Global surface temperatures
(Chapter in Energy and Climate Change)

Elizabeth C. Kent, Duo Chan, Richard C. Cornes, Pia Nielsen-Englyst,
Jules B. Kajtar, Christopher J. Merchant, Peter W. Thorne, Emily J. Wallis and R. Iestyn Woolway

8 April 2024

ECK, RCC, JBK affiliation: National Oceanography Centre, Southampton, UK

PWT affiliation: National University of Ireland Maynooth: Maynooth, IE

DC affiliation: University of Southampton, Southampton, UK

PE affiliation: Danish Meteorological Institute, Copenhagen, DK

CJM affiliation: Department of Meteorology and National Centre for Earth Observation, University of Reading, UK

EJW affiliation: Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich, UK

RIW affiliation: School of Ocean Sciences, Bangor University, Bangor, UK

Abstract

The Sixth Assessment Report of the Intergovernmental Panel on Climate Change stated that ”it is unequivocal that human influence has warmed the atmosphere, ocean and land”. One key piece of evidence for this is the global average of the instrumental record of surface temperature. A change in surface temperature of 1.5°C relative to a reference “pre-industrial” period is the measure internationally agreed to mark the transition to a climate where dangerous impacts become common. This chapter will discuss the observational evidence that ensures a reliable record of the changing temperature, some of the challenges that have been overcome, and some that remain in improving that record and more fully understanding its uncertainty.

Keywords: temperature, observations, uncertainty, climate, extremes.

1. Introduction

Analyses of global surface temperatures have a rich heritage (Hawkins and Jones, 2013), with the first estimate of a globally averaged surface temperature evolution dating from over 80 years ago (Callendar, 1938). Over time methods, data sources, and computational capabilities have evolved and improved. But even these pioneering efforts at global surface temperature analyses stack up well against modern day estimates (Hawkins and Jones, 2013).

The unequivocal statement of Working Group I to the Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report (IPCC, 2023a) that humans have warmed each of Earth’s domains continued a sequence of assessments that have made similar statements with increasing confidence. This confidence arises both from an improved understanding of how to construct climate records with quantified uncertainty, and more worryingly from the rapid increases in temperatures seen since the first IPCC assessment report in 1990. Global mean surface temperature (GMST) is a commonly used indicator of climate change, and keeping its change from an early 20th century benchmark to below 2°C (and ideally 1.5°C) has been designated by the United Nations Framework Convention on Climate Change as the measure which, if achieved, will avoid the worst impacts of climate change. However, it is becoming increasingly clear that there is no safe limit and conditions will become more and more unsafe as concentrations of greenhouse gases increase in the atmosphere and surface temperatures continue to increase across the Earth.

Global surface temperatures are a family of related variables (Merchant et al., 2013) which are designated as an Essential Climate Variable (ECV) by the Global Climate Observing System (Bojinski et al., 2014). Over the ocean there is a choice of generating estimates of GMST using near surface temperature within the ocean (SST) or the atmosphere over the ocean (Marine Air Temperature, MAT hereafter). Historically it had been assumed that anomalies (differences from averages over a chosen reference period usually of 30-years length) of SST and MAT show similar variability and trends at large space and time scales (Folland et al., 1984; Jones et al., 1988). It was further argued that global surface temperature products, and hence global mean surface temperature (GMST) should be constructed using SST. GMST has been derived by combining the air temperature measured near the surface at fixed meteorological sites typically known as "stations" over land or ice (land surface air temperatures, LSAT hereafter), with SST measured aboard ships and also more recently at a range of moored buoys and drifting autonomous platforms (together usually referred to as "platforms"). In contrast, model-based projections generally use surface air temperature everywhere (Global Surface Air Temperature, GSAT) which would be equivalent to combining observed LSAT and MAT. However, with increasingly rapid changes in surface temperature in recent years, along with a focus on regional changes, the assumption of equivalence has been challenged as discussed further later in this chapter.

2. Observations of surface temperature

By necessity the climate record for every ECV is built from the observations we have available, which are a subset of the observations that were made. Motivations for making temperature measurements vary enormously and include: for personal or scientific interest, providing economic advantage for ship routing or agriculture and, more recently, as part of a designed network supporting numerical weather prediction and climate monitoring and climate services. For the ocean a summary of the evolving observing system can be found in Kent and Kennedy (2021) and over land in (Parker, 1994). The most obvious difference between the observations over land and ocean is the predominance of mobile marine platforms: mainly ships and drifting buoys. Only a few marine observations come from fixed stations, in the past from Ocean Weather Stations established during or after World War II and existing until the 1990s, and more recently a fairly sparse network of moored buoys concentrated in the tropics and near coasts (Centurioni et al., 2019). Over land many of the longest-established stations originated at astronomical observatories. However, with the development of the National Meteorological Services in the mid-nineteenth century, the role of record-keeping moved away from these observatories and towards dedicated meteorological enclosures that were managed in a systematic manner. This led to improved consistency across the network in terms of instrumentation and observation schedule.

Figure 1 shows the changing coverage of observations on a 5° latitude-longitude grid from the global gridded product HadCRUT5 (Morice et al., 2021). This gives an optimistic picture of sampling, as any observation in a monthly 5° grid-cell counts toward this coverage. Smaller grid-cells will give lower coverage estimates, and for the ocean there may be insufficient observations to calculate a reliable value for a particular grid-cell. At the start of the record coverage is low, around 20% of the 5° grid-cells are sampled in the 1850s, mostly over the Atlantic and Indian Oceans and the European continent (Figure 1a). By the 1900s North America and most of Europe are well sampled, the Suez Canal has opened, and coverage has doubled to 40% (Figure 1b). By the 1950s global coverage has increased to 60% (Figure 1c) but high latitudes remain largely unsampled, as well as parts of Africa, Asia, South America, Australia and oceans outside the main shipping lanes. Modern coverage (Figure 1d) shows poor sampling remains at high latitudes (Figure 2) but most of the oceans north of about 40 °S are now sampled to some extent. Africa is the region with the lowest coverage. Figure 1e shows that ocean coverage varies more unevenly over time than for the land with noticeable dips in coverage after the 1850s and during the two world wars (1914 to 1918 and 1939 to 1945). More recently there has been a decrease in coverage for MAT (Kent et al., 2019), and the COVID pandemic also affected coverage of ocean observations (Boyer et al., 2023).

The character of observations made over land and ocean is also different, beyond obvious differences between fixed stations and mobile platforms. Marine observations are typically based on reports of conditions over a short period (typically either 10-minutes or instantaneous spot-measurements) made several times a day. Some observations over land are made in a similar manner but also frequently are either daily or monthly summaries. These summaries can be based on daily maximum and minimum temperatures or the average of a number of measurements across the day (Menne et al., 2012). Where observations are summaries of daily maximum and minimum temperatures, at what local hour these are made can matter (Karl et al., 1986). Prior to the 1950s most marine observations were made at the end of each 4-hour watch and have been converted to a local standard time and then to UTC. As observations started to be transmitted in near-real-time to numerical weather prediction centres for use in forecasts there was a transition to reporting in UTC, typically every 3 or 6 hours.

Many of the still-existing observations that have been made historically are not yet available for inclusion in global analyses. Recently digitised data from ships have been shown to have improved the reconstruction of historical storms (Hawkins et al., 2023) and Noone et al. (2020) made recommendations for prioritisation of records according to quality, content and other measures of their value. Many data remain in paper form in archives around the world, some have been converted to images but the information has not yet been extracted in a useable form from those images (Brönnimann et al., 2019). Identification and cataloguing of these records in archives and other stores around the world are key skills requiring expert knowledge both of the archives and of the science requirements for data and metadata (information about the observing platform or station and how the measurement was made). These data include observations prior to 1850 that would be valuable for extending the record further back in time (Brönnimann et al., 2018). Data rescue efforts tend to be driven by particular opportunities, or by national or institutional priorities, and in particular for marine data rescue activities are not currently fully integrated with the international archive used to construct surface temperature estimates for the ocean. Improved data infrastructure providing an end-to-end system linking images to the recovered observations and to documentation and metadata describing the observations and protocols would be hugely beneficial to recovering data and improving the historical records. Such a system could standardize approaches to interpreting location, date and time information and the conversion of units and extraction of information from coded parameters and improve the consistency of input data to analyses. Data extraction and processing should not be considered to be a one-time activity. Issues with data are continually being identified, and availability of the raw data and information on formats and methods is key to correcting past problems and improving climate records.

3. Observation characteristics, identification and reduction of data artefacts

3.1 Overview

Artefacts in the observations arise for many reasons and need to be identified and removed or their effects reduced before uncertainties are estimated and the data can be used in gridded analyses. Because data systems have evolved alongside the observing networks there are some data sources, periods, and regions where observations must be excluded, or adjusted. Quality control is challenging in data sparse regions and periods, and, despite recent progress, some analyses still show artefacts that are clear to the eye, such as "ship tracks" for marine data or "bulls-eyes" for land. Methods for homogenisation developed differently for the land and ocean domains, driven by the character of the data. For the land often only temperature may be available, and the record may be a monthly or daily summary. This has led to the development of sophisticated statistical methods to compare timeseries from neighbouring stations. In contrast the marine record, at least from ships, is multi-variate, and platforms are moving. This has led to the development of physically-based models making use of co-incident variables such as wind and cloud cover.

More recently this clear distinction between methods for identifying biases has blurred. Physics-based models have been developed for land data (Auchmann and Brönnimann, 2012; Wallis et al., 2024), in a similar manner to those used to reduce diurnal-heating biases in marine data (Berry et al., 2004). Such approaches are being developed to adjust for biases in early air temperature measurements over land where thermometers were screened from the sun using non-standard methods (Wallis et al., 2024). Statistical models have been developed for land (Menne and Williams, 2009; Chan et al., 2024) and marine data (Chan and Huybers, 2019), and co-evaluation methods using coastal data can diagnose biases (Cowtan et al., 2018; Chan et al., 2023).

3.2 Observations over land

There are several archives of LSAT station data used in the construction of GMST records. The CRUTEM station database (Osborn et al., 2021) comprises nationally-homogenised monthly records obtained from global or regional data collections and records from national services. The NOAA Global Historical Climatology Network GHCN has monthly (Menne et al., 2018) and daily versions (Menne et al., 2012). Other compilations of station data (e.g. Rohde et al., 2013; Xu et al., 2017) supplement these archives with additional station data.

LSAT station series may contain inhomogeneities that need to be identified and potentially corrected before the data can be reliably used for assessing changes or variations in temperature. These inhomogeneities can arise from many reasons, most notably: station relocations, instrument changes, observer and observing protocol changes, increasing automation, time-of-observation biases, microclimate and exposure changes and urbanisation (Trewin, 2010). These data artefacts can cause step-like changes in the data series, spurious trends or a combination of both of these features, and can lead to trends in the data that are much larger than the natural and/or anthropogenic signal (Figure 3b).

The homogeneity assessment typically involves two stages. Firstly, the identification of data artefacts; and secondly the adjustment of those inhomogeneities if they are deemed to have caused a significant artificial change in the data. Most early homogenization methods focused on changes in the data averages (e.g. Maronna and Yohai, 1978). However, in many cases the inhomogeneities may manifest as a combination of changes in the average alongside changes in the extreme values (Della-Marta and Wanner, 2006). This can lead, for example, to unreliable conclusions being reached regarding changes in the number of hot days recorded at a station. Most homogeneity assessments therefore now seek to determine changes in these extremes alongside changes in the averages.

The determination of absolute homogeneity of a given station series is difficult to achieve in practice, and therefore relative assessments are generally made. This takes the form of comparing a candidate series against neighboring series from a similar climatic zone. This can either be done by averaging the neighboring series to construct a single comparison series (e.g., Menne and Williams, 2005), or using a pairwise comparison whereby the candidate series is assessed against each neighbouring series in turn (e.g., Caussinus and Mestre, 2004; Menne and Williams, 2009). This latter procedure can be particularly beneficial when the neighboring series themselves contain inhomogeneities. Assuming that the inhomogeneity does not occur simultaneously across the network, an inhomogeneity in the candidate series can be assessed against segments in the reference series that are considered homogeneous. Fig. 3 shows an example of pair-wise homogenization when applied to an Angolan station.

An alternative to relative homogenization, which is increasingly being used to homogenize air temperature measurements over land, is the use of physics-based models. Physics-based models estimate adjustments for known inhomogeneities based on the physical relationships between the size of an inhomogeneity and the conditions at the time of observation. Their use is not reliant on the availability of neighbouring reference series and, as a result, they are particularly useful to address inhomogeneities which occur simultaneously across observing networks or in data sparse regions or periods. One example of an inhomogeneity in LSAT records which has been addressed using this method is the exposure bias. Exposure biases are a pervasive network-wide inhomogeneity present in LSAT records due to changes in the way thermometers have been protected from solar radiation and other environmental elements over time. Since the early-twentieth century, the majority of observing networks have used slatted screens to protect thermometers; previously a range of often insufficient methods were used (Parker, 1994). As a result, early LSAT observations are often biased too warm, particularly in the summer months, due to sunlight reaching the thermometer. The availability of near-simultaneous, co-located observations (known as parallel measurements) from different thermometer exposures allows a physical understanding of the variables which influence the bias to be built and enables the development and evaluation of physics-based models to address the bias. Physical models were first used at individual stations or on a national scale: for example, Auchmann and Brönnimann (2012) for Basel, Switzerland and Brunet et al. (2011) for Spain. More recently, however, the approach has been extended to address exposure biases across the mid-latitudes. Wallis et al. (2024) derived physics-based adjustments to minimize the bias arising from three commonly-used early thermometer exposures at mid-latitude stations: open exposures such as Glaisher stands; wall-mounted exposures and closed exposures such as Wild's cylindrical shield and screen. Each model was developed based on an analysis of the relationship between the magnitude of the monthly mean exposure bias observed in parallel measurements and three temperature and solar radiation variables. The impact of the exposure bias adjustments derived by Wallis et al. (2024) on an extended version of the CRUTEM5 gridded global temperature dataset (Osborn et al., 2021) are shown in Figure 4.

A particularly important component of LSAT data homogenization is the incorporation of metadata (information about the measurement or station) into the homogeneity assessment (Williams et al., 2012). For example, an identified break-point in a series may be linked to a documented station relocation and this adds credibility to the statistically-determined inhomogeneity (Chan et al., 2024). However, the metadata record for a given station is unlikely to be complete and therefore the combination of statistical detection/correction procedures along with metadata analyses provides an optimal homogenization procedure. This can be done relatively easily when the homogenization is performed manually, but this approach is limited to analysing a small number of sites. A large network of stations requires an automated approach but the incorporation of metadata to these assessments is currently limited. Advances in artificial intelligence techniques may pave the way for performing large-scale, metadata-included assessments.

Uncertainties in LSAT station data, and for gridded datasets derived from them are considered in detail by Brohan et al. (2006) and Lenssen et al. (2019) who consider uncertainties in measurement, in the bias adjustment applied, arising from gaps in the station timeseries, and in the climatology used as a reference for the calculation of anomalies.

There are five principal global terrestrial datasets (Osborn et al., 2021; Menne et al., 2018; Lenssen et al., 2019; Rohde and Hausfather, 2020; Xu et al., 2017) and these datasets take different approaches to station selection, quality control and homogenization, interpolation and area averaging. Other global datasets that are more experimental in nature are also available (e.g. Gillespie et al., 2022), but are not typically used in climate assessments. The use of these different approaches highlights the degree of sensitivity of the results to methodological choices. However, it should be noted that there is often an overlap between the methods used to construct the datasets and particularly in the source data that are used. The different estimates are generally in agreement throughout the record, with larger differences earlier in the record when observations were sparser and there was little standardisation of temperature scales or observing practices. Around the turn of the 20th century efforts began to standardise temperature scales and methods of observation, so earlier in the record uncertainties will be larger due to differences in instrumentation and measurement protocols. Nonetheless, differences between estimates in the global mean are substantively smaller than the long-term warming trend common to all estimates.

3.3 Observations over the ocean -sea-surface temperature

The largest collation of historical sea surface temperature (SST) observations is the International Comprehensive Ocean-Atmosphere Data Set (ICOADS, Freeman et al., 2017) which is the main input source for all of the marine gridded analyses. ICOADS is sometimes supplemented with other sources but remains the primary input source and provides a measure of continuity between the different analyses. Ship reports are multi-variate and may contain rich metadata giving details about the platform and the measurements.

Unfortunately, past data management restrictions can mean that much of this richness has been lost, and in some parts of the record most of the reports don't even have ship identifiers (Carella et al., 2017).

Aside from inhomogeneities in sampling, the largest source of uncertainty in current GMST estimates comes from the biases inherent in observing historical SST. Changes to the marine observing system relevant to surface temperatures are reviewed by Kent and Kennedy (2021) and for SST in particular by Kennedy (2014). In summary the earliest observations of SST from the early 19th century were probably made from samples of water collected in wooden buckets, transitioning over time to canvas buckets for ease of use. Later the issue of evaporative cooling from canvas buckets was tackled by improved bucket designs tested in wind tunnels (e.g. Ashford, 1948). For ships with engines a pumped seawater system was required for cooling, opening up the opportunity to report the temperature of the pumped system as SST. Such "engine-room intake" temperatures showed an overall warm bias and large scatter (Kent and Kennedy, 2021), but newer technology, including "hull contact sensors" means that more accurate measurements are now made (Kent and Berry, 2005). Because of differences in mean offsets between measurements of different types that can also depend on environmental conditions (Kent et al., 2017), knowing the measurement method is critical to developing adjustment for biases and estimating their uncertainty (e.g. Carella et al., 2018; Chan and Huybers, 2021).

The European Space Agency's Climate Change Initiative for SST (Merchant et al., 2019) carefully reprocessed satellite thermal measurements to extract surface temperature information from trillions of radiances from around twenty satellite missions. These measurements have a far greater density than *in situ* measurements. With care for cross-satellite harmonisation and methods of cloud screening and retrieval, the long-term stability of a satellite record relative to largely independent *in situ* data is suitable for climate analysis. Such records form a stable climate record covering the past forty years of sustained satellite observations and can be used to describe the spatio-temporal modes of variation of SST at resolution of less than 1 degree latitude-longitude at daily timesteps (e.g. Reynolds et al., 2013).

Bias adjustment for SST is typically applied to the gridded monthly fields rather than to the individual observations (e.g. Kennedy et al., 2019; Huang et al., 2017). Physically-based patterns of bias expected for measurements made using buckets (Bottomley et al., 1990; Smith and Reynolds, 2002) are weighted to give broad agreement with global averages of NMAT (Kent et al., 2013) either over the entire ship record (ERSSTv5, Huang et al., 2017) or for the period 1850 to 1920 (HadSST4, Kennedy et al., 2019). However, methods have been developed that can be applied per-ship that can account for conditions at the time of observation (Carella, 2017). These are challenging to apply but are likely to produce better results at regional scales. Statistical relative homogenisation has presently only been applied to collections of ships thought to have similar characteristics (Chan and Huybers, 2019) but could also be applied to individual ships. These techniques have not yet been applied to the data products that are currently available, doing so would enable the generation of estimates of historical SST independent of NMAT.

Centennial scale datasets of historical SST must use the available sometimes sparse sampling, including with large areas of ocean that in a given year are void of any measurements. In this context, two approaches are available: develop products capturing SST variations only where there have been data (e.g. Kennedy et al., 2019); or develop reconstructions on a consistent spatial gridding over times, with gap-filling (e.g. Huang et al., 2017; Hirahara et al., 2014). Approaches to gap-filling may use spatial modes of co-variability of SST (e.g. Reynolds et al., 2013; Rayner et al., 2003) or parameterize the spatial structure of the co-variability using simple geometry (e.g. Karspeck et al., 2011; Hirahara et al., 2014) to extend the coverage to unobserved grid-cells.

The three main global SST datasets (HadSST4, ERSSTv5 and COBE-SST2) are compared by (Kennedy et al., 2019) who concluded that differences among the global trends over the period 1990 to 2018 in the three data sets is not larger than the estimated uncertainty. Recent trends, from 2000 to 2018, also agree within their uncertainty ranges, and are all consistent with instrumentally homogeneous records for this period from moored buoys, Argo floats and satellites, as previously shown by Hausfather et al. (2017). Nevertheless they note several regions and periods of difference. Trends since the 1960s are greater in HadSST4 than ERSSTv5 which Kennedy et al. (2019) attribute to the use of NMAT as a reference for bias adjustments in ERSSTv5. SST during World War 2 remains very uncertain and likely too warm (Chan and Huybers, 2021) and it also seems likely that SST, and NMAT, is too cool in the years around the start of the Twentieth Century (Cowtan et al., 2018; Chan et al., 2023).

3.4 Observations over the ocean -marine air temperature

The observing of marine air temperature (MAT) from instruments onboard ships has a longer history than the recording of SST values. The earliest documented MAT reading was taken by Edmond Halley on board HMS Paramour on its voyage to the south Atlantic in 1699. Ships have continued to record air temperature measurements and this continues today thanks mainly to the Voluntary Observing Ship (VOS) network. Marine air temperature measurements tend to require fewer bias adjustments than SST. However, there is a general lower density of such observations than SST after about 1900. The issues associated with adjustment for biases in the nighttime only reports are summarised by Kent and Kennedy (2021) and are similar to those for SST described earlier. The main adjustment that needs to be applied to MAT data is for variations due to different observing heights. The height of the observing platform varies between different ships and as ships have increased in length, the height of the observing platform has also increased on average (Kent et al., 2007; Kent et al., 2013). As a result, the MAT observations need to be adjusted from the respective observing height to a common reference height, which is typically 10m. In addition, a warm bias typically occurs in MAT observations due to the warming of the ship superstructure throughout the daytime. The traditional approach to dealing with that bias is to exclude values recorded during the daytime and to construct NMAT datasets. However, recent developments have been made to derive adjustments for the diurnal heating bias (Cropper et al., 2023; Berry et al., 2004), which will allow these values to be used in global temperature assessments (Morice et al., 2024). Other biases in MAT data are restricted to certain regions or periods. Notably a warm bias exists in the observations taken onboard certain ships passing through the Suez Canal in the late 19th Century and in measurements taken during World War II, due to the method of recording temperature measurements under cover at night to avoid drawing attention to the ship (Kent et al., 2013).

As with the SST data, the individual NMAT observations are aggregated to form gridded datasets. There are two such datasets currently being regularly updated: CLASSnmat (Cornes et al., 2020) and UAHNMAT (Junod and Christy, 2020). Both NMAT datasets are available on a 5° latitude-longitude monthly grid without infilling. Because observations of MAT are mostly from ships, and the coverage of ship observations has declined in recent years (e.g. Berry and Kent, 2017; Kent et al., 2019) the uncertainty in NMAT gridded datasets has been increasing since the 1990s.

Surface temperatures can also be retrieved from satellites and the EUSTACE project (Rayner et al., 2020) estimated air temperature over all surfaces (ocean, land, cryosphere and lakes) from relationships derived from *in situ* data between the surface temperature and the temperature of the air above. The air temperature in the few metres above the surface has a minimal effect on the radiances measured at the top of the atmosphere, and only indirect air temperature estimates from satellites have been made (e.g. Yu and Jin, 2018), although ground-based radiometric measurements of air temperature can be made (e.g. Hanafin and Minnett, 2005).

3.5 Observations for the cryosphere

The cryosphere is highly sensitive to warming temperatures so it is critical to monitor the temperature of the cryosphere to understand and predict the local as well as global effects of climate change. Historically, near-surface air temperatures of the cryosphere have been measured from automatic weather stations located on the ice sheets, and from ice drifting buoys, stations and ships frozen into the sea ice. The measurements usually provide near-surface air temperatures at a height of 1-3 meters above the surface, depending on snow accumulation, drift and melt.

For the Greenland Ice Sheet the largest collation of surface temperature observations is obtained through the Programme for Monitoring of the Greenland Ice Sheet (PROMICE) and the Greenland Climate Network (GC-Net) operated by the Geological Survey of Denmark and Greenland (GEUS) in collaboration with the Technical University of Denmark and Asiaq. GC-Net was the first year-round monitoring programme. It was initiated in 1990 and extended in 1995 with sites mainly located in the accumulation zone of the Greenland Ice Sheet (Steffen et al., 1996). PROMICE was initiated in 2007 and its monitoring sites are located mainly in the ablation area to best complement the location of existing stations (Ahlstrøm, PROMICE Project Team, et al., 2008). The automatic weather stations are designed to endure extreme temperatures and winds, and the highly variable snow or ice surface while still allowing it to be transportable by helicopter, snowmobile, or dogsled (Fausto et al., 2021). Surface air temperature observations for the Antarctic Ice Sheet are mainly available through the Antarctic Meteorological Research Center (AMRC) and the National Climatic Data Center (NCDC). For both hemispheres, the main source of sea-ice surface temperature observations are drifting buoys and ice mass balance buoys. In 1937, the former Soviet Union established the manned North Pole drifting stations, providing surface air temperature observations over the Arctic Ocean. In 1978, the Arctic Ocean Buoy Program was established, succeeded by the International Arctic Buoy Programme (IABP) in 1991, to maintain a network of drifting buoys on the Arctic Ocean. Since 1993, Ice Mass Balance Buoys have been developed at the US Army Cold Regions Research and Engineering Laboratory (CRREL) and deployed during a variety of observational efforts in the Arctic. In parallel, similar efforts have been initiated and operated in the Southern Ocean mainly through the International Programme for Antarctic Buoys which was initiated in 1994.

Due to the harsh environment and poor accessibility air surface temperature observations of the cryosphere are generally very sparse (Reeves Eyre and Zeng, 2017). This is particularly the case over sea ice, where the drifting buoys often get buried in snow, which means that the measured temperature no longer represents the surface air temperature but the temperature within the snow, which can differ by several degrees from the actual surface air temperature.

Despite their importance, the surface temperatures of the cryosphere are heavily under-sampled causing large uncertainties and differences in the temperature fields of these regions. Recent findings show that the Arctic is warming about 3-4 times the global average based on observational datasets (Rantanen et al., 2022), significantly more than previously reported (Walsh, 2014). Even now, large differences in the regional trends are observed (Marquardt Collow et al., 2020; Rantanen et al., 2022), which are likely explained by the varying observational coverage and the use of different methods to account for the incomplete sampling. Consequently, there is a great potential in the combination of *in situ* measurements and satellite-derived temperatures, which can provide a much more complete surface temperature field. This is illustrated in Figure 2, which shows the observed surface air temperatures from *in situ* stations and infrared satellites during one day for the northern and southern hemisphere, respectively (Nielsen-Englyst et al., 2021; Rayner et al., 2020).

3.6 Estimates of sea-ice concentration

The marginal ice zone is often defined as the area of the ice-covered ocean that is adjacent to the open ocean and where the ice concentration is less than 80% and more than 15% (Dumont, 2022). In the context of global surface temperatures the seaward edge of the marginal ice zone defines the transition between the use of SST anomalies over the ocean and LSAT anomalies over the ice. Shifts between the use of SST and LSAT due to the temporally varying sea ice cover can cause artefacts in GMST (Gulev et al., 2023).

All of the GMST datasets use estimates of sea ice concentration from the HadISST2 dataset (Titchner and Rayner, 2014) to mark this transition, in what is typically a very data sparse region. HadISST2 provides monthly mean sea ice concentration on 1° grid generated from a combination of satellite observations available since the late 1970s with information from various ice charts and atlas compilations. These different data sources have been harmonised to provide the most consistent record possible (Titchner and

Rayner, 2014). Even with these wide-ranging sources of data, much of the variability prior to the 1970s is based solely on seasonally-varying climatology.

3.7 Observations for air temperature over lakes

Accurately measuring air temperature over lakes is critical for studying how these bodies of water interact with the atmosphere and influence regional climate. However, such observations have historically been scarce due to the logistical difficulties of deploying and maintaining monitoring stations on large water bodies, coupled with the historical focus of meteorological networks on land-based data. Neglecting air temperature over lakes can lead to inaccurate and incomplete air temperature datasets, hindering our ability to assess the true impact of climate change. This is particularly critical in regions dominated by large or numerous lakes, where their influence on the surrounding and overlying air temperature can be significant (Balsamo et al., 2012; Thiery et al., 2015). While readily available land-based air temperature data is a valuable resource, it cannot provide a complete picture without incorporating measurements over lakes, particularly with the increasing spatial resolution of gridded air temperature data products. This is because air temperature above lakes can be significantly different from surrounding land areas due to the unique thermal properties of water. Lake surface temperature and air temperature variations often exhibit a time lag (Toffolon et al., 2020), with large lakes typically being cooler in summer and warmer in winter compared to the surrounding air due to their thermal inertia. Additionally, the air-water temperature relationship can vary significantly based on lake size and location (Verburg and Antenucci, 2010; Woolway et al., 2017).

Recent advancements in satellite-based remote sensing offer promising solutions to overcome some of the challenges of monitoring air-lake temperature relationships. These technologies provide valuable information on surface temperatures over large water bodies, complementing existing *in situ* observations and filling critical gaps in our understanding. However, further research is needed to translate these surface water temperature observations into accurate estimates of over-lake air temperature. Encouragingly, air temperature data over some lakes does exist, particularly for large water bodies like the North American Great Lakes, where measurements are collected on lighthouses and dedicated monitoring stations and made openly available via the National Data Buoy Center (NDBC; https://www.ndbc.noaa.gov). Recent decades have also seen the positive trend of deploying *in situ* monitoring stations on smaller lakes worldwide. Examples include the UK Lake Ecosystem Observation Network (UKLEON) and the Global Lake Ecological Observatory Network (GLEON), which provide detailed water and air temperature measurements at the lake surface. Some recent efforts have also demonstrated the use of citizen science to investigate the relationships between air and lake surface water temperature (Weyhenmeyer et al., 2017). These initiatives hold significant promise for establishing global-scale relationships between lake surface and air temperature observations. However, despite these positive developments, significant knowledge gaps remain in our understanding of air temperature over most lakes. Closing this gap is crucial. Continued deployment of monitoring stations, coupled with analyzing existing data, is essential for building a comprehensive and accurate picture of global surface air temperatures.

4. Global fields of surface temperature

Typically surface temperature datasets are developed separately for either marine or land domains and then merged. None of the presently available GMST records make explicit treatment of the air temperature over lakes, and observations for the cryosphere, where available, are included as part of the land component. It was argued that anomalies of sea-surface temperature and marine air temperature were similar enough at large space and time scales to be considered interchangeable for global surface temperature datasets (e.g. Folland et al., 1984; Huang et al., 2017). GMST data products (e.g. Morice et al., 2021; Zhang et al., 2019; Lenssen et al., 2019) therefore combine anomalies of LSAT with anomalies of SST rather than MAT. This choice was made because SST anomalies were believed to more reliable and less variable than MAT. This is despite NMAT being used to estimate the bias adjustments required for SST. More recently this assumption of near-equivalence has been questioned (Cowtan et al., 2015; Richardson et al., 2016; Richardson et al., 2018) as MAT is estimated to warm slightly more than SST in climate models (Richardson et al., 2018; IPCC, 2023a), but not in observations (Cornes et al., 2020; Junod and Christy, 2020), as discussed in the most recent IPCC report (Gulev et al., 2023).

Recently a GSAT dataset, GloSAT, has been developed combining LSAT and MAT anomalies for the period since 1781 (Morice et al., 2024) using the same methods as HadCRUT5 (Morice et al., 2021). This required the adjustment of MAT observations for daytime heating biases (Cropper et al., 2023) and early LSAT observations for exposure biases (Wallis et al., 2024). Further developments were required to estimate a reference climatology for stations where data were missing during the reference period (Taylor et al., 2024). This has enabled the extension of the record back beyond the 1850 start of the current GMST datasets to cover the late pre-industrial period (Hawkins et al., 2017) and shows the cooling of the atmosphere due to several large volcanic eruptions in this early period, as also seen in the Berkeley Earth LSAT dataset (Rohde et al., 2013).

The IPCC AR6 used four combined global analyses to estimate trends in temperature since 1850: HadCRUT5 (Morice et al., 2021), NOAAGlobalTemp (Zhang et al., 2019), BerkeleyEarth (Rohde et al., 2013) and Kadow et al. (2020). There was substantial commonality between the gridded domain inputs to these datasets. HadCRUT5 combines CRUTEM5 (Osborn et al., 2021) and HadSST4 (Kennedy, 2014), NOAAGlobalTemp combines GHCNv4 (Menne et al., 2018) and ERSSTv5 (Huang et al., 2017), BerkeleyEarth (Rohde and Hausfather, 2020) uses their own analysis for LSAT combined with the superseded HadSST3 (Kennedy et al., 2011) and Kadow et al. (2020) is a reprocessing of HadCRUT5. GISTEMPv4 (Lenssen et al., 2019) uses the same components as NOAAGlobalTemp but starts later, in 1880.

Assessments have been made of the dynamical consistency of global temperature analyses by comparing variability over the land and the ocean (Cowtan et al., 2018; Chan et al., 2023). These studies concluded that the long-term record of LSAT is more reliable, even after the application of bias adjustment for SST. In particular the period at the start of the Twentieth Century shows SST anomalies that are unexpectedly cold relative to the LSAT anomalies, and that are also inconsistent with variability shown in climate models (Sippel et al., 2024).

5. Global Changes

The IPCC AR6 concluded that it was likely that every year from 2015 to 2020 was warmer than any other year between 1850 and 2014. Observed warming was assessed relative to the average between 1850 and 1900 (to approximate changes since the pre-industrial era). Using this measure GMST increased by 0.85 [0.69 to 0.95] ˚C between 1850-1900 and 1995-2014 and by 1.09 [0.95 to 1.20] ˚C between 1850-1900 and 2011-2020. Temperature increases over land have been faster than over the oceans (Figure 5) and fastest in polar regions (Figure 6). Increases in GMST have been robustly attributed to human influence (e.g. Gillett et al., 2021).

The State of the Climate Report for 2022 (Blunden et al., 2023) updates the picture from the IPCC. GMST is modulated by the El Niño Southern Oscillation (ENSO) with warmer conditions under the El Niño phase when SSTs in the central and eastern tropical Pacific Ocean are warmer than average in contrast to cooler La Niña conditions. In 2022 GMST was affected by the La Niña state of the Pacific and was ranked either fifth or sixth highest in the instrument record. Despite this, the Indicators of Global Climate Change 2022 paper (Forster et al., 2023) reported that GMST had risen by a further 0.06 ˚C in the 2 years since the IPCC assessment, to give an updated GMST increase of 1.15 [1.00 to 1.25] ˚C. By 2023 El Niño conditions meant that the World Meteorological Organisation Provisional State of the Climate in 2023 (WMO, 2023) describes the warmest year to date with an average GMST across five datasets (HadCRUT5, GlobalTemp, GISTEMP, ERA5 (Hersbach et al., 2020) and JRA55 (Kobayashi et al., 2015)) of 1.40 ± 0.12 ˚C above the 1850-1900 average. Monthly mean temperatures for June through September 2023 all broke previous records in all of the datasets considered, by between 0.1 and 0.5 ˚C (WMO, 2023). The European Union Copernicus Climate Change

Service (C3S) provides information to support climate change adaptation and mitigation policies. Based on ERA5 surface temperatures C3S highlighted the exceptional warmth of the North Atlantic SST during June 2023 (Figure 7, Copernicus, 2023).

6. Characterisation of Extremes and Variability

Although GMST is an important indicator of climate change, the changes in mean temperature are small relative to its variability. This means that the rise in mean temperatures is only important in certain circumstances, for example near the melting point of ice, or for ecosystems that are already at the limit of their range in terms of temperature tolerance. Typically of far more importance is the impact that changing temperatures have on variability and extremes (IPCC, 2023b). More frequent and intense extreme events beyond natural climate variability have already caused widespread adverse impacts to nature and people (IPCC, 2023b). The Expert Team on Climate Change Detection and Indices (ETCCDI) was established in the late 1990s, and the set of indices developed by the group has become the principal tool for monitoring changes in temperature extremes over land (Zhang et al., 2011). The terrestrial indices, calculated from daily values of maximum, minimum or mean temperature values recorded at global terrestrial weather stations have been combined in the HadEX3 dataset to form spatially complete fields of temperature extremes (Dunn et al., 2020). On a global average basis, the annual number of warm days has increased by around 30 days since the late 1970s, whereas the number of cool days has decreased by around 15 days over the same time period. These global annual average values mask important regional and seasonal differences. For example, across the central United States, daytime maximum temperature values have displayed a weak and slightly negative trend over the last 60 years, which contrasts with significant warming across the west and east of the country. This has been attributed to a complex ocean-atmosphere interactions in relationship to the hydrological cycle (Eischeid et al., 2023).

Until fairly recently analysis of temperature extremes was restricted to LSAT. Less attention was paid to extended period of extremely warm water known as marine heatwaves (MHW). MHW events have led to widespread, abrupt and extensive mortality of key habitat-forming species among tropical corals, kelps, seagrasses and mangroves (IPCC, 2023b). One commonly adopted approach identifies a MHW when the water temperature exceeds the seasonally varying 90th percentile threshold for at least 5 days (Hobday et al., 2016). The climatology and threshold are usually calculated from a 30-year baseline period commencing in the early 1980s, which coincides with the advent of daily and high spatial resolution satellite data availability. Other definitions exist based on fixed (Frölicher et al., 2018) or cumulative (Eakin et al., 2010) thresholds, which can serve different purposes, depending on whether the applications are ecological or physical in nature. MHWs have typically been identified in SST data (e.g. Oliver et al., 2018; Frölicher et al. 2018; Holbrook et al., 2019) since it is most readily available at the spatial and temporal resolution required. But MHWs can and do extend to deeper layers of the ocean, and they sometimes occur entirely below the surface (e.g. Schaeffer and Roughan, 2017; Elzahaby and Schaeffer, 2019; Zhang et al., 2023).

Coupled with increasing global mean SST (Figure 5b), MHWs are increasing by all measures, including their intensity, duration, and frequency (Oliver et al., 2018), as well as their spatial extent. However, the trends are not uniform across the global oceans, and some regions have even experienced decreases over the satellite era (Wang et al., 2022). The relationships between the rates of change in mean SST and the extremes is poorly understood, and determining such relationships will depend on understanding how temperature distributions and variability has changed over time, and how they might continue in future.

As global temperatures have risen it is increasingly being asked whether particular extreme events can be attributed to the human impact on climate. That humans have affected temperatures globally was demonstrated by detecting the distinctive pattern of anthropogenically-forced temperature change long before the recent more rapid global temperature increases (Hegerl et al., 1996; Stott et al., 2001). Over land historical warming has increased the severity and probability of the hottest monthly and daily events (Diffenbaugh et al., 2017) and over the ocean the frequency, intensity and duration of MHWs has increased markedly (Oliver et al., 2018; Laufkötter et al., 2020). The risk of individual events such as the European summer heatwave of 2003 have been shown to be substantially increased (Stott et al., 2004). More recently the likely impact of human activity on the occurrence and severity of extreme events has developed to give rapid assessments (Otto et al., 2022). In particular if human-induced climate change is shown to play an important role then it is likely that extreme events will worsen over time, and past observations become increasingly less relevant (Raju et al., 2022).

7. Future Research Directions

Underpinning efforts to improve our records of global surface temperature change must be improvements and extensions to the collections of raw observations on which the gridded records are based. More data are needed, particularly to extend the instrumental record back in time, and to improve coverage in data sparse periods and regions. Particularly important are high quality observations to understand biases, and high-resolution observations to improve variability estimates. There should be a prioritization of long-term observing stations, especially for the cryosphere and other data sparse regions.

The complexity of the individual records, along with the presence of known, but unresolved, data problems in the holdings are a barrier to extending participation and the development of a wider range of analysis methods and data products. Lack of resources have meant that when issues are identified with particular observations they are either simply excluded from any analyses, or *ad hoc* solutions are implemented, rather than solved in periodic reprocessing of the archived data. Routine evaluation of new and existing data sources will reduce the possibility of new problems creeping into the datasets. Approaches to quality control and uncertainty estimation need to be compared, assessed, and integrated into the evaluation of raw and bias adjusted observations.

Ideally the design of the data system for historical observations should be treated in the same way as for a new observing network, following international conventions for data and metadata, where possible, whilst retaining access to the full information from the raw data sources. The lack of a designed data system means that there is no clear pipeline for observations from marine data rescue efforts into the archives so there is a delay in observations being included in global analyses. Additional observations will come from several sources. Historical observations will come from data rescue efforts, but there is already a backlog of data waiting to be integrated into the archives and many ongoing efforts will provide more observations. It is likely that before long Artificial Intelligence (AI) approaches to data rescue will become successful giving a step-change in the amount of historical observations needing to be assessed and integrated meaning that improved data systems will become even more important.

Such a reprocessing of the data archives and modernization of the data system will make it easier to work with the observations and may enable a wider range of contributors and approaches. It will also facilitate the development of AI and machine learning approaches to the evaluation and quality control of observations.

Improvements to the quantification of biases is a priority, particularly for SST and MAT, but also for all data early in the record. Comparisons of coastal data derived from land and ocean sources typically suggests that there are residual biases in SST after adjustment when compared with LSAT which appears to be more consistent. The main issue appears to be at the start of the Twentieth Century which is a period when ship identifiers are largely missing making it harder to understand the cause. Unresolved differences between SST and NMAT also need to be addressed, this is particularly important as SST bias adjustments presently rely on NMAT — breaking this dependence to produce independent estimate of SST and MAT needs to be a priority. There has already been substantial progress in methods to homogenize both SST and LSAT that will improve the next generation of global temperature products.

It is clear that better approaches are needed to capture changes in the data sparse and rapidly changing polar regions. The requirement for higher resolution data products, driven in part by requirements for adaptation and climate services, will mean that the treatment of temperature over lakes will become more critical. This is likely to require integration of temperature data from satellite and *in situ*, which will also feed into improved estimates of variability and co-variability needed for improvements to uncertainty estimation, gridding and interpolation to unobserved regions.

A more structured approach to evaluation would drive rapid improvements to global surface temperature gridded datasets. Dynamical evaluation of the consistency between observations for land, ocean and sea ice, should be integrated into the development of datasets and agreeing protocols for software and data formats would permit benchmarking of approaches to enable efficient comparison of methods and data.

8. Conclusions

It is unequivocal that the global surface temperatures have warmed over the period for which instrumental records are available. This change has not been linear and has varied substantially geographically. Important uncertainties and challenges remain to be addressed regarding, for example: data and metadata availability, measurement understanding, improved techniques for dataset construction and uncertainty quantification, and providing high spatial and temporal resolution data suitable for many applications including climate services to support adaptation to the impacts of climate change. These challenges may alter important aspects of our understanding of surface temperatures but will not affect the bottom-line conclusion that the world has warmed.

9. Funding acknowledgements

ECK and RCC were supported by UKRI NERC Large Grant GloSAT [grant number NE/S015647/2]. RBK was supported by the UKRI NERC CLASS program [grant number NE/R015953/1]. EJW was supported by UKRI NERC Large Grant GloSAT [grant number NE/S015582/1]. RIW was supported by a UKRI NERC Independent Research Fellowship [grant number NE/T011246/1].

References

Ahlstrøm, A. P., PROMICE Project Team, et al. (2008). A new programme for monitoring the mass loss of the Greenland ice sheet. *GEUS Bulletin* 15, pp. 61-64. DOI: 10.34194/geusb.v15.5045.

Alexandersson, H. (1986). A homogeneity test applied to precipitation data. *Journal of Climatology* 6.6, pp. 661-675. DOI: 10.1002/joc.3370060607.

Ashford, O. M. (1948). A new bucket for measurement of sea surface temperature. *Quarterly Journal of the Royal Meteorological Society* 74.319, pp. 99-104. DOI: 10.1002/qj.49707431916.

Auchmann, R. and S. Brönnimann (2012). A physics-based correction model for homogenizing sub-daily temperature series. *Journal of Geophysical Research: Atmospheres* 117.D17. DOI: https://doi.org/10.1029/2012JD018067.

Balsamo, G., R. Salgado, E. Dutra, S. Boussetta, T. Stockdale, and M. Potes (2012). On the contribution of lakes in predicting near-surface temperature in a global weather forecasting model. *Tellus A: Dynamic Meteorology and Oceanography* 64.1, p. 15829. DOI: 10.3402/tellusa.v64i0.15829.

Berry, D. I. and E. C. Kent (Nov. 2017). Assessing the health of the in situ global surface marine climate observing system. *International Journal of Climatology* 37.5, pp. 2248-2259. DOI: 10.1002/joc.4914.

Berry, D. I., E. C. Kent, and P. K. Taylor (2004). An Analytical Model of Heating Errors in Marine Air Temperatures from Ships. *Journal of Atmospheric and Oceanic Technology* 21.8, pp. 1198-1215. DOI: 10.1175/1520-0426(2004)021<1198:AAMOHE>2.0.CO;2.

Blunden, J., T. Boyer, and E. Bartow-Gillies (2023). State of the Climate in 2022. *Bulletin of the American Meteorological Society* 104.9, S1-S516. DOI: 10.1175/2023bamsstateoftheclimate.1.

Bojinski, S., M. Verstraete, T. C. Peterson, C. Richter, A. Simmons, and M. Zemp (2014). The Concept of Essential Climate Variables in Support of Climate Research, Applications, and Policy. *Bulletin of the American Meteorological Society* 95.9, pp. 1431-1443. DOI: 10.1175/bams-d-13-00047.1.

Bottomley, M., C. K. Folland, J. Hsiung, R. E. Newell, and D. E. Parker (1990). *Global Ocean Surface Temperature Atlas (GOSTA)*. Meteorological Office (Bracknell) and Massachusetts Institute of Technology.

Boyer, T., H.-M. Zhang, K. O'Brien, J. Reagan, S. Diggs, E. Freeman, H. Garcia, E. Heslop, P. Hogan, B. Huang, L.-Q. Jiang, A. Kozyr, C. Liu, R. Locarnini, A. V. Mishonov, C. Paver, Z. Wang, M. Zweng, S. Alin, L. Barbero, J. A. Barth, M. Belbeoch, J. Cebrian, K. J. Connell, R. Cowley, D. Dukhovskoy, N. R. Galbraith, G. Goni, F. Katz, M. Kramp, A. Kumar, D. M. Legler, R. Lumpkin, C. R. McMahon, D. Pierrot, A. J. Plueddemann, E. A. Smith, A. Sutton, V. Turpin, L. Jiang, V. Suneel, R. Wanninkhof, R. A. Weller, and A. P. S. Wong (2023). Effects of the Pandemic on Observing the Global Ocean. *Bulletin of the American Meteorological Society* 104.2, E389-E410. DOI: 10.1175/bams-d-21-0210.1.

Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones (2006). Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *Journal of Geophysical Research: Atmospheres* 111.D12. DOI: 10.1029/2005jd006548.

Brönnimann, S., R. Allan, L. Ashcroft, S. Baer, M. Barriendos, R. Brázdil, Y. Brugnara, M. Brunet, M. Brunetti, B. Chimani, R. Cornes, F. Domíınguez-Castro, J. Filipiak, D. Founda, R. G. Herrera, J. Gergis, S. Grab, L. Hannak, H. Huhtamaa, K. S. Jacobsen, P. Jones, S. Jourdain, A. Kiss, K. E. Lin, A. Lorrey, E. Lundstad, J. Luterbacher, F. Mauelshagen, M. Maugeri, N. Maughan, A. Moberg, R. Neukom, S. Nicholson, S. Noone, O. Nordli, K. B. Ólafsdóttir, P. R. Pearce, L. Pfister, K. Pribyl, R. Przybylak, C. Pudmenzky, D. Rasol, D. Reichenbach, L. Rˇezníc ˇková, F. S. Rodrigo, C. Rohr, O. Skrynyk, V. Slonosky, P. Thorne, M. A. Valente, J. M. Vaquero, N. E. Westcottt, F. Williamson, and P. Wyszyn´ski (2019). Unlocking Pre-1850 Instrumental Meteorological Records: A Global Inventory. *Bulletin of the American Meteorological Society* 100.12, ES389-ES413. DOI: 10.1175/bams-d-19-0040.1.

Brönnimann, S., Y. Brugnara, R. J. Allan, M. Brunet, G. P. Compo, R. I. Crouthamel, P. D. Jones, S. Jourdain, J. Luterbacher, P. Siegmund, M. A. Valente, and C. W. Wilkinson (2018). A roadmap to climate data rescue services. *Geoscience Data Journal* 5.1, pp. 28-39. DOI: 10.1002/gdj3.56.

Brunet, M., J. Asin, J. Sigró, M. Bañón, F. García, E. Aguilar, J. E. Palenzuela, T. C. Peterson, and P. Jones (2011). The minimization of the screen bias from ancient Western Mediterranean air temperature records: an exploratory statistical analysis. *International Journal of Climatology* 31, pp. 1879-1895. DOI: 10.1002/joc.2192.

Callendar, G. S. (1938). The artificial production of carbon dioxide and its influence on temperature. *Quarterly Journal of the Royal Meteorological Society* 64.275, pp. 223-240. DOI: 10.1002/qj.49706427503.

Carella, G., J. J. Kennedy, D. I. Berry, S. Hirahara, C. J. Merchant, S. Morak-Bozzo, and E. C. Kent (2018). Estimating Sea Surface Temperature Measurement Methods Using Characteristic Differences in the Diurnal Cycle. *Geophysical Research Letters* 45.1, pp. 363-371. DOI: 10.1002/2017gl076475.

Carella, G. (2017). "New estimates of uncertainty in the marine surface temperature record". Available at https://eprints.soton.ac.uk/415483. PhD thesis. University of Southampton.

Carella, G., E. C. Kent, and D. I. Berry (2017). A probabilistic approach to ship voyage reconstruction in ICOADS. *International Journal of Climatology* 37.5, pp. 2233-2247. DOI: 10.1002/joc.4492.

Caussinus, H. and O. Mestre (2004). Detection and correction of artificial shifts in climate series. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 53.3, pp. 405-425. DOI: https://doi.org/10.1111/j.1467-9876.2004.05155.x.

Centurioni, L. R., J. Turton, R. Lumpkin, L. Braasch, G. Brassington, Y. Chao, E. Charpentier, Z. Chen, G. Corlett, K. Dohan, C. Donlon, C. Gallage, V. Hormann, A. Ignatov, B. Ingleby, R. Jensen, B. A. Kelly-Gerreyn, I. M. Koszalka, X. Lin, E. Lindstrom, N. Maximenko, C. J. Merchant, P. Minnett, A. O'Carroll, T. Paluszkiewicz, P. Poli, P.-M. Poulain, G. Reverdin, X. Sun, V. Swail, S. Thurston, L. Wu, L. Yu, B. Wang, and D. Zhang (2019). Global in situ Observations of Essential Climate and Ocean Variables at the Air-Sea Interface. *Frontiers in Marine Science* 6. DOI: 10.3389/fmars.2019.00419.

Chan, D., G. Gebbie, and P. Huybers (2023). Global and Regional Discrepancies between Early-Twentieth-Century Coastal Air and Sea Surface Temperature Detected by a Coupled Energy-Balance Analysis. *Journal of Climate* 36.7, pp. 2205-2220. DOI: 10.1175/jcli-d-22-0569.1.

Chan, D., G. Gebbie, and P. Huybers (2024). An improved ensemble of land-surface air temperatures since 1880 using revised pair-wise homogenization algorithms accounting for autocorrelation. *Journal of Climate*, in press.

Chan, D. and P. Huybers (2019). Systematic Differences in Bucket Sea Surface Temperature Measurements among Nations Identified Using a Linear-Mixed-Effect Method. *Journal of Climate* 32.9, pp. 2569-2589. DOI: 10.1175/jcli-d-18-0562.1.

Chan, D. and P. Huybers (2021). Correcting Observational Biases in Sea Surface Temperature Observations Removes Anomalous Warmth during World War II. *Journal of Climate* 34.11, pp. 4585-4602. DOI: 10.1175/jcli-d-20-0907.1.

Copernicus (2023). *Record-breaking North Atlantic Ocean temperatures contribute to extreme marine heatwave*. https://climate.copernicus.eu/record-breaking-north-atlantic-oceantemperatures-contribute-extreme-marine-heatwaves. Accessed: 13-03-2024.

Cornes, R. C., E. Kent, D. Berry, and J. J. Kennedy (2020). CLASSnmat: A global night marine air temperature data set, 1880-2019. *Geoscience Data Journal* 7.2, pp. 170-184. DOI: 10.1002/ gdj3.100.

Cowtan, K., Z. Hausfather, E. Hawkins, P. Jacobs, M. E. Mann, S. K. Miller, B. A. Steinman, M. B. Stolpe, and R. G. Way (2015). Robust comparison of climate models with observations using blended land air and ocean sea surface temperatures. *Geophysical Research Letters* 42.15, pp. 6526-6534. DOI: 10.1002/2015gl064888.

Cowtan, K., R. Rohde, and Z. Hausfather (2018). Evaluating biases in sea surface temperature records using coastal weather stations. *Quarterly Journal of the Royal Meteorological Society* 144.712, pp. 670-681. DOI: 10.1002/qj.3235.

Cropper, T. E., D. I. Berry, R. C. Cornes, and E. C. Kent (2023). Quantifying Daytime Heating Biases in Marine Air Temperature Observations from Ships. *Journal of Atmospheric and Oceanic Technology* 40.4, pp. 427-438. DOI: 10.1175/jtech-d-22-0080.1.

Della-Marta, P. M. and H. Wanner (2006). A Method of Homogenizing the Extremes and Mean of Daily Temperature Measurements. *Journal of Climate* 19.17, pp. 4179-4197. DOI: 10.1175/JCLI3855.1.

Diffenbaugh, N. S., D. Singh, J. S. Mankin, D. E. Horton, D. L. Swain, D. Touma, A. Charland, Y. Liu, M. Haugen, M. Tsiang, and B. Rajaratnam (2017). Quantifying the influence of global warming on unprecedented extreme climate events. *Proceedings of the National Academy of Sciences* 114.19, pp. 4881-4886. DOI: 10.1073/pnas.1618082114.

Dumont, D. (2022). Marginal ice zone dynamics: history, definitions and research perspectives. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 380.2235. DOI: 10.1098/rsta.2021.0253.

Dunn, R. J. H., L. V. Alexander, M. G. Donat, X. Zhang, M. Bador, N. Herold, T. Lippmann, R. Allan, E. Aguilar, A. A. Barry, M. Brunet, J. Caesar, G. Chagnaud, V. Cheng, T. Cinco, I. Durre, R. de Guzman, T. M. Htay, W. M. Wan Ibadullah, M. K. I. Bin Ibrahim, M. Khoshkam, A. Kruger, H. Kubota, T. W. Leng, G. Lim, L. Li-Sha, J. Marengo, S. Mbatha, S. McGree, M. Menne, M. de los Milagros Skansi, S. Ngwenya, F. Nkrumah, C. Oonariya, J. D. Pabon-Caicedo, G. Panthou, C. Pham, F. Rahimzadeh, A. Ramos, E. Salgado, J. Salinger, Y. Sané, A. Sopaheluwakan, A. Srivastava, Y. Sun, B. Timbal, N. Trachow, B. Trewin, G. van der Schrier, J. Vazquez-Aguirre, R. Vasquez, C. Villarroel, L. Vincent, T. Vischel, R. Vose, and M. N. Bin Hj Yussof (2020). Development of an Updated Global Land In Situ-Based Data Set of Temperature and Precipitation Extremes: HadEX3. *Journal of Geophysical Research: Atmospheres* 125.16. DOI: 10.1029/2019jd032263.

Eakin, C. M., J. A. Morgan, S. F. Heron, T. B. Smith, G. Liu, L. Alvarez-Filip, B. Baca, E. Bartels, C. Bastidas, C. Bouchon, et al. (2010). Caribbean corals in crisis: record thermal stress, bleaching, and mortality in 2005. *PloS one* 5.11, e13969. DOI: 10.1371/journal.pone.0013969.

Eischeid, J. K., M. P. Hoerling, X.-W. Quan, A. Kumar, J. Barsugli, Z. M. Labe, K. E. Kunkel, C. J. Schreck, D. R. Easterling, T. Zhang, J. Uehling, and X. Zhang (Oct. 2023). Why Has the Summertime Central U.S. Warming Hole Not Disappeared? *Journal of Climate* 36.20, pp. 7319-7336. DOI: 10.1175/jcli-d-22-0716.1.

Elzahaby, Y. and A. Schaeffer (2019). Observational Insight Into the Subsurface Anomalies of Marine Heatwaves. *Frontiers in Marine Science* 6.December, pp. 1-14. DOI: 10.3389/fmars.2019.00745.

Fausto, R. S., D. van As, K. D. Mankoff, B. Vandecrux, M. Citterio, A. P. Ahlstrøm, S. B. Andersen, W. Colgan, N. B. Karlsson, K. K. Kjeldsen, N. J. Korsgaard, S. H. Larsen, S. Nielsen, A. Ø. Pedersen, C. L. Shields, A. M. Solgaard, and J. E. Box (2021). Programme for Monitoring of the Greenland Ice Sheet (PROMICE) automatic weather station data. *Earth System Science Data* 13.8, pp. 3819-3845. DOI: 10.5194/essd-13-3819-2021.

Folland, C. K., D. E. Parker, and F. E. Kates (1984). Worldwide marine temperature fluctuations 1856-1981. *Nature* 310.5979, pp. 670-673. DOI: 10.1038/310670a0.

Forster, P. M., C. J. Smith, T. Walsh, W. F. Lamb, R. Lamboll, M. Hauser, A. Ribes, D. Rosen, N. Gillett, M. D. Palmer, J. Rogelj, K. von Schuckmann, S. I. Seneviratne, B. Trewin, X. Zhang, M. Allen, R. Andrew, A. Birt, A. Borger, T. Boyer, J. A. Broersma, L. Cheng, F. Dentener, P. Friedlingstein, J. M. Gutiérez, J. Gütschow, B. Hall, M. Ishii, S. Jenkins, X. Lan, J.-Y. Lee, C. Morice, C. Kadow, J. Kennedy, R. Killick, J. C. Minx, V. Naik, G. P. Peters, A. Pirani, J. Pongratz, C.-F. Schleussner, S. Szopa, P. Thorne, R. Rohde, M. Rojas Corradi, D. Schumacher, R. Vose, K. Zickfeld, V. Masson-Delmotte, and P. Zhai (2023). Indicators of Global Climate Change 2022: Annual update of large-scale indicators of the state of the climate system and human influence. *Earth System Science Data* 15.6, pp. 2295-2327. DOI: 10.5194/essd-15-2295-2023.

Freeman, E., S. D. Woodruff, S. J. Worley, S. J. Lubker, E. C. Kent, W. E. Angel, D. I. Berry, P. Brohan, R. Eastman, L. Gates, W. Gloeden, Z. Ji, J. Lawrimore, N. A. Rayner, G. Rosenhagen, and S. R. Smith (2017). ICOADS Release 3.0: A major update to the historical marine climate record. *International Journal of Climatology* 37.5, pp. 2211-2232. DOI: 10.1002/joc.4775.

Frölicher, T. L., E. M. Fischer, and N. Gruber (2018). Marine heatwaves under global warming. *Nature* 560.7718, pp. 360-364. DOI: 10.1038/s41586-018-0383-9.

Gillespie, I., L. Haimberger, G. P. Compo, and P. W. Thorne (2022). Assessing homogeneity of land surface air temperature observations using sparse-input reanalyses. *International Journal of Climatology* 43.2, pp. 736-760. DOI: 10.1002/joc.7822.

Gillett, N. P., M. Kirchmeier-Young, A. Ribes, H. Shiogama, G. C. Hegerl, R. Knutti, G. Gastineau, J. G. John, L. Li, L. Nazarenko, N. Rosenbloom, Ø. Seland, T. Wu, S. Yukimoto, and T. Ziehn (2021). Constraining human contributions to observed warming since the pre-industrial period. *Nature Climate Change* 11.3, pp. 207-212. DOI: 10.1038/s41558-020-00965-9.

GISTEMP Team (2024). *GISS Surface Temperature Analysis (GISTEMP), version 4*. https://data.giss.nasa.gov/gistemp/. Accessed: 13-03-2024.

Gulev, S., P. Thorne, J. Ahn, F. Dentener, C. Domingues, S. Gerland, D. Gong, D. Kaufman, H. Nnamchi, J. Quaas, J. Rivera, S. Sathyendranath, S. Smith, B. Trewin, K. von Schuckmann, and R. Vose (2023). "Changing State of the Climate System". *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, pp. 287-422. DOI: 10.1017/9781009157896.004.

Hanafin, J. A. and P. J. Minnett (2005). Measurements of the infrared emissivity of a wind-roughened sea surface. *Applied Optics* 44.3, pp. 398-411. DOI: 10.1364/ao.44.000398.

Hausfather, Z., K. Cowtan, D. C. Clarke, P. Jacobs, M. Richardson, and R. Rohde (2017). Assessing recent warming using instrumentally homogeneous sea surface temperature records. *Science Advances* 3.1. DOI: 10.1126/sciadv.1601207.

Hawkins, E., P. Brohan, S. N. Burgess, S. Burt, G. P. Compo, S. L. Gray, I. D. Haigh, H. Hersbach, K. Kuijjer, O. Martínez-Alvarado, C. McColl, A. P. Schurer, L. Slivinski, and J. Williams (2023). Rescuing historical weather observations improves quantification of severe windstorm risks. *Natural Hazards and Earth System Sciences* 23.4, pp. 1465-1482. DOI: 10.5194/nhess-23-1465-2023.

Hawkins, E. and P. D. Jones (2013). On increasing global temperatures: 75 years after Callendar. *Quarterly Journal of the Royal Meteorological Society* 139.677, pp. 1961-1963. DOI: 10.1002/ qj.2178.

Hawkins, E., P. Ortega, E. Suckling, A. Schurer, G. Hegerl, P. Jones, M. Joshi, T. J. Osborn, V. Masson-Delmotte, J. Mignot, P. Thorne, and G. J. van Oldenborgh (2017). Estimating Changes in Global Temperature since the Preindustrial Period. *Bulletin of the American Meteorological Society* 98.9, pp. 1841-1856. DOI: 10.1175/bams-d-16-0007.1.

Hegerl, G. C., H. von Storch, K. Hasselmann, B. D. Santer, U. Cubasch, and P. D. Jones (1996). Detecting Greenhouse-Gas-Induced Climate Change with an Optimal Fingerprint Method. *Journal of Climate* 9.10, pp. 2281-2306. DOI: 10.1175/1520-0442(1996)009<2281:dggicc>2.0.co;2.

Hersbach, H., B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J. Bidlot, M. Bonavita, G. De Chiara, P. Dahlgren, D. Dee, M. Diamantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. Geer, L. Haimberger, S. Healy, R. J. Hogan, E. Hólm, M. Janisková, S. Keeley, P. Laloyaux, P. Lopez, C. Lupu, G. Radnoti, P. de Rosnay, I. Rozum, F. Vamborg, S. Villaume, and J.-N. Thepaut (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* 146.730, pp. 1999-2049. DOI: 10.1002/qj.3803.

Hirahara, S., M. Ishii, and Y. Fukuda (2014). Centennial-Scale Sea Surface Temperature Analysis and Its Uncertainty. *Journal of Climate* 27.1, pp. 57-75. DOI: 10.1175/jcli-d-12-00837.1.

Hobday, A. J., L. V. Alexander, S. E. Perkins-Kirkpatrick, D. A. Smale, S. C. Straub, E. C. Oliver, J. A. Benthuysen, M. T. Burrows, M. G. Donat, M. Feng, N. J. Holbrook, P. J. Moore, H. A. Scannell, A. Sen Gupta, and T. Wernberg (2016). A hierarchical approach to defining marine heatwaves. *Progress in Oceanography* 141, pp. 227-238. DOI: 10.1016/j.pocean.2015.12.014.

Holbrook, N. J., H. A. Scannell, A. Sen Gupta, J. A. Benthuysen, M. Feng, E. C. Oliver, L. V. Alexander, M. T. Burrows, M. G. Donat, A. J. Hobday, P. J. Moore, S. E. Perkins-Kirkpatrick, D. A. Smale, S. C. Straub, and T. Wernberg (2019). A global assessment of marine heatwaves and their drivers. *Nature Communications* 10. Publisher: Springer US ISBN: 4146701910, p. 2624. DOI: 10.1038/s41467-019-10206-z.

Huang, B., P. W. Thorne, V. F. Banzon, T. Boyer, G. Chepurin, J. H. Lawrimore, M. J. Menne, T. M. Smith, R. S. Vose, and H.-M. Zhang (2017). Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5): Upgrades, Validations, and Intercomparisons. *Journal of Climate* 30.20, pp. 8179-8205. DOI: 10.1175/jcli-d-16-0836.1.

IPCC (2023a). *Climate Change 2021 -The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press. DOI: 10.1017/9781009157896.

IPCC (2023b). "Summary for Policymakers". *Climate Change 2022 -Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, pp. 1-2. DOI: 10.1017/9781009325844.001.

Jones, P. D., T. M. L. Wigley, C. K. Folland, D. E. Parker, J. K. Angell, S. Lebedeff, and J. E. Hansen (1988). Evidence for global warming in the past decade. *Nature* 332.6167, pp. 790-790. DOI: 10.1038/332790b0.

Junod, R. A. and J. R. Christy (2020). A new compilation of globally gridded night-time marine air temperatures: The UAHNMATv1 dataset. *International Journal of Climatology* 40.5, pp. 2609-2623. DOI: 10.1002/joc.6354.

Kadow, C., D. M. Hall, and U. Ulbrich (2020). Artificial intelligence reconstructs missing climate information. *Nature Geoscience* 13.6, pp. 408-413. DOI: 10.1038/s41561-020-0582-5.

Karl, T. R., C. N. Williams, P. J. Young, and W. M. Wendland (1986). A Model to Estimate the Time of Observation Bias Associated with Monthly Mean Maximum, Minimum and Mean Temperatures for the United States. *Journal of Climate and Applied Meteorology* 25.2, pp. 145-160. DOI: 10.1175/1520-0450(1986)025<0145:amtett>2.0.co;2.

Karspeck, A. R., A. Kaplan, and S. R. Sain (2011). Bayesian modelling and ensemble reconstruction of mid-scale spatial variability in North Atlantic sea-surface temperatures for 1850-2008. *Quarterly Journal of the Royal Meteorological Society* 138.662, pp. 234-248. DOI: 10.1002/qj.900.

Kennedy, J. J., N. A. Rayner, C. P. Atkinson, and R. E. Killick (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, pp. 7719-7763. DOI: 10.1029/2018jd029867.

Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby (2011). Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization. *Journal of Geophysical Research* 116.D14, p. D14104. DOI: 10.1029/2010jd015220.

Kennedy, J. J. (2014). A review of uncertainty in in situ measurements and data sets of sea surface temperature. *Reviews of Geophysics* 52.1, pp. 1-32. DOI: 10.1002/2013rg000434.

Kent, E. C. and D. I. Berry (2005). Quantifying random measurement errors in Voluntary Observing Ships' meteorological observations. *International Journal of Climatology* 25.7, pp. 843-856. DOI: 10.1002/joc.1167.

Kent, E. C. and J. J. Kennedy (2021). Historical Estimates of Surface Marine Temperatures. *Annual Review of Marine Science* 13.1, pp. 283-311. DOI: 10.1146/annurev-marine-042120-111807.

Kent, E. C., J. J. Kennedy, T. M. Smith, S. Hirahara, B. Huang, A. Kaplan, D. E. Parker, C. P. Atkinson, D. I. Berry, G. Carella, Y. Fukuda, M. Ishii, P. D. Jones, F. Lindgren, C. J. Merchant, S. Morak-Bozzo, N. A. Rayner, V. Venema, S. Yasui, and H.-M. Zhang (2017). A Call for New Approaches to Quantifying Biases in Observations of Sea Surface Temperature. *Bulletin of the American Meteorological Society* 98.8, pp. 1601-1616. DOI: 10.1175/bams-d-15-00251.1.

Kent, E. C., N. A. Rayner, D. I. Berry, R. Eastman, V. G. Grigorieva, B. Huang, J. J. Kennedy, S. R. Smith, and K. M. Willett (2019). Observing Requirements for Long-Term Climate Records at the Ocean Surface. *Frontiers in Marine Science* 6. DOI: 10.3389/fmars.2019.00441.

Kent, E. C., N. A. Rayner, D. I. Berry, M. Saunby, B. I. Moat, J. J. Kennedy, and D. E. Parker (2013). Global analysis of night marine air temperature and its uncertainty since 1880: The HadNMAT2 data set. *Journal of Geophysical Research: Atmospheres* 118.3, pp. 1281-1298. DOI: 10.1002/ jgrd.50152.

Kent, E. C., S. D. Woodruff, and D. I. Berry (2007). Metadata from WMO Publication No. 47 and an Assessment of Voluntary Observing Ship Observation Heights in ICOADS. *Journal of Atmospheric and Oceanic Technology* 24.2, pp. 214-234. DOI: 10.1175/jtech1949.1.

Kobayashi, S., Y. Ota, Y. Harada, A. Ebita, M. Moriya, H. Onoda, K. Onogi, H. Kamahori, C. Kobayashi, H. Endo, K. Miyaoka, and K. Takahashi (2015). The JRA-55 Reanalysis: General Specifications and Basic Characteristics. *Journal of the Meteorological Society of Japan. Ser. II* 93.1, pp. 5-48. DOI: 10.2151/jmsj.2015-001.

Laufkötter, C., J. Zscheischler, and T. L. Frölicher (2020). High-impact marine heatwaves attributable to human-induced global warming. *Science* 369.6511, pp. 1621-1625. DOI: 10.1126/science. aba0690.

Lenssen, N. J. L., G. A. Schmidt, J. E. Hansen, M. J. Menne, A. Persin, R. Ruedy, and D. Zyss (2019). Improvements in the GISTEMP Uncertainty Model. *Journal of Geophysical Research: Atmospheres* 124.12, pp. 6307-6326. DOI: 10.1029/2018jd029522.

Lund, R. B., C. Beaulieu, R. Killick, Q. Lu, and X. Shi (2023). Good Practices and Common Pitfalls in Climate Time Series Changepoint Techniques: A Review. *Journal of Climate* 36.23, pp. 8041-8057. DOI: 10.1175/JCLI-D-22-0954.1.

Maronna, R. and V. J. Yohai (1978). A Bivariate Test for the Detection of a Systematic Change in Mean. *Journal of the American Statistical Association* 73.363, pp. 640-645. DOI: 10.1080/01621459.1978.10480070.

Marquardt Collow, A. B., R. I. Cullather, and M. G. Bosilovich (2020). Recent Arctic Ocean surface air temperatures in atmospheric reanalyses and numerical simulations. *Journal of Climate* 33.10, pp. 4347-4367. DOI: 10.1175/JCLI-D-19-0703.1.

Menne, M. J. and C. N. Williams (2005). Detection of undocumented changepoints using multiple test statistics and composite reference series. *Journal of Climate* 18.20, pp. 4271-4286. DOI: 10.1175/JCLI3524.1.

Menne, M. J. and C. N. Williams (2009). Homogenization of temperature series via pairwise comparisons. *Journal of Climate* 22.7, pp. 1700-1717. DOI: 10.1175/2008JCLI2263.1.

Menne, M. J., I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston (2012). An Overview of the Global Historical Climatology Network-Daily Database. *Journal of Atmospheric and Oceanic Technology* 29.7, pp. 897-910. DOI: 10.1175/jtech-d-11-00103.1.

Menne, M. J., C. N. Williams, B. E. Gleason, J. J. Rennie, and J. H. Lawrimore (2018). The Global Historical Climatology Network Monthly Temperature Dataset, Version 4. *Journal of Climate* 31.24, pp. 9835-9854. DOI: 10.1175/jcli-d-18-0094.1.

Merchant, C. J., S. Matthiesen, N. A. Rayner, J. J. Remedios, P. D. Jones, F. Olesen, B. Trewin, P. W. Thorne, R. Auchmann, G. K. Corlett, P. C. Guillevic, and G. C. Hulley (2013). The surface temperatures of Earth: steps towards integrated understanding of variability and change. *Geoscientific Instrumentation, Methods and Data Systems* 2.2, pp. 305-321. DOI: 10.5194/gi-2-305-2013.

Merchant, C. J., O. Embury, C. E. Bulgin, T. Block, G. K. Corlett, E. Fiedler, S. A. Good, J. Mittaz, N. A. Rayner, D. Berry, S. Eastwood, M. Taylor, Y. Tsushima, A. Waterfall, R. Wilson, and C. Donlon (2019). Satellite-based time-series of sea-surface temperature since 1981 for climate applications. *Scientific Data* 6, p. 223. DOI: 10.1038/s41597-019-0236-x.

Morice, C. P., J. J. Kennedy, N. A. Rayner, J. P. Winn, E. Hogan, R. E. Killick, R. J. H. Dunn, T. J. Osborn, P. D. Jones, and I. R. Simpson (2021). An Updated Assessment of Near-Surface Temperature Change From 1850: The HadCRUT5 Data Set. *Journal of Geophysical Research: Atmospheres* 126.3, e2019JD032361. DOI: 10.1029/2019jd032361.

Morice et al. (2024). *An observational record of global near surface air temperature change over land and ocean from 1781*. In preparation.

Nielsen-Englyst, P., J. L. Høyer, K. S. Madsen, R. T. Tonboe, G. Dybkjær, and S. Skarpalezos (2021). Deriving Arctic 2m air temperatures over snow and ice from satellite surface temperature measurements. *The Cryosphere* 15.7, pp. 3035-3057. DOI: 10.5194/tc-15-3035-2021.

Noone, S., C. Atkinson, D. I. Berry, R. J. H. Dunn, E. Freeman, I. Perez Gonzalez, J. J. Kennedy, E. C. Kent, A. Kettle, S. McNeill, M. Menne, A. Stephens, P. W. Thorne, W. Tucker, C. Voces, and K. M. Willett (2020). Progress towards a holistic land and marine surface meteorological database and a call for additional contributions. *Geoscience Data Journal* 8.2, pp. 103-120. DOI: 10.1002/gdj3.109.

Oliver, E. C., M. G. Donat, M. T. Burrows, P. J. Moore, D. A. Smale, L. V. Alexander, J. A. Benthuysen, M. Feng, A. Sen Gupta, A. J. Hobday, N. J. Holbrook, S. E. Perkins-Kirkpatrick, H. A. Scannell, S. C. Straub, and T. Wernberg (2018). Longer and more frequent marine heatwaves over the past century. *Nature Communications* 9.1, p. 1324. DOI: 10.1038/s41467-018-03732-9.

Osborn, T. J., P. D. Jones, D. H. Lister, C. P. Morice, I. R. Simpson, J. P. Winn, E. Hogan, and I. C. Harris (2021). Land Surface Air Temperature Variations Across the Globe Updated to 2019: The CRUTEM5 Data Set. *Journal of Geophysical Research: Atmospheres* 126.2, e2019JD032352. DOI: 10.1029/2019jd032352.

Otto, F. E. L., S. Kew, S. Philip, P. Stott, and G. J. V. Oldenborgh (2022). How to Provide Useful Attribution Statements: Lessons Learned from Operationalizing Event Attribution in Europe. *Bulletin of the American Meteorological Society* 103.3, S21-S25. DOI: 10.1175/bams-d-21-0267.1.

Parker, D. (1994). Effects of changing exposure of thermometers at land stations. *International Journal of Climatology* 14, pp. 1-31. DOI: 10.1002/joc.3370140102.

Raju, E., E. Boyd, and F. Otto (2022). Stop blaming the climate for disasters. *Communications Earth & Environment* 3.1. DOI: 10.1038/s43247-021-00332-2.

Rantanen, M., A. Y. Karpechko, A. Lipponen, K. Nordling, O. Hyvärinen, K. Ruosteenoja, T. Vihma, and A. Laaksonen (2022). The Arctic has warmed nearly four times faster than the globe since 1979. *Communications Earth & Environment* 3.1. DOI: 10.1038/s43247-022-00498-3.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan (2003). 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. DOI: 10.1029/2002jd002670.

Rayner, N. A., R. Auchmann, J. Bessembinder, S. Brönnimann, Y. Brugnara, F. Capponi, L. Carrea, E. M. A. Dodd, D. Ghent, E. Good, J. L. Høyer, J. J. Kennedy, E. C. Kent, R. E. Killick, P. van der Linden, F. Lindgren, K. S. Madsen, C. J. Merchant, J. R. Mitchelson, C. P. Morice, P. Nielsen-Englyst, P. F. Ortiz, J. J. Remedios, G. van der Schrier, A. A. Squintu, A. Stephens, P. W. Thorne, R. T. Tonboe, T. Trent, K. L. Veal, A. M. Waterfall, K. Winfield, J. Winn, and R. I. Woolway (2020). The EUSTACE Project: Delivering Global, Daily Information on Surface Air Temperature. *Bulletin of the American Meteorological Society* 101.11, E1924-E1947. DOI: 10.1175/bams-d-19-0095.1.

Reeves Eyre, J. and X. Zeng (2017). Evaluation of Greenland near surface air temperature datasets. *The Cryosphere* 11.4, pp. 1591-1605. DOI: 10.5194/tc-11-1591-2017.

Reynolds, R. W., D. B. Chelton, J. Roberts-Jones, M. J. Martin, D. Menemenlis, and C. J. Merchant (2013). Objective Determination of Feature Resolution in Two Sea Surface Temperature Analyses. *Journal of Climate* 26.8, pp. 2514-2533. DOI: 10.1175/jcli-d-12-00787.1.

Richardson, M., K. Cowtan, E. Hawkins, and M. B. Stolpe (2016). Reconciled climate response estimates from climate models and the energy budget of Earth. *Nature Climate Change* 6.10, pp. 931-935. DOI: 10.1038/nclimate3066.

Richardson, M., K. Cowtan, and R. J. Millar (2018). Global temperature definition affects achievement of long-term climate goals. *Environmental Research Letters* 13.5, p. 054004. DOI: 10.1088/1748-9326/aab305.

Rohde, R., R. Muller, R. Jacobsen, S. Perlmutter, A. Rosenfeld, J. Wurtele, J. Curry, C. Wickham, and S. Mosher (2013). Berkeley Earth temperature averaging process - an overview. *Geoinformatics Geostatistics An Overview* 1.2, pp. 20-100. DOI: 10.4172/2327-4581.1000103.

Rohde, R. A. and Z. Hausfather (2020). The Berkeley Earth Land/Ocean Temperature Record. *Earth System Science Data* 12.4, pp. 3469-3479. DOI: 10.5194/essd-12-3469-2020.

Schaeffer, A. and M. Roughan (2017). Subsurface intensification of marine heatwaves off southeastern Australia: The role of stratification and local winds. *Geophysical Research Letters* 44.10, pp. 5025-5033. DOI: 10.1002/2017GL073714.

Sippel, S., N. Meinshausen, D. Chan, C. Kadow, R. Neukom, E. M. Fischer, V. Humphrey, I. de Vries, and R. Knutti (2024). *An apparent multi-decadal global ocean cold anomaly in the early twentieth century temperature record*. in preparation.

Smith, T. M. and R. W. Reynolds (2002). Bias Corrections for Historical Sea Surface Temperatures Based on Marine Air Temperatures. *Journal of Climate* 15.1, pp. 73-87. DOI: 10.1175/1520-0442(2002)015<0073:bcfhss>2.0.co;2.

Steffen, K., J. Box, and W. Abdalati (1996). Greenland climate network: GC-Net. *US Army Cold Regions Reattach and Engineering (CRREL), CRREL Special Report*, pp. 98-103.

Stott, P. A., S. F. B. Tett, G. S. Jones, M. R. Allen, W. J. Ingram, and J. F. B. Mitchell (2001). Attribution of twentieth century temperature change to natural and anthropogenic causes. *Climate Dynamics* 17.1, pp. 1-21. DOI: 10.1007/pl00007924.

Stott, P. A., D. A. Stone, and M. R. Allen (2004). Human contribution to the European heatwave of 2003. *Nature* 432.7017, pp. 610-614. DOI: 10.1038/nature03089.

Taylor et al. (2024). *GloSAT LAT sdb: a global compilation of land air temperature station records with updated climatological normals from local expectation kriging*. In preparation.

Thiery, W., E. L. Davin, H.-J. Panitz, M. Demuzere, S. Lhermitte, and N. van Lipzig (2015). The Impact of the African Great Lakes on the Regional Climate. *Journal of Climate* 28.10, pp. 4061-4085. DOI: 10.1175/jcli-d-14-00565.1.

Titchner, H. A. and N. A. Rayner (2014). The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations: HADISST.2.1.0.0 SEA ICE CONCENTRATIONS. *Journal of Geophysical Research: Atmospheres* 119.6, pp. 2864-2889. DOI: 10.1002/2013jd020316.

Toffolon, M., S. Piccolroaz, and E. Calamita (2020). On the use of averaged indicators to assess lakes' thermal response to changes in climatic conditions. *Environmental Research Letters* 15.3, p. 034060. DOI: 10.1088/1748-9326/ab763e.

Trewin, B. (2010). Exposure, instrumentation, and observing practice effects on land temperature measurements. *Wiley Interdisciplinary Reviews: Climate Change* 1.4, pp. 490-506. DOI: 10.1002/ wcc.46.

Verburg, P. and J. P. Antenucci (2010). Persistent unstable atmospheric boundary layer enhances sensible and latent heat loss in a tropical great lake: Lake Tanganyika. *Journal of Geophysical Research: Atmospheres* 115.D11, p. D11109. DOI: 10.1029/2009jd012839.

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

Walsh, J. E. (2014). Intensified warming of the Arctic: Causes and impacts on middle latitudes. *Global and Planetary Change* 117, pp. 52-63. DOI: 10.1016/j.gloplacha.2014.03.003.

Wang, Y., J. B. Kajtar, N. J. Holbrook, L. V. Alexander, and G. S. Pilo (2022). Understanding the Changing Nature of Marine Cold-Spells. *Geophysical Research Letters* 49, e2021GL097002. DOI: 10.1029/2021GL097002.

Weyhenmeyer, G. A., M. Mackay, J. D. Stockwell, W. Thiery, H.-P. Grossart, P. B. Augusto-Silva, H. M. Baulch, E. de Eyto, J. Hejzlar, K. Kangur, G. Kirillin, D. C. Pierson, J. A. Rusak, S. Sadro, and R. I. Woolway (2017). Citizen science shows systematic changes in the temperature difference between air and inland waters with global warming. *Scientific Reports* 7.1, p. 43890. DOI: 10.1038/srep43890.

Williams, C. N., M. J. Menne, and P. W. Thorne (2012). Benchmarking the performance of pairwise homogenization of surface temperatures in the United States. *Journal of Geophysical Research: Atmospheres* 117.D5, p. D05116. DOI: 10.1029/2011JD016761.

WMO (2023). *Provisional State of the Global Climate in 2023*. https://wmo.int/publicationseries/provisional-state-of-global-climate-2023. Accessed: 13-03-2024.

Woolway, R. I., P. Verburg, C. J. Merchant, J. D. Lenters, D. P. Hamilton, J. Brookes, S. Kelly, S. Hook, A. Laas, D. Pierson, A. Rimmer, J. A. Rusak, and I. D. Jones (2017). Latitude and lake size are important predictors of over-lake atmospheric stability. *Geophysical Research Letters* 44.17, pp. 8875-8883. DOI: 10.1002/2017gl073941.

Xu, W., Q. Li, P. Jones, X. L. Wang, B. Trewin, S. Yang, C. Zhu, P. Zhai, J. Wang, L. Vincent, A. Dai, Y. Gao, and Y. Ding (June 2017). A new integrated and homogenized global monthly land surface air temperature dataset for the period since 1900. *Climate Dynamics* 50.7-8, pp. 2513-2536. DOI:10.1007/s00382-017-3755-1.

Yu, L. and X. Jin (2018). A regime-dependent retrieval algorithm for near-surface air temperature and specific humidity from multi-microwave sensors. *Remote Sensing of Environment* 215, pp. 199-216. DOI: 10.1016/j.rse.2018.06.001.

Zhang, H.-M., B. Huang, J. Lawrimore, M. Menne, and T. M. Smith (2019). *NOAA Global Surface Temperature Dataset, Version 5*. DOI: 10.25921/9QTH-2P70.

Zhang, X., L. Alexander, G. C. Hegerl, P. Jones, A. K. Tank, T. C. Peterson, B. Trewin, and F. W. Zwiers (Oct. 2011). Indices for monitoring changes in extremes based on daily temperature and precipitation data. *WIREs Climate Change* 2.6, pp. 851-870. DOI: 10.1002/wcc.147.

Zhang, Y., Y. Du, M. Feng, and A. J. Hobday (Oct. 2023). Vertical structures of marine heatwaves. *Nature Communications* 14.1, p. 6483. DOI: 10.1038/s41467-023-42219-0.

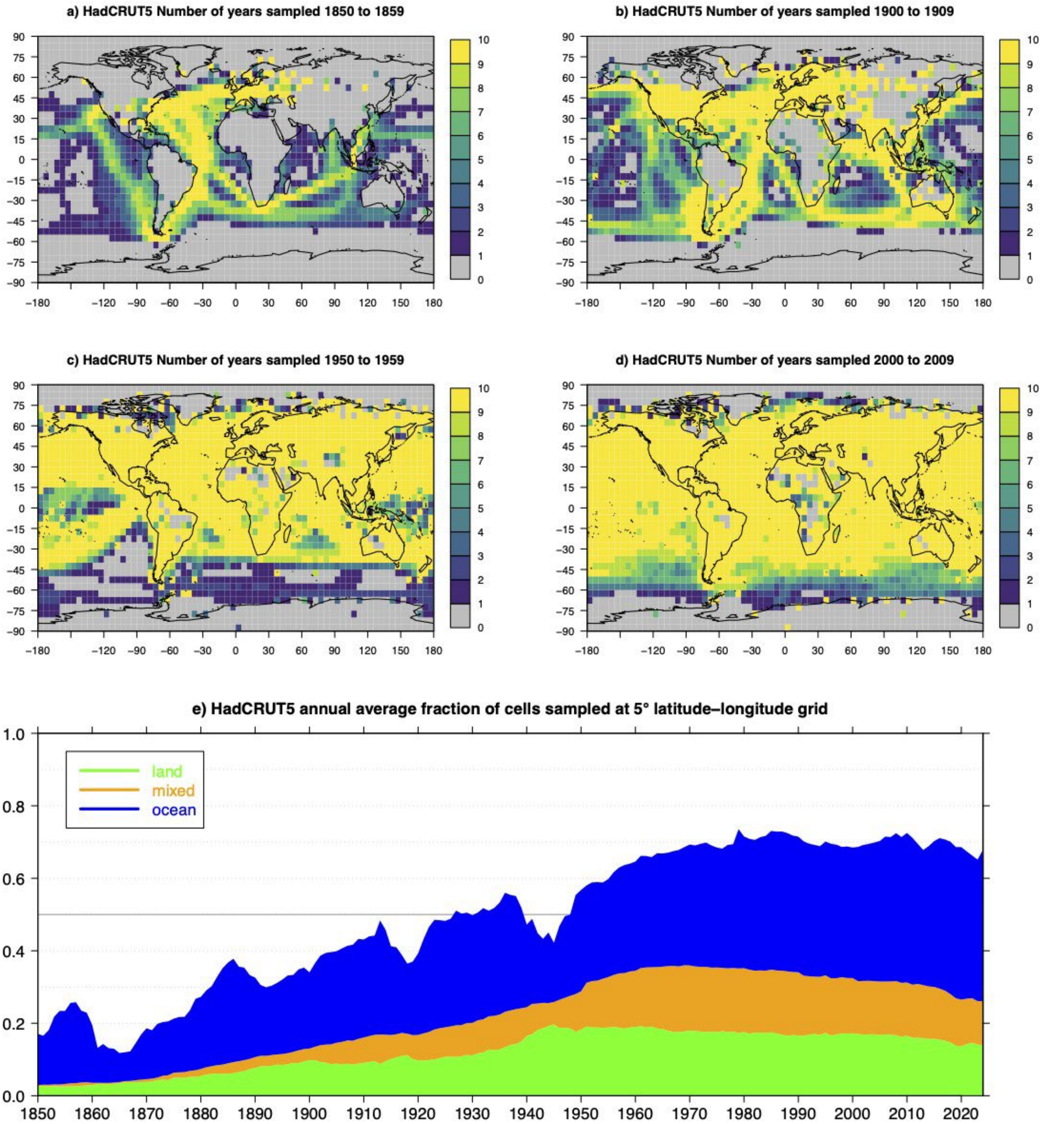

Figure 1: Illustration of HadCRUT5 coverage for 5° latitude-longitude grid-cells for sample decades starting a) 1850; b) 1900; c) 1950; d) 2000. Panel e) shows the fraction of the globe covered annually between 1850 and 2023 for land, ocean and mixed land and ocean grid-cells.

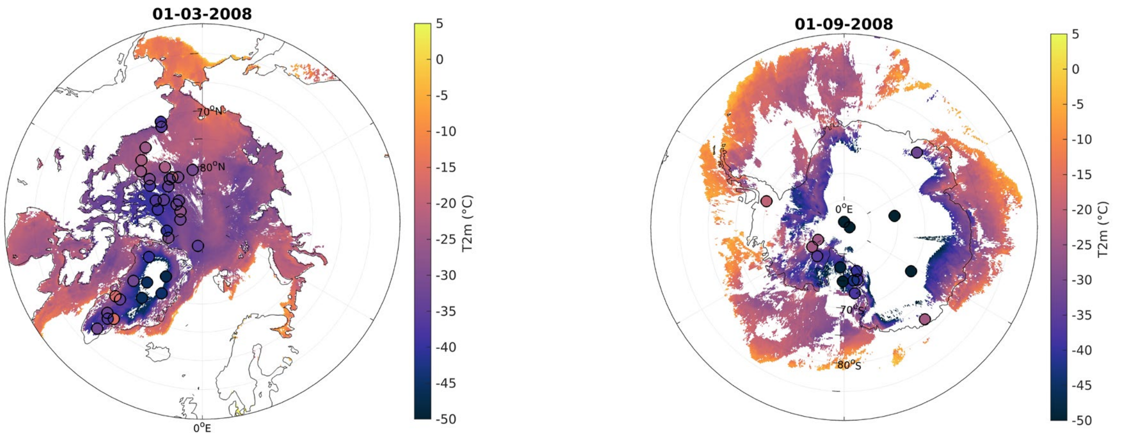

Figure 2: Daily mean 2 m air temperature over land ice and sea ice from infrared (clear-sky) satellite observations and *in situ* measurements (circles) for a) the northern hemisphere March 1, 2008 and b) the southern hemisphere September 1, 2008.

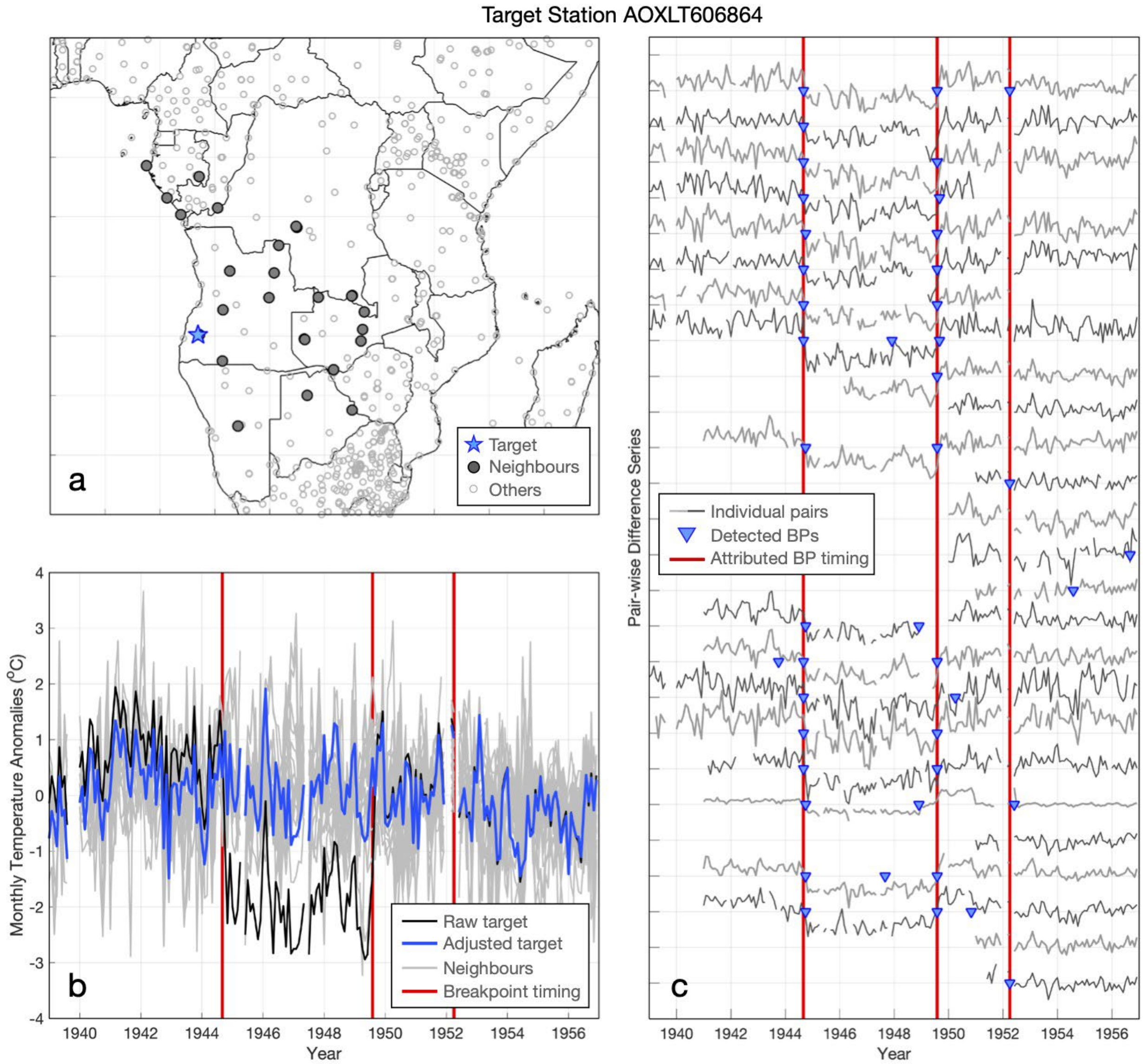

Figure 3: An anatomy of GHCNmV4's pair-wise homogenization at a station in Southwest Angola (AOXLT606864; blue star in panel a). Among all possible stations (open dots in panel a), the pair-wise homogenization algorithm (PHA; Chan et al., 2024) first picks out neighboring stations (filled dots in panel a) based on distance in space and correlation of temperature changes across months. Seasonal cycles are then removed from individual stations. Whereas neighboring stations all indicate increasing temperature from 1939 to 1942, followed by a decrease from 1942 to 1956 (grey curves in panel b), unhomogenized temperature anomalies of the target station contain an apparent discontinuous segment from Aug. 1944 to Jul. 1949 (black curve). The algorithm then uses a Penalized Likelihood Method (Lund et al., 2023) to detect breakpoints (blue triangles in panel c) in the difference series between the target and each of its neighbors (grey curves in panel c). Variants of PHA (Menne and Williams, 2009) could use other statistical tools, for example: a standard normal homogeneity test (Alexandersson, 1986) for breakpoint detection. Identified breakpoints in pair-wise comparison are then attributed to individual stations based on how frequently they are associated with a station at a specific timing. For station AOXLT606864, three breakpoints are attributed (red vertical lines in panels b and c). Whereas the first two breaks are apparent, the third break in Feb. 1952 is less obvious. As a final step, PHA estimates the magnitude of each break by comparing against homogeneous neighbors and removes them to obtain the adjusted temperature (blue curve in panel b).

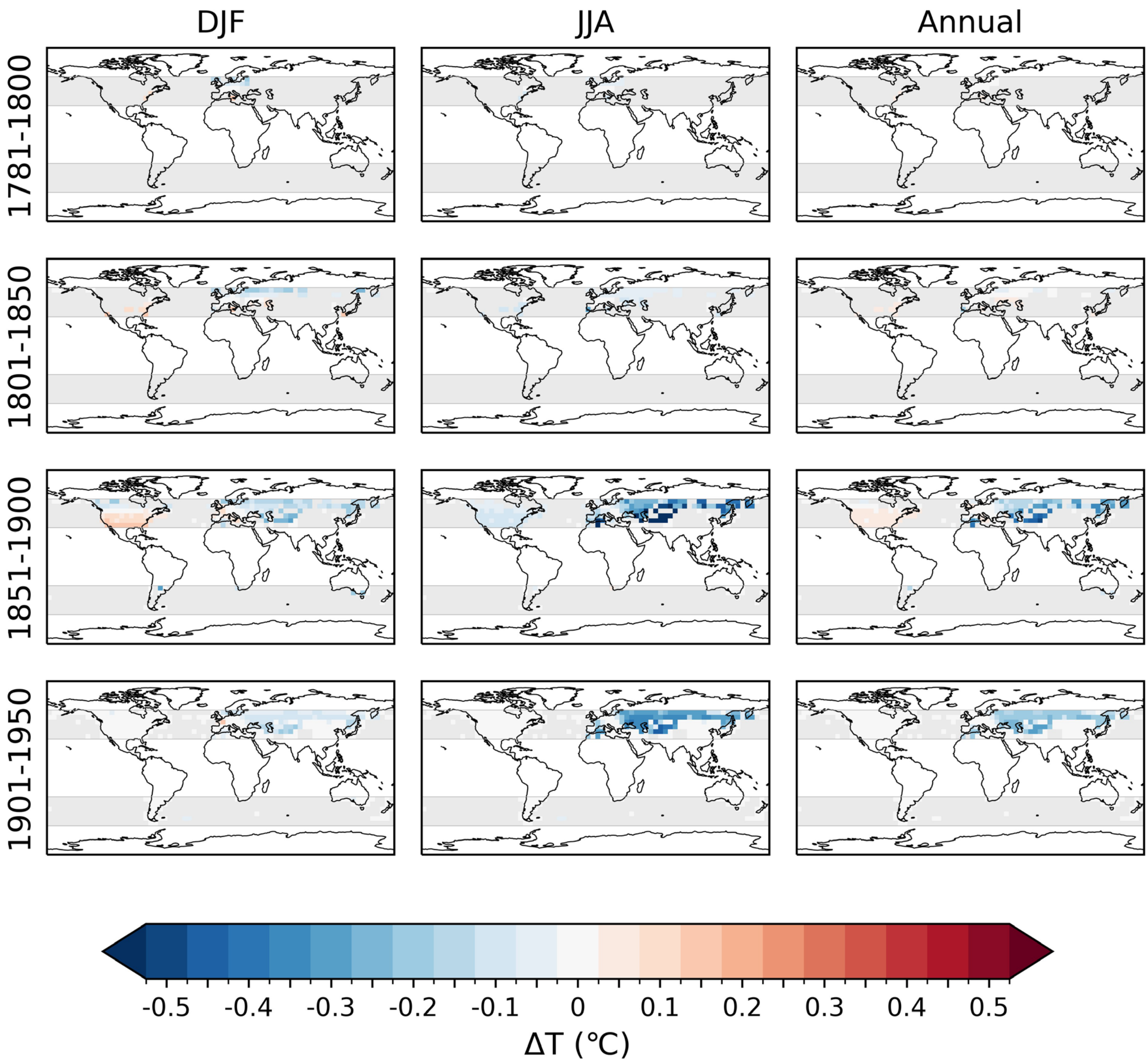

Figure 4: Impact of the exposure bias adjustments derived by Wallis et al. (2024) on the seasonal and annual mean temperatures in an extended version of CRUTEM5 which starts in 1781. ΔT is adjusted minus unadjusted CRUTEM5; the grey shading delineates the mid-latitude region the adjustments were applied to; DJF refers to the December-to-February mean and JJA the June-to-August mean.

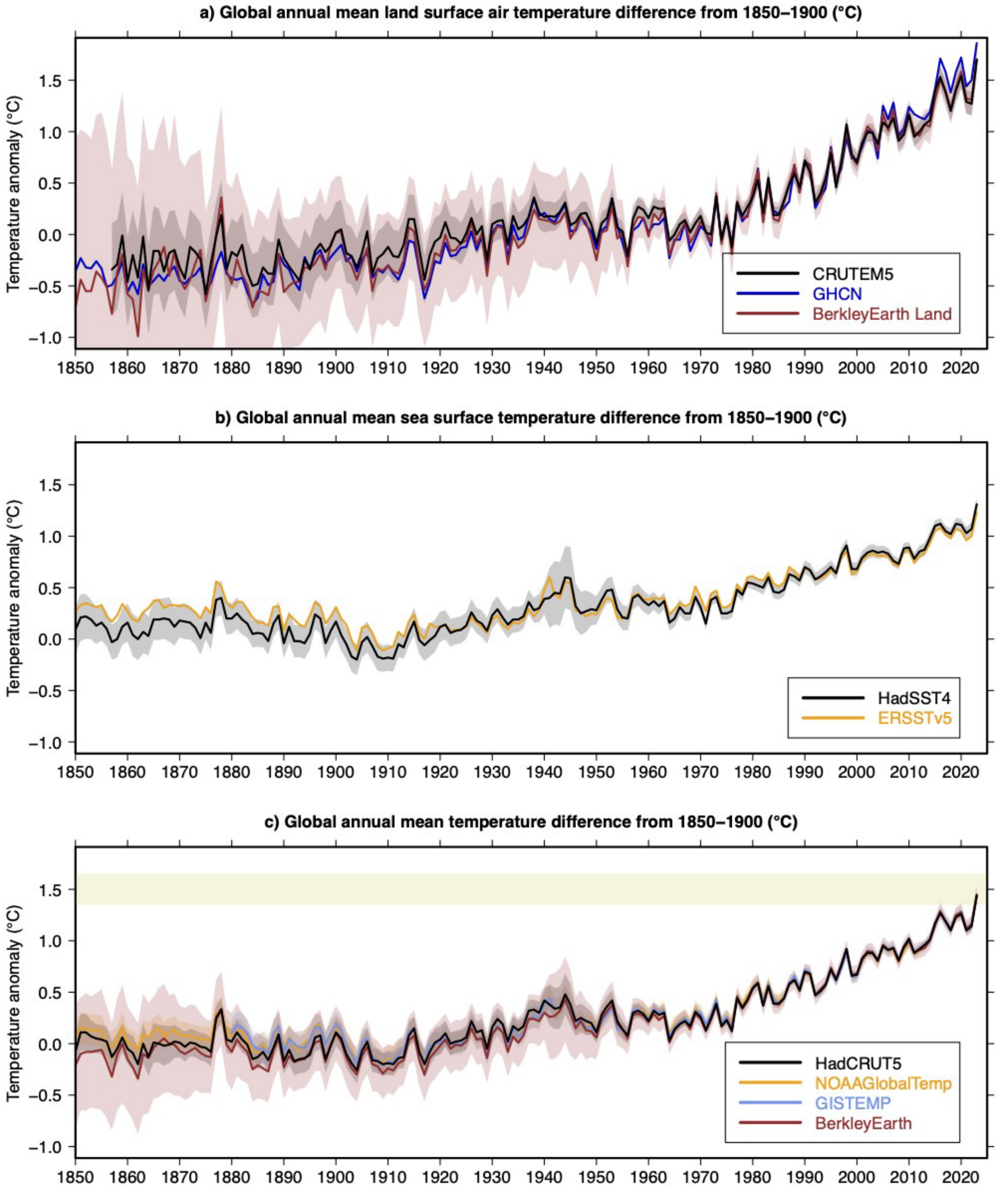

Figure 5: Global annual average timeseries of a) land surface air temperature; b) Sea-surface temperature; c) global surface temperature. Data have been downloaded from the Met Office Climate Dashboard (https://climate.metoffice.cloud/dashboard.html) under an Open Government Licence (https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/). See text for further information about the datasets used. Anomalies have been calculated relative to the average temperature from 1981 to 2010 and offset by 0.69 ˚C which is the best estimate of warming since the reference period 1850 to 1900 from IPCC (2023a). The shaded band centred on 1.5 ˚C in panel c) represents the limit from the Paris Agreement with the estimated uncertainty of the differences between the temperatures in 1850 to 1900 and the period 1986 to 2005 (±0.15˚C, IPCC, 2023a).

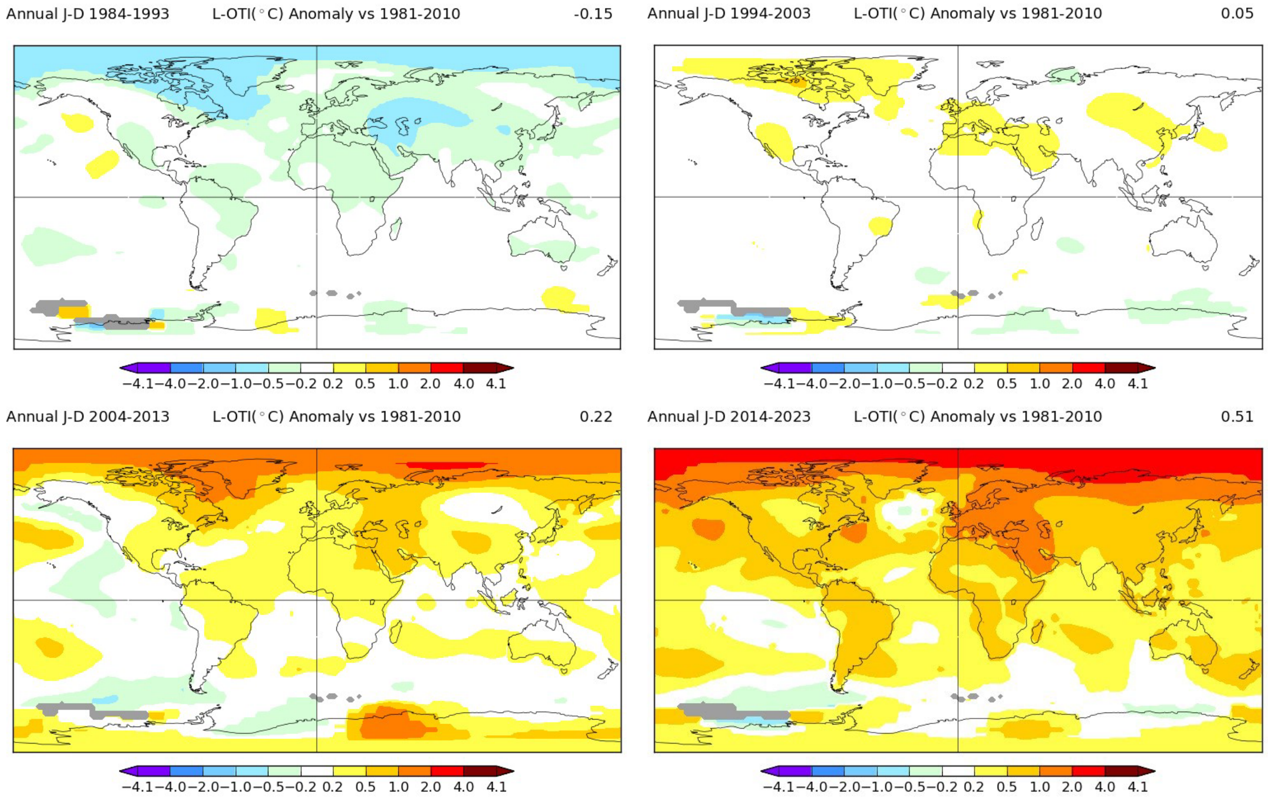

Figure 6: Decadal mean anomalies from a 1981-2010 base period for 1984-1993 (top left); 1994-2003 (top right); 2004-2013 (lower left); 2014-2023 (lower right). Figure generated from GISTEMP (Lenssen et al., 2019; GISTEMP Team, 2024) using NASA's Goddard Institute for Space Studies data viewer. Smoothing radius is 1200km, grey regions are grid-cells without data.

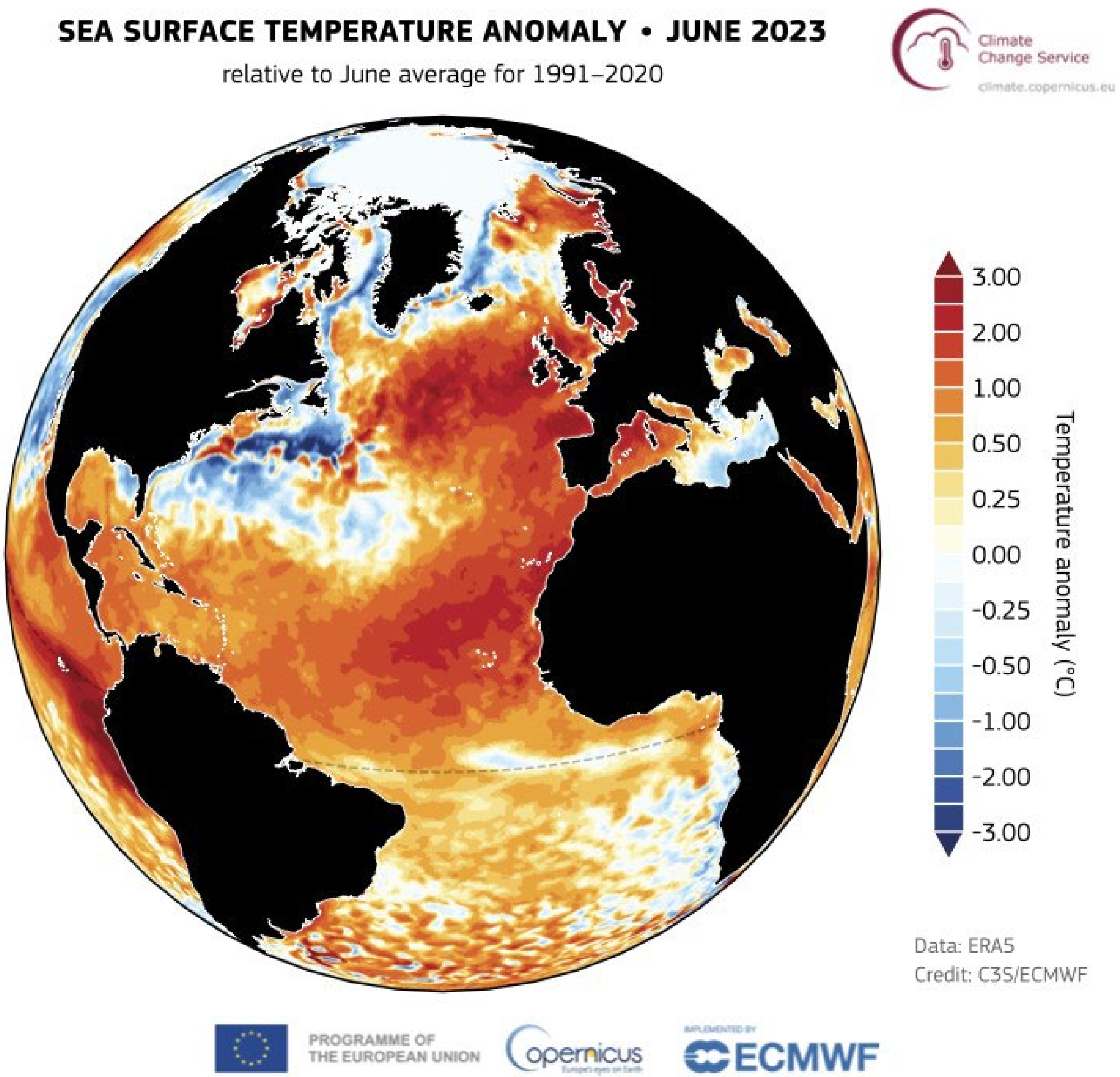

Figure 7: ERA5 (Hersbach et al., 2020) SST anomalies for June 2023 from https://climate.copernicus.eu/global-sea-surface-temperature-reaches-record-high. Credit: Copernicus Climate Change Service/ECMWF.