https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Open Access Earth System Science Data Discussions

How well can we quantify when 1.5°C of global warming has been exceeded?

Peter W. Thorne[1], John M. Nicklas[2], John J. Kennedy[3], Bruce Calvert[4], Baylor Fox-Kemper[2], Mark T. Richardson[5], Adrian Simmons[6], Ed Hawkins[7], Robert Rohde[8], Kathryn Cowtan[9], Nerilie J. Abram[10], Axel Andersson[11], Simon Noone[1], Phillipe Marbaix[12, 55], Nathan Lenssen[13] , Dirk Olonscheck[14], Tristram Walsh[15], Stephen Outten[16], Ingo Bethke[17], Bjorn H. Samset[18], Chris Smith[19,57], Anna Pirani[20], Jan Fuglestvedt[18], Lavanya Rajamani[21], Richard A. Betts[22], Elizabeth C. Kent[23], Blair Trewin[24], Colin Morice[22], Tim Osborn[25], Samantha N Burgess[6], Oliver Geden[26], Andrew Parnell[27], Piers M. Forster[28], Chris Hewitt[23,29], Zeke Hausfather[30], Valerie Masson-Delmotte[31], Jochem Marotzke[14], Nathan Gillett[32], Sonia I. Seneviratne[33], Gavin A. Schmidt[34], Duo Chan[35], Stefan Brönnimann[36], Andy Reisinger[37], Matthew Menne[38], Maisa Rojas Corradi[39], Christopher Kadow[40], Peter Huybers[41], David B. Stephenson[42], Emily Wallis[25], Joeri Rogelj[43], Andrew Schurer[44], Karen McKinnon[45], Panmao Zhai[46], Fatima Driouech[47], Wilfran Moufouma Okia[29], Saeed Vazifehkhah[29], Sophie Szopa[48], Christopher J. Merchant[49], Shoji Hirahara[50], Masayoshi Ishii[50,58], Francois A. Engelbrecht[51], Qingxiang Li[52], June-Yi Lee[53], Alex J. Cannon[54], Christophe Cassou[56], Karina von Schuckmann[59], Amir H. Delju[29], Ellie Murtagh[1]

[1] ICARUS Climate Research Centre, Maynooth University, Maynooth, Co. Kildare, Ireland
[2] Department of Earth, Environmental, and Planetary Sciences, Brown University, Providence, RI, USA
[3] Independent researcher, France
[4] Independent Researcher, Ottawa, Canada
[5] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
[6] European Centre for Medium-Range Weather Forecasts, Reading, UK.
[7] National Centre for Atmospheric Science, Department of Meteorology, University of Reading, Reading, UK.
[8] Berkeley Earth, Berkeley, CA, USA
[9] Department of Chemistry, University of York, York, UK
[10] ARC Centre of Excellence for the Weather of the 21st Century, Research School of Earth Sciences, The Australian National University, Canberra ACT 2601, Australia.
[11] German Meteorological Service, Hamburg, Germany
[12] Division of Thermodynamics and Fluid Dynamics, Université catholique de Louvain, Louvain-la-Neuve, Belgium
[13] Colorado School of Mines, Golden Colorado, USA & NSF National Center for Atmospheric Research, Boulder, Colorado, USA
[14] Max Planck Institute for Meteorology, Hamburg, Germany
[15] Environmental Change Institute, University of Oxford, Oxford, UK
[16] Nansen Environmental and Remote Sensing Center, Bjerknes Centre for Climate Research, Bergen, Norway https://orcid.org/0000-0002-4883-611X
[17] University of Bergen, Bjerknes Centre for Climate Research, Bergen, Norway
[18] CICERO Center for International Climate Research, Oslo, Norway
[19] Department of Water and Climate, Vrije Universiteit Brussel, Brussels, Belgium
[20] Euro-Mediterranean Centre on Climate Change (CMCC Foundation), Venice, Italy
[21] Faculty of Law, University of Oxford
[22] Met Office Hadley Centre, Fitzroy Road, Exeter, UK
[23] National Oceanography Centre, Southampton, UK
[24] Bureau of Meteorology, Melbourne, Australia
[25] Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich, UK

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

[26] German Institute for International and Security Affairs, Berlin, Germany
[27] School of Mathematics and Statistics, University College Dublin, Dublin, Ireland
[28] Priestley Centre for Climate Futures, University of Leeds, Leeds, UK
[29] WMO Secretariat, World Meteorological Organization, Geneva, Switzerland
[30] Stripe Inc., South San Francisco, CA, USA and Berkeley Earth, Berkeley, CA, USA
[31] Laboratoire des Sciences du Climat et de l'Environnement (UMR 8212 CEA-CNRS-UVSQ), Institut Pierre Simon Laplace, Université Paris-Saclay, France
[32] Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada, Victoria, BC, Canada.
[33] Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland
[34] NASA Goddard Institute for Space Studies, New York, NY, USA
[35] School of Ocean and Earth Science, University of Southampton, UK
[36] Institute of Geography and Oeschger Centre for Climate Change Research, University of Bern, Switzerland
[37] Institute for Climate, Energy and Disaster Solutions, Australian National University, Acton ACT, Australia
[38] NOAA's National Centers for Environmental Information, 151 Patton Avenue, Asheville, NC, USA
[39] Department of Geophysics, Faculty of Engineering, University of Chile, Santiago, Chile.
[40] German Climate Computing Center (DKRZ), Hamburg, Germany
[41] Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA, USA
[42] Department of Mathematics and Statistics, University of Exeter, Exeter, UK
[43] Grantham Institute - Climate Change and the Environment, and Centre for Environmental Policy, Imperial College, London, UK
[44] School of GeoSciences, University of Edinburgh, UK
[45] Department of Statistics and Data Science, Department of Atmospheric and Oceanic Sciences, Institute of the Environment and Sustainability, University of California, Los Angeles
[46] State Key Laboratory of Severe Weather Meteorological Science and technology, Chinese Academy of Meteorological Sciences, Beijing, China
[47] University Mohammed VI Polytechnic, Ben Guerir, Morocco
[48] Laboratoire des Sciences du Climat et de l'Environnement (UMR 8212 CEA-CNRS-UVSQ), Institut Pierre Simon Laplace, Université Paris-Saclay, France
[49] National Centre for Earth Observation, University of Reading, Reading, UK
[50] Meteorological Research Institute, Tsukuba, Japan
[51] Global Change Institute, University of the Witwatersrand, Johannesburg, South Africa
[52] Sun Yat-Sen University, Guangzhou/Zhuhai, China
[53] Research Center for Climate Sciences, Pusan National University and Center for Climate Physics, Institute for Basic Science, Busan, Republic of Korea
[54] Climate Research Division, Environment and Climate Change Canada, Victoria, BC, Canada.
[55] Department of Geography, University of Liège, Belgium
[56]LMD-IPSL, CNRS, Ecole Normale Superieure, PSL Research University, Paris, 75005, France
[57] Energy, Climate and Environment Program, International Institute for Applied Systems Analysis, Laxenburg, Austria
[58] Geoenvironmental Sciences, University of Tsukuba, Tsukuba, Japan
[59] Mercator Ocean international, Toulouse, France

Correspondence to: Peter W. Thorne, peter@peter-thorne.net

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Abstract

Parties to the 2015 Paris Agreement agreed to limit the long-term increase in global average temperature to well below 2°C and pursue efforts to keep temperatures below 1.5°C relative to pre-industrial levels. As the world is fast approaching the 1.5°C warming level on a sustained basis, and with 2024 likely the first year that was over 1.5°C warmer than 1850-1900, there is ever increasing interest in how we will know whether and when 1.5°C warming since pre-industrial has been reached or exceeded with respect to a long-term average. This paper represents a comprehensive community methodological overview, building on the IPCC 6th assessment. It explains why there is no straightforward answer and proposes clear and reasoned ways forward. Existing challenges are as follows. Firstly, the Paris Agreement text contains definitional ambiguities around 'pre-industrial', 'global average temperature', whether the assessment should be on realised or long-term human-induced warming, and over what time frame the long-term temperature goal applies. Then, there are intrinsic limitations of observational records which get more uncertain further back in time due to data sparsity and measurement heterogeneity. Finally, in a non-stationary climate, multidecadal mean indicators of global temperature change will either lag behind the change or must rely on expected future temperature changes (based on extrapolation, initialized predictions, or scenario-based and constrained projections). Our analysis shows that knowing 'whether we are there yet' is a multifaceted and inherently probabilistic problem that includes information on the definition of a specific level of global warming, temperature changes over multiple timescales, and also potentially includes unpacking the attribution of human-caused changes from observed variations. Given the policy relevance of understanding where the world stands relative to 1.5°C, or any other level of global warming since pre-industrial, there are a number of practical steps which could be taken to increase specificity in answering this critical question in a timely manner, and inform future monitoring and assessment activities. This paper reviews a broad range of approaches, identifies the most pragmatic, robust and transparent, and clarifies requirements for use in real time including how to handle and represent remaining uncertainties. We show that it is possible by combining lines of evidence and several methodologies to estimate the present long-term warming level without delay in a manner that is robust both in retrospective validation of crossing past warming levels and, critically, to divergent warming futures including potential wildcard impacts of large volcanoes which can mask underlying warming for several years. Results are benchmarked against historical exceedances of 0.5°C and 1°C warming. Long-term warming as assessed using the approaches developed herein and data up to and including 2024 stands at 1.40 [1.23-1.58]°C, and underlying human-caused warming stands at 1.34 [1.18-1.50]°C. In IPCC quantified likelihood language this means that it was *unlikely* that long-term realised warming had exceeded 1.5°C by the end of 2024 and *very unlikely* that human-induced warming had exceeded 1.5°C.

Short Summary

We reassess the basis for determining the present level of long-term global warming. Unbiased estimates of both realised warming and anthropogenic warming are possible that approximate a 20-year retrospective mean. Our resulting estimates of 1.40 [1.23-1.58]°C (realised) and 1.34 [1.18-1.50]°C (anthropogenic) as at end of 2024 highlight the urgency of immediate, far-reaching and sustained climate mitigation actions if we are to meet the long term temperature goal of the Paris Agreement.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

1 Introduction

The Paris Agreement codified, for the first time in an international treaty, a quantified long-term temperature goal (LTTG hereafter) (UNFCCC, 2016)[1] as follows:

> *Article 2*
>
> *1. This Agreement, in enhancing the implementation of the Convention, including its objective, aims to strengthen the global response to the threat of climate change, in the context of sustainable development and efforts to eradicate poverty, including by:*
>
> *(a) Holding the increase in the global average temperature to well below 2°C above pre-industrial levels and pursuing efforts to limit the temperature increase to 1.5°C above pre-industrial levels, recognizing that this would significantly reduce the risks and impacts of climate change; [...]*

This LTTG text is scientifically ambiguous in at least three key respects: i) what constitutes pre-industrial levels; ii) what constitutes the global average temperature; and iii) what metric should be used to determine the warming level that has been reached at any given time. This paper is motivated by the first annual exceedance of 1.5°C in the annual global average temperature in 2024 in many but not all datasets and the imminent proximity of the likely longer-term exceedance of 1.5°C warming, but the issue we address is more general: how best to quantify present sustained global warming relative to the pre-industrial era?

Warming over land regions was recognised in the 1930s (Kincer 1933), including an attribution of global land warming to increases in carbon dioxide concentrations (Callendar 1938), but the first representative global temperature estimates including land and marine temperatures were published in the 1980s and 1990s (Jones et al. 1986; Hansen et al., 1996, Smith and Reynolds, 2005[2]). These global estimates mostly combine land surface air temperatures (LSAT) with sea surface temperatures (SSTs) of the ocean (Jones et al. 1999), and are termed global mean surface temperature (GMST). Only very recently has an instrumental dataset of global surface air temperature (GSAT) change been produced by combining marine air temperatures (MAT) over the ocean with LSAT (Morice et al., 2025).

[1] This followed a considerable body of work that had examined the impacts and risks of different warming levels, including the Intergovernmental Panel on Climate Change (IPCC) 5th Assessment Report (IPCC, 2012, 2013, 2014).
[2] Note that this is the first paper fully describing the NOAA product which had been operationally produced for some years prior to its description in the peer reviewed literature.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

With the adoption of the LTTG there has been a proliferation of activities to monitor and communicate long-term temperature change. Since the 1990s, the World Meteorological Organization (WMO) has provided annual monitoring through its Statement on the Status of the Global Climate (later State of the Climate) issued each March, complemented by annual statements from the three established centres[3] producing long-term global temperature products on their own data, that are typically released in mid January. These agency-based communication statements and reports are expanded in peer-reviewed publications by the detailed annual State of the Climate report series in the Bulletin of the American Meteorological Society[4], as well as more recently annual updates to global climate change indicators that replicate the IPCC methodology between the major assessment reports (Forster et al., 2024, 2025). Recent years have seen more entities including new dataset producers (e.g. Berkeley Earth), climate service providers (e.g. Copernicus Climate Change Service), and specialised climate media outlets (e.g. Carbon Brief), starting to provide reports on global temperature and climate change, with a wider variety of datasets and approaches. Monthly updates have become the norm, often supplemented when interesting events occur. Recently, even daily global temperature variations have been prominently discussed in the media (e.g. AP, 2023, Guardian 2023, 2024).

Throughout this paper we use the IPCC 6th Assessment Report of Working Group I (AR6 WGI) definition of global warming (IPCC, 2021c):

> *Global warming refers to the increase in global surface temperature relative to a baseline reference period, averaging over a period sufficient to remove interannual variations (e.g., 20 or 30 years). A common choice for the baseline is 1850–1900 (the earliest period of reliable observations with sufficient geographic coverage).*

At any given level of global warming it is expected that the Arctic, global land areas, specific high elevation regions and urban areas will have warmed more, and the global ocean less, than the global mean. Natural climate variability will also cause global annual fluctuations of a few tenths of a degree around the long-term global mean with much larger anomalies regionally. It is critical to recognise that exceeding the *long-term* temperature goal at one location or in any *short-term* period such as a calendar year does not necessarily mean that the LTTG has been exceeded, as has sometimes been erroneously implied (e.g. McCulloch et al., 2024; Esper et al., 2024).

Global warming is associated with sustained multi-decadal changes in the distributions of rainfall, drought, snow and ice, extreme weather, and multiple climate impact drivers, which are collectively referred to as climate change[5]. Climate change-

---

[3] NOAA NCEI (was NCDC until 2012), NASA GISS, and the UK Met Office / Climatic Research Unit at the University of East Anglia.

[4] https://www.ametsoc.org/index.cfm/ams/publications/bulletin-of-the-american-meteorological-society-bams/state-of-the-climate/

[5] The IPCC glossary definition is: "A change in the state of the climate that can be identified (e.g., by using statistical tests) by changes in the mean and/or the variability of its properties and that persists for an extended period, typically decades or longer. Climate change may be due to natural internal processes or external forcings such as modulations of the solar cycles, volcanic

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Earth System Science Data
Open Access
Discussions

related impacts are generally estimated relative to a more recent baseline than is used for the LTTG, due to the shorter historical records that are available to quantify these climate impacts. This means that changes in estimated global warming during earlier periods, such as due to changes in the 1850–1900 baseline (IPCC, 2021a) do not, on their own, mean that current impacts are greater or that projected climate impacts will necessarily occur either sooner or later than previously anticipated (Figure 1).

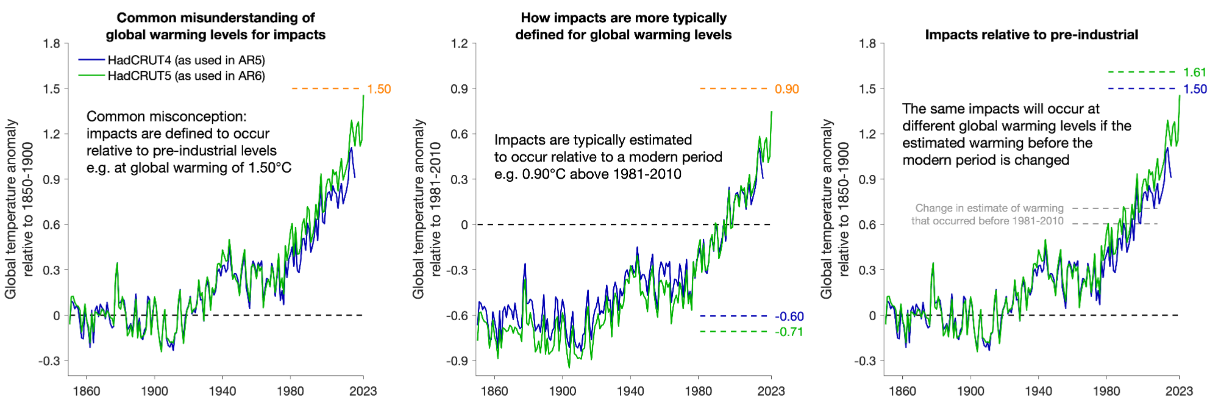


Figure 1: (left) A common misconception is that an increase in the estimate of historical warming (due to scientific progress) brings us closer to a level of warming at which some projected impacts may occur. For example, moving from HadCRUT4 (used in AR5, which informed the Paris Agreement) to HadCRUT5 (as used in AR6) increases the estimate of warming since 1850-1900 by about 0.1°C, so it might appear that impacts projected to occur at 1.5°C above pre-industrial levels (orange dashed line) would occur earlier (middle). However, in practice, impacts are generally assessed relative to a more modern, better quantified and more societally relevant baseline (schematic uses 1981-2010). In this example, an impact is estimated to occur at 0.90°C above 1981-2010 levels (orange dashed line) which, using HadCRUT4, is 0.60°C warmer than pre-industrial levels (1850-1900), producing an impact at 1.5°C above 1850-1900. However, HadCRUT5 estimates that 1981-2010 had warmed 0.71°C above the pre-industrial 1850-1900 period (right). This shift in estimated historical warming means that the same impact would actually occur at 1.61°C above pre-industrial levels using HadCRUT5, meaning that the impact is not necessarily going to occur earlier than previously understood. By contrast, any change in estimated warming that has occurred since the modern reference period (1981-2010 in this example) used to assess impacts might mean the impact would occur earlier. This assumes that the assessment of the level of warming at which the impact occurs does not change; however, AR6 SYR concluded that risks associated with the 'Reasons For Concern' (and thus dangerous anthropogenic interference) generally increase at lower temperatures than previously assessed in the IPCC AR5, due to updated scientific understanding (see Section 6). Note that any anthropogenic warming that occurred before 1850-1900 (likely 0.0-0.2°C, according to IPCC AR6 WGI CCBox 1.2) does not generally matter in the context of defining when particular impacts might occur for the same reason.

A rich history of assessment of impacts on societal and natural systems at different levels of global warming has been comprehensively addressed by IPCC Working Group II (WGII) over successive assessment cycles. The evolution of the science-policy nexus on avoiding dangerous anthropogenic interference with the climate system that informs both the 1.5°C

---

eruptions and persistent anthropogenic changes in the composition of the atmosphere or in land use. Note that the United Nations Framework Convention on Climate Change (UNFCCC), in its Article 1, defines climate change as: 'a change of climate which is attributed directly or indirectly to human activity that alters the composition of the global atmosphere and which is in addition to natural climate variability observed over comparable time periods'. The UNFCCC thus makes a distinction between climate change attributable to human activities altering the atmospheric composition and climate variability attributable to natural causes."

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

and well below 2°C global warming levels in the LTTG stretches over several decades (e.g. O'Neill and Oppenheimer, 2004, Cointe et al., 2011, Cointe and Guillemot, 2023). The third IPCC WGII assessment report in 2001 was the first to explicitly relate aggregated climate change impacts to levels of global average warming, so that policymakers could consider what are unacceptable levels of risk of key impacts ('dangerous climate change') (IPCC, 2001). By 2015, following the fifth IPCC cycle reports, the Structured Expert Dialogue (SED) process (UNFCCC, 2015) concluded that the 'guardrail concept' that had been associated with limiting warming to a certain level was inadequate, since severe impacts were already occurring at current levels of warming, thus motivating efforts to strengthen the long-term global goal as agreed later that year in the Paris Agreement. This understanding was subsequently confirmed by the IPCC SR1.5 which showed that every increment of warming matters (IPCC, 2018). Livingston and Rummukainen (2020) contend that, in turn, the Paris Agreement and resulting request to IPCC for the SR1.5 reframed the science in important ways by reorientating scientific priorities for funders and investigators.

Using IPCC AR6 WGI methods and datasets, Earth's surface warmed by 1.24 [1.11 to 1.35]°C from 1850–1900 to 2015–2024, of which 1.22 [1.0 to 1.5]°C was human-induced (Forster et al., 2025). Ongoing human-caused warming is currently continuing at 0.27 [0.2 to 0.4]°C per decade from continuing greenhouse gas emissions (Forster et al., 2025). Across a range of datasets, 2024 was likely the first calendar year above 1.5°C[6] with a central estimate of 1.52 [1.39 to 1.65]°C using IPCC datasets and approach (Forster et al., 2025) or 1.55°C using WMO datasets and approach. However, this was well above the best estimate of human-caused warming (1.36°C), highlighting the role of internal variability in short-term climate (Forster et al., 2025).

Until the Paris Agreement, changes in GMST dataset versions between successive IPCC reports introduced changes of the order of hundredths of a °C for their common periods of record and were thus deemed negligible and unremarked. Subsequent to the Paris Agreement, however, both new and updated GMST datasets accounted for new knowledge of biases in the raw data and the impacts of spatial coverage. The GMST temperature change estimates for the period 1880-2012 using the methods in IPCC WGI AR5 were increased in AR6 by approximately a tenth of a °C (c.10% of the long-term warming to date at that time) (Gulev et al., 2021) due to improved quantification of GMST in the early instrumental period and better accounting for warming in data sparse regions in recent decades (Fig. 1). A further historical assumption had been that changes in GMST and GSAT were broadly equivalent (e.g. Folland et al., 1984) but this is not the case in Earth System Models (Cowtan et al., 2015). These insights have highlighted the uncertainties inherent in estimating long-term global surface temperature changes which must be accounted for in tracking the current global warming level and adherence to the LTTG.

[6] https://wmo.int/media/news/wmo-confirms-2024-warmest-year-record-about-155degc-above-pre-industrial-level

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

The precise estimate of warming above pre-industrial levels therefore depends upon the choices of baselines, metrics and datasets (e.g. Gulev et al., 2021, Betts et al., 2023). Different choices can change the best-estimate by tenths of a degree Celsius with substantial implications for determining whether, and if so when, the 1.5°C warming level referred to in the LTTG will have been reached or exceeded[7]. Herein we attempt to provide a clear and practical basis for determining the 'current' level of global warming with a view to tracking adherence to the LTTG. The approach chosen should both have broad consensus and endorsement from the scientific community and be easily usable and understandable for scientists, policymakers, climate change negotiators, and end users, including the general public. Multiple studies exist looking at specific parts of the problem of reliably estimating global surface temperatures and informing adherence to the LTTG (e.g. Betts et al., 2023, Hawkins et al., 2017). The risk is that in looking at only specific subsets of the problem a false sense of certainty ensues. Here, instead, we deliberately aim to provide a comprehensive solution-oriented review and assessment of the state of knowledge and propose pathways towards consistent monitoring and messaging to aid public understanding, climate change research and policy responses.

This is far from a purely academic exercise. The surface temperature record has been extensively referenced in the UN Climate treaties and decisions thereunder[8], and is also increasingly relied on in climate litigation across national, regional and international courts. The International Court of Justice (ICJ) has considered the rights and responsibilities of states in relation to climate change in the context of an Advisory Opinion that was referred to it by the UN General Assembly in 2024[9], finding that states have extensive obligations both in and beyond the Paris Agreement to prevent, mitigate and adapt to climate harm, and that a breach of these obligations can lead to legal consequences for the offending states, including a duty to make reparations[10]. And, a growing number of human rights-based cases are being brought against states for inadequately ambitious climate action, for example a judgment from the European Court of Human Rights in the Klimaseniorennen case, in 2024, holding the Swiss Government in breach of its human rights obligations[11].

The remainder of the paper is structured as follows: Section 2 discusses the definition of the pre-industrial baseline. Section 3 explores various global temperature metrics as well as the latest datasets, their uncertainties, and remaining limitations. Section 4 considers possible solutions to the challenge of estimating the underlying long-term global warming in a given year, such as 2024, when that year's temperature results from a combination of human forcings, natural forcings (e.g. solar variability and

[7] We note that IPCC AR6 WGI in SPM B.1.3 refers to exceeding 1.5°C or, in the specific context of SSP1-1.9, reaching 1.5°C before returning below, and we use this language of reached or exceeded herein for consistency. The IPCC AR6 SYR also uses the same two terms associated with 1.5°C, although sometimes in different contexts to that in WGI.
[8] UNFCCC preambular recital 16, PA, preambular recital 4, Art 4.1, 14.1, Glasgow Climate Pact, para 23, UAE consensus, paras 6, 28(d), 39, 55, 61 and 149
[9] https://documents.un.org/doc/undoc/ltd/n23/063/82/pdf/n2306382.pdf
[10] https://www.icj-cij.org/case/187
[11] ECtHR, *Verein KlimaSeniorinnen Schweiz and Others v. Switzerland*, no. 53600/20, judgment (Grand Chamber) of 9 April 2024.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Open Access Earth System Science Data Discussions

volcanism) and internal climate variability (e.g. El Niño Southern Oscillation, Atlantic multidecadal variability). Section 5, taking insights from the prior three sections, proposes practical solutions which can be implemented now that may avoid confusion while respecting the irreducible scientific uncertainties. Section 6 discusses potential implications and interpretation in the context of the UNFCCC and Paris Agreement. Finally, Section 7 reflects on the broader context and meaningful actions that could be taken to better quantify global warming relative to the LTTG. The Supplement contains a table of acronyms used to support the non-expert reader alongside further technical details pertaining to Sections 4 and 5.

2 What constitutes pre-industrial levels?

The Paris Agreement does not specify a pre-industrial time period for the purposes of the LTTG. There are numerous historical interpretations of the true start of the industrial era (Ashton, 1997, Hawkins et al., 2017, Chen et al., 2021). The IPCC AR4 glossary stated: '*In this report the terms pre-industrial and industrial refer, somewhat arbitrarily, to the periods before and after 1750, respectively*' (IPCC, 2007a), and the Cancun Agreements at COP16 in 2010[12] were based upon this. The definition in the AR5 remained the same (IPCC, 2013), preceding the Paris Agreement in 2015. The IPCC AR6 revised the definition of the pre-industrial period to '*The multi-century period prior to the onset of large-scale industrial activity around 1750. The reference period 1850–1900 is used to approximate pre-industrial global mean surface temperature*'[13]. Across successive IPCC reports, radiative forcings have been indexed relative to 1750, while observed global surface temperature changes have, since AR4, been reported relative to 1850–1900 (or 1850–1899) as an 'approximation to pre-industrial' (IPCC, 2007b, 2013, 2018, 2021a). The 1850–1900 average has thus been broadly (but not unanimously) adopted as an approximation to pre-industrial conditions. Throughout the remainder of the paper, this baseline has been assumed, unless otherwise specified.

The 1850–1900 definition for the pre-industrial period originated partly because the global temperature datasets used by the IPCC across successive reports begin during that time window. Figure 2 illustrates the rapidly changing spatial availability of temperature observations through time and highlights difficulties in producing 'global' estimates from earlier temperature observations. GloSAT extends instrumental estimates of global temperature change back to 1781 (Morice et al., 2025), exploiting MAT data from ships, which is more common than SST data during the earlier period. The Berkeley Earth global land-only temperature product extends back to 1750, with 239 or fewer temperature stations contributing data before 1850, compared with thousands since 1950. Large annual uncertainties prior to 1850 are reported by both Berkeley Earth and GloSAT (~0.6 °C, 95% confidence interval for the 1800–1850 period, and much higher values prior to 1800), with high interannual and interdecadal variability. This increased uncertainty is from a combination of relatively poor data coverage and the effects of unusually large recurrent volcanic eruptions in the early 19th Century (Morice et al., 2025). Despite greater uncertainties in the early parts of these historical temperature datasets, both report some warming prior to and during the 1850–1900 period.

[12] https://unfccc.int/tools/cancun/index.html accessed 5/1/25
[13] https://apps.ipcc.ch/apps/glossary/ accessed 5/1/25

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

For example, Berkeley reports 0.11 °C warming of global land temperatures from 1750—1800 to 1850—1900. This early warming is also in agreement with regional paleoclimate evidence from the continents and tropical oceans (Abram et al., 2016; Henley et al., 2024, McCulloch et al., 2024) and recent model simulations (Ballinger et al. 2025).

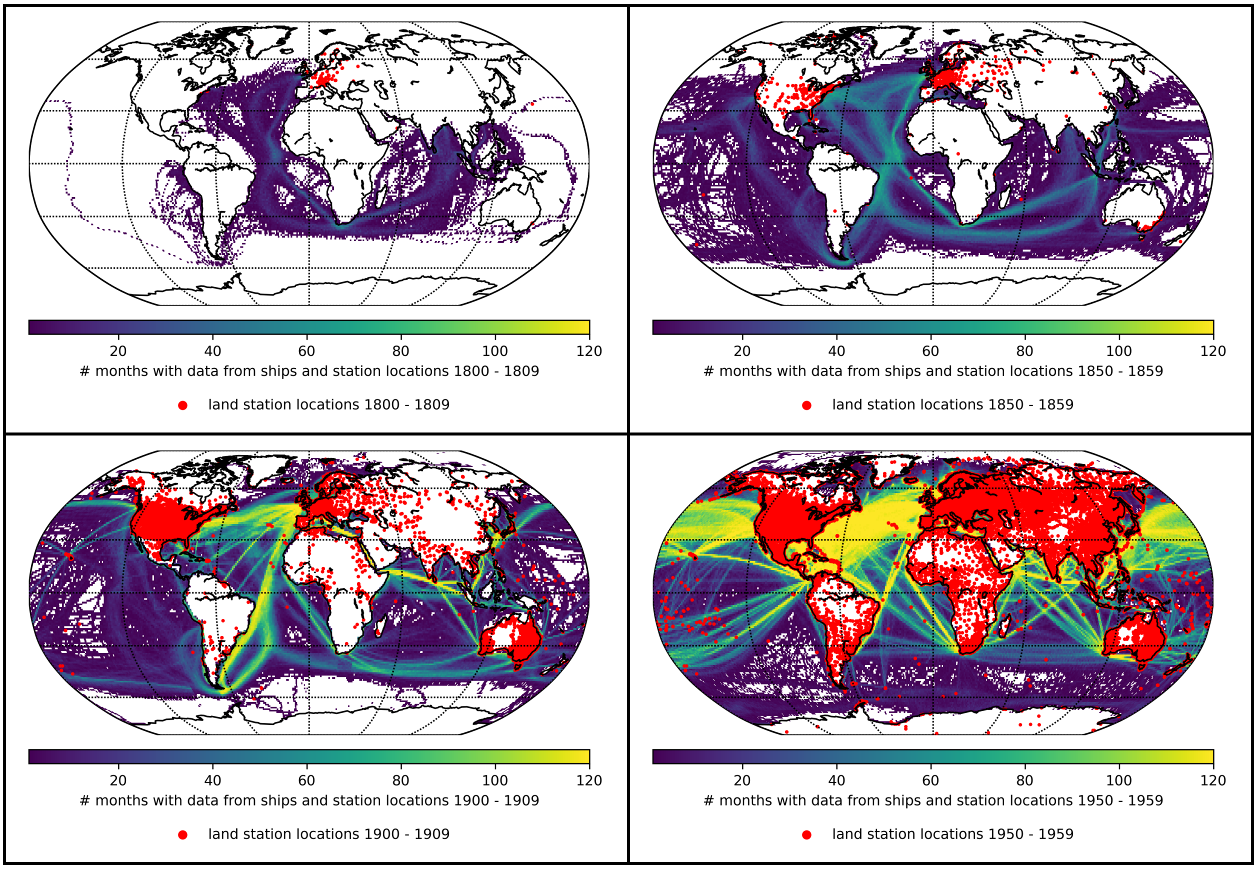


Figure 2. Land stations reporting any data (red dots) and number of reports with MAT and / or SST observations arising from the ship data underlying ICOADS (colours in colour bars) for decades separated by 50 years from 1800-1809 to 1950-1959. Data arise from the latest versions of data release from the Copernicus Climate Change Service contract covering data rescue and access to land and marine in-situ holdings. Similar overall implications would arise from other international repositories and data availability worsens still further prior to 1800. Prior to 1850, land stations were almost exclusively over Europe, while marine data were dominated by measurements in the Atlantic and Indian Oceans, and were mainly of MAT rather than SST. Sampling issues are particularly acute in the Southern Ocean and the Pacific Ocean, especially prior to the opening of the Panama Canal in 1914. Poor Pacific coverage is problematic given the importance of the El Niño-Southern Oscillation (ENSO) in determining global surface temperature variability (Trewin 2022, Thompson et al. 2009, Trenberth et al. 2002). Similarly, the Suez Canal opening in 1869 changed shipping routes in the Indian Ocean and South Atlantic.

Various considerations complicate the interpretation of the period that might best represent pre-industrial conditions. In relation to climate forcings through the last millennium, evidence indicates a possible small human footprint on atmospheric

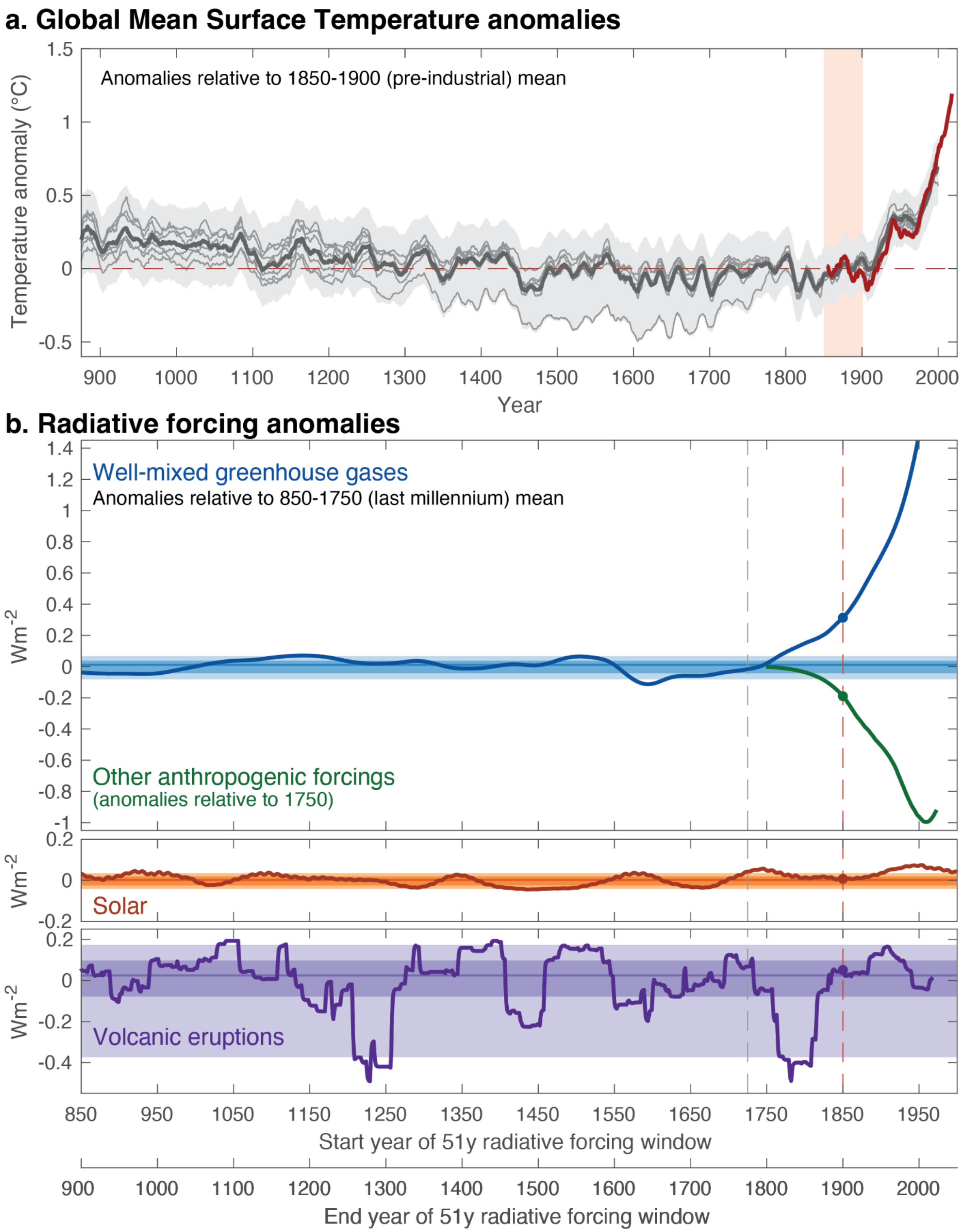
https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

greenhouse gas levels, principally through population and land use changes[14], even before 1750 (Koch et al., 2019, King et al., 2024), while early industrial activity occurred during 1750–1850 (Gulev et al., 2021, Forster et al., 2021, Chen et al., 2021). If we consider 850–1750 variability as a reasonable representation of preindustrial, then Figure 3b (red dashed line) shows that 1850–1900 volcanic and solar irradiance forcings are representative of background pre-industrial natural forcing. In contrast, well-mixed greenhouse gas (WMGHG) forcing already clearly exceeds the long-term preindustrial range (Figure 3b) and in 1850–1900 was +0.33 (0.28 to 0.37) Wm^{-2} relative to 1750 (Chen et al., 2021). Other anthropogenic forcings (mainly aerosols and albedo change due to land use change) were also already evolving during the 1850–1900 climate baseline period, and are estimated to have a -0.15 (-0.33 to -0.03) Wm^{-2} radiative forcing effect relative to 1750 (Chen et al., 2021). A 51-year period centred around 1750 is the most recent interval where the 51-year mean total radiative forcing was representative of typical pre-industrial climate forcing (Figure 3b grey dashed line). However, it was followed by unusually strong and frequent volcanic forcing (Fang et al., 2023), which may have a lasting effect on global temperature variations (Ballinger et al., in press), meaning that the late 18th and early 19th centuries do not represent typical pre-industrial conditions in climate forcing (Hawkins et al., 2017).

[14] These land use changes in addition have had a direct radiative impact through changes in albedo and surface heating characteristics.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Open Access Earth System Science Data Discussions



https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Figure 3: a) Reconstructed temperatures over the past millennium from PAGES2K reconstructions[15] (PAGES 2k consortium, 2019), showing the median for each of the 7 different reconstruction methods (thin grey curves), as well as the overall ensemble median (thick black curve) and 5-95% range across all ensemble members. Also shown are observational GMST anomalies (red curve) since 1850 (Foster et al., 2025). All data shown as decadal moving averages relative to 1850–1900 mean (horizontal dashed line) for the commonly used pre-industrial reference period (light red shading). b) Natural and anthropogenic radiative forcing for 51-year moving windows (denoted by start year) since 850 CE (thick curves). Typical pre industrial last millennium distributions are shown for the median (thin horizontal lines), 25-75% range (dark shading) and 5-95% range (light shading), assessed across the 850-1700 interval. Solid circles represent the mean radiative forcing for the 1850-1900 interval (dashed vertical red line) that is typically used as the preindustrial reference interval to define global mean temperature change (a). Grey dashed line indicates mean radiative forcing for the 51-year interval centred around 1750 which approximates the most recent period where radiative forcing was representative of typical preindustrial conditions. Datasets are based on radiative forcing for well-mixed greenhouse gases (blue), solar (orange) and volcanic (purple) forcing as used for PMIP4/CMIP6 last millennium climate simulations (e.g. AR6 chapters 2 and 7) and for other anthropogenic forcings (green) are derived from the 2023 Indicators of Climate Change (Foster et al 2025). Note that the dataset for other anthropogenic forcings (mostly aerosols and land use albedo change) are only available from 1750, and so are shown as deviations from the 1750 value.

The IPCC AR6 concluded that between 1750 and 1850–1900 the global average temperature warmed by 0.1 [-0.1 to 0.3] °C with an assessed *likely* anthropogenic contribution of 0.1 [0.0 to 0.2]°C, albeit with only *medium confidence* (Chen et al. 2021). Supporting evidence included: very early observations from principally European locations, proxy reconstructions, climate models, and the AR6 emulators[16] used across the IPCC reports (Hawkins et al., 2017, PAGES 2k consortium, 2019, Schurer et al., 2017, Haustein et al., 2017, Forster et al., 2021). Subsequent model simulations add to this evidence. For example, a control simulation using 1750 radiative forcings was 0.1°C cooler than an 1850 based control simulation in the same model (Ballinger et al., in press). There has also been a proposal to re-baseline the early temperature record using the observed linearity between the record of atmospheric carbon dioxide concentration and global surface temperatures (Jarvis and Forster, 2024). This method equates departures of CO_2 from its pre 1700 baseline with a given amount of global surface warming and may help reduce uncertainty in warming level estimates. Based on current knowledge, adopting a pre-industrial reference period before 1850–1900 would boost estimated global warming by approximately 0.1°C but would also increase uncertainty by even more. On balance, we consider that policymakers for now are better supported by a continued use of 1850–1900 as an approximation to the 'pre-industrial' level, whilst recognising that it probably does not truly represent pre-industrial conditions. Were scientific understanding to improve sufficiently in the future then an earlier period estimate better representing 'true pre-industrial' could plausibly be used.

[15] https://figshare.com/articles/dataset/Reconstruction_ensembles/8137706?backTo=/collections/Global_mean_temperature_reconstructions_over_the_Common_Era/4507043 last accessed 24/3/25

[16] IPCC AR6 glossary defines emulators as '*A broad class of heavily parametrized models ('simple climate models'), statistical methods like neural networks, genetic algorithms or other artificial intelligence approaches, designed to reproduce the responses of more complex, process-based Earth system models (ESMs). The main application of emulators is to extrapolate insights from ESMs and observational constraints to a larger set of emission scenarios*.'

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

3 Defining and estimating global surface temperature

3.1 Many ways to define surface temperatures

Surface temperature can refer to skin temperature, near-surface air temperature, or near-surface ice, land or ocean temperature[17] (Merchant et al., 2013). These quantities co-vary and all increase in response to positive radiative forcing, but potentially by different amounts (Merchant et al., 2013, Cowtan et al., 2015). Given: i) the ambiguous language in the LTTG; ii) that much of the nuanced understanding of differences between measurands has arisen after the Paris Agreement; and iii) that ascertaining the absolute basis for the LTTG is itself ambiguous (Sections 1 and 6); it is not currently clear which metric(s) should be used.

The only available direct observations with sufficient spatial coverage and longevity combine LSAT over land with SSTs or MATs over the ocean. Sea ice regions lack long-term continuous weather monitoring stations and this introduces challenges in interpretation, especially when sea ice cover changes (see Section 3.2.3). Observational products to date generally are of GMST (Section 1) whereas model and reanalysis assessments tend to use GSAT, or occasionally GMST (Figure 4).

Cowtan et al. (2015) first drew attention to non-trivial differences in model-derived warming between GMST and GSAT. Earth System Models consistently predict that global-mean MATs should warm by 2-12% more than SSTs (Cowtan et al., 2015; Richardson et al., 2018; Beusch et al., 2020; Gillett et al., 2021). The physical causes appear to be a combination of the direct effect of CO_2 preferentially heating air over the surface, and strong oceanic evaporation feedback which limits surface water warming (Richardson, 2023). However, all model air-sea flux parameterisations are based upon Monin-Obukhov similarity theory (e.g., Businger et al., 1971), so a common bias across models cannot be ruled out (Gulev et al., 2021). The ERA-Interim, ERA5 and JRA-55 reanalyses show 2-4% more warming in GSAT than in GMST over periods since 1979 (Simmons et al., 2017; Soci et al., 2024), but they share ESMs' parameterisations of the boundary layer fluxes.

[17] Defining 'near-surface' is ambiguous with no universal adoption of a given level to define near-surface for measurement purposes across all surfaces (although there are standards e.g. over land specified in WMO regulatory materials, levels of adherence are unclear and standards do not exist over all domains). Frequently sharp vertical gradients in temperature occur near the surface both over and below all surfaces.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

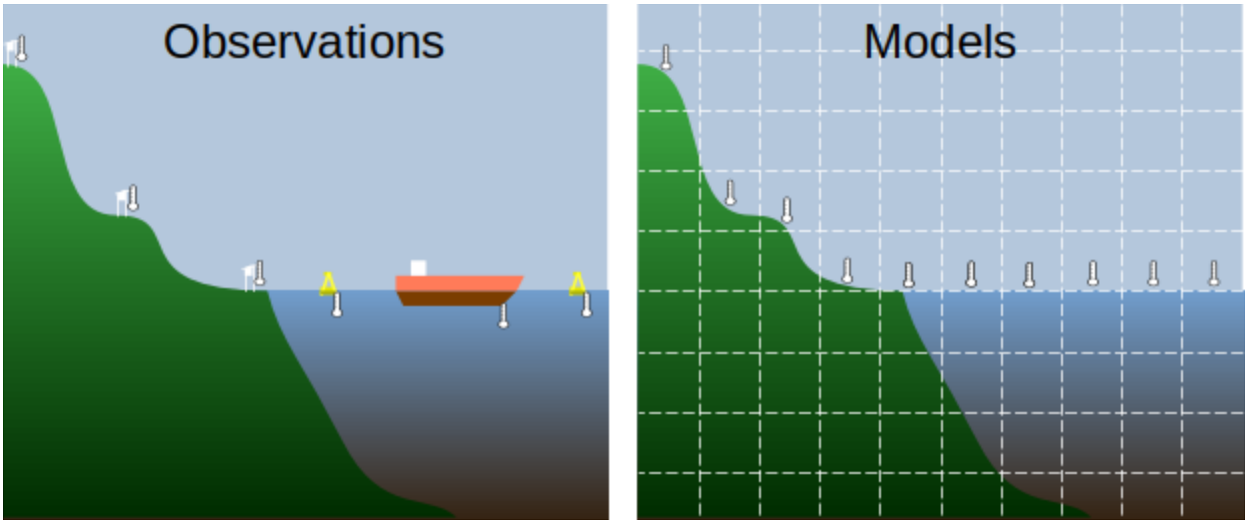


Figure 4. Illustration of data sources for most observational datasets (left panel), which are based on weather station observations of air temperature over land combined with sea surface temperatures measured from diverse platforms including buckets, ship engine room intakes and buoys. These observations have generally been compared to simulated air temperatures from climate models and reanalyses from a fixed height above the surface over both land and ocean (although models do also simulate sea surface temperatures). Comparisons are generally made of anomalies rather than absolutes that can ameliorate to an extent any geophysical differences between the different measurands.

Pacific buoy data show differing short-term regional trends in SST and MAT (Christy et al., 2001; Rubino et al., 2020), but regional data are strongly affected by local ocean heat transport, and may not represent the long-term global response. In the small sample of global observational estimates on the longest timescales, SSTs appear to have warmed somewhat faster than MATs (Gulev et al., 2021). However, uncertainties are plausibly larger than model-observation differences and conclusions depend upon the sub-period and choice of SST and MAT products (Junod & Christy, 2020; Cornes et al. 2020; Gulev et al., 2021). Over 1850 to 2010 (Huang et al. 2017) or 1850 to 1941 (Kennedy et al. 2019) there is also some circularity as these SST products depend on nighttime-MAT (NMAT) measurements for their long-term homogenisation. Furthermore, MAT records are adjusted to account for how ships have overall become taller with time using similar boundary-layer assumptions to those made in models and reanalyses (e.g. Cornes et al., 2020). The challenge of resolving apparent model-observation disagreements is compounded by limitations in our theoretical knowledge of boundary layer processes over the ocean on climate timescales (Gulev et al., 2021).

In the absence of a more recent comprehensive assessment on this topic, we use the IPCC AR6 WGI assessment here (Gulev et al., 2021; their CCB2.3), in which warming uncertainty was expanded by ±10% of changes to account for lack of knowledge about long-term SST-MAT trend differences. Further research on all aspects of MAT-SST differences, but particularly theoretical and observational aspects, could reduce uncertainty in the differences between global temperature estimates.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

3.2 Observational uncertainties

Uncertainties in reconstructions of historical global surface temperature changes are unavoidable. Uncertainties arise from three principal sources : i) spatio-temporal incompleteness and the extent to which interpolation to impute estimates for missing regions is undertaken; ii) inherent measurement uncertainties; and iii) changes in instrumentation, siting, microclimate and measurement practices. Research groups adopt diverse approaches to dataset construction and uncertainty quantification, leading to a range of warming estimates (Figure 5).

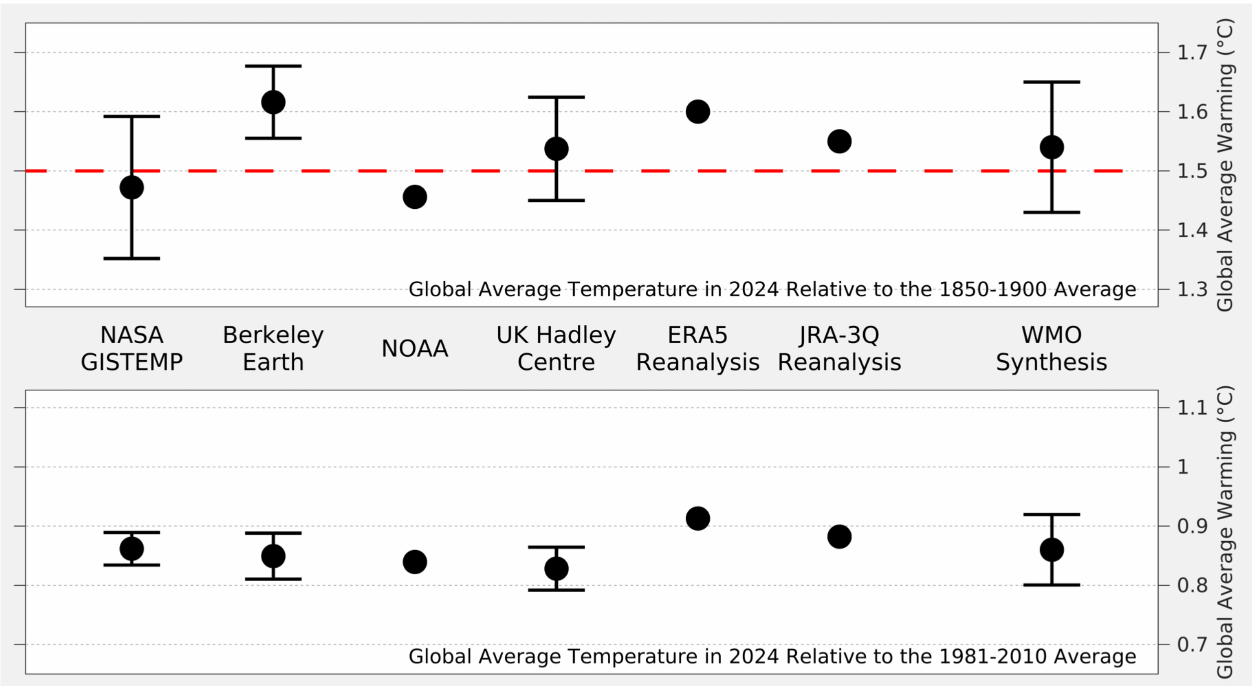


Figure 5. Illustration of how different research groups, taking different approaches, yield a range of estimates as to how warm 2024 was relative to 1850-1900 (top panel) or 1981-2010 (lower panel). This highlights the uncertainty in estimating long-term change arising from data selection and methodological choices. Error bars correspond to the 95% uncertainty range as reported by each dataset, where provided. ERA5, JRA-3Q, and NASA GISTEMP do not extend back to 1850; however, their originating research groups have provided an estimate of the implied change since the 1850-1900 average by adding an adjustment factor to account for the early changes seen in other datasets.

Observational uncertainties arise from effects that can be "systematic" (highly correlated between consecutive measured values), "random" (entirely uncorrelated), or in between (BIPM, 2008). Random components reduce upon averaging across space and/or time, whereas systematic uncertainties persist. Recent reference observation networks show promise to robustly quantify both random and systematic uncertainties in modern data (Thorne et al., 2018). However, historical uncertainties must be derived from the characteristics of the data themselves (e.g. Kent et al. 1999, Williams et al., 2012, Rhines et al., 2015, Carella et al. 2018, Chan and Huybers 2020), or from experiments with parallel measurements (e.g. de Valk and Brandsma, 2023, Wallis et al. 2024, Ashford 1948, Carella et al. 2017b) informed by any known changes in measurement technology or practices (e.g. Folland and Parker 1995, Venema et al., 2012, Camuffo, 2002, Kent and Kennedy 2021, Wallis et al. 2024).

Changes in instrumentation and observing practices can introduce non-climatic changes in reported temperature (often termed inhomogeneities, Trewin, 2010; Kent and Kennedy, 2021). Commonly, metadata such as the timing and the type of changes in instrumentation, siting, methods or observers, are unknown (Kent and Kennedy, 2021; Trewin, 2010). For SST and MAT, uncertainty quantification is further complicated as ships commonly lack vessel identifiers, so uncertainties correlated along a vessel track are often treated as if they are uncorrelated rather than systematic (Kennedy 2014; Kennedy et al., 2019).

Lastly, the geographical availability of observations diminishes in earlier periods and the sampled regions systematically change over time (Figure 2). In-situ observations, which are the basis for all century and longer datasets, are consistently sparser in the Southern Hemisphere than the Northern Hemisphere, and sparser at high latitudes and in the tropics than in the mid-latitudes. This significantly complicates the estimation of a true global value even in the modern era. Sampling effects may be quasi-systematic. For example high-altitude stations have on average warmed more than low altitude stations (Mountain Research Initiative EDW Working Group, 2015) but many high altitude regions are undersampled.

The application of differing reasonable methods generates a range of plausible estimates of global surface temperature change. This pertains to all aspects of dataset construction: data selection, quality control, homogenisation, and the creation of (infilled) gridded products. It is important that this structural uncertainty in dataset construction is adequately sampled to ensure robust inferences (Thorne et al., 2005). The remaining subsections consider in turn the state-of-the-art of these aspects and whether the structural uncertainty is adequately sampled currently.

3.2.1 Land Surface Air Temperature Uncertainties

Datasets of land surface air temperature (LSAT) take somewhat distinct approaches to station selection, quality control, break detection, and homogenisation (e.g. Osborn *et al.*, 2021; Menne *et al.*, 2018; Sun *et al.*, 2021; Rohde *et al.*, 2013; Chan *et al.*, 2024a; Ishii *et al.*, 2025; Chen *et al.*, 2025). There are, however, some overlaps, for example because some datasets draw upon the same national-level and regional homogenised products (Osborn *et al.*, 2021, Xu *et al.*, 2018).

Homogenisation of LSAT records involves accounting for breaks in temperature records stemming from the factors listed in Section 3.2, for example, changes in instrumentation or in the local environment such as through urbanisation. The most common approach is pairwise comparison of a station's monthly aggregated series[18] with its neighbours to identify apparent

[18] Daily and monthly aggregation can be undertaken in multiple ways and this can impart both systematic and random differences (Bonacci and Zeljkovic, 2018). Often monthly aggregation occurs upstream of the dataset teams and depends upon national practices. In cases where dataset providers do the aggregation they use a consistent approach to minimise biases.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

breakpoints, in conjunction with available metadata. For most datasets, an adjustment is calculated for each breakpoint, and the process is iterated until no further significant discontinuities are detected (e.g. Menne and Williams, 2009). The resulting homogenized temperature data are next typically converted to anomalies relative to a common baseline. Anomalies vary much less across a given spatial scale than absolute temperatures (which are a function of surface type, orography and other factors) and hence are less sensitive to changes in the underlying observing network. An averaging algorithm then converts the anomalies into global land surface air temperature changes. Berkeley Earth is distinct in that the breakpoint associated bias corrections occur within the averaging process (Rohde et al. 2013). Benchmarking against plausible test data raises confidence in the effectiveness of homogenisation approaches while highlighting that neighbour-based homogenisation can be limiting if i) the network density is low (a particular challenge in the early record); or ii) quasi-contemporaneous changes occur over large regions (e.g. Williams et al., 2012; Venema et al., 2012). In many cases, changes occurred at the national level almost simultaneously which can lead to challenges for neighbour-based approaches (Williams et al., 2012). Homogenisation generally increases the long-term warming by 0.1-0.2°C over global land with much larger changes in some regions (Chan et al., 2024a, Menne et al., 2018).

There are also novel techniques which are more experimental in nature and have not yet been operationalised. Gillespie et al (2022) used sparse input reanalyses (Section 3.2.4) as an independent background estimate to identify and adjust for breakpoints. They found less warming than in existing products, but the approach is critically dependent upon the reanalysis used as a reference. The EUSTACE project (Rayner et al., 2020) used an ensemble break detection method for station data (Brugnara et al., 2019). The station data and breaks as well as estimates of air temperatures across land, ocean and ice areas derived from satellite retrievals were then statistically combined. There are also several regional (e.g. Cornes et al., 2018) and national temperature products. These additional experimental, regional and national products collectively imply that the available global datasets likely undersample the true structural uncertainty in reconstructions of global LSAT (e.g. Way et al., 2017), and that there is a substantial need to develop and operationalise additional LSAT products to better sample structural uncertainty.

The early period is a particular challenge. Prior to the early 20th Century and standardisation of instrumentation and methods of observation under the International Meteorological Organization (IMO, the precursor to the WMO) there existed substantial heterogeneity in instrumentation, mounting, sheltering and measurement techniques (Parker, 1994). A recent comprehensive review using both contemporaneous and modern parallel measurement series comparing older to modern techniques highlights complex biases with substantial seasonality and geographical variability (Wallis et al., 2024). Given the sparsity of station records (Figure 2) during this early period, existing neighbour-based homogenisation techniques may be insufficient, with some techniques unable to assess early periods of record. Proxy records over 1850-1900 suggest that the available LSAT series may be systematically warm biased in this period during summer (Schneider et al., 2023). This would be consistent with the

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

overall tendency found in Wallis et al. (2024) between older and more modern instrumentation and point to potential under-adjustment for bias in the early records.

3.2.2 Marine Temperature Uncertainties

IPCC AR6 assessed three principal SST datasets: HadSST4 (Kennedy et al., 2019), ERSSTv5 (Huang et al., 2017), and COBE-SST2 (Hirahara et al., 2014). Each dataset is corrected for changes in instrumentation, principally the use of buckets from 1850 to present, the introduction of engine room measurements starting in the 1930s and the recent transition from ships to buoys. The difference between the SST datasets constitutes one of the largest contributions to the assessed uncertainty in GMST changes, and with just HadSST4 and ERSSTv5 used in the AR6 generation of global surface temperature products (although COBE-SST2 is used in some reanalyses; Section 3.2.4) it is likely that structural uncertainty is underestimated.

While HadSST4 and ERSSTv5 account for broad scale changes in the observing system, some issues are not explicitly dealt with by either. These include errors that correlate at the level of data collections, or countries over multiple years (Chan and Huybers, 2019), such as a truncation error in Japanese data over 1930-1960 (Chan et al., 2019). There are also errors in recorded ship positions (Carella et al., 2017a, Dai et al., 2021) and inconsistencies between coastal land stations and sea-surface temperatures (Cowtan et al., 2018; Chan et al., 2023). Integration of various lines of evidence, including paleoclimate data, indicate that global SST estimates have potential cold biases during the early twentieth century (Hegerl et al., 2018; Chan et al. 2024; Sippel et al. 2024) and warm biases during World War II (Pfeiffer et al., 2017, Chan et al. 2021; Ishii et al. 2025) owing to changes in methods of measurements and national practices. In addition, HadSST4 and ERSSTv5 diverge somewhat starting in the early 1990s (Kennedy et al. 2019) for reasons which are still not well understood. These insights do not necessarily affect warming estimates since 1850-1900, but do affect our understanding of multi-decadal variability. Observational uncertainties for SST for the modern observing system have been estimated (Xu and Ignatov, 2016), but the estimates of historical uncertainty for SST (see e.g. Kennedy 2014) and NMAT predate recent improvements in understanding and need to be re-evaluated to account for new approaches to detecting biases, bias adjustment and quality control.

Since IPCC AR6 there have been three substantive new / revised datasets of SST produced. DCSST (Chan et al., 2024b) represents the first entirely new and novel SST dataset produced for several decades. It compares with coastal stations to derive adjustments, which is structurally distinct from many existing methods. This product is also substantially cooler over 1850-1900 than existing SST products. COBE-SST3 (Ishii et al., 2025) is a new 0.25° resolution version of the long-standing JMA product, which uses correlations between LSAT, NMAT, and SST anomalies to improve SST bias corrections, incorporates other improvements to SST estimates, and includes a rigorous uncertainty model represented by 300 ensemble members. In addition, Chen *et al.* (2022; 2025) produced a new dataset (CMA-SST), which is similar to ERSSTv5 but using more data sources and improved quality control checks. More recently, ERSST has been updated to version 6 (Huang *et al.*, 2025a,b) by using neural networks to improve its infilling technique, improved neighbour checks and innovations to large-scale filtering.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

This update improves estimated SST variability relative to the EOT method, but does not change estimates of global mean temperature.

In addition to datasets of SST, there are also datasets of NMAT (CLASSnmat starting 1880, Cornes et al. 2020, and UAHNMATv1 starting 1900, Junod and Christy 2020). NMAT datasets start later than SST datasets since earlier measurement practices included measurement at local noon, and NMAT datasets exclude observations affected by solar heating. Currently, these NMAT are not used directly in the creation of operational global temperature data sets, however HadNMAT2 (Kent et al., 2013) was used to adjust sea surface temperatures during 1850-1941 (HadSST),1850-2010 (ERSST), 1890-2018 (COBE-SST3), and 1850-2023 (CMA-SST), so NMAT data directly affect almost all current GMST datasets. More recently, adjustments for solar heating have been applied (Cropper et al. 2023), allowing all-hours MAT observations to be used in GloSATref (Morice et al., 2025; Section 2).

3.2.3 Interpolation to account for data void regions

The process of estimating the global average temperature from a finite number of point observations introduces additional uncertainty. The approaches fall into 6 methodological families summarised in Figure 6. There is no single correct averaging method, and the range of methods introduces a range of warming estimates. Ad hoc methods, such as linear interpolation, natural neighbor interpolation, and distance weighted interpolation, played an early role in spatial temperature analysis. More advanced techniques which use the field's estimated spatial covariance to optimally weight nearby observations are generally described, somewhat interchangeably, as Kriging or Gaussian process regression are generally favoured in modern products.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Open Access Earth System Science Data Discussions

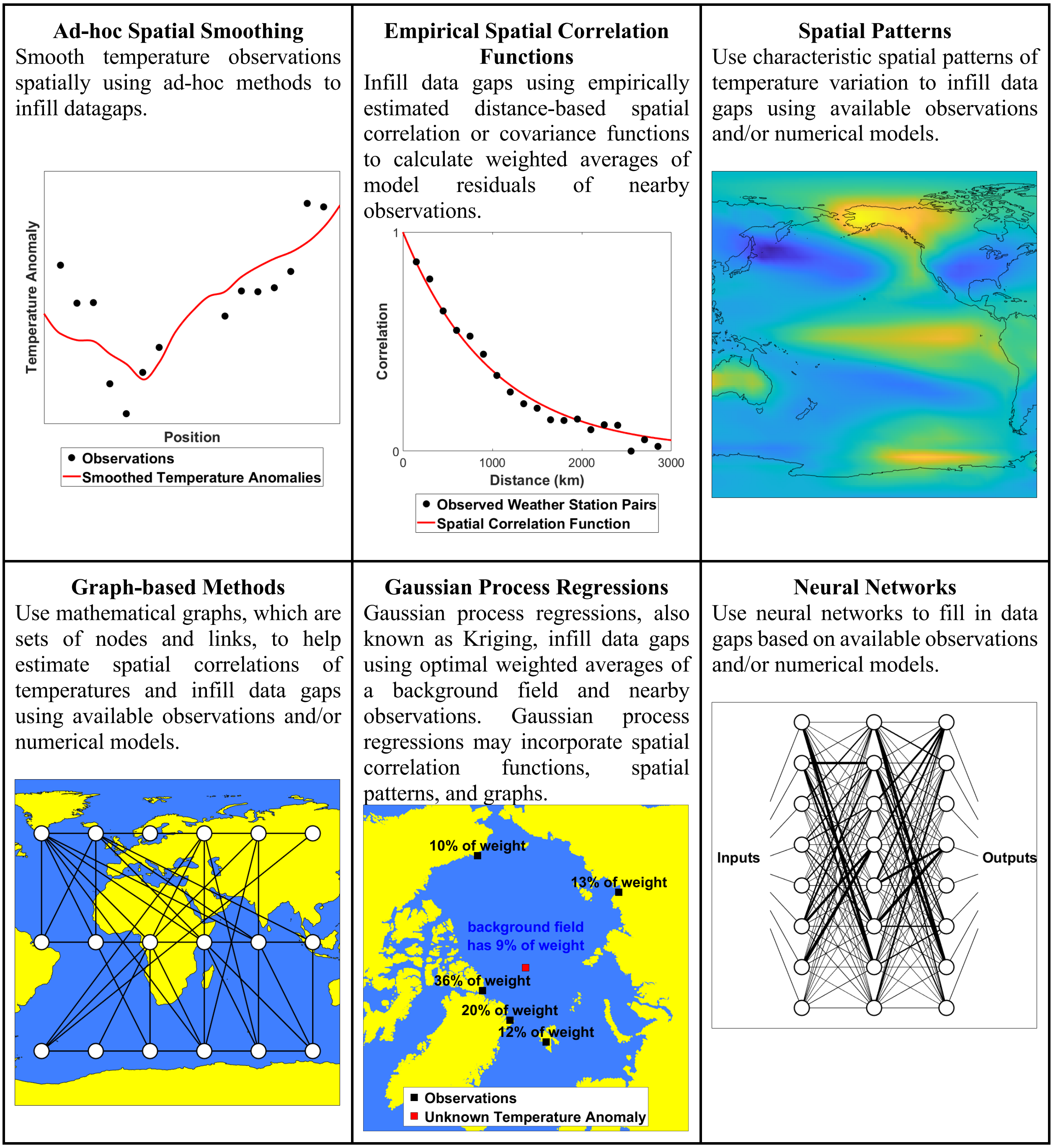


Figure 6. Visual summary of the 6 broad families of approaches to interpolating to create (more) globally complete estimates from the globally incomplete and time-varying observational fields.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Berkeley Earth and Cowtan & Way (2014) were the first to explicitly use Kriging. HadCRUT5 and GloSATref also use Kriging, implemented with a single correlation scale for each surface type (land and sea-ice vs ocean, Morice et al. 2021, 2025). HadCRU_MLE (Calvert, 2024a) and DCENT_MLE (Calvert, 2024b) combined a Kriging-like method with a single EOF to represent ENSO. Generally, a small set of characteristic temperature correlation length scales are used in Kriging-based methods, split by surface type (e.g. land versus ocean). GETQUOCS (Ilyas et al., 2021) applied multi-resolution lattice Kriging to generate a large HadCRUT4-based ensemble that samples uncertainties associated with both infilling and HadCRUT4 bias adjustments. DCENT-I (Chan et al, 2025) used ordinary Kriging with anisotropic and heterogeneous kernels. CMA-GMST (Chen *et al.*, 2025) combined a Kriging-like method with spline interpolation to infill land regions without using distance-based spatial correlation functions. Vaccaro et al. (2021) used a "GraphEM" method that better represents anisotropic features like land–ocean contrasts. They found that more-sophisticated methods had a particularly relevant impact when reconstructing ENSO patterns.

NOAAGlobalTemp v6 (Yin *et al.*, 2024) uses a two step process for both land and sea, first calculating a low-resolution, slowly-varying background field. Residual differences over land are estimated using a Neural Network and the residuals over ocean are estimated using EOTs[19]. A similar EOT-based approach is used in China-MST3.0-Imax (Sun *et al.*, 2022; Li *et al.*, 2025). CMA-GMST (Chen *et al.*, 2025) also uses a similar EOT-based approach for ocean and sea ice regions. GISTEMP uses land observations with distance-weighted spatial smoothing to infill temperature anomalies over land and sea ice (Lenssen *et al.*, 2019). COBE-STEMP3 uses EOFs to simultaneously reconstruct both low and high frequency fields for SSTs (COBE-SST3) and LSATs (COBE-LSAT3). A neural network approach adapted from a method for repairing damaged photographs and trained on reanalysis or model output, was used by Kadow *et al.* (2020) to fill gaps in the HadCRUT4 grid. An update used in IPCC AR6 (Gulev *et al.*, 2021) and subsequent annual updates (Forster *et al.*, 2024) applied the same method to HadCRUT5.

Methodological similarities and differences discussed above, including choice of land and SST datasets are summarised in Table 1. This highlights that despite some diversity there are also similarities between products in both underlying data series and the method of global interpolation. This could potentially lead to a mis-specification of structural uncertainty.

[19] EOTs are conceptually similar to the more commonly used EOFs but are spatially truncated so that they only provide regional inferences.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Earth System Science Data Discussions
Open Access

Global Surface Temperature Dataset	Sea Surface Temperature Dataset	Land Surface Air Temperature Dataset	Sea Ice Concen-tration Dataset	Uses Ad-hoc Spatial Smoothing	Uses Spatial Correlation Functions	Uses Spatial Patterns	Uses Graph-based Methods	Uses Gaussian Process Regression	Uses Neural Networks	Accounts for Climatolo-gical Differences between Sea Ice and Open Sea
Cowtan and Way (2014)	HadSST3	CRUTEM4	HadISST1		✓			✓		
GETQUOCS	HadSST3	CRUTEM4	None		✓			✓		
Vaccaro et al. (2021)	HadSST3	CRUTEM4	None				✓	✓		
HadCRUT5 Analysis	HadSST4	CRUTEM5	HadISST2		✓			✓		
Update to Kadow et al. (2020)	HadSST4	CRUTEM5	None						✓	✓
HadCRU_MLE	HadSST4	CRUTEM5	HadISST2		✓	ENSO Only		✓		✓
Berkeley Earth	HadSST4	Berkeley Earth	HadISST2		✓			✓		✓
GloSAT Reference Analysis	GloSATMAT (MAT not SST)	GloSATLAT	HadISST2		✓			✓		
DCENT_MLE	DCSST	DCLSAT	HadISST2		✓	ENSO Only		✓		✓
DCENT-I	DCSST	DCLSAT	HadISST2		✓			✓		
GISTEMPv4	ERSSTv5	GHCNv4	HadISST2	✓	Land and Ice Only	Sea Only				
NOAAGlobalTempv5.1	ERSSTv5	GHCNv4	HadISST2	✓		✓				Poleward of 65°
NOAAGlobalTempv6.0	ERSSTv6	GHCNv4	HadISST2	✓		Non-Polar Sea Only			Land and Poleward of 65°	Poleward of 65°
China-MST3.0-Imax	ERSSTv6	China-LSAT2.1	HadISST2	✓		Land and North of 65°N			Sea South of 65°N	North of 65°N
CMA-GMST	CMA-SST	CMA-GLST	HadISST2	✓		Sea and Ice Only		Land Only		
COBE-STEMP3	COBE-SST3	COBE-LSAT3	COBE-SST3			✓		✓		

Table 1. Summary of similarities and distinctions between the various extant infilled dataset versions at the time of paper preparation. See main text for citations and further details.

Figure 7 shows differences in global mean estimates arising from interpolation of HadCRUTv5 via a variety of techniques. All approaches tend to increase the long-term warming compared to the non-infilled product, partly associated with rapid warming in the poorly-sampled Arctic (Cowtan and Way, 2014; note that the recent rapid retreat in Antarctic sea-ice may yield similar challenges moving forwards if sustained). However, infilling also affects the area ratio of land and ocean or low and high latitudes, which have likely warmed at different rates. The approach to infilling temperatures over regions of sea ice losses

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

can impact estimates of GMST change since the late nineteenth century by up to 0.1°C (Richardson et al., 2018; Gulev et al. 2021; Calvert 2024a). However, datasets that account for climatological differences but use discontinuous temperature models may overestimate the impact of sea ice loss on GMST change (Calvert 2024a). In contrast, there is no available analysis of sea ice loss effects on GMST trends for datasets where infilling uses EOFs, EOTs, or neural networks trained on reanalysis or climate model output.

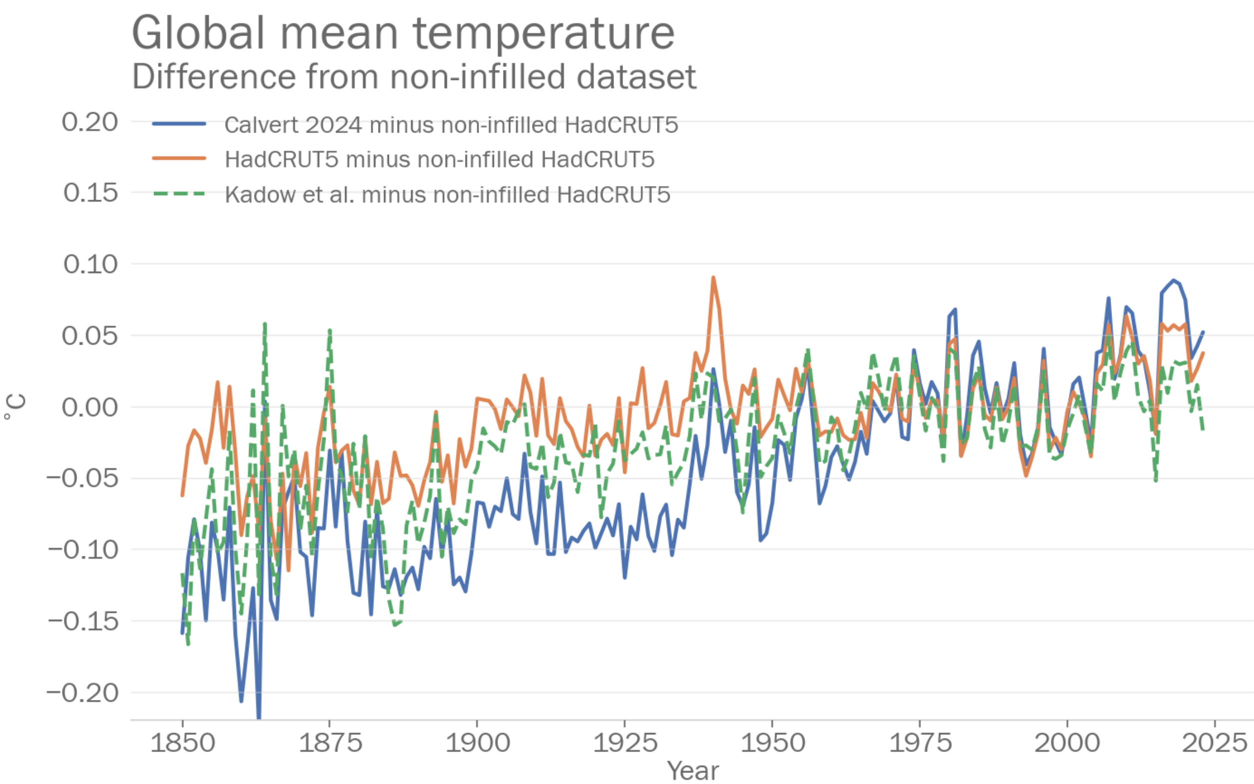


Figure 7: Differences in global mean estimated between 3 distinct infilled versions and the non-infilled version of HadCRUT5, showing the effect of interpolation on estimated global temperature estimates. A value of zero denotes the value of the non-infilled product for a given year and negative deviations are associated with a cooler value in the infilled than the non-infilled version (and viceversa).

3.2.4 Role of Reanalyses

Dynamical reanalyses combine observations with modern data assimilation and numerical weather prediction techniques to create a retrospective analysis. In principle, the result should be a physically coherent climate analysis across multiple variables. However, model and other biases may persist through data assimilation cycles, particularly in sparsely observed regions. With increasing quality (Soci et al., 2024 and references therein), reanalysis products are now gaining widespread usage in monitoring of global surface temperature changes including in the latest IPCC assessment (Gulev et al., 2021) and in WMO climate monitoring statements[20].

Full-input atmospheric reanalyses such as ERA5 (Hersbach et al., 2020) and JRA-3Q (Kosaka et al., 2024) use as many observations as possible to constrain the 3-D atmosphere. They are primarily designed for the period since the mid-20th century due to availability of radiosonde and then satellite observations. They may use SST datasets based on satellite as well as in situ observations, or SST data based only on in situ observations discussed in Section 3.2.2. To estimate temperature changes since 1850-1900 they must be merged with the longer global temperature datasets. This can be achieved by comparing the two types of dataset over a common period (such as 1991-2020) to generate an 'offset' (and its uncertainty) for the change since the pre-industrial period[21]. Full-input reanalyses generally have a latency of days, allowing near-real time monitoring. Unlike traditional datasets which have generally been monthly, these reanalyses enable daily and even sub-daily characterisation of global surface temperature variations.

There also exist sparse-input reanalysis products (such as the 20th Century Reanalysis and ERA-20C) which are constrained only by observed boundary conditions such as SSTs, observed surface pressure, surface wind, and selected time varying forcings such as GHG concentrations. These produce a global warming signal since the mid-19th century (e.g. Compo et al., 2013, Poli et al. 2016; Slivinski et al. 2019; 2021), independent of land-based temperature observations, and are less prone to observation-type changes causing spurious trends than full-input reanalyses. However, they generally track existing in-situ based estimates substantially less well than full-input reanalyses in the modern period.

It should be noted that none of the available reanalyses are produced by data assimilation systems that include a complete representation of known radiative forcings. For example, ERA5 does not represent long-term changes in land use, and 20CRv3 assumes a constant tropospheric aerosol burden. Biases will result, depending on the extent to which the assimilated observations can fully represent the impacts on temperature. Most reanalyses also perform their data assimilation in uncoupled mode such that atmosphere-ocean feedbacks are not explicitly accounted for which can lead to complications (de Rosnay et al., 2022).

3.2.5 Remaining uncertainties and challenges in the estimation of observed global surface temperature changes

3.2.5.1 Do existing global surface temperature products adequately sample the structural uncertainty?

[20] https://wmo.int/publication-series/state-of-global-climate (last accessed: 22/6/24)
[21] e.g. https://climate.copernicus.eu/climate-bulletin-about-data-and-analysis (last accessed: 27/12/24)

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

It is important to recognise that there are fewer methodological degrees of freedom in available global surface temperature change estimates than would be implied by simply counting the number of contributing datasets (Table 1). Datasets generally rely on the same underlying temperature observations. Some global products go further and either share common SST or LSAT products, or the principal differentiation arises solely from post-processing and interpolation method differences. The spread of datasets with shared features inevitably understates the uncertainty compared to having truly independent estimates where all aspects of the processing chain are methodologically distinct. Various possible ways to create 'family trees' to highlight the methodological overlaps are possible (Figure 8 gives an example).

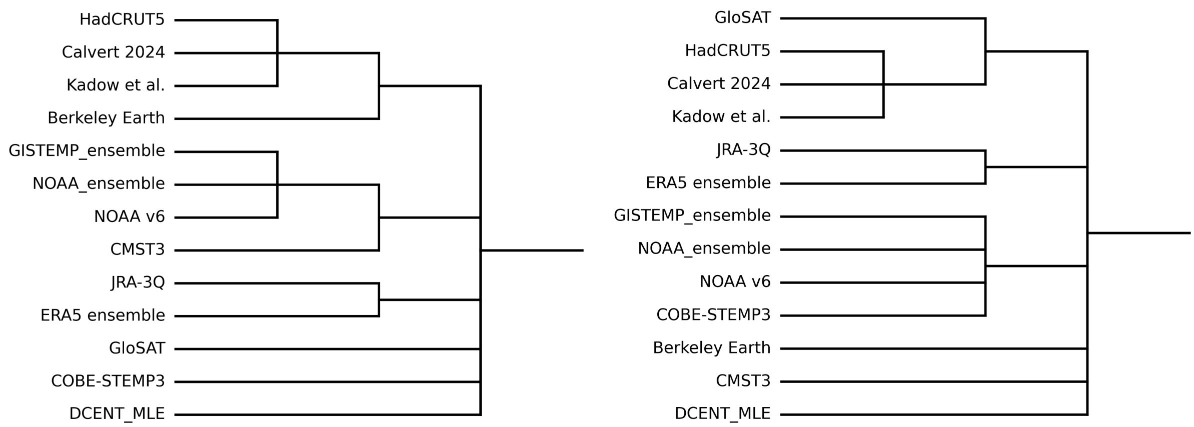


Figure 8: Schematic "family tree" indicating which datasets are more and less closely related at the time of paper preparation (left) based on grouping primarily by SST dataset and then by LSAT dataset and (right) primarily by LSAT dataset in the first instance. In each case reanalyses are grouped separately. Other approaches to such family tree aggregation are possible but would lead to a similar conclusion of fewer degrees of freedom than implied by a naive counting of available products.

To exacerbate the issue of underestimation of methodological degrees of freedom in creating merged products, until very recently those datasets with larger LSAT trends had been merged with SST products exhibiting less warming and vice-versa. This created a false impression of broad agreement in GMST estimates. A simple mixing and matching of LSAT and SST datasets in several long-standing data products leads to a considerably wider spread in estimates (Figure 9).

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

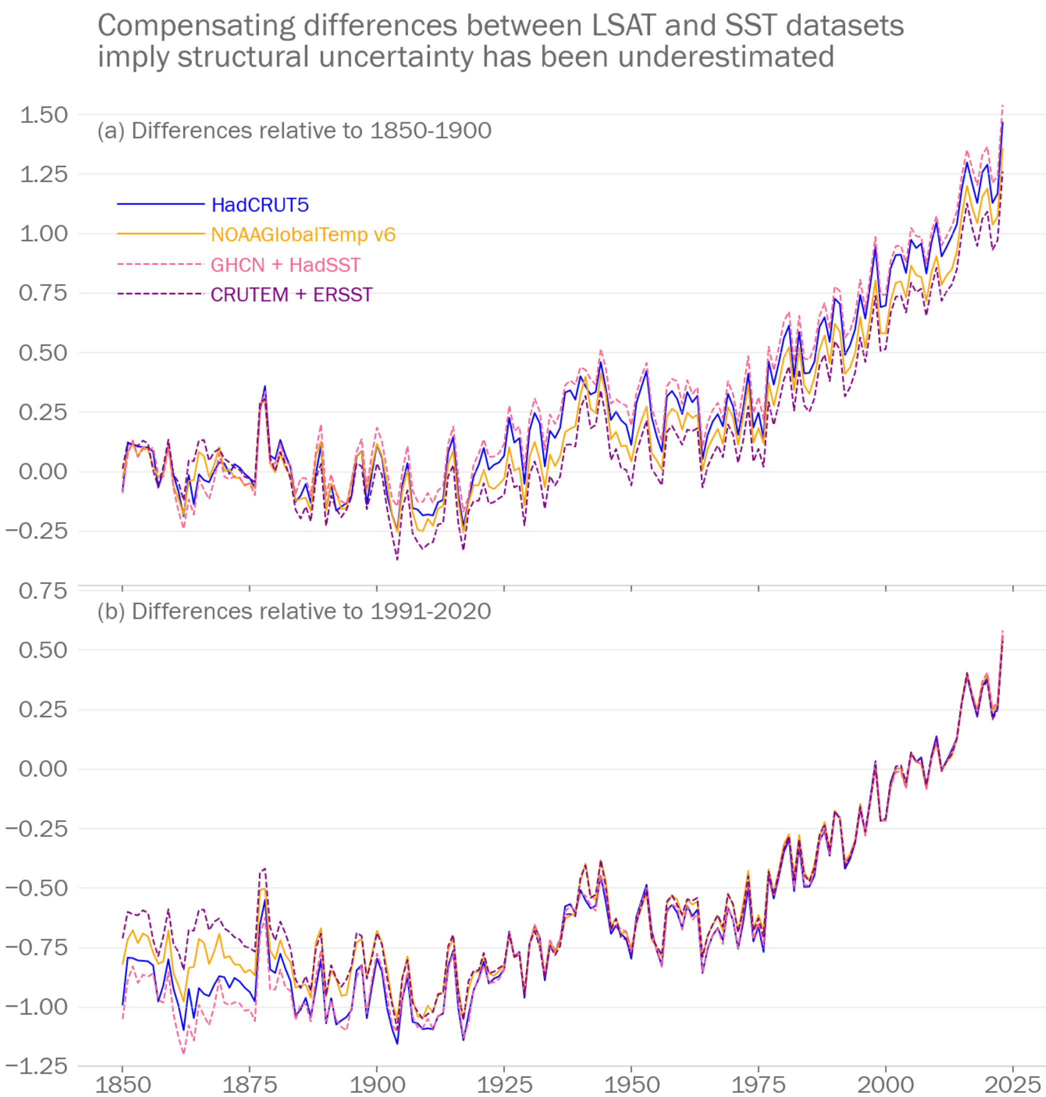


Figure 9. Annual (1850-2023) global mean near-surface temperature expressed as an anomaly relative to (top) the 1850-1900 average and (bottom) 1991-2020. Four datasets are shown, including two standard datasets - HadCRUT5, and NOAAGlobalTemp v6 (solid lines) - and two combinations taking the land and ice contribution from one dataset and combining it with the ocean contribution from another as indicated in the legend (dashed lines). The spread from the two operational combinations (HadCRUT5, NOAA GlobalTemp v6 - solid lines) is considerably smaller than that from the 2 recombinations (dashed lines).

One way to gauge uncertainties in homogenized datasets is to compare LSAT and SST data where they meet along coastlines. Cowtan et al. (2018) found that LSAT records imply SSTs were a couple tenths of a degree cooler during 1850-1900 than HadSST3 and ERSSTv4 estimates (Huang *et al.*, 2015; Kennedy *et al.*, 2011a and 2011b; Liu *et al.*, 2015). Chan et al. (2023) used a more-detailed energy balance model to compare SSTs with their coastal LSAT counterparts and found discrepancies of

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

a similar magnitude prior to the 1940s, depending upon how the LSAT records are homogenized. These discrepancies are not readily reconciled given stated LSAT and SST uncertainties, implying that uncertainty quantification remains incomplete and overly confident. Two recent datasets using coastal LSAT to homogenise SST (Chan et al. 2024; Ishii et al., 2025) support the conclusions of Sippel et al. (2024) that the structural uncertainty was significantly underestimated (in addition to the differences illustrated in Figure 9) at least in the first half of the twentieth century for datasets used in IPCC AR6.

In some cases, dataset developers produce parametric uncertainty estimates given the dataset methodological choices (e.g. Morice et al. 2021, Huang et al. 2020; Ilyas et al. 2021; Sun et al. 2022; Calvert 2024a). Parametric uncertainty estimation involves accounting for the impact of uncertain choices within the chosen methodological framework upon the resulting estimates. There are a very broad range of possible approaches which include: propagation of uncertainty from the individual measurements, assessment of the parametric uncertainty by varying analysis parameters and choices, comparison of subsampled and fully sampled reanalysis fields, maximum likelihood estimation, Bayesian methods, and bootstrapping (e.g. Morice et al., 2021, Lenssen et al, 2019; Lenssen et al., 2024). These parametric uncertainty estimates are therefore rarely, if ever, directly comparable. They either account for a different combination of sources of uncertainty, or use methodologically distinct approaches, or both. Some create ensembles of ‘equi-probable’ estimates whilst others provide probabilistic distributions (Section 5 discusses our use of these ensembles). It is critical to note that tighter parametric uncertainty estimates may simply reflect a more restricted consideration of possible uncertainties (either fewer uncertainties being considered and / or conservative estimates of their magnitude) rather than a higher product quality. Equally, very broad parametric uncertainty estimates may reflect unduly pessimistic assumptions around the possible magnitude of uncertainties. Despite these caveats, there is agreement that these estimates generally increase back in time with a marked spike in many estimates around the time of World War 2 (Figure 10). None of the estimates are large enough to call into question large-scale centennial warming.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Open Access Earth System Science Data Discussions

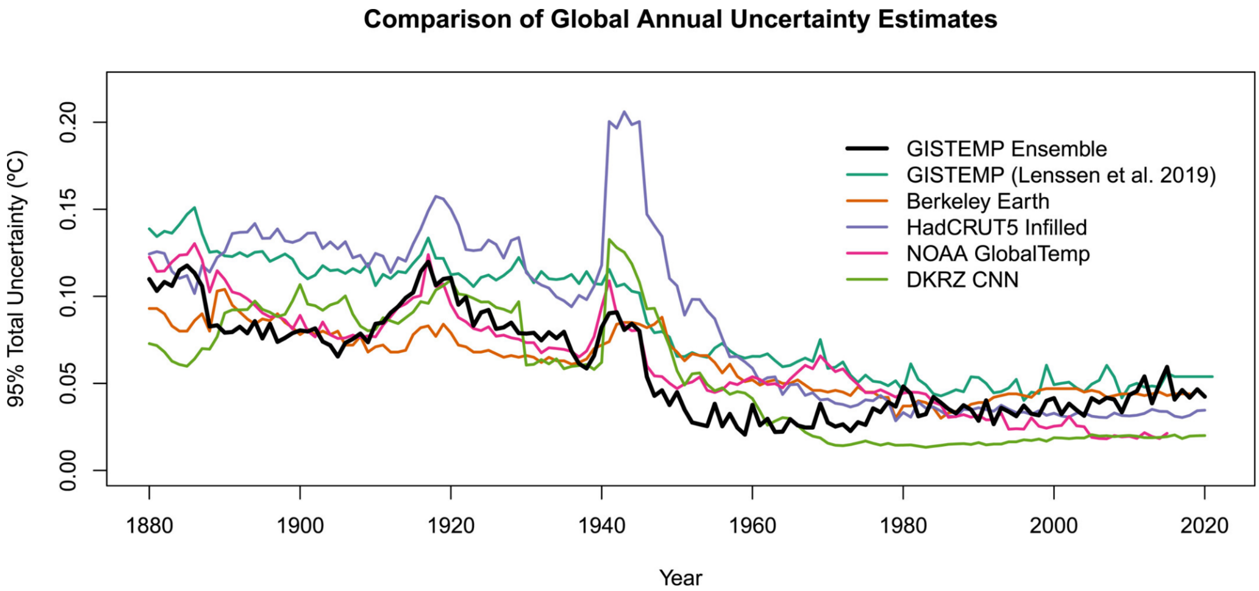


Figure 10. Comparison of quantified estimates of parametric uncertainty from those observational datasets that have provided them. Each trace shows the 95% uncertainty range through time for annual means. Differences arise from the assessment of distinct sources of uncertainty using distinct approaches. All estimates increase back in time with increasing data sparsity. DKRZ CNN is the Kadow et al. dataset. Two versions of GISTEMP uncertainty estimates are shown. Figure from Lenssen et al., 2024 (their Figure 4).

Given the above considerations, we conclude that observational structural uncertainty is currently undersampled and therefore likely misspecified in the available ensemble of in situ based estimates. It is important to stress that this misspecification of structural uncertainty only changes our understanding of precisely how much Earth has warmed, it does not call into question the unequivocal finding that the world has warmed, a finding that rests on multiple lines of evidence of changes arising from diverse observational and proxy sources across the atmosphere, oceans, cryosphere and biosphere (Gulev et al., 2021). The updated datasets used in IPCC AR6 increased estimated warming for 1850-1900 to 2013-2012 by 0.12°C compared with the datasets used in IPCC AR5. They also shifted and increased somewhat the estimated uncertainty range, particularly in the upper bound (Gulev et al., 2021 their CCB 2.3 Figure 1, see also SPM A1.2 and associated footnote 10 in IPCC, 2021). Similarly, the estimated warming from 1850-1900 to the 1986-2005 reference period used in AR5, and which formed the basis for many of the impact studies in AR5 WGII, shifted to be 0.08°C warmer.

Shifts by a similar or larger magnitude could occur in the future. Indeed, new estimates from Chan et al. (2024b) and Calvert (2024a) may result in similar shifts in future assessments. Equally importantly, better sampling of the structural uncertainty may expand or contract the apparent uncertainties, impacting our confidence in any assessment. In this context it is important to screen datasets to remove estimates that persistently do not adequately account for known sources of bias which would unduly widen the spread of estimates (e.g. Gulev et al., 2021; see Section 5 for further discussion on this point).

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Given the importance of monitoring global surface temperatures, a reasonable goal would be to increase the number of broadly independent datasets to numbers commensurate with those for Ocean Heat Content and Global Mean Sea Level each of which had >10 datasets assessed in IPCC AR6 (Gulev et al., 2021). This will require sustained investments with long-term support to create and maintain structurally distinct products.

Better sampling of structural uncertainty can be facilitated by innovation in quality control and homogenisation techniques. There are more techniques that have been developed than are currently deployed in estimation of global surface temperature changes (e.g. Venema et al., 2012, Gillespie et al., 2022, Joelsson et al., 2022). There is also the potential to better benchmark the performance of existing and candidate algorithms against realistic test data sets (Section 3.2.1). Currently lacking are global benchmarks for LSAT and relevant evaluation criteria for the processing of marine observations.

3.2.5.2 Potential for residual common biases across all products

There is a general need for improved understanding of historical data issues. Identified issues that are common to most in situ marine datasets are described in Section 3.2.2. Modern hourly data can aid in quantifying the effect of observation time changes in historical data and of different methods to estimate daily mean temperatures (Gough et al., 2020), though metadata describing observation times is not always available digitally. Rescuing historical parallel measurements series (Wallis et al., 2024 and references therein) and recreating known changes in modern field experiments (e.g. Brunet et al., 2011) can help to assess the impacts of changes in historical observing techniques and practices. There are many fewer such observations for marine observing platforms than for land observing stations despite calls for their recovery (Kent et al., 2017). The issues uncovered over the past couple of decades have highlighted potential data artefacts of the order of a tenth of a degree globally, and a systematic error of such magnitude across all datasets would have a substantial impact upon any assessment of whether 1.5°C has been passed or not. Any hitherto unrecognised data artefacts will, by definition, be common to all datasets to some extent. It would be unduly optimistic to believe that all remaining data artefacts have been identified and quantified to date.

3.2.5.3 Challenges around accounting for sea-ice effects

Uncertainty in sea ice concentrations contributes substantial but poorly quantified uncertainty to instrumental temperature datasets and reanalyses, particularly prior to the availability of satellite records. Gridded Arctic sea ice concentration datasets that go back until 1850, including COBE-SST2, HadISST2 (Titchner and Rayner, 2014), G10010v2 (Walsh *et al.*, 2019), and COBE-SST3, are based on climatologies for much of the period and do not reflect true variability. In particular, for the southern hemisphere, COBE-SST2 and COBE-SST3 use no data prior to the satellite era, whereas HadISST2 may substantially overestimate sea ice concentrations during this period as it relies on a 1929-1939 German climatology and 1947-1962 Soviet climatology, the first of which represents "where ice is known to come farthest towards the open sea" during the climatological period (Orvig, 1955). Calvert (2024a) estimated that the impact of southern hemisphere sea ice on GMST change from 1850-1900 to 2023 is 0.05 °C less warming when using COBE-SST2 rather than HadISST2.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Most sea-ice concentration datasets do not provide uncertainty estimates. ERA5 uses differences between HadISST1 (Rayner *et al.*, 2003) and HadISST2 to represent uncertainty in sea ice concentrations prior to 1979 (Hirahara *et al.*, 2016). But these differences largely reflect adjustments to account for summer melt ponds. Relevant infilled instrumental temperature datasets and reanalysis almost exclusively use HadISST2 prior to the satellite era (Table 1) and thus undersample structural uncertainty in sea ice concentrations; two exceptions are JRA-3Q and COBE-STEMP3. Overall, sea ice contributes an uncertainty on the order of 0.1°C to GMST change since the late 19th century (Section 3.2.3).

Historic sea ice concentration estimates could be improved to better account for infilling and measurement uncertainties. Satellite-based estimates could be improved to better account for summer melt ponds and thin ice (Ivanova *et al.*, 2015). Furthermore, datasets could incorporate additional data from northern hemisphere sea ice observations (Hill, 1999; Vinje, 1999; ACSYS, 2003; Ruffman *et al.*, 2003; Shapiro *et al.*, 2006), whaling log books (Cotté and Guinet, 2007; de la Mare, 2009; Teleti *et al.*, 2019; Walsh *et al.*, 2019), early satellites (Gallaher *et al.*, 2014), and Antarctic sea ice observations (Edinburgh and Day, 2016; Love and Bigg, 2023; Martin *et al.*, 2023). In addition, there have been recent improvements to sea ice reconstructions that combine temperature, pressure or paleoclimate records with statistical or data assimilation methods (Brennan *et al.*, 2020; Samakinwa *et al.*, 2021; Yang *et al.*, 2021; Brennan and Hakim, 2022; Fogt *et al.*, 2022; Dalaiden *et al.*, 2023; Goosse *et al.*, 2024; Maierhofer *et al.*, 2024; Semenov *et al.*, 2024; Dalaiden *et al.*, 2025; Cooper *et al.*, 2025). Figure 11 shows sea ice extent anomalies for a subset of available datasets and reconstructions. Panel b of Figure 11 shows that the large decline in HadISST2 Antarctic sea ice since 1939 is inconsistent with five reconstructions, suggesting that its use of the German climatology prior to the satellite era is unreliable.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Open Access Earth System Science Data Discussions

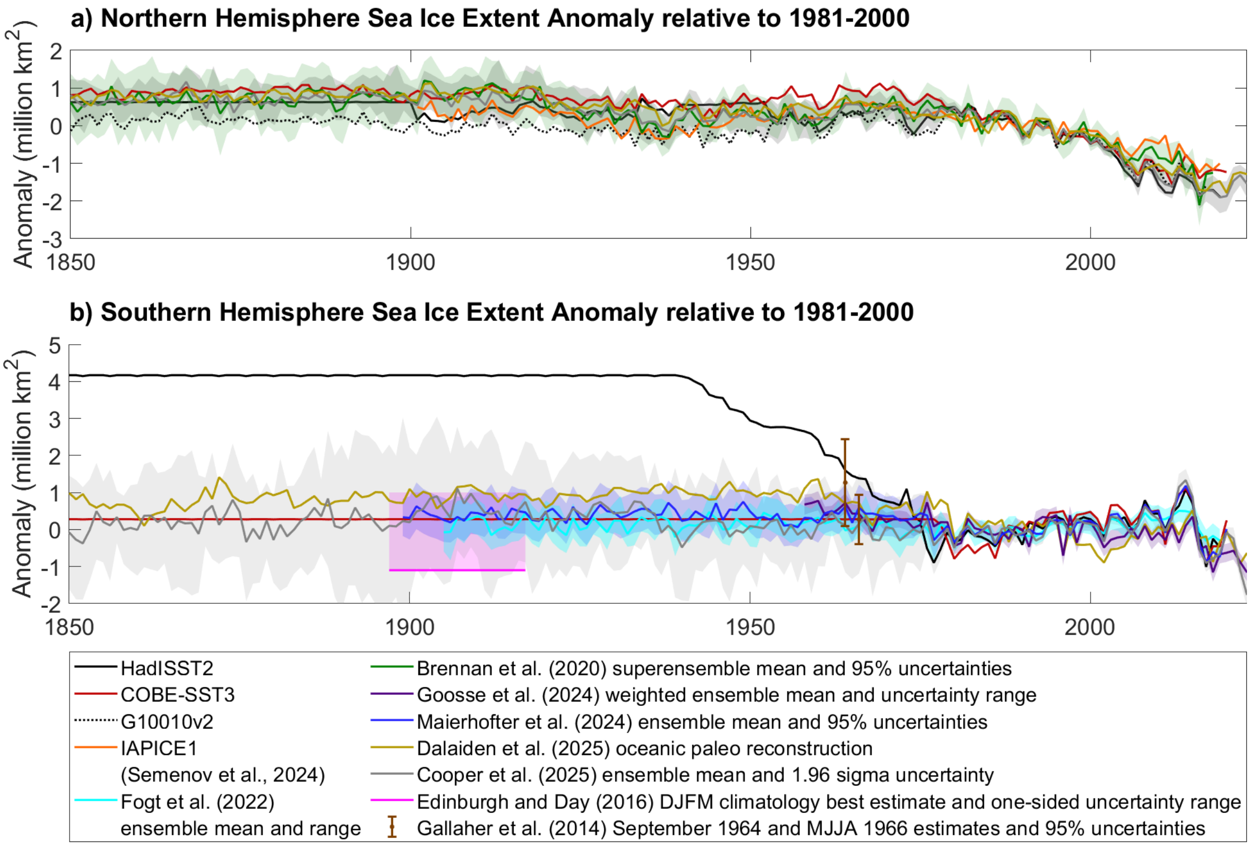


Figure 11: Sea ice extent anomalies from 1850 to 2023. Anomalies are annual mean anomalies unless specified otherwise. Gallaher et al. (2014) anomalies are calculated relative to the NSIDC sea ice extent index (Fetterer et al., 2017) and its uncertainties are calculated by assuming uncorrelated uncertainties between months. Edinburgh and Day (2016) anomalies are calculated relative to the bootstrap algorithm results of the NSIDC Climate Data Record (Meier et al., 2021). The Dalaiden et al. (2025) oceanic paleo reconstruction is shown as their preferred reconstruction; their atmospheric reconstructions are not shown, but include a 95% uncertainty on the order of ±1 million km² for the Antarctic. Cooper et al. (2025) mean and uncertainties were calculated from the publicly available 14 member subensemble.

3.2.6 The potential to improve the observational basis for all products

For reasons relating to historical and ongoing data policies and data management a vast volume of observational data are not currently openly available (e.g. Noone et al., 2021), particularly for earlier records (Bronnimann et al. 2019b). To address this, the WMO has adopted a new unified data policy, which for the first time requires NMHSs to share a minimum subset of historical holdings and encourages sharing of all holdings (WMO, 2022). If all NMHSs were to share these holdings then many data gaps would be filled. The WMO has also endorsed and is starting to make operational the Global Basic Observing Network, in part through the associated Systematic Observations Financing Facility, which will fill additional critical gaps if fully enacted (WMO, 2021), but presently does not consider many observations important for climate, including MAT.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Many data records, particularly prior to 1950, remain in hard copy or image format only (e.g. Allan et al., 2011, Brönnimann et al., 2019b). Efforts are required to expedite the rescue of these records, from inventorying and cataloguing through to digitisation[22]. Citizen science projects (e.g. Teleti et al. 2023, Lakkis et al., 2023, Gergis et al., 2023, Hawkins et al., 2023) (Figure 12), and classroom-based projects (e.g. Ryan et al., 2018, Noone et al., 2024) have delivered progress, but not yet at the scale required. AI-techniques for digitisation may prove useful but are as yet unproven at scale (Prasad et al. 2020, Ziomek and Middleton, 2021, Li et al. 2023, Vercruysse et al., submitted, Lorrey et al. 2022). Even when AI-based techniques become mature, human input will still be required. This can be mitigated by recent advances in semi-supervised training regimes for AI models (Singh and Middleton 2024) or deployment into human-machine teaming environments where AI-generated data is reviewed in efficient "check and correct" type quality assurance cycles, allowing AI data rescue errors to be managed effectively prior to integration with other data sources and downstream use.

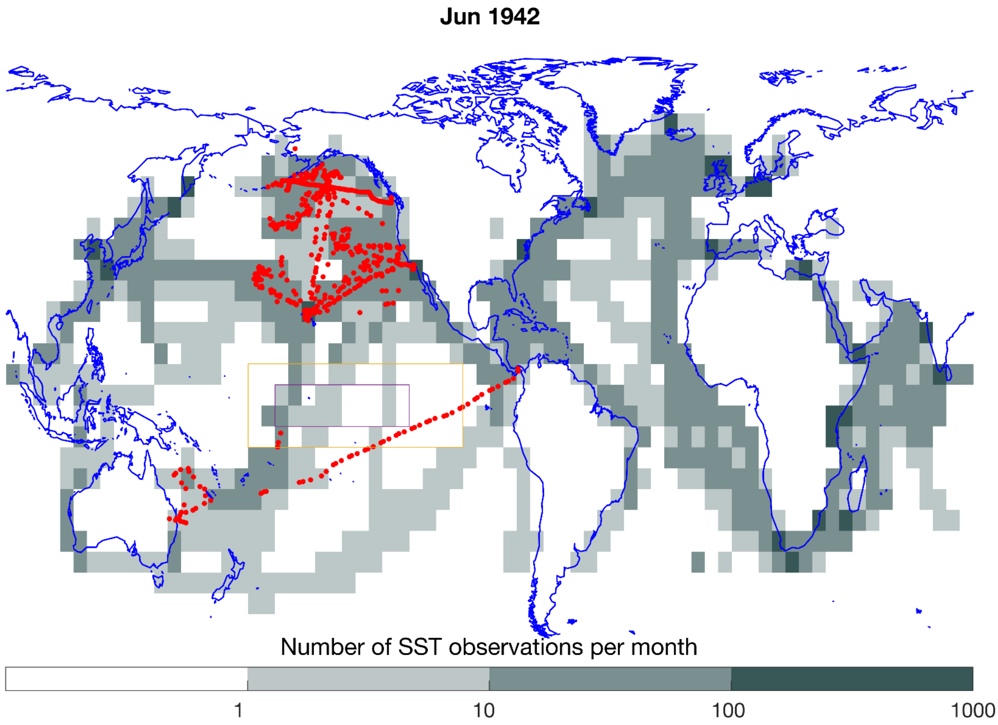


Figure 12: Example of how data recovery can help fill gaps in existing reconstructions. For June 1942, thousands of newly rescued SST observations are now available in the Pacific (red dots), including in regions with no or very few existing observations (Teleti et al. 2023). Shaded greys are the number of SST observations used in HadSST4 (Kennedy et al. 2019) on a log scale. The digitisation of these US ship logs for 1941-1945 has provided millions of new sub-daily SST, MAT and pressure observations but many still non-digitised logbooks exist for the same time period.

[22] https://datarescue.climate.copernicus.eu/en/c3s-data-rescue-service last accessed 3/2/25

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Improved and sustained management of the fundamental data record is imperative. Today's state-of-the-art datasets may seem antiquated in 20 years, whereas curation of the fundamental data record is forever. Without sustained stewardship following "FAIR" principles (Wilkinson et al. 2016), data can be permanently lost. Recent efforts to internationalise support for data stewardship (Noone et al., 2021) need to be deepened and entrenched to make this stewardship less prone to single points of failure, more robust and traceable and to ensure that issues are corrected, newly rescued data and metadata are readily available for dataset improvements, and appropriate credit and support is provided to the data originators.

4 Approaches for estimating the current global warming level

4.1 Global surface temperatures have already exceeded 1.5°C above 1850-1900 levels on at least a temporary basis

Owing to natural variability, individual days, weeks, months and years will differ from the longer-term climate mean state (Gulev et al., 2021, IPCC, 2021b). In a warming climate, any temperature level will typically be temporarily exceeded multiple times before that level of global warming is exceeded on a more sustained decadal or multi-decadal basis (e.g. Rogelj et al., 2017, Lee et al., 2021, Nicklas et al. 2025, Trewin, 2022). A global mean temperature anomaly of 1.5°C relative to 1850–1900 has been exceeded on a daily basis multiple times in the ERA5 reanalyses since late 2015 (Figure 13). This same dataset indicates that daily global surface temperature anomalies approached and possibly exceeded 2°C in November 2023 and February 2024 (Figure 13). Several months have also exceeded 1.5°C across multiple datasets, most notably from July 2023 onwards. There is general agreement that 2024 was the warmest calendar year on record, but in the WMO assessment 2 datasets have reported values for the year that do not exceed 1.5°C, while the remaining 4 do exceed 1.5°C, with a consensus estimate of 1.55°C and a range of 1.46 to 1.62°C[23] (Figure 5).

The recent sustained period above 1.5°C has been analysed in combination with CMIP6 simulations to infer that the long-term mean has probably now exceeded 1.5°C (Cannon, 2025) or that we are now somewhere within the 20-year period during which 1.5°C will be exceeded (Bevacqua et al., 2025). Note the distinct framings here - Cannon argues the 20-year centred mean in 2024 will likely have exceeded 1.5°C, based on the fact that the first instance of 12 consecutive months above 1.5°C typically occurs after the long-term mean exceeds this level in simulations, whereas Bevacqua et al. argue that 2024 will lie within the first such 20-year period and show that historically first exceedances have typically preceded long-term exceedance. The latter result would be consistent with IPCC (2021) whereas the former would imply much earlier exceedance than assessed by IPCC. Both results also assume that there have been no significant deviations in the radiative forcings relative to what was used in CMIP6.

[23] Note the slight distinction with the Forster et al. 2025 numbers given in Section 1 which results from slight differences in dataset selection and methods. Adopting a consistent method, as suggested in this paper, would remove this potential for confusion.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Internal interannual-to-multidecadal variability can cause individual years to be of the order 0.2°C (Trewin, 2022, based on observations) / 0.25°C (IPCC, 2021, based on CMIP6) warmer or cooler than the long-term mean state. Accounting for internal variability is crucial for climate impacts projections in the near-term (Schwarzwald and Lenssen 2022). There are also potential complications arising from variations in natural and anthropogenic short-lived climate forcings including, for the current period, the Hunga Tonga volcanic eruption, the marine bunker fuel regulations and rapid sulphur emission reductions in individual countries (Jenkins et al., 2023, Yuan et al., 2024, Bevacqua et al., 2025, Forster et al, 2025, Samset et al., 2025). While reporting of annual global surface temperatures serves an important role, such short-term fluctuations can significantly complicate interpretation and need to be accounted for in a systematic way. Annual temperatures alone are therefore imperfect for tracking progress towards a specific longer term temperature goal. While methodological approaches and definitions for assessing the current long-term temperature in the context of the 1.5°C target exist (e.g. IPCC SR1.5) and are operationally updated (e.g. Forster et al., 2025), the current proximity and rate of approach towards the LTTG motivate a thorough and broad assessment of which approaches or family of approaches are most robust under a variety of warming scenarios. Section 4 reviews published approaches for separating the long-term warming "signal" from the "noise" arising from observed temperature fluctuations and assesses them against a range of test cases to attempt to answer this need.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

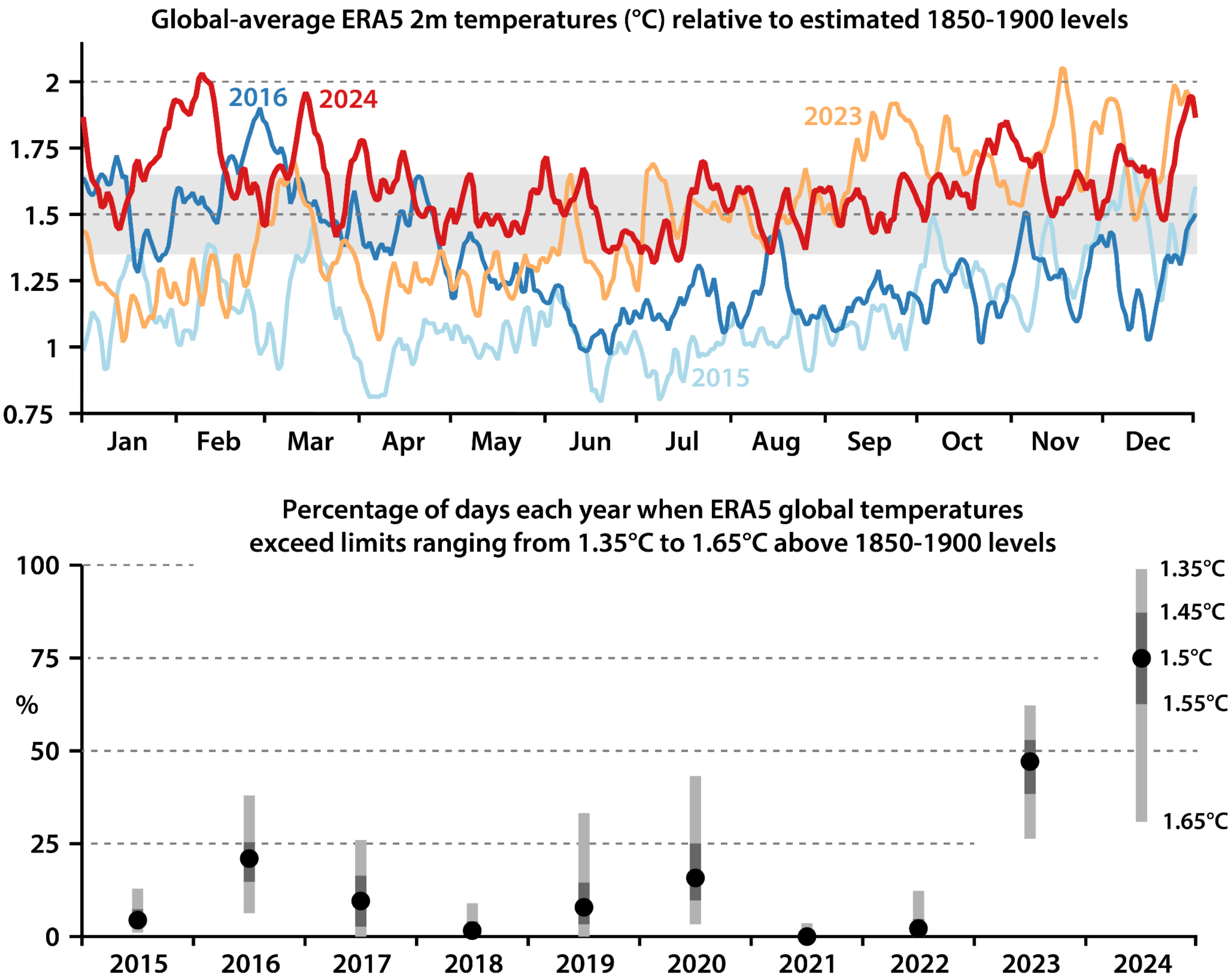


Figure 13. Top-panel: Daily global average temperatures for 2015, 2016, 2023 and 2024 showing occurrences of exceedance of 1.5°C warming according to the ERA5 reanalysis (rebased to 1850-1900 using a range of existing long-term surface temperature estimates). The first such occurrence was in late 2015. The grey-shaded band shows an indicative uncertainty range from 1.35°C to 1.65°C. Lower panel: Percentage of days each year when the ERA5 global-mean temperature exceeds 1.35°C, 1.45°C, 1.5°C, 1.55°C and 1.65°C above 1850-1900. The figure can also be interpreted as indicating the uncertainty in the percentage of days exceeding 1.5°C, for a total uncertainty range of ±0.15°C (light grey bars) and for an uncertainty range of ±0.05°C (dark grey bars), the latter being indicative of uncertainty relative to a modern baseline rather than 1850-1900. Source: ERA5. Credit: C3S/ECMWF

4.2 Ambiguity in the Paris Agreement's specification of the long-term temperature goal (LTTG)

Section 1 articulated several reasonable interpretations of Article 2 (see also Section 6), but the general interpretation is that the LTTG refers to multidecadal levels of globally averaged warming. The Paris Agreement does not provide an operational definition, but one is needed to assess whether a given LTTG is exceeded. The following are three key considerations for defining a specific warming level, above and beyond what constitutes pre-industrial (see Section 2).

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Consideration 1: Over what timescale and using what metric can progress towards the LTTG be assessed?

There is no formal definition of 'longterm' in the Paris Agreement or the UNFCC process. COP 27 (UNFCCC, 2022) comes closest in noting that global average temperature in the context of the LTTG "is assessed over a period of decades". A 30-year mean surface temperature was first reported in 1935 by the IMO as a "standard climate normal" and was used as one part of the IPCC SR1.5 warming definition in 2018. In 2021, the IPCC AR6 used a 20-year centred average to define when projections under various scenarios reach and then remain above a particular Global Warming Level (GWL; Lee et al., 2021, IPCC, 2021). Both definitions are, at least broadly, consistent with the COP 27 clarification.

Multi-decadal averaging filters out a large portion of high-frequency natural variability and is simple to compute (Trewin, 2022). However, such averages do not produce smooth trend estimates and can be systematically biased by longer-term variability (e.g. Medhaug et al., 2017 and references therein for a discussion of the 'hiatus'), and nonlinear or natural forcing (see Consideration 2). They are also inherently lagging. Sections 4.2–4.4 describe approaches for estimating warming over a "period of decades" which may avoid some or all of these limitations.

Consideration 2: Are we assessing the underlying anthropogenic warming or the actual warming irrespective of cause?

Article 1 of the Paris Agreement states that the Agreement inherits the definitions of Article 1 of the Convention (UNFCCC, 1992) which defines climate change as:

> *"a change of climate which is attributed directly or indirectly to human activity that alters the composition of the global atmosphere and which is in addition to natural climate variability observed over comparable time periods"*.

As a result, the Paris Agreement's LTTG is often interpreted as referring to human-induced warming. The IPCC's Special Report on Global Warming of 1.5°C (IPCC, 2018) used human-induced warming as a key feature of its definition of "warming" to determine when 1.5°C would be reached, although the IPCC has long defined climate change and global warming as encompassing both natural and human-induced drivers. Nonetheless, both SR1.5 and AR6 WGI reported human-induced warming levels in their SPMs alongside the estimates of observed warming to date.

Assessing human-induced warming could help to avoid critical misperceptions about the level of sustained warming, which can be temporarily masked by natural volcanic activity, solar forcing[24], or extreme manifestations of internal variability (see for example, Allen et al 2018, Eyring et al., 2021, Otto et al. 2015, Haustein et al., 2017). It does, however, inevitably, require an additional modelling step (even if just a simple linear operator on forcings) which can increase complexity. While human-

[24] Note that in most future climate scenario runs some repeating solar cycle is present, albeit which may vary from the future observed solar behaviour in important ways

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

induced warming very closely matched total warming averaged over 2014–2023, this has not been the case for much of the last century, as naturally-forced components have often exceeded 0.1°C for multi-decadal periods (Forster et al. 2024, their Figure S5).

The counter-argument to use of a human-induced warming metric is that climate impacts and adaptation requirements depend on the actual sustained warming irrespective of its underlying cause(s). If the LTTG was motivated primarily to avoid specific impacts, then the realised temperature change is the relevant metric. In summary, both realised warming and underlying anthropogenic warming have intrinsic value from a policy perspective in the context of the LTTG.

Consideration 3: How timely does the assessment need to be?

If the policy aim is to hold the temperature increase to below some level, it is important to know when that level is reached in near real-time rather than years later. However, the estimated warming must also be robust to avoid criticism that it is either alarmist or conservative and to ensure it serves as a reliable basis for informing policy changes. If a delayed estimate improves robustness, this could be an advantage. Since effective mitigation policies take several years to a decade to materially shift the emissions trajectory, timely detection of a 1.5°C warming level could help minimize further delays in policy adjustments — though the link between providing warming estimates and policy responses remains uncertain. Additionally, public support for mitigation efforts may benefit from access to reliable, up-to-date information about climate change, reinforcing the need for timely and accurate warming assessments.

It will be shown in the remainder of Section 4 that multiple methods are in principle able to provide accurate estimates of both global warming and when global warming crosses any particular level in a timely manner. The IPCC AR6 WGI approach of using a centred 20-year running mean of realised global temperature will be used as a baseline comparator, but this does not imply a normative statement that this is the best possible approach that all methods *should* approximate; indeed it is the purpose of this analysis to review appropriate methods.

4.3 Approaches for estimating the present warming level using observed temperature records

Observed global temperature change at time t can be separated in either a physical or a statistical way (Glaser et al., 2024). An assumed form for a physical separation into components representing forced change (F) and internal climate variability (ICV) with some observational error is:

$$\Delta T_{obs}(t) = \Delta T_{F}(t) + \Delta T_{ICV}(t) + \epsilon(t) \quad (1)$$

The forced and ICV components could also be further decomposed based on the forcing agent or mode of ICV. Reasonable interpretations of the LTTG generally imply that it refers to ΔT_F, or some component thereof (e.g. anthropogenic $\Delta T_{F,anthro}$).

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Meanwhile, statistical decompositions will typically return some fit estimate and a residual term, regardless of which $\Delta T(t)$ term in (1) is being fitted.

$$\Delta T(t) = \Delta T_{fit}(t) + \Delta T_{rsd}(t) \tag{2}$$

Once again, the fit component may be further decomposed, for example into a time-dependent part and components that are a function of ICV indexes. To understand statistical results in the context of a LTTG, the terms in Eqs. (1) and (2) must be related to each other (Glaser et al., 2024). For example, $\Delta T_{fit}(t)$ is sometimes estimated by smoothing or a trend and it is implicitly assumed that $\Delta T_{ICV}(t) + \epsilon(t) \approx \Delta T_{rsd}(t)$, and that $\Delta T_{fit}(t) \approx \Delta T_F(t)$.

The following subsections introduce a large, but non-exhaustive, set of approaches to using observed records. Many of these are then assessed in Section 4.6.

4.3.1 Moving averages

Moving averages have been broadly used including in assessments by IPCC (IPCC, 2021) and WMO. Such measures are, however, by definition lagging indicators. Averaging over only the preceding 5 or 10 years provides a more recent estimate at the cost of higher uncertainty, although a 10-year average has only slightly higher uncertainty than a 30-year average (Trewin, 2022); while 20- or 30- year centered means have been used as retrospective assessments (see supplement for details). A simple average is constant within its averaging window so any non-constant $\Delta T_F(t)$ is added to $\Delta T_{rsd}(t)$, which will inappropriately inflate standard error estimates. An alternative approach, as used by the IPCC, is to define the uncertainty in the (multi-)decadal average using the spread of observationally based estimates (Section 3). This is equivalent to assuming that the error in the mean is only generated by $\epsilon(t)$ in Eq. (1), and that the contribution of $\Delta T_{ICV}(t)$ is negligible over the averaging timescale. Trewin (2022) proposed applying an offset to an 11-year trailing mean so that it provides a less biased estimate of the 20-year mean centred on the current year. The proposed offset, however, is based on post-1970 observed differences between 11-year trailing and 20-year centred means, so it is valid only if there are no changes in the underlying warming rate (Trewin, 2022).

4.3.2 Long-Term Fits

One approach is to fit a $T_{fit}(t)$ to the complete $T(t)$ time series using a statistical optimisation procedure. The method used should appropriately account for the fact that $T_{ICV}(t) + \epsilon(t)$ is strongly autocorrelated. While a single linear fit may be used, $T_F(t)$ is nonlinear since the preindustrial period so polynomials (e.g. Hawkins and Sutton 2009; Lehner et al. 2020) or multiple continuous piecewise linear fits (e.g. Livezey et. al. 2007, Cahill et al., 2015, Ruggieri and Antonellis 2016, Yu and Ruggieri, 2019) have been applied. With long-term fits the less reliable and older observations can unduly influence estimates of the recent trend. Some long-term fits greatly down-weight data from years far from the one for which the current warming level is being estimated. As such the boundary between long- and short-term fits (next subsection) is not distinct.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

4.3.3 Short-term (Local) Fits

"Local" fits heavily weight recent observations to address the issue of influence arising from distant past observations. For instance, IPCC SR1.5 (Allen et al., 2018, their Section 1.2.1) uses the 30-year mean of human-induced warming centred on the present day by extrapolating the recent 15-year secular trend; recently, human-induced warming has been approximately linear, so linear fits of the most recent 15-years of human-induced warming have been used in annual global warming assessments (Forster et al., 2023, 2024, 2025). Another proposal is continuous local regression (LOWESS or a Savitzky–Golay filter) that allows both differential weighting of temperature measurements and higher-order polynomial splines. Implementations for GMST records include those by Clarke and Richardson (2021) and Mann (2004, 2008). We also include an adaptation of the method implied by Cannon (2025), which extrapolates from the coolest monthly temperature achieved in any 12-month window, as detailed in the supplement.

4.3.4 Smoothing spline trend estimates with serially correlated residuals

The regression methods above assume a given functional form for $\Delta T_{fit}(t)$. A more flexible approach is to model $\Delta T_{fit}(t)$ non-parametrically using smoothing splines. Such an approach can be easily implemented for annual temperature data by fitting Generalised Additive Models (see Box 2.2 (pages 179-180) of Chapter 2 of the IPCC 5th assessment report (Hartmann et al., 2013)). Visser et. al. (2018) also used cubic splines (with 7 degrees of freedom and Monte Carlo simulations) to analyze GMST.

4.3.5 Simple state space models

Trends can also be represented using non-stationary stochastic processes such as random walks or integrated random walks rather than by deterministic functions of time (Chapter 5, Chandler and Scott, 2011). If the observations can be considered to be noisy measurements of the underlying state, then time series can be modelled using what are known as state space models (Durbin and Koopman, 2021). A simple example of such an approach, and how to estimate its components using a Kalman filter and smoother, is presented in Example 6.2 in Chapter 6 of Shumway and Stoffer (2016). More realistic dynamic linear models based on energy balance models of the climate system are now being shown to be reliable for estimating the climate trend and climate sensitivity from historical data, e.g. Bennedsen et al. (2023) and Section 4.4.3. below.

4.3.6: Methods for filtering out influences from internal variability

Other methods attempt to filter out known influences from modes of internal variability, volcanic eruptions or similar. The filtering may be purely observations based, or taken from emergent relations in Earth System Models. A simple example is to use multiple linear regression to remove observed temperature contributions that correlate with the Multivariate El Niño Index index (MEI) (Foster and Rahmstorf 2011). Approaches include joint influences of Atlantic and Pacific modes of variability (Wu et al. 2011, 2019), pattern recognition (Wills et al. 2020), spatial EOFs (Chen and Tung, 2018), and statistical learning

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

(regularized linear models) based on historical model output (Sippel et al. 2019). A recent attempt used model-derived Green's functions to estimate and remove short-term $\Delta T_{ICV}(t)$ contributions (Samset et al., 2022, 2023).

4.3.7 Summary of statistically-based methods considered

Figure 14 presents one or more exemplar techniques from each class outlined in Sections 4.3.1 through 4.3.6. Relatively simple techniques outlined in Sections 4.3.1 and 4.3.2 are illustrated by the first three panels (30yr/20yr centred, 10yr lagging and OLS fit). The 10yr lagging and OLS fit methods are outliers compared to the remaining techniques which better approximate the underlying data series. While the centred means closely resemble the other techniques where these estimates are available, they do not extend within the current decade. The approaches are assessed further in Section 4.6.

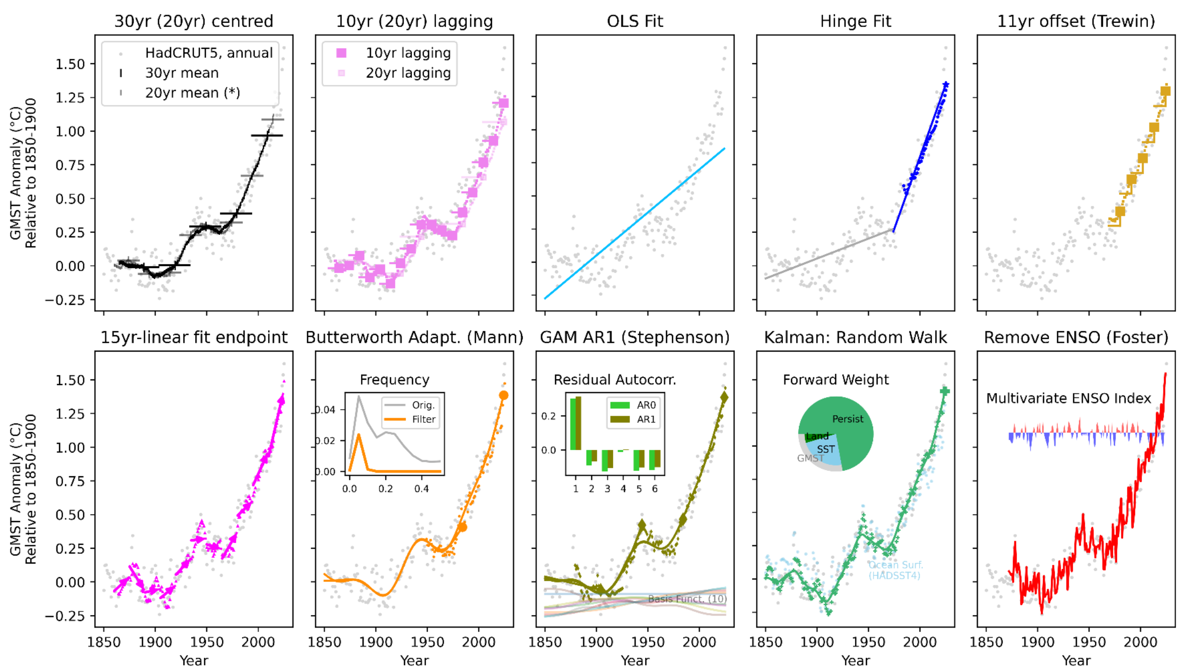


Figure 14. Illustrative example methods (one or more per class) introduced and discussed in Section 4.3. All panels use the HadCRUT5 dataset over 1850-2024 (selected solely for illustrative purposes) with annual values shown in grey. The method results are shown in colour in each panel. For description of methods see sections 4.3.1-4.3.6 text. Within most panels, there are smaller coloured dots showing all results for each method, as well as an enlarged dot and associated lines demonstrating how that method works for selected years. The OLS Fit panel only shows one fitted line through 2024. As of March 2025, HadCRUT5 has released an annual assessment for 2025, but due to the uncertainty in this assessment the latest selected year for all methods covers through 2024.

4.4 Approaches that use both forcing and observed temperature records

Most of the recent long-term temperature trend is attributable to anthropogenic forcings (IPCC, 2021, Forster et al. 2024). However, as discussed in Section 4.2, natural forcings and internal variability also contribute to $\Delta T_{obs}(t)$. The second major

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

class of approaches examined here additionally incorporate information on radiative forcings, and can be used to attempt to attribute a quantifiable fraction of the long-term $\Delta T_{obs}(t)$ to anthropogenic activity. The physical disaggregation assumed is:

$$\Delta Tobs(t) = \sum_i \Delta T_{F,i}(t) + \sum_j \Delta T_{ICV,j}(t) + \epsilon(t) \qquad (3)$$

which follows eq. (1) but with the forced component split into terms $\Delta T_{F,i}(t)$ and the ICV component into terms $\Delta T_{ICV,j}(t)$. Methods described in this section attempt to estimate individual terms, such as the warming due to natural forcings, well-mixed greenhouse gases and/or other human forcings (such as aerosols), and/or their combinations, such as warming due to all anthropogenic activity, or warming due to all forcing regardless of the cause.

Methods for attributing warming contributions are non-exhaustively introduced below. The methods differ substantially in complexity and with regard to the level of granularity with which they are capable of disaggregating $\Delta T_{obs}(t)$. In contrast to temperature-only methods, these approaches use multiple streams of data which may increase or decrease the overall uncertainty depending on their quality, and may involve latency tradeoffs[25].

4.4.1 The quasi-linear relationship between carbon dioxide concentrations and global temperatures

An extensive history of studies have established that, despite the complex dynamics at play in the Earth system, there is a near linear relationship between transient global mean temperature and cumulative CO_2 emissions (e.g. Matthews et al., 2009, MacDougall, 2015, MacDougall & Friedlingstein, 2015), which is (mostly) path independent (Seshadri 2015); this relationship is encapsulated by the transient climate response to cumulative emissions of carbon dioxide (TCRE) and is the basis for calculation of (remaining) carbon budgets. In addition, the airborne fraction, which relates cumulative CO_2 emissions and atmospheric CO_2 concentration increases, has also been approximately stationary to date , although this may not continue at higher levels of warming and CO_2 concentration (Canadell et al. 2021). These two observations indicate a quasi-linear relationship between GMST and CO_2 concentration changes, which is useful owing to these being well-established and directly observable quantities that are operationally updated with little lag. Jarvis and Forster (2024) exploited this apparent linearity, assuming atmospheric CO_2 concentrations as a proxy for all anthropogenic activity; a simple linear regression can then be used to estimate the level of anthropogenic warming. This approach results in an estimate of anthropogenic warming for 2023 relative to 1850–1900 that is statistically similar to existing well-established attribution assessments. While this linear relationship appears to hold well over the observational record, it may break down in the future (for example due to reductions in SLCFs or changes in airborne fraction). Indeed, in Section 4.6.2 this relationship is shown to misestimate warming by 2050 in multiple scenarios of the MPI-ESM1-2-LR ensemble: it overshoots in SSP1-2.6 and undershoots in SSP3-7.0.

[25] For example, many attribution methods produce theoretically non-lagging annual timeseries for the current level of global warming (a benefit over purely statistical approaches), though the methods practically lag according to the timelines for forcing datasets to become available (a potential risk compared to purely statistical approaches); in reality, the latter lag tends to be on the order of months for the most critical forcings (but longer for several more minor forcings).

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

4.4.2 Energy Balance Models within a Kalman Filter

An intermediate complexity approach is *n*-layer (or *n*-box) energy balance models (Geoffroy et al. 2013; Gregory 2000; Held et al. 2010). Each layer has a single temperature, then energy balance is enforced with inter-layer exchange and parameterized forcings and climate feedbacks. Examples include Cummins et al. (2020) who developed a Kalman filter model of the climate with coefficients tuned on CMIP5 simulations. Bennedsen et al. (2023) used a 2-layer energy balance model to estimate and predict future global warming based on historical observations of global mean temperature and ocean temperatures with results comparable to ESMs. Nicklas et. al. (2025) also employed such a model, but modified to yield a posterior "climate state" that continuously assimilates both forcing data and GMST and ocean heat content anomaly (OHCA) estimates together with uncertainties. In the historical tests of Section 4.6.1 the Zanna et al. 2024 OHCA dataset is used, while in Section 4.6.2 the OHCA data is extracted from the model projections used. In Section 5, an ensemble of OHCA data is constructed to be used together with an ensemble of multiple surface temperature records (Supplement Section 2.4). This approach incorporates autocorrelated error into the climate state, filtered to have a smoothness that matches the 30-year running mean (or 20-year running mean in the implementation tested here), which enables timely estimates of the present climate state.

Such approaches offer a probabilistic estimate of whether a particular temperature level has been crossed as current observations are assimilated and can connect metrics such as OHCA, GMST, SST and LSAT. These systems can also account for potential future effects such as large volcanic eruptions were they to occur (Nicklas et al., 2025), reproducing similar results to ESMs with intermittent volcanism (Bethke et al., 2017, Chim et al., 2023; Section 4.6.2). Because these systems ingest both temperature and forcing observations, they can also be used to measure attributable warming (Section 4.4.4). The action of the Kalman filter step is included in Figure 16: it shows that as each observation is collected, the forward model and the observations are compared and the model state is set to a weighted mean of the two distributions.

4.4.3 Estimating temperature contributions from constrained emulators

Climate model emulators can be used to derive estimates of temperature responses to individual forcers (Forster et al., 2021, their Figure 7.8). As emulators are highly parameterisable and quick to run, large perturbed parameter ensembles that are trained on ESMs can be produced, and the outputs constrained to observed climate phenomena and climate system knowledge (Nicholls et al., 2021; Smith et al., 2021; Smith et al., 2024). Emulators and EBMs are not direct replacements for ESMs as they do not simulate physical grid-level processes, and introduce additional structural uncertainty in how emissions translate to radiative forcing and climate outcomes (Nicholls et al., 2021).

Warming attributed to anthropogenic and natural influences can be estimated by running an emulator with emissions of greenhouse gases and short-lived climate forcers (SLCFs) and natural forcing time series of solar and volcanic influences. The emulator, for example FaIR v2.2, produces effective forcing time series of all anthropogenic influences on the climate from source emissions (Smith et al., 2018; Leach et al., 2021), in effect attempting to link the IPCC's radiative forcing assessment (Forster et al., 2021; Fig. 7.6) back to emissions. Emissions of long lived GHGs and SLCFs are typically available to the recent past (up to 1 or 2 years lag) from sources including CO_2 (Friedlingstein et al. 2023), other GHGs (Gütschow et al. 2016), SLCFs from anthropogenic sources (Hoesly et al. 2018) and SLCFs from biomass burning (van der Werf et al., 2017). A small extrapolation can be performed beyond the end of the historical data by extension to a current-policies emission scenario (e.g. Rogelj et al., 2023). Such an extension is most problematic for SLCFs which can vary rapidly, and many of which are poorly observed (Szopa et al., 2021). Natural forcing estimates are obtained from the CMIP time series for solar forcing (Funke et al., 2024) and combined estimates from ice core records, and satellite aerosol optical depth and water vapour retrievals for volcanic forcing (Toohey & Sigl 2017, Kovilakam et al. 2020; Jenkins et al. 2023). Updates to all forcings except solar are described in Forster et al. (2025).

An ensemble of FaIR runs is produced by sampling values for a large set of prior coefficients and deriving a posterior that matches the best estimate time series of historical warming within uncertainty and distributions of eight observed or assessed climate metrics[26] incorporating assessed uncertainty bounds. Anthropogenic and natural forcings calculated by FaIR are also scaled to match assessed uncertainty ranges from the IPCC (Forster et al., 2021). From the constrained ensemble, the anthropogenic and natural forcings can be separated and run in isolation in FaIR, providing attributed warming to each factor.

4.4.4 Attributable Global Warming / Human-Induced Warming

The three attribution methods underlying the assessment of human-induced warming in the IPCC AR6 WGI assessment of global warming (Eyring et al., 2021) were the Global Warming Index (GWI, Haustein et al., 2017), Regularised Optimal Fingerprinting (ROF, Gillett et al., 2021), and Kriging for Climate Change (KCC, Ribes et al., 2021). In addition, attribution using FaIR (Section 4.4.3) was also used as a line of evidence in AR6 (Forster et al., 2021), and EBM-KF (Section 4.4.2) is also used for attribution herein. Attribution methods estimate both total-forced warming $\Delta T_F(t)$, human-induced warming $\Delta T_{F,anthro}(t)$, and naturally-forced warming $\Delta T_{F,natural}(t)$. Methods can quantify uncertainty from observed warming estimates (Section 3), radiative forcing, internal variability, and climate system properties, though most studies do not sample across all uncertainties. The resultant uncertainty ranges (Forster et al. 2024) are similar in magnitude to uncertainty in the observed warming for any given year[27].

[26] Equilibrium climate sensitivity (ECS); transient climate response (TCR); present-day global mean surface temperature (last 20-year mean) relative to 1850-1900; aerosol effective radiative forcing 2005-2014 relative to 1750 for direct, indirect, and total; present-day CO_2 concentration; ocean heat content change change between 1971 and the present day.
[27] For context, the uncertainty range of the level of observed warming in 2023 is 1.06-1.30°C, and the uncertainty range of the level of human-induced warming for 2023 is 1.0-1.4°C (Forster et al. 2024); the latter is a 0.1°C precision envelope over the uncertainty ranges for the three constituent attribution methods.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

At the time the AR6 WGI report was published, decadal-mean attributed warming based on the ROF and KCC methods, and annual-mean attributed warming based on the GWI, had been published. For consistency with the assessed observed warming, the AR6 assessment of anthropogenic warming was given as a decade average. However, single-year estimates for all three methods are now available (Forster et al. 2023, 2024, 2025) and increasingly robust as an indicator (Ribes et al., 2025). To estimate warming from forcing, the KCC and ROF methods depend on climate model projections after historical CMIP6 simulations end in 2014. It would be scientifically desirable for modelling groups to annually update their historical runs in near-real-time to obviate the need to use projections (Naik et al., 2025; Hewitt et al., 2025). This would require significant effort to update observed forcing series more operationally and in closer to near-real-time (see discussion in Section 4.4.3) as well as agreed modelling protocols from CMIP-contributing centres (Schmidt et al., 2023). This could in turn be used to update constrained projections and emulators (Ribes et al., 2025).

Human-induced warming estimates may optionally be additionally “time filtered” as described in Section 4.3. For example, AR6 WGI used a trailing 10-year average (Section 4.3.1) of human-induced warming, and SR1.5 used a 30-year mean centered using extrapolations of current warming (similar to the approach in Section 4.5). However, an especially hot/cold year of observations only advances/retards the single-year annual mean level of human-induced warming by around 1 year at current warming rates (depending on the attribution method) (Forster et al., 2024). Both framings used in the AR6 cycle lie well within the overall uncertainty range for the single-year annual mean attributed level of human-induced warming, and this has led to increased confidence in the robustness of this indicator (e.g. Forster et al. 2024, Haustein et al. 2017, Otto et al, 2015, Ribes et al., 2025).

4.4.4.1 Global Warming Index

The GWI was introduced by Otto et al. (2015) by building on the well-established “multi-fingerprinting” approach from Hasselman (1997). It was updated with uncertainty estimates by Haustein et al. (2017) and expanded by Forster et al. (2023, 2024, 2025) by further disaggregating the attributed warming components, using an updated climate emulator and parameterisations, and incorporating the newest available datasets for forcings, observed warming, and internal variability samplings. The GWI was designed to be as simple, transparent, and explainable as possible while still accurately and robustly representing the evolution of the climate, with a full sampling across sources of uncertainty for GMST observations, radiative forcing, internal variability, and climate system properties. In brief, time series for anthropogenic and natural radiative forcing components are converted to initial estimates of their corresponding $\Delta T_{F,i}(t)$ using a simple climate emulator. The initial time series are then adjusted by fitting them against the observed GMST timeseries and internal variability samplings, with the result representing the attributed anthropogenic and natural $\Delta T_{F,i}(t)$ (Otto et al. 2015, Haustein et al. 2017, Forster et al. 2023). Uncertainty from all sources mentioned above is incorporated through a very large Monte Carlo ensemble. The climate emulators employed are well established, transparent, and simple (GWI calculations first used the IPCC AR5 impulse response model (Myhre et al. 2013) and most recently FaIR v2.0 (Leach et al. 2021)). Crucially, few prior assumptions about the climate

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

response to each forcing are required, since the emulators are tuned to represent a wide range of possible climate behaviours. The GWI value is directly constrained to observations by the regression, resulting in a reduced dependence on the size of each uncertainty component incorporated into the method. The GWI assumes that the long-term temperature responses to the various forcings sum approximately linearly. While comprehensive re-assessments of the GWI are now provided on an annual basis (Forster et al., 2023, 2024, 2025), the indicator is also available as a real-time index[28] (Haustein et al., 2017).

4.4.4.2 Kriging for Climate Change

KCC is a sophisticated approach for utilizing ESM simulations rather than emulators to inform the contribution of forcings (Ribes et al. 2021, Qasmi and Ribes 2022). In this Bayesian approach, the filtered ESM response to combined anthropogenic and natural forcings is used as a prior on the real forced response. The response to natural forcings is estimated from an energy balance model. The observed GMST timeseries is then used in a Bayesian framework to estimate a posterior distribution for anthropogenic and natural attributable temperature changes and to update constrained projections (Ribes et al., 2025).

4.4.4.3 Regularised Optimal Fingerprinting

ROF uses global surface temperature series from observations and ESM simulations of the response to combined anthropogenic and natural forcings, and natural forcings alone (Gillett et al., 2016). It uses frequentist total least squares optimal regression (Stott & Tett 1998, Allen and Stott, 2003), but with a regularized approach to representing the covariance matrix first introduced by Ribes et al. (2013). The implementation used here follows that in Gillett et al. (2021) and Forster et al. (2024).

4.4.5 Summary of attribution methods considered

Figure 15 compares key examples of each approach, when applied to the HadCRUT5 global temperature record. Warming attributed separately to greenhouse gases and to other human forcing (dominated by direct and indirect cooling effects of aerosols) has an uncertainty range of about 0.5°C. However, because they combine to equal the total warming, these individual uncertainties are anti-correlated, such that the warming attributed to the sum of all human causes can be more tightly constrained. Similarly, methods that are low within the set for warming attributed to greenhouse gases will tend to be high within the set for warming attributed to natural and other human forcings. Throughout the historic period, the lowest estimate is often obtained by the FAIR constrained emulator, while the approach that provides the highest estimate varies over time (for the final 20 years it is the regularised optimal fingerprinting method).

[28] www.globalwarmingindex.org; last accessed 01/12/25

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Open Access Earth System Science Data Discussions

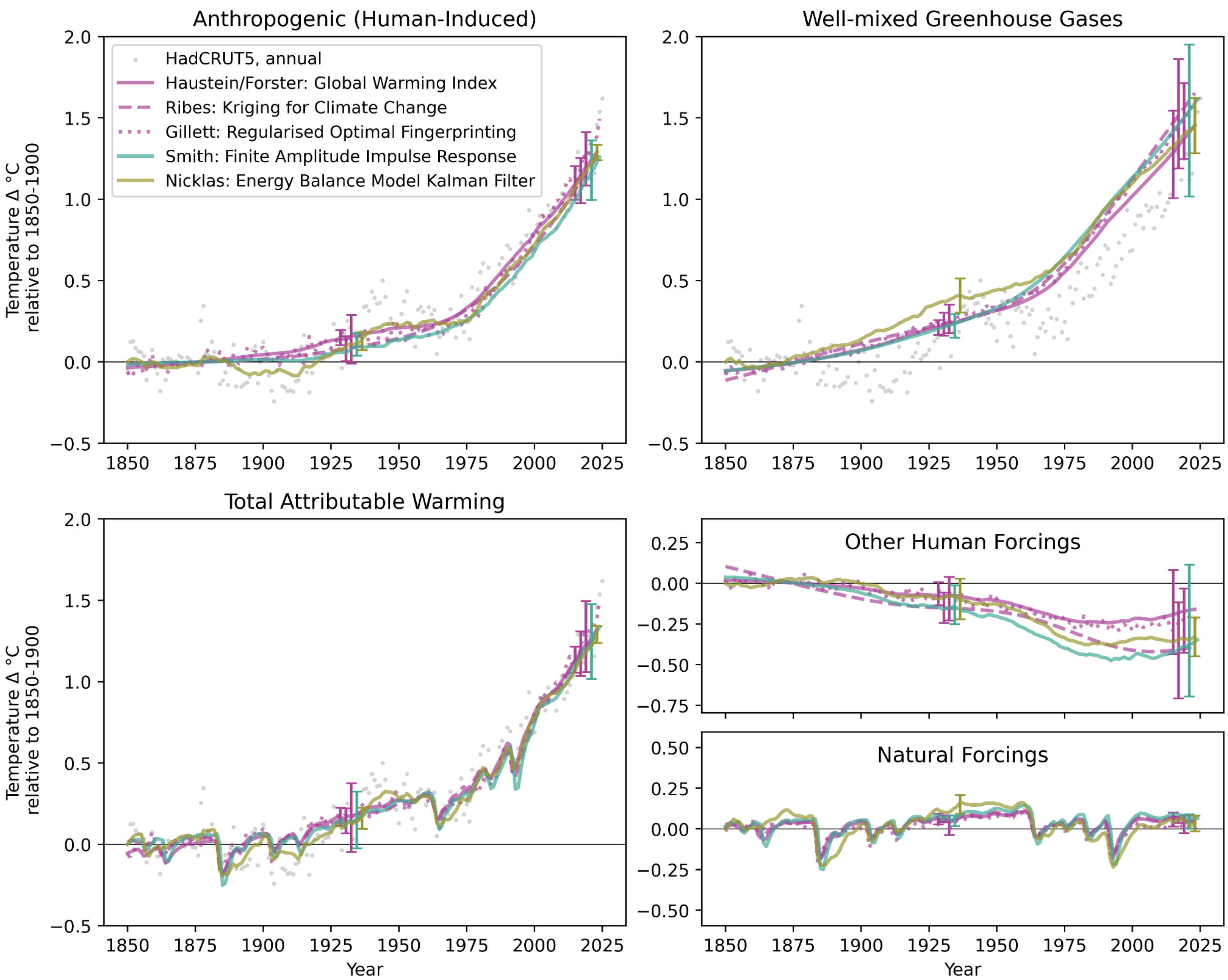


Figure 15. A breakdown of attribution results to different combinations of forcings. To show all of the methods from Sections 4.4.2-4.4.4, EBM-KF-ta4 and FaIR are added to the KCC, ROF, and (Haustein/Forster) GWI results that are each adapted from the Supplement to Forster et al. (2025). There is broadscale consistency in the attributed anthropogenic warming and natural forcing responses among the methods which vary substantively in their underlying methodological assumptions (see main text), but they differ substantially in their uncertainty ranges (shown for selected periods in the panels). Note that the EBM-KF-ta4 method shown here assimilates a handful of ESM simulations (CESM2, ESM1-2-LR, BCC-CSM2-MR) as "observations" on a world with only well-mixed greenhouse gases, other human forcings, or natural forcings (BCC-CSM2-MR excluded due to drift from 1850 baseline).

Figure 16 illustrates schematically the similarities and differences between all of the techniques outlined in Sections 4.4.1 through 4.4.5. The linear correlation between cumulative carbon emissions and warming has a simple dynamical premise, and is based on linear regression (Jarvis & Forster, 2024, 2025). EBM-KF uses a dynamical energy balance model to predict a prior projection and then a Kalman Filter update to arrive at a posterior distribution for the climate state at each annual update (Nicklas et al. 2025). For attributing warming, EBM-KF can be used to expand a single-forcing ESM run into a set of simulated single-forcing ESM ensembles. FaIR uses a dynamical model including a complex set of forcing agents with mostly assumed

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

linear responses, a three-layer ocean, and then a sampling across a Monte Carlo ensemble spanning different prior coefficients followed by a posterior selected to best match the observed temperature record (Smith et al. 2024). The posterior coefficients in FaIR can then be used to attribute natural and anthropogenic warming. The GWI generates a large Monte Carlo ensemble of prior warming component estimates using FaIR, and constrains them against observations using multiple linear fits to attribute warming. The ROF and KCC methods use ensembles of ESMs over multiple scenarios to attribute warming, with a
1150 linear fitting method or a Bayesian estimate, respectively.

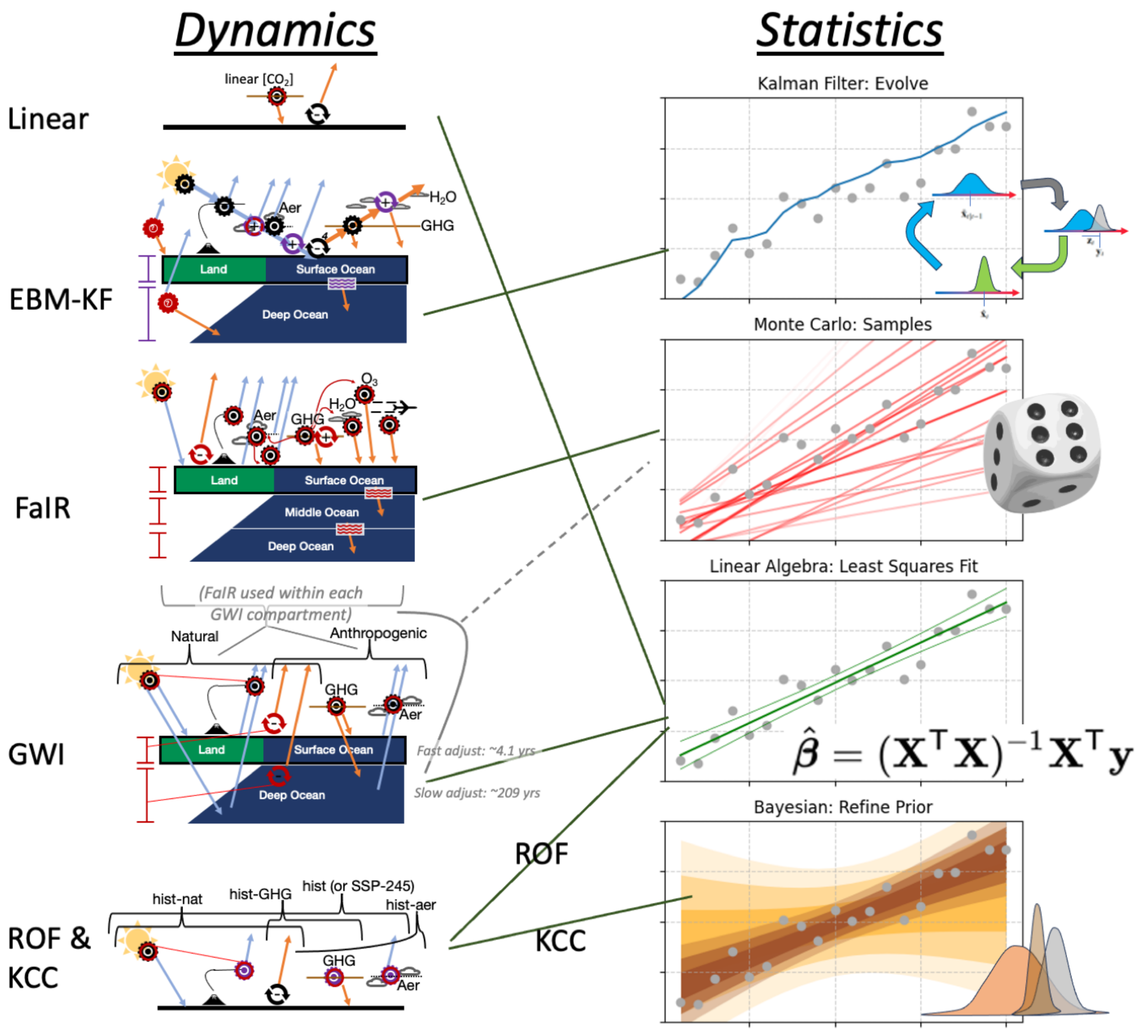


Figure 16 Schematic of the different attributable warming methods from section 4.4. The dynamics of these models are illustrated in the left column, and the mathematical and statistical underpinnings are illustrated in the right column. The symbols on the left
1155 **indicate forcing timeseries (cog/gear), feedbacks (circular arrows), and energy fluxes (straight arrows, blue for shortwave and orange for longwave; radiator symbol for inter-layer exchanges). Colours indicate types of evidence including observed or measured behaviours (black), fits (red), CMIP abstractions (purple). Ranges on the left indicate heat capacity of different layers. Overbraces**

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

indicate simulation scenarios. Thin connected lines in FaIR indicate assumed relationships among mechanisms. Green lines connect dynamics to statistics for a particular method.

4.5 Combining historical observations with ESM projections or initialised decadal predictions

An indicator consistent with the IPCC AR6 approach for projections (Section 4.2), dubbed the "Current Global Warming Level (CGWL)", can be calculated as a 20-year mean centred on the current year, using observations from the last 10 years and projections or forecasts for the next 10 years, and taking the mean across the combined 20-year period (Betts et al., 2023). Projections of future global mean temperature are subject to uncertainty from three sources: i) future forcing; ii) long-term climate responses (climate sensitivity and model error); and iii) internal climate variability. Out to a decade iii) and ii) dominate the uncertainty, while i) becomes more important over longer timescales (Hawkins and Sutton, 2009), although rapid changes in short-lived climate forcings or large volcanoes can impact trends earlier.

Projected or predicted future temperature change can be obtained from several sources:

1. Initialised forecasts using ESMs, such as the WMO decadal forecast multi-model ensemble (Hermanson et al., 2022). These take initial conditions from recent observations, so the internal variability is more consistent with expectations. Newly initialised forecasts from the WMO ensemble are currently issued on an annual basis[29]. The sampling of climate sensitivity uncertainty is dependent on the models used in the WMO forecast system, and some of these have high climate sensitivities so may lead to rates of warming that are overestimated.
2. Uninitialized scenario-based projections, e.g. the CMIP6 multi-model ensemble. The CMIP models cover a wide range of climate sensitivity in an unstructured way; for several models in the CMIP6 ensemble, the climate sensitivity is outside the ranges assessed as "likely" or "very likely" by IPCC AR6 (Lee et al., 2021). The simulated phasing and magnitude of internal variability will likely diverge more strongly from reality than initialised projections. However, since a 10-year average has been shown to closely represent the 30-year average centred on that decade since the 1970s (Trewin, 2022), much of the internal variability may already be averaged out over 10 years.
3. Constrained or assessed future warming rates, drawing on multiple lines of evidence, which can provide a more structured approach to the sampling of uncertainty in climate sensitivity, while also providing an assessment of likelihood. One example of this is to downweight models that are less consistent with observations. IPCC AR6 assessed warming rates using the CMIP6 multi-model ensemble constrained by climate sensitivity based on multiple lines of evidence, alongside emulators that systematically explore climate sensitivity guided by a probabilistic assessment (Lee et al., 2021). This provides projections of the forced climate response, without internal variability. Another example is the UKCP18 global probabilistic projections (Murphy et al, 2019) which used a perturbed parameter ensemble (PPE) of a version of the HadGEM3 climate model to simulate a wide range of different climate outcomes, with the ensemble then being expanded to 10^6 members using a statistical emulator, and the distribution of outcomes adjusted with reference to structurally different climate models in CMIP5 and weighted by comparison with observations.

[29] While the publicly accessible predictions via WMO extend to only 5 years the actual model analyses extend to a decade. The choice of 5 years is a presentational one by WMO.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Regarding the sensitivity to future emissions trajectories, IPCC AR6 assessed the global mean temperature change in 2021-2040 to be within one decimal place, i.e.: 1.5°C for all scenarios except the very high emissions scenario SSP5-8.5[30]. Betts et al (2023) noted that the CGWL calculated using the UKCP18 global probabilistic projections with three emissions scenarios (RCP2.6, RCP4.5 and RCP6.0) differed by 0.02°C (Figure 17). If realised warming is the metric of interest then this should include volcanic influences when a (series of) large eruption(s) of likely climatic importance occurred (Bilbao et al., 2024, Sospedra-Alfonso et al., 2024).

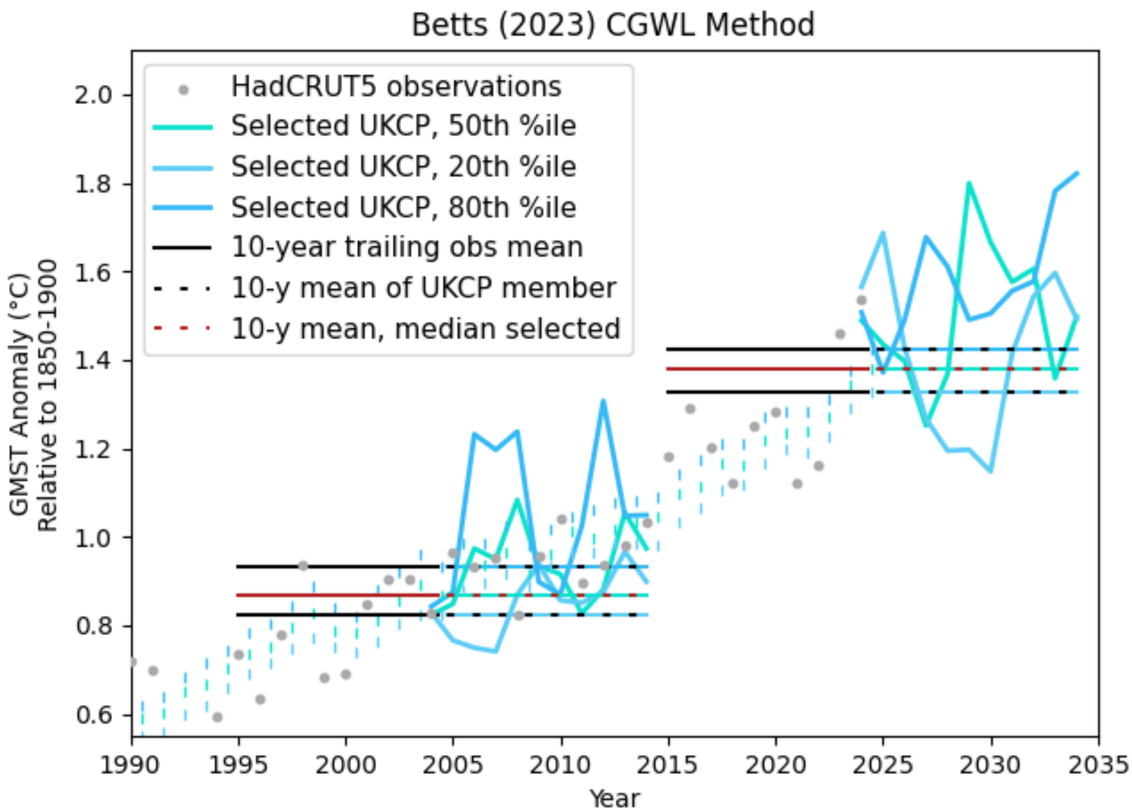


Figure 17. Illustration of Current Global Warming Level estimation. Three distinct probabilistic ensemble projections from UKCP18 are chosen to resemble ESM decadal initialised forecasts over 2025-2034. An average of these projections serves as an estimate of the current global warming level centred on 2024 following Betts et al. (2023). Of the 3000 published UKCP ensemble members, 100 were chosen in 2024 that are both a) close to the HadCRUT5 observations' previous 10-year mean (2015-2024) and b)

[30] While there is substantial divergence between scenarios from mid-century onwards (Lee et al., 2021), as noted in IPCC (2021) "Under these contrasting scenarios, discernible differences in trends of global surface temperature would begin to emerge from natural variability within around 20 years". Therefore assuming that a CGWL technique iteratively updates to use new CMIP generation (or similar) historical runs and future scenarios (and those scenarios are plausible in accounting for enacted policies in the short to medium-term) it will always be the case that scenario dependency is a minor source of uncertainty (Hawkins and Sutton, 2009).

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

close to the last-year observation (2024). Of these 100 analogues to the ESM decadal initialised forecasts, they were then ranked by their average temperature over the 2025-2034 period. The 50th percentile is shown in teal and the 20th and 80th percentiles in shades of blue. The HadCRUT5 10-year trailing mean (2015-2024) is combined with the ensemble statistics to yield the horizontal lines illustrating a 2024 current global warming level (in red for the median and black for the 20th and 80th percentiles). This procedure is also illustrated using 2004 as a current year. Repeating the procedure for each year yields the envelope of small teal and blue vertical lines.

4.6 Characterisation of different approaches

The lack of a formal definition of 'longterm' in the context of the LTTG (Section 4.2) means that determining the 'true' crossing time has a strong normative component. Here we use the centre of the retrospective 20-year moving average for consistency with the IPCC AR6 WGI approach for projections (Lee et al., 2021) as our working benchmark. This is consistent with the choices made in Cannon (2025) and Bevacqua et al. (2025) as to the target metric (and, in the same vein, Scherrer et al. (2024) used a 30-year average). Such a multidecadal average has over a century of heritage in the WMO in terms of defining long-term climate normals (Scherrer et al., 2024). By definition any choice of benchmark will favour those methods which best approximate it. This metric best represents the realised warming from all causes ($\Delta T_{fit}(t)\ in\ eq.2$), and not the underlying anthropogenic component (see Consideration 2 in Section 4.2).

The 20-year centred mean is effectively an estimate of the realised warming at the end of year 10, so this metric can only be formally evaluated retrospectively, and essential differences among the methods from the preceding three sections stem from how they would estimate the realised warming at year 10, without waiting for years 11-20. It is desirable that any operational metric provides a timely estimate (Consideration 3 in Section 4.2; also Scherrer et al., 2024 and references therein). We therefore assess how well the approaches outlined in Sections 4.3 through 4.5 can approximate the eventual 20-year mean at year 10, and how accurately they estimate the time of exceedance (Betts et al., 2023). Our 3 test cases are as follows:

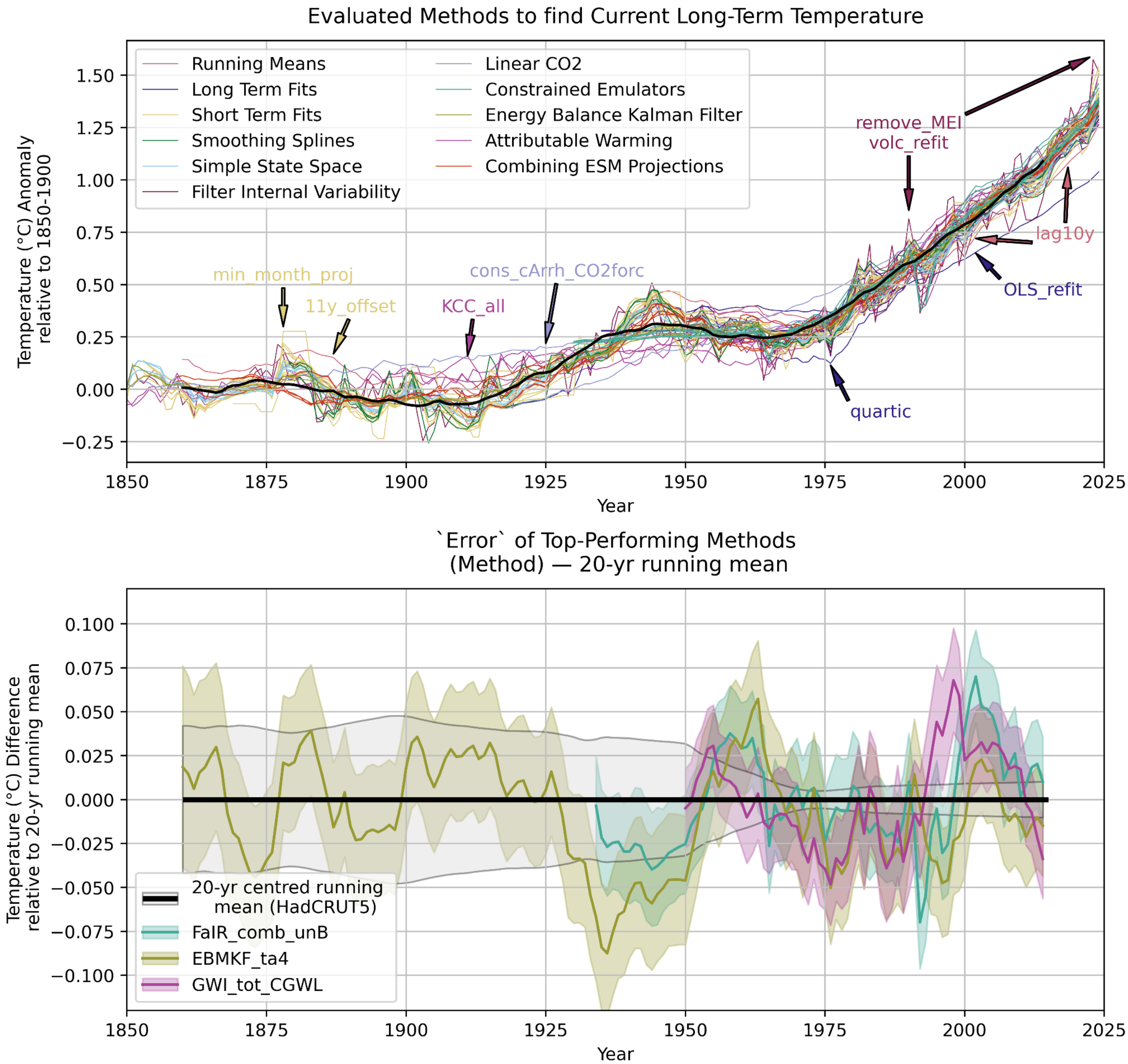
1. In HadCRUT5, how well do different techniques estimate the historical crossings of 0.5°C and 1°C? (Section 4.6.1)
2. In a single ESM Single Model Initialised Large Ensemble (SMILE), where the simulations only differ in terms of ICV, how well do the techniques determine future 1.5°C crossing times? (Section 4.6.2)
3. In an ESM ensemble with major future potential volcanic eruptions, how sensitive are the techniques for crossing 1.5°C to possible multiannual to multidecadal volcanic impacts? (Section 4.6.2)

4.6.1 Assessment of historical exceedances of warming thresholds

The 20-year moving average exceeded 0.5°C in the late 20th Century and 1°C in the 2010s relative to 1850-1900 in most datasets. Here we focus upon the exceedance of 1°C, with a discussion of exceedance of 0.5°C, along with further details, given in the Supplement. We arbitrarily selected HadCRUT5 as an example, since a single time series is needed to assess each method and some methods require spatial information which requires use of a single dataset. Overall, the techniques bracket the 20-year retrospective moving average throughout the series (Figure 18, top panel). Hence in principle there is a potential

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

for almost[31] real-time assessment to be able to determine exceedance. However, there are times when some candidate techniques clearly struggle (see labels) and the trailing averages and full-series linear fit are particularly likely to be outliers. We eliminated any method with an RMSE of 0.06 °C or greater from subsequent consideration, including all labelled outlier methods in Figure 18 (top panel). For the four most skillful methods the reported uncertainty range mostly includes the HadCRUT5 20-yr centred mean temperature (Figure 18, lower panel).



[31] In principle this assessment could be done in real-time, but there may be up to a several months' delay in validating observations of relevant forcings or running decadal forecast models.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Open Access Earth System Science Data Discussions

Figure 18. Top panel: Illustration of how the range of methods able to be operationalised for testing from Sections 4.3 through 4.5 (see Supplement for further method details) perform throughout the timeseries against the retrospective centred 20-year average (solid black line) used herein as a working benchmark. Noteworthy departures are denoted by annotated arrows for ease of interpretation with labels abbreviating specific methods (see Figure 20 and Supplement Section 2 for details of methods and further analysis). Note that different methods have different start dates arising from differences in the necessary constraints available for their calculation. Lower panel: For the 4 methods with best overall performance (here assessed using maximum likelihood), the nature of their departures from the retrospective 20-year centred mean (estimate minus actual retrospectively calculated value). In both panels methods are always calculated using observed data only up to year 10 in the 20-year mean being assessed against. The shaded ranges indicate 95% uncertainty ranges as reported by each of these best-performing methods.

The HadCRUT5 central estimate crosses the 1°C level when the centre of its 20-year moving average moves past 2010, but this central crossing time hides a more complex picture of when individual days and years exceeded 1°C. The impact of interannual and sub-annual variability in global temperature is illustrated in Figure 19 using ERA5 rebased to 1850-1900 using HadCRUT5, NOAAGlobalTempv6 and Berkeley Earth, taking into account uncertainty in their warming since this baseline. More than a decade before the crossing time, daily exceedances of 1°C in ERA5 were rare, with the exception of the large El Nino year of 1998. Daily exceedances became more common but stayed below 50% for all years before 2015 for the best estimate. By contrast, every year since 2015 has >50% of days exceeding this level of global warming by a considerable margin.

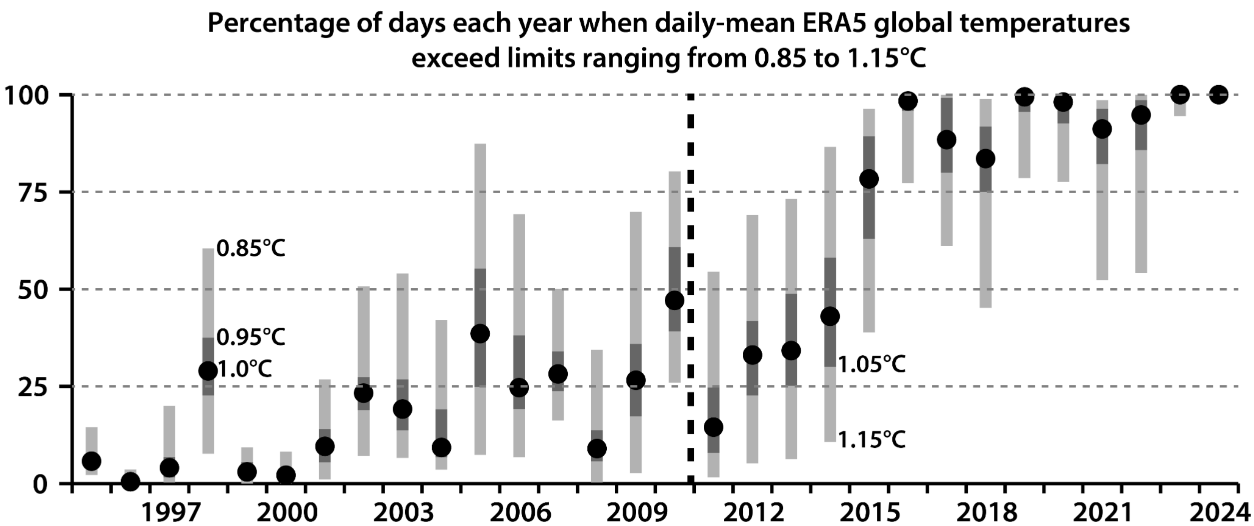


Figure 19. Plot as in the lower panel of Figure 13 but for 1995-2024 illustrating the days exceeding 0.85°C, 0.95°C, 1°C, 1.05°C and 1.15°C relative to 1850-1900. 1995 is the first year with at least one day more than 1.2°C warmer than 1850-1900. 1990 is the only earlier year with at least 5% of days more than 1.0°C warmer than 1850-1900. A further eight years prior to 1995 had at least one day more than 1.0°C warmer. Coincidentally 2010 had almost 50% of days exceed the 1.0°C value. The dashed vertical line denotes December 2010, when the 20-year centred mean equals 1.0°C. The 20-year median does not reach 1.0°C until March 2012. For further methodological details see Figure 13 caption. Source: ERA5. Credit: C3S/ECMWF

Most methods tend to report exceedance of 1.0°C later than the retrospective 20-year centred mean (Figure 20), but the methods are more balanced around the 20-year centred mean when it exceeds 0.5°C (see Supplement Figure S1). This is perhaps unsurprising given the nature of the timeseries around 1.0°C shown in Figure 19 that clearly highlights the back-loaded nature of this value's exceedance. Nearly all of the short-term fit methods report 1.0°C central estimate threshold crossing in 2011-2012, whereas most of the instantaneous attributable warming methods report 2009[32]. At 0.5 °C this behaviour is largely reversed (see Supplement). Similarly, while the performance of the CGWL methods (Section 4.5) appears outstanding as an entire class, this is only true at the 1.0 °C threshold, and several of these methods are 3 years too late or 2 years too early at the 0.5 °C threshold (see Supplement).

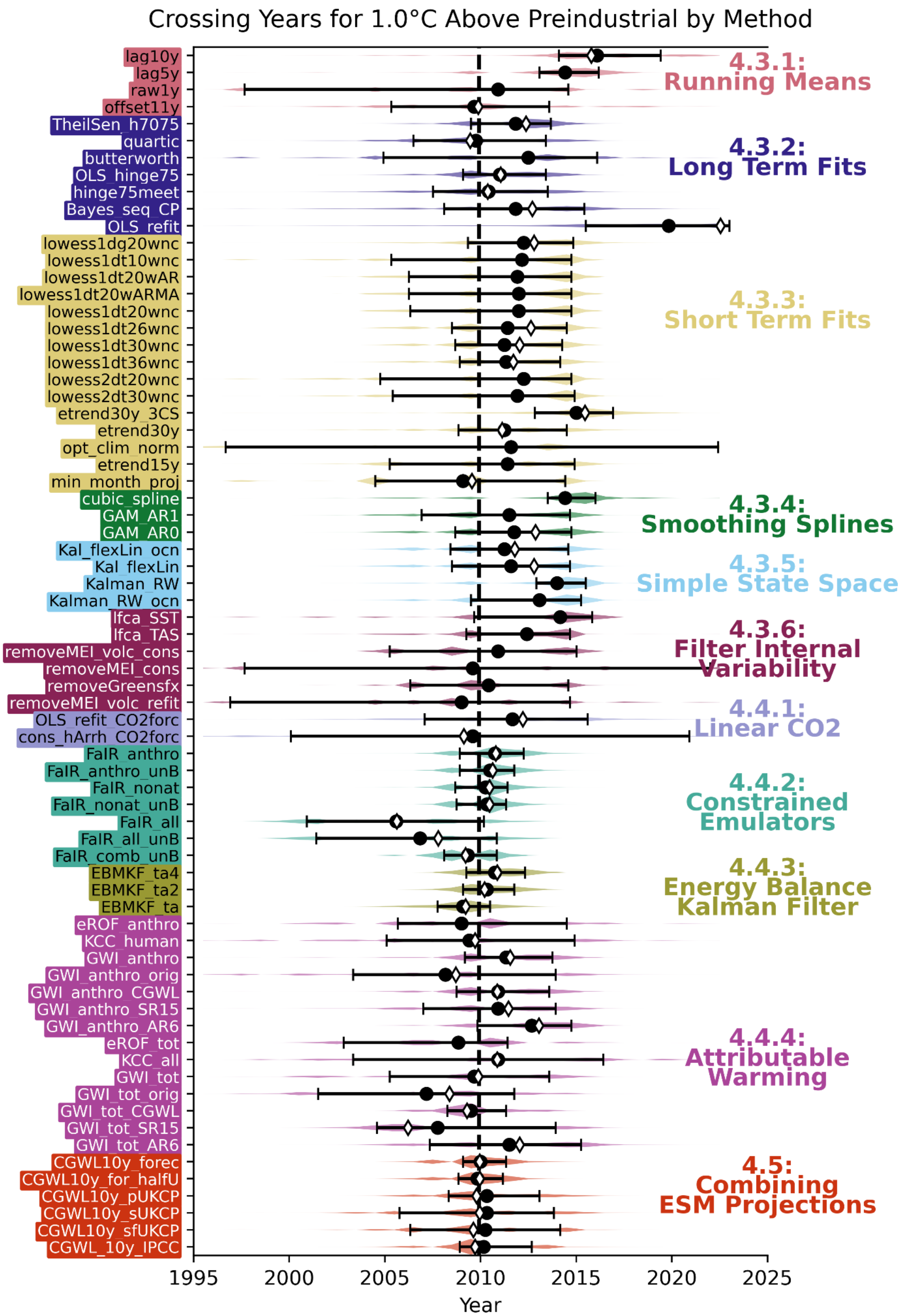
It is desirable that techniques have properly specified uncertainties. In our analysis some methods have extremely large uncertainties, while others with tight uncertainties lie very far from the target value, as the original uncertainty specified by each method may not be intended for this particular purpose. We address this issue of uncertainties by first pre-scaling each of the uncertainty time series to optimize the log-likelihood across the time window 1925-2024. From within each model, if a series of temperatures are drawn from the pre-scaled model distributions for each year, the HadCRUT4 20-year running means are as likely to appear from these distributions as they could be without altering the model timeseries of central estimates.

The running means and OLS_refit methods perform very poorly, with large biases that substantially exceed their reported uncertainties, representing a statistical "miss". The conclusions about which methods perform least well are similar for the 0.5°C crossing (see the Supplement), which occurs during the 20-year period centred on 1985 and is notably affected by volcanism from both El Chichon (1982) and Pinatubo (1991). At the 0.5°C threshold crossing time, the evaluated methods are distributed more symmetrically about the crossing time of the retrospective 20-year centred mean, but there is still a slight overall late bias.

[32] Note however the distinction between attributed total forced warming (see attribution rows with "tot") and attributed anthropogenic forced warming (see attribution rows with "anthro"); across every attribution method, total forced warming reaches 1.0°C a few years earlier than anthropogenic forced warming due to a contribution from naturally forced warming of roughly 0.05°C around 2010. Also the lagged decade-average filter applied to attribution methods (to mimic the AR6 WGI approach) is by definition lagging.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Open Access Earth System Science Data Discussions



https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Figure 20. Summary of threshold-crossing time performance of all methods able to be operationalised (see Supplement for details on each method) in the specific determination of when the centre of 20-year retrospective mean crosses 1°C in HadCRUT5, which occurred in 2010 (vertical black dashed line). For each method a statistical average occurrence of exceedance is shown (black dot), see Supplement. Uncertainty pre-scaling has been applied to all methods (see Table 2). The first time that the probability of exceedance is greater than 16% to the last time that the probability of exceedance is less than 84% is also shown (bar). Shown in background are violin plots of exceedance probabilities. Cases where the estimated exceedance timing does not overlap with the actual exceedance represent a statistical miss. Some methods (for instance removeMEI) never have a unique, monotonic crossing due to noise. Other "stable" methods always cross 50% for any thresholds throughout the series monotonically and uniquely, and yet other "semi-stable" methods have had a limited number of reversal years (<=10) since 1975 when the method estimate temporarily cools. These unique crossings are shown for 1°C with white diamonds to highlight that in these stable and semi-stable methods the threshold crossing could be reported immediately rather than retrospectively (as is the case for the black circles) for a noisy method with a high probability of a reversal year. To aid interpretation, the methods are categorized by method family class (denoted on the right hand side). Violin plots are a fit with a smooth function to the probability of exceedance, as described in Supplement Section 2.2.

Generally each method's behaviour is unique to its particular implementation and cannot be predicted from other examples of the same class. However, Figure 20's white diamonds show that many method classes can give instantaneous crossing-time estimates because their results change quasi-monotonically so once they pass a given global warming level, they tend to stay above it. For example, all of the constrained emulator, energy-balance Kalman filter, and CGWL methods would have provided instantaneous 1 °C crossing times.

We note that as of the end of 2024, only 4 of the 56 evaluated methods exceed 1.5°C based on the HadCRUT5 temperature record (Figure 18). Those 4 methods are 2 implementations of MEI removal (Foster and Rahmstorf 2011), the piecewise linear "opt_clim_norm" fit of Livezey et. al. (2007), and our "min_month_proj" implementation of the Cannon (2025) 12-month continuous exceedance window. These 4 are noisy methods (see Figure 20, the Cannon "min_month_proj" increases monotonically but is still noisy by construction), sometimes briefly exceeding a threshold and then dropping back below it, which results in wide threshold crossing windows and low confidence that any initial exceedance might constitute permanent exceedance.

Overall, the benchmarking against the historical record highlights a broad range of methodologically distinct techniques that can approximate the trailing 20-year average working benchmark. More simple methods such as a trailing decade average (used in IPCC AR6) and OLS linear fit calculated for the full period (used in IPCC AR5) are systematically biased. In particular the OLS approach is problematic (RMSE 0.134°C) because the full time series is increasingly non-linear. Following the application of a threshold RMSE <0.06°C over the historical period (1850-2024) a total of 40 methods were considered further. At least one method from each class (excepting the Linear CO_2 projections 4.4.1 from Jarvis) introduced in Sections 4.3 through 4.5 remains (Table 2). Remaining methods have a first crossing instant lead/lag distribution that is 56% within 1 year of crossing 0.5°C in 1985, and 78% within 2 years. Similarly, of the remaining methods, 46% are within 1 year of crossing 1.0°C in 2010, and 70% within 2 years (see Supplement Figure S2).

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Method Class	Fraction with Historical RMSE <0.06°C	Best in Class by RMSE Performing Historical Method or Selected Realised/Attributable Method	This Method's RMSE	This Method's log -likelihood	This Method's # of q-values <0.5
4.3.1: Running Means	¼	lag5y	0.0563	1.344	0 (54)
4.3.2: Long Term Fits	4/7	TheilSen_h7075	0.0366	2.048	0
4.3.3: Short Term Fits	8/15	lowess1dt30wnc	0.0505	1.463	2 (110)
		lowess1dt36wnc	0.0510	1.429	0
4.3.4: Smoothing Splines	1/3	GAM_AR1	0.0581	1.241	17 (42)
4.3.5: Simple State Space	4/4	Kal_flexLin_ocn	0.0442	1.687	0
		Kal_flexLin	0.0481	1.610	0
4.3.6: Filter Internal Variability	2/6	removeGreensfx	0.0493	1.590	0
4.4.1: Linear CO2	0/2	OLS_refit_CO2forc	0.0864	-3.842	59 (123)
4.4.2: Constrained Emulators	5/7	FaIR_nonat_unB	0.0247	2.308	2
		FaIR_comb_unB	0.0271	2.165	4
4.4.3: Energy Balance Kalman Filter	3/3	EBMKF_ta2	0.0277	2.146	0
		EBMKF_ta4	0.0302	2.056	0
4.4.4: Attributable Warming	6/14	GWI_tot_CGWL	0.0263	2.222	0
		GWI_anthro_CGWL	0.0401	1.815	0
4.5: Combining ESM Projections	6/6	CGWL10y_for_halfU	0.0333	1.988	4
		CGWL10y_sfUKCP	0.0411	1.759	3

Table 2. Method classes and number and statistics of 40 methods remaining after an RMSE<0.06°C selection versus the 20-year running mean based on the historical record (1850-2024 from HadCRUT5) was applied. The "Linear CO2" class of methods is highlighted in red because all methods within it failed to pass this selection criteria, and the GWI_tot_CGWL method is highlighted in green because it has the lowest RMSE in this evaluation. Methods are not penalized for not reporting any estimate in years prior to 1960, and the RMSE is averaged with respect to reported years. For each class of methods, the specific method with the lowest

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

RMSE is named. In the rightmost columns additional attributes of this best-in-class method are listed: as the average log-likelihood (of the HadCRUT5 observation) across all reported years, and then number of years for which there is a statistically suggested difference between the HadCRUT5 observations and a random drawing from each model distribution (note that q-values are used because we would expect 5% of all years under such a random drawing to have p<0.05). For these two right-most columns, the uncertainties provided by each method have been pre-scaled to optimize the log-likelihood across the time window 1925-2024, which often widens the uncertainty reported naively, for instance the standard error of a lagged 5y mean and the default LOWESS uncertainty. This wider uncertainty reduces the number of years with low q-values (tally without uncertainty pre-scaling is listed in parentheses). Note that the LinearCO2 row is worse than all other rows in these statistical aspects, and a substantial number (59) of the 20-year running mean observations fall outside of the LOWESS uncertainty distribution even with this uncertainty pre-scaling.

4.6.2 Assessment of potential future performance based upon modelled future climate

Future global surface temperature trajectories could vary between: i) a world in which aggressive mitigation actions cause a temperature rise deceleration, stabilisation and even beginning to cool; and ii) a world where mitigation actions are grossly insufficient and there is a marked acceleration in the rate of warming. There is also the potential for substantial multi-annual to decadal departures from the underlying human-induced climate change, such as from increased volcanic activity (Bethke et al., 2017; Chim et al., 2025). An operational monitoring capability will need to be robust to such future plausible climate system behaviours. We therefore go on to assess the methods which can be extended to such test cases, noting that it was not possible to extend all methods assessed in 4.6.1 (see Supplement for details).

We turn to Single Model Initial-condition Large Ensembles (SMILE) simulations to explore the interplay of modelled forced response and internal variability. We use the MPI-ESM1.2-LR ensemble (Mauritsen et al., 2019, Olonscheck et al., 2023) under different forcing scenarios; we select SSP1-2.6 and SSP3-7.0 to capture a wide range of potential future warming. The MPI-ESM1.2-LR has been tuned to better match the instrumental period warming (Mauritsen and Roeckner, 2020) and has an ECS of around 3K (see Supplement). This is similar to the best estimate ECS of the IPCC AR6 (2021a). Some forcing-based methods from the preceding analyses that performed well are finely tuned to climate parameter estimates for the real Earth, and may perform differently when applied to an ESM with biases. These methods were re-tuned or modified to the model for the purpose of this evaluation, as described in the Supplement; this step is not implemented for use on actual observed records.

Results for SSP1-2.6 and SSP3-7.0 (Figure 21) cover the potential for stabilisation or acceleration of warming, respectively, at the time of exceedance of 1.5°C. As expected, some methods that were successful in reproducing recent observational records (Section 4.6.1) do not perform well in this test. Methods that account for the forcings (Sections 4.4 and 4.5) rather than being purely statistical tend to handle such cases better. Methods that assume quasi-linearity in warming (e.g., hinge fit methods) break down in the case of SSP1-2.6 where GSAT stabilises and then slowly cools. These methods systematically overstate the degree of warming after the mid-century (Figure 21, middle panel). For SSP3-7.0, which experiences accelerating warming throughout the century, the same methods are systematically biased cool (and hence biased late in their estimate of when a warming level has been exceeded). Several forcing-based methods tuned to the middle of the IPCC-assessed ECS range rather than the IPCC best-estimate ECS (particularly EBM-KF-ta and CGWL-10y-pUKCP) exceed the warming

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

predicted by MPI-ESM1.2-LR in both SSP experiments. We additionally consider the method of removing the MEI ENSO signal with constant coefficients (a variation on the methods of Foster and Rahmstorf, 2011) to be an outlier because it sometimes differs from the 20-yr centred mean by >0.3 °C. However, not all statistically based approaches suffer from such issues. In particular, one of the state space Kalman filtering approaches of section 4.3.5, as well as several LOWESS methods of section 4.3.3 using a 20-year (or 30-year) kernel, are sufficiently sensitive to capture changes in the long-term climate trajectory. Those methods which show unsatisfactory performance against SSP1-2.6 and/or SSP3-7.0 (RMSE>0.06°C for either SSP over 2025-2100) in the MPI ensemble are also removed from the set of techniques for inclusion (Table 3) resulting in a set of 24 methods to carry forward[33].

[33] Three methods, such as GAM_AR0, performed worse in the historical context (RMSE of 0.0652) than in the future context (RMSE of 0.489, 0.418, 0.450 as in Table 3). The others that performed worse in the historical evaluation were etrend30y_3CS variants (a Short Term Fits of 4.3.3) and GWI_tot (without any combination approach). This may be due to subtle volcanic fluctuations in the 20-year centred mean target, which were not present in these initial future test cases, as this method also performed worse than GAM_AR1 in the future volcanic test (realized RMSE of 0.073 and attributable RMSE of 0.1274)

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Open Access Earth System Science Data Discussions

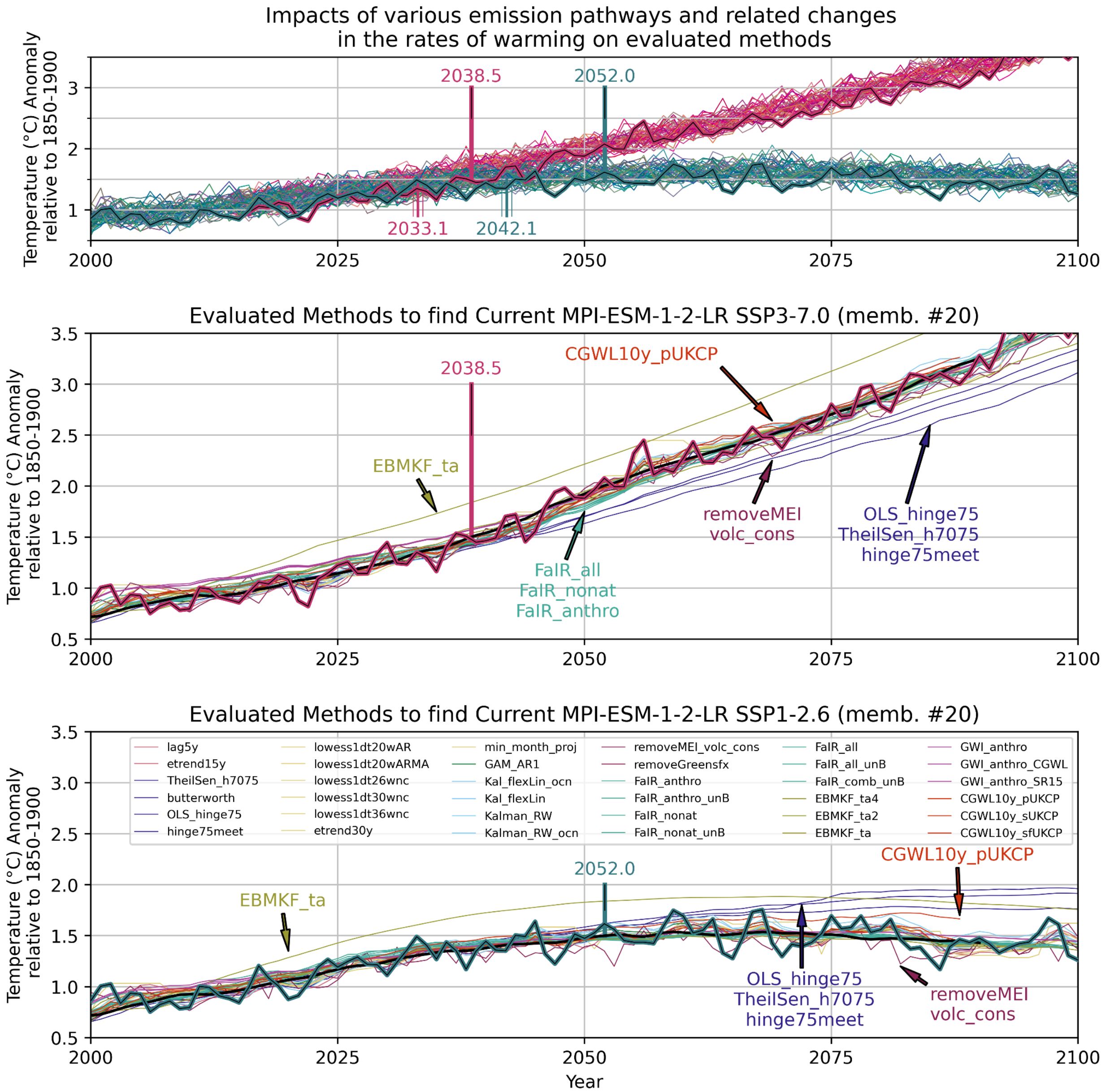


Figure 21. Results applying the subset of methods that passed historical observations tests and are able to be deployed for future test cases against MPI ESM1.2-LR SMILE ensemble member 11 for SSP1-2.6 (middle panel) and SSP3-7.0 (lower panel) across the 21st Century. The top panel illustrates the behaviour of the full SMILE ensemble and the years when the ensemble-mean 20-yr running mean passes 1.5°C are noted below the temperature time series. The 11th ensemble member is highlighted (heavy lines) because is at the cooler end of the ensemble for most years after 2000 and highlights the potential role of natural variability in delaying exceedance. The years noted above indicate that the 20-yr running mean of this ensemble member crosses 1.5 °C considerably later than the ensemble mean does. In the lower panels the heavy green and red lines denote the annual mean ensemble member timeseries from member 20 which lies close to the ensemble mean, the solid black the 20-year centred mean (the target statistic) and a selected group (following removal of methods as outlined in 4.6.1) of the techniques introduced in Sections 4.3 through 4.5 as coloured lines. See Supplement for further details.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Method Class	Surviving: RMSE 2025-2100<0.06°C for both SSP1-2.6 and SSP3-7.0	Untested methods (that passed Historical Evaluation)	Best in Class by RMSE for SSP1-2.6 and SSP3-7.0 combined	SSP1-2.6 RMSE	SSP2-4.5 RMSE	SSP3-7.0 RMSE
4.3.1: Running Means	0	0	lag5y	0.0549	0.0613	0.0835
4.3.2: Long Term Fits	0	0	butterworth	0.0519	0.00597	0.0805
4.3.3: Short Term Fits	9	0	lowess1dt36wnc	0.0455	0.0397	0.0424
4.3.4: Smoothing Splines	2	0	GAM_AR1	0.0575	0.0422	0.0483
4.3.5: Simple State Space	1	0	Kal_flexLin	0.0465	0.0421	0.0428
4.3.6: Filter Internal Variability	0	2 (0)	removeGreensfx	0.0648	0.0701	0.0893
4.4.2: Constrained Emulators	4	0	FaIR_nonat_unB	0.0572	0.0391	0.0456
			FaIR_comb_unB	0.0572	0.0406	0.0497
4.4.3: Energy Balance Kalman Filter	1	0	EBMKF_ta4	0.0358	0.0325	0.0295
4.4.4: Attributable Warming	5	6 (0)	GWI_tot_CGWL	0.0396	0.0400	0.0573
			GWI_anthro_CGWL	0.0422	0.0429	0.0595
4.5: Combining ESM Projections	2	3 (3)	CGWL10y_sfUKCP	0.0443	0.0563	0.0547

Table 3. Method classes and number and statistics of 24 methods remaining after an RMSE<0.06°C selection versus the 20-year running mean based on the SMILE simulations of the MPI-ESM1.2-LR (2025-2090) was applied. Several methods could not yet be applied to the future test case (for instance true decadal forecasts initialized from the climate model), these are tallied in the third column and further documented in the Supplement. Again, red highlighting indicates that 3 further classes of methods did not pass this selection step. Also again, the specific method with the lowest RMSE within each class is named, and rightmost columns show

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

the RMSE performance of this method for each future emissions scenario. The GWI_tot_CGWL and EBMKF_ta4 methods are highlighted in green because they have the lowest RMSE in these three future scenario evaluations.

Volcanic activity with high Volcanic Explosivity Index (VEI) values (generally >5) causes substantial short-term multi-annual cooling in GMST and could delay the long-term exceedance of any given warming level in an otherwise warming climate (Bethke et al., 2017, Nicklas et al., 2025; Chim et al., 2025). Long-term ice core records highlight substantial variability in past volcanic activity (Sigl et al., 2015). The potential impacts on decadal predictions (Wu et al., 2023, Sospedra-Alfonso et al., 2024, Bilbao et al., 2024) and long-term projections (Chim et al., 2023, Nicklas et al. 2025) and how these could be handled is increasingly recognised. Our assessment uses the NorESM1 model ensemble in Bethke et al. (2017), which provides 60 simulations under the RCP4.5 scenario combined with plausible volcanic futures consistent with the Sigl et al ice core reconstructions.

Volcanism affects realised warming, but not anthropogenic warming (which excludes volcanism and other natural contributions). The periodic cooling excursions caused by eruptions lead to delayed projections of the year of exceedance of any level of global warming by 1-5 years during periods affected by those eruptions (Nicklas et al. 2025). Importantly, capturing pdfs precise enough to sample over a wide range of realistic volcanic futures requires many ensemble members (e.g., Nicklas et al. (2025) use 6000 emulator ensemble members) making traditional ESM-based approaches prohibitively expensive. Emulator-based methods (such as FaIR, GWI and EBM-KF) are practical for assessing randomized, intermittent volcanic eruptions, provided that the radiative forcing of the eruption is accurately estimated. Unless ESM ensembles are annually reinitialized with updated aerosol measurements, as called for by Sospedra-Alfonso et al. (2024), they cannot adequately account for recent eruptions.

The issue is of particular relevance in the context of the LTTG and deriving and tracking mitigation policies for achieving it, since if cooling from volcanism is included in the estimate of warming, this may create a misperception about the temperature distance to the LTTG and the size of the remaining carbon budget (Chim et al., 2025). This is clearly depicted in Fig. 22(c) in the difference between the SR1.5 definition of warming (30-year mean *anthropogenic* warming, “GWI_anthro_SR15”, purple line) and AR6 WGI definition of warming (20-year mean *realised* warming, black line): both lines are extremely similar except during the volcanic period. Volcanism temporarily suppresses or “masks” the underlying and continuing anthropogenic warming.

The full ensemble is displayed in Fig. 22(a), which shows the different volcanic realisations and how the 1.5°C central exceedance year derived from the retrospective 20-year mean ranges from 2027—2052. The Fig. 22(b) simulation has the least volcanic activity with no major eruptions until the 2090s. The Fig. 22(c) simulation has substantial eruptions starting around

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Open Access Earth System Science Data Discussions

2040. In the Fig. 22(c) case, even the trailing 20-year average does not constitute a reasonable characterisation of realised warming, and almost all methods struggle in the vicinity of a substantial eruption.

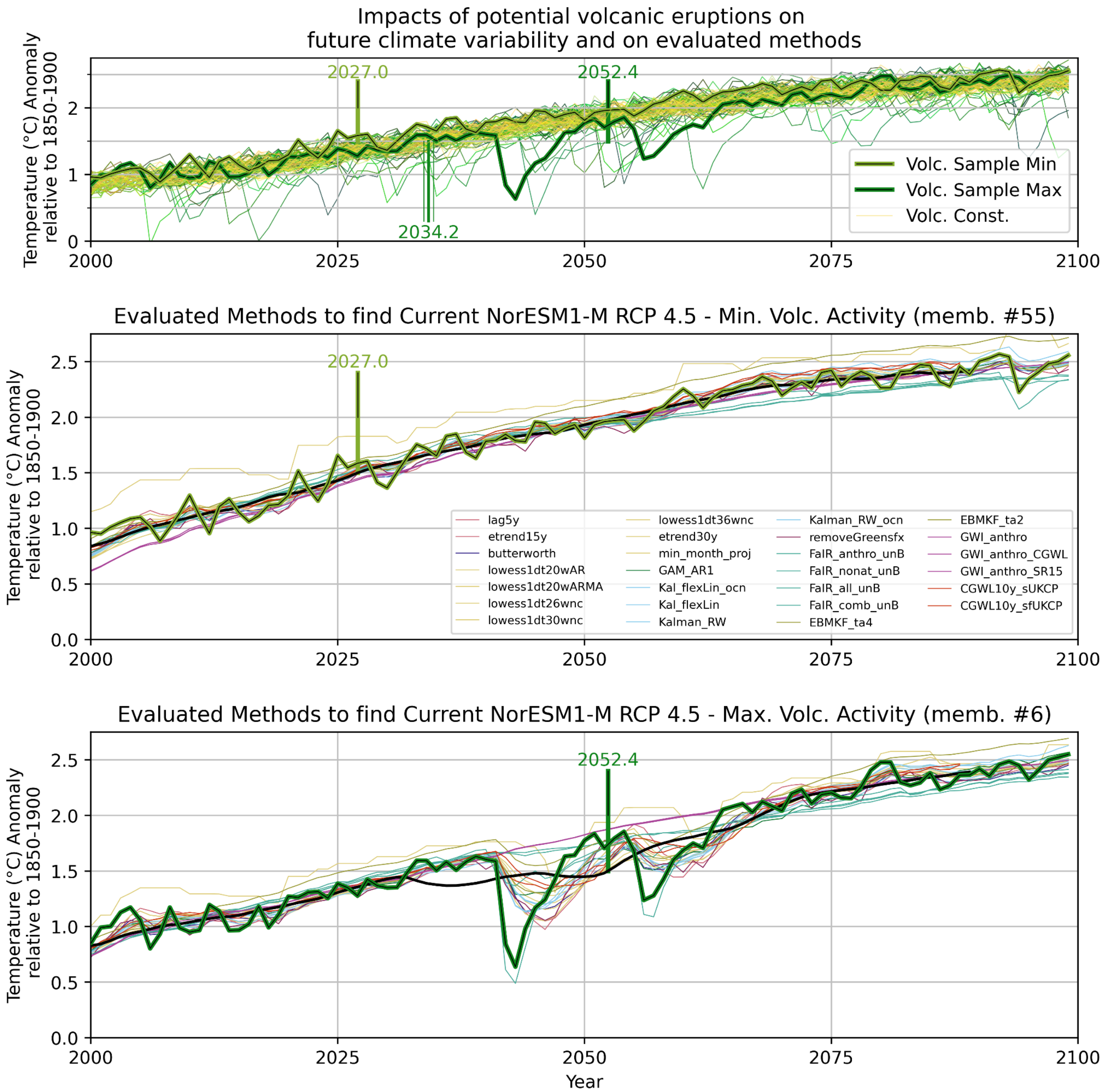


Figure 22. Robustness of climate state detection under volcanic events using a NorESM random-timing volcanic ensemble (Bethke et al. 2017) for the subset of methods skillful for both observations (Figure 18), including 0.5°C and 1°C thresholds (Figures 20, S1, S2), and SSP1-2.6 and SSP3-7.0 (Figure 21). The top panel illustrates the behaviour of the full NorESM ensemble and the ensemble-mean year of exceeding 1.5 °C (below temperature record) and two exceptional ensemble members (heavy lines) with their

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

exceedance years given above the record (the min/earliest and max/latest within the ensemble). In the lower panels the heavy green lines again show the annual time series of those exceptional ensemble members, the solid black the 20-year centred mean for those ensemble members time series (the target statistic), and coloured lines indicate the techniques introduced in Sections 4.3 through 4.5 that have proven skilful thus far.

Table 4 shows all methods that still remain as plausible robust methods for estimating realised and anthropogenic warming after removing methods that have RMSE>0.077°C in the volcanic future tests. Methods remain from all three major categories (statistical, attribution, and combination with projections). However, statistically-based methods only can be applied to realised warming estimation. Five flavours of the GWI approach remain and to avoid over-emphasising of this method class the best performing methods GWI_tot_CGWL for realised warming, and GWI_anthro_CGWL for anthropogenic warming are retained in subsequent analysis for consistency with the proposed approach for combining observational estimates described in Section 5.1.

Method Class	Final Surviving Methods by volc RMSE <0.077°C (2025-2090)	Realised: RMSE wrt member 20-yr cent. mean	Attributable: RMSE wrt ensemble 20-yr cent. mean
4.3.3: Short Term Fits	lowess1dt36wnc	0.0758	0.1304
4.3.4: Smoothing Splines	GAM_AR1	0.0655	0.1197
4.3.5: Simple State Space	Kal_flexLin	0.0727	0.1291
4.4.2: Constrained Emulators	FaIR_comb_unB	0.0677	0.0935
	FaIR_nonat_unB	0.0824	0.0681
	FaIR_anthro_unB	*0.0845*	*0.0688*
4.4.3: Energy Balance Kalman Filter	EBMKF_ta4	0.0601	0.0972
4.4.4: Attributable Warming	GWI_anthro_CGWL	0.0913	0.0412
	GWI_anthro_SR15	*0.0921*	*0.0430*
	GWI_anthro	*0.0916*	*0.0416*
	GWI_tot_CGWL	0.0557	0.0847

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Open Access Earth System Science Data Discussions

	GWI_tot_SR15	*0.0727*	*0.1100*
4.5: Combining ESM Projections	CGWL10y_sfUKCP	0.0765	0.1259

Table 4. Method classes and number and statistics of 13 methods remaining after an RMSE<0.077°C selection versus the 20-year running mean based on the NorESM1 volcanic simulations of Bethke et. al. (2017) (2025-2090) There are two potential targets: the 20-year running mean of that particular simulation (middle column) which approximates the realized warming, and the 20-year running mean of the entire ensemble (which approximates the human-attributable warming). The methods which are ill-suited for each of these respective targets are highlighted in red. Note that there are 8 methods that approximate realized warming, 4 that approximate attributable warming, and none that are proficient at both. Cells within the rows describing the GWI_tot_CGWL and GWI_anthro_CGWL methods are highlighted in green because they have the lowest RMSE with respect to these two targets. In subsequent analyses the 3 other GWI methods and 1 other FaIR method (italicised) are discounted to avoid over-weighting on this methodological approach.

4.6.3 Combining of candidate methods

The best performing overall methods across all tests in Sections 4.6.1-4.6.2 are shown in Figure 23 for exceedance of 1.5°C in each set of simulations. While different methods could be optimized in future for realised warming (left side) or anthropogenic warming (right side), the spread of methods' performance is broadly similar regardless of target. There are slight differences: e.g. GWI-anthro is slightly better at predicting the ensemble average (approximately the attributable anthropogenic warming) in the context of future volcanic eruptions. In contrast, the CGWL10y_sfUKCP (red) and lowess1dg20wnc (yellow) methods are better at predicting realised warming (especially apparent in the SSP3-7.0 and NorESM VolcConst rows). In the absence of volcanism almost all forced warming is anthropogenic, so the choice between realised or anthropogenic warming is less important than the emissions trajectory[34].

[34] SSP1-2.6 yields more difficulty in predicting the precise time of 1.5°C crossing as the GMST levels out at around 1.5°C in the MPI ESM, thus internal climate variability can substantially shift when this threshold is crossed.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

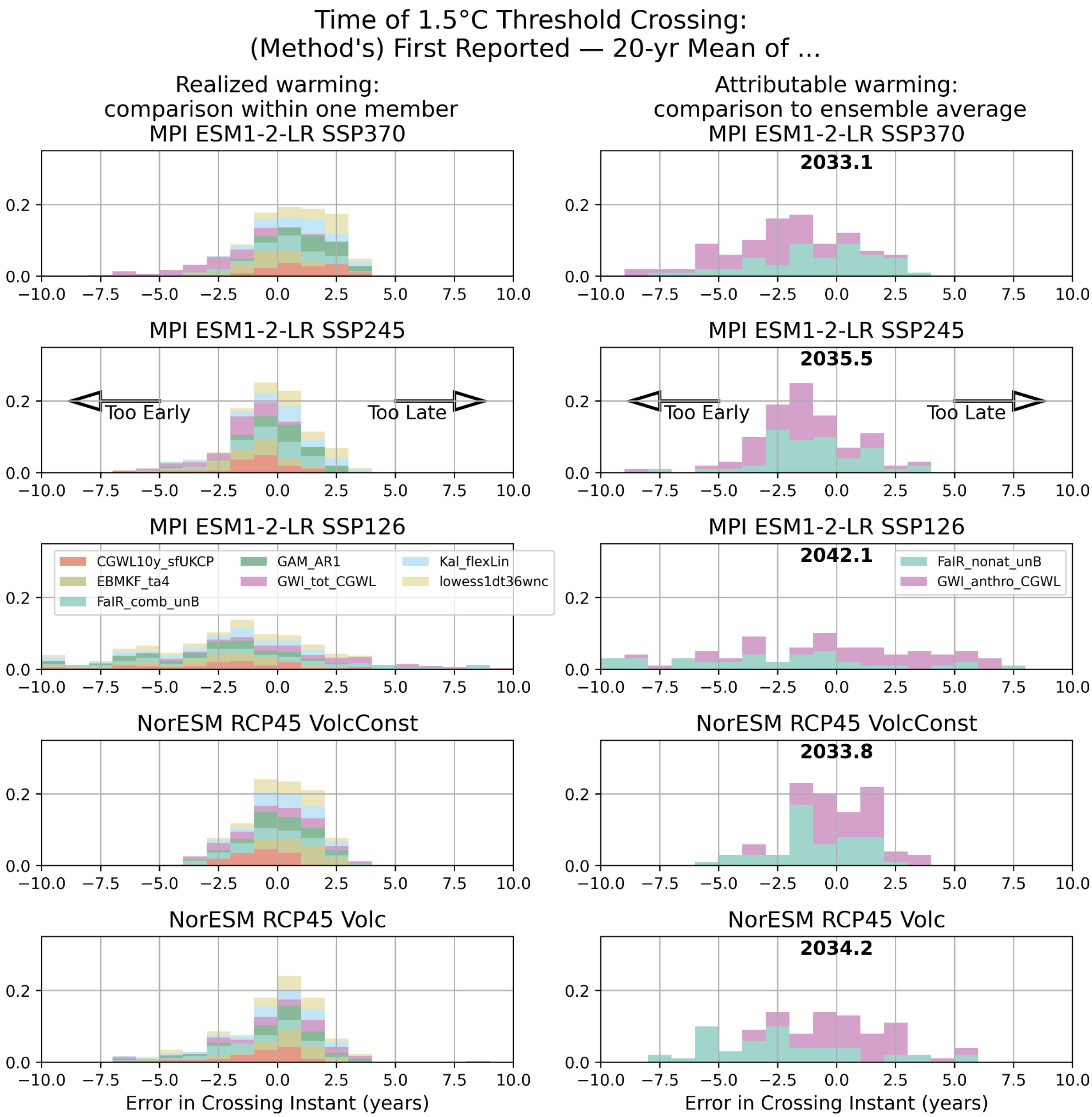


Figure 23. Apparent skill in synthetic cases considered in Section 4.6.2 for crossing of 1.5°C shown as probability density functions for the top performing techniques selected. The column on the left considers the target of observed warming: the year that the 20-yr centered running mean crosses 1.5°C in each ensemble member i.e. 'realised warming'. A difference of 0 represents that the method "hit" that target as it crossed the 1.5°C threshold, a positive year difference means too late, negative means too early. This figure does not consider the internal uncertainty reported by each method, so we use stacked histograms rather than smoothed pdfs. The right column is the same but for when the ensemble average of 20-yr centered running means cross 1.5°C (one target for all ensemble members - note these dates match Figures 21 and 22), i.e. approximately the total forced warming. Results are shown for three distinct SSP scenarios in the MPI ESM1.2-LR SMILE model (see also Figure 21) and the constant volcanic background and synthetic volcano ensembles in Bethke et al. (2017) for NorESM1 (see also Figure 22).

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

With this set of selected methods, the next consideration is whether a combination of such methods can be identified that outperforms any of them individually. If errors arising from different methods are largely uncorrelated, which may be the case at least between distinct classes of approach, they will tend to cancel out (Krogh and Vedelsby, 1994). Formally, this is known as a “M-open” setting, where none of the "current" methods is assumed to be true, and the combination must be optimized for predictive performance rather than a single best method identified (e.g. Le and Clarke, 2017, Yao et al., 2018, Hoge et al., 2020). The current setting differs slightly from more classical posings of this problem in two ways: firstly, the realised warming (of the 20-year centered mean) is not known precisely (see section 3.2.5) and thus is a tight distribution rather than a single number (see Figure 18), secondly and more importantly, it is not necessary to limit the combination method to be a simple mixture of methods.

Our goal is to combine the predictions from multiple methods in a way that accounts for their relative uncertainties. This is challenging, since the methods provide error estimates which may use different assumptions, may be non-Gaussian, or may be slightly biased (although our selection procedure removed methods with extreme bias). Given this diversity, two conceptually distinct combination approaches, primarily differing in their attention to the probability distributions provided by each of the tested methods, are tested.

The pointwise inverse-variance weighting (PIVW) approach (Cochran, 1954, Lee et al., 2016) provides a mean estimate equal to the weighted average of the predictive means, where the weights are inversely proportional to each method’s mean-squared error with respect to the pointwise 20-year centred mean. Other authors call this general approach quantile averaging (Fakoor et al., 2023) or linear pooling (Gneiting and Ranjan, 2013). The combined PIVW predictive distribution is described by this mean and a single, fixed standard deviation that maximizes the likelihood of the target distribution. Thus, the assumption is that the constituents and the total are all normally distributed. While straightforward and computationally efficient, it discards rich information about the methods’ probability distributions. For instance, the EBMKF uncertainty widens at times when observations drift from its energy balance model prediction.

We develop another approach by combining existing techniques, which we term compressed, stacked, calibrated mixture (CSCM) approach (further details provided in Supplement and software documentation). This approach exploits knowledge of the distributions output by each method, rather than using only the mean as in PIVW. Each method’s predictive distribution first undergoes a pre-scaling step, the same as was applied for evaluations within section 4.6.1 (a single, fixed variance-scaling parameter per method; Gneiting et al., 2007). Bayesian stacking (a specific type of probability averaging; Fakoor et al., 2023) is then used to form a weighted mixture of the calibrated predictive distributions, retaining their scaled probabilistic structure (Le and Clarke, 2017; Yao et al., 2018; Hoge et al., 2020). A similar approach has been used in weather forecasting (Hoeting et al., 1999; Raftery et al., 2005). To further sharpen this mixture, a fractional sampling scheme is used: samples are drawn

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

from the Bayesian stacked distribution, generating a distribution of the mean whose likelihood with respect to the target is maximized (pulled tighter by drawing more samples). This approach is more computationally intensive than PIVW, but the resultant temperature uncertainty increases if either the constituent methods' uncertainties increase or if the methods disperse (as they would after a volcano, see Figure 22).

PIVW prioritizes simplicity, while CSCM attempts fidelity to model uncertainty with tunable sharpness. As expected, both improve performance over the best individual method, and unexpectedly yield remarkably similar combined products. Both approaches require choices that determine performance, listed as follows in order of decreasing importance. We first considered which "calibration sets" of the climate data should be used to set the variance-scaling and weight parameters. Using only the historical record leads to more equal method weighting and lower uncertainty ranges both retrospectively and if the emissions continue on roughly their present course (approximately SSP2-4.5)[35] without major volcanic eruptions. Conversely, using all future 270 ensemble members in more standard "leave one out" cross-validation yielded estimates that were more robust to different emissions pathways and volcanic future scenarios, which we present in Table 5 below. Second, we found that both the PIVW and CSCM approaches were more robust to different future scenarios (especially SSP1-2.6) if the weights were trained on 90 years of error (2000-2090) with respect to the centred 20-year running mean, rather than just 65 years of error (2025-2090). Both the RSME and average-year Kullback–Leibler divergence are presented for both intervals in Table 5. Third, the number of methods that are combined matters very slightly: we started with the 7 best-performing ones (at achieving the realized warming target) selected in section 4.6.2, but also included a wider set of the 37 methods which passed the historical warming test in 4.6.1 (note a few of the full 40 which passed could not be evaluated in the future), and an intermediate set of 18 methods. Including 18 methods (or 37, see Supplement Table S4) in the combination rather than 7 produced almost imperceptibly different results: the RMSE improved by 0.0002 on average (maximum change of 0.0026) across the 5 future ensemble cases when switching from PIVW with 7 methods to CSCM with 18 methods (see Table 5).

	First crossing instant of 1.5°C			KL Divergence		# crosses	RMSE (20yr cent)	
Ensemble	within 1 yr	within 2 yrs	within 5yrs	2000-90	2025-90	of 1.5°C	2000-90	2025-90
	Compressed Mixture (CSCM), 18 methods							
MPI ESM1-2-LR_SSP3-7.0	55.7%	85.1%	100.0%	0.345	0.269	1.060	0.0365	0.0332
MPI ESM1-2-LR_SSP2-4.5	55.3%	83.5%	99.1%	0.474	0.465	1.080	0.0322	0.0300
MPI ESM1-2-LR_SSP1-2.6	20.6%	39.1%	76.0%	0.555	0.589	1.320	0.0354	0.0345
NorESM_RCP45_VolcConst	53.5%	86.7%	100.0%	0.568	0.475	1.017	0.0297	0.0295
NorESM_RCP45_Volc	30.1%	55.2%	88.5%	0.721	0.710	1.100	0.0492	0.0508

[35] The match is not exact. In particular the current policies have much greater controls on aerosol pollutants than assumed in SSP2-4.5

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

	Inverse Variance (PIVW), 7 methods							
MPI ESM1-2-LR_SSP3-7.0	54.6%	85.0%	100.0%	0.352	0.271	1.080	0.0368	0.0335
MPI ESM1-2-LR_SSP2-4.5	55.7%	85.1%	99.2%	0.488	0.481	1.040	0.0322	0.0301
MPI ESM1-2-LR_SSP1-2.6	20.8%	39.3%	75.5%	0.568	0.602	1.460	0.0356	0.0348
NorESM_RCP45_VolcConst	53.3%	87.1%	100.0%	0.591	0.623	1.017	0.0300	0.0300
NorESM_RCP45_Volc	29.6%	54.3%	87.9%	0.711	0.771	1.100	0.0495	0.0510
	EBMKF_ta4							
MPI ESM1-2-LR_SSP3-7.0	64.2%	92.1%	100.0%	0.583	0.365	1.060	0.0391	0.0295
MPI ESM1-2-LR_SSP2-4.5	51.0%	81.0%	98.7%	0.564	0.391	1.020	0.0396	0.0325
MPI ESM1-2-LR_SSP1-2.6	21.2%	40.4%	76.9%	0.637	0.525	1.540	0.0415	0.0358
NorESM_RCP45_VolcConst	49.9%	85.2%	100.0%	0.572	0.425	1.000	0.0291	0.0249
NorESM_RCP45_Volc	35.3%	63.0%	92.1%	0.695	0.653	1.117	0.0596	0.0601
	GWI_tot_CGWL							
MPI ESM1-2-LR_SSP3-7.0	15.4%	32.1%	67.7%	0.723	0.653	1.000	0.0479	0.0396
MPI ESM1-2-LR_SSP2-4.5	32.3%	61.0%	92.6%	0.712	0.610	1.000	0.0515	0.0400
MPI ESM1-2-LR_SSP1-2.6	16.5%	34.2%	82.4%	0.755	0.655	1.000	0.0682	0.0573
NorESM_RCP45_VolcConst	34.1%	63.5%	98.9%	0.733	0.674	1.000	0.0480	0.0425
NorESM_RCP45_Volc	24.2%	46.0%	84.2%	0.789	0.730	1.050	0.0601	0.0557

Table 5: Comparison of the performance of two different approaches to combining the methods, CSCM and PIVW, with the two methods that had the best RMSE in one of the 5 future ensembles, GWI_tot_CGWL (NorESM_RCP45_Volc) and EBMKF_ta4 (all others). Note that even though GWI_tot_CGWL has better RMSE in the 2025-2090 interval for the NorESM Volc experimental case, it usually has a lower chance than EBMKF_ta4 of being within 1, 2, or 5 years of the realized warming 20-year centered 1.5°C crossing year (in all ensembles except the NorESM ensemble without volcanoes). Both combination methods yield similar results, and are better than the best individual methods in the SSP2-4.5 case (and comparable to the best individual method in other future cases).

4.7 Summary of assessment of approaches to estimating present-day warming

Given the difficulty in removing stochastic internal variability to produce real time estimates rather than retrospective ones — a plurality of methods is warranted. We analyzed 71: a variety of statistical, observational, and physically-informed techniques, including short- and long-term fits, Kalman filter-based, energy balance models, constrained emulators, and combined projections using ESMs, many of which can reliably approximate the IPCC's 20-year centred mean used as a working benchmark. These methods vary in their responsiveness to internal variability, their robustness to extreme events such as volcanic eruptions, and their ability to attribute observed warming to anthropogenic causes.

While methods such as simple lagging means or full-period OLS linear trends remain in common use, they were found to be systematically biased, particularly when changes are non-linear. By contrast, 5 emulator-based / ESM-based methods (versions

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

of FaIR, GWI, CGWL, and EBMKF) demonstrated low error, high likelihood, and responsiveness to emerging climate signals, as they are attentive to climate forcings in addition to the temperature record. For the aim of approximating realized warming, we chose to include only 1 variation of GWI: GWI-tot-CGWL and not GWI-tot-SR15, to give equal representation across the 4 method classes in this "top-tier" final set. Out of the approaches that are blind to climate forcings, only 3 of 36 were of comparable quality (GAM_AR1, Kal_flexLin, and linear LOWESS smoother with a 20-year Gaussian kernel). These methods proved resilient under plausible future scenarios, including diverging emissions pathways and sequences of intense volcanic eruptions.

Of these final selected methods, 4 are primarily and inflexibly estimating total observed warming: all temperature-only methods (GAM_AR1, Kal_flexLin, lowess1dtwnc) and any implementation of CGWL (either with decadal forecasts or UKCP ensembles). The remaining selected methods (FaIR, GWI, and EBMKF) are all capable of separating the temperature signal into its anthropogenic and natural components (see Figure 15), and thus can all be tailored to estimate either total observed or human-induced warming definitions. In our present collection, EBMKF_ta4 and FaIR_comb_unB are better suited to total observed warming, whereas FaIR_nonat_unB and especially GWI_anthro_SR15 are (by design) better suited to human-induced warming.

Finally, we perform inverse variance weighting (PIVW) and Bayesian Stacking (CSCM) to combine the predictions of several of these realized warming methods (both the 7 "top-tier" realised warming methods and larger sets meeting less strict requirements). We find that this Bayesian Stack is sufficiently accurate to predict when the 20-year centered mean crosses 1.5°C within 1 year with probability >53% in most future scenarios, with the exception of SSP1-2.6 (~21%) which in the MPI model stabilises around this level of global warming. In the volcanic ensemble, there was a ~30% chance that this combination was within 1 year of the realized warming 1.5°C level. We chose to not combine the best attributable warming methods, but these 2 (5 including GWI and FaIR variants) "top-tier" definitions of attributable forced warming were highly consistent.

5 A proposed approach to monitoring present-day warming

Based on the considerations discussed previously, there are multiple potential ways to construct a scientifically robust synthesis that combines lines of evidence for tracking progress in a probabilistic manner, fully accounting for a range of uncertainties. A transparent and well-documented approach is needed, one that can provide both a best estimate of warming-to-date together with an uncertainty interval that can be used to provide probabilistic answers or an IPCC-like calibrated confidence statement. Such an estimate would be helpful for communicating with key audiences that have diverse needs. Here we propose the basis for such an approach building upon the preceding analysis, for a unified data estimate. It explicitly does not preclude individual dataset providers or others updating and communicating their own estimates as they see fit.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

5.1 Methodological approach

If we wish to provide an observation-informed global warming estimate, some choices are *de facto* based on the current state of scientific knowledge. Firstly, given the present preponderance of GMST products, estimates will principally arise from SST-based GMST metrics rather than MAT-based GSAT metrics. However, with the prevalence of MAT-based bias adjustments in SST datasets, it is not possible to provide a purely GMST-based assessment of global warming either (Section 3). Secondly, as was explained in Section 2, less complete instrumental data prior to 1850 makes it necessary to use a period such as 1850-1900 as a proxy for pre-industrial levels. These choices should be revisited and assessed periodically to ensure they remain valid (Section 7).

The proposed approach is outlined in the following 4 steps:

1. Split the observational record into two sub-periods to maximise the available lines of evidence
2. Objectively select observational data products
3. Combine data products to give a consensus ensemble
4. Apply methods that passed the tests in Section 4 to derive estimates of present-warming that approximate the 20-year centred mean to this consensus ensemble

Step 1: Split the observational record into two sub-periods to minimise the impact of early period of record uncertainties and maximise the available lines of evidence

Currently, estimates of present warming are based on only a subset of available information. Not all datasets run from 1850 to present nor are they all updated in a timely manner. Both limitations would have made them ineligible for inclusion in the IPCC AR6 approach for estimating long-term change. Excluding modern reanalysis and satellite-based products results in loss of information on recent warming. Similarly, datasets that are not routinely updated can inform our assessment of early period warming and, crucially, its structural uncertainty. Splitting the observational record in two would allow a much wider range of groups to contribute their analyses, including groups who cannot commit to producing data for the full period or to providing continual updates.

The first proposed step is therefore to split the period in two: the first part running from the pre-industrial proxy period (1850-1900) to a common modern baseline period (1981-2010), and the second from the same baseline period to present. Such a baseline matches the WMO definition of 30 years for calculating a climate normal[36]. The period 1981-2010 encompasses the start of the satellite era when reanalyses become better constrained and represents a peak in the number of available surface observations. It also closely matches within hundredths of a degree (depending upon dataset choice) the global average for the

[36] The WMO climatological standard normal is the most recent 30-year period finishing with a year ending in 0 (1991-2020 at the time of writing), but 1961-1990 is retained as a reference period for long-term climate change assessment.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

1986-2005 subperiod used as a baseline in many impact studies that informed the 2015 SED and therefore the Paris Agreement, providing a tie to the LTTG.

Step 2: Objective selection of data products that can be used for each period

Based upon Section 3 some form of dataset selection is required, as was done in IPCC AR6 WGI (Gulev et al., 2021). Criteria for inclusion will differ in each sub-period.

We propose the following qualification criteria:

1. The data product should be freely available without restrictions for this use.
2. A peer reviewed paper should describe the dataset creation. Subsequent updates or partial replacement of source data, and minor changes to accommodate source data changes, are acceptable without requiring a new peer reviewed paper.
3. The data product should account for major sources of bias and quantify uncertainties that were documented in the peer-reviewed literature prior to acceptance of the paper describing the dataset.
4. The data product makes efforts to account for the bias of incomplete global observational coverage, such as by statistical infilling.
5. For the first sub-period, data should exist from 1850 to 2010 inclusive. For the second sub-period, data should exist from 1981 to the last full year inclusive.
6. Exclude datasets that differ only in minor variations used for sensitivity tests, for example different treatments of sea ice by Berkeley Earth, HadCRU_MLE, and China-MST2; or GMST time series of reanalyses which are reported as GSAT. In addition, include the latest central estimate and the newest uncertainty estimates, even if the uncertainty estimates refer to an older version. This can occur for ensemble-based uncertainty estimates, when the ensembles are only run periodically.
7. For inclusion in the modern period updates, the data product update for the prior year should be available by 31 January.

A data product should in general be removed from the mix when any of the following conditions are met:

1. A newer version becomes available
2. It no longer reflects the current state of the art.
3. Under the advice of the data product producer.

Products that were deemed by the author team to meet these criteria in September 2025 for the period covering 1850-2010 (or later) are:

- HadCRUT5.1.0.0 analysis,
- HadCRU_MLE v1.3,
- Kadow et al. (updated using HadCRUT.5.1.0.0),
- Berkeley Earth,
- NOAAGlobalTempv6.0,

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

- NOAAGlobalTempv5.1[37],
- China-MST3.0-Imax,
- DCENT_MLE[38],
- COBE-STEMP3, and
- GloSATref.

Products that as of September 2025 match the criteria for modern era monitoring are:

- HadCRUT5.1.0.0 analysis,
- Berkeley Earth,
- NOAAGlobalTempv6.0,
- GISTEMPv4,
- China-MST3.0-Imax
- ERA5, and
- JRA-3Q.

Note that this does not exclude the use of other datasets that meet the criteria in future and a mechanism is required to enable regular reviews and updates (see Section 7).

It is proposed that the estimate of historical warming from 1850-1900 to 1981-2010, and its associated uncertainty, be updated only if new information is likely to lead to a material change in the estimate or its uncertainty. To this end, an update would be triggered by any one of the following criteria being met:

1. Two or more contributing datasets issue new versions that change their estimated historical warming to date relative to the version used to previously form this composite by more than 0.05°C; Or
2. A new dataset is produced whose estimated historical warming differs from the current best estimate by more than 0.05°C; Or
3. More than half of the contributing series have new versions or more than two new potential series that meet the qualification criteria have become available since the last update to this estimate; Or
4. If it has not been updated for at least 5 years

It is therefore necessary to keep time-stamped versions of the datasets used for the early period, as some datasets are updated through their entire length at various intervals. For example, NOAAGlobalTemp and GISTEMP change every month and version updates from any product (e.g. HadCRUT5) can affect the whole record.

Step 3: Combination of the qualifying datasets to yield a consensus-estimate and its uncertainty

[37] Note that both versions are used because they use the same underlying datasets but employ distinct interpolation schemes such that they are equivalent in methodological freedom to HadCRUTv5 vs its various interpolated by third party versions. Only v6.0 is updated in NRT

[38] DCENT-I would also satisfy our criteria, but we did not have time to include it in our analysis.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

The aim is to generate an ensemble that is built from merged series for the two sub-periods. Like Gulev *et al.* (Subsection 2.3.1.1.3, 2021) and Craigmile and Guttorp (2021), our approach is intended to account for both structural and non-structural uncertainties of the temperature datasets. As discussed in Section 3.2.5 and shown in Figure 8 there are far fewer degrees of freedom than implied by a simple count of available products. An appropriate way to handle issues of non-independence is therefore required (Step 3.1). In addition, not all datasets have uncertainty information, so a way is required to estimate uncertainty for all datasets (Step 3.2). A large ensemble (10,000 members) is generated to ensure the uncertainties are well sampled (Step 3.3) and then a representative sample of practical size (100 members) is derived from it (Step 3.4).

Step 3.1: Define hierarchical models using tree structures for each period

For both periods, we construct hierarchical models that effectively assign subjective probabilities to each GMST/GSAT dataset using a family tree of temperature data products. We assign 100% probability to the base of the tree and then, for each branching of the tree, we assign equal probability to each branch. Because SST/MAT uncertainties are the dominant sources of both structural and non-structural uncertainties, we use the family tree structure defined by Figure 24, which assigns equal probability to each SST dataset.

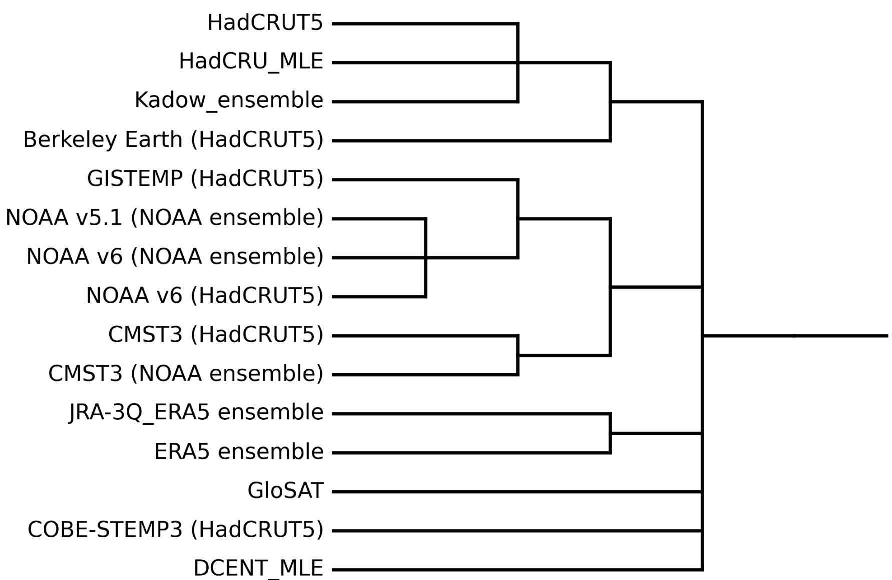


Figure 24. Family tree approach adopted in the production of the large ensemble estimate in the current paper. Datasets are merged hierarchically based upon similarity in the sea surface air temperature dataset / approach. The dataset names in parentheses are those used to provide an ensemble for non-ensemble datasets. Alternative merging hierarchies are shown in the Supplement.

For example, four datasets used HadSST4: HadCRUT5, HadCRUT_MLE, Kadow *et al.*, and Berkeley Earth. Of these four, there is a further natural grouping, with HadCRUT5, HadCRUT_MLE and Kadow *et al.* all based on a non-infilled version of

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

HadCRUT5. Berkeley Earth differs from these in having its own LSAT analysis, so it is placed on a separate sub-branch. Reanalyses were given their own grouping, partly reflecting their very different construction, but also because they represent estimates of GSAT rather than GMST. An alternative approach that groups the two reanalysis datasets with the most closely related SST dataset was also tried and has little effect on the outcome.

The aim is to ensure that excessive weight is not given to those datasets which have the most variants. This is particularly a concern for HadCRUT-based datasets, which are numerous because it is a popular basis for testing new interpolation methods (HadCRUT_MLE, Kadow *et al.*, GETQUOCS, and Vaccaro *et al.* are all based on various versions of HadCRUT). Alternative family tree structures were also tested including trees that grouped datasets based on similarities / distinctions in other choices (see sensitivity tests in Supplement).

Step 3.2: Impute missing uncertainty information

There are multiple challenges when assessing uncertainty in GMST/GSAT time series. Temperature product uncertainties vary over time and are autocorrelated. Some datasets provide no error estimate, or only a confidence interval over a given averaging period (e.g. monthly or annually). Confidence intervals account for changing error variance, but not changing autocorrelation. Some products provide ensembles of possible realisations, which account for both time-varying variance and autocorrelation. We prefer ensemble-based estimates, and we keep the non-ensemble products by using an "ensemble-donor" approach for their uncertainty estimates.

The ensemble mean is subtracted from each donor ensemble to obtain an uncertainty ensemble and three ensemble datasets were used as donors. We primarily used the HadCRUT5 Analysis ensemble as it covers the period 1850 to present. We used the NOAAGlobalTempv5.0 ensemble for ERSST-based datasets for the 1850 to 1995 period. The Bessel-corrected ERA5 ensemble was used for the two reanalysis datasets: ERA5 and JRA-3Q. While the ERA5 ensemble substantially underestimates uncertainty (ECMWF, 2024), grouping ERA5 with JRA-3Q helps account for structural uncertainty of reanalysis datasets.

For datasets without uncertainty estimates, the donor uncertainty ensemble was added to the dataset's best estimate time series to generate an ensemble. For datasets with uncertainty estimates but no ensemble, the uncertainty ensemble was first scaled by the dataset's uncertainty SD time series. The SD was derived from confidence intervals by assuming multivariate normal distributions, e.g. Berkeley Earth's 95 % confidence interval was divided by 1.96 to obtain the SD. If only monthly confidence intervals are provided, we assume perfect correlation on sub-annual scales so use the derived monthly SD as the annual SD. This rescaling was also performed for datasets with ensembles to account for datasets with ensembles that did not include all sources of uncertainty (e.g., HadCRUT5 Analysis, GloSAT). Further details are given in the Supplement.

Step 3.3: Generate an ensemble of temperature time series using the hierarchical models

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

We generated 10,000 ensemble members that covered the whole period 1850-2024 by combining a tail dataset covering the period 1850-1995 and a head dataset covering the period 1996-2024.The 1995-1996 transition was chosen because it is the centre of the 1981-2010 modern baseline used for both sub-periods. Tail and head datasets were selected at random with probability given using the appropriate dataset family tree (Step 3.1). At each branching of the tree, a branch was selected at random with each branch having equal probability. Once a dataset was selected, a random ensemble member (or pseudo ensemble member) was chosen for that dataset. The averages of anomalies for the 1981-2010 period were subtracted from each of the head and tail datasets so that the average of each one was zero during the overlap.

Step 3.4 Choosing a representative sample from the large ensemble

10,000 ensemble members give a stable estimate of the mean and standard deviation of the ensemble. However, some of the methods used in Section 4.6 are computationally expensive and using 10,000 ensemble members is impractical. Consequently, a 100-member subset of the larger ensemble was chosen for input to the final calculation using balanced k-means clustering. Balanced k-means (Malinen and Fränti, 2014) clustering generates equal sized groups of ensemble members which are close together (as measured by the sum of squared differences between series). A single member of each cluster is then chosen at random. This ensures that the 100-member ensemble is representative of the larger ensemble and its members preserve the time-varying variance and autocorrelation structure of the products. Alternative approaches were tested (see sensitivity tests, Supplement).

Figure 25 shows our generated annual time series ensemble members. The median estimate of GMST/GSAT of the 2024 annual anomaly relative to 1850-1900 is 1.55 °C (with a 95% range of [1.37, 1.77]°C), and the mean estimate is 1.55°C. This is identical to the headline estimate in WMO State of the Climate results (WMO, 2025), and the ranges are consistent.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

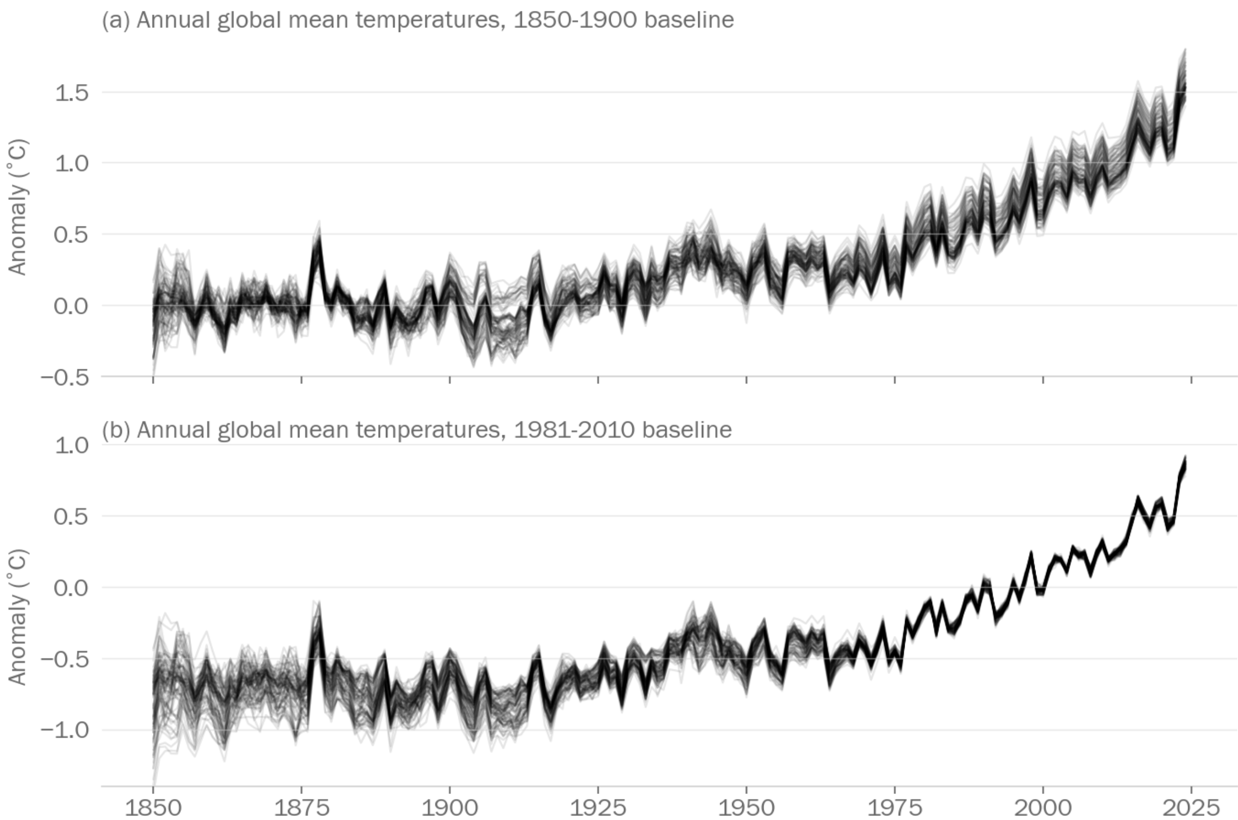


Figure 25. Annual global mean temperature anomalies from the SST family tree relative to (a) the 1850-1900 baseline and (b) the 1981-2010 baseline. The 100 ensemble members selected from the much larger ensemble are shown as individual lines.

Step 4: Applying approaches to estimate global warming to date described in Section 4

Two examples were given at the end of Section 4 as to how to combine the multiple methods in Section 4 and their estimated probability distributions into a synthesis across all methods given an observed temperature and forcing record. Combining a set of low-cost, high-skill methods, such as the 7 selected at the end of Section 4, provides a synthesis probability distribution for the method uncertainty of realised warming level and attributed anthropogenic warming level.

The methods described in Section 4.4 require forcing data and also ocean heat content data in the case of the EBM-KF method. While there is substantial uncertainty for each of the individual forcings, these are already addressed internally within each method. An ensemble of top-of-atmosphere forcing observations is exported from a version of FaIR calibrated on just HadCRUT5 observations to the EBM-KF method. The ocean heat content ensemble is created by a similar method to the tree method of selecting among temperature records above, selecting from available reanalysis and observational datasets. The details are provided in the Supplement Section 2.4.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Figure 26 shows our estimates of realised and anthropogenic warming and their uncertainty relative to an 1850-1900 baseline. The PIVW estimates combining 7 realised and 2 anthropogenic records of GMST/GSAT change from 1850-1900 to 2024 are 1.40°C (with a 95% credible interval of [1.23 - 1.58]°C) and 1.34 °C (with a 95% credible interval of [1.18 - 1.50]°C), for realised and anthropogenic respectively, when the preindustrial uncertainties are included (Figure 26 upper panel). The PIVW weights for the climate state temperature as of the end of 2024 are based on the statistics of each method as evaluated over the HadCRUT5 historical record, and are as follows: realized warming: lowess1dt36wnc 7.2%, GAM_AR1 6.7%, Kal_flexLin 7.8%, FaIR_comb_unB 25.5%, EMBKF_ta4 15.5%, GWI_tot_CGWL 26.9%, CGWL10y_sfUKCP 10.5%; anthropogenic warming: FaIR_nonat_unB 72.4%, GWI_anthro_CGWL 27.6%. An alternative weighting (see Section 4.6.3 and Table 5) is possible that also includes the methods' skill in analyzing future projections, but those projections are based upon ESMs with irreducible biases and uncertainties.

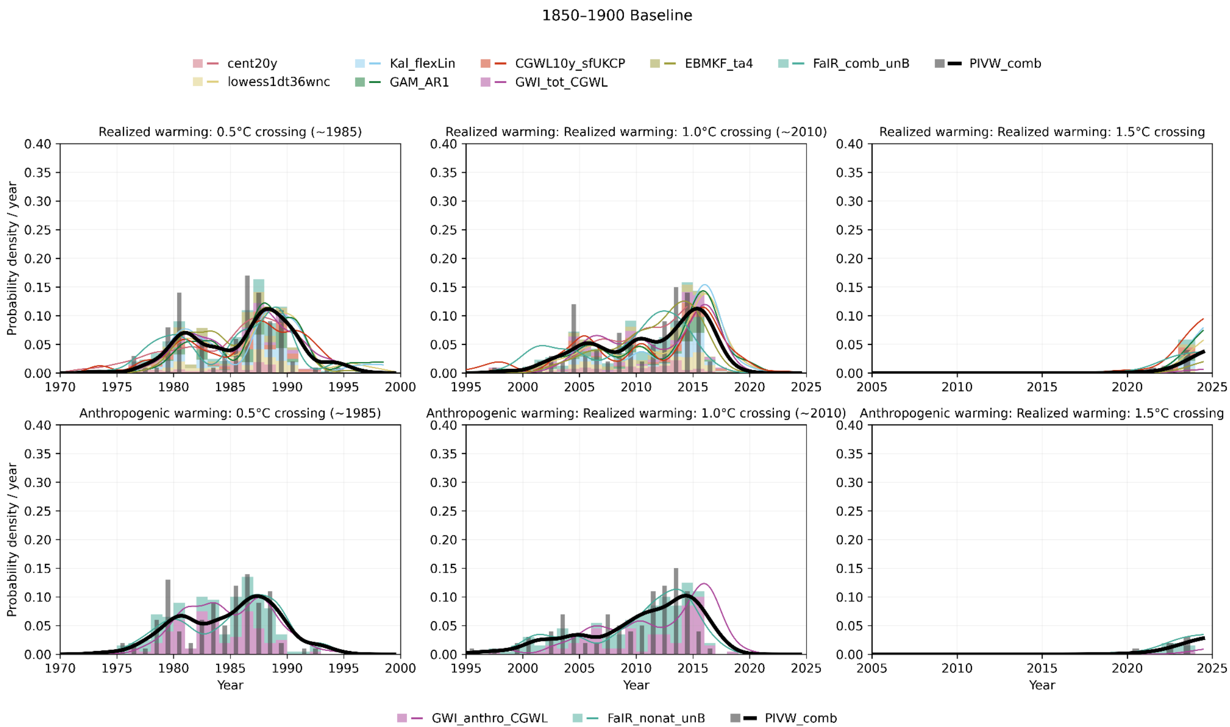


Figure 26. Outcome of examining the crossing times with a 1850-1900 baseline period for realised and anthropogenic warming using multiple methods and multiple temperature records simultaneously. The upper row shows the crossing times for 0.5°C, 1.0°C, and 1.5°C thresholds for realized warming. This row shows all 7 gold-colored methods in Table 4 in two ways, using an ensemble of 100 k-weighted mean samples of plausible temperature time series drawn from the observational syntheses from 1850-2024 described above. The stacked histograms indicate the 7 methods being applied to these long records, so the most likely crossing year weighting all methods equally is the tallest histogram bar. The smooth curves include a Bayesian estimation of other uncertainties internal to each method, in addition to the ensemble estimated uncertainty illustrated by the histograms. Each smooth curve indicates only one method, but all 100 ensemble samples. The thick smooth black lines and black histogram combine all 7 methods using the PIVW

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Open Access Earth System Science Data Discussions

combination, as described above. The lower row shows the crossing times for the 0.5°C, 1.0°C, and 1.5°C thresholds for anthropogenic warming with the 2 blue-colored methods in Table 4, using the same ensemble of 100 k-weighted mean samples of the temperature timeseries. Note that 67/700 individual methods*ensemble samples predict the realized warming crossing of 1.5°C by 2024 using the 1850-1900 baseline, and 6/100 PIVW realized warming ensemble samples crossed 1.5°C. Similarly 20/200 individual methods*ensemble samples predict the anthropogenic warming crossing of 1.5°C by 2024 using the 1850-1900 baseline, and 6/100 PIVW anthropogenic warming ensemble samples crossed 1.5°C.

If, instead, we assume a warming to 1981-2010 of exactly 0.7°C, the PIVW estimate combining 7 realized and 2 anthropogenic records of GMST/GSAT change from the 1850-1900 assumed temperature datum to 2024 are then 1.36 °C (with a 95% credible interval of [1.32 –1.41]°C) and 1.33 °C (with a 95% credible interval of [1.28 – 1.38]°C), for realised and anthropogenic respectively (Figure 27). It follows that the lack of knowledge of the preindustrial conditions dominates the uncertainty in both estimates and also decreases the mismatch between best estimates. The uncertainty ranges on both metrics reduce approximately 5-fold. If warming to 1981-2010 is assumed to be precisely 0.7°C then as at the end of 2024 a level of 1.5 °C above preindustrial levels would be at most *exceptionally unlikely* (more than 5 standard deviations away from the central warming estimate) if not certain not to have been reached for either metric. With the 1981-2010 baseline there was not a single observed 1.5°C crossing to date among the 100 ensemble members and 7+2 methods. Although some realised methods rise above 1% crossing probability in the lower-right panel of Figure 26 (GAM_AR1 at 2.6% and CGWL10y_sfUKCP at 6.2%), these methods carry relatively little weight (6.7% and 10.5%) in the PIVW combination, which moreover has greater certainty (allowed by its construction) so probability is concentrated around 1.36°C rather than spilling up to above 1.5°C.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

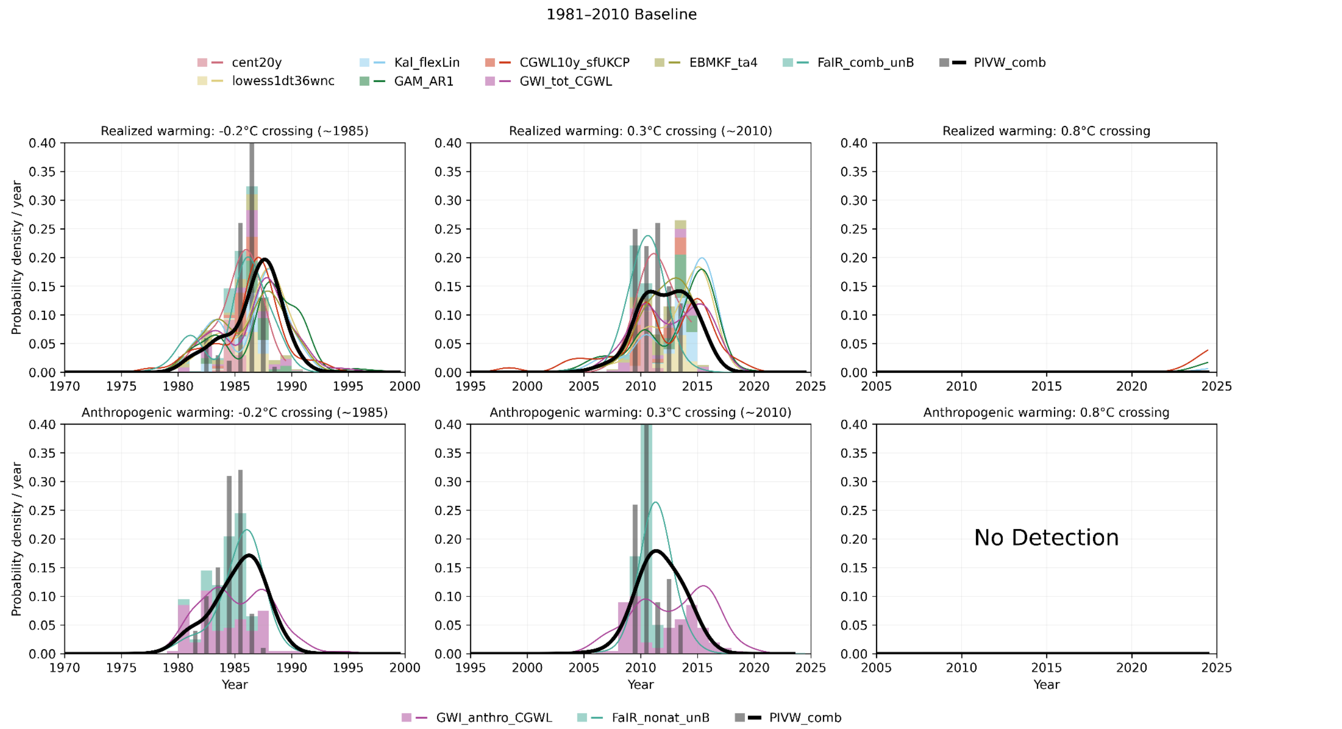


Figure 27. Outcome of examining the crossing times for realised and anthropogenic warming using two methods and multiple temperature records simultaneously using a baseline of 1981-2010. As in Figure 26 above, the upper row shows the crossing times for 0.5°C, 1.0°C, and 1.5°C thresholds for realized warming with the 7 gold methods in Table 4. The lower row shows the crossing times for the 0.5°C, 1.0°C, and 1.5°C thresholds for anthropogenic warming with the 2 blue methods in Table 4, using the same ensemble of 100 k-weighted mean samples of the temperature timeseries. These panels are all distinguished from their counterparts in Figure 26, by assuming that that baseline has *exactly* the temperature of 0.7°C above an arbitrary preindustrial datum. The decrease in uncertainty relative to Figure 26 results only from the changed baseline, that is, the wider range of the upper row indicates the challenge in establishing a clear preindustrial temperature. Note that no anthropogenic methods predict the crossing of 1.5°C using the 1981-2010 baseline to have occurred by 2024, hence the histogram does not appear in the lower right panel.

Both realised warming and anthropogenic warming are policy relevant and can differ by amounts of the order 0.1°C over decades (Section 4.2), although Figures 26 and 27 show very similar threshold crossing statistics for 0.5°C and 1.0°C. We argue that a best estimate and uncertainty for both of these metrics should be consistently provided and communicated. Our estimates may be missing some sources of uncertainty, such as highly correlated structural uncertainty common to all GMST/GSAT datasets, or uncertainty due to differences between GMSTs and GSATs, but the incorporation of emulators (Sections 4.4, 4.5 and Figures 26 and 27) that also draw on OHCA and forcing datasets reduces this risk somewhat. Since we use a hybrid GMST/GSAT approach (albeit dominated by GMST-based estimates), we recommend that our results should be compared to both GMST and GSAT simulations of climate models.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

5.2 Communication and visualisation

Given the inherently probabilistic nature of the estimation of the current warming it is essential to prominently communicate uncertainty information. One option is to use the likelihood language used in IPCC assessments (Mastraneda et al., 2010). This would have the benefit of being easy to communicate and is well known to policymakers, although care is needed to avoid misinterpretation of likelihood terminology by users not familiar with the IPCC definition of likelihood terms (e.g. Budescu et al 2014, Ho and Budescu 2019). Stating a *likely* or *very likely* range would also help to avoid the potential for undue focus upon the exact median value of the range of possible values by users. In Section 5.1 the proposed method yields two ranges: one for realised warming and one for anthropogenic warming. The resulting PDFs can be used to infer directly a probability of exceedance of any given warming level under either metric. The proportion of the distribution lying above the target warming level corresponds directly to the likelihood of exceedance (conditional upon the assumptions underlying the construction of the PDF). Nevertheless, some quantitative information most sensibly as an estimate of current warming +/- a range will be required for both warming to date metrics, with the range potentially described in qualitative (likelihood) and probabilistic terms for more robust communication.

While an operational website is beyond the scope of the current unfunded work Figure 28 provides a potential starting basis for communicating on an ongoing and operational basis the results arising from the methods suggested in Section 5.1. WMO are currently assessing options for ongoing operationalisation. As well as providing diagnostics of the current warming level it would be important to communicate certain principles and potential outcomes clearly. Firstly, showing via a simplified example how with continued warming a given warming level progresses from not being possible through exceptionally unlikely and on through subsequent stages to being unequivocal would be useful. Next, possible scenarios and how they might impact the two metrics would be useful. This might include the potential impacts of one or more large volcanic eruptions (Section 4.6.3), or the difference between reaching, exceeding and returning to a specified warming level (Section 4.6.2).

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Open Access Earth System Science Data Discussions

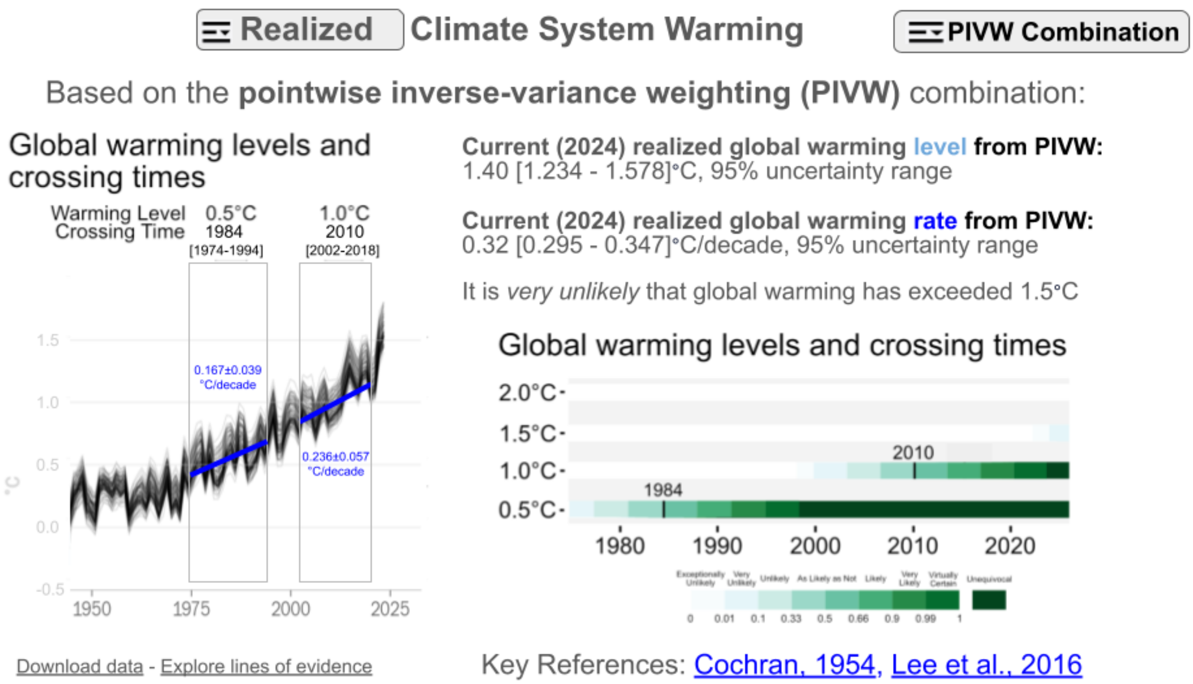


Figure 28. Mock up of a potential approach to communicating the current trajectory towards the LTTG on an interactive website. The pulldown menu at top left allows the user to select between realised and anthropogenic warming, with text and figures responding–without altering the basic layout so that understanding on one setting is transferable among multiple settings of these menus. Basic indicators of level and rate are in text and illustrated on figures, and updated with the pulldowns selected. The pulldown menu on right can select among using only subsets of lines of evidence rather than all methods combined, and descriptions of individual lines of evidence can be "explored" through the link at the bottom left: this leads the user to references and descriptions of each method.

6 Implications and interpretation in the context of the UNFCCC

In line with Article 31(1) of the Vienna Convention on the Law of Treaties (United Nations,1969), of which the Paris Agreement is one, treaties are to be interpreted *'[...] in good faith in accordance with the ordinary meaning to be given to the terms of the treaty in their context and in the light of its object and purpose*.' As noted earlier, the SED of the 2013-2015 Review provided the key science-policy context for the Paris Agreement. The preamble of the Paris Agreement talks of '*Recognizing the need for an effective and progressive response to the urgent threat of climate change on the basis of the best available scientific knowledge*'. In this context, all action under the UNFCCC process is based on the best available science - the latest research and observations from bodies and organisations as assessed by the Intergovernmental Panel on Climate Change (IPCC) and the World Meteorological Organisation (WMO)[39].

[39] Modified for contextual clarity from the original at https://unfccc.int/topics/science/the-big-picture/introduction-science; last accessed 9/12/24

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

The basis upon which individual countries believed the agreed text to have been made arguably differed (i.e. some of the ambiguity noted in Section 1 and assessed in prior sections may have helped make the LTTG acceptable to all Parties). There are two logical extremes for interpreting the text according to its 'ordinary meaning' in the absence of any clarifications (see later) with, undoubtedly, many other potential interpretations in between these:

1. **Interpreting in the context of evolving best available science:** Under this approach any update to the assessed temperature change since pre-industrial levels would necessarily lead to a commensurate change in our proximity to the LTTG. That is to say, if our estimate of change in global surface temperature from 1850-1900 to date increased by 0.1°C then we would be 0.1°C closer and if it decreased by 0.1°C we would be 0.1°C further away from the LTTG warming levels, which would be interpreted literally.
2. **Considering the best scientific understanding at the time of the PA**: Under this approach the interpretation of 1.5°C in the LTTG would be fixed at the level that corresponds to 1.5°C given the understanding of the surface temperature change when the Paris Agreement was agreed in 2015. In that case the *de facto* interpretation would be that it was to avoid an amount of additional warming beyond that which informed the LTTG negotiations e.g. the headline warming in IPCC AR5 up until either the 1986-2005 period used for many impact studies or the warming until around the early 2010s using either the OLS linear trend or trailing decadal average method.

The intention of Parties is only relevant if an ordinary meaning cannot be determined from the words and context. The LTTG does not stand in isolation and it is important to understand the legal and policy context within which the LTTG was agreed (e.g. Rogelj et al., 2017). Equally, an evolutionary method of interpretation is plausible and even desirable in contexts where the purpose would be thwarted if a static interpretation is imposed. It has been argued that the LTTG is not '1.5°C' or '2°C' but one goal with two distinct elements (Rajamani and Werksman, 2018). In this context the recent advisory opinion from the International Court of Justice is worth noting, in particular that it finds that '*The 1.5°C threshold of Article 2 is the parties' agreed primary temperature goal for limiting the global average temperature increase*'[40]. It is not possible, or appropriate, for us to do more than note the issues and challenges here.

The IPCC AR6 WGI assessment in Cross-chapter Box 2.3 of Gulev et al. (2021) *de facto* took the 'evolving best available science' approach described above, taking account of changes to global temperature estimates. But only partially so as it reflected the updated understanding of the observationally-based estimates since 1850-1900 but excluded the assessment of additional warming since 1750 carried out in Box 1.2 of Chen et al. (2021). These choices had knock-on implications for the assessments undertaken in WGII and WGIII. The choice to provide updated estimates of change since 1850-1900 was in part a practical choice in that all datasets used in AR5 had been superseded and thus an update to warming estimates based upon AR5 methods and sources was not possible. The changes in available observationally-based estimates had increased the estimated warming since the late 19th century by 0.08°C for the warming to 1986-2005 (Gulev et al., 2021, CCB2.3 therein).

[40] https://www.icj-cij.org/case/187 advisory opinion available at https://www.icj-cij.org/sites/default/files/case-related/187/187-20250723-adv-01-00-en.pdf

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

All else being equal, the new understanding of the warming to 1986-2005 would shift impacts that had been expected at 1.5°C in the AR5 assessment to then be expected at 1.58°C. However, the shift in our understanding of warming to date was overwhelmed by the advances in understanding of risk between IPCC AR5 and AR6 cycles by IPCC WGII as illustrated through the 5 Reasons for Concern (RfCs) (Figure 29). The risks were assessed to occur at lower levels of global warming in spite of the revised assessment of warming to date. Therefore if the LTTG was premised upon the global warming level at which unacceptable risks may occur then the updates since IPCC AR5 would, overall, point to the need for increased ambition and tighter LTTG limits under an approach of 'evolving best available science' (Table 6).

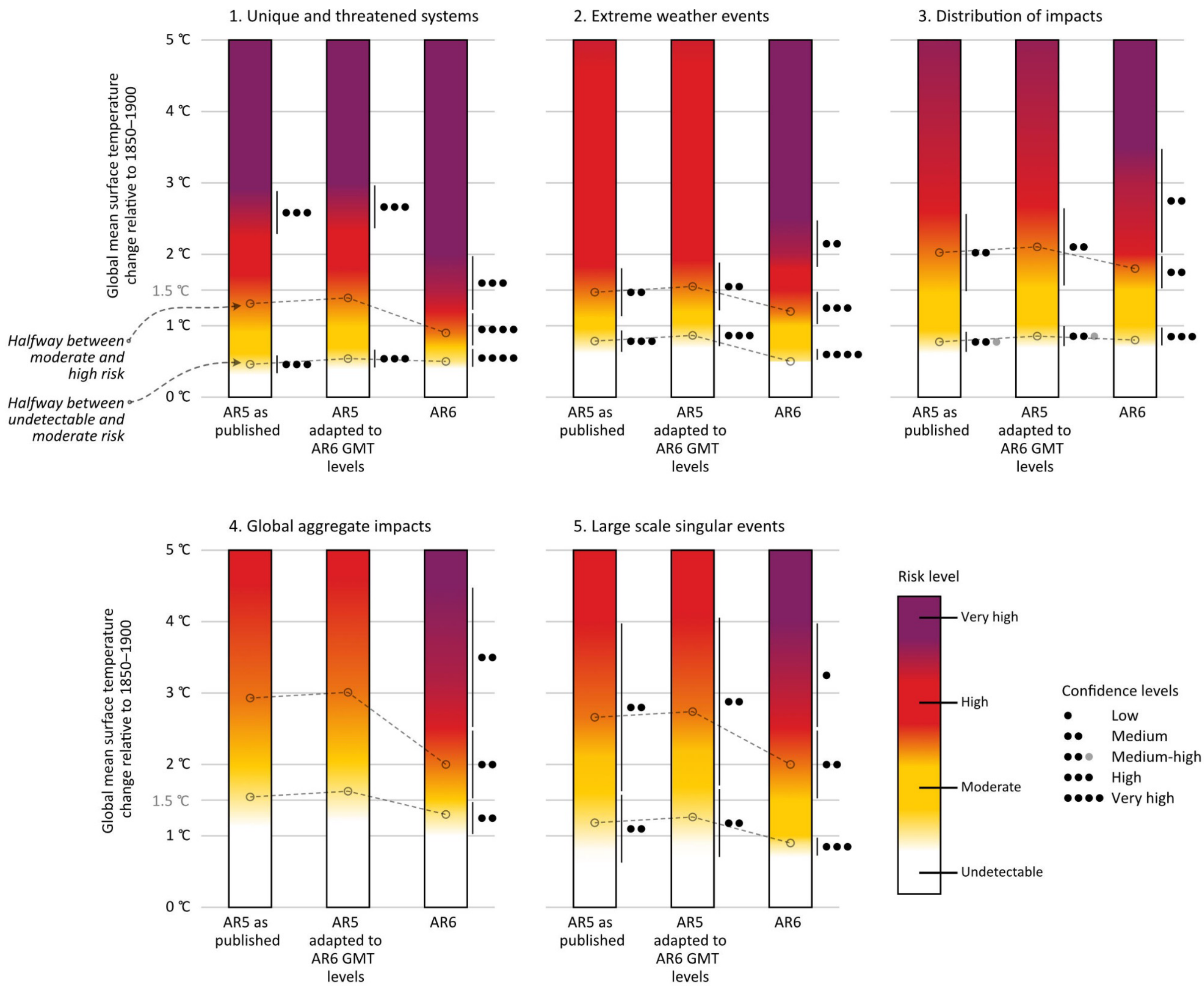


Figure 29. Comparison of the risks associated with the Reasons for Concern (RFC), illustrated as "burning embers" diagrams (Zommers et al. 2020, O'Neill et al., 2022), between IPCC AR5 and IPCC AR6 with an additional step reflecting the impact on the AR5 embers of the shift in understanding of global warming from 1850-1900 to the recent reference period used in AR5 (all risk changes are shifted by 0.08°C upwards, see text). The change in the assessed understanding of when a given risk could be faced is of the opposite sign and dominates over the effect of the change in understanding of long-term warming for most risk levels and across

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

all five RFCs. The AR5 and AR6 burning embers correspond to IPCC AR6 SYR Figure SPM.4 panel a (IPCC, 2023). Figure produced with https://climrisk.org/emberfactory using the excel file provided in Data and Materials.

	Risk level assessed for 1.5°C above pre-industrial in AR5	**Mean warming above pre-industrial that would lead to this level of risk in AR6**
RFC1 – Unique and threatened systems	Moderate to high (closer to high risk)	1.0°C
RFC2 – Extreme weather events	Moderate to high (middle of the risk transition)	1.2°C
RFC3 – Distribution of impacts	Moderate	1.5°C
RFC4 – Global aggregate impacts	Undetectable to moderate (middle of the transition)	1.3°C
RFC5 – Large scale singular events	Undetectable to moderate (almost moderate)	1.0°C

Table 6: Global mean warming above pre-industrial, as assessed in AR6, that would result in the level of risk that was assessed for 1.5°C in AR5. For RFC3, there is no significant change. All other risk estimates now correspond to lower warming above pre-industrial, even with the revised baseline – which was taken into account in WGII risk estimates. Calculations are linear interpolations based on data in O'Neill et al., 2022 and Marbaix et al., 2025.

Parties have the ability through consensus-based decisions under the Paris Agreement to resolve or substantively reduce the ambiguities noted in sections 2 (pre-industrial), 3 (SST/MAT and datasets to include) and 4 (metrics) and the issues raised in the shift from AR5 to AR6 understanding. They could offer further guidance in the form of an explicit clarification on how the PA LTTG is to be interpreted through a Conference of Parties decision. Although Conference of Parties decisions are not, save in a few exceptional situations, legally binding, they nevertheless offer authoritative guidance on the interpretation and application of treaty provisions. Article 31(3) of the Vienna Convention on the Law of Treaties explicitly recognizes the salience of *'subsequent agreement between the parties regarding the interpretation of the treaty or the application of its provisions'* for interpreting a treaty provision. Indeed it is as a result of interpreting COP decisions as reflecting such subsequent agreement between Parties that the ICJ read Article 2 of the Paris Agreement as containing the 'agreed primary temperature goal' of 1.5°C.

There is precedent for such clarificatory guidance on Paris Agreement Articles being offered through subsequent Conference of Parties decisions. For example, Article 4.1 was subsequently clarified at COP-24 (UNFCCC, 2019 (Decision 18/CMA.1, annex, paragraph 37)) to specify how national aggregate greenhouse gas emissions should be reported and thus how Article

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

4.1 was to be interpreted and its adherence or otherwise by Parties to the Paris Agreement should be determined. To clarify the assessment of progress to meet the LTTG we would suggest consideration by UNFCCC of the following aspects covered in detail in our analysis which would greatly clarify matters from a scientific perspective:

1. Clarify whether adherence should be tracked using an approach of best evolving scientific understanding or frozen understanding. If best evolving understanding then consider how new scientific insights are to be handled if they alter our collective understanding of warming to date such as, but not limited to, availability of new or revised datasets. This would be minimised if a recent period reference were to be used (point 3).
2. Clarify whether the adherence should be against realised warming or anthropogenic warming (or both) and whether it should be against GMST, GSAT or both. This should have regard to availability of different data sources and methods and maturity of their understanding.
3. Provide a specific definition of pre-industrial for surface temperature change monitoring that balances policy relevance and scientific limitations. The obvious candidate based upon scientific precedence would be the global surface temperature over 1850-1900 (see Figure 26). A further option which would considerably reduce uncertainty would be to agree that for verification purposes the LTTG actually corresponds to a smaller amount of additional warming relative to a more recent, better observed, period[41] such as 1981-2010 (see Figure 27). This alternative approach would reduce the uncertainty for verification purposes approximately 5-fold.

Preceding sections, and in particular Section 5, have provided the state-of-the-art scientific evidence and a proposed set of potential pathways that have attempted to gather broad community consensus. If adopted, alongside a robust governance mechanism to periodically reassess the approach based upon new scientific knowledge (Section 7), these would ultimately serve to reduce scientific uncertainty in making an assessment of progress in meeting the LTTG.

7 Discussion and conclusions

This community paper has comprehensively reviewed all aspects of determining the current level of global warming and hence adherence to the Paris Agreement LTTG, with a focus upon the proximate 1.5°C warming level. We have reviewed all known sources of uncertainties and shown that it is scientifically possible to robustly estimate the current total warming and anthropogenic warming relative to 1850-1900. The approaches identified were found to be robust across a range of possible future climate trajectories, some of which included potential future explosive volcanic events. The approaches have been fully documented herein and in the associated Supplement and code and data are made available without restriction to enable replication and further development.

We propose that two metrics should be consistently calculated and communicated when discussing long-term (for example, the 20-year centred-mean) as opposed to annual warming estimates: the realised warming (arising from all sources including internal variability) and the anthropogenic warming (arising solely from human perturbations). Realised warming and anthropogenic warming are both policy-relevant in terms of understanding impacts and associated adaptation needs, and

[41] This would make warming from pre-industrial to the more recent period moot for verification processes but would risk a disconnect between the headline value of warming since 1850-1900 and the LTTG verification process so is not without issue.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

understanding what portion of the underlying long-term warming is due to our historical and ongoing emissions in the context of mitigation. At any point in time natural external forcings and / or internal variability can cause the two numbers to diverge by amounts that would have practical policy implications. Currently these metrics through 2024 (based on the PIVW combination of 1850 to 2024 records) stand at 1.40 [1.23-1.58] and 1.34 [1.18-1.50]°C respectively, above 1850-1900 average temperatures where the range given is a 95% confidence range. It is *unlikely* that long-term realised warming of 1.5°C has been reached and *very unlikely* that long-term anthropogenic warming of 1.5°C has been reached.

The year 2024 was the first year in which long-term warming of 1.5°C was no longer certain not to have been reached and we note that 0.5°C and 1°C took under 2 decades to transition from being certain not to have been reached to being certain to have been exceeded under rapidly increasing anthropogenic GHG emissions. In both cases the retrospective 20-year mean passed the threshold within a decade of first being detected as being *extremely unlikely*. This underscores the urgency of immediate, deep and sustained emissions reductions as noted in IPCC AR6 (IPCC, 2023) if the LTTG is to remain within reach.

Our knowledge of changes in historical warming has evolved substantially over the last decade or so as the LTTG has spurred renewed interest in improving our monitoring of surface temperature changes (see Section 3). These insights have led to increases in the estimates of warming since 1850-1900 to the recent past that are the result of improved scientific understanding of the historical record, and not a result of additional warming that has occurred subsequent to the PA LTTG agreement. It is very likely that future work will further modify estimates of warming to date, with both the direction and magnitude of any future adjustments unclear. We have therefore also considered aspects of legal and policy interpretation of the LTTG which are necessary for a holistic assessment. We have proposed potential pathways which UNFCCC could adopt to clean-up definitional ambiguities in the LTTG text which would aid scientific monitoring efforts and minimise uncertainties in tracking adherence while allowing for inevitable future revisions to our estimates of long-term warming.

Issues around communication to key stakeholders, including the general public, have also been considered. A first key step is to distinguish individual daily, monthly or even annual exceedances from sustained long-term exceedance. We have developed an approach that could be used as a basis for consistent communication on where we are today based upon the tools and approaches outlined in Section 5. There are also broader aspects around incorrectly treating these warming levels as thresholds, around countdown to "doomsday" and other communications concerns that were beyond the remit of this analysis. We would stress, as outlined in communications around the IPCC Special Report on 1.5°C (IPCC, 2018), that every bit of warming matters.

Robust scientific information is vital to ensure timely tracking of this contentious policy issue. Without a timely consensus-based approach, by the time we've agreed across multiple competing methods and datasets that 1.5°C has been passed, it will

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

be far in the rear view mirror, given the current warming trend of 0.3 [0.29 - 0.35] °C/decade.[42] A consensus based approach could also help to reduce uncertainty in related key metrics such as the calculation of remaining carbon budgets. As such, it is important to stress that tracking the current warming level better has myriad potential scientific and policy benefits beyond simply knowing how close we are to the LTTG.

More philosophically, an open question remains of what use policymakers and the general public might make of the information that 1.5°C has been crossed. This could encourage efforts in adaptation and mitigation, could also promote defeatism or have little effect. The recent ICJ legal opinion further outlines possible legal implications by making clear that 1.5°C is the reference target, that states have obligations in the Paris Agreement and beyond, including a stringent, objective and demanding duty of due diligence, and that breach of these obligations could give rise to legal consequences, including a potential duty to make repatriations. In addition, there is a major question whether and under what conditions it might be possible to bring global warming back to 1.5°C once this level has been crossed, i.e. whether the crossing of 1.5°C is essentially irreversible over the next several human generations, or whether determined policy actions could ensure that any exceedance of 1.5°C is limited in both magnitude and duration (also referred to as 'overshoot' by IPCC (2023); see also Schleussner et al 2024, Reisinger et al 2025). These questions are beyond the scope of our analysis and proposal, but they provide important context for how information about exceeding or returning to 1.5°C might be interpreted and linked to actions by policymakers and civil society. Any eventual decline in temperatures could be monitored and tracked by the same methods and techniques used in this paper, with only slight modification of some probability statements (integrating the method-provided probability densities below a reference target rather than, as we do, above it).

The paper serves to highlight the criticality of international cooperation and sustained support for research and technical developments. The analysis furthermore highlights that there are a number of practical scientific steps that could be taken in the short-term to better quantify, and potentially reduce, the uncertainty in estimation of present-day warming since the pre-industrial period. These include, but are not limited to (subsections with extra detail are given in parentheses):

- Reducing the uncertainty in estimates of early-industrial period warming via advanced methods that build on historical, proxy and early instrumental records to continue to improve our understanding of warming prior to 1850-1900 (2);
- Better understanding of the theoretical basis for differences between SST and MAT changes and their expression in models and observational records, regionally and globally (3.1);
- Substantially increasing the number of land and, in particular, marine (SST, MAT and sea-ice) datasets using methodologically distinct approaches to better sample the structural uncertainty inherent in global surface temperature estimates. In particular increasing the number of GSAT datasets and ensuring the regular reprocessing of existing datasets to capture the latest scientific understanding (3.2.5; 5.1);
- Ensuring the continuing operation of key observing capabilities necessary to monitor long-term climate changes (3);

[42] Based on the PIVW realized warming combination of records from 1850 through 2024.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

- Operationalisation of the WMO unified data policy - specifically countries sharing their historical digital data holdings with recognised global repositories without restriction. Alongside full implementation of the WMO Global Basic Observing Network on a sustained basis supported by the Systematic Observations Funding Facility to fill persistent data gaps moving forwards (3.2.6);
- Data rescue from numerous national and regional holdings of early meteorological records still in paper or image format only and ensuring these rescued data and associated metadata reach global data repositories expeditiously (3.2.6);
- Delivering annual forcing estimates including SLCFs leading to updating of CMIP historical forcing runs on an annual basis to support the production of a human-induced warming metric (4.4);
- The ability to quickly update estimates following the occurrence of a major volcanic eruption which would require access to emissions and forcing estimates and reinitialisation of decadal predictions (4.4, 4.6).

Finally, the current study and resulting proposal is a snapshot in time assessment and will, eventually, inevitably be superseded. There is hence a need for sustained long-term support and governance which allows for both operational support of the approaches developed herein and periodically revisiting and updating all aspects of the current study. For example, in Section 4 we omitted the rapidly emerging area of machine learning methods. The kind of pattern recognition and signal identification tasks being asked of the statistical and dynamical systems in Section 4 are similar to a variety of machine learning methods presently in development and prototyped in climate applications (Bône et al., 2023, 2024, Diffenbaugh and Barnes, 2023). Their current exclusion should not be interpreted as an assessment that these tools do not have promise in the future. Rather, they were deemed currently too experimental in nature. Once more mature we would expect such approaches to be included if they were proven to add value. Similarly, key assumptions such as use of GMST observed series (Section 3) or 1850-1900 as an approximation to pre-industrial (Section 2) might require revisiting in future. WMO is likely best placed to provide such continuous governance and support, whilst recognising the important role of both advances in scientific understanding both through the peer reviewed literature and the periodic IPCC assessment products in informing any decisions.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Open Access Earth System Science Data Discussions

Code and data availability

All model data has been downloaded from ESGF (https://esgf.github.io and federated servers, e.g., https://esgf-data1.llnl.gov). A slimmed-down version containing just global averages for the specific ensemble members chosen here (which can be used in conjunction with the code below) is at: https://github.com/jnickla1/climate_data (Thorne et al., 2025). A repository of the observational data as used here, including the ensemble members chosen, etc., can be downloaded from https://github.com/jjk-code-otter/global_temperature_merge. The code used can be found at:
https://github.com/jnickla1/Thorne_15 (also includes results)
https://github.com/tristramwalsh/global-warming-index
https://github.com/jjk-code-otter/global_temperature_merge

Supplement link

The link to the supplement will be included by Copernicus, if applicable.

Author contributions

PWT conceptualised the paper, drew up an initial structure and convened regular meetings of the author team leading the overall drafting. JN led the production of Section 4.6 analysis with input from BFK, processed climate data, collated existing code and prepared all new code, ensured its archival and provided all figures and tables therein except Figure 19 which was produced by AS. JN also produced Figures 14, 16, 26 and 27, Tables 1-5 as well as various Figures in the Supplement. JN and TW jointly produced Figure 15. BHS and TW ran computations for, and CS provided support for specific methods. JJK and BC lead the production of the unified surface temperature change approach in Section 5.1. AP, JF and JJK aided the initial structure and facilitation of the process. JJK in addition provided Figures 7, 8 and 9. EH provided Figures 1 and 12. NA provided Figure 2. AA and SN provided Figure 3. KC provided Figure 4. RR provided Figure 5. BC provided Figures 6 and 11 and Table 1. NL provided Figure 10. AS provided Figure 13. PM provided Figure 29 and Table 6. DO provided access to the MPI SMILE ensemble and Figure S3. IB and SO provided access to the NorESM volcanic ensemble. VMD provided conceptual framing suggestions that underpinned much of Section 4.6. All authors contributed actively to drafting and review of the manuscript at various stages with leadership in drafting particular sections at different drafting cycle points from variously: NA, EH, AS, JJK, TW, BFK, BC, LR, MTR, ECK, VMD, AP and JN. EM provided support on all aspects of manuscript finalisation.

Competing interests

At least one of the (co-)authors is a member of the editorial board of Earth System Science Data. The authors declare that they have no other competing interests.

Disclaimer

Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher. All authors contributed in a personal capacity and the results presented should be interpreted as a personal and not an institutional view.

Acknowledgements

Thanks to Leon Hermanson for providing the WMO decadal forecasts

Financial support

Peter Thorne and Ed Hawkins were supported by Co-Centre award number 22/CC/11103. The Co-Centre award is managed by Research Ireland (RI), Northern Ireland's Department of Agriculture, Environment and Rural Affairs (DAERA) and UK Research and Innovation (UKRI) and supported via the UK's International Science Partnerships Fund (ISPF) and the Irish Government's Shared Island initiative. Baylor Fox-Kemper and John Nicklas were supported by NSF OIA-2316271 and Brown University seed funding via the Equitable Climate Futures initiative. Elizabeth Kent was supported by the UKRI Natural Environment Research Centre (NERC) National Capability Program AtlantiS (grant reference NE/Y005589/1). Tim Osborn and Emily Wallis were supported by NERC (NE/S015582/1). Stephen Outten was supported by the Nansen Center basic funding, through the Research Council of Norway (grant 342624). Gavin Schmidt is supported by the NASA Earth Science program. Mark T. Richardson's part of the research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004).

Review statement

The review statement will be added by Copernicus Publications listing the handling editor as well as all contributing referees according to their status anonymous or identified.

References

Abram, N. J., McGregor, H. V., Tierney, J. E., Evans, M. N., McKay, N. P., Kaufman, D. S., and PAGES 2k Consortium: Early onset of industrial-era warming across the oceans and continents, *Nature*, 536, 411–418, https://doi.org/10.1038/nature19082, 2016.

ACSYS: Arctic Climate System Study Historical Ice Chart Archive, 1553–2002, IACPO Informal Rep. 8, Arctic Climate System Study, Tromsø, Norway, 2003.

Allan, R., Brohan, P., Compo, G. P., Stone, R., Luterbacher, J., and Brönnimann, S.: The international Atmospheric Circulation Reconstructions over the Earth (ACRE) initiative, *Bull. Am. Meteorol. Soc.* **92**, 1421–1425, https://doi.org/10.1175/2011BAMS3218.1, 2011.

Allen, M. R. and Stott, P. A.: Estimating signal amplitudes in optimal fingerprinting, Part I: theory, *Clim. Dyn.* **21**, 477–491, https://doi.org/10.1007/s00382-003-0313-9, 2003.

Allen, M. R., Dube, O. P., Solecki, W., Aragón-Durand, F., Cramer, W., Humphreys, S., Kainuma, M., Kala, J., Mahowald, N., Mulugetta, Y., Perez, R., Wairiu, M., and Zickfeld, K.: Framing and Context, in: Global Warming of 1.5 °C. An IPCC Special Report on the impacts of global warming of 1.5 °C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty, edited by: Masson-Delmotte, V., Zhai, P., Pörtner, H.-O., Roberts, D., Skea, J., Shukla, P. R., Pirani, A., Moufouma-Okia, W., Péan, C., Pidcock, R., Connors, S., Matthews, J. B. R., Chen, Y., Zhou, X., Gomis, M. I., Lonnoy, E., Maycock, T., Tignor, M., and Waterfield, T., Cambridge University Press, Cambridge, UK and New York, NY, USA, pp. 49–92, https://doi.org/10.1017/9781009157940.003, 2018.

Ashford, O. M.: A new bucket for measurement of sea surface temperature, *Q. J. R. Meteorol. Soc.* 74, 99–104, https://doi.org/10.1002/qj.49707431916, 1948.

Ashton, T. S.: The Industrial Revolution 1760–1830, Oxford University Press, Oxford, UK, 162 pp., 1997.

Ballinger, A.P., Schurer, A., Hegerl, G., Dittus, A., Hawkins, E., Cornes, R., Kent, E., Marshall, L., Morice, C., Osborn, T., Rayner, N. and Rumbold, S. Importance of beginning industrial-era climate simulations in the eighteenth century. *Environ. Res. Lett.* https://doi.org/10.1088/1748-9326/ae1bbc (in press).

Bennedsen, M., Hillebrand, E., and Lykke, J. Z.: Global temperature projections from a statistical energy balance model using multiple sources of historical data, J. Climate, 36, 6817–6838, https://doi.org/10.1175/JCLI-D-22-0620.1, 2023.

Bethke, I., Outten, S., Otterå, O., et al.: Potential volcanic impacts on future climate variability, Nat. Clim. Change, 7, 799–805, https://doi.org/10.1038/nclimate3394, 2017.

Betts, R. A., Belcher, S.E., Hermanson, L., Tank, A.K., Lower, J.A., Jones, C.D., Morice, C.P., Rayner, N.A., Rayner, N.A., Scaife, A.A. and Stott, P.A..: Approaching 1.5 °C: how will we know we've reached this crucial warming mark?, Nature, 624, 33–35, https://doi.org/10.1038/d41586-023-03775-z, 2023.

Beusch, L., Gudmundsson, L., and Seneviratne, S. I.: Crossbreeding CMIP6 Earth System Models with an Emulator for Regionally Optimized Land Temperature Projections, Geophys. Res. Lett., 47, e2019GL086812, https://doi.org/10.1029/2019GL086812, 2020.

Bevacqua, E., Schleussner, C.-F., and Zscheisler, J.: A year above 1.5 °C signals that Earth is most probably within the 20-year period that will reach the Paris Agreement limit, Nat. Clim. Change, https://doi.org/10.1038/s41558-025-02246-9, 2025.

Bilbao, R., Ortega, P., Swingedouw, D., Hermanson, L., Athanasiadis, P., Eade, R., Devilliers, M., Doblas-Reyes, F., Dunstone, N., Ho, A.-C., Merryfield, W., Mignot, J., Nicoli, D., Samsó, M., Sospedra-Alfonso, R., Wu, X., and Yeager, S.: Impact of volcanic eruptions on CMIP6 decadal predictions: a multi-model analysis, Earth Syst. Dynam., 15, 501–525, https://doi.org/10.5194/esd-15-501-2024, 2024.

BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML: Evaluation of measurement data – Guide to the expression of uncertainty in measurement, Joint Committee for Guides in Metrology, JCGM 100:2008, https://www.bipm.org/documents/20126/2071204/JCGM_100_2008_E.pdf/cb0ef43f-baa5-11cf-3f85-4dcd86f77bd6, 2008.

Bonacci, O. and Željković, I.: Differences between true mean temperatures and means calculated with four different approaches: a case study from three Croatian stations, Theor. Appl. Climatol., 131, 733–743, https://doi.org/10.1007/s00704-016-1993-5, 2018.

Bône, C., Gastineau, G., Thiria, S., Gallinari, P., and Mejia, C.: Detection and attribution of climate change using a neural network, J. Adv. Model. Earth Syst., 15, e2022MS003475, https://doi.org/10.1029/2022MS003475, 2023.

Bône, C., Gastineau, G., Thiria, S., Gallinari, P., and Mejia, C.: Separation of internal and forced variability of climate using a U-Net, J. Adv. Model. Earth Syst., 16, e2023MS003964, https://doi.org/10.1029/2023MS003964, 2024.

Brennan, M. K., Hakim, G. J., and Blanchard-Wrigglesworth, E.: Arctic sea-ice variability during the instrumental era, Geophys. Res. Lett., 47, e2019GL086843, https://doi.org/10.1029/2019GL086843, 2020.

Brennan, M. K. and Hakim, G. J.: Reconstructing Arctic sea ice over the Common Era using data assimilation, J. Climate, 35, 1231–1247, https://doi.org/10.1175/JCLI-D-21-0099.1, 2022.

Brönnimann, S., Franke, J., Nussbaumer, S. U., Zumbühl, H. J., Steiner, D., Trachsel, M., Hegerl, G. C., Schurer, A., Worni, M., Malik, A., Flückiger, J., and Raible, C. C.: Last phase of the Little Ice Age forced by volcanic eruptions, Nat. Geosci., 12, 650–656, https://doi.org/10.1038/s41561-019-0402-y, 2019a.

Brönnimann, S., and Coauthors: Unlocking pre-1850 instrumental meteorological records: a global inventory, Bull. Am. Meteorol. Soc., 100, ES389–ES413, https://doi.org/10.1175/BAMS-D-19-0040.1, 2019b.

Brugnara, Y., Good, E., Squintu, A. A., van der Schrier, G., and Brönnimann, S.: The EUSTACE global land station daily air temperature dataset, Geosci. Data J., 6, 189–204, https://doi.org/10.1002/gdj3.81, 2019.

Brunet, M., Asin, J., Sigro, J., Banon, M., Garcia, F., Aguilar, E., Palenzuela, J., Peterson, T. C., and Jones, P. D.: The minimization of the screen bias from ancient Western Mediterranean air temperature records: an exploratory statistical analysis, Int. J. Climatol., 31, 1879–1895, https://doi.org/10.1002/joc.2192, 2011.

Budescu, D. V., Por, H.-H., Broomell, S. B., and Smithson, M.: The interpretation of IPCC probabilistic statements around the world, Nat. Clim. Change, 4, 508–512, https://doi.org/10.1038/nclimate2194, 2014.

Businger, J. A., Wyngaard, J. C., Izumi, Y., and Bradley, E. F.: Flux-profile relationships in the atmospheric surface layer, J. Atmos. Sci., 28, 181–189,https://doi.org/10.1175/1520-0469(1971)028%3C0181:FPRITA%3E2.0.CO;2, 1971.

Cahill, N., Rahmstorf, S., and Parnell, A. C.: Change points of global temperature, Environ. Res. Lett., 10, 084002, https://doi.org/10.1088/1748-9326/10/8/084002, 2015.

Callendar, G.S.: The artificial production of carbon dioxide and its influence on temperature. Q.J.R. Meteorol. Soc., 64: 223-240. https://doi.org/10.1002/qj.49706427503, 1938.

Calvert, B. T. T.: Improving global temperature datasets to better account for non-uniform warming, Q. J. R. Meteorol. Soc., 150, 3672–3702, https://doi.org/10.1002/qj.4791, 2024a.

Calvert, B. T. T.: Maximum likelihood estimates of temperatures using data from the Dynamically Consistent Ensemble of Temperature (Version 1.0), World Data Center for Climate (WDCC) at DKRZ, Hamburg, Germany, https://doi.org/10.26050/WDCC/DCENT_MLE_v1_0, 2024b.

Camuffo, D.: Calibration and instrumental errors in early measurements of air temperature, Clim. Change, 53, 297–329, https://doi.org/10.1023/A:1014914707832, 2002.

Canadell, J. G., Monteiro, P. M. S., Costa, M. H., Cotrim da Cunha, L., Cox, P. M., Eliseev, A. V., Henson, S., Ishii, M., Jaccard, S., Koven, C., Lohila, A., Patra, P. K., Piao, S., Rogelj, J., Syampungani, S., Zaehle, S., and Zickfeld, K.: Global carbon and other biogeochemical cycles and feedbacks, in: Climate Change 2021: The Physical Science Basis, Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 673–816, https://doi.org/10.1017/9781009157896.007, 2021.

Cannon, A. J.: Twelve months at 1.5 °C signals earlier than expected breach of Paris Agreement threshold, Nat. Clim. Change, https://doi.org/10.1038/s41558-025-02247-8, 2025.

Carella, G., Kent, E. C., and Berry, D. I.: A probabilistic approach to ship voyage reconstruction in ICOADS, Int. J. Climatol., 37, 2233–2247, https://doi.org/10.1002/joc.4838, 2017a.

Carella, G., Morris, A. K. R., Pascal, R. W., Yelland, M. J., Berry, D. I., Morak-Bozzo, S., Merchant, C. J., and Kent, E. C.: Measurements and models of the temperature change of water samples in sea-surface temperature buckets, Q. J. R. Meteorol. Soc., 143, 2198–2209, https://doi.org/10.1002/qj.3078, 2017b.

Carella, G., Kennedy, J. J., Berry, D. I., Hirahara, S., Merchant, C. J., Morak-Bozzo, S., and Kent, E. C.: Estimating sea surface temperature measurement methods using characteristic differences in the diurnal cycle, Geophys. Res. Lett., 45, 363–371, https://doi.org/10.1002/2017GL076475, 2018.

Chan, D. and Huybers, P.: Systematic differences in bucket sea surface temperature measurements among nations identified using a linear-mixed-effect method, J. Climate, 32, 2569–2589, https://doi.org/10.1175/JCLI-D-18-0562.1, 2019.

Chan, D., Kent, E. C., Berry, D. I. and Huybers, P..: Correcting datasets leads to more homogeneous early-twentieth-century sea surface warming, Nature, 571, 393–397, https://doi.org/10.1038/s41586-019-1349-2, 2019.

Chan, D. and Huybers, P.: Systematic differences in bucket sea surface temperatures caused by misclassification of engine room intake measurements, J. Climate, 33, 7735–7753, https://doi.org/10.1175/JCLI-D-19-0972.1, 2020.

Chan, D. and Huybers, P.: Correcting observational biases in sea surface temperature observations removes anomalous warmth during World War II, J. Climate, 34, 4585–4602, https://doi.org/10.1175/JCLI-D-20-0907.1, 2021.

Chan, D., Gebbie, G., and Huybers, P.: Global and regional discrepancies between early-twentieth-century coastal air and sea surface temperature detected by a coupled energy-balance analysis, J. Climate, 36, 2205–2220, https://doi.org/10.1175/JCLI-D-22-0569.1, 2023.

Chan, D., Gebbie, G., and Huybers, P.: An improved ensemble of land-surface air temperatures since 1880 using revised pair-wise homogenization algorithms accounting for autocorrelation, J. Climate, 37, 2325–2345, https://doi.org/10.1175/JCLI-D-23-0338.1, 2024a.

Chan, D., Gebbie, G., Huybers, P., and Kent, E. C.: A dynamically consistent ensemble of temperature at the Earth surface since 1850 from the DCENT dataset, Sci. Data, 11, 953, https://doi.org/10.1038/s41597-024-03742-x, 2024b.

Chan, D., Chan, S.C., Siddons, J.T., Cable, A., Faulkner, A., Cornes, R.C., Kent, E.C., Gebbie, G. & Huybers, P. (2025). DCENT-I: A Globally Infilled Extension of the Dynamically Consistent ENsemble of Temperature Dataset. *Geoscience Data Journal*, https://eartharxiv.org/repository/view/10091/, accepted.

Chen, D., Rojas, M, Samset, B.H., Cobb, K., Diongue Niang, A., Edwards, P., Emori,S., S.H. Faria, S.H., Hawkins,E., Hope, P. Huybrechts, P., Meinshausen, M., Mustafa,S.K.,Plattner, G.K., and Tréguier, A.M.: Framing, Context, and Methods. InClimate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change[Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 147–286, doi:10.1017/9781009157896.003, 2021.

Chen, L., Cao, L., Zhou, Z., Zhang, D., and Liao, J.: A new globally reconstructed sea surface temperature analysis dataset since 1900, J. Meteorol. Res., 35, 911–925, https://doi.org/10.1007/s13351-021-1098-7, 2022.

Chen, L., Xu, W., Zhou, Z., Cao, L., Yang, S., and Xu, C.: A new global land–ocean merged surface temperature dataset since the 1850s: the CMA-GMST dataset, Clim. Dyn., 63, 187, https://doi.org/10.1007/s00382-025-07614-x, 2025.

Cheng, L., Pan, Y., Tan, Z., Zheng, H., Zhu, Y., Wei, W., Du, J., Yuan, H., Li, G., Ye, H., Gouretski, V., Li, Y., Trenberth, K., Abraham, J., Jin, Y., Reseghetti, F., Lin, X., Zhang, B., Chen, G., Mann, M. and Zhu, J. IAPv4 ocean temperature and ocean heat content gridded dataset. *Earth System Science Data*, 16(8), 3517-3546. https://doi.org/10.5194/essd-16-3517-2024, 2024.

Chandler, R. and Scott, M.: Statistical methods for trend detection and analysis in the environmental sciences, Wiley, 400 pp., 2011.

Chen, X. and Tung, K. K.: Global-mean surface temperature variability: space–time perspective from rotated EOFs, Clim. Dyn., 51, 1719–1732, https://doi.org/10.1007/s00382-017-3979-0, 2018.

Chen, D., Rojas, M., Samset, B. H., Cobb, K., Diongue Niang, A., Edwards, P., Emori, S., Faria, S. H., Hawkins, E., Hope, P., Huybrechts, P., Meinshausen, M., Mustafa, S. K., Plattner, G.-K., and Tréguier, A.-M.: Framing, context, and methods, in: Climate Change 2021: The Physical Science Basis, Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 147–286, https://doi.org/10.1017/9781009157896.003, 2021.

Chim, M. M., Aubry, T. J., Abraham, N. L., Marshall, L., Mulcahy, J., Walton, J., and Schmidt, A.: Climate projections very likely underestimate future volcanic forcing and its climatic effects, Geophys. Res. Lett., 50, e2023GL103743, https://doi.org/10.1029/2023GL103743, 2023.

Chim, M.M., Aubry, T.J., Smith, C. *et al.* Neglecting future sporadic volcanic eruptions underestimates climate uncertainty. *Commun Earth Environ* 6, 236, https://doi.org/10.1038/s43247-025-02208-1, 2025.

Christy, J. R., Parker, D. E., Brown, S. J., Macadam, I., Stendel, M., and Norris, W. B.: Differential trends in tropical sea surface and atmospheric temperatures since 1979, Geophys. Res. Lett., 28, 183–186, https://doi.org/10.1029/2000GL011167, 2001.

Clarke, D. C. and Richardson, M.: The benefits of continuous local regression for quantifying global warming, Earth Space Sci., 8, e2020EA001082, https://doi.org/10.1029/2020EA001082, 2021.

Cochran, W. G.: The Combination of Estimates from Different Experiments. Biometrics, 10(1), 101–129. https://doi.org/10.2307/3001666, 1954.

Cointe, B., Ravon, P.-A., and Guerin, E.: 2 °C: The history of a policy-science nexus, Working Papers N°19/11, IDDRI, Paris, France, 28 pp., https://www.iddri.org/sites/default/files/import/publications/wp-1911_bc-par-eg_2-degrees.pdf, 2011.

Cointe, B. and Guillemot, H.: A history of the 1.5 °C target, WIREs Clim. Change, 14, e824, https://doi.org/10.1002/wcc.824, 2023.

Compo, G. P., Sardeshmukh, P. D., Whitaker, J. S., Brohan, P., Jones, P. D., and McColl, C.: Independent confirmation of global land warming without the use of station temperatures, Geophys. Res. Lett., 40, 3170–3174, https://doi.org/10.1002/grl.50425, 2013.

Cooper, V. T., Hakim, G.J. and Armour, K.C.: Monthly sea-surface temperature, sea ice, and sea level pressure over 1850–2023 from coupled data assimilation, J. Climate, 38 (19), 5461–5490, https://doi.org/10.1175/JCLI-D-25-0021.1, 2025.

Cornes, R., van der Schrier, G., van den Besselaar, E. J. M., and Jones, P. D.: An ensemble version of the E-OBS temperature and precipitation datasets, J. Geophys. Res.-Atmos., 123, https://doi.org/10.1029/2017JD028200, 2018.

Cornes, R. C., Kent, E. C., Berry, D. I., and Kennedy, J. J.: CLASSnmat: A global night marine air temperature data set, 1880–2019, Geosci. Data J., 7, 170–184, https://doi.org/10.1002/gdj3.100, 2020.

Cotté, C. and Guinet, C.: Historical whaling records reveal major regional retreat of Antarctic sea ice, Deep-Sea Res. Pt. I, 54, 243–252, https://doi.org/10.1016/j.dsr.2006.11.001, 2007.

Cowtan, K. and 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, https://doi.org/10.1002/qj.2297, 2014.

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, https://doi.org/10.1002/2015GL064888, 2015.

Cowtan, K., Rohde, R., and Hausfather, Z.: Evaluating biases in sea surface temperature records using coastal weather stations, Q. J. R. Meteorol. Soc., 144, 670–681, https://doi.org/10.1002/qj.3245, 2018.

Craigmile, P. F. and Guttorp, P.: A combined estimate of global temperature, Environmetrics, 33, e2706, https://doi.org/10.1002/env.2706, 2021.

Cropper, T. E., Berry, D. I., Cornes, R. C., and Kent, E. C.: Quantifying daytime heating biases in marine air temperature observations from ships, J. Atmos. Ocean. Technol., 40, 427–438, https://doi.org/10.1175/JTECH-D-22-0080.1, 2023.

Cummins, D. P., Stephenson, D. B., and Stott, P. A.: A new energy-balance approach to linear filtering for estimating effective radiative forcing from temperature time series, Adv. Stat. Clim. Meteorol. Oceanogr., 6, 91–102, https://doi.org/10.5194/ascmo-6-91-2020, 2020.

Dai, C., Chan, D., Huybers, P., and Pillai, N.: Late 19th-century navigational uncertainties and their influence on sea surface temperature estimates, *Ann. Appl. Stat.*,15, 22–40, doi:10.1214/20-AOAS1367, 2021.

Dalaiden, Q., Rezsöhazy, J., Goosse, H., Thomas, E. R., Vladimirova, D. O., and Tetzner, D.: An unprecedented sea ice retreat in the Weddell Sea driving an overall decrease of the Antarctic sea-ice extent over the 20th century, *Geophys. Res. Lett.*, 50 (21), https://doi.org/10.1029/2023GL104666, 2023.

Dalaiden, Q., Goosse, H., Holland, P.R. and Barthelemy, A. Dynamical reconstruction of Southern Ocean and Antarctic climate variability since 1700. *Sci Data* 12, 1574 (2025). https://doi.org/10.1038/s41597-025-05808-w

de la Mare, W. K.: Changes in Antarctic sea-ice extent from direct historical observations and whaling records, *Clim. Change*, 92, 461–493, https://doi.org/10.1007/s10584-008-9473-2, 2009.

de Rosnay, P., Browne, P., de Boisséson, E., Fairbairn, D., Hirahara, Y., Ochi, K., et al.: Coupled data assimilation at ECMWF: current status, challenges and future developments, *Q. J. R. Meteorol. Soc.*, 148 (747), 2672–2702, https://doi.org/10.1002/qj.4330, 2022.

de Valk, C., and Brandsma, T.: Homogenization of daily temperatures using covariates and statistical learning – the case of parallel measurements, *Int. J. Climatol.*, 43, 7170–7182,https://doi.org/10.1002/joc.8258, 2023.

Diffenbaugh, N. S., and Barnes, E. A.: Data-driven predictions of the time remaining until critical global warming thresholds are reached, *Proc. Natl. Acad. Sci. USA*, 120, e2207183120, doi:10.1073/pnas.2207183120, 2023.

Domingues, C.M., Church, J., White, N., Gleckler, P., Wijfels, S., Barker, P. and Dunn, J.: Improved estimates of upper-ocean warming and multi-decadal sea-level rise. *Nature*, 453(7198), 1090-1093. https://doi.org/10.1038/nature07080, 2008.

Durbin, J., and Koopman, S. J.: *Time series analysis by state space methods*, 2nd Edn., Oxford University Press, 364 pp., 2012.

Edinburgh, T., and Day, J. J.: Estimating the extent of Antarctic summer sea ice during the Heroic Age of Antarctic Exploration, *Cryosphere*, 10 (6), 2721–2730, https://doi.org/10.5194/tc-10-2721-2016, 2016.

Ellis, E. C.: Land use and ecological change: a 12 000-year history, *Annu. Rev. Environ. Resour.*, 46, 1–33, https://doi.org/10.1146/annurev-environ-012220-010822,, 2021.

Esper, J., et al.: 2023 summer warmth unparalleled over the past 2 000 years, *Nature*, https://doi.org/10.1038/s41586-024-07512-y , 2024.

Eyring, V., Gillett, N. P., Achutarao, K., Barimalala, R., Barreiro Parrillo, M., Bellouin, N., Cassou, C., Durack, P., Kosaka, Y., McGregor, S., Min, S.-K., Morgenstern, O., and Sun, Y.: Human influence on the climate system, in: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the IPCC*, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 423–552, doi:10.1017/9781009157896.005, 2021.

European Centre for Medium-Range Weather Forecasts (ECMWF): *ERA5: uncertainty estimation*. [Online] Available at:https://confluence.ecmwf.int/display/CKB/ERA5%3A+uncertainty+estimation (last access: 29 June 2025), 2024.

Fakoor, R., Kim, T., Mueller, J., Smola, A.J. and Tibshirani.R.J.: Flexible model aggregation for quantile regression. J. Mach. Learn. Res. 24, 1, Article 162, 2023.

Fang, S.-W., Sigl, M., Toohey, M., Jungclaus, J., Zanchettin, D., and Timmreck, C.: The role of small to moderate volcanic eruptions in the early 19th century climate, *Geophys. Res. Lett.*, 50, https://doi.org/10.1029/2023GL105307, 2023.

Fetterer, F., Knowles, K., Meier, W. N., Savoie, M., and Windnagel, A. K.: *Sea Ice Index, Version 3*, National Snow and Ice Data Center (NSIDC), Boulder, CO, USA, https://doi.org/10.7265/N5K072F8, 2017.

Fogt, R. L., Sleinkofer, A. M., Raphael, M. N., and Handcock, M. S.: A regime shift in seasonal total Antarctic sea ice extent in the twentieth century, *Nat. Clim. Change*, 12, 54–62, https://doi.org/10.1038/s41558-021-01254-9,2022.

Folland, C. K., Parker, D. E., and Kates, F. E.: Worldwide marine temperature fluctuations 1856–1981, *Nature*, 310, 670–673, doi:10.1038/310670a0, 1984.

Folland, C. K., and Parker, D. E.: Correction of instrumental biases in historical sea surface temperature data, *Q. J. R. Meteorol. Soc.*, 121, 319–367, 1995.

Foster, G., and Rahmstorf, S.: Global temperature evolution 1979–2010, *Environ. Res. Lett.*, 6, 044022, doi:10.1088/1748-9326/6/4/044022, 2011.

Forster, P., Storelvmo, T., Armour, K., Collins, W., Dufresne, J.-L., Frame, D., Lunt, D. J., Mauritsen, T., Palmer, M. D., Watanabe, M., Wild, M., and Zhang, H.: The Earth's energy budget, climate feedbacks, and climate sensitivity, in: *Climate*

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the IPCC, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 923–1054, doi:10.1017/9781009157896.009, 2021.

Forster, P. M., Smith, C. J., Walsh, T., Lamb, W. F., Lamboll, R., Hauser, M., Ribes, A., Rosen, D., Gillett, N., Palmer, M. D., Rogelj, J., von Schuckmann, K., Seneviratne, S. I., Trewin, B., Zhang, X., Allen, M., Andrew, R., Birt, A., Borger, A., Boyer, T., Broersma, J. A., Cheng, L., Dentener, F., Friedlingstein, P., Gutiérrez, J. M., Gütschow, J., Hall, B., Ishii, M., Jenkins, S., Lan, X., Lee, J.-Y., Morice, C., Kadow, C., Kennedy, J., Killick, R., Minx, J. C., Naik, V., Peters, G. P., Pirani, A., Pongratz, J., Schleussner, C.-F., Szopa, S., Thorne, P., Rohde, R., Rojas Corradi, M., Schumacher, D., Vose, R., Zickfeld, K., Masson-Delmotte, V., and Zhai, P.: Indicators of Global Climate Change 2022: annual update of large-scale indicators of the state of the climate system and human influence, *Earth Syst. Sci. Data*, 15, 2295–2327, https://doi.org/10.5194/essd-15-2295-2023, 2023.

Forster, P. M., Smith, C., Walsh, T., Lamb, W. F., Lamboll, R., Hall, B., Hauser, M., Ribes, A., Rosen, D., Gillett, N. P., Palmer, M. D., Rogelj, J., von Schuckmann, K., Trewin, B., Allen, M., Andrew, R., Betts, R. A., Borger, A., Boyer, T., Broersma, J. A., Buontempo, C., Burgess, S., Cagnazzo, C., Cheng, L., Friedlingstein, P., Gettelman, A., Gütschow, J., Ishii, M., Jenkins, S., Lan, X., Morice, C., Mühle, J., Kadow, C., Kennedy, J., Killick, R. E., Krummel, P. B., Minx, J. C., Myhre, G., Naik, V., Peters, G. P., Pirani, A., Pongratz, J., Schleussner, C.-F., Seneviratne, S. I., Szopa, S., Thorne, P., Kovilakam, M. V. M., Majamäki, E., Jalkanen, J.-P., van Marle, M., Hoesly, R. M., Rohde, R., Schumacher, D., van der Werf, G., Vose, R., Zickfeld, K., Zhang, X., Masson-Delmotte, V., and Zhai, P. 2024: Indicators of Global Climate Change 2023: annual update of key indicators of the state of the climate system and human influence, Earth Syst. Sci. Data, 16, 2625–2658, https://doi.org/10.5194/essd-16-2625-2024, 2024.

Forster, P. M., Smith, C., Walsh, T., Lamb, W. F., Lamboll, R., Cassou, C., Hauser, M., Hausfather, Z., Lee, J.-Y., Palmer, M. D., von Schuckmann, K., Slangen, A. B. A., Szopa, S., Trewin, B., Yun, J., Gillett, N. P., Jenkins, S., Matthews, H. D., Raghavan, K., Ribes, A., Rogelj, J., Rosen, D., Zhang, X., Allen, M., Aleluia Reis, L., Andrew, R. M., Betts, R. A., Borger, A., Broersma, J. A., Burgess, S. N., Cheng, L., Friedlingstein, P., Domingues, C. M., Gambarini, M., Gasser, T., Gütschow, J., Ishii, M., Kadow, C., Kennedy, J., Killick, R. E., Krummel, P. B., Liné, A., Monselesan, D. P., Morice, C., Mühle, J., Naik, V., Peters, G. P., Pirani, A., Pongratz, J., Minx, J. C., Rigby, M., Rohde, R., Savita, A., Seneviratne, S. I., Thorne, P., Wells, C., Western, L. M., van der Werf, G. R., Wijffels, S. E., Masson-Delmotte, V., and Zhai, P. (2025). Indicators of Global Climate Change 2024: annual update of key indicators of the state of the climate system and human influence, Earth System Science Data, https://doi.org/10.5194/essd-2025-250, 2025.

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Bakker, D. C. E., Hauck, J., Landschützer, P., Le Quéré, C., Luijkx, I. T., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Anthoni, P., Barbero, L., Bates, N. R., Becker, M., Bellouin, N., Decharme, B., Bopp, L., Brasika, I. B. M., Cadule, P., Chamberlain, M. A., Chandra, N., Chau, T.-T.-T., Chevallier, F., Chini, L. P., Cronin, M., Dou, X., Enyo, K., Evans, W., Falk, S., Feely, R. A., Feng, L., Ford, D. J., Gasser, T., Ghattas, J., Gkritzalis, T., Grassi, G., Gregor, L., Gruber, N., Gürses, O., Harris, I., Hefner, M., Heinke, J., Houghton, R. A., Hurtt, G. C., Iida, Y., Ilyina, T., Jacobson, A. R., Jain, A., Jarníková, T., Jersild, A., Jiang, F., Jin, Z., Joos, F., Kato, E., Keeling, R. F., Kennedy, D., Klein Goldewijk, K., Knauer, J., Korsbakken, J. I., Körtzinger, A., Lan, X., Lefèvre, N., Li, H., Liu, J., Liu, Z., Ma, L., Marland, G., Mayot, N., McGuire, P. C., McKinley, G. A., Meyer, G., Morgan, E. J., Munro, D. R., Nakaoka, S.-I., Niwa, Y., O'Brien, K. M., Olsen, A., Omar, A. M., Ono, T., Paulsen, M., Pierrot, D., Pocock, K., Poulter, B., Powis, C. M., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Rosan, T. M., Schwinger, J., Séférian, R., Smallman, T. L., Smith, S. M., Sospedra-Alfonso, R., Sun, Q., Sutton, A. J., Sweeney, C., Takao, S., Tans, P. P., Tian, H., Tilbrook, B., Tsujino, H., Tubiello, F., van der Werf, G. R., van Ooijen, E., Wanninkhof, R., Watanabe, M., WimartRousseau, C., Yang, D., Yang, X., Yuan, W., Yue, X., Zaehle, S., Zeng, J., and Zheng, B.: Global Carbon Budget 2023, Earth System Science Data, 15, 5301–5369, https://doi.org/10.5194/essd-15-5301-2023, https://essd.copernicus.org/articles/15/5301/2023/, 2023.

Funke, B., Dudok de Wit, T., Ermolli, I., Haberreiter, M., Kinnison, D., Marsh, D., Nesse, H., Seppälä, A., Sinnhuber, M., and Usoskin, I.: Towards the definition of a solar forcing dataset for CMIP7, Geosci. Model Dev., 17, 1217–1227, https://doi.org/10.5194/gmd-17-1217-2024, 2024.

Gallaher, D. W., Campbell, G. G., and Meier, W. N.: Anomalous variability in Antarctic sea ice extents during the 1960s with the use of Nimbus data, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 7, 881–887, https://doi.org/10.1109/JSTARS.2013.2264391, 2014.

Geoffroy, O., Saint-Martin, D., Bellon, G., Voldoire, A., Olivié, D. J. L., and Tytéca, S.: Transient climate response in a two-layer energy-balance model. Part II: Representation of the efficacy of deep-ocean heat uptake and validation for CMIP5 AOGCMs, J. Climate, 26, 1859–1876, https://doi.org/10.1175/JCLI-D-12-00196.1, 2013.

Gergis, J., Baillie, Z., Ashcroft, L.,Trewin, B and Allan, R.J.: Consolidating historical instrumental observations in southern Australia for assessing pre-industrial weather and climate variability, Clim. Dynam., 61, 1063–1087, https://doi.org/10.1007/s00382-022-06573-x, 2023.

Gillespie, I., Haimberger, L., Compo, G and Thorne, P.W..: Assessing homogeneity of land surface air temperature observations using sparse-input reanalyses, Int. J. Climatol., https://doi.org/10.1002/joc.7822, 2022.

Gillett, N. P., Shiogama, H., Funke, B., Hegerl, G., Knutti, R., Matthes, K., Santer, B. D., Stone, D., and Tebaldi, C.: The Detection and Attribution Model Intercomparison Project (DAMIP v1.0) contribution to CMIP6, Geosci. Model Dev., 9, 3685–3697, https://doi.org/10.5194/gmd-9-3685-2016, 2016.

Gillett, N. P., Kirchmeier-Young, M., Ribes, A., Shiogama, H., Hegerl, G. C., Knutti, R., Gastineau, G., John, J. G., Li, L., Nazarenko, L., Rosenbloom, N., Seland, Ø., Wu, T., Yukimoto, S., and Ziehn, T.: Constraining human contributions to observed warming since the pre-industrial period, Nature Climate Change, 11, 207–212, https://doi.org/10.1038/s41558-020-00965-9, 2021.

Glaser, Z. I., Richardson, M. T., and Landerer, F. W.: Large-ensemble Monte Carlo: a researcher's guide to better climate trend uncertainties, Environ. Res.: Climate, 3, 045007, https://doi.org/10.1088/2752-5295/ad69b6, 2024.

Gneiting, T. and Ranjan, R.: "Combining predictive distributions." Electron. J. Statist. 7 1747 - 1782, https://doi.org/10.1214/13-EJS823, 2013.

Gneiting, T., Balabdaoui, F. and Raftery, A.E.: Probabilistic Forecasts, Calibration and Sharpness, *Journal of the Royal Statistical Society Series B: Statistical Methodology*, Volume 69, Issue 2, Pages 243–268, https://doi.org/10.1111/j.1467-9868.2007.00587.x, 2007.

Good, S.A., Martin, M.J. & Rayner, N.A.:. EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, 118(12), 6704-6716. https://doi.org/10.1002/2013JC009067, 2013.

Goosse, H., Dalaiden, Q., Feba, F., Mezzina, B., and Fogt, R. L.: A drop in Antarctic sea ice extent at the end of the 1970s, Commun. Earth Environ., 5, 628, https://doi.org/10.1038/s43247-024-01793-x, 2024.

Gough, W. A., Žaknić-Ćatović, A., and Zajch, A.: Sampling frequency of climate data for the determination of daily temperature and daily temperature extrema, Int. J. Climatol., 40, 5451–5463, https://doi.org/10.1002/joc.6528, 2020.

Gregory, J.: Vertical heat transports in the ocean and their effect on time-dependent climate change, Clim. Dynam., 16, 501–515, https://doi.org/10.1007/s003820000059, 2000.

Guardian (The): UN climate change: hottest week in the world, The Guardian, 7 July 2023, available at: https://www.theguardian.com/environment/2023/jul/07/un-climate-change-hottest-week-world (last access: 27 July 2024).

Guardian (The): World temperature records shattered: hottest day in climate crisis, The Guardian, 23 July 2024, available at: https://www.theguardian.com/environment/article/2024/jul/23/world-temperature-records-shattered-hottest-day-climate-crisis (last access: 27 July 2024).

Gulev, S.K., P.W. Thorne, J. Ahn, F.J. Dentener, C.M. Domingues, S. Gerland, D. Gong, D.S. Kaufman, H.C. Nnamchi, J. Quaas, J.A. Rivera, S. Sathyendranath, S.L. Smith, B. Trewin, K. von Schuckmann, and R.S. Vose, 2021: Changing State of the Climate System. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change[Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 287–422, doi:10.1017/9781009157896.004.

Gütschow, J., Jeffery, M. L., Gieseke, R., Gebel, R., Stevens, D., Krapp, M., and Rocha, M.: The PRIMAP-hist national historical emissions time series, *Earth Syst. Sci. Data*, 8, 571–603, https://doi.org/10.5194/essd-8-571-2016, 2016.

Hansen, J., Ruedy, R., Sato, M., and Reynolds, R.: Global surface air temperature in 1995: Return to pre-Pinatubo level, Geophys. Res. Lett., 23, 1665–1668, https://doi.org/10.1029/96GL01040, 1996.

Hartmann, D. L., Klein Tank, A. M. G., Rusticucci, M., Alexander, L. V., Brönnimann, S., Charabi, Y., Dentener, F. J., Dlugokencky, E. J., Easterling, D. R., Kaplan, A., Soden, B. J., Thorne, P. W., Wild, M., and Zhai, P.: Observations:

Atmosphere and surface, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

Hasselmann, K. Multi-pattern fingerprint method for detection and attribution of climate change. *Climate Dynamics* 13, 601–611, https://doi.org/10.1007/s003820050185, 1997.

Haustein, K., Otto, F. E. L., Uhe, P., Schaller, N., Allen, M. R., Hermanson, L., Christidis, N., McLean, P., Gordon, M., and Perkins-Kirkpatrick, S. E.: A real-time global warming index, Sci. Rep., 7, 15417, https://doi.org/10.1038/s41598-017-14828-5, 2017.

Hawkins, E. and Sutton, R.: The potential to narrow uncertainty in regional climate predictions, Bull. Am. Meteorol. Soc., 90, 1095–1108, https://doi.org/10.1175/2009BAMS2607.1, 2009.

Hawkins, E., Ortega, P., Suckling, E. B., Schurer, A. P., Hegerl, G. C., Jones, P. D., Joshi, M., Osborn, T. J., Masson-Delmotte, V., Mignot, J., Thorne, P., and van Oldenborgh, G. J.: Estimating changes in global temperature since the preindustrial period, Bull. Am. Meteorol. Soc., 98, 1841–1856, https://doi.org/10.1175/BAMS-D-16-0007.1, 2017.

Hawkins, E., Burt, S., McCarthy, M., Murphy, C., Ross, C., Baldock, M., Hannaford, J., and Thorne, P. W.: Millions of historical monthly rainfall observations taken in the UK and Ireland rescued by citizen scientists, Geosci. Data J., 10, 246–261, https://doi.org/10.1002/gdj3.157, 2023.

Hegerl, G. C., Brönnimann, S., Schurer, A., and Cowan, T.: The early 20th century warming: Anomalies, causes, and consequences, Wiley Interdiscip. Rev. Clim. Change, 9, e522, https://doi.org/10.1002/wcc.522, 2018.

Held, I. M., Winton, M., Takahashi, K., Delworth, T., Zeng, F., and Vallis, G. K.: Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing, J. Climate, 23, 2418–2427, https://doi.org/10.1175/2009JCLI3466.1, 2010.

Henley, B. J., McGregor, H. V., King, A. D., Hoegh-Guldberg, O., Arzey, A. K., Karoly, D. J., Lough, J. M., DeCarlo, T. M., and Linsley, B. K.: Highest ocean heat in four centuries places Great Barrier Reef in danger, Nature, 632, 320–326, https://doi.org/10.1038/s41586-024-07672-x, 2024.

Hermanson, L., Smith, D., Seabrook, M., Bilbao, R., Doblas-Reyes, F. J., Tourigny, E., Lapin, V., Kharin, V. V., Merryfield, W. J., Sospedra-Alfonso, R., Athanasiadis, P., Nicoli, D., Gualdi, S., Dunstone, N., Eade, R., Scaife, A., Collier, M., O'Kane, T., Kitsios, V., Sandery, P., Pankatz, K., Früh, B., Pohlmann, H., Müller, W., Kataoka, T., Tatebe, H., Ishii, M., Imada, Y., Kruschke, T., Koenigk, T., Karami, M. P., Yang, S., Tian, T., Zhang, L., Delworth, T., Yang, X., Zeng, F., Wang, Y., Counillon, F., Keenlyside, N., Bethke, I., Lean, J., Luterbacher, J., Kolli, R. K., and Kumar, A.: WMO global annual to decadal climate update: A prediction for 2021–25, Bull. Am. Meteorol. Soc., 103, E1117–E1129, https://doi.org/10.1175/BAMS-D-20-0311.1, 2022.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Q. J. R. Meteorol. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hewitt, H. T., Flato, G., O'Rourke, E., Dunne, J. P., Adloff, F., Arblaster, J., Bonou, F., Boucher, O., Cavazos, T., Dingley, B., Durack, P. J., Fox-Kemper, B., Ilyina, T., Miyakawa, T., Mizielinski, M., Naik, V., Nicholls, Z., Pincus, R., Taylor, K. E., Tegtmeier, S., and Wedi, N.: Towards provision of regularly updated climate data from the Coupled Model Intercomparison Project, *PLoS Clim.*, submitted 2025.

Hill, B.: Historical record of sea ice and iceberg distribution around Newfoundland and Labrador, in: Proceedings of the 18th International Conference on Offshore Mechanics and Arctic Engineering, ASME, St. John's, Canada, 1999.

Hirahara, S., Ishii, M., and Fukuda, Y.: Centennial-scale sea surface temperature analysis and its uncertainty, J. Climate, 27, 57–75, https://doi.org/10.1175/JCLI-D-12-00837.1, 2014.

Hirahara, S., M. A. Balmaseda, M. A, de Boisseson, E, Hersbach, H.: Sea surface temperature and sea ice concentration for ERA5, ERA Report Series, 26, European Centre for Medium-Range Weather Forecasts, Reading, UK, available at: https://www.ecmwf.int/en/elibrary/79624-sea-surface-temperature-and-sea-ice-concentration-era5 (last access: 22 August 2025), 2016.

Ho, E. H. and Budescu, D. V.: Climate uncertainty communication, Nat. Clim. Change, 9, 802–808, https://doi.org/10.1038/s41558-019-0606-6, 2019.

Hoesly, R., Smith, S., Prime, N., Ahsan, H., and Suchyta, H.: Global anthropogenic emissions inventory of reactive gases and aerosols (1750–2022): An update to the Community Emissions Data System (CEDS), in: AGU Fall Meeting 2023, available at: https://agu.confex.com/agu/fm23/meetingapp.cgi/Paper/1315334 (last access: 22 August 2025), 2023.

Hoeting, J.A., Madigan, D., Raftery, A.E. and Volinsky, C.T.: Bayesian Model Averaging: A Tutorial. Statistical Science. Vol. 14, No. 4, pp. 382-401, 1999.

Höge, M.; Guthke, A.; Nowak, W. : Bayesian Model Weighting: The Many Faces of Model Averaging. Water, 12, 309. https://doi.org/10.3390/w12020309, 2020.

Huang, B., Banzon, V., Freeman, E., Lawrimore, J., Liu, W., Peterson, T., Smith, T., Thorne, P., Woodruff, S., Zhang, H. Extended Reconstructed Sea Surface Temperature Version 4 (ERSST.v4). Part I: Upgrades and Intercomparisons. J. Climate, 28, 911-930, https://doi.org/10.1175/JCLI-D-14-00006.1, 2015.

Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., Menne, M. J., Smith, T. M., Vose, R. S., and Zhang, H.-M.: Extended reconstructed sea surface temperature, version 5 (ERSSTv5): Upgrades, validations, and intercomparisons, J. Climate, 30, 8179–8205, https://doi.org/10.1175/JCLI-D-16-0836.1, 2017.

Huang, B., Menne, M. J., Boyer, T., Freeman, E., Gleason, B. E., Lawrimore, J. H., Liu, C., Rennie, J. J., Schreck, C., Sun, F., Vose, R. S., Williams, C. N., Yin, X., and Zhang, H.-M.: Uncertainty estimates for sea surface temperature and land surface air temperature in NOAAGlobalTemp version 5, J. Climate, 33, 1351–1379, https://doi.org/10.1175/JCLI-D-19-0395.1, 2020.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Huang, B., Yin, X., Boyer, T., Liu, C., Menne, M., Rao, Y. D., Smith, T., Vose, R., and Zhang, H.-M.: Extended Reconstructed Sea Surface Temperature Version 6 (ERSSTv6): Part I. An Artificial Neural Network Approach, J. Climate, 38, 1105–1121, https://doi.org/10.1175/JCLI-D-23-0707.1, 2025a.

Huang, B., Yin, X., Boyer, T., Liu, C., Menne, M., Rao, Y. D., Smith, T., Vose, R., and Zhang, H.-M.: Extended Reconstructed Sea Surface Temperature Version 6 (ERSSTv6): Part II. Upgrades on quality control and large-scale filter, J. Climate, 38, 1123–1136, https://doi.org/10.1175/JCLI-D-24-0185.1, 2025b.

Ilyas, M., Nychka, D., Brierley, C., and Guillas, S.: Global ensemble of temperatures over 1850–2018: Quantification of uncertainties in observations, coverage, and spatial modeling (GETQUOCS), Atmos. Meas. Tech., 14, 7103–7121, https://doi.org/10.5194/amt-14-7103-2021, 2021.

IPCC: Summary for policymakers, in: Climate Change 2001: Impacts, Adaptation and Vulnerability, Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change, edited by: McCarthy, J. J., Canziani, O. F., Leary, N. A., Dokken, D. J., and White, K. S., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, ISBN 0-521-80768-9, 2001.

IPCC: Annex 1: Glossary, in: Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, available at: https://www.ipcc.ch/site/assets/uploads/2018/02/ar4-wg1-annexes-1.pdf (last access: 2 April 2024), 2007a.

IPCC: Summary for policymakers, in: Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007b.

IPCC: Managing the risks of extreme events and disasters to advance climate change adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change, edited by: Field, C. B., Barros, V., Stocker, T. F., Qin, D., Dokken, D. J., Ebi, K. L., Mastrandrea, M. D., Mach, K. J., Plattner, G.-K., Allen, S. K., Tignor, M., and Midgley, P. M., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 582 pp., 2012.

IPCC: Climate Change 2013: The Physical Science Basis, Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp., 2013.

IPCC: Summary for policymakers, in: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects, Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Field, C. B., Barros, V. R., Dokken, D. J., Mach, K. J., Mastrandrea, M. D., Bilir, T. E., Chatterjee, M., Ebi, K. L., Estrada, Y. O., Genova, R. C., Girma, B., Kissel, E. S., Levy, A. N., MacCracken, S., Mastrandrea, P. R., and White, L. L., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1–32, 2014.

IPCC: Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty, edited by: Masson-Delmotte, V., Zhai, P., Pörtner, H.-O., Roberts, D., Skea, J., Shukla, P. R., Pirani, A., Moufouma-Okia, W., Péan, C., Pidcock, R., Connors, S., Matthews, J. B. R., Chen, Y., Zhou, X., Gomis, M. I., Lonnoy, E., Maycock, T., Tignor, M., and Waterfield, T., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 616 pp., https://doi.org/10.1017/9781009157940, 2018.

IPCC: Climate Change 2021: The Physical Science Basis, Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/9781009157896, 2021a.

IPCC: Annex IV: Modes of variability, in: Climate Change 2021: The Physical Science Basis, Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2153–2192, https://doi.org/10.1017/9781009157896.018, 2021b.

IPCC: Annex VII: Glossary, in: Climate Change 2021: The Physical Science Basis, Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2215–2256, https://doi.org/10.1017/9781009157896.022, 2021c.

IPCC: Climate Change 2023: Synthesis Report, Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Core Writing Team, Lee, H., and Romero, J., IPCC, Geneva, Switzerland, 184 pp., https://doi.org/10.59327/IPCC/AR6-9789291691647, 2023.

Ishii, M., Fukuda, Y., Hirahara, H., Yasui, S., Suzuki, T. & Sato, K. (2017). Accuracy of Global Upper Ocean Heat Content Estimation Expected from Present Observational Data Sets. *SOLA*, 13, 163–167. https://doi.org/10.2151/sola.2017-030

Ishii, M., Nishimura, A., Yasui, S., and Hirahara, S.: Historical high-resolution daily SST analysis (COBE-SST3) with consistency to monthly land surface air temperature, J. Meteorol. Soc. Jpn., 103, 17-44, https://doi.org/10.2151/jmsj.2025-002, 2025.

Ivanova, N., Pedersen, L. T., Tonboe, R. T., Kern, S., Heygster, G., Lavergne, A., Sørensen, A., Saldo, R., Dybkjaer, G., Brucker, L., and Shokr, M.: Inter-comparison and evaluation of sea ice algorithms: Towards further identification of challenges

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

and optimal approach using passive microwave observations, The Cryosphere, 9, 1797–1817, https://doi.org/10.5194/tc-9-1797-2015, 2015.

Jarvis, A. and Forster, P. M.: Estimated human-induced warming from a linear temperature and atmospheric CO_2 relationship, Nat. Geosci., 17, 1222–1224, https://doi.org/10.1038/s41561-024-01580-5, 2024.

Jenkins, S., Smith, C., Allen, M., and Grainger, R.: Tonga eruption increases chance of temporary surface temperature anomaly above 1.5 °C, Nat. Clim. Change, 13, 127–129, https://doi.org/10.1038/s41558-022-01568-2, 2023.

Joelsson, L. M. T., Sturm, C., Södling, J., Engström, E., and Kjellström, E.: Automation and evaluation of the interactive homogenization tool HOMER, Int. J. Climatol., 42, 2861–2880, https://doi.org/10.1002/joc.7394, 2022.

Jones, P. D., Wigley, T. M., and Wright, P. B.: Global temperature variations between 1861 and 1984, Nature, 322, 430–434, https://doi.org/10.1038/322430a0, 1986.

Jones, P. D., New, M., Parker, D.E., Martin, S. and Rigor, I.G.: Surface air temperature and its changes over the past 150 years, Rev. Geophys., 37(2), 173–199, doi:10.1029/1999RG900002, 1999.

Junod, R. A. and Christy, J. R.: A new compilation of globally gridded night-time marine air temperatures: The UAHNMATv1 dataset, Int. J. Climatol., 40, 2609–2623, https://doi.org/10.1002/joc.6354, 2020.

Kadow, C., Hall, D. M., and Ulbrich, U.: Artificial intelligence reconstructs missing climate information, Nat. Geosci., 13, 408–413, https://doi.org/10.1038/s41561-020-0582-5, 2020.

Kennedy, J. J., Rayner, N. A., Smith, R. O., Parker, D. E., and Saunby, M.: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization, J. Geophys. Res.,116, D14104, https://doi.org/10.1029/2010JD015220, 2011a.

Kennedy, J. J., Rayner, N. A., Smith, R. O., Parker, D. E., and Saunby, M.: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties, J. Geophys. Res., 116, https://doi.org/10.1029/2010JD015218, 2011b.

Kennedy, J. J.: A review of uncertainty in in situ measurements and data sets of sea surface temperature, Rev. Geophys., 52, 1–32, https://doi.org/10.1002/2013RG000434, 2014.

Kennedy, J. J., Rayner, N. A., Atkinson, C. P., and 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, https://doi.org/10.1029/2018JD029867, 2019.

Kent, E. C., Challenor, P. G., and Taylor, P. K.: A statistical determination of the random observational errors present in voluntary observing ships meteorological reports, J. Atmos. Oceanic Technol., 16, 905–914, https://doi.org/10.1175/1520-0426(1999)016<0905:ASDOTR>2.0.CO;2, 1999.

Kent, E. C., Rayner, N. A., Berry, D. I., Saunby, M., Moat, B. I., Kennedy, J. J., and Parker, D. E.: Global analysis of night marine air temperature and its uncertainty since 1880: The HadNMAT2 data set, J. Geophys. Res. Atmos., 118, 1281–1298, https://doi.org/10.1002/jgrd.50152, 2013.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Open Access Earth System Science Data Discussions

Kent, E. C., Kennedy, J. J., Smith, T. M., Hirahara, S., Huang, B., Kaplan, A., Parker, D. E., Atkinson, C. P., Berry, D. I., Carella, G., Fukuda, Y., Ishii, M., Jones, P. D., Lindgren, F., Merchant, C. J., Morak-Bozzo, S., Rayner, N. A., Venema, V., Yasui, S., and Zhang, H.-M.: A call for new approaches to quantifying biases in observations of sea surface temperature, Bull. Am. Meteorol. Soc., 98, 1601–1616, https://doi.org/10.1175/BAMS-D-15-00251.1, 2017

Kent, E. C. and Kennedy, J. J.: Historical estimates of surface marine temperatures, Annu. Rev. Mar. Sci., 13, 283–311, https://doi.org/10.1146/annurev-marine-042120-111807, 2021.

Kincer, J. B.: Is our climate changing? A study of long-time temperature trends, *Mon. Weather Rev.*, 61, 251–259, https://doi.org/10.1175/1520-0493(1933)61<251:IOCCAS>2.0.CO;2 , 1933.

King, A. C. F., Bauska, T. K., Brook, E. J., and Coauthors: Reconciling ice core CO_2 and land-use change following New World–Old World contact, Nat. Commun., 15, 1735, https://doi.org/10.1038/s41467-024-45894-9, 2024.

Koch, A., Brierley, C., Maslin, M. M., and Lewis, S. L.: Earth system impacts of the European arrival and Great Dying in the Americas after 1492, Quaternary Sci. Rev., 207, 13–36, https://doi.org/10.1016/j.quascirev.2018.12.004, 2019.

Kosaka, Y., Kobayashi, S., Harada, Y., Kobayashi, C., Naoe, H., Yoshimoto, K., Harada, M., Goto, N., Chiba, J., Miyaoka, K., Sekiguchi, R., Deushi, M., Kamahori, H., Nakaegawa, T., Tanaka, T. Y., Tokuhiro, T., Sata, Y., Matsushita, Y., and Onogi, K.: The JRA-3Q reanalysis, J. Meteorol. Soc. Jpn., 102, 49–109, https://doi.org/10.2151/jmsj.2024-004, 2024.

Kovilakam, M., Thomason, L. W., Ernest, N., Rieger, L., Bourassa, A., and Millán, L.: The Global Space-based Stratospheric Aerosol Climatology (version 2.0): 1979–2018, Earth Syst. Sci. Data, 12, 2607–2634, https://doi.org/10.5194/essd-12-2607-2020, 2020.

Krough, A. and Vedelsby, J. Neural network ensembles, cross validation and active learning. In Proceedings of the 8th International Conference on Neural Information Processing Systems (NIPS'94). MIT Press, Cambridge, MA, USA, 231–238. 1994.

Lakkis, S. G., Canziani, P. O., Rodriquez, J. O., and Yuchechen, A. E.: Early meteorological records from Corrientes and Bahía Blanca, Argentina: Initial ACRE-Argentina data rescue and related activities, Geosci. Data J., 10, 328–346, https://doi.org/10.1002/gdj3.176, 2023.

Le, T and Clarke,B. "A Bayes Interpretation of Stacking for M-Complete and M-Open Settings." Bayesian Anal. 12 (3) 807 - 829, https://doi.org/10.1214/16-BA1023, 2017

Leach, N. J., Jenkins, S., Nicholls, Z., Smith, C. J., Lynch, J., Cain, M., Walsh, T., Wu, B., Tsutsui, J., and Allen, M. R.: FaIRv2.0.0: a generalized impulse response model for climate uncertainty and future scenario exploration, Geosci. Model Dev., 14, 3007–3036, https://doi.org/10.5194/gmd-14-3007-2021, 2021.

Lee, C.H., Cook, S., Lee, J.S and Han, B.: Comparison of Two Meta-Analysis Methods: Inverse-Variance-Weighted Average and Weighted Sum of Z-Scores. Genomics Inform. 14(4):173-180. DOI: https://doi.org/10.5808/GI.2016.14.4.173, 2016.

Lee, J.-Y., Marotzke, J., Bala, G., Cao, L., Corti, S., Dunne, J. P., Engelbrecht, F., Fischer, E., Fyfe, J. C., Jones, C., Maycock, A., Mutemi, J., Ndiaye, O., Panickal, S., and Zhou, T.: Future global climate: Scenario-based projections and near-term

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

information, in: Climate Change 2021: The Physical Science Basis, Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 553–672, https://doi.org/10.1017/9781009157896.006, 2021.

Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E. M., Brunner, L., Knutti, R., and Hawkins, E.: Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6, Earth Syst. Dynam., 11, 491–508, https://doi.org/10.5194/esd-11-491-2020, 2020.

Lenssen, N. J., Schmidt, G. A., Hansen, J. E., Menne, M. J., Persin, A., Ruedy, R., and Zyss, D.: Improvements in the GISTEMP uncertainty model, J. Geophys. Res. Atmos., 124, 6307–6326, https://doi.org/10.1029/2018JD029522, 2019.

Lenssen, N., Schmidt, G., Hansen, J., Menne, M., Persin, A., Reudy, R and Zyss, D.: Improvements in the GISTEMP Uncertainty Model. Journal of Geophysical Research: Atmospheres, Vol 124, 12, pp. 6307-6326. https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2018JD029522, 2019.

Lenssen, N., Schmidt, G. A., Hendrickson, M., Jacobs, O., Menne, M. and Ruedy, R.: A NASA GISTEMPv4 observational uncertainty ensemble, ESS Open Archive, https://doi.org/10.22541/essoar.171893198.89278240/v1, 2024.

Levitus, S., Antonov, J. I., Boyer, T. P., Baranova, O. K., Garcia, H. E., Locarnini, R. A., Mishonov, A. V., Reagan, J. R., Seidov, D., Yarosh, E. S., and Zweng, M. M.: World ocean heat content and thermosteric sea level change (0–2000 m), 1955–2010 Geophys. Res. Lett., 39, L10603, https://doi.org/10.1029/2012GL051106, 2012.

Li, M., Lv, T., Chen, J., Cui, L., Lu, Y., Florencio, D., Zhang, C., Li, Z., and Wei, F.: TrOCR: Transformer-based optical character recognition with pre-trained models, Proc. AAAI Conf. Artif. Intell., 37, 13094–13102, https://doi.org/10.1609/aaai.v37i11.26538, 2023.

Li, Z., Li, Q., Jiao, B., *et al.*: A new uncertainty framework for the China-MST 3.0 global surface temperature dataset, J. Geophys. Res. Atmos., submitted, 2025.

Liu, W., Huang, B., Thorne, P., Bazon, V., Zhang, H., Freeman, E., Lawrimore, J., Peterson, T., Smith, T. and Woodruff, S. Extended Reconstructed Sea Surface Temperature Version 4 (ERSST.v4): Part II. Parametric and Structural Uncertainty Estimations, J. Climate, 28, 931-951, https://doi.org/10.1175/JCLI-D-14-00007.1, 2015.

Livezey, R. E., Vinnikov, K. Y., Timofeyeva, M. M., Tinker, R., and van den Dool, H. M.: Estimation and extrapolation of climate normals and climatic trends, J. Appl. Meteorol. Climatol., 46, 1759–1776, https://doi.org/10.1175/2007JAMC1666.1, 2007.

Livingston, J. E. and Rummukainen, M.: Taking science by surprise: The knowledge politics of the IPCC Special Report on 1.5 degrees, Environ. Sci. Policy, 112, 10–16, https://doi.org/10.1016/j.envsci.2020.05.020, 2020.

Lorrey, A. M.,Pearce, P.R., Allan, R., Wilkinson, C., Wolley, J., Judd, E., Mackay, S., Rawhat, S., Silvinski, L., Wilkinson, S., Hawkins, E., Quesnel, P. and Compo, G.P.: Meteorological data rescue: Citizen science lessons learned from Southern Weather Discovery, Patterns, 3 (6), https://doi.org/10.1016/j.patter.2022.100495, 2022.

Love, E. and Bigg, G. R.: Estimating summer sea ice extent in the Weddell Sea during the early 19th century, Clim. Past, 19, 1905–1917, https://doi.org/10.5194/cp-19-1905-2023, 2023.

MacDougall, A. H.: "The Transient Response to Cumulative CO_2 Emissions: a Review," Curr Clim Change Reports , vol. 2, no. 1, pp. 39–47, doi: 10.1007/s40641-015-0030-6, 2015.

MacDougall, A. H. and Friedlingstein P.: The Origin and Limits of the Near Proportionality between Climate Warming and Cumulative CO_2 EmissionsV J Climate, vol. 28, no. 10, pp. 4217–4230, doi: 10.1175/jcli-d-14-00036.1, 2015.

Maierhofer, T.J., Raphael, M.N., Fogt, R.L. & Handcock, M.S.:. A Bayesian Model for 20th Century Antarctic Sea Ice Extent Reconstruction. *Earth and Space Science*, 11(10), e2024EA003577. https://doi.org/10.1029/2024EA003577, 2024.

Malinen, M. I. and Fränti, P.: Balanced K-means for clustering, in: Structural, Syntactic, and Statistical Pattern Recognition, S+SSPR 2014, Lecture Notes in Computer Science, vol. 8621, edited by: Fränti, P., Brown, G., Loog, M., Escolano, F., and Pelillo, M., Springer, Berlin, Heidelberg, https://doi.org/10.1007/978-3-662-44415-3_4, 2014.

Mann, M. E.: On smoothing potentially non-stationary climate time series, Geophys. Res. Lett., 31, L07214, doi:10.1029/2004GL019569, 2004.

Mann, M. E.: Smoothing of climate time series revisited, Geophys. Res. Lett., 35, L16708, doi:10.1029/2008GL034716, 2008.

Marbaix, P., Magnan, A. K., Muccione, V., Thorne, P. W., and Zommers, Z.: Climate change risks illustrated by the IPCC "burning embers", Earth Syst. Sci. Data, 17, 317–327, https://doi.org/10.5194/essd-17-317-2025, 2025.

Martin, S., Long, D.G., and Schodlokm, M.P.: Comparison of Antarctic iceberg observations by Cook in 1772–75, Halley in 1700, Bouvet in 1739 and Riou in 1789 with modern data. J. Glaciol., 69 (276), 911-918. https://doi.org/10.1017/jog.2022.111, 2023.

Mastrandrea, M. D., Field, C. B., Stocker, T. F., Edenhofer, O., Ebi, K. L., Frame, D. J., Held, H., Kriegler, E., Mach, K. J., Matschoss, P. R., Plattner, G.-K., Yohe, G. W., and Zwiers, F. W.: Guidance note for lead authors of the IPCC Fifth Assessment Report on consistent treatment of uncertainties, Intergovernmental Panel on Climate Change (IPCC), available at: https://www.ipcc.ch/site/assets/uploads/2018/05/uncertainty-guidance-note.pdf, (last access: 22 August 2025), 2010

Matthews, H. D., Gillett, N. P., Stott, P. A., and Zickfeld, K.: "The proportionality of global warming to cumulative carbon emissions," Nature, Vol 459, no. 7248, p. 829,https://doi.org/10.1038/nature08047, 2009.

Mauritsen, T., & Roeckner, E. : Tuning the MPI-ESM1.2 global climate model to improve the match with instrumental record warming by lowering its climate sensitivity. J. Adv Model. Earth Syst., 12, e2019MS002037, https://doi.org/10.1029/2019MS002037, 2020.

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., Brovkin, V., Claussen, M., Crueger, T., Esch, M., Fast, I., Fiedler, S., Fläschner, D., Gayler, V., Giorgetta, M., Goll, D. S., Haak, H., Hagemann, S., Hedemann, C., Hohenegger, C., Ilyina, T., Jahns, T., Jiménez-de-la-Cuesta, D., Jungclaus, J. H., Kleinen, T., Kloster, S., Kracher, D., Kinne, S., Kleberg, D., Lasslop, G., Kornblueh, L., Marotzke, J., Matei, D., Meraner, K., Mikolajewicz, U., Modali, K., Möbis, B., Müller, W. A., Nabel, J. E. M. S., Nam, C. C. W., Notz, D., Nyawira, S.-S., Paulsen, H., Peters, K., Pincus, R., Pohlmann, H., Pongratz, J., Popp, M., Raddatz, T. J., Rast, S., Redler, R., Reick, C. H., Rohrschneider, T., Schemann, V., Schmidt, H., Schnur, R., Schulzweida, U., Six, K. D., Stein, L., Stemmler, I., Stevens, B., von Storch, J.-S., Tian, F., Voigt, A., Vrese, P., Wieners, K.-H., Wilkenskjeld, S., Winkler, A., and Roeckner, E.: Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and its response to increasing CO_2, *J. Adv. Model. Earth Syst.*, 11, 998–1038, https://doi.org/10.1029/2018MS001400, 2019.

McCulloch, M. T., Winter, A., Sherman, C. E., and Trotter, J. A. 300 years of sclerosponge thermometry shows global warming has exceeded 1.5 °C Nat. Clim. Chang., 14, 171–177, https://doi.org/10.1038/s41558-023-01919-7, 2024.

Medhaug, I., Stolpe, M., Fischer, E. and Knutti, R.: Reconciling controversies about the 'global warming hiatus'. Nature 545, 41–47. https://doi.org/10.1038/nature22315, 2017.

Meier, W.N., Fetterer, F., Windnagel, A.K., and Stewart, S.: NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration, Version 4. National Snow and Ice Data Center, Boulder, Colorado, U.S.A. https://doi.org/10.7265/efmz-2t65, 2021.

Menne, M. J., and Williams, C. N.: Homogenization of temperature series via pairwise comparisons. *Journal of Climate*, *22*(7), 1700-1717, https://journals.ametsoc.org/view/journals/clim/22/7/2008jcli2263.1.xml, 2009.

Menne, M. J., Williams, C. N., Gleason, B. E., Rennie, J. J., and Lawrimore, J. H.: The Global Historical Climatology Network monthly temperature dataset, version 4, J. Climate, 31, 9835–9854, https://doi.org/10.1175/JCLI-D-18-0094.1, 2018.

Merchant, C. J., Matthiesen, S., Rayner, N. A., Remedios, J. J., Jones, P. D., Olesen, F., Trewin, B., Thorne, P. W., Auchmann, R., Corlett, G. K., Guillevic, P. C., and Hulley, G. C.: The surface temperatures of Earth: steps towards integrated understanding of variability and change, Geosci. Instrum. Method. Data Syst., 2, 305–321, https://doi.org/10.5194/gi-2-305-2013, 2013

Morice, C. P., Kennedy, J.J., Rayner, N.A., Winn, J.P., Hogan, E., Killick, R.E., Dunnd, R.J.H., Osborn, T.J., Jones, P.D., Simpson, I.R.:An updated assessment of near-surface temperature change from 1850: the HadCRUT5 data set. Journal of Geophysical Research: Atmospheres, 126, e2019JD032361. https://doi.org/10.1029/2019JD032361, 2021.

Morice, C. P., Berry, D. I., Cornes, R. C., Cowtan, K., Cropper, T., Hawkins, E., Kennedy, J. J., Osborn, T. J., Rayner, N. A., Recinos Rivas, B., Schurer, A. P., Taylor, M., Teleti, P. R., Wallis, E. J., Winn, J., and Kent, E. C.: An observational record of global gridded near-surface air temperature change over land and ocean from 1781, Earth Syst. Sci. Data, 17, 7079–7100, https://doi.org/10.5194/essd-17-7079-2025, 2025.

Mountain Research Initiative EDW Working Group. Elevation-dependent warming in mountain regions of the world. Nature Clim Change 5, 424–430, https://doi.org/10.1038/nclimate2563, 2015.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Murphy, J. M., Harris, G. R., Sexton, D. M. H., Kendon, E. J., Bett, P. E., Clark, R. T., Eagle, K. E., Fosser, G., Fung, F., Lowe, J. A., McDonald, R. E., McInnes, R. N., McSweeney, C. F., Mitchell, J. F. B., Rostron, J. W., Thornton, H. E., Tucker, S., & Yamazaki, K., 2018. UKCP18 Land Projections: Science Report (V2*)*. Met Office Hadley Centre. Available at: https://www.metoffice.gov.uk/pub/data/weather/uk/ukcp18/science-reports/UKCP18-Land-report.pdf [Accessed 09/03/25].

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., Koch, D., Lamarque, J.-F., Lee, D., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Takemura, T., and Zhang, H.: Anthropogenic and Natural Radiative Forcing – Supplementary Material, in: Climate Change 2013: The Physical Science Basis, Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, Available at: https://www.ipcc.ch/report/ar5/wg1/chapter-8sm-anthropogenic-and-natural-radiative-forcing-supplementary-material/, 2013

Naik, V., Durack, P. J., Nicholls, Z., Buontempo, C., Dunne, J. P., Hewitt, H. T., Macintosh, C., and O'Rourke, E.: Climate models need more frequent releases of input data — here's how to do it, *Nat.*, 644, 874–875, https://doi.org/10.1038/d41586-025-02642-3, 2025.

Nicholls Z, Meinshausen M, Forster P, Armour K, Berntsen T, Collins W, Jones C, Lewis J, Marotzke J, Milinski S, Rogelj J, Smith C.: Cross-Chapter Box 7.1: Physical emulation of Earth System Models for scenario classification and knowledge integration in AR6. Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. 2021:962-7, 2021.

Nicholls, Z., Meinshausen, M., Forster, P., Armour, K., Berntsen, T., Collins, W., Jones, C., Lewis, J., Marotzke, J., Milinski, S., Rogelj, J., and Smith, C.: Cross-Chapter Box 7.1: Physical emulation of Earth system models for scenario classification and knowledge integration in AR6, in: Climate Change 2021: The Physical Science Basis, Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 962–967, 2021.

Nicklas, J.M., Fox-Kemper, B. and Lawrence, C.: Efficient Estimation of Climate State and Its Uncertainty Using Kalman Filtering with Application to Policy Thresholds and Volcanism. J. Climate, 38(5):1235--1270. https://doi.org/10.1175/JCLI-D-23-0580.1, 2025.

Noone, S., Atkinson, C., Berry, D. I., Dunn, R. J. H., Freeman, E., Perez Gonzalez, I., Kennedy, J. J., Kent, E. C., Kettle, A., McNeill, S., Menne, M., Stephens, A., Thorne, P. W., Tucker, W., Voces, C., and Willett, K. M.: Progress towards a holistic land and marine surface meteorological database and a call for additional contributions, Geosci. Data J., Vol 8, (2), PP. 103-120, https://doi.org/10.1002/gdj3.109, 2021.

Noone, S., D'Arcy, C., Donegan, S., Durkan, W., Essel, B., Healion, K., Hersbach, H., Madden, S., Marshall, J., McConnell, L., Mensah, I., Scroxton,N., Thiesen, S., Thorne, P.: Investigating the potential for students to contribute to climate data rescue:

Introducing the Climate Data Rescue Africa project (CliDaR-Africa). Geosci Data J, Vol 11, Issue 4, pp. 758-774, https://doi.org/10.1002/gdj3.248, 2024.

Olonscheck, D., Suarez-Gutierrez, L., Milinski, S., Beobide-Arsuaga, G., Baehr, J., Fröb, F., Ilyina, T., Kadow, C., Krieger, D., Li, H., Marotzke, J., Plésiat, É., Schupfner, M., Wachsmann, F., Wallberg, L., Wieners, K.-H., and Brune, S.: The new Max Planck Institute Grand Ensemble with CMIP6 forcing and high-frequency model output, *J. Adv. Model. Earth Syst.*, 15, e2023MS003790, https://doi.org/10.1029/2023MS003790, 2023.

O'Neill, B. C. and Oppenheimer, M.: Climate change impacts are sensitive to the concentration stabilization path, Proc. Natl. Acad. Sci. U.S.A., 101 (47) 16411-16416, https://doi.org/10.1073/pnas.0405522101, 2004.

O'Neill, B., M. van Aalst, Z. Zaiton Ibrahim, L. Berrang Ford, S. Bhadwal, H. Buhaug, D. Diaz, K. Frieler, M. Garschagen, A. Magnan, G. Midgley, A. Mirzabaev, A. Thomas, and R. Warren, 2022: Key Risks Across Sectors and Regions. In: Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Pörtner, H.-O., Roberts, D. C., Tignor, M., Poloczanska, E. S., Mintenbeck, K., Alegría, A., Craig, M., Langsdorf, S., Löschke, S., Möller, V., Okem, A., and Rama, B., Cambridge University Press, Cambridge, UK and New York, NY, USA, 2411–2538, https://doi.org/10.1017/9781009325844.025, 2022.

Orvig, S.:Atlas Der Eisverhaltnisse Des Nordatlantischen Ozeans Und Ubersichts-Karten Der Eisverhaltnisse Des Nord- Und Sudpolar - Gebietes. Deutsches Hydrographisches Institut. *Arctic*, 8(2), 128-130. https://doi.org/10.14430/arctic3920, 1955.

Osborn, T. J., Jones, P. D., Lister, D. H., Morice, C. P., Simpson, I. R., Winn, J. P., Hogan, E., and Harris, I. C.: Land surface air temperature variations across the globe updated to 2019: the CRUTEM5 dataset, J. Geophys. Res. Atmos., 126, e2019JD032352, https://doi.org/10.1029/2019JD032352, 2021.

Otto, F., Frame, D., Otto, A. and Allen, M.R.: Embracing uncertainty in climate change policy. Nature Clim Change 5, 917–920. https://doi.org/10.1038/nclimate2716, 2015.

PAGES 2k Consortium: Consistent multidecadal variability in global temperature reconstructions and simulations over the Common Era, Nat. Geosci., 12(8), 643–649, doi: 10.1038/s41561-019-0400-0, 2019.

Parker, D. E. : Effects of changing exposure of thermometers at land stations, Int. J. Climatol., 14, 1–31, https://doi.org/10.1002/joc.3370140102, 1994.

Pfeiffer, M., Zinke, J., Dullo, W.-C., Garbe-Schönberg, D., Latif, M., and Weber, M. E. Indian Ocean corals reveal crucial role of World War II bias for twentieth century warming estimates, Sci. Rep., 7, Article 14434, https://doi.org/10.1038/s41598-017-14352-6, 2017.

Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., Laloyaux, P., Tan, D.G.H., Peubey, C., Thepaut, J.N., Tremolet, Y., Holm, E.V., Bonavita, M., Isaken, L. and Fisher, M.: ERA-20C: An Atmospheric Reanalysis of the Twentieth Century, J. Clim., 29, 4083–4097, https://doi.org/10.1175/JCLI-D-15-0556.1, 2016.

Prasad, D. Gadpal, A. Kapadni, K. Visave, M. Sultanpure, K.: CascadeTabNet: An approach for end to end table detection and structure recognition from image-based documents, *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, WA, USA, pp. 2439-2447, doi: 10.1109/CVPRW50498.2020.00294, 2020.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Qasmi, S and Ribes, A.: Reducing uncertainty in local temperature projections.Sci. Adv.8,eabo6872.DOI:10.1126/sciadv.abo6872, 2022.

Quaglia, I. and Visioni, D.: Modeling 2020 regulatory changes in international shipping emissions helps explain anomalous 2023 warming, Earth Syst. Dynam., 15, 1527–1541, https://doi.org/10.5194/esd-15-1527-2024, 2024.

Raftery, A. E., Gneiting, T., Balabdaoui, F. and Polakowski, M.: Using Bayesian Model Averaging to Calibrate Forecast Ensembles. Mon. Wea. Rev., 133, 1155–1174, https://doi.org/10.1175/MWR2906.1, 2005.

Rajamani, L. and Werksman, J.: The legal character and operational relevance of the Paris Agreement's temperature goal. Phil. Trans. R. Soc. A. 376: 20160458, http://doi.org/10.1098/rsta.2016.0458, 2018.

Rayner, N.A., Parker, D.E., Horton, E.B., Folland, C.K., Alexander, L.V., Rowell, D.P. and Kaplan, A.: Global Analyses of Sea Surface Temperature, Sea Ice, and Night Marine Air Temperature since the Late Nineteenth Century. *Journal of Geophysical Research: Atmospheres*, 108(D14), 4407. https://doi.org/10.1029/2002JD002670, 2003.

Rayner, N. A., Auchmann, R., Bessembinder, J., Brönnimann, S., Brugnara, Y., Capponi, F., Carrea, L., Dodd, E. M. A., Ghent, D., Good, E., Høyer, J. L., Kennedy, J. J., Kent, E. C., Killick, R. E., van der Linden, P., Lindgren, F., Madsen, K. S., Merchant, C. J., Mitchelson, J. R., Morice, C. P., Nielsen-Englyst, P., Ortiz, P. F., Remedios, J. J., van der Schrier, G., Squintu, A. A., Stephens, A., Thorne, P. W., Tonboe, R. T., Trent, T., Veal, K. L., Waterfall, A. M., Winfield, K., Winn, J., and Woolway, R. I.: The EUSTACE Project: Delivering Global, Daily Information on Surface Air Temperature, Bull. Amer. Meteorol. Soc., 101, E1924–E1945, https://doi.org/10.1175/BAMS-D-19-0095.1, 2020.

Reisinger, A., Fuglestvedt, J. S., Pirani, A., Geden, O., Jones, C. D., Maharaj, S., Poloczanska, E., Morelli, A., Johansen, T. G., Adler, C., Betts, R. A., and Seneviratne, S. I. Overshoot: a conceptual review of exceeding and returning to global warming of 1.5 °C Annu. Rev. Environ. Resour., 50, 1.1–1.33, https://doi.org/10.1146/annurev-environ-111523-102029, 2025.

Rhines, A., Tingley, M.P., McKinnon, K.A. and Huybers, P.: Decoding the precision of historical temperature observations. Q.J.R. Meteorol. Soc., 141: 2923-2933. https://doi.org/10.1002/qj.2612, 2015.

Ribes, A. and Terray, L.: Application of regularised optimal fingerprinting to attribution. Part II: application to global near-surface temperature. Clim Dyn 41, 2837–2853, https://doi.org/10.1007/s00382-013-1736-6, 2013.

Ribes A., Tessiot, O., Forster, P., Gillet, N., Masson-Delmotte, V., Vautard, R., and Walsh, T. Towards annual updating of forced warming to date and constrained climate projections. Nat Commun, 16, 9214, https://doi.org/10.1038/s41467-025-63026-9, 2025.

Ribes, A., Qasmi, S., and Gillett, N. P.: Making climate projections conditional on historical observations, Sci. Adv., 7, eabc0671, https://doi.org/10.1126/sciadv.abc0671, 2021.

Richardson, M.,Cowtan, K. and Millar, R.J.: Global temperature definition affects achievement of long-term climate goals. Environmental Research Letters, 13(5), 54004, doi: 10.1088/1748-9326/aab305, 2018.

Richardson, M. T.,: A Physical Explanation for Ocean Air–Water Warming Differences under CO2-Forced Warming. J. Climate, 36, 2857–2871, https://doi.org/10.1175/JCLI-D-22-0215.1, 2023.

Rohde, R., Muller, R., Jacobsen, R., Perlmutter, S., Rosenfeld, A., Wurtele, J., Curry, J., Wickham, C., and Mosher, S.: Berkeley Earth Temperature Averaging Process, Geoinfor. Geostat.: An Overview, 1, 1–13, https://doi.org/10.4172/2327-4581.1000103, 2013

Rogelj, J., Schleussner, C.-F. and Hare, W.: Getting It Right Matters: Temperature Goal Interpretations in Geoscience Research. Geophys. Res. Lett., 44 (20), 10662-10665. https://doi.org/10.1002/2017gl075612, 2017.

Rogelj, J., Fransen, T., den Elzen, M. G. J., Lamboll, R. D., Schumer, C., Kuramochi, T., Hans, F., Mooldijk, S., and Portugal-Pereira, J.: Credibility gap in net-zero climate targets leaves world at high risk, Science, 380 (6649), 1014–1016, https://doi.org/10.1126/science.adg6248, 2023.

Ruddiman, W.F. and J.S. Thomson,J.S.: The case for human causes of increased atmospheric CH4 over the last 5000 years. Quaternary Science Reviews, 20(18), 1769–1777, doi: 10.1016/s0277-3791(01)00067-1, 2001.

Ruffman, A., Hill, B. and Drinkwater, K.: A two-hundred-year historical record of sea ice extent for the Scotian Shelf, Gulf of St. Lawrence and its relationship to the sea ice record for the Grand Banks and the North Atlantic Oscillation. Proceedings of Canadian Quaternary Association: Canadian Geomorphological Research Group 2003. Halifax, Canada: The Canadian Geomorphology Research Group, 2003.

Ruggieri, E. and Antonellis, M.: An exact approach to Bayesian sequential change point detection, Computational Statistics & Data Analysis, 97, 71-86, https://doi.org/10.1016/j.csda.2015.11.010, 2016.

Ryan, C., Duffy, C., Broderick, C., Thorne, P. W., Curley, M., Walsh, S., Daly, C., Treanor, M., and Murphy, C.: Integrating Data Rescue into the Classroom, Bull. Amer. Meteorol. Soc., 99 (9), 1757–1764, https://doi.org/10.1175/BAMS-D-17-0147.1, 2018.

Samakinwa, E., Valler, V., Hand, R., Neukom, R., Gómez-Navarro, J.J., Kennedy, J., Rayner, N.A. and Brönnimann, S.: An ensemble reconstruction of global monthly sea surface temperature and sea ice concentration 1000–1849. Scientific Data, 8, 261. https://doi.org/10.1038/s41597-021-01043-1, 2021.

Samset, B. H., Zhou, C., Fuglestvedt, J. S., Lund, M. T., Marotzke, J., and Zelinka, M. D.: Earlier emergence of a temperature response to mitigation by filtering annual variability, Nat. Commun., 13 (1), 1578, https://doi.org/10.1038/s41467-022-29247-y, 2022.

Samset, B. H., Zhou, C., Fuglestvedt, J. S., Lund, M. T., Marotzke, J., and Zelinka, M. D.: Steady global surface warming from 1973 to 2022 but increased warming rate after 1990, Commun. Earth Environ., 4: 400, https://doi.org/10.1038/s43247-023-01061-4, 2023.

Samset, B. H., Wilcox, L. J., Allen, R. J., Stjern, C. W., Lund, M. T., Ahmadi, S., Ekman, A., Elling, M. T., Fraser-Leach, L., Griffiths, P., Keeble, J., Koshiro, T., Husher, P., Lewinscha, A., Makkonen, R., Merikanto, J., Nabat, P., Narazenzko, L., O'Donnell, D., Oshima, N., Rumbold, St., Takemura, T., Tsigardis, K. and Westervelet, M.: East Asian aerosol cleanup has likely contributed to the recent acceleration in global warming Commun. Earth Environ., 6, Article 543, https://doi.org/10.1038/s43247-025-02527-3, 2025.

Scherrer, S.C., de Valk, C., Begert, M., Gubler, S., Kotlarski, S., Croci-Maspoli, M.: Estimating trends and the current climate mean in a changing climate, Climate Services, 33, 100428, https://doi.org/10.1016/j.cliser.2023.100428, 2024.

Schleussner, C.-F., Ganti, G., Lejeune, Q., Zhu, B., Pfleiderer, P., Prütz, R., Ciais, P., Frölicher, T. L., Fuss, S., Gasser, T., Gidden, M. J., Kropf, C. M., Lacroix, F., Lamboll, R., Martyr, R., Maussion, F., McCaughey, J. W., Meinshausen, M., Mengel, M., Nicholls, Z., Quilcaille, Y., Sanderson, B., Seneviratne, S. I., Sillmann, J., Smith, C. J., Steinert, N. J., Theokritoff, E., Warren, R., Price, J., and Rogelj, J.: Overconfidence in climate overshoot, Nature, 634 (8033), 366–373, https://doi.org/10.1038/s41586-024-08020-9, 2024.

Schmidt, G. A., Andrews, T., Bauer, S. E., Durack, P. J., Loeb, N. G., Ramaswamy, V., Arnold, N. P., Bosilovich, M. G., Cole, J., Horowitz, L. W., Johnson, G. C., Lyman, J. M., Medeiros, B., Michibata, T., Olonscheck, D., Paynter, D., Raghuraman, S. P., Schulz, M., Takasuka, D., Tallapragada, V., Taylor, P. C., and Ziehn, T.: CERESMIP: a climate modeling protocol to investigate recent trends in the Earth's Energy Imbalance, Front. Clim., 5, 1202161, https://doi.org/10.3389/fclim.2023.1202161, 2023.

Schneider, L., Konter, O., Esper, J. and Anchukaitis, K.J.: Constraining the Nineteenth-Century Temperature Baseline for Global Warming. J. Climate, 36, 6261–6272, https://doi.org/10.1175/JCLI-D-22-0806.1, 2023.

Schurer, A. P., Mann, M. E., Hawkins, E., Tett, S. F. B., and Hegerl, G. C.: Importance of the pre-industrial baseline for likelihood of exceeding Paris goals, Nat. Clim. Change, 7 (8), 563–567, https://doi.org/10.1038/nclimate3345, 2017.

Schwarzwald, K and Lenssen, N.,: The importance of internal climate variability in climate impact projections, Proc. Natl. Acad. Sci. U.S.A. 119 (42) e2208095119, https://doi.org/10.1073/pnas.2208095119, 2022.

Semenov, V.A., Aldonina, T.A., Li, F., Keenlyside, N.S. and Wang. L.: Arctic Sea Ice Variations in the First Half of the 20th Century: A New Reconstruction Based on Hydrometeorological Data. Advances in Atmospheric Sciences, 41(8), 1483-1495. https://doi.org/10.1007/s00376-024-3320-x, 2024.

Seshadri, A. K.:Origin of path independence between cumulative CO2 emissions and global warming," Clim Dynam, vol. 49, no. 9–10, pp. 3383–3401, doi: 10.1007/s00382-016-3519-3, 2017.

Shapiro, I., Colony, R. and Vinje, T.: April sea ice extent in the Barents Sea, 1850–2001. Polar Research, (22)1, 5-10. https://doi.org/10.3402/polar.v22i1.6437, 2006.

Shumway R.H. and .Stoffer,D.S.: Time Series Analysis and Its Applications
With R Examples, 5th edition, Springer, 689pp.
http://www.stat.ucla.edu/~frederic/415/S23/tsa4.pdf, 2024.

Sigl, M., Winstrup, M., McConnell, J. R., Welten, K. C., Plunkett, G., Ludlow, F., Büntgen, U., Caffee, M., Chellman, N., Dahl-Jensen, D., Fischer, H., Kipfstuhl, S., Kostick, C., Maselli, O. J., Mekhaldi, F., Mulvaney, R., Muscheler, R., Pasteris, D. R., Pilcher, J. R., Salzer, M., Schüpbach, S., Steffensen, J. P., Vinther, B. M., and Woodruff, T. E.: Timing and climate forcing of volcanic eruptions for the past 2,500 years, Nature, 523 (7562), 543–549, https://doi.org/10.1038/nature14565, 2015.

Simmons, A.J. et al., 2017: A reassessment of temperature variations and trends from global reanalyses and monthly surface climatological datasets. Quarterly Journal of the Royal Meteorological Society, 143(702), 101–119, doi: 10.1002/qj.2949.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Simmons, A. J., Berrisford, P., Dee, D. P., Hersbach, H., Hirahara, S. and Thépaut, J.-N.: A reassessment of temperature variations and trends from global reanalyses and monthly surface climatological datasets. Quarterly Journal of the Royal Meteorological Society, 143 (702). pp. 101-119, DOI: 10.1002/qj.2949, 2017.

Sippel, S., Meinshausen, N., Merrifield, A., Lehner, F., Pendergrass, A. G., Fischer, E., and Knutti, R.: Uncovering the forced climate response from a single ensemble member using statistical learning, J. Clim., 32 (17), 5677–5699, https://doi.org/10.1175/JCLI-D-18-0882.1, 2019.

Sippel, S., Kent, E. C., Meinshausen, N., Chan, D., Kadow, C., Neukom, R., Fischer, E. M., Humphrey, V., Rohde, R., de Vries, I., and Knutti, R.: Early-twentieth-century cold bias in ocean surface temperature observations, Nature, 635 (8039), 618–624, https://doi.org/10.1038/s41586-024-08230-1, 2024.

Singh, L. G., and Middleton, S. E.: Data Rescue for Historical Document Tables Using Semi-Supervised Learning, IJDAR (under review), preprint,https://doi.org/10.21203/rs.3.rs-4391424/v1, 2024.

Slivinski, L. C., Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Giese, B. S., McColl, C., Allan, R., Yin, X., Vose, R., Titchner, H., Kennedy, J., Spencer, L. J., Ashcroft, L., Brönnimann, S., Brunet, M., Camuffo, D., Cornes, R., Cram, T. A., Crouthamel, R., Domínguez-Castro, F., Freeman, J. E., Gergis, J., Hawkins, E., Jones, P. D., Jourdain, S., Kaplan, A., Kubota, H., Le Blancq, F., Lee, T.-C., Lorrey, A., Luterbacher, J., Maugeri, M., Mock, C. J., Moore, G. W. K., Przybylak, R., Pudmenzky, C., Reason, C., Slonosky, V. C., Smith, C. A., Tinz, B., Trewin, B., Valente, M. A., Wang, X. L., Wilkinson, C., Wood, K., and Wyszynski, P.: Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system, Q. J. R. Meteorol. Soc., 145 (724), 2876–2908, https://doi.org/10.1002/qj.3598, 2019.

Slivinski, L. C., Compo, G. P., Sardeshmukh, P. D., Whitaker, J. S., McColl, C., Allan, R. J., Brohan, P., Yin, X., Smith, C. A., Spencer, L. J., Vose, R. S., Rohrer, M., Conroy, R. P., Schuster, D. C., Kennedy, J. J., Ashcroft, L., Brönnimann, S., Brunet, M., Camuffo, D., Cornes, R., Cram, T. A., Domínguez-Castro, F., Freeman, J. E., Gergis, J., Hawkins, E., Jones, P. D., Kubota, H., Lee, T.-C., Lorrey, A. M., Luterbacher, J., Mock, C. J., Przybylak, R. K., Pudmenzky, C., Slonosky, V. C., Tinz, B., Trewin, B., Wang, X. L., Wilkinson, C., Wood, K., and Wyszynski, P.: An evaluation of the performance of the Twentieth Century Reanalysis Version 3, J. Clim., 34 (4), 1417–1438, https://doi.org/10.1175/JCLI-D-20-0505.1, 2021.

Smith, T. M. and R. W. Reynolds, R.W.: A Global Merged Land–Air–Sea Surface Temperature Reconstruction Based on Historical Observations (1880–1997). J. Climate, 18, 2021–2036, https://doi.org/10.1175/JCLI3362.1, 2005.

Smith, C., Nicholls, Z. R. J., Armour, K., Collins, W., Forster, P., Meinshausen, M., Palmer, M. D., and Watanabe, M.: *The Earth's Energy Budget, Climate Feedbacks, and Climate Sensitivity Supplementary Material*, In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. 36 pp., Available online at:https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Chapter07_SM.pdf, 2021.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Smith, C., Cummins, D. P., Fredriksen, H.-B., Nicholls, Z., Meinshausen, M., Allen, M., Jenkins, S., Leach, N., Mathison, C., and Partanen, A.-I.: fair-calibrate v1.4.1: calibration, constraining, and validation of the FaIR simple climate model for reliable future climate projections, Geosci. Model Dev., 17, 8569–8592, https://doi.org/10.5194/gmd-17-8569-2024, 2024.

Smith, C. J., Forster, P. M., Allen, M., Leach, N., Millar, R. J., Passerello, G. A., and Regayre, L. A.: FAIR v1.3: a simple emissions-based impulse response and carbon cycle model, *Geosci. Model Dev.*, 11, 2273–2297, https://doi.org/10.5194/gmd-11-2273-2018, 2018.

Soci, C., Hersbach, H., Simmons, A., Poli, P., Bell, B., Berrisford, P., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Radu, R., Schepers, D., Villaume, S., Haimberger, L., Woollen, J., Buontempo, C., and Thépaut, J.-N.: The ERA5 global reanalysis from 1940 to 2022, Q. J. R. Meteorol. Soc., 150 (764), 4014–4048, https://doi.org/10.1002/qj.4803, 2024.

Sospedra-Alfonso, R., Merryfield, W. J., Toohey, M., Timmreck, C., Vernier, J. P., Bethke, I., Wang, Y., Bilbao, R., Donat, M. G., Ortega, P., Cole, J., Lee, W.-S., Delworth, T. L., Paynter, D., Zeng, F., Zhang, L., Khodri, M., Mignot, J., Swingedouw, D., Torres, O., Hu, S., Man, W., Zuo, M., Hermanson, L., Smith, D., Kataoka, T., and Tatebe, H.: Decadal prediction centers prepare for a major volcanic eruption, Bull. Amer. Meteor. Soc. 105, E2496–E2524, https://doi.org/10.1175/BAMS-D-23-0111.1, 2024.

Stott, P. A., and Tett, S.F.B.:Scale-Dependent Detection of Climate Change. J. Climate, 11, 3282–3294, https://doi.org/10.1175/1520-0442(1998)011<3282:SDDOCC>2.0.CO;2, 1998.

Sun, W. B., Li, Q., Huang, B., Cheng, J., Song, Z., Li, H., Dong, W., Zhai, P., and Jones, P.: The assessment of global surface temperature change from 1850s: The C-LSAT2.0 ensemble and the CMST-Interim datasets, *Adv. Atmos. Sci.*, 38, 875–888, https://doi.org/10.1007/s00376-021-1012-3, 2021.

Sun, W. Yang, Y., Chao,C., Dong, W., Huang, B., Jones, P. and Li, Q.: Description of the China global Merged Surface Temperature version 2.0. *Earth System Science Data*, 14(4), 1677-1693. https://doi.org/10.5194/essd-14-1677-2022, 2022.

Szopa, S., Naik, V., Adhikary, B., Artaxo, P., Berntsen, T., Collins, W. D., Fuzzi, S., Gallardo, L., Kiendler-Scharr, A., Klimont, Z., Liao, H., Unger, N., and Zanis, P.: Short-Lived Climate Forcers, in: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, UK and New York, NY, USA, 817–922, https://doi.org/10.1017/9781009157896.008, 2021.

Teleti, P.R., Rees, W.G., Dowdeswell, J.A. and Wilkinson, C.: A historical Southern Ocean climate dataset from whaling ships' logbooks. *Geoscience Data Journal*, 6(1), 30-40. https://doi.org/10.1002/gdj3.65, 2019.

Teleti, P., Hawkins, E. and Wood, K.R.: Digitizing weather observations from World War II US naval ship logbooks. *Geoscience Data Journal*, 00, 1–16. Available from: https://doi.org/10.1002/gdj3.222, 2023.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

Thompson, D.W.J, Wallace, J.M., Jones, P.D. and Kennedy, J.J.:Identifying signatures of natural climate variability in time series of global-mean surface temperature: methodology and insights. Journal of Climate, 22, 6120-6141, https://doi.org/10.1175/2009JCLI3089.1, 2009.

Thorne, P. W., Parker, D. E., Christy, J. R., and Mears, C. A.: Uncertainties in Climate Trends: Lessons from Upper-Air Temperature Records, *Bull. Amer. Meteor. Soc.* **86**, 1437–1442, https://doi.org/10.1175/BAMS-86-10-1437, 2005.

Thorne, P. W., Diamond, H. J., Goodison, B., Harrigan, S., Hausfather, Z., Ingleby, N. B., Jones, P. D., Lawrimore, J. H., Lister, D. H., Merlone, A., Oakley, T., Palecki, M., Peterson, T. C., de Podesta, M., Tassone, C., Venema, V., and Willett, K. M.: Towards a global land surface climate fiducial reference measurements network, *Int. J. Climatol.*, **38**, 2760–2774, https://doi.org/10.1002/joc.5458, 2018.

Thorne, P.W. et al.: https://github.com/jnickla1/climate_data (note: to be updated as a long-term repository with doi prior to publication), 2025

Titchner, H.A. and Rayner, N.A: The Met Office Hadley Center Sea Ice and Sea Surface Temperature Data Set, Version 2: 1. Sea Ice Concentrations. *Journal of Geophysical Research: Atmospheres*, 119(6), 2864-2889. https://doi.org/10.1002/2013JD020316, 2014.

Toohey, M. and Sigl, M.: Volcanic stratospheric sulfur injections and aerosol optical depth from 500 BCE to 1900 CE, Earth Syst. Sci. Data, 9, 809–831, https://doi.org/10.5194/essd-9-809-2017, 2017.

Trenberth, K.E., Caron, J.M., Stepaniak, D.P. & Worley, S.: Evolution of El Niño-Southern Oscillation and global atmospheric surface temperatures. Journal of Geophysical Research, 107 (D8), https://doi.org/10.1029/2000JD000298, 2002.

Trewin, B.: Exposure, instrumentation, and observing practice effects on land temperature measurements. WIREs Clim Change, 1: 490-506. https://doi.org/10.1002/wcc.46, 2010.

Trewin, B.:Assessing Internal Variability of Global Mean Surface Temperature From Observational Data and Implications for Reaching Key Threshold. *Journal of Geophysical Research Atmospheres*, 127, e2022JD036747, https://doi.org/10.1029/2022JD036747, 2022.

United Nations: Vienna Convention on the Law of Treaties (1969), United Nations, Treaty Series, 1155, 331, available at: https://legal.un.org/ilc/texts/instruments/english/conventions/1_1_1969.pdf (last access: 2 April 2024), 1969.

UNFCCC: Report on the structured expert dialogue on the 2013–2015 review, available at: http://unfccc.int/resource/docs/2015/sb/eng/inf01.pdf (last access: 2 April 2024), 2015.

UNFCCC: The Paris Agreement, available at: https://unfccc.int/sites/default/files/resource/parisagreement_publication.pdf (last access: 2 April 2024), 2016.

UNFCCC: Report of the Conference of the Parties serving as the meeting of the Parties to the Paris Agreement on the third part of its first session, held in Katowice from 2 to 15 December 2018. Addendum 2. Part two: Action taken by the Conference of the Parties serving as the meeting of the Parties to the Paris Agreement, available at: https://unfccc.int/documents/193408 (last access: 2 April 2024), 2019.

https://doi.org/10.5194/essd-2025-825
Preprint. Discussion started: 28 January 2026

UNFCCC: Report of the Conference of the Parties on its twenty-seventh session, held in Sharm el-Sheikh from 6 to 20 November 2022. Addendum. Part two: Action taken by the Conference of the Parties at its twenty-seventh session, available at: https://unfccc.int/documents/626563 (last access: 2 April 2024), 2022.

Vaccaro, A., Emile-Geay, J., Guillot, D., Verna, R., Morice, C., Kennedy, J., and Rajaratnam, B.: Climate Field Completion via Markov Random Fields: Application to the HadCRUT4.6 Temperature Dataset, *J. Clim.*, 34, 4169–4188, https://doi.org/10.1175/JCLI-D-19-0814.1, 2021.

van der Werf, G. R., Randerson, J. T., Giglio, L., van Leeuwen, T. T., Chen, Y., Rogers, B. M., Mu, M., van Marle, M. J. E., Morton, D. C., Collatz, G. J., Yokelson, R. J. and Kasibhatla, P. S.: Global fire emissions estimates during 1997–2016, Earth System Science Data, 9, 697–720, https://doi.org/10.5194/essd-9-697-2017, https://essd.copernicus.org/articles/9/697/2017/, 2017.

Venema, V. K. C., Mestre, O., Aguilar, E., Auer, I., Guijarro, J. A., Domonkos, P., Vertacnik, G., Szentimrey, T., Stepanek, P., Zahradnicek, P., Viarre, J., Müller-Westermeier, G., Lakatos, M., Williams, C. N., Menne, M. J., Lindau, R., Rasol, D., Rustemeier, E., Kolokythas, K., Marinova, T., Andresen, L., Acquaotta, F., Fratianni, S., Cheval, S., Klancar, M., Brunetti, M., Gruber, C., Prohom Duran, M., Likso, T., Esteban, P., and Brandsma, T.: Benchmarking homogenization algorithms for monthly data, Clim. Past, 8, 89–115, https://doi.org/10.5194/cp-8-89-2012, 2012.

Vercruysse, B., Thirukokaranam Chandrasekar, K. k., Muheki, D., Thiery, W., Verbeeck, H., Hufkens, K., Jacobsen, K., Birkholz, J. M., and Verbruggen, C.: A critical evaluation of tabular data extraction methods on 19th century climate data, Springer Nature, submitted, 2024.

Vinje, T.: *Barents Sea Ice Edge Variation Over the Past 400 Years*. Proceedings of the
Workshop on Sea-Ice Charts of the Arctic, 949, 4-6. Geneva, Switzerland: World Meteorological Organization, 1999.

Visser, H., Dangendorf, S., van Vuuren, D. P., Bregman, B. and Petersen, A. C.: Signal detection in global mean temperatures after "Paris": an uncertainty and sensitivity analysis, Clim. Past, 14, 139–155, https://doi.org/10.5194/cp-14-139-2018, 2018.

Walling, M. and Borenstein, S.: Tuesday set an unofficial record for the hottest day on Earth. Wednesday may break it. *AP News*, https://apnews.com/article/global-record-breaking-heat-july-27069b5380117534d78f1f40a6edc7a0 (last access: 27 July 2024), 4 July 2023.

Wallis, E. J., Osborn, T. J., Taylor, M., Jones, P. D., Joshi, M. and Hawkins, E.:Quantifying exposure biases in early instrumental land surface air temperature observations. *International Journal of Climatology*, early view. https://doi.org/10.1002/joc.8401, 2024.

Walsh, J. E., Chapman, W. L., Fetterer, F., and Stewart, J. S.: Gridded Monthly Sea Ice Extent and Concentration, 1850 Onward (G10010, Version 2) [Data Set], Boulder, Colorado, USA: National Snow and Ice Data Center, https://doi.org/10.7265/jj4s-tq79, 2019.

Way, R.G., Oliva, F. and Viau, A.E.: Underestimated warming of northern Canada in the Berkeley Earth temperature product. Int. J. Climatol., 37: 1746-1757. https://doi.org/10.1002/joc.4808, 2017.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltrán, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B.: The FAIR Guiding Principles for Scientific Data Management and Stewardship, *Sci. Data*, **3**, 160018, https://doi.org/10.1038/sdata.2016.18, 2016.

Williams, C. N., Menne, M. J., and Thorne, P. W.: Benchmarking the performance of pairwise homogenization of surface temperatures in the United States, *J. Geophys. Res.*, 117, D05116, https://doi.org/10.1029/2011JD016761, 2012.

Wills, R. C. J., Battisti, D. S., Armour, K. C., Schneider, T., and Deser, C.: Pattern Recognition Methods to Separate Forced Responses from Internal Variability in Climate Model Ensembles and Observations, *J. Clim.*, 33, 8693–8719, https://doi.org/10.1175/JCLI-D-19-0855.1, 2020

WMO: Cg-Ext(2021) World Meteorological Congress Abridged Final Report of the Extraordinary Session, virtual session, 11–21 October 2021, WMO-No. 1281, available at: https://library.wmo.int/records/item/57850-world-meteorological-congress (last access: 2 April 2024), 2021.

WMO: WMO Unified Data Policy, WMO-No. 1275, available at: https://library.wmo.int/records/item/58009-wmo-unified-data-policy (last access: 2 April 2024), 2022.

WMO: State of the Global Climate 2024, WMO-No. 1368, available at: https://library.wmo.int/idurl/4/69455 (last access: 29 June 2025), 2025.

Wu, Z., Huang, N. E., Wallace, J. M., Smoliak, B. V., and Chen, X.: On the time-varying trend in global-mean surface temperature, *Clim. Dyn.*, **37**, 759–773, https://doi.org/10.1007/s00382-011-1128-8, 2011.

Wu, T., Hu, A., Gao, F., Zhang, J., and Meehl, G. A.: New insights into natural variability and anthropogenic forcing of global/regional climate evolution, *npj Clim. Atmos. Sci.*, **2**, Article 18, https://doi.org/10.1038/s41612-019-0075-7, 2019.

Wu, X., Yeager, S. G., Deser, C., Rosenbloom, N., and Meehl, G. A.: Volcanic forcing degrades multiyear-to-decadal prediction skill in the tropical Pacific, *Sci. Adv.*, **9**, Article e-add9364, https://doi.org/10.1126/sciadv.add9364, 2023.

Xu, F. and Ignatov, A.:Error characterization in *i*Quam SSTs using triple collocations with satellite measurements, *Geophys. Res. Lett.*, 43, 10,826–10,834, doi:10.1002/2016GL070287, 2016.

Xu, W., Li, Q., Jones, P., Wang, X. L., Trewin, B., Yang, S., Zhu, C., Zhai, P., Wang, J., Vincent, L., Dai, A., Gao, Y., and Ding, Y.: A new integrated and homogenized global monthly land surface air temperature dataset for the period since 1900, *Clim. Dyn.*, 50, 2513–2536, https://doi.org/10.1007/s00382-017-3755-1, 2018.

Yang, J., Xiao, C., Liu, J., Li, S. and Qin, D.: Variability of Antarctic sea ice extent over the past 200 years. *Science Bulletin*, 66(23), 2394-2404. https://doi.org/10.1016/j.scib.2021.07.028, 2021.

Yao, Y., Vehtari, A., Simpson, D. and Gelman, A.: "Using Stacking to Average Bayesian Predictive Distributions (with Discussion)." Bayesian Anal. 13 (3) 917 - 1007, https://doi.org/10.1214/17-BA1091, 2018.

Yin, X., Huang, B., Menne, M., Vose, R., Zhang, H.-M., Adeyeye, A., Applequist, S., Gleason, K., Liu, C., and Sanchez-Lugo, A.: NOAAGlobalTemp Version 6: An AI-Based Global Surface Temperature Dataset, *Bull. Amer. Meteor. Soc.*, 105, E2184–E2193, https://doi.org/10.1175/BAMS-D-24-0012.1, 2025.

Yu, M. and Ruggieri, E.: Change point analysis of global temperature records. Int J Climatol. 39: 3679–3688. https://doi.org/10.1002/joc.6042, 2019.

Yuan, T., Song, H., Oreopoulos, L., Wood, R., Bian, H., Breen, K., Chin, M., Yu, H., Barahona, D., Meyer, K., and Platnick, S.: Abrupt reduction in shipping emission as an inadvertent geoengineering termination shock produces substantial radiative warming, *Commun. Earth Environ.*, **5**, Article 281, https://doi.org/10.1038/s43247-024-01442-3, 2024.

Ziomek. J. and Middleton, S.E.: GloSAT Historical Measurement Table Dataset: Enhanced Table Structure Recognition Annotation for Downstream Historical Data Rescue, 6th International Workshop on Historical Document Imaging and Processing (HIP-2021), ICDAR-2021, Lausanne, Switzerland, https://dl.acm.org/doi/10.1145/3476887.3476890, 2021.

Zommers, Z., Marbaix, P., Fischlin, A., Ibrahim, Z. Z., Grant, S., Magnan, A. K., Pörtner, H.-O., Howden, M., Calvin, K., Warner, K., Thiery, W., Sebesvari, Z., Davin, E. L., Evans, J. P., Rosenzweig, C., O'Neill, B. C., Patwardhan, A., Warren, R., van Aalst, M. K., and Hulbert, M.: Burning embers: towards more transparent and robust climate-change risk assessments, Nat Rev Earth Environ, https://doi.org/10/gg985p, 2020.