Quantifying exposure biases in early instrumental land surface air temperature observations

Article

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: http://dx.doi.org/10.1002/joc.8401

Publisher: Wiley

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

RESEARCH ARTICLE

International Journal of Climatology | RMetS

Quantifying exposure biases in early instrumental land surface air temperature observations

Emily J. Wallis[1] | **Timothy J. Osborn**[1] | **Michael Taylor**[1] |
Philip D. Jones[1] | **Manoj Joshi**[1] | **Ed Hawkins**[2]

[1]Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich, UK

[2]National Centre for Atmospheric Science, Department of Meteorology, University of Reading, Reading, UK

Correspondence
Emily J. Wallis, Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich, UK.
Email: emily.wallis@uea.ac.uk

Abstract

Exposure biases are a pervasive non-climatic change in land air temperature records which have been introduced as a result of changes in the way thermometers were sheltered from solar radiation and the elements over time. Exposure biases have not been widely accounted for in observational records, due to difficulties detecting/correcting the bias using traditional homogenisation techniques; therefore, exposure biases still contribute significant uncertainty to the early period in global temperature compilations. Here, an empirical approach to address the bias arising from the introduction of Stevenson-type screens from the late-19th century is presented. The approach consists of: (1) an empirical analysis of 54 parallel measurement series to identify the characteristics of the exposure bias in four exposure classes; (2) the development of bias-estimation models based on an analysis of which variables influence the bias; and (3) the application of the models to an extended version of CRUTEM5 (CRUTEM5_ext), based on exposure metadata, to quantify and reduce the bias. Step one identified differences between the temperatures recorded in Stevenson screens and early exposures, which vary seasonally, diurnally, and with location and exposure class. The largest biases (in mean temperatures) were found in freestanding exposures (up to −0.78°C annually) and in summer, while the smallest biases were generally found in wall-mounted exposures (near-0°C annually) and in winter. Significant relationships between the bias and temperature, downward top of atmosphere and/or received shortwave downward solar radiation were found in each exposure class and led to the development of three regression-based bias-estimation models. Application of these models to 1,960 mid-latitude stations in CRUTEM5_ext, resulted in small (≤0.016°C) positive adjustments to the Northern Hemisphere mid-latitude mean before 1880, and larger (≤−0.1°C) negative adjustments to the Northern and Southern Hemisphere mid-latitude means between 1882–1934 and 1856–1900, respectively. Larger adjustments were estimated regionally: up to −0.57°C annually and −0.79°C seasonally in individual grid cells.

1 | INTRODUCTION

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). LSAT records can be compromised by non-climatic changes in the data—known as inhomogeneities—which can result from changes in station location or surroundings and from changes to instrumentation or observing practices (Jones, 2016; Trewin, 2010). Inhomogeneities can be large in magnitude, of a similar scale to true climatic responses to forcing at individual stations, therefore affected observational records require correction or consideration before they can be used to study climate variability and change (World Meteorological Organization, 2020). Where inhomogeneities affect individual stations, correction is often possible using homogenisation algorithms (Venema et al., 2012); however, where inhomogeneities—known as biases—systematically affect a large proportion of the observations in a region, 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. Prior to the development and widespread adoption of variants of the Stevenson screen (Figure 1a) in the late-19th and early-20th centuries, 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 (referred to as exposures) varied regionally and included mounting thermometers on poleward-facing walls, stands, and within various freestanding screens (Figure 1b–e). Each type of exposure influenced temperature readings differently, by altering the influence of solar radiation on the thermometer, thus introducing inhomogeneities into station temperature records when the transition to Stevenson-type screens was made (Parker, 1994).

The impact of differing thermometer exposures on temperature readings has been investigated previously, including by: Chenoweth (1992) in North America; Böhm et al. (2010), Brunet et al. (2006, 2011), Butler et al. (2005), Gaster (1882), Margary (1924), Marriott (1879) and Nordli et al. (1996, 1997) in Europe; Gill (1882) in South Africa; 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 conditions and type of exposure. Despite these well-documented differences, and assessments (e.g., Frank et al., 2007; Moberg et al., 2003) presenting evidence of the likely presence of the exposure bias in early temperature observations, relatively few exposure bias-specific corrections have been incorporated into global temperature datasets. In HadCRUT5, for example, only records from Australia (Ashcroft et al., 2012), the Greater Alpine Region (Böhm et al., 2010) and Spain
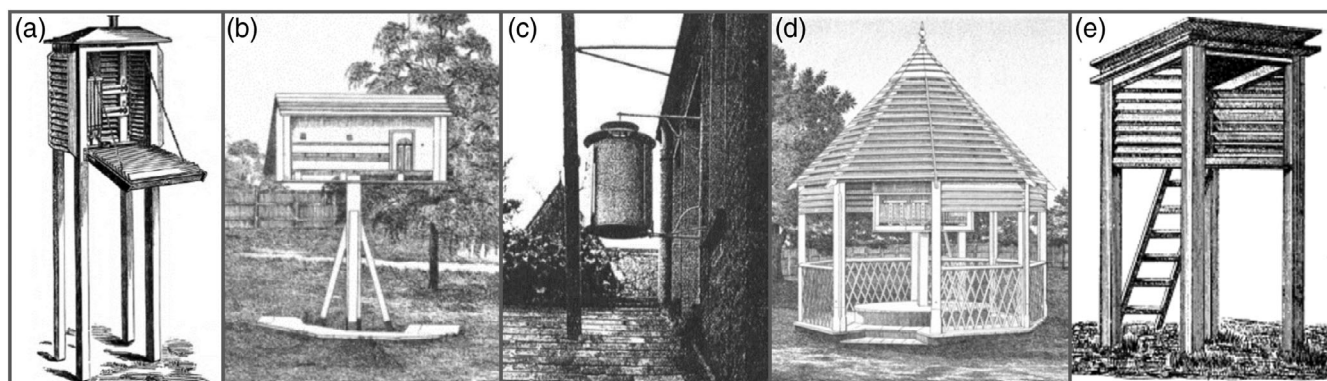

FIGURE 1 Examples of common thermometer exposures. (a) Stevenson screen, (b) Glaisher stand (Open), (c) a type of wall-mounted screen, (d) Summerhouse (Intermediate) and (e) Wild Hut (Closed). Image sources: (a) Gaster (1882); (b), (d) Royal Society of New South Wales; (c) Mawley (1897); (e) Wild (1891).

(Brunet et al., 2006) are known to have been explicitly adjusted to account for exposure biases. 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 homogenisation, such as neighbour comparisons, 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 (homogenisation algorithms often operate on annual timescales) 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 (Trewin, 2010).

As such, it is necessary to account for exposure biases in long observational records, including in global temperature compilations. 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 realizations 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) $20° S–20° N$ (Morice et al., 2012). Current knowledge, however, suggests that this is an over-simplistic 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). Since the development of the HadCRUT error model, additional parallel measurements and assessments of the bias have become available, and metadata has become more easily accessible as a result of digitisation efforts (e.g., Allan et al., 2011). This paper therefore aims to address some of the limitations identified above by (a) updating Parker's (1994) assessment of the characteristics of the exposure bias (Section 3); (b) developing models to estimate the magnitude and seasonal nature of the exposure bias (Section 4); and (c) applying the developed models to the stations in an extended version (CRUTEM5_ext) of

CRUTEM5 (Osborn et al., 2021), using information about historic exposures compiled from a review of metadata and relevant literature (Section 5).

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 1; Figure 2). 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:

$$\Delta T = T_{SS} - T_{\text{Hist}}, \tag{1}$$

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 (from which the values for the other exposure were calculated).

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, and how they relate to one another, the reader is directed to Thorne et al. (2016).

Each parallel measurement series was categorized 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 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;

TABLE 1 Details of the parallel measurement series collated and analysed in this paper.

#	Source	Location	Duration	T_x	T_n	T_m	DTR	ΔT_x	ΔT_n	ΔT_m	ΔDTR	Exposure category	Included in the model development?	Surrogate data
1	AEMET Brunet, Pers. Comms.	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	C	Wall	No—$T_{mean} \neq \frac{1}{2}(T_{max} + T_{min})$	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—Outside of 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
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

TABLE 1 (Continued)

#	Source	Location	Duration	Variables								Exposure category	Included in the model development?	Surrogate data
				T_x	T_n	T_m	DTR	ΔT_x	ΔT_n	ΔT_m	ΔDTR			
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
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—$T_{mean} \neq \frac{1}{2}(T_{max} + T_{min})$	N/A
29	Muller (1984)	46.8° N, 7° E; Payerne, Switzerland	07/1979–06/1981			C	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—$T_{mean} \neq \frac{1}{2}(T_{max} + T_{min})$	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		

(Continues)

TABLE 1 (Continued)

#	Source	Location	Duration	Variables								Exposure category	Included in the model development?	Surrogate data
				T_x	T_n	T_m	DTR	ΔT_x	ΔT_n	ΔT_m	ΔDTR			
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 for Valencia	N/A
38		57.2° N, 2.2° W; Aberdeen, UK	01/1871–12/1900			C		Y	Y	C	C	Wall	Yes	CRUTEM5
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—Outside of 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

TABLE 1 (Continued)

#	Source	Location	Duration	Variables								Exposure category	Included in the model development?	Surrogate data
				T_x	T_n	T_m	DTR	ΔT_x	ΔT_n	ΔT_m	ΔDTR			
51		63.4° N, 20.3° W; Vestmannaeyjar, Iceland	1948–1962							Y		Wall		N/A
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

*For T_{ss} only

Note: In the 'Variables' columns, Y denotes variables available in the source, and C where they were calculated here. An asterisk in the 'Included in the model development?' column denotes series which were used to inform the model development, but, due to missing or incomplete T_{Hist} values, were not included as input in the final model.
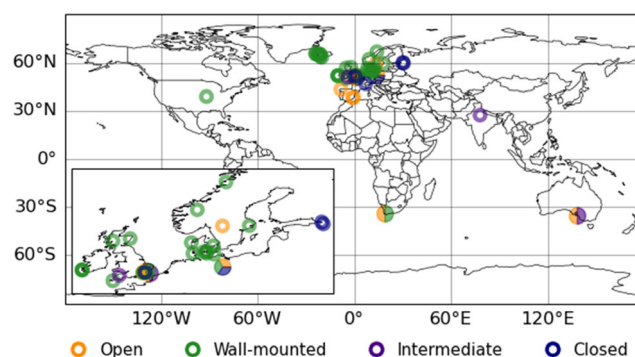

FIGURE 2 Locations of the parallel measurement series collated and analysed. [Colour figure can be viewed at wileyonlinelibrary.com]

ii. Wall-mounted exposures: any exposure where a thermometer is mounted on a wall (Figure 1c), fence or window, either screened or unscreened;

iii. Intermediate exposures: freestanding exposures such as thermometer sheds or summerhouses (Figure 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 1e), which fully enclose the thermometer.

Of the 54 parallel measurements series obtained, 15 were categorized as open (8 Glaisher, 6 Montsouris and 1 other), 30 as wall-mounted, 4 as intermediate (2 summerhouses and 2 thatched shelters) and 5 as closed (all Wild huts).

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 3–6 and are discussed below.

3.1 | Open exposures

Monthly mean ΔT values (Figure 3) 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
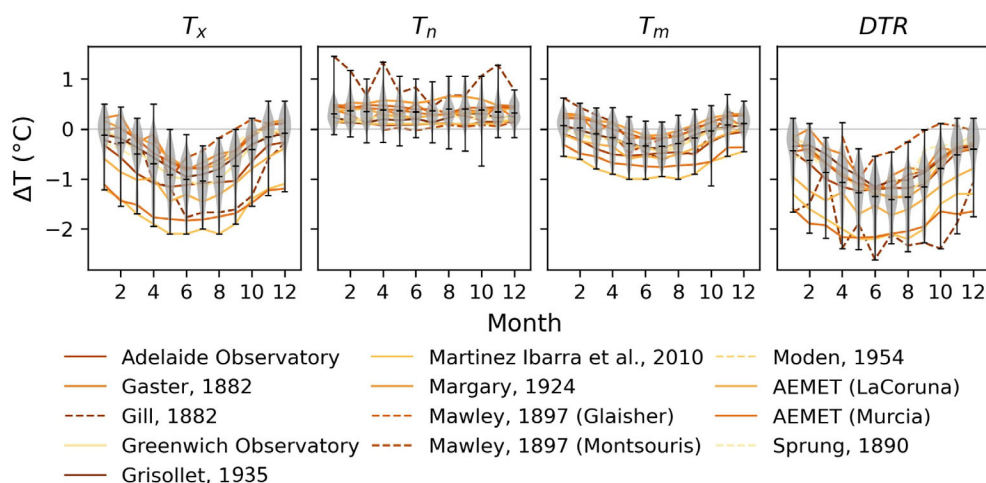
FIGURE 3 Differences between monthly mean temperatures recorded in Stevenson screens and open exposures. The violin plots (black with grey shading) sometimes extend beyond the range of the individual series (coloured lines) because the violin plots show the mean and distribution of all individual monthly ΔT values whereas the coloured lines show 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. [Colour figure can be viewed at wileyonlinelibrary.com]

−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.
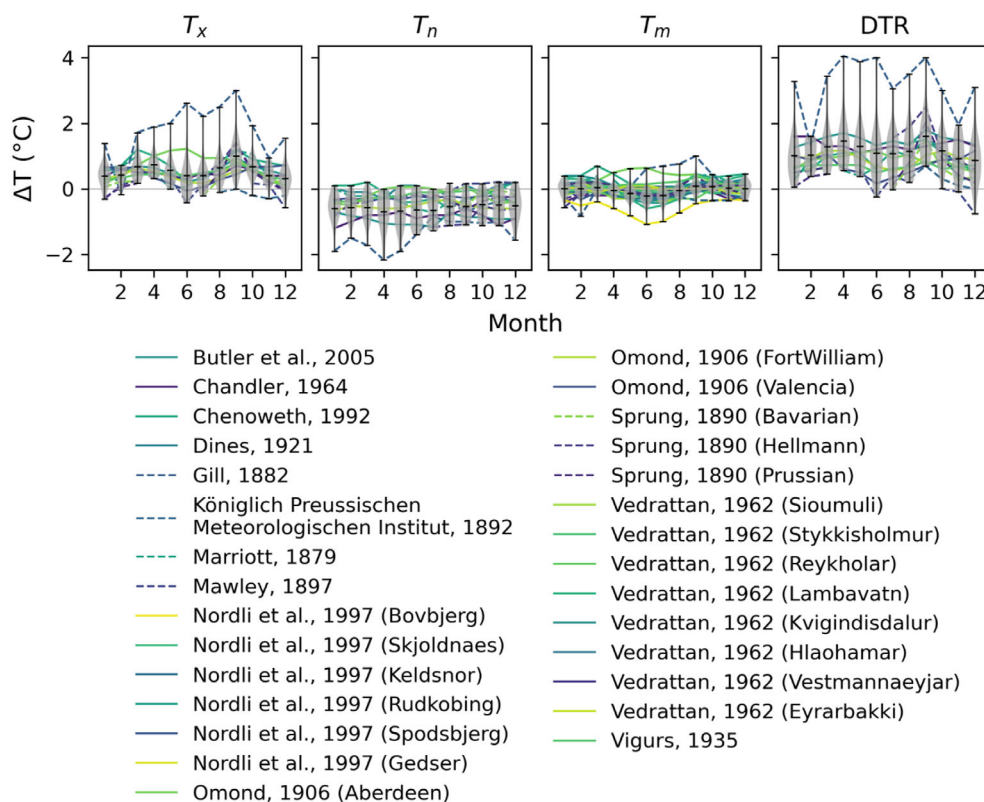
These differences are broadly consistent with those outlined by Parker (1994) who concluded that Glaisher

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 has 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).

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 4). 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

FIGURE 4 As Figure 3, but for differences between Stevenson screens and wall-mounted exposures. [Colour figure can be viewed at wileyonlinelibrary.com]

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 signs 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.

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 4, 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.

3.3 | Intermediate exposures

The comparison between intermediate exposures and Stevenson screens (Figure 5) also shows differences between
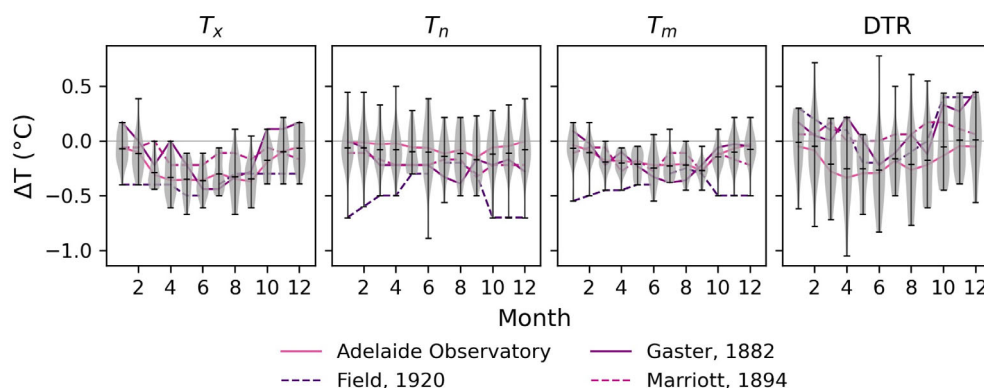
FIGURE 5 As Figure 3, but for differences between Stevenson screens and intermediate exposures. [Colour figure can be viewed at wileyonlinelibrary.com]

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 1 year of data and are therefore noisier.

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 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 mean ΔDTR is 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.

3.4 | Closed exposures

Maximum temperatures in closed exposures are consistently warmer than in Stevenson screens (Figure 6), 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 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.

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 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.

ΔDTR 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 DTRs, 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/locations.

4 | MODELLING THE EXPOSURE BIAS

The findings detailed in Section 3 reinforce the need to account for the exposure bias. In this section, the development of models to estimate the exposure bias in T_m, for each exposure class, is outlined.

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)
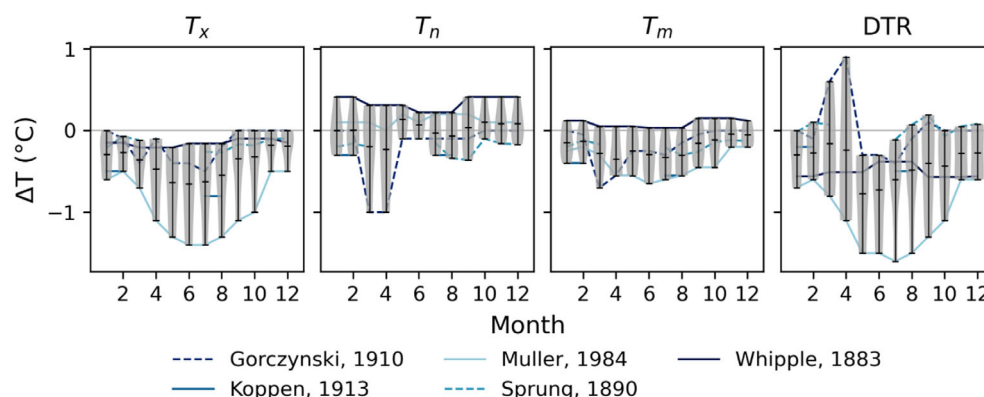

FIGURE 6 As Figure 3, but for differences between Stevenson screens and closed exposures. [Colour figure can be viewed at wileyonlinelibrary.com]

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_ext.

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., Ashcroft et al., 2022; Margary, 1924). As the intention is to use any identified relationship(s) to estimate monthly mean exposure biases in CRUTEM5_ext, 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 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). 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 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, the bias-adjusted version of ERA5 (Cucchi et al., 2020). 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_ext. 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} were obtained to maximize the number of series available for analysis (see Table 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):

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

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 estimates (and the relationship between the bias and the selected predictor) were robust to the choice of input data.

4.1 | Open exposures

The results of the robust linear regressions (Table 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 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; 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 normalizing 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 and Figure 7a show the details of these regressions.

Applying Equations 3 to 5 results in superior estimations of ΔT_m 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 3; Figure 7b) and seasonal cycle of the bias (Figure 7c). This is particularly evident in Figure 7c which shows the close agreement between the observations and estimates for the AEMET (La Coruna) series.

The deviations between the observed and estimated monthly and annual mean biases (Figure 7b) 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.

4.2 | Wall-mounted exposures

Our physically-based reasoning (Section 3.2) suggests the bias in wall-mounted exposures may be related to solar radiation, however, the linear regression analyses (Table 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 (Section 3.2) of a biannual cycle in ΔT_m suggest a quadratic relationship may be more appropriate than linear regression.

Quadratic regression (Table 2; Figure 8a) shows a significant 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 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 8a. 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:

TABLE 2 Results of the robust linear regression analyses for each class of exposure and explanatory variable.

Exposure category		T_{Hist} Int	Slope	r	p < 0.05	TOA Int	Slope	r	p < 0.05	SWD 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[a]	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	✓

Exposure category										T_a Int	Slope	r	p < 0.05
Open	Maximum bias (maxΔT_m)									0.881	−0.058	−0.57	✓
	Minimum bias (minΔT_m)									0.390	−0.062	−0.88	✓

Note: 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/m²) except first term for the Quadratic which is °C/(W/m²)². Final regression coefficients for the model, with the series from Gill (1882) excluded, are given in Equation 6.

[a] Results including the series from Gill (1882) as input.

$$\widehat{\Delta T_m}(m,i) = -8.38\times10^{-6}\times TOA(m,i)^2 + 0.005 \qquad (6)$$
$$\times TOA(m,i) - 0.469.$$

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 8c which shows the correct timing of the 'double-peak' of the bias, but an overestimated amplitude, and Figure 8b which shows greater variance in the observations than the estimates. This is not unexpected. As noted in Section 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 8a). In addition, although lower than for the open model, the skill scores are positive in all-but-one series analysed (Table 3), 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.

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), 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 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
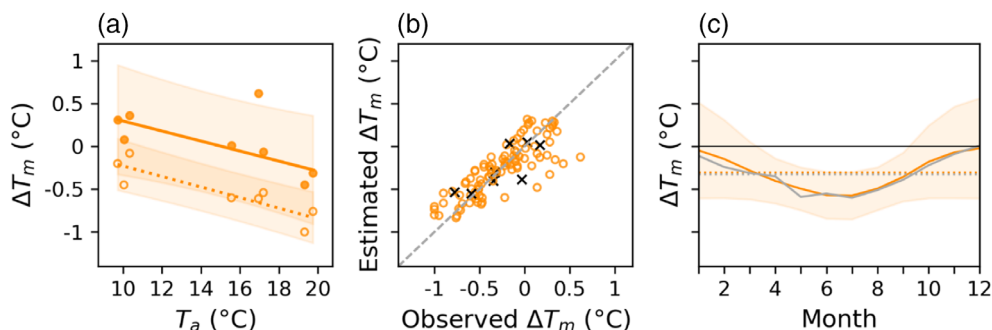
FIGURE 7 (a) Relationship between annual maximum ΔT_m (solid line and filled markers), annual minimum ΔT_m (dotted line and unfilled markers) and annual mean T_{Hist} (T_a) with shaded 95% confidence intervals for open exposures; (b) observed versus estimated monthly (orange circles) and annual mean (black crosses) ΔT_m; and (c) observed (grey) and estimated (orange) ΔT_m with shaded 95% confidence interval for the AEMET La Coruna series (Brunet, pers. comms). 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). [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 3 Lower and upper quartiles of the key performance indicators for the best-performing statistical model in each exposure class.

	RMSE (°C)	Skill	Annual ΔT_m minus Annual $\Delta \hat{T}_m$ (°C)
Open	0.09–0.21	0.39–0.96	−0.08 to 0.08
Wall-mounted	0.14–0.24	0.09–0.44	−0.12 to 0.11
Intermediate	0.07–0.09	0.78–0.88	−0.02 to 0.01
Closed	0.1–0.24	0.47–0.75	−0.08 to 0.07

($r = -0.66$) 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.

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 9b,c), with low RMSE scores for *TOA* (*SWD*) between 0.07°C and 0.1°C (0.09°C–0.1°C) and reasonably high, positive, skill scores between 0.71 and 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 3; Figure 9). 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 5). 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.

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), 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 3.4, as well as with our findings that the bias is largest in the summer.

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 3; Figure 10b,c) obtained using:
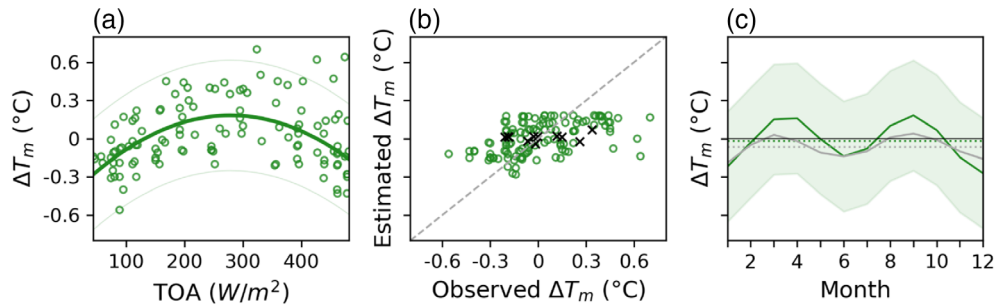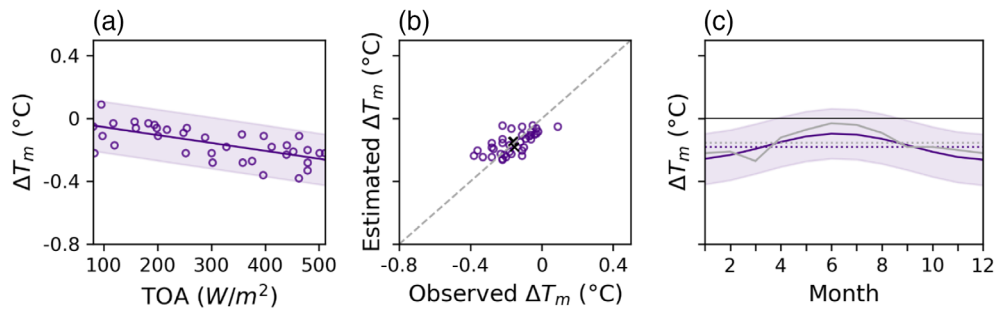
FIGURE 8 (a) Relationship between monthly ΔT_m and top of atmosphere solar radiation with 95% confidence interval for wall-mounted exposures; (b) observed versus estimated monthly (green circles) and annual mean (black crosses) ΔT_m; and (c) observed (grey) and estimated (green) ΔT_m with shaded 95% 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). [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 9 (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 Figures 3-6, 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). [Colour figure can be viewed at wileyonlinelibrary.com]

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

As can be seen in Figure 10c, 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. There is slightly larger variation between observed and estimated ΔT_m at a monthly resolution (Figure 10b), likely for the reasons noted above, however, the greater variance in the observations is accounted for by the confidence intervals (Figure 10a,c). The primarily high, positive, skill scores (Table 3) suggest applying the model would be beneficial for reducing the presence of the exposure bias from closed exposures in most cases.

5 | QUANTIFYING THE EXPOSURE BIAS IN THE EXTENDED CRUTEM5 STATION DATABASE (CRUTEM5_ext)

Three of the models outlined above: the open, wall-mounted and closed models, produce skilful estimations of the exposure bias and are considered robust enough for broader application. Here the application of the models to stations in CRUTEM5_ext is outlined and the results discussed.

5.1 | Exposure metadata

To apply the bias-estimation models to CRUTEM5_ext, an important new database of historic exposures was compiled for many of the stations in CRUTEM5_ext. Metadata were gathered detailing: (a) if/when a Stevenson screen was introduced, (b) which (if any) historic
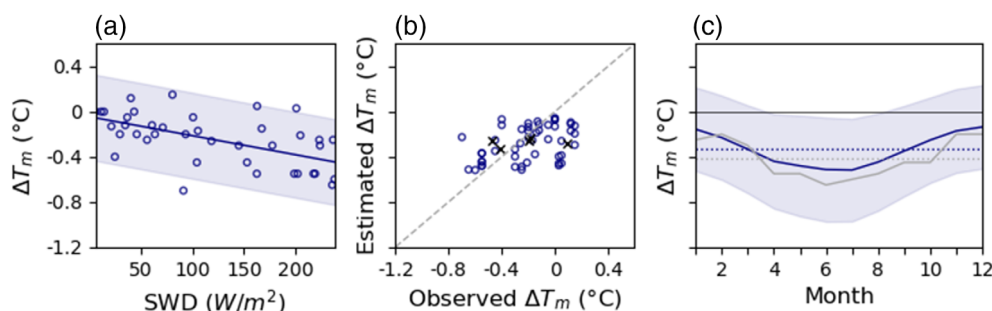
FIGURE 10 (a) Relationship between monthly ΔT_m and shortwave downward received solar radiation with shaded 95% confidence interval for closed exposures; (b) observed versus estimated monthly (navy circles) and annual mean (black crosses) ΔT_m; and (c) observed (grey) and estimated (navy) ΔT_m with shaded 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). [Colour figure can be viewed at wileyonlinelibrary.com]

exposures were in use prior to the introduction, and (c) whether a series had previously been adjusted for the exposure bias. This database was used to identify the stations and time periods affected by the exposure bias as well as the appropriate model(s) to apply. Note that some of the historic exposures identified in the metadata— including previously common (very early) exposures such as hanging the thermometer in a poleward-facing well-ventilated room (Jurin, 1723)—do not fall into any of the categories defined here. In such instances, the exposure was categorized as miscellaneous in the database and no bias-estimation model was applied.

Figure S1 and Video S1 illustrate the metadata gathered; further details of the metadata collation process and results are in Appendix S1.

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 4, and the station absolute 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 ($n \geq 10$) and an estimate of T_a used instead.

The models were only applied to stations within $30°$ to $60°$ latitude, for the reasons outlined in Section 4, and to stations which were not known to have been adjusted for the exposure bias previously. Although Section 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.

As the 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 their calibrated ranges yields implausible results (i.e., overextrapolation). A comparison of the predictors and estimated biases for CRUTEM5_ext stations with the parallel measurement observations and predictors used to develop the models found no evidence of overextrapolation for wall-mounted and closed exposures, but some evidence for open exposures. For open exposures, the range of CRUTEM5_ext 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 $0°C$, respectively, which are not generally observed. Note that these constraints affected only 36 months of data for one station in CRUTEM5_ext 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; see Appendix S2), and the resulting bias estimates were combined with the CRUTEM5_ext data to produce an exposure bias adjusted version, referred to here as: CRUTEM5_eba.

5.3 | An exposure bias adjustment for the CRUTEM5 station database

The metadata identified 2,519 mid-latitude stations in CRUTEM5_ext 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 (Table 4). 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 4; Figure 11). Despite this, the metadata gathered, and the adjustments applied, indicate that the bias has now been accounted for at 75.1% of mid-latitude stations, representing 86.3% of the mid-latitudinal data in terms of monthly values. This is compared with just 37.7% of mid-latitude stations in CRUTEM5_ext (of which only 1.5% were known to have been adjusted).

At the hemispheric scale the impact of the bias adjustments is relatively small (Figure 12). In the Northern Hemisphere (NH) the mid-latitude annual mean is $\leq 0.016°C$ 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 ($\approx 0.05°C$) and summer/winter ($\approx -0.03°$); 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 biannual seasonal cycle (Section 3.2). The change in the direction and seasonal structure of the bias after 1870 (and the increase in

TABLE 4 Number of stations and months which have been adjusted for the exposure bias, as well as the numbers which still require adjustment.

Input: CRUTEM5_ext			**Output: CRUTEM5_eba**
All stations Mid-latitude stations in CRUTEM5_ext (located 30°–60° North or South of the equator) with data prior to 1961 **Stations**: 5,031 **Months**: 5,708,463	**Exposure bias not present** **Stations**: 1,898 (37.7%)	**Series flagged as already exposure bias adjusted** **Stations**: 75 (1.5%) **Months**: 159,523 (2.8%) **Metadata suggests 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%)	**Exposure bias adjusted** **Stations**: 3,776 (75.1% of stations) **Months**: 4,923,593 (86.3% of data)
	Exposure bias present Metadata suggests a Stevenson screen was introduced part-way through a station record **Stations**: 2,519 (50.1%)	**Adjusted**: Stations which have been adjusted for the exposure bias **Stations**: 1878 (37.3%) **Months**: 2,836,051 (49.7%) (*Adjusted: 486,148; No adjustment required: 2,349,903*)	
		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*) **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*)	**Reason for missing bias estimates/ adjustments:** **Missing exposure metadata** **Stations**: 1,209 (24%)[a] **Months**: 769,453 (13.5%) **No bias-estimation model** **Stations**: 44 (0.9%)[a] **Months**: 13,389 (0.2%) **Missing predictor(s)** **Stations**: 10 (0.2%)[a] **Months**: 2,028 (0.04%)	**May require further exposure bias adjustment** **Stations**: 1,255 (24.9% of stations) **Months**: 784,870 (13.7% of data)
	No exposure metadata Series without any metadata entries in the database **Stations**: 614 (12.2%)		

[a]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).
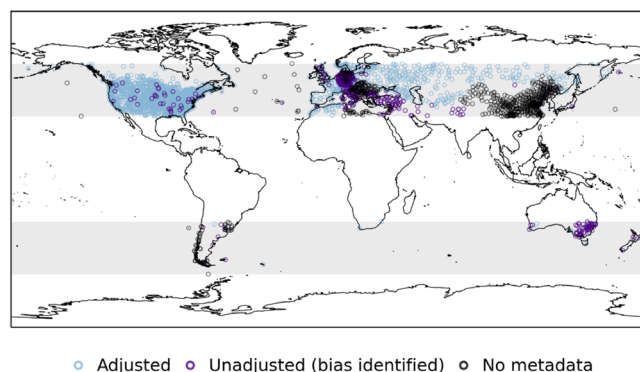
FIGURE 11 Location of stations which have been adjusted for the exposure bias, have not been adjusted but contain probable biases (for reasons given in Table 4), or do not have metadata. Stations with metadata, but which were not adjusted, are not shown. [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 12 Difference between CRUTEM5_ext and CRUTEM5_eba annual and seasonal means (ΔT = CRUTEM5_eba − CRUTEM5_ext) for the Northern and Southern Hemisphere mid-latitudes (coloured lines). The light blue line shows the number of stations with a non-zero bias estimation (adjustment). The dark grey shading represents the range of annual mean exposure bias adjustments present in each CRUTEM5_eba grid cell. The light grey shading with dashed black outline represents the approximate range of 100 realizations of the exposure bias component in the HadCRUT5 error model for comparison. [Colour figure can be viewed at wileyonlinelibrary.com]

magnitude) occurs as a result of more series requiring correction for freestanding exposures, which produce larger biases with a single peak in the seasonal cycle (Sections 3.1 and 3.4). The largest adjustments, however, are relatively geographically constrained (Figure 13) meaning the overall effect on the mid-latitude mean remains small.

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 months) 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. 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-latitude means and exhibits significant spatiotemporal variability (Figure 13). Between 1851 and 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-
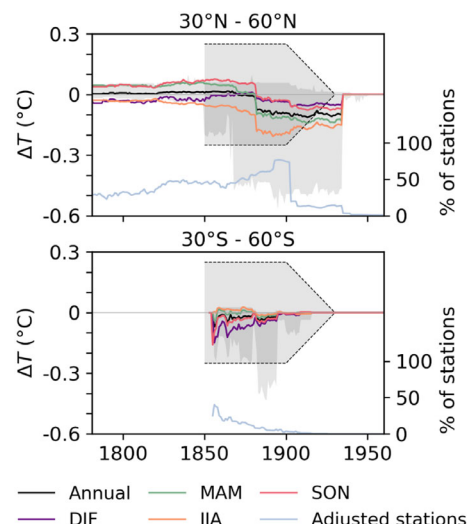
temporal heterogeneity of the historic exposures in use (Video S1), as well as the influence of solar radiation/ temperature. The large negative adjustments in Mediterranean Africa and central Asia, for example, reflect the use of freestanding exposures in those regions combined with the influence of 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 mid-latitude 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 12) shows reasonable agreement annually over large spatial scales, but regionally the comparisons reinforce the limitations of the current representation identified in Section 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 spatiotemporal differences in the use of historic exposures and
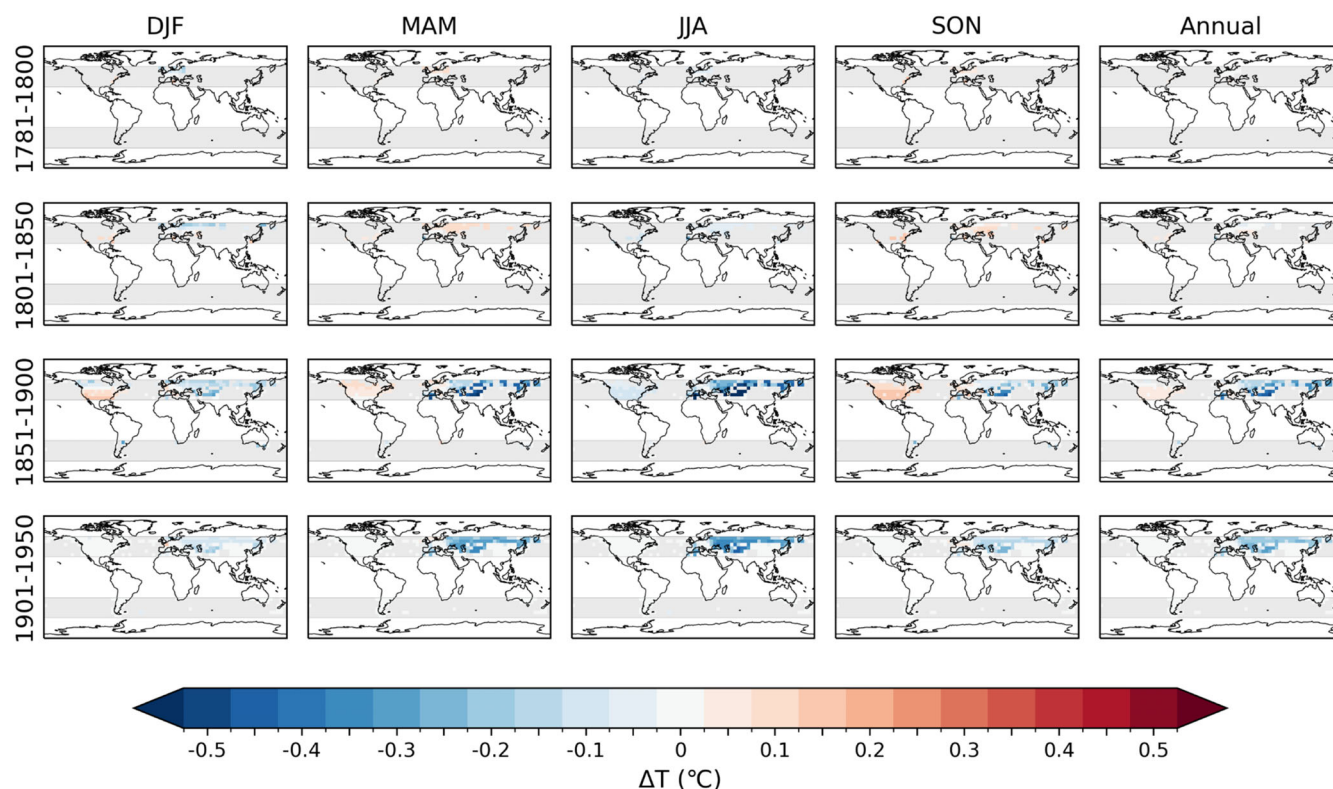
FIGURE 13 Difference between CRUTEM5_ext and CRUTEM5_eba seasonal and annual means over time (ΔT = CRUTEM5_eba − CRUTEM5_ext). [Colour figure can be viewed at wileyonlinelibrary.com]

timing of the introduction of the Stevenson screen. These limitations are highlighted by the large discrepancies between the HadCRUT5 realizations and the grid cell adjustments (particularly between 1868 and 1934 in the NH and the late-1800s in the SH) and by the static nature of the realizations 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 incorporated corrections for other inhomogeneities (e.g., Ashcroft et al., 2012), did not adjust monthly T_m (e.g., Ashcroft et al., 2022; Auchmann & Brönnimann, 2012) and/or are based on the parallel measurements used here (e.g., Brunet et al., 2006). Comparisons with the few independent assessments available produce mixed results. Our estimates of a −0.44°C summer bias in Uppsala (which was not applied because this series in CRUTEM5 had already been adjusted), between 1858 and 1864, and ≈−0.2°C summer bias in the NH mid-latitudes, between 1880 and 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) 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 variables. As such, the bias-estimation models cannot take into account station-specific factors or differences within categories of exposure (although these are partially reflected in the confidence intervals). 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 (Appendix S1). 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 4) and the favourable comparison of our regional results 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-19th century, between 0°C 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 and 1910, in line with the estimates here.

6 | SUMMARY AND CONCLUSIONS

This study has advanced our understanding of the characteristics of the exposure bias introduced by the transition to Stevenson-type screens and has presented an empirical approach to address the exposure bias in an extended version of CRUTEM5, using regression-based bias-estimation models and exposure metadata. To the best of our knowledge, this is the first time such an approach has been used to produce a near-global assessment of the exposure bias (although similar approaches have been used locally, e.g., Brunet et al., 2006, 2011).

The compilation and analysis of 54 parallel measurement series comparing the temperatures recorded in Stevenson-type screens with those recorded in four categories of historic exposure identified seasonally and diurnally varying biases which differed according to the class of exposure. The largest biases in mean temperatures were generally found in freestanding exposures (up to −0.78°C annually) and in the summer months, while the smallest biases were generally found in the winter months and in wall-mounted exposures (near 0°C annually, though the biases arising from wall-mounted exposures showed large inter-site variability). These results are generally consistent with those of Parker (1994), but the inclusion of additional parallel measurement series here highlighted possible regional differences in the magnitude of the exposure bias and confirmed a bi-annual cycle is a common (and distinct) characteristic of the exposure bias arising from wall-mounted exposures. The results also reinforced the need for a monthly-resolved and exposure-specific assessment of the exposure bias in global datasets.

The identification of significant relationships between the magnitude of the bias in monthly mean temperatures and received shortwave downward solar radiation, downward top of atmosphere solar radiation and/or absolute temperature in the historic exposure for each category of exposure facilitated the development of three exposure-specific, bias-estimation models. The models were found to skilfully reproduce the direction and seasonal cycle of the exposure bias, but with reasonably large confidence intervals (particularly for wall-mounted exposures) which reflect the influence of site-specific factors on the magnitude of

the bias. The development of these models builds on the results of previous studies which identified relationships between the magnitude of the exposure bias and absolute temperature and/or solar radiation (e.g., Ashcroft et al., 2022; Auchmann and Brönnimann, 2012; Brunet et al., 2011; Margary, 1924), but is, to our knowledge, the first-time these relationships have been quantified for multiple classes of exposure using a compilation of parallel measurement series from across the mid-latitudes.

The bias adjustments produced using the bias-estimation models and exposure metadata showed reasonable agreement with the representation of the bias in the HadCRUT5 error model annually and over large spatial scales. However, the estimates produced here refine the representation of the bias seasonally and at regional scales by taking into account the annual cycle of the bias and the different exposure history of individual stations and regions. This refinement is reflected in the spatio-temporal variability evident in our bias adjustments compared to the static nature of the bias represented in the HadCRUT5 error model. This improved representation is important to allow a more complete understanding of the impact of exposure biases on global temperatures, particularly the affect they have on regional and seasonal climate trends.

The presented approach is not without limitations. The development of the bias-estimation models is limited by the relatively small number of parallel measurements available to analyse and the accuracy of the individual bias estimates is contingent on the accuracy of the exposure metadata used to inform the model application. Nevertheless, the approach represents a first step toward the more accurate assessment of the (monthly-resolved) impact of exposure biases in mid-latitude weather stations and it is hoped the approach can be continually updated and improved in future, possibly via the inclusion of additional exposure metadata and parallel measurement series. The latter, in particular, would help to refine the relationships between the bias and the relevant predictor(s) and constrain (or more accurately capture) the uncertainty associated with the bias estimates.

A number of avenues for further work exist. The approach outlined above specifically addresses the exposure biases arising from three classes of historic exposure in the mid-latitudes. It does not address the exposure biases present in the tropics and high latitudes or the biases arising from other categories of exposure (i.e., intermediate exposures; very early indoor exposures) or changes to the design of the Stevenson screens in use over time (see Mawley, 1884; Naylor, 2019). At present the availability of parallel measurements prevents the extension of the presented approach to address these areas, however it is hoped additional parallel

measurements may become available in future. Further work is therefore required to address these areas and to more fully account for the exposure biases present in global temperature compilations. Future work could also combine the exposure bias adjustments with the (post-adjustment) application of pair-wise station homogenisation algorithms to assess the broader homogeneity of the CRUTEM5_ext station data.

AUTHOR CONTRIBUTIONS

Emily J. Wallis: Conceptualization; methodology; software; data curation; formal analysis; investigation; writing – original draft; visualization; writing – review and editing; project administration. **Timothy J. Osborn:** Conceptualization; methodology; data curation; writing – review and editing; funding acquisition; supervision; project administration. **Michael Taylor:** Methodology; software; data curation; writing – review and editing; supervision; conceptualization. **Philip D. Jones:** Methodology; writing – review and editing; supervision; conceptualization. **Manoj Joshi:** Methodology; writing – review and editing; supervision; conceptualization. **Ed Hawkins:** Conceptualization; writing – review and editing; supervision; funding acquisition.

ACKNOWLEDGEMENTS

DATA AVAILABILITY STATEMENT

All of the datasets used in this study are freely accessible. The CRUTEM5 station database is available via: https://crudata.uea.ac.uk/cru/data/temperature and the compilation of parallel measurements and the bias estimates produced for CRUTEM5_ext (including the relevant exposure metadata) are available via Zenodo: 10.5281/zenodo.10551236 and 10.5281/zenodo.10551197, respectively.

ORCID

Emily J. Wallis https://orcid.org/0000-0002-2475-0899
Timothy J. Osborn https://orcid.org/0000-0001-8425-6799
Michael Taylor https://orcid.org/0000-0002-3473-3478
Philip D. Jones https://orcid.org/0000-0001-5032-5493
Manoj Joshi https://orcid.org/0000-0002-2948-2811
Ed Hawkins https://orcid.org/0000-0001-9477-3677

REFERENCES

Allan, R., Brohan, P., Compo, G., Stone, R., Luterbacher, J. & Brönnimann, S. (2011) The international atmospheric circulation reconstructions over the earth (ACRE) initiative. *Bulletin of the American Meteorological Society*, 92(11), 1421–1425. Available from: https://doi.org/10.1175/2011BAMS3218.1

Anchukaitis, K.J., Wilson, R., Briffa, K.R., Büntgen, U., Cook, E.R., D'Arrigo, R. et al. (2017) Last millennium Northern Hemisphere summer temperatures from tree rings: part II, spatially resolved reconstructions. *Quaternary Science Reviews*, 163, 1–22. Available from: https://doi.org/10.1016/j.quascirev.2017.02.020

Ashcroft, L., Karoly, D.J. & Gergis, J. (2012) Temperature variations of southeastern Australia, 1860-2011. *Australian Meteorological and Oceanographic Journal*, 62(2012), 227–245 Available from: http://www.bom.gov.au/jshess/docs/2012/ashcroft.pdf

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. Available from: https://doi.org/10.1002/joc.7385

Auchmann, R. & Brönnimann, S. (2012) A physics-based correction model for homogenizing sub-daily temperature series. *Journal of Geophysical Research Atmospheres*, 117(17), D17119. Available from: https://doi.org/10.1029/2012JD018067

Awe, S.O., Mahoney, M., Michaud, E., Murphy, C., Noone, S.J., Venema, V.K.C. et al. (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. Available from: https://doi.org/10.5194/CP-18-793-2022

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

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. Available from: https://doi.org/10.1007/s10584-009-9649-4

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

Brunet, M., Jesús, A., Sigró, J., Bañón, M.G., Aguilar, E. et al. (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. Available from: https://doi.org/10.1002/joc.2192

Butler, C.J., García-Suárez, 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. Available from: http://www.climate.armagh.ac.uk/calibrated/airtemp/yellow-textD.pdf

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

Chenoweth, M. (1992) A possible discontinuity in the U.S. historical temperature record. *Journal of Climate*, 5(10), 1172–1179. Available from: https://doi.org/10.1175/1520-0442(1992)005<1172:APDITU>2.0.CO;2

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

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*. London: His Majesty's Stationary Office, pp. 30–40 Available at: https://digital.nmla.metoffice.gov.uk/file/io_8b9a812f-db8d-4cbb-9412-fddd2cc9a149

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. Available from: https://doi.org/10.1002/qj.4970178006

Field, J.H. (1920) On exposures of thermometers in India. *India Meteorological Memoirs*, XIV, 21–73.

Folland, C.K., Rayner, N.A., Brown, S.J., Smith, T.M., Shen, S.S.P., Parker, D.E. et al. (2001) Global temperature change and its uncertainties since 1861. *Geophysical Research Letters*, 28(13), 2621–2624. Available from: https://doi.org/10.1029/2001GL012877

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. Available from: https://doi.org/10.1016/j.quascirev.2007.08.002

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.

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. Available from: https://doi.org/10.1002/qj.4970084404

Gorczynski, W. (1910) Klatka angielska nowego typu w porownaniu z ochronami termetrycznemi innych systemow. *Wiadomosci Matematyczne*, 14, 205–236 Available at: https://crispa.uw.edu.pl/object/files/322226/display/Default

Grissolet, M.H. (1935) Comparaison entre les temperatures observees dans l'abri anglais O.N.M petit modele et sous l'abri Montsouris. In: *Annales des services techniques d'hygeine de la ville de Paris, XVII*, Paris: Gauthier-Villars et Cie, pp. 256–267 Available at: https://gallica.bnf.fr/ark:/12148/bpt6k11777428/f276.item

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

Jurin, J. (1723) VII. Invitatio ad observationes meteorologicas communi consilio instituendas. *Philosophical Transactions of the Royal Society of London*, 32(379), 422–427. Available from: https://doi.org/10.1098/RSTL.1722.0082

Königlich Preussischen Meteorologischen Institut. (1892) *Deutsches Meteorologisches Jarbuch fur 1889. Ergebnisse der Meteorologischen Beobachtungen im Jahre 1889*. Berlin: A. Asher & Co. Available at: https://hdl.handle.net/2027/osu.32435067383471

Koppen, W. (1913) Uniform thermometer set-up for meteorological stations for the determination of air temperature and humidity. *Meteorologische Zeitschrift*, 30, 474–523.

Lenssen, N.J.L., Schmidt, G.A., Hansen, J.E., Menne, M.J., Persin, A., Ruedy, R. et al. (2019) Improvements in the GISTEMP uncertainty model. *Journal of Geophysical Research: Atmospheres*, 124(12), 6307–6326. Available from: https://doi.org/10.1029/2018JD029522

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. Available from: 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. Available from: https://doi.org/10.1002/qj.4970053204

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. Available from: https://doi.org/10.1002/qj.4970209005

Martinez Ibarra, E., Azorín Molina, C., Bañón García, M., Olcina Cantos, J., Estrela Navarro, M.J. et al. (2010) 'Intercomparacion de las temperaturas extremas en tres tipos de garita meteorologica: Montsouris, Stevenson y Young', in VII Congress of the Spanish Association of Climatology: climate, city and ecosystems. Madrid: Spanish Association of Climatology. Available from: http://hdl.handle.net/20.500.11765/8516

Mawley, E. (1884) Report on temperatures in two different patterns of stevenson screens. *Quarterly Journal of the Royal Meteorological Society*, 10(49), 1–7. Available from: https://doi.org/10.1002/qj.4970104902

Mawley, E. (1897) Shade temperature. *Quarterly Journal of the Royal Meteorological Society*, 23(102), 69–87. Available from: https://doi.org/10.1002/qj.49702310202

Menne, M.J., Williams, C.N., Gleason, B.E., Rennie, J.J. & Lawrimore, J.H. (2018) The global historical climatology network monthly temperature dataset, version 4. *Journal of Climate*, 31(24), 9835–9854. Available from: https://doi.org/10.1175/JCLI-D-18-0094.1

Meteorological Observations made at the Adelaide Observatory, and other places in South Australia and the Northern Territory, during the year 1888. (1890) Adelaide. Available at: 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) Adelaide. Available at: 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) Adelaide. Available at: 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) Adelaide. Available at: 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) Adelaide. Available at: 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) Adelaide. Available at: 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) Adelaide. Available at: 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) Adelaide. Available at: 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) Adelaide. Available at: 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) Adelaide. Available at: 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) Adelaide. Available at: 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) Adelaide. Available at: 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) Adelaide. Available at: https://hdl.handle.net/2027/uc1.c2999508

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. Available from: https://doi.org/10.1002/joc.945

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) Available from: https://doi.org/10.1029/2011JD017187

Morice, C.P., Kennedy, J.J., Rayner, N.A., Winn, J.P., Hogan, E., Killick, R.E. et al. (2021) An updated assessment of near-surface temperature change from 1850: the HadCRUT5 dataset. *Journal of Geophysical Research: Atmospheres*, 126(3), e2019JD032361. Available from: https://doi.org/10.1029/2019JD032361

Muller, G. (1984) Vergleich der Temperaturen verschiedener Wetterhutten an einigen Stationen des Anetzes. In: *Arbeitsberichte der Schweizerischen Meteorologischen Zentralanstalt, (No. 119)*, Zurich: Schweizerischen Meteorologischen Zentralanstalt, pp. 1–36.

Naylor, S. (2019) Thermometer screens and the geographies of uniformity in nineteenth-century meteorology. *Notes and Records*, 73, 203–221. Available from: https://doi.org/10.1098/rsnr.2018.0037

Nicholls, N., Tapp, R., Burrows, K. & Richards, D. (1996) Historical thermometer exposures in Australia. *International Journal of Climatology*, 16(6), 705–710. Available from: https://doi.org/10.1002/(SICI)1097-0088(199606)16:6<705::AID-JOC30>3.0.CO;2-S

Nordli, P.Ø., Alexandersson, H., Frisch, P., Førland, E.J., Heino, R., Jónsson, T. et al. (1996) *The effect of radiation screens on Nordic temperature measurements*. Oslo: Det Norske Meteorologiske Institutt. Available at: https://www.met.no/publikasjoner/met-report/met-report-1996

Nordli, P.Ø., Alexandersson, H., Frisch, P., Førland, E.J., Heino, R., Jónsson, T. et al. (1997) The effect of radiation screens on Nordic time series of mean temperature. *International Journal of Climatology*, 17(15), 1667–1681. Available from: https://doi.org/10.1002/(SICI)1097-0088(199712)17:15<1667::AID-JOC221>3.0.CO;2-D

Omond, R.T. (1906) Temperatures in thermograph and Stevenson screens. *Journal of the Scottish Meteorological Society*, XIV (XXIII), 15–20 Available at: https://archive.org/details/journalscottish00socigoog/page/15/mode/1up

Osborn, T.J., Jones, P.D., Lister, D.H., Morice, C.P., Simpson, I.R., Winn, J.P. et al. (2021) Land surface air temperature variations across the globe updated to 2019: the CRUTEM5 data set. *Journal of Geophysical Research: Atmospheres*, 126(2), e2019JD032352. Available from: https://doi.org/10.1029/2019JD032352

PAGES2k Consortium. (2017) A global multiproxy database for temperature reconstructions of the Common Era. *Scientific Data*, 4(170088), 1–33. Available from: https://doi.org/10.1038/sdata.2017.88

Parker, D.E. (1994) Effects of changing exposure of thermometers at land stations. *International Journal of Climatology*, 14(1), 1–31. Available from: https://doi.org/10.1002/joc.3370140102

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1887. (1889) London. Available at: http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/YBpdf/YB_GRW_1887.pdf

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1888. (1890) London. Available at: http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/YBpdf/YB_GRW_1888.pdf

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1889. (1891) London. Available at: 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) London. Available at: 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) London. Available at: 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) London. Available at: 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) London. Available at: 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) Edinburgh. Available at: 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) Edinburgh. Available at: 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) Edinburgh. Available at: 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) Edinburgh. Available at: 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) Edinburgh. Available at: 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) Edinburgh. Available at: http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/YBpdf/YB_GRW_1899.pdf

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1900. (1902) Edinburgh. Available at: 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) Edinburgh. Available at: http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/YBpdf/YB_GRW_1901.pdf

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1902. (1904) Edinburgh. Available at: http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/YBpdf/YB_GRW_1902.pdf

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1903. (1904) Edinburgh. Available at: http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/YBpdf/YB_GRW_1903.pdf

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1904. (1905) Edinburgh. Available at: http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/YBpdf/YB_GRW_1904.pdf

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1905. (1906) Edinburgh. Available at: http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/YBpdf/YB_GRW_1905.pdf

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1906. (1907) Edinburgh. Available at: http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/YBpdf/YB_GRW_1906.pdf

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1907. (1909) Edinburgh. Available at: http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/YBpdf/YB_GRW_1907.pdf

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1908. (1909) Edinburgh. Available at: http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/YBpdf/YB_GRW_1908.pdf

Results of the Magnetical and Meteorological Observations made at The Royal Observatory, Greenwich, in the year 1909. (1911) Edinburgh. Available at: http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/YBpdf/YB_GRW_1909.pdf

Sen Roy, S., Saha, S.B., 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. Available from: https://doi.org/10.1002/joc.4108

Sparks, W.R. (1972) *The effect of thermometer screen design on the observed temperature*. Geneva: World Meteorological Organization. Available at: https://library.wmo.int/idurl/4/29937

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 Available at: https://hdl.handle.net/2027/nyp.33433069059404

Thorne, P.W., Menne, M.J., Williams, C.N., Rennie, J.J., Lawrimore, J.H., Vose, R.S. et al. (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. Available from: https://doi.org/10.1002/2015JD024583

Trewin, B. (2010) Exposure, instrumentation, and observing practice effects on land temperature measurements. *Wiley Interdisciplinary Reviews: Climate Change*, 1(4), 490–506. Available from: https://doi.org/10.1002/wcc.46

Veðráttan. (1962) *Arsyfirlit samid a vedurstofunni*. Reykjavik: Veðráttan, p. 122 Available at: https://timarit.is/page/3128806

Venema, V.K.C., Mestre, O., Aguilar, E., Auer, I., Guijarro, J.A., Domonkos, P. et al. (2012) Benchmarking homogenization algorithms for monthly data. *Climate of the Past*, 8(1), 89–115. Available from: https://doi.org/10.5194/cp-8-89-2012

Vigurs, C.C. (1935) Temperature at Falmouth. *The Meteorological Magazine*, 70(832), 91–92 Available at: https://archive.org/details/sim_meteorological-magazine_1935-05_70_832/page/91/mode/1up

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 for 1880*, 13–18.

Wild, H. (1887) Further tests on the determination of the true air temperature. *Repertorium fur Meteorologie*, X(10), 1–24. Available from: https://library.oarcloud.noaa.gov/foreign_climate.lib/FCD_002_pdf/QC989R8M48t10.pdf

Wild, H. (1891) Über den einfluss der aufstellung auf die angaben der thermometer zur bestimmung der lufttemperatur. *Repertorium fur Meteorologie*, XIV(9), 1-73. Available from: https://library.oarcloud.noaa.gov/foreign_climate.lib/FCD_002_pdf/QC989R8M48t14.pdf

Willett, K., Williams, C., Jolliffe, I.T., Lund, R., Alexander, L.V., Brönnimann, S. et al. (2014) A framework for benchmarking of homogenisation algorithm performance on the global scale. *Geoscientific Instrumentation, Methods and Data Systems*, 3(2), 187–200. Available from: https://doi.org/10.5194/gi-3-187-2014

World Meteorological Organization. (2020) *Guidelines on homogenization*, 2020 edition. Geneva: World Meteorological Organisation. Available at: https://library.wmo.int/idurl/4/57130

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.