

Causes of irregularities in trends of global mean surface temperature since the late 19th century

Chris K. Folland^{1,2,3,4*}, Olivier Boucher⁵, Andrew Colman¹, David E. Parker¹

The time series of monthly global mean surface temperature (GST) since 1891 is successfully reconstructed from known natural and anthropogenic forcing factors, including internal climate variability, using a multiple regression technique. Comparisons are made with the performance of 40 CMIP5 models in predicting GST. The relative contributions of the various forcing factors to GST changes vary in time, but most of the warming since 1891 is found to be attributable to the net influence of increasing greenhouse gases and anthropogenic aerosols. Separate statistically independent analyses are also carried out for three periods of GST slowdown (1896–1910, 1941–1975, and 1998–2013 and subperiods); two periods of strong warming (1911–1940 and 1976–1997) are also analyzed. A reduction in total incident solar radiation forcing played a significant cooling role over 2001–2010. The only serious disagreements between the reconstructions and observations occur during the Second World War, especially in the period 1944–1945, when observed near-worldwide sea surface temperatures (SSTs) may be significantly warm-biased. In contrast, reconstructions of near-worldwide SSTs were rather warmer than those observed between about 1907 and 1910. However, the generally high reconstruction accuracy shows that known external and internal forcing factors explain all the main variations in GST between 1891 and 2015, allowing for our current understanding of their uncertainties. Accordingly, no important additional factors are needed to explain the two main warming and three main slowdown periods during this epoch.

INTRODUCTION

The recent slowdown in the warming of global mean surface temperature (GST) around 1998–2013 has caused much discussion about its causes (1, 2), including the sequestration of heat in the deeper oceans (3), as many climate models appear to overestimate recent warming of GST (4). Here, we analyze the causes of GST variations from 1891 to 2015, placing the causes of the recent slowdown in a longer-term context. We do not use the words “pause” or “hiatus” as, particularly in the recent slowdown, GST warming did not entirely cease on decadal time scales. Therefore, pause and hiatus cause semantic difficulties that are best avoided (5). However, the recent slowdown was one of at least three slowdown or cooling periods lasting about 15 years or more since 1891. The best-known apparent multidecadal cooling episode lasted from the early 1940s to 1975 (6). This has been called the big pause (6), so we will term this the big slowdown. An earlier cooling period occurred around 1896–1910, partly associated with volcanic eruptions in the West Indies (7), although it has not been fully explained. Not surprisingly, the uncertainty in GST data was appreciably greater at that time than during the recent slowdown (8). GST also shows two periods of strong warming around 1911–1940 and 1976–1997, although as throughout the instrumental GST record, temperatures showed considerable variability on subdecadal time scales. Current understanding of the veracity of the leading instrumental records of GST has recently been thoroughly reviewed (9).

Turning first to observational papers, a number of reconstructions of instrumental GST, based on forms of multiple regression against forcing factors, have been recently reviewed (10, 11), although some reconstructions only go back to about 1950 (12–14). One review (11)

analyzed annual mean GST data from these papers and concluded that detection of an anthropogenic signal was robust when signals from the El Niño–Southern Oscillation (ENSO), the Atlantic Multidecadal Oscillation (AMO) (15), total incident solar radiation (TSI), and volcanic (VOLC) forcing signals were included. The motivation for these studies was sometimes forecasting rather than mechanistic explanation or detection and attribution, leading, for example, to skillful real-time forecasts of GST 1 year ahead since 2000 (12) (www.metoffice.gov.uk/news/releases/2016/global-forecast-2017, as of 30 January 2018). Alternative methodologies more concerned with low-frequency GST variations use empirical mode decomposition (16) or low-pass filtering (17). Both methods enhance the magnitude of multidecadal AMO-like influences on GST compared to those found in regression-based papers quoted above that also include the contribution of the AMO to GST (11, 12), although the reason is not obvious. These alternative methodologies also reduce the inferred anthropogenic warming component of GST in the last few decades and explain the big slowdown of 1941–1975 as being largely due to 60- to 80-year natural variability associated with the AMO.

Dynamical climate models with external and internal forcings, averaged over many ensemble members over the past century and earlier, inevitably remove the signals of internal climate variability and so only respond to external forcings. These studies include an early, although quite skillful, attempt to reconstruct GST variations and trends since 1860 (18). Since then, many climate models have included this period in their analyses, most recently the 40 CMIP5 models or so (10) that collectively provide estimates of reconstructed uncertainties in GST and its mean. Since the late 19th century, multidecadal trends in average GST have been picked up quite well by CMIP5, mainly due to the net warming effect of increasing greenhouse gases and anthropogenic aerosols. The short-term cooling effects of major volcanic eruptions are also reconstructed quite well. Nevertheless, a common problem in CMIP5 simulations is an inability to reproduce the level of observed warmth in GST around the Second World War. Some of this deficiency is likely to be due to a peak in the AMO around 1940, inevitably not

¹Met Office Hadley Centre, Exeter, Devon EX1 3PB, UK. ²School of Environmental Sciences, University of East Anglia, Norwich, Norfolk NR4 7TJ, UK. ³Department of Earth Sciences, University of Gothenburg, Gothenburg 40530, Sweden. ⁴International Centre for Applied Climate Sciences, University of Southern Queensland, Toowoomba, Queensland, Australia. ⁵Institut Pierre-Simon Laplace, Sorbonne Université/CNRS, Paris, France.

*Corresponding author. Email: chris.folland@metoffice.gov.uk

picked up by CMIP5 or the earlier CMIP3 averages (19). Early 20th-century warming is explained as partly due to an increase in TSI, a lack of volcanic eruptions, and a small anthropogenic component due to slowly increasing greenhouse gases, much as in the earliest skillful climate model paper (18). Notably, recent estimates of past TSI variations (20, 21) indicate that the increase in TSI over 1891–1950 is likely to be less than originally estimated (22), especially during minima of the 11-year cycle. Relatively recent climate model results (23–25) also show that the AMO could have contributed appreciably to warming of the Northern Hemisphere and, considerably more weakly, the entire globe, during the years 1975–2005, a period of strong overall warming. However, compared to the first model, the second model study (24) shows about twice the sensitivity of an increase in Northern Hemisphere temperature to these AMO changes (23). The implied sensitivity of GST to the AMO in the second study is similar to that of the observational analyses that used alternative statistical methodologies (16, 17). If true, then these results would also imply a markedly smaller contribution to warming of the net effect of increasing greenhouse gases and anthropogenic aerosols (usually called GA in the remainder of this paper) than shown by CMIP5 models over the same period. The third of these studies (25) identifies internal variability as the time-varying difference between observations and a CMIP5 66-member ensemble mean with anthropogenic and natural forcings during 1920–2013, after scaling down temperature trends in the latter by ~14% to compensate for the models' overestimation of observed warming. The implied internal variability in GST is associated mainly with a pattern resembling the Interdecadal Pacific Oscillation (IPO) and, to a lesser extent, with an AMO-like pattern.

With this background, we reconstruct the monthly GST records using, separately, the main forcing factors at monthly resolution identified by previous authors and the average of a large number of cross-validated regression calculations (see Materials and Methods). Monthly data better represent short-term GST variability than do annual data, especially that due to ENSO. Furthermore, we transform the raw forcing data to represent the delayed response of GST to each forcing. Because these responses are often uncertain, we use a range of response functions, and to incorporate uncertainty in GST into the reconstructions, we use three separate GST data sets. As a result, we create 54 sets of cross-validated multiple regressions. Forcing factors first include GA, as tabulated by the Intergovernmental Panel on Climate Change (20). The greenhouse gas and aerosol forcing components of GA must be combined in a regression model because of their strong collinearity: Their correlation is –0.98 over 1891–1996 and –0.97 over 1896–2011 (20). We then include the AMO, ENSO, IPO (26), and volcanic and TSI forcing (absolute values of volcanic, TSI, and GA forcing are shown in fig. S1 and section S1, normalized values of the two AMO indices are shown in fig. S2A and section S2, and normalized values of the ENSO and IPO indices are separately shown in fig. S2B). Because we use monthly data, we can also include the influence on GST of the Arctic Oscillation (AO) between December and March, as detected in the recent GST slowdown (27). Earlier work (28) used a Northern Hemisphere Cold Ocean Warm Land (COWL) index that is negatively correlated with the AO, but the COWL depends on temperature observations. For completeness, an adjustment is added to GA forcing to account for a deep-ocean impact on sea surface temperature (SST) operating on time scales of about 200 years (see Materials and Methods). This is currently very small but could well become considerably larger in the future.

We divide the instrumental record into five periods, three exhibiting GST slowdowns or actual cooling (1896–1910, 1941–1975, and 1998–2013 and its subperiods 2001–2010 and 2001–2013) and two

strong warming periods (1911–1940 and 1976–1997). We selected slowdown and warming periods with approximately maximum contrast in trends according to visual inspection of the GST series. Piecewise continuous trends (section S3) show that the trends changed quickly around the dates we selected for the beginnings and ends of subperiods, although we did not carry out a formal change-point analysis in view of the absence of true discontinuities and the slight differences between observational data sets. Because the main focus of the paper is on slowdown periods, these are analyzed using regression model training data independent of the slowdown period being studied. However, the results are rather insensitive to whether these independent periods are used or not, as can be deduced from the overview discussion below using Fig. 1 where no independent periods are used compared to later sections where they are. Our reconstructions of the two warming periods use all cross-validated monthly data from 1891 to 2015. The GST records are taken from three key data sets in current use for climate monitoring [HadCRUT4.6, updated from HadCRUT4 (8), Goddard Institute for Space Studies Surface Temperature Analysis (GISTEMP) (29), and the National Climatic Data Center (NCDC) data set (5)]. We call the average of these three series the World Meteorological Organization (WMO) data set as it is used by WMO to assess global and large-scale temperature changes (for example, for 2016, <https://public.wmo.int/en/resources/library/wmo-statement-state-of-global-climate-2016>, as of 30 January 2018). Further details are given in Materials and Methods.

The purpose of the regression analyses is to determine their skill at reproducing GST in each period and to explain the causes of its variations and trends, in principle down to the monthly time scale. We also compare the skill of the reconstructions over the five periods with that of the average of simulations from 40 CMIP5 models (table S1 and section S4). The reconstructions have an important advantage over CMIP5 in that they incorporate internal variability via the observed climate modes such as the AMO, IPO, and ENSO, which the models cannot do. We also calculate the linear component of changes in GST over each period shown by observations, reconstructions, and CMIP5. This is achieved using a method called restricted maximum likelihood regression or REML. This is well adapted to estimating trends over short periods and allows for the sometimes very large serial correlation in trend residuals and data uncertainties (Materials and Methods and section S3). Finally, we show reconstructions of worldwide maps of GST to further investigate some key results and problems. For TSI, we often use the word "solar" in the diagrams. We mainly highlight changes, trends, or correlations significant at the 1% level or better, rather than 5%, as true statistical significance could be easily overestimated given the many statistics in this paper.

RESULTS

Overview

Figure 1A shows monthly WMO GST and its constituent HadCRUT4.6, NCDC, and GISTEMP GST anomalies relative to a 1961–1990 average, together with 95% confidence ranges around the WMO average for each month. The year 1891 is chosen as the first year analyzed as uncertainties in GST increase quite rapidly before then. On decadal time scales, the only periods of obvious observed GST cooling after 1891 are around 1896–1910 and 1941–1950. Figure 1B shows the average influences of the various forcings on GST using the multiple regression reconstructions where cross-validation is carried out over the whole period 1891–2015. This particular reconstruction uses the average of 81,000 fairly similar regression equations throughout the

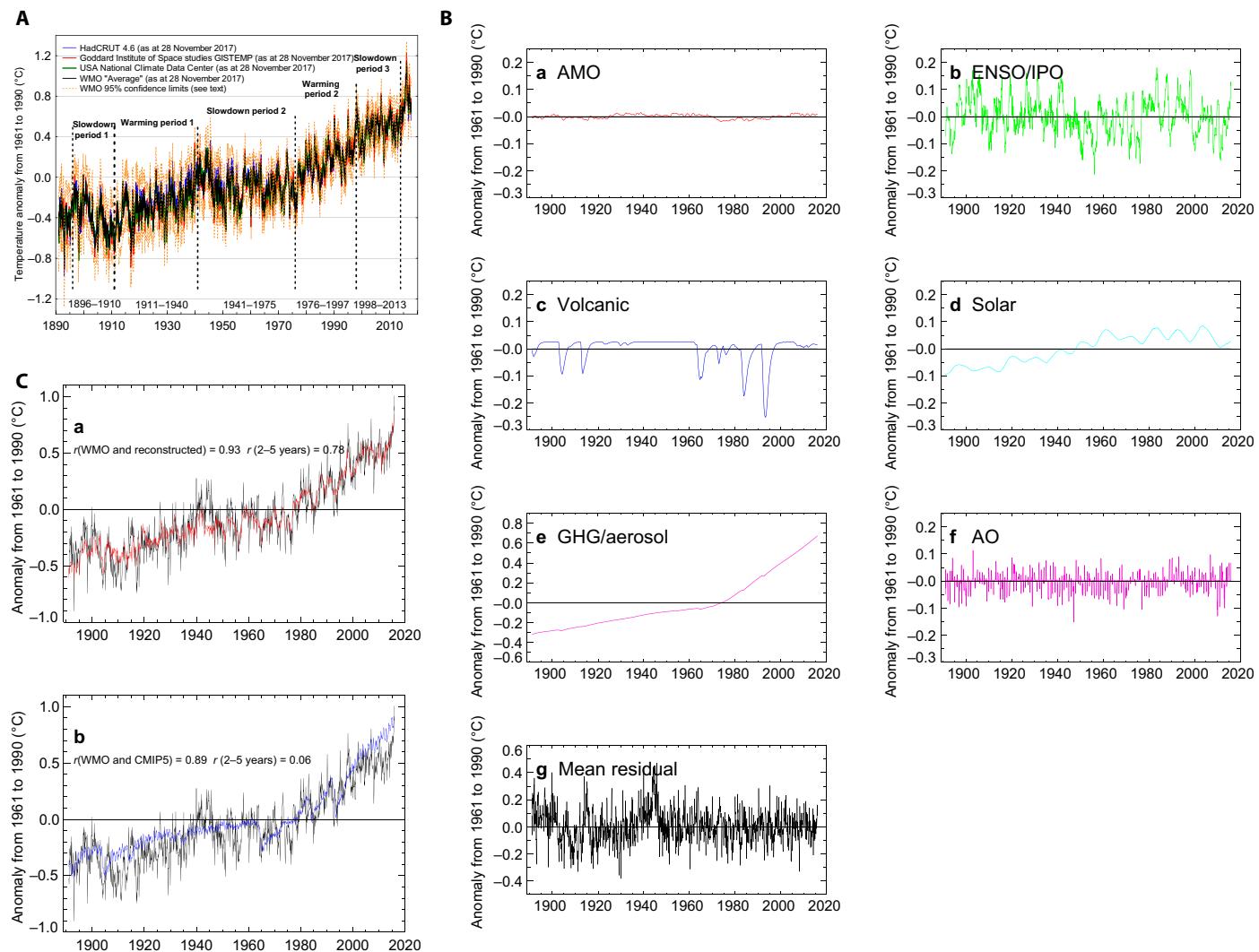


Fig. 1. Observed, reconstructed, and CMIP5 simulated GST for the whole period 1891–2015. (A) Monthly GST anomalies (January 1891 to September 2017) from the three analyses that make up the WMO data set and the WMO average. Shown are the two warming and three slowdown periods. Estimated 95% confidence limits on the WMO data set are shown in brown. (B) Reconstructions of monthly GST 1891–2015 from each of the various forcing factors. This uses the average of 81,000 cross-validated regression equations calculated from data covering this whole period. (C) (a) Average cross-validated reconstruction (red) of observed monthly WMO GST (black) for the whole period 1891–2015. (b) As in (a) but using the CMIP5 40 model average (blue) of monthly GST.

record (see Materials and Methods). The total of 81,000 equations first results from the use of separate regression equations for each of the 1500 months over 1891–2015. In addition, there are 18 predictor combinations for each of the three GST series. The 18 predictor combinations are derived from two or three variants of the ENSO or IPO, volcanic, TSI, and GA forcings, as well as fixed AMO and AO forcings. These combinations are shown in table S2 of section S5. Equation 1 is an example, calculated from the average of 1272 cross-validated equations (training period 1891–1996) for 1998–2013, based on a typical set of the 54 GST and predictor combinations used for this period

$$\text{GST} = -0.0595 + 0.0172 \cdot \text{AMO} + 0.0617 \cdot \text{ENSO} + 0.0441 \cdot \text{VOLC} + 0.0457 \cdot \text{TSI} + 0.2802 \cdot \text{GA} + 0.0219 \cdot \text{AO} \quad (1)$$

Monthly forcing values are normalized over 1891–2015 so that regression coefficient magnitudes can be compared directly. The other

53 equations vary sufficiently to be used to estimate uncertainties in the reconstructions. ENSO and unfiltered IPO are highly correlated (monthly $r = 0.76$) and thus are each used in 9 (of 18) different predictor combinations for each version of GST (Materials and Methods and section S5). Furthermore, VOLC has a positive value at times when negative volcanic forcing is less than the average for 1961–1990, while the AO is set to zero from April to November (Materials and Methods) as its sign and magnitude do not relate clearly to higher-latitude Northern Hemisphere surface temperature anomalies at these times of the year. The dominant factor in all equations is GA. Equation 1 is quite similar to that published for standardized annual forcing and GST data over 1947–2006 (12), although the AO was not included.

In Fig. 1B (b), ENSO influences are combined with those of the IPO. Much of the influence of the December–March AO is intermonthly, but interannual fluctuations also occur, including a decadal decline in the Northern Hemisphere winter AO around 2000–2010 previously highlighted (27). A larger than usual influence of the TSI 11-year cycle

that reduces GST from about 2001 to 2010 by about 0.08°C (Fig. 1B, d) is also notable. Changing GA increases GST at all times, most slowly in the two decades after the Second World War, but increases GST at its strongest and nearly constant rate from the late 1970s. In the early 21st-century slowdown of 1998–2013, the SD of the monthly regression residuals (observed GST minus reconstructed GST) using all the chosen forcing factors is very similar to the assessed uncertainties in monthly observed GST itself (section S6 and fig. S5). However, the structure of the mean residuals (Fig. 1B, g) is complex before 1950, reflecting poorly reconstructed cold conditions around 1907–1910, a short sharp peak of unexplained observed warmth in the sparse observed data of the First World War and a peak in unexplained observed warmth during the Second World War (see the “Using global maps to investigate reconstruction biases in 1907–1910 and 1944–1945” section). After 1950, mean residuals appear quasi-random, although they have an inter-monthly correlation near 0.5 falling to near zero at 6 months lag.

Figure 1C (a) provides an overview of the GST reconstructions using all data. It shows high reconstruction skill on interannual time scales (2- to 5-year high-pass-filtered period; correlation, 0.78 with observed GST; significance, $<0.01\%$) and for the whole spectrum (correlation, 0.93; significance, 0.5%). Materials and Methods describes correlation significance calculations. For the average GST from the 40 CMIP5 models (Fig. 1C, b), its correlation with observed GST over 1891–2015, dominated by trend, is only modestly less at 0.89 (significance, 0.5%), but there is relatively little high-pass 2- to 5-year skill (correlation, 0.06; significance, 2%). High-pass skill results almost entirely from skillful simulations of episodic volcanic forcings of GST around 1902–1906, 1965–1969, 1983–1985, and 1991–1995. The cool period before 1915 is missed by CMIP5 but is somewhat better in the reconstructions (Fig. 1C, a), although the cold conditions in 1907–1910 are underestimated by the reconstructions. The observed rate of warming from 1911 to 1940 is markedly better reconstructed than simulated by CMIP5, with CMIP5 simulating this warming trend too weakly. The CMIP5 models reproduce the observed slowing of GST warming from about 1940 to 1975, but the observed peak of warmth during the Second World War is missed, just as it is by the reconstructions (see the “Using global maps to investigate reconstruction biases in 1907–1910 and 1944–1945” section). The warming from around 1976 to 1997 is well reconstructed and well simulated by CMIP5, but the slowdown in GST around 1998–2013, particularly 2001–2013, is missed entirely by CMIP5 but is very well reconstructed. The rate of simulated warming by CMIP5 of $0.22^{\circ} \pm 0.06^{\circ}\text{C}$ per decade over the period 1976–2015 is somewhat greater than that observed at $0.17^{\circ} \pm 0.02^{\circ}\text{C}$. By contrast, the reconstruction gives the same rate of warming as observed, $0.17^{\circ} \pm 0.01^{\circ}\text{C}$ per decade. This CMIP5 overestimate of recent observed warming in GST, although not quite significant at the 10% level using CMIP5 uncertainties, is shown below to derive from a relatively large overestimate of warming in the recent slowdown since 1998. This overestimate is at least partly due to missing natural forced and internal variability. For obvious reasons, CMIP5 models were unable to simulate the observed changes in the IPO/ENSO, AMO, and the AO, and they were also not provided with observed solar and volcanic forcings for much of this period.

Period 1896–1910

Reconstructions are made for 1891–1913 to give a longer perspective. This period shows a marked overall cooling in GST (Fig. 2A). The reconstruction is surprisingly good for 1891–1906, but overestimates observed GST thereafter, especially during 1907–1910. CMIP5 aver-

age GST is also skillful until about 1902 but much less skillful thereafter (section S6 discusses apparent annual cycles in CMIP5 GST anomalies). As for other slowdown and warming periods, Fig. 2B shows the level of agreement between the 54 main regression reconstructions. Figure 2B (a to j) shows how the contributions to reconstructed GST of the forcing factors build up. For example, the joint influence of ENSO and the IPO is very important for reconstructing interannual and subannual GST over this period (Fig. 2B, g). Figure 2C shows the linear change in GST over 1896–1910 and changes derived from the components and all the forcing factors. These reconstructions are generally less successful than for other periods. However, although the observed linear cooling of GST over 1896–1910 is large at $-0.34^{\circ} \pm 0.28^{\circ}\text{C}$, its uncertainty is also large. Accordingly, the observed cooling just fails significance at the 1% level, partly due to relatively large observed GST uncertainties included in the REML trend method. Reconstructed cooling is underestimated at $-0.08^{\circ} \pm 0.07^{\circ}\text{C}$, which is also not significant at the 1% level, but the differential cooling error of $-0.27^{\circ} \pm 0.09^{\circ}\text{C}$ is significant at the 1% level. CMIP5 gives only a nominal cooling of $-0.05^{\circ} \pm 0.19^{\circ}\text{C}$. Although the reconstruction underestimates observed cooling, Fig. 2C shows that ENSO and the IPO contribute significantly to cooling ($-0.09^{\circ} \pm 0.05^{\circ}\text{C}$ over 1896–1910), with perhaps a very small, if significant, cooling contribution from the AMO. The IPO fluctuated quite strongly but moved to a negative phase toward the end of the period (30). Volcanic eruptions in 1902 also cool GST shortly afterward (7) and perhaps, to a small extent, after an eruption in 1907 (Fig. 2B) (7), but volcanic cooling over 1896–1910 is only significant at the 2% level. The fact that observed GST becomes somewhat cooler than the reconstruction from 1902 to 1905 hints that the 1902 volcanic forcing might be underestimated, although the larger 1907–1910 underestimate likely requires additional causes (see the “Using global maps to investigate reconstruction biases in 1907–1910 and 1944–1945” section).

Table 1A shows the correlation skill and significance of reconstructed and CMIP5 GST with observed GST over 1896–1910 using low- and high-pass correlations on several time scales. All reconstruction correlations are significant at the 2% level or better, whereas for CMIP5, only the high-pass-filtered seasonal time scale shows significant skill. CMIP5 GST is anticorrelated with observed GST for all low-pass correlations and for multiannual high-pass correlations. However, CMIP5 GST does show a slight overall cooling trend over 1896–1910, although it is much smaller in magnitude than observed. The CMIP5 GST behavior may be explained by a combination of a weak anthropogenic forcing during this period and some of the high-frequency variability being driven by ENSO/IPO and volcanic forcings (that is, the 1902 Santa Maria eruption). In contrast to CMIP5 GST and despite problems with the magnitude of the trend, reconstructed multiannual low- and high-pass correlations (0.78 and 0.80) are nearly as high as values in 1976–1997 and 1998–2013 discussed below.

Period 1911–1940

In this first main warming period, observed GST warmed steadily (Fig. 3A), with the observed linear component of warming over 1911–1940 of $0.38^{\circ} \pm 0.18^{\circ}\text{C}$ being more than that reconstructed ($0.25^{\circ} \pm 0.04^{\circ}\text{C}$) but less certain in magnitude. Both changes are highly significant. The estimated differential warming trend (observed minus reconstructed) of $0.08^{\circ} \pm 0.07^{\circ}\text{C}$ is not significant at the 1% level, but slower reconstructed warming results from warmer than observed GST at the beginning of the period. Warming simulated by CMIP5 was also less than observed ($0.19^{\circ} \pm 0.16^{\circ}\text{C}$) and is only significant

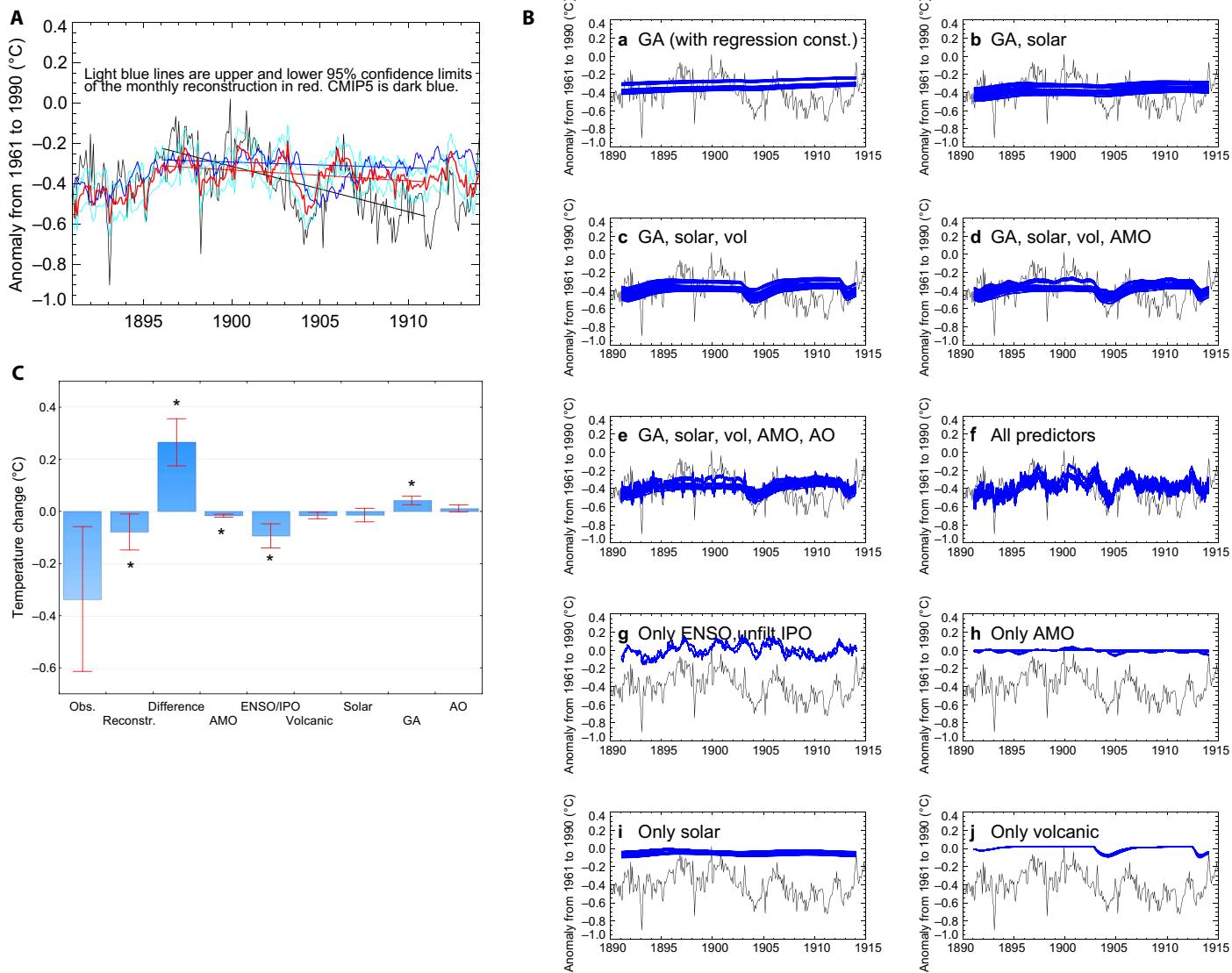


Fig. 2. Reconstructions for slowdown period 1, 1896–1910, where time series include the overlapping longer period 1891–1913 and the observed series 1890–1914. (A) Average reconstruction of GST for 1891–1913. Shown are reconstructions (red), 95% confidence range of GST for the 54 regression equations (light blue), and average CMIP5 simulation of GST (blue) and WMO GST (black). Linear trends are shown from 1896 to 1910 using the REML trend method. (B) Reconstruction of WMO GST for 1891–1913. Shown for each panel are averaged observed GST (black) and each of the 54 regression equation estimates of GST (blue) for the different forcing factors and their combinations and for the three component WMO global temperature data sets. (C) Linear component of total temperature change for slowdown period 1: 1896–1910, for WMO average GST (observations), the reconstruction, its difference from WMO GST, and component forcings. Red lines show $\pm 1\sigma$ uncertainties in the temperature changes. Stars denote significance at the 1% level or better. The REML trend method is used.

at the 2% level. This is explained partly by CMIP5 GST also being warmer than observed at the beginning of the period; significance is further reduced by a high serial correlation of CMIP5 trend residuals. Figure 3B (a to j) shows reconstructions from the 54 main regression equations for individual or combinations of forcing factors for 1911–1940. Most reconstructed warming arises from a combination of GA (a), AMO (h), TSI (i), and volcanic forcing (j), while most of the interseasonal and interannual variability arises from ENSO and the IPO (g). Figure 3C shows that these factors, except the last, appear to contribute significantly to warming over 1911–1940, but none dominates. Allowing for uncertainties, the largest warming factor appears to be GA, giving a warming of $0.12^\circ \pm 0.01^\circ\text{C}$. The AMO contributes $0.05^\circ \pm 0.01^\circ\text{C}$, with a TSI contribution of $0.04^\circ \pm$

0.01°C . Volcanic forcing is smaller at $0.03^\circ \pm 0.01^\circ\text{C}$ but is still significant at the 1% level. Accordingly, increases of greenhouse gases contribute significantly to the early 20th-century GST warming trend but by no means fully explain it, with a more modest TSI influence than in early climate model simulations (18). An AMO warming due to the increasing positive state of the AMO, which a CMIP5 multimodel average cannot capture, is consistent with previous suggestions (31), at least for the Northern Hemisphere component of AMO warming. ENSO and the IPO appear to contribute little to overall warming (Fig. 3C), consistent with alternative instrumental IPO reconstructions (30) not used here. However, this conclusion may disagree with paleoclimatic reconstructions of the IPO (32) where the suggestion was made that the relative lack of Pacific SST data at this time creates an

Table 1. Correlation skill of average reconstructed GST and CMIP5 average GST with WMO GST.

Period	Low pass (r)	Significance (%)	Period	High pass (r)	Significance (%)
A. Slowdown period 1, 1896–1910					
Reconstruction, ≥ 2 months	0.66	0.1	2–10 months	0.32	<0.01
Reconstruction, ≥ 2 seasons	0.72	0.2	2–10 seasons	0.32	2.0
Reconstruction, ≥ 2 years	0.78	2.0	2–10 years	0.80	0.05
CMIP5, ≥ 2 months	-0.34		2–10 months	0.14	
CMIP5, ≥ 2 seasons	-0.47		2–10 seasons	0.34	1.0
CMIP5, ≥ 2 years	-0.75		2–10 years	-0.47	
B. Warming period 1, 1911–1940					
Reconstruction, ≥ 2 months	0.67	<0.01	2–10 months	0.39	<0.01
Reconstruction, ≥ 2 seasons	0.71	0.02	2–10 seasons	0.39	<0.01
Reconstruction, ≥ 2 years	0.78	0.5	2–10 years	0.59	0.1
CMIP5, ≥ 2 months	0.38	0.5	2–10 months	0.16	0.5
CMIP5, ≥ 2 seasons	0.43		2–10 seasons	0.19	
CMIP5, ≥ 2 years	0.52		2–10 years	-0.52	
C. Slowdown period 2, 1951–1975					
Reconstruction, ≥ 2 months	0.64	<0.01	2–10 months	0.13	5.0
Reconstruction, ≥ 2 seasons	0.72	<0.01	2–10 seasons	0.36	0.05
Reconstruction, ≥ 2 years	0.80	<0.01	2–10 years	0.81	<0.01
CMIP5, ≥ 2 months	0.24		2–10 months	-0.02	
CMIP5, ≥ 2 seasons	0.27		2–10 seasons	0.03	
CMIP5, ≥ 2 years	0.31		2–10 years	0.28	
D. Warming period 2, 1976–1997					
Reconstruction, ≥ 2 months	0.80	<0.01	2–10 months	0.28	<0.01
Reconstruction, ≥ 2 seasons	0.86	<0.01	2–10 seasons	0.51	<0.01
Reconstruction, ≥ 2 years	0.94	0.01	2–10 years	0.87	<0.01
CMIP5, ≥ 2 months	0.58	0.5	2–10 months	0.16	1.0
CMIP5, ≥ 2 seasons	0.65	0.5	2–10 seasons	0.20	
CMIP5, ≥ 2 years	0.76	0.1	2–10 years	0.50	5.0
E. Slowdown period 3, 1998–2013					
Reconstruction, ≥ 2 months	0.60	<0.01	2–10 months	0.25	0.05
Reconstruction, ≥ 2 seasons	0.69	0.05	2–10 seasons	0.37	0.5
Reconstruction, ≥ 2 years	0.84	0.1	2–10 years	0.80	0.02
CMIP5, ≥ 2 months	0.34	5.0	2–10 months	0.09	
CMIP5, ≥ 2 seasons	0.41		2–10 seasons	0.20	
CMIP5, ≥ 2 years	0.54		2–10 years	0.26	

erroneous estimate of IPO changes. Figure 3 (A and C) leaves room for additional warming from the IPO and ENSO should future digitized historic instrumental SST data support this (see Discussion and Conclusions).

Table 1B shows the correlation skill and significance of reconstructed and CMIP5 GST with observed GST. Reconstruction skill levels are similar to those for 1896–1910 for low- and high-pass skill (for example, low-pass correlation skill for multiannual periods is

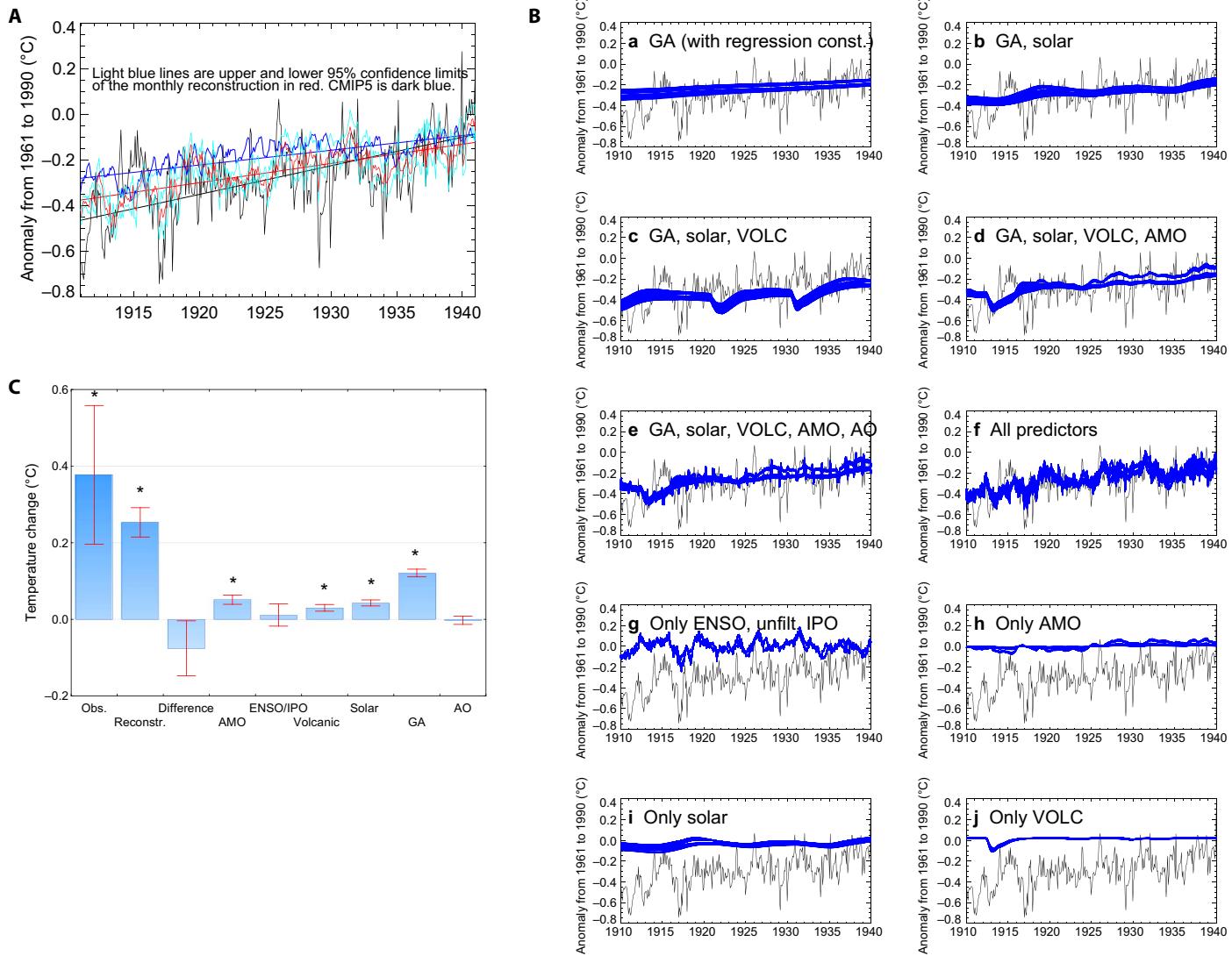


Fig. 3. Reconstructions for warming period 1, 1911–1940. (A) Average reconstruction of GST for warming period 1: 1911–1940. Linear trends are shown from 1911 to 1940 using the REML trend method. Otherwise as for Fig. 2A. (B) Reconstruction of WMO GST for warming period 1: 1911–1940. Otherwise as for Fig. 2B. (C) Linear component of total temperature change for warming period 1: 1911–1940. Stars denote significance at the 1% level or better. Otherwise as for Fig. 2C.

0.78), although high-pass reconstruction interannual skill is less than that for 1896–1910 at 0.59. All reconstruction skill measures are nevertheless significant or highly significant. The smaller high-pass interannual skill in these reconstructions than in 1896–1910 mainly arises from three periods of about a year where reconstructions are poor (Fig. 3A), one of the worst being in much of 1911. Observed GST was markedly cooler than reconstructions, continuing the warm bias in the reconstructions over 1907–1910 noted above. Apparent poor skill around 1915 is also notable when observations were markedly warmer than reconstructions. CMIP5 GST shows some correlation skill (Table 1B) as it picks up the warming trend in observed GST moderately well, but skill as a whole is considerably less than for reconstructions. CMIP5 high-pass skill is low or negative on interannual time scales. Overall, the reconstructions are much more skillful than CMIP5 simulations over 1911–1940 and demonstrate the several causes of observed early 20th-century warming, including increasing greenhouse gases, but none dominates.

Periods 1941–1975 and 1951–1975

The period 1941–1975 (the big slowdown) is the longest (Fig. 4A), with an observed linear cooling of $-0.08^\circ \pm 0.10^\circ\text{C}$, although this is not significant. CMIP5 simulates the big slowdown quite well with almost no overall warming at $0.02^\circ \pm 0.28^\circ\text{C}$. Most other details of the big pause simulated by CMIP5 are poor except for cooling due to the eruption of Mt. Agung in 1963 (Fig. 4A). Reconstructions also show slight overall warming at $0.03^\circ \pm 0.04^\circ\text{C}$ (see also below) but reproduce many of the pronounced annual to interannual fluctuations, including the cooling due to the eruption of Mt. Agung (Fig. 4B, a to j). Figure 4B also shows, as for 1911–1940, that ENSO and the IPO (g) are important for reconstructing interannual GST variations. However, as with CMIP5, a major problem occurs during much of the Second World War. The reconstruction is skillful in 1940, but the observations were warmer than the reconstruction for the next 5 years, especially in 1944 and 1945. This likely arises from observational biases owing to a major change from mainly bucket-based

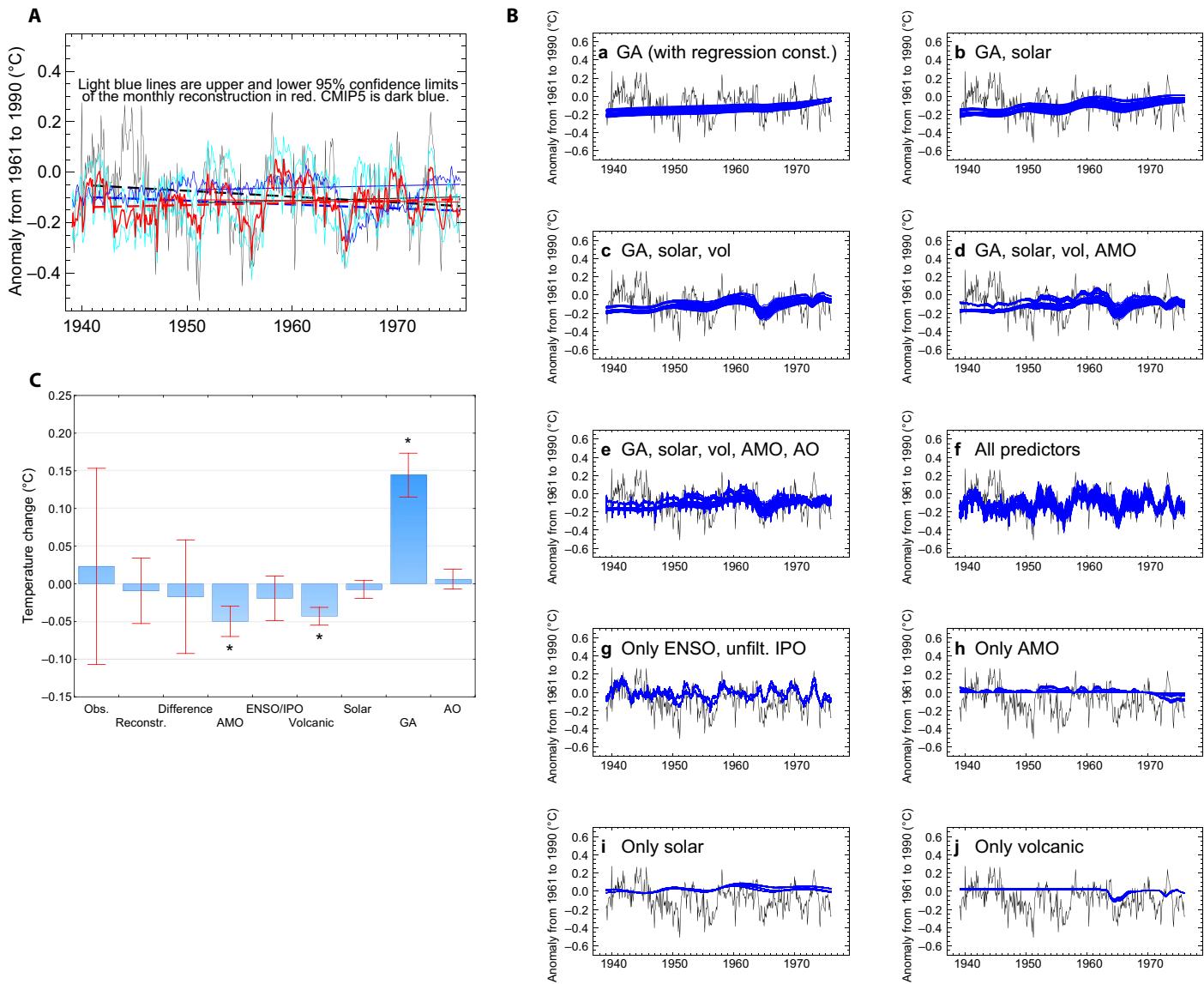


Fig. 4. Reconstructions for slowdown period 2, 1941–1975, where the time series include the lead-in period 1939–1940. (A) Average reconstruction of GST for slowdown period 2: 1941–1975. Thick trend lines are for 1941–1975, and thin lines are for 1951–1975. Otherwise as for Fig. 2A. (B) Reconstruction of WMO GST for slowdown period 2: 1941–1975. Otherwise as for Fig. 2B. (C) Linear component of total temperature change for slowdown period 2: 1951–1975. Stars denote significance at the 1% level or better. Otherwise as for Fig. 2C.

measurements of SST early in the war, adjusted to be warmer than raw observations, to unadjusted ships' engine intake measurements (28, 33). The latter likely became dominant quite suddenly from December 1941 (28, 34) and are likely to be biased warm. This provides the most prominent discrepancy between observed and reconstructed GST. The warm bias in the observations ends suddenly in late 1945 when there was also a sudden partial reversion to bucket-based measurements (28). We return to this problem in the section, “Using global maps to investigate reconstruction biases in 1907–1910 and 1944–1945,” below.

Because of doubtful Second World War data and sparse data immediately after the Second World War, we concentrate the remaining discussion on 1951–1975. Figure 4C shows that the linear component of the observed GST temperature change was almost zero over 1951–1975 ($0.02^\circ \pm 0.13^\circ\text{C}$). The reconstruction also shows no overall warming at $-0.01^\circ \pm 0.04^\circ\text{C}$, as does CMIP5 GST at $0.02^\circ \pm 0.31^\circ\text{C}$.

Thus, both very well reconstruct and simulate the lack of trend in the more secure part of the big slowdown. Figure 4C shows that an essentially zero change in reconstructed GST is the residual of a significant reconstructed warming of $0.14^\circ \pm 0.03^\circ\text{C}$ due to GA and small but significant reconstructed cooling influences due to the AMO ($-0.05^\circ \pm 0.02^\circ\text{C}$) and volcanic forcing ($-0.04^\circ \pm 0.01^\circ\text{C}$), with perhaps a small insignificant cooling due to ENSO and the IPO ($-0.02^\circ \pm 0.03^\circ\text{C}$). Thus, net warming of GST from GA continued very slowly during much of the big slowdown due to an increasing influence of anthropogenic aerosols at that time (20) but accelerated after 1970 [Figs. 1B (e) and 4B (a)]. GA warming just cancelled the small signals from several natural forcing factors that collectively modestly cooled GST.

The good correlation skill of reconstructed GST over 1951–1975 is reflected in Table 1C, which also shows the poor skill of CMIP5 GST. Most reconstruction correlation statistics with observed GST

are significant at better than the 1% level, but no CMIP5 statistics reach this level. Reconstruction statistics are particularly good for low-pass multiannual and for high-pass interannual time scales at 0.80 and 0.81, respectively, although much smaller, high-pass seasonal correlations are still significant at the 0.05% level with a correlation of 0.36, reflecting interseasonal skill contributed by ENSO and the IPO. Table S3 shows correlations for 1941–1975 that are, as expected, generally lower but still show some skill.

Period 1976–1997

This warming period shows rather larger multiannual fluctuations in the rate of warming of GST than does 1911–1940 (compare Fig. 5A with Fig. 3A). Observed fluctuations in GST in the early 1980s and 1990s are caused by volcanic forcing (Fig. 5B, j) due to the eruptions of El Chichón and Mt. Pinatubo (7, 35). Cooling in the late 1980s arises from ENSO and the IPO (Fig. 5B, g), particularly because a moderately

strong El Niño in 1987 was followed by a very strong La Niña in 1988–1989 (www.cpc.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml, as of 16 March 2017). CMIP5 GST reflects the volcanic cooling signals skillfully and the stronger GA influence than in 1911–1940, but of course, it cannot capture the ENSO-related cooling between 1987 and 1989.

Figure 5C shows that the warming of observed and reconstructed GST over 1976–1997 ($0.38^\circ \pm 0.15^\circ\text{C}$ and $0.36^\circ \pm 0.04^\circ\text{C}$, respectively) was both significant at the 0.01% level and similar. CMIP5 gives a slightly smaller warming of $0.31^\circ \pm 0.22^\circ\text{C}$ to that observed, a rate of $0.14^\circ \pm 0.10^\circ\text{C}$ per decade significant at the 1% level, compared to the observed rate of $0.17^\circ \pm 0.07^\circ\text{C}$ per decade. A key result (Fig. 5, B and C) is that, although there was a small but significant reconstructed warming influence from the AMO ($0.02^\circ \pm 0.01^\circ\text{C}$) and a somewhat larger but small overall reconstructed cooling influence from volcanic forcing ($-0.05^\circ \pm 0.03^\circ\text{C}$), much of the dominant influence was net

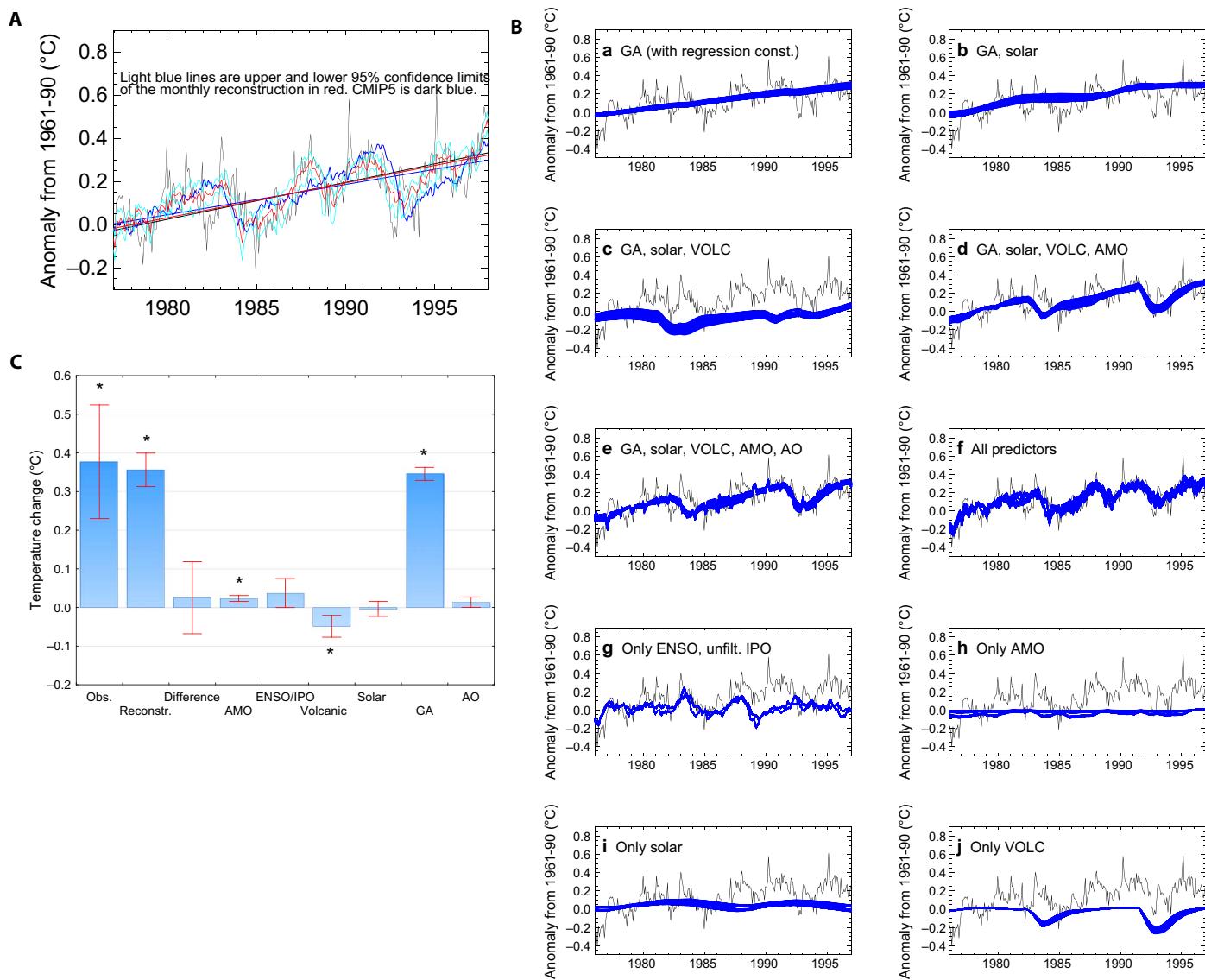


Fig. 5. Reconstructions for warming period 2, 1976–1997. (A) Average reconstruction of GST for warming period 2: 1976–1997. The linear trends are for 1976–1997. Otherwise as for Fig. 2A. (B) Reconstruction of WMO GST for warming period 2: 1976–1997. Stars denote significance at the 1% level or better. Otherwise as for Fig. 2B. (C) Linear component of total temperature change over warming period 2: 1976–1997. Stars denote significance at the 1% level or better. Otherwise as for Fig. 2C.

warming due to GA at $0.35^\circ \pm 0.02^\circ\text{C}$. This is a reconstructed rate of $0.16^\circ \pm 0.01^\circ\text{C}$ per decade. Reconstruction correlation skill with observed GST was high or very high (Table 1D) and generally higher for both low-pass and interannual high-pass correlations than in the previous periods. Thus, low-pass multiannual correlation of reconstructed and observed GST reached 0.94, with a high-pass interannual correlation of 0.87. Shorter time scale high-pass skill estimates were much lower but still significant. CMIP5 correlation skill is higher than that for previous periods but markedly less than the reconstruction skill, with a low-pass multiannual correlation with an observed GST of 0.76 and an interannual high-pass correlation of 0.50.

Period 1998–2013 and subperiods 2001–2010 and 2001–2013

This third slowdown period is often known as the pause or hiatus. It is not easy to define, but it started in about 1998 and can currently be regarded as finishing in 2013 (Fig. 1A). Besides taking an overview of the slightly longer period 1997–2015, we investigate the three overlapping slowdown periods 1998–2013, 2001–2010, and 2001–2013. Figure 6A shows that reconstructed GST from the beginning of 1997 until the end of 2015 follows observed GST quite closely on interannual and some subannual time scales. Monthly observed GST often deviates strongly from the reconstructions, and as elsewhere

in the record, observed monthly GST values have a much larger uncertainty than annual values [1σ uncertainties in HadCRUT4 average near 0.09° and 0.05°C , respectively, over 1997–2015 (8)]. CMIP5 GST, by contrast, fails to capture the slowdown at all, much as previously noted (2). CMIP5 average GST also shows artificial fluctuations on the annual time scale in this period (Fig. 6A). Again, see section S6 for a discussion of apparent annual cycles in CMIP5 GST anomalies. These can cause artificial high-frequency correlations between observed and CMIP5 GST.

Figure 6B first shows that reconstruction uncertainty tends to increase after 2010. Figure 6B (a) reflects a steady average reconstructed warming due to GA averaging 0.17°C per decade, about the same as 1976–1997 despite the observed slowdown. This indicates that no slowdown occurred in background GA warming. Figure 6B (b) shows that adding reconstructed TSI forcing markedly improves the shape of reconstructed GST from GA forcing alone. Adding all other forcing factors, especially ENSO and the IPO, considerably increases this correspondence on all time scales down to the subannual. The match between observed and average reconstructed GST is now quite close throughout 1997–2015 (Fig. 6A), except on monthly time scales and for some multimonth periods.

Figure 6C shows the reconstructed GST changes. In all three periods (1998–2013, 2001–2010, and 2001–2013), the observed but insignificant GST changes ($0.11^\circ \pm 0.17^\circ\text{C}$, $0.07^\circ \pm 0.15^\circ\text{C}$, and

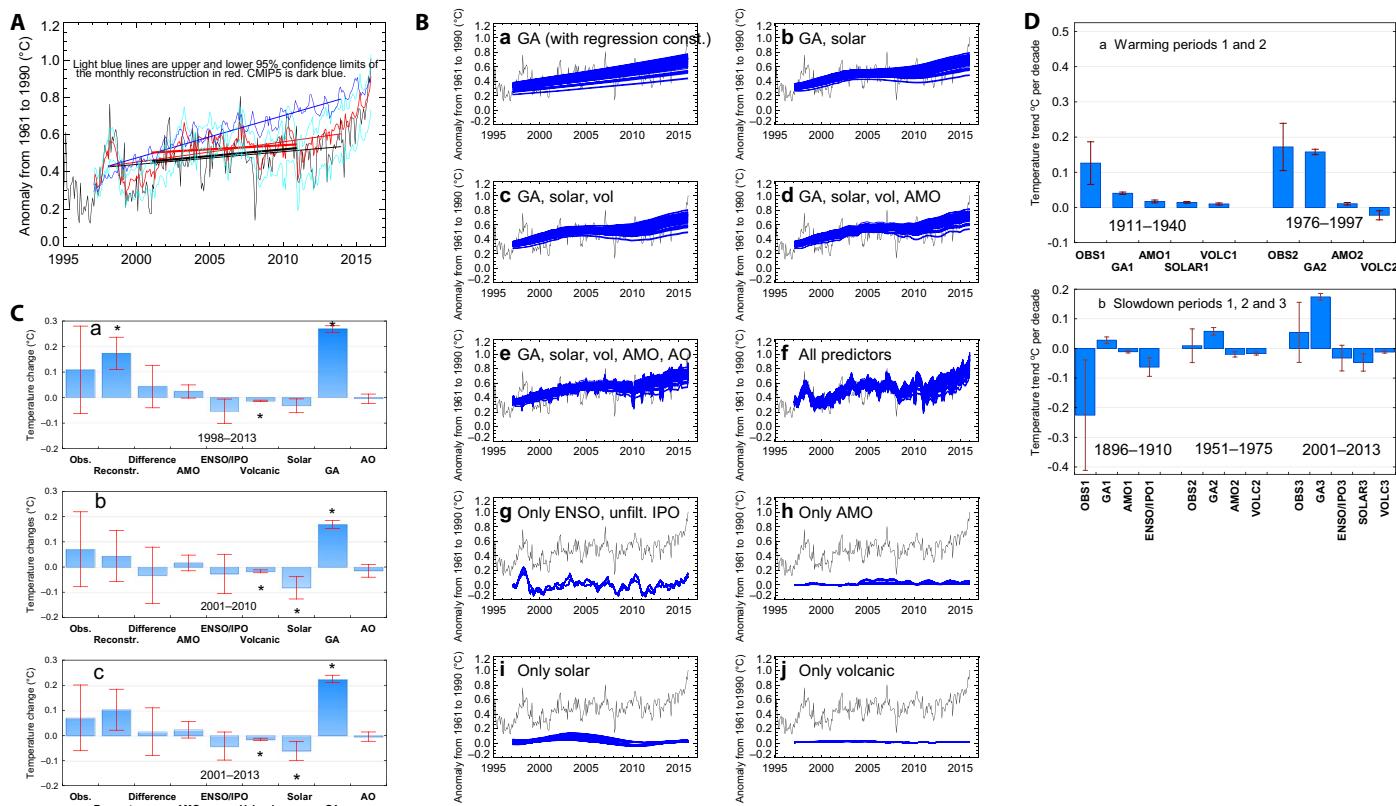


Fig. 6. Reconstructions for slowdown period 3, 1997–2015 and its three main sub-periods, where the observed time series include the lead-in period 1995–1996. (A) Average reconstruction of GST for slow-down period 3: 1997–2015. The linear trends are for 1998–2013 (thin lines) and 2001–2010 (thick lines). Otherwise as for Fig. 2A. (B) Reconstruction of WMO GST for slow-down period 3: 1997–2015. Otherwise as for Fig. 2B. (C) Linear components of total temperature change over slow-down period 3. (a) 1998–2013; (b) 2001–2010; (c) 2001–2013. Stars denote significance at the 1% level or better. Otherwise as for Fig. 2C. (D) Summary of the contributions of significant forcing (at <1% level) factors to GST trends ($^\circ\text{C}$ per decade) during (a) warming periods and (b) slowdown periods. The period 2001–2013 is used to represent slow-down period 3 where the appreciable but not significant ENSO-induced trend is shown.

$0.07^\circ \pm 0.13^\circ\text{C}$, respectively) are quite well (1998–2013) or very well (2001–2010 and 2001–2013) matched by corresponding reconstructed changes ($0.17^\circ \pm 0.06^\circ\text{C}$, $0.04^\circ \pm 0.10^\circ\text{C}$, and $0.10^\circ \pm 0.08^\circ\text{C}$), although reconstructed warming over 1998–2013 is significant. The CMIP5 GST change over 1998–2013 is, by contrast, highly significant at $0.36^\circ \pm 0.10^\circ\text{C}$, a rate of $0.23^\circ \pm -0.06^\circ\text{C}$ per decade that varies little over the period (Fig. 6A). This is markedly higher than the 1976–1997 rate of GST increase shown by CMIP5. The reconstructed component of GST change due to GA alone over 1998–2013 is $0.27^\circ \pm 0.01^\circ\text{C}$. This is not only much more than the total observed warming of 0.11°C but also distinctly less than the total warming shown by CMIP5. The latter, of course, will have been unable to simulate changes in the IPO, ENSO, and the AO, and CMIP5 models did not have observed volcanic and solar forcings for much of this period. A key published finding for the 1998–2013 slowdown, partly based on model simulations, is a cooling influence of ENSO/IPO (36, 37). However, over the early part of the period 1999–2008, Fig. 6B (g) is consistent with a slight overall warming of GST from ENSO/IPO (38). Nevertheless, reconstructed cooling due to ENSO/IPO is seen in the three subperiods (1998–2013, $-0.05^\circ \pm -0.05^\circ\text{C}$; 2001–2010, $-0.03^\circ \pm -0.08^\circ\text{C}$; 2001–2013, $-0.04^\circ \pm -0.06^\circ\text{C}$). However, none of these ENSO/IPO cooling influences is significant at the 1% level, although the 1998–2013 cooling is significant at the 5% level. By contrast, the TSI cooling influence is significant at the 0.1% level or better in 2001–2010 and 2001–2013 and at the 2% level in 1998–2013. Except in 1998–2013, when solar forcing actually causes some warming of GST between 1998 and 2002, overall solar cooling is larger than that due to the ENSO/IPO [Figs. 6B (i) and 1B (d)]. Over 1998–2013, 2001–2010, and 2001–2013, TSI cooling of GST is reconstructed as $-0.03^\circ \pm 0.03^\circ\text{C}$, $-0.08^\circ \pm 0.04^\circ\text{C}$, and $-0.06^\circ \pm -0.04^\circ\text{C}$, respectively, with the largest TSI cooling influence over 2001–2010. The “Using global maps to investigate reconstruction biases in 1907–1910 and 1944–1945” section provides further evidence of a TSI cooling signal. A consistent but small significant volcanic cooling influence in each subperiod (Fig. 6C) agrees with previously reported volcanic cooling influences (1). Finally, although the influence of the AO is always small and not significant, it gives a visible cooling of $-0.015^\circ \pm 0.025^\circ\text{C}$ in 2001–2010 (Fig. 6C). This is caused by a distinct trend in December to March toward a more negative AO over this period, creating increasingly cold winter conditions over mid- and high-latitude Eurasia (27).

There remain doubts about the measurement of GST during the recent slowdown because rapidly increasing Arctic warming may be underestimated (39), possibly even relegating the recent slowdown to statistical insignificance (40). An offsetting problem is that Antarctica as a whole has been warming much more slowly than the global average, and it is also poorly sampled. The GST data sets do not represent either region very well. This problem has recently been investigated (41) using the ERA Interim Reanalysis (42) to estimate better recent values of GST. ERA Interim estimates near-surface air temperature in data-sparse regions in a physically consistent way with the help of a weather forecasting model. Use of ERA Interim, however, does not lead to significantly different conclusions about the causes of the 1998–2013 slowdown (see Materials and Methods and section S8). The observed warming of ERA Interim GST over 1998–2013 is rather larger than that for WMO GST at $0.20^\circ \pm 0.20^\circ\text{C}$, although still not significant at the 1% level, mainly because it represents the rapidly warming Arctic better and agrees somewhat better with the reconstructed GST warming of $0.17^\circ \pm -0.06^\circ\text{C}$. However, ERA Interim GST warming increases after 2001 only slightly more than WMO GST at $0.11^\circ \pm 0.26^\circ\text{C}$ (2001–2010)

and $0.10^\circ \pm 0.20^\circ\text{C}$ (2001–2013) (see figs. S6 and S7), so our analysis does not detect any significant differences in the influences of the forcing factors on the 1998–2013 slowdown when using ERA Interim GST in place of WMO GST, although ERA Interim does hint at more observed warming.

Our main conclusion is that the most robust influence on the 1998–2013 slowdown resulted from the cooling effects of reduced TSI forcing between 2003 and 2011 (Fig. 6B, i) followed by a trend toward increasing La Niña conditions and a negative IPO in its second half. The negative IPO is in turn likely to have been enhanced by regional anthropogenic aerosol forcing (43). Cooling throughout the slowdown was slightly but significantly enhanced by persistent but small increases in volcanic forcing. Previous studies (1, 44) pointed out this modest contribution to the recent slowdown from small volcanic eruptions, and our results add to this evidence. Despite this, the background net influence of GA warming continued at a nearly constant rate of $0.17^\circ \pm 0.01^\circ\text{C}$ per decade.

Using global maps to investigate reconstruction biases in 1907–1910 and 1944–1945

For the first slowdown period 1896–1910, reconstructions for the subperiod 1907–1910 were too warm. To investigate this further, global maps combining surface air temperature anomalies over land with SST were reconstructed for 1907–1910 using the monthly WMO data set at 5° latitude $\times 5^\circ$ longitude resolution (Fig. 7A). For each $5^\circ \times 5^\circ$ box, significance of the differences between reconstructions and observations was calculated at the 1% level using a two-sided *t* test. Figure 7A (a to c) shows that the main cause of the reconstruction warm bias is relative observed coolness of much of the global oceans. Accordingly, most of the oceans are observed to be significantly (at the 1% level) cooler than reconstructions. By contrast, land surface temperatures do not show large-scale differences of either sign. Besides possible observed SST cold biases and most SSTs being strongly positively bias-corrected at this time (8, 33, 34), these differences might reflect a missing or poorly estimated forcing factor in reconstructed SST. Possibilities are inadequate estimates of AMO and IPO (32) forcings or perhaps underestimated volcanic cooling, for example, from the apparently modest 1907 eruption of Ksudach, Kamchatka (7), although this might be expected to be more apparent in land surface temperatures.

A bigger problem, but of opposite sign, occurs during the Second World War where reconstructed GST is much colder than observations, especially in 1944 to 1945. Figure 4A suggests that this offset lasted from about 1942 until late 1945 or the start of 1946. Figure 7B (a to c) shows similar global maps for reconstructions and observations averaged over 1944 and 1945. Observed SST, which is currently not bias-adjusted, is warmer than reconstructions almost worldwide. As mentioned above, a possible reason is a major increase in engine intake SSTs that are often biased warm (5, 33). This may have been a particular problem with the mix of warships and merchant ships reporting SST during the Second World War. Soon after the war, an opposite problem rapidly developed due to an increase in the fraction of SSTs measured using cold-biased uninsulated buckets (28, 45). Therefore, the fact that the warm bias problem did not last beyond 1946 is consistent with a warm SST bias caused by dominant engine intake data. Another possibility is that reconstructed AMO warming was underestimated during the Second World War. Against this possibility, AMO warming usually affects the North Atlantic and North Pacific, while the Southern Hemisphere tends to be near or slightly cooler than

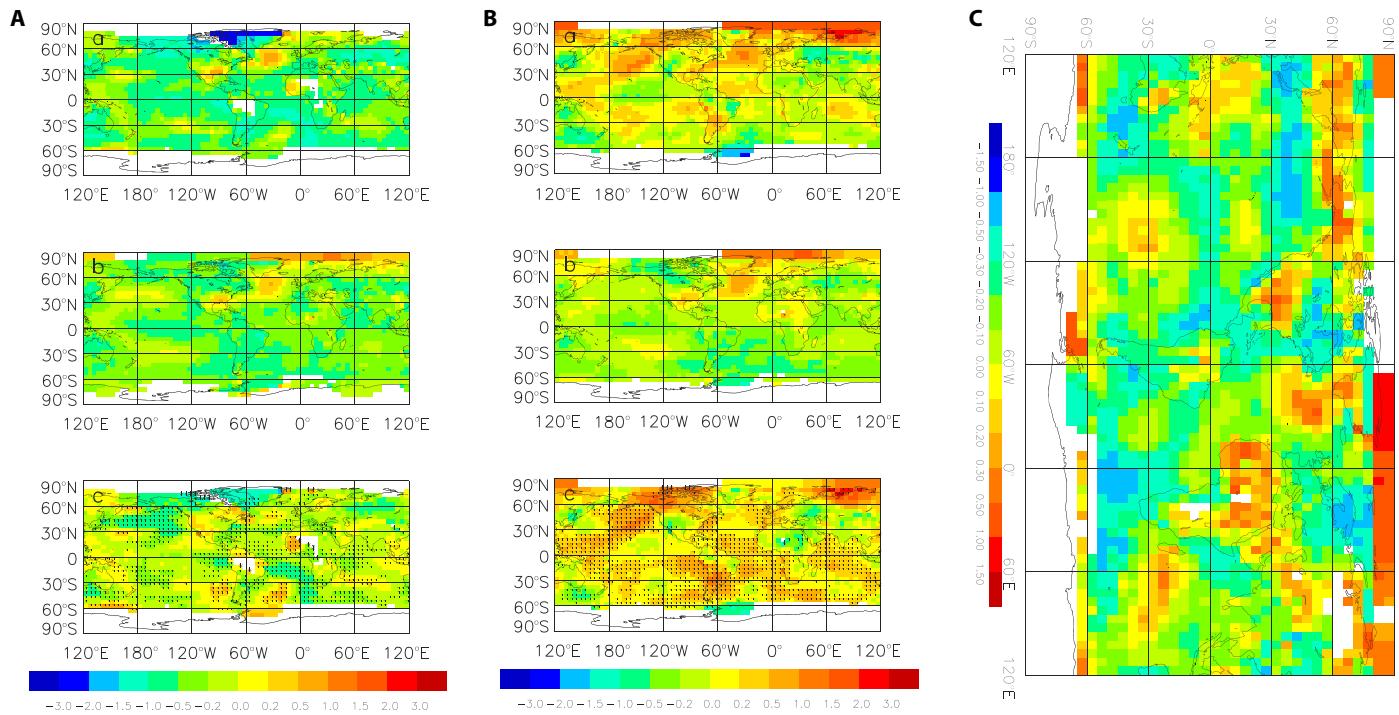


Fig. 7. Global maps of observed and reconstructed temperature anomalies for some key results. (A) (a) Observed maps of temperature anomalies ($^{\circ}\text{C}$) in 1907–1910. (b) Reconstructed anomalies. (c) Observed minus reconstructed anomalies. Significant differences at the 1% level in each $5^{\circ} \times 5^{\circ}$ box for observed minus reconstructed anomalies (two-sided t test) are shown by stars. Dark red indicates warmest, and dark blue denotes coldest. (B) (a) Observed maps of temperature anomalies in 1944–1945. (b) Reconstructed anomalies. (c) Observed minus reconstructed anomalies. Otherwise as for (A). (C) Reconstructed linear component of worldwide temperature change due to TSI forcing, 2001–2010. Calculated changes for $5^{\circ} \times 5^{\circ}$ boxes are ordinary least-squares trends multiplied by the period length. Dark red indicates warmest, and dark blue denotes coldest.

normal (26). However, evidence for a potential observed warm bias in Fig. 7B appears to be truly global.

Further investigation of TSI forcing

Figure 7C shows a map of least-squares regression temperature changes due to TSI forcing for $5^{\circ} \times 5^{\circ}$ boxes in 2001–2010, the decade when we reconstructed the strongest TSI cooling influence on the recent slowdown. REML calculations are too time-consuming for such a large data set. A worldwide cooling signal is fairly consistent over the oceans but very patchy over land. Its average global magnitude of -0.13°C is larger than the REML estimate of -0.08°C , partly because calculations are made separately for each $5^{\circ} \times 5^{\circ}$ box, and REML estimates are likely more accurate for short periods. Nevertheless, Fig. 7C supports the plausibility of a substantial cooling influence of TSI forcing on GST over 2001–2010. The reconstructions described above of solar cooling of GST using REML are consistent in magnitude with the larger responses of GST to TSI forcing simulated by CMIP5 models over 11-year cycles (46).

DISCUSSION AND CONCLUSIONS

We have shown that much of the variability and trends in GST since 1891 can be reconstructed skillfully down to interannual, annual, and, sometimes, subannual time scales. We have highlighted two warming periods, 1911–1940 and 1976–1997, and three slowdown periods, 1896–1910, 1941–1975, and 1998–2013. For each slowdown, we reconstructed slightly longer periods to show how the periods started and ended. For each slowdown, the regression equations used train-

ing data that were independent of the period studied. For warming periods, we relied on the same cross-validation procedure used for slowdowns but used all the GST and forcing data for training the regression method. Although we have allowed for many uncertainties, a full analysis requires a better understanding of the uncertainties in forcings. In this regard, estimates of TSI forcing before the satellite era have evolved considerably (20), but the increasing trend in TSI before 1960 is probably still not known well. Finally, we have not used recent estimates of a small estimated cooling in recent decades due to global greening (47). This effect is small over land; globally, it is smaller still because the ice-free land surface covers only about 25% of the globe.

From soon after the Second World War, there are no serious failures of the reconstructions. However, many monthly and some multi-monthly variations in observed GST are not reconstructed well. Some of these apparent errors are due to observational noise, with observed monthly GST uncertainties throughout the period since 1891 being about twice as large as simultaneous annual GST uncertainties (8). The lack of observed warming of GST over 1951–1975 is well reconstructed, as is strong warming from 1976 to 1997 and the observed warming peak in 1998 due to the 1997–1998 El Niño and the subsequent GST slowdown. Renewed warming in 2014–2015 is also reconstructed well. Less expected is the relatively large reconstructed influence of TSI forcing on the 1998–2013 slowdown, especially when compared to ENSO and IPO forcing. Figure 6 (B and C) shows that uncertainties arising from the use of three quite different assumptions about the delayed GST response time to TSI forcing did not stop reconstructed TSI influences from being significant from 2001 onward,

although uncertainties arising from these markedly varying response times are included in REML calculations of temperature change and trend. Recent research (46) tends to favor a 4-year GST response time, the middle value of our three response times (see also fig. S8), as this corresponds well to the approximately 2-year lag of GST to TSI forcing found there.

Despite important influences of natural variability, since 1891, the net warming effects of increasing greenhouse gases and anthropogenic aerosols on GST have been dominant, shown most clearly by Fig. 1B (e). Even over 1891–1950, net anthropogenic warming was clearly important, although natural effects played a comparable role. Figure 6D compares drivers of each warming (a) and slowdown period (b) in the form of reconstructed GST trends to aid comparison. The first warming period 1911–1940 is unlike the second, with warming in the first period being about equally due to GA forcing on the one hand and four natural forcings on the other. However, in 1976–1997, GA forcing dominated the warming. Figure 6D (b) shows the greatly increasing influence of GA warming on GST trends from the first slowdown period to the third. In all slowdowns, GA warming is fully or partly offset by natural cooling. The first slowdown, a nominal pronounced cooling, was probably dominated by natural ENSO/IPO cooling, but its magnitude and that of the overall cooling remain uncertain. The second slowdown, at least in 1951–1975, is much more securely reconstructed; here, GA warming was mostly offset by natural cooling owing to a decline in the AMO and increased volcanic forcing. In the third slowdown, as measured over 2001–2013, the strongest of all GA warming signals is partly offset by several natural cooling factors: solar cooling, an increasing tendency to La Niña events compared to El Niño events, and a small signal due to increasing volcanic activity. By contrast, since the end of the 20th century, the CMIP5 GST warming rate is likely to be almost all caused by GA warming where its forcing of CMIP5 GST may be a little too fast at $0.23^\circ \pm 0.06^\circ\text{C}$ per decade. Given, say, another decade of observed GST data, REML trend methods could be used to test stringently whether CMIP5, or forthcoming CMIP6, rates of early 21st-century warming really significantly differ from observations globally or hemispherically.

Progress to better understand worldwide surface temperature variations partly depends on improving the instrumental database. It is clear that a large number of historical observations not yet digitized exist back to before 1800 (48) and especially in much of the period studied here. They await extraction and digitization through projects like the international Atmospheric Circulation Reconstructions over the Earth Project (48), which has already delivered a useful fraction of these data (www.met-acre.net/). In conclusion, the mostly high reconstruction skill of observed monthly GST shown here is consistent with the general veracity of the GST record since 1891 and a good understanding of the multiple causes of its variations and trends.

MATERIALS AND METHODS

We used cross-validated linear regression to relate monthly GST de-seasonalized anomalies relative to 1961–1990 from each of three records (5, 8, 29) to a set of monthly forcing factors. These were also expressed relative to 1961–1990 but standardized over 1891–2015. Linearity in the response of GST to the sum of combinations of forcing factors remains a widespread assumption in climate change attribution (10). The forcing factors were transformed to reflect the character of GST responses to these factors, often using a delayed e-folding response profile. We included a range of response times

to account for their uncertainties, and some new estimates of the forcings, to create 54 main regression equations for each period studied. Each main equation for warming periods 1911–1940 and 1976–1997 averages 1500 constituent cross-validated regression equations, the number of months in the training period 1891–2015. Table S2 shows the 18 combinations of forcing factors used in each main equation for a given record of the three records contributing to WMO GST. For the slowdown period 1896–1910, part of a longer study period 1891–1913, training data for 1914–2015 were used in the same way to create 54 main regression equations, each the average of 1224 cross-validated regression equations. For 1941–1975, training data combined the periods 1891–1938 and 1976–2015 in a similar way, and for 1998–2013, the training period was 1891–1996. This procedure minimized biases in the calculations of regression coefficients. The symmetrical cross-validation window centered on a given month was 49 months long for all equations, although necessarily asymmetrical at the beginning and end of GST data sets.

We only chose those predictors known to physically affect GST using published literature. The GA predictor (20) was based on the sum of all the anthropogenic forcing factors listed and thus included all factors listed such as land surface forcing estimates. Because these GA data ended in 2011, they were updated to 2015 in two ways, consistent with the evolution of the RCP4.5 and RCP8.5 forcing time series, respectively, and the average was taken. The TSI predictor used new data compiled for CMIP6 using data for the CMIP6 Historical Simulation until 2014 and the first year of data (2015) for simulations of the future (21). For the AMO predictor, we used an average of an index based on an eigenvector analysis of worldwide SST (26) and one calculated from North Atlantic minus global SST using HadISST1 (33). Twelve-month running means of the AMO indices were used, leading monthly GST by 1 year (12). For ENSO, we used a Niño 3.4 index leading GST by 4 months [leads in the range 3 to 6 months were observed (49) and gave similar regression coefficients], calculated using HadISST2.1.0.0 (50) until 2010. This contains more tropical Pacific data than HadISST1 but had to be updated with HadISST1 from 2011 to 2015. For the IPO, whose time series is highly correlated with that of the Pacific Decadal Oscillation (26, 30), we created an unfiltered monthly index using a published IPO pattern (26) and HadISST1. This also leads GST by 4 months. Monthly Niño 3.4 (ENSO) and IPO indices were separately used in 9 of the 18 main regression equations for each period (section S5). The AMO and IPO (and approximately the related ENSO) indices are independent of global mean SST via the eigenvector analysis that mainly created them and thus are largely independent of GST, minimizing circularity with GST in regression equations (26). The volcanic index is an update (51) of data used to study the recent GST slowdown (1). This finishes in 2014; thus, values for months in 2015 were persisted, using the value for December 2014. The average delayed e-folding time of the response of GST to tropical volcanic eruptions was estimated at 8 months from published data (35). We have also used 6 and 12 months; this range of short e-folding times reflects the likelihood that climate responses to short-term sharp cooling are substantially less than those for the slower GA or TSI forcing changes. The average observed GST response time to major volcanoes is generally longer at near 15 months; this is consistent because negative volcanic forcing takes time to develop fully after an eruption (Fig. 1B, c). We have used the same delayed e-folding times for periods when negative radiative volcanic forcing increased in magnitude and when it decreased. The AO index was calculated using the 20th Century Reanalysis (52) as the time series of

the first eigenvector of pressure at mean sea level over 30° to 90°N in December to March. It was set to zero otherwise.

In the regression procedure, 18 of the 54 equations each use e-folding times of GST to TSI forcing of 2 years, 4 years, and 10 years (table S2). There are widely varying estimates of this response time from less than a year (13) to near a decade (53). Section S9 shows the effects on TSI forcing of using different e-folding response times. For the response to GA forcing, our delayed e-folding time of 4 years partly derives from a published model (54) forced with changing greenhouse gases alone. We also used a 10-year GA response time; supporting this are estimated e-folding times of near 8 years for the HadCM3 model (54) and also of 10 years (14). An e-folding GST response time of 4 years is nevertheless consistent with the most comprehensive study so far (46) where the average lag of CMIP5 model responses of near 2 years of GST to TSI forcing is also consistent with this value. Finally, we added a “recalcitrant response” of GST to multi-century changes in forcing (54), using an e-folding time of 200 years; an e-folding time of 500 years gives little difference. Recalcitrant forcing was added to transformed GA series with both 4- and 10-year e-folding response times. Recalcitrant forcing r_i was calculated from all external forcing factors, including volcanic forcing, using data back to 1750 (20) and projected back to 1290 using average forcing for 1760–1800 to represent forcing throughout the Little Ice Age. The response of GST to recalcitrant forcing r_i was added linearly to that of GA forcing g_i to give a total response T

$$T = (1/c_1)\sum_i g_i \exp((-t_i + 0.5)/d_1) + \\ 0.02(1/c_2)\sum_i r_i \exp((-t_i + 0.5)/d_2) \quad (2)$$

where $c_j = \sum_{i=1,n} \exp((-t_i + 0.5)/d_j)$. Here, n is large enough (that is, the integration goes far enough in the past) such that $\sum_{i=n+1,\infty} \exp((-t_i + 0.5)/d_j)$ is small and can be neglected. Here, t is the time in months, $d_1 = 48$ or 120, consistent with the chosen GST response times to GA forcing of 4 and 10 years, and $d_2 = 2400$, consistent with a recalcitrant response time of 200 years. Here, GA values used in the calculations are equal to T . Equation 2 gives slightly more weight to recalcitrant forcing than previously published (54) where a factor of 0.014 replaced our value of 0.02. However, a much larger factor of 0.1 has very little influence on our regression results.

For modeled GST, we used all available CMIP5 models. The model global means were not masked by the observational coverage. The GST were first averaged across the members of each model ensemble (for models for which such an ensemble was available) and then across the multimodel ensemble. The RCP8.5 CMIP5 simulations were used to cover the 2006–2015 period not covered by the historical CMIP5 simulations.

For observed GST, we separately used three data sets: HadCRUT4.6 updated from HadCRUT4 (8) (www.metoffice.gov.uk/hadobs/hadcrut4, as of 28 November 2017), GISTEMP (29) (https://data.giss.nasa.gov/gistemp/tabledata_v3/GLB.Ts+dSST.txt, as of 28 November 2017), and the NCDC National Centers for Environmental Information data set (5) (www.ncdc.noaa.gov/monitoring-references/faq/anomalies.php, as of 28 November 2017). We also used the average of these three data sets, known as the WMO average. Further details are given in section S10. Furthermore, we used an adjusted version of published ERA Interim GST (41) during 1997–2015, which has a physically consistent coverage of the whole globe. We adjusted the ERA Interim GST anomalies,

published relative to 1981–2010, to make them consistent with those of WMO GST anomalies relative to 1961–1990, using overlapping monthly data for 1983–1996. ERA Interim GST data in 1979–1982 were not used as they are inhomogeneous with later data (41).

To create global regression maps, we used the mean of the three temperature data sources (5, 8, 29) with a $5^\circ \times 5^\circ$ spatial resolution. For a given period mapped, we used the same training period and cross-validation method as for the corresponding time series analysis. We simplified the calculations by using a single set of cross-validation equations taking appropriate averages of predictors used in the 18 equations (not 54 because we are using just average WMO GST). A separate regression was created in this way for each $5^\circ \times 5^\circ$ region, whether land or ocean; this simplification was adequate as uncertainties were not used. We then reconstructed $5^\circ \times 5^\circ$ box temperatures over periods longer than a month by averaging constituent months. Finally, the contribution of each predictor to temperature in each $5^\circ \times 5^\circ$ box was mapped by multiplying the spatially varying regression coefficients for that predictor by the average predictor time series for that box for the period covered.

Cross-correlation significance calculations account for autocorrelations (55). Successive autocorrelations with increasing lag for each series were used to calculate cross-correlation degrees of freedom until one time series has a lagged autocorrelation of <0.10 , approximately the threshold of autocorrelation significance at the 5% level. Calculation of trends or linear temperature changes and their significance uses REML, particularly appropriate for short data sets (section S3) (56). REML allows for persistence in monthly residuals through their first serial correlation values. Uncertainties in the monthly GST reconstructions resulting from using 54 different equations or from uncertainties in observed GST were included in REML calculations. For observed GST, uncertainties for HadCRUT4.6 GST were used (8), increased by taking account of differences between the three WMO data sets for each month. For the CMIP5, GST average published uncertainties were used. Data or model uncertainties only slightly affect the calculation of GST trend or change but reduce trend significance, often quite appreciably, when compared to regression methods neglecting this source of uncertainty. Finally, section S11 and tables S4 (A and B) show cross-correlations between predictor forcings for 1891–2015 and for 1951–2015. Appreciable collinearity between TSI and GA forcing disappears after 1950 but is appreciable over the period as a whole, although it does not usually cause large uncertainties. The other predictors do not have correlations large enough to cause significant problems.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/4/6/eaao5297/DC1>

section S1. Key forcings used expressed in absolute units

section S2. Separating the two AMO time series and ENSO from the IPO

section S3. The REML linear temperature change method

section S4. CMIP5 model experiments used

section S5. Structure of the regression equations and cross-correlation of predictors

section S6. Comparison of annual reconstruction and CMIP5 root mean square errors with estimated WMO GST annual uncertainties

section S7. Reconstruction and CMIP5 correlation statistics for 1941–1975

section S8. Substituting ERA Interim GST for WMO GST for 1997–2015

section S9. Response times of GST to TSI forcing

section S10. Further details of the WMO data sets

section S11. Structure of multiple regression residuals and cross-correlation of predictors

fig. S1. Volcanic, solar, and GA forcings in original units without smoothing, expressed as anomalies from their 1961–1990 averages.

fig. S2. Additional details of monthly time series of the predictors.

- fig. S3. Maximum likelihood plot of the serial correlation of the residuals and their variance for the reconstructed linear component of WMO temperature change over 2001–2013.
- fig. S4. Identifying the boundaries between consecutive slowdown and warming periods using trend analysis.
- fig. S5. Reconstructed and CMIP5 root mean square errors for annual data compared to 1σ uncertainties in annual WMO GST values including additional uncertainties due to the spread of the three component data sets.
- fig. S6. Comparing the reconstructions using ERA Interim and WMO GST during slowdown period 3.
- fig. S7. As in the Fig. 6C panels for 1998–2013, 2001–2010, and 2001–2013 in the main text but for ERA Interim.
- fig. S8. Original (unfiltered) and e-folded monthly TSI time series January 1981 to December 2015 reflecting the three sets of delayed e-folding response times used in the paper.
- table S1. CMIP5 models and historical experiments used in this paper.
- table S2. Combinations of predictors used in the 18 main regression equations.
- table S3. Correlation skill of average reconstruction GST and CMIP5 average GST with WMO GST in slowdown period 2, 1941–1975.
- table S4. Cross-correlation of monthly predictors 1891–2015.
- table S5. Cross-correlation of monthly predictors 1951–2015.
- References (57–59)

REFERENCES AND NOTES

- B. D. Santer, C. Bonfils, J. F. Painter, M. D. Zelinka, C. Mears, S. Solomon, G. A. Schmidt, J. C. Fyfe, J. N. S. Cole, L. Nazarenko, K. E. Taylor, F. J. Wentz, Volcanic contribution to decadal changes in tropospheric temperature. *Nat. Geosci.* **7**, 185–189 (2014).
- G. A. Schmidt, D. T. Shindell, K. Tsigaridis, Reconciling warming trends. *Nat. Geosci.* **7**, 158–160 (2014).
- K. E. Trenberth, J. T. Fasullo, An apparent hiatus in global warming? *Earth's Future* **1**, 19–32 (2013).
- C. D. Roberts, M. D. Palmer, D. McNeall, M. Collins, Quantifying the likelihood of a continued hiatus in global warming. *Nat. Clim. Chang.* **5**, 337–342 (2015).
- T. R. Karl, A. Arguez, B. Huang, J. H. Lawrimore, J. R. McMahon, M. J. Menne, T. C. Peterson, R. S. Vose, H.-M. Zhang, Possible artifacts of data biases in the recent global surface warming hiatus. *Science* **348**, 1469–1472 (2015).
- 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, N. C. Swart, Making sense of the early-2000s warming slowdown. *Nat. Clim. Chang.* **6**, 224–228 (2016).
- A. Robock, Volcanic eruptions and climate. *Rev. Geophys.* **38**, 191–219 (2000).
- C. P. Morice, J. J. Kennedy, N. A. Rayner, P. D. Jones, Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 dataset. *J. Geophys. Res. Atmos.* **117**, D08181 (2012).
- P. Jones, The reliability of global and hemispheric temperature records. *Adv. Atmos. Sci.* **33**, 269–282 (2016).
- N. L. Bindoff, P. A. Stott, K. M. AchutaRao, M. R. Allen, N. Gillett, D. Gutzler, K. Hansingo, G. Hegerl, Y. Hu, S. Jain, I. I. Mokhov, J. Overland, J. Perlitz, R. Sebbani, X. Zhang, Detection and attribution of climate change: From global to regional, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, T. F. Stocker, D. Qin, G.-K. Plattner, M. M. B. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, P. M. Midgley, Eds. (Cambridge Univ. Press, 2013).
- J. Imbers, A. Lopez, C. Huntingford, M. R. Allen, Testing the robustness of the anthropogenic climate change detection statements using different empirical models. *J. Geophys. Res. Atmos.* **118**, 3192–3199 (2013).
- C. K. Folland, A. W. Colman, D. M. Smith, O. Boucher, D. E. Parker, J.-P. Vernier, High predictive skill of global surface temperature a year ahead. *Geophys. Res. Lett.* **40**, 761–767 (2013).
- M. Lockwood, Recent changes in solar outputs and the global mean surface temperature. III. Analysis of contributions to global mean air surface temperature rise. *Proc. R. Soc. A* **464**, 1387–1404 (2008).
- J. L. Lean, D. H. Rind, How natural and anthropogenic influences alter global and regional surface temperatures: 1889 to 2006. *Geophys. Res. Lett.* **35**, L18701 (2008).
- M. E. Schlesinger, N. Ramankutty, An oscillation in the global climate system of period 65–70 years. *Nature* **367**, 723–726 (1994).
- M. Wei, F. Qiao, J. Deng, A quantitative definition of global warming hiatus and 50-year prediction of global-mean surface temperature. *J. Atmos. Sci.* **72**, 3281–3289 (2015).
- K.-K. Tung, J. Zhou, Using data to attribute episodes of warming and cooling in instrumental records. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 2058–2063 (2013).
- P. A. Stott, S. F. B. Tett, G. S. Jones, M. R. Allen, J. F. B. Mitchell, G. J. Jenkins, External control of 20th century temperature by natural and anthropogenic forcings. *Science* **15**, 2133–2137 (2000).
- J. R. Knight, The Atlantic multidecadal oscillation inferred from the forced climate response in coupled general circulation models. *J. Clim.* **22**, 1610–1625 (2009).
- M. Prather, G. Flato, P. Friedlingstein, C. Jones, J.-F. Lamarque, H. Liao, P. Rasch, IPCC 2013: Annex II: Climate system scenario tables, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, T. F. Stocker, D. Qin, G.-K. Plattner, M. M. B. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, P. M. Midgley, Eds. (Cambridge Univ. Press, 2013).
- K. Matthes, B. Funke, M. E. Andersson, L. Barnard, J. Beer, P. Charbonneau, M. A. Chilvers, T. D. de Wit, M. Haberreiter, A. Hendry, C. H. Jackman, M. Kretschmar, T. Kruschke, M. Kunze, U. Langematz, D. R. Marsh, A. C. Maycock, S. Misios, C. J. Rodger, A. A. Scaife, A. Seppälä, M. Shangguan, M. Sinnhuber, K. Tourpali, I. G. Usoskin, M. van de Kamp, P. T. Verronen, S. Versick, Solar forcing for CMIP6 (v 3.2). *Geosci. Model Dev.* **10**, 2247–2302 (2017).
- J. Lean, D. Rind, Climate forcing by changing solar radiation. *J. Clim.* **11**, 3069–3094 (1998).
- J. R. Knight, R. J. Allan, C. K. Folland, M. Vellinga, M. E. Mann, A signature of persistent natural thermohaline circulation cycles in observed climate. *Geophys. Res. Lett.* **32**, L20708 (2005).
- V. A. Semenov, M. Latif, D. Dommenget, N. S. Keenlyside, A. Strehz, T. Martin, W. Park, The impact of North Atlantic–Arctic multidecadal variability on Northern Hemisphere surface air temperature. *J. Clim.* **23**, 5668–5677 (2010).
- A. Dai, J. C. Fyfe, S.-P. Xie, X. Dai, Decadal modulation of global surface temperature by internal climate variability. *Nat. Clim. Chang.* **5**, 555–559 (2015).
- D. Parker, C. Folland, A. Scaife, J. Knight, A. Colman, P. Baines, B. Dong, Decadal to multidecadal climate variability and the climate change background. *J. Geophys. Res.* **112**, D18115 (2007).
- J. L. Cohen, J. C. Furtado, M. Barlow, V. A. Alexeev, J. E. Cherry, Asymmetric seasonal temperature trends. *Geophys. Res. Lett.* **39**, L04705 (2012).
- D. W. J. Thompson, J. J. Kennedy, J. M. Wallace, P. D. Jones, A large discontinuity in the mid-twentieth century in observed global-mean surface temperature. *Nature* **453**, 646–649 (2008).
- J. Hansen, R. Ruedy, M. Sato, K. Lo, Global surface temperature change. *Rev. Geophys.* **48**, RG4004 (2010).
- B. J. Henley, J. Gergis, D. J. Karoly, S. Power, J. Kennedy, C. K. Folland, A Tripole Index for the Interdecadal Pacific Oscillation. *Clim. Dyn.* **45**, 3077–3090 (2015).
- R. Zhang, T. L. Delworth, I. M. Held, Can the Atlantic Ocean drive the observed multidecadal variability in Northern Hemisphere mean temperature? *Geophys. Res. Lett.* **34**, L02709 (2007).
- D. M. Thompson, J. E. Cole, G. T. Shen, A. W. Tudhope, G. A. Meehl, Early twentieth-century warming linked to tropical Pacific wind strength. *Nat. Geosci.* **8**, 117–121 (2015).
- N. A. Rayner, D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, A. Kaplan, Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.* **108**, 4407 (2003).
- C. K. Folland, D. E. Parker, Correction of instrumental biases in historical sea surface temperature data. *Q. J. R. Meteorol. Soc.* **121**, 319–367 (1995).
- D. E. Parker, H. Wilson, P. D. Jones, J. R. Christy, C. K. Folland, The impact of Mount Pinatubo on world-wide temperatures. *Int. J. Climatol.* **16**, 487–497 (1996).
- Y. Kosaka, S.-P. Xie, Recent global-warming hiatus tied to equatorial Pacific surface cooling. *Nature* **501**, 403–407 (2013).
- M. H. England, S. McGregor, J. P. Spence, G. A. Meehl, A. Timmermann, W. Cai, A. S. Gupta, M. J. McPhaden, A. Purich, A. Santoso, Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus. *Nat. Clim. Chang.* **4**, 222–227 (2014).
- J. Knight, J. J. Kennedy, C. Folland, G. Harris, G. S. Jones, M. Palmer, D. Parker, A. Scaife, P. Stott, Do global temperature trends over the last decade falsify climate predictions? In: State of the Climate in 2008. *Bull. Am. Meteorol. Soc.* **90**, S22–S23 (2009).
- K. Cowtan, R. G. Way, Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Q. J. R. Meteorol. Soc.* **140**, 1935–1944 (2014).
- J. Huang, X. Zhang, Q. Zhang, Y. Lin, M. Hao, Y. Luo, Z. Zhao, Y. Yao, X. Chen, L. Wang, S. Nie, Y. Yin, Y. Xu, J. Zhang, Recently amplified arctic warming has contributed to a continual global warming trend. *Nat. Clim. Chang.* **7**, 875–879 (2017).
- A. J. Simmons, P. Berrisford, D. P. Dee, H. Hersbach, S. Hitahara, J.-N. Thépaut, A reassessment of temperature variations and trends from global reanalyses and monthly surface climatological datasets. *Q. J. R. Meteorol. Soc.* **143**, 101–119 (2017).
- D. P. Dee, S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hölm, L. Isaksen, P. Källberg, M. Köhler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut, F. Vitart, The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**, 553–597 (2011).

43. D. M. Smith, B. B. B. Booth, N. J. Dunstone, R. Eade, L. Hermanson, G. S. Jones, A. A. Scaife, K. L. Sheen, V. Thompson, Role of volcanic and anthropogenic aerosols in the recent global surface warming slowdown. *Nat. Clim. Chang.* **6**, 936–940 (2016).
44. B. D. Santer, J. C. Fyfe, G. Pallotta, G. M. Flato, G. A. Meehl, M. H. England, E. Hawkins, M. E. Mann, J. F. Painter, C. Bonfill, I. Cvijanovic, C. Mears, F. J. Wentz, S. Po-Chedley, Q. Fu, C.-Z. Zou, Causes of differences in model and satellite tropospheric warming rates. *Nat. Geosci.* **10**, 478–485 (2017).
45. J. J. Kennedy, A review of uncertainty in in situ measurements and data sets of sea surface temperature. *Rev. Geophys.* **52**, 1–32 (2013).
46. S. Misios, D. M. Mitchell, L. J. Gray, K. Tourpali, K. Matthes, L. Hood, H. Schmidt, G. Chiodo, R. Thiéblemont, E. Rozanov, A. A. Krivolutsky, Solar signals in CMIP-5 simulations: Effects of atmosphere–ocean coupling. *Q. J. R. Meteorol. Soc.* **142**, 928–941 (2016).
47. Z. Zeng, S. Piao, L. Z. X. Li, L. Zhou, P. Ciais, T. Wang, Y. Li, X. Lian, E. F. Wood, P. Friedlingstein, J. Mao, L. D. Estes, R. B. Myneni, S. Peng, X. Shi, S. I. Seneviratne, Y. Wang, Climate mitigation from vegetation biophysical feedbacks during the past three decades. *Nat. Clim. Chang.* **7**, 432–436 (2017).
48. R. Allan, P. Brohan, G. P. Compo, R. Stone, J. Luterbacher, S. Brönnimann, The international Atmospheric Circulation Reconstructions over the Earth (ACRE) initiative. *Bull. Am. Meteorol. Soc.* **92**, 1421–1425 (2011).
49. K. E. Trenberth, J. M. Caron, D. P. Stepaniak, S. Worley, Evolution of El Niño–Southern Oscillation and global atmospheric surface temperatures. *J. Geophys. Res.* **107**, 4065 (2002).
50. H. A. Titchner, N. A. Rayner, The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations. *J. Geophys. Res. Atmos.* **119**, 2864–2889 (2014).
51. L. W. Thomason, N. Ernest, L. Millán, L. Rieger, A. Bourassa, J.-P. Vernier, G. Manney, B. Luo, F. Arfeuille, T. Peter, A global, space-based stratospheric aerosol climatology: 1979–2016. *Earth Syst. Sci. Data* **10**, 469–492 (2018).
52. G. P. Compo, J. S. Whitaker, P. D. Sardeshmukh, N. Matsui, R. J. Allan, X. Yin, B. E. Gleason Jr., R. S. Vose, G. Rutledge, P. Bessemoulin, S. Brönnimann, M. Brunet, R. I. Crouthamel, A. N. Grant, P. Y. Groisman, P. D. Jones, M. C. Kruk, A. C. Kruger, G. J. Marshall, M. Maugeri, H. Y. Mok, Ø. Nordli, T. F. Ross, R. M. Trigo, X. L. Wang, S. D. Woodruff, S. J. Worley, The twentieth century reanalysis project. *Q. J. R. Meteorol. Soc.* **137**, 1–28 (2011).
53. T. M. L. Wigley, S. C. B. Raper, Climatic change due to solar irradiance changes. *Geophys. Res. Lett.* **17**, 2169–2172 (1990).
54. O. Boucher, J. A. Lowe, C. D. Jones, Implications of delayed actions in addressing carbon dioxide emission reduction in the context of geo-engineering. *Clim. Chang.* **92**, 261–273 (2009).
55. M. S. Bartlett, Some aspects of the time-correlation problem in regard to tests of significance. *J. R. Stat. Soc. B* **98**, 536–543 (1935).
56. P. J. Diggle, K-Y. Liang, S. L. Zeger, *Analysis of Longitudinal Data* (Oxford Univ. Press, 1999).
57. IPCC 2001, *Climate Change 2001: Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, J. T. Houghton, Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell, C. A. Johnson, Eds. (Cambridge Univ. Press, 2001).
58. J. R. Lanzante, Resistant, robust and non-parametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station data. *Int. J. Climatol.* **16**, 1197–1226 (1996).
59. H. von Storch, F. W. Zwiers, *Statistical Analysis in Climate Research* (Cambridge Univ. Press, 1999).

Acknowledgments: We thank I. Macadam for the original software and document on the REML method. This paper is a contribution to the CLIVAR International Climate of the Twentieth Century Plus project. <http://cola.gmu.edu/c20c/>. **Funding:** C.K.F. and A.C. were funded by the Joint UK Department for Business, Energy and Industrial Strategy/Defra Met Office Hadley Centre Climate Programme (GA01101). O.B. was funded by the CNRS. D.E.P., formerly funded as for C.K.F. and A.C., is now an independent researcher. **Author contributions:** C.K.F. carried out the bulk of the work, including calculations, and wrote the drafts of the paper. O.B. helped with the development of the exponential response times, ideas about the recalcitrant response, and the presentation of the paper, while A.C. carried out the map-based calculations and advised on some of the diagrams. D.E.P. advised on the presentation of results and on modifications to and interpretation of the REML software and made some revisions of the paper. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The monthly transformed, normalized forcing data and the observed GST are available on the CLIVAR [Climate and Ocean: Variability, Predictability and Change (one of the four core projects of the World Climate Research Programme)] Climate of the Twentieth Century Plus web site at <http://cola.gmu.edu/c20c/> through the sidebar “Experiments.” The full HadISST2.1.0.0 data set is available from the corresponding author. All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 31 July 2017

Accepted 2 May 2018

Published 6 June 2018

10.1126/sciadv.ao5297

Citation: C. K. Folland, O. Boucher, A. Colman, D. E. Parker, Causes of irregularities in trends of global mean surface temperature since the late 19th century. *Sci. Adv.* **4**, eaao5297 (2018).