

Identification of anthropogenic climate change using a second-generation reanalysis

Benjamin D. Santer,¹ Tom M. L. Wigley,² Adrian J. Simmons,³ Per W. Kållberg,³ Graeme A. Kelly,³ Sakari M. Uppala,³ Caspar Ammann,² James S. Boyle,¹ Wolfgang Brüggemann,⁴ Charles Doutriaux,¹ Mike Fiorino,¹ Carl Mears,⁵ Gerald A. Meehl,² Robert Sausen,⁶ Karl E. Taylor,¹ Warren M. Washington,² Michael F. Wehner,⁷ and Frank J. Wentz⁵

Received 31 May 2004; revised 2 August 2004; accepted 19 August 2004; published 4 November 2004.

[1] 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 m 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 data sets are more highly correlated with changes in ERA-40 than with those in a first-generation reanalysis, 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. *INDEX TERMS:*

0350 Atmospheric Composition and Structure: Pressure, density, and temperature; 0370 Atmospheric Composition and Structure: Volcanic effects (8409); 1620 Global Change: Climate dynamics (3309); 1640 Global Change: Remote sensing; *KEYWORDS:* climate change, detection, reanalysis

Citation: Santer, B. D., et al. (2004), Identification of anthropogenic climate change using a second-generation reanalysis, *J. Geophys. Res.*, 109, D21104, doi:10.1029/2004JD005075.

1. Introduction

[2] 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.

[3] 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 Centers 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 Reanalysis). NCEP and the National Center for Atmospheric Research (NCAR) jointly produced a similar 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, 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

¹Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, Livermore, California, USA.

²National Center for Atmospheric Research, Boulder, Colorado, USA.

³European Centre for Medium-Range Weather Forecasts, Reading, UK.

⁴School of Mathematics and Statistics, University of Birmingham, Birmingham, UK.

⁵Remote Sensing Systems, Santa Rosa, California, USA.

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

⁷Lawrence Berkeley National Laboratory, Berkeley, California, USA.

change detection studies [Gillett *et al.*, 2003; Santer *et al.*, 2003a].

[4] The present paper focuses 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:

[5] 1. 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].

[6] 2. 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], particularly for less well-observed variables. (Sampling complex, time-evolving patterns of climate change with in situ observational networks that are spatially incomplete and vary over time can introduce nontrivial biases in global estimates of surface and atmospheric temperature changes [Duffy *et al.*, 2001; Santer *et al.*, 2000a].)

[7] 3. 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].

[8] These advantages must be weighed against several deficiencies:

[9] 1. Climate data from reanalyses, especially the “first generation” ERA-15 and NCEP-50 reanalyses, exhibit inhomogeneities related to temporal changes in the distribution, availability, and quality of assimilated satellite and radiosonde information [see, e.g., Källberg, 1997; Basist and Chelliah, 1997; Pawson and Fiorino, 1998; Santer *et al.*, 1999; Randel *et al.*, 2000; Trenberth *et al.*, 2001]. Observing system changes introduce spurious nonclimatic 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 [Pawson and Fiorino, 1999]. Inhomogeneities related to observing system changes are not restricted to the presatellite era [Trenberth *et al.*, 2001; Santer *et al.*, 2004].

[10] 2. 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 procedures used to adjust for biases in the assimilated data [Harris and Kelly, 2001], and the properties and implementation of the data assimilation system. In the latter case, this 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].

[11] 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 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 SAN03 (this point was noted earlier and assessed by Santer *et al.* [2003b] and SAN03), and argue that this discrepancy invalidates the SAN03 detection results.

[12] Here we address these criticisms using data from the second-generation ERA-40 reanalysis [Simmons and Gibson, 2000]. ERA-40 provides new estimates of atmospheric variability over more than four decades, based on a data assimilation system that is more advanced than those used for ERA-15 and NCEP-50 [Andersson *et al.*, 1998; Simmons and Hollingsworth, 2002]. The ERA-40 project was completed by ECMWF in 2003, and spans the period from September 1957 to August 2002. We employ monthly mean ERA-40 analyses to address the following scientific questions:

[13] 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 ERA-15 and NCEP-50 reanalyses?

[14] 2. Are the tropopause height increases in climate model and reanalysis data driven by similar large-scale changes in the temperature of the free atmosphere?

[15] 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?

[16] 4. Are decadal-timescale changes in p_{LRT} , the pressure of the lapse-rate tropopause, sensitive to the vertical resolution of the atmospheric temperature data used for calculating p_{LRT} ?

[17] The structure of this paper is as follows. In section 2, we provide a brief introduction to the reanalysis data sets employed in our detection study, and to the observational satellite data sets 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 tropo-

pause 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 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 data sets. 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

[18] 2.1 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 (<http://wesley.wvb.noaa.gov> (NCEP-50), <http://www.ecmwf.int/research/era/ERA-15> (ERA-15), and <http://www.ecmwf.int/research/era> (ERA-40)). 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 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. ERA-40 employs a hybrid sigma-pressure coordinate system with 60-level vertical resolution; 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).

[19] 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 assimilated retrievals based on a less-extensive set of satellite radiances that had undergone several preprocessing steps [*McNally et al.*, 2000]. These steps are not required in the case of direct assimilation of raw radiances.

[20] 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: MSU, the Stratospheric Sounding Unit (SSU), and the High-Resolution Infrared Radiation Sounder (HIRS). 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].

[21] 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 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 data sets 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).

[22] 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

[23] 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. (The maxima of the weighting functions for MSU channels 4 and 2 are at roughly 74 and 595 hPa.) 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 data sets 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.

[24] RSS and UAH use different procedures to adjust the raw MSU radiances for intersatellite biases, uncertainties in instrument calibration coefficients, changes in

Table 1. Statistics for Time Series of Changes in p_{LRT} , T4, and T2^a

		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
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

^aResults 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 (G) or spatial averages over 90°N–60°S (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., 2000b]. Owing to their high AR-1 values and small effective sample sizes (<5), 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.

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°C/decade in RSS, while the T2 trend in UAH is close to zero (Table 1). Discrepancies between the RSS and UAH T2 changes influence the detectability of model-predicted T2 fingerprints [Santer et al., 2003c].

2.3. Climate Model Data

[25] 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; T. M. L. Wigley et al., Effect of climate sensitivity on the response to volcanic forcing, submitted to *Journal of Geophysical Research*, 2004] and the differential responses to solar and greenhouse-gas forcing [Meehl et al., 2003].

[26] 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.

[27] 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

[28] 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°C/km, and then remains less than this critical value for a vertical distance of 2 km [World Meteorological Organization (WMO), 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

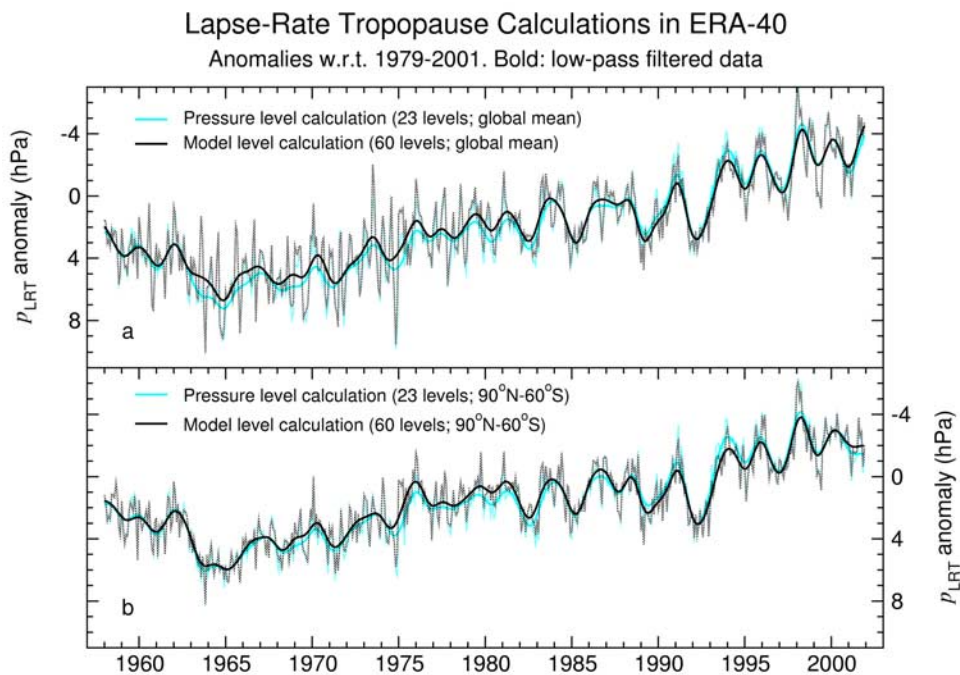


Figure 1. Effect of vertical resolution of the input atmospheric temperature data on estimates of global-scale p_{LRT} changes in ERA-40. p_{LRT} was calculated from ERA-40 temperatures archived at two different vertical resolutions: 23 pressure levels (L23) and 60 model levels (L60; see section 3.1). 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.

layers are present [Reichler *et al.*, 2003]. To avoid unrealistically high or low p_{LRT} values, search limits are restricted to pressure levels between roughly 600 and 75 hPa; the search proceeds upward from 600 hPa. The algorithm is applied in a consistent way to monthly mean profiles of atmospheric temperature in PCM, ECHAM, ERA-40, and NCEP-50.

[29] 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 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 (‘L23’) and the full 60 model levels (‘L60’). The latter data set has higher vertical resolution in the vicinity of the tropopause. Note that in the L60 case, model-level pressures were calculated from hybrid sigma-pressure coordinates using monthly mean values of surface pressure and the vertical coordinate definition specified by Källberg *et al.*, [2004]. Model-level pressures near the tropopause depend only weakly on surface pressure.

[30] The L60 and 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 (Figure 1a and Table 1). (A decrease in p_{LRT} signifies an increase in tropopause height.) 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 (Figures 2a and 2b), 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 (Figure 2). The mean height of the tropical tropopause (102.7 hPa for L60 and 105.7 hPa for L23) differs by less than 3% in the two calculations. (These values were computed using climatological annual-mean p_{LRT} data over 1979 to 2001, averaged over 20°N – 20°S .)

[31] 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 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 from L23 calculations.

[32] 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 Figure 3). The ‘high tropopause’ problem becomes less severe in the satellite era, when improved observational data constraints are introduced. The time-

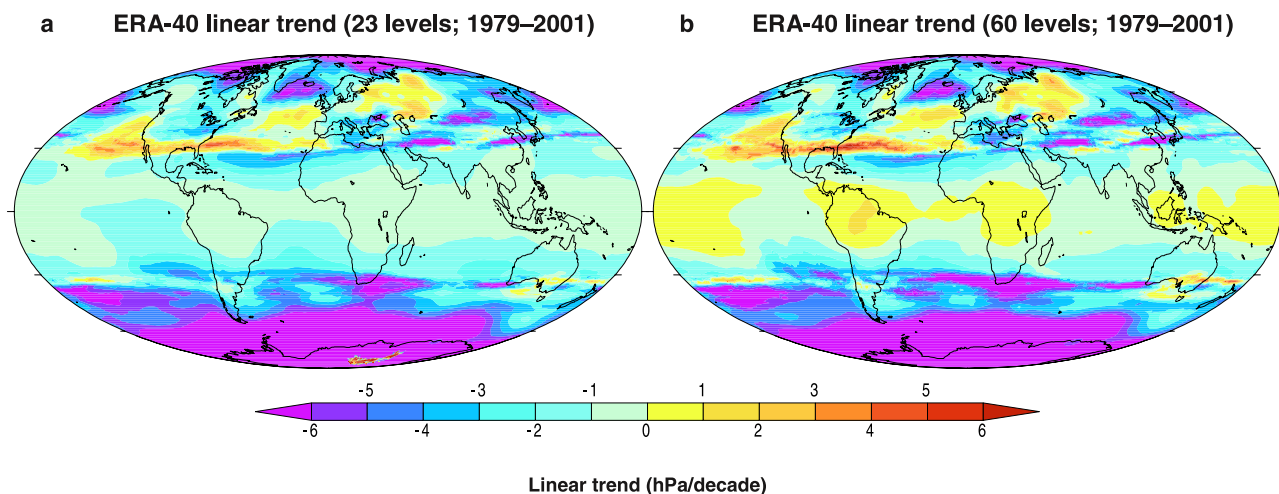


Figure 2. Effect of vertical resolution of the input temperature data on the estimated patterns of p_{LRT} change in ERA-40. p_{LRT} calculations were performed with both L23 and L60 atmospheric temperature data (a, b). Linear p_{LRT} trends over 1979–2001 were calculated using monthly mean anomaly data,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

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 (Figure 1a). For this reason, data poleward of 60S 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 (see Figures 1a and 1b), 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 (Figure 2).

[33] 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 very 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

[34] 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].

[35] 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

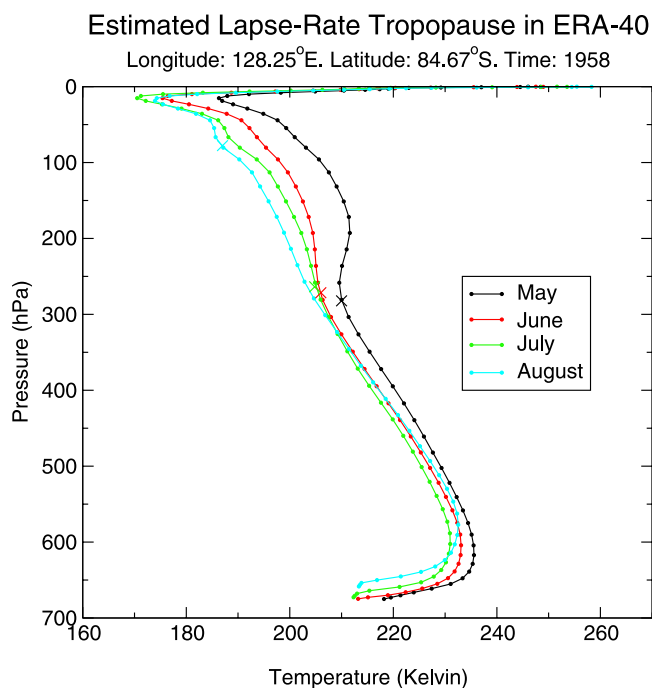


Figure 3. Atmospheric temperature profiles in ERA-40 at a selected Antarctic grid point (84.67°S, 128.25°E). Temperatures are monthly mean values for May, June, July, and August of 1958. Crosses denote the pressure of the lapse-rate tropopause estimated with the standard WMO criterion [WMO, 1957; Reichler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

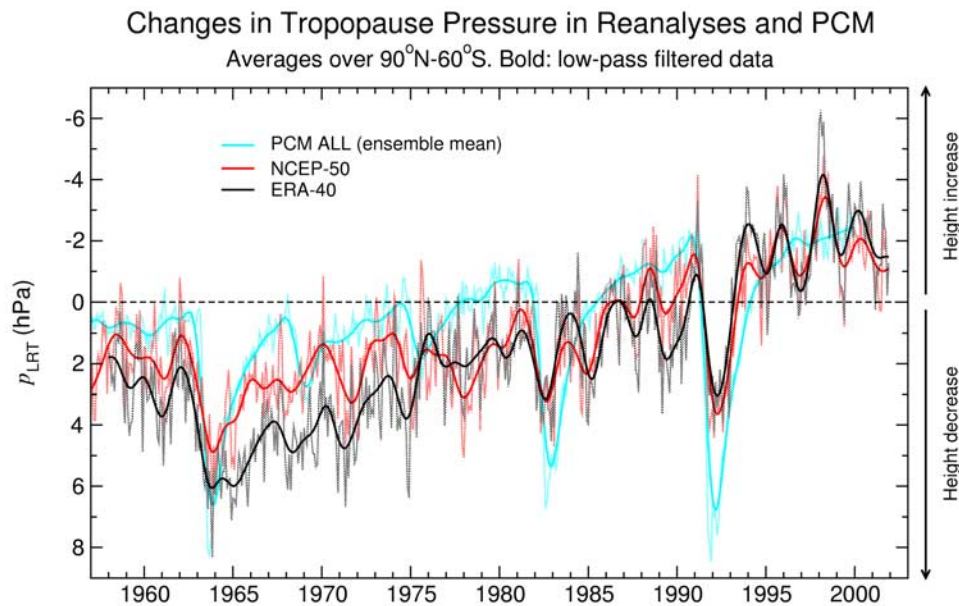


Figure 4. Time series of monthly mean p_{LRT} anomalies from the NCEP-50 and ERA-40 reanalyses and the ensemble mean of the PCM ALL experiment (section 2). Results are spatial averages over 90°N–60°S. Bold lines denote data that were low-pass filtered to highlight changes on 5–10 year timescales; thin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

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

[36] The height of the tropopause shows a sustained multidecadal increase since the early 1960s (Figure 4). This overall increase is evident in ERA-40, NCEP-50, and the PCM ALL experiment. (Note 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.) 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 (Figure 5). This is a common deficiency in models with coarse vertical resolution in the stratosphere [see, e.g., Bengtsson *et al.*, 1999].

[37] 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 presatellite 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 differences between reanalyses and PCM.

[38] ERA-40 shows a pronounced lower stratospheric warming in 1975 (Figure 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 for the period 1975 through mid-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 Figure 9 below).

[39] Figure 6 provides a simple conceptual model for interpreting the low- and high-frequency p_{LRT} changes shown in Figure 4 (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. (Depletion of stratospheric ozone cools the stratosphere and the troposphere. These changes have effects of opposite sign on tropopause height. The stratospheric cooling influence

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

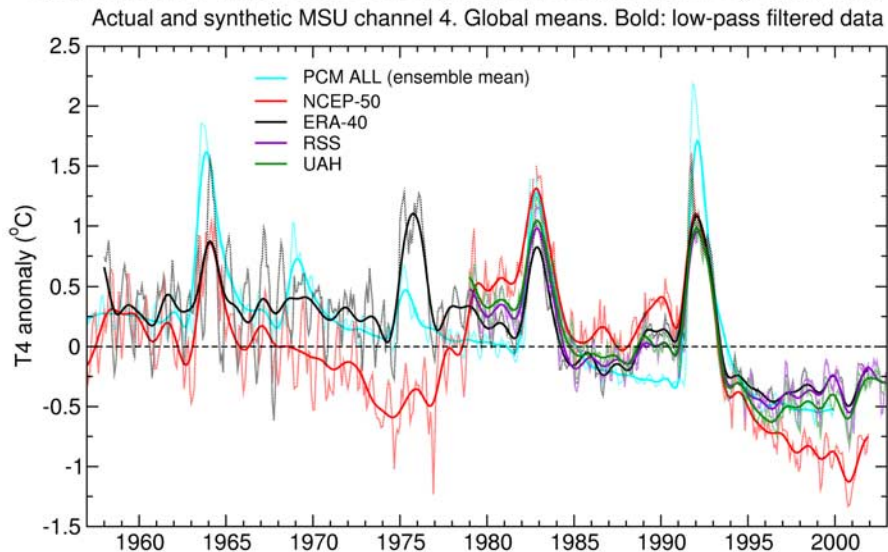


Figure 5. Time series of global mean, monthly mean anomalies in lower stratospheric temperatures (MSU T4). Results are processed MSU T4 measurements (UAH, RSS) and synthetic T4 temperatures calculated from NCEP-50, ERA-40, and the ensemble mean of the PCM ALL experiment (section 2.2). For definition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

dominates, 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.) In contrast, 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 (Figure 6).

[40] 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 Figure 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.

[41] 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. (Recall that the ALL experiment ends in 1999.) 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 m. The p_{LRT} decrease of 1.79 hPa/decade in NCEP-50 corresponds to a height increase of approximately 170 m. 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.

4.2. Spatial Patterns

[42] 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 (Figures 7a–7d). Both show increases in height 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

Effects of Different Forcing Mechanisms on Tropopause Height

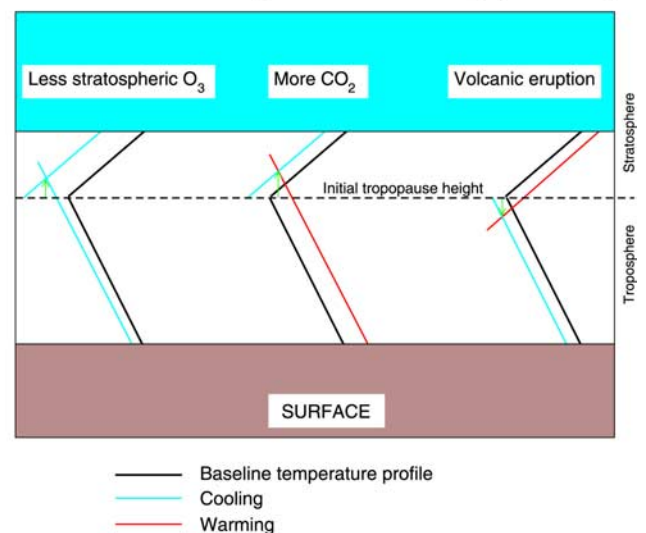


Figure 6. Conceptual model for the effect of three different forcings on tropopause height. The solid black lines are the baseline atmospheric temperature profiles. Forcing by stratospheric ozone depletion, increases in well-mixed greenhouse gases, and volcanic eruptions can perturb this base state.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

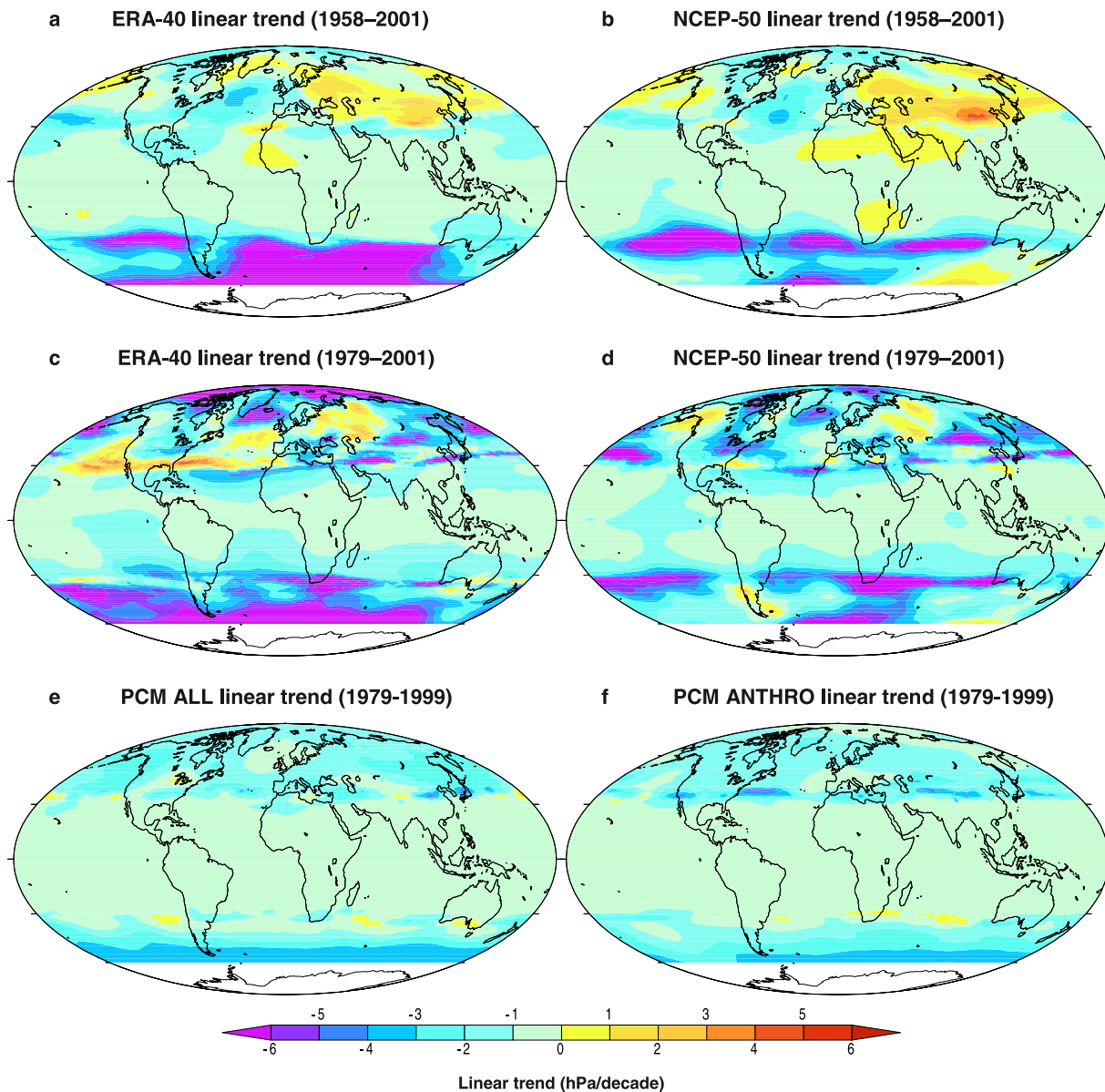


Figure 7. Tropopause pressure changes in reanalyses and PCM. Least squares linear trends in monthly mean p_{LRT} data (in hPa/decade) were computed over 1958–2001 for ERA-40 and NCEP-50 (a, b). ERA-40 and NCEP-50 trends are also shown for the shorter period 1979–2001 (c, d). For PCM, p_{LRT} trends over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

(Figures 7c and 7d), 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 (Figures 7a and 7b). 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 (compare Figures 7a and 7b and Figures 7c and 7d). There are also prominent differences off the coast of California (Figures 7c and 7d).

[43] 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 (compare Figures 7e and 7f and Figures 7c and 7d). As in ERA-40 and NCEP-50, increases are small and relatively unstruc-

tured 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 (Figures 7e and 7f) arises because both patterns are driven primarily by anthropogenic forcing, at least in PCM (see SAN03 and section 5).

5. Fingerprint Detection Results

[44] 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

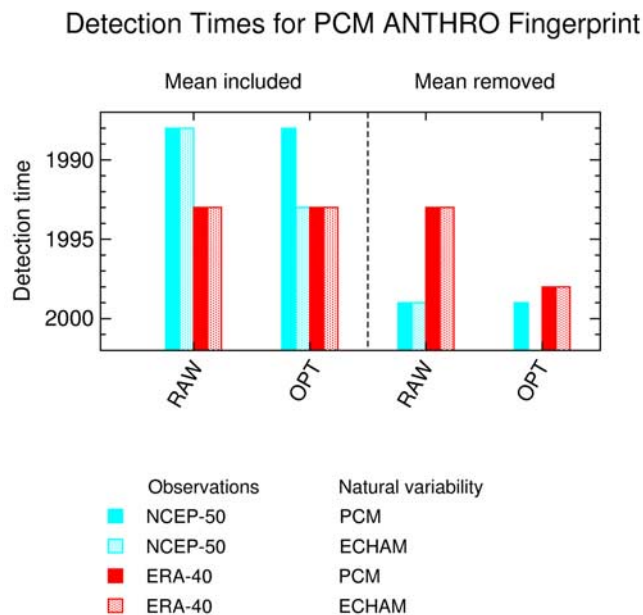


Figure 8. Detection times for PCM tropopause height fingerprints in NCEP-50 and ERA-40 reanalyses. The detection analysis uses both the ‘mean included’ and ‘mean removed’ fingerprints calculated from the PCM ANTHRO experiment, with a 5% significance level as the detection threshold (section 5). The longer the colored bar, the earlier the detection time. If no bar is present, the fingerprint could not be identified before the final year of the reanalyses (2001). ‘RAW’ denotes detection times for nonoptimized 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.

\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 the following: (1) Reanalysis data set (ERA-40 or NCEP-50). (2) 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. (3) The model control run (ECHAM or PCM) used for optimizing \vec{f} and assessing statistical significance. (4) 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 focuses 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].

[45] As noted previously, large, abrupt changes in the availability of satellite-based data on atmospheric temperature, moisture, and winds can introduce nonclimatic 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 problems of poorly defined and unrealistically high p_{LRT} values over the Antarctic continent (section 3.1). This area was included in the SAN03 fingerprint analysis, which used a spatial domain of 85°N – 85°S . A further difference relative to SAN03 relates to the fingerprint, which was defined here with the PCM ANTHRO experiment, while SAN03 relied on the PCM ALL experiment to specify \vec{f} (see below). These differences in the study area and searched-for fingerprint explain why the current study and SAN03 obtain slightly different detection times with the same NCEP-50 p_{LRT} data.

[46] 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 (Figure 8). For each reanalysis data set, 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 (see Figures 7c–7f).

[47] Positive detection of \vec{f} is not due solely to the large global mean height increases in PCM and reanalyses (Figure 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 raw ‘mean removed’ version of \vec{f} is identifiable six 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 (see Figures 7c and 7d), and are more similar to p_{LRT} changes in PCM. Removing the mean significantly degrades detection times for NCEP-50.

[48] 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 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

[49] As discussed above, both stratospheric cooling and tropospheric warming tend to increase tropopause height (Figure 6 and 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

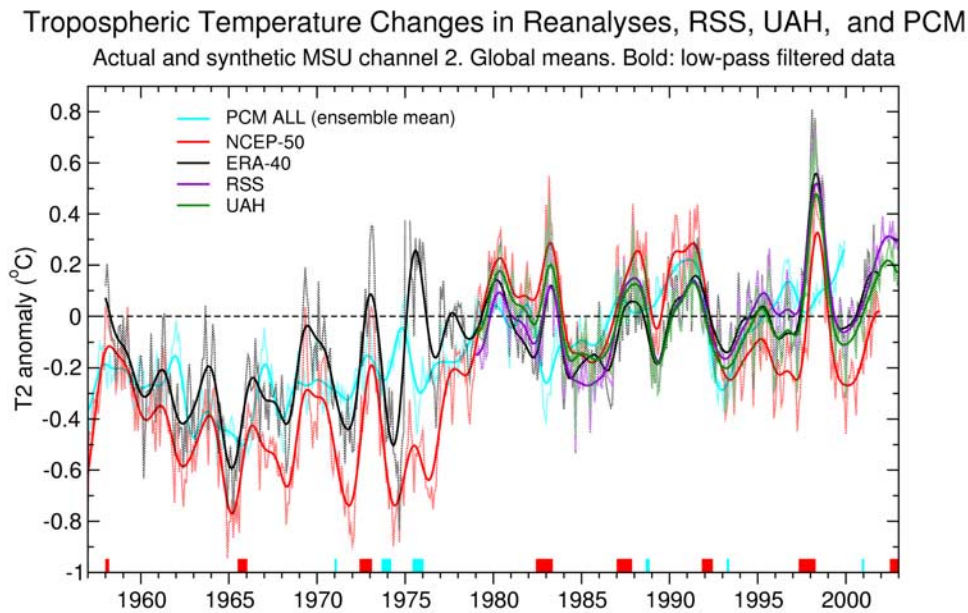


Figure 9. Time series of global mean, monthly mean anomalies in processed and synthetic tropospheric temperatures (MSU T2). For further details, refer to Figure 5. The colored rectangles on the time axis use a composite index of SST and circulation changes [Smith and Sardeshmukh, 2000] to indicate the 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].

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.

[50] 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 and Figures 5 and 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.

[51] It is also instructive to compare the low-frequency variability of T2 changes in ERA-40 and the ALL experiment (Figure 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. This constitutes a more stringent test of PCM's

performance than comparison of trends alone. Note that during 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 (Figure 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

[52] 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. For related comparisons of the statistical properties of atmospheric temperature changes in RSS, UAH, and various radiosonde data sets, refer to Seidel et al. [2004].

6.2.1. Temporal Correlations

[53] 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 time series in RSS and UAH are more highly correlated with each other than with the synthetic MSU time series from either reanalysis (Table 2). This conclusion holds for both T2 and T4 changes, and for correlations calculated with and

Tropospheric Temperature Changes: Consistency Between ERA-40 and PCM

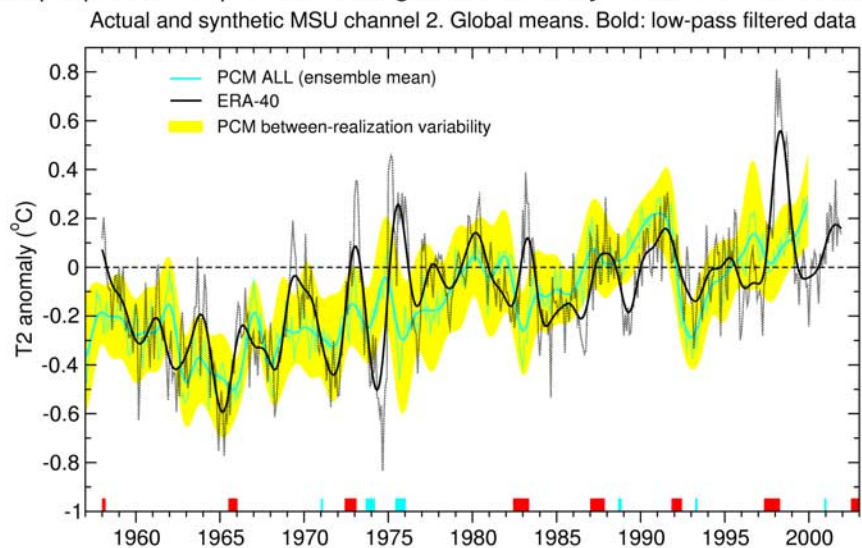


Figure 10. Consistency between changes in synthetic T2 temperatures in ERA-40 and the PCM ALL experiment. Bold lines denote the low-pass filtered T2 data in ERA-40 (black) and in the ALL ensemble mean (blue). The yellow envelope defines the range between the highest and lowest T2 anomalies in the four 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 Figure 4. The colored rectangles provide information on observed ENSO events (see Figure 9).

without the overall linear trend (which emphasize low- and high-frequency components of the time series, respectively).

[54] Another general result is that atmospheric temperature changes in the two observational MSU data sets 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-data set’ correlations always involve NCEP data (Table 2). This is probably due to spuriously large cooling in NCEP’s lower stratospheric temperatures (Figure 5 and 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 data set ($-0.11^{\circ}\text{C}/\text{decade}$; Figure 9 and Table 1).

6.2.2. Spatial Correlations

[55] Patterns of linear trends in T2 are qualitatively similar in ERA-40, RSS, and UAH (Figures 11a–11c). 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 data sets 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 (Figures 12a–12c). NCEP’s T2 and T4 changes (Figures 11d and 12d) are distinctly different from those in ERA-40 and the two satellite data sets, 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 (Figures 11e and 12e) are more similar to those in ERA-40, RSS and UAH than to NCEP’s trend patterns.

[56] 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

Table 2. Correlations Between Time Series of Global Mean Monthly Mean Atmospheric Temperature Anomalies for Four Different Data Sets^a

	UAH	RSS	NCEP-50	ERA-40
<i>T4 Results</i>				
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
<i>T2 Results</i>				
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

^aCorrelations 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).

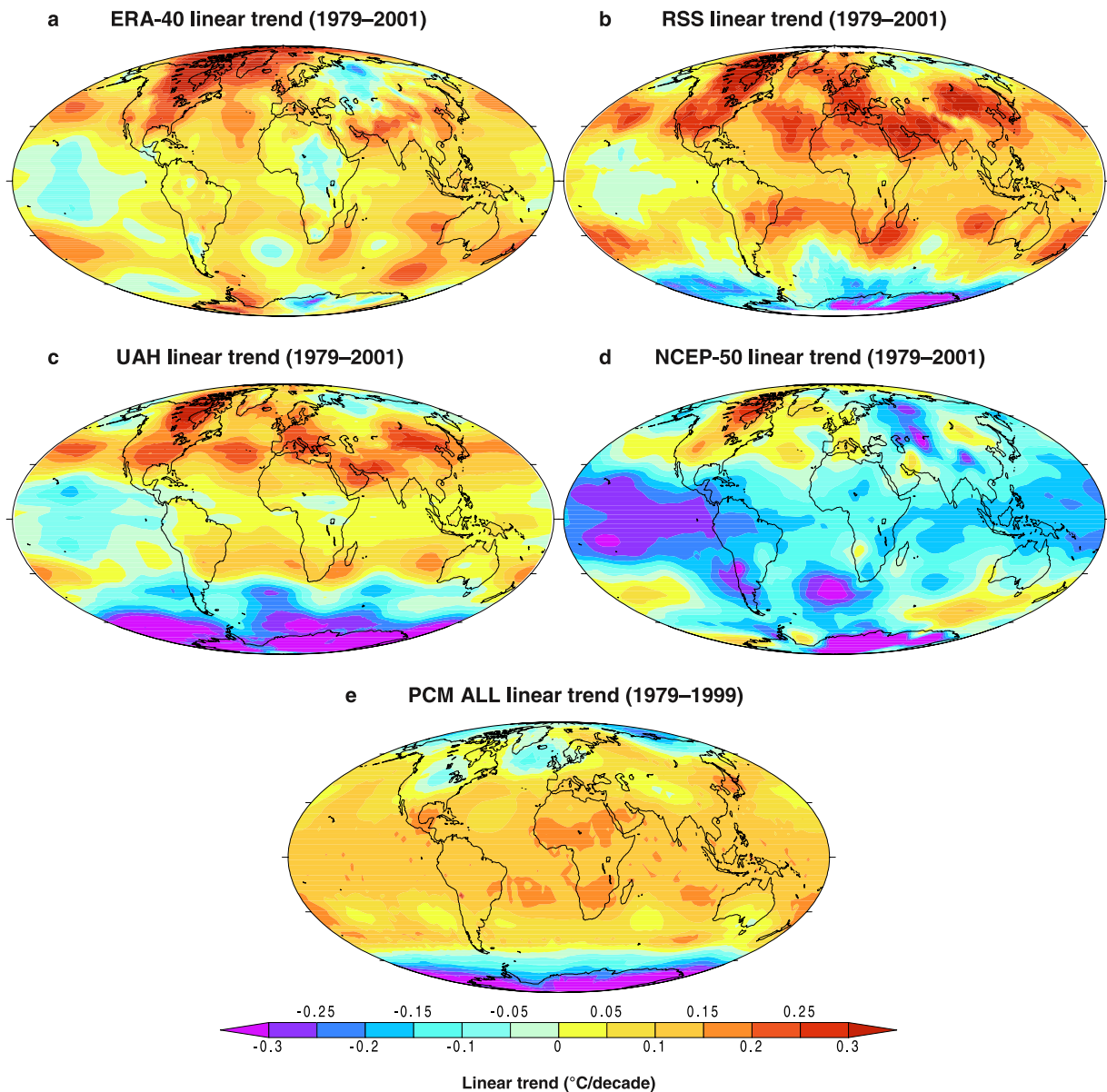


Figure 11. Tropospheric temperature changes in reanalyses and PCM. Least squares linear trends over 1979–2001 in monthly mean processed or synthetic MSU T2 temperatures from ERA-40 (a), RSS (b), UAH (c), and NCEP-50 (d). Also shown are T2 trends over 1979–1999 in the ensemble mean of the PCM ALL experiment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

pattern similarities and/or differences that may be obscured by global mean trend differences between two data sets. This is the case with T2 changes in the two reanalyses, for which $c_{\{NCEP:ERA\}} = -0.08$, while $r_{\{NCEP:ERA\}} = 0.74$. 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 ($c_{\{NCEP:ERA\}} = 0.86$, $r_{\{NCEP:ERA\}} = 0.41$; Figures 12a and 12d).

[57] 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 data set; (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

[58] 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.* [2003a] (“SAN03”). SAN03 claimed that they could identify a model-predicted “fingerprint” of externally forced tropo-

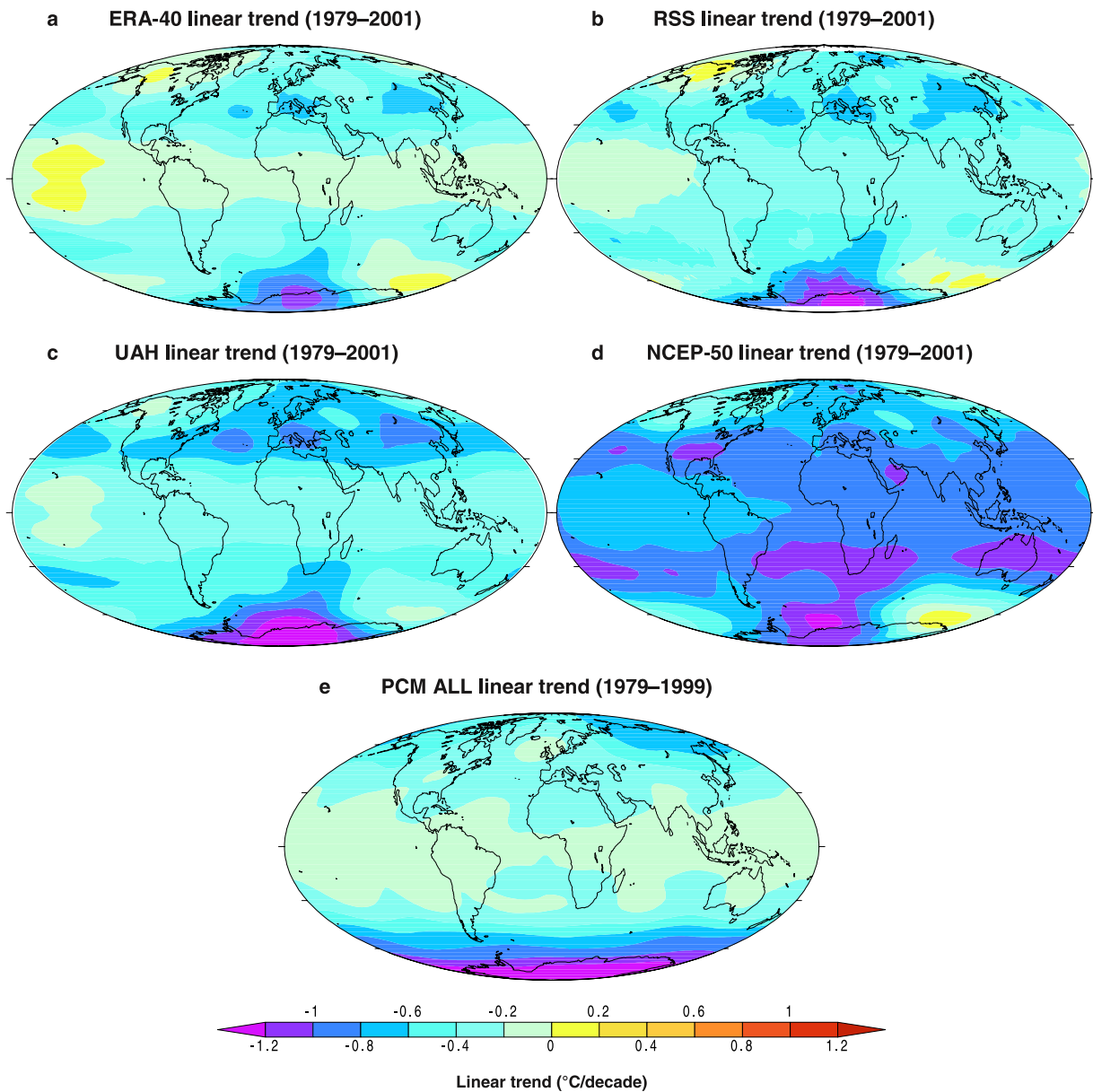


Figure 12. Stratospheric temperature changes in reanalyses and PCM. Least squares linear trends over 1979–2001 in monthly mean processed or synthetic MSU T4 temperatures from ERA-40 (a), RSS (b), UAH (c), and NCEP-50 (d). Also shown are T4 trends over 1979–1999 in the ensemble mean of the PCM ALL experiment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

pause 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).

[59] 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).

[60] The second question focused 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; Seidel *et al.*, 2001; Santer *et al.*, 2003b] (Figure 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].

Table 3. Correlations Between the Spatial Patterns of Atmospheric Temperature Change in Reanalysis and Satellite Data Sets^a

	UAH	RSS	NCEP-50	ERA-40
<i>T4 Results</i>				
UAH	1.00	0.99	0.91	0.98
RSS	0.97	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
<i>T2 Results</i>				
UAH	1.00	0.80	0.27	0.42
RSS	0.95	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

^aPattern correlations were calculated with the linear trend data in Figures 11 and 12 (for T2 and T4, respectively). All data sets were transformed to the RSS grid and masked with RSS coverage. Two forms of correlation are given: with spatial means included (*e*) and spatial means subtracted (*r* [Barnett and Schlesinger, 1987]). The latter are underlined.

[61] 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.

[62] 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 data sets. 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).

[63] 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 between 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 might lead to differences between the synthetic MSU temperatures in reanalyses and the MSU temperatures processed by RSS and UAH.

[64] 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.

[65] We note, however, that ERA-40 still manifests inhomogeneities, particularly the unrealistically large stratospheric warming in the mid-1970s, which is related to an error in the bias correction of the NOAA VTPR radiances (Figure 5). Comparisons with radiosonde data, and with an "AMIP-style" SST experiment performed with the ERA-40 model, indicate that the ERA-40 analyses are generally biased cold in the Southern Hemisphere prior to the availability of satellite sounding data. This is consistent with the global mean cold bias in the early years of ERA-40 inferred by Bengtsson *et al.* [2004]. This bias does not affect our detection analysis, which uses only 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.

[66] Even the post-1978 period, however, has data homogeneity problems. For example, the comparisons with radiosonde data and the above-mentioned AMIP simulation indicate that ERA-40's tropical temperatures at 100 hPa and neighboring levels 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 (Figure 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).

[67] 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 (Figures 2a and 2b). Our detection conclusions are not sensitive to this source of uncertainty. (We 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)).

[68] 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 observed 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.

Appendix A: Forcings and Model Details

[69] 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 SO_2 emissions is time invariant (except over the seasonal cycle), and can be scaled by estimates of historical changes in global mean SO_2 emissions. This 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 SO_2 emissions are larger over the 20th century than over the past 25 years.

[70] 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].

[71] 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 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

B1. Definition of Fingerprint

[72] 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 Figure 1). We define \vec{f} as the first EOF, which explains a substantial fraction of the overall variance of $\vec{s}(t)$: 62% for the ‘mean included’ analysis, and 33% for the ‘mean removed’ case.

[73] Note that our reanalysis data sets end in 2001, while the ANTHRO experiment ends in 1999. This slight mismatch in the time periods covered by ANTHRO, ERA-40, and NCEP-50 does not unduly affect our analysis. Extending the ANTHRO run by two years (i.e., by extending the applied anthropogenic forcings from January 2000 through to December 2001) would result in only minor changes to the fingerprint, since \vec{f} is defined over the full period of the ANTHRO experiment, and primarily captures the large anthropogenically forced changes in the mean height of the tropopause over 1890 to 1999. Our fingerprint analysis searches for an increasing expression of this (time-invariant) mean change pattern in time-varying reanalysis p_{LRT} data. 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].

B2. Estimation of Detection Time

[74] 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.

[75] 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.

[76] 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 Figure 8 shows results for the $m = 15$ case only.

[77] 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 data sets with more temporal structure has employed both spatial and temporal information for fingerprint optimization [Stott *et al.*, 2000].

Signal-to-Noise Ratios for PCM and Reanalysis Tropopause Height Data

Start date for signal trend calculations: 1979

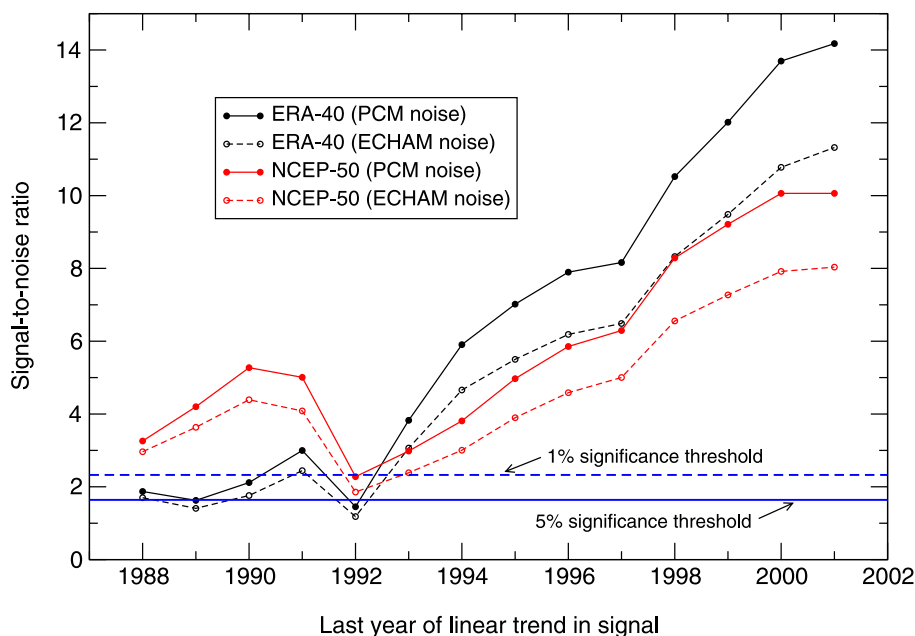


Figure B1. Sensitivity of estimated detection times to choice of significance threshold. Signal-to-noise ratios (SNRs) used for the definition of detection times were computed as described in Appendix B, using the raw (i.e., nonoptimized) PCM ANTHRO fingerprint, with spatial means included. This fingerprint is searched for in both ERA-40 and NCEP-50 p_{LRT} data. ‘Signals’ are trends in the test statistic $Z(t)$. Both PCM and ECHAM control run data were used to estimate noise trends arising from natural internal variability. The nominal 1% and 5% significance thresholds are indicated by dashed and solid blue lines (respectively). The assumed start date for calculation of signal trends is 1979, and the assumed minimum length of record for characterizing the signal is 10 years. SNR results are plotted on the final year of the signal trend.

Optimization leaves detection times unchanged in four of the eight cases shown in Figure 8, and actually degrades detection time in the other four cases.

[78] One possible explanation for the failure of optimization to improve detection times is that important components of the fingerprint may be lost in projecting \vec{f} onto the subspace of 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.

B3. Analysis With Mean Removed

[79] 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 $\vec{c}_1(t)$.

B4. Sensitivity to Significance Level

[80] Our fingerprint analysis uses a nominal 5% significance threshold for estimating detection times. Our conclusions regarding the detectability of the PCM ANTHRO fingerprint do not depend on this choice. There is, however, a weak sensitivity of the estimated detection times to the stipulated significance threshold.

[81] For example, in the case where the raw, ‘mean included’ fingerprint is searched for in NCEP-50 data, use

of a 5% significance threshold yields detection in 1988; detection is achieved five years later (in 1993) with a more conservative 1% significance threshold (Figure 8). This sensitivity arises because of the large effect of Pinatubo on signal trends ending in 1992 (Figure B1). Signal-to-noise ratios (SNRs) for NCEP-50 data are consistently above the 5% significance threshold for signal trends of all lengths L (see section B2 above), but dip below the 1% significance level in 1992, when Pinatubo’s influence leads to a decrease in tropopause height (Figures 4 and 6).

[82] Detection times obtained for ERA-40 data do not show a similar sensitivity, since ERA-based SNRs are initially lower than NCEP-based SNRs, and do not continuously exceed either the 5% or 1% significance thresholds until 1993, when the effect of Pinatubo on signal trends diminishes (Figure B1). These systematic differences in the initial SNR levels obtained with NCEP-50 and ERA-40 data may be related to ERA-40’s likely underestimation of the stratospheric cooling over 1979 to the mid-1980s (see section 7).

[83] **Acknowledgments.** Work at Lawrence Livermore National Laboratory (LLNL) was supported under the auspices of the Office of Science, U.S. Department of Energy, at the University of California Lawrence Livermore National Laboratory, under 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 data sets (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 NA87GP0105, and from the U.S. Department of Energy, grant DE-FG02-98ER62601. A portion of this study was supported by the U.S. Department 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 (University 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 two anonymous reviewers for their comments. Gary Strand, Julie Arblaster, Lawrence Buja, and Adrienne Middleton (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 data set in simulations of 20th century climate, *Geophys. Res. Lett.*, *30*(12), 1657, doi:10.1029/2003GL016875.
- Andersson, E., et al. (1998), The ECMWF implementation of three-dimensional variational assimilation (3D-Var). Part III: Experimental results, *Q. J. R. Meteorol. Soc.*, *124*, 1831–1860.
- Andrae, U., N. Sokka, and K. Onogi (2004), ERA-40 Project report series. 15. The radiosonde temperature bias corrections used in ERA-40, 37 pp., Eur. Cent. for Medium-Range Weather Forecasts, Reading, U. K.
- 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. Weather Rev.*, *127*, 1157–1186.
- Barnett, T. P., and M. E. Schlesinger (1987), Detecting changes in global climate induced by greenhouse gases, *J. Geophys. Res.*, *92*, 14,772–14,780.
- 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. Am. Meteorol. Soc.*, *7*, 1431–1447.
- Bengtsson, L., and J. Shukla (1988), Integration of space and in situ observations to study global climate change, *Bull. Am. Meteorol. 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.
- Bengtsson, L., S. Hagemann, and K. I. Hodges (2004), Can climate trends be calculated from reanalysis data?, *J. Geophys. Res.*, *109*, D11111, doi:10.1029/2004JD004536.
- 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. Oceanic Technol.*, *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, *Tech. Memo. 377*, 37 pp., Eur. Cent. for Medium-Range Weather Forecasts, Reading, U. K.
- 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, K., et al. (1997), ECMWF Reanalysis Project report series. 1. ERA description, 66 pp., Eur. Cent. for Medium-Range Weather Forecasts, Reading, U. K.
- 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., et al. (1984), Climate sensitivity: Analysis of feedback mechanisms, in *Climate Processes and Climate Sensitivity*, *Geophys. Monogr. Ser.*, vol. 29, edited by J. Hansen and T. Takahasi, pp. 130–163, AGU, Washington, D. C.
- Hansen, J., et al. (2002), Climate forcings in Goddard Institute for Space Studies S12000 simulations, *J. Geophys. Res.*, *107*(D18), 4347, doi:10.1029/2001JD001143.
- Harris, B. A., and G. A. Kelly (2001), A satellite radiance bias correction scheme for data assimilation, *Q. J. R. Meteorol. Soc.*, *127*, 1453–1468.
- Hasselmann, K. (1979), On the signal-to-noise problem in atmospheric response studies, in *Meteorology of Tropical Oceans*, edited by D. B. Shaw, pp. 251–259, R. Meteorol. Soc., London.
- 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. Meteorol. Soc.*, *124*, 1579–1604.
- Highwood, E. J., B. J. Hoskins, and P. Berrisford (2000), Properties of the Arctic tropopause, *Q. J. R. Meteorol. 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*, 18,895–18,906.
- Källberg, P. (1997), ECMWF Reanalysis Project report series. 2. Aspects of the reanalysed climate, 89 pp., Eur. Cent. for Medium-Range Weather Forecasts, Reading, U. K.
- Källberg, P., A. Simmons, S. Uppala, and M. Fuentes (2004), ERA-40 Project report series. 17. The ERA-40 archive, 35 pp., Eur. Cent. for Medium-Range Weather Forecasts, Reading, U. K.
- Kalnay, E., et al. (1996), The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. 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. Weather Rev.*, *119*, 1866–1880.
- Khariin, 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.
- 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*, 31,239–31,254.
- Kistler, E., et al. (2001), The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation, *Bull. Am. Meteorol. Soc.*, *82*, 247–267.
- 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.
- McNally, A. P., J. C. Derber, W. Wu, and B. B. Katz (2000), The use of TOVS level-1b radiances in the NCEP SSI analysis system, *Q. J. R. Meteorol. Soc.*, *126*, 689–724.
- 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. Weather 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 presatellite data era, *Clim. Dyn.*, *15*, 241–250.
- Pielke, R. A., Sr., and T. N. Chase (2004), Comment on "Contributions of anthropogenic 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., et al. (2001), Stratospheric temperature trends: Observations and model simulations, *Rev. Geophys.*, *39*, 71–122.
- Randel, W. J., F. Wu, and D. Gaffen (2000), Interannual variability of the tropical tropopause derived from radiosonde data and NCEP reanalyses, *J. Geophys. Res.*, *105*, 15,509–15,523.
- Reichler, T., M. Dameris, and R. Sausen (2003), Determining the tropopause height from gridded data, *Geophys. Res. Lett.*, *30*(20), 2042, 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., et al. (1995), Ocean variability and its influence on the detectability of greenhouse warming signals, *J. Geophys. Res.*, *100*, 10,693–10,725.
- Santer, B. D., et al. (1999), Uncertainties in observationally based estimates of temperature change in the free atmosphere, *J. Geophys. Res.*, *104*, 6305–6333.
- Santer, B. D., et al. (2000a), Interpreting differential temperature trends at the surface and in the lower troposphere, *Science*, *287*, 1227–1232.
- Santer, B. D., et al. (2000b), Statistical significance of trends and trend differences in layer-average atmospheric temperature time series, *J. Geophys. Res.*, *105*, 7337–7356.
- Santer, B. D., et al. (2003a), Contributions of anthropogenic and natural forcing to recent tropopause height changes, *Science*, *301*, 479–483.
- Santer, B. D., et al. (2003b), Behavior of tropopause height and atmospheric temperature in models, reanalyses, and observations: Decadal changes, *J. Geophys. Res.*, *108*(D1), 4002, doi:10.1029/2002JD002258.
- Santer, B. D., et al. (2003c), Influence of satellite data uncertainties on the detection of externally forced climate change, *Science*, *300*, 1280–1284.
- Santer, B. D., et al. (2004), Response to Comment on “Contributions of anthropogenic and natural forcing to recent tropopause height changes”, *Science*, *303*, 1771c.
- Sausen, R., and B. D. Santer (2003), Use of changes in tropopause height to detect human influences on climate, *Meteorol. 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.
- Seidel, D. J., et al. (2004), Uncertainty in signals of large-scale climate variations in radiosonde and satellite upper-air temperature data sets, *J. Clim.*, *17*, 2225–2240.
- Simmons, A. J., and J. K. Gibson (2000), ERA-40 Project report series. 1. The ERA-40 project plan, 62 pp., Eur. Cent. for Medium-Range Weather Forecasts, Reading, U. K.
- 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. Climatol.*, *20*, 1543–1557.
- Sperber, K. R. (2003), Propagation and vertical structure of the Madden-Julian Oscillation, *Mon. Weather Rev.*, *131*, 3018–3037.
- Stott, P. A., et al. (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., and J. G. Olson (1988), An evaluation and intercomparison of global analyses from NMC and ECMWF, *Bull. Am. Meteorol. Soc.*, *69*, 1047–1057.
- Trenberth, K. E., D. P. Stepaniak, J. W. Hurrell, and M. Fiorino (2001), Quality of reanalyses in the tropics, *J. Clim.*, *14*, 1499–1510.
- Washington, W. M., et al. (2000), Parallel Climate Model (PCM) control and transient simulations, *Clim. Dyn.*, *16*, 755–774.
- Wigley, T. M. L. (2000), ENSO, volcanoes, and record-breaking temperatures, *Geophys. Res. Lett.*, *27*, 4101–4104.
- 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.
- World Meteorological Organization (WMO) (1957), Meteorology—A three-dimensional science: Second session of the commission for aerology, *WMO Bull.*, *IV*(4), 134–138.
- C. Ammann, G. A. Meehl, W. M. Washington, and T. M. L. Wigley, National Center for Atmospheric Research, Boulder, CO 80303, USA.
- J. S. Boyle, C. Doutriaux, M. Fiorino, B. D. Santer, and K. E. Taylor, Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA. (santer1@llnl.gov)
- W. Brüggemann, School of Mathematics and Statistics, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK.
- P. W. Källberg, G. A. Kelly, A. J. Simmons, and S. M. Uppala, European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading RG2 9AX, UK.
- C. Mears and F. J. Wentz, Remote Sensing Systems, 438 First Street, Suite 200, Santa Rosa, CA 95401, USA.
- R. Sausen, Institut für Physik der Atmosphäre, Deutsches Zentrum für Luft- und Raumfahrt, Oberpfaffenhofen, D-82234 Wessling, Germany.
- M. F. Wehner, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.