Recommended temperature metrics for carbon budget estimates, model evaluation and climate policy

Author(s):
Tokarska, Katarzyna B.; Schleussner, Carl-Friedrich; Rogelj, Joeri; Stolpe, Martin B.; Matthews, H. Damon; Pfleiderer, Peter; Gillett, Nathan P.

Publication date:
2019-12

Permanent link:
https://doi.org/https://doi.org/10.3929/ethz-b-000382300

Originally published in:
Nature Geoscience 12(12), https://doi.org/10.1038/s41561-019-0493-5

Funding acknowledgement:
- LC-CLA-08-2018 | RIA | Constraining uncertainty of multi decadal climate projections ()

PERSPECTIVE
https://doi.org/10.1038/s41561-019-0493-5

nature geoscience

Recommended temperature metrics for carbon budget estimates, model evaluation and climate policy

Katarzyna B. Tokarska[1,2*], Carl-Friedrich Schleussner[3,4,5], Joeri Rogelj[1,6,7], Martin B. Stolpe[1], H. Damon Matthews[8], Peter Pfleiderer[3,4,5] and Nathan P. Gillett[9]

Recent estimates of the amount of carbon dioxide that can still be emitted while achieving the Paris Agreement temperature goals are larger than previously thought. One potential reason for these larger estimates may be the different temperature metrics used to estimate the observed global mean warming for the historical period, as they affect the size of the remaining carbon budget. Here we explain the reasons behind these remaining carbon budget increases, and discuss how methodological choices of the global mean temperature metric and the reference period influence estimates of the remaining carbon budget. We argue that the choice of the temperature metric should depend on the domain of application. For scientific estimates of total or remaining carbon budgets, globally averaged surface air temperature estimates should be used consistently for the past and the future. However, when used to inform the achievement of the Paris Agreement goal, a temperature metric consistent with the science that was underlying and directly informed the Paris Agreement should be applied. The resulting remaining carbon budgets should be calculated using the appropriate metric or adjusted to reflect these differences among temperature metrics. Transparency and understanding of the implications of such choices are crucial to providing useful information that can bridge the science–policy gap.

Carbon budgets provide a tool for clearly communicating that limiting global warming to a particular level implies a cap on global total CO_2 emissions[1]. Defined as the total amount of CO_2 that can be emitted while keeping global warming below a given level with some probability, carbon budgets emerge from an approximately linear relationship between warming and cumulative CO_2 emissions, known as the transient climate response to cumulative CO_2 emissions (TCRE)[2–5]. The TCRE and related carbon budgets were initially derived under idealized CO_2-only emission scenarios[2]. However, under real-world conditions, several factors complicate the simplicity and clarity of the carbon budget concept. Emissions other than CO_2 (such as methane, soot or sulphate aerosols) also affect both global temperature and the state of carbon sinks (albeit to a smaller extent than CO_2 itself[6–10]), and hence the size of the remaining carbon budget. In addition to CO_2 emissions from fossil fuels (which are well known), CO_2 emissions from other types of land-use change represent a quarter of historical CO_2 emissions: these emissions are difficult to diagnose, and are subject to large uncertainty both in models[11,12] and in estimates derived from historical data based on energy and industry statistics and land-use bookkeeping methods[13]. To further complicate matters, estimates of historical warming since pre-industrial times come with uncertainties due to limited observational coverage[14], instrumental uncertainty and uncertainties associated with constructing long-term temperature datasets. Global warming can also be expressed in different ways — for example, as near-surface air temperatures covering the entire globe or as a combination of sea surface temperatures over open ocean and near-surface air temperatures elsewhere[15,16], averaged over locations where observations are present. Finally, interannual and decadal variability adds further complications[17].

Several recent studies[18–20] and the Special Report on global warming of 1.5 °C (SR1.5)[21] of the Intergovernmental Panel on Climate Change (IPCC) introduced a new approach to estimate the remaining carbon budget. These studies report model-based remaining carbon budgets for the additional warming from today until we reach 1.5 °C or 2 °C of anthropogenic warming. This was a departure from the previous approach of estimating the total carbon budget since pre-industrial times, and then reporting the remaining budget by subtracting emissions to date. The new approach in SR1.5 is a kind of bias correction, as it corrects for any inconsistencies in simulated and observed warming as a function of cumulative emissions over the historical period, and can potentially decrease uncertainties in estimates of the remaining carbon budget, especially for levels of warming relevant to the Paris Agreement[22]. As the remaining carbon budgets for 1.5 °C or 2 °C are small, even adjustments that are limited in absolute terms result in large relative changes. For example, recent estimates of the remaining carbon budget for 1.5 °C are larger by more than a factor of two when compared with those reported in the IPCC Fifth Assessment Report (AR5)[4,23] (see fig. 2 in ref. [24] and their supplementary table 2 for a comprehensive comparison of the remaining carbon budget estimates from different studies). This difference can be partly understood as a result of

[1]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland. [2]School of Geosciences, University of Edinburgh, Edinburgh, UK. [3]Climate Analytics, Berlin, Germany. [4]Integrative Research Institute on Transformations of Human-Environment Systems (IRI THESys), Humboldt-Universität zu Berlin, Berlin, Germany. [5]Potsdam Institute for Climate Impact Research, Potsdam, Germany. [6]Grantham Institute, Imperial College London, London, UK. [7]International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria. [8]Concordia University, Montréal, Quebec, Canada. [9]Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada, University of Victoria, Victoria, British Columbia, Canada. *e-mail: kasia.tokarska@env.ethz.ch

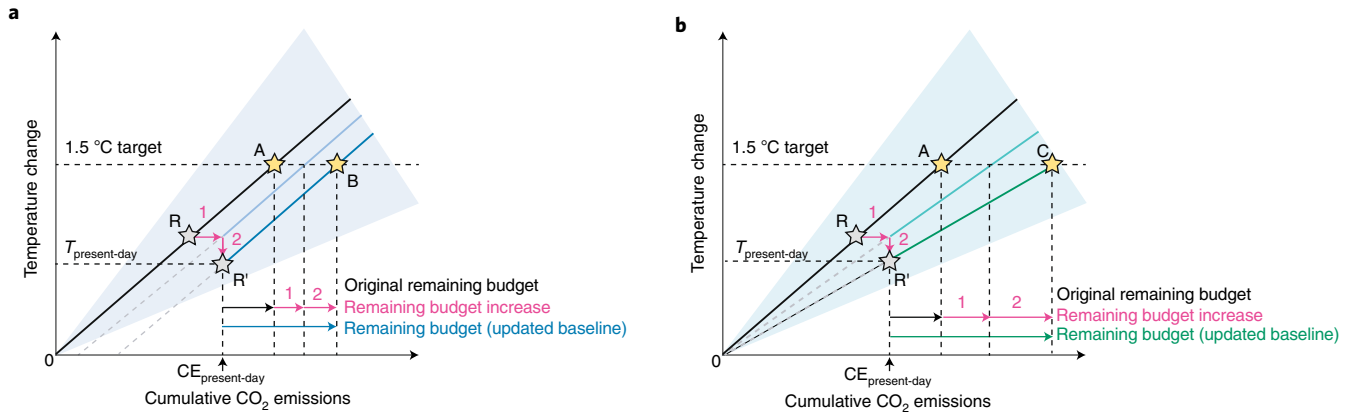


Fig. 1 | Schematic representation of the effects of updating the baseline with respect to the cumulative CO_2 emissions and temperature change. **a**,**b**, Remaining carbon budgets after updating the baseline (**a**) and with scaling of future warming (**b**). Arrow 1 represents the carbon cycle effect (a correction for model biases in historical CO_2 emissions); arrow 2 represents the temperature effect (arising from the differences between modelled and observed warming). The first yellow star (A) indicates the initial carbon budget at the 1.5 °C warming level with the original reference period (R). The second yellow star (B or C) indicates the final (and larger) remaining carbon budget, calculated after updating the baseline to a present-day reference period (R′). The shaded area represents the spread of the relationship between temperature (*T*) and cumulative CO_2 emissions (CE). The present-day level of warming and cumulative CO_2 emissions is indicated by the dashed lines, as labelled, although the figure is for illustrative purposes only.

a higher temperature response to cumulative CO_2 emissions in the Coupled Model Intercomparison Project Phase 5 (CMIP5)[25] models used to inform the AR5 carbon budgets, compared with estimates of historical CO_2 emissions and warming[16,26]. However, recent insights related to uncertainty in the observational temperature record also suggest that part of the difference among carbon budget estimates is related to the method of calculating historical warming that is used in the analysis[27].

Here we explain the reasons why carbon budget estimates expressed relative to a more recent reference period differ from previous ones, and separate these into differences caused by carbon cycle and temperature-driven components. We then clarify how the choice of temperature metric affects the size of remaining carbon budget estimates, and emphasize the need for transparency and clarity about its implications. Finally, we provide recommendations for future estimates of remaining carbon budgets, along with remaining challenges.

Effects underlying adjustments of the baseline

The effect of changing the baseline to a more recent period (from R to R′; Fig. 1), can be separated into carbon cycle effects (arrow 1), and temperature effects (arrow 2). First, the Earth system models (ESMs) that were used to estimate the carbon budgets reported in IPCC AR5, on average, underestimated carbon uptake (by land and ocean carbon sinks) in prescribed CO_2 concentration simulations. These models therefore, on average, estimated lower cumulative CO_2 emissions over the historical period compared with CO_2 emissions estimated from independent fossil fuel use and other data[18,19]. Updating the baseline to account for this carbon cycle bias thus leads to an increase in the remaining carbon budget compared with those reported in IPCC AR5 (arrow 1 in Fig. 1). Second, accounting for a possible difference in warming over the historical period results in a second offset (arrow 2 in Fig. 1). As the global mean temperature has already increased by about 1 °C above pre-industrial levels[28], even minor corrections arising from methodological adjustments or model biases can have a sizeable effect on the remaining 1.5 °C budget.

Remaining carbon budgets are often based on the likely (>66% probability) TCRE range assessed by IPCC AR5[29] of 0.8 to 2.5 °C per 1,000 PgC (where 1 PgC = 3.67 $GtCO_2$). Several recent studies[18,19] that updated the baseline did not alter the resulting TCRE range: that is, they used the same slope for the relationship between temperature and cumulative emissions (TCRE) before and after changing the baseline, as illustrated in Fig. 1a. Another approach would be to adjust the slope of TCRE relationship to align the TCRE with the lower temperature response to emissions implied by updating the baseline to a more recent period. In principle, both carbon cycle and temperature adjustments could lead to changes in the rate of warming as a function of cumulative emissions, as illustrated in Fig. 1b. Whether such an adjustment is warranted depends on the assessment of the validity of extrapolation of historical to future warming as a function of cumulative emissions. Little correlation exists between cumulative emissions at present-day warming and at 1.5 °C across the CMIP5 ensemble[19], probably due to differences in response to non-CO_2 forcing across models. Hence, we would caution against scaling simulated 1.5 °C carbon budgets using the ratio of simulated to observed historical warming as a function of cumulative CO_2 emissions, given the important and uncertain role played by non-CO_2 forcings in historical climate change. Identifying the conditions under which the slope of TCRE would require an adjustment needs further research. Expressing carbon budgets relative to a recent reference period (for example, using the 2006–2015 reference period instead of the pre-industrial baseline) is intended to minimize the effect of uncertainties arising from mismatches between modelled and observed cumulative CO_2 emissions and warming in the historical period. However, such adjustment of the baseline does not involve a correction for the model processes that led to those discrepancies in the historical period.

Temperature metric choices

While the correction for carbon cycle effects is relatively straightforward, attempts to assess consistency between warming estimates based on model output and observations have highlighted questions surrounding the choice of the method used to estimate changes in global mean temperature[30]. One way of expressing the global mean temperature is global mean surface air temperature (here referred to as GSAT), usually estimated in models by calculating the modelled global average SAT — the temperature at about 2 m above the Earth's surface. By contrast, the observed global mean temperature is constructed by combining observational measurements of surface air temperature over land and sea ice (SAT) with sea surface temperature (SST) measurements for open ocean locations. This blended temperature is referred to as GBST, or global mean blended

surface temperature. It is important to note that GBST estimates based on observational measurements do not sample the full globe. Some datasets use statistical infilling techniques to account for this and estimate the global temperature implied by nearly full observational coverage (for example, GISTEMP[31], HadCRUT-CW[32] and Berkeley Earth[33]). Others provide estimates using only data where measurements are available (such as HadCRUT[34]). Estimates that use observations thus reflect the blended (SST + SAT), and in some cases masked (incomplete coverage without statistical infilling) estimates of global mean temperature. Relative to GSAT, both blending and masking in the GBST metric reduce the estimated warming[15,26], and statistical infilling might not always alleviate the masking bias when instrumental coverage is low[14]. Furthermore, the effects of both masking and blending are time-dependent: (1) the observational mask will change over time as the distribution of measurements changes, and (2) the use of SST versus SAT measurements can also change as a result of changing sea-ice coverage, leading (in general) to more open water (and hence SST measurements) over time. This time-dependent blended-masking effect lowers warming since pre-industrial times by about 0.1 °C during the 10 yr average reference period used in SR1.5 (2006–2015). This difference between GBST and GSAT increases with additional warming[16,30].

To estimate remaining carbon budgets relative to a present-day reference period, an estimate of the present-day level of warming is needed to determine the amount of continuing warming until 1.5 °C or any other temperature level would be reached. Given a median estimate of TCRE[4,29], a difference in global mean temperature of 0.1 °C, either as a result of a different temperature limit or as a result of a different estimate of warming to date, would alter carbon budget estimates by about 200 $GtCO_2$ (refs. [21,30]).

Beyond blending-masking adjustments

The multimodel mean GSAT change of the CMIP5 ensemble[25] matched well with GBST observations (HadCRUT4.6)[34] up to the 1986–2005 period, which is the reference period used by IPCC AR5[35] (and see table 1.1 therein). However, the mean of the simulated CMIP5 GSAT warming between 1986–2005 and 2006–2015 (the updated SR1.5 reference period) lies above observation-based estimates. The observed GBST warming in HadCRUT4.6 of 0.22 °C between these periods was lower than the CMIP5 multimodel mean GSAT increase of 0.38 °C and only barely within the range of simulated warming. Accounting for the difference in GBST and GSAT definitions and incomplete coverage of observations can only partly explain this difference (accounting for coverage and blending of SST and SAT reduces modelled warming to 0.33 °C, Fig. 2b).

Several other reasons have been suggested to reconcile the remaining mismatch between the multimodel mean and observations[36]. We identify three main groups of effects that might contribute to the differences between models and observations of GBST (Fig. 2b). First, the SST dataset of HadCRUT4.6, HadSST3, shows a significant cooling bias from around year 2005 onwards, when compared with instrumentally homogeneous SST records from drifting buoys, Argo floats and satellites[37]. This and other biases in the SST record have been recently addressed in HadSST4[38]. However, the increase in GBST between the two reference periods, 1986–2005 and 2006–2015, is virtually unchanged as HadSST4 is warmer during both reference periods than HadSST3 (compared with the pre-industrial baseline). The choice of the SST dataset, therefore, seems to exert only a small influence on the divergence between modelled and observed warming, but uncertainties in the temperature record remain. Second, from the early 1990s, Pacific trade winds intensified, enhancing equatorial upwelling in the central and eastern Pacific. This reduced the SSTs in that region, thereby also reducing the pace of global mean temperature increase[39,40]. These effects of internal variability in the Pacific region lower the observed global mean temperature increase between the two reference periods by roughly 0.08 °C (with a range of −0.03 to −0.20 °C across published estimates; see the Pacific variability effect bars in Fig. 2b). Third, a series of small-to-moderate-magnitude volcanic eruptions have led to an increase in stratospheric aerosols after the year 2004[41,42], which is neglected in CMIP5 model projections. Furthermore, CMIP5 radiative forcing

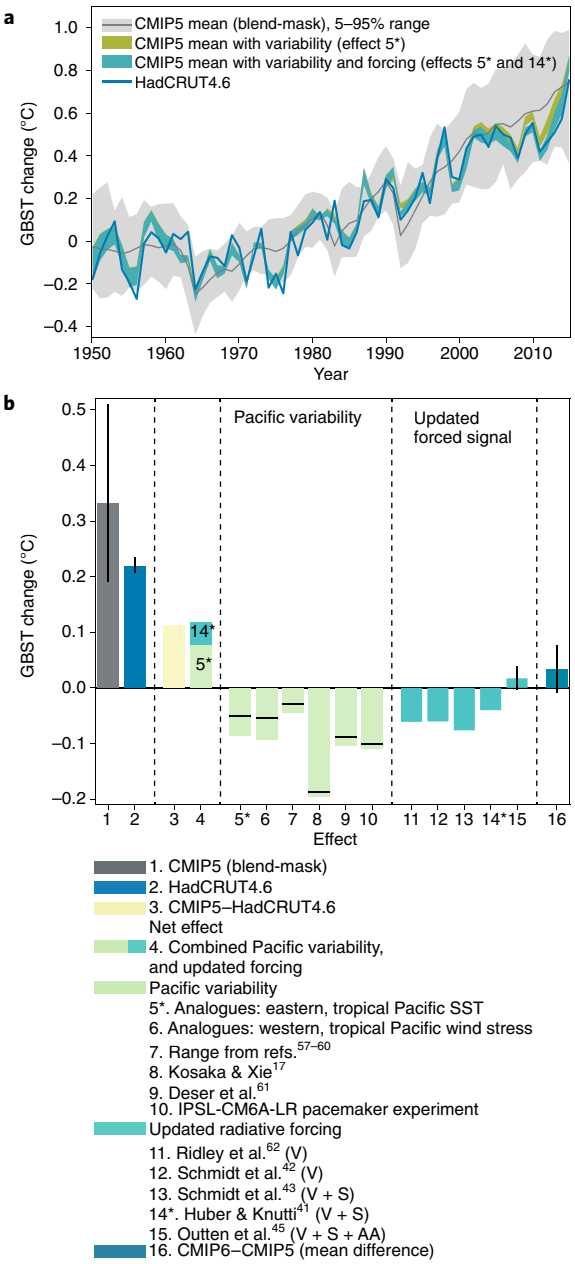


Fig. 2 | Contributions to differences in recent observed and modelled warming. a,**b**, Time series of modelled and observed warming (**a**) and different effects leading to adjustments in observed and modelled GBST (**b**). The length of the bars (horizontal black lines in **b**) shows upper (lower) estimates of the influence of Pacific variability on warming. The spread arises from uncertainty in both observations and the forced signal (effects 5 and 6), from missing years (effects 8 to 10), and reflects the range across four studies (effect 7). Vertical black lines in **b** indicate 5–95% uncertainty ranges. Effects indicated by an asterisk are used for the net effect (bar 4). The global mean temperature base period is 1961–1990 in **a**, and 2006–2015 relative to 1986–2005 in **b** (see Methods for details). AA, anthropogenic aerosols; S, solar; V, volcanic.

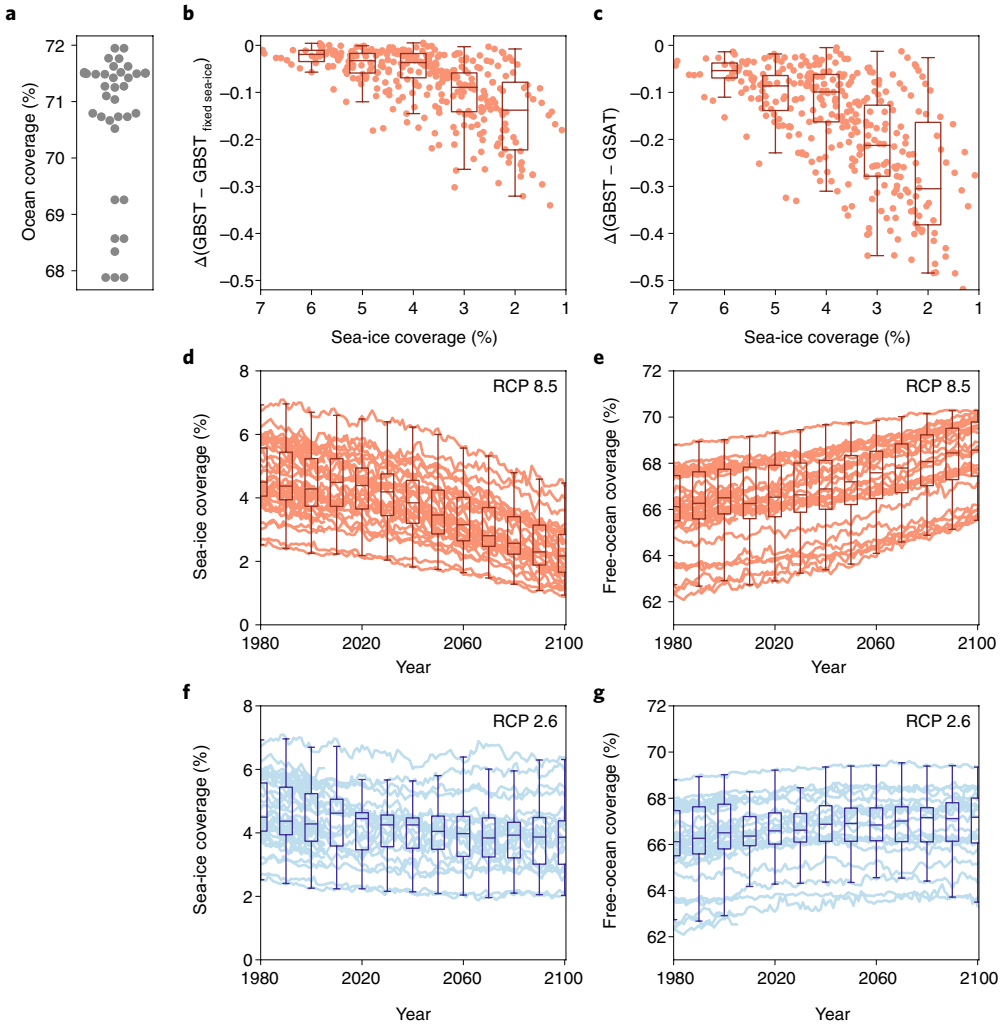


Fig. 3 | Differences in ocean and sea ice coverage in CMIP5 models, and related differences between GBST and GSAT metrics, under different future emission scenarios. a–c, Swarm plots of the time-invariant, constant field defining ocean grid-cells (the sftof CMIP variable, **a**); the sea-ice effect, shown as a difference between GBST and GBST with fixed sea-ice mask (**b**); and the overall blending effect, shown as a difference between GBST and GSAT, as a function of sea-ice coverage (**c**). **d–g**, Time series of the evolution of the sea-ice fraction (**d**) and free-ocean area (**e**) in historical and RCP 8.5 scenarios., and the corresponding sea-ice fraction (**f**) and free-ocean area (**g**) in historical and RCP 2.6 scenarios. In **b** and **c**, box plots are shown for five sea-ice coverage intervals: [6.5–5.5)%, [5.5–4.5)%, [4.5–3.5)%, [3.5–2.5)% and [2.5–1.5)%. In **d–g**, the box plots show the interquartile ranges for 10-yr time slices. Historical and RCP scenarios are from ref. [56].

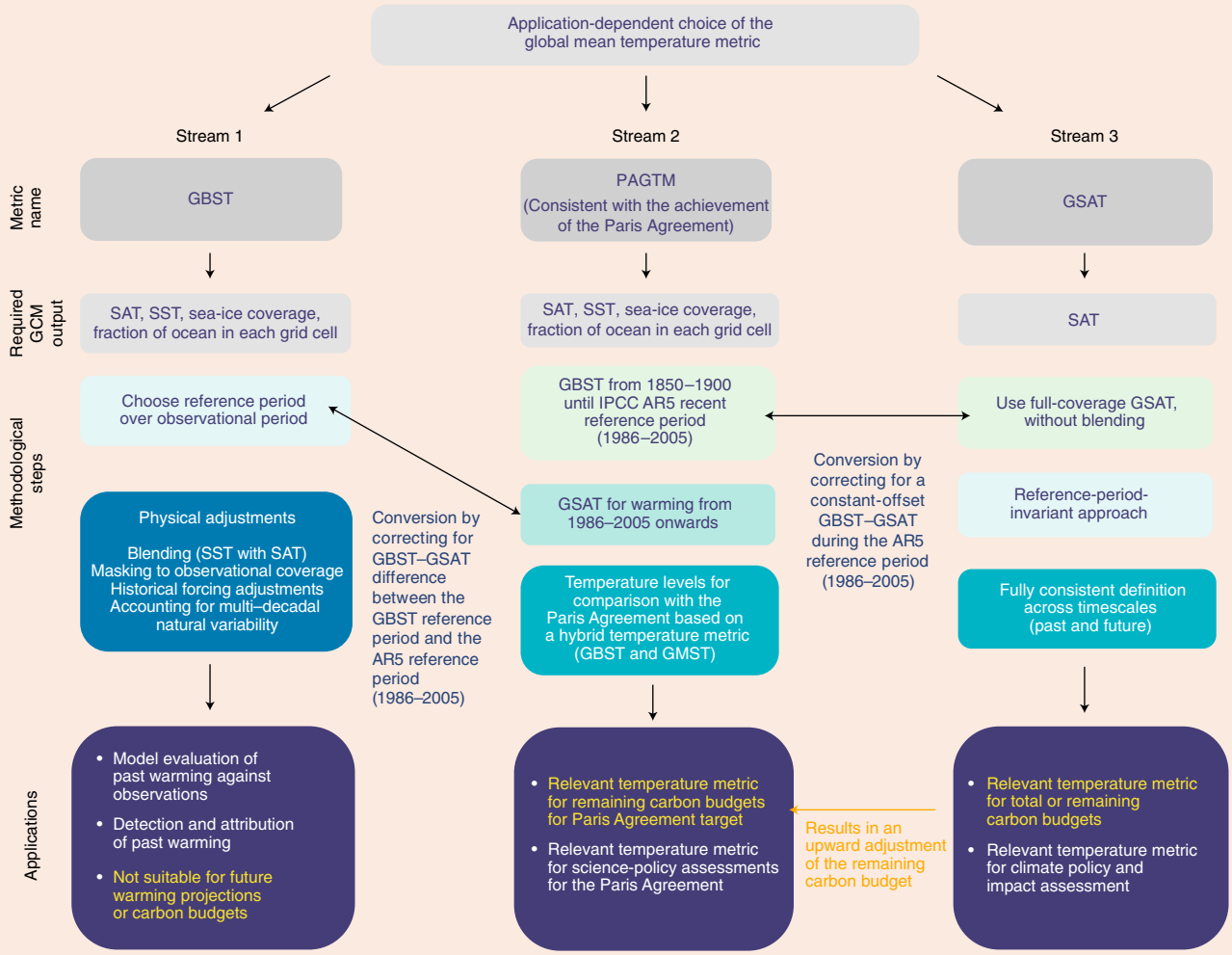
projections also assume that the last solar cycle before 2005 is repeated in the subsequent period. As a result, the assumed recent solar forcing in the model projections is too large when compared with observations[36,41,43]. Correcting models to account for both the updated solar forcing and updated volcanic forcing reduces the modelled global mean temperature increase between the two reference periods, but effects from revised anthropogenic tropospheric aerosols[44] are uncertain and might have reduced[43] or increased the warming[45]. Overall, the assessed studies indicate that warming changes by −0.08 to +0.02 °C from updated forcing between the two reference periods (Updated forced signal bars in Fig. 2b). The CMIP6 models[46] are forced with updated radiative forcings, and while some models indicate reduced warming in the early twenty-first century, explained partly by updated forcing[47], the set of available models simulates slightly more warming between the two reference periods than CMIP5. Major changes in representation of different processes in the models might have led to an increase in climate sensitivity[48], which might increase the warming between the two reference periods[49].

While the strength of the effects is considerably uncertain, and there might be further aspects not considered here, we note that modelled and observed GBST warming between the 1986–2005 and 2006–2015 periods can be fully reconciled within the uncertainty ranges of the different contributing effects (Fig. 2). Moreover, multimodel mean GBST warming in 2006–2015 relative to the 1850–1900 base period is very close to the best observational estimates[35]. This highlights that warming expressed in two different temperature metrics (GBST and GSAT) can be made internally consistent by carefully accounting for various effects, and used to compare models and observations for the historical period.

Application and consistency

Different temperature metrics come with their respective strengths and weaknesses. A GSAT estimate will, by definition, draw from the SAT field everywhere across all models. In contrast, GBST is a composite of land SAT and SST, and GBST estimates depend on the ratio of land and sea ice versus ocean across the Earth' surface. The share of free-ocean coverage differs between models by about 7 percentage

Box 1 | Different choices and recommendations for the use of global mean temperature metrics



The three streams illustrate our recommendations depending on the application domain, with the appropriate use of temperature metrics for carbon budget calculations shown in yellow. Stream 1, using the GBST temperature metric alone, allows a consistent comparison with global mean temperature estimates currently provided by observational temperature products (such as the HadCRUT4.6 dataset[34]). Unless observational products routinely also provide estimates of GSAT, the GBST metric is so far the best choice for applications related to model evaluation of historical warming with the observations or detection and attribution of historical warming[57]. However, this metric of choice for Stream 1 presents challenges when applied to future warming projections (see above discussion of Fig. 3) and is therefore not recommended for calculating remaining carbon budgets (that use future warming projections).

Stream 3, using only the GSAT temperature metric, provides a consistent estimate of global mean temperature increase in model simulations for both the historical period and into the future. Estimating global mean temperature increase based on GSAT alone with full global coverage allows such consistency over time to be achieved. Therefore, we recommend using GSAT as the primary temperature metric for Stream 3 applications, including remaining carbon budget calculations for future levels of warming. This would also ensure consistency with some impact assessment studies that use model simulations from a pre-industrial baseline and use a spatially complete temperature metric across timescales.

Between Stream 1 and Stream 3 lies Stream 2, with applications that intend the assessments of global mean temperature and carbon budgets to be consistent with the achievement of the Paris Agreement target. The Paris Agreement did not specify explicitly which temperature metric applies to the warming levels of 1.5 °C and well below 2 °C. However, this does not mean that the temperature metric is unknown. The temperature goal of the Paris Agreement needs to be read in the context of the accompanying decisions under the UNFCCC and the science, as reflected in the most recent IPCC reports at the time[50]. We therefore propose a Paris Agreement global temperature metric (PAGTM) following the approach applied in the AR5: a hybrid product with GBST until 1986–2005 and GSAT for warming from 1986–2005 onwards.

For a direct comparison of studies using only the GBST metric (Stream 1) with the PAGTM (Stream 2), the difference between the GBST and GSAT metrics over the period between the GBST study's reference period and the AR5 recent reference period (1986–2005) has to be accounted for (indicated by the blue arrow between Stream 1 and Stream 2). For the 2006–2015 reference period, this adjustment is about 0.16 °C and is the difference between modelled GSAT and the observed masked GBST evolution applied to the same model runs (see Methods and table 1.1 in SR1.5).

We do not recommend using GBST for future projections, because this would require implementing model-specific and time-varying adjustments (due to changing sea-ice coverage; see Fig. 3

Box 1 | Different choices and recommendations for the use of global mean temperature metrics (Continued)

and its discussion) to bring these estimates in line with the Paris Agreement-compatible Stream 2 metric. On the other hand, for a direct comparison of results from studies using only the GSAT metric (Stream 3) with the PAGTM (Stream 2), a constant adjustment for the difference between GSAT and GBST during the 1986–2005 period (that is, the AR5 reference period) relative to the 1850–1900 reference period in HadCRUT4.6 needs to be made (indicated by the blue arrow between Stream 3 and Stream 2). In the CMIP5 multimodel mean, this offset is very small (up to about 0.036 °C) compared with the 5–95% uncertainty range of the observational product (HadCRUT4.6 observed warming from 1850–1900 to 1986–2005 is reported to be 0.57–0.66 °C, with a central estimate of 0.6 °C; see table 1.1 of ref. [35]). The transition from Stream 3 to Stream 2 is independent of the chosen baseline or period of interest. For studies using CMIP5, translating results obtained with the full GSAT approach (Stream 3) to the Paris Agreement-consistent metric (Stream 2) results in a constant upward adjustment of the remaining carbon budget by about 80 $GtCO_2$ (for a middle-of-the-range TCRE estimate of 1.65 °C per 1,000 PgC), but can depend on the precise assumptions. For studies using CMIP6 models[46], climate model emulators or other approaches, this adjustment would need to be calculated according to those models.

points (Fig. 3e,g) due to differences in present-day sea ice (Fig. 3d,f) and the land–sea share in the model grid (Fig. 3a, Methods).

The land and sea-ice versus ocean ratio does not only differ among models, but also among various runs from the same model due to internal variability, as well as over time as a result of differences and changes in sea-ice cover. Therefore, the GBST metric is dependent on model, time and even realization within the model ensemble itself. Such differences complicate the comparison of GBST estimates among models or even within ensemble members of the same model. Owing to the combination of these challenges surrounding the GBST temperature metric, the GBST metric is not well-suited for projections of future warming levels (for example, 1.5 °C or 2.0 °C), for which remaining carbon budgets are calculated.

Given the various possible choices regarding methods of calculating global mean temperature rise and their effect on estimates of remaining carbon budgets, we summarize recommended approaches in Box 1. We identify three main streams of application and for each we recommend an appropriate metric for estimating the global mean temperature level and remaining carbon budgets. These streams depend on the purpose of the application: (1) model evaluation of global mean temperature against observations or detection and attribution analysis of global mean temperature (Stream 1); (2) assessments of temperature estimates and carbon budgets for the Paris Agreement goal (Stream 2); and (3) assessing carbon budgets or impacts across time and for future levels of warming with a consistent definition of temperature change (Stream 3).

The differences between temperature metrics such as GBST and GSAT were not thoroughly discussed in the literature available for the AR5, and thus could not be assessed by the IPCC before SR1.5 was published in 2018. It hence cannot be expected that the 2015 Paris Agreement would be specific about the temperature metrics underlying its temperature goal. The same holds for other scientific concepts developed and assessed after the adoption of the Paris Agreement. However, the literature available at the time of AR5 can provide guidance on the metric consistent with the achievement of the Paris Agreement global mean temperature target.

The adoption of the Paris Agreement was informed by a multi-year process reviewing the temperature goal under the UNFCCC. This review process concluded in 2015, adopting a long-term global goal under the Conference of the Parties (COP) that is identical to the Paris Agreement's Article 2.1(a)[22]. The process included a scientific arm, the so-called structured expert dialogue[50], which provided a comprehensive assessment of the impacts of climate change at 1.5 °C and 2 °C based predominantly on AR5. The long-term temperature goal of the Paris Agreement is directly linked to this assessment and thereby the AR5 methodology[51,52]. The Working Groups 1 and 2 of AR5 used GBST from 1850–1900 until the reference period 1986-2005 and GSAT for warming from the reference period onwards. We propose this hybrid temperature metric (PAGTM; Box 1, Stream 2) as being compatible with the Paris Agreement target. Paris Agreement compatibility is linked to the policy context and does not imply that such a hybrid temperature metric (GBST and GSAT) holds any specific scientific merit. As our scientific understanding progresses, new temperature metrics based on either new observational products or new analysis metric will become available, and could be scientifically superior. To avoid misguiding policy by unintentionally shifting baselines, however, we recommend that any assessments aiming to inform the science–policy interface and the Paris Agreement should be expressed in, or at least provide a conversion to, the metric that is consistent with the achievement of the Paris Agreement (that is, the hybrid of GBST and GSAT) presented in Stream 2[24,30,51,52]. This will require the conversion of temperature metrics (either in Stream 1 or Stream 3) to Stream 2 metric, illustrated in Box 1 by the double-headed arrows. Such conversion (from Stream 3 to Stream 2) would lead to upward adjustments of carbon budgets (that is, more allowable CO_2 emissions) (Box 1). This transition to Stream 2 is not exclusive to CMIP5 models, and could be applied, in principle, to any model-based temperature projections or carbon budgets that use the GSAT metric (Stream 3), and aim to report their results in the light of the Paris Agreement[22] (Stream 2).

Remaining challenges for the total carbon budget

Calculating the remaining carbon budget relative to a present-day reference period makes these estimates more accurate, as shown by recent studies[18–20] (see also ref. [24] for a comprehensive summary of recent carbon budget estimates). However, changing the baseline to a more recent period is only a partial solution that does not address the underlying issue of discrepancies between CMIP5 models and observations in the historical period, particularly in their cumulative CO_2 emissions (as the temperature discrepancy between the models and observations can be addressed by comparing models and observations in a like-for-like manner). Moreover, changing the baseline does not help with constraining estimates of the total carbon budget for a given level of warming (that is, including historical and future CO_2 emissions), which may be useful for assessing aspects of historical responsibility for past CO_2 emissions[53].

Implications for the science–policy interface

Using a recent reference period to calculate remaining carbon budgets, rather than first calculating total carbon budgets relative to pre-industrial levels and then subtracting historical emissions, makes these estimates more accurate, providing a physically compelling reason to do so. However, such changes of the baseline to a more recent period also come with political implications that one should be mindful of. Changing the reference period from pre-industrial times to the present day shifts the focus of the study from estimating total carbon budgets and their relevance for the assessment of historical responsibilities and intergenerational or international equity, towards questions of our collective ability to avoid the exceedance of certain warming limits in line with the Paris Agreement.

Given the relevance of carbon budgets for climate policy, we recommend that methodological choices made in their estimation be fully transparent and traceable. Moreover, we recommend that assessments of the progress towards the Paris Agreement goals, including the carbon budgets for 1.5 °C, should provide a comparison with the temperature metric that is consistent with the achievement of the Paris Agreement (that is, PAGTM in Stream 2; Box 1). Owing to different definitions of the temperature metrics discussed in this Perspective, carbon budgets calculated in Stream 2 are expected to be larger than carbon budgets calculated using temperature metrics in Stream 3. Finally, although it may be challenging to constrain all the sources of uncertainty in estimating carbon budgets (for example, refs. [7,9,10,21,54,55]), the large spread in carbon budgets should not be used as an excuse to delay mitigation actions.

Ultimately, more than a decade of research on carbon budgets and the cumulative emissions framework demonstrate very clearly that reaching any global mean warming target that avoids dangerous climate change will require CO_2 emissions to be reduced to net-zero or net-negative[21] levels this century. The sooner this transition to declining emission rates begins, the smaller reliance on net-negative emissions is required in the future[21].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41561-019-0493-5.

Received: 4 May 2019; Accepted: 23 October 2019;
Published online: 28 November 2019

References

1. Zickfeld, K., Eby, M., Matthews, H. D. & Weaver, A. J. Setting cumulative emissions targets to reduce the risk of dangerous climate change. *Proc. Natl Acad. Sci. USA* **106**, 16129–16134 (2009).
2. Matthews, H. D., Gillett, N. P., Stott, P. A. & Zickfeld, K. The proportionality of global warming to cumulative carbon emissions. *Nature* **459**, 829–832 (2009).
3. Gillett, N. P., Arora, V. K., Matthews, D. & Allen, M. R. Constraining the ratio of global warming to cumulative CO_2 emissions using CMIP5 simulations. *J. Clim.* **26**, 6844–6858 (2013).
4. IPCC: Technical Summary. In *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. et al.) 33–115 (Cambridge Univ. Press, 2013).
5. MacDougall, A. H. The transient response to cumulative CO_2 emissions: a review. *Curr. Clim. Change Rep.* **2**, 39–47 (2016).
6. Gillett, N. P. & Matthews, H. D. Accounting for carbon cycle feedbacks in a comparison of the global warming effects of greenhouse gases. *Environ. Res. Lett.* **5**, 034011 (2010).
7. Tokarska, K. B., Gillett, N. P., Arora, V. K., Lee, W. G. & Zickfeld, K. The influence of non-CO_2 forcings on cumulative carbon emissions budgets. *Environ. Res. Lett.* **13**, 034039 (2018).
8. Matthews, H. D. et al. Estimating carbon budgets for ambitious climate targets. *Curr. Clim. Change Rep.* **3**, 69–77 (2017).
9. MacDougall, A. H., Zickfeld, K., Knutti, R. & Matthews, H. D. Sensitivity of carbon budgets to permafrost carbon feedbacks and non-CO_2 forcings. *Environ. Res. Lett.* **10**, 125003 (2015).
10. Mengis, N., Partanen, A.-I., Jalbert, J. & Matthews, H. D. 1.5 °C carbon budget dependent on carbon cycle uncertainty and future non-CO_2 forcing. *Sci. Rep.* **8**, 5831 (2018).
11. Arora, V. K. et al. Carbon–concentration and carbon–climate feedbacks in CMIP5 Earth system models. *J. Clim.* **26**, 5289–5314 (2013).
12. Arora, V. K. et al. Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases. *Geophys. Res. Lett.* **38**, L05805 (2011).
13. Quéré, C. L. et al. Global carbon budget 2017. *Earth Syst. Sci. Data* **10**, 405–448 (2018).
14. Benestad, R. E., Erlandsen, H. B., Mezghani, A. & Parding, K. M. Geographical distribution of thermometers gives the appearance of lower historical global warming. *Geophys. Res. Lett.* **46**, 7654–7662 (2019).
15. 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).
16. Schurer, A. P. et al. Interpretations of the Paris climate target. *Nat. Geosci.* **11**, 220–221 (2018).
17. Kosaka, Y. & Xie, S.-P. The tropical Pacific as a key pacemaker of the variable rates of global warming. *Nat. Geosci.* **9**, 4–6 (2016).
18. Millar, R. J. et al. Emission budgets and pathways consistent with limiting warming to 1.5 °C. *Nat. Geosci.* **10**, 741–747 (2017).
19. Tokarska, K. B. & Gillett, N. P. Cumulative carbon emissions budgets consistent with 1.5 °C global warming. *Nat. Clim. Change* **8**, 296–299 (2018).
20. Goodwin, P. et al. Pathways to 1.5 °C and 2 °C warming based on observational and geological constraints. *Nat. Geosci.* **11**, 102–107 (2018).
21. Rogelj, J. et al. in *Global Warming of 1.5 °C* (eds Masson-Delmotte, V. et al.) Ch. 2 (IPCC, 2018).
22. *Adoption of the Paris Agreement* FCCC/CP/2015/L.9/Rev.1 (UNFCCC, 2015).
23. IPCC *Climate Change 2014: Synthesis Report* (eds Core Writing Team, Pachauri, R. K. & Meyer, L. A.) (IPCC, 2014).
24. Rogelj, J., Forster, P. M., Kriegler, E., Smith, C. J. & Séférian, R. Estimating and tracking the remaining carbon budget for stringent climate targets. *Nature* **571**, 335–342 (2019).
25. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93**, 485–498 (2011).
26. Richardson, M., Cowtan, K., Hawkins, E. & Stolpe, M. B. Reconciled climate response estimates from climate models and the energy budget of Earth. *Nat. Clim. Change* **6**, 931–935 (2016).
27. Richardson, M., Cowtan, K. & Millar, R. J. Global temperature definition affects achievement of long-term climate goals. *Environ. Res. Lett.* **13**, 054004 (2018).
28. Hawkins, E. et al. Estimating changes in global temperature since the preindustrial period. *Bull. Am. Meteorol. Soc.* **98**, 1841–1856 (2017).
29. IPCC: Summary for Policymakers. In *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. et al.) (Cambridge Univ. Press, 2013).
30. Pfleiderer, P., Schleussner, C.-F., Mengel, M. & Rogelj, J. Global mean temperature indicators linked to warming levels avoiding climate risks. *Environ. Res. Lett.* **13**, 064015 (2018).
31. Lenssen, N. J. L. et al. Improvements in the GISTEMP Uncertainty Model. *J. Geophys. Res. Atmos.* **124**, 6307–6326 (2019).
32. 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).
33. Rohde, R. et al. A new estimate of the average earth surface land temperature spanning 1753 to 2011. *Geoinform. Geostat.* **1**, https://doi.org/10.4172/2327-4581.1000101 (2013).
34. 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**, D08101 (2012).
35. Allen, M. R. et al. in *Global Warming of 1.5 °C* (eds Masson-Delmotte, V. et al.) Ch. 1 (IPCC, 2018).
36. Medhaug, I., Stolpe, M. B., Fischer, E. M. & Knutti, R. Reconciling controversies about the 'global warming hiatus'. *Nature* **545**, 41–47 (2017).
37. Hausfather, Z. et al. Assessing recent warming using instrumentally homogeneous sea surface temperature records. *Sci. Adv.* **3**, e1601207 (2017).
38. Kennedy, J. J., Rayner, N. A., Atkinson, C. P. & Killick, R. E. An ensemble data set of sea surface temperature change from 1850: the Met Office Hadley Centre HadSST.4.0.0.0 data set. *J. Geophys. Res. Atmos.* **124**, 7719–7763 (2019).
39. Kosaka, Y. & Xie, S.-P. Recent global-warming hiatus tied to equatorial Pacific surface cooling. *Nature* **501**, 403–407 (2013).
40. England, M. H. et al. Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus. *Nat. Clim. Change* **4**, 222–227 (2014).
41. Huber, M. & Knutti, R. Natural variability, radiative forcing and climate response in the recent hiatus reconciled. *Nat. Geosci.* **7**, 651–656 (2014).
42. Schmidt, A. et al. Volcanic radiative forcing from 1979 to 2015. *J. Geophys. Res. Atmos.* **123**, 12491–12508 (2018).
43. Schmidt, G. A., Shindell, D. T. & Tsigaridis, K. Reconciling warming trends. *Nat. Geosci.* **7**, 158–160 (2014).
44. Myhre, G. et al. Multi-model simulations of aerosol and ozone radiative forcing due to anthropogenic emission changes during the period 1990–2015. *Atmos. Chem. Phys.* **17**, 2709–2720 (2017).
45. Outten, S., Thorne, P., Bethke, I. & Seland, Ø. Investigating the recent apparent hiatus in surface temperature increases: 1. Construction of two 30-member Earth System Model ensembles. *J. Geophys. Res. Atmos.* **120**, 8575–8596 (2015).
46. Eyring, V. et al. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* **9**, 1937–1958 (2016).

47. Volodin, E. & Gritsun, A. Simulation of observed climate changes in 1850–2014 with climate model INM-CM5. *Earth Syst. Dynam.* **9**, 1235–1242 (2018).
48. The CMIP6 landscape. *Nat. Clim. Change* **9**, 727 (2019).
49. Jiménez-de-la-Cuesta, D. & Mauritsen, T. Emergent constraints on Earth's transient and equilibrium response to doubled CO_2 from post-1970s global warming. *Nat. Geosci.* **12**, 902–905 (2019).
50. *Report on the Structured Expert Dialogue on the 2013–2015 Review* FCCC/SB/2015/INF.1 (UNFCCC, 2015).
51. Rogelj, J., Schleussner, C.-F. & Hare, W. Getting it right matters: temperature goal interpretations in geoscience research. *Geophys. Res. Lett.* **44**, 10662–10665 (2017).
52. Schleussner, C.-F. et al. Science and policy characteristics of the Paris Agreement temperature goal. *Nat. Clim. Change* **6**, 827–835 (2016).
53. Matthews, H. D. et al. National contributions to observed global warming. *Environ. Res. Lett.* **9**, 014010 (2014).
54. Lowe, J. A. & Bernie, D. The impact of Earth system feedbacks on carbon budgets and climate response. *Phil. Trans. R. Soc. A* **376**, 20170263 (2018).
55. Comyn-Platt, E. et al. Carbon budgets for 1.5 and 2 °C targets lowered by natural wetland and permafrost feedbacks. *Nat. Geosci.* **11**, 568–573 (2018).
56. Meinshausen, M. et al. The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change* **109**, 213–241 (2011).
57. Bindoff, N. L. et al. in *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. et al.) Ch. 10 (IPCC, Cambridge Univ. Press, 2013).

Methods

We make use of CMIP5 and CMIP6 models[25,46], as detailed in each subsection regarding Figs. 2 and 3. The sets of models used in Figs. 2 and 3 are different, as described below.

Contributions of different effects to the observed and modelled warming. Figure 2 summarizes the effects that explain why observed and modelled global mean temperature might differ between the two reference periods (1986–2005 and 2006–2015). The CMIP5 ensemble is that of ref. [15] and consists of 38 models with 86 realizations (bcc-csm1-1-m and CMCC-CESM show unphysical features in the difference between GBST and GSAT in the late twenty-first century and were excluded in ref. [15], but are included here as we are interested in the period up to 2015). We first average the ensemble members of each model to obtain the multimodel mean. Uncertainties in the observed GBST arising from SSTs are assessed by comparing the warming of the HadCRUT-CW dataset[32] when it is constructed using three different SST datasets, but is otherwise unchanged: HadSST3[58], COBE-SST2[59] and HadSST4. With both HadSST3 and HadSST4, the GBST increase between 1986–2005 and 2006–2015 is 0.26 °C, whereas it is 0.28 °C with COBE-SST2. The choice of the SST dataset has therefore only a relatively small influence on the GBST increase. Using GISTEMPv4 as an alternative GBST dataset produces a warming of 0.26 °C between the two reference periods. Bar 2 of Fig. 2b displays the 5–95% range across the 100-member HadCRUT4.6 ensemble and indicates data uncertainty.

We use variability analogues[41] to quantify how Pacific variability altered the warming between the two reference periods (M. B. Stolpe et al., manuscript in preparation). Therefore, we search for periods from 33 CMIP5 and 18 CMIP6 control simulations (29,950 model years in total) in which the modelled variability agrees with the observed variability (based on the root mean squared error between the time series over a period of 40 months, and we keep the 20 best matching analogues within each period). We standardize both the observed and modelled variability time series. The GSAT anomaly in the analogues is a measure of the contribution of the observed Pacific variability to the observed GBST evolution. To describe internal variability, we take area-weighted SSTs in the Nino3.4 region (5° S–5° N, 170° W–120° W) and from a larger region in the central to eastern tropical Pacific (15° N–15° S, 180°–90° W) using two spatially interpolated SST datasets, ERSSTv5[60] and COBE-SST2. The SSTs in these regions also include a forced signal that we remove before selecting the analogues. We estimate the forced signal using the method of ref. [61] (that is, a linear trend over observed tropical ocean SST from 1962 to 2011), and the ensemble means of the CMIP5 and CMIP6 models for the respective regions. Figure 2 shows the range across the resulting 12 combinations of region, SST dataset and forced signal correction. We also select analogues on the basis of the observed zonal wind stress in the western tropical Pacific over two regions (180°–150° W, 6° S–6° N, and 150° E–150° W, 10° S–10° N) from 49 control simulations (31 CMIP5 and 18 CMIP6 models with 29,084 years). These regions are based on refs. [40,62]. We take observed wind stress from two reanalyses, ERA-Interim[63] and MERRA2[64] and in Fig. 2b we display the range across the resulting four wind-stress estimates.

The contribution of tropical Pacific variability to GBST is quantified in refs. [65–68] using multiple linear regression. These studies describe tropical Pacific variability using the Nino3.4 and Multivariate El Niño/Southern Oscillation (ENSO) indices[69,70]. We use an updated and modified version of ref. [67], where a second ENSO lag term was added. Other works[17,71] and the simulations with IPSL-CM6A-LR that follow the 'Decadal Climate Prediction Project' protocol of ref. [72] quantify the Pacific contribution to GSAT as the difference between two climate model experiments. In a first experiment, a freely evolving initial condition ensemble is forced with historical radiative forcings, while a second experiment is driven by the same radiative forcings, but the modelled central to eastern tropical Pacific SSTs are nudged towards observed anomalies. These so-called pacemaker experiments end in 2013 and 2014, respectively. We use the variability analogues to approximately extend the estimates to 2015. Alternatively, we assume that the complete year-to-year HadCRUT4.6 GBST variability during the missing years was caused by Pacific variability. Figure 2b shows the spread arising from these two assumptions. The pacemaker experiments indicate a larger Pacific induced global temperature decrease between the two reference periods than studies using multiple regression. This could be related to a timescale dependence of the imprint of tropical Pacific variability on GSAT, which in climate model simulations is larger on a decadal than on an interannual time scale[17,73]. Regression models constructed on interannual variability might underestimate the Pacific influence on a decadal timescale[73]. Whether and how the forced signal is removed from tropical Pacific SSTs plays a role. If it is not fully removed, the cooling from internal variability is underestimated, and vice versa. The spread in the Pacific contribution to the GSAT change between the two reference periods is also substantial across the pacemaker studies (Fig. 2b, effects 8–10) and this is probably related to how strongly the tropical Pacific variability projects onto higher latitudes on a decadal timescale[73].

We use the forcing corrections of refs. [41–43,45,74]. For ref. [41], we combine the forcing corrections of updated solar variability (based on the dataset by the Physikalisch-Meteorologisches Observatorium Davos; PMOD) and of stratospheric aerosols (not including their correction for background stratospheric aerosols from 1960 to 1990); ref. [43] and ref. [45] also estimate the effects of updated well-mixed greenhouse gas concentrations, which is very small in both studies, and human-made tropospheric aerosols. While ref. [43] finds underestimated aerosol cooling during the first decade of the twenty-first century, ref. [45] argues for overestimated aerosol cooling, presumably related to primary organic matter aerosols. For the ref. [43] forcing correction, we show only the GSAT influence of updated solar and volcanic forcing. In refs. [42,43], the radiative forcing of the Mount Pinatubo eruption is downgraded, making the 1986–2005 period warmer and thereby also decreasing the GSAT increase between the two reference periods. On the other hand, ref. [45] increases volcanic forcing during the early reference more than from 2006 onwards, and thus increases the simulated warming between the two reference periods. This and the reduced cooling from tropospheric aerosols lead to slightly increased warming between the two reference periods compared with the control experiment with CMIP5 forcings in ref. [45]. Different from the other forcing corrections, some internal variability is left in the estimate of ref. [45] as it is the difference between two 30-member climate model ensembles. Effect 15 in Fig. 2b shows the difference between the two ensemble means (with the 90% confidence interval using data until 2012) and the central estimate is from assuming that the anomaly comes back to zero by 2015. Furthermore, we display the volcanic aerosol GSAT corrections of ref. [74] and ref. [42], which account for volcanic aerosols in the lowermost stratosphere below 15 km that are not included in the other stratospheric aerosol corrections (for ref. [74] we use the AERONET mean GSAT estimate that we digitized from their fig. 3b). Except for ref. [42], which fully covers the period 2006–2015, the other studies include data until 2012–2013 and for the missing years we assume that the GSAT anomaly of stratospheric aerosols remains constant and that the adjustment from updated solar irradiance comes back to zero anomaly by 2015. Not all forcing corrections fully cover the early 1986–2005 reference period, and for missing years we assume a zero GSAT anomaly.

The CMIP6 models are forced with updated radiative forcing until 2014 (we extrapolate until 2015 by repeating the warming of the previous year), but as model physics also changed, and the set of models is not the same, the difference in the GSAT increase compared with CMIP5 cannot solely be attributed to changes in radiative forcings. The CMIP6 ensemble of historical simulations consists of (number of members in parentheses): BCC-CSM2-MR (3), BCC-ESM1 (3), CAMS-CSM1-0 (2), CanESM5 (50), CESM2 (11), CESM2-WACCM (3), CNRM-CM6-1 (10), CNRM-ESM2-1 (5), E3SM-1-0 (5), EC-Earth3 (6), EC-Earth3-Veg (4), FGOALS-g3 (3), GFDL-CM4 (1), GFDL-ESM4 (1), GISS-E2-1-G (20), GISS-E2-1-G-CC (1), GISS-E2-1-H (10), HadGEM3-GC31-LL (4), IPSL-CM6A-LR (32), MIROC6 (10), MIROC-ES2L (3), MRI-ESM2-0 (5), NESM3 (5), NorCPM1 (30), NorESM2-LM (1), SAM0-UNICON (1) and UKESM1-0-LL (6). We compare the CMIP6 ensemble mean with the CMIP5 mean for GSAT (with RCP 8.5 from 2006 onwards) and estimate the uncertainty of the difference in the ensemble means using Welch's *t*-test (Fig. 2b shows the 90% confidence interval). Overall, the warming simulated by the CMIP6 ensemble mean between the two reference periods is slightly higher than that of the CMIP5 ensemble (Fig. 2b).

For the net effect (bar 4 in Fig. 2b), we combine the Pacific variability estimated by analogues from the central to eastern tropical Pacific with the CMIP5 mean removed and averaged across ERSSTv5 and COBE-SST2 (for Fig. 2a we show the range across all combinations of SST-based analogues), and the updated radiative forcing of ref. [41]. We stress, however, that this only one possible combination and that the individual components are uncertain. There might be further effects not accounted for by our analysis (such as Atlantic multidecadal variability, but that effect on GSAT is probably small during the period examined[75]). The forcing and variability corrections are estimated for GSAT and not GBST, which might also cause a small bias.

Differences in the ocean and sea-ice coverage, and related differences between GBST and GSAT. Figure 3 displays the global free-ocean fraction and the influence of changes in sea-ice coverage on the difference between GBST and GSAT. Free-ocean coverage is the area fraction of ocean cells in each model minus the sea-ice coverage. While the number of ocean cells is constant, sea-ice coverage declines with global warming. In the computation of GBST, SAT are taken over land and sea ice and surface ocean temperatures are used for ocean cells. In grid cells partially covered by sea ice, surface air and ocean temperatures are blended respective to the sea-ice fraction. We follow ref. [15] for the computation of GBST and GBST with fixed sea ice. Fixed sea-ice coverage is based on monthly sea-ice coverage between 1961 and 2014: cells that have not been covered in that period (and in the respective month) are considered as sea-ice free, the remaining cells are considered as fully covered by sea ice. Figure 3 includes 28 CMIP5 models: ACCESS1-0, ACCESS1-3, CCSM4, CESM1-BGC, CMCC-CMS, CMCC-CM, CSIRO-Mk3-6-0, CanESM2, GFDL-CM3, GFDL-ESM2G, GFDL-ESM2M, GISS-E2-H-CC, GISS-E2-H, GISS-E2-R-CC, GISS-E2-R, HadGEM2-CC, HadGEM2-ES, IPSL-CM5A-LR, IPSL-CM5A-MR, IPSL-CM5B-LR, MIROC-ESM-CHEM, MIROC-ESM, MIROC5, MPI-ESM-LR, MPI-ESM-MR, MRI-CGCM3, NorESM1-ME and NorESM1-M.

Transitions between GBST and the Paris Agreement-consistent method. The magnitude of the first arrow in Box 1 between Stream 1 and Stream 2 (that is, the difference between the GBST and Paris-consistent temperature method for 2006-2015) is based on the values from table 1.1 of SR1.5[35]. It is calculated as the

difference between the CMIP5 GSAT for the period 1850–1900 to 2006–2015 and the CMIP5 GSAT for the period 1850–1900 to 1986–2005, minus the difference between HadCRUT4.6 for the period 1850–1900 to 2006–2015 and HadCRUT4.6 for the period 1850–1900 to 1986–2005. Using values from table 1.1 of ref. [35] results in: (0.99 − 0.62) − (0.84 − 0.60) = 0.13 °C, or more precisely, taking the values in brackets directly from column 4 (that is, the difference between the CMIP5 GSAT and the HadCRUT4.6 GBST change from 1986–2005 to 2006–2015) of table 1.1 results in: 0.38 − 0.22 = 0.16 °C (note that the difference between these two estimates comes from rounding).

Data availability

The Cowtan and Way[32] GBST datasets with different SST reconstructions are available at: https://www-users.york.ac.uk/~kdc3/papers/coverage2013/. The HadCRUT4.6 data is available at: https://www.metoffice.gov.uk/hadobs/hadcrut4/. GISTEMPv4 is available at: https://data.giss.nasa.gov/gistemp/. COBE-SST2 and ERSSTv5 data is provided by the NOAA/OAR/ESRL PSD via https://www.esrl.noaa.gov/psd/data/gridded/. ERA-Interim is available at: https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim. MERRA2 was downloaded from: https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/. CMIP5 and CMIP6 model output is available at: http://pcmdi9.llnl.gov/. CESM1 pacemaker experiments are available at: https://www.earthsystemgrid.org/.

References

58. Kennedy, J. J., Rayner, N. A., Smith, R. O., Parker, D. E. & Saunby, M. Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization. *J. Geophys. Res. Atmos.* **116**, D14103 (2011).
59. Hirahara, S., Ishii, M. & Fukuda, Y. Centennial-scale sea surface temperature analysis and its uncertainty. *J. Clim.* **27**, 57–75 (2013).
60. Huang, B. et al. Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5): upgrades, validations, and intercomparisons. *J. Clim.* **30**, 8179–8205 (2017).
61. Turkington, T., Timbal, B. & Rahmat, R. The impact of global warming on sea surface temperature based El Niño–Southern Oscillation monitoring indices. *Int. J. Climatol.* **39**, 1092–1103 (2019).
62. Saenko, O. A., Fyfe, J. C., Swart, N. C., Lee, W. G. & England, M. H. Influence of tropical wind on global temperature from months to decades. *Clim. Dynam.* **47**, 2193–2203 (2016).
63. Dee, D. P. et al. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**, 553–597 (2011).
64. Gelaro, R. et al. The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *J. Clim.* **30**, 5419–5454 (2017).
65. Folland, C. K., Boucher, O., Colman, A. & Parker, D. E. Causes of irregularities in trends of global mean surface temperature since the late 19th century. *Sci. Adv.* **4**, eaao5297 (2018).
66. Lean, J. L. Observation-based detection and attribution of 21st century climate change. *WIREs Clim. Change* **9**, e511 (2018).
67. Foster, G. & Rahmstorf, S. Global temperature evolution 1979–2010. *Environ. Res. Lett.* **6**, 044022 (2011).
68. Rypdal, K. The life and death of the recent global surface warming hiatus parsimoniously explained. *Climate* **6**, 64 (2018).
69. Trenberth, K. E. The definition of El Niño. *Bull. Am. Meteorol. Soc.* **78**, 2771–2778 (1997).
70. Wolter, K. & Timlin, M. S. El Nino/Southern Oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index (MEI.ext). *Int. J. Climatol.* **31**, 1074–1087 (2011).
71. Deser, C., Guo, R. & Lehner, F. The relative contributions of tropical Pacific sea surface temperatures and atmospheric internal variability to the recent global warming hiatus. *Geophys. Res. Lett.* **44**, 7945–7954 (2017).
72. Boer, G. J. et al. The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. *Geosci. Model Dev.* **9**, 3751–3777 (2016).
73. Wang, C.-Y., Xie, S.-P., Kosaka, Y., Liu, Q. & Zheng, X.-T. Global influence of tropical Pacific variability with implications for global warming slowdown. *J. Clim.* **30**, 2679–2695 (2017).
74. Ridley, D. A. et al. Total volcanic stratospheric aerosol optical depths and implications for global climate change. *Geophys. Res. Lett.* **41**, 7763–7769 (2014).
75. Haustein, K. et al. A limited role for unforced internal variability in twentieth-century warming. *J. Clim.* **32**, 4893–4917 (2019).

Acknowledgements

We are thankful for the discussions at the workshop on carbon budgets, co-organized by J.R. and attended by K.B.T., N.P.G., H.M.D. and J.R., with the support of the Global Carbon Project, the CRESCENDO project, Stanford University, the University of Melbourne and Simon Fraser University. We thank E. Bush and A. Schurer for helpful insights. We thank K. Cowtan for providing data and the computer code for blending SAT and SST estimates. We thank I. Bethke, G. Foster, C. K. Folland, M. Huber, Y. Kosaka, J. L. Lean, K. Rypdal and A. Schmidt for providing data used in Fig. 2. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling groups for producing and making available their model output. The US Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support for CMIP and led the development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. K.B.T, C-F.S. and J.R. were supported by the European Union's Horizon 2020 research and innovation programme under grant agreement no. 820829 (CONSTRAIN project). K.B.T. was also supported by the UK NERC-funded SMURPHs project (grant no. NE/N006143/1). C.F.S. and P.P. acknowledge support from the German Federal Ministry of Education and Research (grant no. 01LN1711A).

Author contributions

C.-F.S. initiated the study. K.B.T. wrote the manuscript with substantial input from C.-F.S., J.R., M.B.S., H.D.M. and N.P.G. Figure 2 was produced by M.B.S., Fig. 3 by P.P. and the remaining figures by K.B.T., with suggestions from other authors. All authors participated in manuscript editing and revisions.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence should be addressed to K.B.T.

Peer review information Primary Handling Editor(s): Heike Langenberg.