

Causes of differences in model and satellite tropospheric warming rates

Article

Accepted Version

Santer, B. D., Fyfe, J. C., Pallotta, G., Flato, G. M., Meehl, G. A., England, M. H., Hawkins, E. ORCID: https://orcid.org/0000-0001-9477-3677, Mann, M. E., Painter, J. F., Bonfils, C., Cvijanovic, I., Mears, C., Wentz, F. J., Po-Chedley, S., Fu, Q. and Zou, C.-Z. (2017) Causes of differences in model and satellite tropospheric warming rates. Nature Geoscience, 10 (7). pp. 478-485. ISSN 1752-0894 doi: 10.1038/ngeo2973 Available at https://centaur.reading.ac.uk/71264/

It is advisable to refer to the publisher's version if you intend to cite from the work. See <u>Guidance on citing</u>.

To link to this article DOI: http://dx.doi.org/10.1038/ngeo2973

Publisher: Nature Publishing Group

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the <u>End User Agreement</u>.

www.reading.ac.uk/centaur



CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Causes of Differences in Model and Satellite Tropospheric Warming Rates

³ Benjamin D. Santer¹, John C. Fyfe², Giuliana Pallotta¹, Gregory M. Flato², Gerald

A. Meehl³, Matthew H. England⁴, Ed Hawkins⁵, Michael E. Mann⁶, Jeffrey F.

⁵ Painter¹, Céline Bonfils¹, Ivana Cvijanovic¹, Carl Mears⁷, Frank J. Wentz⁷, Stephen

6

1

2

Po-Chedley¹, Qiang Fu⁸ & Cheng-Zhi Zou⁹

⁷ ¹Program for Climate Model Diagnosis and Intercomparison (PCMDI), Lawrence
⁸ Livermore National Laboratory, Livermore, CA 94550, USA.

⁹ ²Canadian Centre for Climate Modelling and Analysis, Environment and Climate
¹⁰ Change Canada, Victoria, British Columbia, V8W 2Y2, Canada.

¹¹ ³National Center for Atmospheric Research, Boulder, Colorado 80307, USA.

¹² ⁴ARC Centre of Excellence for Climate System Science, University of New South
¹³ Wales, New South Wales 2052, Australia.

¹⁴ ⁵National Centre for Atmospheric Science, Department of Meteorology, University of
¹⁵ Reading, Reading RG6 6BB, UK.

¹⁶ ⁶Department of Meteorology and Earth and Environmental Systems Institute, Penn-

¹⁷ sylvania State University, University Park, Pennsylvania, USA.

- ¹⁸ ⁷Remote Sensing Systems, Santa Rosa, CA 95401, USA.
- ¹⁹ ⁸Dept. of Atmospheric Sciences, University of Washington, Seattle, WA 98195, USA.
- ⁹Center for Satellite Applications and Research, NOAA/NESDIS, College Park, MD
 ²¹ 20740, USA.

Submitted to Nature Geoscience Revised: May 25, 2017

In the early 21st century, satellite tropospheric warming trends were gen-24 erally smaller than trends estimated from a large multi-model ensemble. 25 Because observations and coupled model simulations do not have the same 26 phasing of natural internal variability, such decadal differences in sim-27 ulated and observed warming rates invariably occur. Here we analyse 28 global-mean tropospheric temperatures from satellites and climate model 29 simulations to determine whether warming rate differences over the satel-30 lite era can be explained by internal climate variability alone. We find 31 that in the last two decades of the 20th century, differences between mod-32 eled and observed tropospheric temperature trends are broadly consistent 33 with internal variability. Over most of the early 21st century, however, 34 model tropospheric warming is substantially larger than observed; warm-35 ing rate differences are generally outside the range of trends arising from 36 internal variability. There is a low probability (between zero and $\approx 9\%$) 37 that multi-decadal internal variability fully explains the asymmetry be-38 tween the late 20th and early 21st century results. It is also unlikely that 39 this asymmetry is due to the combined effects of internal variability and 40 a model error in climate sensitivity. We conclude that model overestima-41 tion of tropospheric warming in the early 21st century is partly due to 42 systematic deficiencies in some of the post-2000 external forcings used in 43 the model simulations.

The Fifth Assessment Report of the Intergovernmental Panel on Climate Change 45 (IPCC) contained prominent discussion of differences between warming rates in ob-46 servations and model simulations [1, 2]. The focus of the discussion was on two issues: 47 the causes of a putative "slowdown" in observed surface and tropospheric warming 48 during the early 21st century, and the reasons for the inability of most climate model 49 simulations to capture this behavior. The IPCC defined the "slowdown" as a sub-50 stantially reduced surface warming trend over 1998 to 2012 relative to the long-term 51 warming over 1951 to 2012 [2]. 52

Since publication of the Fifth Assessment Report, at least three different interpre-53 tations of the "slowdown" have emerged. One interpretation is that this phenomenon 54 is largely an artifact of residual errors in surface temperature data sets [3, 4, 5]. A 55 second school of thought holds that the "slowdown" is primarily a routine decadal 56 fluctuation in temperature [6], and is not statistically distinguishable from previous 57 manifestations of internal variability [7, 8, 9]. A third interpretation is that the "slow-58 down" is attributable to the combined effects of different modes of internal variability 59 [10, 11, 12, 13, 14] and multiple external forcings [15, 16, 17]. 60

It is of interest to examine some implications of these schools of thought. If the reduction in early 21st century warming is mainly an artifact of errors in surface temperature data [3, 5], independent, satellite-based measurements of tropospheric temperature should show little evidence of a recent "slowdown" in warming – consistent with corrected surface results. Current satellite datasets, however, provide support
for a reduced rate of tropospheric warming in the early 21st century [15, 16, 18].

If the "slowdown" is predominantly a routine manifestation of internal variability (and if model-based estimates of the forced temperature signal and internal variability are realistic), then the differences between simulated and observed warming rates arise solely from different phasing of internal variability in "model world" and in the real world. Under this interpretation, model-versus-observed warming rate differences should be fully consistent with internal variability.

In the third school of thought, both internal variability and external forcing con-73 tribute to the "slowdown" [2, 19]. The externally forced contribution is due to the 74 combined cooling effects of a succession of moderate early 21st century eruptions 75 [15, 20, 21, 22, 23, 24], a long and anomalously low solar minimum during the last so-76 lar cycle [25], increased atmospheric burdens of anthropogenic sulfate aerosols [17, 26], 77 and a decrease in stratospheric water vapor [27]. There are known systematic errors 78 in these forcings in model simulations performed in support of the IPCC Fifth As-79 sessment Report [2, 17, 19, 20, 27]. These errors arise in part because the simulations 80 were performed before more reliable estimates of early 21st century forcing became 81 available [20, 27]. The net effect of the forcing errors is that the simulations underes-82 timate some of the cooling influences contributing to the observed "slowdown". 83

We find that for tropospheric temperature, model-versus-observed warming rate

differences during most of the early 21st century cannot be fully explained by natural internal variability of the climate system. We consider whether this result provides support for the third school of thought, or if it could be plausibly explained by the combined effects of a model error in climate sensitivity [28] and different phasing of modeled and observed internal variability [10, 11, 12, 13, 14].

Our focus is on satellite- and model-based estimates of tropospheric temperature. 90 There are two reasons for this choice. First, satellite tropospheric temperature mea-91 surements have time-invariant, near-global coverage [29, 30, 31]. In contrast, there 92 are large, non-random temporal changes in spatial coverage in the observed surface 93 temperature datasets used in most "slowdown" studies [3, 19, 32]. Second, satellite 94 tropospheric temperature datasets have been a key component of recent claims that 95 current climate models are too sensitive (by a factor of three or more) to human-96 caused changes in greenhouse gases [28, 33]. Errors of this magnitude would diminish 97 confidence in model projections of future climate change. It is therefore critically 98 important to evaluate the validity of such claims.

¹⁰⁰ Satellite and model temperature data

¹⁰¹ Our analysis primarily relies on satellite-based measurements of global-scale changes ¹⁰² in the temperature of the mid- to upper troposphere (TMT). TMT data with near-¹⁰³ global coverage are available from three groups: Remote Sensing Systems (RSS) [29], the Center for Satellite Applications and Research (STAR) [31], and the University of Alabama at Huntsville (UAH) [34]. Older and more recent dataset versions are provided by each of these groups (see Methods). A fourth group (the University of Washington; UW) [30] produces TMT data for a tropical domain. We briefly discuss both tropical TMT changes and global-scale changes in the temperature of the lower troposphere (TLT); the latter are provided by RSS and UAH only.

Model TMT data are from simulations of historical climate change (HIST) and 110 of 21st century climate change under Representative Concentration Pathway 8.5 111 (RCP8.5). These simulations yield information on the tropospheric temperature re-112 sponse to combined anthropogenic and natural external forcing. To compare models 113 and observations over the full satellite temperature record (January 1979 to Decem-114 ber 2016), HIST and RCP8.5 temperatures were spliced together ("HIST+8.5"). We 115 also analyze control runs with no changes in external forcings. Control runs are one 116 of a number of different sources of information on natural internal climate variability 117 [35, 36, 37, 38]. The HIST, RCP8.5, and control simulations were performed under 118 phase 5 of the Coupled Model Intercomparison Project (CMIP5) [39]. 119

Because TMT receives a contribution from the cooling of the stratosphere, a standard regression-based approach was employed to correct for this influence [40]. Correction yields a more representative measure of bulk changes in tropospheric temperature [41, 42, 43], and was performed for both satellite and model TMT data. ¹²⁴ Further information on the correction method and the satellite and model tempera-¹²⁵ ture data is provided in the Methods section and the Supplementary Information.

¹²⁶ Tropospheric temperature time series

The multi-model average (MMA) of TMT changes in the HIST+8.5 simulations is 127 smoother than any individual observational TMT time series (see Fig. 1A). This 128 difference in the amplitude of variability is expected [12, 15, 44]. In "free running" 129 simulations with coupled models of the climate system, the phasing of internally gen-130 erated climate variability is random. By averaging over 49 realizations of HIST+8.5 131 (performed with 37 different climate models), the amplitude of random variability is 132 reduced, more clearly revealing the underlying temperature response to external forc-133 ings. The real world, however, has only one sequence of internal climate variability. 134

Tropospheric warming is larger in the MMA than in the satellite data [45] (Figs. 135 1A, B). Another prominent feature of the observed results is the large interannual 136 temperature variability arising from the internally generated El Niño/Southern Os-137 cillation (ENSO). The positive (El Niño) phase of ENSO causes short-term warming. 138 The large 1982/83 El Niño partly obscured cooling caused by the 1982 eruption of 139 El Chichón. Because of the above-described noise reduction arising from averaging 140 over realizations and models, the cooling signatures of El Chichón and Pinatubo are 141 clearer in the MMA [15, 46]. Removal of temperature variability induced by ENSO 142

¹⁴³ improves the agreement between volcanic cooling signals in the MMA and in satel¹⁴⁴ lite tropospheric temperature data, but does not fully explain mismatches between
¹⁴⁵ simulated and observed tropospheric warming during the early 21st century [15].

¹⁴⁶ Significance of individual difference series trends

¹⁴⁷ Next, we assess whether there are statistically significant differences between tropo-¹⁴⁸ spheric temperature changes in models and individual satellite temperature datasets. ¹⁴⁹ We operate on the difference series $\Delta T_{f-o}(k,t) = \overline{T}_f(t) - T_o(k,t)$, where k is an index ¹⁵⁰ over the number of satellite datasets, t is an index over time (in months), $\overline{T}_f(t)$ is the ¹⁵¹ MMA, and $T_o(k,t)$ is an individual observational temperature time series. The sub-¹⁵² scripts f and o denote results from forced simulations and observations (see Methods ¹⁵³ and statistical terminology section in the Supplementary Information).

Our significance testing procedure rests on two assumptions. First, we assume 154 that the MMA provides a credible, "noise free" estimate of the true (but unknown) 155 externally forced tropospheric temperature signal in the real world. If this assump-156 tion is valid, the difference series $\Delta T_{f-o}(k,t)$ should reflect the departures of the 157 observed realization of internal variability from the externally forced signal. A sec-158 ond necessary assumption is that the CMIP5 control runs provide unbiased estimates 159 of the amplitude, period, and frequency of major modes of natural internal variability, 160 particularly on interannual to multi-decadal timescales. Whether this assumption is 161

¹⁶² justifiable is discussed in the final section of the paper.

Under these two assumptions, we formulate the null hypothesis that departures 163 between the expected and observed tropospheric temperature trends are consistent 164 with internal climate noise. Rejection of the null hypothesis can have multiple ex-165 planations: systematic deficiencies in the external forcings applied in the HIST+8.5 166 simulations (such as neglect of moderate volcanic eruptions in the early 21st century 167 [20, 21, 22, 23]), errors in the climate sensitivity to external forcings, errors in the sim-168 ulated spectrum of internal variability, and residual inhomogeneities in the satellite 169 temperature measurements. These explanations are not mutually exclusive. 170

Most previous studies of differences between simulated and observed warming rates in the early 21st century focused on changes over specific periods [3, 16, 47, 48]. The appropriateness of different analysis period choices has been the subject of debate [3, 16, 19]. To avoid such debate, we focus instead on *L*-year analysis timescales. We consider five timescales here: L = 10, 12, 14, 16, and 18 years. For each timescale, an *L*-year "window" is advanced by one month at a time through $\Delta T_{f-o}(k, t)$. A least-squares linear trend is calculated for each individual window.

These maximally overlapping trends are plotted in the left column of Fig. 2. As expected, shorter *L*-year trends are noisier. For example, 10-year windows ending close to the peak tropospheric warming caused by the 1997/98 El Niño have large negative trends in the difference series. The use of longer trend-fitting periods damps ¹⁸² such end-point effects. Another noteworthy feature of Fig. 2 is that most *L*-year ¹⁸³ windows which sample a substantial portion of the early 21st century have large ¹⁸⁴ positive trends in $\Delta T_{f-o}(k,t)$. During this period, the average simulated warming is ¹⁸⁵ larger than the tropospheric warming in each satellite dataset. We use CMIP5 control ¹⁸⁶ runs to estimate the probability that trends in $\Delta T_{f-o}(k,t)$ are either unusually large or ¹⁸⁷ unusually small relative to unforced temperature trends (see Methods). The resulting ¹⁸⁸ empirical *p*-values are plotted in the right-hand column of Fig. 2.

For most L-year trends ending after 2005, model-versus-observed differences in 189 tropospheric warming are significantly larger (at the 10% level or better) than can be 190 explained by natural internal variability alone. This result holds for all six satellite 191 TMT datasets examined here. In contrast, L-year difference series trends ending 192 before 2005 are generally not significantly larger than unforced TMT trends in the 193 CMIP5 control runs. Qualitatively similar results are obtained for TMT averaged 194 over the tropics, as well as for near-global changes in TLT (see Supplementary Figs. 195 S1 and S2, respectively). 196

In each panel in the right-hand column of Fig. 2, there are upper and lower rejection regions for our stipulated null hypothesis. The upper (lower) rejection regions are for significant negative (positive) trends in $\Delta T_{f-o}(k,t)$. Under the null hypothesis, significant negative and positive trends in $\Delta T_{f-o}(k,t)$ should be equally likely. We find, however, that significant positive trends dominate. There is only one small ²⁰² group of significant negative trends in $\Delta T_{f-o}(k,t)$ – the group with end points close ²⁰³ to the anomalous warmth of the 1997/98 El Niño.

Other features of Fig. 2 are also of interest. Consider, for example, the group of 204 positive 10-year trends ending between approximately 1990 and 1993 (Fig. 2B). As 205 noted above, El Chichón's cooling signal is larger and clearer in the MMA than in 206 satellite TMT data, where it was partly masked by the 1982/83 El Niño. This explains 207 why simulated TMT trends commencing close to the Chichón eruption tend to show 208 a larger post-eruption recovery (and larger warming) than in the observations (Figs. 209 1A and B). The influence of the 1982/83 El Niño on trends in $\Delta T_{f-o}(k,t)$ diminishes 210 as the trend fitting period is increased. 211

The large tropospheric warming caused by the 2015/16 El Niño event also has a 212 pronounced effect. As shorter (10- to 12-year) sliding windows sample this observed 213 warming spike, the size of trends in the $\Delta T_{f-o}(k,t)$ difference series decreases, and p-214 values increase (Figs. 2B, D). However, as the longer 16- and 18-year sliding windows 215 approach the end of the TMT records, even the anomalous observed warmth of late 216 2015 and early 2016 does not negate the larger simulated warming during most of 217 the "slowdown" period – *i.e.*, trends in $\Delta T_{f-o}(k,t)$ remain significantly larger than 218 unforced trends (Figs. 2H, J). 219

Figure 2 reveals large structural uncertainties in satellite TMT datasets. These uncertainties reflect different choices in dataset construction, primarily related to the treatment of orbital drift, the impact of orbital drift on sampling the diurnal cycle of atmospheric temperature [29, 30, 31, 34, 49], and the influence of instrument body temperature [50, 51]. For example, versions 5.6 and 6.0 of the UAH TMT dataset have pronounced differences in tropospheric warming in the first third of the satellite record. These differences (which are probably due to an update in how the UAH group deals with instrument bias correction) are large enough to lead to different decisions regarding the statistical significance of initial trends in $\Delta T_{f-o}(k, t)$.

Our use of older and newer versions of satellite TMT records highlights the evolutionary nature of these datasets. This evolutionary understanding is not always well understood outside of the scientific community [33], which is why we choose to illustrate it in Fig. 2. In the following analysis, however, we focus on newer dataset versions, which incorporate adjustments for recently identified inhomogeneities, and are likely to be improved relative to earlier dataset versions [29, 30].

²³⁵ Significance of asymmetry statistics

The analysis in Fig. 2 focuses on the significance of individual trends in $\Delta T_{f-o}(k, t)$. It does not consider whether overall asymmetries in *p*-values (such as the preponderance of significant positive trends in the difference series) could be due to internal variability alone. To address this question, we define three asymmetry statistics. The first is γ_1 , which measures asymmetry in the numbers of significant positive and significant negative trends in $\Delta T_{f-o}(k,t)$. The second and third are the γ_2 and γ_3 statistics, which provide information on asymmetries in the temporal distribution of individual *p*-values. To calculate γ_2 and γ_3 , we split the number of maximally overlapping difference series trends into a first and second set of approximately equal size (SET 1 and SET 2; see Fig. 2). This is done for each value of the trend length *L*. The difference in the total number of significant positive trends in SET 1 and SET 2 is γ_2 . The difference in "set-average" *p*-values is γ_3 (see Methods).

Figure 3 shows asymmetry statistics for the specific case of maximally overlapping 10-year trends in $\Delta T_{f-o}(k,t)$. The actual values of γ_1 , γ_2 and γ_3 reveal a preponderance of significant positive trends in $\Delta T_{f-o}(k,t)$, a larger number of significant positive trends in SET 2 than in SET 1, and a sharp decrease in average *p*-values between SET 1 and SET 2 (see Figs. 3A, C, and E, respectively). We seek to estimate the likelihood that these actual values could be due to multi-decadal internal variability alone. We refer to these probabilities subsequently as p_{γ_1} , p_{γ_2} and p_{γ_3} .

²⁵⁵ We begin by randomly selecting 5,000 surrogate "observed" TMT time series ²⁵⁶ from the CMIP5 control runs (see Methods and Supplementary Figs. S3 and S4). ²⁵⁷ For each surrogate time series, maximally overlapping *L*-year trends are compared ²⁵⁸ with control run distributions of unforced *L*-year trends; *p*-values are calculated for ²⁵⁹ each individual trend, and asymmetry statistics are computed from the *p*-values. This ²⁶⁰ procedure yields 5,000-member null distributions of γ_1 , γ_2 and γ_3 . We know *a priori* that the statistical properties of these distributions are solely influenced by natural internal variability. Actual values of the asymmetry statistics are compared with the null distributions to estimate p_{γ_1} , p_{γ_2} and p_{γ_3} (see Figs. 3B, D, and F).

Figure 4 summarizes these probability estimates. By averaging over satellite datasets and analysis timescales, we obtain the overall probabilities $\overline{\overline{p_{\gamma_1}}}$, $\overline{\overline{p_{\gamma_2}}}$ and $\overline{\overline{p_{\gamma_3}}}$ (the magenta lines in Fig. 4). For the statistic gauging the asymmetry in the numbers of positive and negative difference series trends, $\overline{\overline{p_{\gamma_1}}} \approx 0.005$. On average, therefore, there is only a 1 in 200 chance that the actual preponderance of significant positive trends in $\Delta T_{f-o}(k,t)$ could be due to internal variability alone (Fig. 4A).

Consider next the temporal asymmetries between the properties of difference series trends in SET 1 and SET 2 (Figs. 4B and C). The likelihood is very small ($\overline{p_{\gamma_2}} \approx$ 0.004) that random internal fluctuations in climate could fully explain why the number of significant positive trends in $\Delta T_{f-o}(k,t)$ is larger in SET 2 than in SET 1. For the third asymmetry statistic, there is less than a 1 in 10 chance ($\overline{p_{\gamma_3}} \approx 0.09$) that the actual decline in average *p*-values between SET 1 and SET 2 is due to internal variability alone.

The probabilities in Fig. 4 are calculated separately for each asymmetry statistic. We also considered the joint behavior of γ_1 , γ_2 and γ_3 . We estimated $p_{\gamma_{123}}$, the likelihood that internal variability alone can simultaneously produce values of γ_1 , γ_2 and γ_3 that are more extreme than their "satellite average" actual values (the brown vertical lines in Figs. 3B, D and F). The calculation of $p_{\gamma_{123}}$ was performed with the same Monte Carlo-generated sampling distributions employed for computing the individual probabilities p_{γ_1} , p_{γ_2} and p_{γ_3} .

For each of the five analysis timescales, $p_{\gamma_{123}}$ is zero. This indicates that in the 284 5,000 realizations of surrogate observations, there is not a single realization in which 285 multi-decadal internal variability can simultaneously explain the actual asymmetries 286 in the sign and temporal distribution of significant trends in $\Delta T_{f-o}(k,t)$. We cau-287 tion, however, that our estimate of $p_{\gamma_{123}}$ relies on non-independent information, and 288 is therefore likely to be biased: γ_1 , γ_2 , and γ_3 are all calculated from the same set 289 of p-values for maximally overlapping trends in $\Delta T_{f-o}(k,t)$. Nevertheless, our find-290 ings suggest that there is real value in considering the joint behavior of γ_1 , γ_2 and 291 γ_3 , and that each statistic provides some unique information about the asymmetric 292 distribution of difference series trends. 293

²⁹⁴ "Perfect model" analysis

It has been posited that the differences between modeled and observed tropospheric warming rates are solely attributable to a fundamental error in model sensitivity to anthropogenic greenhouse gas increases [28]. Several aspects of our results cast doubt on the "sensitivity error" explanation. First, it is difficult to understand why significant differences between modeled and observed warming rates should be preferentially concentrated in the early 21st century (see Fig. 2). A fundamental model sensitivity error should be manifest more uniformly in time. Second, a large sensitivity error should appear not only in trend behavior, but also in the response to major volcanic eruptions [46]. After removal of ENSO variability, however, there are no large systematic model errors in tropospheric cooling following the eruptions of El Chichón in 1982 and Pinatubo in 1991 [15].

We performed a "perfect model" analysis to further investigate this issue. We 306 consider whether asymmetries in the sign and temporal distribution of significant 307 trends in $\Delta T_{f-o}(k,t)$ could be solely due to the combined effects of a large model 308 sensitivity error and different realizations of modeled and observed internal variabil-309 ity. The "perfect model" study emulates our analysis of the "MMA minus satellite" 310 difference series. Now, however, the difference series $\Delta T_{f-f}(j,t)$ is formed between 311 the MMA and each individual HIST+8.5 realization. We calculate "perfect model" 312 values of the γ_1 , γ_2 and γ_3 statistics not only over 1979 to 2016, but also over three 313 earlier and two later 38-year analysis periods (see Methods). 314

For each asymmetry statistic, our "perfect model" analysis yields 288 individual samples. This allows us to explore how γ_1 , γ_2 and γ_3 behave over a large range of inter-model differences in climate sensitivity and phasing of low-frequency modes of variability (Supplementary Fig. S5). Because consistently derived estimates of Equilibrium Climate Sensitivity (ECS) are not available for all CMIP5 models, we ³²⁰ use a simple ECS proxy to study relationships between climate sensitivity and the ³²¹ "perfect model" values of γ_1 , γ_2 and γ_3 . This proxy, $\Delta T_{8.5}$, is the global-mean change ³²² in corrected TMT over 2006 to 2095; $\Delta T_{8.5}$ can be calculated from all 37 models for ³²³ which we have RCP8.5 simulations (see Supplementary Fig. S6).

Relationships between the "perfect model" results and $\Delta T_{8.5}$ are shown in Sup-324 plementary Fig. S7. Results are partitioned into two groups. The first group is for 325 the three earlier analysis periods (1862 to 1899, 1900 to 1937, and 1940 to 1977). The 326 second group contains results for three later analysis periods (1979 to 2016, 2020 to 327 2057 and 2058 to 2095). For both groups of results, there are only weak relationships 328 between $\Delta T_{8.5}$ and the statistics capturing temporal asymmetries in trend behavior 329 $(\gamma_2 \text{ and } \gamma_3)$. In contrast, the statistic reflecting asymmetries in trend sign (γ_1) is 330 highly correlated with $\Delta T_{8.5}$, but only during the three later analysis periods. 331

The latter result has several explanations. First, inter-model differences in ECS 332 become more pronounced as greenhouse gas forcing increases. These sensitivity dif-333 ferences are manifest as a time-increasing spread in tropospheric warming rates (Sup-334 plementary Fig. S5). As this spread grows in the 21st century, high-ECS (low ECS) 335 models yield a larger number of significant negative (positive) trends in the $\Delta T_{f-f}(j,t)$ 336 difference series, and γ_1 becomes more highly correlated with $\Delta T_{8.5}$. Second, as trends 337 in $\Delta T_{f-f}(j,t)$ become larger, the correlation between $\Delta T_{8.5}$ and γ_1 is less affected by 338 natural decadal variability (Supplementary Fig. S8). 339

Despite the fact that our "perfect model" analysis encompasses a large range 340 of inter-model climate sensitivity differences, the average actual values of the three 341 asymmetry statistics (the brown vertical lines in Figs. 3B, D, and F) remain unusual. 342 For γ_1 , there are only 12 out of 288 cases where the "perfect model" result exceeds 343 the actual value (Supplementary Fig. S9A). This yields a probability of $p_{\gamma_1} = 0.042$ 344 that the actual γ_1 value could be due to the combined effects of a model error in 345 climate sensitivity and different phasing of modeled and observed internal variability. 346 For the statistics gauging temporal asymmetry, this likelihood is even smaller: $p_{\gamma_2} =$ 347 0.010, and $p_{\gamma_3} = 0.038$ (Supplementary Figs. S9B, C). Finally, if the behavior of the 348 asymmetry statistics is examined jointly rather individually, there is only one out of 349 288 cases in which the "perfect model" values of γ_1 , γ_2 and γ_3 are simultaneously 350 more extreme than the average actual values, and $p_{\gamma_{123}} = 0.003$. 351

In contrast, statistically unusual values of all three asymmetry statistics could have 352 been plausibly generated by the temporal coincidence of multiple externally forced 353 and internally generated cooling influences in the early 21st century. Internally driven 354 contributions to the "warming slowdown" arise from the transition to a negative 355 phase of the Interdecadal Pacific Oscillation (IPO) in roughly 1999 [11, 13, 16, 52], 356 and from changes in the phasing of other internal variability modes [14, 53]. Our 357 statistical results are best explained by the combined effects of these known phase 358 changes and by previously identified systematic model forcing errors in the early 21st 359 century [2, 17, 20, 25, 27]. 360

³⁶¹ Reliability of model variability estimates

The credibility of our findings depends on the reliability of model-based estimates of natural variability. If CMIP5 models systematically underestimated the amplitude of tropospheric temperature variability on 10- to 18-year timescales, it would spuriously inflate the significance of individual difference series trends. In previous work, we found no evidence of such a systematic low bias. On average, CMIP5 models slightly overestimated the amplitude of decadal variability in TMT [54].

It is more difficult to assess the credibility of our estimated probabilities for the 368 overall asymmetry statistics shown in Figs. 3 and 4. Such an evaluation requires 369 information on model performance in capturing the "real-world" variability of tro-370 pospheric temperature on longer 30- to 40-year timescales. This information is not 371 directly available from relatively short satellite TMT records, and must instead be 372 inferred from other sources (see Supplementary Information). Such indirect sources 373 do not support a systematic model underestimate of tropospheric temperature vari-374 ability on 30- to 40-year timescales [55]. Note also that a low bias in model estimates 375 of longer-timescale variability is physically inconsistent [56] with the above-mentioned 376 claim of a high bias in model climate sensitivity [28]. 377

³⁷⁸ A related issue is the fidelity with which models capture the periods of multi-³⁷⁹ decadal oscillations. Underestimates of these periods could bias the sampling dis-³⁸⁰ tributions of the γ_2 and γ_3 statistics, in both the "perfect model" analysis and the analysis with surrogate observations. There is some evidence that such an error may exist for the IPO [57], although it is difficult to make a reliable assessment of this type of error given relatively short observational record lengths and the obfuscating effects of low-frequency changes in external forcings [26].

In conclusion, the temporary "slowdown" in warming in the early 21st century 385 has provided the scientific community with a valuable opportunity to advance under-386 standing of internal variability and external forcing, and to develop improved climate 387 observations, forcing estimates, and model simulations. Further work is necessary to 388 reliably quantify the relative magnitudes of the internally generated and externally 389 forced components of temperature change. It is also of interest to explore whether 390 surface temperature yields results consistent with those obtained here for tropospheric 391 temperature. 392

Our analysis is unlikely to reconcile divergent schools of thought regarding the causes of differences between modeled and observed warming rates in the early 21st century. However, we have shown that each hypothesized cause may have a unique statistical signature. These signatures should be exploited in improving understanding. While scientific discussion about the causes of short-term differences between modeled and observed warming rates is likely to continue [19], this discussion does not cast doubt on the reality of long-term anthropogenic warming.

$_{400}$ Methods

401 Satellite temperature data

We use satellite estimates of tropospheric temperature change produced by RSS [29,
58], STAR [31, 59, 60], UAH [34], and the University of Washington (UW) [30]. The
UW group supplies TMT data for the tropics only. All other groups have near-global
coverage of TMT measurements.

RSS, UAH, and STAR produce satellite measurements of the temperature of the lower stratosphere (TLS), which is used to correct TMT for the influence it receives from stratospheric cooling. Only RSS and UAH supply measurements of the temperature of the lower troposphere (TLT), which we briefly discuss in the main text.

UAH provides two different versions (5.6 and 6.0) of their TLS, TMT, and TLT datasets. RSS currently has only one version (3.3) of their TLS and TLT datasets, but two versions (3.3 and 4.0) of their TMT product. Two versions were available for the STAR TLS and TMT datasets (3.0 and 4.0). At present, there is only one version (1.0) of the UW tropical TMT dataset.

Satellite datasets are in the form of monthly means on $2.5^{\circ} \times 2.5^{\circ}$ latitude/longitude grids. Near-global averages of TMT and TLT were calculated over areas of common coverage in the RSS, UAH, and STAR datasets (82.5°N to 82.5°S for TMT, and ⁴¹⁸ 82.5°N to 70°S for TLT). All tropical averages are over 20°N to 20°S. At the time this
⁴¹⁹ analysis was performed, satellite temperature data were available for the 456-month
⁴²⁰ period from January 1979 to December 2016.

421 Method used for correcting TMT data

Trends in TMT estimated from microwave sounders receive a substantial contribution from the cooling of the lower stratosphere [40, 41, 61, 62]. In ref. [40], a regressionbased method was developed for removing the bulk of this stratospheric cooling component of TMT. This method has been validated with both observed and model atmospheric temperature data [41, 63, 64]. Here, we refer to the corrected version of TMT as TMT_{cr} . The main text discusses corrected TMT only, and does not use the subscript cr to identify corrected TMT.

For calculating tropical averages of TMT_{cr} , ref. [61] used:

$$TMT_{cr} = a_{24}TMT + (1 - a_{24})TLS$$
 (1)

where $a_{24} = 1.1$. For the near-global domain considered here, lower stratospheric cooling makes a larger contribution to TMT trends, so a_{24} is larger [40, 62]. In refs. [40] and [62], $a_{24} \approx 1.15$ was applied directly to near-global averages of TMT and TLS. Since we are performing corrections on local (grid-point) data, we used $a_{24} =$ 1.1 between 30°N and 30°S, and $a_{24} = 1.2$ poleward of 30°. This is approximately equivalent to use of the $a_{24} = 1.15$ for globally-averaged data.

436 Details of model output

We used model output from phase 5 of the Coupled Model Intercomparison Project 437 (CMIP5) [39]. The simulations analyzed here were contributed by 19 different re-438 search groups (see Supplementary Table S1). Our focus was on three different types 439 of numerical experiment: 1) simulations with estimated historical changes in human 440 and natural external forcings; 2) simulations with 21st century changes in green-441 house gases and anthropogenic aerosols prescribed according to the Representative 442 Concentration Pathway 8.5 (RCP8.5), with radiative forcing of approximately 8.5 443 W/m^2 in 2100, eventually stabilizing at roughly 12 W/m^2 ; and 3) pre-industrial con-444 trol runs with no changes in external influences on climate. 445

Most CMIP5 historical simulations end in December 2005. RCP8.5 simulations 446 were typically initiated from conditions of the climate system at the end of the histori-447 cal run. To avoid truncating comparisons between modeled and observed atmospheric 448 temperature trends in December 2005, we spliced together synthetic satellite temper-449 atures from the historical simulations and the RCP8.5 runs. Splicing allows us to 450 compare actual and synthetic temperature changes over the full 38-year length of the 451 satellite record. We use the acronym "HIST+8.5" to identify these spliced simula-452 tions. Some issues related to splicing are discussed in the Supplementary Information. 453

⁴⁵⁴ Supplementary Table S2 provides information on the external forcings in the ⁴⁵⁵ CMIP5 historical simulations. Details of the start dates, end dates, and lengths of the ⁴⁵⁶ historical integrations and RCP8.5 runs are given in Supplementary Table S3. Corre⁴⁵⁷ sponding information for the pre-industrial control runs is supplied in Supplementary
⁴⁵⁸ Table S4. In total, we analyzed 49 individual HIST+8.5 realizations performed with
⁴⁵⁹ 37 different CMIP5 models. Our climate noise estimates rely on pre-industrial control
⁴⁶⁰ runs from 36 CMIP5 models.

⁴⁶¹ Calculation of synthetic satellite temperatures

We use a local weighting function method developed at RSS to calculate synthetic satellite temperatures from model output [54]. At each model grid-point, simulated temperature profiles were convolved with local weighting functions. The weights depend on the grid-point surface pressure, the surface type (land or ocean), and the selected layer-average temperature (TLS, TMT, or TLT).

467 Statistical analysis

We analyze the statistical significance of trends in the temperature difference time series $\Delta T_{f-o}(k,t)$:

$$\Delta T_{f-o}(k,t) = \overline{T}_f(t) - T_o(k,t)$$
(2)

$$k = 1, \ldots, N_{obs}; \quad t = 1, \ldots, N_t$$

where $\overline{T}_{f}(t)$ is the multi-model average atmospheric temperature time series calculated from the forced HIST+8.5 simulations, and $T_{o}(k,t)$ is the temperature time series of the k^{th} observational dataset. Positive (negative) trends in $\Delta T_{f-o}(k,t)$ indicate model-average tropospheric warming that is larger (smaller) than observed. We seek to determine whether internal variability alone can explain large differences between expected and observed warming rates (both positive and negative).

All trends are calculated with monthly-mean TMT or TLT data. Rather than focusing on one specific period or timescale, we perform a comprehensive analysis of difference series trends on timescales ranging from 10 to 18 years, in increments of two years. These are typical record lengths used for study of the "warming slowdown" in the early 21st century [16, 19].

⁴⁸¹ Our analysis relies on maximally overlapping trends. "Maximally overlapping" ⁴⁸² indicates that an *L*-year sliding window is used for trend calculations. This window ⁴⁸³ advances in increments of one month until the end of the current window reaches the ⁴⁸⁴ final month of the $\Delta T_{f-o}(k, t)$ difference series.

In calculating the HIST+8.5 multi-model average (MMA), we specify that j is a combined index over models and HIST+8.5 realizations. The first averaging step is over HIST+8.5 realizations, and the second is over models. For processing the pre-industrial control runs, each model has only one control run, so j is an index over the number of models only. Anomalies in the satellite observations and HIST+8.5 runs were defined relative to climatological monthly means calculated over the 38-year period from January 1979 to December 2016. Control run anomalies were defined relative to climatological monthly means over the full length of each model's control integration.

⁴⁹⁴ Calculating p-values for individual difference series trends

⁴⁹⁵ We assess trend significance using weighted *p*-values, which account for inter-model ⁴⁹⁶ differences in control run length [45].

⁴⁹⁷ The weighted *p*-value, $\overline{p_c}(i, k, l)'$, is defined as:

$$\overline{p_c}(i,k,l)' = \sum_{j=1}^{N_{model}} p_c(i,j,k,l) / N_{model}$$
(3)

498

$$i = 1, \dots, N_{f-o}(l); \quad j = 1, \dots, N_{model}; \quad k = 1, \dots, N_{obs}; \quad l = 1, \dots, N_L$$

where *i* is over $N_{f-o}(l)$, the total number of maximally overlapping *L*-year trends in $\Delta T_{f-o}(k,t)$; *j* is over N_{model} , the number of model control runs; *k* is over N_{obs} , the total number of satellite datasets; and *l* is over N_L , the number of values of the trend length *L*. Here, $N_{f-o}(l) = 337$ for 10-year (120-month) trends; $N_{model} = 36$; $N_{obs} = 6$; and $N_L = 5$ (10, 12, 14, 16, and 18 years).

The individual $p_c(i, j, k, l)$ values for each model pre-industrial control run are calcu-

⁵⁰⁵ lated as follows:

$$p_c(i, j, k, l) = K_c(i, j, k, l) / N_c(j, l)$$
(4)

506

$$i = 1, \dots, N_{f-o}(l); \quad j = 1, \dots, N_{model}; \quad k = 1, \dots, N_{obs}; \quad l = 1, \dots, N_L$$

where $K_c(i, j, k, l)$ is the number of *L*-year trends in the j^{th} pre-industrial control run (for the l^{th} value of the trend length *L*) that are larger than the current *L*-year trend in $\Delta T_{f-o}(k, t)$. The sample size $N_c(j, l)$ is the number of maximally overlapping *L*-year trends in the j^{th} control run.

Use of maximally overlapping trends has the advantage of reducing the impact 511 of seasonal and interannual noise on atmospheric temperature trends, both in the 512 $\Delta T_{f-o}(k,t)$ difference series and in the control runs. It has the disadvantage of de-513 creasing the statistical independence of trend samples. Non-independence of samples 514 is an important issue in formal statistical significance testing, but is not a serious 515 concern here. This is because $\overline{p_c}(i,k,l)'$ is not used as a basis for formal statistical 516 tests. Instead, it simply provides useful information on whether trends in $\Delta T_{f-o}(k,t)$ 517 are unusually large or small relative to model estimates of unforced trends. 518

519 Calculating actual values of asymmetry statistics

The *p*-values in the right-hand column of Fig. 2 reveal pronounced asymmetries. Three asymmetries are of interest here. The first type of asymmetric behavior relates to the numbers of significant positive and significant negative trends. For each analysis timescale in Fig. 2, the overlapping trends computed from the $\Delta T_{f-o}(k,t)$ difference series display a preponderance of significant positive results. We use the γ_1 statistic to quantify this asymmetry:

$$\gamma_1(k,l) = K_{+ve}(k,l) - K_{-ve}(k,l)$$
(5)

where

$$K_{+ve}(k,l) = \sum_{i=1}^{N_{f-o}(l)} M(i,k,l)$$

$$M(i,k,l) = 1 \text{ if } \overline{p_c}(i,k,l)' \leq 0.1$$

$$M(i,k,l) = 0 \text{ if } \overline{p_c}(i,k,l)' > 0.1$$
(6)

and

$$K_{-ve}(k,l) = \sum_{i=1}^{N_{f-o}(l)} M(i,k,l)$$

$$M(i,k,l) = 1 \text{ if } \overline{p_c}(i,k,l)' \ge 0.9$$

$$M(i,k,l) = 0 \text{ if } \overline{p_c}(i,k,l)' < 0.9$$
(7)

The summation variables $K_{+ve}(k, l)$ and $K_{-ve}(k, l)$ in equation (6) are the total numbers of significant positive and significant negative trends in $\Delta T_{f-o}(k, t)$ (respectively). M(i, k, l) in equations (7) and (8) is an integer counter, and $\overline{p_c}(i, k, l)'$ is the weighted *p*-value for the current maximally overlapping trend, satellite dataset, and trend length. The significance of individual trends is assessed at the 10% level. ⁵³² bution of significant positive trends in $\Delta T_{f-o}(k, t)$. If we split the total number of ⁵³³ maximally overlapping difference series trends into two equally sized sets, there are ⁵³⁴ noticeably fewer significant positive trends in the first set (SET 1) than in the sec-⁵³⁵ ond set (SET 2). With the γ_2 statistic, we seek to determine whether this temporal ⁵³⁶ asymmetry is unusual:

$$\gamma_2(k,l) = K_{\text{SET1}}(k,l) - K_{\text{SET2}}(k,l)$$
 (8)

where

$$K_{\text{SET1}}(k,l) = \sum_{i=1}^{N(l)} M(i,k,l)$$

$$M(i,k,l) = 1 \quad \text{if} \quad \overline{p_c}(i,k,l)' \leq 0.1$$

$$M(i,k,l) = 0 \quad \text{if} \quad \overline{p_c}(i,k,l)' > 0.1$$

$$N(l) = [N_{f-o}(l) - 1] / 2$$

$$(9)$$

and

$$K_{\text{SET2}}(k,l) = \sum_{i=N(l)+1}^{N_{f-o}(l)} M(i,k,l)$$

$$M(i,k,l) = 1 \quad \text{if} \quad \overline{p_c}(i,k,l)' \leq 0.1$$

$$M(i,k,l) = 0 \quad \text{if} \quad \overline{p_c}(i,k,l)' > 0.1$$
(10)

The γ_3 statistic is analogous to γ_2 , but relies on differences between the average values of $\overline{p_c}(i, k, l)'$ in SET 1 and SET 2:

$$\gamma_3(k,l) = \overline{\overline{p_{c_1}}}(k,l)' - \overline{\overline{p_{c_2}}}(k,l)'$$
(11)

where the average SET 1 and SET 2 *p*-values, $\overline{\overline{p_{c_1}}}(k,l)'$ and $\overline{\overline{p_{c_2}}}(k,l)'$, are given by:

B. D. Santer et al. 31

$$\overline{\overline{p_{c_1}}}(k,l)' = \sum_{i=1}^{N(l)} \overline{p_c}(i,k,l)' / N(l)$$
(12)

$$\overline{\overline{p_{c_2}}}(k,l)' = \sum_{i=N(l)+1}^{N_{f-o}(l)} \overline{p_c}(i,k,l)' / N(l)$$
(13)

$$N(l) \approx N_{f-o}(l) / 2$$

Unlike γ_1 and γ_2 , the γ_3 statistic is not sensitive to the selected level for assessing the significance of individual trends in $\Delta T_{f-o}(k, t)$.

541 Overall significance of asymmetry statistics

To determine the significance of the actual values of these asymmetry statistics, we require null distributions of γ_1 , γ_2 and γ_3 , where we know *a priori* that changes in the statistics are solely due to random realizations of natural internal variability. We obtain null distributions of γ_1 , γ_2 and γ_3 using surrogate observational temperature time series from the CMIP5 control runs. The processing steps are as follows:

⁵⁴⁷ 1. Randomly select one of the 36 CMIP5 pre-industrial control runs.

From the selected control run, randomly choose the initial month of a 456-month
segment of temperature anomaly data. Ensure that the selected initial month
is valid (*i.e.*, that there are still at least 455 months between the selected initial
month and the end of the current control run). If this condition is not satisfied,

continue random selection of an initial month until the first valid month is obtained. The time series of surrogate observations is comprised of the first valid month and the next 455 months.

- 3. With the current surrogate observational time series, $T_{surr}(m, t)$, calculate the 555 weighted p-values, $\overline{p_c}(i, k, l)'$, as in equation (3). Since we are interested in how 556 $\gamma_1,\,\gamma_2$ and γ_3 behave in the presence of natural variability alone, the surrogate 557 observations are not used to form a difference series -i.e., they are not sub-558 tracted from $\overline{\overline{T}}_{f}(t)$ (the multi-model average), as was the case with the actual 559 satellite temperature data. Instead, individual maximally overlapping L-year 560 trends in the surrogate observations are compared directly with distributions 561 of control run L-year trends. In computing $\overline{p_c}(i,k,l)'$, the current surrogate 562 observational time series is excluded from the control runs used to calculate 563 unforced L-vear temperature trends, and the summation in equation (3) is over 564 $N_{model} - 1$ rather than over N_{model} . 565
- 4. From the values of $\overline{p_c}(i, k, l)'$ obtained from step 3, calculate the asymmetry statistics γ_1 , γ_2 and γ_3 , as in equations (5), (8), and (11).

5. Store these asymmetry statistics in $\gamma_1(l,m)^*$, $\gamma_2(l,m)^*$ and $\gamma_3(l,m)^*$, where the index *m* is over the total number of time series of randomly selected surrogate observations, and * denotes a statistic calculated with surrogate observational temperature data. 6. Return to step 1; repeat steps 1 through 5 until 5,000 surrogate observational time series have been selected, and 5,000-member distributions of $\gamma_1(l,m)^*$, $\gamma_2(l,m)^*$ and $\gamma_3(l,m)^*$ have been generated.

7. For each observational dataset, and for each of the five trend lengths considered (10, 12, ... 18 years), compare the actual values of $\gamma_1(k, l)$, $\gamma_2(k, l)$ and $\gamma_3(k, l)$ with their corresponding null distributions – *i.e.*, with $\gamma_1(l, m)^*$, $\gamma_2(l, m)^*$ and $\gamma_3(l, m)^*$, respectively. Examples of such comparisons are shown in Figs. 3B, D, and F of the main text for the case of 10-year trends. Determine the probability that the actual values of $\gamma_1(k, l)$, $\gamma_2(k, l)$ and $\gamma_3(k, l)$ could be due to internal variability alone. These overall probabilities are $p_{\gamma_1}(k, l)$, $p_{\gamma_2}(k, l)$ and $p_{\gamma_3}(k, l)$.

⁵⁸² "Perfect model" results

⁵⁸³ Our "perfect model" analysis considers whether an error in model Equilibrium Cli-⁵⁸⁴ mate Sensitivity (ECS), coupled with different phasing of internal climate variability ⁵⁸⁵ in the real world and in model HIST+8.5 simulations, could plausibly explain the ⁵⁸⁶ actual values of the three asymmetry statistics. To address this question, we form ⁵⁸⁷ difference series between tropospheric temperature changes in the HIST+8.5 MMA ⁵⁸⁸ and in individual model realizations of HIST+8.5:

$$\Delta T_{f-f}(j,t) = \overline{\overline{T}}_f(t) - T_f(j,t) \tag{14}$$

$$j = 1, \dots, N_{model}; \quad t = 1, \dots, N_t$$

where j is an combined index over HIST+8.5 realizations and models used to perform 589 the HIST+8.5 simulation. We calculate $\Delta T_{f-f}(j,t)$ for six different non-overlapping 590 456-month periods: the same January 1979 to December 2016 period used for comput-591 ing the "MMA minus observed" difference series in equation (2), three earlier periods 592 (1862 to 1899, 1900 to 1937, and 1940 to 1977), and two later periods (2020 to 2057)593 and 2058 to 2095). Because two of the three HadGEM2-CC HIST+8.5 realizations com-594 mence in December 1959, the sample size is not identical for the six analysis periods: 595 $N_{model} = 47$ (49) for the first three (last three) periods, yielding a total number of 596 288 $\Delta T_{f-f}(j,t)$ time series from which asymmetry statistics can be calculated. 597

We process these 288 "MMA minus individual model" difference time series in 598 the same way we treat the "MMA minus observed" difference series -i.e., we fit 599 maximally overlapping L-year trends to each $\Delta T_{f-f}(j,t)$ series, estimate weighted 600 *p*-values for each overlapping trend (by comparing with control run distributions of 601 unforced L-year trends), and then use these p-values to calculate asymmetry statistics. 602 The resulting "perfect model" asymmetry statistics are $\gamma_1(j, l)$, $\gamma_2(j, l)$ and $\gamma_3(j, l)$; 603 the statistics are indexed over HIST+8.5 realizations and models (the j index) and 604 over the number of values of the trend timescale (the l index). Distributions of these 605 statistics are shown in Supplementary Fig. S9 for the 10-year analysis timescale. 606

607 Proxy for ECS

ECS information is typically obtained from a $4 \times CO_2$ simulation [65]. Not all mod-608 eling groups participating in CMIP5 performed this simulation. Here, we have ECS 609 information for only 23 of the 37 CMIP5 models employed in our "perfect model" 610 analysis. To study underlying relationships between Equilibrium Climate Sensitivity 611 (ECS) and the "perfect model" results, we require a proxy for ECS. Our selected 612 proxy is $\Delta T_{8.5}$, the total linear change in near-global averages of corrected TMT in 613 the RCP8.5 simulation. For each realization and model, $\Delta T_{8.5}$ is calculated over the 614 1,080-month period from January 2006 to December 2095 – the longest common pe-615 riod in the RCP8.5 simulations analyzed here (see Supplementary Table S3). For the 616 23 models with $4 \times CO_2$ simulations, ECS is highly correlated with $\Delta T_{8.5}$ (Supplemen-617 tary Fig. S6). This provides justification for our use of $\Delta T_{8.5}$ as an ECS proxy in 618 Supplementary Fig. S7. For the models analyzed here, $\Delta T_{8.5}$ ranges from 3.28°C in 619 GISS-E2-R (p1) to 6.28° C in GFDL-CM3. 620

⁶²¹ Sample sizes in tests of asymmetry statistics

In assessing the statistical significance of our asymmetry statistics, we have greater confidence in our ability to rule out internal variability than in our ability to rule out the combined effects of internal variability and a model sensitivity error. This is because the sample size used to test the "internal variability only" explanation (5,000 time series of surrogate observations) is much larger than the sample size in
the "perfect model" analysis (288 time series of differences between the MMA and
individual model HIST+8.5 realizations). The analysis using surrogate observations
explores a much larger phase space in the timing and amplitude of the IPO and other
modes of internal variability.

Author contributions

B.D.S., J.C.F., G.P., G.M.F., and E.H. designed the analysis. B.D.S. performed all
statistical analyses. J.F.P. calculated synthetic satellite temperatures from model
simulation output and provided assistance with processing of observed temperature
data. C.M., F.J.W., S.P.-C., Q.F., and C.-Z.Z. provided satellite temperature data.
I.C., C.B., and J.F.P. assisted with the processing of the CMIP5 simulations analyzed
here. All authors contributed to the writing and review of the manuscript.

638 Competing financial interests

⁶³⁹ The authors declare no competing financial interests.

640 Corresponding author

⁶⁴¹ Correspondence should be sent to B. D. Santer (santer1@llnl.gov).

References 642

647

- [1] IPCC. Summary for Policymakers. In T. F. Stocker, D. Qin, G.-K. Plattner, 643
- M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midg-644
- ley, editors, Climate Change 2013: The Physical Science Basis. Contribution of 645
- Working Group I to the Fifth Assessment Report of the Intergovernmental Panel 646 on Climate Change. Cambridge University Press, 2013. 29 pages.
- [2] G. Flato, J. Marotzke, B. Abiodun, P. Braconnot, S. C. Chou, W. Collins, P. Cox, 648
- F. Driouech, S. Emori, V. Eyring, C. Forest, P. Gleckler, E. Guilyardi, C. Jakob, 649
- V. Kattsov, C. Reason, and M. Rummukainen. Evaluation of climate models. 650
- In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, 651
- A. Nauels, Y. Xia, V. Bex, and P. M. Midgley, editors, *Climate Change 2013:* 652
- The Physical Science Basis. Contribution of Working Group I to the Fifth Assess-653
- ment Report of the Intergovernmental Panel on Climate Change, pages 741–866. 654
- Cambridge University Press, 2013. 655
- [3] T. R. Karl, A. Arguez, B. Huang, J. H. Lawrimore, J. R. McMahon, M. J. Menne, 656 T. C. Peterson, R. S. Vose, and H.-M. Zhang. Possible artifacts of data biases 657 in the recent global surface warming hiatus. Science, 348:1469–1472, 2015.
- [4] K. Cowtan, Z. Hausfather, E. Hawkins, P. Jacobs, M. E. Mann, S. K. Miller, B. A. 659
- Steinman, M. B. Stolpe, and R. G. Way. Robust comparison of climate models 660

661		with observations using blended land air and ocean sea surface temperatures.
662		Geophys. Res. Lett., $42(15)$:6526–6534, 2015.
663	[5]	Z. Hausfather, K. Cowtan, D. C. Clarke, P. Jacobs, M. Richardson, and R. Ro-
664		hde. Assessing recent warming using instrumentally homogeneous sea surface
665		temperature records. <i>Sci. Adv.</i> , 3, e1601207, 2017.
666	[6]	S. Lewandowsky, J. S. Risbey, and N. Oreskes. The "pause" in global warming:
667		Turning a routine fluctuation into a problem for science. Bull. Amer. Meteor.
668		Soc., 97(5):723–733, 2016.
669	[7]	N. Cahill, S. Rahmstorf, and A. C. Parnell. Change points of global temperature.
670		Environ. Res. Lett., 10, 084002, 2015.
671	[8]	B. Rajaratnam, J. Romano, M. Tsiang, and N. S. Diffenbaugh. Debunking the
672		climate hiatus. Clim. Change, 133:129–140, 2015.
673	[9]	S. Rahmstorf, G. Foster, and N. Cahill. Global temperature evolution: recent
674		trends and some pitfalls. Env. Res. Lett., 12, 2017.
675	[10]	Y. Kosaka and SP. Xie. Recent global-warming hiatus tied to equatorial Pacific
676		surface cooling. <i>Nature</i> , 501:403–407, 2013.
677	[11]	G. A. Meehl, H. Teng, and J. M. Arblaster. Climate model simulations of the
678		observed early-2000s hiatus of global warming. Nat. Clim. Change, 4:898–902,

679 2014.

680	[12]	J. S. Risbey, S. Lewandowsky, C. Langlais, D. P. Monselesan, T. J. O'Kane,
681		and N. Oreskes. Well-estimated global surface warming in climate projections
682		selected for ENSO phase. Nat. Clim. Change, 4:835–840, 2014.
683	[13]	M. H. England, S. McGregor, P. Spence, G. A. Meehl, A. Timmermamm, W. Cai,
684		A. Sen Gupta, M. J. McPhaden, A. Purich, and A. Santoso. Recent intensifica-
685		tion of wind–driven circulation in the Pacific and the ongoing warming hiatus.
686		Nat. Clim. Change, 4:222–227, 2014.

- [14] B. A. Steinman, M. E. Mann, and S. K. Miller. Atlantic and Pacific multidecadal
 oscillations and Northern Hemisphere temperatures. *Science*, 347:988–991, 2015.
- [15] B. D. Santer, C. Bonfils, J. Painter, M. Zelinka, C. Mears, S. Solomon, G. A.
 Schmidt, J. C. Fyfe, J. N. S. Cole, L. Nazarenko, K. E. Taylor, and F. J. Wentz.
 Volcanic contribution to decadal changes in tropospheric temperature. *Nat. Geosci.*, 7:185–189, 2014.
- ⁶⁹³ [16] J. C. Fyfe, G. A. Meehl, M. H. England, M. E. Mann, B. D. Santer, G. M. Flato,
- E. Hawkins, N. P. Gillett, S.-P. Xie, Y. Kosaka, and N. C. Swart. Making sense
 of the early-2000s warming slowdown. *Nat. Clim. Change*, 6:224–228, 2016.
- [17] G. A. Schmidt, D. T. Shindell, and K. Tsigaridis. Reconciling warming trends.
 Nat. Geosci., 7:1–3, 2014.

698	[18]	H. Gleisner, P. Thejll, B. Christianson, and J. K. Nielsen. Recent global warming
699		hiatus dominated by low-latitude temperature trends in surface and troposphere
700		data. Geophys. Res. Lett., 42:510–517, 2014.
701	[19]	I. Medhaug, M. B. Stolpe, E. M Fischer, and R. Knutti. Reconciling controversies
702		about the 'global warming hiatus'. Nature, 545:41–47, 2017.
703	[20]	S. Solomon, J. S. Daniel, R. R. Neely, JP. Vernier, E. G. Dutton, and L. W.
704		Thomason. The persistently variable "background" stratospheric aerosol layer
705		and global climate change. Science, 333:866–870, 2011.
706	[21]	JP. Vernier, L. W. Thomason, JP. Pommereau, A. Bourassa, J. Pelon,
707		A. Garnier, A. Hauchecorne, L. Blanot, C. Trepte, D. Degenstein, and
708		F. Vargas. Major influence of tropical volcanic eruptions on the strato-
709		spheric aerosol layer during the last decade. Geophys. Res. Lett., 38, 2011.
710		http://dx.doi.org/10.1029/2011GL047563.
711	[22]	R. R. Neely, O. B. Toon, S. Solomon, JP. Vernier, C. Alvarez, J. M. English,

anthropogenic increases in SO_2 from Asia have minimal impact on stratospheric aerosol. *Geophys. Res. Lett.*, 40:1–6, 2013.

712

K. H. Rosenlof, M. J. Mills, C. G. Bardeen, J. S. Daniel, and J. P. Thayer. Recent

- [23] D. A. Ridley, S. Solomon, J. E. Barnes, V. D. Burlakov, T. Deshler, S. I. Dolgii,
 A. B. Herber, T. Nagai, R. R. Neely III, A. V. Nevzorov, C. Ritter, T. Sakai,
- B. D. Santer, M. Sato, A. Schmidt, O. Uchino, and J.-P. Vernier. Total volcanic

718		stratospheric aerosol optical depths and implications for global climate change.
719		Geophys. Res. Lett., 41:7763–7769, 2014.
720	[24]	B. D. Santer, S. Solomon, C. Bonfils, M. D. Zelinka, J. F. Painter, F. Beltran,
721		J. C. Fyfe, G. Johannesson, C. Mears, D. A. Ridley, JP. Vernier, and F. J.
722		Wentz. Observed multivariable signals of late 20th and early 21st century vol-
723		canic activity. <i>Geophys. Res. Lett.</i> , 42:500–509, 2015.
724	[25]	G. Kopp and J. L. Lean. A new, lower value of total solar irradi-
725		ance: Evidence and climate significance. Geophys. Res. Lett., 38, 2011.
726		http://dx.doi.org/10.1029/2010GL045777.
727	[26]	D. M. Smith, B. B. B. Booth, N. J. Dunstone, R. Eadie, L. Hermanson, G. S.
728		Jones, A. A. Scaife, K. L. Sheen, and V. Thompson. Role of volcanic and an-
729		thropogenic aerosols in the recent global surface warming slowdown. Nat. Clim.
730		Change, 6:936–940, 2016.
731	[27]	S. Solomon, K. H. Rosenlof, R. W. Portman, J. S. Daniel, S. M. Davis, T. J. San-
732		ford, and GK. Plattner. Contributions of stratospheric water vapor to decadal
733		changes in the rate of global warming. Science, 327:1219–1223, 2010.
734	[28]	J. R. Christy. Testimony in Hearing before the U.S. Senate Committee on
735		Commerce, Science, and Transportation, Subcommittee on Space, Science, and
736		Competitiveness, December 8, 2015, 2015. [Available online at http://www.
737		commerce.senate.gov/public/index.cfm/2015/12/data-or-dogma-pro

739		<pre>impact-on-earth-s-climate].</pre>
740	[29]	C. Mears and F. J. Wentz. Sensitivity of satellite-derived tropospheric temper-
741		ature trends to the diurnal cycle adjustment. J. Clim., 29:3629–3646, 2016.
742	[30]	S. Po-Chedley, T. J. Thorsen, and Q. Fu. Removing diurnal cycle contami-
743		nation in satellite-derived tropospheric temperatures: Understanding tropical
744		tropospheric trend discrepancies. J. Clim., 28:2274–2290, 2015.
745	[31]	CZ. Zou and W. Wang. Inter-satellite calibration of AMSU-A observa-
746		tions for weather and climate applications. J. Geophys. Res., 116, 2011.
747		http://dx.doi.org/10.1029/2011JD016205.
748	[32]	K. Cowtan and R. G. Way. Coverage bias in the HadCRUT4 temperature series
749		and its impact on recent temperature trends. Quart. J. Roy. Met. Soc., 140:1935–
750		1944, 2014.
	[]	

moting-open-inquiry-in-the-debate-over-the-magnitude-of-human-

[33] US Senate. Data or Dogma? Promoting open inquiry in the debate over the
magnitude of human impact on Earth's climate. Hearing before the U.S. Senate
Committee on Commerce, Science, and Transportation, Subcommittee on Space,
Science, and Competitiveness, One Hundred and Fourteenth Congress, first session, December 8, 2015. [https:clio.columbia.edu/catalog/12267036].

756	[34]	J. R. Christy, W. B. Norris, R. W. Spencer, and J. J. Hnilo. Tropospheric tem-
757		perature change since 1979 from tropical radiosonde and satellite measurements.
758		J. Geophys. Res., 112, 2007. http://dx.doi.org/10.1029/2005JD006881.
759	[35]	P. Bloomfield and D. Nychka. Climate spectra and detecting climate change.
760		Clim. Change, 21:275–287, 1992.
761	[36]	P. T. Brown, W. Li, E. C. Cordero, and S. A. Mauget. Comparing the model-
762		simulated global warming signal to observations using empirical estimates of
763		unforced noise. Nature Sci. Rep., 5, 9957, 2016.
764	[37]	M. R. Allen and S. F. B. Tett. Checking for model consistency in optimal
765		fingerprinting. Cli. Dyn., 15:419–434, 1999.
766	[38]	M. E. Mann, S. Rahmstorf, B. A. Steinman, M. Tingley, and S. K. Miller. The
767		likelihood of recent warmth. Nat. Sci. Rep., 6, 19831, 2016.

- [39] K. E. Taylor, R. J. Stouffer, and G. A. Meehl. An overview of CMIP5 and the
 experiment design. *Bull. Amer. Meteor. Soc.*, 93:485–498, 2012.
- [40] Q. Fu, C. M. Johanson, S. G. Warren, and D. J. Seidel. Contribution of stratospheric cooling to satellite-inferred tropospheric temperature trends. *Nature*,
 429:55–58, 2004.
- ⁷⁷³ [41] Q. Fu and C. M. Johanson. Stratospheric influences on MSU-derived tropospheric
- temperature trends: A direct error analysis. J. Clim., 17:4636–4640, 2004.

- [42] Q. Fu, S. Manabe, and C. M. Johanson. On the warming in the tropical upper troposphere: Models versus observations. *Geophys. Res. Lett.*, 38, 2011.
 http://dx.doi.org/10.1029/2011GL048101.
- [43] S. Po-Chedley and Q. Fu. Discrepancies in tropical upper tropospheric warming
 between atmospheric circulation models and satellites. *Environ. Res. Lett.*, 7,
 2012. http://dx.doi.org/10.1088/1748-9326/7/4/044018.
- [44] B. D. Santer, C. Mears, C. Doutriaux, P. Caldwell, P. J. Gleckler, T. M. L.
 Wigley, S. Solomon, N. P. Gillett, D. Ivanova, T. R. Karl, J. R. Lanzante, G. A. Meehl, P. A. Stott, K.E. Taylor, P. W. Thorne, M. F. Wehner,
 and F. J. Wentz. Separating signal and noise in atmospheric temperature changes: The importance of timescale. *J. Geophys. Res.*, 116, 2011.
 http://dx.doi.org/10.1029/2011JD016263.
- [45] B. D. Santer, S. Solomon, G. Pallotta, C. Mears, S. Po-Chedley, Q. Fu, F. Wentz,
 C.-Z. Zou, J. Painter, I. Cvijanovic, and C. Bonfils. Comparing tropospheric
 warming in climate models and satellite data. J. Clim., 30:373–392, 2017.
- [46] T. M. L. Wigley, C. M. Ammann, B. D. Santer, and S. C. B. Raper. The effect
 of climate sensitivity on the response to volcanic forcing. *J. Geophys. Res.*, 110,
 2005. http://dx.doi.org/10.1029/2004/JD005557.
- [47] J. C. Fyfe, N. P. Gillett, and F. W. Zwiers. Overestimated global warming over
 the past 20 years. *Nat. Clim. Change*, 3:767–769, 2013.

- [48] D. J. A. Johansson, B. C. O'Neill, C. Tebaldi, and O. Häggström. Equilibrium
 climate sensitivity in light of observations over the warming hiatus. *Nat. Clim. Change*, 5:449–453, 2015.
- [49] F. J. Wentz and M. Schabel. Effects of orbital decay on satellite-derived lower tropospheric temperature trends. *Nature*, 394:661–664, 1998.
- [50] C. A. Mears, M. C. Schabel, and F. J. Wentz. A reanalysis of the MSU channel
 2 tropospheric temperature record. J. Clim., 16:3650–3664, 2003.
- [51] S. Po-Chedley and Q. Fu. A bias in the mid-tropospheric channel warm target
 factor on the NOAA-9 Microwave Sounding Unit. J. Atmos. Oceanic Technol.,
 29:646–652, 2012.
- ⁸⁰⁵ [52] K. E. Trenberth. Has there been a hiatus? *Science*, 349:791–792, 2015.
- ⁸⁰⁶ [53] X. Chen and K. K. Tung. Varying planetary heat sink led to global-warming
 ⁸⁰⁷ slowdown and acceleration. *Science*, 345:897–903, 2014.
- 808 [54] B. D. Santer, J. F. Painter, C. A. Mears, C. Doutriaux, P. Caldwell, J. M. Ar-
- blaster, P. J. Cameron-Smith, N. P. Gillett, P. J. Gleckler, J. Lanzante, J. Perl-
- witz, S. Solomon, P. A. Stott, K. E. Taylor, L. Terray, P. W. Thorne, M. F.
- Wehner, F. J. Wentz, T. M. L. Wigley, L. J. Wilcox, and C.-Z. Zou. Identifying
- human influences on atmospheric temperature. Proc. Nat. Acad. Sci., 110:26–33,
- 813 2013.

814	[55] J. Imbers, A. Lopez, C. Huntingford, and M. R. Allen. Testing the robustness
815	of anthropogenic climate change detection statements using different empirical
816	models. J. Geophys. Res., 118:3192–3199, 2013.

- [56] T. M. L. Wigley and S. C. B. Raper. Natural variability of the climate system 817 and detection of the greenhouse effect. Nature, 344:324–327, 1990. 818
- [57] B. J. Henley, G. Meehl, S. B. Power, C. K. Folland, A. D. King, J. N. Brown, 819
- D. J. Karoly, F. Delage, A. J. E. Gallant, M. Freund, and R. Neukom. Spatial 820 and temporal agreement in climate model simulations of the interdecadal pacific 821 oscillation. Env. Res. Lett., 12, 2017. 822
- [58] C. Mears, F. J. Wentz, P. Thorne, and D. Bernie. Assessing uncer-823 tainty in estimates of atmospheric temperature changes from MSU and 824 AMSU using a Monte-Carlo technique. J. Geophys. Res., 116, 2011.825 http://dx.doi.org/10.1029/2010JD014954. 826
- [59] C.-Z. Zou, M. D. Goldberg, Z. Cheng, N. C. Grody, J. T. Sullivan, C. Cao, 827 Recalibration of microwave sounding unit for climate and D. Tarpley. 828 studies using simultaneous nadir overpasses. J. Geophys. Res., 111, 2006. 829 http://dx.doi.org/10.1029/2005JD006798. 830
- [60] C.-Z. Zou, M. Gao, and M. Goldberg. Error structure and atmospheric tem-831 perature trends in observations from the Microwave Sounding Unit. J. Clim., 832 22:1661-1681, 2009.

- [61] Q. Fu and C. M. Johanson. Satellite-derived vertical dependence of
 tropical tropospheric temperature trends. *Geophys. Res. Lett.*, 32, 2005.
 http://dx.doi.org/10.1029/2004GL022266.
- ⁸³⁷ [62] C. M. Johanson and Q. Fu. Robustness of tropospheric temperature trends from ⁸³⁸ MSU Channels 2 and 4. J. Clim., 19:4234–4242, 2006.
- [63] N. P. Gillett, B. D. Santer, and A. J. Weaver. Quantifying the influence of stratospheric cooling on satellite-derived tropospheric temperature trends. *Nature*, 432,
 2004. http://10.1038/nature03209.
- [64] J. T. Kiehl, J. Caron, and J. J. Hack. On using global climate model simulations
 to assess the accuracy of MSU retrieval methods for tropospheric warming trends.
 J. Clim., 18:2533-2539, 2005.
- [65] T. Andrews, J. M. Gregory, M. J. Webb, and K. E. Taylor. Forcing, feedbacks and
 climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. *Geo- phys. Res. Lett.*, 39, 2012. http://10.1029/2012GL051607.

848 Acknowledgments

We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling groups for producing and making available their model output. For CMIP, the U.S.

Department of Energy's Program for Climate Model Diagnosis and Intercomparison 852 (PCMDI) provides coordinating support and led development of software infrastruc-853 ture in partnership with the Global Organization for Earth System Science Portals. 854 We thank Mark Zelinka (PCMDI) for providing CMIP5 climate sensitivity results, 855 Susan Solomon (M.I.T.) for helpful discussions, and three anonymous reviewers for 856 constructive comments. The views, opinions, and findings contained in this report 857 are those of the authors and should not be construed as a position, policy, or decision 858 of the U.S. Government, the U.S. Department of Energy, or the National Oceanic and 859 Atmospheric Administration. 860

Figure 1: Time series (panel A) and difference series (panel B) of simulated and 861 observed tropospheric temperature. Results are monthly-mean TMT anomalies for 862 the 456-month period from January 1979 to December 2016, spatially averaged over 863 82.5°N-82.5°S and corrected for lower stratospheric cooling [40]. Multi-model average 864 (MMA) temperature data are from HIST+8.5 simulations performed with 37 different 865 CMIP5 models; satellite TMT data are for RSS version 4.0 [29]. Model TMT data 866 were computed using vertical weighting functions that approximate the satellite-based 867 vertical sampling of the atmosphere [54]. The time series of differences between the 868 MMA and the RSS data is shown in both raw form and smoothed with a 12-month 869 running mean (panel B). All anomalies are relative to climatological monthly means 870 calculated over January 1979 to December 2016. The vertical purple line is plotted at 871 the time of the maximum global-mean tropospheric warming during the 1997/98 El 872 Niño. The vertical green lines denote the eruption dates of El Chichón and Pinatubo. 873 Trends in the MMA and RSS over the full 456 months (the grey and pink lines in panel 874 A) are 0.291 and 0.199°C/decade, respectively. The corresponding trends over the 875 early 21st century (January 2000 to December 2016) are 0.286 and 0.191°C/decade. 876

Figure 2: Trends (left column) and trend significance (right column) for TMT difference series. The six difference series are for near-global averages of corrected TMT, and were computed by subtracting each of the six individual satellite TMT records from the HIST+8.5 multi-model average TMT time series (see Fig. 1). Maximally overlapping trends were fit to each 456-month difference series. Results are for trend

lengths of L = 10, 12, 14, 16, and 18 years; the overlap between successive L-year 882 trends is by all but one month. The *p*-values associated with each *L*-year difference 883 series trend were obtained by testing against multi-model distributions of unforced 884 L-year TMT trends from 36 different CMIP5 control runs. Results are plotted on 885 the last month of the trend-fitting period. Grev shading denotes the rejection region 886 (at a stipulated 10% significance level) for the null hypothesis that the difference be-887 tween modeled and observed TMT trends is due to internal variability alone. Each 888 panel in the right-hand column has a lower (upper) rejection region for large posi-889 tive (large negative) trends in the model-minus-observed difference series. The lower 890 (upper) rejection region spans the p-value range 0 to 0.1 (0.9 to 1.0). The y-axis 891 range was extended to -0.06 to facilitate visual display of *p*-values at or close to 892 zero. To calculate the actual values of the γ_2 and γ_3 statistics in Figs. 3D and F, 893 the maximally overlapping L-year trends were divided into two sets of approximately 894 equal size ("SET 1" and "SET 2"; see Methods). The dashed vertical lines in the 895 right-hand column panels denote the final month of the last L-year trend in SET 1. 896

Figure 3: Asymmetries in the statistical significance of differences between modeled and observed tropospheric temperature trends. Results are for maximally overlapping 10-year trends in near-global averages of corrected TMT. We calculate three asymmetry statistics. The first compares the numbers of significant positive and significant negative trends in the $\Delta T_{f-o}(k,t)$ difference time series (panel A). Subtracting the number of significant negative trends from the number of significant positive trends

yields the γ_1 statistic (panel B). The second statistic gauges asymmetry in the tem-903 poral distribution of positive trends in the difference series (panel C). To quantify 904 this asymmetry, we split the number of maximally overlapping 10-year trends into 905 two sets of approximately equal size. Trends sampling earlier (later) portions of the 906 difference series are in SET1 (SET 2). The difference in the number of positive trends 907 (SET1 minus SET2) is the γ_2 statistic (panel D). The third asymmetry statistic re-908 lies on the average p-values of the individual trends in SET1 and SET2 (panel E). 909 The difference between these set-average p-values is γ_3 (panel F). The vertical lines 910 in panels B, D, and F are the actual values of γ_1 , γ_2 and γ_3 . The grey histograms 911 in panels B, D, and F are null distributions of the asymmetry statistics, which were 912 generated using 5,000 realizations of surrogate observations (see Methods). 913

Figure 4: Overall statistical significance of the γ_1 , γ_2 and γ_3 asymmetry statistics as 914 a function of the analysis timescale and the satellite data used to compute the "MMA" 915 minus observed" difference time series. Results are estimates of p_{γ_1} , p_{γ_2} and p_{γ_3} , the 916 probabilities that the actual value of the asymmetry statistic could have been obtained 917 by natural internal variability alone (panels A, B, and C, respectively). The magenta 918 lines in panels A, B, and C are the averages (over the three recent observational 919 datasets and the five analysis timescales) of p_{γ_1} , p_{γ_2} and p_{γ_3} . Zero values of the 920 probabilities are indicated by colored arrows. The y-axis range in panels A and B is 921 substantially smaller than in panel C. For further details refer to the caption of Fig. 922 3 and the Methods section. 923



Figure 1: Santer et al.





Figure 2: Santer *et al.*



Asymmetries in Significance of Model-Minus-OBS TMT Differences (10-yr trends)

Figure 3: Santer *et al.*



Figure 4: Santer et al.