Contributions of Anthropogenic and Natural Forcing to Recent Tropopause Height Changes

B.D. Santer,[1 *] M.F. Wehner,[2] T.M.L. Wigley,[3] R. Sausen,[4] G.A. Meehl,[3] K.E. Taylor,[1] C. Ammann,[3] J. Arblaster,[3] W.M. Washington,[3] J.S. Boyle,[1] W. Brüggemann[5]

[1] Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA. [2] Lawrence Berkeley National Laboratory Berkeley, CA 94720, USA. [3] National Center for Atmospheric Research, Boulder, CO 80303, USA. [4] Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, D-82234 Wessling, Germany. [5] University of Birmingham, Edgbaston, Birmingham B15 2TT, U.K.

Submitted to *Science*

March 3, 2003

* To whom correspondence should be addressed. Email: santer1@llnl.gov

Abstract

Observations indicate that the average height of the tropopause – the transition zone between the stratosphere and troposphere – has increased by several hundred meters since 1979. Comparable increases are evident in climate model experiments. The latter show that human-induced changes in ozone and well-mixed greenhouse gases account for over 80% of the simulated rise in tropopause height over 1979-1999. Their primary contributions are through cooling of the stratosphere (ozone) and warming of the troposphere (well-mixed greenhouse gases). Tropopause height changes simulated over 1900-1949 are smaller than in recent decades, and are driven largely by variations in volcanic aerosols and solar irradiance. A model-predicted fingerprint of tropopause height changes is statistically detectable in two different observational ("reanalysis") datasets. This positive detection result allows us to attribute overall tropopause height changes to a combination of anthropogenic and natural forcings. Our study shows that the increase in tropopause height over the second half of the 20th century was predominantly due to human activity, and provides independent support for claims of recent tropospheric warming.

The tropopause is the transition zone between the troposphere and stratosphere, marked by large changes in the thermal, dynamical, and chemical structure of the atmosphere (*1*, *2*). Increases in tropopause height over the last several decades have been identified in radiosonde data (*2*), in observationally-constrained numerical weather forecasts (reanalyses) (*3*), and in climate models forced by combined natural and anthropogenic effects (*4*). Model experiments suggest that this increase cannot be explained by natural climate variability alone (*4*).

To date, no study has quantified the contributions of different anthropogenic and natural forcing mechanisms to tropopause height changes over the 20th century. We estimate these contributions here, and demonstrate the usefulness of tropopause height as an integrated indicator of human-induced climate change. We also make the first identification of a model-predicted "fingerprint" of tropopause height changes in reanalysis data. Detection of this fingerprint enables us to attribute tropopause height changes to the combined effects of anthropogenic and natural forcing.

Our anthropogenic forcings are changes in well-mixed greenhouse gases (G), the direct scattering effects of sulfate aerosols (A), and tropospheric and stratospheric ozone (O). The natural external forcings considered are changes in solar irradiance (S) and volcanic aerosols (V). All of these factors are likely to have modified the thermal structure and static stability of the atmosphere (*4–8*), thus affecting tropopause height. To isolate the thermal responses that drive tropopause height changes, we analyze the effects of G, A, O, S, and V on temperatures averaged over broad layers of the stratosphere and troposphere.

We employ the Department of Energy Parallel Climate Model (PCM) developed by the National Center for Atmospheric Research and Los Alamos National Laboratory (*9*). PCM has been used to perform a wide range of experiments, of which seven have been analyzed specifically for tropopause height changes. In the first five experiments, only a single forcing is changed at a time – *e.g.*, G varies according to historical estimates of greenhouse-gas concentration changes, while A, O, S, and V are all held constant at pre-industrial levels (*10*, *11*). In the sixth experiment (ALL), all five forcings are varied simultaneously. Both natural forcings are changed in the seventh experiment (SV). G and A commence in 1872, while O, S, V, SV, and ALL start in 1890. The experiments end in 1999. To obtain better estimates of the underlying responses to the imposed forcings, four realizations of each experiment were performed. Each realization of a given experiment has identical forcing, but commences from different initial conditions of an unforced control run.

The detection part of our study requires estimates of observed tropopause height changes, which are obtained here from two reanalyses (*12*, *13*). The first is from the National Center for Environmental Prediction and the National Center for Atmospheric Research (NCEP), the second from the European Centre for Medium-Range Weather Forecasts (ERA). Reanalyses use an atmospheric numerical weather forecast model with no changes over time in either the model itself or in the observational data assimilation system (*14*). NCEP data were available from 1948 to 2001, but data prior to January 1979 were excluded due to well-documented inhomogeneities (*14*). ERA spans the shorter period from 1979 to 1993.

To estimate tropopause height in PCM, NCEP, and ERA, we use a standard thermal definition of p_{LRT}, the pressure of the lapse-rate tropopause. At each grid-point, we compute p_{LRT} from the monthly-mean temperature profile at discrete pressure levels (*4*). For PCM, we also calculate stratospheric and tropospheric temperatures equivalent to those monitored by channels 4 and 2 ("T4", "T2") of the satellite-based Microwave Sounding Unit (MSU) (*14*, *15*). The simulated MSU data are useful in relating changes in p_{LRT} to broad changes in atmospheric thermal structure. Direct comparisons with observed MSU T4 and T2 data are given elsewhere (*16*).

For any specified forcing, the ensemble members define an 'envelope' of climate trajectories (Fig. 1A). The time at which the ALL and SV envelopes separate completely (and remain separated) is a simple qualitative measure of the detectability of an anthropogenic signal in PCM data. For tropopause height, this separation occurs in the mid-1980s. Divergence of the ALL and SV solutions occurs earlier for T4 data (before the eruption of Mt. Agung in 1963; Fig. 1C) and later for T2 (in the early 1990s; Fig. 1E).

In ALL, the total linear p_{LRT} decrease over 1979-1999 is 2.9 hPa, corresponding to a tropopause height increase of roughly 120 meters (Fig. 1A). Tropopause height in NCEP rises by *ca.* 190 meters. The smaller increase in PCM is due primarily to the model's overestimate of volcanically-induced stratospheric cooling, and hence tropopause height changes. Large explosive volcanic eruptions warm the lower stratosphere and cool the troposphere (Figs. 1C,E), leading to a decrease in tropopause height and an increase in p_{LRT} (*4*). The short-term (3-4 year) p_{LRT} signatures of major volcanic eruptions are clearly evident in reanalyses and the SV and ALL experiments (Fig. 1A). In PCM, p_{LRT} changes after the eruptions of Santa Maria, Agung, El Chichón, and Pinatubo are large relative to both the 'between realization' variability of p_{LRT} and the interdecadal p_{LRT} variability in volcanically-quiescent periods.

The largest contributions to the simulated tropopause height increases are from changes in well-mixed greenhouse-gases (GHGs) and ozone (Fig. 1B). To quantify these contributions, we examine the total linear p_{LRT} trends in the ensemble means of the G, A, O, S, and V experiments (*17*). Linear trends are computed over four different time horizons: the 20th century, 1900-1949, 1950-1999, and the satellite era (1979-1999; Fig. 2A). Since some forcings increase tropopause height and others lower it, the contributions of individual forcings are expressed as percentages of the sum of absolute changes in p_{LRT} (*18*).

Over the 20th century, anthropogenic forcings (G, A, and O) account for 80% of the total trend in tropopause height. Well-mixed GHGs and ozone explain 31%

and 36% of this change (respectively). The relative contributions of G and O vary with time. While G and O make roughly equivalent contributions to tropopause height change over 1900-1949, ozone becomes more important in the second half of the century and over the satellite era (Fig. 2A). The contributions of natural external forcings (S and V) also change markedly with time. S and V explain 60% of the total tropopause height change over 1900-1949, but only 24% of the p_{LRT} change over 1950-1999. The large influence of V over 1900-1949 arises from the Santa Maria eruption in 1902 (Fig. 1A).

Decadal-scale increases in the height of the tropopause are driven by lapse-rate changes above and below the tropopause. Lapse rates in this region are influenced primarily by ozone- and GHG-induced cooling of the stratosphere and GHG-induced warming of the troposphere. Both of these effects tend to raise tropopause height (*4*). A key question, therefore, is whether the simulated height increase in ALL could have occurred without significant anthropogenically-induced warming of the troposphere. We address this question by estimating the contributions of individual forcings to T4 and T2 changes.

Previous model-based work suggests that ozone forcing is the major driver of recent stratospheric cooling (*7*, *19*). A similar result holds in PCM: ozone changes account for 68% of the total linear change in T4 over 1950-1999, and 78% of the T4 decrease during the satellite era (Figs. 1D, 2B). The corresponding contributions of GHG forcing to T4 changes are only 5% and 2%, respectively (*4*). In contrast, well-mixed GHGs are the major contributor to PCM's tropospheric warming, and explain 63% of the total absolute change in T2 over 1950-1999 (Figs. 1F, 2C). The ozone component of T2 changes is small (*20*). In PCM, therefore, the main effect of well-mixed GHGs on tropopause height is through warming of the troposphere rather than cooling of stratosphere.

Several additional features are noteworthy. Anthropogenic sulfate aerosols decrease tropopause height by cooling the troposphere (Figs. 1F, 2C). In the model, sulfate aerosols have little impact on stratospheric temperature (Figs. 1D, 2B). Solar irradiance changes over the 20th century warm both the troposphere and the stratosphere with offsetting effects on tropopause height. The sign of the solar effect on p_{LRT} must therefore depend on the relative magnitudes of solar-induced stratospheric and tropospheric warming. The small rise in tropopause height in S (Fig. 1B) suggests that for solar forcing, tropospheric warming is more important.

A key assumption in many detection studies is that the sum of the individual climate responses to several different forcing mechanisms is equal to the response obtained when these forcings are varied simultaneously (*5*, *6*, *21*). This implies that there are no strong interactions between individual forcings. We tested this assumption for p_{LRT}, T4, and T2 by comparing ALL results with the sum of responses to G, A, O, S, and V (SUM). ALL and SUM show very similar global-mean changes

(Figs. 1B,D,F). For the three variables considered here, the estimated linear changes in SUM are within 10% (*22*) of the corresponding ALL values (Fig. 2). For these global-scale changes, additivity is a reasonable assumption.

We next use a standard detection method (*23*, *24*) to determine whether a model-predicted spatial pattern of externally-forced p_{LRT} changes can be identified in reanalysis data. Our detection method assumes that the searched-for signal (the "fingerprint", $\vec{f}$) is the first Empirical Orthogonal Function (EOF) of the ALL ensemble mean (*16*). We search for an increasing expression of $\vec{f}$ in the NCEP and ERA p_{LRT} data, and estimate the "detection time" – the time at which $\vec{f}$ becomes consistently identifiable at a stipulated 5% significance level. The climate noise estimates required for assessing statistical significance are obtained from 300-year control runs performed with PCM and the ECHAM model of the Max-Planck Institute for Meteorology in Hamburg (*25*).

Detection times are computed both with and without the global-mean component of tropopause height change (*16*). If the spatial mean is included, detection results can be dominated by large global-scale changes. Removing the spatial mean focusses attention on smaller-scale model-data pattern similarities, and provides a more stringent test of model performance.

We first describe the patterns of tropopause height change in PCM and reanalyses. Maps of linear p_{LRT} trends over 1979-1999 are similar in NCEP and PCM (Figs. 3A,C).

Both show spatially-coherent increases in tropopause height (*i.e.*, negative trends in p_{LRT}), with smallest changes in the tropics and largest increases towards the poles, particularly in the Southern Hemisphere (*4*). Tropopause height changes in ERA (Fig. 3B) are less coherent than in NCEP or PCM, with slight decreases in height in the tropics. This largely reflects the shorter record length of ERA, which ends at a time when Pinatubo decreased tropopause height (Figs. 1A,B). The ALL 'mean included' p_{LRT} fingerprint is similar to the PCM and NCEP linear trend patterns, with uniform sign, largest loadings at high latitudes in the Southern Hemisphere, and strong zonal structure (Fig. 3D). Removal of the spatial means emphasizes the strong equator-to-pole gradients and hemispheric asymmetry of the fingerprint (Fig. 3E).

When our detection method is applied with spatial means included, the ALL tropopause height fingerprint is consistently identifiable in reanalyses. Although ERA tropopause height increases are less coherent than in NCEP, and visually less similar to the ALL fingerprint, $\vec{f}$ is still readily detectable. In all four cases considered (*26*), fingerprint detection occurs in 1988, only 10 years after the start of our analysis period (1979). With our strategy, 1988 is the earliest time at which detection can be achieved.

Removal of spatial means still yields an identifiable fingerprint, but only in NCEP. This positive result arises from model-data similarities in the equator-to-pole gradient and hemispheric asymmetry of tropopause height changes (Figs. 3A,E). Detection of

$\vec{f}$ in NCEP data occurs in 1995 (*27*), seven years later than in the 'mean included' case, and after the end of the ERA record. This suggests that the ERA record is simply too short to achieve positive detection of sub-global features of the predicted tropopause height changes.

Our results are relevant to the issue of whether the 'real world' troposphere has warmed over the satellite era (*16*). PCM provides both direct and indirect evidence in support of a warming troposphere. The direct evidence is that in the ALL experiment, the troposphere warms by 0.07°C/decade over 1979-1999 (*16*). Over 60% of this warming is due to increases in well-mixed GHGs (Fig. 2C). We have shown previously (*16*) that the T2 fingerprint estimated from ALL is identifiable in a satellite dataset with a warming troposphere (*28*), but not in a satellite dataset with little overall tropospheric temperature change (*15*).

The second (and more indirect) line of evidence relies on the relationship between changes in tropopause height and changes in tropospheric temperature. Our detection results indicate consistency between the patterns of tropopause height increase in ALL and reanalyses. The PCM individual forcing experiments help to identify the main drivers of this change. Over 1979-1999, roughly one-third of the increase in tropopause height in ALL is explained by GHG-induced warming of the troposphere (Fig. 2A). Anthropogenically-driven tropospheric warming is therefore an important factor in explaining modeled changes in tropopause height. The inference is that human-induced tropospheric warming may also be an important driver of observed increases in tropopause height. Both the direct and indirect lines of evidence support the contention that the troposphere has warmed markedly over the satellite era.

References and Notes

1. K. P. Hoinka, *Mon. Weath. Rev.* **126**, 3303 (1998).
2. D. J. Seidel, R. J. Ross, J. K. Angell, G. C. Reid, *J. Geophys. Res.* **106**, 7857 (2001).
3. W. J. Randel, F. Wu, D. J. Gaffen, *J. Geophys. Res.* **105**, 15509 (2000).
4. B. D. Santer *et al.*, *J. Geophys. Res.* **108(D1)**, 4002, doi:10.1029/2002JD002258 (2003).
5. B. D. Santer *et al.*, *Nature* **382**, 39 (1996).
6. S. F. B. Tett, J. F. B. Mitchell, D. E. Parker, M. R. Allen, *Science*, **274**, 1170 (1996).
7. V. Ramaswamy *et al.*, *Rev. Geophys.* **39**, 71 (2001).
8. J. E. Hansen *et al.*, *J. Geophys. Res.* **107**, ACL-2, doi:10.1029/2001JD001143 (2002).
9. W. M. Washington *et al.*, *Clim. Dyn.* **16**, 755 (2000).
10. C. Ammann, G. A. Meehl, W. M. Washington, C. Zender, *Geophys. Res. Lett.* (submitted).
11. G. A. Meehl, W. M. Washington, T. M. L. Wigley, J. M. Arblaster, A. Dai, *J. Clim.* **16**, 426 (2003).

12. E. Kalnay *et al.*, *Bull. Amer. Meteorol. Soc.* **77**, 437 (1996).

13. J. K. Gibson *et al.*, ERA description, *ECMWF Re-anal. Proj. Rep. Ser.* **1**, European Centre for Medium-Range Weather Forecasts, Reading, U.K., 66 pp. (1997).

14. B. D. Santer *et al.*, *J. Geophys. Res.* **104**, 6305 (1999).

15. J. R. Christy, R. W. Spencer, W. D. Braswell, *J. Atmos. Ocean. Tech.* **17**, 1153 (2000).

16. B. D. Santer *et al.*, *Science* (submitted).

17. The total linear change is defined as $b \cdot n$, where b is the slope parameter of the linear trend (in hPa/month or °C/month) fitted by the standard least-squares method over a specified period of n months.

18. For forcing component x, this is calculated as $\frac{|x|}{y} \cdot 100$, where $|x|$ is the absolute value of the linear change in p_{LRT} over a stipulated time interval (due to x) and $y = |\text{G}| + |\text{A}| + |\text{O}| + |\text{S}| + |\text{V}|$.

19. V. Ramaswamy, M. D. Schwarzkopf, W. Randel, *Nature* **382**, 616 (1996).

20. Since the O experiment does not separate tropospheric and stratospheric ozone changes, we cannot accurately quantify their individual contributions to overall changes in p_{LRT} and atmospheric temperature. We know, however, that stratospheric ozone depletion raises tropopause height through strong cooling of the

lower stratosphere (*4*). Increases in tropospheric ozone should warm the lower troposphere and could contribute to the recent rise in tropopause height. In PCM, tropospheric warming due to near-surface ozone changes is partly offset by the cooling effect of stratospheric ozone depletion on the upper troposphere. This explains why the net contribution of ozone forcing to T2 changes is only 2% over 1950-1999 (Fig. 2C). Total ozone forcing is small in the first half of the 20th century, which explains why the ozone contribution to p_{LRT}, T4, and T2 changes is negligible over 1900-1949 (Figs. 2A-C).

21. P. A. Stott *et al.*, *Science* **290**, 2133 (2000).

22. Linear changes in tropopause height over 1979-1999 are the sole exception: SUM results are 19% smaller than in ALL. This difference is largely attributable to noise differences. SUM is a linear combination of five climate responses plus five different (ensemble-averaged) noise estimates, and therefore must be noisier than ALL. ALL-SUM noise differences are likely to have greatest impact on trends calculated over relatively short periods.

23. K. Hasselmann, in *Meteorology of Tropical Oceans*, D. B. Shaw (Ed.), *Roy. Met. Soc. London*, pp. 251-259 (1979).

24. B. D. Santer *et al.*, *J. Geophys. Res.* **100**, 10693 (1995).

25. E. Roeckner, L. Bengtsson, J. Feichter, J. Lelieveld, H. Rodhe, *J. Climate* **12**, 3004 (1999).

26. Two reanalyses (NCEP, ERA) combined with two noise datasets for calculation of natural variability statistics (PCM, ECHAM) yields four non-optimized detection time estimates.

27. The detection time is the same for both ECHAM and PCM noise estimates.

28. C. A. Mears, M. C. Schabel, F. W. Wentz, *J. Climate* (submitted).

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. T.M.L. Wigley was supported by the NOAA Office of Global Programs ("Climate Change Data and Detection") grant no. NA87GP0105, and by 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).

Figure 1: Time series of global-mean monthly-mean anomalies in tropopause pressure (p_{LRT}; **A,B**), stratospheric temperature (T4; **C,D**), and mid- to upper tropospheric temperature (T2; **E,F**). Model results are from seven different PCM ensemble experiments (*9*, *10*, *11*). Five experiments use a single forcing only (G, A, O, S, or V). Two integrations involve combined forcing changes, either in natural external forcings (SV), or in all forcings (ALL). There are four realizations of each experiment. For individual forcings (**B,D,F**), only low-pass filtered ensemble means are shown. For ALL and SV results (**A,C,E**), both the low-pass filtered ensemble mean and the (unfiltered) range between the highest and lowest values of the realizations are given. All model anomalies are defined relative to climatological monthly means computed over 1890-1909. Reanalysis-based p_{LRT} estimates from NCEP (*12*) and ERA (*13*) were filtered in the same way as model data (**A**). NCEP p_{LRT} data are available from 1948-2001, but pre-1960 data were ignored due to deficiencies in the coverage and quality of assimilated radiosonde data (*14*). The ERA record spans 1979-1993. NCEP (ERA) was forced to have the same mean as ALL over 1960-1999 (1979-1993). The SUM results (**B,D,F**) are the sum of the filtered ensemble-mean responses from G, A, O, S, and V.

Figure 2: Total linear changes in global-mean, monthly-mean tropopause height (**A**), stratospheric temperature (**B**), and tropospheric temperature (**C**) in PCM experiments with individual forcings (G, A, O, S, V) and combined natural and anthropogenic forcings (ALL). Linear changes are computed over four different time

intervals (*17*) using the (unfiltered) ensemble-mean data from Fig. 1. For each time period, anomalies were defined relative to climatological monthly means computed over 1900-1999. SUM denotes the sum of the linear changes in G, A, O, S, V.

Figure 3: Least-squares linear trends and "fingerprints" of tropopause height change. Trends in global-mean, monthly-mean p_{LRT} data (in hPa/decade) were computed over 1979-1999 for NCEP (**A**) and the PCM ALL experiment (**C**), and over 1979-1993 for ERA (**B**). The p_{LRT} climate-change fingerprints were computed from the ensemble-mean ALL results. Results are shown for analyses with the spatial mean included (**D**) and with the mean removed (**E**).

Supporting Online Material

Reanalysis details

The operational numerical weather forecast models used in the two reanalyses had different horizontal resolution: T62 spectral truncation for NCEP versus T106 for ERA. Both models had comparable vertical resolution (17 levels). The assimilation schemes in NCEP and ERA were markedly different, particularly in their treatment of satellite data (*14*, *12*, *13*).

Model details

Both PCM and ECHAM were run with T42 spectral truncation in their atmospheric model components, corresponding to a horizontal resolution of roughly 2.8° latitude × 2.8° longitude. 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 $2/3° \times 2/3°$ 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° ($2.8° \times 2.8°$). Like PCM, ECHAM's ocean resolution decreases to 0.5° at the equator.

Calculation of tropopause pressure

The algorithm that we use to estimate p_{LRT} from reanalysis and model data involves linear 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. We identify the model level that satisfies the WMO criterion for the thermal tropopause, and then interpolate using the layers immediately above and below the threshold level (*4*).

Calculation of equivalent MSU temperatures

We use a static global-mean weighting function to compute equivalent MSU T4 and T2 temperatures from model data. For global and hemispheric means, this approach yields results similar to those obtained with a complex radiative transfer code (*14*).

Definition of fingerprint

Let $\vec{s}(t)$ represent annual-mean p_{LRT}, T4, or T2 data from a realization of the PCM ALL experiment, expressed as anomalies relative to the smoothed ALL 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 after regridding to a 10° latitude × 10° longitude grid.

We define $\vec{f}$ as the first EOF, which explains a substantial fraction of the overall variance of $\vec{s}(t)$: 65% for the 'mean included' analysis (Fig. 3D), and 40% for the 'mean removed' case (Fig. 3E).

Estimation of detection time

We begin with annual-mean 'observational' data, $\vec{o}(t)$ (NCEP or ERA), and control integrations, $\vec{c}(t)$ and $\vec{c}_1(t)$ (PCM and ECHAM). Reanalysis data are expressed as anomalies relative to either 1979-2001 (NCEP) or the 1979-1993 (ERA). Control anomalies are defined relative to the mean of the full 300-year integration. Two forms of detection time are computed: non-optimized ('raw') and optimized. To simplify and shorten the discussion, only non-optimized results are presented. For completeness, however, we show both raw and optimized detection times in Fig. S1. The primary conclusions of the paper do not depend on whether raw or optimized results are used.

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 trends of increasing length L to $Z(t)$, and then compare these 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, the start date of ERA and the more reliable portion of the NCEP reanalysis (*14*). 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 choices of m (5, 10, and 15). Full details of the detection method are given elsewhere (*24*).

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 (*21*).

Fig. S1 illustrates that optimization does not always improve detection times. One possible explanation for this behavior is that $\vec{f}$ may miss important components of the underlying response to the imposed forcing. A second explanation is that even if $\vec{f}$ captures most of the variance of the original data, important components of the fingerprint may be lost in projecting $\vec{f}$ onto the subspace of only the first m control run EOFs. Finally, 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.

Analysis with mean removed

In the 'mean removed' case, spatial means of the ensemble-mean ALL 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 $\vec{c}_1(t)$.

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.

Caption for Supporting Online Material

Figure S1: Detection times for PCM tropopause height fingerprints in NCEP and ERA reanalyses. The detection analysis uses both 'mean included' (**A**) and 'mean removed' (**B**) fingerprints (Figs. 3D,E), with a 5% significance level as the detection threshold (*24*). 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 for NCEP, 1993 for ERA). 'RAW' denotes detection times for non-optimized fingerprints. Optimized detection times are given for three different choices of the truncation dimension m. 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.