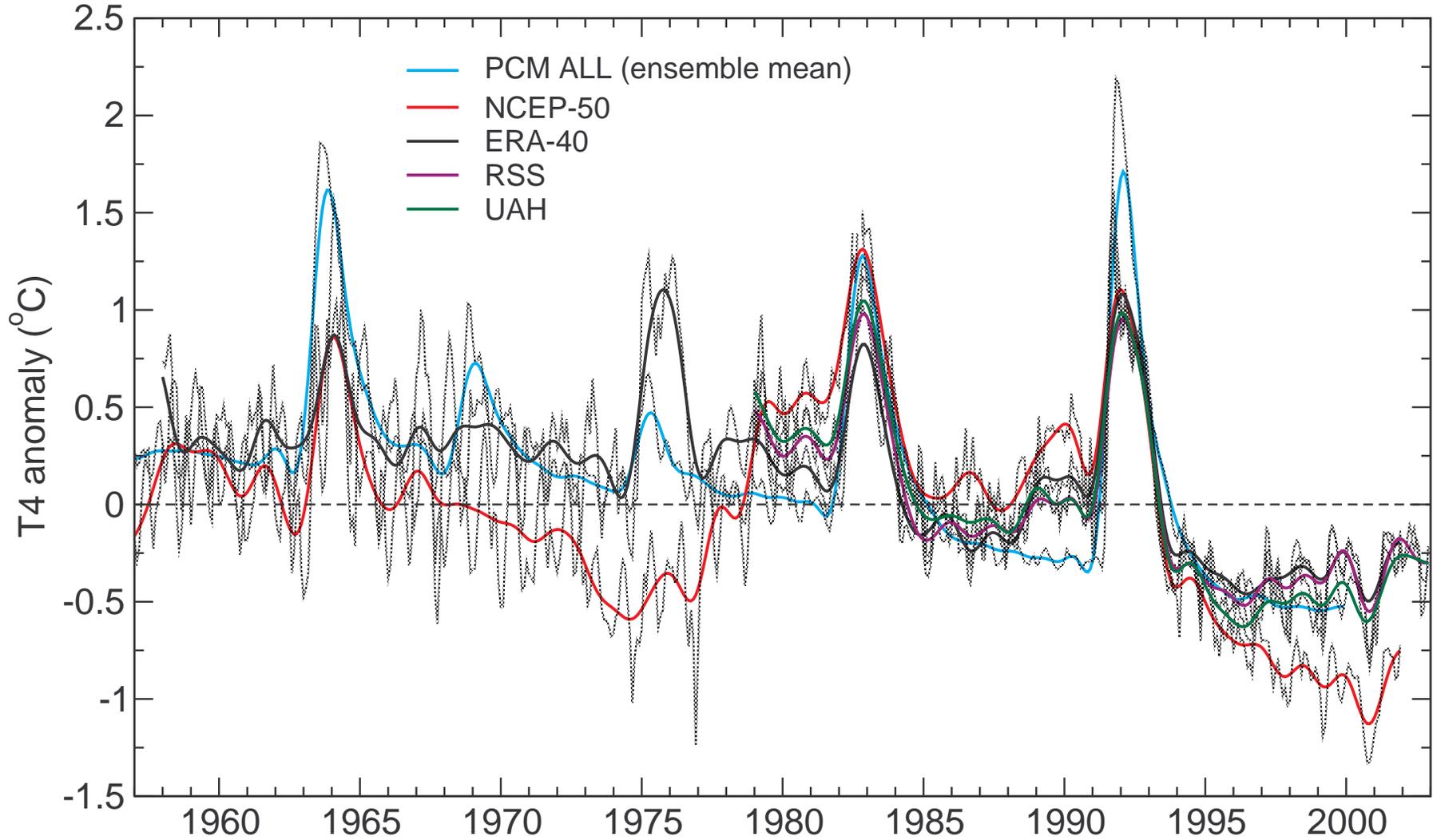
Changes in Tropopause Pressure in Reanalyses and PCM

Averages over 90°N-60°S. Bold: low-pass filtered data

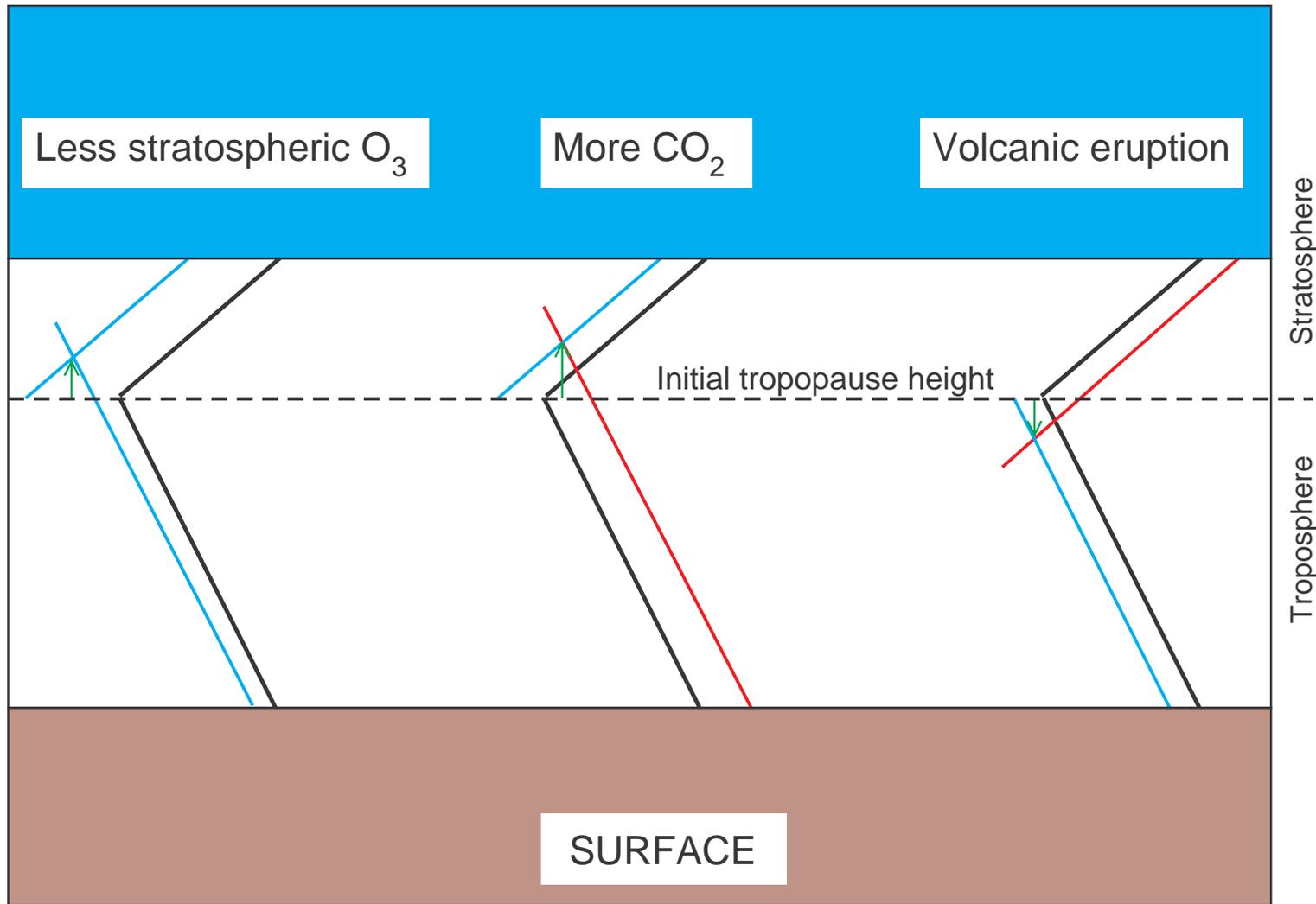


Stratospheric Temperature Changes in Reanalyses, RSS, UAH, and PCM

Actual and synthetic MSU channel 4. Global means. Bold: low-pass filtered data

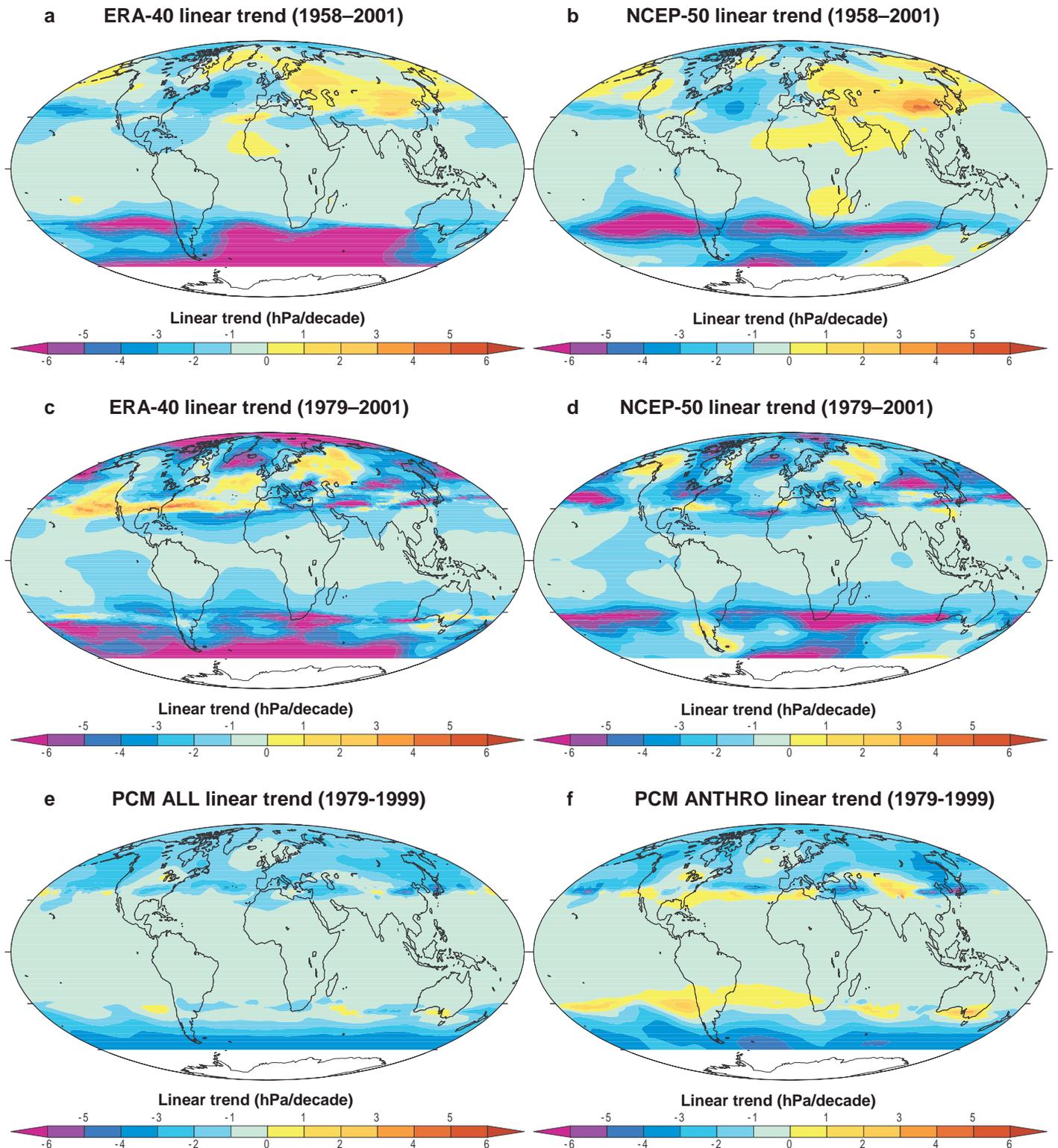


Effects of Different Forcing Mechanisms on Tropopause Height

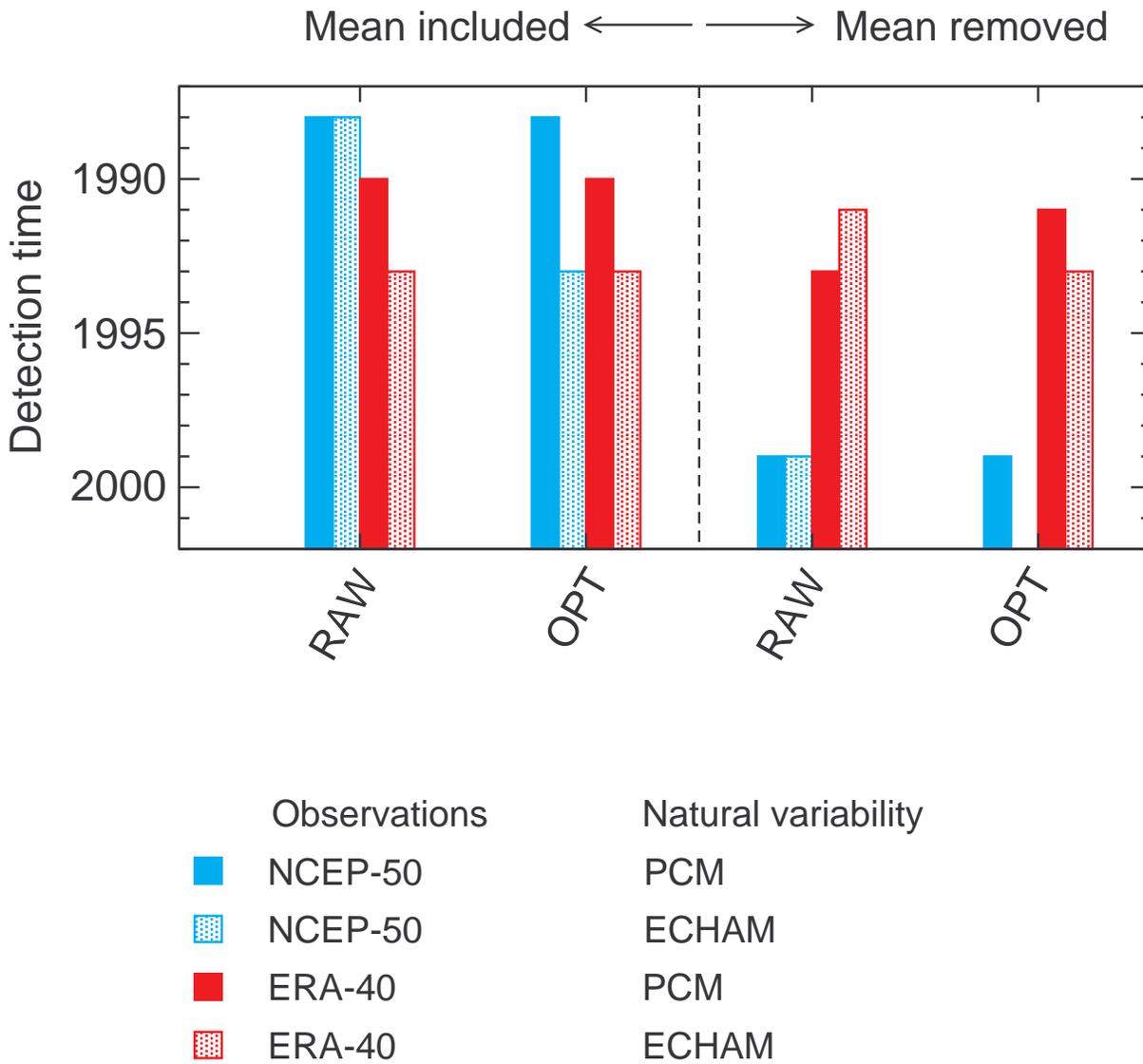


- Baseline temperature profile
- Cooling
- Warming

Tropopause Pressure Changes in Reanalyses and PCM

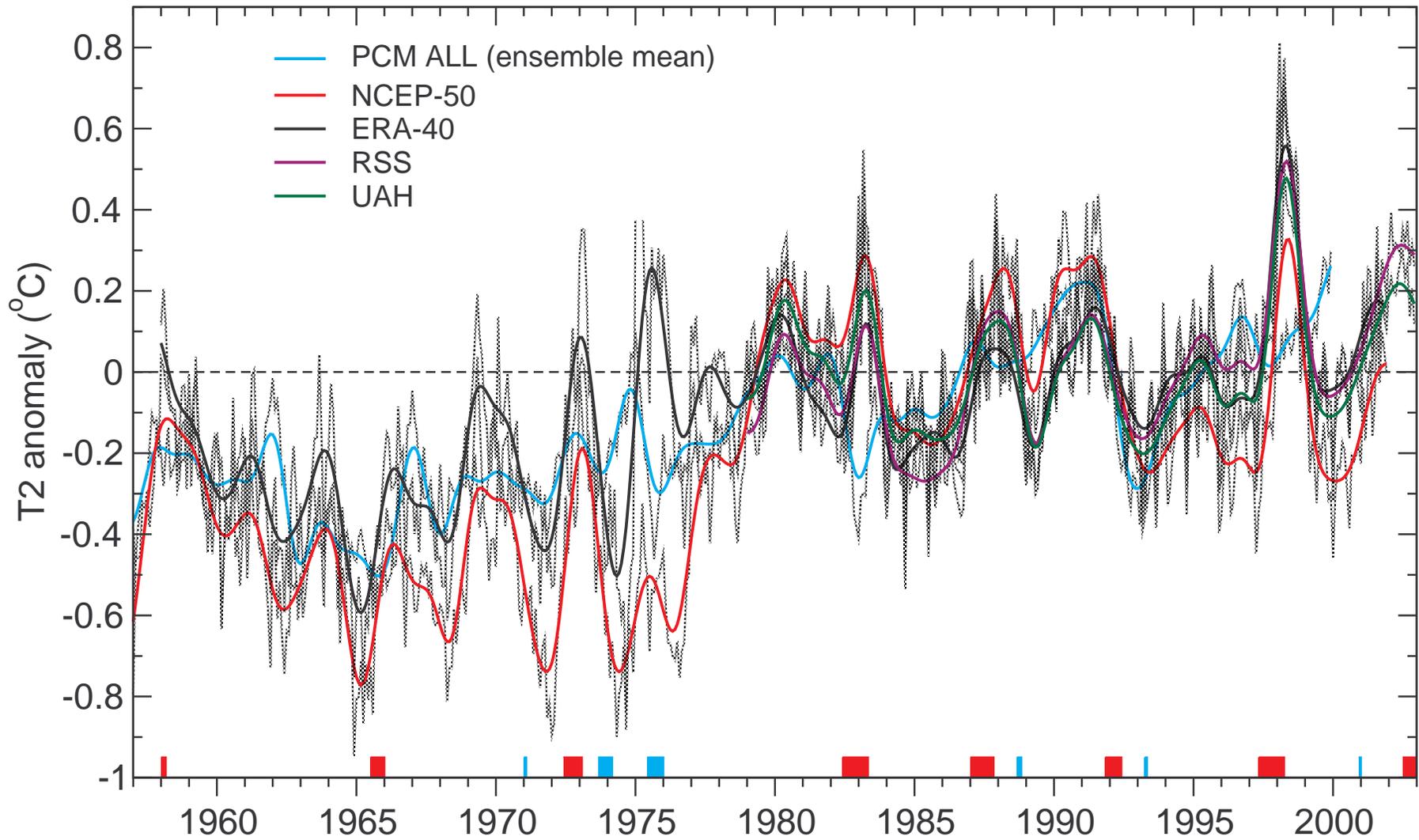


Detection Times for PCM ANTHRO Fingerprint



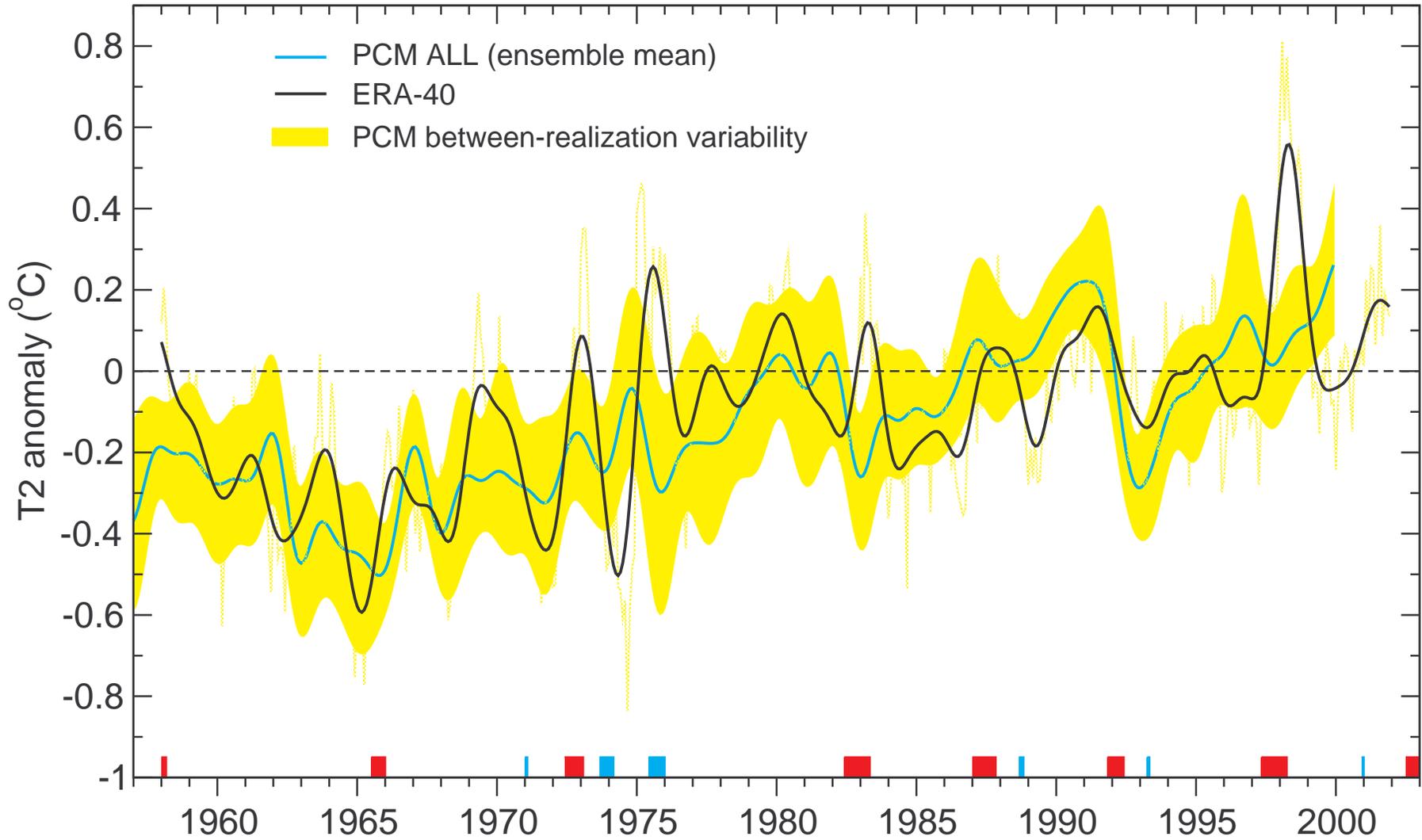
Tropospheric Temperature Changes in Reanalyses, RSS, UAH, and PCM

Actual and synthetic MSU channel 2. Global means. Bold: low-pass filtered data



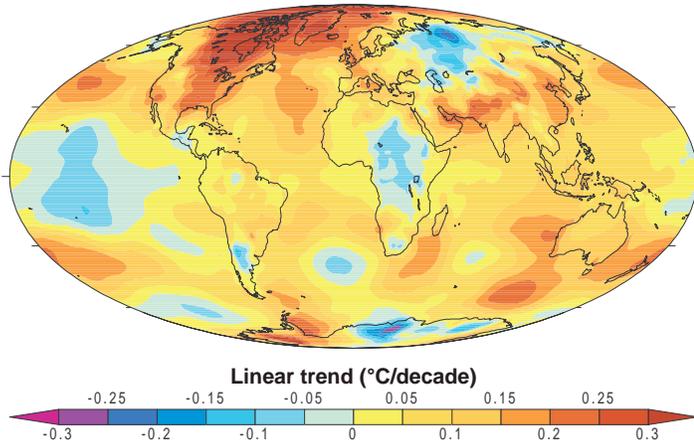
Tropospheric Temperature Changes: Consistency Between ERA-40 and PCM

Actual and synthetic MSU channel 2. Global means. Bold: low-pass filtered data

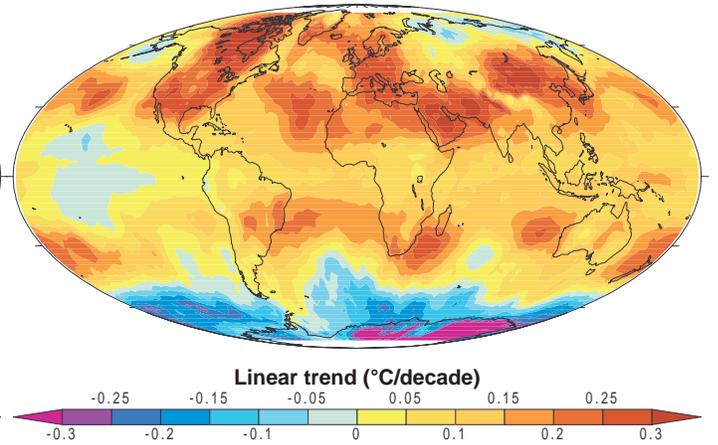


T2 Changes in Reanalyses, MSU and PCM

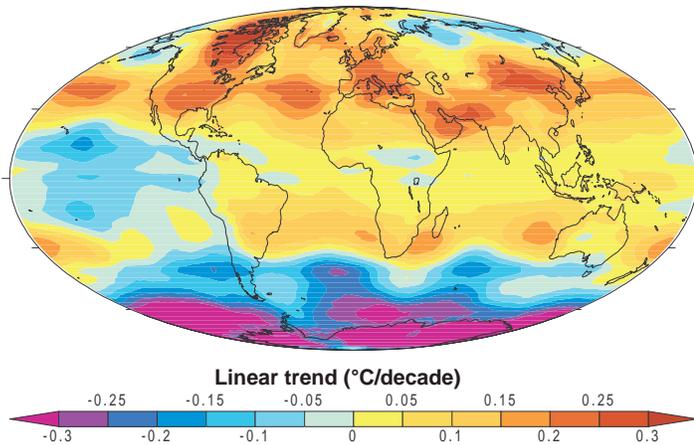
a ERA-40 linear trend (1979–2001)



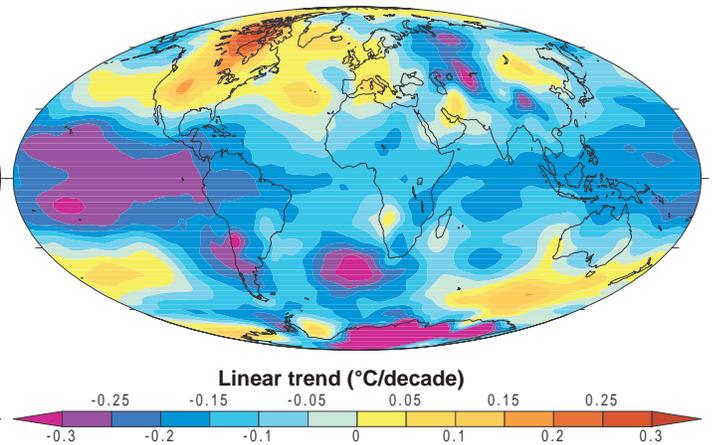
b RSS linear trend (1979–2001)



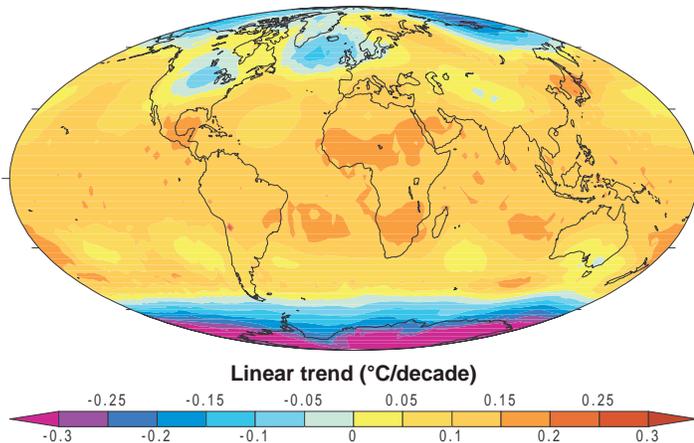
c UAH linear trend (1979–2001)



d NCEP-50 linear trend (1979–2001)

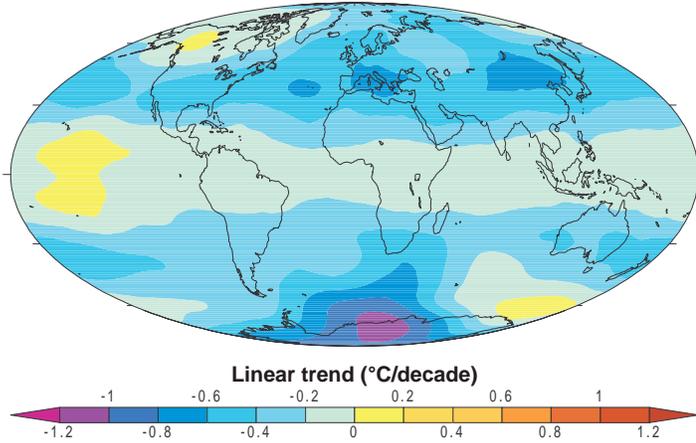


e PCM ALL linear trend (1979–1999)

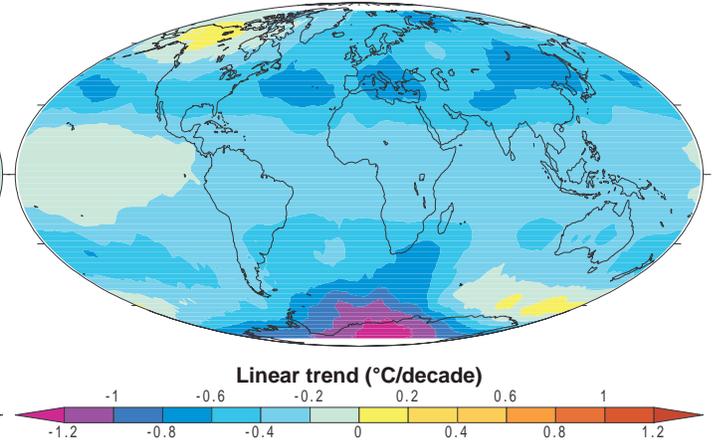


T4 Changes in Reanalyses, MSU and PCM

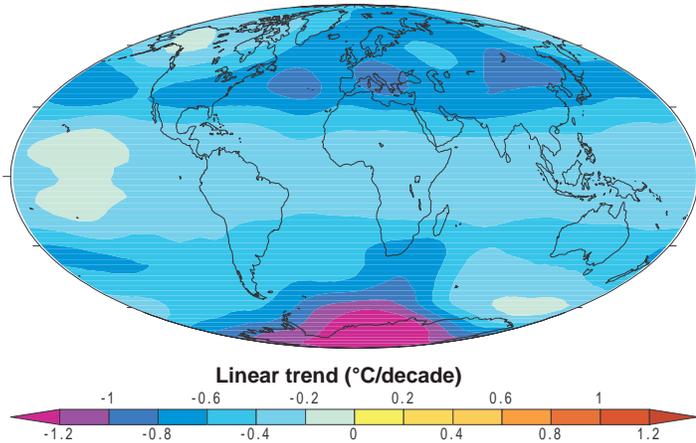
a ERA-40 linear trend (1979–2001)



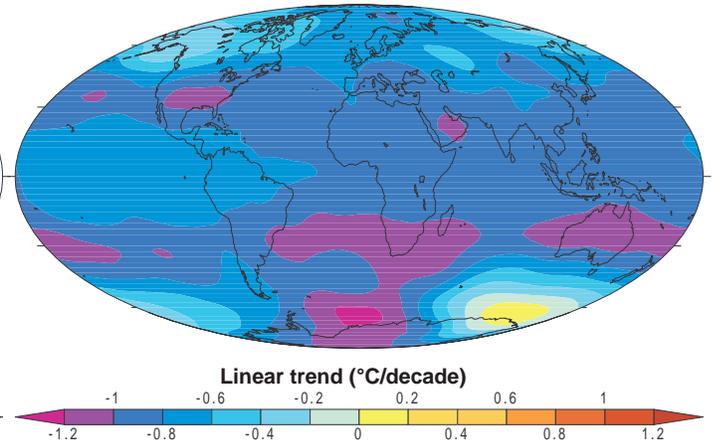
b RSS linear trend (1979–2001)



c UAH linear trend (1979–2001)



d NCEP-50 linear trend (1979–2001)



e PCM ALL linear trend (1979–1999)

