Quantifying biases and exploring natural climate variability in early instrumental temperature records

by

Emily Jayne Wallis

A thesis submitted to the School of Environmental Sciences
of the University of East Anglia in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

May 2024

Abstract

Early instrumental land surface air temperature (LSAT) records are valuable resources for climate science; however, many are compromised by the presence of non-climatic influences (inhomogeneities) which must be addressed before records can be used for climate analysis. This thesis presents an empirical approach to address a pervasive inhomogeneity affecting LSAT records – exposure bias – before using the exposure-bias-adjusted data, two reanalyses and model simulations to explore the climatic influence of volcanic eruptions; specifically, whether there is early instrumental evidence that explosive volcanism causes post-eruption Eurasian winter warming.

Exposure biases occur due to changes in the way thermometers have been exposed to the elements over time and have not been widely accounted for in global LSAT compilations. This thesis addresses the exposure bias in an extended version of CRUTEM5 (starting in 1781) by (1) analysing 54 parallel measurement series to better characterise the bias; (2) developing three statistical models to predict the bias from temperature and radiation variables; and (3) applying the models to mid-latitude stations in CRUTEM5 using a compilation of exposure metadata collated here. The models are applied to 1,960 stations, cooling summer temperatures in the Northern Hemisphere mid-latitudinal mean by ~0.2°C between 1882–1934, with smaller adjustments in other seasons and time periods.

Analysis of the first and second winters following eight explosive volcanic eruptions finds a consistent and significant volcanic signal of warmer winter temperatures in Europe, cooler near Greenland and a tendency for a positive NAO index in both the observations and reanalyses. In contrast, the UKESM1.1 model simulates a similar response only following the largest eruption, Tambora. The consistency of the volcanic signal in the observations and reanalyses, plus the modelled response following Tambora, suggests explosive volcanism does contribute to observed post-eruption Eurasian winter warming and the eruptions in the early instrumental period strengthen this evidence.

Contents

List of Tables

List of Figures

Acknowledgements

First, I would like to thank my supervisors: Tim Osborn, Michael Taylor, Manoj Joshi and Ed Hawkins, for their feedback, suggestions, encouragement and patience throughout the project. Special thanks particularly to Tim for being an excellent and supportive primary supervisor and to Michael for his help with Python and JASMIN. Thanks also to Phil Jones, David Lister and Ian 'Harry' Harris for welcoming me into their weekly 'CRU datasets' meetings and for providing valuable discussion and suggestions.

I would also like to thank all members of the GloSAT team and advisory board, firstly for making this PhD possible by devising the wider project and securing NERC funding (grant number NE/S015582), but mostly for providing me with an excellent introduction to academic group projects. Thanks particularly to Colin Morice for his work to incorporate the exposure bias adjustments into the GloSAT dataset; Stefan Brönnimann for providing access to the ModE-RA data and to Andrea Dittus, Andrew Schurer and Andrew Ballinger for running the GloSAT model simulations and for answering many of my volcano-model-related questions!

Thank you also to everyone who provided data for use in this thesis; particularly to those who provided parallel measurement series or who were involved in the production of the CRUTEM, GloSAT, ModE-RA or 20CRv3 datasets. Support for the Twentieth Century Reanalysis Project version 3 dataset is provided by the U.S. Department of Energy, Office of Science Biological and Environmental Research (BER), by the National Oceanic and Atmospheric Administration Climate Program Office, and by the NOAA Physical Sciences Laboratory.

To my fellow ENV, CRU and ARIES PhD cohort, and particularly everyone in office 01.04b, thanks for your camaraderie throughout the PhD, impromptu lunches, cake, badminton sessions and ongoing friendship!

It has been a huge privilege to get to know and work with you all.

On a personal level I would like to thank my friends, family and particularly my husband Daniel, for their unwavering support and belief in me. Thanks for always being my biggest cheerleaders and for all of your encouragement throughout the last four years!

Lastly, thank you to the early observers (and all those since) who devotedly observed and recorded the climate with little knowledge of the importance their measurements would have today. Their dedication is an inspiration and climate science owes them a huge debt.

Chapter 1

Introduction

1.1. A brief history of land surface air temperature observations

Observations of land surface air temperature (LSAT) have been recorded using instruments since the invention of the thermometer in the early seventeenth century and have been reported in scientific journals from at least 1669 (Knowles Middleton, 1966; Le Treut et al., 2007; Wallis & Beale, 1669). With the exception of the Medici Network, which was active between 1654-1667 (Camuffo & Bertolin, 2012), the majority of early observations were independent efforts made by interested individuals, scholars and religious figures, usually at universities, astronomical observatories or monasteries (Brönnimann et al., 2019a; Freeman et al., 2017; Hiebl et al., 2006; Rennie et al., 2014; Thorne et al., 2017). Observations were generally not made for the purposes of climate science but were used to assess the influence of climate on such things as health, agriculture and prosperity. The first coordinated observing networks did not appear until the eighteenth century; one of the earliest was that coordinated by James Jurin and the Royal Society in 1723, who organised standardised weather reports from locations in Europe, North America and India (Jurin, 1723). To many, however, the establishment of the *Societas Meteorologica Palatina* (Palatine Meteorological Society) in 1780 in Mannheim, (present-day) Germany, represents the first coordinated international meteorological network due to its focus on standardization (Cassidy, 1985; Hiebl et al., 2006; Kington, 1974; Pappert et al., 2021). The network was fully funded, supplied both instruments and instructions for making observations to observers and brought together information from 37 weather stations in the Northern Hemisphere (NH), stretching from eastern North America to Russia (Cassidy, 1985). These early networks were the predecessors to the first National Meteorological Services (NMSs), which were established from the 1850s (Brönnimann et al., 2019b; Hiebl et al., 2006), and the first formal International Meteorological Organisation (precursor to the World Meteorological

Organisation) which was established in 1873 (Meteorological Committee, 1874). The establishment of NMSs marks the start of greater standardisation of observation times, methods and instrumentation and any instrumental observations made before their advent are considered to be 'early instrumental records' (Brönnimann et al., 2019b).

Thanks to the dedication of early observers and the coordination of early observing networks, a surprising number of early instrumental records exist. The longest continuous records of monthly mean LSAT are the Paris Montsouris (Rousseau, 2015) and Central England Temperature (Manley, 1974; Parker et al., 1992) series which start in June 1658 and January 1659, respectively, but a number of other long records exist. Some of the longest continuous records are from Berlin, Germany (1701), De Bilt, Netherlands (1706) (van Engelen & Nellestijn, 1995), Uppsala, Sweden (1722) (Bergström & Moberg, 2002; Moberg & Bergström, 1997), St Petersburg, Russia (1743) (Jones & Lister, 2002) and the St Lawrence Valley, Canada (1743) (Slonosky, 2014). In total, meteorological observations at more than 2,250 locations are believed to have been made prior to the establishment of NMSs in 1850 (1890 for Africa and the Arctic), although not all have been located or digitised (Brönnimann et al., 2019b).

1.2. The value of early instrumental temperature observations

Observations of surface air temperature are a hugely valuable resource. They form the longest available instrumental record of climate (Chen et al., 2021), are designated an essential climate variable (World Meteorological Organization, 2022) and underpin much of our knowledge of the climate system (Thorne et al., 2017). Long, continuous records stretching from the early instrumental period (EIP) to present are of particular value as they provide the opportunity to assess temperature variability and change over long (centennial) timescales, including a period in the nineteenth century of high volcanic activity (Toohey & Sigl, 2017) and of little anthropogenic impact on the climate (Hegerl et al., 2019). Such records facilitate an improved (longer-term) understanding of the climate system, including its internal variability and how it interacts with natural and anthropogenic forcing, and provide an opportunity to more accurately quantify the anthropogenic contribution to recent warming. Long instrumental records also provide an important link between past and future climates via their use to calibrate and validate proxy reconstructions (i.e. PAGES2k Consortium (2017)) and to assess and constrain climate models which can be used to project future climate (Eyring et al., 2019; Waliser et al., 2020).

The value of early instrumental series is not purely academic. Weather and climate have important impacts on societies (i.e., via their influence on agricultural productivity, water availability, heath and mortality, migration and tourism) and in a changing climate the severity and frequency of these impacts is expected to increase (IPCC, 2022). In this context, long early instrumental records are extremely valuable to a) better understand how climate has evolved and impacted societies in the past (Ljungqvist et al., 2021), b) place current weather events, and particularly extreme events, in a longer-term context and determine the likelihood of recurrence (Hawkins et al., 2023; Yule et al., 2023) and c) to improve projections of future climate and climate-related impacts via improved parameterization, initialization and assessment of models (Gettelman et al., 2022; Hawkins & Sutton, 2009; Ortega et al., 2022). This is vital to provide the best possible information about the range of future climate impacts to stakeholders and policymakers and thus to effectively plan adaptation and mitigation measures (Dee et al., 2021).

Finally, the Paris Agreement, which commits signatories to "[holding] *the increase in the global average temperature to well below 2°C above pre-industrial levels* (…)" (United Nations, 2015, p. 3) places additional importance on long records of surface air temperature by linking legally binding targets to global temperature change. While improved quantification of post-industrial warming is unlikely to affect the target itself (as the target is likely to be measured against the best estimate of post-industrial warming at the time of ratification (0.85°C; (IPCC, 2013)), improved analysis of warming could help to refine estimates of the transient climate response to cumulative carbon emissions (Rogelj et al., 2019; Schurer et al., 2018; Tokarska et al., 2019). This in turn will improve the accuracy of the remaining global carbon budget and thus allow a more accurate understanding of the global and national mitigation efforts required to meet the Paris Agreement (Rogelj et al., 2019; Schurer et al., 2017).

1.3. Inhomogeneities in early instrumental LSAT records

Section 1.2 highlights the numerous areas early instrumental records can benefit both climate science and climate adaptation and mitigation. As a result, there has been a recent move to temporally extend global temperature compilations (which generally start between 1850-1880 when NMSs were established (Menne et al., 2018; Osborn et al., 2021)) via the incorporation of early instrumental records (e.g. Rohde et al. (2013)). Thanks to national and international data rescue efforts (such as Alcoforado et al. (2012); Allan et al. (2011);

Ashcroft et al. (2014); Brugnara et al. (2020); Brunetti et al. (2006); Camuffo & Jones (2002); Cappelen (2021a, 2021b); Demarée et al. (2002); Domínguez-Castro et al. (2014, 2017); Gergis et al. (2022); Pappert et al. (2021); Westcott et al. (2011) and Williamson et al. (2018)), as well as data cataloguing efforts (Lundstad et al., 2023; Rennie et al., 2014), numerous early instrumental series are now accessible to climate science, including nearly 300 series in the CRUTEM5 station database (Osborn et al., 2021). The incorporation of such records into global temperature compilations, however, is not straightforward.

One of the key challenges associated with using early instrumental temperature records is ensuring the data are homogenous. For a series to be homogenous it should exclusively reflect variations in weather and climate and should not contain artificial changes arising from non-climatic influences, known as breakpoints or inhomogeneities (Conrad & Pollak, 1950). Unfortunately, this is rarely the case for long series, particularly those originating from the EIP which were not initially recorded for the purposes of climate science and whose records encompass multiple advancements in instrumentation and a move toward greater standardization of observing practises (Section 1.1). Previous homogeneity assessments of European series suggest (detectable) inhomogeneities occur once every 15 to 20 years; if this metric holds true for early instrumental series, then the longest records could contain close to 20 breakpoints or non-climatic influences (Venema et al., 2012; World Meteorological Organization, 2020).

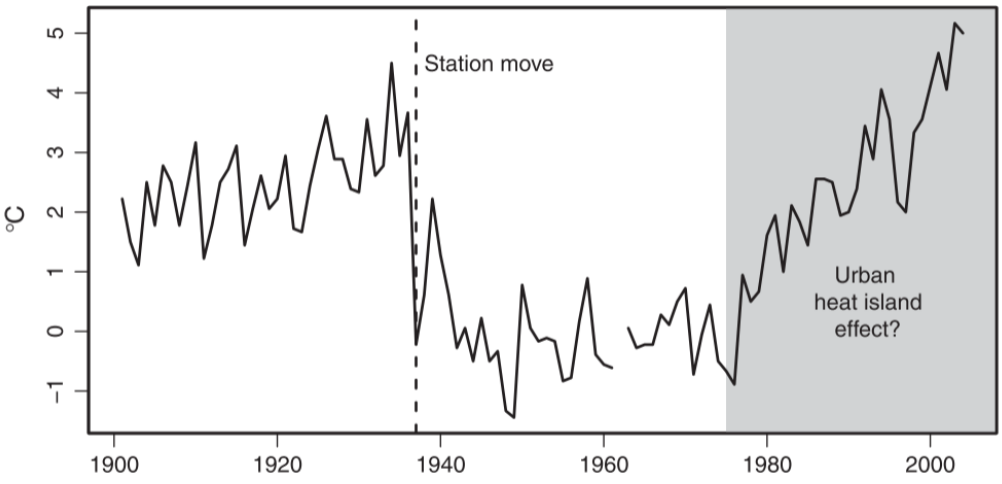


Figure 1.1. An example of two inhomogeneities present in the annual mean minimum temperatures recorded at a weather station in Reno, Nevada. The vertical dashed line indicates the timing of a known station move and the shading indicates where the temperatures may be affected by urbanisation. Figure from Chandler et al. (2012).

Table 1.1. Sources of inhomogeneity in LSAT records.

Source	Description
Site relocation	Any relocation can introduce inhomogeneities into timeseries, but inhomogeneities can be especially large when weather stations are relocated between **urban and rural** locations (Dienst et al., 2017) or relocated in regions with a **steep temperature gradient** (i.e. coastal or mountainous regions). The former move was common in Australia and the United States: weather stations were often established in urban areas and later moved to airports on the outskirts of cities (Hansen et al., 2001; Hausfather et al., 2013; Trewin, 2018).
Changes in the local environment	Changes in the local environment can include small scale changes, such as the gradual **deterioration of the thermometer screen,** the **growth of vegetation** or the **addition of buildings**/structures nearby, as well as larger scale changes. **Urbanization** is the best-known example of the latter (Hausfather et al., 2013; Wang & Yan, 2016) and has led to an estimated warm bias in global datasets of 0.006°C per decade since 1900 (Jones et al., 1990). Another large-scale example is changes to **irrigation** practises in the surrounding area - irrigation can artificially lower temperatures leading to a cool bias (Cook et al., 2014; Lobell & Bonfils, 2008).
Instrumentation	Any changes in the instruments used to record temperatures can introduce non-climatic changes into timeseries. Inhomogeneities related to changes to **thermometer exposure**, differences between **spirit and mercury thermometers** (Camuffo, 2002a), **thermometer drift** (Camuffo, 2002a; Winkler, 2009) and differences between thermometers and **temperature sensors** (Burt & Podesta, 2020; Hannak et al., 2020), have all been widely documented.
Observing practices	Inhomogeneities can be introduced due to any change in observing practice. Some of the most common inhomogeneities in this category arise from changes to the **time of observation** (Karl et al., 1986; Vincent et al., 2009), the method used to **calculate the daily-mean** (Trewin, 2004), the **units** of measure used, or the level of **precision** recorded (see Trewin (2010) and Camuffo (2002b) for discussion).

Inhomogeneities can arise from numerous sources, but largely fall into four categories: changes in the location or surroundings of a weather station, changes to instrumentation and changes to observing practises (see Table 1.1). They can take the form of abrupt changes (e.g., due to a change in site location), gradual changes or trends (e.g., urbanisation or thermometer drift) and can also affect the variance and seasonal or diurnal cycle of series (see Figure 1.1) (Willett et al., 2014). Regardless of source, inhomogeneities can be large in magnitude, of a similar size to true climatic changes, and thus must be accounted for before series can be used for climate assessment (World Meteorological Organization, 2020). There

are two main steps to account for inhomogeneities: *detection*, which identifies the presence of inhomogeneities in a series, and *adjustment*, which adjusts the data to minimize the inhomogeneity (World Meteorological Organization, 2020). Where inhomogeneities randomly affect individual stations, they can often be detected and adjusted via comparison with nearby observational data (Jones et al., 1986) and/or by using homogenisation algorithms (see Domonkos et al. (2021) and Venema et al. (2012) for an overview). As a result, random inhomogeneities are generally well-accounted for and, even where they are not, they generally do not result in large errors in global temperature compilations as their random nature means they mostly cancel out over large spatial scales (Jones, 2016; Venema et al., 2012). Where inhomogeneities are more problematic is where they affect a large proportion of the weather stations in a region or observing network. These inhomogeneities - referred to as biases - lead to highly spatially-correlated errors which are hard to detect and which compound over large spatial scales. Such errors present a significant and pervasive source of uncertainty in global temperature compilations and pose a major challenge to using early instrumental temperature records for climate assessment.

1.3.1. Exposure bias

One of the most significant biases affecting land surface air temperature records in the EIP stems from the differential exposure of thermometers to solar radiation. Prior to the widespread adoption of the louvred Stevenson screen (Stevenson, 1866) in the late-nineteenth and early-twentieth centuries, various methods were employed to protect thermometers from exposure to solar radiation and the elements (Knowles Middleton, 1966; Parker, 1994; Trewin, 2010). Each method exposed the thermometer to differing levels of solar radiation (both direct and indirect) and affected temperature readings differently, thus introducing a bias into global temperature records when the transition to Stevenson screens was made (Jones, 2016; Parker, 1994; Trewin, 2010).

Despite the exposure bias being a widely documented issue (i.e. Böhm et al. (2010); Brunet et al. (2006, 2011); Nicholls et al. (1996); Parker (1994); Trewin (2010) and Jones (2016)), relatively few adjustments have been applied or incorporated into global temperature compilations. This is because Stevenson screens were often introduced across NMSs quasi-simultaneously and often without documentation at individual stations, making the bias both difficult to detect using traditional homogenisation techniques (as the bias affects the majority of stations in a region) and making it difficult to determine the appropriate bias

adjustment even when a bias is known to have occurred (discussed further in Chapter 2). As a result, explicit adjustments for the exposure bias have, to date, only been applied to a relatively small number of individual weather stations or regional networks where a) the time of Stevenson screen introduction was known and b) where parallel measurements comparing the Stevenson screen and the exposure it replaced allowed the appropriate bias adjustment to be determined, either directly (Ashcroft et al., 2022; Brunet et al., 2006; Butler, Garcıa-Suarez, et al., 2005), or via the development of statistical (Böhm et al., 2010; Brunet et al., 2011) or physical models (Auchmann & Brönnimann, 2012). The lack of more widespread adjustment means the exposure bias likely still represents a significant source of uncertainty in LSAT records, particularly in long records from the EIP.

This uncertainty is acknowledged in existing global temperature datasets. The HadCRUT5 (Morice et al., 2021) (and CRUTEM5 (Osborn et al., 2021)) error model, for example, includes a distinct term to account for the uncertainty arising from the transition to Stevenson screens. This term makes the assumption that errors associated with the exposure bias are present in the underlying station database until 1930 (1950 in the tropics) and that the biases are correlated across all tropical and (separately) all extratropical weather stations (Morice et al., 2021; based on Folland et al. (2001)). The representation of the bias in the error model, however, was designed primarily to capture the global-scale effect of the exposure bias on the annual mean surface temperature and therefore makes some oversimplistic assumptions. For example, the error model assumes a fixed annual error, despite the well-documented seasonal nature of the bias (e.g. Parker (1994)), and applies blanket error terms to the tropics and extratropics which do not account for regional differences in the pre-Stevenson screen exposures in use or regional differences in the timing of the transition to the Stevenson screen. With the recent increased focus on the nineteenth century and the early instrumental data (Section 1.2), and the move to incorporate such data into global temperature datasets, this issue deserves additional attention to better characterise the exposure bias and to develop more representative models of the bias. To do this, an approach to both detect the bias in large scale datasets and to model the appropriate bias adjustments for individual stations/regions is required. This is one of the two key aims of this thesis.

1.4. Natural climate variability in the early instrumental period

Once inhomogeneities and biases (along with other uncertainties) have been accounted for in early instrumental temperature series, extended global temperature compilations can be

used to address some of the scientific and societal challenges outlined in Section 1.2. As noted in Section 1.2, one of the defining (climate-related) elements of the early instrumental period is the presence of four explosive volcanic eruptions in the early nineteenth century (Figure 1.2): two unknown eruptions in 1809 and 1831, Tambora in 1815 (which is the largest eruption since the thirteenth century (Raible et al., 2016)), and Cosiguina in 1835 (National Geophysical Data Center / World Data Service (NGDC/WDS), n.d.; Toohey & Sigl, 2017). These eruptions were the dominant cause of natural, externally forced, climate change in the EIP and the new extended global temperature compilations provide an exciting new opportunity to better understand their climatic impact using instrumental observations, rather than models or palaeoclimate data (Crowley, 2000; Hegerl et al., 2019; Schurer et al., 2014).

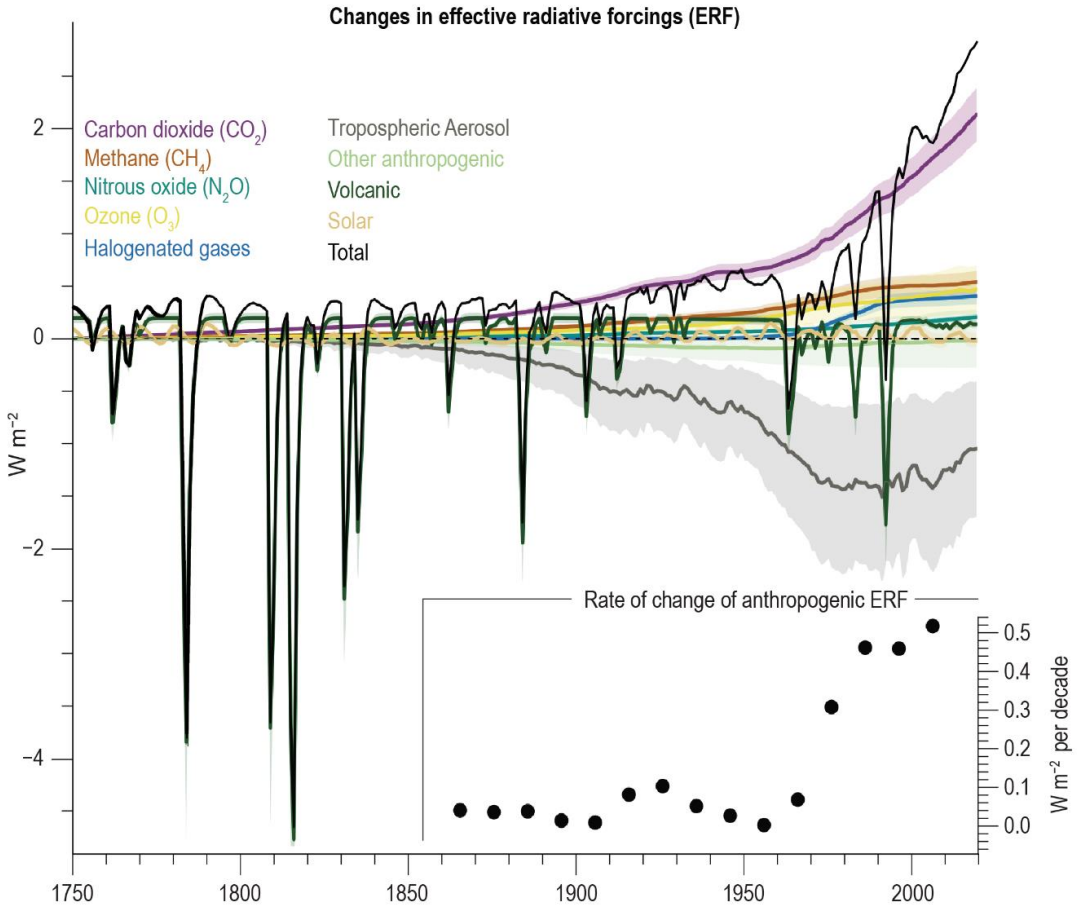


Figure 1.2. Temporal evolution of effective radiative forcing (ERF) over the instrumental record. The ERF associated with volcanic eruptions is shown in dark green. Note the four negative spikes in the early nineteenth century; these are the result of the eruptions noted in Section 1.4. Figure is Fig. 2.10 from Gulev et al. (2021).

Explosive volcanic eruptions influence the climate through the ejection of magmatic material (e.g. ash) and gases, including sulphur species, into the atmosphere. The magmatic material

rapidly falls out of the atmosphere (in minutes-to-weeks), causing only localized short-term climatic impacts. However, the sulphur-containing gases (primarily sulphur dioxide) oxidise and condense to form sulphate aerosols which, if they reach the stratosphere, rapidly disperse globally and can persist in the atmosphere and perturb the climate for several years (Bluth et al., 1992; Deshler, 2008; Mass & Robock, 1982; Niemeier et al., 2009). Sulphate aerosols are well-known to lead to cooling at the Earth's surface - following the eruption of Pinatubo in 1991, global surface air temperatures were up to 0.5°C cooler than pre-eruption and remained anomalously cool for a further two summers (Parker et al., 1996), and 1816 (the year after the eruption of Tambora) is famously known as the 'Year Without a Summer' due to the anomalously cold and wet conditions experienced in Europe and North America (Cole-Dai et al., 2009; Luterbacher & Pfister, 2015; Schurer et al., 2019). However, the presence of sulphate aerosols in the atmosphere is also thought to interact with natural modes of variability (e.g., the El Niño Southern Oscillation, ENSO (Dee et al., 2020; Dee & Steiger, 2022; Zhu et al., 2022) and the North Atlantic Oscillation, NAO (Christiansen, 2008)) and lead to dynamic climatic effects which are less well-understood.

At present there is discussion in the literature regarding whether explosive volcanic eruptions play a role in the warmer winter temperatures observed over continental Eurasia following recent explosive eruptions. Some schools of thought suggest they do via an interaction with the polar vortex and the NAO (e.g., Azoulay et al. (2021); Christiansen (2008); Stenchikov et al. (2002)), however others (e.g. Polvani et al. (2019); Polvani & Camargo (2020)) have argued the warmer temperatures are merely the result of internal variability and unrelated to the preceding volcanic eruptions. The majority of previous studies which assess this issue have been based on the results of modelling studies or reanalyses (e.g., Polvani et al. (2019); Polvani & Camargo (2020); Zambri et al. (2017)), palaeoclimate data (e.g., Fischer et al. (2007); Shindell et al. (2004)) or based on a small number of eruptions which are encompassed by current global temperature compilations (e.g., Kelly et al. (1996); Robock & Mao (1995)). The early instrumental data provide an opportunity to expand the observational evidence for (or against) a dynamic atmospheric response to explosive volcanism by analysing the post-eruption winter temperatures following the four explosive eruptions noted earlier, which have not previously been assessed using a global gridded observation-based dataset. Thus, the second key aim of this thesis is to answer the question: 'Do the early nineteenth century eruptions support the evidence for volcanically induced winter warming over Eurasia?'.

The two key aims of this thesis are linked as the main temperature dataset used to address this latter question incorporates the improved exposure bias adjustments that are an outcome of achieving the first key aim of the thesis.

1.5. Thesis structure

This thesis is divided into two main parts: Part I outlines an approach to quantify the exposure bias in global temperature compilations, using a version of CRUTEM5 extended back to 1781 (CRUTEM5_ext) as an example, and Part II uses the exposure-bias-adjusted data to explore natural climate variability in the early instrumental period, particularly looking at the impact of early nineteenth century volcanic eruptions on winter climate over continental Eurasia and the North Atlantic.

Part I is composed of Chapters 2 and 3. Chapter 2 starts by describing the collation and empirical analysis of 54 parallel measurement series to better characterise the exposure bias arising from the transition to the Stevenson screen in four categories of early exposure. The development and assessment of four regression-based bias-estimation models, based on an analysis of the relationship between the magnitude of the monthly mean exposure bias and three variables considered *a priori* to influence it, is then presented.

Chapter 3 begins with an assessment of the different methods which could be used to identify the exposure bias in global LSAT compilations before outlining the development of a database of exposure metadata for the CRUTEM5 station database (CRUTEM5_sdb). The second half of Chapter 3 then presents the application of three of the bias-estimation models developed in Chapter 2 to the CRUTEM5_sdb, using the newly collated exposure metadata, to quantify the bias present and to produce an exposure bias adjusted version of CRUTEM5_ext (CRUTEM5_eba). Elements of Chapters 2 and 3 have been published in the *International Journal of Climatology*.

Part II of this thesis is composed of Chapter 4, which exploits the newly exposure-bias-adjusted data - in the form of a new global surface air temperature dataset (GloSAT) which uses CRUTEM5_eba as the land component along with a new dataset of marine air temperature anomalies over the ocean (not produced as part of this PhD, but as part of the wider GloSAT project that this PhD is part of) - to explore the natural climate variability present in the early industrial period. Specifically, Chapter 4 uses the GloSAT data, which starts in 1781, and an observation-based NAO index (Jones et al., 1997) to assess the impact of eight explosive low-latitude eruptions on post-eruption winter temperatures

and the NAO index. The results from analysing these instrumental datasets are compared with those obtained from two reanalyses and six model runs to assess whether there is evidence that explosive volcanism leads to winter warming over Eurasia.

Chapter 5 concludes the thesis with a summary of the key findings, limitations and suggestions for further research.

Several observed surface temperature datasets are used or developed throughout this thesis; a description of each is included in Table 1.2, below, for information.

Table 1.2. Global observed temperature datasets developed and/or used in this thesis.

Dataset	Description	Start	Resolution	Infilled?
CRUTEM5_sdb	Database of weather station temperature data which forms the land component of the following datasets	1658	Monthly	N/A
CRUTEM5 Osborn et al. (2021)	Global, gridded LSAT anomalies w.r.t. 1961-90	1850	5° x 5° Monthly	No
CRUTEM5_ext	An extended version of CRUTEM5 starting in 1781	1781	5° x 5° Monthly	No
CRUTEM5_eba Developed in Ch. 3	A version of CRUTEM5_ext, adjusted for the exposure bias at stations within 30° to 60° latitude	1781	5° x 5° Monthly	No
GloSAT Morice *et al.* [in prep]	Global, gridded surface air temperature anomalies w.r.t. 1961-90. CRUTEM5_eba forms the land component of this dataset	1781	5° x 5° Monthly	Yes

Part I

Quantifying the exposure bias in early instrumental temperature records

Chapter 2

Characterising and quantifying the exposure bias

2.1. Introduction

As is outlined in Chapter 1, land surface air temperature (LSAT) observations are vital to advancing knowledge of climate variability and change. They form a key component of global surface air temperature datasets used for climate assessment (e.g., Menne et al. (2018) in Lenssen et al. (2019) and Osborn et al. (2021) in Morice et al. (2021)) and are also used to calibrate many temperature palaeoreconstructions (e.g., Anchukaitis et al. (2017); PAGES2k Consortium (2017)). However, many LSAT records are compromised by non-climatic changes in the data, known as inhomogeneities, which require correction or consideration before they can be used for climate assessment (World Meteorological Organization, 2020). Where inhomogeneities affect individual stations, correction is often possible via comparison with neighbouring series or the use of statistical homogenisation techniques (Domonkos et al., 2021; Jones, 2016; Venema et al., 2012); however, where inhomogeneities - known as biases - systematically affect a large proportion of the observations in a region, these traditional methods may be insufficient to identify and correct them. This means that biases potentially still exist in LSAT records and contribute significant uncertainty to global temperature compilations (e.g., HadCRUT5; Morice et al., (2021)).

One bias affecting LSAT records is the exposure bias which has been introduced into temperature records due to changes in the way thermometers have been exposed over time. Thermometer exposure refers to the way thermometers are sheltered from the elements, principally solar radiation, in order to achieve a temperature reading which reflects the 'true' (shade) temperature of the air (Abbe, 1888; Hazen, 1885; Renou, 1875; Trewin, 2010; Wild, 1873). Since the early-twentieth century, thermometers have been relatively uniformly exposed, principally in variants of the Stevenson screen (Figure 2.1a; Stevenson (1866)) and (more recently) multiplate shields, following a push by the International Meteorological Committee for standardisation (Ellis, 1897; Parker, 1994; Sparks, 1972; Trewin, 2010).

Prior to the widespread adoption of variants of the Stevenson screen in the late-nineteenth and early-twentieth centuries, however, various (often inadequate) methods were employed to protect thermometers from exposure to solar radiation and the elements (Parker, 1994; Sparks, 1972; Trewin, 2010). These early methods varied regionally and included mounting thermometers on poleward-facing walls, stands, and within various freestanding screens (Figure 2.1b-e) (Knowles Middleton, 1966; Parker, 1994). In Great Britain and Commonwealth countries, for example, the Glaisher stand (Figure 2.1b) was popular historically, whereas in tropical countries thatched sheds were more common, and in the former Soviet Union the Wild hut, either freestanding (Figure 2.1e) or wall-mounted, was the predominant exposure in the late-nineteenth century (Gorczynski, 1910; Naylor, 2019; Nicholls et al., 1996; Parker, 1994). Each type of exposure influenced temperature readings differently, by altering the influence of solar radiation and other elements on the thermometer (Figure 2.2), thus introducing inhomogeneities into station temperature records when the transition to Stevenson-type screens was made (Parker, 1994). These inhomogeneities are known as exposure biases.

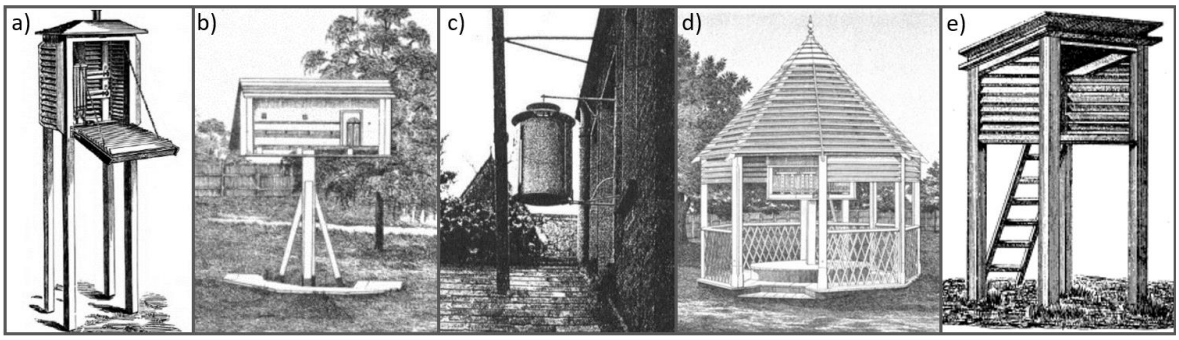


Figure 2.1. Examples of common thermometer exposures. a) Stevenson screen, b) Glaisher stand (Open), c) a type of wall-mounted screen, d) Summerhouse (Intermediate) and e) Wild Hut (Closed). Image sources: a) Gaster (1882); b) d) Royal Society of New South Wales; c) Mawley (1897); e) Wild (1891).

Note that the term ‘exposure bias’ is also used to refer to biases arising from a more recent transition from thermometers exposed in Stevenson screens to automatic temperature sensors exposed in multiplate shields, which has occurred as part of the transition to automatic weather stations (AWS). As the focus here is on improving early instrumental records (in part to support the creation of improved datasets for analysing the climate effect of the early nineteenth century eruptions: see Part II of this thesis), the exposure bias related to this later transition is not considered here but is discussed further in Chapter 5.

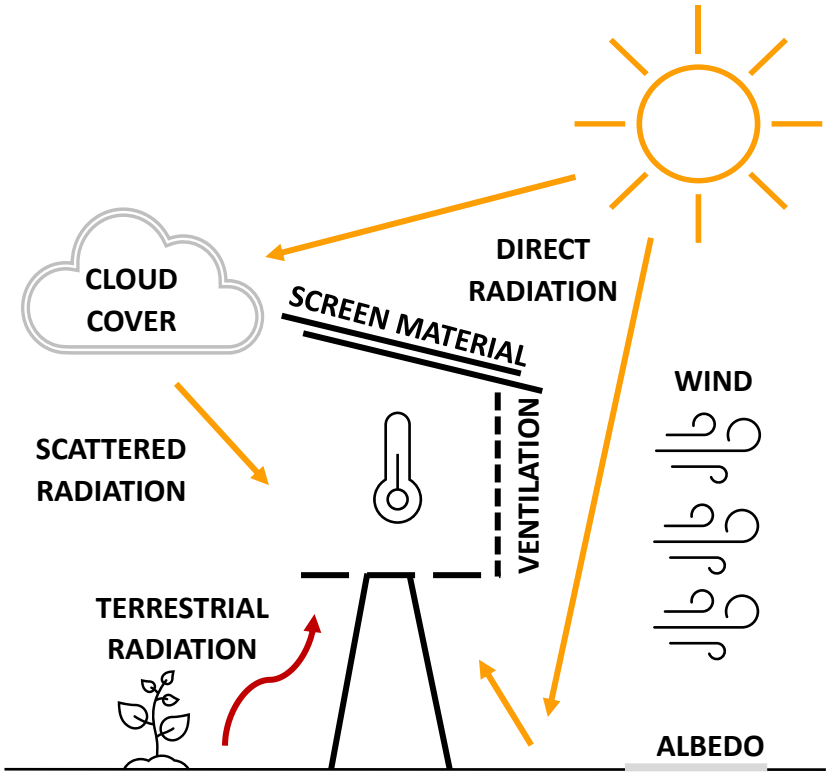


Figure 2.2. Schematic showing some of the factors which can influence the air temperature measured by thermometers. Image adapted from: Venema (2016).

The impact of differing thermometer exposures on temperature readings has been investigated by numerous studies. In the late-nineteenth and early-twentieth centuries, interest in determining the 'true' temperature of the air and agreeing a uniform, 'best', method of thermometer exposure led to multiple comparisons between exposures being conducted (e.g. Marriott (1879, 1894); Gaster (1882); Gill (1882); Whipple (1883); Doberck (1887); Sprung (1890); Ellis (1891); Mawley (1897); Field (1920); Dines (1921); Margary (1924)). One of the earliest and most well-known was Gaster (1882) who compared nine thermometer exposures, over two years, in the United Kingdom. These early studies, along with more recent analyses by Chenoweth (1993) and Slonosky (2014) in North America; Muller (1984), Nordli et al. (1996, 1997), Butler et al. (2005), Brunet et al. (2006, 2011) and Böhm et al. (2010) in Europe; Ashcroft et al. (2022) and Nicholls et al. (1996) in Australia, Awe et al. (2022) in Mauritius, and Parker (1994), globally, all present similar findings – significant differences in temperature readings between Stevenson screens and historic exposures, which vary seasonally, diurnally and according to weather condition and type of exposure.

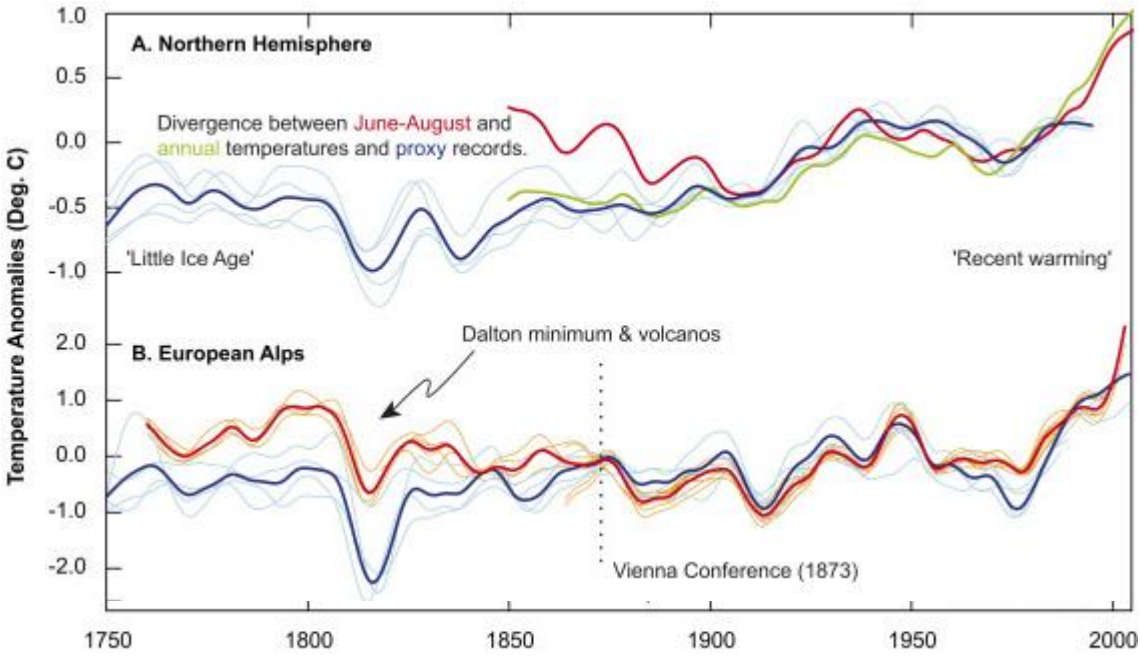


Figure 2.3. A comparison between warmer early instrumental temperatures and cooler proxy reconstructions in a) the Northern Hemisphere and b) the European Alps. Individual proxy reconstructions are shown in light blue (a): Briffa (2000); D'Arrigo et al. (2006); Esper et al. (2002) and Jones et al. (1999); b): Büntgen et al. (2005, 2006) and Frank & Esper (2005)) and their mean in dark blue. In a) instrumental annual mean temperatures are shown in green and June-to-August temperatures are shown in red (both Brohan et al. (2006)) and in b) individual June-to-August instrumental temperatures are shown in orange and their mean in red (both Auer et al. (2007) and Böhm et al. (2001)). Figure adapted from Frank et al. (2007).

Possible evidence of the exposure bias has also been documented in LSAT records. Folland et al. (1990) and Jones et al. (2003), for example, both identified seasonal differences in nineteenth and twentieth century warming rates in the Northern Hemisphere (NH) in early versions of CRUTEM (Jones, 1988; Jones & Moberg, 2003), with the winter half-year warming more than the summer half-year. Moberg et al. (2003) identified a possible warm bias of between 0.5°C to 0.8°C in pre-1860 summer temperature readings in Sweden (Stockholm and Uppsala) by comparing instrumental temperatures with temperature-correlated variables, such as cloud cover and air pressure, as well as long instrumental series from Central Europe. And, finally, Frank et al. (2007) identified a divergence between (cooler) proxy reconstructions and (warmer) instrumental June-to-August temperatures in the nineteenth century in the NH and the European Alps (Figure 2.3). All were suggested to be the result of ineffective thermometer exposures prior to the introduction of the Stevenson screen resulting in summer temperatures that were too warm.

Despite the well-documented differences found in the parallel measurement studies, and assessments presenting evidence of the likely presence of the exposure bias in early temperature observations, relatively few exposure-bias-specific corrections have been made to long temperature series. Of the series in the CRUTEM5 station database, for example, only selected records from Australia (Ashcroft et al., 2012), the Greater Alpine Region in Europe (Böhm et al., 2010) and Spain (Brunet et al., 2006) are known to have been explicitly adjusted to account for exposure biases (Morice et al., 2012). The lack of more widespread adjustment is largely due to the fact stations within regions or meteorological networks introduced the Stevenson screen quasi-simultaneously, and often without documentation, making the bias difficult to identify and rendering traditional approaches to breakpoint detection and adjustment less effective (Brunet et al., 2011; Jones, 2016; Trewin, 2010; World Meteorological Organization, 2020). Even where a bias is known to be present, determining the appropriate adjustment is problematic due to the seasonal nature of the bias and the number of variables which are expected to influence its characteristics (Willett et al., 2014). These factors, combined with a lack of available or accessible metadata, mean a large proportion of long LSAT records likely retain biases related to the introduction of Stevenson-type screens (Morice et al., 2021; Trewin, 2010).

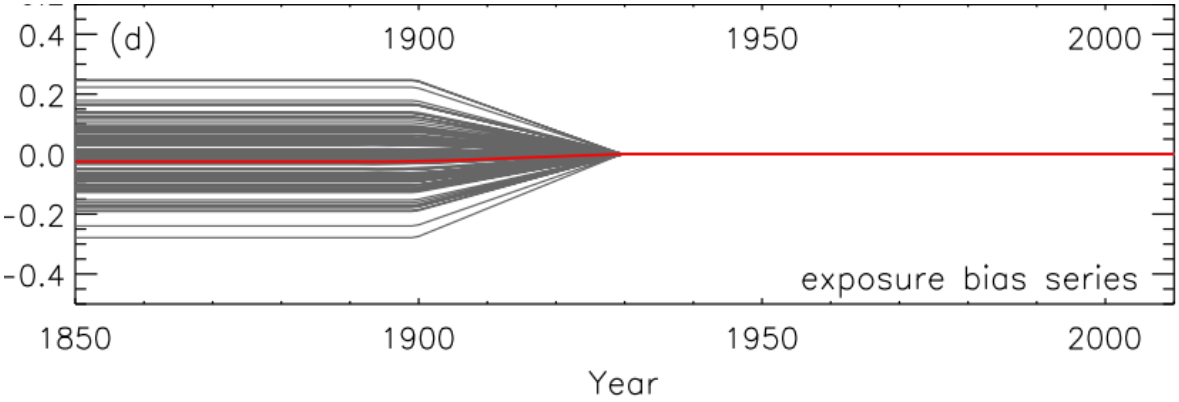


Figure 2.4. One hundred realisations of the exposure component of the HadCRUT5 error model for stations outside 20°S-20°N. Figure from: Morice et al. (2012).

As such, it is necessary to account for exposure biases in long observational records, including in global temperature datasets (gridded and global/hemispheric means). The HadCRUT5 dataset does this by including uncertainties from "nonstandard measurement enclosures" in its error model (Morice et al., 2021). The model, developed by Folland et al. (2001) based on work by Parker (1994), generates an ensemble of exposure bias error realisations based on assumptions of a fixed annual 1σ uncertainty of 0.2°C (0.1°C) prior to 1930 (1900), decreasing linearly to 0°C in 1950 (1930), for stations within (outside of) 20°S-

20°N (see Figure 2.4) (Morice et al., 2012). Current knowledge, however, suggests that this is an oversimplistic representation of the bias. The fixed annual uncertainty does not account for the well-documented seasonal nature of the exposure bias and could lead to inaccurate assessment of season-specific trends. In addition, the HadCRUT5 error model does not account for regional differences in a) the historic exposures in use prior to the introduction of the Stevenson screen, or b) the timing of the transition to the Stevenson screen. Both factors vary independently by region and affect the characteristics and magnitude of the exposure bias (a) and the period of time affected by the bias (b) (Parker, 1994; Sparks, 1972). To address these limitations, and improve the way exposure biases are accounted for in global temperature compilations, the following elements are required:

- an improved understanding of the characteristics of the exposure bias arising from different forms of historic exposure, including how they vary regionally and/or seasonally.
- the ability to model or otherwise quantify the monthly mean exposure bias globally.
- the ability to identify exposure bias affected series.

This chapter aims to address the first two points by a) updating Parker's (1994) assessment of the characteristics of the exposure bias using newly available parallel measurements (Section 2.3) and b) developing models to estimate the magnitude and seasonal nature of the exposure bias based on an assessment of the variables which influence the monthly mean bias in four categories of historic exposure (Section 2.4). The third point is addressed in Chapter 3.

2.2. Data

To study the characteristics of the exposure bias, series of parallel measurements – temperatures recorded near-simultaneously in two or more co-located exposures – were collated from the literature and meteorological yearbooks (Table 2.1; Figure 2.5). As this study is concerned specifically with the transition to Stevenson-type screens, only studies detailing temperatures (or differences) for Stevenson-type screens (T_{SS}) and at least one other historic exposure (T_{Hist}) were collated.

For each series and exposure, the mean monthly (or seasonal) maximum (T_x), minimum (T_n) and/or mean (T_m) temperature readings were recorded and, where necessary, converted to degrees Celsius. Where not given by the source, and where sufficient data were available, the diurnal temperature range (DTR) and monthly mean temperatures were then calculated.

Finally, the difference between the variables: T_x, T_n, T_m and *DTR*, recorded in the Stevenson screen and the historic exposure were calculated according to:

$$\Delta T = T_{SS} - T_{Hist} \quad (1)$$

where T is substituted for each of the previously listed variables. Note that for some series, only monthly differences (ΔT) were available, or ΔT plus the readings for one exposure. In the latter case, the values for the missing exposure were calculated from the two available values (ΔT and T_{Hist} or T_{SS}).

ΔT_x, ΔT_n and ΔDTR were recorded, in addition to ΔT_m (the variable of relevance to CRUTEM5), to provide a more comprehensive picture of the characteristics of the exposure bias as well as to allow an assessment of the elements which contribute to ΔT_m (as, by construction, ΔT_m is the mean of ΔT_x and ΔT_n and cannot have a larger bias than both ΔT_x and ΔT_n). For further theoretical discussion of the nature of biases in T_x, T_n, T_m and *DTR*, including how they relate to one another, the reader is directed to Thorne et al. (2016).

Each parallel measurement series was categorised to allow easier comparison between the main types of historic exposure. The categories used were based on those in Gaster (1882) and are defined as follows:

i. Open exposures: freestanding exposures, such as Glaisher (Figure 2.1b) and Montsouris stands, which, with the exception of protection above and to one side, expose the thermometer fully, or nearly fully, to the air.
ii. Wall-mounted exposures: any exposure where a thermometer is mounted on a wall (Figure 2.1c), fence or window, either screened or unscreened.
iii. Intermediate exposures: freestanding exposures such as thermometer sheds or summerhouses (Figure 2.1d), which, in addition to the protection offered by i), also provide some lateral protection to the thermometer.
iv. Closed exposures: freestanding exposures, such as the Wild hut and metallic shield (Figure 2.1e), which fully enclose the thermometer.

Of the 54 parallel measurements series obtained, fifteen were categorised as open (eight Glaisher, six Montsouris and one other), 30 as wall-mounted, four as intermediate (two summerhouses and two thatched shelters) and five as closed exposures (all Wild huts).

Table 2.1. Details of the parallel measurement series collated and analysed in Chapter 2.

#	Source	Location	Duration	Variables								Exposure Category	Inc. in model development?	Surrogate Data
				T_x	T_n	T_m	DTR	ΔT_X	ΔT_n	ΔT_m	ΔDTR			
1	AEMET M. Brunet, pers. comm.	43.4°N, 8.4°W; La Coruna, Spain	2003 - 2008	C	C	C	C	C	C	C	C	Open	Yes	N/A
2		38°N, 1.1°W; Murcia, Spain	2003 - 2008	C	C	C	C	C	C	C	C	Open	Yes	N/A
3	Butler, et al. (2005)	51.9°N, 10.3°W; Valencia, Ireland	1955 - 1959			C		Y	Y	C	C	Wall	Yes	Met Eireann
4	Chandler (1964)	51.47°N, 0.28°E; Kew, UK	1958 - 1960			C		Y	Y	Y	C	Wall	Yes	Monthly Weather Reports
5	Chenoweth (1992)	38.88°N, 91.93°W Fulton, Missouri, USA	08/1984 - 06/1987					Y	Y	C	C	Wall	Yes	No
6			09/1986 - 06/1987						Y			Wall	No-Insufficient data	N/A
7	Dines (1921)	51.9°N, 10.3°W; Valencia, Ireland	1909-1916							Y		Wall	No - $T_{mean} \neq {}^1/_2(T_x+T_n)$	N/A
8	Ellis (1891)	51.5°N, 0°E; Greenwich, UK	1887-1889					Y	Y	Y	C	Open	No - Duplicate of series 39	N/A
9	Field (1920)	27.2°N, 78°E; Agra, India	1917 - 1918					Y	Y	C	C	Intermediate	No – Not in mid-latitudes	N/A
10	Gaster (1882)	51.33°N, 1°E; Stratfield Turgis, UK	1868 - 1870	Y	Y	C	C	Y	Y	C	C	Open	Yes	N/A
11			1868 - 1870	Y	Y	C	C	Y	Y	C	C	Intermediate	Yes	N/A
12	Gill (1882)	33.9°S, 18.5°E; Cape of Good Hope, South Africa	1881	Y	Y	C	Y	C	C	C	C	Open	Yes	N/A
13			1881	Y	Y	C	Y	C	C	C	C	Wall	Included in initial analysis	N/A

#	Source	Location	Duration	Variables								Exposure Category	Inc. in model development?	Surrogate Data
				T_x	T_n	T_m	DTR	ΔT_X	ΔT_n	ΔT_m	ΔDTR			
14	Gorczynski (1910)	60°N, 30.1°E St Petersburg, Russia	1906			C		Y	Y	C	C	Closed	Yes	CRUTEM5
15	Grissolet (1935)	Unknown	1934-1935					Y	Y	C	C	Open	No-Insufficient data	N/A
16	Königlich Preussischen Meteorologischen Institut (1892)	Wyk au Fohr, Germany	08/1891 - 12/1891					Y	Y	C	C	Wall	No - Atypical exposure	N/A
17	Koppen (1913) (in Parker, 1994)	59.7°N, 30.4°E; Pavlovsk, Russia						Y	Y	C	C	Closed	Yes	No
18	Margary (1924)	51.5°N, 0.12°E; London, UK	1881 - 1920			C		Y	Y	C	C	Open	No - Duplicate of series 19	N/A
19			1881 - 1915	Y	Y	Y	C	Y	Y	Y	C	Open	Yes	N/A
20	Marriott (1879)	51.4°N, 0.08°W; Norwood, UK (Based on where Marriott was observer at the time)	04/1878 - 03/1879	Y	Y	Y	Y	Y	Y	Y	Y	Wall	Yes	N/A
21	Marriott (1894)	51°N, 4°W; Ilfracombe, UK	1893	Y	Y	Y	Y	Y	Y	Y	Y	Intermediate	Yes	N/A
22	Martinez Ibarra et al. (2010)	38.37°N, 0.5°W; Alicante, Spain	09/2008 - 04/2010			C		Y	Y	C	C	Open	Yes	CRUTEM5
23	Mawley (1897)	51.37°N, 0.08°E; Croydon, UK	1877 - 1881					Y	Y	Y	Y	Open	Yes[×]	No
24		51.75°N, 0.57°E Berkhamsted, UK	04/1896 - 12/1896					Y	Y	Y	Y	Open	Yes[×]	No
25			04/1896 - 12/1896					Y	Y	Y	Y	Wall	Yes	No

#	Source	Location	Duration	Variables								Exposure Category	Inc. in model development?	Surrogate Data
				T_x	T_n	T_m	DTR	ΔT_X	ΔT_n	ΔT_m	ΔDTR			
26	Meteorological Observations […] Adelaide Observatory (1890 - 1902)	35°S, 138.5°E; Adelaide, Australia	1887 - 1899	Y	Y	Y	C	C	C	C	C	Open	Yes	N/A
27			1887 - 1899	Y	Y	Y	C	C	C	C	C	Intermediate	Yes	N/A
28	Moden (1954) (in Nordli et al., (1997))	59.37°N, 13.38°E Henstad, Sweden	Unknown (7 years)							Y		Open	No - $T_{mean} \neq {}^1/_2(T_x+T_n)$	N/A
29	Muller (1984)	46.8°N, 7°E; Payerne, Switzerland	07/1979 - 06/1981			C		Y	Y	C	C	Closed	Yes	CRUTEM5
30	Nordli et al. (1997)	56.52°N, 8.12°E Bovbjerg Fyr, Denmark	1971 - 1987							Y		Wall	No - $T_{mean} \neq {}^1/_2(T_x+T_n)$	N/A
31		54.97°N, 10.2°E Skjoldnaes Fyr, Denmark	1971 - 1983							Y		Wall		
32		54.73°N, 10.72°E Keldsnor, Denmark	1971 - 1987							Y		Wall		
33		54.95°N, 10.72°E Rudkobing, Denmark	1971 - 1987							Y		Wall		
34		55.98°N, 11.85°E Spodsbjerg Fyr, Denmark	1971 - 1974							Y		Wall		
35		54.57°N, 11.97°E Gedser Fyr, Denmark	1971 - 1982							Y		Wall		
36	Omond (1906)	56.8°N, 5.1°W; Fort William, UK	08/1890 - 07/1904			C		Y	Y	C	C	Wall	Yes	CRUTEM5
37		51.9°N, 10.3°W; Valencia, Ireland	01/1871 - 12/1900					Y	Y	C	C	Wall	No – Series 3 used	N/A
38		57.2°N, 2.2°W; Aberdeen, UK	01/1871 - 12/1900			C		Y	Y	C	C	Wall	Yes	CRUTEM5

#	Source	Location	Duration	Variables								Exposure Category	Inc. in model development?	Surrogate Data
				T_x	T_n	T_m	DTR	ΔT_X	ΔT_n	ΔT_m	ΔDTR			
39	Results of [...] Observations [...], Greenwich (1889-1911)	51.5°N, 0°E; Greenwich, UK	1887 - 1909	Y	Y	C	C	Y	Y	C	C	Open	Yes	N/A
40	Sprung (1890)	52.4°N, 13.3°E Gross Lichterfelde, Germany	07/1886 - 03/1887	Y	Y	C	Y	C	C	C	C	Open	Yes[×]	N/A
41			07/1886 - 03/1887	Y	Y	C	Y	C	C	C	C	Wall	Yes	N/A
42			07/1886 - 03/1887	Y	Y	C	Y	C	C	C	C	Wall	Yes	N/A
43			07/1886 - 03/1887	Y	Y	C	Y	C	C	C	C	Wall	Yes	N/A
44			07/1886 - 03/1887	Y	Y	C	Y	C	C	C	C	Closed	Yes	N/A
45	Veðráttan (1962)	64.7°N, 21.4°W; Sioumuli, Iceland	1948 - 1962							Y		Wall	No – Not in mid-latitudes	N/A
46		65.1°N, 22.7°W Stykkisholmur, Iceland	1948 - 1962							Y		Wall		N/A
47		65.5°N, 22.2°W Reykholar, Iceland	1948 - 1962							Y		Wall		N/A
48		65.5°N, 24.1°W Lambavatn, Iceland	1948 - 1962							Y		Wall		N/A
49		65.6°N, 24°W Kvigindisdalur, Iceland	1948 - 1962							Y		Wall		N/A
50		65.3°N, 21.2°W Hlaohamar, Iceland	1948 - 1962							Y		Wall		N/A
51		63.4°N, 20.3°W Vestmannaeyjar, Iceland	1948 - 1962							Y		Wall		N/A

#	Source	Location	Duration	Variables								Exposure Category	Inc. in model development?	Surrogate Data
				T_x	T_n	T_m	DTR	ΔT_X	ΔT_n	ΔT_m	ΔDTR			
52		63.9°N, 21.2°W Eyrarbakki, Iceland	1948 - 1962							Y		Wall		N/A
53	Vigurs (1935)	50.1°N, 5.1°W; Falmouth, UK	1893-1912	Y	Y	C	C	Y	Y	C	C	Wall	Yes	N/A
54	Whipple (1883)	51.5°N, 0.3°E; Kew, UK	06/1879 - 12/1881	Y*	Y*	C	Y*	Y	Y	C	Y	Closed	Yes	N/A

In the 'Variables' columns, Y denotes variables available in the source, and C where they were calculated here. ˣ Denotes series which were used to inform the model development, but, due to missing or incomplete T_{Hist}, were not used as input in the final model.

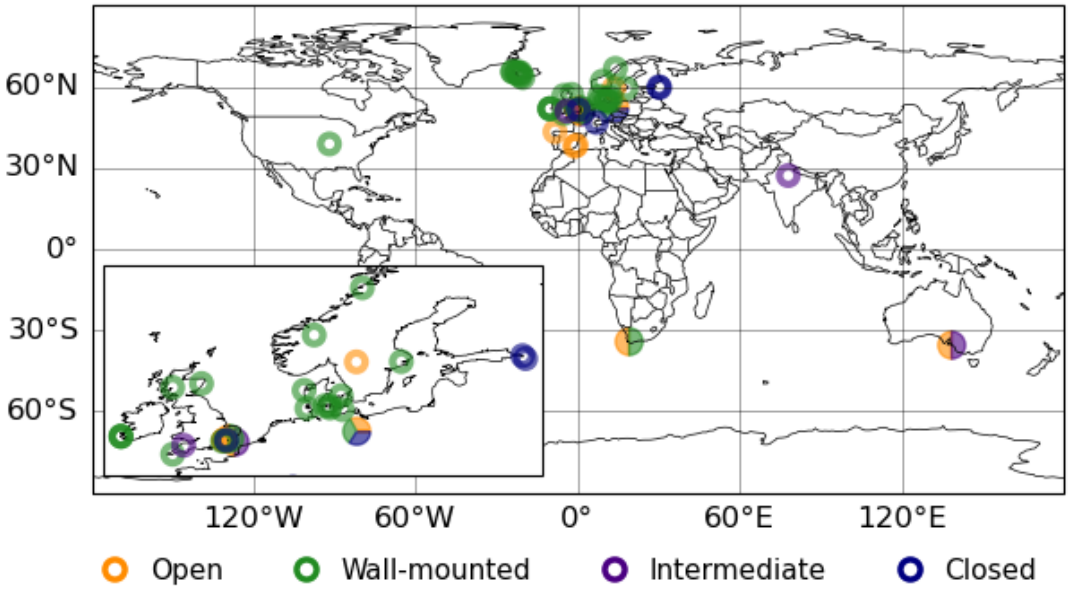


Figure 2.5. Locations of the parallel measurement series collated and analysed in Chapter 2.

2.3. Characteristics of the Exposure Bias

The characteristics of the exposure bias (the seasonal and diurnal structure) were identified for each exposure class by compositing the available ΔT values for each variable. The key findings are shown in Figures 2.6 to 2.9 and are discussed below.

2.3.1. Open exposures

Monthly mean ΔT values (Figure 2.6) reveal clear differences between the temperatures recorded in open exposures and Stevenson screens, across all four variables. The predominantly negative ΔT_x indicates maximum temperatures tend to be cooler in Stevenson screens than in open exposures, with mean annual ΔT_x ranging from -0.27°C in Gaster (1882) to -1.68°C in Martinez Ibarra et al. (2010). ΔT_x shows a clear seasonal cycle, increasing from an average warm bias of -0.08°C in winter to -1.04°C in summer. In contrast, the minimum temperatures recorded in open exposures are generally cooler than in Stevenson screens – on average by 0.36°C annually (range: 0.11°C–0.92°C) – with no obvious seasonal cycle. These differences can be explained by the larger quantity of reflected shortwave solar radiation and longwave terrestrial radiation which influence thermometers in open exposures during the day, causing a warm bias in T_x, and the greater radiative heat loss from the open exposures at night, causing a cool bias in T_n. The larger deviations in ΔT_x during the summer months are due to the increased strength of solar radiation.

The opposite signs of ΔT_x and ΔT_n mean the largest differences occur in the diurnal temperature range, with an exaggerated *DTR* in open exposures compared to Stevenson

screens. On average the *DTR* is 1.15°C larger annually in open exposures, but with mean annual (monthly) differences as large as -1.92°C (-2.61°C) in individual series. Despite the opposite signs of ΔT_x and ΔT_n, the bias in the mean does not cancel out, instead the larger magnitude of ΔT_x leads to warmer mean temperatures in open exposures compared to Stevenson screens, on average by 0.21°C annually, but with substantial variation in annual mean ΔT_m between the individual series (range: -0.78°C to 0.11°C). The strong seasonal cycle in ΔT_x is apparent in both ΔT_m and ΔDTR with average monthly differences in ΔT_m (ΔDTR) varying between 0.07°C (-0.4°C) in winter and -0.35°C (-1.4°C) in summer.

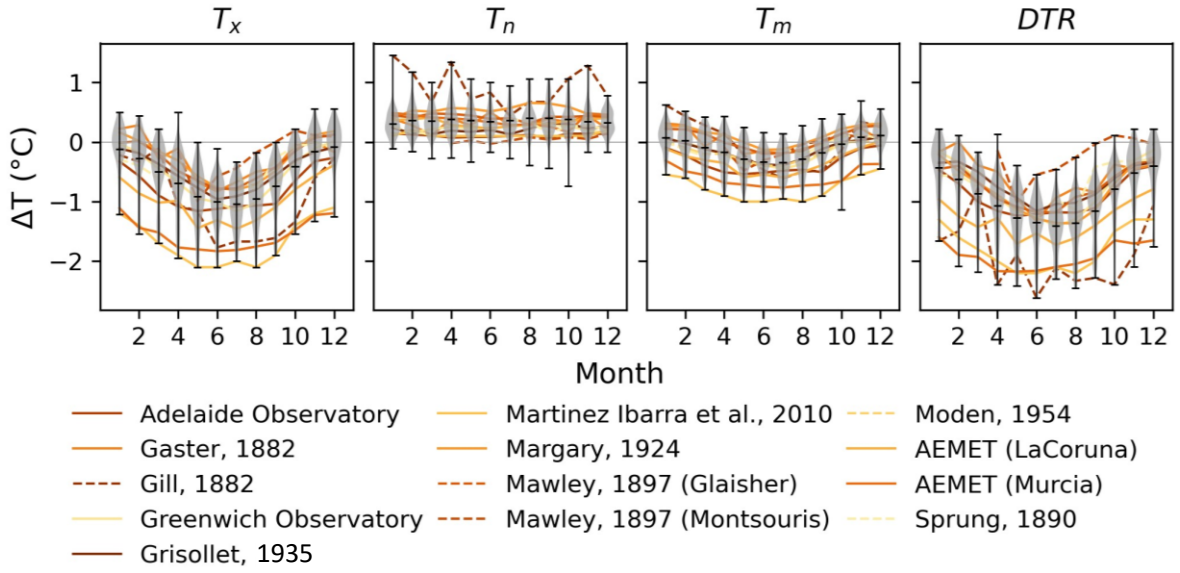


Figure 2.6. Differences between monthly mean temperatures recorded in Stevenson screens and open exposures. The violin plots (black and grey shading) show the mean and distribution of all individual monthly ΔT values; coloured lines represent monthly ΔT averaged over all years available for each individual series (lines are dashed where series have ≤12 months of data). Series located in the Southern Hemisphere have been shifted by 6 months to allow comparison with the Northern Hemisphere series.

These differences are broadly consistent with those outlined by Parker (1994) who concluded that Glaisher stands and Montsouris screens record annual mean temperatures 0.0°C-0.2°C warmer than Stevenson screens, with warmer and slightly cooler monthly mean summer and winter temperatures, respectively. Although the series analysed here and by Parker (1994) have similar characteristics in terms of the direction and seasonal cycle of ΔT, the inclusion of additional studies here highlights regional differences in the magnitude of the bias. Series from Spain and Australia, for example, show larger annual mean ΔT_m (range: -0.33°C to -0.78°C) than the UK series (range: 0.11°C to -0.16°C), suggesting location may have an influence on the magnitude of the bias. This finding is in agreement

with Nicholls et al. (1996) who found larger biases in Adelaide than in the higher latitude series analysed by Parker (1994).

2.3.2. Wall-mounted exposures

In contrast to open exposures, wall-mounted thermometers tend to record maximum temperatures which are 0.0°C–0.5°C cooler annually than those recorded in Stevenson screens (Figure 2.7). This is likely primarily due to the more shaded position of wall-mounted exposures, which are protected from solar radiation by the poleward facing wall, in comparison to Stevenson screens which are exposed in the open and receive solar radiation year-round (Omond, 1906). The thermal lag of the wall, and the greater height of some wall-mounted exposures, may also contribute to the difference (Chandler, 1964). There is evidence of a biannual seasonal cycle in ΔT_x, with the majority of series showing a 'double-peak' of larger values in early spring and autumn. This seasonal variation can be explained by the strength and angle of the incoming solar radiation (Omond, 1906). During spring and autumn, insolation is reasonably strong, but due to the angle of insolation mostly influences the Stevenson screen. This leads to larger values of ΔT_x in comparison to the rest of the year when solar radiation is able to influence both exposures (summer) or does not have a large influence on either (winter). Two series – Gill (1882) and Omond (1906; Aberdeen) – do not have the same biannual cycle in ΔT_x; this may be due to differences in their exposure which mean they are protected from solar radiation year-round.

Minimum temperatures in wall-mounted exposures tend to be warmer than in Stevenson screens, with no obvious seasonal cycle present in the differences. The warm bias is likely due to the thermal capacity of the walls which retain heat during the day, and release it as longwave radiation at night, keeping minimum temperatures warmer in wall-mounted exposures than Stevenson screens. The average difference across all series is -0.44°C annually, however there is significant variation between series, with the smallest mean annual differences close to 0°C and the largest up to -1.42°C. This variation is likely the result of site-specific differences between wall-mounted exposures, including building-type, thermometer orientation and height.

The cooler maximum and warmer minimum temperatures lead to a reduced *DTR* in wall-mounted exposures in comparison to Stevenson screens. On average, ΔDTR is 1.26°C annually, however there is significant variation between series with mean annual differences ranging from 0.68°C to 3.2°C, largely as a result of the variation in ΔT_n. The opposite sign

but similar magnitude of ΔT_x and ΔT_n result in little difference in annual mean temperatures between the wall-mounted exposures and the Stevenson screens (-0.02°C on average); however, larger differences are present in individual series (range: -0.58°C to 0.26°C) and on monthly timescales (up to ±1°C). The biannual seasonal cycle present in ΔT_x is apparent in the overall mean values for both ΔDTR and ΔT_m, albeit with reduced amplitude in ΔT_m and less well-defined in ΔDTR for some individual series.

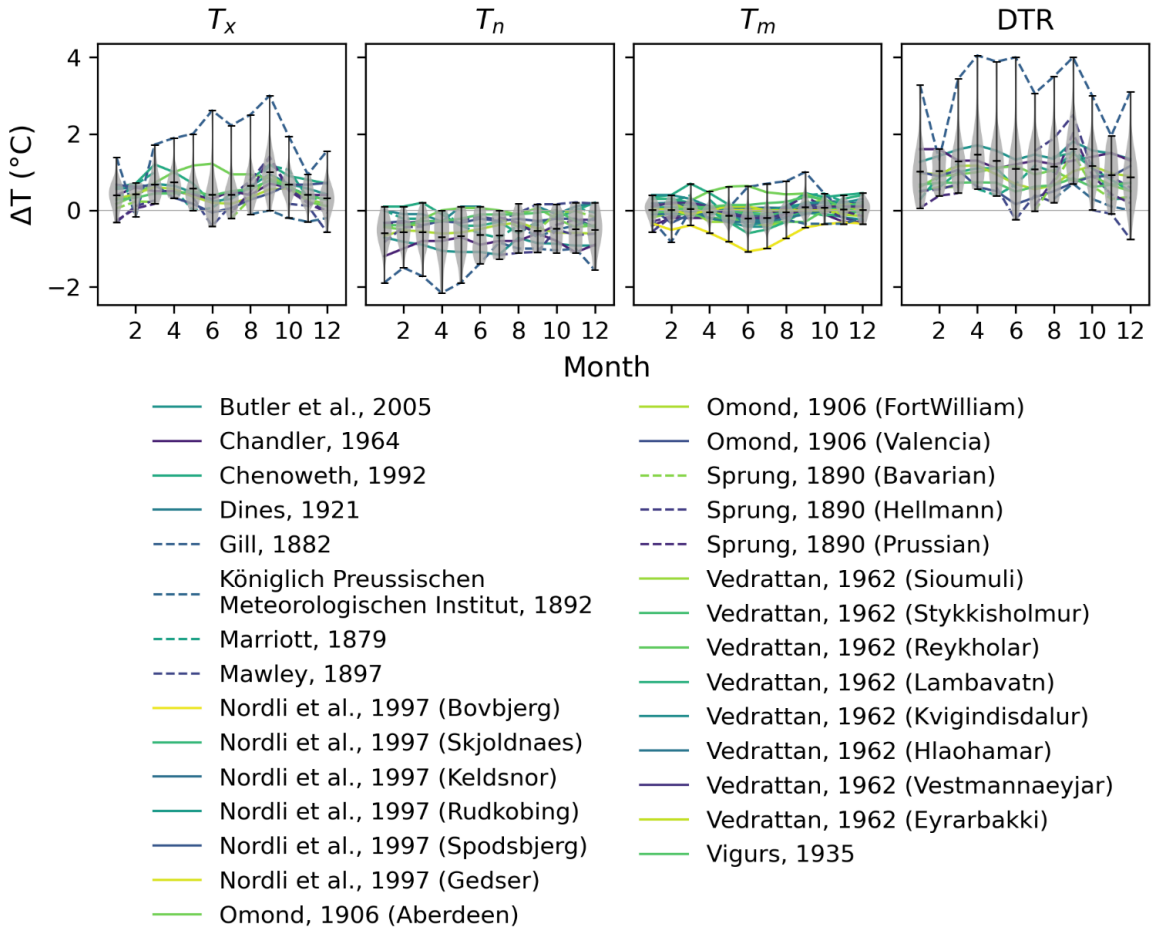


Figure 2.7. As Figure 2.6, but for differences between Stevenson screens and wall-mounted exposures.

These findings are consistent with Parker (1994) who concluded that mean temperatures in wall-mounted exposures do not consistently differ substantially from Stevenson screens, but that some larger differences are likely dependent on site-specific factors. The biannual cycle discussed above, and clearly evident in Figure 2.7, is only present in one of the series analysed by Parker (1994). The presence of a biannual cycle in the majority of the series analysed here, and the proposed physical explanation for it, suggests that it is a common feature of the bias in wall-mounted exposures, rather than an isolated occurrence.

2.3.3. Intermediate exposures

The comparison between intermediate exposures and Stevenson screens (Figure 2.8) also shows differences between the temperatures recorded in each. Maximum temperatures in intermediate exposures are generally warmer than in Stevenson screens – on average by 0.21°C annually, with little variation between series (range: -0.11°C to -0.38°C). The warmer maxima are likely the result of more reflected radiation from the surrounding unshaded ground reaching the thermometer in intermediate exposures, as they provide less lateral and basal protection than Stevenson screens. The stagnation of warm air in the eaves of some intermediate exposures, due to the roof structures impeding airflow, may also contribute to the warm bias. A seasonal cycle in ΔT_x is present in two series - Gaster (1882) and Adelaide Observatory - with larger differences in summer (up to -0.44°C), and smaller negative or positive differences in winter (up to 0.17°C). The same seasonal cycle is not obvious in Field (1920) or Marriott (1894), perhaps because they are based on only one year of data and are therefore noisier.

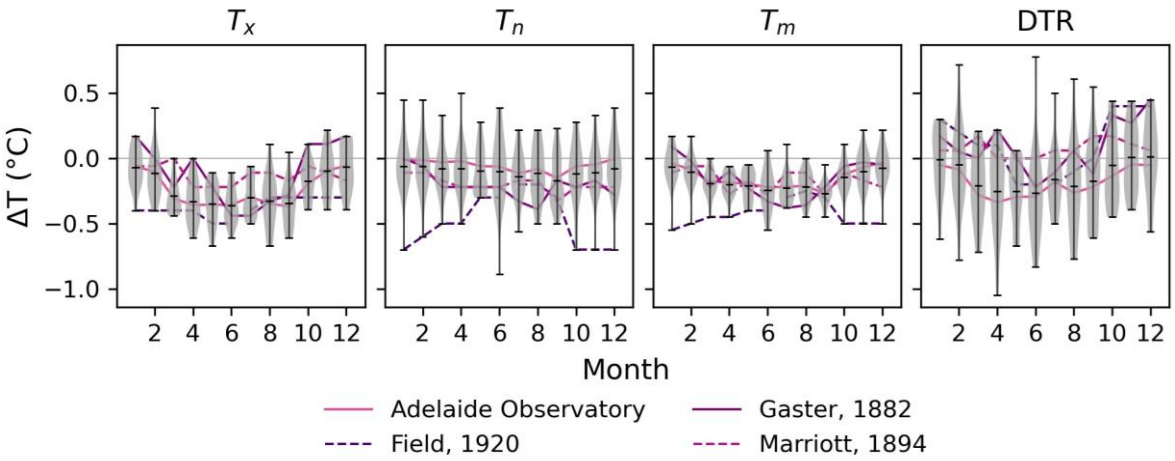


Figure 2.8. As Figure 2.6, but for differences between Stevenson screens and intermediate exposures.

Unlike open and wall-mounted exposures, the direction of the bias in intermediate exposures is the same for both T_x and T_n, with minimum temperatures also warmer in intermediate exposures than in Stevenson screens, on average by 0.23°C annually, but with significant variation within and between the series. The tendency for warmer minima is thought to be the result of the thermal properties of the intermediate exposures' (usually tiled or thatched) roof structures, which retain heat during the day and radiate it at night, preventing or slowing the cooling of the thermometer (Parker, 1994). The greater variation, and occasional opposite sign of ΔT_n may arise from varying weather conditions. In cloudy conditions, for example,

intermediate exposures will not absorb as much heat during the day, reducing the warming effect overnight. Without this effect, the thermometer is likely to cool more quickly overnight than thermometers in more enclosed Stevenson screens, leading to cooler minima in some conditions. The series do not show a consistent seasonal cycle in ΔT_n; three series - Adelaide Observatory, Gaster (1882) and Marriott (1894) – show a slightly larger bias in the summer and autumn, whereas Field (1920) shows a more pronounced, but reverse, seasonal cycle with larger differences in winter (-0.7°C) and smaller differences in summer (-0.2°C). The larger biases in summer and autumn may be the result of stronger solar radiation in those seasons leading to increased daytime heat retention, whereas the reverse seasonal cycle in the latter may be linked to increased cloud cover in summer over the Indian subcontinent (Sen Roy et al., 2015).

The warm biases in both T_x and T_n cause ΔDTR to be more muted in intermediate exposures than for open and wall-mounted exposures. In three of the series, the mean annual *DTR* is marginally smaller in the intermediate exposure than in the Stevenson screen (by 0.08°C–0.1°C), and in the fourth (Adelaide Observatory) it is slightly larger (by 0.19°C). Monthly deviations are larger (ΔDTR varies between -0.34°C and 0.45°C) and show a seasonal cycle with small differences in winter and larger negative differences in summer. Within series there is significant variation between individual monthly ΔDTR (-1.05°C to 0.78°C) due mostly to the variation in ΔT_n.

The warmer maximum and minimum temperatures in intermediate exposures also lead to warmer mean temperatures than in Stevenson screens. On average, annual ΔT_m is -0.22°C but varies between -0.43°C (Field, 1920) and -0.15°C (Adelaide Observatory). There is some evidence of a seasonal cycle in ΔT_m in Gaster (1882) and Adelaide Observatory, with near 0°C differences in winter and larger mean differences in summer (up to -0.38°C in Gaster (1882)). There is little evidence of a seasonal cycle in Marriott (1894) (likely due to noise due to a single year of data) and some evidence of the inverse seasonal cycle in Field (1920), with larger mean monthly differences of up to -0.55°C in winter, driven largely by the seasonal cycle in ΔT_n.

These results differ slightly with those in Parker (1994) who only analysed the results of Field (1920) and annual mean values from a study in Sri Lanka (Bamford, 1928): both tropical series. As a result, Parker (1994) concluded that annual mean differences between

the exposures were larger (0.4°C) than found here (0.22°C) and had a weak seasonal cycle in the opposite direction to two of the series analysed here.

2.3.4. Closed exposures

Maximum temperatures in closed exposures are consistently warmer than in Stevenson screens (Figure 2.9), though with large inter-site variation in magnitude. Annual mean ΔT_x varies between -0.16°C and -0.95°C, with an overall mean difference of -0.43°C. A seasonal cycle of smaller negative ΔT_x in winter and more negative ΔT_x in summer is present in the overall mean values, driven mostly by Muller (1984), with less pronounced seasonal variation in the remaining studies. As the closed exposures plotted here are all Wild huts (Figure 2.1e), it is likely the warmer maxima are due to the inner metal shield limiting air flow around the thermometer or becoming heated by indirect radiation. The former theory is supported by Wild (1887) who found ventilation reduces daytime overheating.

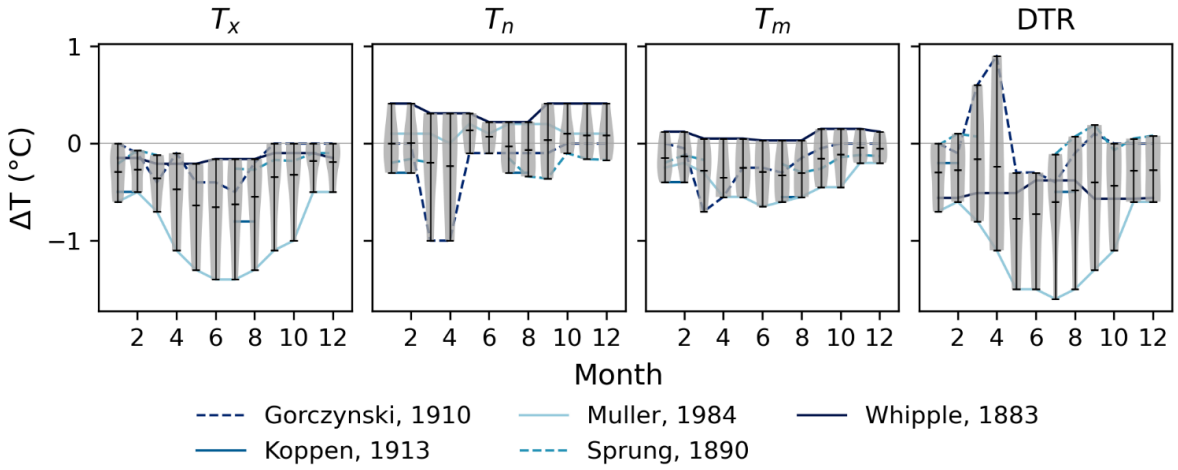


Figure 2.9. As Figure 2.6, but for differences between Stevenson screens and closed exposures.

There is no consensus between the series regarding the bias in minimum temperatures – three series observed warmer minimum temperatures in the closed exposure than the Stevenson screen, and two cooler. Warmer minimum temperatures may be explained by the larger thermal mass of the Wild hut (Figure 2.1e) cooling more slowly than the Stevenson screen, whereas cooler minima may occur if there is less radiative heating of the Wild hut in the day, followed by more rapid cooling overnight, via the open (poleward-facing) side of the hut, compared to the more enclosed Stevenson screen (Auchmann & Brönnimann, 2012). In all cases, however, annual and monthly differences are relatively small, not more than ±0.34°C and ±0.41°C, respectively (with the exception of the two larger monthly differences in Gorczynski (1910)). ΔT_n does not have a clear seasonal cycle in any of the series analysed,

with the possible exception of Whipple (1883) which shows slightly smaller differences in summer than in winter.

Δ*DTR* shows more consistency between series. Annually, the majority of closed exposures show a larger *DTR* than the Stevenson screen – on average 0.52°C larger. Two series have slightly smaller *DTR*s, however, both are based on ≤12 months of data and are potentially skewed by missing summer values (Sprung (1890)) and/or large potential outliers in ΔT_n (Gorczynski (1910)). ΔT_m also shows greater consistency between the series, with mean temperatures in closed exposures 0.2°C–0.5°C warmer than in the Stevenson screen in four of the exposures, and 0.09°C cooler in one: Whipple (1883). As with ΔT_x there does appear to be a seasonal cycle present in ΔT_m, with larger differences between the two exposures present in the summer months (up to -0.7°C) than in winter when differences are closer to zero. Again, the seasonal cycle is most pronounced in Muller (1984).

These findings are similar to those in Parker (1994) who found annual mean temperatures in Wild huts to be 0.1°C–0.2°C warmer than in Stevenson screens, with larger differences in summer than winter and an enhanced *DTR*. The magnitudes of ΔT_m found by Parker (1994) are at the lower end of the range found here (0.2°C–0.5°C), again highlighting variability in the magnitude of the exposure bias between series.

2.4. Modelling the Exposure Bias

The findings detailed in Section 2.3 reinforce the need to account for the exposure bias in a more realistic manner than the current blanket error estimates used in HadCRUT. In this section, the development of models to estimate the exposure bias in T_m, for each exposure class, is presented. This development is guided by the characteristics and physical reasoning about the causes of the bias that were explored in the previous section.

To develop models to estimate the exposure bias, an understanding of which variables influence the bias in each class of historic exposure is required. This understanding was developed by examining the relationship(s) between ΔT_m and three potential explanatory variables, using Pearson's correlation coefficient (r) and robust regression analysis. Robust regression was chosen to reduce the influence of possible outliers or atypical observations of ΔT_m on the model fit and the focus is on ΔT_m because it is the variable relevant to CRUTEM5 (the land component of HadCRUT5).

The three potential explanatory variables considered were: downward top of atmosphere solar radiation (*TOA*), shortwave downward solar radiation received at the Earth's surface (*SWD*) and the absolute temperature recorded in the historic exposure (T_{Hist}). The former two were chosen because the exposure bias, in its simplest form, stems from differences in the quantities of solar radiation which are able to influence the thermometer in each exposure; the latter was chosen due to the link between temperature and longwave radiation and because of suggestions in the literature of a relationship between temperature and the magnitude of the bias (e.g. Margary (1924); Ashcroft et al. (2022)). As the intention is to use any identified relationship(s) to estimate monthly mean exposure biases in CRUTEM5_sdb, the explanatory variables also have to be available, or calculable, for all series.

Monthly mean values of ΔT_m were sourced from the parallel measurement series detailed in Table 2.1. Not all series listed were included. Series outside 30° to 60° latitude were excluded because additional factors such as snow cover become more important at higher latitudes and because too few tropical series were available for analysis. Series where T_m was not calculated using $\frac{1}{2}(T_x + T_n)$ were also excluded because the exposure bias is sensitive to the method used to calculate daily-mean temperature (Böhm et al. (2010); Appendix A). Where more than 12 months of data were available for an individual series, the multi-year mean for each calendar month bias was used to avoid weighting any relationship toward the parallel measurements with the longest time series. Duplicate series were also excluded for similar reasons. Details of the excluded series, and reasoning, are given in Table 2.1.

For each included series, corresponding monthly mean values of *TOA* were calculated using the Python package 'climlab' and of *SWD* were sourced from a 30-year climatology (1981-2010) of WFDE5 (Cucchi et al., 2020). WFED5 is a version of ERA5 (Hersbach et al., 2020) which has been adjusted for biases using observations. The SWD data, used here, are bias-adjusted for cloud-radiation and aerosol-radiation interactions using cloud cover from CRU TS4.03 (Harris et al., 2020) and tropospheric and stratospheric aerosol optical depths from Bellouin et al. (2011) and Sato et al. (1993), respectively. Ideally *SWD* would have been sourced from observations (or derived from cloud cover) for each series location and month, however, neither were available for each parallel measurement series, nor more generally for stations within CRUTEM5_sdb. The use instead of a modern climatology was deemed sufficient to capture the large spatio-temporal variations in *SWD* which may explain differences in exposure biases between sites or seasons. Monthly mean values of T_{Hist} were

obtained or calculated from the series source; where this was not possible, 'surrogate' values of T_{Hist} (temperature readings for the same location, but obtained from an alternative source, i.e. an Observatory Yearbook) were obtained to maximise the number of series available for analysis (see Table 2.1 for details). Where 'surrogate' values were used, care was taken to ensure they did not significantly alter the results of the analyses.

Where a significant relationship between the bias and an explanatory variable was identified, simple statistical models were developed, using the variable as a predictor, and applied to the data to determine whether the relationship could be used to estimate ΔT_m. Model performance was assessed by comparing the observed and estimated monthly ΔT_m for each input series. Key performance indicators included the root-mean-square error (RMSE), skill score versus no adjustment (where 1.0 indicates perfect skill):

$$Skill = 1 - \left(\frac{\sum \left(\widehat{\Delta T_m} - \Delta T_m \right)^2 / n}{\sum (0 - \Delta T_m)^2 / n} \right) \quad (2)$$

and the closeness of the observed and estimated annual mean biases. Due to the limited number of series available as input, data were not held back from the initial analyses for validation purposes. However, data were held back during the assessment of the final statistical models - using a leave one out approach - to ensure the bias-estimations (and the relationship between ΔT_m and the selected predictor) were robust to the choice of input data.

2.4.1. Open exposures

The results of the robust linear regressions (Table 2.2) show the magnitude of ΔT_m in open exposures is significantly related to all three of the explanatory variables assessed. Each variable has a negative relationship with the magnitude of the bias, with increasing temperature and solar radiation both resulting in a larger warm bias in open exposures relative to Stevenson screens. This relationship is consistent with our understanding of the mechanisms which cause the exposure bias in open exposures (outlined in Section 2.3.1), as well as with previous studies which found larger biases in open exposures to correspond with stronger solar radiation (Brunet et al., 2011) and/or warmer absolute temperatures (Ashcroft et al., 2022; Brunet et al., 2011; Margary, 1924).

Of the three explanatory variables, T_{Hist} and *SWD* produced the best estimates of ΔT_m when used as predictors in simple linear regression models (not shown, see Table 2.2 for values);

however, neither was able to sufficiently capture both the timing of the annual cycle and the magnitude of the bias. T_{Hist} skilfully captured the magnitude of the bias but not the timing of the peak in the annual cycle, whereas the opposite was true for *SWD*. This makes physical sense as the annual cycle of ΔT_m is likely to be primarily controlled by the amplitude and strength of received solar radiation, whereas the magnitude of the bias is more likely to be dependent on a combination of local climatic factors which are better captured by T_{Hist}.

To exploit the strengths of each, the variables were combined to form a model which uses *SWD* to estimate the shape of the seasonal cycle of the bias, and T_{Hist} to estimate its magnitude and amplitude. This was achieved by normalising the *SWD* seasonal cycle at location i (using its minimum and maximum values) and then scaling it to fit the minimum ($\widehat{min\Delta T_m}$) and maximum ($\widehat{max\Delta T_m}$) bias estimated for i using T_{Hist}:

$$\widehat{\Delta T_m}(m,i) = \frac{\widehat{max\Delta T_m}(i)[S(m,i) - \min S(i)] - \widehat{min\Delta T_m}(i)[S(m,i) - \max S(i)]}{\max S(i) - \min S(i)} \tag{3}$$

where $\widehat{\Delta T_m}(m,i)$ is the estimate of the exposure bias for month m at location i, S is inverse *SWD*, and $\widehat{max\Delta T_m}$ and $\widehat{min\Delta T_m}$define the magnitude and amplitude of the bias and are estimated by regression on annual mean T_{Hist} (T_a):

$$\widehat{max\Delta T_m}(i) = -0.058 \times T_a(i) + 0.881 \tag{4}$$

$$\widehat{min\Delta T_m}(i) = -0.062 \times T_a(i) + 0.390 \tag{5}$$

Table 2.2 and Figure 2.10a show the details of these regressions.

Applying Equations 3 to 5 results in superior estimations of ΔT_m (based on the aforementioned performance indicators) in comparison to using the predictors in isolation. A comparison between the observations and estimates reveals the combination of predictors skilfully captures both the magnitude (Table 2.3 and Figure 2.10b) and seasonal cycle of the bias (Figure 2.10c). This is particularly evident in Figure 2.10c which shows the close agreement between the observations and estimates for the AEMET (La Coruna) series. The estimates for all of the open exposure parallel measurement series can be seen in Appendix A (Figure A2).

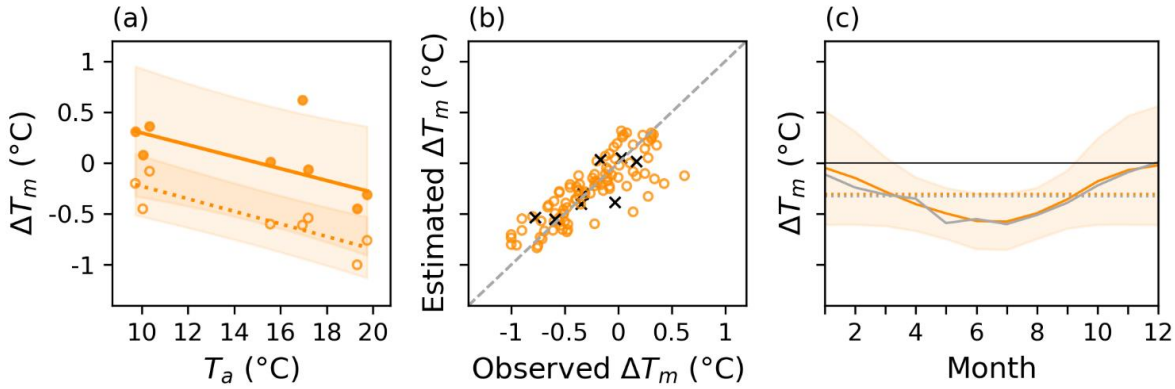


Figure 2.10. a) Relationship between annual maximum ΔT_m (solid line and filled markers), annual minimum ΔT_m (dotted line and unfilled markers) and annual mean T_{Hist} (T_a) with shaded 95% confidence intervals for open exposures; b) observed versus estimated monthly (orange circles) and annual mean (black crosses) ΔT_m; and c) observed (grey) and estimated (orange) ΔT_m with shaded 95% confidence interval for the AEMET La Coruna series (M. Brunet, pers. comm.). Observed (grey) and estimated (orange) annual mean biases are given by the dotted lines. Each panel has the same y-axis, so it is only labelled in panel a).

The deviations between the observed and estimated monthly and annual mean biases (Figure 2.10b) are generally small. Except for the series from Gill (1882), which behaves differently, the RMSE remains below 0.25°C in all series assessed, and the differences between the annual mean biases are generally below 0.2°C. Where the deviations from the observed annual and monthly mean biases are largest, the modelled value tends to be an underestimate, but in all cases the observed values are captured within the 95% confidence interval. The skill score is also positive in all assessed series (again, except Gill (1882)), and more than 0.9 in 50% of them. These statistics suggest the model can provide skilful estimations of the exposure bias in open exposures, and, if applied to observations, would reduce the bias associated with the transition from open exposures to Stevenson-type screens.

Table 2.2. Results of the robust linear regression analyses for each class of exposure and explanatory variable. Units: Int (Intercept) is °C, Slope for explanatory variables T_{Hist} and T_a is °C/°C, and Slope for explanatory variables *TOA* and *SWD* is °C/(W/m^2) except first term for the Quadratic which is °C/(W/m^2)^2.

Exposure Category		T_{Hist}				TOA				SWD			
		Int	**Slope**	**r**	***p*<0.05**	**Int**	**Slope**	**r**	***p*<0.05**	**Int**	**Slope**	**r**	***p*<0.05**
Open		0.305	-0.039	-0.71	✓	0.278	-0.002	-0.6	✓	0.244	-0.030	-0.68	✓
Wall-mounted*	**Linear**	-0.097	0.010	0.25	✓	-0.060	0	0.20	×	-0.073	0.001	0.26	✓
	Quadratic	-0.121	-6×10^{-4} 0.020		×	-0.412	-6.66×10^{-6} 0.004		✓	-0.101	-2.56×10^{-6} 0.001		×
Intermediate		-0.029	-0.010	-0.52	✓	-0.002	-0.0005	-0.66	✓	-0.041	-0.0007	-0.61	✓
Closed		-0.138	-0.012	-0.33	✓	-0.037	-0.001	-0.49	✓	-0.043	-0.002	-0.55	✓
		T_a											
Open	**Maximum Bias ($max\Delta T_m$)**	0.881	-0.058	-0.57	✓								
	Minimum Bias ($min\Delta T_m$)	0.390	-0.062	-0.88	✓								

*Results including the series from Gill (1882) as input. Final regression coefficients for the model, with Gill (1882) excluded, are given in Equation 6.

Table 2.3. Lower and upper quartiles of the key performance indicators for the best-performing statistical model in each exposure class.

	RMSE (°C)	Skill	Observed minus estimated annual ΔT_m (°C)
Open	0.09 - 0.21	0.39 - 0.96	-0.08 - 0.08
Wall-mounted	0.14 - 0.24	0.09 - 0.44	-0.12 - 0.11
Intermediate	0.07 - 0.09	0.78 - 0.88	-0.02 - 0.01
Closed	0.10 – 0.24	0.47 – 0.75	-0.08 – 0.07

2.4.2. Wall-mounted exposures

Our physically based reasoning (Section 2.3.2) suggests the bias in wall-mounted exposures may be related to solar radiation, however, the linear regression analyses (Table 2.2) show only weak ($r \leq 0.26$) positive relationships (with T_{Hist} and *SWD*) or no significant relationship (with *TOA*). Scatter plots, plus the indication of a biannual cycle in ΔT_m (Section 2.3.2), suggest a quadratic relationship may be more appropriate than linear regression.

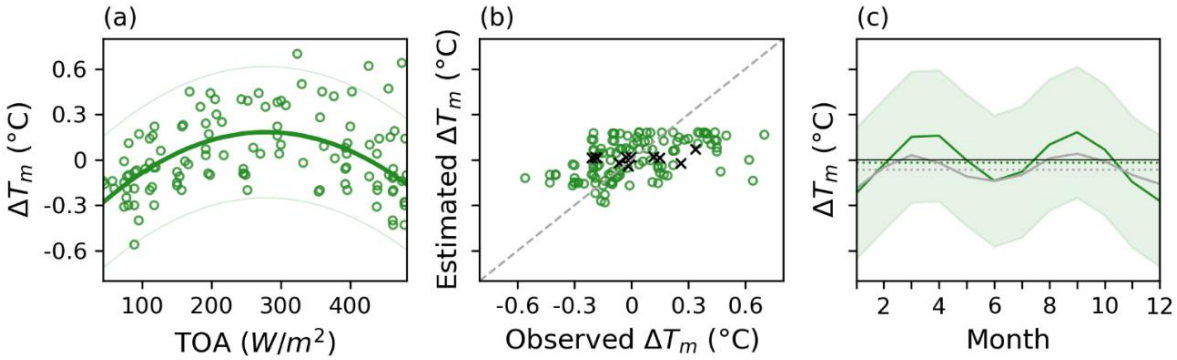


Figure 2.11. a) Relationship between monthly ΔT_m and top of atmosphere solar radiation with 95% confidence interval for wall-mounted exposures; b) observed versus estimated monthly (green circles) and annual mean (black crosses) ΔT_m; and c) observed (grey) and estimated (green) ΔT_m with shaded 95% confidence interval for the Fort William series from Omond (1906). Observed (grey) and estimated (green) annual mean biases are given by the dashed lines. Each panel has the same y-axis, so it is only labelled in panel a).

Quadratic regression (Table 2.2; Figure 2.11a) shows a significant quadratic relationship between ΔT_m and *TOA*, but not with T_{Hist} or *SWD* which appear to be more affected by outliers. The relationship suggests increasing *TOA* leads to increasing values of ΔT_m until a threshold, after which the relationship becomes negative, leading to biases of a similar magnitude for both high and low levels of *TOA*. This relationship is consistent with Section 2.3.2 which found the largest values of ΔT_m in spring and autumn (when *TOA* is mid-strength

but only influences the Stevenson screen), and smaller values in summer and winter (when *TOA* is strong but affects both screens, or is weak and has little effect on either). Analysis of the individual parallel measurement series confirms similar relationships between ΔT_m and *TOA*, with the same timing of the peak, are present in all-but-one of the input series, giving confidence that the relationship is robust despite the wide scatter between sites evident in Figure 2.11a. The series which did not have a significant relationship (p=0.79), and which behaved markedly differently (Gill (1882)), was excluded from the final calculation of the regression coefficients due to concerns about its validity. The final model is:

$$\widehat{\Delta T_m}(m,i) = -8.38 \times 10^{-6} \times TOA(m,i)^2 + 0.005 \times TOA(m,i) - 0.469 \quad (6)$$

Applying Equation 6 to the input series, and comparing the results with the observations, reveals *TOA* is able to skilfully estimate the shape of the annual cycle of the exposure bias in wall-mounted exposures, but that the magnitude of the estimates and their annual means can deviate from the observations. This is illustrated in Figure 2.11c which shows the correct timing of the 'double-peak' of the bias, but an overestimated amplitude, and Figure 2.11b which shows greater variance in the observations than the estimates. This is not unexpected. As noted in section 2.3.2, there is significant variation in the magnitude of the bias between wall-mounted series, due to the wide variety of exposures which fall into the category and the number of factors which influence the bias in addition to solar radiation. As a result, estimates based on one variable will capture only a small part of the variance observed in the biases arising from wall-mounted exposures. Despite this, the deviations between the observed and estimated values tend to be relatively small – the majority (all) of the series analysed had an RMSE below 0.2°C (0.37°C) – and the greater variation in the input data is captured by the size of the confidence intervals (Figure 2.11a; Figure A3). In addition, although lower than for the open model, the skill scores are positive in all-but-one series analysed (Table 2.3; Figure A3), meaning the model still provides estimates that, when applied, can reduce the size of the exposure bias and are thus better than ignoring the bias in most cases.

2.4.3. Intermediate exposures

As with open exposures, ΔT_m in intermediate exposures displays strong, negative correlations (r >-0.5; p <0.05) with all three variables (Table 2.2), with increases in T_{Hist}, *TOA* and *SWD* all leading to larger warm biases (more negative ΔT_m). This is consistent with

the findings in Section 2.3.3 of larger biases in the summer months in the extratropical series, as well as with our understanding of the processes which contribute to the bias. Intermediate exposures record warmer mean temperatures relative to the Stevenson screen due to the influence of reflected shortwave radiation during the day (controlled by *SWD*/*TOA*) and the emittance of longwave radiation from the roof structure at night (which is influenced by T_{Hist} and linked to daytime heat retention which varies according to *SWD* and *TOA*). The weaker correlation between T_{Hist} and ΔT_m (r=-0.52), in comparison to ΔT_m and *TOA* (r=-0.66) and *SWD* (r=-0.61), is also consistent with this understanding and suggests the relationship with T_{Hist} may partly be an artefact of the correlation between T_{Hist} and solar radiation.

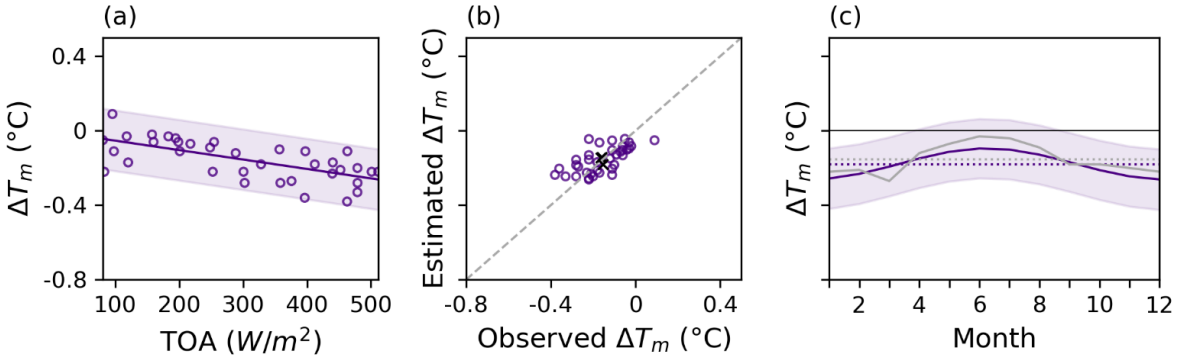


Figure 2.12. a) Relationship between monthly ΔT_m and top of atmosphere solar radiation with shaded 95% confidence interval for intermediate exposures; b) observed versus estimated monthly (purple circles) and annual mean (black crosses) ΔT_m; and c) observed (grey) and estimated (purple) ΔT_m with shaded 95% confidence interval for the Adelaide Observatory series (in contrast to Figure 2.6 and Figure 2.8, the Adelaide data are not shifted by 6 months here). Observed (grey) and estimated (purple) annual mean biases are given by the dashed lines. Each panel has the same y-axis, so it is only labelled in panel a).

When used as predictors in simple linear regression models, both *TOA* and *SWD* produced similar results, and more skilful estimations than T_{Hist}. Both solar radiation predictors were able to accurately reproduce the seasonal cycle and the magnitude of the biases (Figure 2.12b,c), with low RMSE scores for *TOA* (*SWD*) between 0.07°C-0.10°C (0.09°C–0.10°C) and reasonably high, positive, skill scores between 0.71-0.84 (0.63–0.83). Using each model, the sign of the annual bias is also always correctly estimated, and, in each case, the estimated biases are within 0.16°C of the observed annual biases.

Overall, using *TOA* as predictor gives marginally better estimates than *SWD* (Table 2.3 and Figure 2.12). See Appendix A (Figure A4) for the individual series estimates produced using *TOA* as predictor. However, further validation of both models, holding data back, revealed

the reproduction of the seasonal cycle was dependent on the series from Gaster (1882) being included as input (Figure A5). As a result, the broad application of either model is not advised due to the limited input data used to fit the model and the overreliance on one series to replicate the seasonal cycle.

2.4.4. Closed exposures

In keeping with the other freestanding exposures analysed, monthly ΔT_m in closed exposures is significantly negatively correlated with all three explanatory variables (Table 2.2), with increasing T_{Hist}, *TOA* and *SWD* corresponding with larger warm biases. As with intermediate exposures, the strongest correlations are with *SWD* (r=-0.55) and *TOA* (r=-0.49), and the correlation with T_{Hist} (r=-0.33) slightly weaker. These correlations are consistent with our understanding of the cause(s) of the warm bias in closed exposures, outlined in Section 2.3.4, as well as with our findings that the bias is largest in the summer.

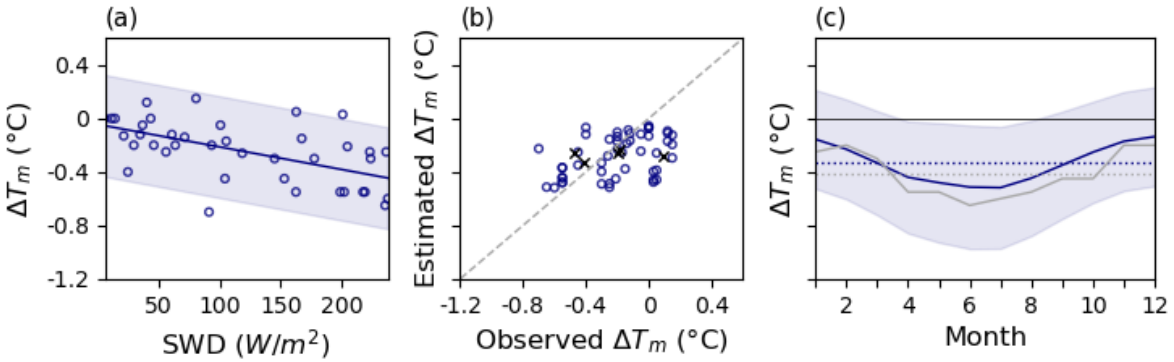


Figure 2.13. a) Relationship between monthly ΔT_m and shortwave downward received solar radiation with shaded 95% confidence interval for closed exposures; b) observed versus estimated monthly (navy circles) and annual mean (black crosses) ΔT_m; and c) observed (grey) and estimated (navy) ΔT_m with shaded 95% confidence interval for the series from Muller (1984). Observed (grey) and estimated (navy) annual mean biases are given by the dashed lines. Each panel has the same y-axis, so it is only labelled in panel a).

Overall, the correlation coefficients are slightly lower than for the other freestanding exposures analysed here. This makes physical sense as these exposures are more enclosed. As a result, the influence of solar radiation on ΔT_m is comparatively smaller, and other variables, such as wind speed, become more important. This is supported by previous studies (outlined in Parker (1994)) which found the largest biases in closed exposures during clear, calm, weather when received solar radiation is greatest and ventilation is reduced.

Despite the slightly weaker correlations, skilful estimations of the monthly exposure bias were still obtained using one predictor. Of the three variables assessed, *SWD* produced the more skilful estimations, with reasonable agreement between the observed and estimated values of ΔT_m (Table 2.3 and Figure 2.13b,c) obtained using:

$$\widehat{\Delta T_m}(m,i) = -0.002 \times SWD(m,i) - 0.043 \quad (7)$$

As can be seen in Figure 2.13c, Equation 7 is able to skilfully reproduce the seasonal cycle of the bias, particularly the timing of the peak, as well as capturing the magnitude of the bias to within a few tenths of a degree. With the exception of the series from Whipple (1883), which shows a cool bias in contrast to the other input series, the model has relatively low RMSE values, below 0.24°C for each series, and is able to estimate the annual mean bias to within ±0.2°C of the observed values (Table 2.3; Figure A6). There is slightly larger variation between observed and estimated ΔT_m at a monthly resolution (Figure 2.13b), likely for the reasons noted above, however, the greater variance in the observations is accounted for by the confidence intervals (Figure 2.13a,c). The primarily high, positive, skill scores (Table 2.3) suggest applying the model would be beneficial for reducing the presence of the exposure bias from closed exposures in most cases.

2.5. Chapter summary

Exposure biases are a pervasive non-climate change introduced into land surface air temperature records as a result of changes in the way thermometers have been sheltered from solar radiation and the elements over time. Exposure biases have not been widely accounted for in observational records due to difficulties both in identifying the bias and determining the appropriate adjustment to apply; therefore, exposure biases are still believed to contribute significant uncertainty to the early part of global temperature compilations. Current representations of this uncertainty are considered oversimplified as they do not account for the seasonal nature of the bias, regional differences in the types of historic exposure in use prior to the introduction of the Stevenson screen, or regional differences in the timing of the transition to the Stevenson screen. This chapter has addressed two aspects of these limitations by a) updating Parker's (1994) assessment of the characteristics of the exposure bias and by b) developing three regression-based models to estimate the monthly mean bias (in T_m) arising from three categories of historic thermometer exposure.

An empirical analysis of 54 parallel measurement series, comparing temperature readings recorded in the Stevenson screen and one of four categories of historic exposure, found seasonally varying biases which differ according to the class of exposure and potentially according to location (particularly in open exposures). The largest biases were generally present in the maximum temperatures and the diurnal temperature range, but significant biases, which require consideration before series can be used for climate assessment, were also found in mean temperatures. The largest biases in the mean temperature were found in freestanding exposures (up to -0.78°C annually) and in the summer months, while the smallest biases were generally found in wall-mounted exposures (near 0°C annually) and in the winter, though the wall-mounted exposures showed the greatest site-specific variability.

An analysis of the relationship between the magnitude of the bias in T_m and three variables which were considered *a priori* to influence the bias – received shortwave downward solar radiation, downward top of atmosphere solar radiation and absolute temperature – led to the identification of statistically significant relationships and the development of four bias-estimation models: one for each category of exposure analysed. Three of the models – the open, wall-mounted and closed models – were found to skilfully reproduce the magnitude and direction of the monthly exposure bias and are considered robust enough for wider application. The application of these three models to series in CRUTEM5_sdb, to quantify the exposure bias and to produce an exposure bias adjusted version of the data (CRUTEM5_eba), will be presented in the following chapter.

Chapter 3

Developing an exposure bias adjusted version of CRUTEM5

3.1. Introduction

Chapter 2 confirms the introduction of Stevenson screens in the late-nineteenth and early-twentieth centuries would have introduced biases into long temperature series and thus confirms the need to account for exposure biases in global temperature compilations. In this chapter, the application of three of the bias-estimation models developed in Chapter 2 to the CRUTEM5 database of station temperature records (CRUTEM5_sdb), to quantify the exposure bias and produce an extended and exposure-bias-adjusted version of CRUTEM5, will be presented.

In order to apply the bias-estimation models, the final point highlighted in Section 2: 'the ability to identify exposure bias affected series' needs to be addressed. This information is required to determine which series to apply each of the models to, and during which time periods. Therefore, the first half of this chapter (Sections 3.2 and 3.3) will outline and assess the options for identifying the exposure bias in stations within CRUTEM5_sdb. Section 3.4 will then describe the final approach used to identify the bias: the development of a new metadata database detailing the pre-Stevenson screen thermometer exposures in use at selected stations and countries within CRUTEM5_sdb. In Section 3.5 the application of the models to CRUTEM5_sdb is then outlined and the resulting exposure-bias-adjusted dataset is discussed. Note, as the models were developed using parallel measurement series within 30° to 60° degrees latitude (North and South of the equator), and as the literature suggests the majority of weather stations would have had Stevenson screens by the 1960s (Parker, 1994), only series located within 30° to 60° latitude and with data prior to 1961 will be considered in this chapter.

3.2. Identifying exposure biases in CRUTEM5_sdb

The most commonly employed approaches to detect or determine the location of non-climatic influences (inhomogeneities, biases or breakpoints) in temperature records fall into two broad categories: a) direct approaches, which use metadata to identify documented changes and b) indirect approaches, which use statistical tests and/or comparisons with reference series to detect undocumented changes (Aguilar et al., 2003; Peterson et al., 1998; Ribeiro et al., 2016). Each approach is outlined below and its suitability to identify exposure bias-affected stations in CRUTEM5_sdb is briefly evaluated.

3.2.1. Direct approach: metadata

Metadata are 'data about data' and, in the context of temperature records, include any information pertaining to how temperatures have been observed or processed. This could include: the name of the observer, the location and surroundings of a weather station, the instruments used and their exposure, the time-of-day observations were made, the methods used to calculate the daily mean temperature or information about any adjustments applied to the observations following measurement (Aguilar et al., 2003; Brunet et al., 2020; World Meteorological Organization, 2020). When accurate, metadata are valuable sources of information for identifying inhomogeneities as they provide an insight into how observing practices, and the conditions under which observations were made, have changed over time (Brunet et al., 2020; World Meteorological Organization, 2020).

The use of metadata for breakpoint identification (either directly or in combination with indirect methods) is often considered the optimal approach to homogenise climate records because: metadata generally provide more precise and accurate information than statistical tests or series comparisons allow, metadata can identify breakpoints which are less detectable using indirect methods (such as the exposure bias; Chapter 2) and metadata provide information about the cause of any identified breaks (Domonkos, 2022; O'Neill et al., 2022; World Meteorological Organization, 2020). For these reasons, metadata have been widely employed to identify inhomogeneities, including exposure biases, in long temperature series. Brunet et al. (2006, 2011) and Böhm et al. (2010), for example, used metadata to identify exposure biases at selected weather stations in Spain and in the Greater Alpine Region, respectively. To date, metadata have not been used (in isolation) to identify inhomogeneities in LSAT records on a global scale in most global temperature datasets (though Berkeley Earth use incomplete metadata alongside statistical methods to identify

undocumented changes (Rohde et al., 2016)); however, the approach has long been used to address inhomogeneities in global sea surface temperature (SST) datasets, such as HadSST (e.g. Folland & Parker (1995); Kennedy et al. (2011, 2019); Rayner et al. (2006)).

The lack of implementation of direct breakpoint detection to LSAT records on a global scale may be because there are a number of challenges to using metadata, particularly on a global scale. Despite the increasing recognition of the value of metadata - guidelines for climate observation and data rescue now include the importance of recording and/or preserving metadata (World Meteorological Organization, 2016, 2019, 2021) - metadata are often not available or easily accessible to climate scientists. This can be especially true for early instrumental records where metadata may not have been recorded at the time of observation, may have been lost, misplaced or damaged over time, or may only be available in paper format buried within meteorological or national archives (Trewin et al., 2020; Vincent, 1998; World Meteorological Organization, 2020). Throughout station histories, borders may also have changed, meaning the metadata for individual weather stations (such as in Poland (Pospieszyńska & Przybylak, 2019)) may be dispersed between multiple national libraries or archives. Additionally, even where station metadata records exist, there is no guarantee they are complete or accurate. As a result, gathering metadata for breakpoint detection (at both a local and global scale) is a hugely time-consuming process, with large amounts of irrelevant information having to be sifted through to find the required information (with no guarantee that the information exists) (Vincent, 1998; World Meteorological Organization, 2020).

To apply the exposure bias-estimation models to all stations in CRUTEM5_sdb within 30° to 60° latitude, and with observations before 1961, exposure metadata for more than 5,000 weather stations would be required. This information has not previously been collated for the majority of stations within CRUTEM5_sdb. Therefore, while the use of metadata to identify breakpoints globally is possible (as shown by Kennedy et al. (2019) and others for SSTs), and provides many advantages to using indirect methods, alternative approaches to identify exposure biases in CRUTEM5_sdb will first be explored.

3.2.2. Indirect approaches: statistical tests and series comparisons

Indirect approaches to breakpoint detection fall into two categories: absolute and relative (Aguilar et al., 2003; Peterson et al., 1998; Ribeiro et al., 2016). Absolute breakpoint detection refers to methods which use only the series being assessed (the candidate series)

to detect a breakpoint (e.g., by applying a statistical test) whereas relative breakpoint detection approaches compare the candidate series to at least one other, called a reference series.

The use of absolute approaches to detect breakpoints is no longer recommended where use of a direct or relative approach is possible (Auer et al., 2001; Guijarro, 2011; Ribeiro et al., 2016). This is because indirect breakpoint detection is largely a signal-to-noise challenge and individual climate series contain substantial noise arising from both climate variability and from non-climatic influences. As a result, absolute methods are often unable to distinguish between breaks which have occurred for climatic and non-climatic reasons, nor can they detect small non-climatic breaks which are masked by series variability (Auer et al., 2001; Caussinus & Mestre, 2004; Peterson et al., 1998; Ribeiro et al., 2016; Wijngaard et al., 2003). For a temperature series with a standard deviation of 1°C, for example, the World Meteorological Organization (2020; p.15) states "*it will be difficult or impossible for any statistical method to detect a 0.4°C inhomogeneity*". The standard deviation of individual series in CRUTEM5_sdb is generally more than 1°C (for individual monthly and annual series), and the magnitudes of the monthly and annual mean exposure biases identified in Chapter 2 (Section 2.3) are frequently less than 0.4°C; therefore, absolute breakpoint detection is not considered appropriate for use here and will not be considered further.

Relative approaches to breakpoint detection address the signal-to-noise challenge, outlined above, by comparing the candidate series (the series requiring homogenisation) to a reference series which contains the same climate signal (Conrad & Pollak, 1950). This approach is predicated on the principle that a comparison between two series with the same climate signal will remove the shared climate variability and any trends, but will retain the effect of any inhomogeneities (unless they are common to both series), thereby increasing the signal-to-noise ratio and allowing the detection of smaller breakpoints with increased accuracy (Conrad & Pollak, 1950; Ducré-Robitaille et al., 2003; Ribeiro et al., 2016; World Meteorological Organization, 2020). Relative approaches vary from simple visual comparisons between pairs of neighbouring stations, as utilized by Jones et al. (1986), to the application of increasingly sophisticated statistical techniques and homogenisation software (see Domonkos et al. (2021) and Venema et al. (2012) for an overview). Indirect, relative approaches, now form the basis of most breakpoint detection (Cao & Yan, 2012; Lindau & Venema, 2013; Venema et al., 2012) and have been widely used to detect and adjust for

breakpoints in regional and global temperature compilations (e.g., Joelsson et al. (2023); Menne et al. (2009, 2018); Trewin et al. (2020) and Vincent et al. (2020)).

Relative breakpoint detection is most effective when the reference series is highly correlated with the candidate (ideally $r \geq 0.8$, according to Peterson & Easterling (1994)) and homogenous throughout the period requiring assessment (Domonkos, 2011; Ducré-Robitaille et al., 2003; Reeves et al., 2007). The latter requirement is rarely met (Lindau & Venema, 2013); therefore, a dense network of reference series is also required, in order to distinguish the breaks which occur only in the candidate series (breaks in the candidate should appear in all comparisons, breaks in one reference should appear in only one comparison) (Gubler et al., 2017; Menne & Williams, 2009; Peterson & Easterling, 1994). As a result of these requirements, temperature data from surrounding weather stations are most commonly used as reference series for relative homogenisation. However, the use of neighbouring data is not effective for identifying spatially pervasive biases, such as the exposure bias. This is because Stevenson screens were introduced quasi-simultaneously within observing networks (Section 2.1), meaning the majority of stations within a region will contain co-occurring exposure biases which will not show up in any comparisons.

In such instances, independent references series, which meet the requirements outlined above, but which do not have co-occurring breakpoints, are required. Where a candidate series borders another country or NMS, it might be possible to obtain an independent reference series from weather stations across the border, which are less likely to have co-occurring biases. This approach highlighted the timing of the transition to the Stevenson screen in northern Italy (Böhm et al., 2001; Brunetti et al., 2006). Alternatively, a closely related variable, which is not affected by the same biases, could be used. This approach was used by Moberg et al. (2003) who used a number of temperature-correlated variables (including cloud cover and air pressure) to identify possible exposure biases in summer temperatures in Stockholm and Uppsala, as well as by Cowtan, Rohde and Hausfather (2018) who assessed SSTs for biases using LSAT from coastal weather stations. Each method provides a potential solution for a proportion of the series in CRUTEM5_sdb, however neither provides a solution for the majority. Cross-border station data cannot be used to identify breakpoints at isolated weather stations or those which do not closely border another NMS, and, as air temperature is more spatially and temporally complete than other climate variables (Chen et al., 2023), additional high quality, temperature-correlated, variables are

unlikely to be available for many stations within CRUTEM5_sdb, particularly in the early period when exposure biases occurred.

One additional solution, which has been proposed by previous studies, and which may be applicable here, is to use an alternative source of temperature data, such as reanalysis or proxy reconstructions, to create independent reference series (Guijarro, 2011; Trewin, 2010; World Meteorological Organization, 2020).

3.2.2.1. Reanalyses

Reanalysis (retrospective analysis) datasets use data assimilation techniques to combine climate observations with forecasts from numerical weather prediction (NWP) models to produce globally complete atmospheric circulation fields which are consistent with observed conditions and with the physical/dynamical processes represented by the NWP model (Compo et al., 2011). Multiple reanalysis products exist, varying from relatively short (post-1940) datasets which assimilate large amounts of climatic information (e.g., Kobayashi et al. (2015) and Hersbach et al. (2020)), to longer datasets which assimilate fewer variables (e.g. Laloyaux et al. (2018); Poli et al. (2016); Slivinski et al. (2019, 2020)). The latter are referred to as 'sparse-input reanalyses' and may provide useful reference series for breakpoint detection because: the reanalysis data should contain the same climate signal as the observations but should not be affected by the same biases/inhomogeneities (assuming the potentially inhomogeneous variable being assessed is not assimilated), and reanalyses are globally-complete, so corresponding reference series will be available for all weather stations in CRUTEM5_sdb.

The use of reanalysis data as reference series to address network-wide breaks is supported by the World Meteorological Organization (2020) and, although still a relatively novel approach, a few studies have demonstrated its potential. Haimberger (2007) and Haimberger et al. (2012) successfully used reanalysis data to identify breakpoints in radiosonde temperatures; Azorin-Molina et al. (2014) used reanalysed wind speeds to homogenise wind speed data over Spain and Portugal and Gillespie et al. (2021) demonstrated the potential to use sparse-input reanalyses to detect breakpoints in observed LSAT data, before applying the technique to assess the homogeneity of the International Surface Temperature Initiative database (Gillespie et al., 2023).

The NOAA-CIRES-DOE Twentieth Century Reanalysis version 3 (20CRv3) dataset is an example of a sparse-input reanalysis which could help to detect exposure biases in

CRUTEM5_sdb. The dataset spans the period 1806-2015 (which covers the period likely to contain breaks related to the introduction of Stevenson screens), has a relatively high spatial and temporal resolution (approximately 0.75 degrees and 3-hourly, respectively), is forced by observation-based SST and radiative forcing and assimilates only surface pressure observations (Slivinski et al., 2019, 2020), meaning it is entirely independent of CRUTEM5_sdb and should not be affected by the exposure bias. A preliminary assessment suggests 20CRv3 is also highly and significantly correlated with CRUTEM5, especially over the region between 30° to 60° latitude (Figure 3.1). Therefore, using 20CRv3 as a reference series will be explored further in Section 3.3.

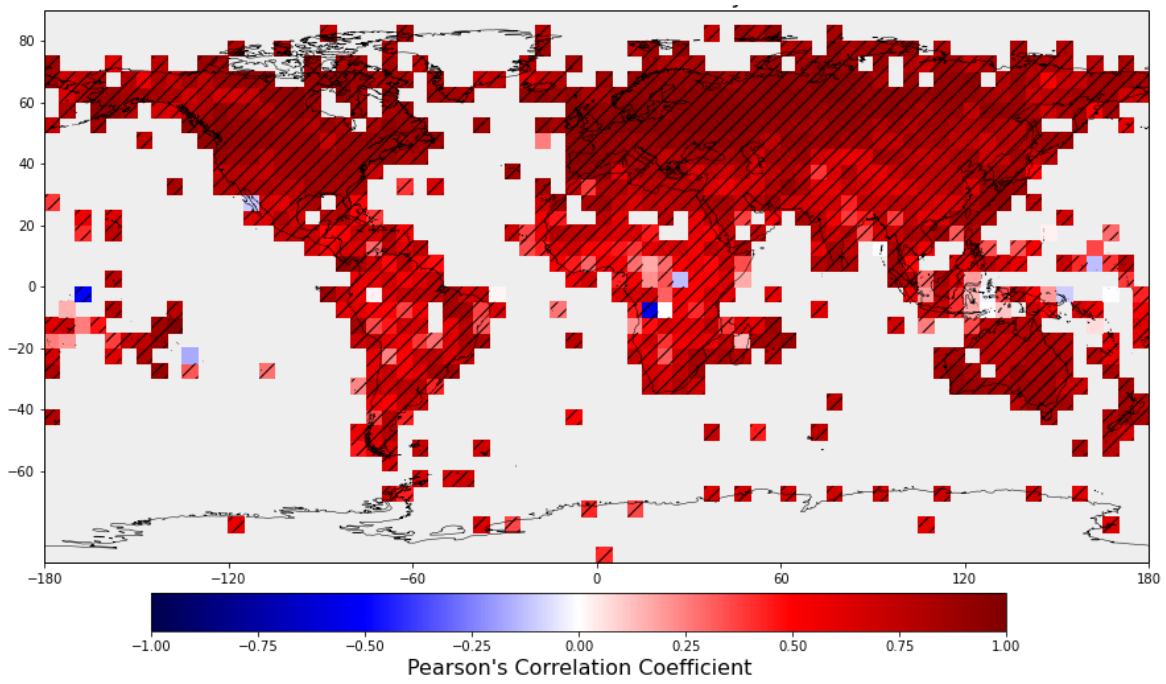


Figure 3.1. Correlation between CRUTEM5 and 20CRv3 for the period 1900 - 2015. The hashing identifies significance at the 95% level.

3.2.2.2. Proxy reconstructions

Climate proxies are indirect measures of past climatic conditions which have been recorded in historical documents or preserved in natural archives, such as: tree rings, corals and ice sheets, by climate-dependent physical, chemical or biological processes (Bradley, 2014). Proxy temperature reconstructions are potentially useful as reference series for relative breakpoint detection for similar reasons as reanalyses (Section 3.2.2.1). High quality proxy reconstructions should contain the same climate signal as the observations, proxies are independent of instrumental LSAT observations (with the possible exception of during their calibration period) and therefore should not be affected by the same biases/inhomogeneities,

and in the early instrumental period proxy data often have broader spatial coverage than the instrumental observations (Figure 3.2).

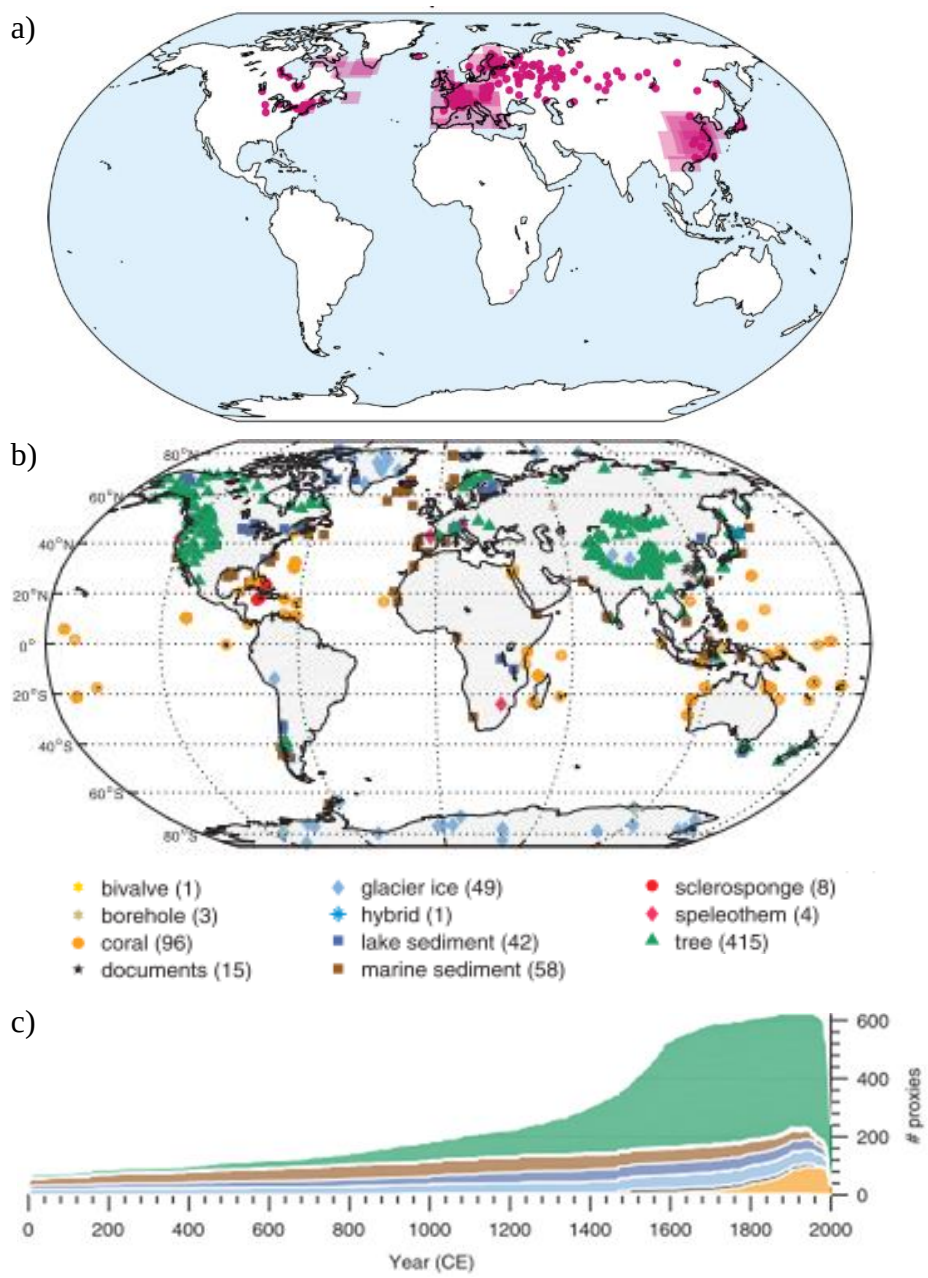


Figure 3.2. Spatial availability of a) documentary reconstructions of temperature (at least 30 years in length) for the period 1400-1880 and b) temperature-sensitive proxy data in the PAGES2k database. c) Temporal availability of the proxies in the PAGES2k database. Figure a) is from Burgdorf (2022); b) and c) are from PAGES2k Consortium (2017).

The use of proxy reconstructions to assess instrumental datasets for potential inhomogeneities and biases is well established. Pfeiffer et al. (2017), for example, used a coral-based proxy reconstruction of temperature in the Indian Ocean to identify a bias in observed sea surface temperatures and several European studies (including: Böhm et al.

(2010); Büntgen et al., (2006); Frank et al. (2007); Hiebl et al. (2006) and Meier et al. (2007)) have used proxy temperature reconstructions to highlight potential instances of the exposure bias in LSAT observations. The use of proxy reconstructions to identify exposure biases in CRUTEM5_sdb will therefore be explored further in the following section (3.3).

3.3. Exploring the potential to use 20CRv3 and proxies to identify exposure biases

Section 3.2.2 highlights the limitations of using neighbouring weather station data as reference series for relative breakpoint detection and outlines the potential to use two alternative sources of temperature data (reanalysis and proxy reconstructions) to identify pervasive inhomogeneities (such as the exposure bias). This section explores this potential in more detail via a comparison between 20CRv3, four proxy reconstructions and one early instrumental series from CRUTEM5_sdb: Berlin Dahlem. Note that other early instrumental series from CRUTEM5_sdb were also assessed, but only the assessment of the Berlin Dahlem series is presented here.

3.3.1. Data & Methods

3.3.1.1. Berlin Dahlem weather station

Berlin Dahlem weather station (referred to going forward as Dahlem) is located on the outskirts of Berlin, Germany, at approximately: 52.5°N, 13.3°E. It is one of the longest continuous records in CRUTEM5_sdb, starting in January 1769 and continuing until the present day. The station was chosen for assessment here because a) the record begins almost a century before the invention of the Stevenson screen in the 1860s (Stevenson, 1866), meaning changes in exposure are likely to have occurred, and b) the series is not known to have been adjusted for the exposure bias, meaning any bias should still be present in the observations.

3.3.1.2. 20CRv3

The absolute temperature series from the grid cell containing the Dahlem weather station (centred on 52.281°N, 13.359°E) was extracted and the correlation between the reanalysis ensemble mean and the observations was tested using Pearson's correlation coefficient applied on an annual 30-year rolling basis. This assessment was conducted in addition to the assessment shown in Figure 3.1, as the correlation may not remain stable over time (e.g. due to the quantity and quality of assimilated data) and because a good correlation between the gridded data does not guarantee the same for an individual station series (Dee et al., 2014;

Slivinski et al., 2020). The two series were found to be sufficiently correlated ($r \geq 0.8$) from 1850.

A 'best fit' version of 20CRv3 was then created by linearly regressing the 20CRv3 series onto the Dahlem observations (on a monthly basis) over the period 1950-2015. This was done to compensate for any biases which may affect 20CRv3, or any microclimatic factors which may produce an offset between 20CRv3 and the Dahlem observations. The period 1950-2015 was chosen for the regression because the data is more certain during that period and, importantly, because the data is unlikely to be affected by the exposure bias (arising from the introduction of a Stevenson screen) during that period (Parker, 1994). A comparison between the 'best fit' 20CRv3 series and the Dahlem observations was then undertaken. As this is an initial exploratory analysis, no statistical breakpoint detection techniques were applied at this stage. Note that all future references to 20CRv3 (in Section 3.3), including Figures 3.3 to 3.5, refer to the 'best fit' version of the data.

3.3.1.3. Proxy reconstructions

A review of the literature and databases of palaeoclimate data (e.g. Burgdorf (2022); Burgdorf et al. (2023); National Oceanic and Atmospheric Administration (n.d.); PAGES2k Consortium (2017) and St. George (2014)) was conducted to identify high quality proxy temperature reconstructions which were: a) independent of the Dahlem observations (at least during the period of interest, see Table 3.1: 'Calibration period'), b) located no more than 1,200km from Dahlem weather station (the correlation decay length for monthly mean air temperature (Harris et al., 2014)) and c) sufficiently temporally resolved (monthly-to-annually) and of sufficient duration to be used as reference series. More than ten proxy reconstructions met these criteria and four were selected for further assessment: one tree ring-based reconstruction (Anchukaitis et al. (2017)) and three documentary-based reconstructions (Dobrovolný et al. (2010); Možný et al. (2016) and Tarand and Nordli (2001)) (see Table 3.1 for details). The former three reconstructions were chosen because of their proximity to Dahlem weather station - all are within 400km - and the latter was chosen to enable an assessment of the Dahlem winter temperatures. Note that Anchukaitis et al. (2017) is a five-degree gridded dataset for the Northern Hemisphere; going forward when Anchukaitis et al. (2017) is referred to, it is the series from the grid cell containing the Dahlem weather station (centred on 52.5°N, 12.5°E) that is being discussed.

Each proxy reconstruction selected for use here preferentially represents the temperature in a sub-section of the year: for example, Anchukaitis et al. (2017) is most strongly correlated with May-to-August (MJJA) temperatures and Možný et al. (2016) with April-to-August (AMJJA) (see Table 3.1: 'Season' column). To enable a comparison with the proxy reconstructions, corresponding seasonal means were produced from the Dahlem series and (where necessary) both the observations and reconstructions were converted to temperature anomalies relative to the 1961-1990 mean. Pearson's correlation coefficient was then used to assess the correlation between the proxy reconstructions and Dahlem observations between 1950-2000 and further comparisons undertaken if the proxy reconstruction was deemed sufficiently correlated. Note that, as proxies can be influenced by factors other than temperature, a lower correlation threshold of $r \geq 0.6$ was considered sufficient for use here (analysis by Briffa et al. (2002) suggests only the top quartile of tree-ring-density-based temperature reconstructions have a correlation coefficient of ≥0.6). As in Section 3.3.1.2, no statistical breakpoint detection techniques were applied at this stage.

Table 3.1. Details of the proxy reconstructions used as reference series. The 'Season' column details the months the reconstruction represents and the 'Calibration period' column details the instrumental data and time period used to calibrate the reconstructions.

Study	Location	Duration	Proxy	Season	Calibration period
Anchukaitis et al. (2017)	Grid cell centre: 52.5°N, 12.5°E	750-2014	Tree rings	MJJA	1945-1988 Data: Cowtan and Way (2014)
Dobrovolný et al. (2010)	Central Europe 49°N, 13°E	1500-2007	Documentary	JJA	1771-1816 Auer, Böhm and Schöner (2001)
Možný et al. (2016)	North of Prague, Czech Republic	1499-2015	Grape harvest dates	AMJJA	1811-2010 Data: Brázdil et al. (2012)
Tarand and Nordli (2001)	Tallin, Estonia 59.4°N, 24.8°E	1500-2000	Ice break-up dates	DJFM	1757 onwards Data: Tallinn, Stockholm & St Petersburg series

MJJA = March-to-August; JJA = June-to-August; AMJJA = April-to-August; DJFM = December-to-March.

3.3.2. Results

3.3.2.1. 20CRv3

A comparison between the Dahlem observations and 20CRv3 (Figure 3.3a) illustrates close agreement between the annual means from approximately 1910, and in the interannual variability from approximately 1850, but highlights an offset between the two series prior to

1910 when the observed temperatures are approximately 1°C warmer than the 20CRv3 ensemble mean. This offset is particularly evident in the difference between the two series (Figure 3.3b) and suggests a breakpoint occurred in 1908, when there is a transition from generally warmer observed temperatures to similar or cooler, and when the difference between the series jumps by -0.76°C. The correlation between the Dahlem observations and the 20CRv3 ensemble mean also supports the occurrence of a breakpoint in 1908. Figure 3.4 shows high seasonal correlations between the two series for the period 1850-1907 and 1908-2015 ($r \geq 0.87$) but highlights an offset in the regressions between each period, indicative of a step-change in the mean temperatures in one series, as may occur due to a breakpoint.

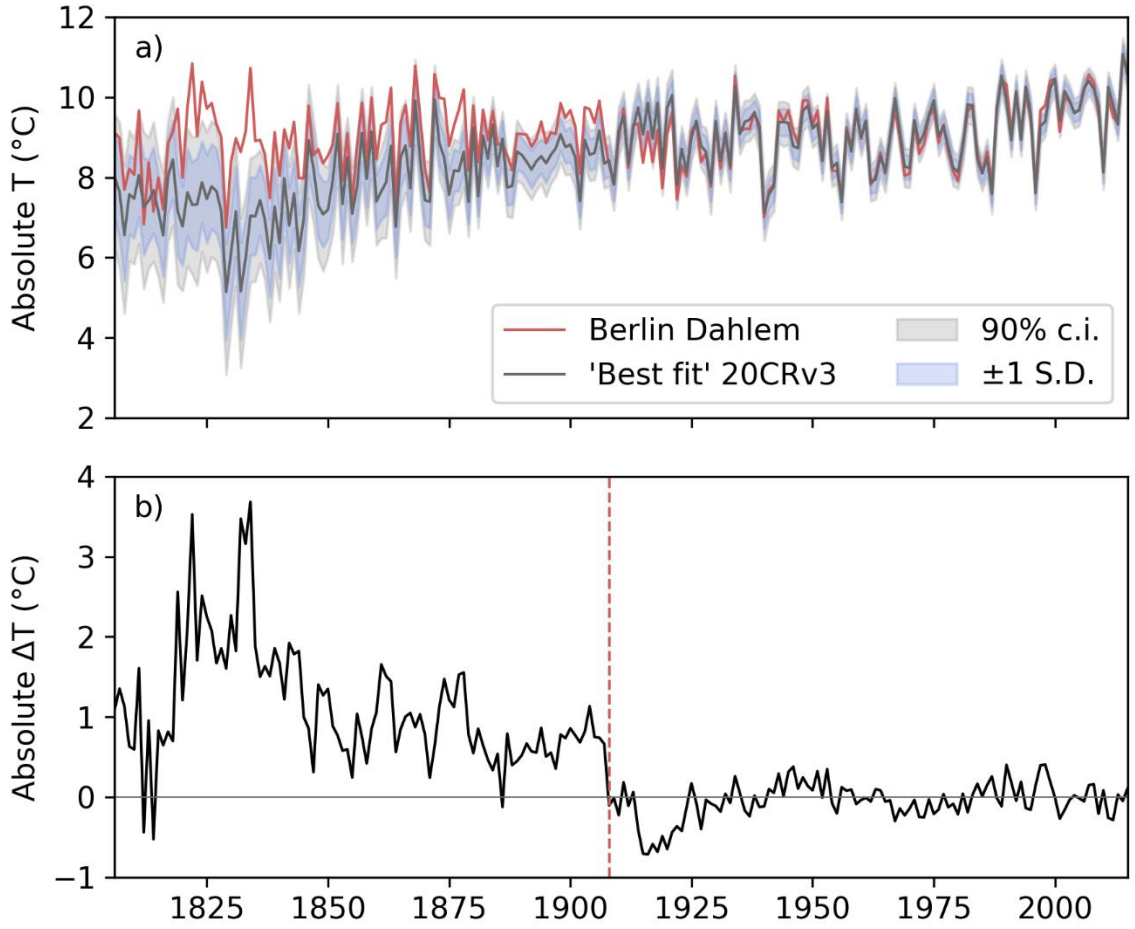


Figure 3.3. a) Annual mean Berlin Dahlem observations compared to the 'best fit' version of 20CRv3. The grey and blue shading show the 90% confidence interval and ±1 standard deviation of the 20CRv3 ensemble, respectively. b) Difference between the observations and 20CRv3 (Berlin Dahlem minus 20CRv3). 1908 is marked by the red dashed line.

An examination of the frequency and magnitude with which the Dahlem observations exceed the 20CRv3 ensemble mean and 95th percentile (Figure 3.5) further strengthens the evidence for a breakpoint in 1908. Visually, there is a clear change in the frequency and magnitude of

exceedance in all seasons in 1908 (marked by the dashed red line in Figure 3.5), and statistically the frequency of exceedance indicates a breakpoint in the data. If 20CRv3 provides an accurate representation of the climate in Europe (which previous studies and the strong correlations found in Figure 3.1 and Figure 3.4 suggest it does (Gillespie et al., 2021; Slivinski et al., 2020)) then, by definition, only 5% of the Dahlem observations might be expected to fall outside of the 20CRv3 95th percentile. This is approximately the case post-1908 when 2%-7% of seasons exceeds the 95th percentile, however, between 1850-1907, a substantially higher percentage of seasons exceed it: ranging from 31% of winter mean temperatures to 74% of autumn mean temperatures. These percentages are substantially higher than 5% and are unlikely to have occurred by chance. Thus, these results strongly suggest a breakpoint occurred in 1908 in either the Dahlem observations or the 20CRv3 data.

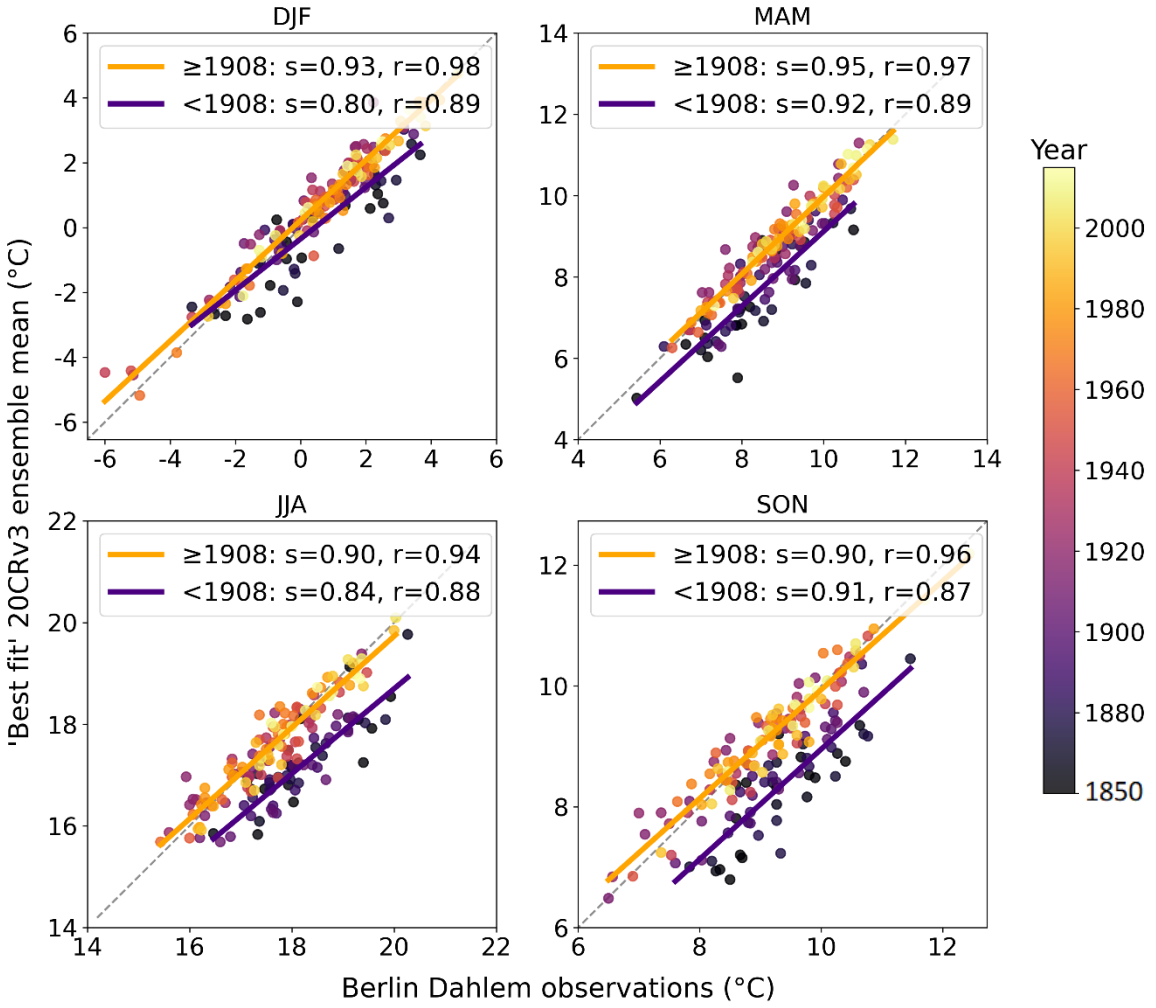


Figure 3.4. Seasonal correlations and regressions between the 20CRv3 ensemble mean and the Berlin Dahlem observations. The colours of the markers delineate the observation year (only the period 1850-2015 is shown), 's' is the slope of the regression line and 'r' is Pearson's correlation coefficient.

Further assessment of Figure 3.4 and Figure 3.5 highlights a possible seasonal component to the breakpoint. Between 1908-2015 the correlations shown in Figure 3.4 are relatively stable between seasons, confirming there are no inherent seasonal differences between the two series; however, prior to 1908, differences between the seasons become apparent. The offset in the regressions between the pre- and post-1908 periods is larger in summer and autumn than in the winter, suggesting the cause of the breakpoint has a seasonal cycle and affects the summer months more than the winter. This suggestion is supported by Figure 3.5 which clearly illustrates that the observed summer and autumn mean temperatures exceed the 20CRv3 ensemble mean (left-hand panel) and 95th percentile (right-hand panel) more frequently, and by a greater magnitude, than the winter mean temperatures. In summer, for example, the observations exceed the 20CRv3 ensemble mean in every year (between 1850-1907), with a median exceedance of 1.01°C, whereas exceedance occurs in only 72% of winters during the same period, with a substantially lower median exceedance of 0.65°C.

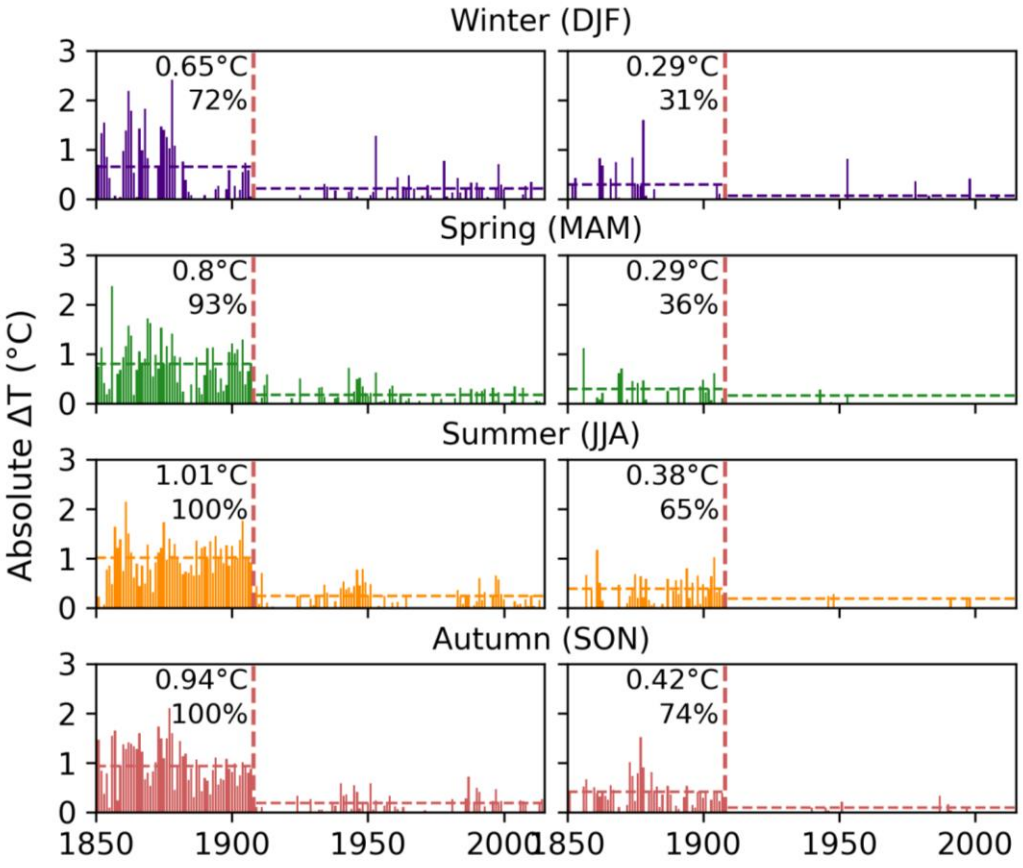


Figure 3.5. Occasions the Berlin Dahlem seasonal mean temperatures exceed the 20CRv3 ensemble mean (left panel) and 95th percentile (right panel). The upper number in each plot is the median observed exceedance between 1850-1907 (inclusive) and the lower number is the percentage of observed seasonal means which exceed 20CRv3 during the same period. The vertical dashed red lines mark 1908.

The timing and characteristics of the offset between the observations and 20CRv3 suggest the breakpoint could have been caused by a change in thermometer exposure. The literature suggests the transition to Stevenson screens occurred in the early-twentieth century in Prussia (which incorporated present-day Berlin) (Gorczynski, 1910; Parker, 1994) and the parallel measurements assessed in Chapter 2 indicate earlier exposures caused a warm bias which was often largest in the summer months. Note, however, that the magnitude of the offsets identified here are larger than the mean June-to-August biases identified in Chapter 2, which were between -0.15°C and -0.32°C dependent on category of exposure, though one parallel measurement series does show a June-to-August bias which is almost as large (-0.98°C compared to approximately -1°C here).

Care, however, needs to be taken when interpreting these results. The 20CRv3 data may not be homogenous: changes in the quantity and quality of assimilated observations and SST boundary conditions can lead to biases in reanalysis data and biases may also exist in the underlying NWP model (Dee et al., 2014). Ferguson and Villarini (2014), for example, found inhomogeneities in an earlier version of the Twentieth Century Reanalysis (Compo et al., 2011). More recent assessments (e.g. Gillespie et al. (2021)) that suggest version 3 is homogenous from at least 1850 (which concurs with the rolling correlations conducted here) give confidence in the homogeneity of the series; however, the fact remains that this comparison, in isolation, cannot definitively attribute the breakpoint to either series.

3.3.2.2. Proxy reconstructions

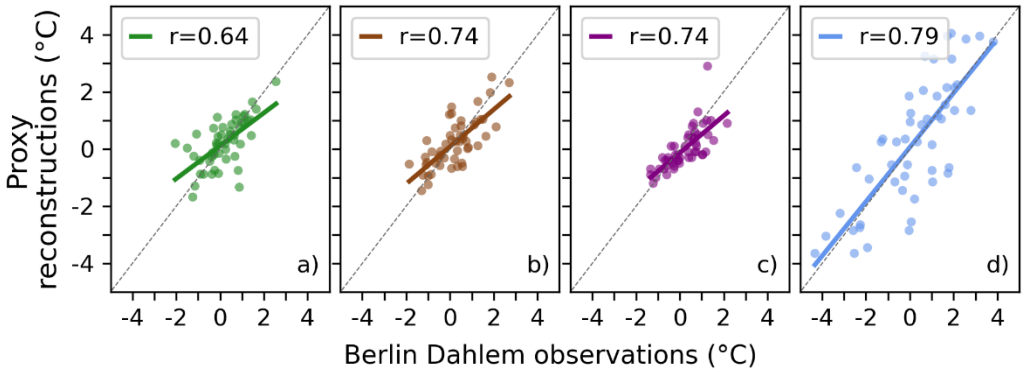


Figure 3.6. Correlation between the corresponding Berlin Dahlem observations and a) Anchukaitis et al. (2017), b) Dobrovolný et al. (2010), c) Možný et al. (2016) and d) Tarand and Nordli (2001) for the period 1950-2000. Pearson's correlation coefficient is given by '*r*'.

Figure 3.6 illustrates reasonably good comparability between the four proxy reconstructions and the Dahlem seasonal means during the recent period (1950-2000). The correlation coefficients for the three documentary-based reconstructions are stronger than for the tree

ring-based reconstruction (0.74-0.79 compared to 0.64, respectively), however all are sufficiently well correlated ($r \geq 0.6$) to be used as reference series. Note that the lower correlation for Anchukaitis et al. (2017) may arise from the fact the reconstruction incorporates information from more distant proxies than the other three reconstructions - each grid cell can include information from up to 2000km away - thus potentially weakening the local climate signal.

The agreement between the four proxy reconstructions and the Dahlem observations is also apparent in Figure 3.7 which shows good agreements between the series in the recent period (from at least 1950) and reasonably consistent interannual variability throughout the records. Prior to 1900, however, the three warm season proxies: Dobrovolný et al. (2010); Možný et al. (2016) and Anchukaitis et al. (2017), appear to diverge from the Dahlem observations, with the proxies indicating cooler temperatures than the instrumental records. The divergence is most pronounced in Anchukaitis et al. (2017) with the observations exceeding the reconstructed temperatures in 92% of years between 1850-1900, with a median exceedance of 1.06°C (and a maximum exceedance of 2.26°C). Of the three warm season reconstructions, Možný et al. (2016) shows the smallest divergence, however, there does appear to be a shift to generally cooler reconstructed temperatures pre-1900, with warmer observations than reconstructed again present in 92% of years (between 1850-1900), but with a smaller median exceedance of 0.76°C. Although each warm season reconstruction indicates the spring/summer instrumental observations may be too warm in the nineteenth century, the difference series (black lines, Figure 3.7a-c) do not exhibit a clear (or consistent) breakpoint. Možný et al. (2016) presents some evidence for a breakpoint in 1908 – ΔT jumps from 0.46°C in 1907 to -0.79°C in 1908 – but the divergence between the observations and Anchukaitis et al. (2017) and Dobrovolný et al. (2010) appears to be more gradual with no obvious breakpoint evident.

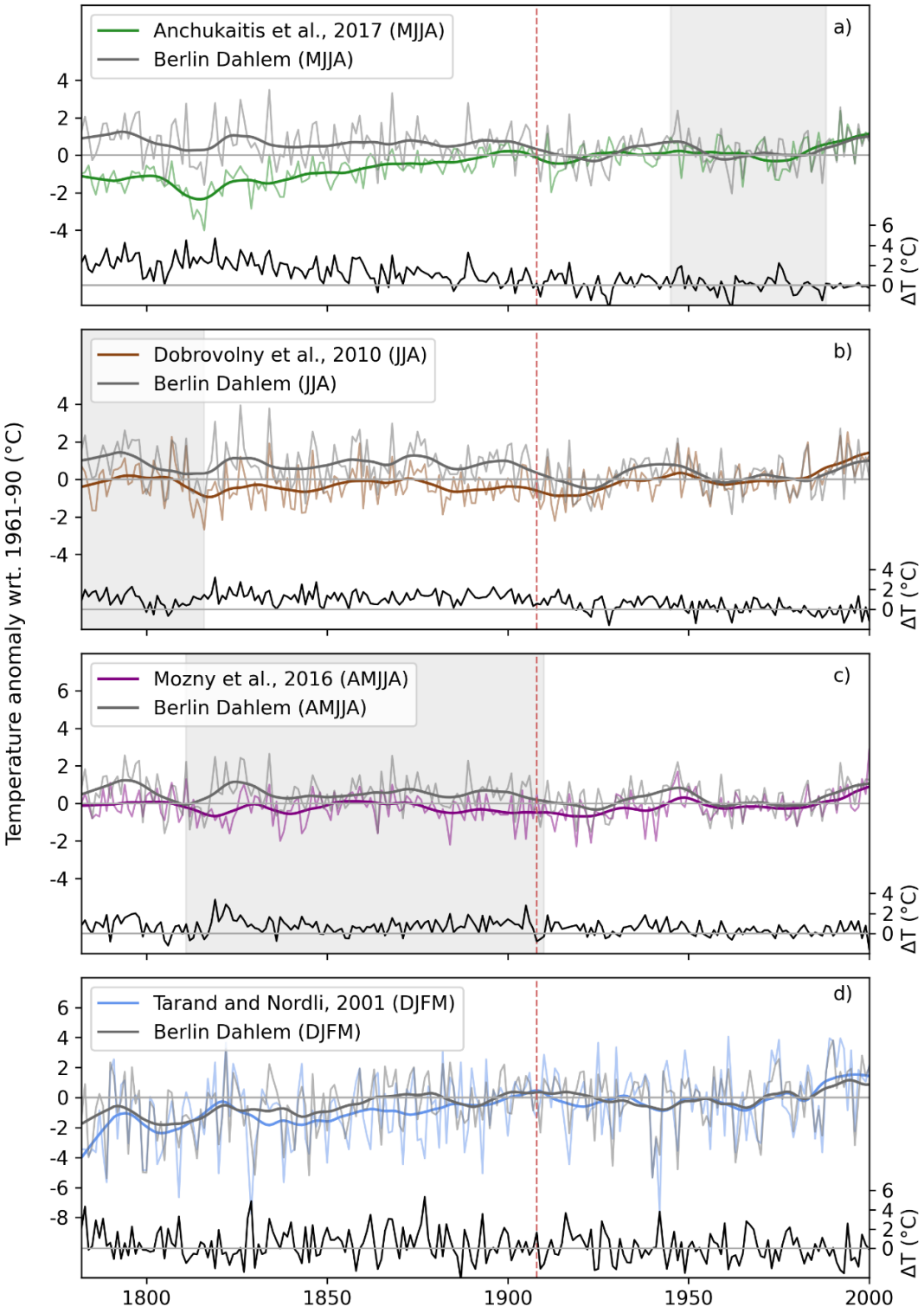


Figure 3.7. Comparison between the Berlin Dahlem observations and a) Anchukaitis et al. (2017), b) Dobrovolný et al. (2010), c) Možný et al. (2016) and d) Tarand and Nordli (2001). The grey and coloured lines represent the observed seasonal means and proxy reconstructions, respectively, and the black line (secondary axis) shows the difference between the two (observations minus proxy reconstruction). The dashed vertical red lines identify 1908 and the grey shading represents the proxy calibration period.

Interestingly, the only winter temperature reconstruction assessed here, Tarand and Nordli (2001), does not show a clear divergence from the observations in either a visual comparison of the two series (Figure 3.7d), or in the magnitude of the difference series which remains relatively stable throughout (black line, Figure 3.7d). As in Section 3.3.2.1, this suggests there may be a seasonal component to the divergence, with the winter months less effected. This is also supported by the fact that Možný et al. (2016) - which represents spring and summer temperatures (AMJJA) - exhibits a smaller divergence from the Dahlem observations than Anchukaitis et al. (2017) and Dobrovolný et al. (2010), which primarily represent the temperature in the summer months: May-to-August and June-to-August, respectively.

Note, however, that the absence of an obvious divergence between the Dahlem observations and the Tarand and Nordli (2001) reconstruction does not necessarily mean the divergence identified in Figure 3.7a-c is seasonal nor that the winter observations are homogenous. Likewise, the presence of a divergence from the proxy reconstructions does not necessarily mean the summer observations are inhomogeneous. In both instances, the presence/absence of a divergence could instead be a reflection of the proxies each reconstruction is based on. All three warm season reconstructions incorporate information from biological proxies: Anchukaitis et al. (2017) is based on tree ring indicators (e.g., ring width and density), Možný et al. (2016) is based on grape harvest dates and Dobrovolný et al. (2010) incorporates information from various sources including phenological data and crop harvest dates, whereas Tarand and Nordli (2001) does not, the reconstruction is based on ice break-up dates. Biological proxies are influenced by factors other than temperature. For example, tree ring width can be affected by soil moisture, CO_2 concentration, tree age, forest management practises and pests (Briffa, 1995; Frank et al., 2022) while crop and grape harvest dates can be affected by changing crop varieties, in addition to many of the previously listed variables (de Cortázar-Atauri et al., 2010). As a result, the divergence evident in Figure 3.7a-c could reflect changes in any one of the listed factors, or changes in the relative influence that each factor has on the proxy through time, rather than a breakpoint in the Dahlem observations. The temperature signal derived from observations of ice break-up is not affected by the majority of the factors listed above - although it may be influenced by others, such as observer presence or subjectivity (Nordli et al., 2007) - therefore the same break would not occur if it were unrelated to temperature. Therefore, the lack of divergence between the observations and Tarand and Nordli (2001) could indicate a seasonal breakpoint

in the observations, consistent with an exposure bias, or it could indicate a non-temperature related divergence specific to the biological proxies/warm season. Thus, these comparisons do not provide definitive evidence of a breakpoint in the Dahlem observations.

3.3.3. Discussion and conclusions

This section has explored the potential to use 20CRv3 and proxy temperature reconstructions as reference series in order to identify exposure biases in CRUTEM5_sdb, using the Berlin Dahlem observed temperature series as a case study. 20CRv3 and proxy-derived temperatures were identified as potentially useful reference series because: a) the series should contain the same regional climate signal as the observations and b) the series are (largely) independent of the temperature observations and thus should not be affected by the exposure bias (as long as the observations are not used as a calibration target for the proxy reconstruction during the period of interest). Additionally, previous studies have highlighted the potential to use reanalyses (e.g. Gillespie et al. (2021, 2023)) and proxy reconstructions (e.g. Frank et al. (2007) and Pfeiffer et al. (2017)) to identify inhomogeneities in temperature series and official guidance (World Meteorological Organization (2020)) now recommends the use of reanalyses for homogenisation where an inhomogeneity affects a large proportion of the observing network (as is the case for the exposure bias).

The preliminary comparisons between the Berlin Dahlem instrumental observations and data from proxy reconstructions and 20CRv3 appear to confirm this potential. The comparisons both suggest temperatures in the nineteenth century were cooler than the Berlin Dahlem observations indicate, particularly in the summer and autumn months, with remarkable agreement between the comparisons evident. Between 1850-1907 (1900 for the proxy reconstructions), for example, the Dahlem JJA means exceeded the 20CRv3 ensemble mean and the Anchukaitis et al. (2017) reconstruction in 100% and 92% of summers, respectively, with a median exceedance of 1.01°C and 1.06°C. The only inconsistency between the comparisons is in regard to whether the winters were also cooler than the instrumental observations indicate. 20CRv3 suggests they were prior to 1908, whereas the winter proxy reconstruction (Tarand and Nordli (2001)) does not provide compelling evidence of this. The timing of the breakpoint is also not conclusive: the 20CRv3 data strongly suggests the breakpoint occurred in 1908 (and Možný et al. (2016) provides some support for this) however, there is no clear breakpoint year present in the difference series with Anchukaitis et al. (2017) and Dobrovolný et al. (2010).

In both the comparisons, the characteristics of the possible breakpoint are generally consistent with the characteristics of the exposure bias indicated in Chapter 2 (Section 2.3). The likely timing of the breakpoint in 1908 falls within the period Stevenson screens were introduced in Prussia; the divergence most likely has a seasonal component; and the magnitude of the differences between the observations and the reference series are similar to those identified in Section 2.3 (although marginally higher). As a result, it may be logical to suggest the divergence between the series occurs due to the introduction of a Stevenson screen. However, as noted previously, while the characteristics of the divergence can provide an indication of the cause of the breakpoint, these comparisons alone cannot confirm this, nor that the breakpoint occurs in the Dahlem observations. Although, in this case, the likelihood of the latter is greatly increased by the presence of the divergence in four different reference series which are unlikely to be affected by the same biases.

To determine whether the breakpoint identified in the comparisons represented a true inhomogeneity in the Dahlem observations, a subsequent review of the station metadata was undertaken. This confirmed the occurrence of a non-climatic break in 1908 when the Berlin Dahlem weather station was relocated from the city of Berlin to the suburb of Berlin-Dahlem and a Stevenson screen was introduced (Figure 3.8) (Cubash & Kadow, 2011; Pelz, 2007; Smithsonian Institution, 1927). Thus, in this instance the use of proxy reconstructions, and particularly 20CRv3, as reference series led to the successful identification of a non-climatic breakpoint arising partially from the introduction of a Stevenson screen.

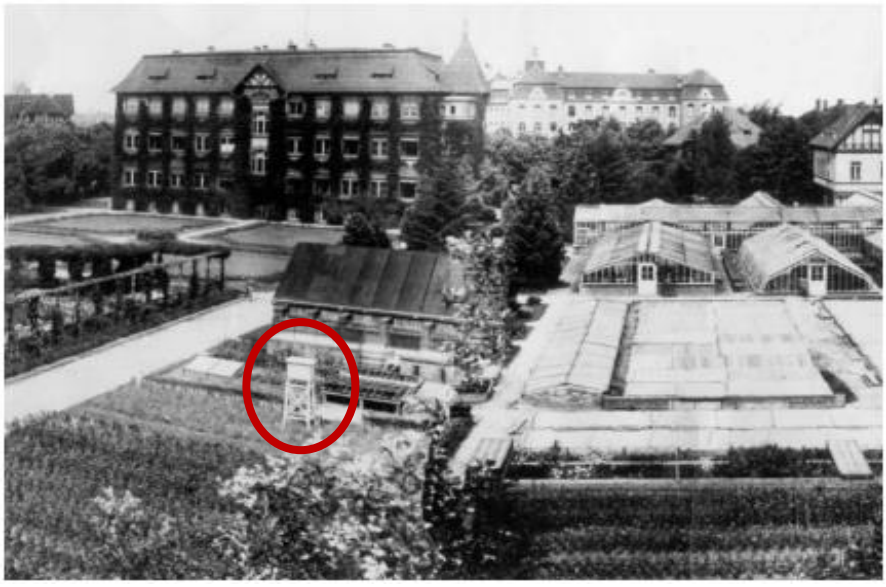


Figure 3.8. Location of the Berlin Dahlem weather station from 1908. Note the Stevenson-type screen circled in red. Image from: Pelz (2007).

However, there are some limitations to applying this method more broadly for the purposes of determining when and where to apply the bias-estimation models to stations in CRUTEM5_sdb. The first limitation concerns data availability. Although 20CRv3 is globally complete, the correlation between 20CRv3 and individual series in CRUTEM5_sdb is not always sufficient for 20CRv3 to be used as a reference series. This is because the quality of 20CRv3 (in terms of reproducing the observed climate signal) depends on factors such as the quality and quantity of assimilated observations (Dee et al., 2014; Sterl, 2003). Europe is an observation-rich region, and thus 20CRv3 is reasonably well-constrained in Berlin throughout its record; however, many regions are not as well-constrained, particularly in the early period (Compo et al., 2019; Slivinski et al., 2020). In Australia and South America, for example, very few surface pressure observations are assimilated until the late-1870s and early-1930s, respectively (Compo et al., 2019), and the lack of observational constraint is reflected in the weaker correlations between 20CRv3 and observations in those regions. A 30-year annual rolling correlation between 20CRv3 and Moruya Heads weather station in southeast Australia, for example, remains below 0.8 from the start of the observed temperature record in 1876 until 1927 and is not consistently above 0.8 until the 1980s. Thus, in some regions and time periods, 20CRv3 is not sufficiently correlated with individual series in CRUTEM5_sdb (for a sufficient duration) to be used as a reference series.

This limitation is not restricted to 20CRv3. Despite the reasonably good spatial coverage of proxies in the early instrumental period (compared to instrumental observations; Figure 3.2), well correlated, independent proxies with sufficient temporal resolution and duration to act as reference series are not available for all series in CRUTEM5_sdb which require assessment. This is because many proxies are unevenly spatially distributed and geographically limited by factors such as climate. Tree-ring-based temperature reconstructions, for example, are generally limited to the high latitudes or altitudes where temperature is the primary limiting factor for tree growth and ice-core-based reconstructions are limited to regions which have (or have recently had) continuously accumulating glaciers or ice sheets (Bradley, 2015). Documentary-based reconstructions are also unevenly spatially distributed; they are only found in regions which have had permanent and literate societies over recent centuries, which have long traditions of recording weather or climate dependent phenomena (e.g. phenological records of cherry tree flowering in Japan (Aono & Kazui, 2008)) and where there have been research efforts to convert the records into reconstructions of temperature (Burgdorf, 2022; Burgdorf et al., 2023; Pfister et al., 2009).

As a result, documentary-based reconstructions are currently mostly confined to the Northern Hemisphere and are most abundant in Europe and East Asia (Figure 3.2a) (Burgdorf, 2022; Burgdorf et al., 2023; Pfister et al., 2009). Due to this uneven distribution, in some regions, particularly regions such as South America and Australia where the capacity of 20CRv3 to act as a reference is also diminished, there may not be any (proxy- or reanalysis-based) reference series available to use. In addition, even where proxy reconstructions are available, independent and well-correlated with the instrumental series, they may not be suitable for identifying breakpoints arising from the introduction of Stevenson screens. If only winter temperature reconstructions are available, for example, it may be more difficult (or impossible) to identify relevant breakpoints, as the exposure bias is generally small, or negligible, in the winter months (Chapter 2). This is demonstrated by the Dahlem case study: if only the Tarand and Nordli (2001) reconstruction was available for comparison, the 1908 breakpoint wouldn't have been identified.

The second key limitation regards the inability to distinguish whether the breakpoint is present in the candidate or reference series, or to determine the cause of the break. In order to apply the bias-estimation models developed in Chapter 2, confidence that the breakpoint occurs in the observations and that the break arises due to the introduction of a Stevenson-type screen is required. In traditional breakpoint detection the former would usually be addressed by comparing the candidate series to multiple reference series - if the break occurs consistently in multiple comparisons, then it can be attributed to the candidate (Caussinus & Mestre, 2004; Menne & Williams, 2009). For many series in CRUTEM5_sdb, however, it may not be possible to make multiple comparisons. 20CRv3 is currently the only sparse-input reanalysis which starts before 1900, so a second reanalysis reference series could not be used, and many regions may only have one proxy or one source of proxy (i.e., tree rings). Therefore, it will not always be possible to determine, with confidence, whether the breakpoint occurred in the CRUTEM5_sdb observations. In terms of determining the cause of the breakpoint, the characteristics of the break or divergence may give an indication (as discussed above), however a review of the metadata would always be required to confirm this, as is the case in the Berlin Dahlem case study.

Finally, and most crucially for these purposes, even if it were possible to overcome the previously noted limitations and every identified breakpoint arose from a change in exposure, metadata would still be required to confirm the latter and to determine the type of historic exposure in place prior to the introduction of the Stevenson screen. Without this

additional information about the historic exposure in use it would not be possible to determine the appropriate bias-estimation model to apply. Therefore, while the use of reanalyses and/or proxy reconstructions have evident potential to identify breaks and inhomogeneities in observations of temperature *in some regions*, they cannot, in isolation, provide the information necessary to apply the bias-estimation models to CRUTEM5_sdb. Thus, the only presently available option is to use metadata to determine the timing of the introduction of the Stevenson screen, and the historic exposure(s) in use prior to the introduction. The following section will therefore outline the development of a database of historic exposures for CRUTEM5_sdb.

3.4. Developing a database of historic exposures for CRUTEM5_sdb

The creation of a database of historic exposures was a substantial undertaking which has not previously been completed for the majority of series in CRUTEM5_sdb. As this study is concerned specifically with the transition to Stevenson-type screens at stations in the mid-latitudes, metadata collation was prioritised for stations or countries between 30° and 60° latitude (North and South of the equator), metadata were only collated for stations which have data before 1961 (by which time the majority of stations would have had Stevenson screens), and only metadata detailing exposures in use prior to the transition to Stevenson screens were recorded. Unless the metadata suggested otherwise, it was assumed that stations did not revert back to historic exposures once the transition to a Stevenson screen had been made.

The HathiTrust Digital Library, Internet Archive, Met Office Digital Library and Archive and the NOAA Foreign Climate Data Repository all proved to be particularly useful sources of exposure metadata (see Appendix B for further details). Each contains freely available scans of yearbooks and other meteorological sources. These sources were carefully examined for any information related to thermometer exposures: either descriptions of the thermometer exposure in use at a station during a given time period, mentions of when exposure changes were implemented, or more general information regarding the type(s) of exposures found in particular regions. Any gathered metadata was mapped to the stations in the CRUTEM5_sdb using information about the station name, station location and/or station ID number.

The exposure metadata were compiled for station records at an annual resolution and given a source code to indicate the level of confidence attached to the information. A source code

of '1' was applied when a given exposure was known to be in use at the station in the specific year; a source code of '2' was applied when the exposure was estimated using station-specific metadata; a source code of '3' was applied when the exposure was estimated and applied to all stations within a country or region based on country or Meteorological Service-specific metadata and a source code of '4' was applied where no exposure information was found. Where the transition between two screen types (including the introduction of the Stevenson-type screen) was known to occur in a particular period, but could not be pinpointed to a year, a transition period was noted in the database and the details of both screen types entered. A flag was also applied if the particular version of a station's data contained in CRUTEM5_sdb was known to have already been adjusted for the exposure bias - such as the stations adjusted by Brunet et al. (2006, 2011) and Böhm et al. (2010) - to prevent additional adjustments being applied to those already-adjusted stations.

Examples of source code '1' (station-specific) and source code '3' (NMS-/ country-specific) information are outlined in the following sections. Details of all the source code '3' information used to populate the database are also given in Appendix B.

3.4.1. Station-specific metadata (Source codes '1' and '2')

Wherever possible, station-specific metadata were used to populate the database of historic exposures. This form of metadata was considered to be the most accurate and reliable and was always prioritised over higher-level source code '3' metadata. Valuable sources of station-specific metadata included: the published literature, Meteorological Service or Observatory Yearbooks and the World Weather Records books collated by the Smithsonian Institution (1927, 1934, 1947). An example of the source code '1' metadata applied to Poznan-Lawica weather station in Poland is provided below.

3.4.1.1. Poznan-Lawica, Poland

Table 3.2. Exposure metadata for Poznan-Lawica, Poland detailed in Kolendowicz et al. (2019).

Location	Height [m ASL]	Height [m AGL]	Thermometer site	Period
Pocztowa St.	–	–	–	Jan 1848–March 1862
Grobla St. 1	–	2.5	Outside the window without a shade	Apr 1862–Sep 1867
ZielonaSt. 1	–	15	Outside the window without a shade	Oct 1867–Jul 1885
Zielona St. 2	–	6.2	Outside the window without a shade	Aug 1885–Sep 1892
Długa St. 3	58	8.6	Outside the window in the zinc box	Sep 1892–Dec 1911
University I (Wały Wazów)	78	2.0	Bigger English instrumental shelter in the garden	Jan 1912–May 1919
University II	78	2.0	Instrumental shelter in the garden	June 1919–December 1920
Poznań-Ławica	83	2.0	Instrumental shelter next to the runway	January 1921–December 2016

Information about the exposures in use at Poznan-Lawica weather station in Poland, including the timing of the introduction of the Stevenson screen, had previously been collated and published by Kolendowicz et al. (2019). The published information, shown in Table 3.2, was used to assign the exposures to one of the four categories described in Section 2.2, and the assigned category was then used to populate the database.

Based on Table 3.2, Poznan-Lawica was given the exposure classification 'Wall-mounted' between 1881 (the start of the record in CRUTEM5_sdb) and 1911 and 'Stevenson screen' between 1912 and present. Note that from 2017 a source code of '2' was applied to the metadata entry on the basis that the metadata in Table 3.2 is not specific to the years after 2016; the exposure classification from 2017 is therefore an informed estimate based on the station-specific metadata (and thus falls under source code '2' rather than '1').

As the Poznan-Lawica record in CRUTEM5_sdb begins in 1881, no exposure metadata were required prior to 1862. If the record had started before 1862, then either country-/Meteorological Service-specific metadata would have been applied (source code '3') or, if unavailable, the exposure would have been listed as 'Unknown' (source code '4').

In this instance, each exposure fell into one of the categories defined in Section 2.2, however, if an exposure did not - for example, a thermometer exposed in an unheated room or shaded by a tree - a classification of 'Miscellaneous' would have been applied and the details of the miscellaneous exposure recorded in the database.

3.4.2. Country-level metadata (Source code '3')

As noted in Section 3.2.1, records of metadata are not always available or easily accessible for weather stations, particularly in the EIP, and even where metadata exist, information about the exposures in use and/or the timing of changes in exposure were not always documented. As a result, station-specific metadata were not found for all weather stations within CRUTEM5_sdb, or for all time periods. In such instances, higher-level information, specific to the country or Meteorological Service that the station was administered by, was used to populate the database. This information was only applied where more specific information (source codes '1' or '2') was not available and every effort was made to account for changes in national borders and governance over time when applying assumptions.

An example of the source code ‘3’ metadata applied to weather stations in India is included below and all of the source code ‘3’ metadata used to populate the database can be found in Appendix B.

3.4.2.1. India

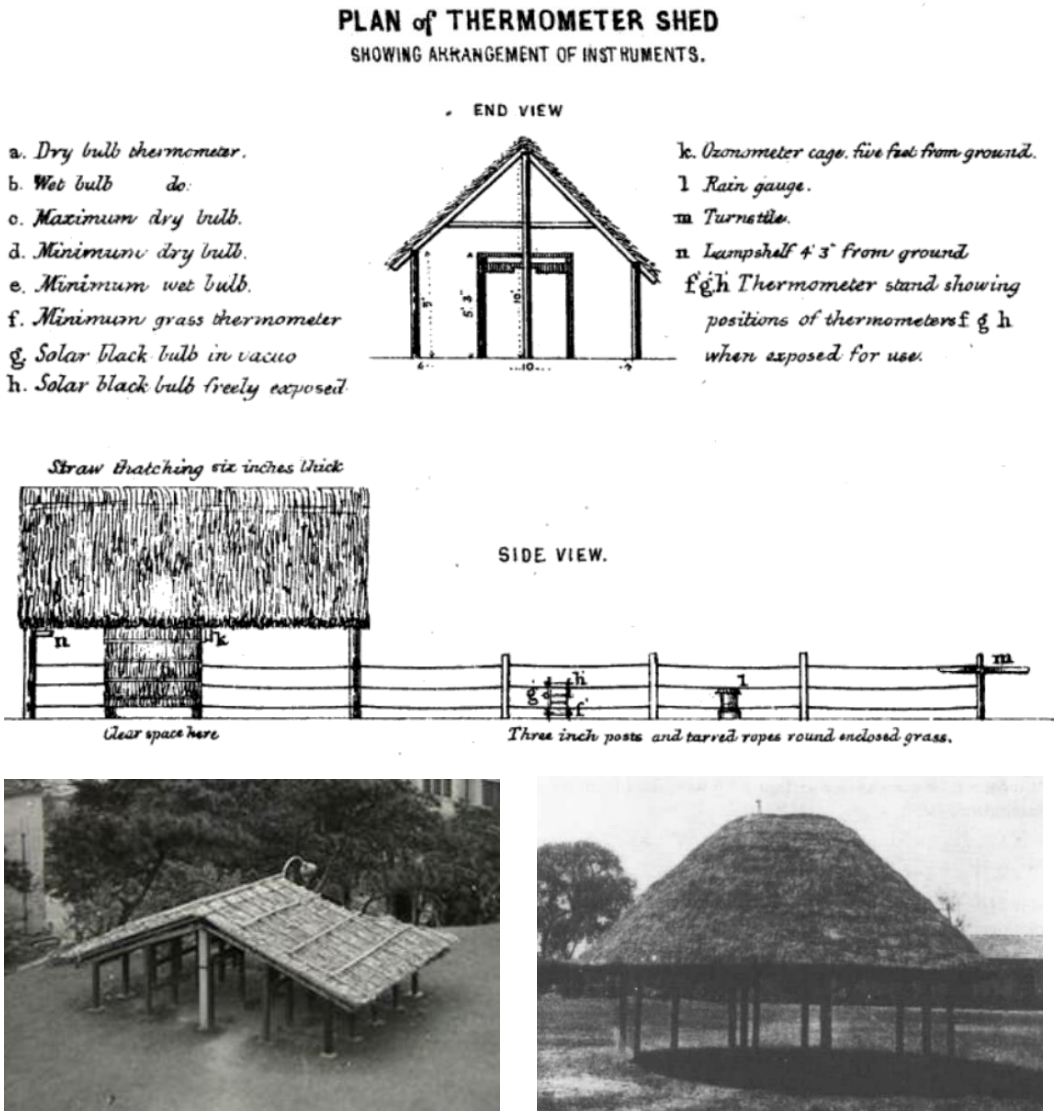
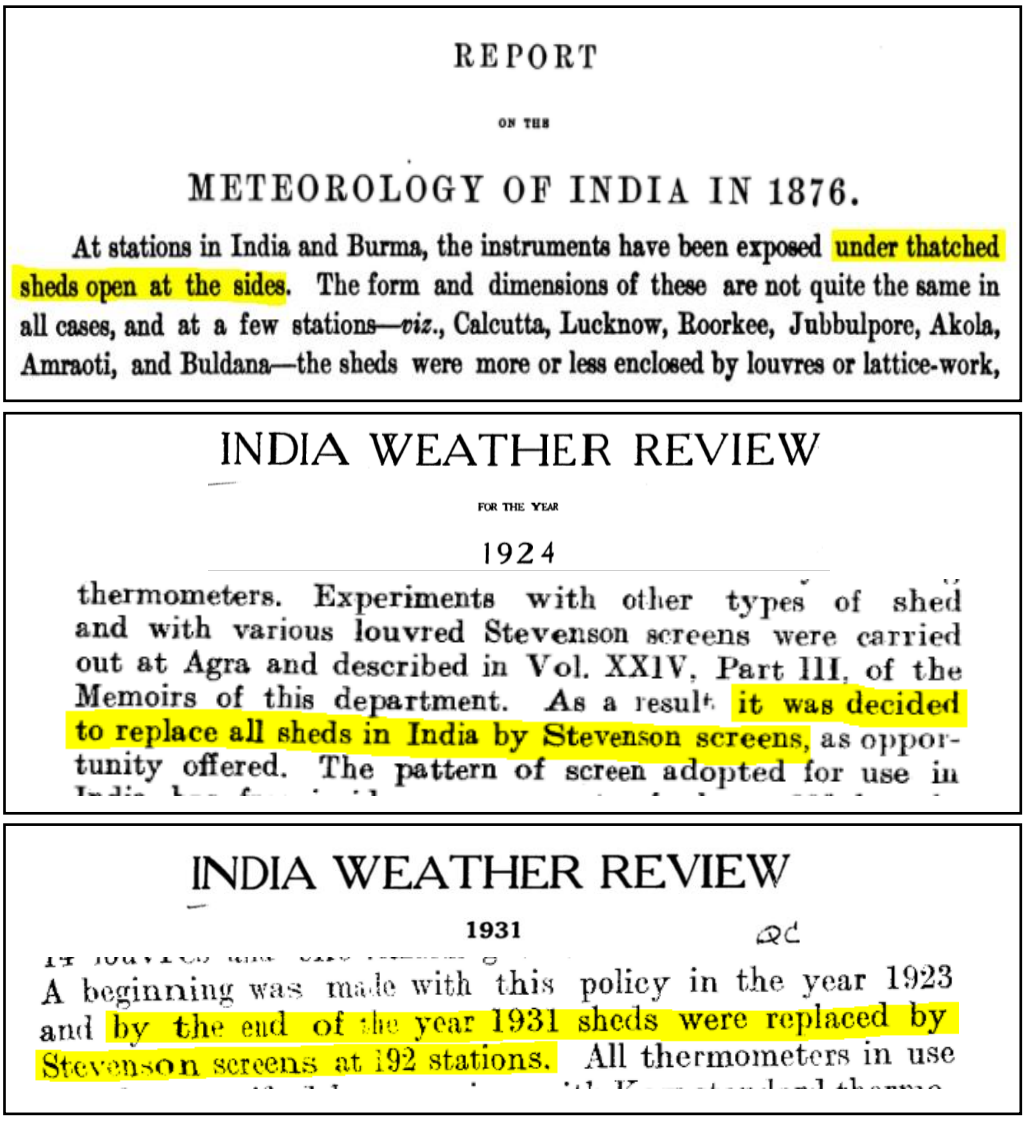


Figure 3.9. Thatched thermometer sheds of the style believed to be in use in India between at least 1876 and 1930. Images (clockwise from top): Report on the Administration of the Madras Presidency during the year 1867-68. (1868), Field (1920) and Lee (2016).

Meteorological observations in India began in the late-1700s, however it was not until 1875 that they were brought under the control of the Government (via the establishment of the Indian Meteorological Network) and observing practises were standardised (India Meteorological Department, n.d.). The first centralised instructions for observers were issued in 1876 (Blanford, 1876b) and recommended exposing thermometers under a

'thatched shed', of the style shown in Figure 3.9. In the same year the 'Report on the Meteorology of India' (Blanford, 1876a) suggested the majority of stations were following this advice (Figure 3.10). The instructions remained the same (regarding thermometer exposure) until 1924 when the 'India Weather Review' (Government of India Meteorological Department, 1926) outlined a decision to replace all thermometer sheds with Stevenson screens. By the end of 1924, 30 stations had reportedly introduced the Stevenson screen, and by 1931 Stevenson screens were in use at 192 weather stations (out of an estimated 200, not all of which are included in CRUTEM5_sdb) (Figure 3.10) (Government of India Meteorological Department, 1926, 1933). The use of Stevenson screens in India continued until at least 1970 (Sparks, 1972).



REPORT

ON THE

METEOROLOGY OF INDIA IN 1876.

At stations in India and Burma, the instruments have been exposed under thatched sheds open at the sides. The form and dimensions of these are not quite the same in all cases, and at a few stations—*viz.*, Calcutta, Lucknow, Roorkee, Jubbulpore, Akola, Amraoti, and Buldana—the sheds were more or less enclosed by louvres or lattice-work,

INDIA WEATHER REVIEW

FOR THE YEAR

1924

thermometers. Experiments with other types of shed and with various louvred Stevenson screens were carried out at Agra and described in Vol. XXIV, Part III, of the Memoirs of this department. As a result it was decided to replace all sheds in India by Stevenson screens, as opportunity offered. The pattern of screen adopted for use in

INDIA WEATHER REVIEW

1931

A beginning was made with this policy in the year 1923 and by the end of the year 1931 sheds were replaced by Stevenson screens at 192 stations. All thermometers in use

Figure 3.10. Descriptions of the thermometer exposures in use in India in the nineteenth and twentieth centuries taken from (top to bottom): Blanford (1876a), Government of India Meteorological Department (1926) and Government of India Meteorological Department (1933).

Based on the information above, the following country-specific (source code '3') exposure classifications were applied to weather stations in India:

- 1876-1923: Intermediate (Indian thatched thermometer shed)
- 1924 - 1930: Transition to the Stevenson screen (Intermediate-to-Stevenson screen)
- ≥1931: Stevenson screen.

Prior to 1876 there are no Indian stations in CRUTEM5_sdb, within 30°- 60° latitude, with observations, so no earlier country-specific assumptions were required or applied.

3.4.3. The CRUTEM5 database of historic exposures

The CRUTEM5 database of historic exposures contains at least one entry of exposure metadata for 88% of stations within 30 to 60 degrees latitude (see Table 3.3 for details of the source codes applied) and illustrates how methods of exposing thermometers have evolved over time. As can be seen in Figure 3.11, wall-mounted exposures were almost exclusively used in the early part of CRUTEM5_sdb, until the introduction of early varieties of freestanding exposures (first open, then intermediate and closed) from the mid-1840s onwards. All four categories of historic exposure were in use in the decades which followed with none universally adopted. This began to change with the invention of the Stevenson screen in the 1860s (Stevenson, 1866). Its use grew gradually at first, but by the early 1900s it was the most commonly used exposure and by the 1930s it had been almost universally adopted in the mid-latitudes.

Table 3.3. Number of stations and months in the CRUTEM5 database of historic exposures which were populated with each source of exposure metadata.

	Stations	Months
Source code '1'	538	146,759
Source code '2'	692	708,851
Source code '3'	4066	4,069,094

The use of, and transition between, exposures was not spatially or temporally homogenous (see Video 3.1). Different countries and Meteorological Services favoured different exposures and introduced the Stevenson screen at different times. In the UK for example, wall-mounted, then open exposures were commonly used before the Stevenson screen was adopted in the early 1870s. In contrast, in India and Southeast Australia, intermediate exposures were the favoured historic exposure, and the Stevenson screen was not introduced

until later – in the 1920s and 1890s respectively. This spatiotemporal heterogeneity reinforces the need to consider the exposure history of individual countries or regions when accounting for the exposure bias.

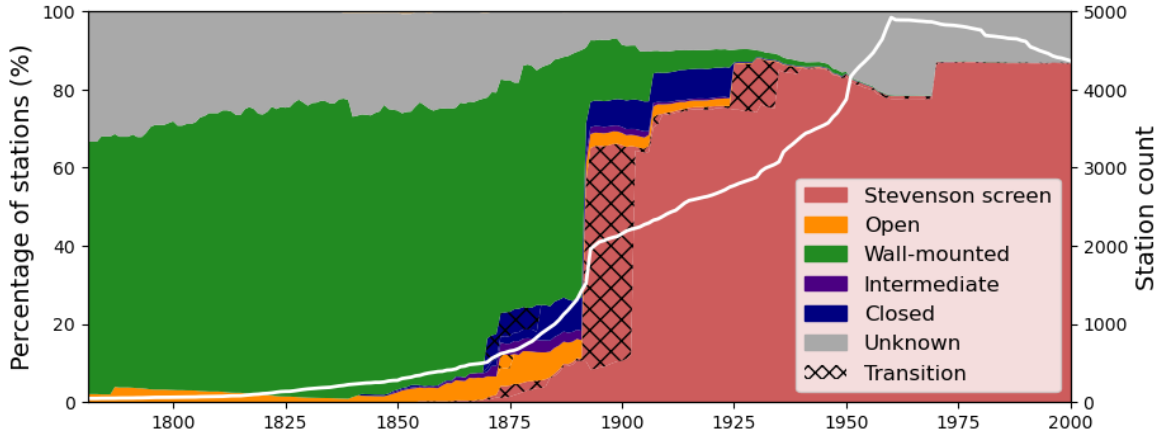


Figure 3.11. Temporal evolution of the thermometer exposures in use at mid-latitude weather stations in CRUTEM5_sdb. The coloured shading represents the percentage of stations with each exposure over time (left-hand axis) and the white line shows the CRUTEM5_sdb station count (right-hand axis). Note, only stations between 30° to 60° latitude, with data before 1961, are included.

The collated exposure metadata for mid-latitude weather stations in the CRUTEM5 station database (with data before 1961) are openly accessible via the Zenodo data repository (zenodo.org/doi/10.5281/zenodo.10551196). One csv file is provided for each monthly timestep from January 1781 to December 2021 and each file contains details of the exposure category and source for each weather station with a temperature value in that month.

3.5. Quantifying the exposure bias in CRUTEM5_ext

Chapter 2 (Section 2.1) highlighted the need to improve the way exposure biases are accounted for in global temperature datasets and identified three elements required to do so:

- an improved understanding of the characteristics of the exposure bias arising from different forms of historic exposure, including how they vary regionally or seasonally.
- the ability to model or otherwise quantify the monthly mean exposure bias globally.
- the ability to identify exposure bias affected series.

Chapter 2 addressed the first two elements by using parallel measurement series to better characterise the bias (Section 2.3) and to develop three bias-estimation models (Section 2.4) and the preceding sections of Chapter 3 have addressed the final element via the creation of a database of historic exposures. In this section the application of these elements to

CRUTEM5_sdb, to quantify the exposure bias in a version of CRUTEM5 which has been extended to start in 1781 (CRUTEM5_ext), is outlined and the results discussed.

3.5.1. Exposure metadata

The CRUTEM5 database of historic exposures, outlined in Section 3.4 was used to identify the stations and time periods in CRUTEM5_sdb affected by the exposure bias as well as the appropriate bias-estimation model to apply. Stations were considered to be affected by the exposure bias if they had: a) exposed thermometers in any form of historic exposure *and* introduced the Stevenson screen at any point during their record and b) had not been previously adjusted for the exposure bias.

3.5.2. Model application

The models were applied to individual CRUTEM5_ext stations using model predictors obtained from: WFDE5 (*SWD*) and 'climlab' (*TOA*), as described in Section 2.4, and the station's own actual temperature record (T_a). Where missing data prevented the calculation of T_a for a given year, the missing months were infilled using a climatology of the neighbouring ±15 years (minimum required years n≥10) and an estimate of T_a used instead to maximise the number of years the model could be applied to.

The models were only applied to stations within 30° to 60° latitude, for the reasons outlined in Section 2.4, and to stations which were not known to have been adjusted for the exposure bias previously. Although Section 2.4 also acknowledges that the method used to calculate daily-mean temperatures influences the bias, the model application did not discriminate based on this due to insufficient metadata. This is a noted limitation of this approach and is discussed in Chapter 5.

As the bias-estimation models were developed using the relatively few parallel measurement series available, a preliminary application of the models was conducted to identify whether applying the models outside of their calibrated ranges yields implausible results (i.e. overextrapolation). The predictors and estimated biases for the stations in CRUTEM5_sdb were compared with the observed biases and with the predictors used as input for the models. The comparison found no evidence of overextrapolation for wall-mounted and closed exposures, but some evidence for open exposures. For open exposures, the range of CRUTEM5_sdb predictors extended beyond the range of predictors used to develop the bias-estimation model, and, at the extremes, the estimated biases did not remain within the

observed range. As we cannot be certain the relationship between the magnitude of the bias and the predictors continues linearly outside of the observed range, Equation 4 was constrained where T_a < 4.84°C and Equation 5 was constrained where T_a < 6.29°C, to prevent winter and summer bias-estimates exceeding 0.6°C, and 0.0°C, respectively, which are not generally observed. Note that these constraints affected only 36 months of data for one station in CRUTEM5_sdb and the constrained bias estimates always fell within the uncertainty range of the unconstrained estimates.

The revised models were then reapplied to produce metadata-based estimates of the monthly mean exposure bias (with 95% confidence intervals) at individual stations within CRUTEM5_sdb. The confidence intervals were produced to inform the uncertainty associated with the new GloSAT dataset (introduced in Chapter 1). For the open and closed bias-estimation models, they were calculated based on equation 7.23 in Wilks (2019) and for the wall-mounted model they were calculated based on the standard deviation of the residuals. Note the magnitude of the confidence intervals varies by station, month and with the model applied, but an approximation of the uncertainty associated with each model can be found in Table 3.4 and in the figures in Chapter 2 (Section 2.4). The monthly bias estimates for each station were then used to adjust the monthly mean values in CRUTEM5_ext to produce an exposure bias adjusted version of the data (CRUTEM5_eba). As per World Meteorological Organization (2003, 2020) guidelines, the adjustments were applied to the observations made in the historic exposures, so the early observations 'match' the most recent homogenous section (the Stevenson screen) (see Figure 3.12). This is to allow series to be updated by appending new measurements without further adjustment.

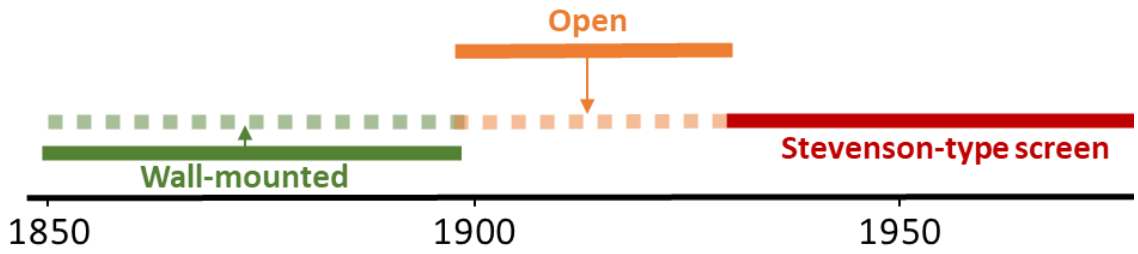


Figure 3.12. Illustration of the application of the exposure bias adjustments to bring the observations made in the historic exposures in-line with those made in the Stevenson screen.

Table 3.4. Uncertainties associated with each of the applied bias-estimation models.

	Coefficient(s)		Constant	Coefficient Standard Error		Constant Standard Error	Covariance			Residual standard deviation
	1	2		1	2		Coefficient 1 & Constant	Coefficients 1 & 2	Coefficient 2 & Constant	
Open: $min\Delta T_m$	-0.062		0.390	0.015834		0.243433	-0.00373			0.153278
Open: $max\Delta T_m$	-0.058		0.881	0.018989		0.291937	-0.00536			0.324863
Wall-mounted	0.005	-8.38×10^{-6}	-0.469	0.000749	1.35×10^{-6}	0.084357	-5.9×10^{-5}	-9.93×10^{-10}	9.79×10^{-8}	0.223306
Closed	-0.002		-0.043	0.000387		0.054691	-1.8×10^{-5}			0.191068

3.5.3. An exposure bias adjustment for CRUTEM5_ext

The metadata identified 2,519 mid-latitude stations in CRUTEM5_sdb with probable biases resulting from the transition to Stevenson screens. Of those, bias-estimates were produced for 1,960 stations (524,894 months) leading to the partial adjustment of 82 stations and the complete adjustment of 1,878 in CRUTEM5_ext (Table 3.5). Unfortunately, not all stations or months could be adjusted - estimates could not be obtained where incomplete metadata prevented the identification of the appropriate bias-estimation model, or the presence of an exposure bias; where predictors could not be obtained, or where intermediate or miscellaneous exposures were identified (Table 3.5; Figure 3.13). Despite this, the metadata gathered, and the adjustments applied, indicate that the bias has now been minimized at 75.1% of mid-latitude stations, representing 86.3% of the mid-latitude data in terms of monthly values. This is compared with just 37.7% of mid-latitude stations in CRUTEM5_ext which were not affected by the exposure bias, either because no transition to a Stevenson screen occurred during the record or because (for 1.5%) the record had already been adjusted.

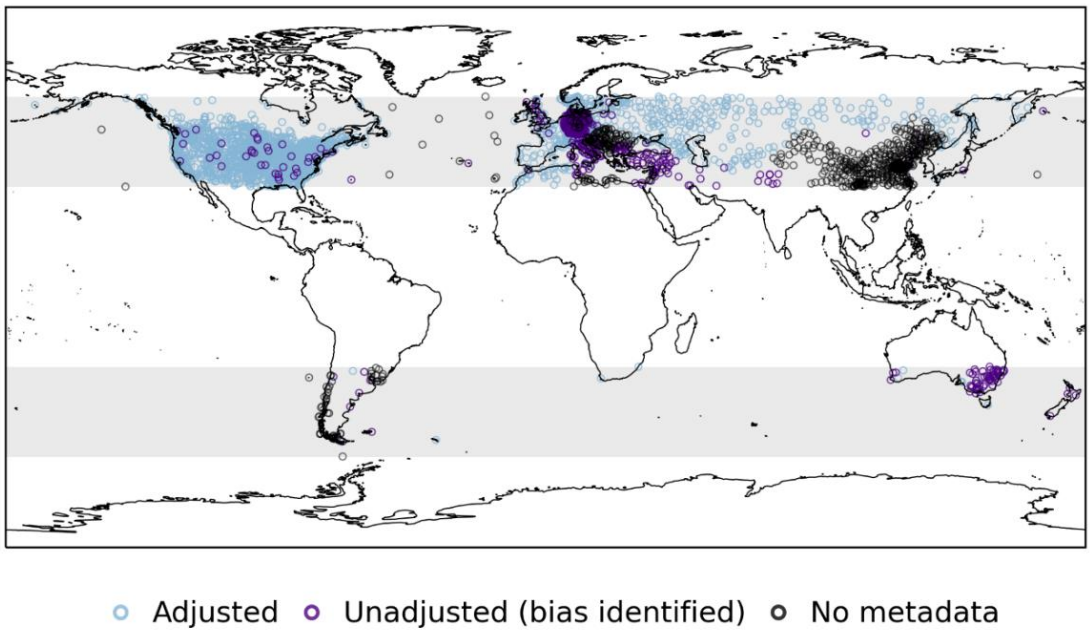


Figure 3.13. Location of stations which have been adjusted for the exposure bias, have not been adjusted but contain probable biases (for the reasons given in Table 3.5), or do not have metadata. Stations with metadata, but which were not adjusted, are not shown.

At the hemispheric scale the impact of the bias adjustments is relatively small (Figure 3.14). In the Northern Hemisphere (NH) the mid-latitudinal annual mean is ≤0.016°C warmer before 1870 in CRUTEM5_eba and up to 0.1°C cooler between 1870 and 1934. In the former period, the seasonal adjustments are all of a similar magnitude, but differ in sign between

spring/autumn (≈0.05°C) and summer/winter (≈-0.03°); whereas in the latter period the seasonal adjustments are all negative, with the largest adjustments in summer (-0.2°C). The small adjustments – and their unique seasonal structure – before 1870, are due to the predominance of wall-mounted exposures, which introduce biases with an annual mean close to 0°C and a bi-annual seasonal cycle (Section 2.3.2). The change in the direction and seasonal structure of the bias after 1870 (and the increase in magnitude) occurs as a result of more series requiring correction for freestanding exposures, which produce larger biases with a single summer peak in the seasonal cycle (Sections 2.3.1 and 2.3.4). The largest adjustments, however, are relatively geographically constrained (Figure 3.15) meaning the overall effect on the mid-latitude mean remains small.

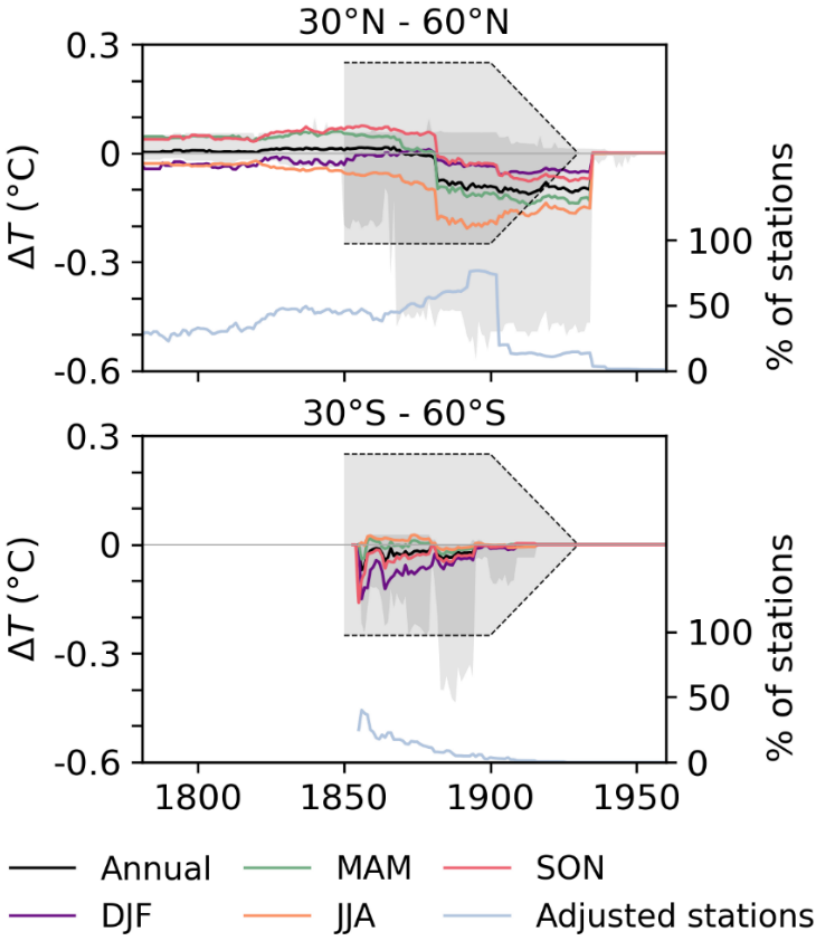


Figure 3.14. Difference between CRUTEM5_ext and CRUTEM5_eba annual and seasonal means (ΔT = CRUTEM5_eba – CRUTEM5_ext) for the NH and SH mid-latitudes (coloured lines). The light blue line shows the number of stations with a non-zero bias estimate (adjustment). The dark grey shading shows the range of annual mean exposure bias adjustments present in each CRUTEM5_eba grid cell. The light grey shading with dashed outline shows the approximate range of 100 realisations of the exposure bias component in the HadCRUT5 error model for comparison.

Table 3.5. Number of stations and months which have been adjusted for the exposure bias, as well as the numbers which still require adjustment.

<table>
<tr><th colspan="2">Input:
CRUTEM5_ext</th><th colspan="3"></th><th>Output: CRUTEM5_eba</th></tr>
<tr><td rowspan="5">All stations
Mid-latitude stations in CRUTEM5_sdb (located 30° – 60° North or South of the equator) with data prior to 1961

Stations: 5,031
Months: 5,708,463</td><td rowspan="2">Exposure bias not present

Stations: 1,898 (37.7%)</td><td colspan="2">Series flagged as already exposure bias adjusted

Stations: 75 (1.5%); Months: 159,523 (2.8%)</td><td rowspan="2">Exposure bias adjusted

Stations: 3,776 (75.1% of stations)
Months: 4,923,593 (86.3% of data)</td></tr>
<tr><td colspan="2">Metadata suggest there is no bias present
- Stevenson screen in place from the start of the station record
- Stevenson screen never introduced

Stations: 1,823 (36.2%); Months: 1,489,136 (26.1%)</td></tr>
<tr><td rowspan="3">Exposure bias present
Metadata suggest a Stevenson screen was introduced part-way through a station record

Stations: 2,519 (50.1%)</td><td colspan="3">Adjusted: Stations which have been adjusted for the exposure bias

Stations: 1,878 (37.3%);
Months: 2,836,051 (49.7%) (Adjusted: 486,148; No adjustment required: 2,349,903)</td></tr>
<tr><td>Partially: Stations which have been partially adjusted for the exposure bias

Stations: 82 (1.6%);
Months: 140,962 (2.5%) (Adjusted: 38,746; No adjustment required: 74,033; Unadjusted: 28,183)</td><td rowspan="2">Reason for missing bias estimates/adjustments:

Missing exposure metadata
Stations: 1,209 (24%) *
Months: 769,453 (13.5%)

No bias-estimation model
Stations: 44 (0.9%) *
Months: 13,389 (0.2%)

Missing predictor(s)
Stations: 10 (0.2%) *
Months: 2,028 (0.04%)</td><td rowspan="2">May require further exposure bias adjustment

Stations: 1,255 (24.9% of stations)
Months: 784,870 (13.7% of data)</td></tr>
<tr><td>Unadjusted: Stations which have not been adjusted for the exposure bias

Stations: 1,173 (23.3%);
Months: 1,082,791 (19%) (No adjustment required: 326,104; Unadjusted: 756,687)</td></tr>
<tr><td>No exposure metadata
Series without any metadata in the database

Stations: 614 (12.2%)</td><td colspan="3"></td></tr>
</table>

*Stations may be counted more than once; for example, if a station has a period with missing metadata and a period with an intermediate exposure (no model).

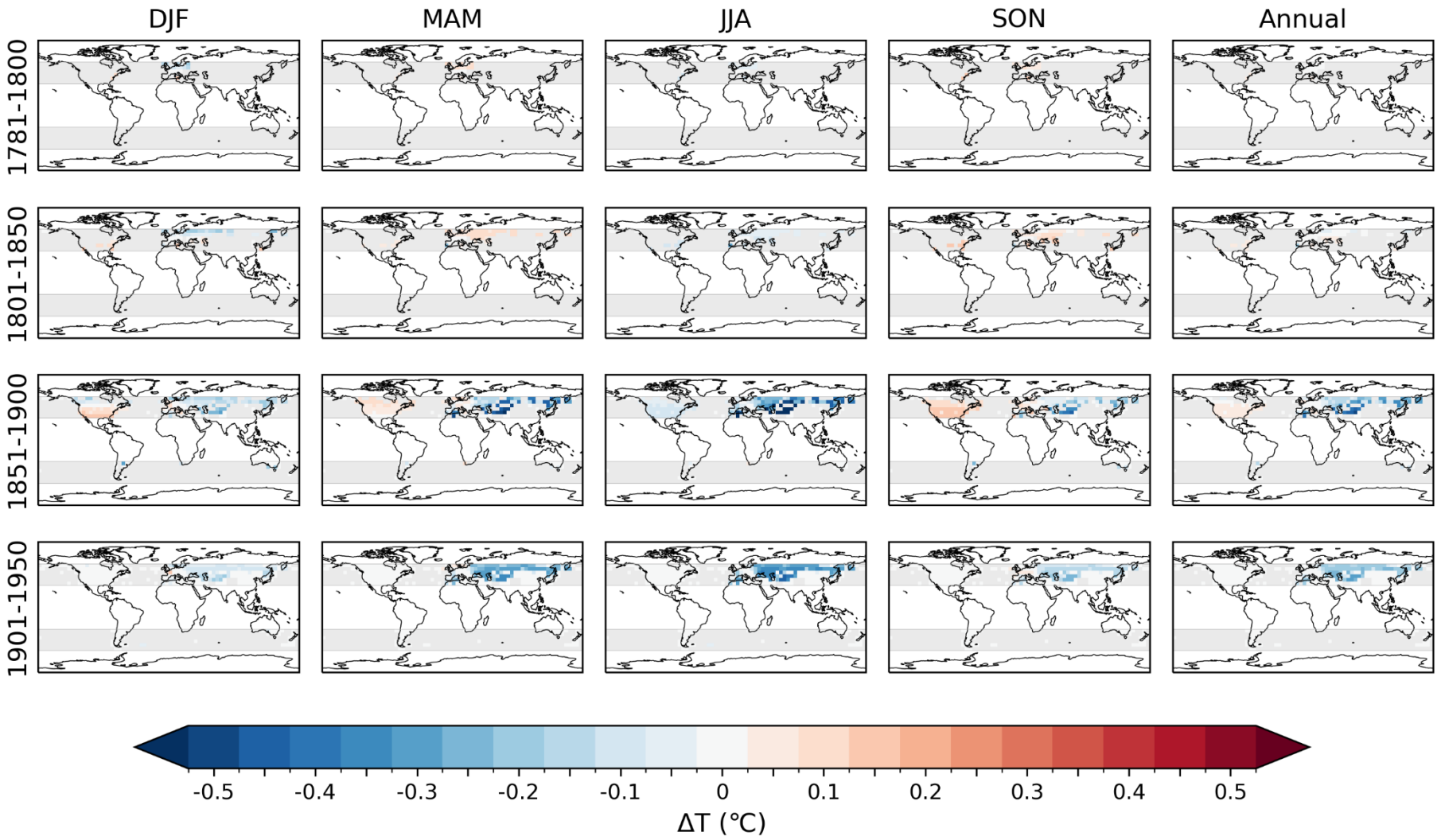


Figure 3.15. Difference between CRUTEM5_ext and CRUTEM5_eba seasonal and annual means over time (ΔT = CRUTEM5_eba – CRUTEM5_ext).

In the Southern Hemisphere (SH) the adjustments are always negative, with a similar seasonal structure to the latter period in the NH, but smaller and more temporally constrained, peaking at -0.07°C in 1856 and decreasing approximately linearly to ≈0°C by 1900. Here the majority of adjustments are also for freestanding exposures, explaining the similar seasonal structure, however the smaller magnitude of the bias is because fewer adjustments were made in the SH: only 11 stations (2,803 months) compared with 1,949 stations (522,091 month) in the NH. This is partly because there is less land, and thus fewer stations, in the SH mid-latitudes, but also because station records often started later, when Stevenson screens were already in place. Not all earlier stations were adjusted either: many lacked accessible metadata, exposed thermometers in intermediate exposures which could not be adjusted, or had been adjusted previously (Figure 3.13). Many also introduced Stevenson screens comparatively early, meaning fewer adjustments were required.

The magnitude of the bias adjustments on a regional basis is much larger than for the mid-latitudinal means and exhibits significant spatiotemporal variability (Figure 3.15). Between 1851-1900, for example, large negative adjustments (up to -0.79°C in summer, and -0.57°C annually) are present in Mediterranean Africa and central Asia, whereas small positive adjustments are present in North America. This variability arises from the spatio-temporal heterogeneity of the historic exposures in use (Video 3.1), as well as the influence of solar radiation/temperature. The large negative adjustments in Mediterranean Africa and central Asia, for example, reflect the use of freestanding exposures in those regions combined with the influence of comparatively strong solar radiation and/or hot temperatures. In contrast, the small positive adjustments in North America reflect the continued use of wall-mounted exposures in the United States and Canada until the 1890s and 1900s, respectively, when many other nations had introduced freestanding exposures or Stevenson screens.

A comparison between the bias adjustments produced here and the representation of the bias in HadCRUT5 (Figure 3.14) shows reasonable agreement annually over large spatial scales, but regionally the comparisons reinforce the limitations of the current representation (in HadCRUT5) identified in Section 2.1. The assumption of a fixed annual bias in HadCRUT5, with no spatio-temporal variation (outside of the two specified latitudinal bands), fails to capture the pronounced seasonal nature of the bias, differences in the magnitude and seasonal structure between (and within) exposure classes, or the spatio-temporal differences in the use of historic exposures and timing of the introduction of the Stevenson screen. These limitations are highlighted by the large discrepancies between the HadCRUT5 realisations

and the grid cell adjustments (particularly between 1868-1934 in the NH and the late-1800s in the SH) and by the static nature of the realisations in comparison to the time-varying adjustments produced here. Note, however, that the simple exposure bias error model used in HadCRUT5 was only designed to capture the large-scale influence of the bias rather than the local variations.

Opportunities for comparison with previous exposure bias adjustments are limited. Many previous studies simultaneously incorporated corrections for other inhomogeneities (e.g. Ashcroft et al. (2012)), did not adjust monthly T_m (e.g. Ashcroft et al. (2022) and Auchmann & Brönnimann (2012)) and/or are based on the parallel measurements used here (e.g. Brunet et al. (2006, 2011)). Comparisons with the few independent assessments available produce mixed results. The estimates produced here of a -0.44°C summer bias in Uppsala (which was not applied because this series in CRUTEM5 had already been adjusted), between 1858–1864, and ≈-0.2°C summer bias in the NH mid-latitudes, between 1880–1900, are of a similar magnitude to assessments by Moberg et al. (2003) and Frank et al. (2007) who estimated biases of -0.5°C to -0.8°C in Uppsala and approximately -0.3°C in the NH (30–90°N), respectively. However, Frank et al. (2007) estimated increasingly large NH biases earlier in the record and Moberg et al. (2003) consistently large biases in Uppsala and Stockholm, Sweden, before 1858, when the estimates here (again, not applied because the series has already been adjusted) suggest smaller biases due to the use of wall-mounted exposures.

Although the cited assessments are themselves uncertain, the mixed results do highlight some potential limitations of this approach. The biases estimated here represent the average bias for a specified exposure and location, based on the identified relationship(s) between the bias and up to two predictor variables. As such, the bias-estimation models cannot take into account station-specific factors or differences within categories of exposure (although these are partially captured by the confidence intervals, Table 3.4). The accuracy of the estimates is also dependent on the accuracy of the exposure metadata collated, which in many cases, is nation-, rather than station-, specific (Table 3.3). This approach, however, is designed to give an estimate of the exposure bias in global temperature compilations in the absence of station-level homogenisation; it is not designed to replace detailed, station-specific, homogenisation, which is always preferable. Overall, therefore, confidence can be taken from the assessed performance of the bias-estimation models (Section 2.4) and the favourable comparison of the regional results produced here with those of Parker (1994) who

anticipated annual mean biases: close to 0°C in Canada, central and Eastern Europe, Russia, the UK and the United States until the late-nineteenth century, between 0 and -0.2°C in France and Australia in the same period, and between -0.1°C and -0.2°C in Russia between 1870–1910, in line with the estimations here.

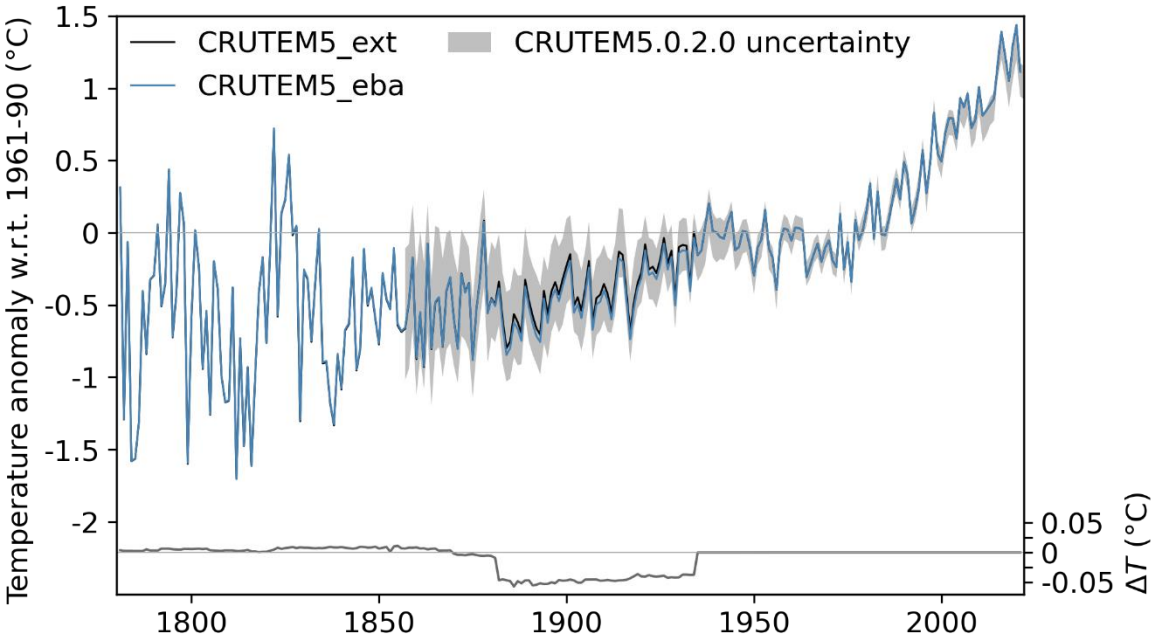


Figure 3.16. Comparison between the CRUTEM5_eba (blue line) and CRUTEM5_ext (black line) global annual mean temperature series (left-hand axis). The grey shading shows the uncertainty associated with the CRUTEM5.0.2.0. global annual mean, for information (left-hand axis), and the grey line shows the magnitude of the difference between the two series (ΔT = CRUTEM5_eba – CRUTEM5_ext) on the right-hand axis.

Finally, for completeness, the impact of the applied exposure bias adjustments on the CRUTEM5 global mean timeseries is shown in Figure 3.16. This figure shows that, on a global and annual scale, the impact of the adjustments is small and, reassuringly, always within the range of uncertainty estimated for the CRUTEM5.0.2.0 dataset (where available, Osborn et al. (2021)). For the majority of the two datasets' duration, the difference between the timeseries is less than ±0.01°C, with larger differences present only between 1882 and 1934, when the CRUTEM5_eba annual mean is 0.03°C to 0.06°C cooler than CRUTEM5_ext due to the increased number of adjustments made for freestanding exposures during that period. Note, however, that the exposure bias adjustments applied in Chapter 3 were only applied to mid-latitude weather stations, could not be applied to all stations, and only address the (average) bias arising from three forms of (well-maintained) early thermometer exposure. As a result, the 'true' impact of the exposure bias on the

CRUTEM5_ext global annual mean is likely larger than shown in Figure 3.16, particularly in the early part of the record (see Chapter 5 for further discussion).

3.6. Chapter summary

This chapter has attempted to address the presence of the exposure bias in an extended version of CRUTEM5 through the application of three exposure-specific bias-estimation models using a new database of historic exposures. The new database represents one of the first attempts to collate thermometer exposure information at scale and this is believed to be the first time metadata and statistical models have been combined to produce a (near-) global assessment of the exposure bias in global temperature compilations.

The approach represents an improvement on the current representation of the exposure bias in HadCRUT5 (and other global temperature compilations) as it accounts for the seasonal nature of the exposure bias, the exposure history of individual stations/regions, and geographic differences in the magnitude of the exposure bias. The approach is not without limitations; however, it has been shown to provide skilful estimates of the exposure bias, consistent with other global analyses. It is also hoped the approach can be built upon in future via the inclusion of additional parallel measurement series and exposure metadata.

Part II

Exploring natural climate variability in early instrumental temperature records

Chapter 4

Do the early nineteenth century eruptions provide evidence for volcanically induced Eurasian winter warming?

4.1. Introduction

Large volcanic eruptions are key drivers of natural climate variability and have been the dominant cause of natural, externally forced, climate change over the instrumental period (Crowley, 2000; Hegerl et al., 2019; Schurer et al., 2014). Explosive volcanic eruptions influence the climate primarily through the ejection of sulphur gases into the stratosphere. Once there, the sulphur gases (primarily sulphur dioxide) oxidise and condense to form sulphate aerosols which are then rapidly transported globally (Bluth et al., 1992). Sulphate aerosol particles interact strongly with solar radiation, scattering it, and leading to a net decrease in the amount of solar radiation reaching the Earth's surface (Robock, 2000). This aerosol-induced negative radiative forcing is the dominant global climatic effect following an eruption, however, analysis of temperature observations suggests that, for tropical eruptions, it is not the dominant temperature response for all regions and seasons.

In the Northern Hemisphere (NH), temperature observations consistently show anomalous winter warming over the Eurasian continent following explosive eruptions. Groisman (1985, 1992) was one of the first to draw attention to this – highlighting the similarities between the unusually warm winter conditions across central Russia in the winter of 1991/92, following the eruption of Pinatubo, and the warming of up to 4°C that had been observed across Europe and western Russia in the two winters following the eruption of El Chichon in 1982. A link between the observed winter warming and explosive volcanism was proposed and subsequently supported by Robock & Mao (1992, 1995) and Kelly et al. (1996) who identified similar Eurasian winter warming (composite temperature anomalies of up to 2°C)

following the largest eruptions of the instrumental period (Askja, 1875 – Pinatubo, 1991). Shindell et al. (2004) and Fischer et al. (2007) later extended these studies, finding additional evidence of Northern Hemisphere continental winter warming following the largest eruptions of the last half millennium (Kelut, 1586 – Pinatubo, 1991) using a combination of observations and proxy temperature reconstructions.

Early modelling studies also replicated the observed winter warming (e.g. Graf et al. (1993); Kirchner et al. (1999); Kirchner & Graf (1995); Shindell et al. (2001, 2004); Stenchikov et al. (2002, 2004)) and helped lead to the identification of a possible mechanism to explain its link to volcanism (illustrated in Figure 4.1) (Graf et al., 1993; Kodera, 1994; Perlwitz & Graf, 1995). In addition to scattering solar radiation, sulphate aerosols absorb outgoing longwave terrestrial and incoming near-infrared radiation, leading to heating of the lower stratosphere. Where aerosol is present in the tropical stratosphere (for example, following low latitude explosive eruptions), the stratospheric heating becomes concentrated in that region, due to the unequal meridional distribution of solar and infrared radiation (particularly during the winter half-year). This concentrated tropical warming leads to an enhanced equator-to-pole stratospheric temperature gradient, which strengthens the zonal winds and, in turn, leads to a strengthened stratospheric polar vortex. A strong polar vortex has long been associated with a positive phase of the North Atlantic Oscillation (NAO) (e.g. Perlwitz & Graf (1995)), which is indicated by a stronger than average pressure gradient between the subtropical (Azores) high and the subpolar (Icelandic) low. This pressure gradient strengthens the northerly winds over Greenland and northeast Canada, leading to cold conditions in those regions, and, crucially, enhances the westerly winds across the North Atlantic, leading to warm conditions over northern Europe, as has been observed following volcanic eruptions (Hurrell et al., 2003). This proposed mechanism – which will be referred to as the ‘stratospheric pathway’ in line with recent studies – is supported by observations of lower stratospheric heating following the eruption of El Chichon in 1982 (Parker & Brownscombe, 1983) and Pinatubo in 1991 (Labitzke & McCormick, 1992) as well as studies which find a positive NAO index is favoured in the winters following explosive eruptions (Christiansen, 2008).

The existence of a plausible mechanism to explain the observed winter warming, plus the agreement between the observations, proxy reconstructions and early modelling studies, led to a general consensus in the literature that the observed winter warming over Eurasia was a forced response to explosive volcanism. A key review paper by Robock (2000), for example,

included the winter warming (and proposed 'stratospheric mechanism') as a largely established climatic response to explosive volcanism, as did the IPCC Fourth Assessment Report (Forster et al., 2007). More recent studies, however, have questioned this conclusion, as well as the proposed stratospheric pathway, based on findings that the newer generations of climate models do not reproduce the observed winter warming over Eurasia (Driscoll et al., 2012; Marshall et al., 2009; Polvani et al., 2019; Stenchikov et al., 2006; Thomas, Timmreck, et al., 2009). This is despite significant advancements in the vertical resolution of the models and their treatment of volcanic aerosols (in comparison to the early modelling studies), and despite the fact some models were found to simulate the lower stratospheric warming and (to varying degrees) the strengthening of the polar vortex (which are central to the stratospheric pathway) (Bittner et al., 2016; Driscoll et al., 2012; Stenchikov et al., 2006; Thomas, Timmreck, et al., 2009). These findings, along with criticism of the methods used in some of the early observational studies (see Polvani & Camargo (2020) for details), have resulted in renewed discussion in the literature regarding whether the observed Eurasian winter warming is volcanically forced, or merely the result of internal variability.

Two recent studies have looked at this issue in detail. Polvani et al. (2019) and Polvani & Camargo (2020) examined the atmospheric response to the eruptions of Pinatubo (1991) and Krakatau (1883), respectively, and concluded there was little evidence to support the Eurasian winter warming being volcanically forced. Their conclusions were based on the findings that a) the observed warming was not 'exceptional', b) there was a wide spread between the model and reanalysis ensemble members, which incorporated both warmer and cooler conditions over Eurasia, despite identical volcanic forcing and c) the forced (ensemble mean) winter temperature anomaly in the models was not significantly different from zero. This, they argued, indicated that neither eruption had a significant effect on the observed winter warming and that any forced response was tiny in comparison to the internal variability present in the Northern Hemisphere winter climate system. Each study also questioned the veracity of the proposed stratospheric pathway, arguing that the polar vortex was not anomalously strong following Pinatubo or Krakatau, that the simulated strengthening is insufficient to account for the observed winter warming, and that Eurasian surface temperatures are not strongly correlated with the strength of the polar vortex.

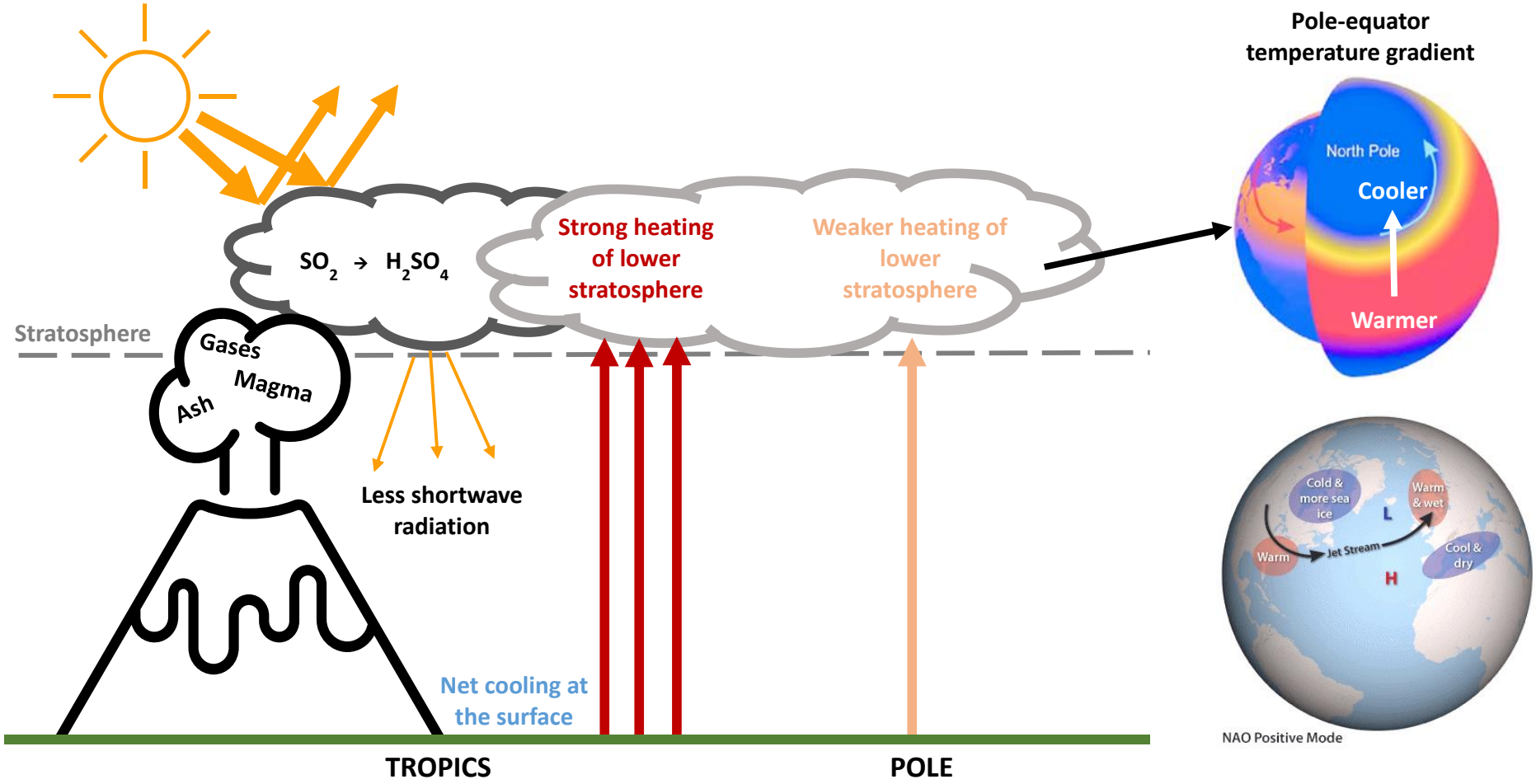


Figure 4.1. Schematic showing selected impacts of volcanic aerosols on the atmosphere; specifically, how they lead to the formation of a stratospheric temperature gradient which is believed to strengthen the polar vortex and induce a positive phase of the NAO. Elements of this schematic are from: Brönnimann (2015) and NOAA Climate (n.d.).

Significantly, the latter of the two studies Polvani & Camargo (2020) concluded: "*the evidence is overwhelming: low-latitude eruptions as large as Pinatubo or Krakatau are unable to cause a forced surface temperature anomaly over Eurasia that can be distinguished from unforced variability*" (p.13697) and placed the onus on future studies to present evidence that suggests otherwise. Much of the evidence cited in support of this conclusion, however, is based on the results of climate models and thus makes the assumption that the present generation of models are able to accurately simulate past eruptions, including the proposed stratospheric pathway. However, there are multiple reasons why this may not be the case (see Timmreck (2012) and Mann et al. (2015) for an overview), and not all modelling studies agree that there is no volcanically forced response. Zambri & Robock (2016), for example, find evidence of forced Eurasian winter warming following the eruptions of Krakatau and Pinatubo in CMIP5 models; Coupe & Robock (2021) find a similar response in the Community Atmospheric Model 5 when observed sea surface temperatures are specified and Azoulay et al. (2021) find evidence for volcanically forced winter warming in northern Eurasia following eruptions with injections of more than 10Tg(S), as well as evidence of a positive correlation between the strength of the stratospheric polar vortex and Northern Eurasian surface temperatures.

Due to the temporal infrequency of explosive volcanic eruptions, and the limited availability of early observational data, to date there has been little choice but to draw conclusions based on the results of models, proxy reconstructions or on the observed responses to a relatively small number of eruptions, including smaller magnitude eruptions such as Mt. Agung (1963). To my knowledge, only Groisman (1985) has examined the observed winter temperature response to large eruptions earlier than Askja in 1875 and that study was based on a relatively small number of individual weather station timeseries. The recent development of a gridded observed global surface air temperature dataset, starting in 1781 (GloSAT; Morice et al. [in prep]) provides a new opportunity to conduct a more in-depth study into the temperature response to the large early nineteenth century eruptions, including Tambora which is the largest eruption to have occurred since 1257 (Raible et al., 2016). The following chapter therefore presents an observation-based assessment of the winter temperature response to the eight largest low latitude eruptions of the nineteenth and twentieth centuries with the aim of determining whether the early nineteenth century eruptions support the evidence for volcanically induced winter warming. To ensure a comprehensive assessment, and because there are inherent uncertainties associated with the

observational data (particularly in the early nineteenth century), this assessment is complemented by three additional surface temperature datasets: two reanalysis products and an Earth System model, as well as an assessment of the winter atmospheric circulation (NAO) response to volcanism.

Sections 4.2 and 4.3 will outline the datasets and methods used; Sections 4.4 and 4.5 will outline the results of the surface temperature and circulation assessments, respectively; Section 4.6 will synthesize and discuss the evidence for volcanic forcing and Section 4.7 will summarise and draw conclusions.

4.2. Data

4.2.1. GloSAT Global Surface Air Temperature

The Global Surface Air Temperature dataset (referred to here as GloSAT) is an observed two metre monthly mean air temperature compilation which spans the period 1781 to 2021 and which incorporates the exposure bias adjustments applied to the LSAT data in Part I of this thesis. It is the first global temperature compilation to use air temperature over both the land and ocean (rather than sea surface temperature, SST) and this allows the dataset to start in the late-eighteenth century (as many more marine air temperature measurements were made before 1850 than SST measurements).

As is outlined in Section 4.1, this extension of the observed surface air temperature record enables the assessment of additional large eruptions, which occurred before 1850 when the earliest global (land and ocean) instrumental temperature datasets currently start. It is a gridded dataset of monthly temperature anomalies on a 5° latitude/longitude grid. Importantly, the dataset also contains 200 ensemble members which sample the uncertainty associated with the observations, allowing the spread of potential winter temperature responses to be assessed and quantified. The ensemble members represent the uncertainty arising from known systematic biases in the observations (and their adjustment), uncertainty in the calculation of the climatological normals, and uncorrelated measurement and sampling uncertainty. The ensemble does not account for structural uncertainties (e.g. in the choice of statistical model to produce a gridded dataset from point observations). A full description of the GloSAT data will be included in (Morice et al. [in prep]) but the spatial analysis methods closely follow those of the HadCRUT5 temperature dataset (C. Morice, pers. comm.; Morice et al., 2021), meaning estimates are made for all grid cells where the available observations

in the neighbourhood are informative, even if there are no observations within the grid cell itself.

4.2.2. NOAA-CIRES-DOE Twentieth Century Reanalysis

The NOAA-CIRES-DOE Twentieth Century Reanalysis version 3 (20CRv3) (Slivinski et al., 2020) is a globally complete reanalysis dataset which spans the period 1835 to 2015, with an experimental back extension to 1806. The dataset is produced by assimilating surface pressure observations into an underlying atmosphere-land model, which is constrained by prescribed sea ice concentrations and SST observations. The model is forced with time-varying solar radiation, atmospheric carbon dioxide, stratospheric ozone concentration and volcanic aerosol; the latter of which is prescribed according to Crowley & Unterman (2013) and produces 3-hourly surface and atmospheric fields.

The dataset was chosen for use here as it is one of the few available reanalyses with monthly coverage over a similar timespan as the GloSAT observations and, more significantly, because the dataset is largely independent of the GloSAT observations (marine and land air temperatures are not assimilated into 20CRv3). It therefore provides a valuable independent check of the results of the observation-based analysis and the associated 80-member ensemble provides an additional measure of uncertainty. The 80-member ensemble samples the uncertainty associated with the assimilated observations (measurement and sampling errors) as well as the uncertainty associated with the underlying model (Compo et al., 2011; Slivinski et al., 2019). It should be noted that the ensemble spread may be an underestimate of the 'true' uncertainty due to a tendency for ensembles produced using an Ensemble Kalman Filter to "overtighten towards the mean"; however, this has partially been addressed in 20CRv3 via the use of an inflationary parameter (Compo et al., 2011; Slivinski et al., 2019, p. 2881). An earlier version of this dataset (version 2c) has been used by previous studies to assess the winter response to more recent explosive volcanic eruptions (e.g. Wunderlich & Mitchell (2017) and Polvani & Camargo (2020)); however, improvements made to version 3 (e.g. significant expansion of the number of early observations assimilated; see Slivinski et al. (2019) for details) and the longer temporal coverage mean it is valuable to revisit here.

4.2.3. Modern Era Reanalysis

The Modern Era Reanalysis (ModE-RA) (Valler et al., 2023, 2024) is a globally complete monthly palaeo-reanalysis which spans the period 1421 to 2008. Similar to 20CRv3, the

dataset is produced by assimilating historical climate information into an underlying atmospheric general circulation model, which is constrained with prescribed SSTs and sea ice concentrations. The model is forced according to PMIP4 past1000 forcings (R. Hand, pers. comm.; Jungclaus et al., 2017).

The dataset is not independent of GloSAT as it assimilates surface air temperature observations; however, it has been chosen for use here because, unlike 20CRv3, it also assimilates data from proxy and documentary sources, including climatic information relevant to the winter season. This additional information should help to constrain the temperature response to volcanic eruptions in time periods and/or regions where observations are limited (Reichen et al., 2022). The dataset also covers all of the major eruptions of the nineteenth and twentieth centuries and has a 20-member ensemble to help quantify uncertainty. Similarly to 20CRv3, the ModE-RA ensemble samples the uncertainty associated with the assimilated climatic information (measurement and sampling error for the instrumental observations and a residuals-based estimate of the error associated with the proxy and documentary reconstructions) as well as the uncertainty associated with the underlying model (Valler et al., 2024). As in 20CRv3 an Ensemble Kalman Filter has been used to produce the ensemble, meaning the spread may underestimate the true uncertainty.

4.2.4. United Kingdom Earth System Model version 1.1

A six-member ensemble of simulations, starting in 1750 and produced by the GloSAT project using the United Kingdom Earth System Model version 1.1, is also used to assess the winter temperature response to volcanic eruptions. The ensemble (referred to here as G_UKESM1.1) includes all forcings - natural and anthropogenic - with the volcanic component based on Toohey & Sigl (2017) before 1850 and CMIP6 historical forcings thereafter (A. Schurer, pers. comm.). Further details of the model simulations will be made available in Ballinger et al. [in prep].

The model simulations are included for analysis, despite previous studies suggesting the present generation of climate models do not always capture the dynamic winter warming response, for two main reasons. Firstly, inclusion of the model simulations will allow the spatially resolved observed and modelled responses to the early nineteenth century eruptions to be compared. And secondly, the earlier initialisation of the G_UKESM1.1 simulations in 1750 (compared to the CMIP models; (Eyring et al., 2016; Taylor et al., 2012)) may allow a more representative climate (and particularly oceanic) state to develop over the nineteenth

and twentieth centuries as the multi-decadal influence of the large early nineteenth century eruptions will be included in the model runs. This may provide a more direct comparison between the models and observations and could influence the responses seen following the later eruptions (Coupe & Robock, 2021; Zanchettin et al., 2012, 2013). The six ensemble members (which have the same forcing but different initial conditions) will also provide a useful indication of the internal variability present in the climate system and thus the relative importance of any potential forced response. Note that, unlike the previously described datasets, the UKESM1.1 ensemble represents different realisations of internal variability; it does not represent observational/model uncertainty.

4.3. Methods

4.3.1. Choice of volcanic eruptions

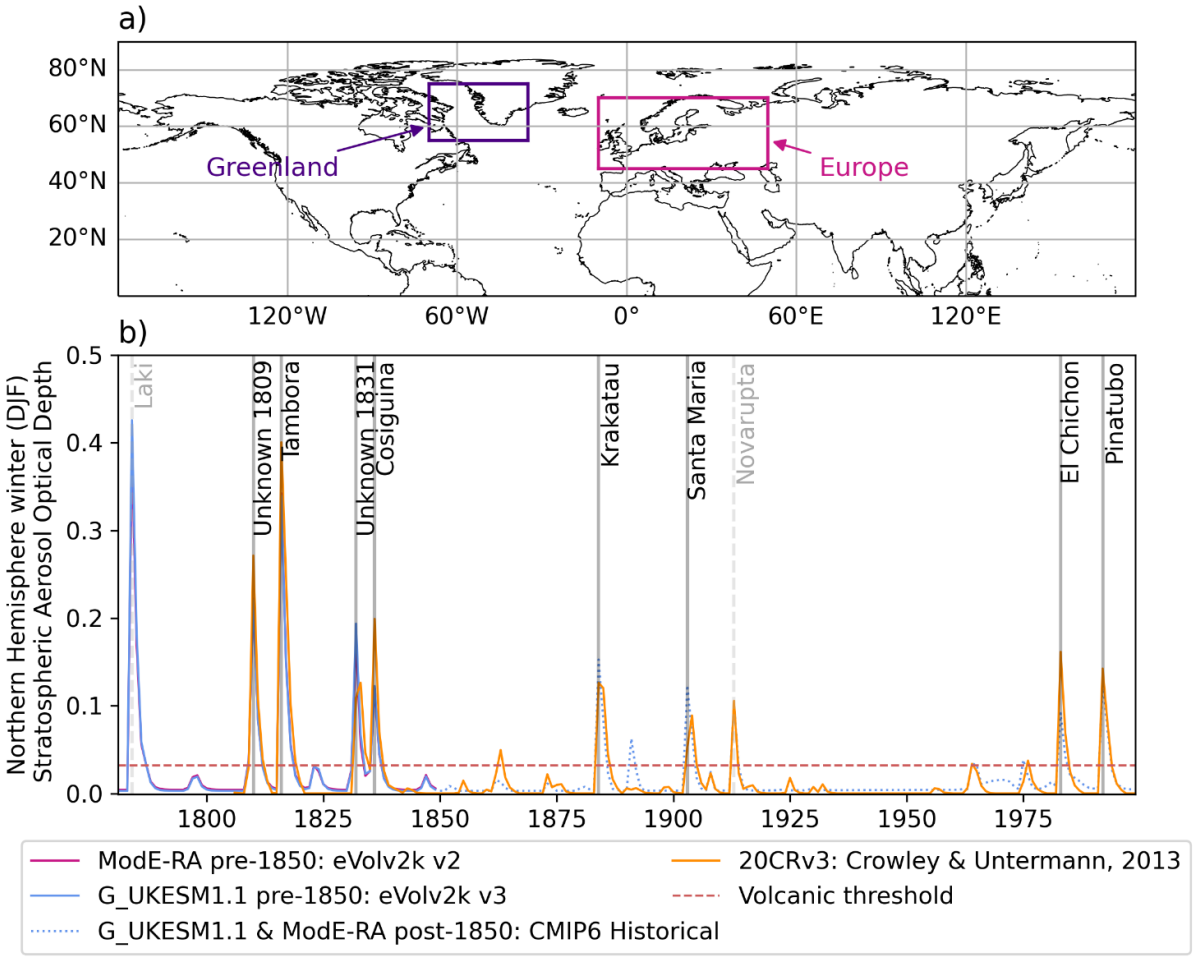


Figure 4.2. a) Location of the Greenland (55-75°N, 35-70°W) and Europe (45-70°N, 10°W-50°E) study regions and b) Northern Hemisphere (December-to-February) mean stratospheric aerosol optical depth associated with each temperature dataset. The eruptions detailed in black are the low latitude eruptions selected for analysis; the dashed red line shows the 0.032 'volcanic' threshold.

To reduce the noise present in the analysis, only explosive volcanic eruptions considered *a priori* to be of a magnitude large enough to impact the Northern Hemisphere winter climate were selected for analysis. As in similar studies, the Volcanic Explosivity Index (VEI) (Newhall & Self, 1986) was used as a guide, however, as the winter climate response is more closely related to the quantity and distribution of sulphate aerosols in the stratosphere than explosivity (Robock, 2000), the final selection of eruptions was based on an examination of the stratospheric aerosol optical depth (SAOD) - a unitless measure of the opacity of the stratosphere, measured at 500nm - in the years following an eruption.

The SAOD was examined in the datasets used to force 20CRv3, ModE-RA and G_UKESM1.1 (noted in Section 4.2 and shown in Figure 4.2b) as these datasets represent the current best understanding of the quantity and distribution of aerosol present in the stratosphere following an eruption. The criteria for selection included: a) the eruption had to be large in magnitude (generally with a VEI ≥5), b) the eruption had to occur during the period covered by the GloSAT observations (1781-2021), c) the eruption had to be located in the tropics (30°S to 30°N), and d) substantial quantities of sulphate aerosol had to be present in the Northern Hemisphere following the eruption. Only tropical eruptions were considered, contrary to previous studies, due to documented differences in the Northern Hemisphere winter temperature and circulation response to extratropical and tropical eruptions (Christiansen, 2008; Oman et al., 2005; Robock & Mao, 1995; Sjolte et al., 2021), including potential differences in the manifestation of each response (Graf & Timmreck, 2001; Shindell et al., 2004).

In total, eight eruptions met the criteria and were chosen for analysis: the unknown 1809 and 1831 eruptions (referred to going forward as U1809 and U1831, respectively), Tambora (1815), Cosiguina (1835), Krakatau (1883), Santa Maria (1902), El Chichon (1982) and Pinatubo (1991). The U1831 eruption was included for analysis despite uncertainty surrounding the location of the eruption (see (Garrison et al., 2021; Garrison et al., 2018)) and the (uncertain) VEI designation of '4?'. This decision was made as the SAOD forcing datasets locate the aerosol injection in the tropics and because the magnitude of the SAOD present in the NH is greater than following the larger VEI eruptions of Krakatau, Santa Maria, and Pinatubo. Similarly, although the location of U1809 is not known, it is also selected for analysis as the simultaneous presence of a similar quantity of sulphur in both Greenland and Antarctic ice cores strongly suggests the eruption was tropical (Cole-Dai et al., 2009).

Table 4.1. Volcanic eruptions selected for analysis.

Volcano	Date	Latitude	Longitude	VEI	DJF1	QBO during eruption	ENSO during eruption	NAO in DJF0
Unknown	1809	Equatorial injection*		6	1809/10		*La Niña*	*-0.89*
Tambora	Apr 1815	8.25°S	118.00°E	7	1815/16		*Neutral to La Niña*	*-0.61*
Unknown	1831	Equatorial injection*		4?	1831/32		*Neutral to La Niña*	-2.32
Cosiguina	Jan 1835	12.98°N	87.57°W	5	1835/36		*Neutral to La Niña*	0.51
Krakatau	Aug 1883	6.10°S	105.42°E	6	1883/84	*Easterly*	Neutral	0.64
Santa Maria	Oct 1902	14.76°N	91.55°W	6	1902/03	*Easterly*	El Niño	-1.23
El Chichon	Apr 1982	17.36°N	93.23°W	5	1982/83	Easterly	Neutral	-0.99
Pinatubo	Jun 1991	15.13°N	120.35°E	6	1991/92	Easterly	Neutral	-0.04

*Equatorial injection in the SAOD forcing datasets. Data sources: (National Geophysical Data Center / World Data Service (NGDC/WDS), n.d.); Quasi-Biennial Oscillation (QBO) at 30hPa: Krakatau: (Hamilton, 2012; Symons, 1888), Santa Maria: (Brönnimann et al., 2007); El Chichon and Pinatubo: (Freie Universitat Berlin, n.d.; Naujokat, 1986); El Niño Southern Oscillation (ENSO) pre-1850: (Dätwyler et al., 2019; Li et al., 2013; F. Zhu et al., 2020), post-1850: ERSSTv5 NINO3.4 (Huang et al., 2017; KNMI, 2023); North Atlantic Oscillation (NAO) in the winter at the start of the eruption year (DJF0) post-1824: (Jones et al., 1997), pre-1824: NAO index calculated from ModE-RA. Information *italicised* when based primarily on proxy reconstructions or documentary sources.

The eruptions of Galunggung (1822) and Mount Agung (1963), which each have a VEI of 5 and have been included in some previous studies, were not selected for analysis here due to the minimal SAOD present in the Northern Hemisphere following each eruption. In the latter case the majority of the aerosol was transported into the Southern Hemisphere, whereas in the former the SAOD datasets suggest only a minor eruption (not comparable with the other ≥5 VEI eruptions) with minimal SAOD present in either hemisphere (and no increase in SAOD present in the 20CRv3 forcing dataset (Crowley & Unterman, 2013)). Eruptions with minimal aerosol present in the NH would not be expected *a priori* to have a large direct influence on the NAO, therefore these eruptions are not included. Further details of each eruption selected for analysis can be found in Table 4.1; the SAOD profiles for each eruption can be found in Figure C1.

4.3.2. Study locations

The winter temperature response to volcanic eruptions was assessed across the Northern Hemisphere, but with a particular focus on two regions: one located over southwest Greenland: 55-75°N; 35-75°W, and one located over Europe: 45-70°N; 10°W-50°E (Figure 4.2a). These regions were chosen because of the high (and opposite) correlation between the surface temperatures in each and a positive phase of the NAO (Visbeck et al., 2001) and because observational data were available in each region following all eight selected eruptions. This allows the same metrics to be assessed following each eruption.

These regions differ from the one most recently used by Polvani et al. (2019) and Polvani & Camargo (2020) to study the winter temperature response to Krakatau and Pinatubo; however, GloSAT data is not available across the majority of the eastern portion of the box they use (40-70°N; 0-150°W) until Krakatau (1883). The decision to analyse two regions here - each of which responds differently to the NAO - also has advantages as it allows an additional dipole metric (Europe minus Greenland) to be analysed and enables an assessment of whether the post-eruption anomalies are consistent with a positive phase of the NAO.

4.3.3. Determining the post-eruption winter temperature response

A commonly employed method to determine the winter temperature response to volcanism is to calculate temperature anomalies based on the difference between a non-volcanic reference period and the winter(s) immediately following an eruption. Previous studies (Polvani et al., 2019; Polvani & Camargo, 2020; Zambri & Robock, 2016) have used a five-winter reference period, however this approach was not deemed appropriate for use here (despite the wish for the results to be directly comparable with earlier studies). The spatial and temporal incompleteness of the GloSAT observations in the early part of the record, combined with the temporal proximity of the early nineteenth century eruptions (Table 4.1; Figure 4.2b), meant it was not always possible to construct a five-winter pre-eruption reference period which had both sufficient data to calculate a representative mean and minimal volcanic influence. In the Greenland study region, for example, a period of more than ten winters would have been required to calculate a pre-eruption reference period (with five winters' worth of data and minimal volcanic influence) for Tambora. In addition, a simple average of the five winters pre-eruption was not considered sufficient to determine a robust non-volcanic reference period due to the large variability (exacerbated by the spatially

and temporally incomplete data in the EIP) present in the NH winter, especially in the relatively small Europe and Greenland study regions.

Instead, a reference period using *all* non-volcanic winters (December-to-February: DJF) between December 1781 and February 1999 (spline detrended to remove the effect of long-term warming) was used to calculate the post-eruption anomalies. This period was chosen to cover the eight selected eruptions (1809-1991), and to maximise the number of winters included, without extending too far into the recent period due to documented changes in the winter variability of the NAO, particularly since 2000 (Hanna et al., 2015, 2022). Note that for the 20CRv3 data, a slightly shorter non-volcanic reference period of December 1806 to February 1999 was used due to the shorter time span of the dataset.

Non-volcanic winters were identified using the NH winter (DJF) mean SAOD values in the forcing datasets (Section 4.2) with a value of 0.032 set as the threshold below which a winter was considered non-volcanic (see Figure 4.2b). This value was chosen to ensure at least two winters following large eruptions were excluded from the non-volcanic reference period (due to evidence in previous studies of climatic impacts for two winters following eruptions (e.g. Fischer et al. (2007)) but without excluding too many winters due to the non-zero background aerosol quantities present in the SAOD datasets (Toohey & Sigl, 2017). In total, 34 of the 218 winters in the period December 1781 to February 1999 had SAOD values of more than 0.032 (in either of the forcing datasets) and were therefore excluded from the non-volcanic reference period.

To ensure long-term trends did not bias the reference period, a quartic spline was fitted to the winter (DJF) mean temperature anomalies and subtracted to remove the trend prior to analysis. Before fitting the spline, the volcanically forced winters (NH DJF mean SAOD ≥0.032) were temporarily removed from the dataset to ensure the (temporally frequent) volcanic eruptions present in the reference period did not influence the spline fit (and thus lead to some of the volcanic signal being removed in addition to the long-term trend). Figure C2 highlights the necessity of this step. If the volcanically forced years are not removed, the long-term trend in Greenland (Europe) is pulled down (up) during volcanically active periods. The detrending process was conducted per grid cell and ensemble member, before the ensemble means were calculated. To ensure stationarity, the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) statistical test was applied following the detrending. Overall, fewer than 0.5% of grid cells (in any individual ensemble member) were flagged as non-stationary

(none within the two study regions) and, where they were flagged, the KPSS test is believed to have failed because of the larger variability in the early part of the record, rather than the presence of a long-term trend.

Post-eruption winter (DJF) temperature anomalies were then calculated for each eruption (using the detrended series) by subtracting the mean of the non-volcanic winter reference period from the temperature anomaly in the first post-eruption winter (DJF1; Table 4.1). Area weighted anomalies were then calculated for the Europe and Greenland study regions (Section 4.3.2) and a dipole metric calculated by subtracting the Greenland anomaly from the anomaly over Europe. Where this dipole metric is large and positive, it suggests a positive phase of the NAO may have occurred. Note that each of the post-eruption responses will be assessed individually, rather than creating a composite. This is to allow differences in the temperature responses to be represented and to avoid damping any potential volcanic signal due to averaging over multiple eruptions (all of which have different eruption seasons and baseline environmental conditions (Table 4.1) which may be important to the individual responses (Toohey et al., 2014)).

Note that, unlike some previous studies (e.g. Wunderlich & Mitchell (2017)), the El Niño Southern Oscillation (ENSO) signal was not explicitly removed from the data prior to analysis. This is because accurately defining the ENSO signal at the grid cell level is problematic and, more importantly, because previous studies (e.g. Robock & Mao (1995); Shindell et al. (2004); Wunderlich & Mitchell (2017)) have shown that the ENSO signal does not significantly affect the (short term) post-eruption winter temperature response over the Europe and Greenland study regions, which are the focus of this study, or the NH extratropics more generally.

4.3.4. Determining the post-eruption winter circulation response

To assess the winter circulation response to volcanism, the observed winter (DJF) NAO index from Jones et al. (1997) was examined, along with the NAO indices for 20CRv3, ModE-RA and G_UKESM1.1, which were calculated using their respective mean sea level pressures (MSLP). To ensure comparability between the datasets, the NAO indices for each were calculated according to the method used by Jones et al. (1997) using the MSLP series from the grid cells over Gibraltar (36°N, 5.5°W) and southwest Iceland (64.1°N, 22°W).

As the 1951-1980 reference period used by Jones et al. (1997) corresponds with an anomalous period of the NAO (and leads to a disproportionate number of years with positive

indices) the indices used here were re-expressed as anomalies from the 1901-1999 mean. This period was chosen to be as long as possible (and thus more representative), without including data from the earlier or later parts of the record, which are more uncertain and more variable, respectively (Hanna et al., 2015, 2022).

4.4. Northern Hemisphere winter temperature response

4.4.1. GloSAT observations

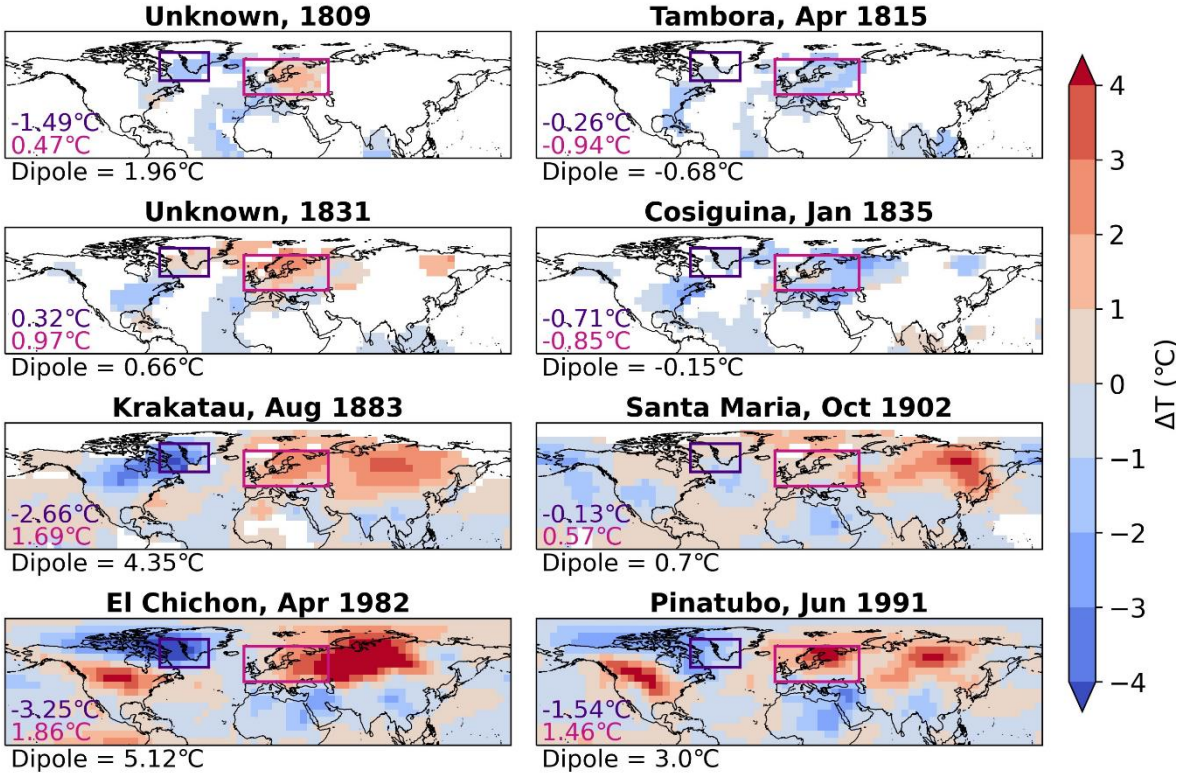


Figure 4.3. Post-eruption surface temperature anomalies in the first winter (DJF1) following each named eruption in the GloSAT observations ensemble mean. The number in purple (pink) is the area-weighted mean temperature anomaly for the Greenland (Europe) box and the dipole is the difference between the two (Europe minus Greenland). Temperature anomalies are calculated with respect to detrended non-volcanic winters between 1781/82 – 1998/99.

The spatially resolved ensemble mean surface temperature anomalies in the GloSAT observations, for the first winter (DJF1) following each eruption, are shown in Figure 4.3. Focusing first on the four most recent eruptions (Krakatau to Pinatubo), the results show a reasonably similar temperature response between eruptions, consistent with a positive phase of the NAO. Following each eruption there is evidence of winter warming over Eurasia (particularly over Scandinavia, the Baltic region and central Russia) and evidence of cooler conditions over northeast North America and the Labrador Sea. The area-weighted mean

temperature anomalies over Greenland and Europe are always negative and positive, respectively, however, some differences in the magnitudes of the anomalies are evident: Santa Maria, for example, has a much smaller regional dipole (0.7°C) than El Chichon (5.12°C). This pattern (and magnitude) of post-eruption warming and cooling is consistent with the findings of previous observation and proxy-based studies (e.g. Fischer et al. (2007); Groisman (1992, 1985); Kelly et al. (1996); Polvani & Camargo (2020); Robock & Mao (1992, 1995); Shindell et al. (2004) and Wunderlich & Mitchell (2017)) and confirms the comparability of the results found here, despite the different approach used to calculate the post-eruption temperature anomalies.

The consistency in the post-eruption (DJF1) temperature response following the four most recent eruptions is not mirrored by the earlier four. Only two of the early nineteenth century eruptions - U1809 and U1831 - present evidence of winter warming over Eurasia in the first post-eruption winter, and only U1809 also has evidence of cooler conditions over Greenland, as is indicative of a positive phase of the NAO. The U1831 eruption instead presents evidence of warmer conditions over Greenland (0.32°C). This is in contrast with the findings of Shindell et al. (2004) who identified cooling over Greenland in the cold season (October to March) immediately following U1831.

Following the remaining two eruptions - Tambora and Cosiguina - a different surface temperature response is seen (Figure 4.3). Both eruptions have negative area-weighted mean temperature anomalies over Europe and Greenland and neither presents compelling evidence of Eurasian winter warming. No warm anomalies are present across the NH following Tambora and following Cosiguina only small regions of warm anomalies are present in northern Germany and central Russia. This response is inconsistent with a positive phase of the NAO and, following Tambora, is in contrast with the findings of previous (proxy-based) studies, including Shindell et al. (2004), who identified warming over Eurasia and central North America in the 1815/16 cold season (October to March). The response following Cosiguina, however, is in keeping with the findings of Shindell et al. (2004) as well as Fischer et al. (2007) who both noted a reduced warming response following Cosiguina (in comparison to other explosive eruptions).

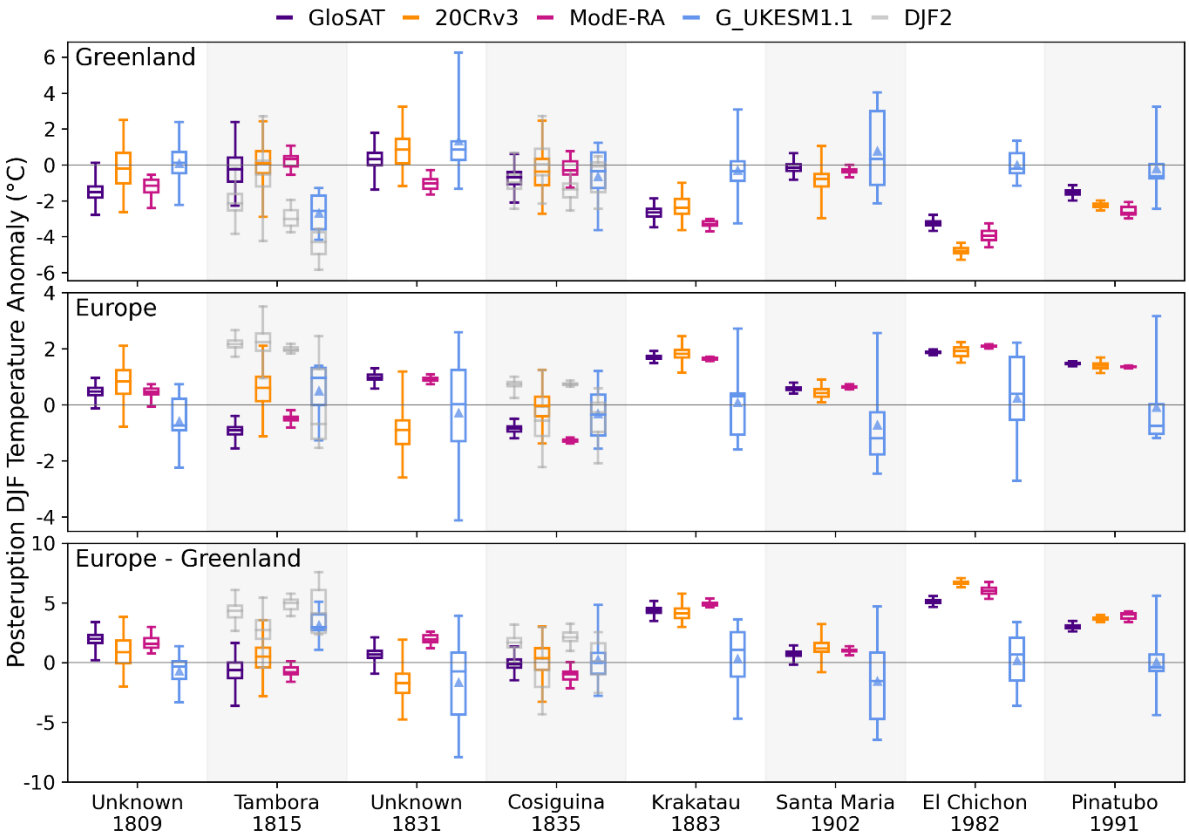


Figure 4.4. Box and whisker plots showing the ensemble spread, interquartile range and median of the post-eruption area-weighted mean temperature anomalies in DJF1 (colour) and DJF2 (grey) for the four surface temperature datasets assessed. DJF2 anomalies are shown only for Tambora and Cosiguina as the DJF2 anomalies are analysed in more depth for these eruptions; triangles indicate the ensemble mean in the G_UKESM1.1 dataset. Note that the individual datasets have not been masked to match the GloSAT coverage (as the complete response is of interest here) though GloSAT coverage is almost complete – see Figure 4.3.

An analysis of the GloSAT ensemble members (Figure 4.4; purple boxplots) indicates there is minimal uncertainty associated with the ensemble mean temperature anomalies following Krakatau and the twentieth century eruptions (ensemble spreads of up to 0.4°C in Europe and 1.6°C in Greenland) but, as anticipated, the observational uncertainty increases further back in time. This is particularly evident over Greenland where the ensemble spread ranges from 2.7°C following Cosiguina to 4.7°C following Tambora. Over Europe, the ensemble spread never alters the conclusions drawn from the ensemble mean, however, over Greenland, the larger ensemble spread suggests that either warm or cool conditions could have occurred following the eruptions of Tambora, U1831, Cosiguina and Santa Maria. This suggests the finding of a positive NAO surface temperature response following Santa Maria

should be treated with caution, but that a positive NAO response following U1831 is also possible within the range of uncertainty.

Overall, the GloSAT observations find conclusive evidence of surface temperatures consistent with a dynamic winter warming response following four eruptions: U1809, Krakatau, El Chichon and Pinatubo, and possible evidence (response within the range of uncertainty) following: Santa Maria and U1831.

4.4.2. NOAA-CIRES-DOE Twentieth Century Reanalysis v3

The 20CRv3 temperature anomalies in the first post-eruption winter (DJF1) are shown in Figure 4.5. At first glance, the data shows a reasonably consistent response to volcanism with warm anomalies present (to differing magnitudes and extents) over the Eurasian continent following all eight eruptions. However, the magnitude and extent of the warm anomalies differs vastly between eruptions (e.g. U1831 vs El Chichon) and following only five (U1809 and Krakatau to Pinatubo) is there overall warming present in Europe accompanied by cooling in Greenland, as would be expected during a positive phase of the NAO.

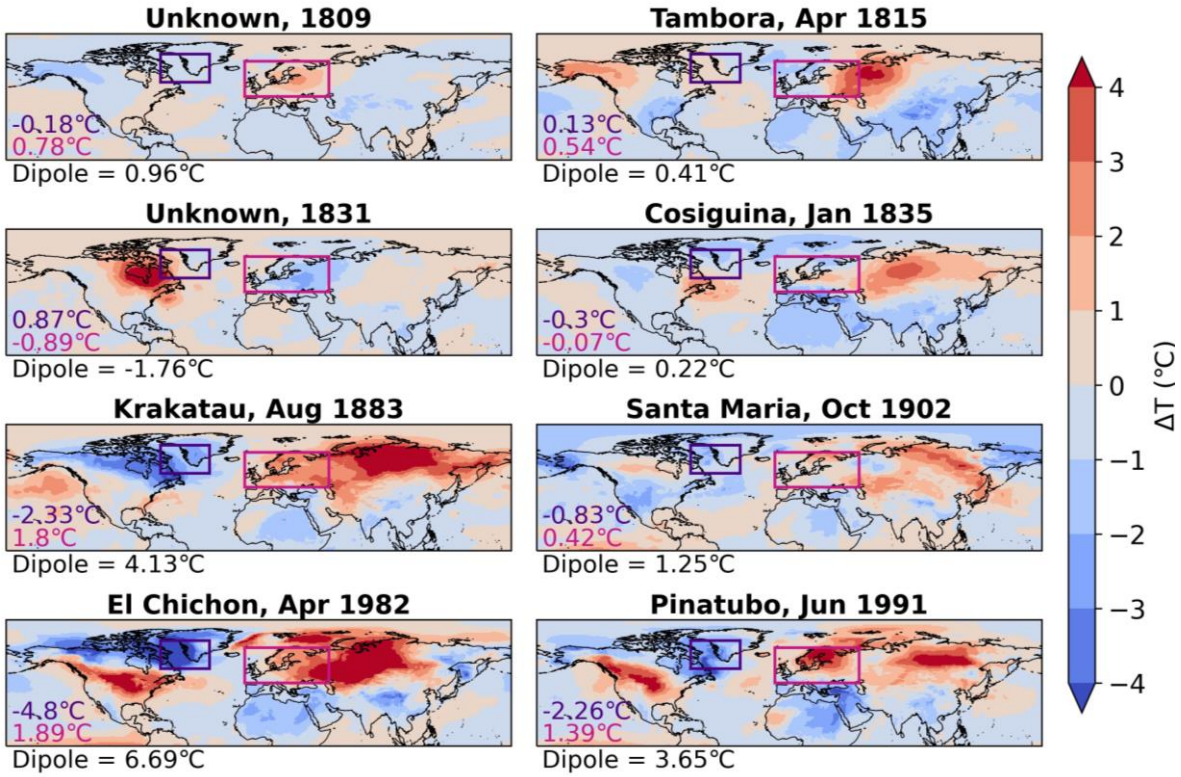


Figure 4.5. As for Figure 4.3, but for the 20CRv3 ensemble mean. Note that the anomalies are calculated with respect to a slightly shorter reference period than for the other datasets due to the 20CRv3 dataset starting in 1806.

Following the five eruptions which show a positive NAO response, the 20CRv3 results (Figure 4.5) are in good agreement with the GloSAT observations (Figure 4.3), thus independently confirming the response. The only inconsistencies between the datasets (following those eruptions) are the greater extent of the cool anomalies across Greenland in 20CRv3 following Santa Maria, the presence of regions of pronounced warming over the Barents and Greenland Seas in 20CRv3 following El Chichon and some differences in the magnitude of the temperature anomalies, most notably over Greenland, following U1831 and El Chichon.

Considerable similarities between the datasets also exist following the eruption of Cosiguina. As in the GloSAT observations (Figure 4.3), the 20CRv3 data (Figure 4.5) shows a small area of warming around the Baltic region, but not sufficient to suggest a positive phase of the NAO occurred. The only large disagreement between the two datasets is the presence of a more extensive region of warm anomalies over central Russia in 20CRv3. The absence of this in the GloSAT observations, however, may partly be an artefact of the spatial coverage of the observations in 1835/36 as there is little data present over central Russia during that time.

The level of agreement noted above, however, is not found following the eruptions of Tambora and U1831. Following Tambora, the 20CRv3 data presents evidence of warmer than normal conditions over Greenland (0.13°C) and Europe (0.54°C), with pronounced warm anomalies (of more than 4°C) in north-western Russia, to the west of the Europe box (Figure 4.5). This is in stark contrast to the GloSAT observations (Figure 4.3) which found no warm anomalies in the NH following Tambora. The 20CRv3 data also shows a vastly different response to the GloSAT observations following the U1831 eruption. In contrast to the cool anomalies found in northeast America and the warm anomalies found in Europe in the GloSAT observations, the 20CRv3 dataset shows the opposite: a substantial (≥4°C) region of warm anomalies over northeast North America and cool anomalies over western Europe (which lead to a negative dipole, inconsistent with a positive phase of the NAO).

An analysis of the 20CRv3 ensemble spread (Figure 4.4; orange boxplots) presents a similar picture to the GloSAT observations, with the range of uncertainty increasing further back in time. Overall, however, the data shows a higher level of uncertainty than the GloSAT observations. This is particularly apparent in the early nineteenth century when the ensemble spreads are substantially larger: up to 3.7°C in Europe and up to 5.3°C in Greenland. This is

thought to be the result of the use of an SST climatology before 1836, as well as the quantity and location of the surface pressure data assimilated into 20CRv3 (Slivinski et al., 2020). Between 1809 and 1836, fewer than 20 stations with pressure data are assimilated, compared to more than 400 in 1883 and 1902 (and several thousand for the later eruptions) (Figure C3; Compo et al. (2019)). Crucially, for this analysis, no pressure observations from Iceland are assimilated into 20CRv3 prior to 1845, meaning the atmospheric circulation anomalies following the early nineteenth century eruptions have little observational constraint and are therefore dominated by internal variability within the 20CRv3 ensemble together with any response of the underlying atmospheric model to the external forcings (Compo et al., 2019; Slivinski et al., 2020). As a result, the post-eruption anomalies in 20CRv3 should be treated with significant caution prior to 1845.

As alluded to above, from Krakatau onwards the dataset is more robust and the ensemble spreads generally do not alter the conclusions drawn from the ensemble means. The only exception to this is (again) Santa Maria, where the ensemble members do not rule out the possibility that Greenland was warmer (in agreement with the GloSAT ensemble). Interestingly, the conclusive warming over Europe following Krakatau is in contrast to Polvani & Camargo (2020), who found that both warming and cooling was evident in the ensemble members in version 2c of the NOAA-CIRES Twentieth Century Reanalysis data. Note that the finding here remains true even when the analysis is repeated using the larger study region used by that study: 40-70°N; 0-150°W (not shown).

Overall, the 20CRv3 ensemble mean anomalies provide evidence of surface temperatures consistent with a positive phase of the NAO following five eruptions: U1809, Krakatau, Santa Maria, El Chichon and Pinatubo. However, analysis of the ensemble members suggests the cool response over Greenland following Santa Maria may not be robust and highlights significant uncertainty in the temperature anomalies following the early nineteenth century eruptions.

4.4.3. Modern Era Reanalysis

The ModE-RA post-eruption temperature anomalies for DJF1 are shown in Figure 4.6. As in 20CRv3, following each of the eight eruptions, there is evidence (to differing extents) of warmer than usual winter conditions over parts of the Eurasian continent. However, following only six eruptions: U1809, U1831, Krakatau, Santa Maria, El Chichon and Pinatubo, is the spatial distribution of the warm anomalies consistent with a positive phase

of the NAO and accompanied by cool anomalies over Greenland. Following Cosiguina, warm anomalies are limited to a few regions in DJF1, and cool conditions prevail over the Europe study region, and following Tambora, cool and warm conditions are dominant over western Europe and the Labrador Sea, respectively: the opposite of the temperature dipole seen during a positive phase of the NAO.

As expected (because the datasets are not fully independent), there is very good agreement between the ModE-RA results and the GloSAT observations following the majority of eruptions. This is particularly true over Eurasia (where the spatial distribution and magnitude of the anomalies are very similar following all of the eruptions) and following the five eruptions which show a positive NAO response in the GloSAT observations. Following the five eruptions with a positive NAO response, the only (notable) differences between the two datasets are the additional pronounced warming in the Barents and Greenland Seas following El Chichon and the larger magnitude of the cool anomalies in ModE-RA over the Labrador Sea, particularly following the more recent eruptions (as is seen in 20CRv3).

Where larger disagreements between the two datasets are seen is over Greenland following the eruptions of Tambora and U1831. In contrast to GloSAT, ModE-RA identifies a surface temperature response characteristic of a positive NAO following U1831 (Figure 4.6) with cool anomalies present over Greenland (-1.03°C) in addition to warming over Europe (0.91°C). This is in contrast to the GloSAT observations (Figure 4.3) which do not provide evidence of cooler conditions over Greenland in the ensemble mean (although a cooling response of up to -1.38°C is within the range of uncertainty). Following Tambora, the opposite disagreement is evident: the ModE-RA data suggests slightly warmer than average conditions (0.21°C) prevailed over Greenland, particularly over the Labrador Sea, whereas the GloSAT observations suggest cooler conditions (-0.26°C) occurred. This discrepancy, however, may partly be an artefact of the spatial coverage of the GloSAT observations, as the datasets agree well where the coverage overlaps. Additionally, as with U1831, the warming identified by ModE-RA following Tambora is within the range of uncertainty in the GloSAT observations, with warming of up to 2.38°C evident in the ensemble.

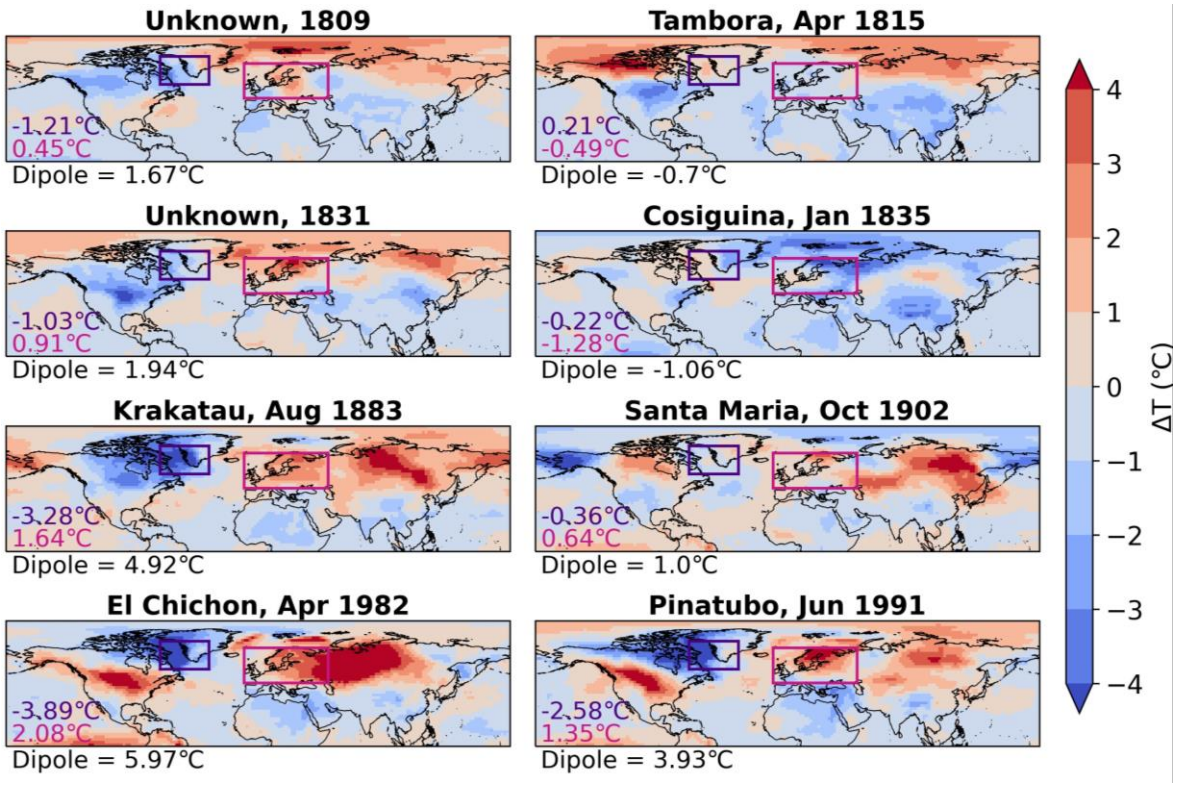


Figure 4.6. As for Figure 4.3, but for the ModE-RA ensemble mean.

An analysis of the ModE-RA ensemble members (Figure 4.4; pink boxplots) shows a relatively small ensemble spread following the more recent eruptions (0.1°C - 0.2°C in Europe and 0.7°C - 1.3°C in Greenland), in line with the GloSAT observations and the 20CRv3 data. However, unlike the latter datasets the spread also remains relatively small following the early nineteenth century eruptions: only up to 0.8°C in Europe and 2°C in Greenland. This is most likely due to the additional constraints introduced by the assimilation of climatic information from winter-responding proxies and documentary sources (Section 4.2.3) as well as the fact that ModE-RA does assimilate pressure data from Iceland prior to 1845, which provides an additional observational constraint. The small ensemble spreads confirm the robustness of the conclusions drawn above for the six eruptions which show a positive NAO response, including Santa Maria and U1831 which are more uncertain in the other datasets.

Overall, the ModE-RA dataset provides evidence of air temperatures consistent with a positive phase of the NAO following six eruptions: U1809, U1831, Krakatau, Santa Maria, El Chichon and Pinatubo. The more constrained nature of ModE-RA also suggests that a response consistent with a positive phase of the NAO is likely to have occurred following U1831 and Santa Maria, even though there is not conclusive evidence of cooler conditions

over Greenland in the GloSAT observations. This suggestion is not inconsistent with the GloSAT ensemble spread, and, for U1831, agrees with the findings of Shindell et al. (2004).

4.4.4. United Kingdom Earth System Model version 1.1

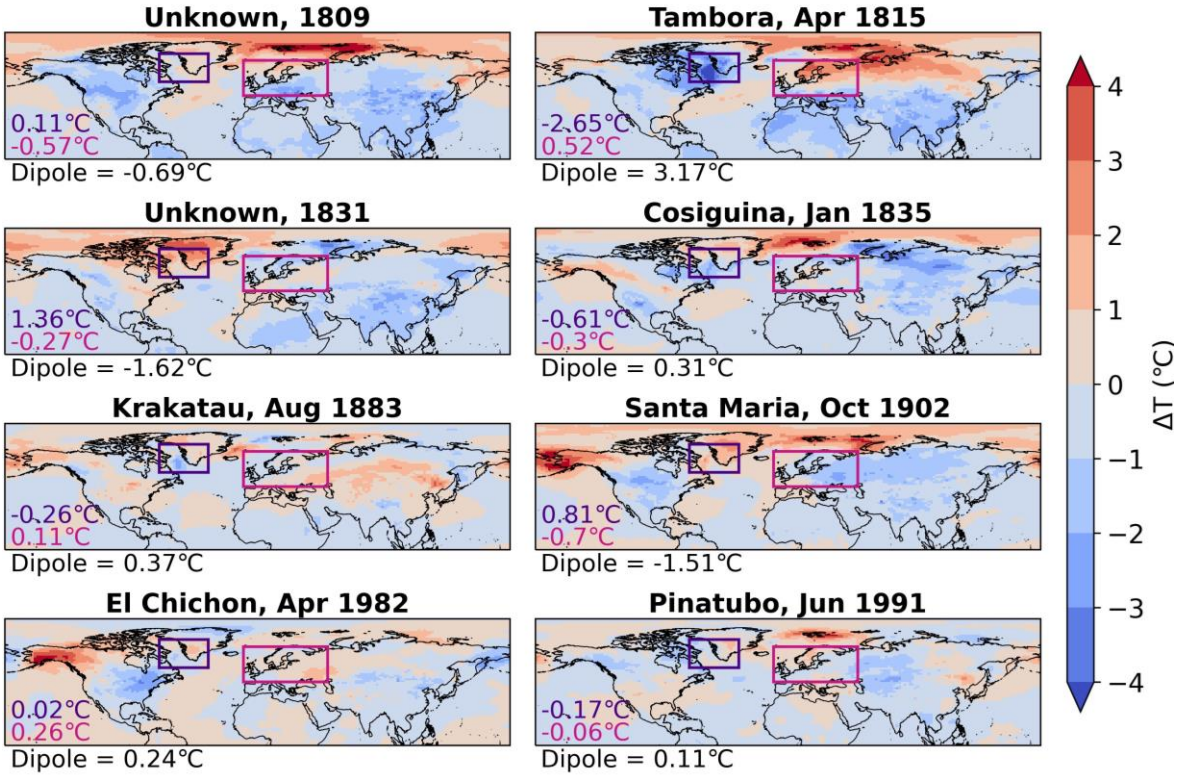


Figure 4.7. As for Figure 4.3, but for the G_UKESM1.1 ensemble mean.

The post-eruption (DJF1) temperature anomalies for the G_UKESM1.1 ensemble mean are shown in Figure 4.7. In contrast to the observations (Figure 4.3) and reanalyses (Figures 4.5 and 4.6) a winter temperature response similar to a positive phase of the NAO (warming over Europe and cooling over Greenland) is only present following Tambora and Krakatau (although with a much weaker magnitude than the observations following the latter) and there is no consistent response following the remaining six eruptions. U1809, U1831 and Santa Maria present evidence of warm conditions over Greenland and cool conditions over Europe; El Chichon shows warm anomalies over both study regions and Cosiguina and Pinatubo show the opposite: cooling over both. In addition, with the exception of Tambora, the magnitudes of the anomalies are generally small (less than ±0.26°C in either study region following Krakatau, El Chichon and Pinatubo which show the largest responses in the observations) and, where the anomalies are larger (e.g., 1.36°C in Greenland following U1831) they are often of the opposite sign to a positive phase of the NAO.

The only possible exception to the above is Tambora. In the first post-eruption winter, the G_UKESM1.1 ensemble means (Figure 4.7) show substantial cooling over southwestern Greenland and the Labrador Sea, as well as extensive warming over northern Eurasia. The European warm anomalies are shifted further north than is evident in many of the observations (as found in modelled responses by Azoulay et al. (2021)), however, the magnitudes of the anomalies (-2.65°C in Greenland, 0.52°C in Europe and a 3.17°C regional dipole) are not dissimilar to those present (following other eruptions) in the observations and reanalyses. The latter is in spite of the fact that the G_UKESM1.1 *ensemble mean* anomalies have much reduced internal variability (reduced by $1/\sqrt{6}$, i.e., about 40% as strong as in a single run of the model) because, unlike the observation and reanalysis ensembles, the G_UKEMS1.1 ensemble members sample over completely independent realisations of internal variability.

These findings are consistent with previous modelling studies (Driscoll et al., 2012; Marshall et al., 2009; Polvani et al., 2019; Polvani & Camargo, 2020) which suggest that climate models typically do not replicate a dynamic winter warming response following eruptions of the magnitude of Krakatau and Pinatubo but may be able to following eruptions of a larger magnitude (Azoulay et al., 2021; Dallasanta & Polvani, 2022).

4.4.5. DJF2 temperature response

Previous observation and proxy-based studies (Kelly et al., 1996; Robock & Mao, 1995; Wunderlich & Mitchell, 2017) have suggested a surface temperature response to explosive volcanism also occurs in the second post-eruption winter (DJF2), with Fischer et al. (2007) even finding a stronger composite response over Eurasia in DJF2 than in DJF1. More recently, however, the plausibility of these findings has been questioned due to the reduced stratospheric aerosol present in the second post-eruption winter and the probable lack of memory associated with the stratospheric pathway (Dallasanta & Polvani, 2022; Polvani et al., 2019). To investigate whether a temperature response is present in DJF2 in the surface temperature datasets analysed here, post-eruption anomalies were also calculated for the second post-eruption winter following each eruption.

As in DJF1, the G_UKESM1.1 simulations show little evidence of a surface temperature response to volcanism in DJF2 (Figure C4). As in Section 4.4.4, the one exception to this is Tambora - which has a substantial quantity of aerosol remaining in the NH in DJF2 (0.34-0.40; Figure C1) - which does show substantial cooling over the Labrador Sea (-4.46°C) and

some localized warming over Eurasia in DJF2 (but with no clear are-averaged warming signal: 0.01°C; Figure 4.8). The remainder of this section will therefore focus on the DJF2 response in the GloSAT observations and the ModE-RA and 20CRv3 datasets.

Focussing first on the six eruptions which likely show evidence of a positive NAO response in DJF1 (U1809, U1831, Krakatau, Santa Maria, El Chichon and Pinatubo), five show evidence of a sustained response in DJF2 with warm and cool anomalies (consistent with a positive phase of the NAO) present over Europe and Greenland, respectively (Figures C5 to C7). As in DJF1, there are some inconsistencies between the three datasets, particularly following the U1831 and U1809 eruptions, however, only U1809 does not show evidence of a sustained response over both Europe and Greenland in at least one dataset. The spatial warming pattern evident in DJF2 is slightly different to DJF1 - the warm anomalies over Eurasia are less widespread, shifted poleward and usually of a lower magnitude. However, overall, the results do support the findings of previous studies that a dynamic winter warming response can also be seen during DJF2. The results do not, however, support the finding of a response that is stronger in DJF2 than in DJF1, in contrast to Fischer et al. (2007).

Perhaps more interesting than the above, however, is the finding that both Tambora and Cosiguina, which do not show evidence of a positive NAO temperature response in DJF1, do in DJF2 (Figure 4.8). The warm anomalies over Europe following Cosiguina (0.7°C) are of a similar magnitude and distribution to those found in DJF1 following other eruptions and are present in every ensemble member in both ModE-RA and GloSAT (Figure 4.4; grey boxplots), meaning they are clearly stronger than observational uncertainty. Similarly, the anomalies found in DJF2 following Tambora have a similar spatial distribution to those found following other eruptions in DJF1, and the response is also present in every ensemble member (Figure 4.4; grey boxplots). The magnitude of the anomalies following Tambora are significantly larger: between 1.98°C and 2.25°C over Europe and between -0.41°C and -2.92°C over Greenland (dependent on the dataset). This is marginally warmer than the largest anomalies seen over Europe in DJF1 (2.08°C following El Chichon) and nearly twice as warm as those seen in DJF2 (the next largest being 1.33°C following Pinatubo). The finding of pronounced warming in DJF2 following Tambora (but not in DJF1) may help to explain the stronger composite Eurasian winter warming response found in DJF2 by Fischer et al., (2007), who included Tambora in their composite.

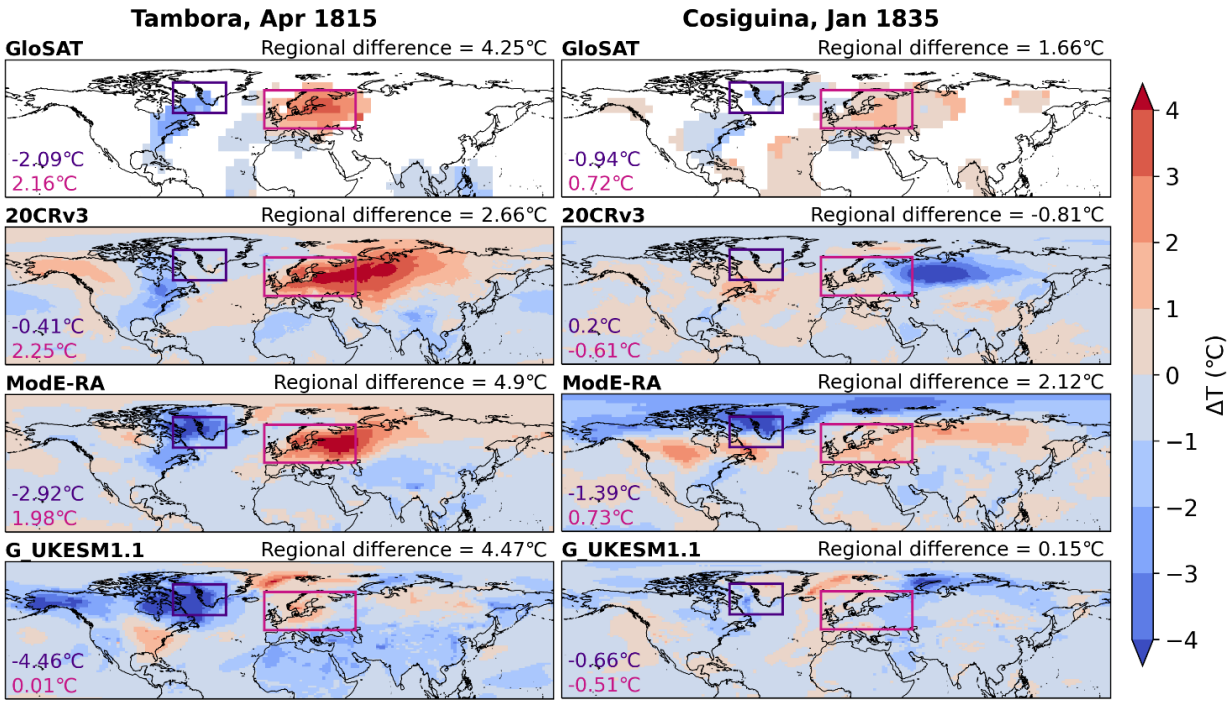


Figure 4.8. Post-eruption DJF2 surface temperature anomalies in each dataset following Tambora (1815) and Cosiguina (1835). The number in purple (pink) is the area-weighted mean temperature anomaly for the Greenland (Europe) box and the regional difference is the difference between the two (Europe minus Greenland).

Overall, the observations and reanalyses suggest that a temperature response, similar in characteristic to a positive phase of the NAO, is present (to differing extents) following seven eruptions in DJF2: Tambora, U1831, Cosiguina, Krakatau, Santa Maria, El Chichon and Pinatubo; including two eruptions which did not show a response in DJF1: Tambora and Cosiguina.

4.5. Post-eruption winter circulation response

To complement the analysis in Section 4.4, the winter circulation response (as indicated by the NAO indices associated with each dataset) to explosive volcanism is also assessed as a separate line of evidence for a dynamic response to explain the European winter warming.

4.5.1. Observations and reanalyses

Figure 4.9 shows the observed NAO index as well as the NAO index from the reanalyses (20CRv3 and ModE-RA). There is good visual agreement between the three datasets from the mid-1800s. This is confirmed by a rolling (10-year) correlation (Figure 4.9) which shows generally high correlation coefficients ($r \geq 0.9$) throughout the overlapping periods except before 1860 in 20CRv3 and a short period in the 1910s in all datasets where the correlations

weaken. As the correlations here and the analysis of the ensemble spread discussed in Section 4.4.2 both suggest the 20CRv3 data is not as reliable before the mid-1800s (due to the use of climatological SSTs and the assimilation of few pressure observations, including none over Iceland), the 20CRv3 dataset will not be included in the assessment of the NAO index following the earlier four eruptions.

The NAO indices suggest a positive phase of the NAO occurred following up to six eruptions in DJF1: U1809, U1831, Krakatau, Santa Maria, El Chichon and Pinatubo, and following up to five eruptions in DJF2: Tambora, U1831, Cosiguina, El Chichon and Pinatubo. In DJF1 these results are consistent with the temperature-based findings in Section 4.4; however, the MSLP data suggest fewer eruptions have a positive NAO in DJF2 than suggested by the temperature anomalies. Krakatau does not have a positive NAO in DJF2 in either dataset and Santa Maria only has a (marginally) positive index in the observations (Figure 4.9).

The presence of a positive NAO index, however, does not tell the whole story. As is noted by Christiansen (2008), the NAO shows periods with some apparent persistence which means there are frequently consecutive winters with the same sign (see Figure 4.9). Although the overall autocorrelation in the NAO index is low - Eade et al. (2022), for example, estimate lag-1 autocorrelations of 0.169 and 0.142 in the HadSLP2r (Allan & Ansell, 2006) and 20CR (Compo et al., 2011) winter NAO indices (for the period 1850-2020), respectively. This fact does raise questions regarding the significance of the occurrence of a positive NAO following a volcanic eruption. Krakatau, for example, has a positive NAO index in the first post-eruption winter, but also has a positive index in the winter prior to the eruption. In such instances it is therefore difficult to determine whether the positive phase is the result of the eruption, or whether it would have occurred anyway due to the apparent persistence. To address this, the indices are assessed again, this time taking into account the phase of the NAO in the winter prior to the eruption (referred to as DJF0).

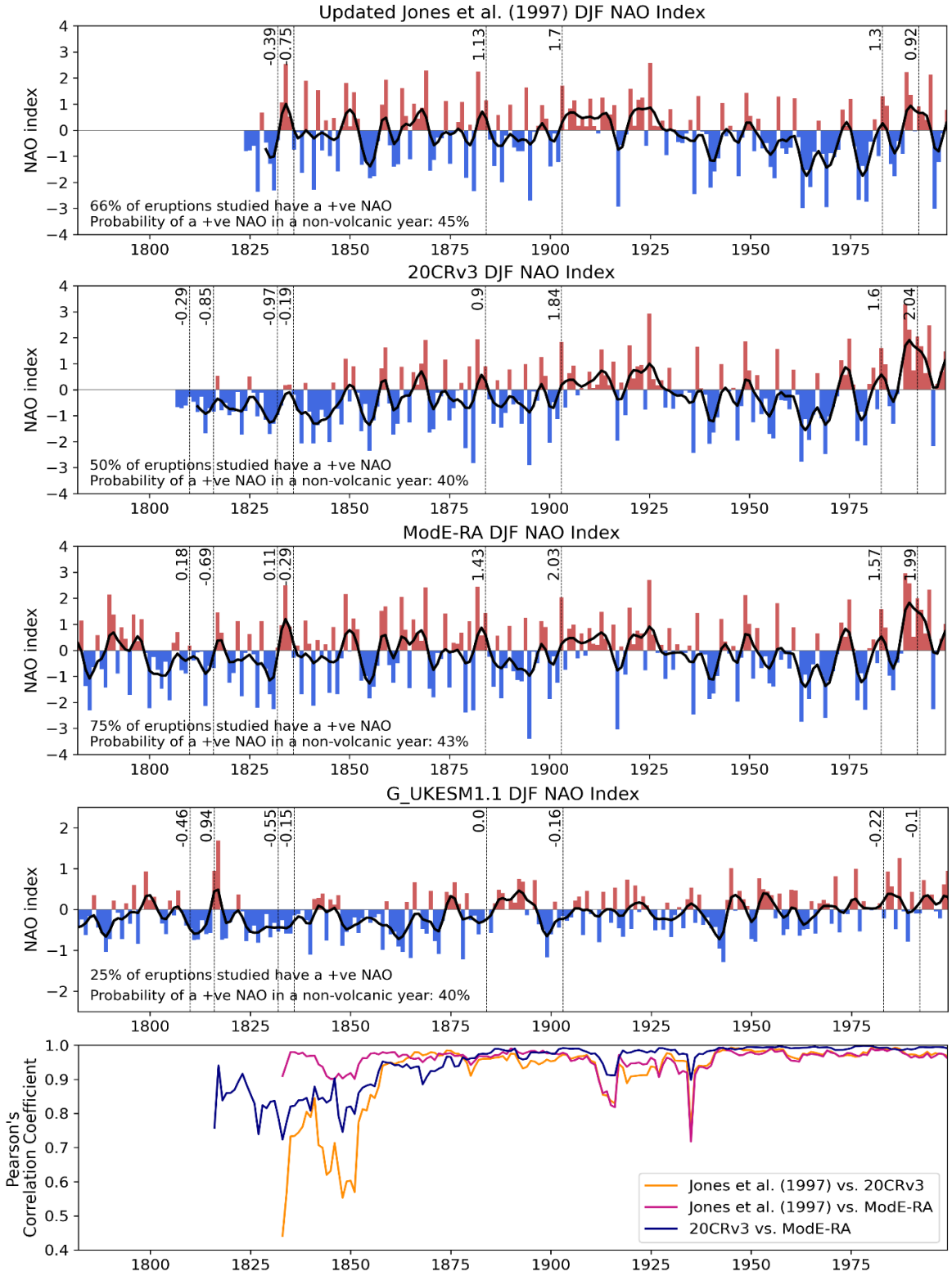


Figure 4.9. Upper four panels: North Atlantic Oscillation indices calculated from observed and simulated mean sea level pressure. The vertical dashed lines and annotated values represent the NAO index in DJF1 following each eruption (where data is available). The indices for the 20CRv3, ModE-RA and G_UKESM1.1 datasets are the ensemble mean values. NAO indices for each G_UKESM1.1 ensemble member can be found in Appendix C (Figure C8). Lower panel: 10-yr rolling correlation between the Jones et al. (1997), 20CRv3 and ModE-RA NAO indices.

Looking at the same three datasets, the results show that up to five eruptions: U1809, U1831, Santa Maria, El Chichon and Pinatubo, present evidence of a shift from a negative phase in DJF0 to a positive phase in DJF1, and that six eruptions (all except Tambora and Cosiguina) show a shift to more positive values (i.e. where a positive NAO occurred in DJF0, as is the case for Krakatau, the magnitude of the index increased in DJF1). The same is also true for Tambora in DJF2 which shows a shift to a positive NAO index following negative indices in DJF0 and DJF1, and to a lesser extent for Cosiguina which shows a shift from a negative index in DJF1 to either a weaker negative value in the observations or a small positive in ModE-RA. Of the other eruptions which show a positive NAO index in DJF2 (U1831, Santa Maria, El Chichon and Pinatubo) only U1831 has a stronger (more positive) index than DJF1. This, however, is not unexpected as the surface temperature response is also generally weaker following these eruptions.

Overall, the findings suggest a shift to a positive index, or a strengthening of an existing positive NAO index, occurred following six eruptions in DJF1 (U1809, U1831, Krakatau, Santa Maria, El Chichon and Pinatubo) and following three eruptions in DJF2 (Tambora, U1831 and Cosiguina). The results in DJF1 concur with Christiansen (2008) who identified a positive NAO following 85% of the largest volcanic eruptions since 1870 as well as the findings of Ortega et al. (2015); Sjolte et al. (2018) and Wunderlich & Mitchell (2017) who all identified a shift toward positive NAO indices following the majority (or all) of the explosive eruptions they assessed. The finding of a positive index following Tambora in DJF2, but not DJF1, is also supported by Hanna et al. (2022).

4.5.2. G_UKESM1.1 model

Similar to the findings in Section 4.4, the NAO index simulated by G_UKESM1.1 does not replicate the same tendency for a positive NAO in the winters following the eruptions as the observations and reanalyses. A positive NAO index in the ensemble mean is identified in DJF1 following only two eruptions: Tambora and (marginally) Krakatau, which is not significant given that a positive NAO index occurs in 40% of non-volcanic winters (Figure 4.9). For completeness, an analysis of the NAO index in the ensemble members was also undertaken (Figure C8). This shows a similar occurrence of positive NAO indices in non-volcanic winters (41%-54%), but with much intra-ensemble variability in the percentages following volcanic eruptions: ranging from 12% in ensemble member four, to 100% in ensemble member three. This illustrates that an individual simulation (i.e., ensemble

member) could by chance show a tendency to simulate a positive NAO even when the model shows no overall NAO response to volcanic forcing. Overall, there is little evidence to suggest the G_UKESM1.1 model consistently simulates a positive NAO in either the first or second post-eruption winter, in contrast to what is seen in the observations and reanalyses. These results are in line with Swingedouw et al. (2017), who found the majority of the PMIP3 models do not produce a significant shift to a positive NAO following volcanic eruptions, as well as Driscoll et al. (2012) who showed the same for the CMIP5 models.

It is worth noting that Tambora is the only eruption for which all six G_UKESM1.1 ensemble members show the same sign (positive) NAO index in DJF1 (Figure C8). All six also show a positive NAO index in DJF2. The G_UKESM1.1 ensemble mean shows a large positive NAO index of 0.94 in DJF1 and an even stronger index of 1.68 in DJF2. The latter value is the largest NAO index (of either sign) present in the ensemble mean during the period of study and both values are of a similar magnitude to the NAO indices found (following other eruptions) in the observations. This is notable because any contribution from internal variability will be reduced in the ensemble mean. This finding concurs with the suggestion of Swingedouw et al. (2017) that only large eruptions are able to produce a positive NAO response in the models, since Tambora is the strongest eruption considered here.

4.6. Assessing the evidence for a dynamic winter response to explosive volcanism

The results of Sections 4.4 and 4.5 suggest a winter temperature and circulation response, consistent with a positive phase of the NAO, occurred in DJF1 following six of the eruptions assessed: U1809, U1831, Krakatau, Santa Maria, El Chichon and Pinatubo, and in DJF2 following five eruptions: Tambora, U1831, Cosiguina, El Chichon and Pinatubo. The identification of temperature and pressure anomalies consistent with a positive NAO, however, is not by itself evidence of volcanic forcing. Robock & Mao (1992), for example, found temperature anomalies of a similar pattern to those identified following volcanic eruptions in 21% of non-volcanic winters (between 1866-1992) and, as the NAO is a natural mode of variability, positive phases of the NAO regularly occur in the absence of volcanic forcing. This is nicely illustrated by the observed NAO index (Figure 4.9; Jones et al. (1997)) which suggests that a positive phase of the NAO index occurs in 45% of non-volcanic years. Thus, this section will assess whether there is evidence that the identified anomalies (with a focus on temperature) were a result of the preceding volcanic eruptions, or how likely it is that they would have occurred even in the absence of an eruption. As the conclusions drawn

in this section are influenced by whether it is accepted that a volcanically forced response is plausible in DJF2, and as this plausibility has been questioned previously (see Section 4.4.5), the ability of volcanic eruptions to have an effect on the NH winter climate in DJF2 will first be discussed.

4.6.1. Is a volcanically forced response in the second post-eruption winter plausible?

As noted in Section 4.4.5, some previous studies have questioned the plausibility of a winter temperature and circulation response in DJF2, suggesting a) that there is not sufficient aerosol present in the stratosphere in DJF2 to cause a response and b) that there is no memory associated with the stratospheric pathway which could sustain a response from the first post-eruption winter. An assessment of the NH SAOD in DJF2 (Figure C1), however, suggests that a) may not be a valid criticism for eruptions of the magnitude assessed here. Figure C1 shows that, following all eight eruptions, the aerosol load in DJF2 in the NH is at least as large as that present in DJF1 following the eruption of Mount Agung (1964), an eruption that was followed by winter warming over Eurasia (Graf et al., 2007; Kelly et al., 1996; Robock & Mao, 1995; Shindell et al., 2004). Thus, the sustained warming (and Greenland cooling) identified in DJF2 is not inconsistent with the quantities of aerosol present in the stratosphere. This is especially true if the aerosol particle size is taken into account. Stenchikov et al. (1998) showed that smaller aerosol particles – which are likely to constitute a larger proportion of the SAOD in DJF2 – absorb more near-IR radiation than larger particles; therefore, the lower stratospheric heating in DJF2 may be comparatively larger than in DJF1, even if the same SAOD were present. This is supported by Marshall et al. (2020) who found the forcing per unit of SAOD is weaker in the first year following an eruption in comparison to the second and third years.

It is also worth noting here that the magnitude of the anomalies following Tambora in DJF2 (similar in size to the largest post-eruption anomalies in DJF1) are also plausibly volcanically forced because, due to the magnitude of Tambora, the reconstructed NH SAOD in DJF2 (0.149 – 0.247, Figure C1) is of a similar magnitude to, or larger than, that present in DJF1 following the majority of the other eruptions. Thus, it is concluded that the quantities of aerosol present following each eruption are not inconsistent with the magnitudes of the temperature anomalies and, based on the results of observation-based studies, are sufficient to produce a dynamic circulation response.

What is perplexing here, however, is the existence of a potential temperature response in DJF2 following Tambora and Cosiguina, when no response was present in DJF1. It could be argued that the less pronounced anomalies following Cosiguina are merely the result of internal variability, however, the magnitude of the anomalies in DJF2, as well as the quantities of aerosol present in the stratosphere in both DJF1 and DJF2 (Figure C1), make this argument less convincing for Tambora. As this finding in itself could lead to questions about the plausibility of a volcanically forced response in DJF2, a few potential explanations for the delayed response are outlined below.

Previous studies have suggested a dynamic winter warming response occurs in the second post-eruption winter following mid- and high-latitude eruptions, due to the additional time taken for the aerosol to reach the tropics and for a strengthened equator-to-pole temperature gradient to form (Robock & Mao, 1995). While the tropical (8.25°S; 118.00°E) location of Tambora is well documented (Stothers, 1984) it is possible that the quantity of aerosol present in the tropics, and/or in the NH, was not sufficient to strengthen the NH equator-to-pole temperature gradient until DJF2. This could have occurred either because the explosivity of the eruption lofted the aerosol high into the atmosphere where it was transported straight to the polar regions by the upper 'deep' branch of the Brewer-Dobson circulation (Plumb, 2002) or because the aerosol remained tightly confined to the Southern Hemisphere and/or equatorial region during DJF1. Neither scenario is represented in the reconstructed SAOD profiles for Tambora used to force the climate or reanalysis models (shown in Figure C1); however, Clyne et al. (2021) found that some models simulate the equatorial confinement of aerosol following Tambora and earlier studies by Arfeuille et al. (2014) and Crowley et al. (2008) both suggest Tambora may have produced an asymmetrical aerosol cloud, with the majority of the aerosol present in the SH. Ultimately, the real-world quantity and distribution of SAOD is uncertain. Large disagreements between the simulated SAOD profiles following Tambora have previously been identified by Marshall et al. (2018) and Clyne et al. (2021) and the estimated quantity and distribution of SAOD are largely based on observations following Pinatubo, a much smaller eruption with potentially different background atmospheric circulation (Table 4.1). As a result, both scenarios are plausible, even if not supported by the reconstructions in Figure C1.

In addition to the above, it is also plausible that the size of the aerosols in the first post-eruption winter following Tambora could have prevented a dynamic circulation response in DJF1. Studies (e.g. Arfeuille et al. (2014)) show that the larger the stratospheric

aerosol injection associated with an eruption, the larger the average aerosol particle size (effective radius) is. Due to the large magnitude of the Tambora eruption, it is therefore likely that the aerosol particles in the first post-eruption winter following Tambora were substantially larger than following the other eruptions assessed. As noted earlier, larger aerosol particles are less efficient at absorbing radiation than smaller particles (Stenchikov et al., 1998), thus the large particle size may have limited the heating of the lower stratosphere and thus the strengthening of the equator-to-pole temperature gradient in DJF1. As larger aerosol particles fall out of the atmosphere more quickly than smaller particles, this limitation would not have been present in DJF2, thus potentially explaining the delayed response following Tambora.

Finally, it is also worth noting that both Tambora and Cosiguina are the only eruptions included in this analysis which are closely preceded by other large low latitude eruptions. U1809 erupted only six years prior to Tambora and U1831 potentially less than four years prior to Cosiguina. Previous modelling studies have suggested the combined climate response to two sequential eruptions differs from the response to one (Zanchettin et al., 2013); therefore, it is possible this factor could have contributed to the lack of Eurasian winter warming present in DJF1 following these eruptions.

While the reasons for the delayed post-eruption temperature and circulation responses following Tambora and Cosiguina remain unknown, this section concludes that there may have been sufficient aerosol present in the stratosphere following all of the eruptions assessed (in this study) to make a volcanically forced response in DJF2 plausible. Thus, the focus will now turn to the question of whether there is evidence that the surface temperatures and NAO indices identified in Sections 4.4 and 4.5 were volcanically forced.

4.6.2. Is there evidence of a volcanic signal in the observations and reanalyses?

Having considered each eruption in turn (Section 4.4), the magnitudes of the post-eruption temperature anomalies will now be compared with the magnitudes of the anomalies that occurred without volcanic forcing. This comparison is made in Figure 4.10, which shows the area-weighted mean post-eruption temperature anomalies for DJF1 and DJF2 (for each eruption) compared to a histogram of the anomalies present in the non-volcanic reference period (Section 4.3.3) in each dataset. Each histogram gives an approximation of the internal variability in the climate system and provides an indication of the likelihood that an anomaly could have occurred by chance.

A comparison between the histograms and the individual post-eruption temperature anomalies (Figure 4.10) proves inconclusive. Following only Krakatau and El Chichon in DJF1 and Tambora and Pinatubo in DJF2, do the Greenland post-eruption temperature anomalies fall outside 2σ (in at least two of the datasets), indicating they were unlikely to have occurred without external forcing. The anomalies in Europe following the same eruptions, and across all three metrics following the other eruptions (Figure 4.10); however, are not especially anomalous, falling within 1σ or 2σ (in line with the findings of other studies, including Polvani & Camargo (2020)). This raises the possibility that the identified anomalous cooling over southwest Greenland (and thus the strong regional dipole) may instead result from a direct radiative (cooling) response to volcanism, rather than a dynamic circulation response via the NAO. Thus, following the majority (all) of the eruptions studied, the magnitude of the anomalies over Greenland (Europe) do not, by themselves, provide conclusive evidence for a volcanic signal as the majority (all) could realistically have occurred in the absence of an eruption.

However, as the stratospheric pathway, which is believed to account for the observed winter warming over Eurasia (described in Section 4.1), essentially excites a natural mode of variability (the NAO), it may not be reasonable to expect the anomalies to be ‘unusual’ or to fall outside of the range of internal variability. Instead, the consistency of the response across the eight eruptions, may be more appropriate to consider.

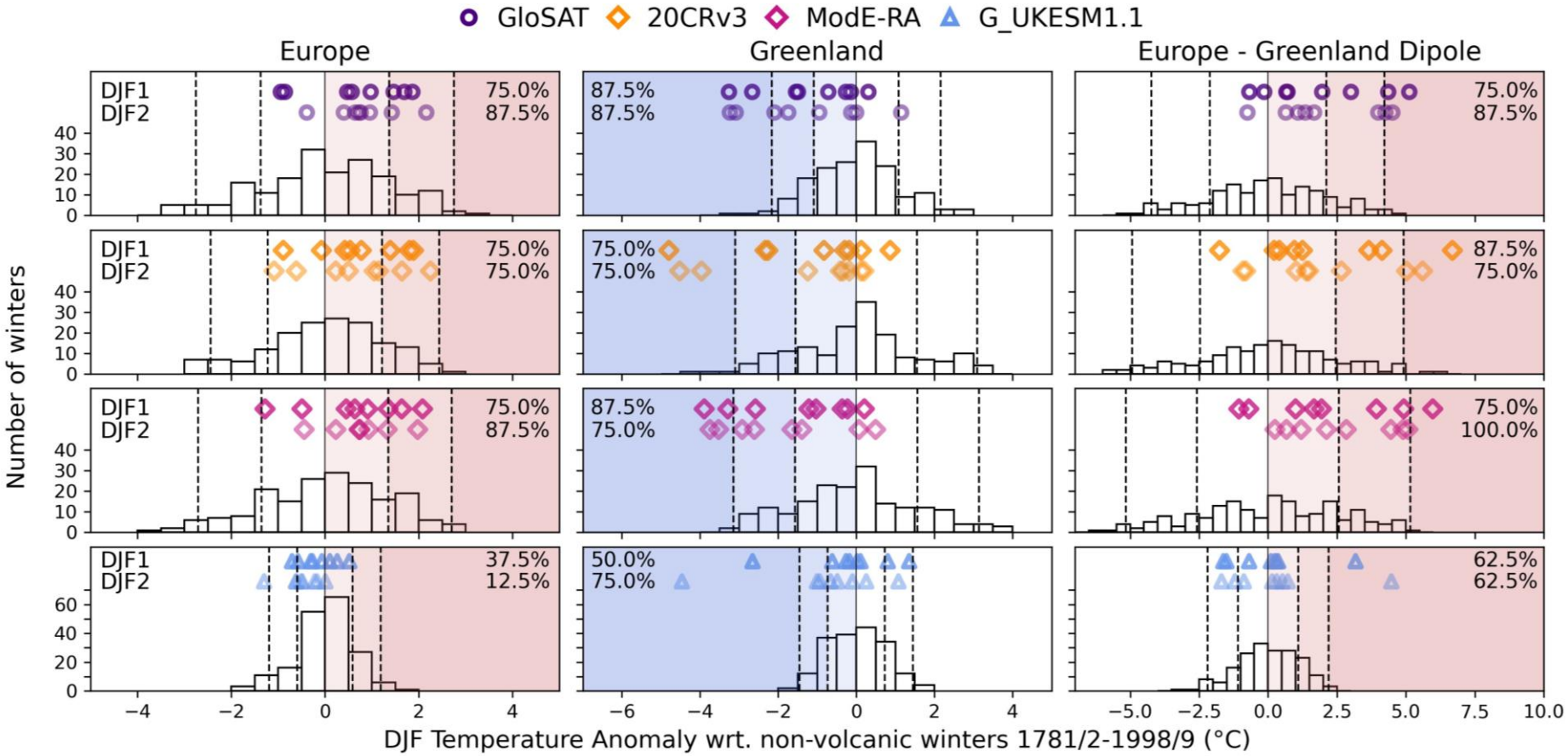


Figure 4.10. Post-eruption ensemble mean temperature anomalies for DJF1 and DJF2 (coloured markers) plotted against histograms of the temperature anomalies in every non-volcanic winter between 1781/82-1998/99. Dashed vertical lines represent 1σ and 2σ of the non-volcanic winters; the shading represents the sign of the anomaly that corresponds with a positive phase of the NAO and the percentage of volcanic winters with an anomaly of this sign is indicated on each panel for DJF1 and DJF2. Each row represents a different dataset: rows from top to bottom are for GloSAT, 20CRv3, ModE-RA and G_UKESM1.1 (histograms for G_UKESM1.1 are narrower because this ensemble member averages over different realisations of internal variability). Note that the anomalies for 20CRv3 are expressed with respect to (wrt.) the non-volcanic winters for 1806/07-1998/99 (and the histograms also represent this slightly shorter period).

In Figure 4.10 it is notable that the histograms are approximately normally distributed around 0°C, however, the post-eruption temperature anomalies in DJF1 and DJF2, in all three datasets, have a clear tendency to be negative in Greenland and positive in Europe. In the latter case, this is despite the known direct negative radiative forcing of explosive eruptions. Even in the relatively few instances where the anomalies are not of the anticipated sign (26/144), only one anomaly exceeds 1σ in magnitude. If there were no external forcing (or no dynamic response), the post-eruption anomalies might be expected to be more centred on zero (or shifted toward more negative temperatures due to the negative radiative forcing). Thus, the shifted distribution (in the anticipated direction) in both study regions and in both post-eruption winters, lends weight to the suggestion that the temperatures are volcanically forced and that there is a dynamic circulation response to explosive volcanism in the NH. This conclusion is supported by a Student's t-test and a Kolmogorov-Smirnov test which find the magnitude and distribution of the post-eruption anomalies in DJF1 and DJF2 (combined) to be significantly different (at the 95% level) from the non-volcanic winter anomalies in all three metrics: Greenland, Europe and the regional dipole, and in all three datasets (except over Europe in 20CRv3) (Table 4.2).

Table 4.2. Results of a Student's t-test and Kolmogorov-Smirnov test comparing the (combined) DJF1 and DJF2 temperature anomalies with the temperature anomalies in non-volcanic winters. The numbers in brackets for G_UKESM1.1 are the results excluding Tambora.

		Student's t-test		**Kolmogorov-Smirnov test**	
		t-statistic	***p* value**	**KS-statistic**	***p* value**
GloSAT	Greenland	-4.26	0.000	0.39	0.017
	Europe	2.13	0.035	0.41	0.010
	Dipole	3.68	0.000	0.42	0.008
20CRv3	Greenland	-3.03	0.003	0.37	0.029
	Europe	2.18	0.031	0.29	>0.05
	Dipole	3.00	0.003	0.35	0.042
ModE-RA	Greenland	-4.22	0.000	0.43	0.005
	Europe	2.09	0.038	0.37	0.027
	Dipole	3.68	0.000	0.40	0.014
G_UKESM1.1	Greenland	-2.30 (-0.21)	0.022* (>0.05)	0.21(0.18)	>0.05(>0.05)
	Europe	-2.01 (-2.39)	0.046 (0.018)ˣ	0.35(0.44)	0.046 (0.007)ˣ
	Dipole	0.58 (-1.15)	>0.05 (>0.05)	0.16 (0.29)	>0.05 (>0.05)

*Significance is due to the inclusion of Tambora in the test. ˣSignificance is for cooler than average conditions in Europe, not warmer as is being tested for.

The suggestion of a volcanically forced response is also supported by the MSLP data, which provides compelling evidence that a positive phase of the NAO index is favoured in the one-to-two years following explosive low latitude eruptions. Up to 75% of eruptions in DJF1 and up to 63% of eruptions in DJF2 have a positive NAO index in at least one dataset, despite a positive NAO occurring in only 40-45% of non-volcanic years. A Student's t-test confirms the significance of this finding (see Table 4.3).

Table 4.3. Results of the Student's t-test comparing the NAO indices in post-eruption and non-volcanic years. Values in brackets for 20CRv3 include the period pre-1850, for information.

	t-statistic		***p*-value**	
	DJF1	**DJF1 & 2**	**DJF1**	**DJF1 & 2**
Jones et al. (1997)	1.68	1.98	0.096	0.049
20CRv3 ensemble mean	2.92 (1.60)	2.70 (1.46)	0.004 (0.114)	0.007 (0.145)
ModE-RA ensemble mean	2.17	2.51	0.031	0.013
G_UKESM1.1 ensemble mean	0.17	1.16	0.863	0.246

4.6.3. Is there evidence of a volcanic signal in the G_UKESM1.1 simulations?

Focusing now on the G_UKESM1.1 simulations, the results outlined in Sections 4.4.4 and 4.5.2 clearly suggest there is little evidence of a dynamic winter warming response present in the models, following the majority of eruptions. These findings are confirmed here by an assessment of the ensemble spread and individual ensemble members.

As is noted in Section 4.2.4, each ensemble member in the G_UKESM1.1 simulations represents a different realisation of internal variability, therefore, the spread of the ensemble members can be considered a measure of the internal variability present in the climate system and the ensemble means an approximation of the forced response. With this in mind, a comparison of the ensemble means and spreads in Figure 4.4 and 4.7 clearly demonstrate that any forced response is minimal in comparison to internal variability. Over Greenland the ensemble spread ranges from 2.5°C to 7.6°C, compared to ensemble means of no more than ±1.36°C (excluding Tambora), and over Europe the ensemble spreads range from 2.6°C to 6.7°C, compared to ensemble means of no more than ±0.7°C (following any eruption).

In addition, as noted in Section 4.4.4, there is no consistency in the direction or spatial distribution of the temperature anomalies, and where the ensemble means are largest, they are often of the opposite sign to the anticipated response. This is demonstrated by the

distribution of the ensemble means in Figure 4.10. In contrast to the observations and reanalyses, the distribution of ensemble means for post-eruption winters in G_UKESM1.1 is more closely centred on 0°C in DJF1, as would be expected if no forced response were present, and shifted slightly toward cooler conditions in all regions in DJF2, the opposite of the anticipated dynamic response in Europe. Note the two winters which show significant cool anomalies in Greenland, and a large, positive Europe-Greenland dipole, are the two post-eruption winters following Tambora; these will be discussed further later in this section.

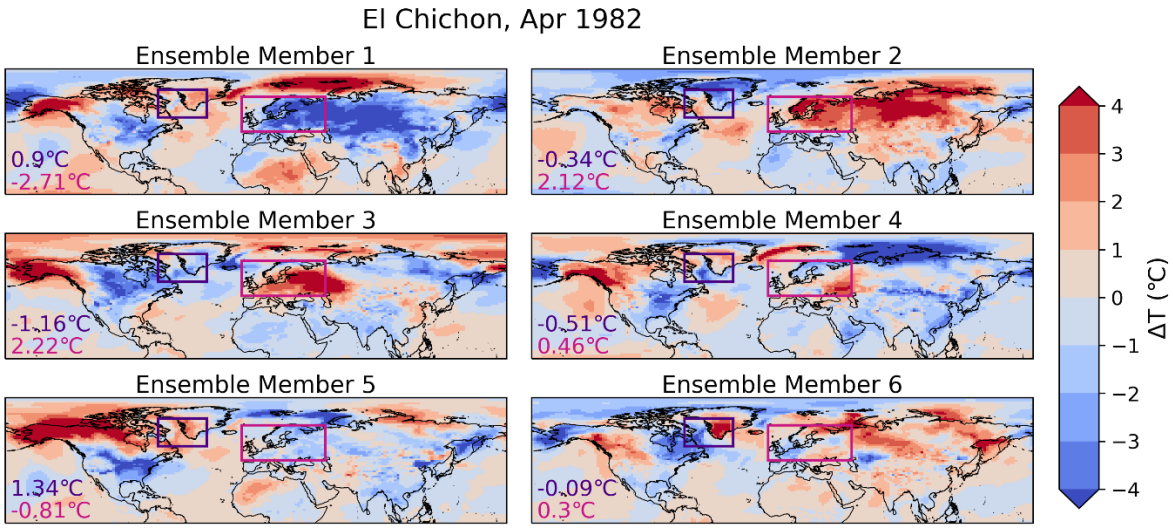


Figure 4.11. Post-eruption surface temperature anomalies in the first winter (DJF1) following El Chichon for each ensemble member in the G_UKESM1.1 dataset. The number in purple (pink) is the area-weighted mean temperature anomaly for the Greenland (Europe) box.

The lack of a consistent or substantial dynamic winter response in the models is further demonstrated by an examination of the post-eruption temperature anomalies in the individual ensemble members, which shows that, under exactly the same volcanic forcing, vastly different post-eruption temperatures can be simulated over Eurasia and Greenland. This is obvious in Figure 4.4, which shows the ensemble spreads generally span 0°C, and in Figure 4.11 which shows that following El Chichon, both unusually cold (member one) and unusually warm (member two) conditions can be simulated over Eurasia, and vice versa over Greenland. This is also true in the MSLP data where both positive and negative NAO indices are simulated by the ensemble members following seven of the eruptions. Thus, both a positive and negative NAO index and warm and cold conditions over Europe and Greenland are consistent with the volcanic forcing present in the model. This strongly suggests that any real-world volcanic effect on the NH temperature or atmospheric circulation is either not sufficiently captured by the model or is overwhelmed by internal variability (for eruptions

of a smaller magnitude than Tambora). This conclusion is supported by a Student's t-test and Kolmogorov-Smirnov test which find no significant difference (in the anticipated direction) between the post-eruption and non-volcanic temperatures in Europe, Greenland or the regional dipole (Table 4.2), as well as a Student's t-test which finds no significant difference between the NAO index in volcanic and non-volcanic winters (Table 4.3).

These results are consistent with the findings of Azoulay et al. (2021), Polvani et al. (2019) and Polvani & Camargo (2020), who do not find evidence in model simulations of a volcanically forced Eurasian winter temperature response (distinguishable from internal variability) following the eruptions of Pinatubo and/or Krakatau, and of Driscoll et al. (2012) who find the same following nine eruptions (including Krakatau, Santa Maria, El Chichon and Pinatubo) in the CMIP5 models. The results do not replicate the significant warming found in model simulations by Zambri & Robock (2016) or Coupe & Robock (2021), respectively; however, the findings are not inconsistent with their results either. Both identified warming further north than the regions studied here, and both assessed ensemble sizes significantly larger than six (thus enhancing the signal-to-noise ratio). Additionally, when smaller eruptions were considered by Zambri & Robock (2016), and when observed sea surface temperatures were not accounted for in Coupe & Robock (2021), the signal was largely lost.

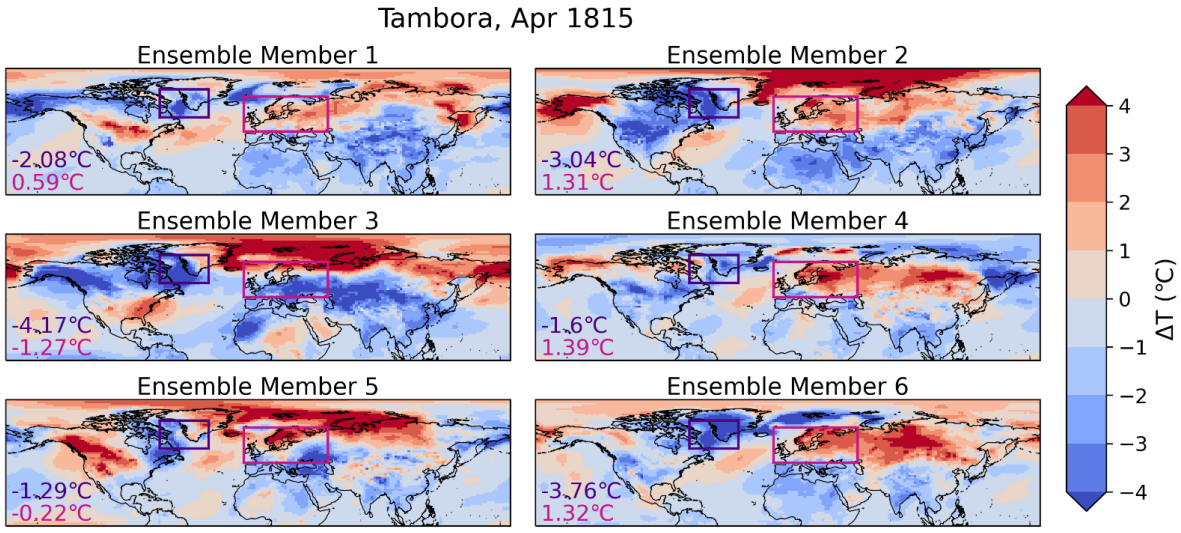


Figure 4.12. As for Figure 4.11, but for Tambora.

The caveat to the findings above is Tambora. The results outlined in Sections 4.4.4 and 4.5.2 suggest there is compelling evidence that the G_UKESM1.1 simulations do produce a volcanically forced dynamic winter warming response following the eruption of Tambora. Tambora is the only eruption where the ensemble members consistently simulate cooler

conditions over Greenland (6/6 ensemble members) and show a tendency for warmer conditions over Europe (4/6 ensemble members) (Figure 4.4; Figure 4.12); where the magnitude and spatial distribution of the temperature anomalies are similar to those seen in the observations (although following other eruptions) and where a positive NAO index is also simulated in the ensemble mean (Figure 4.9), including across all six ensemble members (Figure C8).

The simulation of a potential volcanically forced temperature and circulation response following Tambora is consistent with the findings of Azoulay et al. (2021) and Dallasanta & Polvani (2022) who suggest a volcanically forced response *can* be reproduced by the present generation of models following very large eruptions, even with a relatively small ensemble size. The magnitude of Tambora – which is significantly larger than any of the others assessed here (Figure 4.2b; Figure C1) – also concurs with the thresholds identified by each study: 10Tg(S) based on 100 ensemble members (Azoulay et al., 2021), and 20Tg(S) based on 20 ensemble members (Dallasanta & Polvani, 2022). Tambora injected an estimated 28.1Tg(S) into the stratosphere and is the only eruption assessed here which has a stratospheric sulphur injection of more than 20Tg(S) (Toohey & Sigl, 2017). Thus, these findings suggest the G_UKESM1.1 model is able to reproduce the observed winter warming response following eruptions of the magnitude of Tambora, or larger, but not following eruptions of a smaller magnitude.

It is possible the lack of an apparent volcanic signal following eruptions of a smaller magnitude than Tambora is linked to the signal-to-noise paradox in climate models. The signal-to-noise paradox suggests that the predictable component (signal) in model simulations is erroneously too small relative to the unpredictable component (noise), particularly in the North Atlantic (Eade et al., 2014; Scaife & Smith, 2018; Smith et al., 2020). If the signal-to-noise paradox applies to the modelled response to volcanic forcing, as is suggested by Scaife & Smith (2018), then the role of internal variability may be overestimated in these results and may be overwhelming the volcanic signal following the eruptions of a smaller magnitude than Tambora. This may help to explain the small ensemble mean (forced signal) relative to the ensemble spread (internal variability) evident in Figure 4.4 and will be discussed further in Chapter 5.

4.7. Chapter summary

This study has assessed the evidence for a volcanically forced dynamic winter warming response over Eurasia, following eight of the largest tropical eruptions of the nineteenth and twentieth centuries, using observations, two reanalysis datasets (including a palaeo-reanalysis) and model simulations. In agreement with previous studies, the four datasets do not produce the same results. The assessment of the observations and two reanalyses finds:

a) Eurasian winter warming, consistent with a positive phase of the NAO, occurs in the first post-eruption winter following six eruptions: U1809, U1831, Krakatau, Santa Maria, El Chichon and Pinatubo, and in the second post-eruption winter following seven, including Tambora and Cosiguina.
b) a positive NAO index is favoured in the first and second post-eruption winters. Six eruptions show a positive NAO index in DJF1 (same eruptions as in a)) and five in DJF2 (including Tambora and Cosiguina).
c) the magnitude of the *individual* post-eruption temperature anomalies is generally not unusual (the majority fall within the 1σ or 2σ ranges of internal variability), but the *consistency* across the eight events of positive and negative anomalies in Europe and Greenland, respectively, in either DJF1 or DJF2 is significant in comparison to non-volcanic winters.

In contrast, the assessment of the model simulations finds:

a) the G_UKESM1.1 model does not reproduce a Eurasian winter warming response, similar to that found in the observations and reanalyses, or a shift to a positive phase of the NAO, following seven of the eruptions assessed.
b) G_UKESM1.1 does produce evidence of Eurasian winter warming following the larger eruption of Tambora and this is corroborated by a positive phase of the NAO. The response is reasonably consistent between ensemble members: all six show a positive NAO index and negative anomalies over Greenland and four show positive anomalies over Europe.

Previous studies have concluded that the unexceptional size of the post-eruption temperature anomalies and the lack of a forced response in the models, weakens the evidence for a link between observed Eurasian winter warming and explosive volcanism. While the findings here agree that the models do not reproduce the observed response following the majority of eruptions, and that the post-eruption anomalies are generally not exceptional, it is argued

that these two factors do not preclude the hypothesis that the warming is volcanically forced, for two main reasons. Firstly, the finding that G_UKESM1.1 does produce a potential dynamic winter warming (and NAO) response following Tambora, suggests that there is a causal link between the observed warming and explosive volcanism that can be captured by the model for eruptions of a sufficient magnitude. And, secondly, as the stratospheric pathway essentially excites a natural mode of variability (the NAO), it may be unreasonable to require exceptional post-eruption temperature anomalies or NAO indices as the threshold to conclude the warming is volcanically forced.

Instead, it is argued that it is the consistency of the observed response that is more significant. Despite the uncertainties associated with a) the observations and eruptions themselves; b) the different eruption source parameters (i.e., eruption season, injection height, latitude) and c) the background environmental conditions (i.e. phase of the ENSO and QBO; Table 4.1), all of which can affect the climatic impact of eruptions (Coupe & Robock, 2021; Kirchner et al., 1999; Marshall et al., 2020; Muthers et al., 2014, 2015; Thomas, Giorgetta, et al., 2009; Thomas, Timmreck, et al., 2009; Toohey et al., 2014; Zanchettin et al., 2019), all eight eruptions assessed here present some evidence of a winter temperature and circulation response consistent with a positive phase of the NAO. It is this consistency in the response which is significant and which it is argued here lends support to the earlier conclusions that there is a causal link between the observed winter warming over continental Eurasia and the preceding volcanic eruptions.

In conclusion, and returning to the question originally posed: do the early nineteenth century eruptions support the evidence for volcanically induced winter warming? Taken in isolation, the results of the early nineteenth century eruptions are not sufficient to draw a conclusion; however, taken in combination with the four more recent eruptions, it is suggested the findings of the observations, reanalyses, and the modelled response to Tambora, outlined in this chapter, do lend support to the suggestion that explosive low latitude eruptions lead to warmer winter conditions over Eurasia due to the inducement (or strengthening) of a positive phase of the NAO.

Chapter 5

Summary and Conclusions

5.1. Summary of key findings

5.1.1. Part I: Quantifying the exposure bias in early instrumental temperature records

Exposure biases are a pervasive non-climatic change in land surface air temperature records which have been introduced as a result of changes in the way thermometers have been sheltered from solar radiation and the elements over time. Exposure biases have not been widely accounted for in observational records, particularly on a global scale, due to difficulties detecting and correcting the bias using traditional homogenisation techniques. As a result, exposure biases still contribute significant uncertainty to the early period in global temperature compilations.

Part I of this thesis presented an empirical approach to address the exposure bias arising from the introduction of Stevenson-type screens from the late-nineteenth century. The approach consisted of four main steps: 1) an empirical analysis of 54 parallel measurement series to identify the characteristics of the bias arising from four main classes of historic exposure; 2) the development of three bias-estimation models, based on an analysis of which variables influence the bias in the monthly mean temperature (T_m); 3) the identification of bias-affected stations using exposure metadata and 4) the application of the models to the CRUTEM5 station database, based on the exposure metadata, to quantify and reduce the bias in a version of CRUTEM5 extended back to begin in 1781.

Step one (Chapter 2, Section 2.3) identified differences between the temperatures recorded in Stevenson screens and early exposures which varied seasonally, diurnally, and with location and exposure class. The largest biases were generally present in maximum temperatures and the diurnal temperature range but significant biases, which require consideration before series can be used for climate assessment, were also found in mean

temperatures. The largest biases in mean monthly temperatures were found in freestanding exposures (annual means up to 0.78°C too warm) and in the summer months, while the smallest biases were generally found in wall-mounted exposures (near 0°C annually) and in winter. The biases arising from three classes of exposure were found to have distinct seasonal and diurnal characteristics. The open exposures tended to record warmer maxima and cooler minima than Stevenson screens, with a warm bias in T_m which peaks in summer and is smallest in winter. Wall-mounted exposures generally recorded the opposite: cooler maxima and warmer minima than the Stevenson screen, with a bi-annual cycle in ΔT_m and a near 0°C annual mean bias. In contrast to both, intermediate exposures tended to record warmer maxima and minima than the Stevenson screen, with a warm bias in T_m, which peaks in the summer months (similar to open exposures). The bias arising from closed exposures had similar characteristics to both open and intermediate exposures, with warmer maximum temperatures in closed exposures and a single-peaked warm bias in T_m; however, it is unclear from the parallel measurements analysed here whether the minima in closed exposures are generally warmer or cooler than those recorded in Stevenson screens. The inclusion of additional parallel measurements in Section 2.3 (in comparison to Parker (1994)) highlighted possible regional differences in the magnitude of the exposure bias arising from open (Section 2.3.1) and closed (Section 2.3.4) exposures; confirmed a bi-annual cycle in ΔT_m is a common feature of the bias arising from wall-mounted exposures (Section 2.3.2) and also highlighted possible differences in the seasonal cycle and magnitude of the bias (in T_m) arising from intermediate exposures (Section 2.3.3).

In step two, (Chapter 2, Section 2.4) significant relationships between the bias and the temperature recorded in the historic exposure (T_{Hist}), downward top of atmosphere solar radiation (*TOA*) and/or received shortwave downward solar radiation (*SWD*) were identified in each exposure class and led to the development of three regression-based bias-estimation models. A combination of T_{Hist} and *SWD* were found to produce the best estimates of the bias in open exposures; a quadratic dependence of the bias on *TOA* produced the best estimates for wall-mounted exposures, and *SWD* for closed exposures. This work built on the findings of previous studies which identified possible relationships between the magnitude of the bias, absolute temperature (e.g., Ashcroft et al. (2022); Brunet et al. (2011) and Margary (1924)) and/or solar radiation variables (e.g. Auchmann & Brönnimann (2012); Brunet et al. (2011)), but is, to my knowledge, the first-time these relationships have been

quantified for multiple classes of early exposure using parallel measurement series from a variety of locations globally.

In order to apply the bias-estimation models developed in step two, information regarding which stations and time periods in CRUTEM5_sdb (station database) were affected by the exposure bias was required. Chapter 3 (Sections 3.2 and 3.3) assessed the utility of two approaches to bias identification: a direct approach using metadata and an indirect relative approach using reanalysis and/or climate proxies as reference series. The indirect approach showed promise: both 20CRv3 and three of the proxy reconstructions analysed were able to identify an inhomogeneity in the Berlin Dahlem observed record. However, a number of limitations associated with each data source, as well as the inability for the comparisons, in isolation, to definitively establish the cause of the breakpoint, meant neither met the requirements for use in this thesis. Step three (Chapter 3, Section 3.4), therefore, led to the development of a database of historic exposures for CRUTEM5_sdb. The database contains at least one entry of exposure metadata for 88% of stations within 30° to 60° latitude (with observations before 1961) and represents one of the first attempts to collate exposure information relevant to individual stations and countries/Meteorological Services, at a near-global scale.

Application of the bias-estimation models to 1,960 mid-latitude stations in an extended version of CRUTEM5 (CRUTEM5_ext), in step four (Chapter 3; Section 3.5), resulted in small (≤0.016°C) positive adjustments to the Northern Hemisphere mid-latitudinal mean before 1880, and larger (≤-0.1°C) negative adjustments to the Northern and Southern Hemisphere mid-latitudinal means between 1882–1934 and 1856-1900, respectively. The largest hemispheric adjustments were made to summer temperatures in the Northern Hemisphere between 1882-1934 (~0.2°C) and larger adjustments were estimated regionally: up to -0.57°C annually and -0.79°C seasonally in individual grid cells. The applied adjustments mean the exposure bias has now been minimised at 75.1% of mid-latitude stations in CRUTEM5_eba, representing 86.3% of the mid-latitude monthly values. This represents a significant improvement on the proportion of unaffected stations prior to the application of the bias-estimation models - just 37.7% of mid-latitude weather stations in CRUTEM5_ext were not affected by the exposure bias pre-application, either because no transition to a Stevenson screen occurred during the record or because (for 1.5%) the record had already been adjusted.

Annually, and over large spatial scales, the bias adjustments produced here and the representation of the bias in the HadCRUT5 error model show reasonable agreement - the mid-latitude annual mean biases estimated for each hemisphere fall within the range of the error realisations applied to HadCRUT5 (Figure 3.14). However, the estimates produced in this thesis refine the current representation of the bias seasonally, temporally and at a regional/station-specific scale as the estimates account for seasonal differences in the magnitude of the exposure bias, differences in the magnitude and seasonal structure between (and within) exposure classes and spatio-temporal differences in the use of historic exposures and the timing of the introduction of the Stevenson screen. This refinement is highlighted by the monthly-resolved and time-varying adjustments produced here in comparison to the static annual nature of the HadCRUT5 realisations (Figure 3.14), as well as by the spatially varying magnitude of the bias adjustments applied (Figure 3.15), compared to the blanket latitude-based approach applied in HadCRUT5. This work represents the first-time exposure bias adjustments have been applied to LSAT series on a near-global scale using only station-specific metadata and empirical models. The approach has been demonstrated using CRUTEM5, however, the models and steps applied here are applicable to any monthly mean global LSAT compilation.

5.1.1.1. Key outputs from Part I of the thesis:

- A newly collated dataset of 54 parallel measurement series which can be used to assess the bias introduced by the transition from four main classes of historic exposure to variants of the Stevenson screen.
- Three models which produce skilful and robust estimates of the monthly mean exposure bias introduced by the transition from open, wall-mounted and closed exposures to variants of the Stevenson screen.
- An ‘Historic Exposures Database’ for CRUTEM5_sdb which details the exposures in use at individual stations and meteorological networks prior to the introduction of the Stevenson screen.
- Monthly-resolved, exposure-specific, metadata-based estimates of the exposure bias present at individual stations in the CRUTEM5 station database. These estimates have been incorporated into an extended global surface air temperature dataset: GloSAT (Morice *et al*. [in prep]), which was analysed in Part II of the thesis.

Elements of Part I of this thesis have been published in the *International Journal of Climatology:*

> Wallis, E.J., Osborn, T.J., Taylor, M., Jones, P.D, Joshi, M., Hawkins, E. (2024). 'Quantifying exposure biases in early instrumental land surface air temperature observations.' *International Journal of Climatology, 44*(5), 1611-1635. *https://doi.org/10.1002/joc.8401*

and the parallel measurement data and exposure metadata, bias estimates and uncertainty produced for mid-latitude stations in CRUTEM5_sdb are both openly accessible via the Zenodo data repository: https://zenodo.org/doi/10.5281/zenodo.10551235 and https://zenodo.org/doi/10.5281/zenodo.10551196, respectively.

5.1.2. Part II: Exploring natural climate variability in early instrumental temperature records

Part II of this thesis explored a key driver of natural climate variability in the early instrumental period: volcanic eruptions. Specifically, the evidence for (or against) volcanically induced winter warming over Eurasia following explosive low latitude eruptions was assessed using four sources of surface temperature data: *in situ* instrumental observations, two reanalyses (including a palaeo-reanalysis) and an Earth System Model. The land component of the observational dataset used in Part II incorporated the exposure bias adjustments developed in Part I of this thesis.

Early observation- and model-based studies linked post-eruption winter warming over Eurasia to volcanic aerosols via a dynamic circulation response referred to as the 'stratospheric pathway' (Graf et al., 1993; Kelly et al., 1996; Robock & Mao, 1992, 1995). Present-generation models, however, have struggled to replicate the observations (e.g., Driscoll et al. (2012)) and recent studies have questioned the role eruptions play in the observed winter warming (Polvani et al., 2019; Polvani & Camargo, 2020). The majority of recent studies have either been model-based or have assessed the observed response following the more recent (and often smaller) volcanic eruptions such as El Chichon (1982) and Pinatubo (1991), which are covered by existing global temperature compilations. The recent development of a gridded observed global surface air temperature dataset, starting in 1781 (GloSAT; Morice et al. [in prep]), provided a new opportunity to assess additional eruptions. Chapter 4, therefore built upon earlier studies by assessing the temperature and circulation response to the eight largest low latitude eruptions of the nineteenth and twentieth

centuries to determine whether the (generally larger) early nineteenth century eruptions provide evidence for volcanically induced Eurasian winter warming. The assessment included four eruptions: U1809, Tambora, U1831, and Cosiguina which had not previously been assessed using a global compilation of observed temperatures and was complemented by two reanalysis datasets and an ensemble of Earth System Model (G_UKESM1.1) simulations initialised in 1750 that was provided by the GloSAT project.

The assessment of the observations and reanalyses found evidence of a winter temperature and circulation response, consistent with a positive phase of the NAO, for each of the eight eruptions assessed. Anomalously warm winters in Europe combined with a positive phase of the NAO index were found in the first post-eruption winter following six eruptions: U1809, U1831, Krakatau, Santa Maria, El Chichon and Pinatubo, and in the second post-eruption winter following five, including Tambora and Cosiguina. The magnitudes of the individual post-eruption anomalies were not generally 'exceptional' or outside the range of internal variability, however, the consistency of the response following all eight eruptions, across two post-eruption winters, was found to be significant in comparison to non-volcanic winters. This consistency, despite the varying background conditions at the time of the eruptions and the varying eruption characteristics (see Table 4.1), was considered one of the most important results arising from the assessment in Chapter 4.

As in previous studies, the G_UKESM1.1 model did not replicate the observed and reanalysed conditions for seven of the eruptions studied; however, evidence of Eurasian winter warming (and Greenland cooling) and a positive phase of the NAO was found following Tambora, the largest eruption assessed. This is potentially significant as it suggests there may be a causal link between the observed winter warming and explosive volcanism that can be captured by the present generation of models, but only for eruptions of a sufficient magnitude. This concurs with the latest studies (e.g., Azoulay et al. (2021)) and could also be related to the signal-to-noise paradox in models (discussed in Section 4.6.3 and further in Section 5.2.2 below).

Overall, the comprehensive assessment conducted in Chapter 4 lends weight to the theory that explosive low latitude volcanic eruptions lead to warmer winter temperatures over Eurasia due to the inducement (or strengthening) of a positive phase of the NAO, and the inclusion of the eruptions which occurred in the early instrumental period in this assessment strengthens this evidence.

5.1.2.1. Key outputs from Part II of the thesis:

- The first assessment of the observed post-eruption winter temperature anomalies in Greenland and Europe following the early nineteenth century eruptions conducted using a global dataset of gridded air temperatures over the land and ocean (GloSAT). This dataset incorporates the exposure-bias-adjusted data produced in Part I of this thesis.
- Evidence from observations and reanalyses of winter warming (in DJF1 and/or DJF2), consistent with a positive phase of the NAO, following eight explosive low latitude eruptions, including the four large early nineteenth century eruptions.
- The first assessment of the modelled post-eruption winter temperature anomalies and NAO index using a version of the UKESM1.1 model initialised in 1750.

Elements of Part II of this thesis are being prepared for submission to *Climate of the Past*.

5.2. Limitations and potential for further research

5.2.1. Part I: Quantifying the exposure bias in early instrumental temperature records

One of the key limitations of Part I of this thesis is the number of parallel measurement series that are available for analysis. Although the 54 series analysed here (of which 31 were used to inform the bias-estimation models) represent an improvement on those available to Parker (1994), the number is still relatively low and the series available are geographically limited. Fewer than eleven parallel measurement series informed each bias-estimation model and the majority of series analysed were located in Europe: only six of the 54 collated series were located elsewhere (Figure 2.5). The inclusion of additional parallel measurement series would help to: a) refine the relationship between the bias and the relevant predictor(s), b) constrain (or more accurately capture) the uncertainty associated with the bias estimates and c) would potentially enable more sophisticated testing of the models during development, for example, by allowing more data to be held back for validation purposes. The addition of parallel measurement series from a broader geographical area would also help to confirm that the identified relationships between the bias and the predictor(s) remain stable across space.

The availability of additional parallel measurement series could also facilitate the development of a more robust bias-estimation model for intermediate exposures. The model developed in Section 2.4.3 is based on only three parallel measurement series and is not

robust enough to be used to estimate the bias at stations in CRUTEM5_sdb. This is a key limitation of Part I of the thesis which prevented the adjustment of more than 40 series in CRUTEM5_ext (Table 3.5) and which needs to be addressed in future work.

The addition of parallel measurement series is not an unrealistic prospect. Since the completion of this work, additional parallel measurement series have come to light, and it is likely there are additional series 'hidden' in archive material. Ongoing data rescue initiatives, such as the Atmospheric Circulation Reconstructions over the Earth initiative (Allan et al., 2011), and work by organisations such as the National Oceanic and Atmospheric Administration (NOAA) Central Library and the National Meteorological Archive in England (Ross, 2023) to scan their archives provide a realistic opportunity for additional parallel measurement series to be discovered. Additionally, if resources were made available, there is also the potential to conduct contemporary comparisons between exposures, as done by the SCREEN project (Brunet et al., 2006, 2011).

Another key limitation of the approach presented in Chapters 2 and 3 is that the bias-estimation models are applied to stations irrespective of the method used to calculate the daily-mean, despite the models being based on parallel measurements using $^1/_2(T_x+T_n)$ exclusively. This is a limitation because previous studies, plus analyses conducted using the AEMET La Coruna and Murcia series (see Appendix A), demonstrate differences in the magnitude of the exposure bias dependent on the method used. Böhm et al. (2010), for example, found the June exposure bias in wall-mounted exposures to be 0.6°C smaller when using $^1/_2(T_x+T_n)$ to calculate the daily-mean compared to using observations at fixed hours ($^1/_3$(6h+13h+20h)), while the analysis shown in Figure A1 finds the June bias in open exposures to be approximately 0.15-0.4°C larger when using $^1/_2(T_x+T_n)$ (compared to fixed hours: $^1/_3$(6h+13h+20h)). At present, the parallel measurement series available for analysis prevent the development of models using alternative methods of daily-mean calculation and the lack of collated metadata regarding the methods in use would prevent their application. However, as the use of $^1/_2(T_x+T_n)$ was not predominant globally in the early instrumental period - it was primarily used in English-speaking countries and former British colonies, (Thorne et al., 2016; Trewin, 2004, 2010) - future work should consider incorporating this uncertainty into future quantifications of the bias or developing approaches to apply method-specific adjustments.

The exposure bias metadata present another limitation to this study. As is noted in Chapter 3, the availability of metadata determines whether a series can be adjusted, and the accuracy of the exposure adjustments are dependent on the accuracy of the metadata. If the information in the metadata database is inaccurate - for example, if a weather station used a different exposure to that in use in the rest of the country/NMS - then the bias estimates and adjustments applied will also be inaccurate. Due to the availability and accessibility of metadata, as well as time considerations, the exposures database contains primarily nation-, rather than station-, specific metadata and is not complete (Figure 3.13, Table 3.5, Appendix B); however, it is hoped that the database will be expanded and improved upon by future work. Any improvement in the accuracy of the metadata should lead to an improvement in the bias estimates and the expansion of the database to include new metadata may allow the bias to be estimated at additional stations, thus improving the overall quantification of the bias in global temperature datasets.

In addition to improving and expanding the metadata database, it may also be possible to improve the application of the models (and thus the bias estimates) by incorporating relative breakpoint detection into the bias identification process. There are two key areas where this may be particularly beneficial: a) to confirm the existence of a break identified by the metadata and b) to isolate breakpoints within a transition period. Starting with a), exposure metadata are not always accurately documented, and, in some instances, exposure metadata may have been incorrectly assigned to stations in the historic exposures database due to changes in station names over time or the existence of multiple closely located observing stations (particularly in the very early period). Additionally, it is not always known whether exposure adjustments have already been applied to observations. For example, the bias may have been accounted for by observers at the time of observation, but not explicitly documented, or (unknown) contemporary efforts may have addressed the bias. In such instances it would therefore be beneficial to employ relative breakpoint detection methods to confirm the existence of a break in the data prior to the application of any adjustment(s).

In terms of point b), the majority of source code ‘3’ (country-specific) information used to populate the historic exposures database (Appendix B), does not pinpoint the introduction of new exposures to one year, but designates a transition period over several years. At stations in India for example, the transition period between the intermediate and Stevenson screen exposures is six years (Section 3.4.2.1). In this study, the decision was made to apply the bias-estimation models to the end of the transition period (i.e. for the purposes of the

model application, the early exposure was assumed to be in place until the end of the transition period); however, the incorporation of relative breakpoint detection could help to isolate the timing of the change at individual stations. This would help to reduce the step-changes notable in the hemispheric adjustments in Figure 3.14 and would ensure observations were not adjusted during periods when a Stevenson screen was in use.

In addition to addressing the limitations discussed above, a few possible avenues exist to expand the research undertaken in Part I.

Additional evaluation of the exposure bias adjusted data

The assessment of the exposure bias adjustments conducted in this thesis focussed primarily on the evaluation of the bias-estimation models themselves and comparisons with other (primarily large scale) estimates of the exposure bias. As a result, future assessment of the individual station adjustments would be beneficial. As few independent estimates of the exposure bias exist for comparison, this could be achieved using methods similar to those outlined in Section 3.2.2. For example, for selected stations the adjusted and unadjusted data could be compared with proxy reconstructions and/or reanalysis to assess whether the adjusted station data agrees more closely than the unadjusted data. Alternatively, neighbouring stations which received different bias adjustments (such as those either side of Meteorological Service borders) could be compared to see if the temperatures at each station are more closely correlated following adjustment for the exposure bias.

Further work to assess the early instrumental observations for random inhomogeneities following the application of the bias adjustments would also be worthwhile, as minimizing the effect of the exposure bias may make other inhomogeneities more detectable (World Meteorological Organization, 2020). The application of homogenization techniques, such as Pairwise Homogeneity Assessment (PHA) (Menne & Williams, 2009; Williams et al., 2012), post-adjustment could therefore help to further homogenize the data before its inclusion in global temperature compilations.

Exposure biases in the tropics and high latitudes

One of the key remaining uncertainties in the work presented here is the impact of the exposure bias on temperature series located outside of 30° to 60° latitude. The bias-estimation models presented in Chapter 3 were designed for application in the mid-latitudes, only, as an insufficient number of parallel measurement series were available in the tropics

and high latitudes (Figure 2.5) to determine whether the models were also applicable to those regions. Only one (intermediate) parallel measurement series was available for analysis in the tropics (Field, 1920), and in the high latitudes the only series available were located in Iceland (Veðráttan, 1962), did not use $^1/_2(T_x+T_n)$ to calculate the daily-mean, and were exclusively for wall-mounted exposures. As a result, it was not possible to determine whether the relationships identified between ΔT_m and temperature, *SWD* and/or *TOA* in Section 2.4 also applied to the tropics or whether factors such as snow cover (in the high latitudes) or the angle of insolation would alter the relationships. As a result, the work presented in Part I does not address the exposure bias in the tropics or high latitudes and this requires additional attention.

Annual estimates of the exposure bias in the tropics and high latitudes have been produced previously by Parker (1994) and Folland et al. (2001) (which the current exposure bias term in the HadCRUT5 error model is based on). However, improving upon these estimates - particularly in terms of quantifying the bias at individual weather stations and on a monthly timescale - is likely to be challenging without the availability of additional parallel measurement series. As noted earlier, however, there are likely to be parallel measurements 'hidden' in (as yet) undigitized or unprocessed archive materials which may become available in future. Awe et al. (2022), for example, recently uncovered and analysed a long series of parallel measurements from Mauritius and data rescue initiatives (e.g., Allan et al. (2011)) provide the opportunity for additional series to come to light. At least one contemporary comparison between exposures is also known to be ongoing in the tropics (in Hong Kong; V. Venema, pers. comm.). Thus, a key aim of future work should be to refine estimates of the exposure bias for weather stations located in the tropics and high latitudes. This would allow a more complete exposure bias adjustment of global temperature compilations, such as CRUTEM5/HadCRUT5.

Exposure biases arising from (very) early thermometer exposures and the transition to Automatic Weather Stations

This study has been primarily concerned with quantifying the bias arising from the transition to Stevenson screens; however, other changes in exposure have also occurred throughout the instrumental period, including transitions from (very) early exposures (such as thermometers hung in well-ventilated, unheated rooms, Jurin (1723)) to the four categories of exposure defined in Chapter 2 (Section 2.2), and more recent transitions from the Stevenson screen to

Automatic Weather Stations (AWS) with temperature sensors exposed in multi-plate shields (and other newer forms of exposure) in the 1980s (Trewin, 2010). To fully account for exposure biases in global temperature compilations these changes must also be accounted for.

Few studies have addressed the exposure bias arising from very early exposures; however, it is likely thermometers exposed inside would have had a reduced *DTR* and mean temperatures which were biased warm. Analysis by Camuffo, della Valle & Becherini (2023), for example, compared the measurements taken inside and outside on snowy days in two locations in Italy, and found the outside thermometer recorded predominantly negative and near-zero temperatures, whereas the thermometer exposed indoors recorded primarily positive values. Thus, any temperature series which contains observations from thermometers exposed indoors is likely too warm and this bias must be accounted for in future work.

The introduction of automatic weather stations in the 1980s have also been shown to introduce exposure biases into temperature records. Quayle et al. (1991), for example, found an annual bias in the mean temperature of 0.06°C (AWS minus Stevenson screen) at weather stations in the United States and Brandsma and van der Meulen (2008) found an annual mean bias of 0.032°C at De Bilt in the Netherlands (Stevenson - AWS) following the introduction of AWSs. Although smaller than the bias arising from the transition to the Stevenson screen, for long series to be completely comparable throughout their records (in regard to exposure), future work should also account for this later change. Parallel measurement series related to the transition to AWSs have previously been collated as part of the (discontinued) International Surface Temperature Initiative Parallel Observations Science Team AWS initiative (Thorne & Willett, 2015; Venema et al., 2014) and efforts now continue under the C3S Global Land and Marine Observations Database project. Therefore, future analyses could take advantage of these parallel measurements, along with existing work, to estimate the bias arising from the transition to AWSs. This would allow the complete assessment of the impact of exposure changes in global temperature compilations.

5.2.2. Part II: Exploring natural climate variability in early instrumental temperature records

A number of questions have arisen from Part II of this thesis which would benefit from further research. One of the most intriguing findings in Chapter 4 is the presence (in the

reanalyses and observations) of a winter temperature and circulation response in the second post-eruption winter following Tambora, but not in the first (Section 4.4.5). This finding is interesting because, to my knowledge, a similar (temperature) response has not been identified previously. Section 4.6.1 proposed some potential explanations for the delayed response - related to the quantity, characteristics and distribution of the aerosol particles, as well as the temporal proximity of the preceding eruptions - however, additional research in this area would be beneficial to improve scientific understanding of the impact of very large eruptions on atmospheric circulation and climate, and to improve seasonal forecasts following eruptions.

Chapter 4 also finds marked differences between the magnitude of observed winter warming and the magnitude of the preceding eruption (in terms of Stratospheric Aerosol Optical Depth, SAOD). While this concurs with previous studies which find the relationship between SAOD and climate impact to be non-linear (Raible et al., 2016; Timmreck, 2012); an interesting aspect evident in Figure 4.4 is the difference in the magnitude of the observed post-eruption winter warming between the early nineteenth century eruptions and the more recent eruptions. The magnitude of the observed post-eruption winter warming is greater following the more recent eruptions than following the four early nineteenth century eruptions, despite the SAOD of the latter being generally larger (Figures 4.4 and C1). In DJF1, for example, El Chichon and Pinatubo have a NH SAOD of 0.091 and 0.13 and European warm anomalies of 1.86°C and 1.46°C, respectively, while the U1809 eruption has a larger NH SAOD of 0.221 in DJF1 but a substantially smaller European warm anomaly of only 0.47°C. A number of factors could help to explain the differences. Previous studies have shown the climate impacts of volcanoes and/or the strength of the stratospheric polar vortex can be influenced by the eruption season, latitude and injection height (Marshall et al., 2019; Robock, 2000; Robock & Mao, 1995; Timmreck, 2012; Toohey et al., 2011, 2014) the phase of the ENSO (Bittner et al., 2016; Calvo et al., 2008; Mao & Robock, 1998; Robock & Mao, 1995) and the QBO (Stenchikov et al., 2004; Thomas, Giorgetta, et al., 2009), as well as factors such as atmospheric chemistry (Muthers et al., 2014, 2015; Rozanov et al., 2002; Solomon, 1999). Additionally, it is possible that elements of the stratospheric pathway are being altered by climate change (Aubry et al., 2021). Recent studies, for example, suggest the teleconnection between the QBO and the polar vortex may be changing due to a warmer climate (Karami et al., 2023) and there is evidence the NAO is becoming more extreme (Hanna et al., 2015, 2022). Further research into this area, both to explore why

the dynamic winter warming response appears to be stronger per unit of aerosol following the more recent eruptions, and to assess the impact of the background environmental conditions on volcano-climate impacts would therefore be beneficial. Further research in this area may also help to inform or constrain statistical emulators, such as the one developed recently by Marshall et al. (2019).

While beyond the scope of this thesis (which focusses on the instrumental data), another interesting avenue for future research would be to look in more detail at the modelled response to volcanic eruptions in the G_UKESM1.1 model (and in other climate models). One of the key limitations to the assessment of the model simulations was the relatively small ensemble size of six. Bittner et al. (2016) recommend the use of at least 15 ensemble members to distinguish the dynamic winter response to explosive volcanism from internal variability, therefore, if additional computing time were made available, it would be valuable to re-run the analysis with additional ensemble members. Additional model runs with interactive volcanic aerosols would also be of interest, due to findings which suggest this factor (i.e. the accurate evolution, size distribution and transport of aerosols) is important for models to accurately replicate climate impacts (Mann et al., 2015; Timmreck, 2012).

Additional analysis of the models could also expand the number of variables assessed. Here the focus has been on temperature and mean sea level pressure, however, as the post-eruption winter warming over Eurasia is theorised to be the result of a dynamic stratospheric pathway, the response of other key elements of this pathway could also be assessed. For example, recent modelling studies have also looked at whether the models simulate an increased stratospheric temperature gradient (Polvani et al., 2019), a strengthening of the polar vortex (via the zonal mean zonal wind anomaly at 60°N and 10hPa) (Azoulay et al., 2021; Dallasanta & Polvani, 2022; Polvani et al., 2019; Polvani & Camargo, 2020; Toohey et al., 2014) and whether there is a correlation between the strength of the latter and Eurasian surface temperatures (Azoulay et al., 2021; Polvani et al., 2019).

Finally, a potential area for further research would be to look at the effect of the signal-to-noise paradox on the results obtained here. The signal-to-noise paradox suggests that the magnitude of the predictable component (the signal) is erroneously small in model simulations, particularly in the North Atlantic (Eade et al., 2014; Scaife & Smith, 2018; Smith et al., 2020). If this signal-to-noise paradox also applies to the modelled response to external forcing, then it could help to explain some of the discrepancy between the modelled

and observed responses to the eight volcanic eruptions found in Chapter 4 (and in previous studies). Future work could therefore assess whether this is the case and, if so, attempt to resolve it (i.e., via the approach suggested by Eade et al. (2014)). The discrepancy between the models and observations is also a broader question which remains open and requires further research.

5.3. Concluding remarks

Now, more than ever, it is crucial to better understand our changing climate; one way this is being realized is through the temporal extension of global surface air temperature compilations via the incorporation of early instrumental records. Early instrumental records of LSAT are hugely valuable resources for climate science as they allow climate variability and change to be assessed over long (centennial) timescales (including periods with little anthropogenic influence), thus facilitating an improved understanding of the climate system, its internal variability and how it interacts with natural and anthropogenic forcing. However, early temperature records can be compromised by the influence of non-climatic changes, such as the exposure bias, and thus can only be used for climate science applications where they have been assessed for homogeneity issues and where adjustments or estimates of uncertainty have been applied accordingly.

This thesis makes a key contribution to both the assessment and exploitation of early instrumental records. Part I of the thesis develops and applies an improved approach to account for the exposure bias in early instrumental LSAT observations and Part II exploits the adjusted data (in the form of an extended global surface air temperature dataset) to explore natural climate variability in the nineteenth and twentieth centuries, specifically the influence explosive volcanic eruptions had on winter climate. Part II of the thesis, and the findings therein, are made possible only by the incorporation of early instrumental records into global temperature compilations, which is itself made possible only by continual improvements to the way biases, errors and uncertainties are addressed in early instrumental records and global temperature compilations.

Appendix A

Supplement to Chapter 2

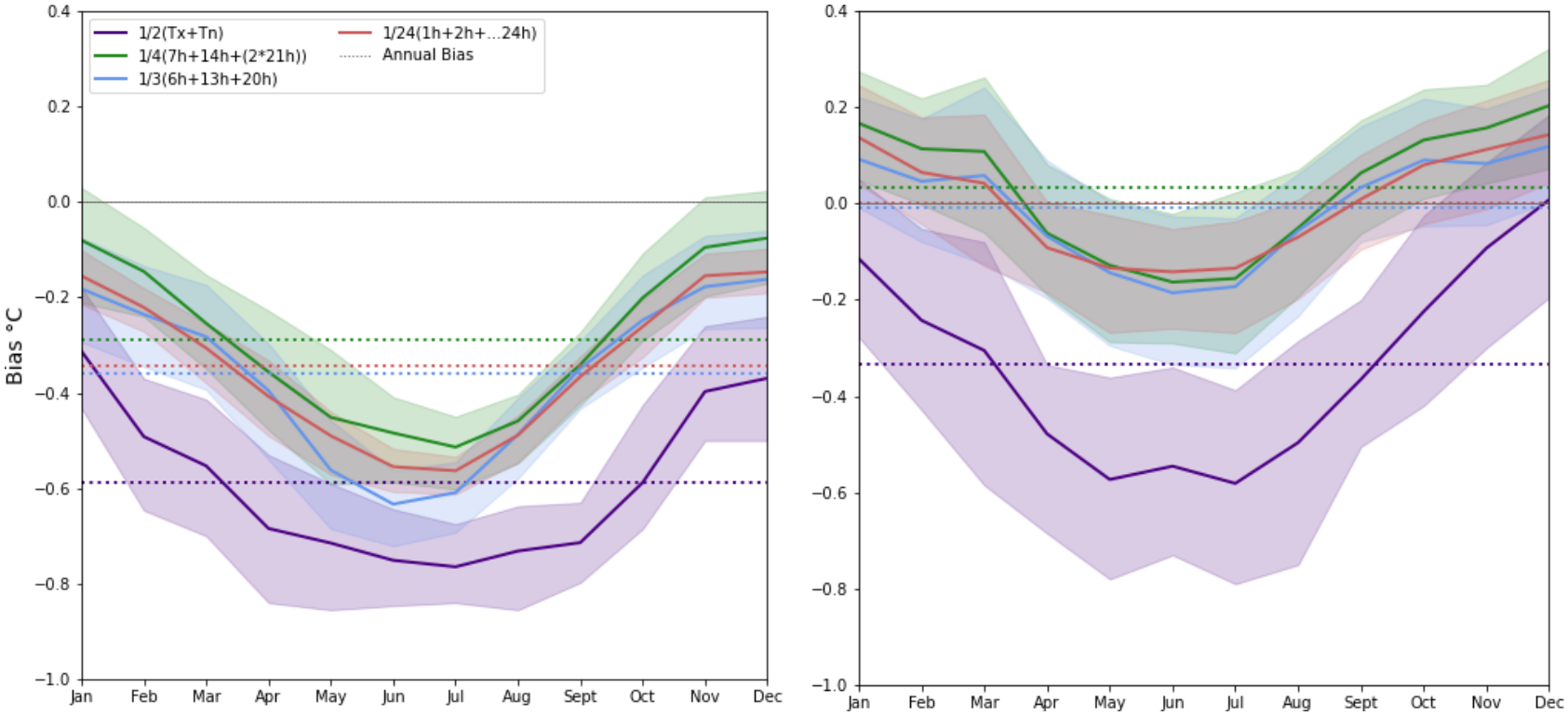


Figure A1. Influence of the daily-mean temperature calculation on the magnitude of the exposure bias, illustrated using the AEMET series from Murcia (left-hand panel) and La Coruna (right-hand panel) (M. Brunet, pers. comm.). The solid lines and shading show the means and ranges of the individual monthly biases and the dashed lines show the annual mean exposure biases (Stevenson screen minus open exposure).

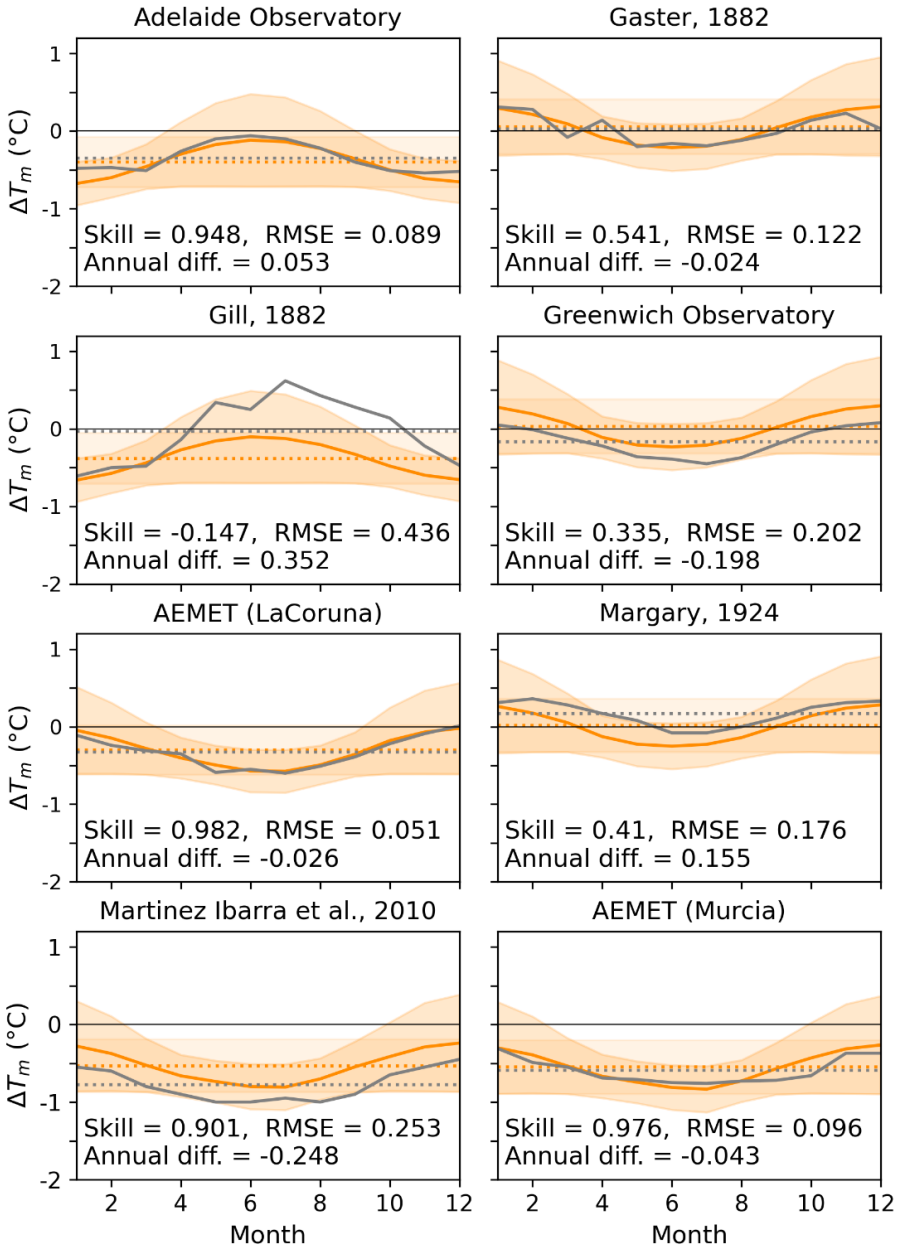


Figure A2. Observed (grey) and estimated (orange) ΔT_m with shaded 95% confidence interval for each parallel measurement series used to develop the open exposure bias-estimation model. Observed (grey) and estimated (orange) annual mean biases are given by the dotted lines.

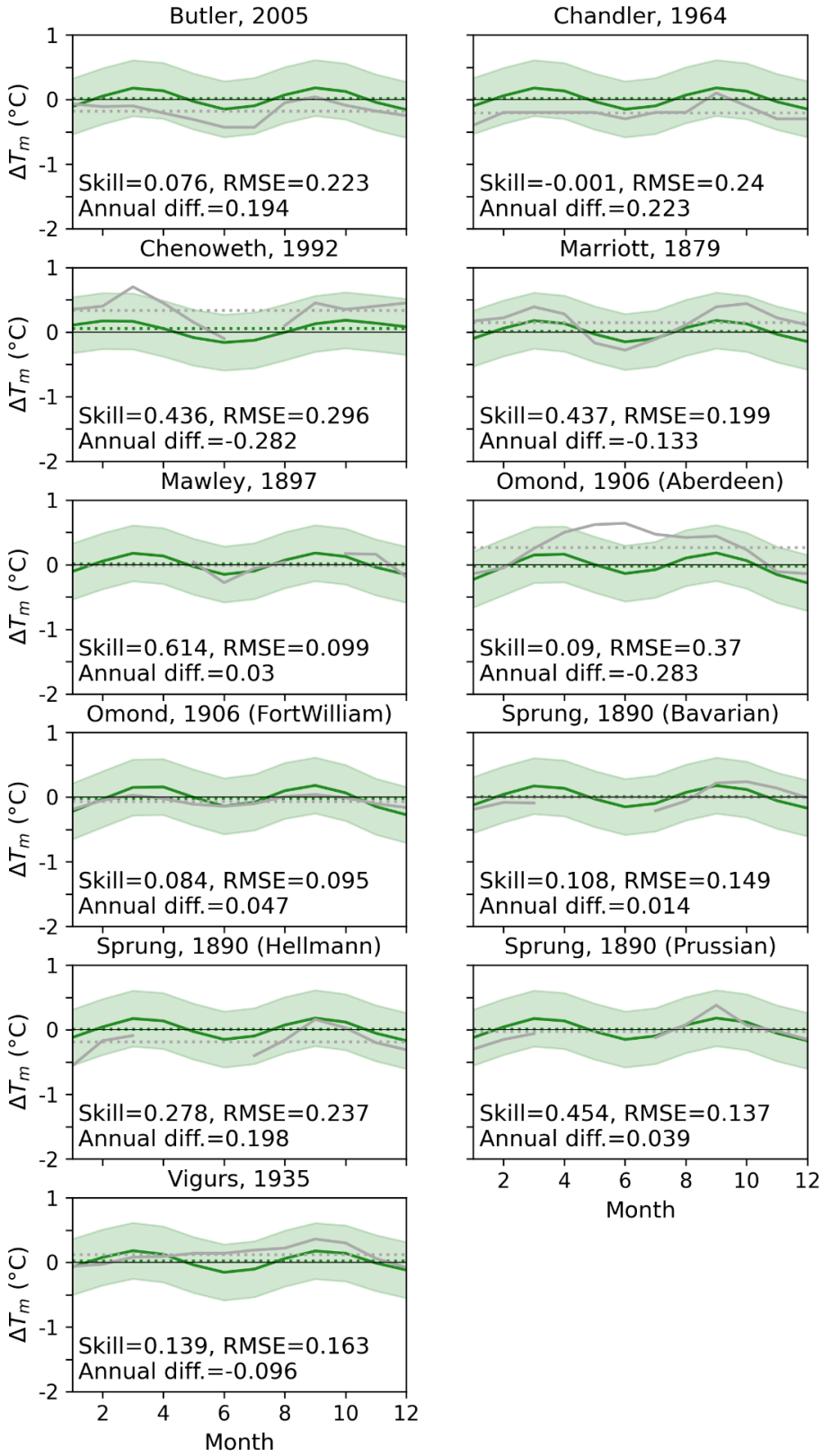


Figure A3. Observed (grey) and estimated (green) ΔT_m with shaded 95% confidence interval for each parallel measurement series used to develop the wall-mounted exposure bias-estimation model. Observed (grey) and estimated (green) annual mean biases are given by the dotted lines.

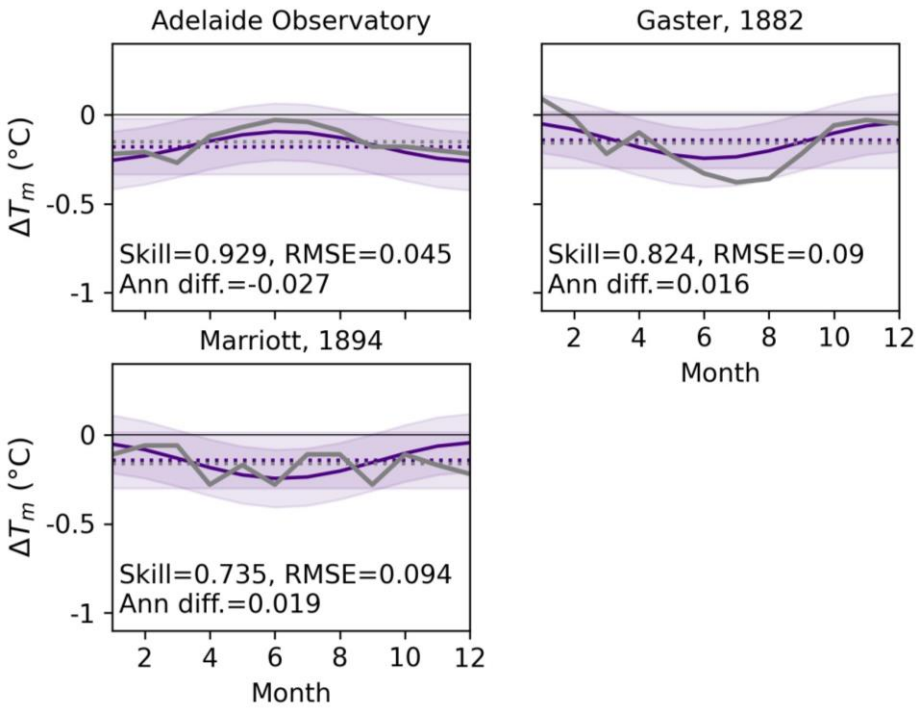


Figure A4. Observed (grey) and estimated (purple) ΔT_m with shaded 95% confidence interval for each mid-latitude intermediate parallel measurement series. Observed (grey) and estimated (purple) annual mean biases are given by the dotted lines. Estimates shown used *TOA* as the predictor.

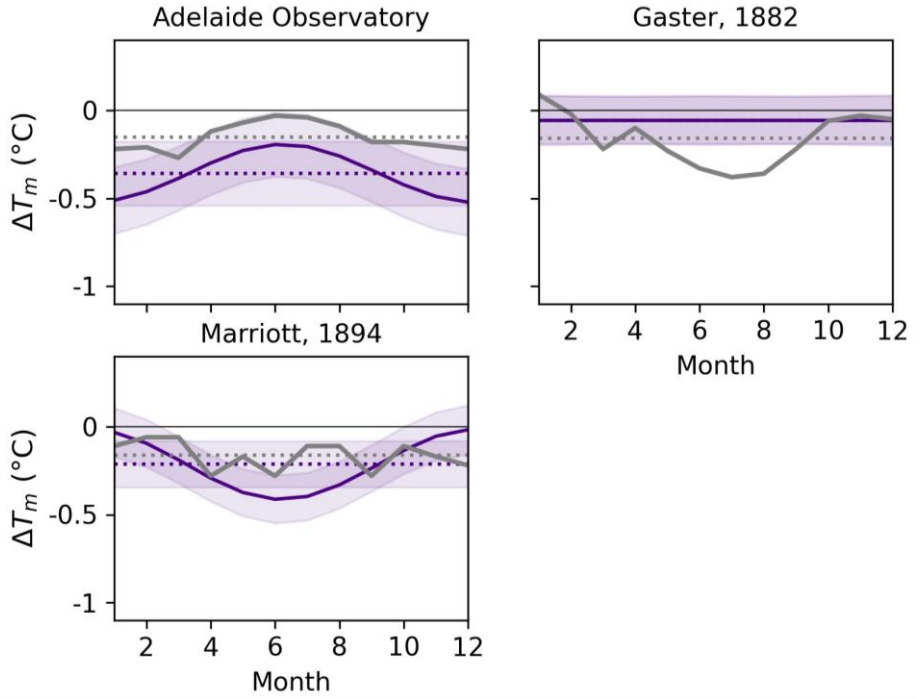


Figure A5. As in Figure A4, except the estimates shown for each parallel measurement series were calculated excluding that series from the model development. Note the lack of seasonal cycle in 'Gaster, 1882'.

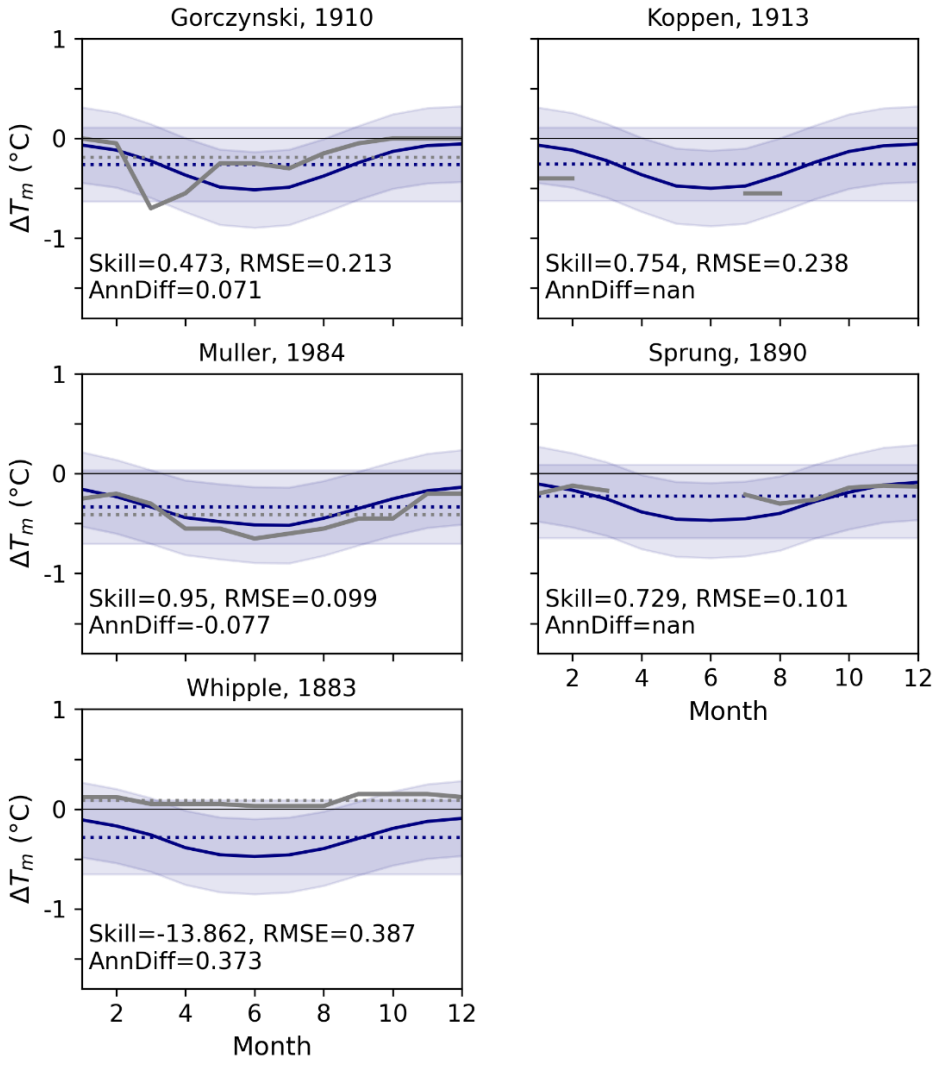


Figure A6. Observed (grey) and estimated (navy) ΔT_m with shaded 95% confidence interval for each parallel measurement series used to develop the closed exposure bias-estimation model. Observed (grey) and estimated (navy) annual mean biases are given by the dotted lines.

Appendix B

Supplement to Chapter 3

This appendix provides additional information about the collation of the 'Historic Exposures Database', including:

- Details of some of the key archives accessed to obtain exposure metadata (Table B1);
- An outline of the higher-level country- or Meteorological Service-specific (source code '3') metadata which were applied to stations without station-specific metadata in order to populate the database (Sections B.1-B.81);
- Details of useful sources of station-specific exposure metadata for each country (Sections B.1-B.81).

All countries (or overseas territories) which have stations located within 30° to 60° latitude (North or South of the equator) with observations prior to 1961 are included in this appendix.

Table B1. Details of some of the key archives accessed to obtain exposure metadata, including examples of the sources each contained.

Archive	URL
ALOCLIM (Austrian-Central European long-term climate) Metadata [CD]	
Bibliotheque Nationale de France • La Météorologie • Meteorologie du Bresil	https://gallica.bnf.fr/
British Geological Survey Geomagnetic Observatory Yearbooks	https://geomag.bgs.ac.uk/data_service/data/yearbooks/yearbooks.html
Climatic Research Unit Library • Word Weather Records • (UK) Quarterly Weather Report	https://crudata.uea.ac.uk/cru/library/datasets.htm
Danish Meteorological Institute (DMI) Publications • DMI Historical Climate Data Collection reports	https://www.dmi.dk/publikationer
HathiTrust • Deutsches Meteorologisches Jahrbuchs	https://www.hathitrust.org/

• India Weather Review • Meteorologische Zeitschrift (published by the Austrian and German Meteorological Societies) • Reports of Meteorological Congresses	
Internet Archive • Report on the Meteorology of India • The Climate of Japan	https://archive.org/
KNMI Library • Het Klimaat van Nederland	https://www.knmi.nl/knmi-bibliotheek/
Met Eireann • Metadata for long term datasets	https://www.met.ie/climate/available-data/long-term-data-sets
Met Office Digital Library and Archive • Instructions in the use of Meteorological Instruments • Notes on the Meteorological Observations made in British Colonies and Protectorates • Meteorological Observations in the Foreign and Colonial Stations of the Royal Engineers and the Army Medical Department, 1852-1886	https://digital.nmla.metoffice.gov.uk/
MeteoSwiss - Weather Archive of Switzerland • Schweizerische Meteorologische Beobachtungen	https://www.meteoswiss.admin.ch/weather/weather-and-climate-from-a-to-z/weather-archive-of-switzerland.html
Midwestern Regional Climate Center's Military Forts and Voluntary Observers Database • Histories of Weather Observation at Selected Locations	https://mrcc.purdue.edu/FORTS
NOAA Central Library Foreign Climate Data Repository • Repertorium fur Meteorologie	https://libguides.library.noaa.gov/weather-climate/foreign-climate
Royal Meteorological Society • Quarterly Journal	

Please note, this table is not exhaustive and not all documents contained within each archive were searched for exposure metadata. Further information about the sources which provided nation- and station-specific metadata can be found in the sections below.

B.1. Albania

≥1940: Stevenson screen

There are only three Albanian stations in CRUTEM5_sdb which have data before 1961. The earliest starts in April 1940, therefore the assumption is made that Stevenson screens were in use by this date.

B.2. Algeria

Algeria was a French colony until 1962, therefore the assumptions applied are the same as for France (Section B.26).

B.3. Andorra

≥1950: Stevenson screen

There are only three Andorran stations in CRUTEM5_sdb which have data before 1961. The earliest starts in January 1950, therefore the assumption is made that Stevenson screens were in use by this date.

B.4. Argentina

<1885 Unknown
1885 - 1895: Transition to Stevenson screen
≥1895: Stevenson screen

The Argentine Meteorological Office was established in 1872, and the first instructions to observers were written in 1873 (Paolantonio, 2014; Rotch, 1894). However, it was not until 1885 that instruments and exposures began to be standardised and that the Stevenson screen was introduced (Smithsonian Institution, 1927). Sparks (1972) confirms that Stevenson screens continued to be used at Argentine stations until at least 1970.

Useful sources of station-specific information for Argentina included: Davis (1887, 1889, 1893, 1895) and Rotch, (1894).

B.5. Armenia

There are only three Armenian stations in CRUTEM5_sdb which have data before 1961. As these records are from the period before Armenia gained independence in 1991, the assumptions applied are the same as for Russia (Section B.63).

B.6. Australia

Thermometer exposures were not standardised across Australia until 1908 when the Bureau of Meteorology was established to consolidate the meteorological activities of the Australian states (Bureau of Meteorology, 2011; Nicholls et al., 1996). The early assumptions for Australia are therefore applied according to state (or territory) based on Nicholls et al. (1996). In 1907 Australian stations were instructed to use Stevenson screens (Hunt, 1907) and it is believed the screen was rapidly adopted from 1908.

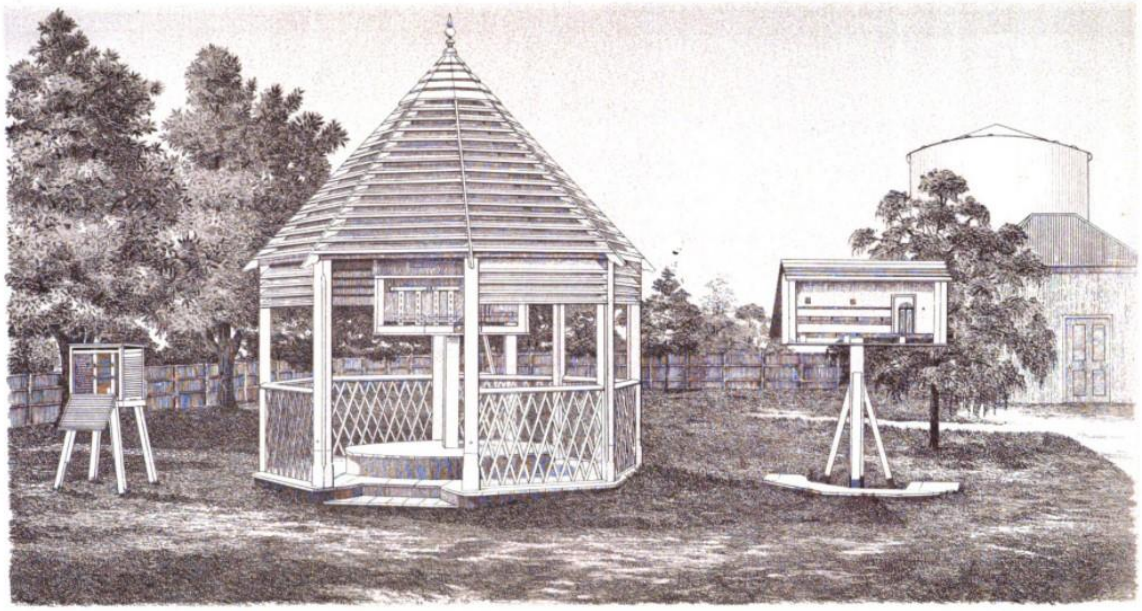


Figure B1. Examples of thermometer exposures in use in Australia in 1889. Left-to-right: Stevenson screen, Summerhouse, Glaisher stand. Image from: Meteorological observations made at the Adelaide Observatory (1891).

B.6.1. New South Wales

<1898: Unknown

1898 - 1908: Transition to the Stevenson screen

≥1909: Stevenson screen

B.6.2. Northern Territories and South Australia

<1892: Glaisher stand (Open exposure)

1892 - 1908: Transition to the Stevenson screen

≥1909: Stevenson screen

B.6.3. Queensland

<1888: Wall-mounted

1888 - 1889: Transition to the Stevenson screen

≥1890: Stevenson screen

B.6.4. Tasmania

<1895: Unknown

≥1895: Stevenson screen

B.6.5. Victoria

<1906: Thermometer shed (Intermediate exposure)

1906-1908: Transition to the Stevenson screen

≥1909: Stevenson screen

B.6.6. Western Australia

<1896: Unknown

1896 - 1897: Transition to the Stevenson screen

≥1898: Stevenson screen

Useful sources of station-specific information for Australia included: Abbott (1872); Ashcroft (2014); Cooke (1897); Hunt (1912); Nicholls et al. (1996); Russell (1875, 1892); Sabine (1850) and the Smithsonian Institution (1927, 1934).

B.7. Austria

<1925: Wall-mounted exposure

1925 - 1939: Transition to the Stevenson screen

≥1940: Stevenson screen

Figure B2. Examples of the wall-mounted exposures used in Austrian weather stations. Image from: Auer, Böhm and Schöner (2001).

In Austria, early observers exposed thermometers on North-facing walls throughout the eighteenth and nineteenth centuries, with cylindrical metal shields (such as Figure B2) commonly used from at least 1830 (Auer et al., 2001; Hann & Koppen, 1886; Harding, 1881). The first Stevenson screens were introduced to the Austrian network from the 1870s, however, wall-mounted exposures remained dominant until approximately 1925 when the number of Stevenson screens in use rapidly increased (Auer et al., 2001). By 1940 more than

90% of stations had a Stevenson screen, and in 1970 100% of Austrian stations still had a Stevenson-type screen (Auer et al., 2001; Sparks, 1972).

The CD included with the ALOCLIM report (Auer et al., 2001) was a very valuable source of station-specific metadata for Austrian stations.

B.8. Azerbaijan

Azerbaijan was part of the Russian Empire, and later the Soviet Union, until 1991, therefore all assumptions applied are the same as for Russia (Section B.63).

B.9. Belarus

Belarus was part of the Russian Empire, and later the Soviet Union, until 1990, therefore all assumptions applied are the same as for Russia (Section B.63).

B.10. Belgium

≥1920: Stevenson screen

At Belgian stations Stevenson screens were introduced in the 1920s (Parker, 1994; Venema, 2016). Only one station in CRUTEM5_sdb has data before 1920, this station has station-specific metadata, therefore no country-specific assumptions were required for the period before 1920.

B.11. Bermuda

Two Bermudan stations in CRUTEM5_sdb have data before 1961. Both stations have station-specific metadata for the majority of their records (James, 1861; Meteorological Office, 1923, 1926, 1927, 1928, 1929, 1930, 1931, 1932, 1933), therefore no country-specific assumptions were required or applied.

B.12. Bosnia and Herzegovina

<1970: Unknown

≥1970: Stevenson screen

Nine stations in Bosnia and Herzegovina have data before 1961. Unfortunately, little information has been found about the exposures in use prior to 1970 when Sparks (1972) found 95% of Yugoslavian stations exposed thermometers in Stevenson screens. Further information about the exposures used by the Austro-Hungarian empire and Yugoslavia are required to improve these assumptions.

B.13. Brazil

<1909: Unknown

≥1909: Closed

In 1909 the climatological activities in Brazil were unified by Henrique Morize and the network of observing stations was vigorously expanded (de Carvalho, 1917). At second class observing stations an adapted version of the Stevenson screen - an outer louvred screen with a smaller louvred box inside (Figure B3) - was used; at third order stations a louvred screen similar to a Stevenson screen was used and in the Sau Paulo network louvred screens with a thatched roof were used (de Carvalho, 1917; Parker, 1994). The former screens appear to still be in use at 80% of stations in 1970 suggesting a transition to a "typical English" Stevenson screen may not have occurred.

Note, these assumptions are only applied to two Brazilian stations in the state of Rio Grande do Sul (which start in 1924 and 1949 and which fall within 30° to 60° latitude).

Figure B3. The version of the Stevenson screen used at Brazilian stations. An additional louvred screen is used within the outer Stevenson-type screen. Image: de Carvalho (1917).

B.14. Bulgaria

<1970: Unknown

≥1970: Stevenson screen

Little information has been found about the exposures in use in Bulgaria prior to 1970 when Sparks (1972) found 99.99%-100% of stations had Stevenson screens. Further information

is required to improve these assumptions. Note, however, that all Bulgarian stations in CRUTEM5_sdb start in 1931, or later, so are less likely to be affected by the exposure bias.

B.15. Canada

<1878: Unknown

1878-1906: Wall-mounted exposure

≥1907: Stevenson screen

In 1878, Kingston (1878) instructed observers in Canada to expose thermometers in a louvred screen attached to the north side of a shed or insulated board (Figure B4) and this advice continued until at least 1897 (Stupart, 1900). In 1907 the Stevenson screen became the official screen in use at the Toronto Observatory and likely became standardised across Canada at that time (Parker, 1994). In 1970, 100% of Canadian stations still exposed thermometers in a Stevenson-type screen (Sparks, 1972).

Useful sources of station-specific information for Canada included: Kingston (1857); Morley (1991) and Sabine (1845, 1853, 1857).

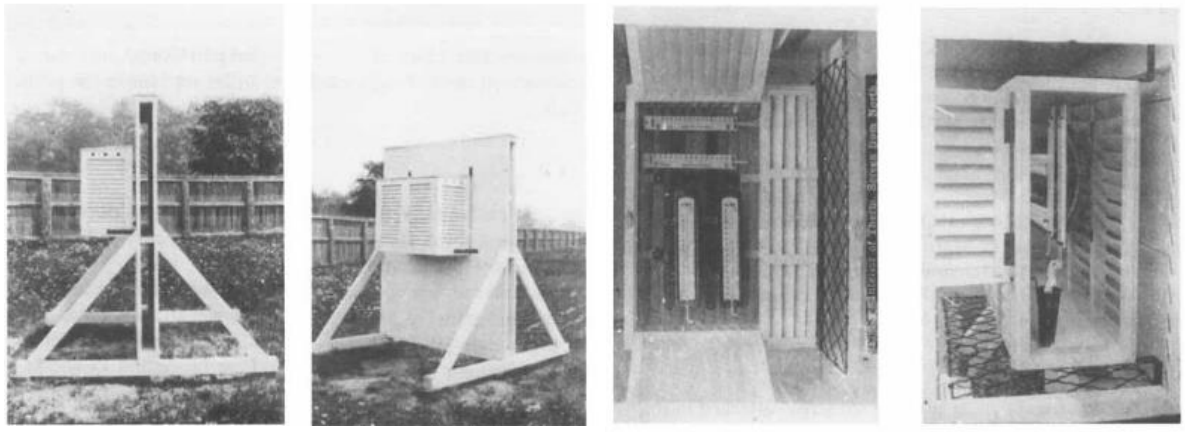


Figure B4. The Canadian screen and 'shed' recommended for use by Kingston (1878). Image: Parker (1994).

B.16. Chile

Unknown

Some information about the meteorological observations made in Chile were found in Oficina Central Meteorolojica (1870); Rotch (1894) and Sparks (1972); however, insufficient metadata was found to make country level assumptions about the exposures in

use. 28 stations (located between 30-60°S) in CRUTEM5_sdb have data before 1961, so further investigation is required.

B.17. China

Unknown

No country-wide assumptions were applied for China due to a lack of information about the early exposures in use; however, assumptions were applied to one early station network: the stations under the control of the Chinese Maritime Customs Service, and to stations in Manchuria which were under the control of Japan.

Useful sources of station-specific information for China included: Li et al. (2020); Shen et al. (2016); Si et al. (2021); von Storch & Grabel (2018) and the Observatory Reports found in the NOAA Central Library Foreign Climate Data repository.

B.17.1. Chinese Maritime Customs Service weather stations

<1887: Unknown
≥1887: Stevenson screen

In 1869 the Chinese Maritimes Customs Service began a meteorological network and from 1879 was routinely supplying the Zikawei Observatory in Shanghai with data (Ladds & Bickers, 2008). In an attempt to standardise the observations, instructions for observers were written in 1880 by Marc Dechevrens and then again in 1883 by William Doberck (Ladds & Bickers, 2008; Zhu, 2012). No copy of the 1880 instructions could be located, however, the 1883 version included instructions to expose thermometers in a double louvred screen (with a description very similar to a Stevenson screen). According to Ladds & Bickers (2008) the 1883 instructions were reprinted as an official Chinese Maritimes Customs Service publication in 1887 and the instructions were followed by weather stations from that year (Doberck, 1887b).

B.17.2. Weather stations in Manchuria

Manchuria was a region in present-day northeast China which was under the control of Japan in the early twentieth century. Metadata (e.g., Meteorological Observatory (1907)) suggest stations in this region had the same instrumentation as Japan (Section B.38).

B.18. Croatia

Unknown

No country-wide assumptions were applied for Croatia; further investigation of the metadata is required.

B.19. Cyprus

≥1887: Stevenson screen

Two Cypriot series starting in 1887 are present in CRUTEM5_sdb. As Cyprus was under British rule between 1878-1960, the stations are assumed to have Stevenson screens throughout their records (Section B.77). By 1970, all stations in Cyprus have Stevenson screens (Sparks, 1972).

B.20. Czech Republic

<1925: Wall-mounted
1925 - 1939: Transition to Stevenson screen
≥1940: Stevenson screen

Country-level assumptions for the Czech Republic are based on the information contained in Brazdil (2012) and are informed by the exposures in use in Austria (which controlled the weather stations in the Czech Lands between 1850 - 1918). 12 stations in the Czech Republic (in CRUTEM5_sdb) have data before 1950 but seven of these end before 1900 and may never have transitioned to a Stevenson screen. By 1970, 100% of stations had Stevenson-type screens (Sparks, 1972).

Useful sources of station-specific information for stations in the Czech Republic included: Bělohlávek (1977); Brazdil (2012) and Brázdil & Budíková, (1999).

B.21. Denmark

<1872: Unknown
1872 - 1912: Wall-mounted
1913 - 1928: Transition to Stevenson screen
≥1929: Stevenson screen

The country-wide assumptions above are based on Brandt (1994) who states: "*Between 1913 and 1928 most Danish climatological stations were equipped with a Stevenson screen*" (p. 23). Before this, most thermometers were mounted on a wall or at a window (Brandt, 1994; Meteorological Committee, 1876; Smithsonian Institution, 1927).

Useful sources of station-specific information for Denmark included: Brandt (1994); Cappelen (2021a) and Laursen (2003).

B.22. Egypt

<1915: Unknown

≥1915: Stevenson screen

Parker (1994) states that around 1910 a large network of Stevenson screens was established in Egypt and that these replaced Russian screens (which had earlier replaced French screens). By 1915 the Meteorological report states that Stevenson-type screens were in place at most stations (Ministry of Public Works Egypt, 1921). No dates for the use of French or Russian exposures were given by Parker (1994) so unfortunately information about the early exposures in use in Egypt could not be included in the country-level assumptions. Note that only three Egyptian stations (between 30° to 60° latitude) in CRUTEM5_sdb have observations before 1915.

B.23. Estonia

Due to the influence of the Russian Empire and the Soviet Union in Estonia throughout the nineteenth and twentieth centuries, the assumptions applied are the same as for Russian stations (Section B.63).

B.24. Falkland Islands

There are only two stations in CRUTEM5_sdb on the Falkland Islands: Mount Pleasant and Port Stanley. Exposure information was known for Port Stanley (Marriott, 1880; Meteorological Office, 1923-1933) therefore high-level assumptions were only required to populate Mount Pleasant. As the Falkland Islands have been under British administration for the duration of the Mount Pleasant record (1895-2021) it was assumed a Stevenson screen was in place throughout (Section B.77).

B.25. Finland

<1882: Wall-mounted exposure

1882 - 1894: Transition to Wild screen

1895-1908: Wild screen (Closed exposure)

1909 - 1911: Transition to Stevenson screen

≥1912: Stevenson screen

The above exposures are based on the description of historic station exposures in Finland in Nordli et al. (1996; p18). Note that only three Finnish stations in CRUTEM5_sdb (that are located between 30-60°N) have data before 1961; two of these stations start in January 1959 and one in January 1906.

B.26. France

<1873: Wall-mounted exposure

1873 - 1875: Transition to Montsouris (Open exposure)

1876 - 1924: Montsouris (Open exposure)

1925 - 1930: Transition to Stevenson screen

≥1931: Stevenson screen

Early meteorological observers in France were advised to expose thermometers, sheltered from the sun, on a northwest to northeast facing wall (Cotte, 1788). This form of exposure was likely dominant in France until the early 1870s when an open shelter - referred to as the French or Montsouris screen (Figure B5) - was designed by Charles Saint-Claire Deville while he was director of the Montsouris Observatory (Knowles Middleton, 1966; Leroy & Lefebvre, 2000). In 1873 Charles Saint-Claire Deville became the Inspector-General of Meteorological stations in France and between 1873 and 1875 the Montsouris shelter was distributed to French weather stations for use (Bureau Central Meteorologique de France, 1881; Knowles Middleton, 1966; Leroy & Lefebvre, 2000). The shelter was widely used until the transition to the Stevenson screen began in 1925 (Harding, 1881; Knowles Middleton, 1966; Leroy & Lefebvre, 2000). By 1931 the Stevenson screen was the exposure in use at the Central Meteorological Bureau, was included in the Bureau's Instructions Meteorologiques (Angot, 1931), and was also in use at the majority of French stations included in the World Weather Records, 1941-1950 (which included French data from 1931) (US Weather Bureau, 1959).

Useful sources of station-specific information for French stations included: Calvet (1985); Detwiller (1978); US Weather Bureau (1959) as well as the journal La Météorologie (available at: https://gallica.bnf.fr/) and the annual reports of the French and German Meteorological Societies (Annuaire de la Société météorologique de France and Deutsches meteorologisches Jahrbuch; both available at: https://www.hathitrust.org/).

Figure B5. The screen designed by Charles Saint-Claire Deville which became known as the French or Montsouris screen. Image: Karim Drici, Meteo-France from Leroy & Lefebvre (2000).

B.27. Georgia

Only two stations in Georgia (in CRUTEM5_sdb): Samtredia and Tbilisi, have observations before 1961. The record at Samtredia does not start until 1936 therefore it is assumed a Stevenson screen was in place from the start of the record. The record for Tbilisi starts in 1844; exposure information was known until 1882 (Waluew & Wild, 1881; Wild, 1881, 1883a) and after 1882 the country-level assumptions for Russia (Section B.63) were applied, as the observatory was under the control of the Russian Empire.

B.28. Germany

<1970: Unknown

≥1970: Stevenson screen

Due to Germany's complex history - between 1848-1934 nine meteorological services were active, each with their own instruments and instructions (Kaspar et al., 2013, 2015) - further research is required to determine the timing of the introduction of the Stevenson screen as well as the early exposures in use at stations in present-day Germany.

Kaspar et al. (2015) indicates observational practises were standardised in Germany in 1936 and the descriptions for the German variant of the Stevenson screen detailed in Sparks (1972) are dated 1933 and 1935, therefore, it is possible the Stevenson screen was introduced in the 1930s. For now, 1970 is set as the date when Stevenson screens were known to be in place, based on Sparks (1972) who found that 100% of German stations were using Stevenson-type screens in 1970.

B.29. Gibraltar

Two stations in Gibraltar in CRUTEM5_sdb have data before 1961: Gibraltar and North Front. The record for North Front does not start until 1951, therefore a Stevenson screen is assumed to be in place. The record for Gibraltar starts in 1852; some station-specific information is available between 1852-1886 (Meteorological Council, 1890) and 1923-1933 (Meteorological Office, 1923-1933). For the periods without station-specific metadata the country-level assumptions for the United Kingdom (Section B.77) were applied as Gibraltar has been a British Overseas Territory throughout the Gibraltar record.

B.30. Greece

≥1951: Stevenson screen

The majority of Greek station records in CRUTEM5_sdb start in 1951. Stevenson screens are assumed to be in use by 1951 in Greece based on a) Sparks (1972) which states that 100% of stations were using a Stevenson-type screen in 1970 and the screen had been "*used for many years with good performance by the Hellenic National Meteorological Service*" (p.56), and b) the fact that the National Observatory in Athens was using a Stevenson-type screen in 1890 (Founda et al., 2009; Karapiperis, 1954).

The Athens National Observatory is the only Greek station in CRUTEM5_sdb which starts earlier than 1951. From 1890 the exposure is believed to be a Stevenson screen (Founda et al., 2009; Karapiperis, 1954) and before 1890 the exposure is unknown.

B.31. Hungary

Unknown

Further information is required to make country-level assumptions. Note that only seven stations in Hungary have data before 1941 and the 1934 Hungarian Weather Report suggests the majority of Hungarian stations had a Stevenson-type screen at that time (Időjárási Jelentés Magyarországól, 1934).

B.32. India

<1924: Indian thatched shelter (Intermediate)
1924 - 1930: Transition to Stevenson screen
≥1931: Stevenson screen

See Section 3.4.2.1 for details.

Useful sources of station-specific information for India included: Chambers (1878) and the annual 'Reports of the Meteorology in India' and 'India Weather Review' available at Hathi Trust: https://catalog.hathitrust.org/Record/007957916 and the National Oceanic and Atmospheric Administration: https://libguides.library.noaa.gov/weather-climate/foreign-climate.

B.33. Iran

<1956: Unknown

≥1956: Stevenson screen

The Iranian Meteorological Department was established in 1956 and the first report of the department details thermometers were exposed in Stevenson screens at all stations (Climatological Branch of the Iranian Meteorological Department, 1958). No country-level information was found detailing the exposures before this date. Note that only five stations in Iran (in CRUTEM5_sdb and located between 30° to 60° latitude) have data before 1956.

B.34. Iraq

≥1923: Stevenson screen

From 1923, all Meteorological work in Iraq was conducted in accordance with the instructions in the British Meteorological Observer's Handbook (Meteorological Office, 1911), which stated that a Stevenson screen should be used (Meteorological Service, 1950).

Two stations in Iraq (in CRUTEM5_sdb) have data before 1923: Baghdad and Basra. Both stations were maintained by the Indian Meteorological service from the start of their records until 1918 and had thermometers exposed in Indian thatched shelters until that time (Meteorological Service, 1950). Between 1919-1922 the exposure 'Transition to Stevenson screen' is assumed at these stations.

B.35. Ireland

The British Meteorological Office administered the weather stations in Ireland until 1936 when Met Eireann was established (and when Stevenson screens would have been in place at most stations). See section B.77 for assumptions.

Useful sources of station-specific information for Ireland included: Mateus et al. (2020); Met Eireann (n.d.) and U.S. Department of Commerce (1966, 1979).

B.36. Israel

<1930: Unknown

≥1930: Stevenson screen

All but three of the station records from Israel (in CRUTEM5_sdb, between 30°-60° latitude) start after 1930 when a variant of the Stevenson screen is assumed to be in use at the majority of stations due to the British influence in the region (Section B.77).

Of the three stations with data before 1930, two short series: Jaffa (1904-1906) and Sarona (1880-1889), were assigned an unknown exposure and Stevenson screen exposure, respectively (the latter based on the date and the fact the observations were conducted by a British Organisation (Goldreich, 2003)). One longer series: Jerusalem (1861-2013) has station-specific information available for the majority of the record (Buchan, 1872; Chaplin, 1883; Meteorological Office, 1927-1933) so no earlier country-level assumption is required.

B.37. Italy

<1900: Wall-mounted

1900 - 1970: Unknown

≥1970: Stevenson screen

In the nineteenth century wall-mounted exposures were most common in Italy with observers instructed to place thermometers outside a North-facing window, ideally on the highest floor of a meteorological tower, above the roofs of surrounding buildings (Figure B6) (Brunetti et al., 2006; Denza, 1882; Hann & Koppen, 1891; Meteorological Committee, 1876; Tacchini, 1879). Specific dates for the transition to the Stevenson screen have not been found - Brunetti et al. (2006) suggests it "*began in the late-nineteenth century and continued in the following decades*" (p.359); however more specific information is required to refine the country-level assumptions. Sparks (1972) confirms 100% of stations in Italy have a Stevenson-type screen in 1970.

Useful sources of station-specific information for Italy include: Camuffo (2002c); Hann & Hellmann (1894); Maugeri et al. (2002) and Ogrin (2015).

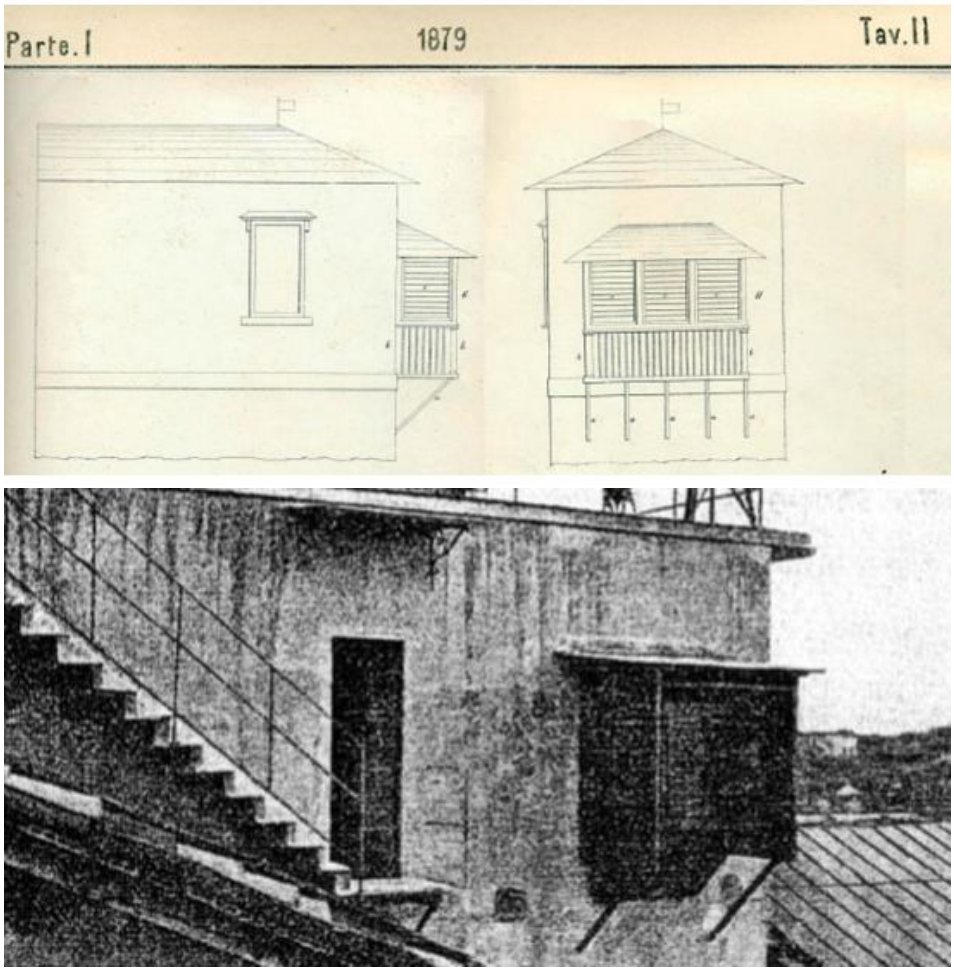


Figure B6. Upper image: illustration of the wall-mounted screen recommended for use by the Italian Central Office for Meteorology and Climate in 1879 (Tacchini, 1879). Lower image: an example of the wall-mounted screen in use at the Milano-Brera weather station (Maugeri et al., 2002).

B.38. Japan

<1880: Unknown

≥1880: Stevenson screen

A National Meteorological network was established in Japan in 1875 with the opening of the Central Meteorological Observatory in Tokyo. Reports suggest the instruments in use were standardised from this time and that the observations were taken according to strict guidelines:

> "*The observations, too, are taken following strictly to the Regulations ordered by the Central Observatory, and moreover observers are generally those who were trained in the Central Observatory. Thus, there is no doubt that the results of observations in all stations may be equally trusted.*" (Central Meteorological Observatory of Japan, 1893, p. Preface).

As the Central Meteorological Observatory exposed thermometers in a Stevenson screen from its establishment (The Imperial meteorological observatory, 1886) it is likely all stations did from this date; however, 1880 is used as the country-level assumption as this is when the first instructions for Meteorological Observers in Japan were printed (Central Meteorological Observatory of Japan, 1893a). Stevenson screens were listed as in use at all stations in 1893 (Central Meteorological Observatory of Japan, 1893b) and were still in use at 91% of stations in 1970 (Sparks, 1972).

Useful sources of station-specific information for Japan included: Central Meteorological Observatory of Japan (1893b, 1893a); Konnen et al. (2003) and The Imperial meteorological observatory (1886).

B.39. Jordan

≥1950: Stevenson screen

Four weather stations in Jordan in CRUTEM5_sdb have data before 1961, however, only two do not have any station-specific metadata. Both series without station-specific metadata start after 1950 and therefore a Stevenson screen is assumed to be in place.

B.40. Kazakhstan

Kazakhstan was part of the Russian Empire, and later the Soviet Union, until 1991, therefore all country-level assumptions applied are the same as for Russia (Section B.63).

B.41. Korea (North and South)

Korea was a Japanese colony between 1910 and 1945, therefore all country-level assumptions applied are the same as for Japan (Section B.38).

B.42. Kyrgyzstan

Kyrgyzstan was part of the Russian Empire, and later the Soviet Union, until 1991, therefore all country-level assumptions applied are the same as for Russia (Section B.63).

B.43. Latvia

Latvia was part of the Russian Empire, and later the Soviet Union, until 1991, therefore all country-level assumptions applied are the same as for Russia (Section B.63).

B.44. Lebanon

Unknown

Further information is required. Note that only four stations in Lebanon have observations before 1961 in CRUTEM5_sdb.

B.45. Libya

Unknown

Further information is required. 17 stations in Libya have observations before 1961 (in CRUTEM5_sdb), but the majority (12) start in 1945 or later.

B.46. Lithuania

Lithuania was part of the Russian Empire, and later the Soviet Union, until 1990, therefore all country-level assumptions applied are the same as for Russia (Section B.63).

B.47. Luxembourg

No country level assumption made. Only one series in CRUTEM5_sdb has data before 1961: Luxembourg Town, and station-specific information was used to determine the exposure (Reis, 2005).

B.48. Macedonia

Unknown

Further information is required. Three stations in Macedonia have data before 1961 (in CRUTEM5_sdb); one is a short fragment (Bitola: 1896-1910) and the remaining two start in 1949 or 1951 so are less likely to be affected by the exposure bias.

B.49. Madeira

Unknown

Further information is required. Two stations in Madeira have data before 1961 (in CRUTEM5_sdb); one starting in 1900 and the other in 1940.

B.50. Malta

No country-level assumptions were made. Only one series in CRUTEM5_sdb has data before 1961: Luqa, and station-specific information was used to determine the exposure (James, 1861; Meteorological Office, 1923-1933).

B.51. Moldova

Only one series in CRUTEM5_sdb has data before 1961: Kisinev (1825-2021). Station-specific information is known for some periods (Waluew & Wild, 1881); where information is not known the Russian assumptions are applied (Section B.63) as the Moldovan Meteorological Office was not established until 1944 (and the station was previously administered by the Russian Empire/Soviet Union) (Serviciul Hidrometeorologic de Stat, n.d.).

B.52. Mongolia

<1926: Unknown

≥1926: Stevenson screen

In the mid-1920's the Mongolian Government established a Meteorological Observing service with the support of the Central Geophysical Observatory in St Petersburg (Fickeler, 1926). As the latter had been using a Stevenson screen since 1915 it was assumed stations in Mongolia would also have been set up using one. Note, that only one series in CRUTEM5_sdb: Ulaanbaatar, has data before 1926.

B.53. Montenegro

Unknown

Further information is required. However, only two stations in Montenegro have data before 1961 (in CRUTEM5_sdb); one starting in 1949 and the other in 1950, so are less likely to be affected by the exposure bias.

B.54. Morocco

Unknown

Further information is required to determine country-wide exposure assumptions for Morocco.

B.54.1. Moroccan weather stations in the French Protectorate

Seven of the Moroccan stations in CRUTEM5 were within the French Protectorate between 1912 - 1956. The French assumptions are applied to these stations from 1912 (Section B.26).

Useful sources of station-specific information for Morocco included: Heidke (n.d.-a, n.d.-b); Neumayer (n.d.-a, n.d.-b) and US Weather Bureau, 1959).

B.55. Netherlands

<1906: Unknown

≥1906: Stevenson screen

Early observers in the Netherlands initially exposed thermometers in front of windows (Hartman, 1918; Meteorological Committee, 1876). Over time stations generally replaced this exposure with shelters, however the timing of the introduction and the design of the shelters varied greatly (Hartman, 1918). Uniformity was gradually established during the early 1900s via the introduction of Stevenson screens, and in 1906 a large climatological network which (generally) exposed thermometers in Stevenson screens was established (Hartman, 1918; Labrijn, 1948; van Ulden et al., 2009). Although the metadata suggest there was an earlier transition from wall-mounted to open or intermediate shelters, further information is required before these transitions can be incorporated into the country-level assumptions.

Useful sources of station-specific information for the Netherlands included: Dirksen et al. (2020); Labrijn (1945); van Engelen & Geurts (1985); van Ulden et al. (2009) and Visser (2005).

B.56. New Zealand

<1870: Unknown

1870 - 1876: Transition to the Stevenson screen

1877: Stevenson screen

Early observations in New Zealand were unstandardised; from 1859 ‘standard instruments’ were introduced in some regions by Colonial Secretary Sir Edward Stafford; however, observations only became standardised across New Zealand from June 1867 when observations came under the direction of James Hector (Salinger, 1981) . Stevenson screens were in use in New Zealand from 1870 (Hector, 1871) and were in use at most stations by 1877 (Symons, 1877).

Useful sources of station-specific information for New Zealand included: Mullan (2012); Salinger (1981) and US Weather Bureau (1959).

B.57. Norway

<1930: Wall-mounted exposure

1930 - 1934: Transition to the Stevenson screen

≥1935: Stevenson screen

Early observers in Norway were instructed to expose thermometers in front of windows from at least the 1860s and this was the most common form of exposure throughout the nineteenth century (Hann & Koppen, 1889; Meteorological Committee, 1876; Nordli et al., 1996). The first freestanding screens were introduced in Norway in 1877, but they did not become widespread and wall screens remained the most common form of exposure until the early 1930s when Stevenson-type screens replaced them (Nordli et al., 1996).

B.58. Pakistan

Until the Partition of India in 1947, Pakistan was part of India, therefore the assumptions applied are the same as sections B.32/3.4.2.1.

B.59. Peru

Unknown

Further information is required. However, only one station in Peru, which has data before 1961 in CRUTEM5_sdb, falls within 30-60°S.

B.60. Poland

<1945: Unknown

≥1945: Stevenson screen

The majority of stations in Poland (70%) which have data in CRUTEM5_sdb before 1961, do not start until 1951. These stations were likely set up by the National Hydrological and Meteorological Institute which was established in 1945 and which was responsible for reconstructing the network of hydrological and meteorological stations in Poland (IMGW-PIB, 2023). Due to the date, and documented use of the Stevenson screen in present-day Poland since the late-nineteenth century, all stations set up after 1945 are assumed to have Stevenson screens (Sparks (1972) confirms that in 1970 100% of stations used a Stevenson screen).

As many of the Polish stations which have records before 1951 were part of the Prussian Meteorological network, separate assumptions are applied to those stations.

B.60.1. Prussian weather stations

<1847: Unknown

1847-1888: Wall-mounted exposure

1889-1909: Transition to the Stevenson screen

≥1910: Stevenson screen

Prussian observers were advised from at least 1847 to expose thermometers in front of a North facing window (Mahlmann, 1847; Parker, 1994). This practise continued throughout the late-nineteenth century, with reports to the International Meteorological Committee in 1876 (Meteorological Committee, 1876) suggesting window-mounted exposures were the most common form of exposure in Prussia and instructions to observers in 1888 (Königlich Preussischen Meteorologischen Institut, 1888) recommending the use of the exposure illustrated in Figure B7. This recommendation is thought to have changed by 1909, with Gorczynski (1910) stating the Stevenson screen was the preferred exposure at stations in the Warsaw Meteorological network at that time. The 'Englishe Hutte' (a modified Stevenson screen) is first mentioned as being in use at Prussian stations in 1889, therefore the beginning of the transition period is set to 1889 (Königlich Preussischen Meteorologischen Institut, 1892a).

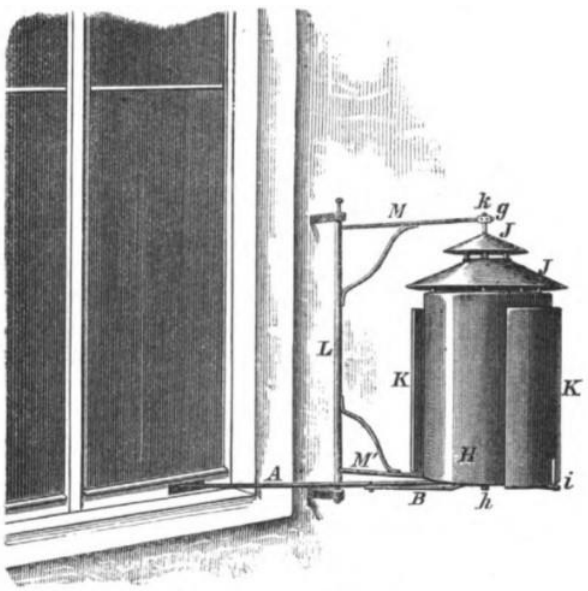


Figure B7. The "Thermometergehause" recommended for use in Prussian weather stations (Königlich Preussischen Meteorologischen Institut, 1888).

Useful sources of station-specific information for Polish stations included: Brys & Brys (2010); Deutsche Seewarte (1878, 1882, 1884, 1889); Institute of Meteorology and Water Management – National Research Institute (n.d.); Kolendowicz et al. (2019); Königlich Preussischen Meteorologischen Institut (1889, 1891, 1892a, 1892b, 1896, 1897, 1898,

1899); Kożuchowski et al. (1994); Mietus (1999, 2001); Mietus et al. (1994); Piotrowicz (2007); Przybylak et al. (2010); Trepinska (2007) and Trepinska & Piotrowicz (n.d.).

B.61. Portugal

≥1939: Stevenson screen

Station-specific metadata are available for all stations with data before 1939 in CRUTEM5_sdb. By 1939 it is assumed the Stevenson screen would have been in use at all stations.

Useful sources of station-specific information for Portugal included: Alcoforado et al. (2012); Bližňák et al. (2015) and Morozova & Valente (2012).

B.62. Romania

Unknown

Further information is required.

B.63. Russia

< 1870: Wall exposure
1870 - 1882: Transition to Wild hut
1882 - 1924: Wild hut
1925-1934: Transition to Stevenson screen
≥1935: Stevenson screen

The first meteorological observations in Russia began in 1726 in St. Petersburg, but it was not until the 1830s that the number of observing stations rapidly increased (Waluew & Wild, 1881). The majority of observing stations in the former Russian Empire reported to the Central Observatory in St. Petersburg who issued frequent instructions to observers regarding the way measurements should be taken (Waluew & Wild, 1881). Initially, observers were instructed to mount thermometers in shade outside a North-facing window (Kamtz, 1860; Kuppfer, 1835). However, in 1869, new instructions were issued (Wild, 1869) which recommended exposing thermometers in a cylindrical shield within a freestanding louvred shelter; an exposure which became known as a ‘Wild Hutte’ (Figure B8) (Waluew & Wild, 1881). These instructions came into force in 1870 and the Wild Hutte remained the predominant exposure in Russia until at least 1910 (Gorczynski, 1910; Parker, 1994). In 1912 the Central Observatory began using a Stevenson screen (Koppen, 1913), however

shortly after the network was adversely affected by war and it is unlikely that the screen was introduced more widely (Parker, 1994; Sternzat, 1967). Sternzat (1967) suggests new instrumentation was unlikely much before 1930 and Rusin and Bespalov (1956) (summarized in Boylan (1969)) suggest unification and standardisation of the Soviet Meteorological network occurred between 1925 and 1935, therefore these dates are set as the likely timing of the transition to the Stevenson screen. Sparks (1972) confirms that by 1970 100% of stations in the Soviet Union exposed thermometers in Stevenson-type screens.

Useful sources of station-specific information for Russia (including stations in the Russian Empire and Soviet Union) included: Frolich (1876); Jones & Lister (2002); Voznessensky (1915); Waluew & Wild (1881) and Wild (1883b, 1885, 1887b, 1888, 1889).

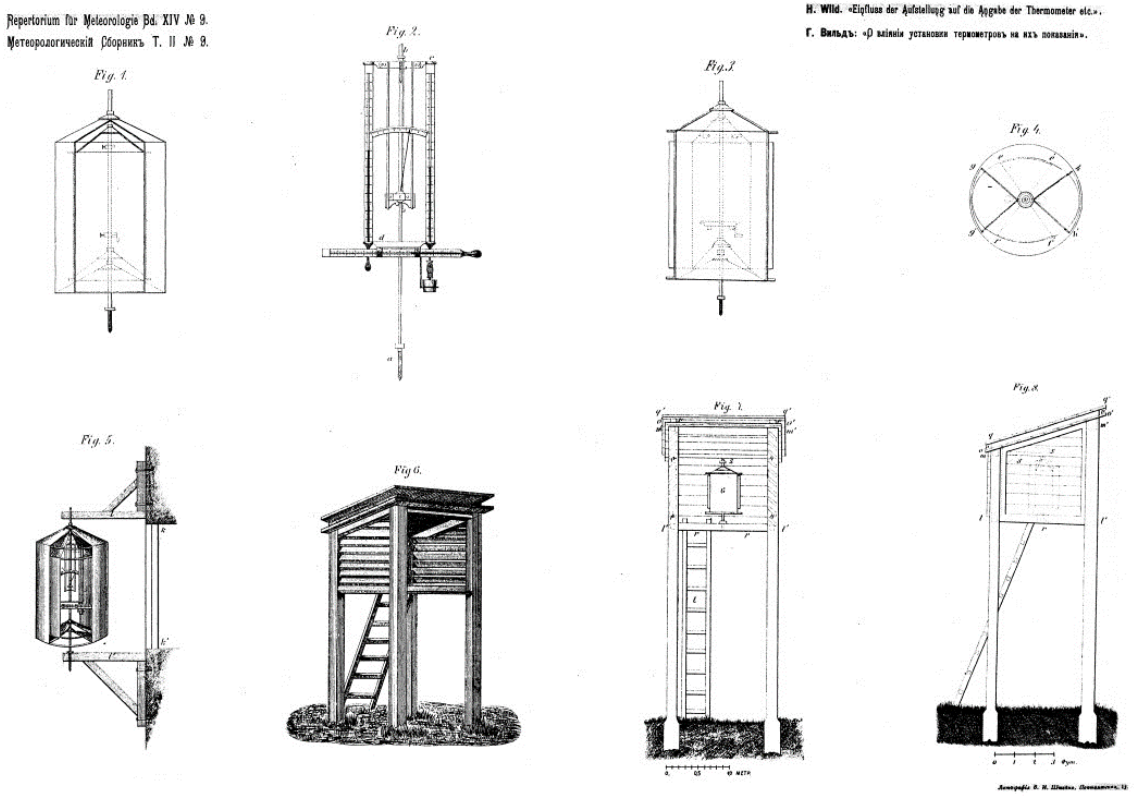


Figure B8. Illustration of the 'Wild Hutte' and cylindrical shield which was common in Russian weather stations from the early 1880s. In many European countries a version of Wild's cylindrical shield was also used without the 'Hutte', mounted on a North-facing wall. Image from: Wild (1891b).

B.64. Serbia

Unknown

Further information is required. However, all stations start in 1939, or later, so are less likely to be affected by the exposure bias.

B.65. Slovakia

Unknown

Further information is required. However, all but two stations start after 1940, so are less likely to be affected by the exposure bias.

B.66. Slovenia

Unknown

Further information is required. However, all but four stations start after 1955, so are less likely to be affected by the exposure bias.

B.67. South Africa

≥1881: Stevenson screen

The Stevenson screen was adopted by the South African Meteorological Commission in 1880 and in 1881 the majority of stations replaced early exposures with the Stevenson screen (Gamble, 1879; Gill, 1882). The Stevenson screen remained in use at most stations until at least 1970 (Sparks, 1972).

Only one station in CRUTEM5_sdb has data before 1881: the Royal Observatory Cape Town. As station-specific information was available for this station (Gill, 1882), no earlier country assumptions were required or applied.

B.68. Spain

<1910: Open exposure
1910-1920: Transition to the Stevenson screen
≥1921: Stevenson screen

These assumptions are based on the information in Brunet et al. (2006, 2011).

B.69. Saint Helena

There is only one series on Saint Helena in CRUTEM5_sdb Tristan de Cunha (1942-1988). Given the dates, and because Saint Helena is a British Overseas Protectorate, a Stevenson screen is assumed to be in place throughout the record (Section B.77).

B.70. Sweden

<1930: Wall exposure

1930 - 1939: Transition to Stevenson type screen

≥1940: Stevenson screen

These assumptions are based on the information contained in Nordli et al. (1996).

Useful sources of station-specific information for Sweden included: Alexandersson & Eriksson (1989); Bergström & Moberg (2002) and Moberg et al. (2002, 2003).

B.71. Switzerland

<1950: Wall-mounted exposure

1950 - 1969: Transition Stevenson screen

≥1970: Stevenson screen

Observers in Switzerland were instructed to mount thermometers on a north facing wall, close to a window, from at least 1825; this advice continued throughout the nineteenth century with instructions in 1863 and 1893 recommending observers exposed thermometers in a 'Zinkblechhutte' (a zinc cylindrical shield) mounted on a north wall (Candolle, 1825; Schweizerische meteorologische Zentralanstalt, 1893). The 'Zinkblechhutte' was not replaced by a Stevenson screen in the majority of stations until the 1950's, however some stations had a Wild Hutte (Figure B8) from the 1890s (Begert et al., 2003, 2005). By 1970, 100% of stations in Switzerland had a Stevenson-type screen (Sparks, 1972).

Useful sources of station-specific information for Switzerland included: Auchmann & Brönnimann (2012); Cassidy (1985); Kuglitsch et al. (2012); Schweizerischen Meteorologischen Central-Anstalt (1891, 1894, 1898) and Schweizerischen Meteorologischen Zentralanstalt (1956).

B.72. Syria

<1970: Unknown

≥1970: Stevenson screen

Fourteen stations in Syria in CRUTEM5_sdb have data before 1961. Little information has been found about the exposures in use prior to 1970 when Sparks (1972) confirms 100% of stations exposed thermometers in Stevenson screens.

B.73. Tajikistan

Tajikistan was part of the Russian Empire, and later the Soviet Union, until 1991, therefore all country-level assumptions applied are the same as for Russia (Section B.63).

B.74. Tunisia

<1970: Unknown

≥1970: Stevenson screen

Four stations in Tunisia in CRUTEM5_sdb have data before 1961 (three before 1949). Little information has been found about the exposures in use prior to 1970 when Sparks (1972) confirms 90% of stations have Stevenson screens.

B.75. Turkey

<1970: Unknown

≥1970: Stevenson screen

Little information has been found about the exposures in use in Turkey prior to 1970 when Sparks (1972) confirms 100% of stations have Stevenson screens. 33 stations in CRUTEM5_sdb have data before 1961, so further investigation is required.

B.76. Turkmenistan

Turkmenistan was part of the Russian Empire, and later the Soviet Union, until 1991, therefore all country-level assumptions applied are the same as for Russia (Section B.63).

B.77. United Kingdom

<1740: Exposed within an unheated room (Miscellaneous)

1740 - 1839: Wall exposure

1840 - 1872: Glaisher (Open) / Wall-mounted

1873 - 1883: Transition to the Stevenson screen

≥1884: Stevenson screen

From the 1740's meteorological observers in the United Kingdom (UK) were advised to expose thermometers on a North-facing wall; prior to this, thermometers would likely have been exposed inside an unheated room, as advised by Jurin (1723) (Knowles Middleton, 1966). The first freestanding shelters, including the Glaisher stand (Figure B1; Figure B9), were used in the UK from the 1840s; however, they did not replace the use of wall-mounted screens with both forms of exposure common between 1840-1880 (Knowles Middleton, 1966; Parker, 1994). In 1873 the Stevenson screen was adopted for use in the UK by the Meteorological Society (Mawley, 1897) and new instructions for observers, detailing the Stevenson screen, were printed in 1875 (Scott, 1875). The use of Stevenson screens rapidly

expanded through the mid-1870s and by 1884 they were in use at the majority of stations in the UK (Naylor, 2019; Parker, 1994).

Useful sources of station-specific information for the UK included: the Quarterly Journal of the Royal Meteorological Society, Burt (2022); Burt & Burt (2019); Butler, Garcia Suarez, et al. (2005); Hawkins et al. (2019); Manley (1941); Plummer & Scott (1873) and Observatory Yearbooks (available from: http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/yearbooks.html).

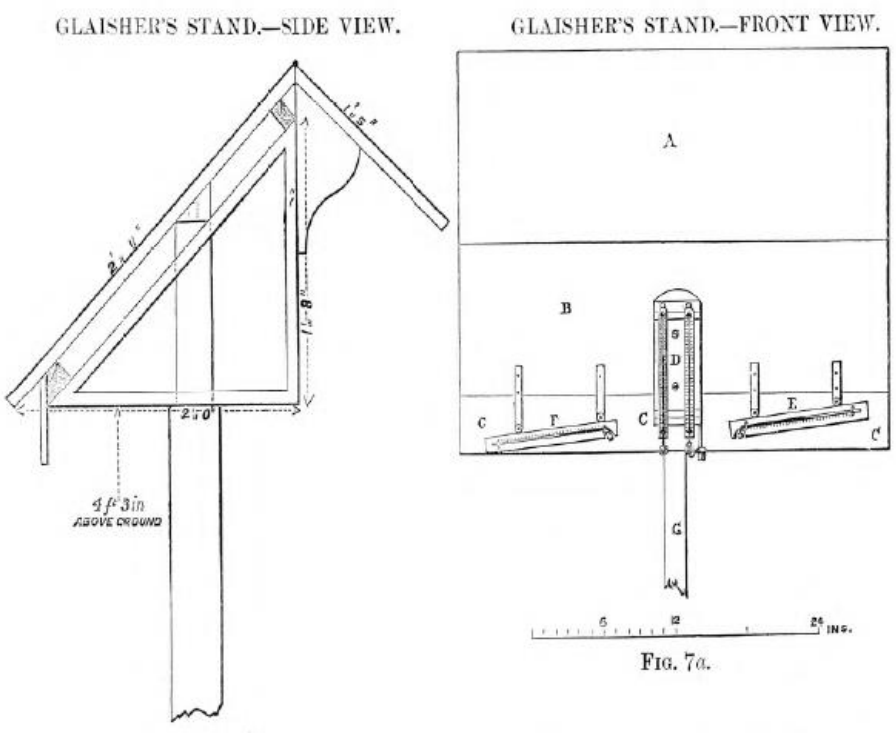


Figure B9. Illustration of the Glaisher stand which was designed by James Glaisher and adopted for use by the Greenwich Observatory, UK in 1841. Image: Gaster (1882).

B.78. Ukraine

Ukraine was part of the Russian Empire, and later the Soviet Union, until 1991, therefore all country-level assumptions applied are the same as for Russia (Section B.63).

Useful sources of station-specific information for Ukraine included: Waluew & Wild (1881) and Wild (1875, 1887b, 1889).

B.79. United States of America

<1892: Wall-mounted exposure

1892 - 1902: Transition to the Stevenson screen

≥1903: Stevenson screen

Early observers in the United States of America were advised to mount thermometers in front of a North-facing wall or window from at least 1860 (Figure B10) (Smithsonian Institution, 1860). This advice continued until 1885, when, after extensive investigation of different thermometer exposures, Hazen (1885) recommended the use of a freestanding screen either over grass or on the roof and advised against exposing thermometers mounted on the walls of buildings (Figure B10). Despite this, wall-mounted exposures are thought to have remained the most common form of exposure until the 1890s when the Cotton Region Shelter (Figure B10), a variant of the Stevenson screen, was adopted by the US Weather Bureau (Chenoweth, 1993; Flora, 1920; Russell, 1892). By 1903 this form of exposure was the most commonly used in the USA (Chenoweth, 1993).

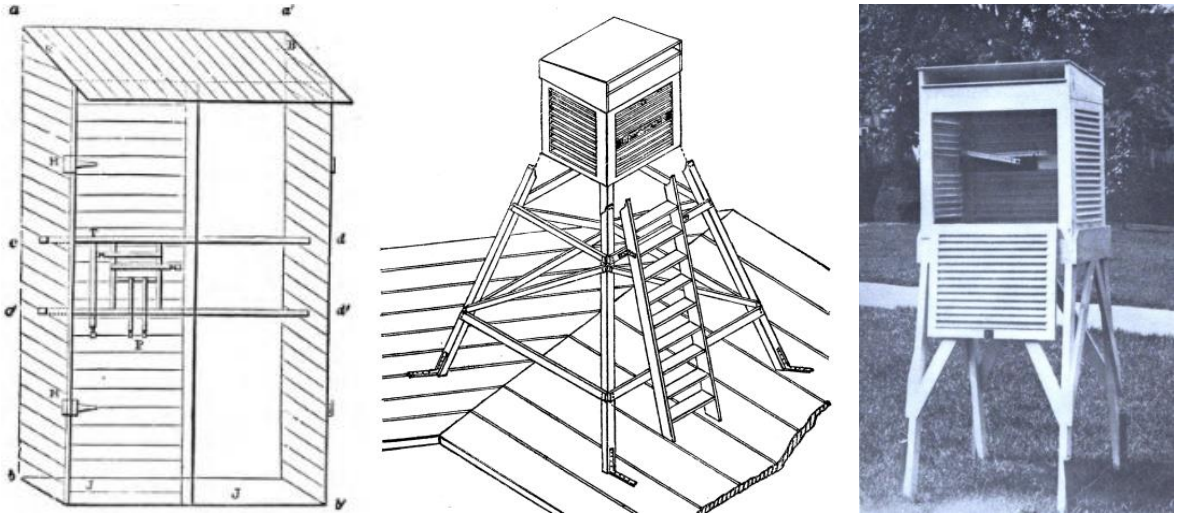


Figure B10. Thermometer exposures recommended for use in the United States of America. Left: Wall-mounted exposure recommended in 1860 (Smithsonian Institution, 1860); Centre: Freestanding Hazen shelter recommended in 1885 (Hazen, 1885; Image from: US Signal Office, 1887) Right: Cotton Region shelter recommended for use from 1892 (Russell, 1892; Image from: Marvin, 1915).

Useful sources of station-specific information for the USA included: Henry (1906); Smithsonian Institution (1927); US Weather Bureau (1959), and the Station Histories from the Midwestern Regional Climate Center's Military Forts and Voluntary Observers Database (Available at: https://mrcc.purdue.edu/FORTS).

B.80. Uruguay

Unknown

Further research is required to determine the early exposures in use in Uruguay and the date of Stevenson screen introduction. Fourteen stations in Uruguay in CRUTEM5_sdb have data before 1961.

B.81. Uzbekistan

Uzbekistan was part of the Russian Empire, and later the Soviet Union, until 1991, therefore all country-level assumptions applied are the same as for Russia (Section B.63).

Appendix C

Supplement to Chapter 4

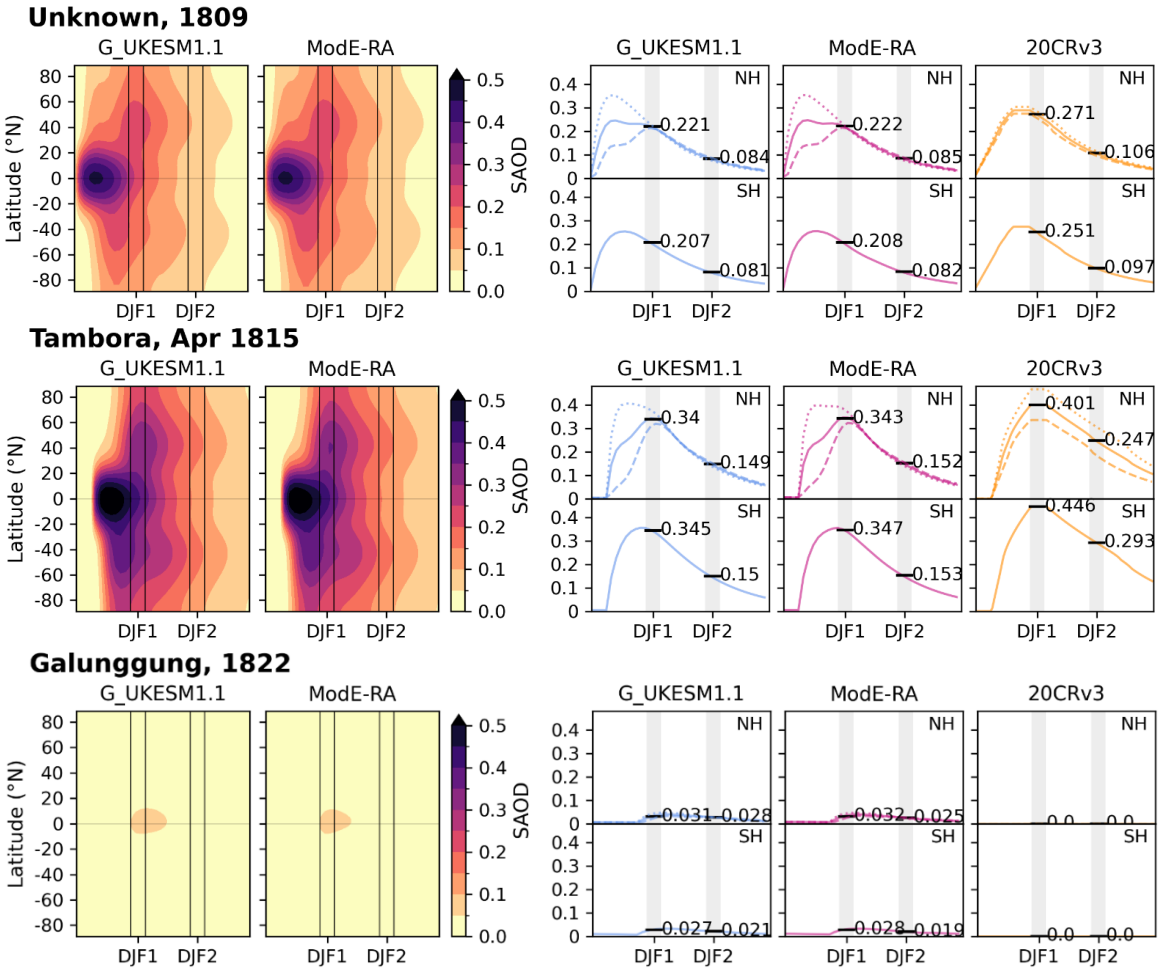


Figure C1. Stratospheric Aerosol Optical Depth (SAOD) profiles for each major low latitude eruption considered for selection. The first two panels show the evolution of SAOD during the year of the eruption and the following two years (with the first and second DJF seasons lying between each pair of vertical lines) as a function of latitude, for the forcing used in G_UKESM1.1 and ModE-RA. The remaining three panels show timeseries of SAOD averaged over each hemisphere (solid lines). In the NH panels, the SAOD in the tropics (0-30°N, dotted) and extratropics (30-90°N, dashed) is also shown. The NH winters are indicated by grey shading and the hemispheric averages during each NH winter are labelled. Note that from 1850, G_UKESM1.1 and ModE-RA use the same volcanic forcing.

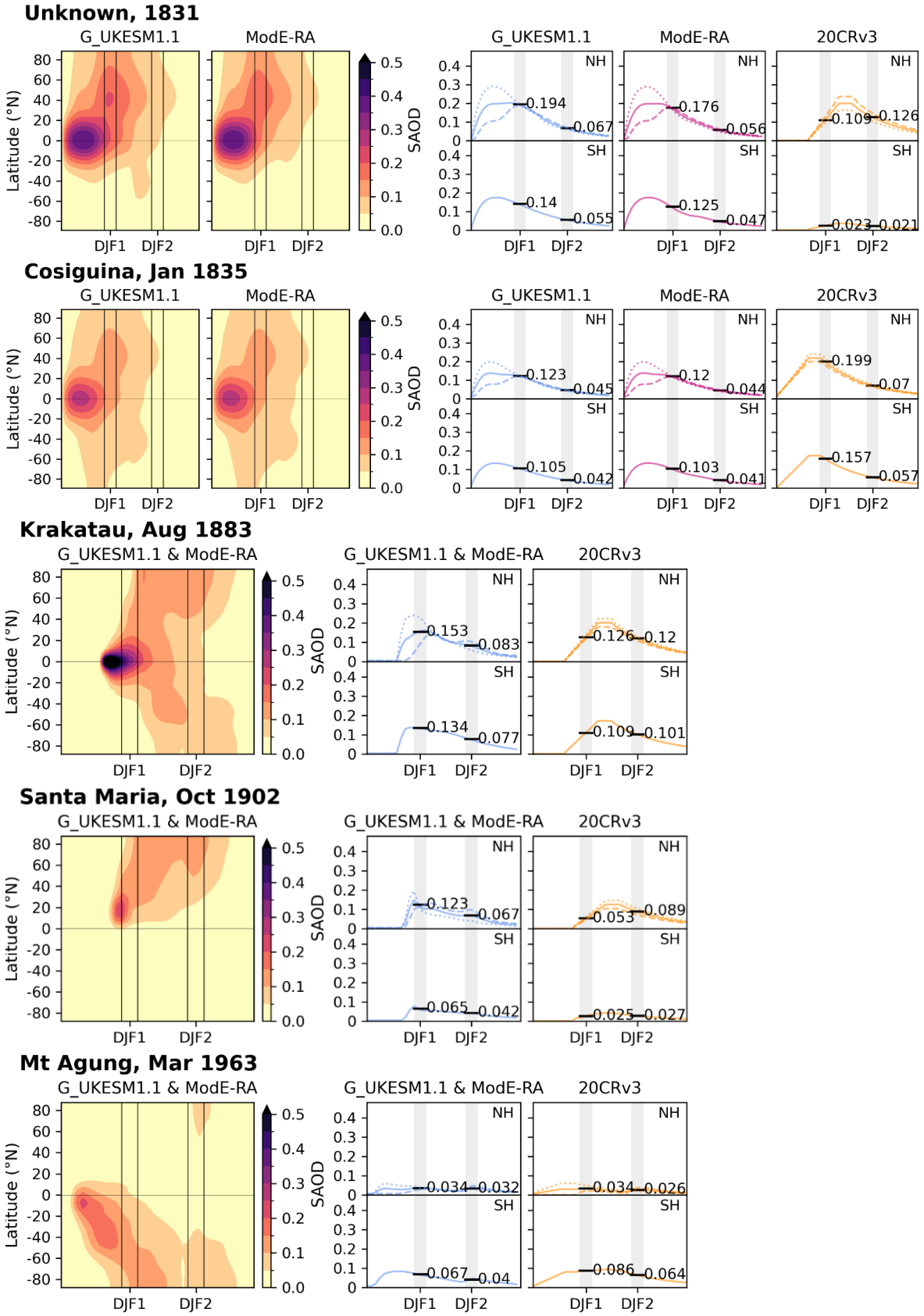


Figure C1 continued.

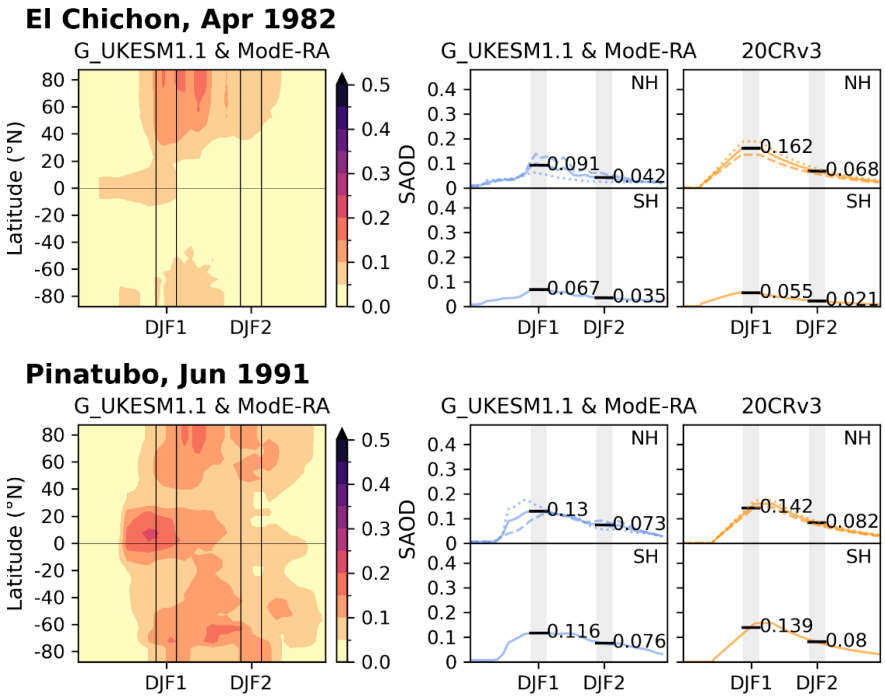


Figure C1 continued.

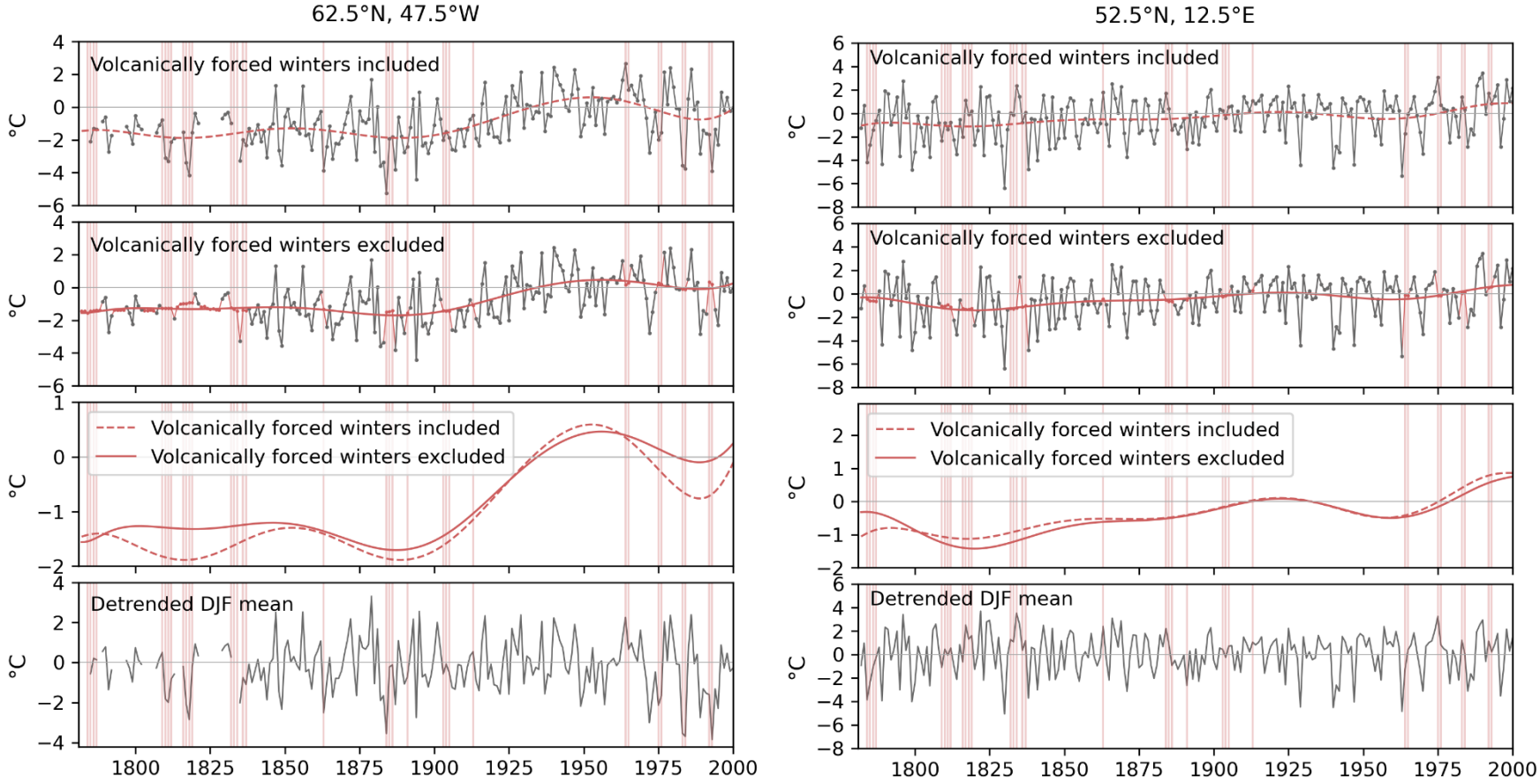


Figure C2. Quartic spline fitted including (dashed line) or excluding (solid line) the volcanically forced winters for two grid cells in the GloSAT observations (temperature anomalies with respect to 1961-1990, °C). The bottom panels show the detrended December-to-January mean using the spline fitted excluding the volcanically forced winters. The grid cell on the left is located in the Greenland study region and the grid cell on the right is located in the Europe study region. Vertical dashed lines represent 'volcanic years' as defined in Section 4.3.3.

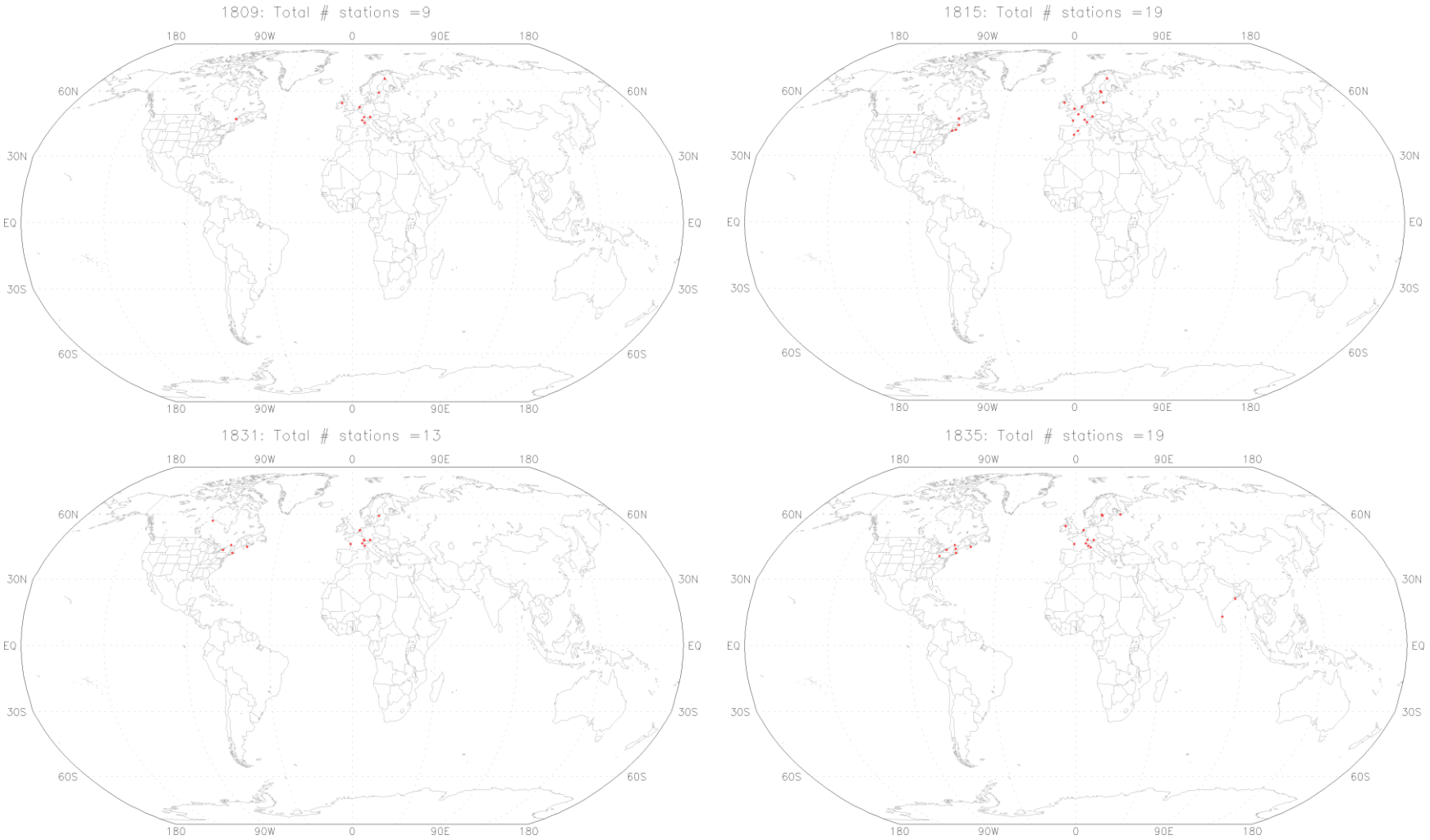


Figure C3. Locations of the surface pressure observations assimilated into the 20CRv3 dataset at the time of the eruptions of U1809, Tambora (1815), U1831 and Cosiguina (1835). Plots from Compo et al. (2019).

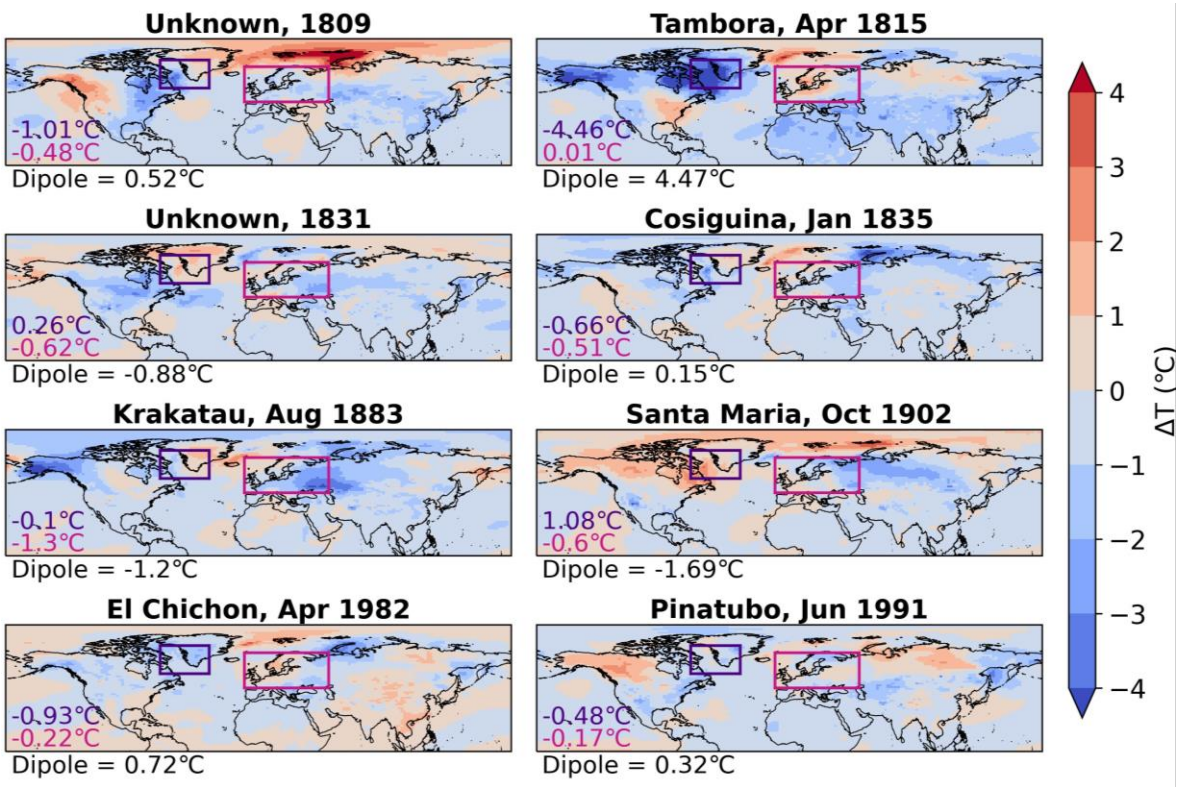


Figure C4. Post-eruption surface temperature anomalies in the second winter (DJF2) following each named eruption in the G_UKESM1.1 ensemble mean. The number in purple (pink) is the area-weighted mean temperature anomaly for the Greenland (Europe) box and the regional difference is the difference between the two (Europe minus Greenland). Temperature anomalies are calculated with respect to detrended non-volcanic winters between 1781/2 – 1998/9.

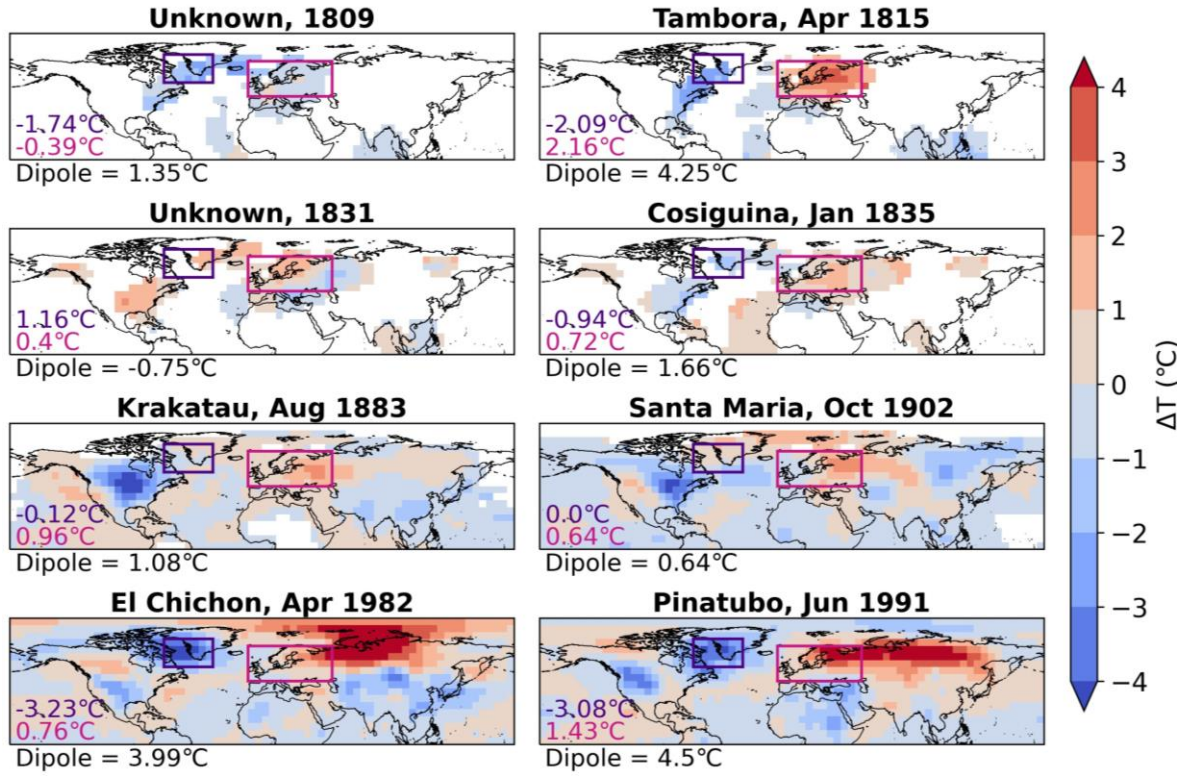


Figure C5. As for Figure C4, but for the GloSAT ensemble mean.

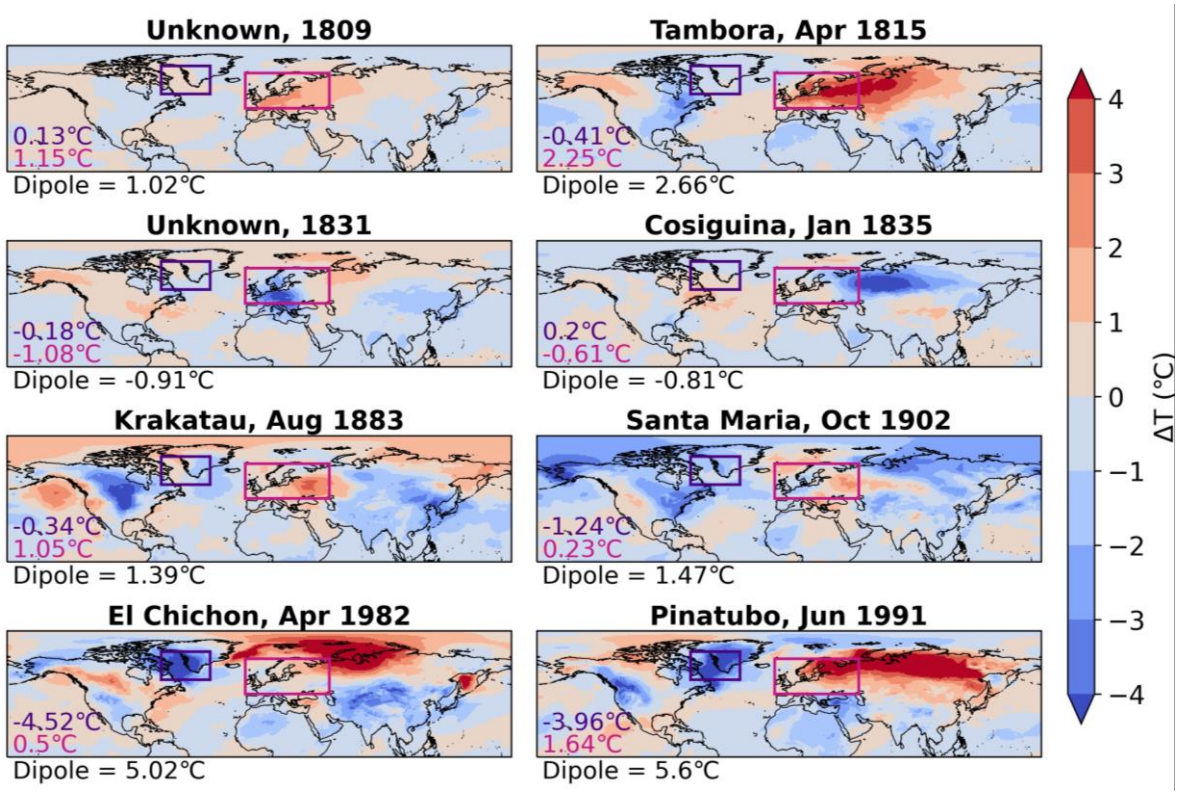


Figure C6. As for Figure C4, but for the 20CRv3 ensemble mean. Note that the anomalies are calculated with respect to a slightly shorter reference period than for the other datasets due to the 20CRv3 dataset starting in 1806.

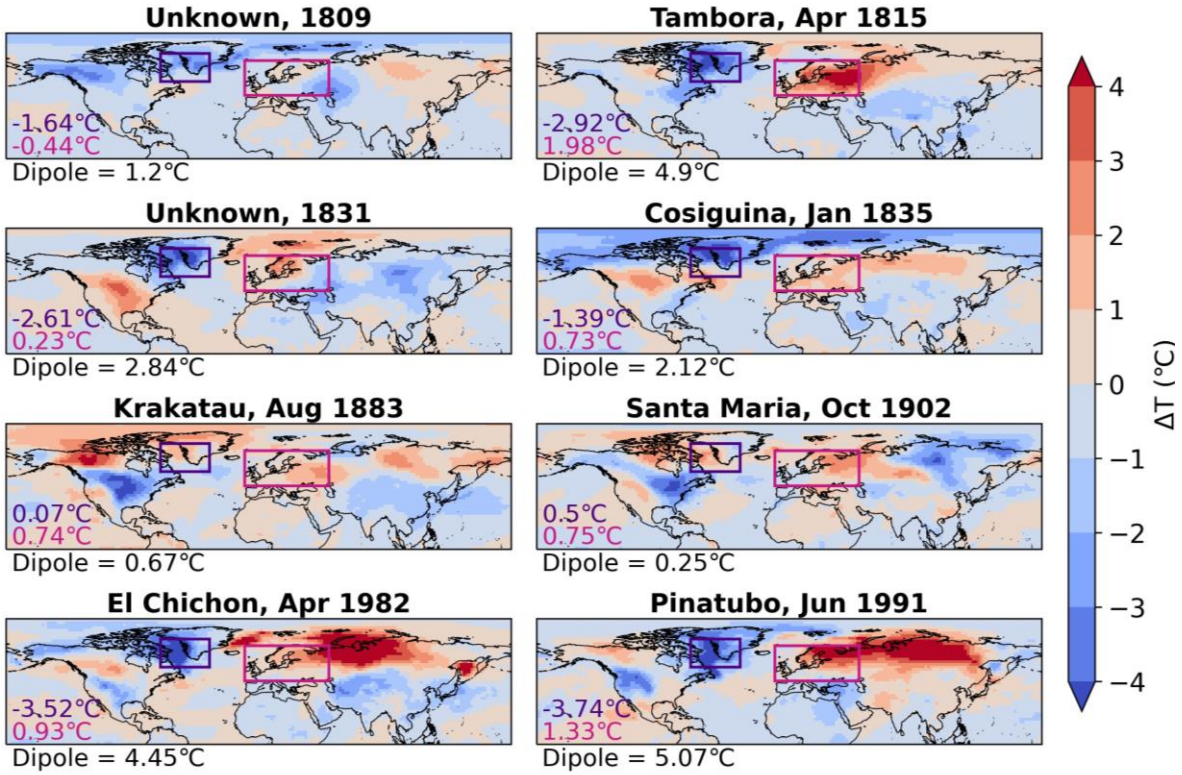


Figure C7. As for Figure C4, but for the ModE-RA ensemble mean.

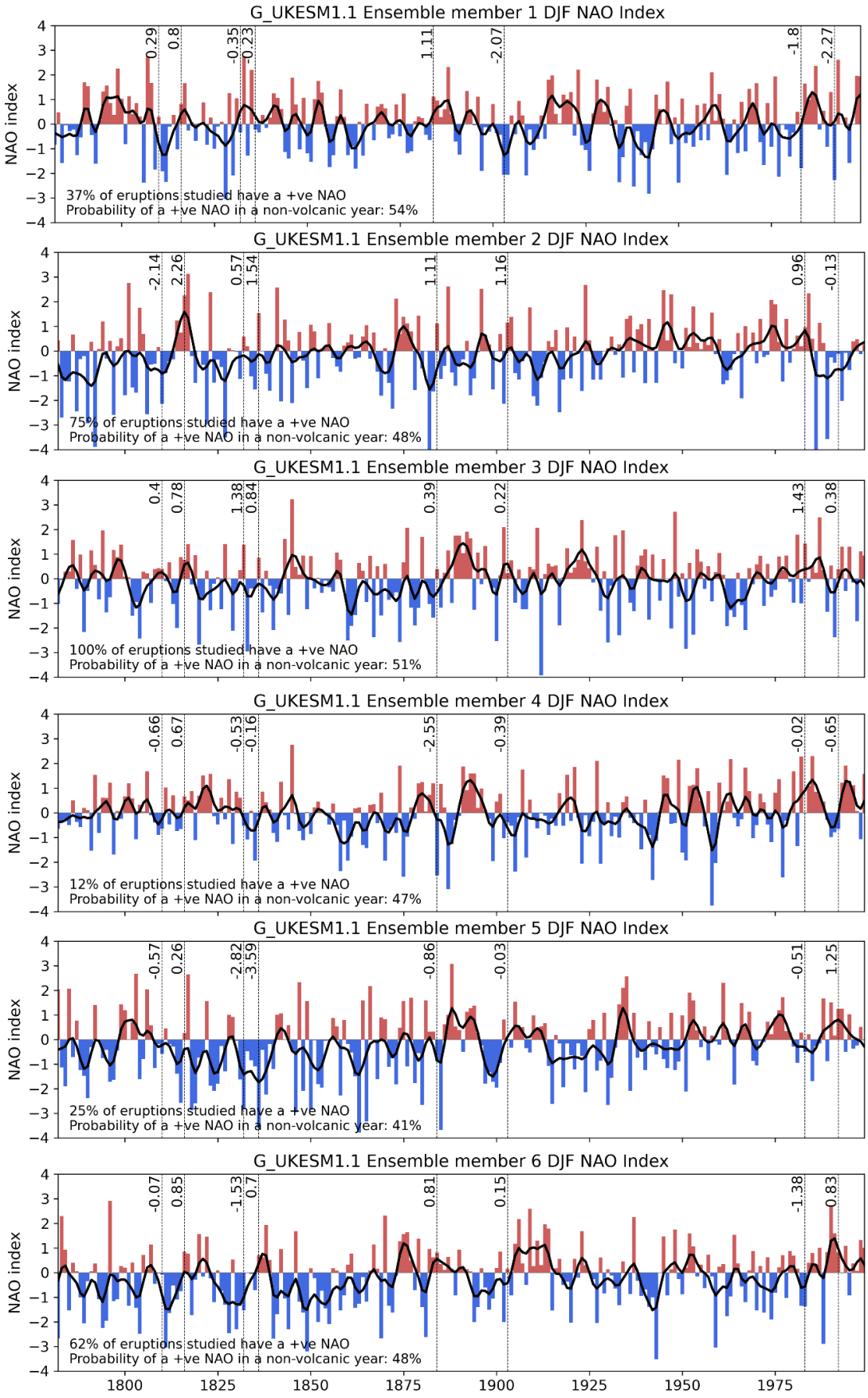


Figure C8. Winter North Atlantic Oscillation Indices for each of the G_UKESM1.1 ensemble members. Values during winters following volcanic eruptions are indicated.

Glossary

20CRv3 Twentieth Century Reanalysis version 3 [dataset]

AMJJA April-to-August

AWS Automatic Weather Station

CMIP Coupled Model Intercomparison Project

CMIP5 Coupled Model Intercomparison Project, phase 5

CMIP6 Coupled Model Intercomparison Project, phase 6

CRUTEM5 Climatic Research Unit Land Surface Air Temperature version 5 [dataset]

CRUTEM5_eba Exposure bias adjusted version of CRUTEM5_ext [dataset]

CRUTEM5_ext Extended version of CRUTEM5, starting in 1781 [dataset]

CRUTEM5_sdb Station database for CRUTEM5 [dataset]

DJF December-to-February

DJF0 The winter directly preceding an eruption (or the winter the eruption occurred)

DJF1 The first post-eruption winter

DJF2 The second post-eruption winter

DJFM December-to-March

DTR Diurnal Temperature Range

EIP Early Instrumental Period

ENSO El Niño Southern Oscillation

HadCRUT5 Hadley Centre/Climatic Research Unit Temperature version 5 [dataset]

HadSST4 Hadley Centre Sea Surface Temperature version 4 [dataset]

IPCC Intergovernmental Panel on Climate Change

GloSAT Global Surface Air Temperature [project and dataset]

JJA June-to-August

KPSS Kwiatkowski–Phillips–Schmidt–Shin statistical test

LSAT Land Surface Air Temperature

MAM March-to-May

MJJA May-to-August

ModE-RA Modern Era Reanalysis [dataset]

MSLP Mean Sea Level Pressure

NAO North Atlantic Oscillation

NH Northern Hemisphere

NMS National Meteorological Service

NOAA National Oceanic and Atmospheric Administration

NWP Numerical Weather Prediction

QBO Quasi-Biennial Oscillation

SAOD Stratospheric Aerosol Optical Depth

SH Southern Hemisphere

SON September-to-November

SST Sea Surface Temperature

SWD Shortwave Downward solar radiation

TOA Top Of Atmosphere solar radiation

U1809 Unknown volcanic eruption in 1809

U1831 Unknown volcanic eruption in 1831

VEI Volcanic Explosivity Index

References

Abbe, C. (1888). Exposure of Thermometers. In *Treatise on meteorological apparatus and methods* (p. 392). https://catalog.hathitrust.org/Record/012304586

Abbott, F. (1872). *Results of five years' meteorological observations for Hobart Town*. https://hdl.handle.net/2027/hvd.hxgnkc

Aguilar, E., Auer, I., Brunet, M., Peterson, T. C., & Wieringa, J. (2003). *Guidelines on climate metadata and homogenization*.

Alcoforado, M. J., Vaquero, J. M., Trigo, R. M., & Taborda, J. P. (2012). Early Portuguese meteorological measurements (18th century). *Climate of the Past*, *8*(1), 353–371. https://doi.org/10.5194/CP-8-353-2012

Alexandersson, H., & Eriksson, B. (1989). *Climate Fluctuations in Sweden 1860-1987*.

Allan, R., & Ansell, T. (2006). A New Globally Complete Monthly Historical Gridded Mean Sea Level Pressure Dataset (HadSLP2): 1850–2004. *Journal of Climate*, *19*(22), 5816–5842. https://doi.org/10.1175/JCLI3937.1

Allan, R., Brohan, P., Compo, G. P., Stone, R., Luterbacher, J., & Bronnimann, S. (2011). The International Atmospheric Circulation Reconstructions over the Earth (ACRE) Initiative. *Bulletin of the American Meteorological Society*, *92*(11), 1421–1425. https://doi.org/10.1175/2011BAMS3218.1

Anchukaitis, K. J., Wilson, R., Briffa, K. R., Büntgen, U., Cook, E. R., D'Arrigo, R., Davi, N., Esper, J., Frank, D., Gunnarson, B. E., Hegerl, G., Helama, S., Klesse, S., Krusic, P. J., Linderholm, H. W., Myglan, V., Osborn, T. J., Zhang, P., Rydval, M., … Zorita, E. (2017). Last millennium Northern Hemisphere summer temperatures from tree rings: Part II, spatially resolved reconstructions. *Quaternary Science Reviews*, *163*, 1–22. https://doi.org/10.1016/j.quascirev.2017.02.020

Angot, A. (1931). *Instructions Meteorologiques*.

Aono, Y., & Kazui, K. (2008). Phenological data series of cherry tree flowering in Kyoto, Japan, and its application to reconstruction of springtime temperatures since the 9th century. *International Journal of Climatology*, *28*(7), 905–914. https://doi.org/10.1002/joc.1594

Arfeuille, F., Weisenstein, D., MacK, H., Rozanov, E., Peter, T., & Brönnimann, S. (2014). Volcanic forcing for climate modeling: A new microphysics-based data set covering years 1600-present. *Climate of the Past*, *10*(1), 359–375. https://doi.org/10.5194/cp-10-359-2014

Ashcroft, L. (2014). *Adjustments made to 1860–2009 southeastern Australian temperature, rainfall and mean sea level pressure data*.

Ashcroft, L., Gergis, J., & Karoly, D. J. (2014). A historical climate dataset for southeastern Australia, 1788–1859. *Geoscience Data Journal*, *1*(2), 158–178. https://doi.org/10.1002/gdj3.19

Ashcroft, L., Karoly, D., & Gergis, J. (2012). Temperature variations of southeastern Australia, 1860–2011. *Australian Meteorological and Oceanographic Journal*, *62*(4), 227–245.

Ashcroft, L., Trewin, B., Benoy, M., Ray, D., & Courtney, C. (2022). The world's longest known parallel temperature dataset: A comparison between daily Glaisher and Stevenson screen temperature data at Adelaide, Australia, 1887–1947. *International Journal of Climatology*, *42*(5), 2670–2687. https://doi.org/10.1002/joc.7385

Aubry, T. J., Staunton-Sykes, J., Marshall, L. R., Haywood, J., Abraham, N. L., & Schmidt, A. (2021). Climate change modulates the stratospheric volcanic sulfate aerosol lifecycle and radiative forcing from tropical eruptions. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-24943-7

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

Auer, I., Böhm, R., Jurkovic, A., Lipa, W., Orlik, A., Potzmann, R., Schöner, W., Ungersböck, M., Matulla, C., Briffa, K., Jones, P., Efthymiadis, D., Brunetti, M., Nanni, T., Maugeri, M., Mercalli, L., Mestre, O., Moisselin, J. M., Begert, M., … Nieplova, E. (2007). HISTALP - Historical instrumental climatological surface time series of the

Greater Alpine Region. *International Journal of Climatology*, *27*(1), 17–46. https://doi.org/10.1002/joc.1377

Auer, I., Böhm, R., & Schöner, W. (2001). *Austrian long-term climate 1767-2000 multiple instrumental climate time series from central Europe* (P. Steinhauser, E. Bruckl, M. Hantel, H. Kromp-Kolb, M. Kuhn, & H. Mauritsch, Eds.). Zentralanstalt für Meteorologie und Geodynamik. http://www.zamg.ac.at/docs/wir_ueber_uns/cv/boehm_reinhard/2001-Auer-etal-OEBMG-aloclim.pdf

Awe, S. O., Mahony, M., Michaud, E., Murphy, C., Noone, S. J., Venema, V., Thorne, T. G., & Thorne, P. W. (2022). Insights from 20 years of temperature parallel measurements in Mauritius around the turn of the 20th century. *Climate of the Past*, *18*(4), 793–820. https://doi.org/10.5194/CP-18-793-2022

Azorin-Molina, C., Vicente-Serrano, S. M., Mcvicar, T. R., Jerez, S., Sanchez-Lorenzo, A., López-Moreno, J. I., Revuelto, J., Trigo, R. M., Lopez-Bustins, J. A., & Espírito-Santo, F. (2014). Homogenization and Assessment of Observed Near-Surface Wind Speed Trends over Spain and Portugal, 1961–2011. *Journal of Climate*, *27*(10), 3692–3712. https://doi.org/10.1175/JCLI-D-13-00652.1

Azoulay, A., Schmidt, H., & Timmreck, C. (2021). The Arctic Polar Vortex Response to Volcanic Forcing of Different Strengths. *Journal of Geophysical Research: Atmospheres*, *126*(11). https://doi.org/10.1029/2020JD034450

Bamford, A. J. (1928). On the exposure of thermometers in Ceylon. *Ceylon Journal of Science*, 153–167.

Begert, M., Schlegel, T., & Kirchhofer, W. (2005). Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000. *International Journal of Climatology*, *25*(1), 65–80. https://doi.org/10.1002/JOC.1118

Begert, M., Seiz, G., Schlegel, T., Musa, M., Baudraz, G., & Moesch, M. (2003). *No Homogenisierung von Klimamessreihen der Schweiz und Bestimmung der Normwerte 1961-1990Title*.

Bellouin, N., Rae, J., Jones, A., Johnson, C., Haywood, J., & Boucher, O. (2011). Aerosol forcing in the Climate Model Intercomparison Project (CMIP5) simulations by

HadGEM2-ES and the role of ammonium nitrate. *Journal of Geophysical Research: Atmospheres*, *116*(D20). https://doi.org/10.1029/2011JD016074

Bělohlávek, V. (1977). The Meteorological Tradition of Prague, Czechoslovakia. *Bulletin of the American Meteorological Society*, *58*(10), 1056–1057. https://doi.org/10.1175/1520-0477-58.10.1056

Bergström, H., & Moberg, A. (2002). Daily air temperature and pressure series for Uppsala (1722-1998). *Climatic Change*, *53*(1–3), 213–252. https://doi.org/10.1023/A:1014983229213

Bittner, M., Schmidt, H., Timmreck, C., & Sienz, F. (2016). Using a large ensemble of simulations to assess the Northern Hemisphere stratospheric dynamical response to tropical volcanic eruptions and its uncertainty. *Geophysical Research Letters*, *43*(17), 9324–9332. https://doi.org/10.1002/2016GL070587

Blanford, H. F. (1876a). *Report on the meteorology of India. 2nd (1876)*.

Blanford, H. F. (1876b). *The Indian meteorologist's vade-mecum. Part I: Instructions to Meteorological Observers in India*. https://babel.hathitrust.org/cgi/pt?id=hvd.hn4plq&view=1up&seq=107&q1=plate

Bližňák, V., Valente, M. A., & Bethke, J. (2015). Homogenization of time series from Portugal and its former colonies for the period from the late 19th to the early 21st century. *International Journal of Climatology*, *35*(9), 2400–2418. https://doi.org/10.1002/JOC.4151

Bluth, G. J. S., Doiron, S. D., Schnetzler, C. C., Krueger, A. J., & Walter, L. S. (1992). Global tracking of the SO2 clouds from the June, 1991 Mount Pinatubo eruptions. *Geophysical Research Letters*, *19*(2), 151–154. https://doi.org/10.1029/91GL02792

Böhm, R., Auer, I., Brunetti, M., Maugeri, M., Nanni, T., & Schöner, W. (2001). Regional temperature variability in the European Alps: 1760-1998 from homogenized instrumental time series. *International Journal of Climatology*, *21*(14), 1779–1801. https://doi.org/10.1002/joc.689

Böhm, R., Jones, P. D., Hiebl, J., Frank, D., Brunetti, M., & Maugeri, M. (2010). The early instrumental warm-bias: A solution for long central European temperature series 1760-2007. *Climatic Change*, *101*(1), 41–67. https://doi.org/10.1007/s10584-009-9649-4

Boylan, L. (1969). *Communist-Bloc Meteorological Systems and Sensors*. https://archive.org/details/DTIC_AD0689767/mode/1up

Bradley, R. S. (2014). Paleoclimatology: Reconstructing Climates of the Quaternary: Third Edition. *Paleoclimatology: Reconstructing Climates of the Quaternary: Third Edition*, 1–675. https://doi.org/10.1016/C2009-0-18310-1

Bradley, R. S. (2015). Chapter 5. Ice Cores. In *Paleoclimatology* (pp. 137–194). Elsevier. https://doi.org/10.1016/B978-0-12-386913-5.00005-3

Brandsma, T., & van der Meulen, J. P. (2008). Thermometer screen intercomparison in De Bilt (the Netherlands)—Part II: description and modeling of mean temperature differences and extremes. *International Journal of Climatology*, *28*(3), 389–400. https://doi.org/10.1002/joc.1524

Brandt, M. L. (1994). *Summary of Meta Data from NACD-stations in Demark, Greenland and the Faroe Islands 1872-1994*. https://www.dmi.dk/fileadmin/user_upload/Rapporter/TR/1994/tr94-20.pdf

Brazdil, R. (2012). *History of Weather and Climate in the Czech Lands* (9th ed.). https://www.academia.edu/20197435/Brázdil_R_et_al_2012_History_of_Weather_and_Climate_in_the_Czech_Lands_IX_Temperature_and_Precipitation_Fluctuation_in_the_Czech_Lands_during_the_Instrumental_Period_Masaryk_University_Brno_236_pp

Brázdil, R., & Budíková, M. (1999). An urban bias in air temperature fluctuations at the Klementinum, Prague, The Czech Republic. *Atmospheric Environment*, *33*(24–25), 4211–4217. https://doi.org/10.1016/S1352-2310(99)00163-6

Brázdil, R., Zahradníček, P., Pišoft, P., Štěpánek, P., Bělínová, M., & Dobrovolný, P. (2012). Temperature and precipitation fluctuations in the Czech Republic during the period of instrumental measurements. *Theoretical and Applied Climatology*, *110*(1–2), 17–34. https://doi.org/10.1007/s00704-012-0604-3

Briffa, K. R. (1995). Interpreting High-Resolution Proxy Climate Data — The Example of Dendroclimatology. In *Analysis of Climate Variability* (pp. 77–94). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-03167-4_5

Briffa, K. R. (2000). Annual climate variability in the Holocene: Interpreting the message of ancient trees. *Quaternary Science Reviews*, *19*(1–5), 87–105. https://doi.org/10.1016/S0277-3791(99)00056-6

Briffa, K. R., Osborn, T. J., Schweingruber, F. H., Jones, P. D., Shiyatov, S. G., & Vaganov, E. A. (2002). Tree-ring width and density data around the Northern Hemisphere: Part 1, local and regional climate signals. *The Holocene*, *12*(6), 737–757. https://doi.org/10.1191/0959683602hl587rp

Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. B., & Jones, P. D. (2006). Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *Journal of Geophysical Research Atmospheres*, *111*(12). https://doi.org/10.1029/2005JD006548

Brönnimann, S. (2015). Climatic Changes Since 1700. *Brönnimann, Stefan (2015). Climatic Changes Since 1700. Advances in Global Change Research: Vol. 55. Cham: Springer 10.1007/978-3-319-19042-6*, *55*. https://doi.org/10.1007/978-3-319-19042-6

Brönnimann, S., Allan, R., Ashcroft, L., Baer, S., Barriendos, M., Brázdil, R., Brugnara, Y., Brunet, M., Brunetti, M., Chimani, B., Cornes, R., Domínguez-Castro, F., Filipiak, J., Founda, D., Herrera, R. G., Gergis, J., Grab, S., Hannak, L., Huhtamaa, H., … Wyszyński, P. (2019a). Unlocking pre-1850 instrumental meteorological records: A global inventory. *Bulletin of the American Meteorological Society*. https://doi.org/10.1175/bams-d-19-0040.1

Brönnimann, S., Allan, R., Ashcroft, L., Baer, S., Barriendos, M., Brázdil, R., Brugnara, Y., Brunet, M., Brunetti, M., Chimani, B., Cornes, R., Domínguez-Castro, F., Filipiak, J., Founda, D., Herrera, R. G., Gergis, J., Grab, S., Hannak, L., Huhtamaa, H., … Wyszyński, P. (2019b). Unlocking Pre-1850 Instrumental Meteorological Records: A Global Inventory. *Bulletin of the American Meteorological Society*, *100*(12), ES389–ES413. https://doi.org/10.1175/BAMS-D-19-0040.1

Brönnimann, S., Annis, J. L., Vogler, C., & Jones, P. D. (2007). Reconstructing the quasi-biennial oscillation back to the early 1900s. *Geophysical Research Letters*, *34*(22). https://doi.org/10.1029/2007GL031354

Brugnara, Y., Pfister, L., Villiger, L., Rohr, C., Alessandro Isotta, F., & Brönnimann, S. (2020). Early instrumental meteorological observations in Switzerland: 1708-1873.

Earth System Science Data, *12*(2), 1179–1190. https://doi.org/10.5194/ESSD-12-1179-2020

Brunet, M., Asin, J., Sigró, J., Bañón, M., García, F., Aguilar, E., Palenzuela, J. E., Peterson, T. C., & Jones, P. (2011). The minimization of the screen bias from ancient Western Mediterranean air temperature records: An exploratory statistical analysis. *International Journal of Climatology*, *31*(12), 1879–1895. https://doi.org/10.1002/joc.2192

Brunet, M., Brugnara, Y., Noone, S., Stephens, A., Valente, M. A., Ventura, C., Jones, P., Gilabert, A., Brönnimann, S., Luterbacher, J., Allan, R., Brohan, P., & Compo, G. P. (2020). *Practice Guidelines for Climate Data and Metadata Formatting, Quality Control and Submission of the Copernicus Climate Change Service Data Rescue Service*. https://doi.org/10.24381/kctk-8j22

Brunet, M., Saladié, O., Jones, P., Sigró, J., Aguilar, E., Moberg, A., Lister, D., Walther, A., Lopez, D., & Almarza, C. (2006). The development of a new dataset of Spanish daily adjusted temperature series (SDATS) (1850-2003). *International Journal of Climatology*, *26*(13), 1777–1802. https://doi.org/10.1002/joc.1338

Brunetti, M., Maugeri, M., Monti, F., & Nanni, T. (2006). Temperature and precipitation variability in Italy in the last two centuries from homogenised instrumental time series. *International Journal of Climatology*, *26*(3), 345–381. https://doi.org/10.1002/joc.1251

Brys, K., & Brys, T. (2010). The First One Hundred Years (1791–1890) of the Wrocław Air Temperature Series. In R. Przybylak, J. Majorowicz, & M. Brazdil, R. Kejna (Eds.), *The Polish Climate in the European Context: An Historical Overview* (pp. 485–524). Springer. https://link.springer.com/book/10.1007/978-90-481-3167-9

Buchan, A. (1872). Remarks on the climate of Jerusalem. *Palestine Exploration Quarterly*, *4*(1), 19–30. https://doi.org/10.1179/PEQ.1872.005/ASSET//CMS/ASSET/899835B0-FF3C-440D-9266-7E4A063BCF7E/PEQ.1872.005.FP.PNG

Büntgen, U., Esper, J., Frank, D. C., Nicolussi, K., & Schmidhalter, M. (2005). A 1052-year tree-ring proxy for Alpine summer temperatures. *Climate Dynamics*, *25*(2–3), 141–153. https://doi.org/10.1007/s00382-005-0028-1

Büntgen, U., Frank, D. C., Nievergelt, D., & Esper, J. (2006). Summer temperature variations in the European Alps, A.D. 755-2004. *Journal of Climate*, *19*(21), 5606–5623. https://doi.org/10.1175/JCLI3917.1

Büntgen, U., Frank, D. C., Schmidhalter, M., Neuwirth, B., Seifert, M., & Esper, J. (2006). Growth/climate response shift in a long subalpine spruce chronology. *Trees - Structure and Function*, *20*(1), 99–110. https://doi.org/10.1007/s00468-005-0017-3

Bureau Central Meteorologique de France. (1881). *Instructions Meteorologiques*. http://bibliotheque.meteo.fr/exl-php/docs/mf_-_internet_bibliotheque_numerique_-_fonds_ancien/12476/iso00009849__PDF.txt

Bureau of Meteorology. (2011). *A short history of the Bureau of Meteorology*. https://media.bom.gov.au/social/blog/10/a-short-history-of-the-bureau-of-meteorology/

Burgdorf, A. M. (2022). A global inventory of quantitative documentary evidence related to climate since the 15th century. *Climate of the Past*, *18*(6), 1407–1428. https://doi.org/10.5194/CP-18-1407-2022

Burt, S. (2022, September 14). Oxford and Durham - The Two Longest Climatological Records in England. *Royal Meteorological Society History Special Interest Group Talk*.

Burt, S., & Burt, T. (2019). *Oxford Weather and Climate Since 1767*.

Burt, S., & Podesta, M. (2020). Response times of meteorological air temperature sensors. *Quarterly Journal of the Royal Meteorological Society*, *146*(731), 2789–2800. https://doi.org/10.1002/qj.3817

Butler, C. J., Garcia Suarez, A. M., Coughlin, A. D. S., & Morrell, C. (2005). AIR TEMPERATURES AT ARMAGH OBSERVATORY, NORTHERN IRELAND,FROM 1796 TO 2002. *International Journal of Climatology*, *25*, 1055–1079. https://doi.org/DOI: 10.1002/joc.1148

Butler, C. J., Garcıa-Suarez, A. M., Coughlin, A. D. S., & Cardwell, D. (2005). *Meteorological Data recorded at Armagh Observatory: Vol 2 - Daily, Mean Monthly, Seasonal and Annual, Maximum and Minimum Temperatures, 1844-2004*. http://www.climate.armagh.ac.uk/calibrated/airtemp/yellow-textD.pdf

Calvet, C. (1985). L'effet thermique d'activite humaine a Paris. *La Meteorologie*, *7*(10), 34–42. https://gallica.bnf.fr/ark:/12148/bpt6k9632452t/f38.item.r=abri

Calvo, N., García-Herrera, R., & Garcia, R. R. (2008). The ENSO signal in the stratosphere. *Annals of the New York Academy of Sciences*, *1146*, 16–31. https://doi.org/10.1196/ANNALS.1446.008

Camuffo, D. (2002a). Calibration and instrumental errors in early measurements of air temperature. *Climatic Change*, *53*(1–3), 297–329. https://doi.org/10.1023/A:1014914707832

Camuffo, D. (2002b). Errors in early temperature series arising from changes in style of measuring time, sampling schedule and number of observations. *Climatic Change*, *53*(1–3), 331–352. https://doi.org/10.1023/A:1014962623762

Camuffo, D. (2002c). History of the long series of daily air temperature in Padova (1725-1998). *Climatic Change*, *53*(1–3), 7–75. https://doi.org/10.1023/A:1014958506923

Camuffo, D., & Bertolin, C. (2012). The earliest temperature observations in the world: The Medici Network (1654-1670). *Climatic Change*, *111*(2), 335–363. https://doi.org/10.1007/s10584-011-0142-5

Camuffo, D., della Valle, A., & Becherini, F. (2023). Instrumental and Observational Problems of the Earliest Temperature Records in Italy: A Methodology for Data Recovery and Correction. *Climate*, *11*(9), 178. https://doi.org/10.3390/cli11090178

Camuffo, D., & Jones, P. (Eds.). (2002). *Improved Understanding of Past Climatic Variability from Early Daily European Instrumental Sources*. Springer Netherlands. https://doi.org/10.1007/978-94-010-0371-1

Candolle, A. P. (1825). *Circulare mit Instruction für die im Namen der allgemeinen schweizerischen naturforschenden Gesellschaft anzustellenden barometrischen und thermometrischen Beobachtungen*. https://www.e-rara.ch/zuz/content/titleinfo/10068212

Cappelen, J. (2021a). *Denmark – DMI Historical Climate Data Collection 1768- 2020*. https://www.dmi.dk/fileadmin/Rapporter/2021/DMIRep21-02.pdf

Cappelen, J. (2021b). *Greenland – DMI Historical Climate Data Collection 1784- 2020*. https://www.dmi.dk/fileadmin/Rapporter/2021/DMIRep21-04.pdf

Cassidy, D. C. (1985). Meterology in Mannheim: The Palatine Meterological Society, 1780-1795. *Sudhoffs Archiv*, *69*(1), 8–25. https://www.jstor.org/stable/20776952

Caussinus, H., & Mestre, O. (2004). Detection and correction of artificial shifts in climate series. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *53*(3), 405–425. https://doi.org/10.1111/j.1467-9876.2004.05155.x

Central Meteorological Observatory of Japan. (1893a). *Organisation of the meteorological system in Japan*. https://catalog.hathitrust.org/Record/009262018

Central Meteorological Observatory of Japan. (1893b). *The Climate of Japan*. https://hdl.handle.net/2027/uc2.ark:/13960/t9x060n4c

Chambers, C. (1878). *The Meteorology of the Bombay Presidency*. Her Majesty's Stationary Office.

Chandler, R. E., Thorne, P., Lawrimore, J., & Willett, K. (2012). Building trust in climate science: Data products for the 21st century. *Environmetrics*, *23*(5), 373–381. https://doi.org/10.1002/env.2141

Chandler, T. J. (1964). North-wall and stevenson screen temperatures at kew observatory. *Quarterly Journal of the Royal Meteorological Society*, *90*(385), 332–333. https://doi.org/10.1002/qj.49709038514

Chaplin, T. (1883). Observations on the climate of Jerusalem. *Palestine Exploration Quarterly*, *15*(1), 8–40. https://doi.org/10.1179/PEQ.1883.15.1.8/ASSET//CMS/ASSET/C3752C17-8A43-4064-9172-4C8C98A30D18/PEQ.1883.15.1.8.FP.PNG

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

Chenoweth, M. (1992). A Possible Discontinuity in the U.S. Historical Temperature Record. *Journal of Climate*, *5*(10), 1172–1179. https://doi.org/10.1175/1520-0442(1992)005<1172:APDITU>2.0.CO;2

Chenoweth, M. (1993). Nonstandard thermometer exposures at US cooperative weather stations during the late nineteenth century. *Journal of Climate*, *6*(9), 1787–1797. https://doi.org/10.1175/1520-0442(1993)006<1787:NTEAUC>2.0.CO;2

Christiansen, B. (2008). Volcanic eruptions, large-scale modes in the Northern Hemisphere, and the El Niño-Southern Oscillation. *Journal of Climate*, *21*(5), 910–922. https://doi.org/10.1175/2007JCLI1657.1

Climatological Branch of the Iranian Meteorological Department. (1958). *Meteorological Yearbook 1956*. https://library.oarcloud.noaa.gov/docs.lib/htdocs/rescue/cd377_pdf/LSN3484.PDF

Clyne, M., Lamarque, J. F., Mills, M. J., Khodri, M., Ball, W., Bekki, S., Dhomse, S. S., Lebas, N., Mann, G., Marshall, L., Niemeier, U., Poulain, V., Robock, A., Rozanov, E., Schmidt, A., Stenke, A., Sukhodolov, T., Timmreck, C., Toohey, M., … Toon, O. B. (2021). Model physics and chemistry causing intermodel disagreement within the VolMIP-Tambora Interactive Stratospheric Aerosol ensemble. *Atmospheric Chemistry and Physics*, *21*(5), 3317–3343. https://doi.org/10.5194/acp-21-3317-2021

Cole-Dai, J., Ferris, D., Lanciki, A., Savarino, J., Baroni, M., & Thiemens, M. H. (2009). Cold decade (AD 1810–1819) caused by Tambora (1815) and another (1809) stratospheric volcanic eruption. *Geophysical Research Letters*, *36*(22). https://doi.org/10.1029/2009GL040882

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., … Worley, S. J. (2011). The Twentieth Century Reanalysis Project. *Quarterly Journal of the Royal Meteorological Society*, *137*(654), 1–28. https://doi.org/10.1002/qj.776

Compo, G., Slivinski, L. C., Whitaker, J. S., Sardeshmukh, P. D., McColl, C., Brohan, P., Allan, R., Yin, X., Vose, R., Spencer, L. J., Ashcroft, L., Bronnimann, S., Brunet, M., Camuffo, D., Cornes, R., Cram, T. A., Crouthamel, R., Dominguez-Castro, F.,

Freeman, J. E., … Wyszynski, P. (2019). *The International Surface Pressure Databank version 4. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory.* https://psl.noaa.gov/data/ISPD/

Conrad, V., & Pollak, C. (1950). *Methods in Climatology*. Harvard University Press.

Cook, B. I., Shukla, S. P., Puma, M. J., & Nazarenko, L. S. (2014). Irrigation as an historical climate forcing. *Climate Dynamics*, *44*, 1715–1730.

Cooke, E. (1897). *Western Australia Meteorological Report for the year 1894*. https://hdl.handle.net/2027/uc1.c2999439

Cotte, L. (1788). *Memoires sur La Meteorologie Pour fervir de Suite et de Supplement Au Traite de Meteorologie publue en 1774*. De L'Imprimerie Royale. http://bibliotheque.meteo.fr/exl-php/util/documents/accede_document.php?1646740743515

Coupe, J., & Robock, A. (2021). The Influence of Stratospheric Soot and Sulfate Aerosols on the Northern Hemisphere Wintertime Atmospheric Circulation. *Journal of Geophysical Research: Atmospheres*, *126*(11). https://doi.org/10.1029/2020JD034513

Cowtan, K., Rohde, R., & Hausfather, Z. (2018). Evaluating biases in sea surface temperature records using coastal weather stations. *Quarterly Journal of the Royal Meteorological Society*, *144*(712), 670–681. https://doi.org/10.1002/qj.3235

Cowtan, K., & Way, R. G. (2014). Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Quarterly Journal of the Royal Meteorological Society*, *140*(683), 1935–1944. https://doi.org/10.1002/qj.2297

Crowley, T. J. (2000). Causes of climate change over the past 1000 years. *Science*, *289*(5477), 270–277. https://doi.org/10.1126/science.289.5477.270

Crowley, T. J., & Unterman, M. B. (2013). Technical details concerning development of a 1200 yr proxy index for global volcanism. *Earth System Science Data*, *5*(1), 187–197. https://doi.org/10.5194/essd-5-187-2013

Crowley, T. J., Zielinski, G., Vinther, B., Udisti, R., Kreutz, K., Cole-Dai, J., & Castellano, E. (2008). Volcanism and the Little Ice Age. *PAGES News*, *16*(2), 22–23. https://doi.org/10.1029/2002GL0166335

Cubash, U., & Kadow, C. (2011). Global Climate Change and Aspects of Regional Climate Change in the Berlin-Brandenburg Region. *Global Change: Challenges for Regional Water Resources*, 3–20.

Cucchi, M., Weedon, G., Amici, A., Bellouin, N., Lange, S., Schmied, H. M., Hersbach, H., & Buontempo, C. (2020). WFDE5: bias adjusted ERA5 reanalysis data for impact studies. *Earth System Science Data Discussions*, *12*, 2097–2120. https://doi.org/10.5194/essd-12-2097-2020

Dallasanta, K., & Polvani, L. M. (2022). Volcanic stratospheric injections up to 160 Tg(S) yield a Eurasian winter warming indistinguishable from internal variability. *Atmos. Chem. Phys*, *22*, 8843–8862. https://doi.org/10.5194/acp-22-8843-2022

D'Arrigo, R., Wilson, R., & Jacoby, G. (2006). On the long-term context for late twentieth century warming. *Journal of Geophysical Research*, *111*(D3), D03103. https://doi.org/10.1029/2005JD006352

Dätwyler, C., Abram, N. J., Grosjean, M., Wahl, E. R., & Neukom, R. (2019). El Niño–Southern Oscillation variability, teleconnection changes and responses to large volcanic eruptions since AD 1000. *International Journal of Climatology*, *39*(5), 2711–2724. https://doi.org/10.1002/joc.5983

Davis, G. (1887). *Anales de la Oficina Meteorológica Argentina Tomo V*. https://hdl.handle.net/2027/mdp.39015086762286

Davis, G. (1889). *Anales de la Oficina Meteorológica Argentina Tomo VII*. https://hdl.handle.net/2027/mdp.39015086762443

Davis, G. (1893). *Anales de la Oficina Meteorológica Argentina Tomo IX*. https://hdl.handle.net/2027/mdp.39015086762609

Davis, G. (1895). *Anales de la Oficina Meteorológica Argentina Tomo X*. https://hdl.handle.net/2027/mdp.39015086762617

de Carvalho, D. (1917). *Meteorologie du Bresil*. John Bale, Sons & Danielsson, Ltd. http://catalogue.bnf.fr/ark:/12148/cb32010925p

de Cortázar-Atauri, I. G., Daux, V., Garnier, E., Yiou, P., Viovy, N., Seguin, B., Boursiquot, J. M., Parker, A. K., van Leeuwen, C., & Chuine, I. (2010). Climate reconstructions

from grape harvest dates: Methodology and uncertainties. *Holocene*, *20*(4), 599–608. https://doi.org/10.1177/0959683609356585

Dee, D. P., Balmaseda, M., Balsamo, G., Engelen, R., Simmons, A. J., & Thépaut, J. N. (2014). Toward a Consistent Reanalysis of the Climate System. *Bulletin of the American Meteorological Society*, *95*(8), 1235–1248. https://doi.org/10.1175/BAMS-D-13-00043.1

Dee, D., Thorne, P., Noone, S., Baddour, O., & Tassone, C. (2021). The Critical Role of Observations in Informing Climate Science, Assessment and Policy. *WMO Bulletin*, *70*(2). https://public.wmo.int/en/resources/bulletin/critical-role-of-observations-informing-climate-science-assessment-and-policy

Dee, S. G., Cobb, K. M., Emile-Geay, J., Ault, T. R., Lawrence Edwards, R., Cheng, H., & Charles, C. D. (2020). No consistent ENSO response to volcanic forcing over the last millennium. *Science*, *367*(6485). https://doi.org/10.1126/science.aax2000

Dee, S. G., & Steiger, N. J. (2022). ENSO's Response to Volcanism in a Data Assimilation-Based Paleoclimate Reconstruction Over the Common Era. *Paleoceanography and Paleoclimatology*, *37*(3). https://doi.org/10.1029/2021PA004290

Demarée, G. R., Lachaert, P.-J., Verhoeve, T., & Thoen, E. (2002). *THE LONG-TERM DAILY CENTRAL BELGIUM TEMPERATURE (CBT) SERIES (1767-1998) AND EARLY INSTRUMENTAL METEOROLOGICAL OBSERVATIONS IN BELGIUM*.

Denza, P. F. (1882). *Instructions for Meteorological Observations and Barometric Altimetry*.

Deshler, T. (2008). A review of global stratospheric aerosol: Measurements, importance, life cycle, and local stratospheric aerosol. *Atmospheric Research*, *90*(2–4), 223–232. https://doi.org/10.1016/j.atmosres.2008.03.016

Detwiller, J. (1978). L'evolution seculaire de la temperature a Paris. *La Météorologie*.

Deutsche Seewarte. (1878). *Meteorologische Beobachtungen in Deutschland angestellt an 17 Stationen Zweiter Ordnung im Jahre 1876*. https://hdl.handle.net/2027/pst.000007229206

Deutsche Seewarte. (1882). *Meteorologische Beobachtungen in Deutschland angestellt an 18 Stationen Zweiter Ordnung im Jahre 1880*. https://hdl.handle.net/2027/pst.000007229220

Deutsche Seewarte. (1884). *Meteorologische Beobachtungen in Deutschland angestellt an 18 Stationen Zweiter Ordnung im Jahre 1882*. https://hdl.handle.net/2027/pst.000007229237

Deutsche Seewarte. (1889). *Deutsches meteorologisches Jahrbuch fur 1888. Ergebnisse der Meteorologischen Beobachtungen an 9 Stationen II. Ordnung, an 9 Normal-Beobachtungs-Stationen in stundlichen Aufzeichnungen und an 43 Signalstellen*. https://hdl.handle.net/2027/pst.000007229268

Dienst, M., Lindén, J., Engström, E., & Esper, J. (2017). Removing the relocation bias from the 155-year Haparanda temperature record in Northern Europe. *International Journal of Climatology*, *37*(11), 4015–4026. https://doi.org/10.1002/JOC.4981

Dines, L. H. G. (1921). *A comparison between the dry bulb temperature in the climatological screen at Valencia Observatory and that in a Stevenson screen exposed in an open field adjoining* (pp. 30–40). His Majesty's Stationary Office. https://digital.nmla.metoffice.gov.uk/file/io_8b9a812f-db8d-4cbb-9412-fddd2cc9a149

Dirksen, M., Knap, W. H., Steeneveld, G. J., Holtslag, A. A. M., & Tank, A. M. G. K. (2020). Downscaling daily air-temperature measurements in the Netherlands. *Theoretical and Applied Climatology*, *142*(1–2), 751–767. https://doi.org/10.1007/S00704-020-03313-1/FIGURES/5

Doberck, W. (1887a). *Observations and Researches made at the Hongkong Observatory in the year 1886*.

Doberck, W. (1887b). *Special Series No. 7 Instructions for Making Meteorological Observations prepared for use in China and The Law of Storms in the Eastern Seas*. http://gis.rchss.sinica.edu.tw/cmcs/wp-content/uploads/2013/09/Instructions-for-Making-Meteorological-Observations-prepared-for-Use-in-China-and-the-Law-of-Storms-in-the-Eastern-Seas.pdf

Dobrovolný, P., Moberg, A., Brázdil, R., Pfister, C., Glaser, R., Wilson, R., van Engelen, A., Limanówka, D., Kiss, A., Halíčková, M., Macková, J., Riemann, D., Luterbacher, J., & Böhm, R. (2010). Monthly, seasonal and annual temperature reconstructions for

Central Europe derived from documentary evidence and instrumental records since AD 1500. *Climatic Change*, *101*(1), 69–107. https://doi.org/10.1007/s10584-009-9724-x

Domínguez-Castro, F., Vaquero, J. M., Gallego, M. C., Farrona, A. M. M., Antunã-Marrero, J. C., Cevallos, E. E., Herrera, R. G., De La Guiá, C., Mejía, R. D., Naranjo, J. M., Del Rosario Prieto, M., Ramos Guadalupe, L. E., Seiner, L., Trigo, R. M. H., & Villacís, M. (2017). Early meteorological records from Latin-America and the Caribbean during the 18th and 19th centuries. *Scientific Data*, *4*(1), 1–10. https://doi.org/10.1038/sdata.2017.169

Domínguez-Castro, F., Vaquero, J. M., Rodrigo, F. S., Farrona, A. M. M., Gallego, M. C., García-Herrera, R., Barriendos, M., & Sanchez-Lorenzo, A. (2014). Early Spanish meteorological records (1780-1850). *International Journal of Climatology*, *34*(3), 593–603. https://doi.org/10.1002/joc.3709

Domonkos, P. (2011). Homogenising time series: beliefs, dogmas and facts. *Advances in Science and Research*, *6*(1), 167–172. https://doi.org/10.5194/asr-6-167-2011

Domonkos, P. (2022). Automatic Homogenization of Time Series: How to Use Metadata? *Atmosphere 2022, Vol. 13, Page 1379*, *13*(9), 1379. https://doi.org/10.3390/ATMOS13091379

Domonkos, P., Guijarro, J. A., Venema, V., Brunet, M., & Sigro, J. (2021). Efficiency of time series homogenization: method comparison with 12 monthly temperature test datasets. *Journal of Climate*, *34*(8), 2877–2891.

Driscoll, S., Bozzo, A., Gray, L. J., Robock, A., & Stenchikov, G. (2012). Coupled Model Intercomparison Project 5 (CMIP5) simulations of climate following volcanic eruptions. *Journal of Geophysical Research Atmospheres*, *117*(17). https://doi.org/10.1029/2012JD017607

Ducré-Robitaille, J. F., Vincent, L. A., & Boulet, G. (2003). Comparison of techniques for detection of discontinuities in temperature series. *International Journal of Climatology*, *23*(9), 1087–1101. https://doi.org/10.1002/JOC.924

Eade, R., Smith, D., Scaife, A., Wallace, E., Dunstone, N., Hermanson, L., & Robinson, N. (2014). Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophysical Research Letters*, *41*(15), 5620–5628. https://doi.org/10.1002/2014GL061146

Eade, R., Stephenson, · D B, Scaife, · A A, & Smith, · D M. (2022). *Quantifying the rarity of extreme multi-decadal trends: how unusual was the late twentieth century trend in the North Atlantic Oscillation? 58*, 1555–1568. https://doi.org/10.1007/s00382-021-05978-4

Ellis, W. (1891). On the comparison of thermometrical observations made in a Stevenson screen with corresponding observations made on the revolving stand at the royal observatory, Greenwich. *Quarterly Journal of the Royal Meteorological Society*, *17*(80), 240–249. https://doi.org/10.1002/qj.4970178006

Ellis, W. (1897). Report on the International Meteorological Conference, held at Paris, September 17 to 23, 1896. *Quarterly Journal of the Royal Meteorological Society*, *23*(101), 1–10.

Emile-Geay, J., McKay, N. P., Kaufman, D. S., Von Gunten, L., Wang, J., Anchukaitis, K. J., Abram, N. J., Addison, J. A., Curran, M. A. J., Evans, M. N., Henley, B. J., Hao, Z., Martrat, B., McGregor, H. V., Neukom, R., Pederson, G. T., Stenni, B., Thirumalai, K., Werner, J. P., … Zinke, J. (2017). A global multiproxy database for temperature reconstructions of the Common Era. *Scientific Data*, *4*(170088), 1–33. https://doi.org/10.1038/sdata.2017.88

Esper, J., Cook, E. R., & Schweingruber, F. H. (2002). Low-frequency signals in long tree-ring chronologies for reconstructing past temperature variability. *Science*, *295*(5563), 2250–2253. https://doi.org/10.1126/science.1066208

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., Collins, W. D., Gier, B. K., Hall, A. D., Hoffman, F. M., Hurtt, G. C., Jahn, A., Jones, C. D., Klein, S. A., Krasting, J. P., Kwiatkowski, L., Lorenz, R., Maloney, E., Meehl, G. A., … Williamson, M. S. (2019). Taking climate model evaluation to the next level. *Nature Climate Change 2019 9:2*, *9*(2), 102–110. https://doi.org/10.1038/s41558-018-0355-y

Ferguson, C. R., & Villarini, G. (2014). An evaluation of the statistical homogeneity of the Twentieth Century Reanalysis. *Climate Dynamics*, *42*(11–12), 2841–2866. https://doi.org/10.1007/S00382-013-1996-1/FIGURES/16

Fickeler, P. (1926). Establishment of Meteorological Stations in Mongolia. *Monthly Weather Review*, *54*(9), 386. https://journals.ametsoc.org/view/journals/mwre/54/9/1520-0493_1926_54_386a_eomsim_2_0_co_2.xml?tab_body=pdf

Field, J. H. (1920). On exposures of thermometers in India. In *India Meteorological Memoirs* (pp. 21–73).

Fischer, E. A., Luterbacher, J., Zorita, E., Tett, S. F. B., Casty, C., & Wanner, H. (2007). European climate response to tropical volcanic eruptions over the last half millennium. *Geophysical Research Letters*, *34*(5). https://doi.org/10.1029/2006GL027992

Flora, S. D. (1920). Shading instrument shelters. *Monthly Weather Review*, *48*(5), 271–271. https://doi.org/https://doi.org/10.1175/1520-0493(1920)48<271:SIS>2.0.CO;2

Folland, C. K., Karl, T. R., & Vinnikov, K. Y. R. (1990). Observed Climate Variations and Change. In J. J. Houghton, G. J. Jenkins, & J. J. Ephraums (Eds.), *Climate change: The IPCC Scientific Assessment*. Cambridge University Press.

Folland, C. K., & Parker, D. E. (1995). Correction of instrumental biases in historical sea surface temperature data. *Quarterly Journal of the Royal Meteorological Society*, *121*(522), 319–367. https://doi.org/10.1002/qj.49712152206

Folland, C. K., Rayner, N. A., Brown, S. J., Smith, T. M., Shen, S. S. P., Parker, D. E., Macadam, I., Jones, P. D., Jones, R. N., Nicholls, N., & Sexton, D. M. H. (2001). Global temperature change and its uncertainties since 1861. *Geophysical Research Letters*, *28*(13), 2621–2624. https://doi.org/10.1029/2001GL012877

Forster, P., Ramaswamy, V., Artaxo, P., Berntsen, T., Betts, R., Fahey, D. W., Haywood, J., Lean, J., Lowe, D. C., Myhre, G., Nganga, J., Prinn, R., Raga, G., Schulz, M., & Van Dorland, R. (2007). Changes in Atmospheric Constituents and in Radiative Forcing. In S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, & H. L. Miller (Eds.), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press. https://archive.ipcc.ch/publications_and_data/ar4/wg1/en/ch2.html

Founda, D., Kambezidis, H. D., Petrakis, M., Zanis, P., & Zerefos, C. (2009). A correction of the recent air-temperature record at the historical meteorological station of the National Observatory of Athens (NOA) due to instrument change. *Theoretical and Applied Climatology*, *97*(3–4), 385–389. https://doi.org/10.1007/S00704-008-0084-7/FIGURES/3

Frank, D., Büntgen, U., Böhm, R., Maugeri, M., & Esper, J. (2007). Warmer early instrumental measurements versus colder reconstructed temperatures: shooting at a moving target. *Quaternary Science Reviews*, *26*(25–28), 3298–3310. https://doi.org/10.1016/j.quascirev.2007.08.002

Frank, D., & Esper, J. (2005). Temperature reconstructions and comparisons with instrumental data from a tree-ring network for the European Alps. *International Journal of Climatology*, *25*(11), 1437–1454. https://doi.org/10.1002/joc.1210

Frank, D., Fang, K., Fonti, P., Frank, D., Fang, K., & Fonti, P. (2022). Dendrochronology: Fundamentals and Innovations. *Tree Physiology*, *8*, 21–59. https://doi.org/10.1007/978-3-030-92698-4_2

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

Freie Universitat Berlin. (n.d.). *The Quasi-Biennial-Oscillation (QBO) Data Serie*. Retrieved 7 August 2023, from https://www.geo.fu-berlin.de/en/met/ag/strat/produkte/qbo/

Frolich, O. (1876). *Uber die warme des himmels, die temperatur des weltraums und die mittlere temperatur der atmosphare*. https://library.oarcloud.noaa.gov/foreign_climate.lib/FCD_002_pdf/QC989R8M48t6.pdf

Gamble, J. G. (1879). XVI. Summer and Winter Temperature in South Africa. *Http://Dx.Doi.Org/10.1080/21560382.1879.9526153*, *2*(3), 104–109. https://doi.org/10.1080/21560382.1879.9526153

Garrison, C., Kilburn, C., Smart, D., & Edwards, S. (2021). The blue suns of 1831: Was the eruption of Ferdinandea, near Sicily, one of the largest volcanic climate forcing events of the nineteenth century? *Climate of the Past*, *17*(6), 2607–2632. https://doi.org/10.5194/cp-17-2607-2021

Garrison, C. S., Kilburn, C. R. J., & Edwards, S. J. (2018). The 1831 eruption of Babuyan Claro that never happened: Has the source of the one of the largest volcanic climate forcing events of the nineteenth century been misattributed? *Journal of Applied Volcanology*, *7*(1), 8. https://doi.org/10.1186/s13617-018-0078-9

Gaster, F. (1882). Report on experiments made at Strathfield Turgiss in 1869 with stands or screens of various patterns, devised and employed for the exposing of thermometers, in order to determine the temperature of the air'. *Appendix II to the Quarterly Weather Report for 1879*.

Gergis, J., Baillie, Z., Ashcroft, L., Trewin, B., & Allan, R. J. (2022). Consolidating historical instrumental observations in southern Australia for assessing pre-industrial weather and climate variability. *Climate Dynamics*. https://doi.org/10.1007/s00382-022-06573-x

Gettelman, A., Geer, A. J., Forbes, R. M., Carmichael, G. R., Feingold, G., Posselt, D. J., Stephens, G. L., van den Heever, S. C., Varble, A. C., & Zuidema, P. (2022). The future of Earth system prediction: Advances in model-data fusion. *Science Advances*, *8*(14), 3488. https://doi.org/10.1126/SCIADV.ABN3488/ASSET/386C78C3-2DD4-46B5-9B17-5EAE9C6A4C57/ASSETS/IMAGES/LARGE/SCIADV.ABN3488-F3.JPG

Gill, D. (1882). On the effect of different kinds of thermometer screens, and of different exposures, in estimating the diurnal range of temperature at the royal observatory, cape of good hope. *Quarterly Journal of the Royal Meteorological Society*, *8*(44), 238–243. https://doi.org/10.1002/qj.4970084404

Gillespie, I., Haimberger, L., Compo, G. P., & Thorne, P. W. (2023). Assessing homogeneity of land surface air temperature observations using sparse-input reanalyses. *International Journal of Climatology*, *43*(2), 736–760. https://doi.org/10.1002/joc.7822

Gillespie, I. M., Haimberger, L., Compo, G. P., & Thorne, P. W. (2021). Assessing potential of sparse-input reanalyses for centennial-scale land surface air temperature

homogenisation. *International Journal of Climatology*, *41*(S1), E3000–E3020. https://doi.org/10.1002/joc.6898

Goldreich, Y. (2003). *The Climate of Israel: Observation, Research and Application*. Springer.

Gorczynski, W. (1910). Klatka angielska nowego typu w porownaniu z ochronami termetrycznemi innych systemow. *Wiadomosci Matematyczne*, *14*, 205–236. https://crispa.uw.edu.pl/object/files/322226/display/Default

Government of India Meteorological Department. (1926). *India Weather Review for the year 1924*.

Government of India Meteorological Department. (1933). *India Weather Review 1931*. https://library.oarcloud.noaa.gov/docs.lib/htdocs/rescue/cd019_pdf/004E9BC3.pdf

Graf, H. F., Li, Q., & Giorgetta, M. A. (2007). Volcanic effects on climate: revisiting the mechanisms. In *Atmos. Chem. Phys* (Vol. 7). www.atmos-chem-phys.net/7/4503/2007/

Graf, H. F., & Timmreck, C. (2001). A general climate model simulation of the aerosol radiative effects of the Laacher See eruption (10,900 B.C.). *Journal of Geophysical Research: Atmospheres*, *106*(D14), 14747–14756. https://doi.org/10.1029/2001JD900152

Graf, H.-F., Kirchner, I., Robock, A., & Schult, I. (1993). Pinatubo eruption winter climate effects: model versus observations. In *Climate Dynamics* (Vol. 9).

Greely, A. W. (1887). *General Instructions to Observers of The Signal Service*. Government Printing Office. https://www.analogweather.com/uploads/7/7/7/5/77750690/signal_service_-_instructions_to_observers_-_1887.pdf

Grissolet, M. H. (1935). Comparaison entre les temperatures observees dans l'abri anglais O.N.M petit modele et sous l'abri Montsouris. *Annales Des Services Techniques d'hygeine de La Ville de Paris*, *XVII*, 256–267. https://gallica.bnf.fr/ark:/12148/bpt6k11777428/f276.item.r=grisollet#

Groisman, P. Y. (1992). Possible regional climate consequences of the Pinatubo eruption: An empirical approach. *Geophysical Research Letters*, *19*(15), 1603–1606. https://doi.org/10.1029/92GL01474

Groisman, P. Ya. (1985). Regional Climatic Consequences of Volcanic Eruptions [in Russian]. *Meteorologiya i Gidrologiya*, *4*, 39–45.

Gubler, S., Hunziker, S., Begert, M., Croci-Maspoli, M., Konzelmann, T., Brönnimann, S., Schwierz, C., Oria, C., & Rosas, G. (2017). The influence of station density on climate data homogenization. *International Journal of Climatology*, *37*(13), 4670–4683. https://doi.org/10.1002/joc.5114

Guijarro, J. A. (2011). Influence of network density on homogenisation performance. In M. Lakatos, T. Szentimrey, & E. Vincze (Eds.), *Seventh Seminar for Homogenization and Quality Control in Climatological Databases jointly organized with the Meeting of COST ES0601 (HOME) Action MC Meeting* (pp. 11–18). World Meteorological Organization.

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

Haimberger, L. (2007). Homogenization of Radiosonde Temperature Time Series Using Innovation Statistics. *Journal of Climate*, *20*(7), 1377–1403. https://doi.org/10.1175/JCLI4050.1

Haimberger, L., Tavolato, C., & Sperka, S. (2012). Homogenization of the Global Radiosonde Temperature Dataset through Combined Comparison with Reanalysis Background Series and Neighboring Stations. *Journal of Climate*, *25*(23), 8108–8131. https://doi.org/10.1175/JCLI-D-11-00668.1

Hamilton, K. (2012). Sereno bishop, Rollo Russell, bishop's ring and the discovery of the 'krakatoa easterlies'. *Atmosphere - Ocean*, *50*(2), 169–175. https://doi.org/10.1080/07055900.2011.639736

Hann, J., & Hellmann, G. (1894). *Meteorologische Zeitschrift. Jahrg. 11*. https://hdl.handle.net/2027/nyp.33433069059388

Hann, J., & Koppen, W. (1886). *Meteorologische Zeitschrift. Jahrg.3*. https://hdl.handle.net/2027/hvd.hxgg8e

Hann, J., & Koppen, W. (1889). *Meteorologische Zeitschrift. Jahrg. 6*. https://hdl.handle.net/2027/hvd.hxgg8b

Hann, J., & Koppen, W. (1891). *Meteorologische Zeitschrift. Jahrg.8* . https://hdl.handle.net/2027/hvd.hxgg89

Hanna, E., Cropper, T. E., Hall, R. J., Cornes, R. C., & Barriendos, M. (2022). Extended North Atlantic Oscillation and Greenland Blocking Indices 1800–2020 from New Meteorological Reanalysis. *Atmosphere*, *13*(3). https://doi.org/10.3390/atmos13030436

Hanna, E., Cropper, T. E., Jones, P. D., Scaife, A. A., & Allan, R. (2015). Recent seasonal asymmetric changes in the NAO (a marked summer decline and increased winter variability) and associated changes in the AO and Greenland Blocking Index. *International Journal of Climatology*, *35*(9), 2540–2554. https://doi.org/10.1002/joc.4157

Hannak, L., Friedrich, K., Imbery, F., & Kaspar, F. (2020). Analyzing the impact of automatization using parallel daily mean temperature series including breakpoint detection and homogenization. *International Journal of Climatology*. https://doi.org/10.1002/joc.6597

Harding, J. S. (1881). The organization of the Meteorological Service in some of the principle countries of Europe. *Symons's Monthly Meteorological Magazine*.

Harris, I., Jones, P. D., Osborn, T. J., & Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset. *International Journal of Climatology*, *34*(3), 623–642. https://doi.org/10.1002/JOC.3711

Harris, I., Osborn, T. J., Jones, P., & Lister, D. (2020). Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific Data*, *7*(1), 109. https://doi.org/10.1038/s41597-020-0453-3

Hartman, M. A. (1918). *Le climat des Pays-Bas*. https://cdn.knmi.nl/knmi/pdf/bibliotheek/knmipubmetnummer/knmipub102-24.pdf

Hausfather, Z., Menne, M. J., Williams, C. N., Masters, T., Broberg, R., & Jones, D. (2013). Quantifying the effect of urbanization on U.S. Historical Climatology Network temperature records. *Journal of Geophysical Research: Atmospheres*, *118*(2), 481–494. https://doi.org/10.1029/2012JD018509

Hawkins, E., Brohan, P., Burgess, S. N., Burt, S., Compo, G. P., Gray, S. L., Haigh, I. D., Hersbach, H., Kuijjer, K., Martínez-Alvarado, O., McColl, C., Schurer, A. P., Slivinski, L., & Williams, J. (2023). Rescuing historical weather observations improves quantification of severe windstorm risks. *Natural Hazards and Earth System Sciences*, *23*(4), 1465–1482. https://doi.org/10.5194/NHESS-23-1465-2023

Hawkins, E., Burt, S., Brohan, P., Lockwood, M., Richardson, H., Roy, M., & Thomas, S. (2019). Hourly weather observations from the Scottish Highlands (1883–1904) rescued by volunteer citizen scientists. *Geoscience Data Journal*, *6*(2), 160–173. https://doi.org/10.1002/GDJ3.79

Hawkins, E., & Sutton, R. (2009). The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society*, *90*(8), 1095–1107. https://doi.org/10.1175/2009BAMS2607.1

Hazen, H. A. (1885). *Thermometer Exposure* (XVIII; Professional Papers of the Signal Service). https://hdl.handle.net/2027/uc1.d0003183480

Hector, J. (1871). *Meteorological Report 1870: Including Returns for 1869 and Abstracts for Previous Years*.

Hegerl, G. C., Bronnimann, S., Cowan, T., Friedman, A. R., Hawkins, E., Iles, C., Muller, W., Schurer, A. P., & Undorf, S. (2019). Causes of climate change over the historical record. *Environmental Research Letters*, *14*(12).

Heidke, P. (n.d.-a). *Deutsche Uberseeische Meteorologische Beobachtungen. Heft XIV*. https://library.oarcloud.noaa.gov/docs.lib/htdocs/rescue/cd012_pdf/002BDFA8.pdf

Heidke, P. (n.d.-b). *Deutsche Uberseeische Meteorologische Beobachtungen. Heft XV und XVI*. https://library.oarcloud.noaa.gov/docs.lib/htdocs/rescue/cd013_pdf/002BDFCA.pdf

Henry, A. J. (1906). *Climatology of the United States*. Government Printing Office. https://hdl.handle.net/2027/coo1.ark:/13960/t88g95j1f

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J. N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/QJ.3803

Hiebl, J., Schöner, W., & Moberg, A. (2006). *The early instrumental climate period (1760-1860) in Europe Evidences from the Alpine region and Southern Scandinavia*.

Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., Menne, M. J., Smith, T. M., Vose, R. S., & Zhang, H. M. (2017). Extended reconstructed Sea surface temperature, Version 5 (ERSSTv5): Upgrades, validations, and intercomparisons. *Journal of Climate*, *30*(20), 8179–8205. https://doi.org/10.1175/JCLI-D-16-0836.1

Hunt, H. A. (1907). *Instructions to country observers*. https://nla.gov.au/nla.obj-37113049/view?partId=nla.obj-37114029#page/n0/mode/1up

Hunt, H. A. (1912). *Results of meteorological observations made in Western Australia during 1908*. https://hdl.handle.net/2027/uc1.$c188057

Hurrell, J. W., Kushnir, Y., Ottersen, G., & Visbeck, M. (2003). An Overview of the North Atlantic Oscillation. *Geophysical Monograph Series*, *134*, 1–35. https://doi.org/10.1029/134GM01

Időjárási jelentés Magyarországól. (1934). https://library.oarcloud.noaa.gov/docs.lib/htdocs/rescue/cd020_pdf/004D8E79.pdf

IMGW-PIB. (2023). *100-lat historii IMGW-PIB*. https://www.imgw.pl/instytut/historia

India Meteorological Department. (n.d.). *History of Meteorological Services in India*. https://mausam.imd.gov.in/imd_latest/contents/history.php

Institute of Meteorology and Water Management – National Research Institute. (n.d.). *Digitization of archived meteorological data collected in Małopolska from the 19th to the mid-20th century*. https://histklim.imgw.pl/pl/metadata/

IPCC. (2013). *Climate Change 2013: The Physical Science Basis. Contribution of the Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (T. F. Stocker, D. Qin, G. K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, & P. M. Midgley, Eds.). Cambridge University Press.

IPCC. (2022). Climate Change 2022: Impacts, Adaptation, and Vulnerability. In H.-O. Pörtner, D. C. Roberts, M. Tignor, E. S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, & B. Rama (Eds.), *Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 1–3056). Cambridge University Press.

James, H. (1861). *Instructions for taking Meteorological Observations with Tables*. https://www.google.co.uk/books/edition/Instructions_for_Taking_Meteorological_O/hHEU-PxQJNYC?hl=en&gbpv=1&printsec=frontcover

Joelsson, L. M. T., Engström, E., & Kjellström, E. (2023). Homogenization of Swedish mean monthly temperature series 1860–2021. *International Journal of Climatology*, *43*(2), 1079–1093. https://doi.org/10.1002/JOC.7881

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

Jones, P. D. (1988). Hemispheric Surface Air Temperature Variations: Recent Trends and an Update to 1987. *Journal of Climate*, *1*(6), 654–660.

Jones, P. D., Briffa, K. R., & Osborn, T. J. (2003). Changes in the Northern Hemisphere annual cycle: Implications for paleoclimatology? *Journal of Geophysical Research D: Atmospheres*, *108*(18). https://doi.org/10.1029/2003jd003695

Jones, P. D., Groisman, P. Y., Coughlan, M., Plummer, N., Wang, W. C., & Karl, T. R. (1990). Assessment of urbanization effects in time series of surface air temperature over land. *Nature*, *347*(6289), 169–172. https://doi.org/10.1038/347169a0

Jones, P. D., Jonsson, T., & Wheeler, D. (1997). Extension to the North Atlantic oscillation using early instrumental pressure observations from Gibraltar and south-west Iceland. *International Journal of Climatology*, *17*(13), 1433–1450.

https://doi.org/10.1002/(sici)1097-0088(19971115)17:13<1433::aid-joc203>3.0.co;2-p

Jones, P. D., & Lister, D. (2002). The Daily Temperature Record for St. Petersburg (1743–1996). *Climatic Change*, *53*(1).

Jones, P. D., & Moberg, A. (2003). Hemispheric and Large-Scale Surface Air Temperature Variations: An Extensive Revision and an Update to 2001. *Journal of Climate*, *16*(2), 206–223.

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

Jones, P. D., Raper, S. C. B., Bradley, R. S., Diaz, H. F., Kelly, P. M., & Wigley, T. M. L. (1986). Northern Hemisphere surface air temperature variations: 1851-1984. *Journal of Climate & Applied Meteorology*, *25*(2), 161–179. https://doi.org/10.1175/1520-0450(1986)025<0161:NHSATV>2.0.CO;2

Jungclaus, J. H., Bard, E., Baroni, M., Braconnot, P., Cao, J., Chini, L. P., Egorova, T., Evans, M., Fidel González-Rouco, J., Goosse, H., Hurtt, G. C., Joos, F., Kaplan, J. O., Khodri, M., Klein Goldewijk, K., Krivova, N., Legrande, A. N., Lorenz, S. J., Luterbacher, J., … Zorita, E. (2017). The PMIP4 contribution to CMIP6 - Part 3: The last millennium, scientific objective, and experimental design for the PMIP4 past1000 simulations. *Geoscientific Model Development*, *10*(11). https://doi.org/10.5194/gmd-10-4005-2017

Jurin, J. (1723). VII. Invitatio ad observationes meteorologicas communi consilio instituendas. *Philosophical Transactions of the Royal Society of London*, *32*(379), 422–427. https://doi.org/10.1098/RSTL.1722.0082

Kamtz, L. F. (1860). *Repertorium fur Meteorologie*.

Karami, K., Garcia, R., Jacobi, C., Richter, J. H., & Tilmes, S. (2023). The Holton-Tan mechanism under stratospheric aerosol intervention. *Atmospheric Chemistry and Physics*, *23*(6), 3799–3818. https://doi.org/10.5194/ACP-23-3799-2023

Karapiperis, P. (1954). Trends in Athens Temperature. *Tenth General Assembly of the International Union of Geodesy and Geophysics*, 212–217.

Karl, T. R., Williams, C. N., Young, P. J., & Wendland, W. M. (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 & Applied Meteorology*, *25*, 145–160.

Kaspar, F., Müller-Westermeier, G., Penda, E., Mächel, H., Zimmermann, K., Kaiser-Weiss, A., & Deutschländer, T. (2013). Monitoring of climate change in Germany-data, products and services of Germany's National Climate Data Centre. *Adv. Sci. Res*, *10*, 2012. https://doi.org/10.5194/asr-10-99-2013

Kaspar, F., Tinz, B., Mächel, H., & Gates, L. (2015). Data rescue of national and international meteorological observations at Deutscher Wetterdienst. *Advances in Science and Research*, *12*(1), 57–61. https://doi.org/10.5194/asr-12-57-2015

Kelly, P. M., Jones, P. D., & Pengqun, J. (1996). The spatial response of the climate system to explosive volcanic eruptions. *International Journal of Climatology*, *16*(5), 537–550. https://doi.org/10.1002/(SICI)1097-0088(199605)16:5<537::AID-JOC23>3.0.CO;2-F

Kennedy, J. J., Rayner, N. A., Atkinson, C. P., & Killick, R. E. (2019). *An ensemble data set of sea-surface temperature change from 1850: the Met Office Hadley Centre HadSST.4.0.0.0 Data Set*. https://doi.org/https://doi.org/10.1029/2018JD029867

Kennedy, J. J., Rayner, N. A., Smith, R. O., Parker, D. E., & Saunby, M. (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). https://doi.org/10.1029/2010jd015220

Kingston. (1857). *General instructions for making the meteorological observations at the senior county grammar schools in Upper Canada*. https://archive.org/details/cihm_35960/page/n7/mode/1up

Kingston, G. T. (1878). *Instructions for Recording Rain, Snow, Weather and Miscellaneous Phenomena, with a supplementary chapter on the temperature of the air*. https://archive.org/details/cihm_92372/page/n6/mode/1up

Kington, J. A. (1974). The Societas Meteorologica Palatina: An eighteenth-century meteorological society. *Weather*, *29*(11), 416–426. https://doi.org/10.1002/J.1477-8696.1974.TB04330.X

Kirchner, I., & Graf, H.-F. (1995). Volcanos and El Nifio: signal separation in Northern Hemisphere winter. In *Climate Dynamics* (Vol. 11).

Kirchner, I., Stenchikov, G. L., Graf, H. F., Robock, A., & Antuña, J. C. (1999). Climate model simulation of winter warming and summer cooling following the 1991 Mount Pinatubo volcanic eruption. *Journal of Geophysical Research Atmospheres*, *104*(D16), 19039–19055. https://doi.org/10.1029/1999JD900213

KNMI. (2023). *Time Series Monthly NINO3.4*. https://climexp.knmi.nl/getindices.cgi?WMO=NCDCData/ersst_nino3.4a&STATION=NINO3.4&TYPE=i&id=

Knowles Middleton, W. E. (1966). *A History of the Thermometer and Its Use in Meteorology*. The John Hopkins Press. https://archive.org/details/thermometer0000unse/page/n5/mode/2up

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., & Kiyotoshi, T. (2015). The JRA-55 Reanalysis: General Specifications and Basic Characteristics. *Journal of the Meteorological Society of Japan. Ser. II*, *93*(1), 5–48. https://doi.org/10.2151/JMSJ.2015-001

Kodera, K. (1994). Influence of volcanic eruptions on the troposphere through stratospheric dynamical processes in the Northern Hemisphere winter. *Journal of Geophysical Research*, *99*(D1), 1273–1282. https://doi.org/10.1029/93JD02731

Kolendowicz, L., Czernecki, B., Półrolniczak, M., Taszarek, M., Tomczyk, A. M., & Szyga-Pluga, K. (2019). Homogenization of air temperature and its long-term trends in Poznań (Poland) for the period 1848–2016. *Theoretical and Applied Climatology*, *136*(3–4), 1357–1370. https://doi.org/10.1007/S00704-018-2560-Z/FIGURES/9

Königlich Preussischen Meteorologischen Institut. (1888). *Instruktion für die Beobachtunger an den meteorologischen Stationen II., III. und IV. Ordnung*. https://catalog.hathitrust.org/Record/102360735

Königlich Preussischen Meteorologischen Institut. (1889). *Deutsches Meteorologisches Jahrbuch für 1887. Ergebnisse der Meteorologischen Beobachtungen im Jahre 1887*. A. Asher & Co. https://hdl.handle.net/2027/osu.32435067383018

Königlich Preussischen Meteorologischen Institut. (1891). *Deutsches Meteorologisches Jahrbuch für 1888. Ergebnisse der Meteorologischen Beobachtungen im Jahre 1888.* A. Asher & Co. https://hdl.handle.net/2027/njp.32101056917113

Königlich Preussischen Meteorologischen Institut. (1892a). *Deutsches Meteorologisches Jahrbuch für 1889. Ergebnisse der Meteorologischen Beobachtungen im Jahre 1889.* A. Asher & Co. https://hdl.handle.net/2027/osu.32435067383471

Königlich Preussischen Meteorologischen Institut. (1892b). *Deutsches Meteorologisches Jarbuch fur 1889. Ergebnisse der Meteorologischen Beobachtungen im Jahre 1889*. A. Asher & Co. https://hdl.handle.net/2027/osu.32435067383471

Königlich Preussischen Meteorologischen Institut. (1896). *Ergebnisse der Beobachtungen an den Stationen II. und III. Ordnung im Jahre 1892.* A. Asher & Co. https://hdl.handle.net/2027/osu.32435067383117

Königlich Preussischen Meteorologischen Institut. (1897). *Ergebnisse der Beobachtungen an den Stationen II. und III. Ordnung im Jahre 1893.* A. Asher & Co. https://hdl.handle.net/2027/osu.32435067383265

Königlich Preussischen Meteorologischen Institut. (1898). *Ergebnisse der Beobachtungen an den Stationen II. und III. Ordnung im Jahre 1894.* A. Asher & Co. https://hdl.handle.net/2027/osu.32435067383570

Königlich Preussischen Meteorologischen Institut. (1899). *Ergebnisse der Beobachtungen an den Stationen II. und III. Ordnung im Jahre 1895.* A. Asher & Co. https://hdl.handle.net/2027/osu.32435067383166

Konnen, G. P., Zaiki, M., Baede, A. P. M., Mikami, T., Jones, P. D., & Tsukahara, T. (2003). Pre-1872 Extension of the Japanese Instrumental Meteorological Observation Series back to 1819. *Journal of Climate*, *16*(1), 118–131. https://doi.org/https://doi.org/10.1175/1520-0442(2003)016<0118:PEOTJI>2.0.CO;2

Koppen, W. (1913). Uniform thermometer set-up for meteorological stations for the determination of air temperature and humidity. *Meteorologische Zeitschrift*, *30*, 474–523.

Kożuchowski, K., Trepińaska, J., & Wibig, J. (1994). The air temperature in cracow from 1826 to 1990: Persistence, fluctuations and the urban effect. *International Journal of Climatology*, *14*(9), 1035–1049. https://doi.org/10.1002/JOC.3370140908

Kuglitsch, F. G., Auchmann, R., Bleisch, R., Brönnimann, S., Martius, O., & Stewart, M. (2012). Break detection of annual Swiss temperature series. *Journal of Geophysical Research: Atmospheres*, *117*(D13), n/a-n/a. https://doi.org/10.1029/2012JD017729

Kuppfer, A. (1835). *Guide to making meteorological and magnetical observations*.

Labitzke, K., & McCormick, M. P. (1992). Stratospheric temperature increases due to Pinatubo aerosols. *Geophysical Research Letters*, *19*(2), 207–210. https://doi.org/10.1029/91GL02940

Labrijn, A. (1945). *The climate of the Netherlands during the last two and a half centuries*.

Labrijn, A. (1948). *The Climate of the Netherlands*. https://cdn.knmi.nl/knmi/pdf/bibliotheek/knmipubmetnummer/knmipub102-53.pdf

Ladds, C. F., & Bickers, R. A. (2008). *The Meteorological Work of the Chinese Maritime Customs Service, 1869-1947*. https://research-information.bris.ac.uk/en/publications/the-meteorological-work-of-the-chinese-maritime-customs-service-1

Laursen, E. V. (2003). *Metadata, selected climatological and synoptic stations, 1750-1996*. https://www.dmi.dk/fileadmin/user_upload/Rapporter/TR/2003/tr03-24.pdf

Le Treut, H., Somerville, R., Cubasch, U., Allen, M., Treut, L., Somerville, R., Cubasch, U., Ding, Y., Mauritzen, C., Mokssit, A., Peterson, T., Prather, M., Marquis, M., Averyt, K., & Tignor, M. (2007). Historical Overview of Climate Change Science. In S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, & H. L. Miller (Eds.), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Lee, T. C. (2016). *Metadata of Surface Meteorological Observations at the Hong Kong Observatory Headquarters 1884-2015*.

Lenssen, N. J. L., Schmidt, G. A., Hansen, J. E., Menne, M. J., Persin, A., Ruedy, R., & Zyss, D. (2019). Improvements in the GISTEMP Uncertainty Model. *Journal of*

Geophysical Research: Atmospheres, *124*(12), 6307–6326. https://doi.org/10.1029/2018JD029522

Leroy, M., & Lefebvre, G. (2000). CARACTÉRISATION DE L'ABRI DU XIXE SIÈCLE. *La Météorologie*, *8*(31). https://lameteorologie.fr/issues/2000/31/meteo_2000_31_67

Li, J., Xie, S. P., Cook, E. R., Morales, M. S., Christie, D. A., Johnson, N. C., Chen, F., D'Arrigo, R., Fowler, A. M., Gou, X., & Fang, K. (2013). El Niño modulations over the past seven centuries. *Nature Climate Change*, *3*(9), 822–826. https://doi.org/10.1038/nclimate1936

Lindau, R., & Venema, V. (2013). On the multiple breakpoint problem and the number of significant breaks in homogenization of climate records. *Quarterly Journal of the Hungarian Meteorological Service*, *117*(1), 1–34.

Ljungqvist, F. C., Seim, A., & Huhtamaa, H. (2021). Climate and society in European history. *Wiley Interdisciplinary Reviews: Climate Change*, *12*(2), e691. https://doi.org/10.1002/WCC.691

Lobell, D. B., & Bonfils, C. (2008). The Effect of Irrigation on Regional Temperatures: A Spatial and Temporal Analysis of Trends in California, 1934–2002. *Journal of Climate*, *21*(10), 2063–2071. https://doi.org/10.1175/2007JCLI1755.1

Lundstad, E., Brugnara, Y., Pappert, D., Kopp, J., Samakinwa, E., Hürzeler, A., Andersson, A., Chimani, B., Cornes, R., Demarée, G., Filipiak, J., Gates, L., Ives, G. L., Jones, J. M., Jourdain, S., Kiss, A., Nicholson, S. E., Przybylak, R., Jones, P., … Brönnimann, S. (2023). The global historical climate database HCLIM. *Scientific Data 2023 10:1*, *10*(1), 1–16. https://doi.org/10.1038/s41597-022-01919-w

Luterbacher, J., & Pfister, C. (2015). The year without a summer. *Nature Geoscience 2015 8:4*, *8*(4), 246–248. https://doi.org/10.1038/ngeo2404

Mahlmann, W. (1847). *Official instructions for the Observers at the Meteorological Stations in the Kingdom of Prussia*. Royal Statistical Bureau.

Manley, G. (1941). The durham meteorological record, 1847-1940. *Quarterly Journal of the Royal Meteorological Society*, *67*(292), 363–380. https://doi.org/10.1002/qj.49706729209

Manley, G. (1974). Central England temperatures: Monthly means 1659 to 1973. *Quarterly Journal of the Royal Meteorological Society*, *100*(425), 389–405. https://doi.org/10.1002/qj.49710042511

Mann, G., Dhomse, S., Deshler, T., Timmreck, C., Schmidt, A., Neely, R., & Thomason, L. (2015). Evolving particle size is the key to improved volcanic forcings. *PAGES Magazine*, *23*(2).

Mao, J., & Robock, A. (1998). Surface Air Temperature Simulations by AMIP General Circulation Models: Volcanic and ENSO Signals and Systematic Errors. *Journal of Climate*, *11*, 1538–1552.

Margary, I. D. (1924). Glaisher stand versus stevenson screen. A comparison of forty years' observations of maximum and minimum temperature as recorded in both screens at Camden Square, London. *Quarterly Journal of the Royal Meteorological Society*, *50*(211), 209–226. https://doi.org/10.1002/qj.49705021109

Marriott, W. (1879). Thermometer exposure - Wall versus Stevenson Screens. *Quarterly Journal of the Royal Meteorological Society*, *5*(32), 217–223. https://doi.org/10.1002/qj.4970053204

Marriott, W. (1880). Results of meteorological observations made at Stanley, Falkland Islands, 1875-77. By William Marriott, F.M.S., Assistant Secretary. *Quarterly Journal of the Royal Meteorological Society*, *6*(36), 199–202. https://doi.org/10.1002/qj.4970063606

Marriott, W. (1894). Comparative observations with two sets of instruments at llfracombe, north devon, during 1893. *Quarterly Journal of the Royal Meteorological Society*, *20*(90), 164–168. https://doi.org/10.1002/qj.4970209005

Marshall, A. G., Scaife, A. A., & Ineson, S. (2009). Enhanced seasonal prediction of European winter warming following volcanic eruptions. *Journal of Climate*, *22*(23), 6168–6180. https://doi.org/10.1175/2009JCLI3145.1

Marshall, L., Johnson, J. S., Mann, G. W., Lee, L., Dhomse, S. S., Regayre, L., Yoshioka, M., Carslaw, K. S., & Schmidt, A. (2019). Exploring How Eruption Source Parameters Affect Volcanic Radiative Forcing Using Statistical Emulation. *Journal of Geophysical Research: Atmospheres*, *124*(2), 964–985. https://doi.org/10.1029/2018JD028675

Marshall, L. R., Smith, C. J., Forster, P. M., Aubry, T. J., Andrews, T., & Schmidt, A. (2020). Large Variations in Volcanic Aerosol Forcing Efficiency Due to Eruption Source Parameters and Rapid Adjustments. *Geophysical Research Letters*, *47*(19). https://doi.org/10.1029/2020GL090241

Marshall, L., Schmidt, A., Toohey, M., Carslaw, K. S., Mann, G. W., Sigl, M., Khodri, M., Timmreck, C., Zanchettin, D., Ball, W. T., Bekki, S., Brooke, J., Dhomse, S., Johnson, C., Lamarque, J. F., Legrande, A. N., Mills, M. J., Niemeier, U., Pope, J. O., … Tummon, F. (2018). Multi-model comparison of the volcanic sulfate deposition from the 1815 eruption of Mt. Tambora. *Atmospheric Chemistry and Physics*, *18*(3), 2307–2328. https://doi.org/10.5194/ACP-18-2307-2018

Martinez Ibarra, E., Azorin-Molina, C., Banon, M., Olcina Cantos, J., Estrela, M. J., & Gil Olcina, A. (2010, November). Intercomparacion de las temperaturas extremas en tres tipos de garita meteorologica: Montsouris, Stevenson y Young. *VII Congress of the Spanish Association of Climatology: Climate, City and Ecosystems*. https://repositorio.aemet.es/handle/20.500.11765/8516

Marvin, C. F. (1915). *Instructions for Cooperative Observers*. https://hdl.handle.net/2027/mdp.39015035467466

Mass, C., & Robock, A. (1982). The Short-Term Influence of the Mount St. Helens Volcanic Eruption on Surface Temperature in the Northwest United States. *Monthly Weather Review*, *110*(6), 614–622. https://doi.org/10.1175/1520-0493(1982)110

Mateus, C., Potito, A., & Curley, M. (2020). Reconstruction of a long-term historical daily maximum and minimum air temperature network dataset for Ireland (1831-1968). *Geoscience Data Journal*, *7*(2), 102–115. https://doi.org/10.1002/GDJ3.92

Maugeri, M., Buffoni, L., & Chlistovsky, F. (2002). *DAILY MILAN TEMPERATURE AND PRESSURE SERIES (1763-1998): HISTORY OF THE OBSERVATIONS AND DATA AND METADATA RECOVERY*.

Mawley, E. (1897). Shade temperature. *Quarterly Journal of the Royal Meteorological Society*, *23*(102), 69–87. https://doi.org/10.1002/qj.49702310202

Meier, N., Rutishauser, T., Pfister, C., Wanner, H., & Luterbacher, J. (2007). Grape harvest dates as a proxy for Swiss April to August temperature reconstructions back to AD

1480. *Geophysical Research Letters*, *34*(20), L20705. https://doi.org/10.1029/2007GL031381

Menne, M. J., & Williams, C. N. (2009). Homogenization of temperature series via pairwise comparisons. *Journal of Climate*, *22*(7), 1700–1717. https://doi.org/10.1175/2008JCLI2263.1

Menne, M. J., Williams, C. N., Gleason, B. E., Jared Rennie, J., & Lawrimore, J. H. (2018). The Global Historical Climatology Network Monthly Temperature Dataset, Version 4. *Journal of Climate, 31*(24), 9835–9854. https://doi.org/10.1175/JCLI-D-18-0094.1

Menne, M. J., Williams, C. N., & Vose, R. S. (2009). The U.S. historical climatology network monthly temperature data, version 2. *Bulletin of the American Meteorological Society*, *90*(7), 993–1007. https://doi.org/10.1175/2008BAMS2613.1

Met Eireann. (n.d.). *Long Term Data Sets*. Retrieved 5 October 2023, from https://www.met.ie/climate/available-data/long-term-data-sets

Meteorological Committee. (1874, September). *Report of the proceedings of the International Meteorological congress at Vienna, 1873*.

Meteorological Committee. (1876). *Report of the Permanent Committee of the First International Congress at Vienna, meeting at London, 1876*. https://hdl.handle.net/2027/hvd.hxgnti

Meteorological Council. (1890). *Meteorological Observations in the Foreign and Colonial Stations of the Royal Engineers and the Army Medical Department, 1852-1886*. https://www.google.co.uk/books/edition/Meteorological_Observations_at_the_Forei/NFwUAAAAYAAJ?hl=en&gbpv=1&dq=Instructions+for+taking+Meteorological+Observations+at+the+principal+Stations+of+the+Royal+Engineers&pg=PR1&printsec=frontcover

Meteorological Observations made at the Adelaide Observatory, and other places in South Australia and the Northern Territory, during the year 1888. (1890). https://hdl.handle.net/2027/uc1.c2999509

Meteorological observations made at the Adelaide Observatory, and other places in South Australia and the Northern Territory during the year 1889. (1891). https://hdl.handle.net/2027/uc1.c2999510

Meteorological Observations made at the Adelaide Observatory, and other places in South Australia and the Northern Territory, during the year 1890. (1892). https://hdl.handle.net/2027/uc1.c2999511

Meteorological Observations made at the Adelaide Observatory, and other places in South Australia and the Northern Territory, during the year 1891. (1894). https://hdl.handle.net/2027/uc1.c2999512

Meteorological Observations made at the Adelaide Observatory, and other places in South Australia and the Northern Territory, during the year 1892. (1894). https://hdl.handle.net/2027/uc1.c2999513

Meteorological Observations made at the Adelaide Observatory, and other places in South Australia and the Northern Territory, during the year 1893. (1896). https://hdl.handle.net/2027/uc1.c2999514

Meteorological Observations made at the Adelaide Observatory, and other places in South Australia and the Northern Territory, during the year 1894. (1897). https://hdl.handle.net/2027/uc1.c2999515

Meteorological Observations made at the Adelaide Observatory, and other places in South Australia and the Northern Territory, during the year 1895. (1898). https://hdl.handle.net/2027/uc1.c2999516

Meteorological Observations made at the Adelaide Observatory, and other places in South Australia and the Northern Territory, during the year 1896. (1899). https://hdl.handle.net/2027/uc1.c2999376

Meteorological Observations made at the Adelaide Observatory, and other places in South Australia and the Northern Territory, during the year 1897. (1900). https://hdl.handle.net/2027/uc1.c2999377

Meteorological Observations made at the Adelaide Observatory, and other places in South Australia and the Northern Territory, during the year 1898. (1901). https://hdl.handle.net/2027/uc1.c2999378

Meteorological Observations made at the Adelaide Observatory, and other places in South Australia and the Northern Territory, during the year 1899. (1902). https://hdl.handle.net/2027/uc1.c2999379

Meteorological observations made at the Adelaide Observatory, and other places in South Australia and the Northern Territory during the years 1886-7. (1893). https://hdl.handle.net/2027/uc1.c2999508

Meteorological Observatory, D. M. (1907). *Annual report of the Meteorological Observations made at the Meteorological Stations of the Government-General of Kuantung*. https://library.oarcloud.noaa.gov/docs.lib/htdocs/rescue/cd120_pdf/LSN0471.PDF

Meteorological Office. (1911). *The Observer's Handbook*. His Majesty's Stationery Office. https://hdl.handle.net/2027/uiug.30112069855077

Meteorological Office. (1923). *Notes on the Meteorological Observations made in British Colonies and Protectorates in 1923 and Summarized in the Annual Reports of Colonial Governments*. https://digital.nmla.metoffice.gov.uk/IO_d3c33368-b1fa-41cf-8a3f-8a3750cf3806/

Meteorological Office. (1926). *Notes on the Meteorological Observations made in British Colonies and Protectorates in 1926 and Summarised in the Annual Reports of Colonial Governments*. https://digital.nmla.metoffice.gov.uk/IO_a41b900f-7ed9-4c08-bfa8-842037d2cc25/

Meteorological Office. (1927). *Notes on the Meteorological Observations made in British Colonies and Protectorates in 1927 and Summarized in the Annual Reports of Colonial Governments*. https://digital.nmla.metoffice.gov.uk/IO_129a8d12-e369-4514-ba78-71f7215c229f/

Meteorological Office. (1928). *Notes on the Meteorological Observations made in British Colonies and Protectorates in 1928 and Summarized in the Annual Reports of Colonial Governments*. https://digital.nmla.metoffice.gov.uk/IO_e863c08d-d789-4fd1-a2a5-6cb193aa401f/

Meteorological Office. (1929). *Notes on the Meteorological Observations made in British Colonies and Protectorates, ETC. in 1929 and Summarized in the Annual Reports of Colonial Governments*. https://digital.nmla.metoffice.gov.uk/IO_80a6a3ce-5ab2-4213-90f4-9fcee928630b/

Meteorological Office. (1930). *Notes on the Meteorological Observations made in British Colonies and Protectorates, ETC. in 1930 and Summarized in the Annual Reports of*

Colonial Governments. https://digital.nmla.metoffice.gov.uk/IO_85e898f3-1850-4598-a751-650d484389b8/

Meteorological Office. (1931). *Notes on the Meteorological Observations made in British Colonies and Protectorates, ETC. in 1931 and Summarized in the Annual Reports of Colonial Governments*. https://digital.nmla.metoffice.gov.uk/IO_3c7bf413-2205-4311-93d1-20ce58a53827/

Meteorological Office. (1932). *Notes on the Meteorological Observations made in British Colonies and Protectorates, ETC. in 1932 and Summarized in the Annual Reports of Colonial Governments*. https://digital.nmla.metoffice.gov.uk/IO_77c7b656-7304-4ef1-b5d0-7f57c035bf4f/

Meteorological Office. (1933). *Notes on the Meteorological Observations made in British Colonies and Protectorates, ETC. in 1933 and Summarized in the Annual Reports of Colonial Governments*. https://digital.nmla.metoffice.gov.uk/IO_efa8a89a-5be5-4999-a02d-06045b9c6a1a/

Meteorological Service. (1950). *Climatological means for Iraq*. https://library.oarcloud.noaa.gov/docs.lib/htdocs/rescue/cd024_pdf/005D830D.pdf

Mietus, M. (1999). Historia obserwacji meteorologicznych w okresie przed rokiem 1945 w Gorzowie Wielkopolskim. *Wiadomości IMGW*, *22*, 153–158. https://www.researchgate.net/publication/282337461_Historia_obserwacji_meteorologicznych_w_okresie_przed_rokiem_1945_w_Gorzowie_Wielkopolskim/citation/download

Mietus, M. (2001). Jednorodność wieloletnich serii pomiarowych. Rzeczywistość czy fikcja? *Annales Universitatis Mariae Curie-Skłodowska Lublin - Polonia*, *55/56*(29), 239–249.

Mietus, M., Wielbinska, D., & Owczarek, M. (1994). Historia obserwacji meteorologicznych na niektórych stacjach polskiego wybrzeża. *Wiadomosci Instytutu Meteorologii I Gospodarki Wodnej*, *17*(4), 149–162. https://www.researchgate.net/publication/282333943_Historia_obserwacji_meteorologicznych_na_niektorych_stacjach_polskiego_wybrzeza

Ministry of Public Works Egypt. (1921). *Meteorological Report for the Year 1915*. https://library.oarcloud.noaa.gov/docs.lib/htdocs/rescue/cd005_pdf/002254CE.pdf

Moberg, A., Alexandersson, H., Bergström, H., & Jones, P. D. (2003). Were southern Swedish summer temperatures before 1860 as warm as measured? *International Journal of Climatology*, *23*(12), 1495–1521. https://doi.org/10.1002/joc.945

Moberg, A., & Bergström, H. (1997). Homogenization of Swedish temperature data. Part III: the long temperature records from Uppsala and Stockholm. *International Journal of Climatology*, *17*(7), 667–699. https://doi.org/10.1002/(SICI)1097-0088(19970615)17:7<667::AID-JOC115>3.0.CO;2-J

Moberg, A., Bergström, H., Ruiz Krigsman, J., & Svanered, O. (2002). Daily air temperature and pressure series for Stockholm (1756-1998). *Climatic Change*, *53*(1–3), 171–212. https://doi.org/10.1023/A:1014966724670

Moden, H. (1954). Termometeruppstallningar - en kort jamforelse mellan fristaende bur och huv. *Notiser Och Preliminara Rapporter SMHI*, *1*(3), 12.

Morice, C. P., Kennedy, J. J., Rayner, N. A., & Jones, P. D. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *Journal of Geophysical Research: Atmospheres*, *117*(D8). https://doi.org/10.1029/2011JD017187

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., Dunn, R. J. H., Osborn, T. J., Jones, P. D., & Simpson, I. R. (2021). An updated assessment of near-surface temperature change from 1850: the HadCRUT5 dataset. *Journal of Geophysical Research: Atmospheres*, *126*(3). https://doi.org/10.1029/2019JD032361

Morley, T. (1991). *The beginnings of Canadian meteorology*. ECW Press. https://archive.org/details/beginningsofcana0000thom

Morozova, A. L., & Valente, M. A. (2012). Homogenization of Portuguese long-term temperature data series: Lisbon, Coimbra and Porto. *Earth System Science Data*, *4*(1), 187–213. https://doi.org/10.5194/ESSD-4-187-2012

Možný, M., Brázdil, R., Dobrovolný, P., & Trnka, M. (2016). April–August temperatures in the Czech Lands, 1499–2015, reconstructed from grape-harvest dates. *Clim. Past*, *12*, 1499–2015. https://doi.org/10.5194/cp-12-1421-2016

Mullan. (2012). Applying the Rhoades and Salinger Method to New Zealand's "Seven-Station" Temperature Series. *Weather and Climate*, *32*(1), 23. https://doi.org/10.2307/26169723

Muller, G. (1984). Vergleich der Temperaturen verschiedener Wetterhutten an einigen Stationen des Anetzes. *Arbeitsberichte Der Schweizerischen Meteorologischen Zentralanstalt, No. 119*, 1–36.

Muthers, S., Anet, J. G., Raible, C. C., Brönnimann, S., Rozanov, E., Arfeuille, F., Peter, T., Shapiro, A. I., Beer, J., Steinhilber, F., Brugnara, Y., & Schmutz, W. (2014). Northern hemispheric winter warming pattern after tropical volcanic eruptions: Sensitivity to the ozone climatology. *Journal of Geophysical Research*, *119*(3), 1340–1355. https://doi.org/10.1002/2013JD020138

Muthers, S., Arfeuille, F., Raible, C. C., & Rozanov, E. (2015). The impacts of volcanic aerosol on stratospheric ozone and the Northern Hemisphere polar vortex: Separating radiative-dynamical changes from direct effects due to enhanced aerosol heterogeneous chemistry. *Atmospheric Chemistry and Physics*, *15*(20), 11461–11476. https://doi.org/10.5194/ACP-15-11461-2015

National Geophysical Data Center / World Data Service (NGDC/WDS). (n.d.). *NCEI/WDS Global Significant Volcanic Eruptions Database*. NOAA National Centers for Environmental Information. https://doi.org/doi:10.7289/V5JW8BSH

National Oceanic and Atmospheric Administration. (n.d.). *Paleo Data Search*. Retrieved 22 October 2023, from https://www.ncei.noaa.gov/access/paleo-search/

Naujokat, B. (1986). An Update of the Observed Quasi-Biennial Oscillation of the Stratospheric Winds over the Tropics. *Journal of the Atmospheric Sciences*, *43*(17), 1873–1877.

Naylor, S. (2019). *THERMOMETER SCREENS AND THE GEOGRAPHIES OF UNIFORMITY IN NINETEENTH-CENTURY METEOROLOGY*. https://doi.org/10.1098/rsnr.2018.0037

Neumayer. (n.d.-a). *Deutsche Uberseeische Meteorologische Beobachtungen. Heft IX - X*. https://library.oarcloud.noaa.gov/docs.lib/htdocs/rescue/cd012_pdf/002BDF90.pdf

Neumayer. (n.d.-b). *Deutsche Uberseeische Meteorologische Beobachtungen. Heft VI - VIII*. https://library.oarcloud.noaa.gov/docs.lib/htdocs/rescue/cd012_pdf/002BDF77.pdf

Newhall, C. G., & Self, S. (1986). *The Volcanic Explosivity Index (VEI): An estimate of explosive magnitude for historical volcanism* (pp. 143–150). American Geophysical Union (AGU). https://doi.org/10.1029/HG002p0143

Nicholls, N., Tapp, R., Burrows, K., & Richards, D. (1996). Historical thermometer exposures in Australia. *International Journal of Climatology*, *16*(6), 705–710. https://doi.org/10.1002/(SICI)1097-0088(199606)16:6<705::AID-JOC30>3.0.CO;2-S

Niemeier, U., Timmreck, C., Graf, H. F., Kinne, S., Rast, S., & Self, S. (2009). Initial fate of fine ash and sulfur from large volcanic eruptions. *Atmospheric Chemistry and Physics*, *9*(22), 9043–9057. https://doi.org/10.5194/ACP-9-9043-2009

NOAA Climate. (n.d.). *Climate Variability: North Atlantic Oscillation*. Retrieved 14 November 2023, from https://www.climate.gov/news-features/understanding-climate/climate-variability-north-atlantic-oscillation

Nordli, Ø., Lundstad, E., & Ogilvie, A. E. J. (2007). A late-winter to early-spring temperature reconstruction for southeastern Norway from 1758 to 2006. *Annals of Glaciology*, *46*, 404–408. https://doi.org/10.3189/172756407782871657

Nordli, P. Ø., Alexandersson, H., Frich, P., Forland, E., Heino, R., Jonsson, T., Steffensen, P., Tuomenvirta, H., & Tveito, O. E. (1996). *The effect of radiation screens on Nordic temperature measurements*. https://www.met.no/publikasjoner/met-report/met-report-1996

Nordli, P. Ø., Alexandersson, H., Frich, P., Førland, E. J., Heino, R., Jónsson, T., Tuomenvirta, H., & Tveito, O. E. (1997). The effect of radiation screens on Nordic time series of mean temperature. *International Journal of Climatology*, *17*(15), 1667–1681. https://doi.org/10.1002/(SICI)1097-0088(199712)17:15<1667::AID-JOC221>3.0.CO;2-D

Oficina Central Meteorologica. (1870). *Anuario de la Oficina Central Meteorologica de Santiago de Chile*. https://hdl.handle.net/2027/hvd.hxg9ki

Ogrin, D. (2015). Long-term air temperature changes in Ljubljana (Slovenia) in comparison to Trieste (Italy) and Zagreb (Croatia). *Moravian Geographical Reports*, *23*(3), 17–26. https://www.geonika.cz/EN/research/ENMGRClanky/2015_3_OGRIN.pdf

Oman, L., Robock, A., Stenchikov, G., Schmidt, G. A., & Ruedy, R. (2005). Climatic response to high-latitude volcanic eruptions. *Journal of Geophysical Research Atmospheres*, *110*(13). https://doi.org/10.1029/2004JD005487

Omond, R. T. (1906). Temperatures in thermograph and Stevenson screens. *Journal of the Scottish Meteorological Society*, *XIV*(XXIII), 15–20. https://archive.org/details/journalscottish00socigoog/page/15/mode/1up

O'Neill, P., Connolly, R., Connolly, M., Soon, W., Chimani, B., Crok, M., de Vos, R., Harde, H., Kajaba, P., Nojarov, P., Przybylak, R., Rasol, D., Skrynyk, O., Skrynyk, O., Štěpánek, P., Wypych, A., & Zahradníček, P. (2022). Evaluation of the Homogenization Adjustments Applied to European Temperature Records in the Global Historical Climatology Network Dataset. *Atmosphere*, *13*(2), 285. https://doi.org/10.3390/ATMOS13020285/S1

Ortega, P., Blockley, E. W., Køltzow, M., Massonnet, F., Sandu, I., Svensson, G., Acosta Navarro, J. C., Arduini, G., Batté, L., Bazile, E., Chevallier, M., Cruz-García, R., Day, J. J., Fichefet, T., Flocco, D., Gupta, M., Hartung, K., Hawkins, E., Hinrichs, C., … Jung, T. (2022). Improving Arctic Weather and Seasonal Climate Prediction: Recommendations for Future Forecast Systems Evolution from the European Project APPLICATE. *Bulletin of the American Meteorological Society*, *103*(10), E2203–E2213. https://doi.org/10.1175/BAMS-D-22-0083.1

Ortega, P., Lehner, F., Swingedouw, D., Masson-Delmotte, V., Raible, C. C., Casado, M., & Yiou, P. (2015). A model-tested North Atlantic Oscillation reconstruction for the past millennium. *Nature*, *523*(7558), 71–74. https://doi.org/10.1038/nature14518

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

PAGES2k Consortium. (2017). A global multiproxy database for temperature reconstructions of the Common Era. *Scientific Data*, *4*(170088). https://doi.org/10.1038/sdata.2017.88

Paolantonio, S. (2014). *Pablo E. Coni editor of the Results of the Argentine National Observatory*. https://historiadelaastronomia.wordpress.com/documentos/coni/#ref2

Pappert, D., Brugnara, Y., Jourdain, S., Pospieszy´nskapospieszy´nska, A., Przybylak, R., Rohr, C., & Brönnimann, S. (2021). Unlocking weather observations from the Societas Meteorologica Palatina (1781-1792). *Climate of the Past*, *17*, 2361–2379.

Parker, D. E. (1994). Effects of changing exposure of thermometers at land stations. *International Journal of Climatology*, *14*(1), 1–31. https://doi.org/10.1002/joc.3370140102

Parker, D. E., & Brownscombe, J. L. (1983). Stratospheric warming following the El Chichon volcanic eruptions. *Nature*, *301*, 406–408.

Parker, D. E., Legg, T. P., & Folland, C. K. (1992). A new daily central England temperature series, 1772–1991. *International Journal of Climatology*, *12*(4), 317–342. https://doi.org/10.1002/joc.3370120402

Parker, D. E., Wilson, H., Jones, P. D., Christy, J. R., & Folland, C. K. (1996). The impact of mount pinatubo on world-wide temperatures. *International Journal of Climatology*, *16*(5), 487–497. https://doi.org/10.1002/(SICI)1097-0088(199605)16:5<487::AID-JOC39>3.0.CO;2-J

Pelz, J. (2007). *Einhundert Jahre Wetteraufzeichnungen in (Berlin)-Dahlem*. https://berliner-wetterkarte.de/Beilagen/2007/Pelz_29_07.pdf

Perlwitz, J., & Graf, H. F. (1995). The statistical connection between Tropospheric and Stratospheric Circulation of the Northern Hemisphere in Winter. *Journal of Climate*, *8*(10), 2281–2295.

Peterson, T. C., & Easterling, D. R. (1994). Creation of homogeneous composite climatological reference series. *International Journal of Climatology*, *14*(6), 671–679. https://doi.org/10.1002/JOC.3370140606

Peterson, T. C., Easterling, D. R., Karl, T. R., Groisman, P., Nicholls, N., Plummer, N., Torok, S., Auer, I., Boehm, R., Gullet, D., Vincent, L., Heino, R., Tuomenvirta, H.,

Mestre, O., Szentimrey, T., Salinger, J., Førland, E. J., Hanssen-Bauer, I., Alexandersson, H., … Parker, D. (1998). Homogeneity adjustments of in situ atmospheric climate data: A review. *International Journal of Climatology*, *18*(13), 1493–1517. https://doi.org/10.1002/(SICI)1097-0088(19981115)18:13<1493::AID-JOC329>3.0.CO;2-T

Pfeiffer, M., Zinke, J., Dullo, W. C., Garbe-Schönberg, D., Latif, M., & Weber, M. E. (2017). Indian Ocean corals reveal crucial role of World War II bias for twentieth century warming estimates. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-14352-6

Pfister, C., Luterbacher, J., Wanner, H., Wheeler, D., Brázdil, R., Ge, Q., Hao, Z., Moberg, A., Grab, S., & Rosario Del Prieto, M. (2009). *Documentary evidence as climate proxies* (Proxy- Specific White Paper Produced from the PAGES/CLIVAR Workshop). http://www.pages-igbp.org/cgi-bin/WebObjects/products.woa/wa/product?id=331

Piotrowicz, K. (2007). Temperatura Powietrza. In P. Redakcja & D. Matuszko (Eds.), *Klimat Krakowa W XX Wieku* (pp. 99–112). Instytut Geografii i Gospodarki Przestrzennej Uniwersytetu Jagiellońskiego. https://www.researchgate.net/profile/Katarzyna-Piotrowicz-2/publication/329983677_Temperatura_powietrza_Air_temperature/links/5c274976a6fdccfc706f910d/Temperatura-powietrza-Air-temperature.pdf

Plumb, A. R. (2002). Stratospheric Transport. *Journal of the Meteorological Society of Japan*, *80*(4B), 793–809.

Plummer, J. J., & Scott, R. H. (1873). On some results of temperature observations at Durham. *Quarterly Journal of the Royal Meteorological Society*, *1*(8), 241–246. https://doi.org/10.1002/qj.4970010802

Polvani, L. M., Banerjee, A., & Schmidt, A. (2019). Northern Hemisphere continental winter warming following the 1991 Mt. Pinatubo eruption: Reconciling models and observations. *Atmospheric Chemistry and Physics*, *19*(9), 6351–6366. https://doi.org/10.5194/acp-19-6351-2019

Polvani, L. M., & Camargo, S. J. (2020). Scant evidence for a volcanically forced winter warming over Eurasia following the Krakatau eruption of August 1883. *Atmospheric*

Chemistry and Physics, *20*(22), 13687–13700. https://doi.org/10.5194/ACP-20-13687-2020

Pospieszyńska, A., & Przybylak, R. (2019). Air temperature changes in Toruń (central Poland) from 1871 to 2010. *Theoretical and Applied Climatology*, *135*(1–2), 707–724. https://doi.org/10.1007/s00704-018-2413-9

Przybylak, R., Majorowicz, J., Brázdil, R., & Kejna, M. (2010). The polish climate in the European context: An historical overview. *The Polish Climate in the European Context: An Historical Overview*, 1–535. https://doi.org/10.1007/978-90-481-3167-9/COVER

Quayle, R. G., Easterling, D. R., Karl, T. R., & Hughes, P. Y. (1991). Effects of Recent Thermometer Changes in the Cooperative Station Network. *Bulletin of the American Meteorological Society*, *72*(11), 1718–1723. https://doi.org/10.1175/1520-0477(1991)072<1718:EORTCI>2.0.CO;2

Raible, C. C., Brönnimann, S., Auchmann, R., Brohan, P., Frölicher, T. L., Graf, H. F., Jones, P., Luterbacher, J., Muthers, S., Neukom, R., Robock, A., Self, S., Sudrajat, A., Timmreck, C., & Wegmann, M. (2016). Tambora 1815 as a test case for high impact volcanic eruptions: Earth system effects. In *Wiley Interdisciplinary Reviews: Climate Change* (Vol. 7, Issue 4, pp. 569–589). Wiley-Blackwell. https://doi.org/10.1002/wcc.407

Rayner, N. A., Brohan, P., Parker, D. E., Folland, C. K., Kennedy, J. J., Vanicek, M., Ansell, T. J., & Tett, S. F. B. (2006). Improved analyses of changes and uncertainties in sea surface temperature measured in Situ since the mid-nineteenth century: The HadSST2 dataset. *Journal of Climate*, *19*(3), 446–469. https://doi.org/10.1175/JCLI3637.1

Reeves, J., Chen, J., Wang, X. L., Lund, R., & Lu, Q. Q. (2007). A Review and Comparison of Changepoint Detection Techniques for Climate Data. *Journal of Applied Meteorology and Climatology*, *46*(6), 900–915. https://doi.org/10.1175/JAM2493.1

Reichen, L., Burgdorf, A. M., Brönnimann, S., Franke, J., Hand, R., Valler, V., Samakinwa, E., Brugnara, Y., & Rutishauser, T. (2022). A decade of cold Eurasian winters reconstructed for the early 19th century. *Nature Communications 2022 13:1*, *13*(1), 1–9. https://doi.org/10.1038/s41467-022-29677-8

Reis, C. M. J. A. , D. G. H. L. , P. L. P. C. M. P. , L. T. (2005). *Contribution à la climatologie du Luxembourg. Analyses historiques, scénarios futurs.* (C. Ries, Ed.). Musée national d'histoire naturelle à Luxembourg.

Rennie, J. J., Lawrimore, J. H., Gleason, B. E., Thorne, P. W., Morice, C. P., Menne, M. J., Williams, C. N., de Almeida, W. G., Christy, J. R., Flannery, M., Ishihara, M., Kamiguchi, K., Klein-Tank, A. M. G., Mhanda, A., Lister, D. H., Razuvaev, V., Renom, M., Rusticucci, M., Tandy, J., … Yin, X. (2014). The international surface temperature initiative global land surface databank: monthly temperature data release description and methods. *Geoscience Data Journal*, *1*(2), 75–102. https://doi.org/10.1002/gdj3.8

Renou, M. E. (1875). Note on the determination of the temperature of the air. *Symons's Monthly Meteorological Magazine, March*.

Report on the Administration of the Madras Presidency during the year 1867-68. (1868). https://www.google.co.uk/books/edition/Report_on_the_Administration_of_the_Madr/TRUTAAAAYAAJ?hl=en&gbpv=1

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1887. (1889). http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/grw.html

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1888. (1890). http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/grw.html

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1889. (1891). http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/YBpdf/YB_GRW_1889.pdf

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1890. (1892). http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/YBpdf/YB_GRW_1890.pdf

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1891. (1893).

http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/YBpdf/YB_GRW_1891.pdf

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1892. (1894). http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/YBpdf/YB_GRW_1892.pdf

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1893. (1896). http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/YBpdf/YB_GRW_1893.pdf

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1894. (1897). http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/YBpdf/YB_GRW_1894.pdf

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1895. (1897). http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/YBpdf/YB_GRW_1895.pdf

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1896. (1898). http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/YBpdf/YB_GRW_1896.pdf

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1897. (1899). http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/YBpdf/YB_GRW_1897.pdf

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1898. (1900). http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/YBpdf/YB_GRW_1898.pdf

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1899. (1901).

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1900. (1902). http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/YBpdf/YB_GRW_1900.pdf

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1901. (1903). http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/grw.html

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1902. (1904). http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/grw.html

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1903. (1904). http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/grw.html

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1904. (1905). http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/grw.html

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1905. (1906). http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/grw.html

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1906. (1907). http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/grw.html

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1907. (1909). http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/grw.html

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1908. (1909). http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/grw.html

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1909. (1911). http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/grw.html

Ribeiro, S., Caineta, J., & Costa, A. C. (2016). Review and discussion of homogenisation methods for climate data. In *Physics and Chemistry of the Earth* (Vol. 94, pp. 167–179). Elsevier Ltd. https://doi.org/10.1016/j.pce.2015.08.007

Robock, A. (2000). *Volcanic Eruptions and Climate*.

Robock, A., & Mao, J. (1992). Winter warming from large volcanic eruptions. *Geophysical Research Letters*, *19*(24), 2405–2408. https://doi.org/10.1029/92GL02627

Robock, A., & Mao, J. (1995). The Volcanic Signal in Surface Temperature Observations. In *Source: Journal of Climate* (Vol. 8, Issue 5). https://about.jstor.org/terms

Rogelj, J., Forster, P. M., Kriegler, E., Smith, C. J., & Séférian, R. (2019). Estimating and tracking the remaining carbon budget for stringent climate targets. *Nature 2019 571:7765*, *571*(7765), 335–342. https://doi.org/10.1038/s41586-019-1368-z

Rohde, R., Muller, R., Jacobsen, R., Muller, E., Perimutter, S., Rosenfield, A., Wurtele, J., Groom, D., & Wickham, C. (2013). A New Estimate of the Average Earth Surface Land Temperature Spanning 1753 to 2011. *Geoinformatics & Geostatistics: An Overview*, *01*(01), 1–7. https://doi.org/10.4172/2327-4581.1000101

Rohde, R., Muller, R., Jacobsen, R., Perlmutter, S., Rosenfeld, A., Wurtele, J., Curry, J., Wickham, C., & Mosher, S. (2016). Berkeley Earth Temperature Averaging Process. *Geoinformatics & Geostatistics: An Overview*, *2013*(02). https://doi.org/10.4172/2327-4581.1000103

Ross, C. (2023, October 11). Sitting on a goldmine: how the National Meteorological Archive is working to improve access to its key metadata and data resources. *Data Rescue - Discovery and Recovery of Historic Climatic Observations*.

Rotch, A. L. (1894). The Meteorological Services of South America. *The American Meteorological Journal*. https://www.google.co.uk/books/edition/The_American_Meteorological_Journal/08IEAAAAYAAJ?hl=en&gbpv=1&dq=THE+METEOROLOGICAL+SERVICES+OF+SOUTH+AMERICA&pg=PA187&printsec=frontcover

Rousseau, D. (2015). Variabilité des températures mensuelles à Paris de 1658 à 2014. *XXVIIIe Colloque de l'Association Internationale de Climatologie*, 597–602.

Rozanov, E. V., Schlesinger, M. E., Andronova, N. G., Yang, F., Malyshev, S. L., Zubov, V. A., Egorova, T. A., & Li, B. (2002a). Climate/chemistry effects of the Pinatubo volcanic eruption simulated by the UIUC stratosphere/troposphere GCM with interactive photochemistry. *Journal of Geophysical Research Atmospheres*, *107*(21), ACL 12-1-ACL 12-14. https://doi.org/10.1029/2001JD000974

Rozanov, E. V., Schlesinger, M. E., Andronova, N. G., Yang, F., Malyshev, S. L., Zubov, V. A., Egorova, T. A., & Li, B. (2002b). Climate/chemistry effects of the Pinatubo volcanic eruption simulated by the UIUC stratosphere/troposphere GCM with interactive photochemistry. *Journal of Geophysical Research: Atmospheres*, *107*(D21), ACL 12-1. https://doi.org/10.1029/2001JD000974

Rusin, N. P., & Bespalov, D. P. (1956). *The Soviet Meteorological Network and Methods of Observation*.

Russell, H. C. (1875). *Results of meteorological observations made in New South Wales during 1873*. https://babel.hathitrust.org/cgi/pt?id=njp.32101058084995&view=1up&seq=2&skin=2021

Russell, H. C. (1892). *Results of meteorological observations made in New South Wales during 1880, 1881, 1882*. https://hdl.handle.net/2027/uc1.$b619223

Russell, T. (1892). *Instructions for Voluntary Observers*. https://hdl.handle.net/2027/uc2.ark:/13960/t9668b56h

Sabine, E. (1845). *Observations made at the Magnetical and meteorological observatory at Toronto in Canada. Vol I - 1840, 1841, 1842*. https://catalog.hathitrust.org/Record/008607532?type%5B%5D=subject&lookfor%5B%5D=%22 Meteorology%22&filter%5B%5D=topicStr%3AMeteorology Observations&filter%5B%5D=ht_availability_intl%3AFull text&ft=

Sabine, E. (1850). *Observations made at the Magnetical and Meteorological Observatory at Hobarton, in Van Diemen Island and by the Antarctic Naval Expedition*. https://hdl.handle.net/2027/njp.32101051662847

Sabine, E. (1853). *Observations made at the Magnetical and Meteorological Observatory at Toronto in Canada. Vol II - 1843, 1844,1845*. https://hdl.handle.net/2027/njp.32101054593171

Sabine, E. (1857). *Observations made at the Magnetical and Meteorological Observatory at Toronto in Canada. Vol III - 1846, 1847,1848*. https://hdl.handle.net/2027/njp.32101054593197

Salinger, M. J. (1981). *New Zealand Climate: The Instrumental Record.* Victoria University.

Sato, M., Hansen, J. E., McCormick, M. P., & Pollack, J. B. (1993). Stratospheric aerosol optical depths, 1850-1990. *Journal of Geophysical Research*, *98*(D12). https://doi.org/10.1029/93jd02553

Scaife, A. A., & Smith, D. (2018). A signal-to-noise paradox in climate science. In *npj Climate and Atmospheric Science* (Vol. 1, Issue 1). Nature Research. https://doi.org/10.1038/s41612-018-0038-4

Schurer, A. P., Hegerl, G. C., Luterbacher, J., Brönnimann, S., Cowan, T., Tett, S. F. B., Zanchettin, D., & Timmreck, C. (2019). Disentangling the causes of the 1816 European year without a summer. *Environmental Research Letters*. https://doi.org/10.1088/1748-9326/ab3a10

Schurer, A. P., Hegerl, G., Ribes, A., Polson, D., Morice, C., & Tett, S. (2018). Estimating the Transient Climate Response from Observed Warming. *Journal of Climate*, *31*(20), 8645–8663. https://doi.org/10.1175/JCLI-D-17-0717.1

Schurer, A. P., Mann, M. E., Hawkins, E., Tett, S. F. B., & Hegerl, G. C. (2017). Importance of the pre-industrial baseline for likelihood of exceeding Paris goals. *Nature Climate Change*, *7*(8), 563–567. https://doi.org/10.1038/NCLIMATE3345

Schurer, A. P., Tett, S. F. B., & Hegerl, G. C. (2014). Small influence of solar variability on climate over the past millennium. *Nature Geoscience*, *7*(2), 104–108. https://doi.org/10.1038/ngeo2040

Schweizerische meteorologische Zentralanstalt. (1893). *Instruktionen für die Beobachter der meteorologischen Stationen der Schweiz*. https://doi.org/https://doi.org/10.3931/e-rara-83098

Schweizerischen Meteorologischen Central-Anstalt. (1891). Der Schweizerischen meteorologischen Beobachtungen, Achtundzwanzigster Jahrgang. *Annalen Der Schweizerischen Meteorologischen Central-Anstalt*. https://www.meteoswiss.admin.ch/assets/weather-archive/annalen-1891.pdf

Schweizerischen Meteorologischen Central-Anstalt. (1894). Der Schweizerischen meteorologischen Beobachtungen, Einunddreissigster Jahrgang. *Annalen Der Schweizerischen Meteorologischen Central-Anstalt*. https://www.meteoswiss.admin.ch/assets/weather-archive/annalen-1894.pdf

Schweizerischen Meteorologischen Central-Anstalt. (1898). Der Schweizerischen meteorologischen Beobachtungen, Funfunddreissigster Jahrgang. *Annalen Der Schweizerischen Meteorologischen Central-Anstalt*. https://www.meteoswiss.admin.ch/assets/weather-archive/annalen-1898.pdf

Schweizerischen Meteorologischen Zentralanstalt. (1956). *Annalen der Schweizerischen Meteorologischen Zentralanstalt, 1955 Zweiundneunzigster Jahrgang*. https://www.meteoswiss.admin.ch/assets/weather-archive/annalen-1955.pdf

Scott, R. H. (1875). *Instructions in the use of Meteorological Instruments* (London). Her Majesty's Stationary Office. https://digital.nmla.metoffice.gov.uk/IO_8dc44768-f95a-4dc8-b85f-edb7268b4507/

Sen Roy, S., Brata Saha, S., Roy Bhowmik, S. K., & Kundu, P. K. (2015). Analysis of monthly cloud climatology of the Indian subcontinent as observed by TRMM precipitation radar. *International Journal of Climatology*, *35*, 2080–2091. https://doi.org/10.1002/joc.4108

Serviciul Hidrometeorologic de Stat. (n.d.). *Service History*. http://www.meteo.md/index.php/en/about/history/

Shindell, D. T., Schmidt, G. A., Mann, M. E., & Faluvegi, G. (2004). Dynamic winter climate response to large tropical volcanic eruptions since 1600. *Journal of Geophysical Research D: Atmospheres*, *109*(5). https://doi.org/10.1029/2003jd004151

Shindell, D. T., Schmidt, G. A., Miller, R. L., & Rind, D. (2001). Northern Hemisphere winter climate response to greenhouse gas, ozone, solar, and volcanic forcing. *Journal of Geophysical Research Atmospheres*, *106*(D7), 7193–7210. https://doi.org/10.1029/2000JD900547

Sjolte, J., Adolphi, F., Guðlaugsdòttir, H., & Muscheler, R. (2021). Major Differences in Regional Climate Impact Between High- and Low-Latitude Volcanic Eruptions. *Geophysical Research Letters*, *48*(8). https://doi.org/10.1029/2020GL092017

Sjolte, J., Sturm, C., Adolphi, F., Vinther, B. M., Werner, M., Lohmann, G., & Muscheler, R. (2018). Solar and volcanic forcing of North Atlantic climate inferred from a process-based reconstruction. *Climate of the Past*, *14*(8), 1179–1194. https://doi.org/10.5194/cp-14-1179-2018

Slivinski, L. C., Compo, G. P., Sardeshmukh, P. D., Whitaker, J. S., McColl, C., Allan, R. J., Brohan, P., Yin, X., Smith, C. A., Spencer, L. J., Vose, R. S., Rohrer, M., Conroy, R. P., Schuster, D. C., Kennedy, J. J., Ashcroft, L., Brönnimann, S., Brunet, M., Camuffo, D., … Wyszyński, P. (2020). An Evaluation of the Performance of the Twentieth Century Reanalysis Version 3. *Journal of Climate*, *34*(4), 1417–1438. https://doi.org/10.1175/jcli-d-20-0505.1

Slivinski, L. C., Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Giese, B. S., McColl, C., Allan, R., Yin, X., Vose, R., Titchner, H., Kennedy, J., Spencer, L. J., Ashcroft, L., Brönnimann, S., Brunet, M., Camuffo, D., Cornes, R., Cram, T. A., Crouthamel, R., … Wyszyński, P. (2019). Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system. *Quarterly Journal of the Royal Meteorological Society*, *145*(724), 2876–2908. https://doi.org/10.1002/qj.3598

Slonosky, V. (2014). Historical climate observations in Canada: 18th and 19th century daily temperature from the St. Lawrence Valley, Quebec. *Geoscience Data Journal*, *1*(2), 103–120. https://doi.org/10.1002/gdj3.11

Smith, D. M., Scaife, A. A., Eade, R., Athanasiadis, P., Bellucci, A., Bethke, I., Bilbao, R., Borchert, L. F., Caron, L. P., Counillon, F., Danabasoglu, G., Delworth, T., Doblas-Reyes, F. J., Dunstone, N. J., Estella-Perez, V., Flavoni, S., Hermanson, L., Keenlyside, N., Kharin, V., … Zhang, L. (2020). North Atlantic climate far more predictable than models imply. *Nature*, *583*(7818), 796–800. https://doi.org/10.1038/s41586-020-2525-0

Smithsonian Institution. (1860). *Directions for meteorological observations, and the registry of periodical phenomena.* https://hdl.handle.net/2027/uiug.30112038054505

Smithsonian Institution. (1927). *World Weather Records Collected From Official Sources. Volume 79* (H. H. Clayton, F. Exner, G. Walker, G. C. Simpson, H. H. Clayton, & R. G. Mossman, Eds.). Smithsonian Institution. https://catalog.hathitrust.org/Record/009800874

Smithsonian Institution. (1934). *World Weather Records, 1921-1930, Volume 90* (Vol. 90).

Smithsonian Institution. (1947). *World Weather Records, 1931 - 1940, Volume 105* (H. H. Clayton & F. L. Clayton, Eds.). Smithsonian Institution.

Solomon, S. (1999). Stratospheric ozone depletion: A review of concepts and history. *Reviews of Geophysics*, *37*(3), 275–316. https://doi.org/10.1029/1999RG900008

Sparks, W. R. (1972). *The effect of thermometer screen design on the observed temperature*. https://library.wmo.int/doc_num.php?explnum_id=8131

Sprung, A. (1890). Bericht uber vergleichende Beobachtungen an verschiedenen Thermometer-Aufstellungen zu Gross-Lichterfelde bei Berlin. *Abhandlungen Des Koniglich Preussuschen Meteorologischen Instituts*, *1*(2), 17–19. https://hdl.handle.net/2027/nyp.33433069059404

St. George, S. (2014). An overview of tree-ring width records across the Northern Hemisphere. *Quaternary Science Reviews*, *95*, 132–150. https://doi.org/10.1016/j.quascirev.2014.04.029

Stenchikov, G., Hamilton, K., Robock, A., Ramaswamy, V., & Schwarzkopf, M. D. (2004). Arctic oscillation response to the 1991 Pinatubo eruption in the SKYHI general circulation model with a realistic quasi-biennial oscillation. *Journal of Geophysical Research: Atmospheres*, *109*(3). https://doi.org/10.1029/2003jd003699

Stenchikov, G., Hamilton, K., Stouffer, R. J., Robock, A., Ramaswamy, V., Santer, B., & Graf, H. F. (2006). Arctic Oscillation response to volcanic eruptions in the IPCC AR4 climate models. *Journal of Geophysical Research Atmospheres*, *111*(7). https://doi.org/10.1029/2005JD006286

Stenchikov, G. L., Kirchner, I., Robock, A., Graf, H. F., Antuña, J. C., Grainger, R. G., Lambert, A., & Thomason, L. (1998). Radiative forcing from the 1991 Mount Pinatubo volcanic eruption. *Journal of Geophysical Research: Atmospheres*, *103*(D12), 13837–13857. https://doi.org/10.1029/98JD00693

Stenchikov, G., Robock, A., Ramaswamy, V., Schwarzkopf, M. D., Hamilton, K., & Ramachandran, S. (2002). Arctic Oscillation response to the 1991 Mount Pinatubo eruption: Effects of volcanic aerosols and ozone depletion. *Journal of Geophysical Research Atmospheres*, *107*(24), ACL 28-1-ACL 28-16. https://doi.org/10.1029/2002JD002090

Sterl, A. (2003). On the (In)Homogeneity of Reanalysis Products. *Journal of Climate*, *17*, 3866–3873.

Sternzat, M. S. (1967). The development of Meteorological Instruments. In M. I. Budyko (Ed.), *The Main Geophysical Observatory during 50 years of the Soviet Regime*.

Stevenson, T. (1866). New description of box for holding thermometers. *Journal of the Scottish Meteorological Society*, *1*(2). https://babel.hathitrust.org/cgi/pt?id=hvd.hxgrmx&view=2up&seq=147

Stothers, R. B. (1984). The Great Tambora Eruption in 1815 and Its Aftermath. *Science*, *224*(4654), 1191–1198. https://doi.org/https://doi.org/10.1126/science.224.4654.1191

Stupart, R. F. (1900). *Report of the Meteorological Service of Canada for the year ended December 31 1897*. https://hdl.handle.net/2027/iau.31858046357228

Swingedouw, D., Mignot, J., Ortega, P., Khodri, M., Menegoz, M., Cassou, C., & Hanquiez, V. (2017). Impact of explosive volcanic eruptions on the main climate variability modes. In *Global and Planetary Change* (Vol. 150). https://doi.org/10.1016/j.gloplacha.2017.01.006

Symons. (1888). The Eruption of Krakatoa and Subsequent Phenomena. *Quarterly Journal of the Royal Meteorological Society*, *14*(68), 253–312.

Symons, G. J. (1877). On the climates of various British Colonies. *Symons's Monthly Meteorological Magazine*. https://babel.hathitrust.org/cgi/pt?id=umn.31951000919119a&view=1up&seq=483

Tacchini, P. (1879). *Annali Dell'Ufficio Centrale di Meteorologia, Serie II*.

Tarand, A., & Nordli, P. Ø. (2001). The Tallinn Temperature Series Reconstructed Back Half a Millennium by Use of Proxy Data. In *The Iceberg in the Mist: Northern Research in pursuit of a "Little Ice Age"* (pp. 189–199). Springer Netherlands. https://doi.org/10.1007/978-94-017-3352-6_9

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

The Imperial meteorological observatory, T. (1886). *Report of the meteorological observations for the ten years 1876 - 1885, made at the Imperial meteorological observatory of Tokio*. https://hdl.handle.net/2027/uc1.$c187292

Thomas, M. A., Giorgetta, M. A., Timmreck, C., Graf, H.-F., & Stenchikov, G. (2009). Simulation of the climate impact of Mt. Pinatubo eruption using ECHAM5-Part 2: Sensitivity to the phase of the QBO and ENSO. In *Atmos. Chem. Phys* (Vol. 9). www.atmos-chem-phys.net/9/3001/2009/

Thomas, M. A., Timmreck, C., Giorgetta, M. A., Graf, H.-F., & Stenchikov, G. (2009). Simulation of the climate impact of Mt. Pinatubo eruption using ECHAM5-Part 1: Sensitivity to the modes of atmospheric circulation and boundary conditions. In *Atmos. Chem. Phys* (Vol. 9). www.atmos-chem-phys.net/9/757/2009/

Thorne, P. W., Allan, R. J., Ashcroft, L., Brohan, P., Dunn, R. J. H., Menne, M. J., Pearce, P. R., Picas, J., Willett, K. M., Benoy, M., Bronnimann, S., Canziani, P. O., Coll, J., Crouthamel, R., Compo, G. P., Cuppett, D., Curley, M., Duffy, C., Gillespie, I., … Worley, S. J. (2017). Toward an Integrated Set of Surface Meteorological Observations for Climate Science and Applications. *Bulletin of the American Meteorological Society*, *98*(12), 2689–2702. https://doi.org/10.1175/BAMS-D-16-0165.1

Thorne, P. W., Menne, M. J., Williams, C. N., Rennie, J. J., Lawrimore, J. H., Vose, R. S., Peterson, T. C., Durre, I., Davy, R., Esau, I., Klein-Tank, A. M. G., & Merlone, A. (2016). Reassessing changes in diurnal temperature range: A new data set and characterization of data biases. *Journal of Geophysical Research: Atmospheres*, *121*(10), 5115–5137. https://doi.org/10.1002/2015JD024583

Thorne, P., & Willett, K. (2015). *The International Surface Temperature Initiative*. https://www.ecmwf.int/sites/default/files/elibrary/2015/13500-international-surface-temperature-initiative-poster.pdf

Timmreck, C. (2012). Modeling the climatic effects of large explosive volcanic eruptions. *Wiley Interdisciplinary Reviews: Climate Change*, *3*(6), 545–564. https://doi.org/10.1002/wcc.192

Tokarska, K. B., Schleussner, C. F., Rogelj, J., Stolpe, M. B., Matthews, H. D., Pfleiderer, P., & Gillett, N. P. (2019). Recommended temperature metrics for carbon budget estimates, model evaluation and climate policy. *Nature Geoscience*, *12*(12), 964–971. https://doi.org/10.1038/s41561-019-0493-5

Toohey, M., Krüger, K., Bittner, M., Timmreck, C., & Schmidt, H. (2014). The impact of volcanic aerosol on the Northern Hemisphere stratospheric polar vortex: Mechanisms and sensitivity to forcing structure. *Atmospheric Chemistry and Physics*, *14*(23), 13063–13079. https://doi.org/10.5194/acp-14-13063-2014

Toohey, M., Krüger, K., Niemeier, U., & Timmreck, C. (2011). The influence of eruption season on the global aerosol evolution and radiative impact of tropical volcanic eruptions. *Atmospheric Chemistry and Physics*, *11*(23), 12351–12367. https://doi.org/10.5194/acp-11-12351-2011

Toohey, M., & Sigl, M. (2017). Volcanic stratospheric sulfur injections and aerosol optical depth from 500 BCE to 1900 CE. *Earth System Science Data*, *9*(2), 809–831. https://doi.org/10.5194/essd-9-809-2017

Trepinska, J. B. (2007). Instrumentalne i wizualne obserwacje pogody w Obserwatorium Astronomicznym Uniwersytetu Jagiellońskiego. In K. Piotrowicz & R. Twardosz (Eds.), *Wahania klimatu w różnych skalach przestrzennych i czasowych* (pp. 31–37). https://ruj.uj.edu.pl/xmlui/bitstream/handle/item/258676/trepinska_instrumentalne_i_wizualne_obserwacje_pogody_2007.pdf?sequence=1&isAllowed=y

Trepinska, J. B., & Piotrowicz, K. (n.d.). *DZIEJE STACJI METEOROLOGICZNEJ UNIWERSYTETU JAGIELLOŃSKIEGO*. https://holmes.iigw.pl/~mbodzion/dydaktyka/mik/pliki/art2.pdf

Trewin, B. (2004). Effects of changes in algorithms used for the calculation of Australian mean temperature. *Australian Meteorological Magazine*, *53*(1), 1–11.

Trewin, B. (2010). Exposure, instrumentation, and observing practice effects on land temperature measurements. *Wiley Interdisciplinary Reviews: Climate Change*, *1*(4), 490–506. https://doi.org/10.1002/wcc.46

Trewin, B. (2018). *The Australian Climate Observations Reference Network-Surface Air Temperature (ACORN-SAT) version 2*.

Trewin, B., Braganza, K., Fawcett, R., Grainger, S., Jovanovic, B., Jones, D., Martin, D., Smalley, R., & Webb, V. (2020). An updated long-term homogenized daily temperature data set for Australia. *Geoscience Data Journal*, *7*(2), 149–169. https://doi.org/10.1002/gdj3.95

United Nations. (2015). Adoption of the Paris Agreement. In *Conference of the Parties on its twenty-first session* (Issue December). https://doi.org/FCCC/CP/2015/L.9

U.S. Department of Commerce. (1966). *World Weather Records 1951-60 Vol 2 Europe*. U.S. Department of Commerce. https://archive.org/details/worldweatherreco02smit/page/n1/mode/2up

U.S. Department of Commerce. (1979). *World Weather Records 1961-70 Vol 2 Europe*. https://archive.org/details/worldweatherreco020smit/page/n1/mode/2up?q=stand

US Weather Bureau. (1959). *World Weather Records, 1941 - 1950*. https://archive.org/details/worldweatherreco4150smit/mode/2up?q=salisbury

Valler, V., Franke, J., Brugnara, Y., Burgdorf, A., Lundstad, E., Hand, R., Samakinwa, E., Lipfert, L., Friedman, A., & Bronnimann, S. (2023). ModE-RA - A global monthly paleo-reanalysis of the modern era (1421 to 2008): Set 1420-3_1850-1. In *World Data Center for Climate (WDCC) at DKRZ*. https://doi.org/https://doi.org/10.26050/WDCC/ModE-RA_s14203-18501

Valler, V., Franke, J., Brugnara, Y., Samakinwa, E., Hand, R., Lundstad, E., Burgdorf, A.-M., Lipfert, L., Friedman, A. R., & Brönnimann, S. (2024). ModE-RA: a global monthly paleo-reanalysis of the modern era 1421 to 2008. *Scientific Data*, *11*(1), 36. https://doi.org/10.1038/s41597-023-02733-8

van Engelen, A. F. V., & Geurts, H. A. M. (1985). *Nicolaus Cruquius (1678-1754) and his meteorological observations*. https://cdn.knmi.nl/knmi/pdf/bibliotheek/knmipubmetnummer/knmipub165_IV.pdf

van Engelen, A. F. V., & Nellestijn, J. W. (1995). *Monthly, seasonal and annual means of the air temperature in tenths of centigrades in De Bilt, Netherlands, 1706–1995*.

van Ulden, A., van Oldenborgh, G. J., & van der Schrier, G. (2009). *The Construction of a Central Netherlands Temperature*.

https://cdn.knmi.nl/system/data_center_publications/files/000/068/325/original/CNT.pdf?1495621137

Veðráttan. (1962). Arsyfirlit samid a vedurstofunni. *Veðráttan*, 122. https://timarit.is/page/3128806

Venema, V. (2016). *Early global warming: How much did the world warm during the transition to Stevenson screens around 1900?* http://variable-variability.blogspot.com/2016/02/early-global-warming-transition-Stevenson-screens.html

Venema, V., Anguilar, E., Auchmann, R., Auer, I., Brandsma, T., Chimani, B., Gilbart, A., Toreti, A., Mestre, O., & Vertacnik, G. (2014). Parallel measurements to study inhomogeneities in daily data. *Homogenisation Seminar*.

Venema, V., Mestre, O., Aguilar, E., Auer, I., Guijarro, J. A., Domonkos, P., Vertacnik, G., Szentimrey, T., Stepanek, P., Zahradnicek, P., Viarre, J., Müller-Westermeier, G., Lakatos, M., Williams, C. N., Menne, M. J., Lindau, R., Rasol, D., Rustemeier, E., Kolokythas, K., … Brandsma, T. (2012). Benchmarking homogenization algorithms for monthly data. *Climate of the Past*, *8*(1), 89–115. https://doi.org/10.5194/cp-8-89-2012

Vigurs, C. C. (1935). Temperature at Falmouth. *The Meteorological Magazine*, *70*(832), 91–92. https://archive.org/details/sim_meteorological-magazine_1935-05_70_832/page/91/mode/1up

Vincent, L. A. (1998). A Technique for the Identification of Inhomogeneities in Canadian Temperature Series. *Journal of Climate*, *11*(5), 1094–1104.

Vincent, L. A., Hartwell, M. M., & Wang, X. L. (2020). A Third Generation of Homogenized Temperature for Trend Analysis and Monitoring Changes in Canada's Climate. *Atmosphere - Ocean*. https://doi.org/10.1080/07055900.2020.1765728

Vincent, L. A., Milewska, E. J., Hopkinson, R., & Malone, L. (2009). Bias in Minimum Temperature Introduced by a Redefinition of the Climatological Day at the Canadian Synoptic Stations. *Journal of Applied Meteorology and Climatology*, *48*(10), 2160–2168. https://doi.org/10.1175/2009JAMC2191.1

Visbeck, M. H., Hurrell, J. W., Polvani, L., & Cullen, H. M. (2001). The North Atlantic oscillation: Past, present, and future. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(23), 12876–12877. https://doi.org/10.1073/PNAS.231391598/ASSET/EB72ACB2-1128-4379-ACED-1394F1604934/ASSETS/GRAPHIC/PQ2313915001.JPEG

Visser, H. (2005). *The significance of climate change in the Nederlands. An analysis of historical and future trends (1901-2020)*. https://www.researchgate.net/publication/27451699_The_significance_of_climate_change_in_the_Nederlands_An_analysis_of_historical_and_future_trends_1901-2020

Voznessensky. (1915). *Annales de L'Observatoire Physique Central Nicolas, 1913*. https://library.oarcloud.noaa.gov/docs.lib/htdocs/rescue/cd116_pdf/LSN0368.PDF

Waliser, D., J. Gleckler, P., Ferraro, R., E. Taylor, K., Ames, S., Biard, J., G. Bosilovich, M., Brown, O., Chepfer, H., Cinquini, L., J. Durack, P., Eyring, V., Mathieu, P. P., Lee, T., Pinnock, S., L. Potter, G., Rixen, M., Saunders, R., Schulz, J., … Tuma, M. (2020). Observations for Model Intercomparison Project (Obs4MIPs): Status for CMIP6. *Geoscientific Model Development*, *13*(7), 2945–2958. https://doi.org/10.5194/GMD-13-2945-2020

Wallis, & Beale. (1669). Some observations concerning the baroscope and thermoscope, made and communicated by Doctor I. Wallis at Oxford, and Dr. I Beale at Yeovil in Somerset, deliver'd here according to the several dates, when they were imparted. Dr. Beale in those letters of his dated Decemb.18. Decemb. 29. 1669. and Januar. 3. 1670. *Philosophical Transactions (1665-1678)*, *4*, 1113–1120.

Waluew, P. A. V., & Wild, H. (1881). *Die Temperatur-Verhaltnisse des Russischen Reiches*. https://library.oarcloud.noaa.gov/foreign_climate.lib/FCD_002_pdf/QC989R8M48suppl-no1.pdf

Wang, J., & Yan, Z.-W. (2016). Urbanization-related warming in local temperature records: a review. *Atmospheric and Oceanic Science Letters*, *9*(2), 129–138. https://doi.org/10.1080/16742834.2016.1141658

Westcott, N., Andsager, K., Stoecker, L., Spinar, M., Smith, R., Obrecht, R., Kruk, M., Truesdell, R., & O'connell, D. (2011). *Quality Control of 19 th Century Weather Data*.

Whipple, G. M. (1883). Report on experiments made at the Kew Observatory with thermometer screens of different patterns during 1879, 1880, and 1881. *Appendix II of the Quarterly Weather Report*, 13–18.

Wijngaard, J. B., Klein Tank, A. M. G., & Können, G. P. (2003). Homogeneity of 20th century European daily temperature and precipitation series. *International Journal of Climatology*, *23*(6), 679–692. https://doi.org/10.1002/joc.906

Wild, H. (1869). *Instruction fur Meteorologische Stationen*. https://library.oarcloud.noaa.gov/foreign_climate.lib/FCD_002_pdf/QC989R8M48t1.pdf

Wild, H. (1873). On the exposure of thermometers for the calculation of air temperature. *International Meteorological Conference, Vienna, 1873*. https://babel.hathitrust.org/cgi/pt?id=hvd.hxgnti&view=1up&seq=94

Wild, H. (1875). *Instruction fur Meteorologische Stationen*. https://library.oarcloud.noaa.gov/foreign_climate.lib/FCD_002_pdf/QC989R8M48t5.pdf

Wild, H. (1881). *Jahresbericht des Physikalischen Central-Observatoriums fur 1879 und 1880.* https://library.oarcloud.noaa.gov/foreign_climate.lib/FCD_002_pdf/QC989R8M48t7.pdf

Wild, H. (1883a). *Jahresbericht des Physikalischen Central-Observatoriums fur 1881 und 1882.* https://library.oarcloud.noaa.gov/foreign_climate.lib/FCD_002_pdf/QC989R8M48t8.pdf

Wild, H. (1883b). *Repertorium fur Meteorologie, Volume VIII*. https://library.oarcloud.noaa.gov/foreign_climate.lib/FCD_002_pdf/QC989R8M48t8.pdf

Wild, H. (1885). *Repertorium fur Meteorologie, Volume IX*. https://library.oarcloud.noaa.gov/foreign_climate.lib/FCD_002_pdf/QC989R8M48t9.pdf

Wild, H. (1887a). Further tests on the determination of the true air temperature. *Meteorol. Sbornik*, *10*(10).

Wild, H. (1887b). *Repertorium fur Meteorologie, Volume X*. https://library.oarcloud.noaa.gov/foreign_climate.lib/FCD_002_pdf/QC989R8M48t10.pdf

Wild, H. (1888). *Repertorium fur Meteorologie, Volume XI*. https://library.oarcloud.noaa.gov/foreign_climate.lib/FCD_002_pdf/QC989R8M48t11.pdf

Wild, H. (1889). *Repertorium fur Meteorologie, Volume XII*. https://library.oarcloud.noaa.gov/foreign_climate.lib/FCD_002_pdf/QC989R8M48t12.pdf

Wild, H. (1891a). *Repertorium fur Meteorologie*.

Wild, H. (1891b). Uber den einfluss der aufstellung auf die angaben der thermometer zur bestimmung der lufttemperatur. *Repertorium Fur Meteorologie*, *14*. https://library.oarcloud.noaa.gov/foreign_climate.lib/FCD_002_pdf/QC989R8M48t14.pdf

Wilks, D. S. (2019). *Statistical Methods in the Atmospheric Sciences* (Fourth). Elsevier. https://doi.org/10.1016/C2017-0-03921-6

Willett, K., Williams, C., Jolliffe, I. T., Lund, R., Alexander, L. V., Brönnimann, S., Vincent, L. A., Easterbrook, S., Venema, V., Berry, D., Warren, R. E., Lopardo, G., Auchmann, R., Aguilar, E., Menne, M. J., Gallagher, C., Hausfather, Z., Thorarinsdottir, T., & Thorne, P. W. (2014). A framework for benchmarking of homogenisation algorithm performance on the global scale. *Geoscientific Instrumentation, Methods and Data Systems*, *3*(2), 187–200. https://doi.org/10.5194/gi-3-187-2014

Williams, C. N., Menne, M. J., & Thorne, P. W. (2012). Benchmarking the performance of pairwise homogenization of surface temperatures in the United States. *Journal of Geophysical Research Atmospheres*, *117*(5). https://doi.org/10.1029/2011JD016761

Williamson, F., Allan, R., Ren, G., Lee, T. cheung, Lui, W. hong, Kubota, H., Matsumoto, J., Luterbacher, J., Wilkinson, C., & Wood, K. (2018). Collating Historic Weather Observations for the East Asian Region: Challenges, Solutions, and Reanalyses.

Advances in Atmospheric Sciences, *35*(8), 899–904. https://doi.org/10.1007/s00376-017-7259-z

Winkler, P. (2009). Revision and necessary correction of the long-term temperature series of Hohenpeissenberg, 1781-2006. *Theoretical and Applied Climatology*, *98*(3–4), 259–268. https://doi.org/10.1007/s00704-009-0108-y

World Meteorological Organization. (2003). *Guidelines on climate metadata and homogenization*. www.wmo.ch/web/wcp/wcdmp/wcdmp.html

World Meteorological Organization. (2016). *Guidelines on Best Practices for Climate Data Rescue*.

World Meteorological Organization. (2019). *WIGOS Metadata Standard*. https://library.wmo.int

World Meteorological Organization. (2020). *Guidelines on Homogenization* (V. Venema, B. Trewin, X. L. Wang, T. Szentrimrey, M. Lakatos, E. Aguilar, I. Auer, J. A. Guijarro, M. Menne, C. Oria, W. S. R. L. Louamba, & G. Rasul, Eds.; 2020 editi). https://library.wmo.int/doc_num.php?explnum_id=10352

World Meteorological Organization. (2021). *Guide to Instruments and Methods of Observation. Volume I - Measurement of Meteorological Variables*.

World Meteorological Organization. (2022). *The 2022 GCOS Implementation Plan*.

Wunderlich, F., & Mitchell, D. M. (2017). Revisiting the observed surface climate response to large volcanic eruptions. *Atmospheric Chemistry and Physics*, *17*(1), 485–499. https://doi.org/10.5194/acp

Yule, E. L., Hegerl, G., Schurer, A. P., & Hawkins, E. (2023). Using early extremes to place the 2022 UK heat waves into historical context. *Atmospheric Science Letters*, *24*(7). https://doi.org/10.1002/asl.1159

Zambri, B., LeGrande, A. N., Robock, A., & Slawinska, J. (2017). Northern Hemisphere winter warming and summer monsoon reduction after volcanic eruptions over the last millennium. *Journal of Geophysical Research: Atmospheres*, *122*(15), 7971–7989. https://doi.org/10.1002/2017JD026728

Zambri, B., & Robock, A. (2016). Winter warming and summer monsoon reduction after volcanic eruptions in Coupled Model Intercomparison Project 5 (CMIP5) simulations. *Geophysical Research Letters*, *43*(20), 10,920-10,928. https://doi.org/10.1002/2016GL070460

Zanchettin, D., Bothe, O., Graf, H. F., Lorenz, S. J., Luterbacher, J., Timmreck, C., & Jungclaus, J. H. (2013). Background conditions influence the decadal climate response to strong volcanic eruptions. *Journal of Geophysical Research Atmospheres*, *118*(10), 4090–4106. https://doi.org/10.1002/jgrd.50229

Zanchettin, D., Timmreck, C., Graf, H. F., Rubino, A., Lorenz, S., Lohmann, K., Krüger, K., & Jungclaus, J. H. (2012). Bi-decadal variability excited in the coupled ocean-atmosphere system by strong tropical volcanic eruptions. *Climate Dynamics*, *39*(1–2), 419–444. https://doi.org/10.1007/S00382-011-1167-1/FIGURES/7

Zanchettin, D., Timmreck, C., Toohey, M., Jungclaus, J. H., Bittner, M., Lorenz, S. J., & Rubino, A. (2019). Clarifying the Relative Role of Forcing Uncertainties and Initial-Condition Unknowns in Spreading the Climate Response to Volcanic Eruptions. *Geophysical Research Letters*, *46*(3), 1602–1611. https://doi.org/10.1029/2018GL081018

Zhu, F., Emile-Geay, J., Anchukaitis, K. J., Hakim, G. J., Wittenberg, A. T., Morales, M. S., Toohey, M., & King, J. (2022). A re-appraisal of the ENSO response to volcanism with paleoclimate data assimilation. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-28210-1

Zhu, F., Emile-Geay, J., Hakim, G. J., King, J., & Anchukaitis, J. (2020). Resolving the differences in the simulated and reconstructed climate response to volcanism over the last millennium. *Geophysical Research Letters*, *47*(8), 1–12. https://doi.org/https://doi.org/10.1029/2019GL086908

Zhu, M. (2012). *Typhoons, meteorological intelligence, and the inter-port mercantile community in nineteenth-century China* [Binghampton University, State University of New York]. https://www.proquest.com/docview/1037995425?pq-origsite=gscholar&fromopenview=true