# Consequential Differences in Satellite-Era Sea Surface Temperature Trends Across Datasets

**Sofia Menemenlis**

smenemenlis@princeton.edu

Princeton University https://orcid.org/0000-0003-4420-3785

**Gabriel Vecchi**

Princeton University https://orcid.org/0000-0002-5085-224X

**Wenchang Yang**

Princeton University https://orcid.org/0000-0003-0053-9527

**Stephan Fueglistaler**

Princeton University https://orcid.org/0000-0002-0419-440X

**Shiv Priyam Raghuraman**

University of Illinois Urbana-Champaign

**Analysis**

**Keywords:**

**Posted Date:** February 4th, 2025

**DOI:** https://doi.org/10.21203/rs.3.rs-5493841/v1



**Additional Declarations:** There is **NO** Competing Interest.

# Consequential Differences in Satellite-Era Sea Surface Temperature Trends Across Datasets

**S. Menemenlis[1], G. A. Vecchi[2], Wenchang Yang[2], S. Fueglistaler[2], SP. Raghuraman[3]**

[1]Program in Atmospheric & Oceanic Sciences, Princeton University

[2]Department of Geosciences, Princeton University

[3]Department of Climate, Meteorology & Atmospheric Sciences, University of Illinois Urbana-Champaign

Corresponding author: Sofia Menemenlis (smenemenlis@princeton.edu)

## Abstract

Global surface temperatures since the 1980s, when near-global satellite-based sea surface temperature (SST) measurements became available, are presumed to be well-known. Satellite-era warming trends in four commonly used global (land+ocean) temperature reconstructions agree closely. However, trends in four commonly used SST datasets show first-order differences, with 1982-2022 60°S to 60°N trends ranging from 0.09-0.17 °C per decade. These large discrepancies are perplexing given the agreement between global temperature datasets and the fact that 70% of Earth's surface is covered by ocean, but are legible upon recognizing that global temperature datasets use SST fields whose trends agree more closely than those of the above four SST datasets. Considering the trend uncertainty across SST datasets widens the range of plausible global temperature trends and impacts interpretations of recent record global temperatures. We discuss some of these findings' wide implications for understanding the character and impacts of global warming.

## 1. Introduction

Although estimates of global temperatures based on direct observations are available dating back to the late 19th century, these reconstructions are assumed to be best known since the 1980s, the advent of satellite-based remote sensing of sea surface temperatures (SSTs). Infrared and microwave measurements from instruments on meteorological and military satellites are used to estimate SSTs. These remote sensing products are combined with *in situ* surface marine observations from ships, drifting and moored buoys, and fixed observing stations, which are collected in the International Comprehensive Ocean-Atmosphere Dataset (ICOADS, Freeman et al., 2017). Each source of temperature information carries particular biases and uncertainties. Each also produces different sampling patterns in space and time, and measures surface temperatures at different depths, from several meters below the surface for ship intake measurements to upper micrometers of the ocean surface for infrared satellites. To produce global reconstructions of SSTs, heterogeneous and spatio-temporally incomplete observations must be combined.

A handful of institutions produce regularly updated, spatially complete, global SST reconstructions for public access. These datasets, with current version numbers, are the Hadley

Centre's sea ice and sea surface temperature (HadISST1, Rayner, 2003) from the UK Met Office, the Extended Reconstructed Sea Surface Temperature version 5 (ERSSTv5, Huang et al., 2017) and Daily Optimum Interpolation Sea Surface Temperature version 2.1 (OISSTv2.1, Huang et al., 2021) from the U.S. National Oceanic and Atmospheric Administration (NOAA), and the Centennial In Situ Observation Based Estimates of the Variability of SST and Marine Meteorological Variables version 2 (COBE2, Hirahara et al., 2014) from the Japan Meteorological Agency (JMA). These reconstructions use much of the same observational data, but employ different techniques to correct for measurement biases, homogenize and interpolate SST data, estimate SSTs in areas with sea ice, and estimate uncertainties (see Methods). Similarly, a few global surface temperature products, which combine SSTs with reconstructions of near-surface land temperature, are regularly updated and made available for public use. These products, with current version numbers where applicable, are the Goddard Institute for Space Studies Surface Temperature product version 4 (GISTEMP v4, Hansen et al., 2010; Lenssen et al., 2019), The UK Met Office Hadley Centre/Climatic Research Unit global surface temperature data set version 5 (HadCRUT5, Morice et al., 2021), the NOAA Global Temperature dataset version 6 (NOAAGlobalTempv6.0.0), and the BerkeleyEarth global temperature dataset (Rohde & Hausfather, 2020) from the California-based nonprofit of the same name.

These SST and global temperature products are a crucial part of climate science infrastructure. They are used to monitor climate change, as boundary conditions for climate models, as input to reanalysis, and in the development of climate models. However, these data are typically used without explicit consideration of uncertainty and without comparison to other datasets. Here, we compare commonly used SST and global temperature reconstructions. We focus on trends in the satellite era when observational coverage is most robust, and consider average SSTs between 60S and 60N to avoid comparing datasets with explicitly different techniques for estimating temperatures in areas with sea ice. We will demonstrate that large differences across dataset in satellite-era global SST trends exist, imply greater uncertainty in global-mean temperature trends than previously apparent, and impact interpretations of past and present climate.

## 2. Results

## 2.1 Temperature Trend Uncertainty

The four publicly available global surface temperature datasets—NOAAGlobalTempv6.0.0, BerkeleyEarth, GISTEMPv4, and HadCRUT5.0.2—exhibit very similar 60°S-60°N 1982-2022 land + ocean warming trends of 0.1693, 0.1725, 0.1733, and 0.1752 °C decade$^{-1}$ respectively. Over this same area and time period, the four SST datasets—HadISST1, ERSSTv5, COBE2, and OISSTv2.1—exhibit trends of 0.0933, 0.1225, 0.1409, and 0.1691 °C decade$^{-1}$, respectively (Fig. 1). That is, satellite-era warming trends differ between SST datasets by up to roughly 80%. The differences between SST datasets cannot be explained by statistical uncertainties associated with the linear fit, which are small enough that the 95% confidence intervals do not overlap. Nor can the differences be explained by any single ocean region or latitude range; they are global-scale features (Figure S1, Figure S2).

The four SST datasets drift in relation to one another at the decadal scale, despite similar spatial warming patterns (Figure S1) and very consistent year-to-year variability. HadISST consistently records higher absolute temperatures than ERSST, the offset between them narrowing slightly over the satellite era. COBE and especially OISST drift from temperatures more similar to ERSST to those more similar to HadISST, such that COBE, OISST, and HadISST record similar temperature values over the past decade. SSTs in MERRA2 reanalysis are most similar to OISST. ERA5 reanalysis temperatures are higher than MERRA2 at the beginning of the satellite era, but converge during the past 2 decades (Figure S3).

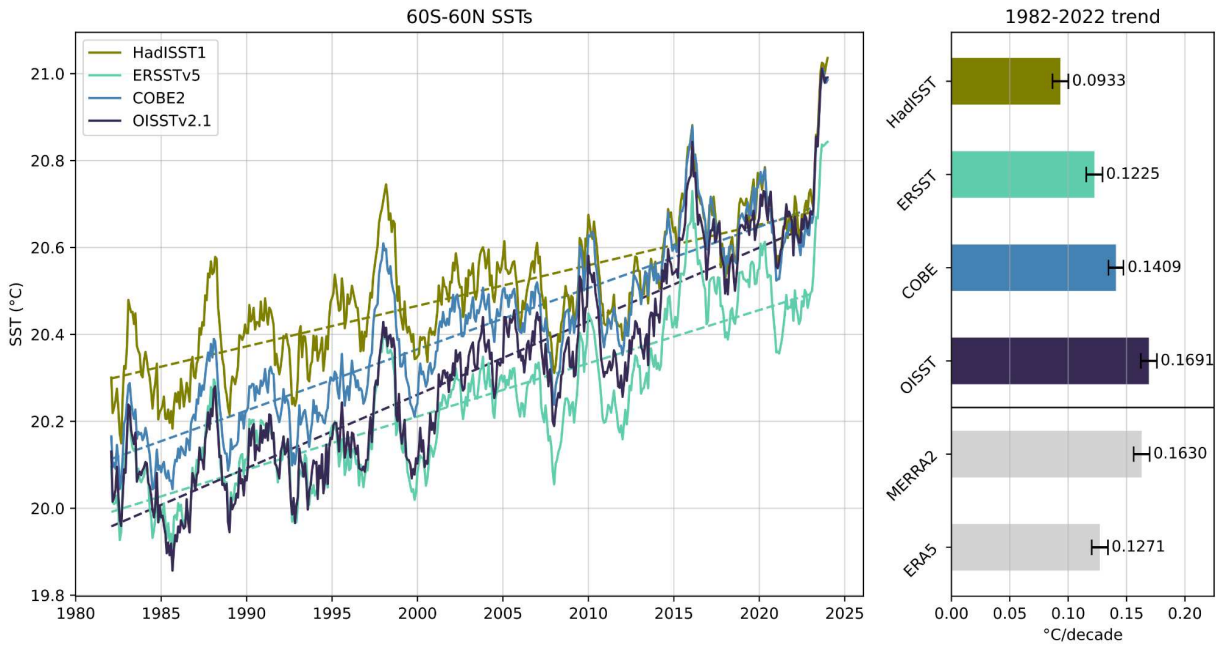


**Figure 1: Differences in satellite-era SST trends.** Left panel: time series of deseasonalized monthly average sea surface temperatures (SSTs) between 60°S and 60°N (°C). Dashed lines show the linear trend between 1982 and 2022. HadISST1, ERSSTv5, COBE2, and OISSTv2.1 are depicted in olive, teal, blue, and indigo respectively. Right panel: trends in SSTs between 60°S and 60°N for the period 1982-2022, in °C per decade, for the same four SST datasets, plus MERRA2 and ERA5 reanalysis.. Error bars represent the 95% confidence interval calculated based on the standard error of the linear fit.

If one produced estimates of 60°S-60°N land+ocean temperature with each of the four SST datasets, using identical land temperature trends, one would expect the range in trends between these datasets to be approximately 0.74*0.076, or 0.056, °C decade$^{-1}$. (The range in trends derived from the four SST datasets is 0.076 °C decade$^{-1}$, and the ocean takes up 74% of the Earth's surface area between 60°S and 60°N.) However, the range in 60°S-60°N trends given by the four global temperature datasets is a mere  0.0062 °C decade$^{-1}$ (Figure 2a).

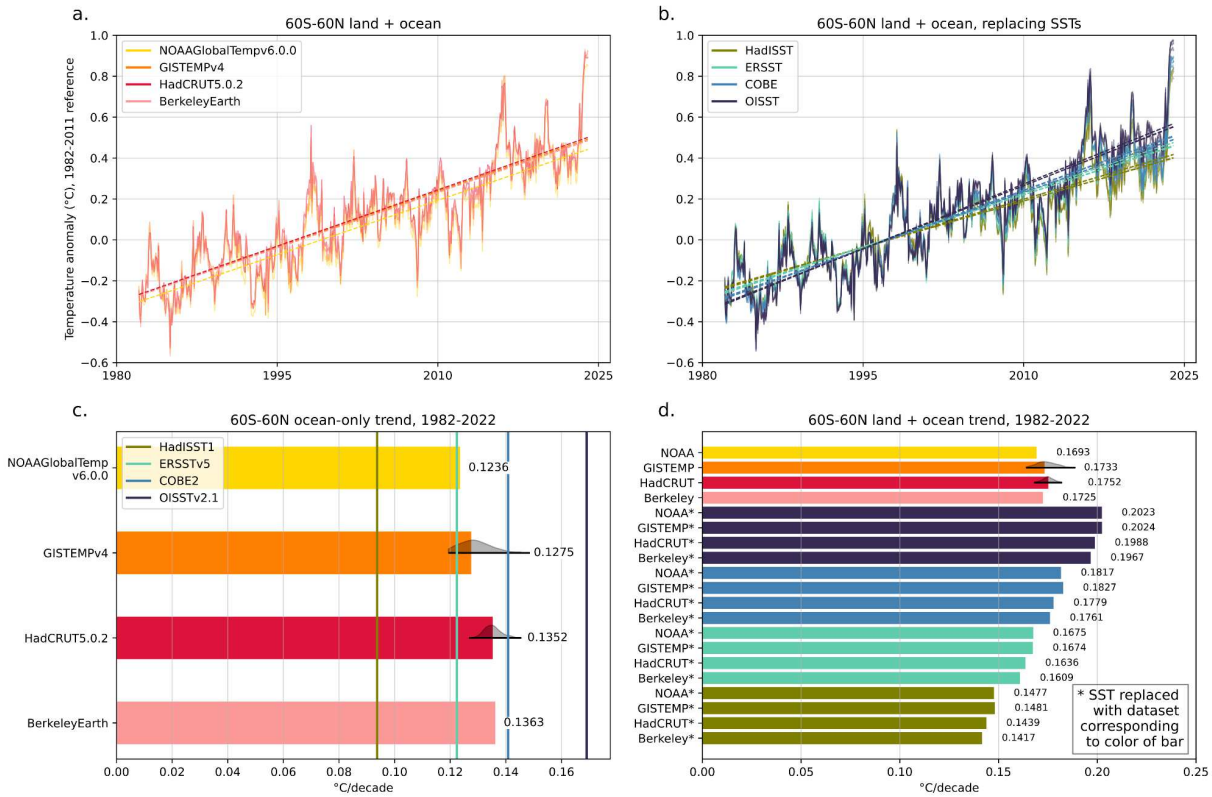


**Figure 2: Comparing global to SST-only datasets.** (a) Time Series of deseasonalized monthly average temperature anomalies, for all land and ocean between 60°S and 60°N. Data from the NOAAGlobalTempv6.0.0, GISTEMPv4, HadCRUT5.0.2, and BerkeleyEarth datasets are shown in yellow, orange, red, and pink respectively. Temperature anomalies, in °C, are against the 1982-2011 reference period. Dashed lines denote the linear trend between 1982-2022. (b) Same as panel (a) but for temperature fields constructed from combinations of SSTs from each SST dataset and land temperatures from each global temperature dataset. A total of 16 time series with corresponding trend lines are shown, color-coded by the SST field used. (c) Trends in the ocean-only 60°S-60°N component of global temperature datasets, from 1982-2022, in °C per decade. The SST dataset trends corresponding to Fig. 1, right panel are marked with vertical colored lines. The black error bars for HadCRUT5.0.2 and GISTEMPv4 span the range of trends calculated from the datasets' respective ensembles, with kernel density estimates shaded in gray to indicate the distribution of ensemble trends. (d) Trends in 60°S-60°N land and ocean temperature in the four global temperature datasets, and in each of these datasets if their SSTs were replaced with each of the four SST datasets. Error bars for HadCRUT5.0.2 and GISTEMPv4 similar to those in panel (c).

Considering only the ocean components of global temperature reconstructions, one finds that trends in SSTs from global temperature datasets are more similar to one another than SST

trends from SST datasets (Fig. 2b, c). The range in SST trends from the global temperature datasets is 0.124-0.136 °C decade$^{-1}$, compared to 0.093-0.169 °C decade$^{-1}$ in the SST datasets. SST trends from global temperature datasets fall in the middle of the range of those from the SST datasets, such that none of the SST trends from global temperature datasets resemble those of the highest and lowest end-members of the SST datasets, HadISST1 and OISSTv2.1. We also find a slight compensating effect between land and ocean temperature trends in the global temperature datasets, such that global temperature reconstructions with relatively higher SST trends tend to have relatively lower land temperature trends, and vice versa.

If one creates a "replacement ensemble" by combining SSTs from SST-only datasets with land temperatures from global temperature datasets, one finds a larger range of plausible global temperature trends than the global temperature datasets alone indicate. The full ensemble of 60°S-60°N trends produced by all possible SST replacements (Fig. 2d) spans from 0.1417 °C decade$^{-1}$ (GISTEMP land + HadISST ocean) to 0.2024 °C decade$^{-1}$ (BerkeleyEarth land + OISST ocean). The range in 60°S-60°N trends across all combinations of reconstructed SSTs and land temperature fields is thus ten times larger than the range given by global temperature datasets only (0.06 vs. 0.006 °C decade$^{-1}$). The uncertainties in SST trends (Fig. 2 c) and land+ocean trends (Fig. 2d) calculated from the 200-member HadCRUT5.0.2 and GISTEMPv4 ensembles, which sample the distribution of known uncertainties in its temperature reconstruction, are smaller than those calculated in the replacement ensemble.

| **Global temperature dataset** | **Sea surface temperature source** | **Land temperature source** |
|---|---|---|
| GISTEMPv4 | ERSSTv5 | GHCNv4 |
| HadCRUTv5.0.2 | HadSST4 | CRUTEM5 |
| NOAAGlobalTempv6.0.0 | ERSSTv5 | GHCNv4 |
| BerkeleyEarth | HadSST4 | Berkeley Earth monthly land temperature |

**Table 1: Dataset interdependency.** Ocean and land temperature data sources for all-surface global temperature datasets.

The SST reconstructions used in the four global temperature datasets are not independent: GISTEMPv4 and NOAAGlobalTempv6.0.0 derive their SST fields from ERSSTv5, while

HadCRUTv5.0.2 and BerkeleyEarth derive theirs from HadSST4 (Table 1, see also Methods). SST trends in reanalysis are also sensitive to the choice of dataset. MERRA2 reanalysis (Gelaro et al. 2017) uses OISST until March 2006, then the Operational SST and Sea Ice Analysis (OSTIA, Donlon et al. 2012). ERA5 (Hersbach et al. 2020) uses HadISST2.1.1.0 until August 2007, then OSTIA. Accordingly, MERRA2 exhibits a relatively high satellite-era trend of 0.1630 °C decade$^{-1}$, while ERA5 exhibits a lower trend of 0.1271 °C decade$^{-1}$ (Figure 1, Figure S3).

## 2.2 Recent record warmth

How does SST uncertainty impact our view of the extreme global temperatures observed in 2023? Although all four SST and all four global temperature datasets agree that the year 2023 was the warmest in the instrumental record, the choice of dataset impacts one's interpretation of how anomalous 2023 appears in the context of internal variability and global warming.

A stacked time series of annual temperatures is a common way to visualize and communicate the magnitude of global SSTs, for example with the University of Maine's "Climate Reanalyzer" tool (https://climatereanalyzer.org/clim/sst_daily/), which visualizes OISSTv2.1. This type of plot allows the viewer to compare recent temperatures to the average and variation of temperatures in some historical period, for example 1982-2011. Fig. 3 plots in this manner SST data from the reconstructions with the lowest and highest trends. With OISSTv2.1, average temperatures within a 1982-2011 reference period are 20.20 °C, with a standard deviation of 0.15 °C. Temperatures exceed the +/-2 standard deviation range from this reference period for more than half the months in the year for each year since 2014. With HadISST1, average temperatures in the reference period are 20.43 °C with a standard deviation of 0.11 °C, and only 6 years exceed the +/-2 standard deviation range from the reference period.

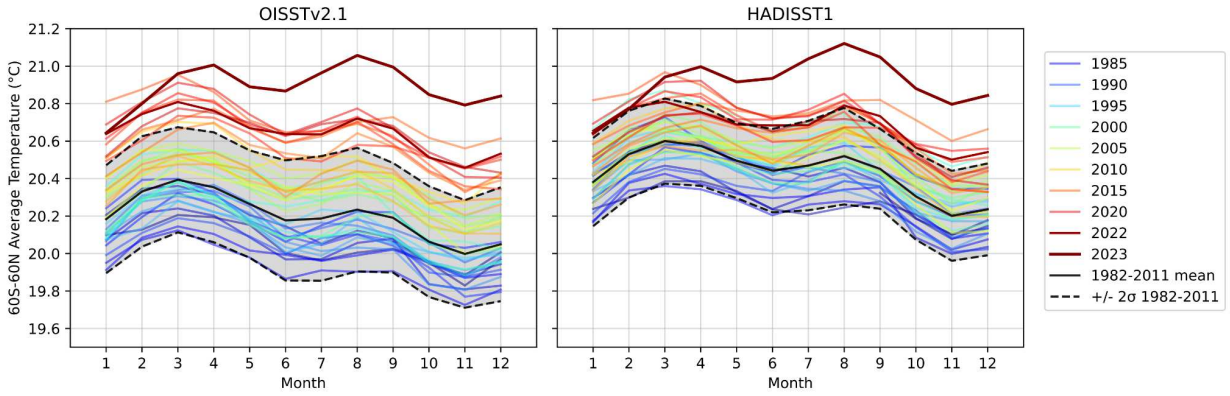


**Figure 3: Choice of dataset impacts interpretation of warming trend & recent anomalies.** Annual cycle of SST since 1982, with years coded by color (°C). Solid black line denotes 1982-2011 monthly mean, and the gray shaded area denotes 2 sigma confidence intervals. Contrast OISSTv2.1 (left panel) and HadISST1 (right panel).

Because spikes in annual global temperature are most likely to coincide with El Niño events in both observations and climate models (Raghuraman et al., 2024), we consider the 2023 temperature spike in the context of previous El Niño events. We calculate monthly record margins, the difference between the monthly temperature and the previous record for that month since 1982, to confirm that the largest record margins indeed occur during El Niño events for all SST datasets (Fig. 4). The highest record margins are during the 1997-1998, 2015-2016, and 2023-2024 El Niño events. Were 2023 record margins similar to previous El Niño events, or do they far exceed records broken by previous spikes? The answer to this question is sensitive to the choice of dataset. In OISSTv2.1, the ratio of the maximum 2023 to maximum 2015/2016 record margin is 0.28 °C to 0.30 °C, or 0.9. In HadISST1, this ratio is 0.27 °C to 0.15 °C, or 1.8.

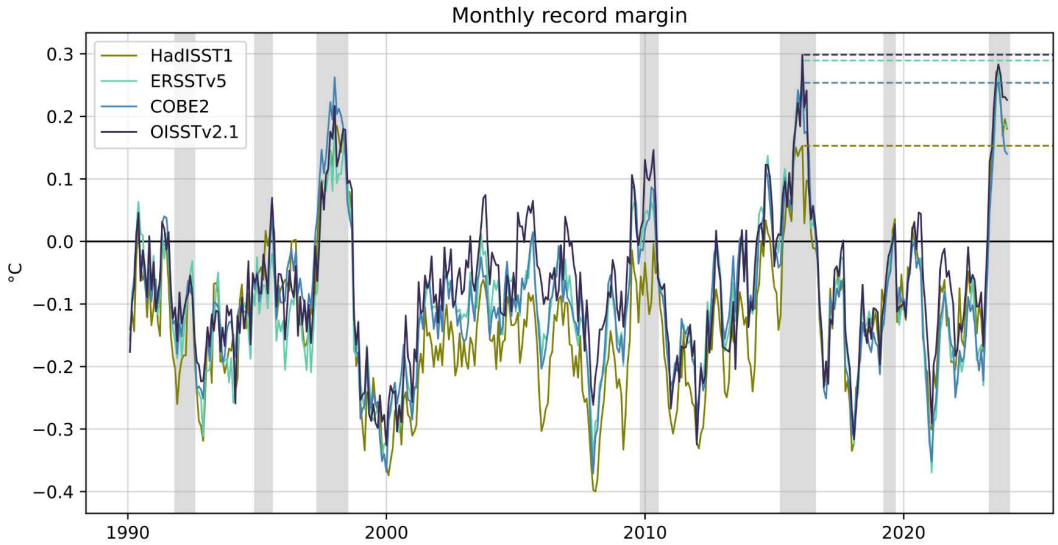


**Figure 4: El Niño record margin unprecedented for HadISST, but comparable to 2015-2016 for ERSST, COBE, & OISST.** Timeseries of monthly record margins, i.e. the difference between the current month's temperatures and the previous highest temperature since 1982, for 60°S-60°N average SSTs in four SST datasets (°C). Values above zero are record-breaking temperatures. Shaded in gray are El Niño events, defined here as periods during which the Oceanic Niño Index (3-month running mean) exceeds 0.5°C for a minimum of 5 consecutive overlapping months. The dashed lines indicate the highest 2015-2015 El Niño record margins for comparison to recent records.

## 3. Discussion

A straightforward analysis of publicly available SST and global temperature datasets reveals greater uncertainty in satellite-era global temperature trends than global temperature datasets and their published uncertainty estimates would suggest. At face value, published global temperature time series appear to agree very closely, especially in the satellite era (Fig. 2a). Reconstructions of global temperature anomalies consistently fall within one anothers' estimated confidence intervals (Huang et al., 2020; Jones, 2016; Lenssen et al., 2019; Lenssen et al., 2024; Morice et al., 2021; Rohde & Hausfather, 2020). This is generally interpreted to signal confidence in our knowledge of global temperature trends, with the implicit assumption that the range of technical choices made by the different data-providing institutions offers a reasonable estimate of the true uncertainty. However, although produced by different groups, the global temperature datasets use overlapping SST and land temperature input data (see Table 1 and

Methods). As noted by the IPCC (2021), global temperature reconstructions have "far fewer methodological degrees of freedom than implied by a straight count of the number of available estimates." Because global temperature datasets use SST fields with trends more similar than trends in SST-only reconstructions, each of which we also consider to be plausible estimates of reality, interpretations that infer confidence in satellite-era warming trends based on the similarity between these datasets are overconfident.

Our results further demonstrate that systematic uncertainty—uncertainty due to the choices made in producing spatially and temporally complete fields from incomplete and inhomogenous observations—dwarfs statistical and parametric uncertainty in reconstructed SST and global temperature trends. Statistical uncertainty associated with the linear fitting of trends is inconsequentially small relative to the spread between datasets (Figure 1). Parametric uncertainty estimates, which sample known uncertainties introduced by measurement error, bias correction, and aspects of the particular reconstruction methods used for each dataset (Lenssen et al., 2019, Lenssen et al., 2024), are also relatively small. The groups producing HadCRUT5 and GISTEMPv4 provide 200-member uncertainty ensembles, allowing for the explicit calculation of an ensemble of trends (BerkeleyEarth provides temperature uncertainties in the form of confidence intervals, and NOAAGlobalTempv6 uncertainties are not yet available). The range of global temperature trends across the HadCRUT5 and GISTEMPv4 ensembles is smaller than the broader uncertainty range obtained by explicitly considering the four publicly available SST datasets (Figure 2), implying that the ensembles sample fewer methodological degrees of freedom than the four SST datasets.

The first-order differences in satellite-era SST trends highlighted here are consequential for many types of analyses. The authors of this paper suspect that these differences are generally underappreciated, especially given the perception that satellite measurements correct known problems with pre-satellite-era temperature reconstructions (e.g. Chan & Huybers 2020). Unless temperature records can be reconciled, the impact of SST trend uncertainties on the scientific literature must be evaluated. Below, we offer a non-exhaustive list of some implications of these uncertainties.

- *Global temperature spike:* In section 3.2, we considered the example of 2023, so far the warmest calendar year in the instrumental record. Raghuraman et al. (2024) find that spikes in annual mean global temperatures are most likely to occur in cases of transitions from prolonged La Niña to El Niño conditions, implying that ENSO variability is a key driver of the 2023 warming spike. Rantanen & Laaksonen (2024), however, highlight that the September 2023 global temperature record margin of 0.5 °C is extremely unlikely to appear in large climate model ensembles, even with anthropogenic greenhouse gas forcing imposed. From a purely observational perspective, different SST reconstructions lend themselves to different interpretations of the 2023 warming spike. In OISSTv2.1, the fastest warming dataset, the 2023 temperature spike is in line with the last El Niño event. In HadISST1, the record margins during the 2023-2024 El Niño are the largest in recent decades, suggesting a relatively larger role for possible additional factors contributing to the warming spike, such as the reduction of aerosol cloud effects due to changes in shipping fuel regulations (Diamond, 2023; Manshausen et al., 2022; Yuan et al., 2022).

- *Transient climate response:* Global mean surface temperature reconstructions are used to estimate measures of climate sensitivity. Satellite-era temperature records are particularly salient for calculating transient climate response (TCR), defined as the temperature change from doubling $CO_2$ following a 1% per year increase, which helps us predict how the Earth System may warm in coming decades. TCR may be estimated using observations or a combination of observations and climate model simulations (e.g. Gillett et al., 2012; Morgenstern, 2024). Model error, unpredictable variability, and observational uncertainty affect our confidence in TCR estimates (Sherwood et al. 2020, Otto et al. 2013). Relevant observational variables include historical effective radiative forcing (ERF), whose uncertainties are dominated by aerosol ERF (Skeie et al. 2024), as well as historical temperatures and ocean heat content. How does observational uncertainty in the satellite-era global temperature record alone affect TCR uncertainty? Following Gregory & Forster (2008), for a climate system with steadily increasing forcing over a time period, and assuming the overall warming

trend is due to well-mixed greenhouse gases, the magnitude of climate change in response to a forcing $F$ (units W m$^{-2}$) can be approximated by the heat balance $\boldsymbol{N = F - \alpha \Delta T}$, where $N$ (W m$^{-2}$) is the net heat flux into the system, $\alpha$ is a climate sensitivity parameter (W m$^{-2}$ K$^{-1}$), and $\Delta$T is the change in temperature. Because $N$ is dominated by ocean heat uptake, we approximate $N = \kappa \Delta T$ where κ is an ocean heat uptake efficiency. Then, $F = \Delta T(\rho)$ where $\rho \stackrel{\text{def}}{=} \alpha + \kappa$. Thus ρ can be calculated from estimated values of  and observed values of $\Delta T$ over an historical period, and used to estimate the transient climate response. Taking the range of temperature combinations from Fig. 2d, but including temperatures poleward of 30°S and 30°N from the global temperature datasets, the end-member combinations of SST and global temperature datasets with the lowest and highest 1982-2022 trends are HadISST+NOAA (0.1748 °C/decade) and OISST+GISTEMP (.2266 °C/decade). Using a radiative forcing of 3.7 Wm$^{-2}$ corresponding to a doubling of $CO_2$, this yields a TCR range of 1.80-2.33 °C due to uncertainties in observed global temperatures alone.

- *SST-forced experiments:* Observational SST reconstructions are imposed as lower boundary conditions in atmospheric general circulation model (AGCM) experiments, which are important tools for studying phenomena such as tropical cyclones (Chan et al., 2021). Results in AGCM studies may be sensitive to the choice of SST dataset. For example, Flannaghan et al. (2014) found that forcing an AGCM with HadISST1 vs. Hurrell SSTs had substantial impacts on tropical atmospheric temperature trends. The standard temperature dataset used to force Atmospheric Model Intercomparison Project (AMIP) simulations is the Merged Hadley-NOAA/OI Sea Surface Temperature & Sea-Ice Concentration dataset (Hurrell et al., 2008), which is produced by merging HadISST1 and OISSTv2, which has a 1982-2021 (we use this range because 2022 temperatures are not yet available at the time of writing) trend of 0.110, on the lower end of the four SST datasets' range in trends over that same time period (these trends are 0.093, 0.123, 0.143, 0.169 for HadISST1, ERSSTv5, COBE2, and OISSTv2.1 respectively).

- *Earth's radiative imbalance:* Capturing the relationship between global temperatures and net TOA radiative imbalances is important for understanding the drivers of global warming. While comparisons between climate models and observations of Earth's Energy Imbalance (EEI) indicate that observed positive trends in EEI are highly unlikely to be a result of internal variability alone, both coupled and atmospheric climate models simulate changes in EEI at the lower end of the CERES confidence interval (Raghuraman et al., 2021) . Schmidt et al. (2023) propose using AGCM experiments to compare simulated EEI trends, in part because revised SST estimates are available; however, such analyses would be sensitive to the choice of SST dataset used, so should account for uncertainty across SST datasets. Furthermore, SST-forced AGCM experiments are used in climate model development and evaluation to test models' ability to reproduce the observed TOA energy imbalance; here, again, the choice of SST dataset matters, and may affect how coupled models ultimately simulate EEI trends.

Climate research centers devote significant resources to the intercomparison of model simulations. Observations of air temperatures have also received much attention, leading to multiple revisions of key observational estimates such as the Microwave Sounding Unit temperature measurements (Christy et al., 2000; Thorne et al. 2011, Po-Chedley & Fu 2012, Po-Chedley et al. 2015). While centers have devoted much time and care to producing best estimates of SSTs for a wide variety of applications, there has been a relative lack of attention directed towards documenting, understanding, and accounting for differences between these widely referenced observational datasets. As a result, the scientific community has yet to arrive at a widely known consensus on the rate of global warming in the satellite era.

Kent et al. (2017) stress the importance of maintaining thorough information about SST measurements, improving bias models, and improving access to information—recommendations we echo here. Making available the code used to produce SST reconstructions would facilitate explorations of uncertainties within datasets and comparisons between datasets. Most broadly, users of SST and global temperature reconstructions must be aware of these uncertainties in satellite-era trends. Global temperature data inform our perceptions of the societal risks and impacts of global warming. Such perceptions inform decisions with social consequences. The

uncertainties described here advise against using these data with undue assumptions of precision, orienting us instead toward greater humility and precaution.

## 4. Data and Methods

### 4.1 Sea Surface Temperature Datasets

We compare four sea surface temperature products that are spatially in-filled and regularly updated. We use the most current version of each dataset as of July 2024. Here we provide brief descriptions of each dataset with references where more detailed information can be found.

**HadISST1:** The UK Met Office Hadley Centre's sea ice and sea surface temperature (HadISST1, Rayner, 2003) provides sea surface temperatures and sea ice concentration on a 1 degree latitude-longitude grid beginning in 1870. SST input data include in situ observations from the Met Office Marine Data Bank and ICOADS, and satellite-based observations from Advanced Very High Resolution Radiometer (AVHRR) sensors. The SST reconstruction method uses reduced-space optimal interpolation (RSOI), with additional measures taken to account for long-term SST changes since the late 19$^{th}$ century. Gridded observations are re-introduced to restore variance on ~500 km scales. Both in situ and satellite observations are used from 1982 onwards. SSTs near sea ice are estimated using statistical relationships between SST and sea ice concentration. Unlike ERSSTv5, COBE2, and OISSTv2.1, no SSTs are given for sea ice areas. HadISST1 data were downloaded from https://hadleyserver.metoffice.gov.uk/hadisst/data/download.html.

**ERSSTv5:** The Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5, Huang et al., 2017) is produced by NOAA and provides SST data on a 2x2 degree latitude-longitude resolution from 1854 to the present. SST input data are primarily from ICOADS and Argo data above 5m. SST data are quality controlled and adjusted for mutual consistency. SST observations are separated into low- and high-frequency components, and high-frequency components are reconstructed using empirical orthogonal teleconnections. SSTs near sea ice are relaxed towards a representative freezing temperature of seawater (-1.8°C) based on sea ice concentration. Uncertainties for ERSSTv4, which used similar methods to ERSSTv5, are

described in Huang et al. (2016). ERSSTv5 data were downloaded from https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ncdc:C00927.

**COBE2:** Centennial In Situ Observation Based Estimates of the Variability of SST and Marine Meteorological Variables version 2 (COBE2, Hirahara et al., 2014) is a product of the Japan Meteorological Agency (JMA) and provides SST data from 1850-present on a 1x1 degree latitude-longitude grid. SST input data include ICOADS release 2.5 and additional SST data compiled by the JMA, plus AVHRR data to reconstruct variability in data-sparse regions. An empirical orthogonal function (EOF) based method is used to reconstruct SSTs. In areas with sea ice, SSTs are computed based on sea ice concentration such that areas around sea ice are close to the climatological salinity-dependent freezing point of seawater. COBE2 data were downloade from https://psl.noaa.gov/data/gridded/data.cobe2.html.

**OISSTv2.1:** The NOAA Daily Optimum Interpolation Sea Surface Temperature version 2.1 (OISSTv2.1, Huang et al., 2021) provides a record of SSTs beginning in September 1981 at 0.25 degree resolution. SST input data are from AVHRR and ICOADS3.0. After adjusting ship SSTs for biases and adjusting AVHRR measurements to match in situ SSTs at 0.2m depth, the input data are combined using optimum interpolation. When sea ice concentrations are higher than 35%, SSTs are replaced by the climatological salinity-dependent freezing point of seawater. Daily OISSTv2.1 data were downloaded from https://www.ncei.noaa.gov/data/sea-surface-temperature-optimum-interpolation/v2.1/access/avhrr/, and monthly means were calculated.

### 4.2 Global Temperature Datasets

We consider four publicly available and regularly updated global (land+ocean) surface temperature products, taking the most current version of each dataset as of July 2024. These global temperature datasets combine SSTs with near-surface (~2m) land temperatures.

**GISTEMPv4:** The Goddard Institute for Space Studies Surface Temperature product version 4 (GISTEMP v4, Hansen et al., 2010; Lenssen et al., 2019) provides global temperatures on a 2x2 grid from 1880 to the present. The input data are ERSSTv5 for SSTs, and NOAA Global Historical Climatology Network version 4 (GHCNv4) for land surface air temperatures. Lenssen et al. (2019) calculate uncertainties in GISTEMP global- and annual-mean temperature

anomalies by combining the ERSSTv4 uncertainty analysis in Huang et al. (2016), which considers EOT reconstruction uncertainties and parametric uncertainties, and land temperature uncertainties from sampling and homogenization. GISTEMP data were downloaded from https://data.giss.nasa.gov/gistemp/.

**HadCRUTv5.0.2:** The UK Met Office Hadley Centre/Climatic Research Unit global surface temperature data set version 5 (HadCRUT5, Morice et al., 2021) is a 5 degree global temperature dataset spanning from 1850 to the present. Input data are the UK Met Office Hadley Centre SST data set (HadSST4, Kennedy et al., 2019) for SSTs, and Climatic Research Unit temperature version 5 (CRUTEM5, Osborn et al., 2021) for surface air temperatures. The central estimate is the average of a 200 member ensemble. The ensemble members are intended to sample the distribution of systematic observational uncertainties and uncertainties in the analysis of observations. The infilled "HadCRUT5 analysis" ensemble members and ensemble mean were downloaded from https://www.metoffice.gov.uk/hadobs/hadcrut5/data/HadCRUT.5.0.2.0/download.html.

**NOAAGlobalTempv6.0.0:** The NOAA Global Temperature dataset (NOAAGlobalTemp, Huang et al., 2022) is provided on a 5x5 degree grid from 1850 to the present. Input data are ERSSTv5 for SSTs and GHCNv4 for land surface air temperatures. Version 6.0.0 uses an artificial neural network method to improve reconstruction of land surface temperatures. Uncertainties at global and local scales are described for NOAAGlobalTemp Version 5 in Huang et al., 2020, but have not yet been estimate for Version 6. NOAAGlobalTempv6.0.0 data were downloaded from https://www.ncei.noaa.gov/data/noaa-global-surface-temperature/v6/access/gridded/.

**BerkeleyEarth:** The Berkeley Earth global temperature dataset (Rohde & Hausfather, 2020) is provided on a 1x1 degree grid from 1850 to the present. Input data are HadSST4 for sea surface temperatures (personal communication with Zeke Hausfather) and the Berkeley Earth monthly land temperature data set for land surface air temperatures. Uncertainty estimates described in Rohde & Hausfather (2020) combine measurement & sampling uncertainty, coverage uncertainty, and bias uncertainty. Berkeley Earth global gridded data on a 1x1 latitude-longitude grid were downloaded from https://berkeleyearth.org/data/.

### 4.3 Processing

In this analysis we consider the areas between 60°S and 60°N. We focus on this area because the SST and global temperature datasets treat temperatures in areas with sea ice in different ways. Some datasets estimate the temperature of the sea surface directly under ice, while others estimate near surface air temperatures over ice. Because of amplified Arctic warming, the choice of how to treat sea ice can have a significant impact on global mean temperatures (Rohde & Hausfather, 2020). We therefore exclude polar regions in this analysis to allow us to more easily compare global temperature datasets to SST datasets, and to focus on the impact of uncertainties in SSTs. The area of the globe between 60S and 60N covers about 86% of the surface area of the Earth.

When plotting most time series and calculating trends, we use deseasonalized monthly average temperatures or temperature anomalies. There is a predictable annual cycle in global temperatures and global SSTs. The deseasonalized data are produced by subtracting the monthly climatology from the temperature each month, then adding the difference to the climatological average.

Decadal temperature trends were calculated using a linear least squares regression. We consider the time period beginning in 1982, during the satellite era. Unless otherwise noted, trends are calculated for 1982-2022. Although the values of temperature trends in a period of this duration (41 years) are sensitive to small shifts in start and end dates, the qualitative results of this analysis are not. Trends for combinations of land surface air temperature data from global temperature datasets and SST data from SST datasets are found by re-gridding the SST data to the global temperature dataset's grid, replacing global temperature dataset grid cells with the SST dataset data where they exist, then calculating the trend using the same method as above.

### Acknowledgments

We acknowledge and thank the groups that produced the reconstructions of SSTs and global temperatures used in this article.

### Data Availability Statement

All the data used in this analysis are publicly available as described in the methods section.

## References


Chan, D., Huybers, P., 2020. Systematic Differences in Bucket Sea Surface Temperatures Caused by Misclassification of Engine Room Intake Measurements. Journal of Climate 33, 7735–7753. https://doi.org/10.1175/JCLI-D-19-0972.1

Chan, D., Vecchi, G. A., Yang, W., & Huybers, P. (2021). Improved simulation of 19th- and 20th-century North Atlantic hurricane frequency after correcting historical sea surface temperatures. *Science Advances*, *7*(26), eabg6931. https://doi.org/10.1126/sciadv.abg6931

Donlon, C.J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E., Wimmer, W., 2012. The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system. Remote Sensing of Environment, Advanced Along Track Scanning Radiometer(AATSR) Special Issue 116, 140–158. https://doi.org/10.1016/j.rse.2010.10.017

Flannaghan, T. J., Fueglistaler, S., Held, I. M., Po-Chedley, S., Wyman, B., & Zhao, M. (2014). Tropical temperature trends in Atmospheric General Circulation Model simulations and the impact of uncertainties in observed SSTs. *Journal of Geophysical Research: Atmospheres*, *119*(23). https://doi.org/10.1002/2014JD022365

Freeman, E., Woodruff, S. D., Worley, S. J., Lubker, S. J., Kent, E. C., Angel, W. E., Berry, D. I., Brohan, P., Eastman, R., Gates, L., Gloeden, W., Ji, Z., Lawrimore, J., Rayner, N. A., Rosenhagen, G., & Smith, S. R. (2017). ICOADS Release 3.0: A major update to the historical marine climate record. *International Journal of Climatology*, *37*(5), 2211–2232. https://doi.org/10.1002/joc.4775

Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A., Darmenov, A., Bosilovich, M.G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., Silva, A.M. da, Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J.E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S.D., Sienkiewicz, M., Zhao, B., 2017. The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). Journal of Climate 30, 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Gillett, N. P., Arora, V. K., Flato, G. M., Scinocca, J. F., & von Salzen, K. (2012). Improved constraints on 21st-century warming derived using 160 years of temperature observations. *Geophysical Research Letters*, *39*(1). https://doi.org/10.1029/2011GL050226

Gregory, J. M., & Forster, P. M. (2008). Transient climate response estimated from radiative forcing and observed temperature change. *Journal of Geophysical Research: Atmospheres*, *113*(D23). https://doi.org/10.1029/2008JD010405

Hansen, J., Ruedy, R., Sato, M., & Lo, K. (2010). Global Surface Temperature Change. *Reviews of Geophysics*, *48*(4). https://doi.org/10.1029/2010RG000345

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., Thépaut, J.-N., 2020. The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society 146, 1999–2049. https://doi.org/10.1002/qj.3803

Hirahara, S., Ishii, M., & Fukuda, Y. (2014). Centennial-Scale Sea Surface Temperature Analysis and Its Uncertainty. *Journal of Climate*, *27*(1), 57–75. https://doi.org/10.1175/JCLI-D-12-00837.1

Huang, B., Liu, C., Banzon, V., Freeman, E., Graham, G., Hankins, B., Smith, T., & Zhang, H.-M. (2021). Improvements of the Daily Optimum Interpolation Sea Surface Temperature (DOISST) Version 2.1. *Journal of Climate*, *34*(8), 2923–2939. https://doi.org/10.1175/JCLI-D-20-0166.1

Huang, B., Menne, M. J., Boyer, T., Freeman, E., Gleason, B. E., Lawrimore, J. H., Liu, C., Rennie, J. J., Schreck, C. J., Sun, F., Vose, R., Williams, C. N., Yin, X., & Zhang, H.-M. (2020). Uncertainty Estimates for Sea Surface Temperature and Land Surface Air Temperature in NOAAGlobalTemp Version 5. *Journal of Climate*, *33*(4), 1351–1379. https://doi.org/10.1175/JCLI-D-19-0395.1

Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., Menne, M. J., Smith, T. M., Vose, R. S., & Zhang, H.-M. (2017). Extended Reconstructed Sea

Surface Temperature, Version 5 (ERSSTv5): Upgrades, Validations, and Intercomparisons. *Journal of Climate*, *30*(20), 8179–8205. https://doi.org/10.1175/JCLI-D-16-0836.1

Huang, B., Thorne, P. W., Smith, T. M., Liu, W., Lawrimore, J., Banzon, V. F., Zhang, H.-M., Peterson, T. C., & Menne, M. (2016). Further Exploring and Quantifying Uncertainties for Extended Reconstructed Sea Surface Temperature (ERSST) Version 4 (v4). *Journal of Climate*, *29*(9), 3119–3142. https://doi.org/10.1175/JCLI-D-15-0430.1

Huang, B., Yin, X., Menne, M. J., Vose, R., & Zhang, H.-M. (2022). Improvements to the Land Surface Air Temperature Reconstruction in NOAAGlobalTemp: An Artificial Neural Network Approach. *Artificial Intelligence for the Earth Systems*, *1*(4). https://doi.org/10.1175/AIES-D-22-0032.1

Hurrell, J. W., Hack, J. J., Shea, D., Caron, J. M., & Rosinski, J. (2008). A New Sea Surface Temperature and Sea Ice Boundary Dataset for the Community Atmosphere Model. *Journal of Climate*, *21*(19), 5145–5153. https://doi.org/10.1175/2008JCLI2292.1

IPCC. (2021). *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (p. 2391). Cambridge University Press. 10.1017/9781009157896

Jones, P. (2016). The reliability of global and hemispheric surface temperature records. *Advances in Atmospheric Sciences*, *33*(3), 269–282. https://doi.org/10.1007/s00376-015-5194-4

Kennedy, J. J., Rayner, N. A., Atkinson, C. P., & Killick, R. E. (2019). An Ensemble Data Set of Sea Surface Temperature Change From 1850: The Met Office Hadley Centre HadSST.4.0.0.0 Data Set. *Journal of Geophysical Research: Atmospheres*, *124*(14), 7719–7763. https://doi.org/10.1029/2018JD029867

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., & Zhang, H.-M. (2017). A Call for New Approaches to Quantifying Biases in Observations of Sea Surface Temperature. *Bulletin of the American Meteorological Society*, *98*(8), 1601–1616. https://doi.org/10.1175/BAMS-D-15-00251.1

Lenssen, N. J. L., Schmidt, G. A., Hansen, J. E., Menne, M. J., Persin, A., Ruedy, R., & Zyss, D. (2019). Improvements in the GISTEMP Uncertainty Model. *Journal of Geophysical Research: Atmospheres*, *124*(12), 6307–6326. https://doi.org/10.1029/2018JD029522

Lenssen, N., Schmidt, G.A., Hendrickson, M., Jacobs, P., Menne, M.J., Ruedy, R., 2024. A NASA GISTEMPv4 Observational Uncertainty Ensemble. JGR Atmospheres 129, e2023JD040179. https://doi.org/10.1029/2023JD040179

Morgenstern, O. (2024). Using historical temperature to constrain the climate sensitivity, the transient climate response, and aerosol-induced cooling. *Atmospheric Chemistry and Physics*, *24*(14), 8105–8123. https://doi.org/10.5194/acp-24-8105-2024

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., Dunn, R. J. H., Osborn, T. J., Jones, P. D., & Simpson, I. R. (2021). An Updated Assessment of Near-Surface Temperature Change From 1850: The HadCRUT5 Data Set. *Journal of Geophysical Research: Atmospheres*, *126*(3), e2019JD032361. https://doi.org/10.1029/2019JD032361

Osborn, T. J., Jones, P. D., Lister, D. H., Morice, C. P., Simpson, I. R., Winn, J. P., Hogan, E., & Harris, I. C. (2021). Land Surface Air Temperature Variations Across the Globe Updated to 2019: The CRUTEM5 Data Set. *Journal of Geophysical Research: Atmospheres*, *126*(2), e2019JD032352. https://doi.org/10.1029/2019JD032352

Otto, A., Otto, F.E.L., Boucher, O., Church, J., Hegerl, G., Forster, P.M., Gillett, N.P., Gregory, J., Johnson, G.C., Knutti, R., Lewis, N., Lohmann, U., Marotzke, J., Myhre, G., Shindell, D., Stevens, B., Allen, M.R., 2013. Energy budget constraints on climate response. Nature Geosci 6, 415–416. https://doi.org/10.1038/ngeo1836

Po-Chedley, S., Fu, Q., 2012. Discrepancies in tropical upper tropospheric warming between atmospheric circulation models and satellites. Environ. Res. Lett. 7, 044018. https://doi.org/10.1088/1748-9326/7/4/044018

Po-Chedley, S., Thorsen, T.J., Fu, Q., 2015. Removing Diurnal Cycle Contamination in Satellite-Derived Tropospheric Temperatures: Understanding Tropical Tropospheric Trend Discrepancies. Journal of Climate 28, 2274–2290. https://doi.org/10.1175/JCLI-D-13-00767.1

Raghuraman, S.P., Paynter, D., Ramaswamy, V., 2021. Anthropogenic forcing and response yield observed positive trend in Earth's energy imbalance. Nat Commun 12, 4577. https://doi.org/10.1038/s41467-021-24544-4

Raghuraman, S. P., Soden, B., Clement, A., Vecchi, G., Menemenlis, S., & Yang, W. (2024). The 2023 global warming spike was driven by El Niño/Southern Oscillation. *EGUsphere*, 1–12. https://doi.org/10.5194/egusphere-2024-1937

Rantanen, M., & Laaksonen, A. (2024). The jump in global temperatures in September 2023 is extremely unlikely due to internal climate variability alone. *Npj Climate and Atmospheric Science*, *7*(1), Article 1. https://doi.org/10.1038/s41612-024-00582-9

Rayner, N. A. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, *108*(D14), 4407. https://doi.org/10.1029/2002JD002670

Rohde, R. A., & Hausfather, Z. (2020). The Berkeley Earth Land/Ocean Temperature Record. *Earth System Science Data*, *12*(4), 3469–3479. https://doi.org/10.5194/essd-12-3469-2020

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., Ziehn, T., 2023. CERESMIP: a climate modeling protocol to investigate recent trends in the Earth's Energy Imbalance. Front. Clim. 5. https://doi.org/10.3389/fclim.2023.1202161

Sherwood, S.C., Webb, M.J., Annan, J.D., Armour, K.C., Forster, P.M., Hargreaves, J.C., Hegerl, G., Klein, S.A., Marvel, K.D., Rohling, E.J., Watanabe, M., Andrews, T., Braconnot, P., Bretherton, C.S., Foster, G.L., Hausfather, Z., von der Heydt, A.S., Knutti, R., Mauritsen, T., Norris, J.R., Proistosescu, C., Rugenstein, M., Schmidt, G.A., Tokarska, K.B., Zelinka, M.D., 2020. An Assessment of Earth's Climate Sensitivity Using Multiple Lines of Evidence. Reviews of Geophysics 58, e2019RG000678. https://doi.org/10.1029/2019RG000678

Skeie, R.B., Aldrin, M., Berntsen, T.K., Holden, M., Huseby, R.B., Myhre, G., Storelvmo, T., 2024. The aerosol pathway is crucial for observationally constraining climate sensitivity

and anthropogenic forcing. Earth System Dynamics 15, 1435–1458. https://doi.org/10.5194/esd-15-1435-2024

Thorne, P.W., Lanzante, J.R., Peterson, T.C., Seidel, D.J., Shine, K.P., 2011. Tropospheric temperature trends: history of an ongoing controversy. WIREs Climate Change 2, 66–88. https://doi.org/10.1002/wcc.80

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SSTanalysissupplement.pdf