nature climate change

LETTERS

PUBLISHED ONLINE: 27 JUNE 2016 | DOI: 10.1038/NCLIMATE3066

Reconciled climate response estimates from climate models and the energy budget of Earth

Mark Richardson[1*], Kevin Cowtan[2], Ed Hawkins[3] and Martin B. Stolpe[4]

Climate risks increase with mean global temperature[1], so knowledge about the amount of future global warming should better inform risk assessments for policymakers. Expected near-term warming is encapsulated by the transient climate response (TCR), formally defined as the warming following 70 years of 1% per year increases in atmospheric CO_2 concentration, by which point atmospheric CO_2 has doubled. Studies based on Earth's historical energy budget have typically estimated lower values of TCR than climate models, suggesting that some models could overestimate future warming[2]. However, energy-budget estimates rely on historical temperature records that are geographically incomplete and blend air temperatures over land and sea ice with water temperatures over open oceans. We show that there is no evidence that climate models overestimate TCR when their output is processed in the same way as the HadCRUT4 observation-based temperature record[3,4]. Models suggest that air-temperature warming is 24% greater than observed by HadCRUT4 over 1861–2009 because slower-warming regions are preferentially sampled and water warms less than air[5]. Correcting for these biases and accounting for wider uncertainties in radiative forcing based on recent evidence, we infer an observation-based best estimate for TCR of 1.66 °C, with a 5–95% range of 1.0–3.3 °C, consistent with the climate models considered in the IPCC 5th Assessment Report.

TCR for the Climate Model Intercomparison Project, phase 5 (CMIP5) models is defined using simulations in which atmospheric CO_2 increases at 1% per year and the multi-model mean is 1.8 °C (1.2–2.4 °C, henceforth bracketed values refer to 5–95% ranges)[6–8]. TCR has also been estimated from Earth's energy budget using:

$$\mathrm{TCR} = \frac{\Delta T}{\Delta F}\Delta F_{2\times\mathrm{CO}_2} \quad (1)$$

where ΔT is the observed change in temperature, ΔF is the change in radiative forcing, and $\Delta F_{2\times\mathrm{CO}_2}$ is the forcing change for doubled atmospheric CO_2. Energy-budget calculations have recently been able to provide more constrained estimates of TCR, due to increased amplitudes of ΔT and ΔF relative to their uncertainties (see Supplementary Information). These energy-budget estimates have typically fallen below the CMIP5 multi-model mean, for example, 1.5 °C from Bengtsson and Schwartz (1.0–1.9 °C)[9], 1.3 °C (0.9–2.0 °C) from Otto *et al.*[2] and 1.3 °C (0.9–2.5 °C) from Lewis and Curry (2015)[10].

The lower best estimates of TCR from these observation-based studies relative to CMIP5 may be due to a combination of: biases in observed temperature series[11], varying efficacy of different forcings[12–16], time and history dependence of TCR[17], internal variability[18], overestimate of forcings[19], efficacy of ocean heat uptake[20–22], structural uncertainties in energy-budget calculations or lower real-world TCR.

We focus on potential biases in temperature series due to geographical incompleteness of the data ('masking') and the combination of air and water measurements ('blending') by applying energy-budget TCR calculations to CMIP5 simulations and observations. We calculate energy-budget TCR with the Otto *et al.* method[2], henceforth 'Otto', which uses differences between an early baseline period and a recent reference period:

$$\mathrm{TCR} = \frac{\overline{T_{2000-2009}} - \overline{T_{1861-1880}}}{\overline{F_{2000-2009}} - \overline{F_{1861-1880}}}\Delta F_{2\times\mathrm{CO}_2} \quad (2)$$

where $\overline{T_{2000-2009}}$ is the mean temperature anomaly over 2000–2009, and the other symbols follow this format. We shift the Otto baseline period by one year to include CMIP5 simulations beginning in 1861, and end at 2009 due to lack of available and consistent forcing data. Our conclusions are robust to the choice of time period and to two other energy-budget calculation methods (see Supplementary Information). As we use published radiative forcing series[2,7], our analysis determines only the effect on calculated TCR due to changes in the ΔT term.

The single largest contribution to the formal error in calculated TCR is, however, due to uncertainty in ΔF. Otto used a Gaussian distribution with a 5–95% range of $\pm 0.58\ \mathrm{W\,m^{-2}}$. The IPCC 5th Assessment Report reports a larger uncertainty range, so we use the Otto median with uncertainties based on Lewis and Curry's more sophisticated 2015 IPCC-based uncertainty distribution, which also accounts for non-Gaussian behaviour and cross correlation between terms (see Methods). This range requires scaling, as it uses slightly different time periods, but our result is not sensitive to this (see Supplementary Information). Although we focus on TCR, the equilibrium climate sensitivity (ECS) is another common metric:

$$\mathrm{ECS} = \frac{\Delta T}{\Delta F - \Delta Q}\Delta F_{2\times\mathrm{CO}_2} \quad (3)$$

where ΔQ is the system heat uptake, which, being positive during warming, means that ECS is larger than TCR. We do not calculate ECS here, to avoid uncertainties associated with ΔQ, and to avoid the assumption of linear climate response, which is less accurate over the longer time periods required for equilibrium[17]. However, as ΔT is in the numerators of equations (1) and (3), any ΔT bias affects each calculation equally in percentage terms.

[1]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA. [2]Department of Chemistry, University of York, York YO10 5DD, UK. [3]National Centre for Atmospheric Science, Department of Meteorology, University of Reading, Reading RG6 6BB, UK. [4]Institute for Atmospheric and Climate Science, ETH Zurich, 8092 Zurich, Switzerland. *e-mail: markr@jpl.nasa.gov

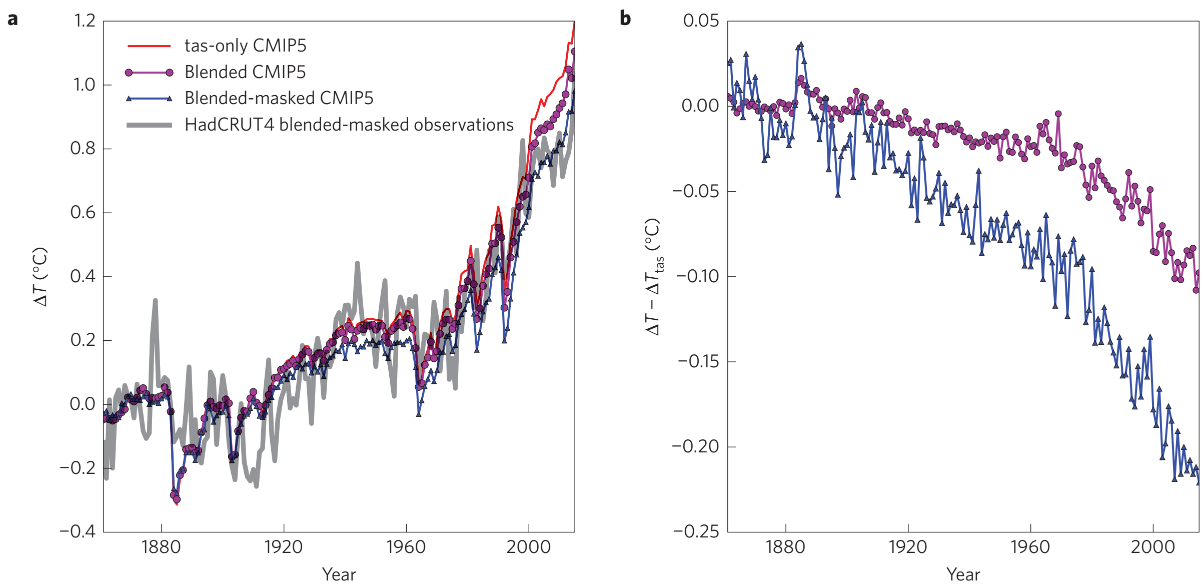

Figure 1 | Median CMIP5-simulated temperature series from the temperature reconstruction method compared with the HadCRUT4 observational series. a, Ensemble-median temperature change relative to an 1861–1880 baseline for tas-only (red line), blended (magenta line with circles) and blended-masked simulations (blue line with triangles), along with HadCRUT4 blended-masked observations (thick grey line). **b**, Blended minus tas-only (magenta line with circles) and blended-masked minus tas-only (blue line with triangles).

Formally, TCR refers to global near-surface air temperature ('tas' in CMIP5 nomenclature) for ΔT, whereas observational temperature records have incomplete and varying geographical coverage and combine air temperatures over land and sea ice with near-surface water temperatures over oceans. These differences introduce biases, as warming is not spatially uniform, sea ice coverage changes, and as air and water temperatures are expected to change differently[4,23]. Although some work accounted for these issues, it has not yet been included in energy-budget analyses[24].

Here we use equation (2) to calculate TCR in a consistent way from both observations and CMIP5 simulations. For observation-based TCR, ΔT is from HadCRUT4 and ΔF is the Otto median with IPCC-like uncertainty, which updated the CMIP5 mean based on observational constraints[10]. For modelled TCR, ΔT is from the Cowtan *et al.*[4] CMIP5 series, ΔF is the CMIP5 forcing series for each model where available, and the multi-model mean otherwise[7] (see Methods and Supplementary Information). Modelled data use the historical scenario from 1861–2005 and the Representative Concentration Pathway 8.5 (RCP8.5) from 2006[25]. Scenario choice has little effect over the short period 2006–2009 used in the TCR calculation, but may diverge from reality in the future.

Model temperatures are reconstructed in three ways: by using global air temperature ('tas-only'), by blending air temperature over land and sea ice with ocean temperatures over water ('blended'), and by blending temperatures and using the historical geographical coverage of observations in HadCRUT4 ('blended-masked'). We assume that the modelled near-surface water temperature over oceans ('tos' in CMIP5 nomenclature) is equivalent to measured sea surface temperatures. Results are similar between models with different ocean layering: for example with a 2.5 m top-layer depth instead of 10 m, suggesting tos is a robust measure of modelled sea surface temperature (see Supplementary Information).

The tas-only reconstructions are used in standard model assessments of TCR, the 'blended' reconstructions represent the same reconstruction techniques as HadCRUT4, but with perfect data coverage, and the 'blended-masked' reconstructions represent HadCRUT4.

Figure 1 shows the ensemble-median global temperature series for each reconstruction. Between 1861–1880 and 2000–2009, HadCRUT4 warms slightly more (0.75 °C) than the multi-model median (0.73 °C) in a like-with-like comparison, although modelled tas-only series warmed more (0.93 °C). This effect exceeds the approximately 10% difference quoted in Cowtan *et al.*[4], which referred to the blending effect only, that is, masking increases the effect further. Supplementary Table 8 shows that the masking bias is largely due to undersampling of rapidly warming polar regions. The blending and masking effects were not accounted for in the energy-budget studies cited here, although masking has been considered in some other analyses[26].

After applying equation (2), Fig. 2 shows that the TCR from the blended-masked HadCRUT4 series of 1.34 °C falls at the 33rd percentile of the blended-masked model distribution, but at the 7th percentile of TCR derived from tas-only model reconstructions.

Figure 3 shows that the energy-budget TCR inferred from tas-only temperature reconstructions is consistently higher than that inferred from blended or blended-masked reconstructions, and that both blending and masking contribute to the median bias of 24% in ΔT. We correct for this bias by updating the blended-masked TCR derived from equation (2) using Otto data for the best estimates of each parameter but a scaled Lewis and Curry forcing distribution accounting for correlation between ΔF and $\Delta F_{2\times CO_2}$ (see Methods). Our blended-masked estimate of 1.34 °C (range 0.8–2.6 °C) is updated by applying our derived correction of 24% (including ±2% Gaussian uncertainty) to this distribution. The observation-based energy-budget calculation implies a best estimate for tas-only TCR of 1.66 °C (range 1.0–3.3 °C, see Methods and Supplementary Information), consistent with the CMIP5 range as shown in Fig. 4. This result is robust to a variety of assumptions and correction approaches (see Supplementary Information). Intrinsic uncertainties in natural variability, model structure and real-world ΔF are large, and improved understanding of these factors may adjust these results in the future. Of the 24% difference between tas-only TCR and the observation-based blended-masked estimate, we report that approximately 9 percentage points are due to blending and 15 percentage points to masking (from Supplementary Table 5).

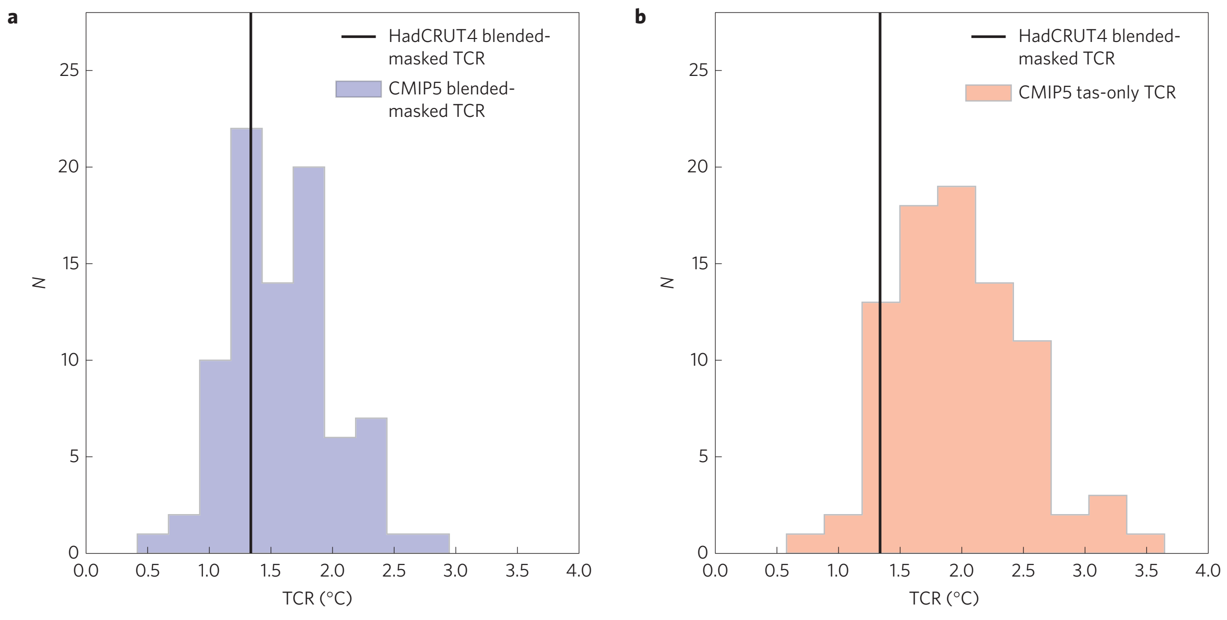

Figure 2 | Histograms of TCR calculated for CMIP5 simulations with the observation-based HadCRUT4-derived value also shown as a vertical line. **a**,**b**, HadCRUT4 used with Otto median forcing, CMIP5 simulations with model-specific forcing where available, multi-model mean otherwise. **a**, Consistent comparison between blended-masked observations and blended-masked CMIP5 simulations, where the observations fall at the 33rd percentile of the model distribution. **b**, Inconsistent comparison between blended-masked observations and global-air-temperature-derived values from CMIP5, where the observations fall at the 7th percentile of the model distribution.

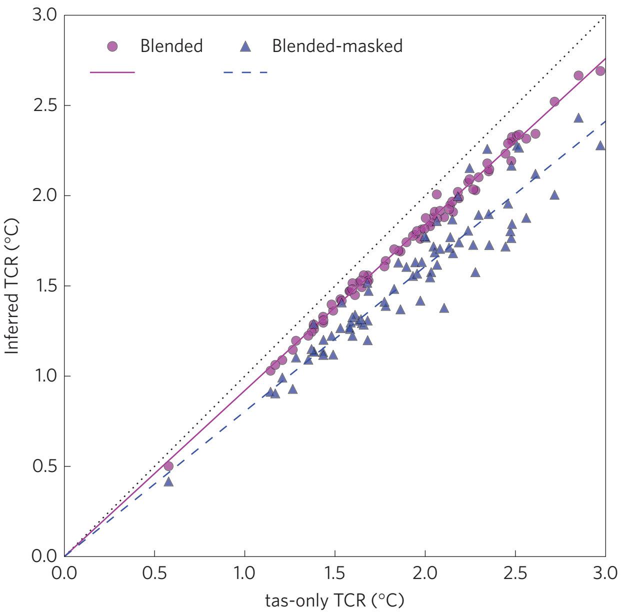

Figure 3 | Energy budget estimates of TCR using the Otto *et al.*[2] energy-budget calculation applied to historical-RCP8.5 simulations. Values calculated from blended reconstructions or blended-masked reconstructions as a function of the tas-only derived TCR for each simulation. Best-fit lines shown for each case: solid magenta for blended and dashed blue for blended-masked, whereas the 1:1 line is shown as a dotted line.

Two further questions can be raised: is the difference in simulated warming rates between water and air realistic, and what does this mean for future research and impacts?

Modelled global air temperatures warm 7–9% faster than blended air–water temperatures, with a component from the faster warming of air relative to water, and the remainder from changes in sea ice redistributing air and water measurements, as discussed by Cowtan and colleagues[4]. We propose that changes in surface energy balance contribute to air temperatures warming faster: radiative equilibrium implies a temperature discontinuity at Earth's surface, with surface temperatures higher than air[27], which drives vertical latent and sensible heat fluxes. The size of this discontinuity depends on atmospheric optical depth, such that more CO_2 and warming-induced increases in water vapour suppress the surface temperature discontinuity, meaning greater air-temperature warming. Further adjustments in surface energy balance associated with non-radiative heat transfer affect the amplitude of this effect: warming increases evaporation at the surface, whereas condensation increases at altitude. The increased latent heat transfer outweighs reductions in sensible heat fluxes in models[28] and is related to the lapse-rate feedback, which acts to reduce surface warming and increase warming of the air aloft.

The blending effect implies a limiting case of a 7–9% bias in model-observation comparisons for perfect geographical data coverage. Alternative measurements of surface and air temperatures over oceans are required to assess this expected bias in observations. The greatest immediate opportunities to reduce bias therefore appear to be in recovery efforts for historical data records[29] and improved spatial interpolation[11], which should reduce the potential 24% bias in observed global mean warming inferred over 1861–2009. Indeed, improved observational coverage has reduced the combined blending-masking bias to approximately 15% over the period 1970–2010 (see Supplementary Information). This implies that future estimates of TCR will be less sensitive to this bias as more data become available.

Other research that uses temperature changes over multidecadal or longer timescales may well be sensitive to the choice of temperature metric, and researchers should be clear about which temperature metric or reconstruction method they are using, to minimize the risk of biases introduced through inconsistent comparisons.

This issue also has considerable implications for policy discussions about limiting global average temperature to some

NATURE CLIMATE CHANGE DOI: 10.1038/NCLIMATE3066

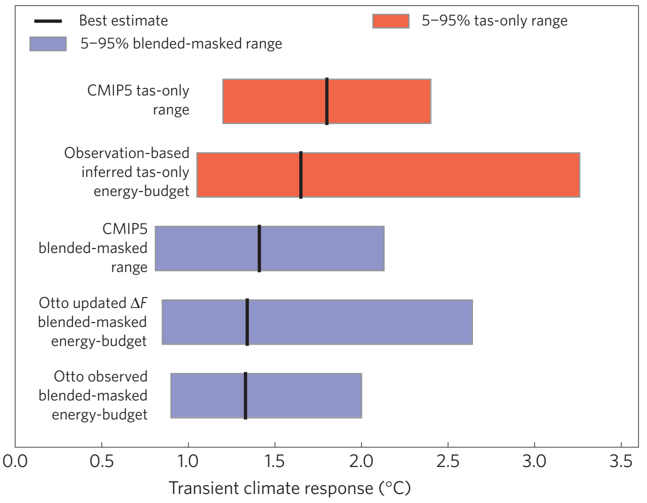

Figure 4 | Comparison of modelled and observed TCR estimated from Earth's historical energy budget. The blue bars show blended-masked results, reported upwards as Otto *et al.*'s results[2] using blended-masked HadCRUT4 observations, the same results using scaled Lewis and Curry forcing, and the range when the same calculation is applied to blended-masked CMIP5 temperature series (one simulation per model). The red bars compare our bias-corrected estimates of tas-only TCR from HadCRUT4 using the Otto calculation with Lewis and Curry forcings, and the canonical CMIP5 model range. The updated observation-based estimate is higher due to the corrected blending-masking bias, and has a wider range than Otto due to the greater uncertainty in radiative forcing series used. Boxes represents 5–95% range and thick vertical lines are the best estimate.

particular level, such as 2 °C above pre-industrial[30]. If our reported air–ocean warming differences are robust, then which global temperatures are relevant for policy? If it is decided that climate targets refer to global near-surface air temperature, then the current warming is probably 24% (range 9–40%, see Supplementary Table 1) larger than reported by HadCRUT4.

Previous energy-budget-based TCR estimates reported TCR values towards the lower end of the climate model range. These studies clearly highlighted their limitations, including issues of spatial coverage[11], time sensitivity and the efficacy of different forcings. Nevertheless Otto stated: 'Our results match those of other observation-based studies and suggest that the TCRs of some of the models in the CMIP5 ensemble with the strongest climate response to increases in atmospheric CO_2 levels may be inconsistent with recent observations'.

However, in our like-with-like comparison, the Otto TCR best estimate of 1.3 °C based on the HadCRUT4 blended-masked observational series falls at the 33rd percentile of the CMIP5 blended-masked ensemble. There is therefore no evidence for significant disagreement between modelled and real-world TCR. This implies a TCR for global air temperature of 1.66 °C (1.0–3.4 °C), in better agreement with the CMIP5 multi-model mean of 1.8 °C (1.2–2.4 °C). We conclude that previous analyses that reported observation-based estimates toward the low end of the model range did so largely because of inconsistencies in the temperature reconstruction methods between models and observations.

Methods

Methods and any associated references are available in the online version of the paper.

Received 17 November 2015; accepted 27 May 2016; published online 27 June 2016

References

1. IPCC *Climate Change 2014: Impacts, Adaptation, and Vulnerability* (eds Field, C. *et al.*) (Cambridge Univ. Press, 2015); https://www.ipcc.ch/pdf/assessment-report/ar5/wg2/WGIIAR5-FrontMatterA_FINAL.pdf
2. Otto, A. *et al*. Energy budget constraints on climate response. *Nature Geosci.* **6,** 415–416 (2013).
3. 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.* **117,** D08101 (2012).
4. Cowtan, K. *et al*. Robust comparison of climate models with observations using blended land air and ocean sea surface temperatures. *Geophys. Res. Lett.* **42,** 6526–6534 (2015).
5. Richter, I. & Xie, S.-P. Muted precipitation increase in global warming simulations: a surface evaporation perspective. *J. Geophys. Res.* **113,** D24118 (2008).
6. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93,** 485–498 (2012).
7. 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,** 1139–1150 (2013).
8. 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 (2012).
9. Bengtsson, L. & Schwartz, S. E. Determination of a lower bound on Earth's climate sensitivity. *Tellus B* **65,** 21533 (2013).
10. Lewis, N. & Curry, J. A. The implications for climate sensitivity of AR5 forcing and heat uptake estimates. *Clim. Dynam.* **45,** 1009–1023 (2015).
11. Cowtan, K. & Way, R. G. Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Q. J. R. Meteorol. Soc.* **140,** 1935–1944 (2014).
12. Hansen, J. Efficacy of climate forcings. *J. Geophys. Res.* **110,** D18104 (2005).
13. Shindell, D. & Faluvegi, G. Climate response to regional radiative forcing during the twentieth century. *Nature Geosci.* **2,** 294–300 (2009).
14. Shindell, D. *et al*. Spatial scales of climate response to inhomogeneous radiative forcing. *J. Geophys. Res.* **115,** D19110 (2010).
15. Kummer, J. R. & Dessler, A. E. The impact of forcing efficacy on the equilibrium climate sensitivity. *Geophys. Res. Lett.* **41,** 3565–3568 (2014).
16. Marvel, K., Schmidt, G. A., Miller, R. L. & Nazarenko, L. S. Implications for climate sensitivity from the response to individual forcings. *Nature Clim. Change* **6,** 386–389 (2015).
17. Knutti, R. & Rugenstein, M. A. A. Feedbacks, climate sensitivity and the limits of linear models. *Phil. Trans. R. Soc. A* **373,** 20150146 (2015).
18. Huber, M., Beyerle, U. & Knutti, R. Estimating climate sensitivity and future temperature in the presence of natural climate variability. *Geophys. Res. Lett.* **41,** 2086–2092 (2014).
19. Schmidt, G. A., Shindell, D. T. & Tsigaridis, K. Reconciling warming trends. *Nature Geosci.* **7,** 158–160 (2014).
20. Rose, B. E. J., Armour, K. C., Battisti, D. S., Feldl, N. & Koll, D. D. B. The dependence of transient climate sensitivity and radiative feedbacks on the spatial pattern of ocean heat uptake. *Geophys. Res. Lett.* **41,** 1071–1078 (2014).
21. Winton, M., Takahashi, K. & Held, I. M. Importance of ocean heat uptake efficacy to transient climate change. *J. Clim.* **23,** 2333–2344 (2010).
22. Armour, K. C., Bitz, C. M. & Roe, G. H. Time-varying climate sensitivity from regional feedbacks. *J. Clim.* **26,** 4518–4534 (2013).
23. Ramanathan, V. The role of ocean-atmosphere interactions in the CO_2 climate problem. *J. Atmos. Sci.* **38,** 918–930 (1981).
24. Santer, B. D. *et al*. Interpreting differential temperature trends at the surface and in the lower troposphere. *Science* **287,** 1227–1232 (2000).
25. Riahi, K. *et al*. RCP 8.5—a scenario of comparatively high greenhouse gas emissions. *Climatic Change* **109,** 33–57 (2011).
26. Jones, G. S., Stott, P. A. & Christidis, N. Attribution of observed historical near-surface temperature variations to anthropogenic and natural causes using CMIP5 simulations. *J. Geophys. Res.* **118,** 4001–4024 (2013).
27. Goody, R. & Yung, Y. L. *Atmospheric Radiation Theoretical Basics* 388–425 (Oxford Univ. Press, 1989).
28. DeAngelis, A. M., Qu, X., Zelinka, M. D. & Hall, A. An observational radiative constraint on hydrologic cycle intensification. *Nature* **528,** 249–253 (2015).
29. Thorne, P. W. *et al*. Guiding the creation of a comprehensive surface temperature resource for twenty-first-century climate science. *Bull. Am. Meteorol. Soc.* **92,** ES40–ES47 (2011).
30. *The Copenhagen Accord FCCC/CP/2009/11/Add.1* (United Nations Framework Convention on Climate Change, 2009).

Acknowledgements

M.R. is funded by the Cloudsat and OCO-2 projects. The research described in this paper was performed at the Jet Propulsion Laboratory, California Institute of Technology, sponsored by NASA. E.H. is funded by the UK Natural Environment Research Council and the National Centre for Atmospheric Science. We thank Piers Forster for providing support regarding CMIP5 radiative forcing time series and R. Knutti, P. Jacobs and P. Kalmus for substantial helpful comments. M.R. thanks G. Stephens for advisory support and helpful scientific discussions.

Author contributions

M.R. performed the main analysis, produced the figures and drafted the article. K.C. provided code for temperature reconstruction methods and performed sensitivity tests. E.H. provided input on experimental design and helped write the article, M.B.S. provided input on experimental design, helped write the article and performed sensitivity tests.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.R.

Competing financial interests

The authors declare no competing financial interests.

Methods

The primary results require five main steps, with further analysis and sensitivity tests available in the Supplementary Information. The main steps are as follows:

(1) Extraction of observed and modelled temperature series.
(2) Best estimates of radiative forcing time series for models and observations.
(3) Application of energy-budget calculation.
(4) Deriving a bias correction for the observation-based TCR calculation.
(5) Applying the bias correction to the blended-masked observation-based value to infer the tas-only TCR from observations.

Temperature series. The observed HadCRUT4 temperature record was taken from http://www.cru.uea.ac.uk/cru/data/temperature/HadCRUT4-gl.dat [downloaded 22 March 2016] whereas the CMIP5 temperature series were for the 84 CMIP5 historical-RCP8.5 simulations reported by Cowtan *et al.*[4] with a small update to account for inconsistencies in how some models handled missing data and to include 2015 data. The code used for the present paper is available at http://www-users.york.ac.uk/~kdc3/papers/reconciled2016 and the simulations used are listed in Supplementary Table 13. The Otto *et al.* uncertainty of $\pm 0.20\,^{\circ}$C for changes from 1860–1879 to 2000–2009 is used, with ΔT assumed to follow a Gaussian distribution based on Otto's analysis of intrinsic measurement uncertainty combined with CMIP5-based estimates of internal variability.

For each simulation, three time series of temperature are calculated.

(1) tas-only—the global mean average air-temperature change.
(2) blended—the global mean average temperature change using near-surface air temperatures (tas) over land and sea ice, and near-surface ocean water temperatures (tos) over ice-free ocean. These are referred to as unmasked/anomaly/variable ice in Cowtan and colleagues[4].
(3) blended-masked—similar to blended, but calculated on a $5^{\circ} \times 5^{\circ}$ with the historical month-by-month HadCRUT4 coverage mask applied. These are referred to as the HadCRUT4 method series in Cowtan and colleagues[4].

Our blended-masked simulations are designed to match the HadCRUT4 methodology as closely as possible, using the same gridding and following the corresponding month-by-month HadCRUT4 data coverage. Global temperature anomaly is determined by taking the arithmetic mean of the Southern- and Northern Hemisphere area-weighted means, as in HadCRUT4.

Each model's own sea ice field is used to determine whether to use air or water temperature measurements: in months where sea ice is present the air temperature is used, otherwise the water temperature is used. As discussed in Cowtan *et al.*[4], this may lead to discontinuities as sea ice area changes. In CMIP5 sea ice retreat occurs mostly in summer, and summer air temperatures warm more quickly than ice-covered water temperatures, which are strongly coupled to the freezing point of water and are insulated by the overlying sea ice. By the time sea ice melts, air temperatures have warmed by notably more relative to water temperatures since the reference period used in the anomaly calculation. The removal of ice therefore leads to an immediate jump downwards in reported temperature anomalies, as discussed in Cowtan and colleagues[4]. The use of tos is taken as the closest equivalent to observational SST records which sample near-surface water temperatures.

Each individual CMIP5 simulation is then baselined such that the 1861–1880 mean temperature anomaly is zero, and the CMIP5 median then comes from the median temperature of the ensemble in each year.

The comparison in temperature changes shown in Fig. 1 is based on the difference between the tas-only, blended and blended-masked simulations. The 24% difference we report for ΔT refers to the median of the set of model tas-only divided by blended-masked ΔT values. The difference seen in Fig. 1 is slightly higher, as Fig. 1 shows the difference of the medians rather than the median of the differences.

Radiative forcing. Radiative forcing used with HadCRUT4 to obtain the 'observation-based' TCR was taken from Otto and colleagues[2]. This series is largely diagnosed from models, but was updated based on some observation-based constraints, so we take it as the best understanding of real-world historical ΔF. It uses a version of the CMIP5 multi-model-mean historical-RCP4.5 forcing with updates to better match observed natural variability and with an upward adjustment of 0.3 W m^{-2} based on evidence for weaker real-world cooling by tropospheric aerosol than that simulated by the CMIP5 simulations. The weaker cooling effect of aerosols leads to an increase in the total forcing and a consequent lower calculated value of TCR, as ΔF is in the denominator of equation (1).

For the CMIP5 simulation ΔF we used the historical-RCP8.5 simulated forcing series from http://www.see.leeds.ac.uk/research/icas/research-themes/climate-change-and-impacts/physical-climate-change/current-research/ipcc-intergovernmental-panel-on-climate-change-reports-and-forcings [last accessed 2016-03-25]. We use all models which provide a full radiative forcing time series from 1861 onwards. Each model uses its own forcing if available ($N = 54$), or the multi-model mean otherwise ($N = 30$). Supplementary Information shows that the TCR best estimate is not sensitive to this choice.

There is substantial uncertainty in ΔF and various values have been calculated for observational series. Otto *et al.*[2] reported 1.95 ± 0.58 W m^{-2} for the change from 1860–1879 to 2000–2009, whereas Lewis and Curry (2014) reported 1.98 W m^{-2} (0.99–2.86 W m^{-2}) for the change from 1859–1882 to 1995–2011. The Otto results represents the 5–95% range of a Gaussian distribution, whereas Lewis and Curry used updated forcing estimates from the IPCC Fifth Assessment Report, accounting for individual forcings and allowing for non-Gaussian distributions in some components. We build on the Lewis and Curry forcing uncertainty, as it more accurately represents the IPCC's best understanding and includes a more sophisticated treatment of the cross correlations between terms.

To produce ΔF and $\Delta F_{2\times CO_2}$ distributions we use the Lewis and Curry (2014) code that is available at https://niclewis.wordpress.com/the-implications-for-climate-sensitivity-of-ar5-forcing-and-heat-uptake-estimates [last accessed 30 April 2016]. We extract 1 million samples from each of the output distributions. These distributions include some correlation due to the correlated uncertainty in the CO_2 component that is present in each.

The Lewis and Curry ΔF values are then scaled such that their medians match those from Otto data for 1861–1880 to 2000–2009, resulting in a distribution with the same shape as that derived in Lewis and Curry, a median of 1.94 W m^{-2} and 5–95% range of 0.97–2.81 W m^{-2}. This scaling is required to ensure that the best estimate matches the period used.

The $\Delta F_{2\times CO_2}$ distribution is then scaled such that it has a median of 3.44 W m^{-2} and a range of $\pm 10\%$, consistent with Otto's values, but maintaining the correlation with the ΔF term as in Lewis and Curry.

Energy-budget calculation to obtain TCR. Temperature and radiative forcing differences were calculated using equation (2) by taking the mean values for ΔT and ΔF from 2000 to 2009 and subtracting the means from 1861 to 1880. The mean forcing at CO_2 doubling was taken to be 3.44 W m^{-2}, from Forster and colleagues[7]. In addition, different time periods and the one-box calculation of Held and colleagues[21] and the trend method of Bengtsson and Schwartz[9] were also assessed in the Supplementary Information, and our results are found to be generally robust to the choice of method.

For the HadCRUT4-based estimate, the distributions of $\Delta F_{2\times CO_2}$, ΔT and ΔF were sampled one million times to obtain the TCR distribution. Our best estimate is $1.34\,^{\circ}$C versus $1.32\,^{\circ}$C in Otto, due to the one-year shift in the baseline period from 1860–1879 to 1861–1880, and possibly differences between HadCRUT4 versions and the skewed forcing distribution. Due to the broader forcing uncertainty, the range in our TCR is 0.8–2.6 °C (see Supplementary Table 12).

For Figs 2 and 3 the best estimates of TCR according to the energy-budget calculation equation (1) are shown using each simulation's temperature reconstructions (tas-only, blended and blended-masked) to calculate ΔT with the model-specific ΔF if available, and the multi-model mean ΔF otherwise. For the model TCRs used in Fig. 4, we use the first simulation of each model in the ensemble.

Resultant TCR bias correction. Energy-budget calculations performed on blended-masked simulations were found to consistently underestimate the tas-only value, and so a correction was determined by performing a linear regression of CMIP5 tas-only TCR against blended-masked TCR for the 84 available historical-CMIP5 simulations. This linear regression was constrained to go through zero, and found to have a gradient of 1.24 ± 0.02 (5–95% error, as throughout). To this precision, the same result is determined when using the 54 simulations for which model forcing is available.

This result suggests that an upward revision of 24% is required to accurately represent tas-only TCR, given the result of a calculation using blended-masked temperature series. This 24% value is appropriate for the time period used, and is found to change with time (see Supplementary Information)—it was larger historically and is now tending towards approximately 15% for HadCRUT4 coverage over 1970–2010, or 7–9% for perfect coverage (that is, blending bias only).

Applying TCR bias correction. Having obtained an adjustment factor, α, of 1.24 ± 0.02 from linear regression, we can apply it to the blended-masked energy-budget TCR to estimate the relevant tas-only TCR from:

$$\mathrm{TCR}_{\text{tas-only}} = \alpha \Delta F_{2\times CO_2} \frac{\Delta T_{\text{blended-masked}}}{\Delta F} \quad (4)$$

We use the distributions described above with the HadCRUT4-based ΔT and broader ΔF range with α taken to be a Gaussian with the mean and error determined from the linear regression fit. Each of these distributions is sampled one milion times to derive a one-million-member set of $\mathrm{TCR}_{\text{tas-only}}$ values from which the median and range statistics are extracted. Our blended-masked TCR of $1.34\,^{\circ}$C (0.8–2.6 °C) becomes $1.66\,^{\circ}$C (range 1.0–3.3 °C, see Supplementary Table 12). Alternatively α could be sampled from the distribution of $N = 84$ ratios of tas-only TCR to blended-masked TCR determined previously. Supplementary Table 13 shows that this would result in $1.67\,^{\circ}$C (range 1.0–3.3 °C).