**Article:**

Marotzke, J and Forster, PM orcid.org/0000-0002-6078-0171 (2015) Forcing, feedback and internal variability in global temperature trends. Nature, 517 (7536). pp. 565-570. ISSN 0028-0836

https://doi.org/10.1038/nature14117

# Forcing, feedback, and internal variability in global temperature trends

*Jochem Marotzke, Max Planck Institute for Meteorology, Bundesstrasse 53, 20146 Hamburg, Germany*; jochem.marotzke@mpimet.mpg.de; *phone: +49-40-41173-440; fax: +49-40-41173-366 (corresponding author)*

*Piers M. Forster, School of Earth and Environment, University of Leeds; p.m.forster@leeds.ac.uk*

## Summary

Most current-generation climate models simulate an increase in global mean surface temperature (GMST) since 1998 while observations suggest a warming hiatus. It is still unclear to what extent this mismatch is caused by incorrect model forcing, by incorrect model response to forcing, or by random factors. To place the hiatus in context, we analyse simulations and observations of GMST from 1900 to 2012 and show that the distribution of simulated 15-year trends shows no systematic bias against the observations. Using a multiple regression approach that is physically motivated by surface energy balance, we isolate the impact of radiative forcing, climate feedback, and ocean heat uptake on GMST – with the residual interpreted as internal variability – and assess all possible 15- and 62-year trends. The differences between simulated and observed trends are dominated by (1) random internal variability at 15 years and (2) variations in the radiative forcings used to drive models at 62 years. For either trend length, spread in simulated climate feedback leaves no traceable imprint on GMST trends and thus on the difference between simulations and observations. The claim that climate models systematically overestimate the response to radiative forcing from increasing greenhouse-gas concentrations therefore appears to be unfounded.

## Introduction

The global-mean surface temperature (GMST) has risen in the past fifteen years at a rate that is only one-third to one-half of the average over the second half of the 20th century (e.g., refs. 1-5). This hiatus is not reproduced in most simulations with current-generation climate models, which instead over the period 1998 to 2012 show a larger GMST trend than observed[5-14]. The difference between GMST observations and simulations is caused in part by quasi-random internal climate variability[5-10,13,14], which arises because of chaotic processes in the climate system. But part of the difference is likely caused by errors in the model radiative forcing[5,12,14-16] or in the model response to radiative forcing[5,14,17,18]. The relative magnitudes of these three contributions are poorly known. Here we quantify how forcing, feedback, and internal climate variability contribute to spread in simulated historical GMST trends and hence to the differences between models and observations.

We use a three-pronged approach. First, we note that due to quasi-random internal climate variability, the difference between observed and simulated trends likewise contains quasi-random contributions. To avoid focusing too strongly on the particular period 1998 to 2012 – which contains some climate extremes relevant for GMST[19-21] and is hence unlikely to be reproduced in a simulation containing quasi-random contributions – we analyse GMST trends of a certain length for the entire period 1900 to 2012 (see ref. 13). Second, we quantify the contributions of forcing, climate feedback, ocean heat uptake, and internal variability to simulated GMST trends, through a multiple linear regression approach that is physically motivated by the global surface energy balance. And third, we investigate trends over both 15 and 62 years, representing decadal and multi-decadal timescales, respectively. We combine these three aspects into a new unified conceptual framework, which allows us to put the GMST trends over the 15-year period 1998 to 2012 into appropriate context.

We first create linear trends from an ordinary-least-squares fit and perform all statistical analyses on these trends. This procedure implies that the analysis must be repeated for each trend length, in contrast to previous work aiming at attributing elements in the observed GMST time series itself. Such elements include effects of volcanic eruptions, solar variability, anthropogenic forcing, El Niño events, and atmospheric dynamic variability including land-sea contrasts[13,14,22-25]. Because the amplitude of internal variability decreases with increasing trend length[3,26], we expect a cleaner breakdown into the individual contributions from forcing, feedback, and internal variability if we focus on one trend length at a time. We analyse trends over 15 and 62 years, because these were the trend lengths primarily considered in the Intergovernmental Panel on Climate Change Assessment Report 5 (IPCC AR5, ref. 5).

## Observed and simulated 15-year trends

To gauge whether the difference between simulations and observations is unusual over the hiatus period, we first compare observed and simulated 15-year trends over the entire period from 1900 to 2012 (Fig. 1, see also ref. 13). We use the HadCRUT4 observational data set[27] and the "historical" simulations conducted under the auspices of the Coupled Model Intercomparison Project Phase 5 (CMIP5, ref. 28), extended for the years 2006 to 2012 with the RCP4.5 scenario runs (Extended Data Fig. 1, Extended

Data Table 1). The simulation output is subsampled using the HadCRUT4 data mask[11], to account for the effects of incomplete observational coverage[29,30].

Figure 1a contains the joint relative frequency distribution of 15-year GMST trends across the 114 available CMIP5 simulations, as a function of start years since 1900 and trend size. Compared to the CMIP5 ensemble, observed trends are distributed in no discernibly preferred way and occur sometimes at the upper end of the ensemble (e.g., start year 1927, best-estimate observed trend larger than 110 of the 114 simulated trends, Fig. 1b) and sometimes at the lower end of the ensemble (e.g., start year 1998, best-estimate observed trend smaller than all 114 simulated trends, Fig. 1c)[5,13,26].

In both cases depicted in Figs. 1b or 1c, fewer than 5% of the simulations lie in one of the tails relative to the observed trend. Hence, if a 5% criterion for statistical significance is used, one would diagnose formal model–observation inconsistency for 15-year trends with start years in 1927 and 1998 (ref. 11). But when the comparison is repeated for all start years, the rank that the observed trend would have as a member of the ensemble of simulated trends[31] shows no apparent bias (Fig. 1e), indicating that the observed and simulated distributions of 15-year trends are broadly consistent with each other. Any position of the observed trend within the ensemble of simulated trends – including a position at or near the margin – is thus dominated by quasi-random effects (although for any particular start year, a non-negligible contribution from systematic errors cannot be excluded).

The marginal distribution of simulated GMST trends as a function of trend size is wider than the observed distribution of trends (Fig. 1d), a finding consistent with that from the previous generation of climate models[32]. The width is exaggerated owing to contributions arising at three distinct periods. Some simulated trends with start years from around 1950 to 1960 are more strongly negative than any observed trends since 1900, and some simulated trends with start years from around 1960 to 1970 and from around 1985 to 1998 are more strongly positive than any observed trends since 1900 (Fig. 1a). All three periods (1950 to 1960, 1960 to 1970, 1985 to 1998) are influenced by volcanic eruptions (Agung in 1963 and Pinatubo in 1991). We speculate that some, though not all, models overestimate the cooling induced by an eruption and the subsequent warming recovery (see, e.g., ref. 12 concerning a confounding role of El Niño).

The mean over all simulated 15-year trends during the period 1900 to 2012 is at (0.086±0.001) °C per decade (mean±s.e.m.; n=11,186) in excellent agreement with the observed (0.088±0.01) °C per decade (mean±s.e.m.; n=99). Furthermore, of all 11,186 pairwise comparisons that are possible between simulated and observed trends, the observed trend is higher in 53.6% of all cases, slightly above the break-even point. Figure 1 demonstrates that when viewed over the entire period 1900 to 2012, the 15-year GMST trends simulated by the CMIP5 ensemble show no systematic deviation from the observations.

Our interpretation of Fig. 1 tacitly assumes that simulated multi-model-ensemble spread accurately characterises internal variability, an assumption shared with other interpretations of the position of observed relative to simulated trends (e.g., the reduction in Arctic summer sea ice[5,33,34]). We now test the validity of this assumption, by identifying deterministic and quasi-random causes of ensemble spread. We exploit the availability of a large number of simulations – 114 realisations with 36 different models, with forcing information available for 75 realisations with 18 different models[35] (Extended Data Figs. 1 and 2 and Extended Data Table 1) – and investigate the contributions of radiative forcing, climate feedback, and ocean heat uptake to all simulated 15-year and 62-year GMST trends during the period 1900 to 2012.

## Energy balance and multiple regression

Our starting point is the globally averaged energy balance for the surface layer[35-37]. An increasing trend ΔF in effective radiative forcing (ERF) causes an increasing trend ΔT in GMST. This in turn leads to increased outgoing radiation, which in linearised form is written as αΔT, where α is the climate feedback parameter. Furthermore, the GMST increase leads to increased heat transfer from the surface layer to the subsurface ocean, written again in linearised form as κΔT, where κ is the ocean heat uptake efficiency. The thermal adjustment of the surface layer to ΔF is expected to occur within a few years[35-37]. This means that for timescales of one to several decades, the surface energy balance is in quasi-steady state and reads

$$(\alpha+\kappa)\Delta T=\Delta F\ , \tag{1}$$

which produces the energy-balance "prediction" for the GMST trend

$$\Delta T=\Delta F/(\alpha+\kappa). \tag{2}$$

Each CMIP5 model simulates its own ERF time series over the historical period. These time series were diagnosed previously[35]; if multiple realisations were available for a model, the ensemble average of the individual diagnosed ERF time series for this model was given[35] and is used here. The individual $\alpha$ and $\kappa$ were previously determined for each CMIP5 model from a regression of global top-of-atmosphere energy imbalance against GMST[5,35,38-41], in turn based on simulations in which the $CO_2$ concentration was quadrupled abruptly. The ranges of $\alpha$ and $\kappa$ are (0.6–1.8) and (0.45–1.52) $Wm^{-2}(°C)^{-1}$, respectively. That $\alpha$ and $\kappa$ in the CMIP5 models might vary with time and climate state[42,43] is ignored here. There is some positive though not statistically significant correlation between $\alpha$ and $\kappa$ (across the 75-member sub-ensemble, correlation is 0.17, $p = 0.14$).

Each model's $\alpha$ is related to its equilibrium climate sensitivity ECS by

$$ECS = F_{2x}/\alpha\,, \tag{3}$$

where $F_{2x}$ is the effective radiative forcing from a doubling of the pre-industrial atmospheric $CO_2$ concentration. The reference value for $F_{2x}$ is 3.7 $Wm^{-2}$ (e.g., ref. 44), but $F_{2x}$ varies between 2.6 $Wm^{-2}$ and 4.3 $Wm^{-2}$ across the CMIP5 ensemble[5,38]. In order not to confound model-response uncertainty with uncertainty from $CO_2$ forcing, we use $\alpha$ and not ECS to characterise model response.

Based on the physical foundation of energy balance (2), we determine the extent to which the *across-ensemble variations* of $\Delta F$, $\alpha$, and $\kappa$ contribute to the ensemble spread of GMST trends $\Delta T$, using the 75-member sub-ensemble of CMIP5 historical simulations for which radiative-forcing information can be obtained from the CMIP5 archive[35] (see Extended Data Table 1). The presence of internal variability is included in our framework by adding a random term to (2), so that our equation is

$$\Delta T = \Delta F/(\alpha+\kappa)+\varepsilon\,. \tag{4}$$

Because (4) assumes an increasing trend in ERF, its validity is somewhat questionable following a volcanic eruption (e.g., ref. 25). On the other hand, Extended Data Figure 3 shows that overall we see a reliable relationship between ERF and GMST trends in the CMIP5 ensemble, even if the ERF trend is negative.

We make the connection to multiple linear regression by writing each quantity as

$$x = \bar{x} + x', \tag{5}$$

where the overbar marks the ensemble average and the prime the across-ensemble variation. Linear expansion of (4) thus produces

$$\Delta\bar{T} + \Delta T' = \frac{\Delta\bar{F}}{\bar{\alpha}+\bar{\kappa}} + \frac{1}{\bar{\alpha}+\bar{\kappa}}\Delta F' - \frac{\Delta\bar{F}}{\left(\bar{\alpha}+\bar{\kappa}\right)^2}\alpha' - \frac{\Delta\bar{F}}{\left(\bar{\alpha}+\bar{\kappa}\right)^2}\kappa' + \varepsilon . \tag{6}$$

This equation holds for each start year separately and suggests the regression model

$$\Delta T'_j = \beta_0 + \beta_1 \Delta F'_j + \beta_2 \alpha'_j + \beta_3 \kappa'_j + \varepsilon_j ; \quad j = 1,...,75 . \tag{7}$$

We thus perform for each start year a multiple linear regression of $\Delta T'$ against $\Delta F'$, $\alpha'$, and $\kappa'$. The regression residual $\varepsilon$ is interpreted as the contribution from internal variability. The complete regression-based prediction for GMST trend is obtained by adding the ensemble-mean trend to the regression for the across-ensemble variations:

$$\Delta\hat{T}_{reg,j} = \Delta\bar{T} + \hat{\beta}_0 + \hat{\beta}_1 \Delta F'_j + \hat{\beta}_2 \alpha'_j + \hat{\beta}_3 \kappa'_j ; \quad j = 1,...,75 , \tag{8}$$

where the caret marks the regression estimate. Note that for a model that has multiple realisations, the same $\Delta F'_j$, $\alpha'_j$ or $\kappa'_j$ is counted multiple times. The regression is performed separately for each period length over which trends are computed. We will interpret the ensemble spread of the regression result $\Delta\hat{T}_{reg,j}$, $j = 1,...,75$, as the deterministic spread and the spread $\hat{\varepsilon}_j$, $j = 1,...,75$, of the residuals as the quasi-random spread.

## Deterministic vs. quasi-random spread

For 15-year GMST trends, deterministic across-ensemble variations are smaller than internal variability, as shown by the comparison of the regression-based ensemble spread with the regression residuals (Figs. 2b and c, respectively). The regression result shows substantial time-dependence in ensemble spread only for 15-year periods influenced by major volcanic eruptions, in particular the Agung eruption in 1963 (Fig. 2b; the deterministic ensemble spread is particularly large in these periods, see Extended Data Fig. 4a). The distribution of residuals shows little time-dependence, as witnessed by spread that is similar for all start years (Figs. 2c–f). The generally weak time-dependence of the spread suggests that we can estimate the magnitudes of deterministic spread and internal variability from the marginal distributions obtained by time-averaging the distributions shown in Figs. 2b and 2c, respectively. The 5–95% range is 0.11 °C per decade for the regression result and 0.26 °C per decade for the residuals; internal variability thus dominates deterministic spread by a factor of two-and-a-half. The dominance of internal variability in the ensemble spread of the 15-year GMST trends indicates that, viewed over the entire period 1900 to 2012, no systematic model error needs to be invoked when trying to explain differences between simulated and observed trends. In particular, the GMST spread due to feedback $\alpha$ is not systematically larger than spread from either ERF trend or ocean heat uptake efficiency and is much smaller than internal variability (Extended Data Fig. 4 and Fig. 2; see also ref. 12).

For any given start year, the residual spread is very similar to the full ensemble spread, implying that we can indeed use the ensemble spread as a measure of internal variability (compare Figs. 1b and c to Figs. 2d and e). Furthermore, identifying the ensemble spread of the regression residuals with internal variability allows us to characterise the component of observational uncertainty that arises from internal variability (Figs. 2a and f). This uncertainty does not concern the construction of the global average from individual station data (which has much smaller uncertainty[5]) but relates to the question of whether an observed trend is statistically significant (detectable) given serial correlation arising from internal variability[18]. Our model-based estimate of 0.26 °C per decade for the 5–95% confidence interval for observed 15-year GMST trends is slightly larger than the AR5 serial-correlation-based estimate for the uncertainty of the observed GMST trend over the hiatus period (0.2 °C per decade, see ref. 4). We deem this an acceptable agreement given that the estimates were obtained through completely different approaches. We further note that the CMIP5 ensemble has been assessed to be generally consistent with observed historical decadal variability in GMST[5], although on average it overestimates somewhat the global variability in the lower troposphere[45].

For most of the historical period, the entire ensemble of regression-based simulated 15-year GMST trends lies within the model-estimated 5–95% confidence interval of the observations (Fig. 2a). The regression-based simulated ensemble partially falls outside this interval during the cooling following the Agung eruption and the subsequent warming recovery, as well as for start dates after 1990, which include the warming recovery following the Pinatubo eruption and the surface-warming hiatus (Fig. 2a). Because the phases of volcanically driven cooling and subsequent warming coincide with larger regression spread due to ERF trend (Extended Data Fig. 4), we speculate that the implementation of volcanic forcing requires improvement in some climate models.

The ensemble spread of 62-year GMST trends is dominated by internal variability for start years early in the 20$^{th}$ century, but for start years from 1910 onward, the deterministic spread increases and dominates for start years 1920 and later (Fig. 3). The 5–95% range of the regression residuals is 0.059 °C per decade, compared to a deterministic range of 0.032 °C per decade for start year 1900 and 0.093 °C per decade for start year 1951. The 5–95% deterministic range for all 62-year trends is 0.081 °C per decade, which is larger by one-third than the 5–95% range from internal variability. Nevertheless, we see a substantial influence of internal variability even for GMST trends over 62 years.

When observational uncertainty is accounted for – based again on the 5–95% confidence interval derived from quasi-random model spread – the ensemble-mean simulated 62-year GMST trend is consistent with the observed trend for all start years after around 1915; before that, the simulations tend to warm too little (Fig. 3a). After around 1945, the ensemble-mean simulated 62-year trend lies above the observed trend, although their difference is smaller than the range of internal variability. From around 1925 onward, both the largest and the smallest individual regression-based simulated trends lie outside the range defined by observations plus internal variability and would hence be judged to be inconsistent with observations (Fig. 3a).

The cause of this inconsistency can be traced almost entirely to the contribution to the regression by the ERF trend (Fig. 3). By contrast, the magnitude of the contributions by $\alpha$ and $\kappa$ is around 0.01 °C per decade or less for all start years (Figs. 3e and f). The deterministic ensemble spread in 62-year GMST trend is hence dominated by the spread in ERF throughout the 20$^{th}$ century (Fig. 3).

## Discussion

Viewed over the entire period since 1900, the differences between simulated and observed 15-year trends in GMST are dominated by internal variability and hence arise largely by coincidence, with a minor contribution from volcanic forcing that is sometimes too strong in some models (Fig. 2). Furthermore, we confirm and extend to all 15-year radiative-forcing trends since 1900, the AR5 assessment for the hiatus period[5] that the CMIP5 models show little systematic bias when comparing against the AR5 best-estimate radiative-forcing trend[46] – despite the substantial scatter about the ensemble mean (Extended Data Fig. 2).

The generally dominant role of internal variability in shaping simulated 15-year GMST trends implies that internal variability also dominates the difference between simulations and observations during the hiatus period. This conclusion sharpens considerably the relative roles of internal variability, forcing error, and response error, compared to the corresponding AR5 assessment[5]. While there is no obvious contribution of forcing bias in the CMIP5 models (Extended Data Fig. 2), the diagnosed radiative forcing is uncertain[35]. Hence our analysis cannot rule out a small contribution from a systematic forcing bias[12,15,16,46-48] in the models. In particular, volcanic forcing is estimated to contribute to the difference between simulations and observations by up to 15% over 1998 to 2012 in ref. 12, with large uncertainty in the magnitude, a contribution that our method cannot detect. Furthermore, the period 1998 to 2012 stands out as the only one during which the HadCRUT4 15-year GMST trend falls entirely outside the CMIP5 ensemble (if only narrowly), suggesting that the CMIP5 models could be missing a cooling contribution from the radiative forcing during the hiatus period[12,15,16,46-48], or that there has been an unusual enhancement of ocean heat uptake not simulated by any model[19].

For 62-year GMST trends since 1900, the difference between simulations and observations is dominated by the spread in radiative-forcing trend in the models, with a smaller yet substantial influence of internal variability (Fig. 3). Our simple regression-based estimate of internal variability in 62-year GMST trends corresponds to a 17–83% range of ±0.11 °C for the temperature change over six decades, which is in excellent agreement with the value of ±0.10 °C that has been found for the period 1951 to 2010 using much more sophisticated formal methods of detection and attribution[18].

There is scientific, political, and public debate regarding the question of whether the GMST difference between simulations and observations during the hiatus period might be a sign of an equilibrium model response to a given radiative forcing that is systematically too strong, or equivalently, of a simulated climate feedback $\alpha$ that is systematically too small (cf., (3)). By contrast, we find no substantive physical or statistical connection between simulated climate feedback and simulated GMST trends over the hiatus or any other period, either for 15- or for 62-year trends (Figs. 2 and 3, Extended Data Fig. 4). The role of simulated climate feedback in explaining the difference between simulations and observations is hence minor or even negligible. By implication, the comparison of simulated and observed GMST trends does not permit inference about which magnitude of simulated climate feedback – ranging from 0.6 to 1.8 $Wm^{-2}(°C)^{-1}$ in the CMIP5 ensemble – better fits the observations. Because observed GMST trends do not allow us to distinguish between simulated climate feedback that varies by a factor of three, the claim that climate models systematically overestimate the GMST response to radiative forcing from increasing greenhouse-gas concentrations appears to be unfounded.

## References


1. Knight, J. *et al.* Do global temperature trends over the last decade falsify climate predictions? [In: State of the Climate in 2008]. *Bull. Am. Met. Soc.* **90**, S22-S23 (2009).
2. Wang, S. *et al.* Does the global warming pause in the last decade: 1999-2008? *Adv. Clim. Change Res.* **1**, 49-54 (2010).
3. Liebmann, B., Dole, R. M., Jones, C., Blade, I. & Allured, D. Influence of choice of time period on global surface temperature trend estimates. *Bull. Am. Met. Soc.* **91**, 1485-1491 (2010).
4. Hartmann, D. L. *et al.* 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* (eds T. F. Stocker *et al.*) 159-254 (Cambridge University Press, 2013).
5. Flato, G. *et al.* 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* (eds T. F. Stocker *et al.*) 741-866 (Cambridge University Press, 2013).
6. Meehl, G. A., Arblaster, J. M., Fasullo, J. T., Hu, A. & Trenberth, K. E. Model-based evidence of deep-ocean heat uptake during surface-temperature hiatus periods. *Nature Clim. Change* **1**, 360-364 (2011).
7. Meehl, G. A. & Teng, H. Case studies for initialized decadal hindcasts and predictions for the Pacific region. *Geophys. Res. Lett.* **39**, L22705, doi: 22710.21029/22012GL053423 (2012).

8. Meehl, G. A., Hu, A., Arblaster, J. M., Fasullo, J. & Trenberth, K. E. Externally Forced and Internally Generated Decadal Climate Variability Associated with the Interdecadal Pacific Oscillation. *J. Climate* **26**, 7298-7310 (2013).

9. Doblas-Reyes, F. J. *et al.* Initialized near-term regional climate change prediction. *Nature Comm.* **4**, 1715, doi:1710.1038/ncomms2704 (2013).

10. Guemas, V., Doblas-Reyes, F. J., Andreu-Burillo, I. & Asif, M. Retrospective prediction of the global warming slowdown in the past decade. *Nature Clim. Change* **3**, 649-653 (2013).

11. Fyfe, J. C., Gillett, N. P. & Zwiers, F. W. Overestimated global warming over the past 20 years. *Nature Clim. Change* **3**, 767-769 (2013).

12. Santer, B. D. *et al.* Volcanic contribution to decadal changes in tropospheric temperature. *Nature Geosci.* **7**, 185-189 (2014).

13. Risbey, J. S. *et al.* Well-estimated global surface warming in climate projections selected for ENSO phase. *Nature Clim. Change* **4**, 835-840 (2014).

14. Huber, M. & Knutti, R. Natural variability, radiative forcing and climate response in the recent hiatus reconciled. *Nature Geosci.* **7**, 651-656 (2014).

15. Solomon, S. *et al.* The persistently variable "background" stratospheric aerosol layer and global climate change. *Science* **333**, 866-870 (2011).

16. Schmidt, G. A., Shindell, D. T. & Tsigaridis, K. Reconciling warming trends. *Nature Geosci.* **7**, 158-160 (2014).

17. Stott, P., Good, P., Jones, G., Gillett, N. & Hawkins, E. The upper end of climate model temperature projections is inconsistent with past warming. *Env. Res. Let.* **8** (2013).

18. Bindoff, N. L. *et al.* 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* (eds T. F. Stocker *et al.*) 867-952 (Cambridge University Press, 2013).

19. England, M. H. *et al.* Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus. *Nature Clim. Change* **4**, 222-227 (2014).

20. Cohen, J. L., Furtado, J. C., Barlow, M., Alexeev, V. A. & Cherry, J. E. Asymmetric seasonal temperature trends. *Geophys. Res. Lett.* **39**, L04705, doi:04710.01029/02011GL050582 (2012).

21. Kosaka, Y. & Xie, S.-P. Recent global-warming hiatus tied to equatorial Pacific surface cooling. *Nature* **501**, 403-407 (2013).

22. Lean, J. L. & Rind, D. H. How will Earth's surface temperature change in future decades? *Geophys. Res. Lett.* **36** (2009).

23. Foster, G. & Rahmstorf, S. Global temperature evolution 1979-2010. *Env. Res. Let.* **6** (2011).

24. Kaufmann, R. K., Kauppi, H., Mann, M. L. & Stock, J. H. Reconciling anthropogenic climate change with observed temperature 1998-2008. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 11790-11793 (2011).

25. Thompson, D. W. J., Wallace, J. M., Jones, P. D. & Kennedy, J. J. Identifying signatures of natural climate variability in time series of global-mean surface temperature: Methodology and insights. *J. Climate* **22**, 6120-6141 (2009).

26. Santer, B. D. *et al.* Separating signal and noise in atmospheric temperature changes: The importance of timescale. *J. Geophys. Res.-Atmos.* **116**, D22105 (2011).

27. Morice, C. P., Kennedy, J. J., Rayner, N. A. & Jones, P. D. Quantifying uncertainties in global and regional temperature change using an ensemble of

observational estimates: The HadCRUT4 data set. *J. Geophys. Res.-Atmos.* **117** (2012).

28. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An Overview of CMIP5 and the Experiment Design. *Bull. Am. Met. Soc.* **93**, 485-498 (2012).

29. Cowtan, K. & Way, R. G. Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Quart. J. Roy. Met. Soc.* **140**, 1935-1944 (2014).

30. Simmons, A. J., Willett, K. M., Jones, P. D., Thorne, P. W. & Dee, D. P. Low-frequency variations in surface atmospheric humidity, temperature, and precipitation: Inferences from reanalyses and monthly gridded observational data sets. *J. Geophys. Res.-Atmos.* **115**, D01110 (2010).

31. Anderson, J. L. A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Climate* **9**, 1518-1530 (1996).

32. Easterling, D. R. & Wehner, M. F. Is the climate warming or cooling? *Geophys. Res. Lett.* **36**, L08706, doi: 08710.01029/02009GL037810 (2009).

33. Stroeve, J. C. *et al.* Trends in Arctic sea ice extent from CMIP5, CMIP3 and observations. *Geophys. Res. Lett.* **39** (2012).

34. Notz, D., Haumann, F. A., Haak, H., Jungclaus, J. H. & Marotzke, J. Arctic sea-ice evolution as modeled by Max Planck Institute for Meteorology's Earth system model. *J. Adv. Model. Earth Sys.* **5**, 173-194 (2013).

35. Forster, P. M. *et al.* Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models. *J. Geophys. Res.-Atmos.* **118**, 1-12, doi: 10.1002/jgrd.50174 (2013).

36. Gregory, J. M. & Forster, P. M. Transient climate response estimated from radiative forcing and observed temperature change. *J. Geophys. Res.-Atmos.* **113** (2008).

37. Held, I. M. *et al.* Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing. *J. Climate* **23**, 2418-2427 (2010).

38. Andrews, T., Gregory, J. M., Webb, M. J. & Taylor, K. E. Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. *Geophys. Res. Lett.* **39**, L09712, doi: 09710.01029/02012GL051607 (2012).

39. Kuhlbrodt, T. & Gregory, J. M. Ocean heat uptake and its consequences for the magnitude of sea level rise and climate change. *Geophys. Res. Lett.* **39** (2012).

40. Vial, J., Dufresne, J.-L. & Bony, S. On the interpretation of inter-model spread in CMIP5 climate sensitivity estimates. *Clim. Dyn.*, 1-24 (2013).

41. Gregory, J. M. *et al.* A new method for diagnosing radiative forcing and climate sensitivity. *Geophys. Res. Lett.* **31** (2004).

42. Block, K. & Mauritsen, T. Forcing and feedback in the MPI-ESM-LR coupled model under abruptly quadrupled $CO_2$. *J. Adv. Model. Earth Sys.* **5**, 676-691 (2013).

43. Meraner, K., Mauritsen, T. & Voigt, A. Robust increase in equilibrium climate sensitivity under global warming. *Geophys. Res. Lett.* **40**, 5944-5948 (2013).

44. Forster, P. *et al.* in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds S. Solomon *et al.*) 129-234 (Cambridge University Press,, 2007).

45. Santer, B. D. *et al.* Identifying human influences on atmospheric temperature. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 26-33 (2013).

46. Myhre, G. *et al.* 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* (eds T. F. Stocker *et al.*) 659-740 (Cambridge University Press, , 2013).

47. Fyfe, J. C., von Salzen, K., Cole, J. N. S., Gillett, N. P. & Vernier, J. P. Surface response to stratospheric aerosol changes in a coupled atmosphere-ocean model. *Geophys. Res. Lett.*, DOI: 10.1002/grl.50156 (2013).

48. Ridley, D. A. *et al.* Total volcanic stratospheric aerosol optical depths and implications for global climate change. *Geophys. Res. Lett.*, in press, doi: 10.1002/2014GL061541 (2014).

## Acknowledgements

We are indebted to John Fyfe for making his CMIP5 GMST dataset available to us, and to Dirk Notz, Ben Santer, and an anonymous reviewer for helpful comments on the manuscript. 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 (names of models listed in Extended Data Table 1) for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. This work was supported by the Max Planck Society for the Advancement of Science (JM) and by a Royal Society Wolfson Merit Award and EPSRC grant EP/1014721/1 (PMF).

## Author contributions

Both authors jointly designed the study. JM analysed the data and wrote the manuscript. Both authors discussed the results and the manuscript.

## Author information

Correspondence and requests for materials should be addressed to jochem.marotzke@mpimet.mpg.de. The authors declare no competing financial interests. Reprints and permissions information is available at www.nature.com/reprints.

## Figure legends

**Figure 1.** Simulated and observed 15-year GMST trends since 1900. (a) Joint relative frequency distribution of GMST trends as a function of start year and trend size, based on the full 114-member ensemble (in bins of 0.025 °C per decade; coloured shading). Circles mark the observed trend from the HadCRUT4 data set[27]. (b) Vertical cross-section of (a) for start year 1927; vertical line marks the observed trend. (c) As (b) but for start year 1998. (d) Marginal distribution of simulated GMST trend as a function of trend size (coloured shading), obtained by time-averaging the joint distribution in (a); observed trend distribution (grey shading). (e) Frequency distribution of the rank that the observed trend would have as a member of the model ensemble (rank 1: observed trend smaller than all simulations; rank 115: observed trend larger than all simulations); bin size is five. All histograms are normalised such that their area integral is unity. In (a), each vertical cross section is normalised.

**Figure 2.** Regression-based and observed 15-year GMST trends since 1900. (a) Shading: Joint relative frequency distribution of regression-based GMST trends (from equation (8)) as a function of start year and trend size (in bins of 0.025 °C per decade), based on the reduced 75-member ensemble for which forcing information is available. Thick red line marks the ensemble average; thick black line the observed trend; whiskers the 5–95% confidence range derived from (f). (b) Joint relative frequency distribution of regression result (from equation (8) but without the ensemble-mean trend) as a function of start year and trend size (in bins of 0.025 °C per decade). The p-value of the regression has a median across start years of 0.075, based on the null hypothesis that all regression coefficients are zero. (c) Joint relative frequency

distribution of regression residual as a function of start year and trend size (in bins of 0.025 °C per decade). (d) Vertical cross-section of (c) for start year 1927. (e) Vertical cross-section of (c) for start year 1998. (f) Marginal distribution of regression residual as a function of trend size, obtained by time-averaging the joint distribution in (c). All histograms are normalised such that their area integral is unity. In (a)–(c), each vertical cross section is normalised. All ordinate ranges are identical.

**Figure 3.** Regression-based and observed 62-year GMST trends since 1900. (a) Shading: Joint relative frequency distribution of regression-based GMST trends (from equation (8)) as a function of start year and trend size, based on the reduced 75-member ensemble for which forcing information is available. Thick red line marks the ensemble average; thick black line the observed trend; whiskers the 5–95% confidence range derived from the marginal distribution of (c). (b) Joint relative frequency distribution of regression result (from equation (8) but without the ensemble-mean trend) as a function of start year and trend size. All p-values of the regression are below 0.001, based on the null hypothesis that all regression coefficients are zero. (c) Joint relative frequency distribution of regression residual as a function of start year and trend size. (d) Joint relative frequency distribution of regression contribution from trend in effective radiative forcing. (e) Joint relative frequency distribution of regression contribution from climate feedback parameter $\alpha$. (f) Joint relative frequency distribution of regression contribution from ocean heat uptake efficiency $\kappa$. In all joint relative frequency distributions, GMST trend is collected in bins of 0.0125 °C per decade, and each vertical cross section is normalised such that its area integral is unity. All ordinate ranges are identical.

## Extended Data figure and table legends

**Extended Data Figure 1.** Observed and simulated time series of the anomalies in annually averaged global-mean surface temperature (GMST), from 1900 to 2012. All anomalies are differences from the 1961–1990 time-mean of each individual time series. GMST is the globally averaged merged surface temperature (2 m height over land and surface temperature over the ocean). The figure shows single simulations for the CMIP5 models (thin lines), the multi-model ensemble mean (thick red line), and the HadCRUT4[27] observations (thick black line). All model results have been sub-sampled using the HadCRUT4 observational data mask[11]. (a) 114 realisations from the CMIP5 archive, obtained with 36 different models. (b) Subset of 75 realisations with the 18 different models for which information on effective radiative forcing (ERF) is available[35] (see Extended Data Table 1). The two model ensembles are nearly indistinguishable.

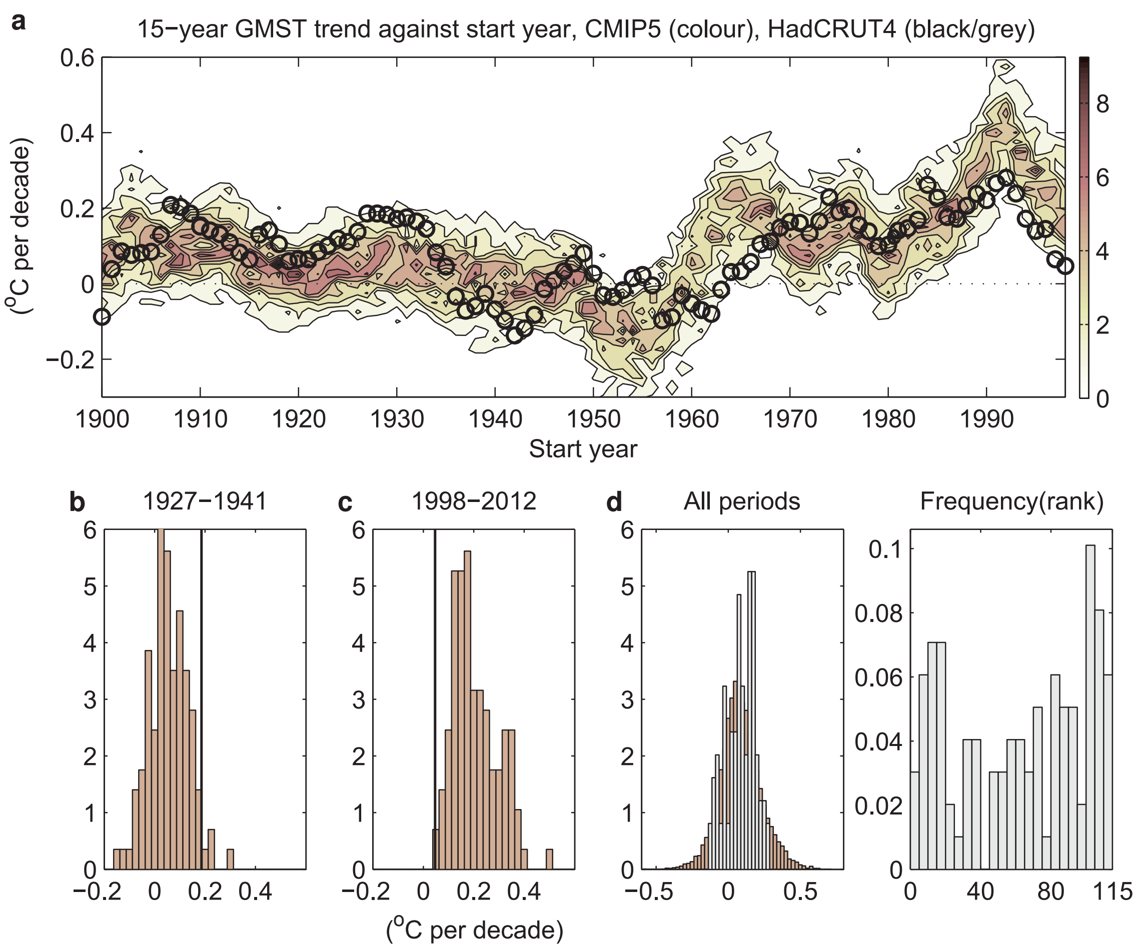
**Extended Data Figure 2.** Time series of trends in effective radiative forcing (ERF), as a function of start year. (a) 15-year trends; (b) 62-year trends. Thin coloured lines: individual models as diagnosed previously[35]; if multiple realisations were available for a model, the ensemble average of the individual diagnosed ERF time series for this model was given[35] and is shown here. Thick red line: ensemble average over all models. Thick black line: best estimate from IPCC AR5 (ref. 46), including for illustration the 5–95% uncertainty range for the periods 1984–1998 (a) and 1951–2011 (b), taken from Fig. 8.19 in ref. 46. These uncertainty ranges, both of which are around 0.2 $Wm^{-2}$ per decade, do not take into account observational biases such as diagnosed in ref. 48. Despite the scatter of the CMIP5 ensemble trends, the ensemble mean is in good agreement with the AR5 best estimate for almost all start years. The IPCC AR5 best-estimate ERF sums time series of forcing across individual forcing terms. Individual time series of AR5 ERF were derived in different ways. Greenhouse-gas concentrations (observed or inferred), stratospheric aerosol optical depth, and total solar irradiance were used to derive estimates of radiative forcing using simple formulae. Surface albedo forcing was derived from estimated anthropogenic vegetation trends. Ozone and aerosol forcings were derived from chemical transport model results with aspects of the forcing constrained by other modelling approaches or observations, or both. ERF sums rapid adjustments with traditional radiative forcings (RFs). Most time series in AR5 were based on traditional radiative forcings, and only $CO_2$ and aerosol forcings included an assessment of the rapid adjustment. In other cases ERF and RFs were assumed to be the same. The AR5 ERF for the most recent 2000–2011 period included updated estimates of volcanic and solar forcing, taking into account the broader 2008/9 solar minimum and post-2000 volcanic activity[46]. These two cooling influences are not included in the CMIP5 ERF; it is hence surprising and unexplained why the CMIP5 ensemble-mean of 15-year ERF trends lies below the best-estimate AR5 ERF trend for the latest start years in (a).

**Extended Data Figure 3.** Joint relative frequency distribution as a function of GMST trend and ERF trend, for the reduced 75-member ensemble for which forcing information is available and all start years. (a) 15-year trends; bin sizes are 0.025 °C per decade and 0.05 $Wm^{-2}$ per decade for GMST and ERF trend, respectively. (b) 62-year trends; bin sizes are 0.0125 °C per decade and 0.025 $Wm^{-2}$ per decade for GMST and ERF trend, respectively. The "climate resistance", $\rho$, is given by $\rho = \alpha + \kappa$ (refs. 35-37).

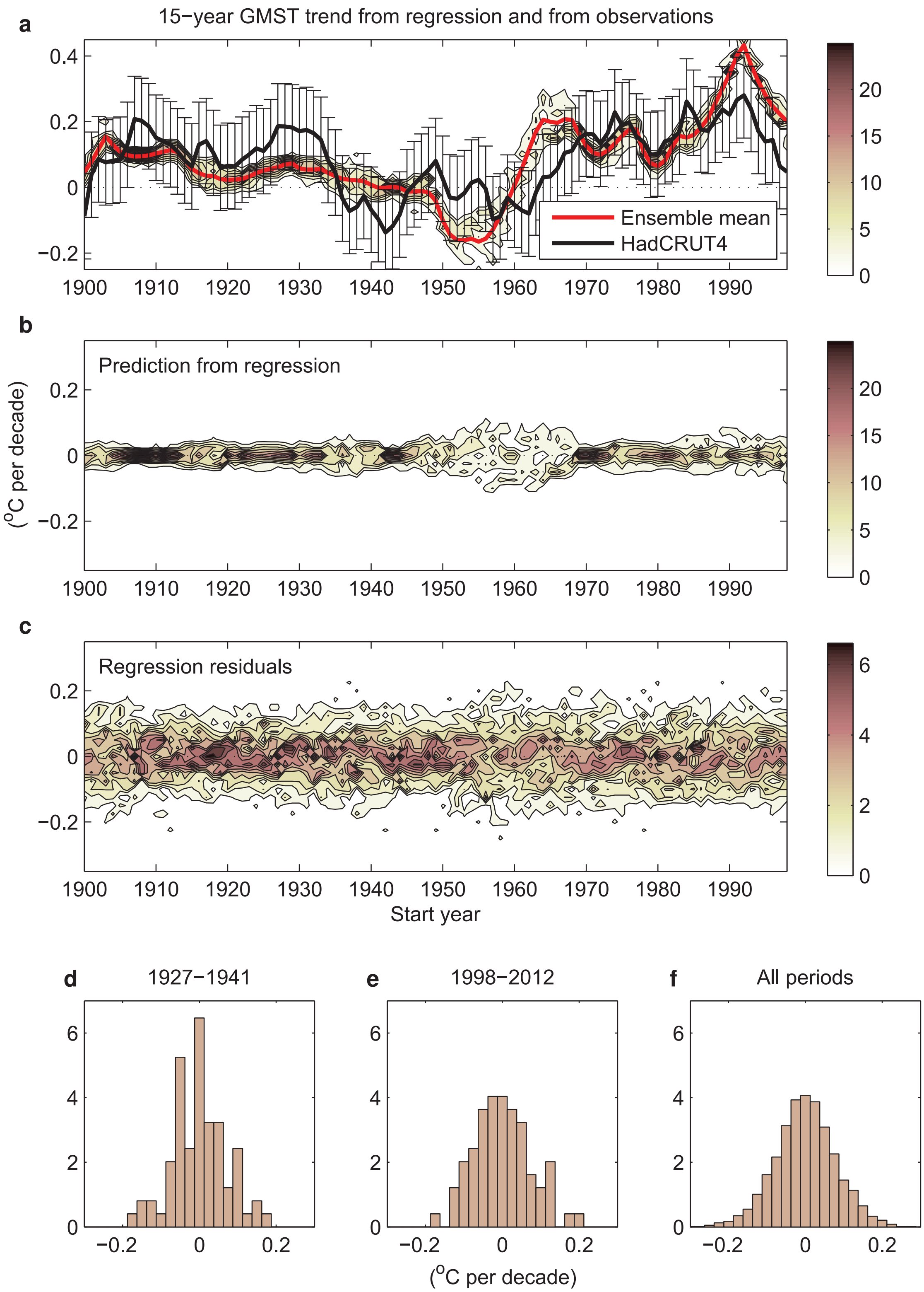
Each joint distribution is normalised such that its area integral is unity. Notice the different axes, reflecting the much tighter correlation of the 62-year trends.

**Extended Data Figure 4.** Regression-based 15-year GMST trends since 1900. (a) Joint relative frequency distribution of regression result (from equation (8) but without the ensemble-mean trend) as a function of start year and trend size. The p-values of the regression have a median across start years of 0.075, based on the null hypothesis that all regression coefficients are zero. (b) Joint relative frequency distribution of regression contribution from trend in effective radiative forcing (ERF). (c) Joint relative frequency distribution of regression contribution from climate feedback parameter $\alpha$. (d) Joint relative frequency distribution of regression contribution from ocean heat uptake efficiency $\kappa$. In all joint relative frequency distributions, GMST trend is collected in bins of 0.025 °C per decade, and each vertical cross section is normalised such that its area integral is unity.

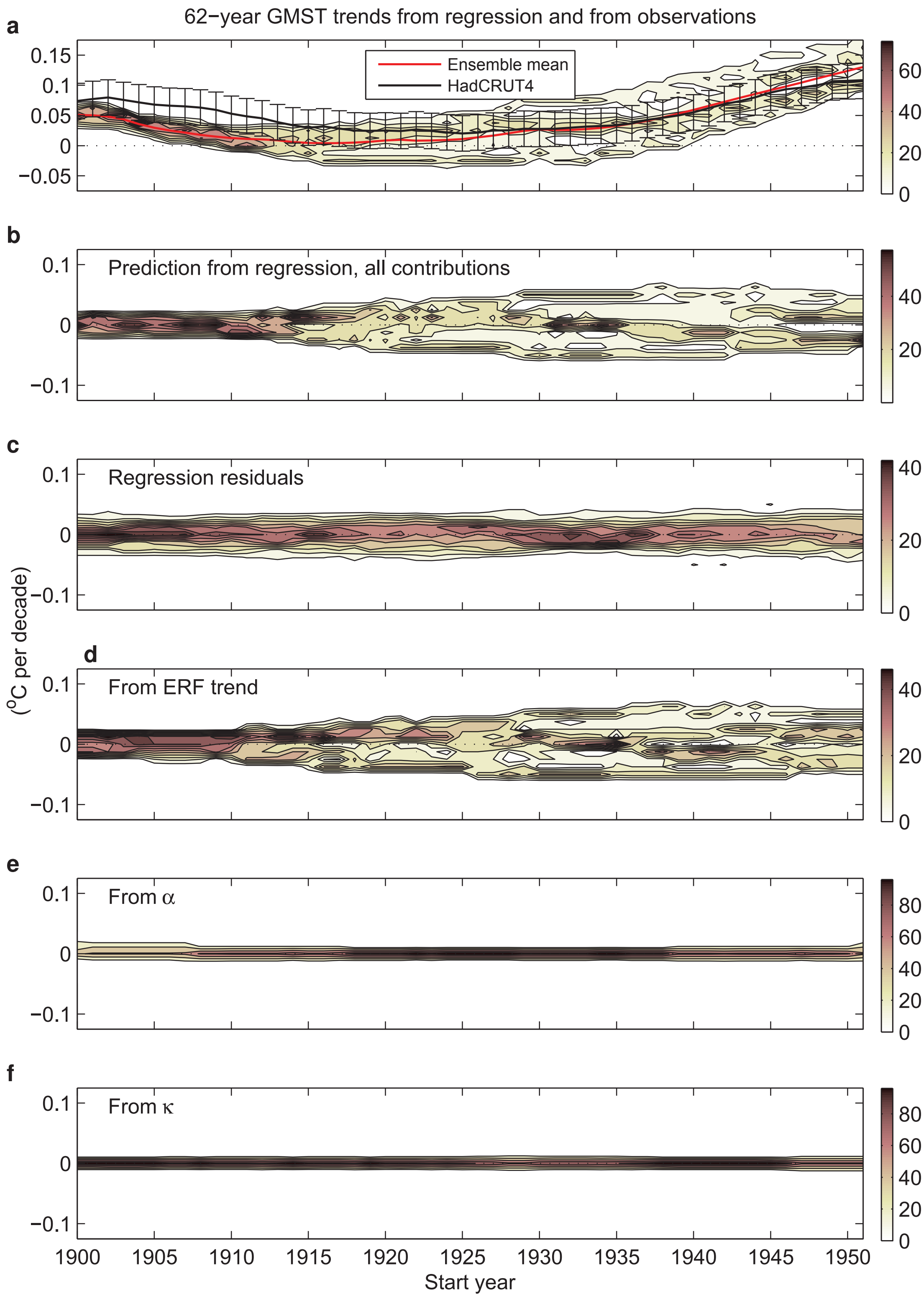
**Extended Data Table 1.** CMIP5 models used in this study. The originating institutions and publications documenting the models are listed comprehensively in Table 9.A1 of ref. 5.



a
15−year GMST trend against start year, CMIP5 (colour), HadCRUT4 (black/grey)
(°C per decade)
Start year
b
1927−1941
c
1998−2012
(°C per decade)
d
All periods
Frequency(rank)

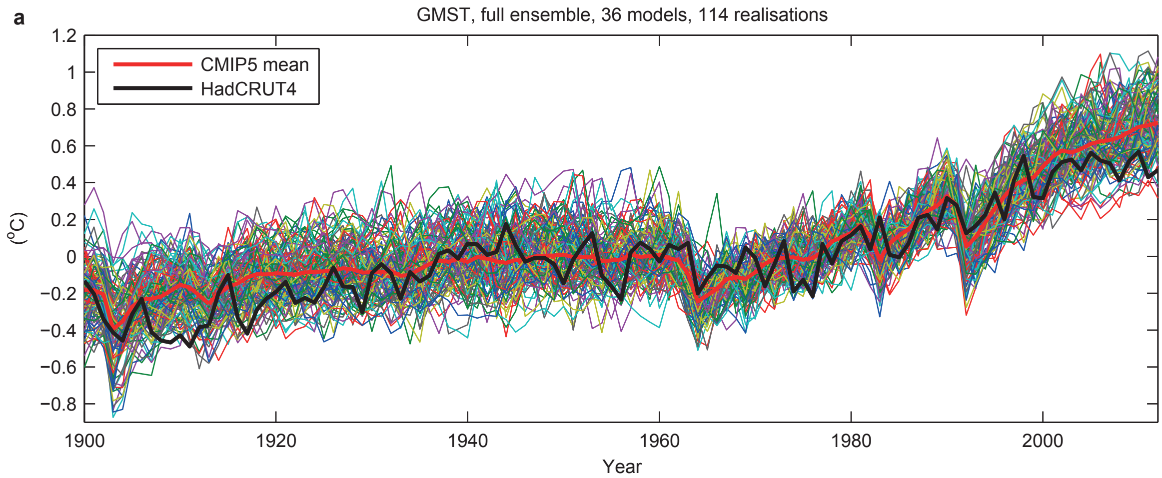


a
15−year GMST trend from regression and from observations
Ensemble mean
HadCRUT4
b
Prediction from regression
(°C per decade)
c
Regression residuals
Start year
d
1927−1941
e
1998−2012
f
All periods
(°C per decade)



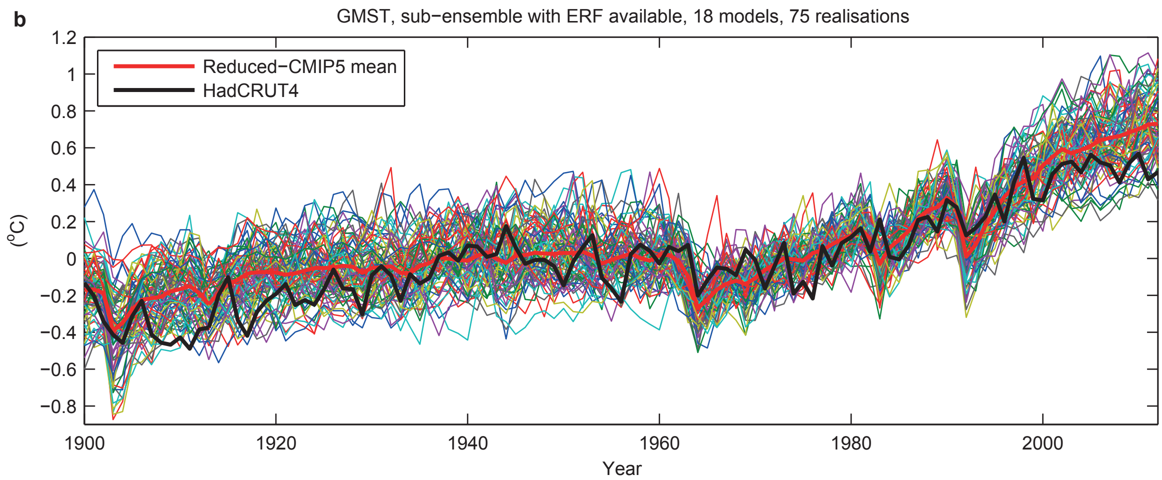
a
62−year GMST trends from regression and from observations
Ensemble mean
HadCRUT4
0.15
0.1
0.05
0
−0.05
60
40
20
0
b
Prediction from regression, all contributions
c
Regression residuals
d
From ERF trend
e
From α
f
From κ
(°C per decade)
1900
1905
1910
1915
1920
1925
1930
1935
1940
1945
1950
Start year

**a**

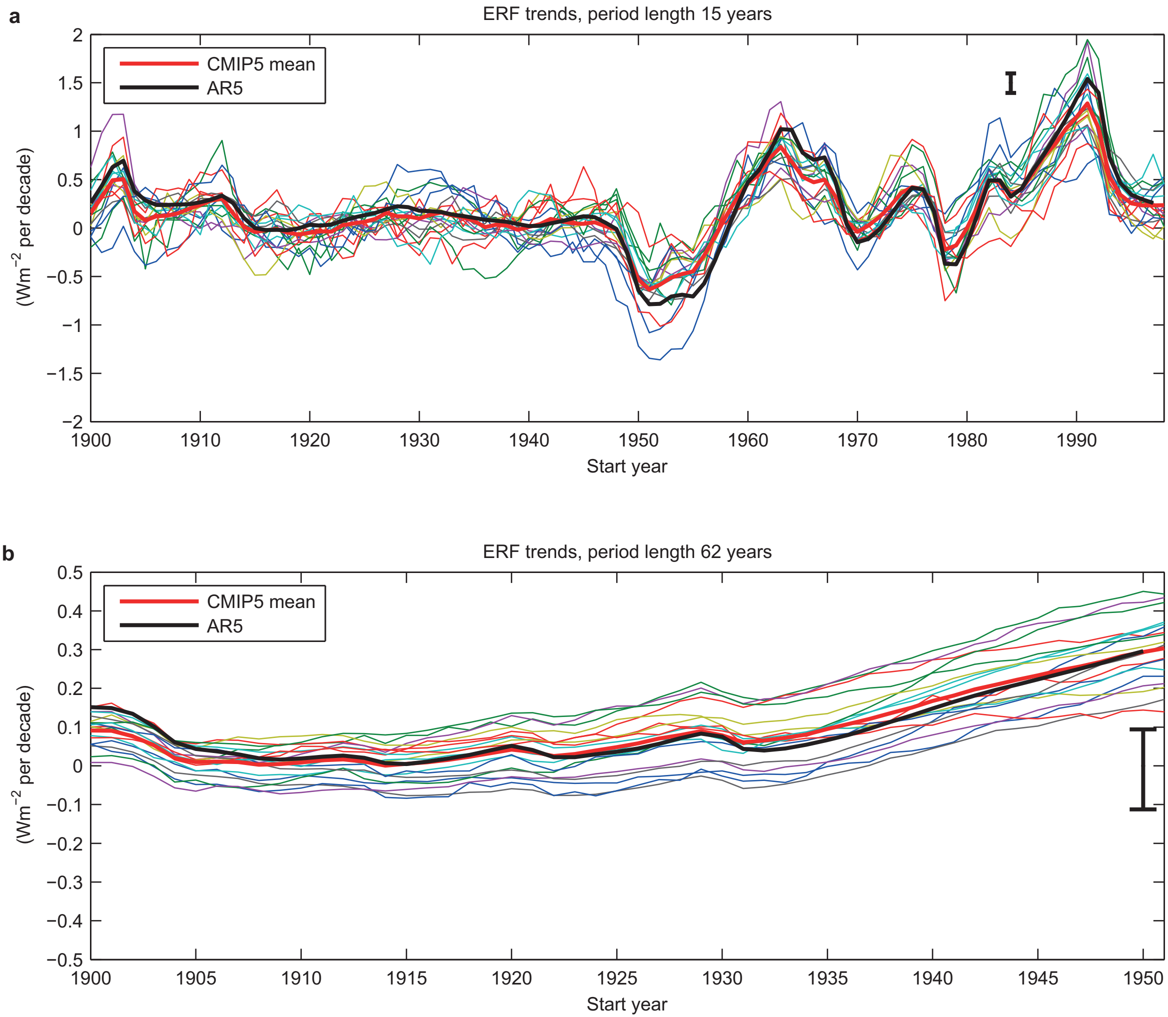


GMST, full ensemble, 36 models, 114 realisations
CMIP5 mean
HadCRUT4
(°C)
1.2
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
1900
1920
1940
1960
1980
2000
Year

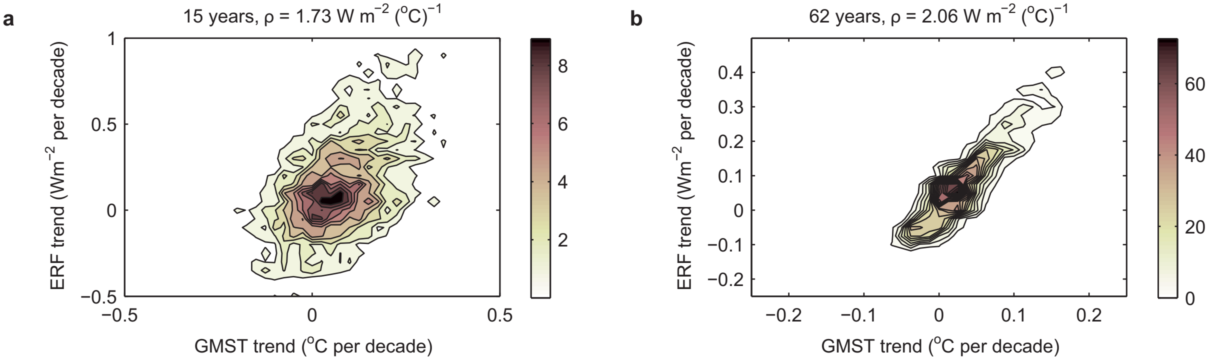

**b**

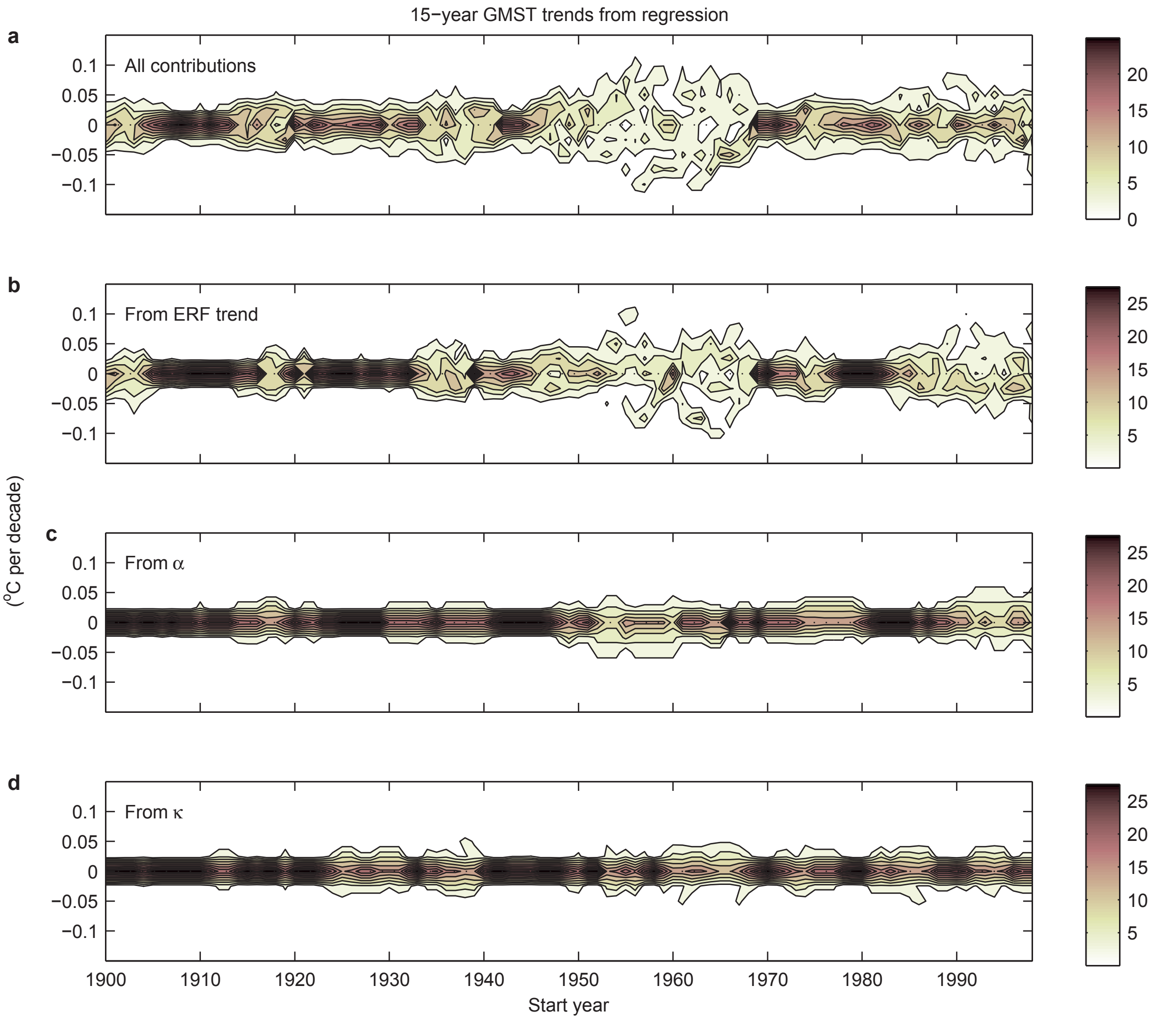


GMST, sub−ensemble with ERF available, 18 models, 75 realisations
Reduced−CMIP5 mean
HadCRUT4
(°C)
1.2
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
1900
1920
1940
1960
1980
2000
Year



a
ERF trends, period length 15 years
CMIP5 mean
AR5
(Wm$^{-2}$ per decade)
Start year
2
1.5
1
0.5
0
−0.5
−1
−1.5
−2
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
b
ERF trends, period length 62 years
CMIP5 mean
AR5
(Wm$^{-2}$ per decade)
Start year
0.5
0.4
0.3
0.2
0.1
0
−0.1
−0.2
−0.3
−0.4
−0.5
1900
1905
1910
1915
1920
1925
1930
1935
1940
1945
1950


a
15 years, ρ = 1.73 W m−2 (°C)−1
ERF trend (Wm−2 per decade)
GMST trend (°C per decade)
1
0.5
0
−0.5
−0.5
0
0.5
8
6
4
2
b
62 years, ρ = 2.06 W m−2 (°C)−1
ERF trend (Wm−2 per decade)
GMST trend (°C per decade)
0.4
0.3
0.2
0.1
0
−0.1
−0.2
−0.2
−0.1
0
0.1
0.2
60
40
20
0


15-year GMST trends from regression
a
All contributions
b
From ERF trend
c
From α
d
From κ
0.1
0.05
0
−0.05
−0.1
(°C per decade)
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
Start year
25
20
15
10
5
0

| Model name | Number of realisations | Forcing available? |
|---|---|---|
| ACCESS1-0 | 1 | Y |
| ACCESS1-3 | 1 | |
| bcc-csm1-1 | 3 | Y |
| bcc-csm1-1-m | 3 | Y |
| BNU-ESM | 1 | |
| CanESM2 | 5 | Y |
| CCSM4 | 6 | Y |
| CESM1-BGC | 1 | |
| CESM1-CAM5 | 3 | |
| CMCC-CM | 1 | |
| CMCC-CMS | 1 | |
| CNRM-CM5 | 10 | Y |
| CSIRO-Mk3-6-0 | 10 | Y |
| FIO-ESM | 3 | |
| GFDL-CM3 | 5 | Y |
| GFDL-ESM2G | 1 | Y |
| GFDL-ESM2M | 1 | Y |
| GISS-E2-H | 5 | |
| GISS-E2-H-CC | 1 | |
| GISS-E2-R | 6 | Y |
| GISS-E2-R-CC | 1 | |
| HadCM3 | 10 | |
| HadGEM2-AO | 1 | |
| HadGEM2-CC | 1 | |
| HadGEM2-ES | 1 | Y |
| IPSL-CM5A-LR | 6 | Y |
| IPSL-CM5A-MR | 3 | |
| IPSL-CM5B-LR | 1 | |
| MIROC5 | 5 | Y |
| MIROC-ESM | 3 | Y |
| MIROC-ESM- | 1 | |
| MPI-ESM-LR | 3 | Y |
| MPI-ESM-MR | 3 | |
| MRI-CGCM3 | 3 | Y |
| NorESM1-M | 3 | |