Identifying exposure biases in early instrumental data

Emily Wallis, Tim Osborn, Michael Taylor, David Lister, Phil Jones

University of East Anglia

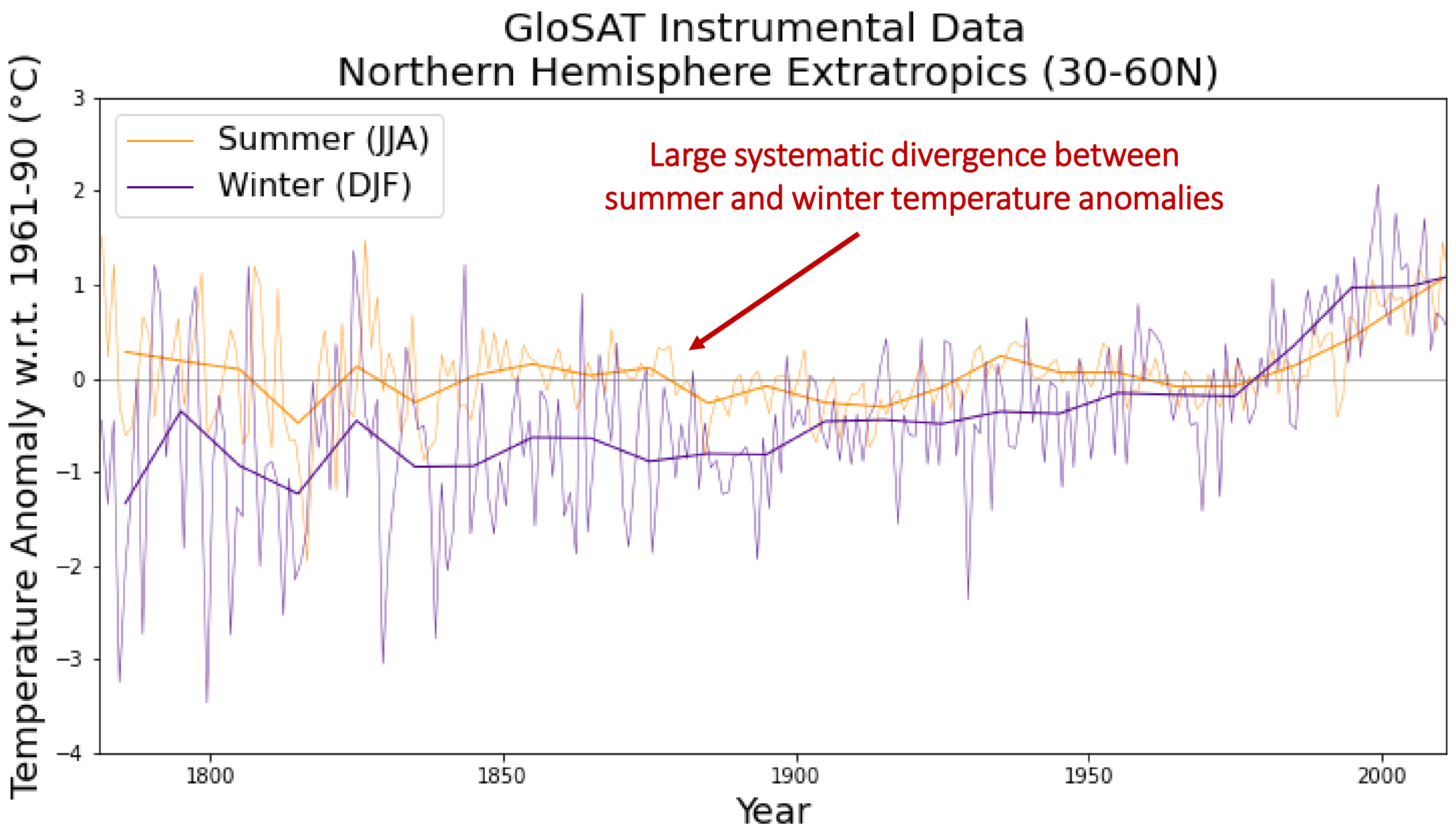
Global Surface Air Temperature

MOTIVATIONS

Why study exposure bias?

Exposure bias contributes significant uncertainty to long instrumental temperature records which are vital to the study of long-term climate variability and change.

Marked seasonal contrast, combined with little summer warming, is evident in the early GloSAT instrumental data – is this real or the product of bias?



EXPOSURE BIAS

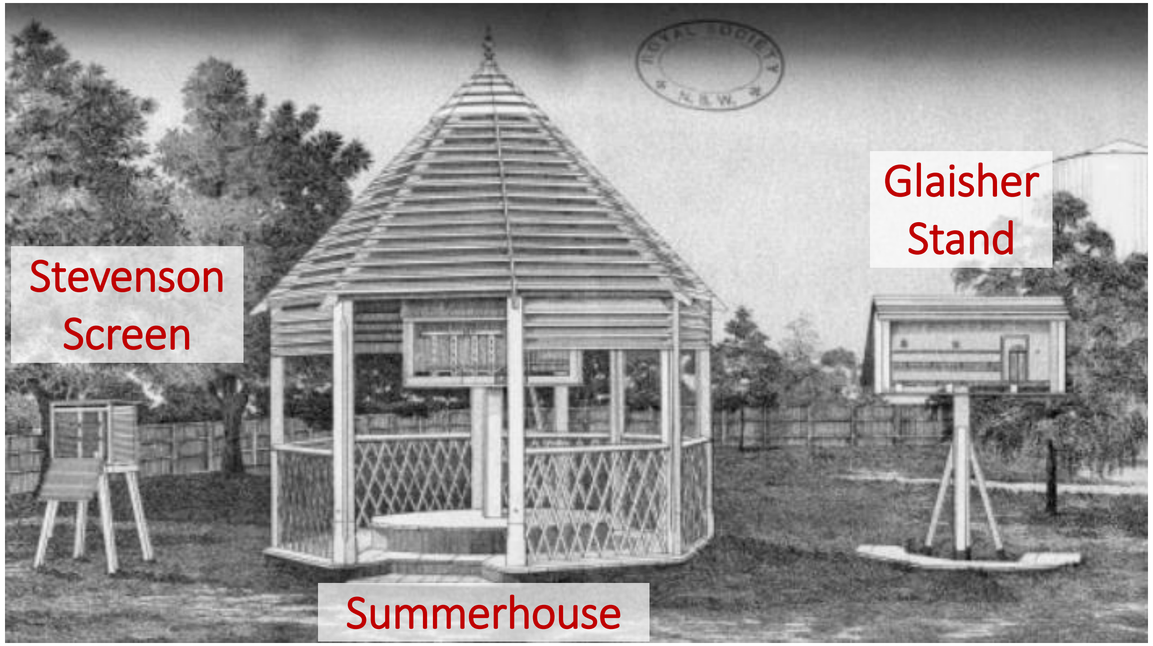
What is the exposure bias?

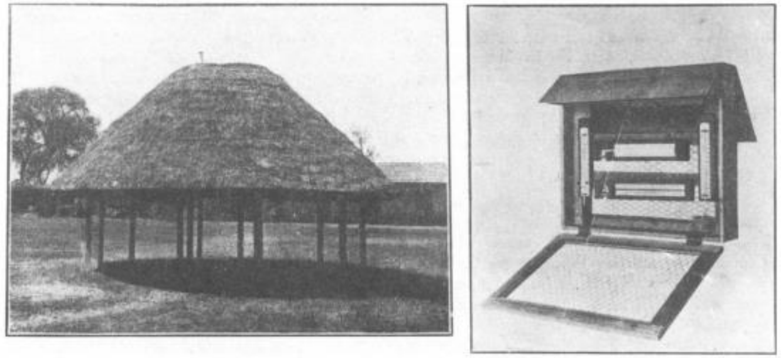
Prior to the widespread adoption of the Stevenson screen in the late-19th/ early-20th century, multiple approaches were taken to protect thermometers from solar radiation.

A few examples are given on the right (the Stevenson screen is pictured in the top-right image).

North Wall/Window exposure

Montsouris (French) Screen





Thatched shelter and 'cage' used in the Tropics

Image source (clockwise): Bohm et al., 2010; Trewin, 2010, Parker, 1994; Brunet et al, 2006.

EXPOSURE BIAS

Each approach to protecting (exposing) thermometers influenced the temperature reading differently. When Stevenson screens were subsequently adopted, this introduced a bias into the station temperature record.

Stations within regions often introduced new screens simultaneously (on unknown dates) making the bias hard to identify and correct using traditional methods (e.g. neighbour comparison).

Ways thermometer screens can influence temperature readings

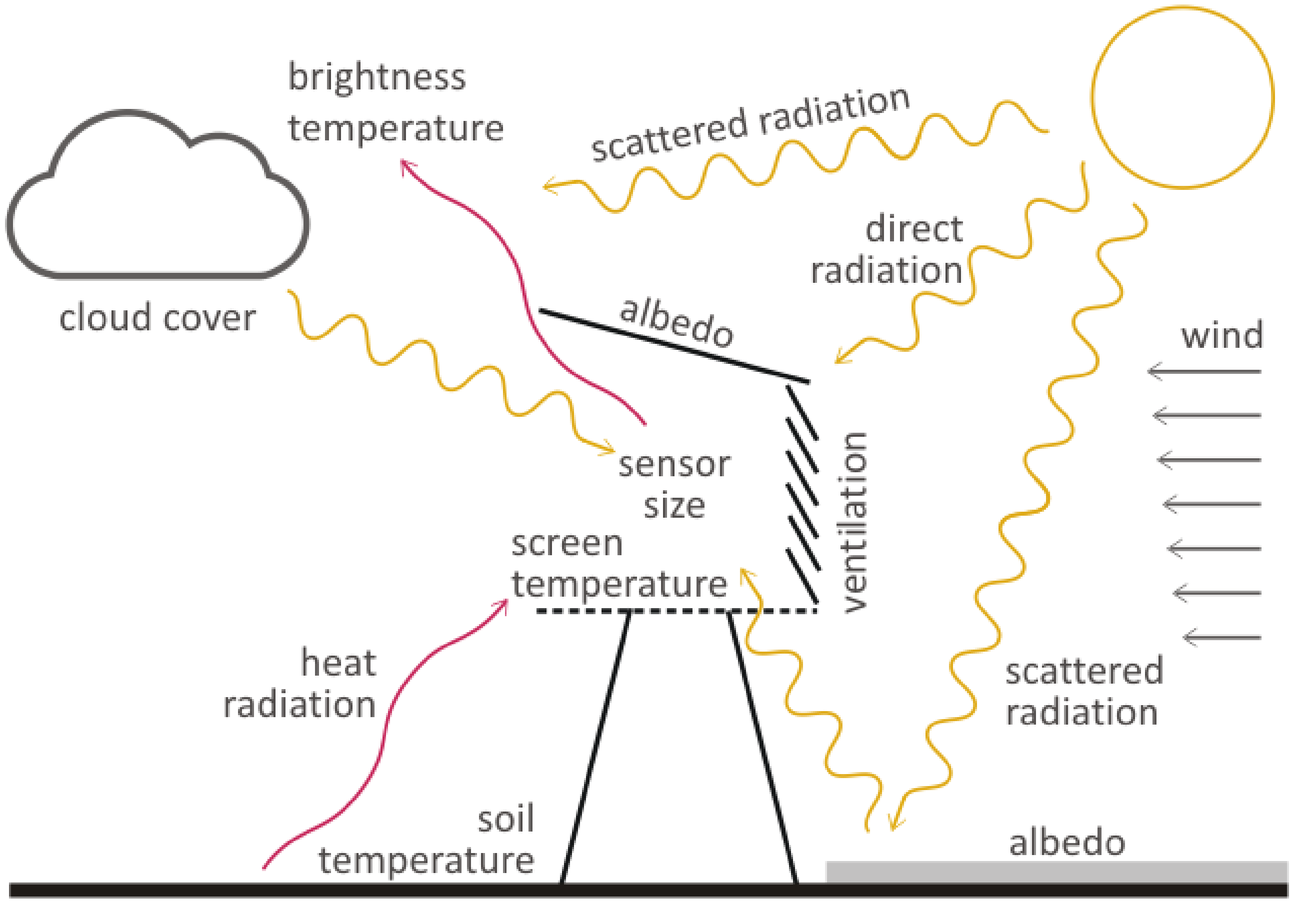


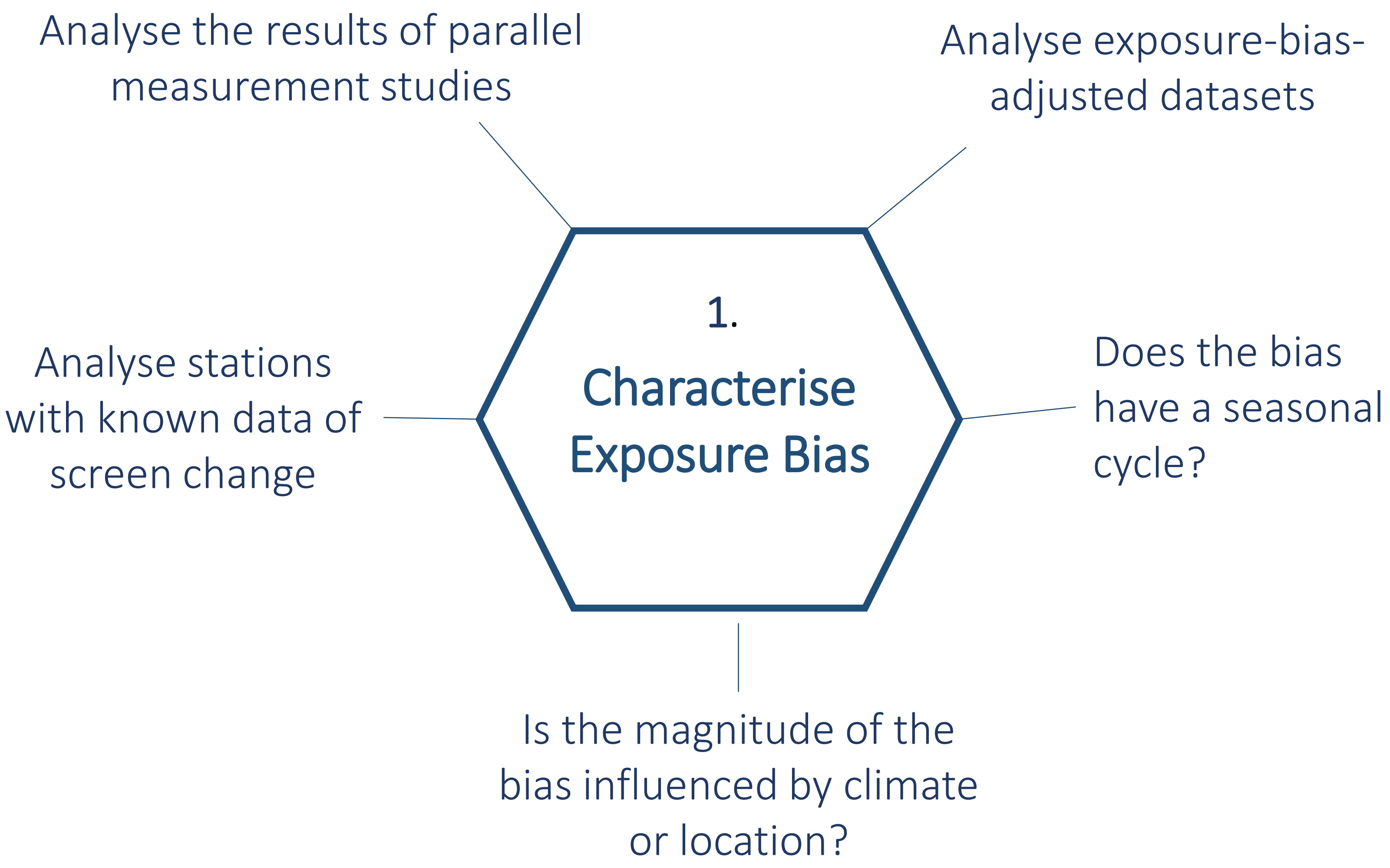
Image: Victor Venema (variable-variability.blogspot.com)

APPROACH

As traditional homogenisation methods are often ineffective for addressing the exposure bias, we are trialling an alternative approach.

Our approach to identifying the bias is to:

1. Better characterise the exposure bias using available data



University of East Anglia

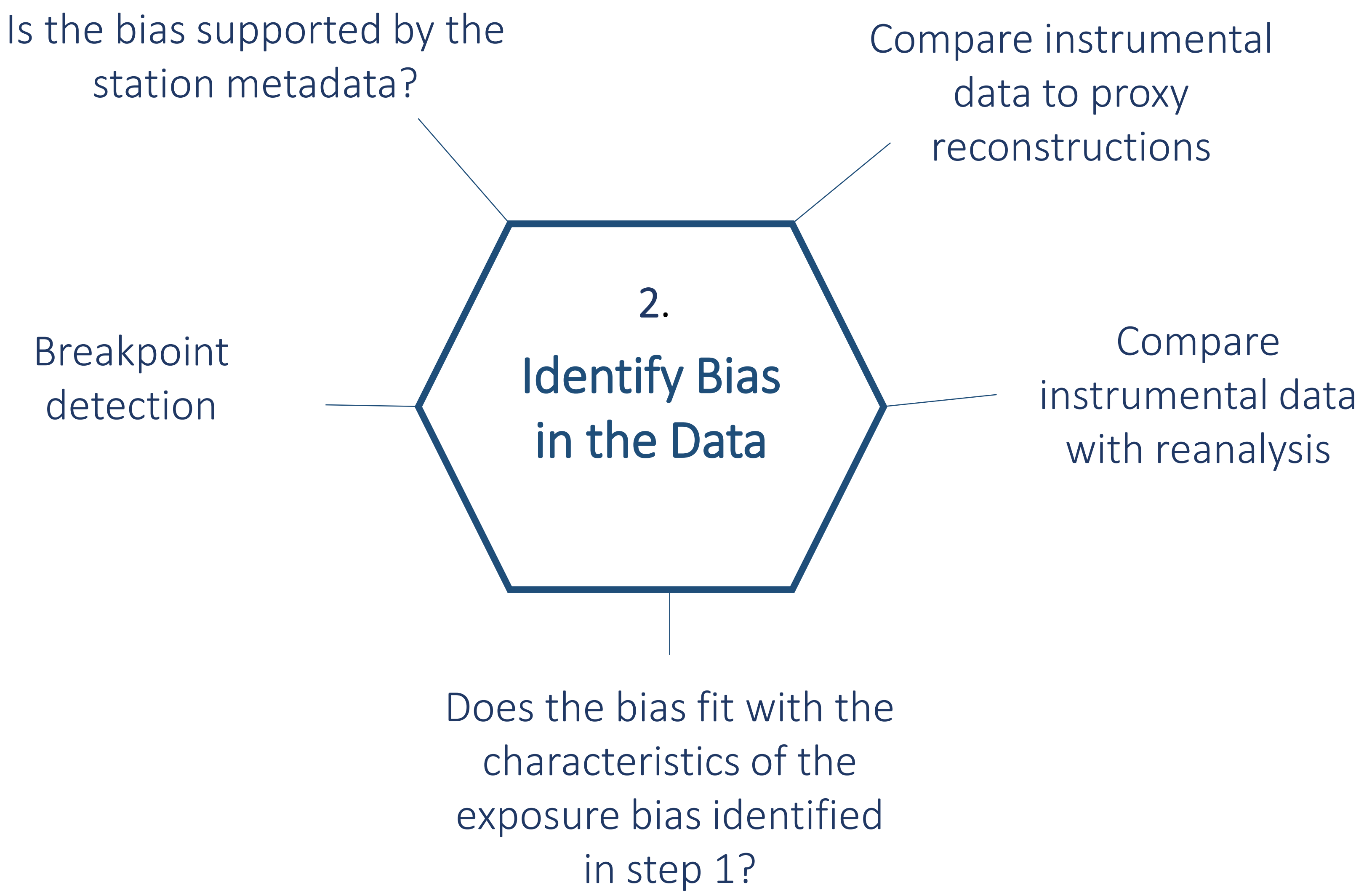
GloSAT Global Surface Air Temperature

APPROACH

Our approach to identifying the bias is to:

1. Better characterise the exposure bias using available data

2. Use the characteristics identified in step 1 to identify possible exposure bias in the instrumental data



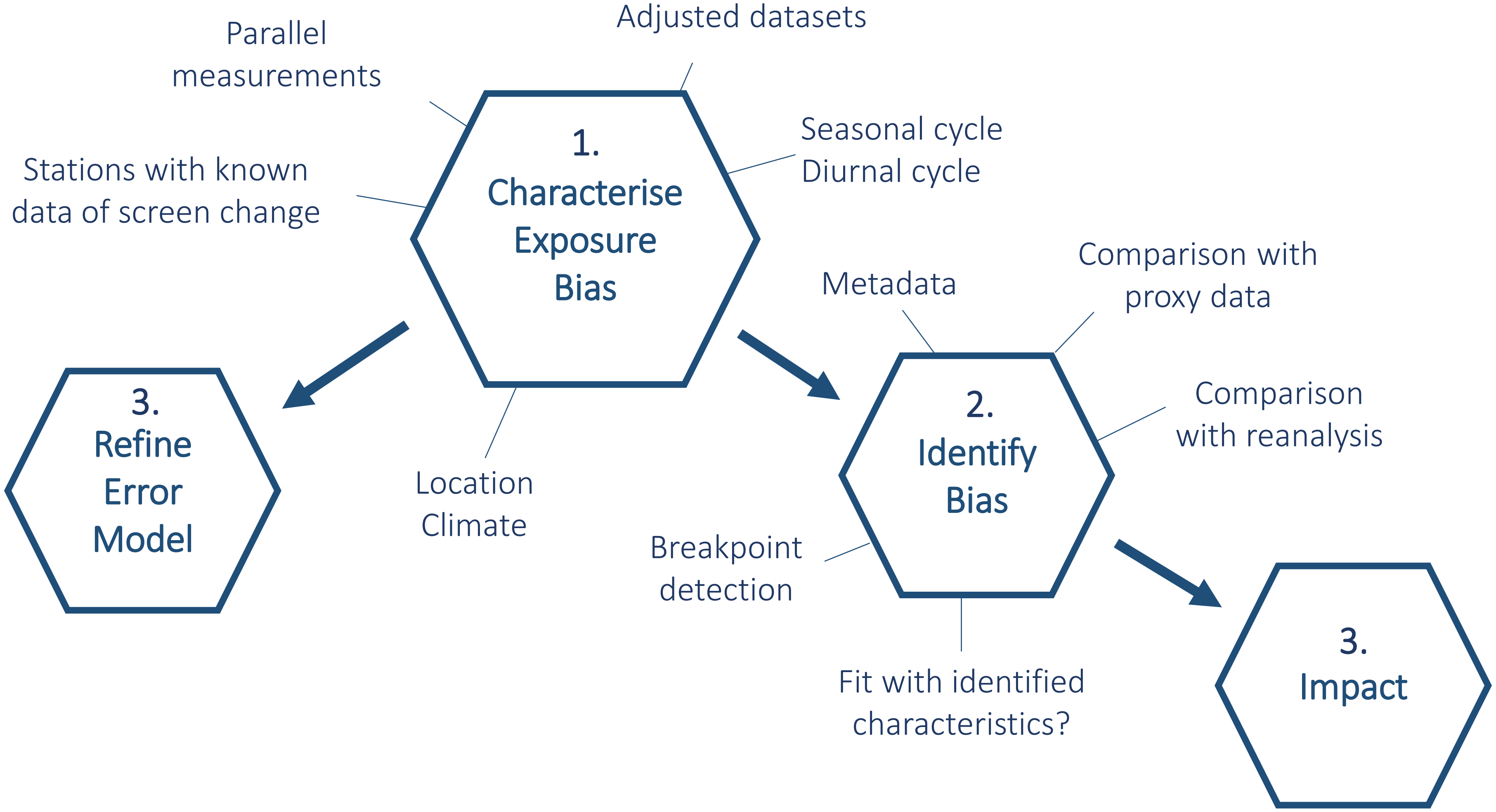
APPROACH

Our approach to identifying the bias is to:

1. Better characterise the exposure bias using available data

2. Use the characteristics identified in step 1 to identify possible exposure bias in the instrumental data

3. Potential to improve the dataset error model? Impact?



University of East Anglia

GloSAT Global Surface Air Temperature

1. CHARACTERISE THE EXPOSURE BIAS

The following slides outline some preliminary results of the work we have been doing in Step 1 of the study

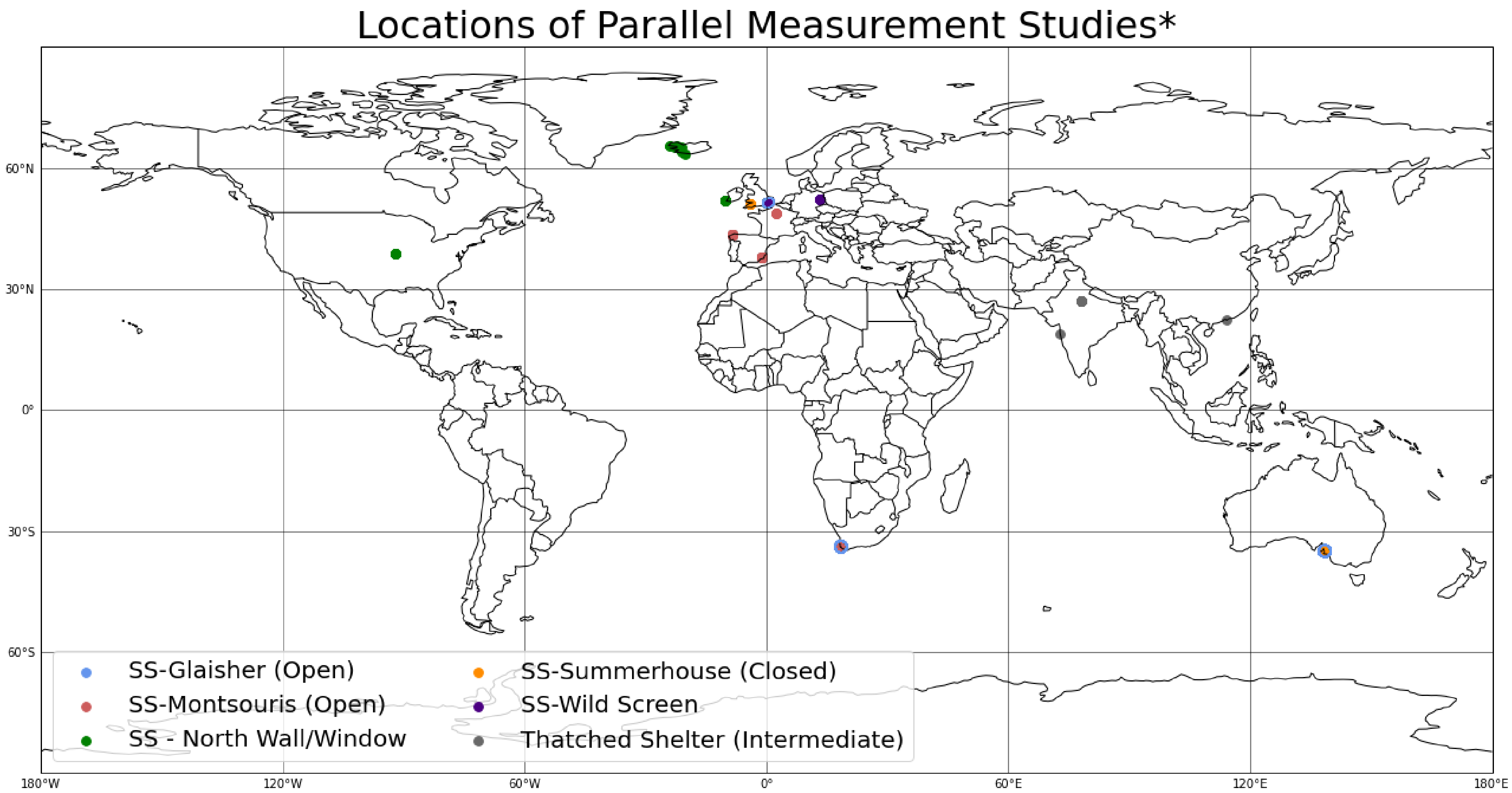
1. CHARACTERISE

Parallel Measurement Studies

Parallel measurement studies - where readings are taken in 2 or more exposures in parallel - can give an indication of the features of the exposure bias.

In the following slides we examine the difference between readings recorded in Stevenson and "open" screens (e.g. Glaisher, Montsouris).

Study locations are displayed in the figure to the right.



*Locations of studies for which data has been obtained for this piece of work, thus far.

University of East Anglia

GloSAT Global Surface Air Temperature

1. CHARACTERISE

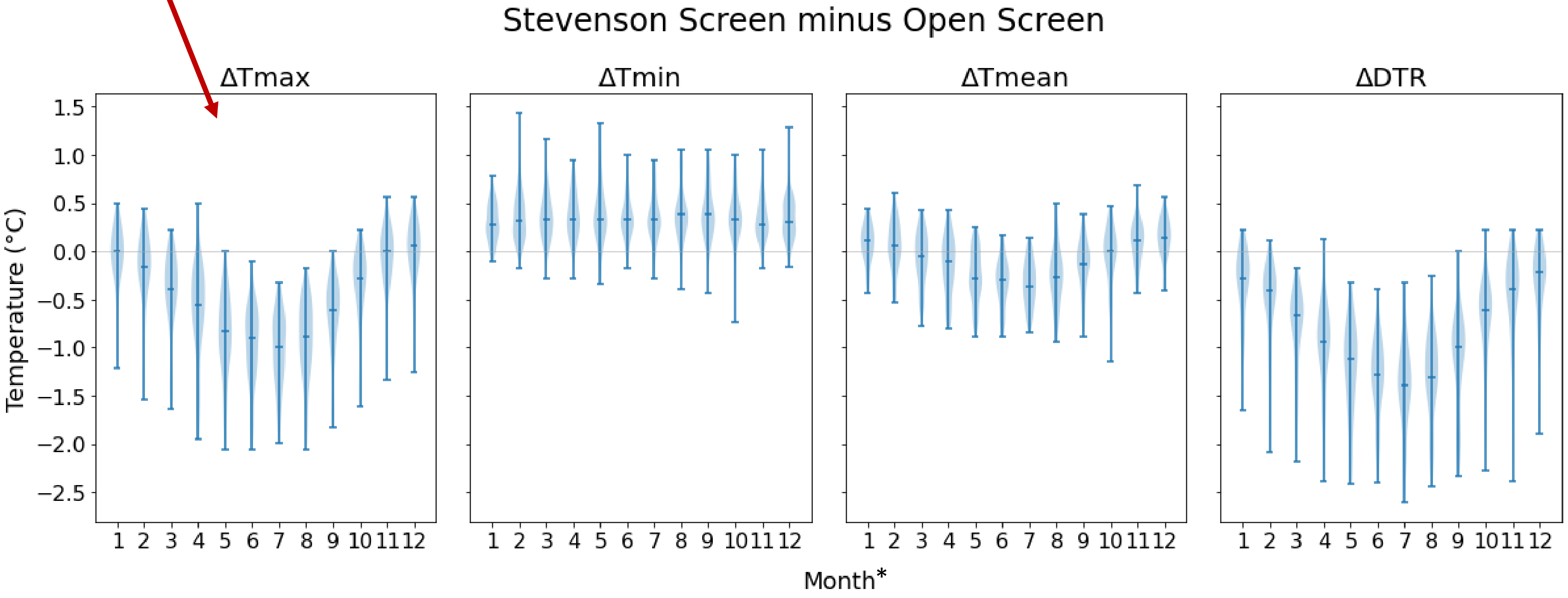
Stevenson screen minus "open" screen

Stevenson screens tends to read cooler Tmax and warmer Tmin than "open" screens

There is a clear seasonal cycle to the bias (except in Tmin)

The bias is greatest in Tmax and DTR , but can lead to a monthly bias in Tmean of up to 1.1°C.

Difference between the thermometer reading in the Stevenson screen and the "open" screen



Data sources: Adelaide Observatory Yearbooks; Detwiller, 1978; Ellis, 1891; Gaster, 1882; Gill, 1882; Greenwich Observatory Yearbooks; Margary, 1924; Mawley, 1897; SDATS/AEMET (Brunet, pers. comms)

**Monthly data for the Southern Hemisphere studies has been shifted 6 months so the seasons align*

University of East Anglia

GloSAT Global Surface Air Temperature

1. CHARACTERISE

Stations with a known date of screen change

Stations with a known date of screen change

When we know the date a Stevenson screen was introduced at a station, as well as the previous method of exposure, analysis of the period pre & post-introduction may give us an indication of the characteristics of the exposure bias.

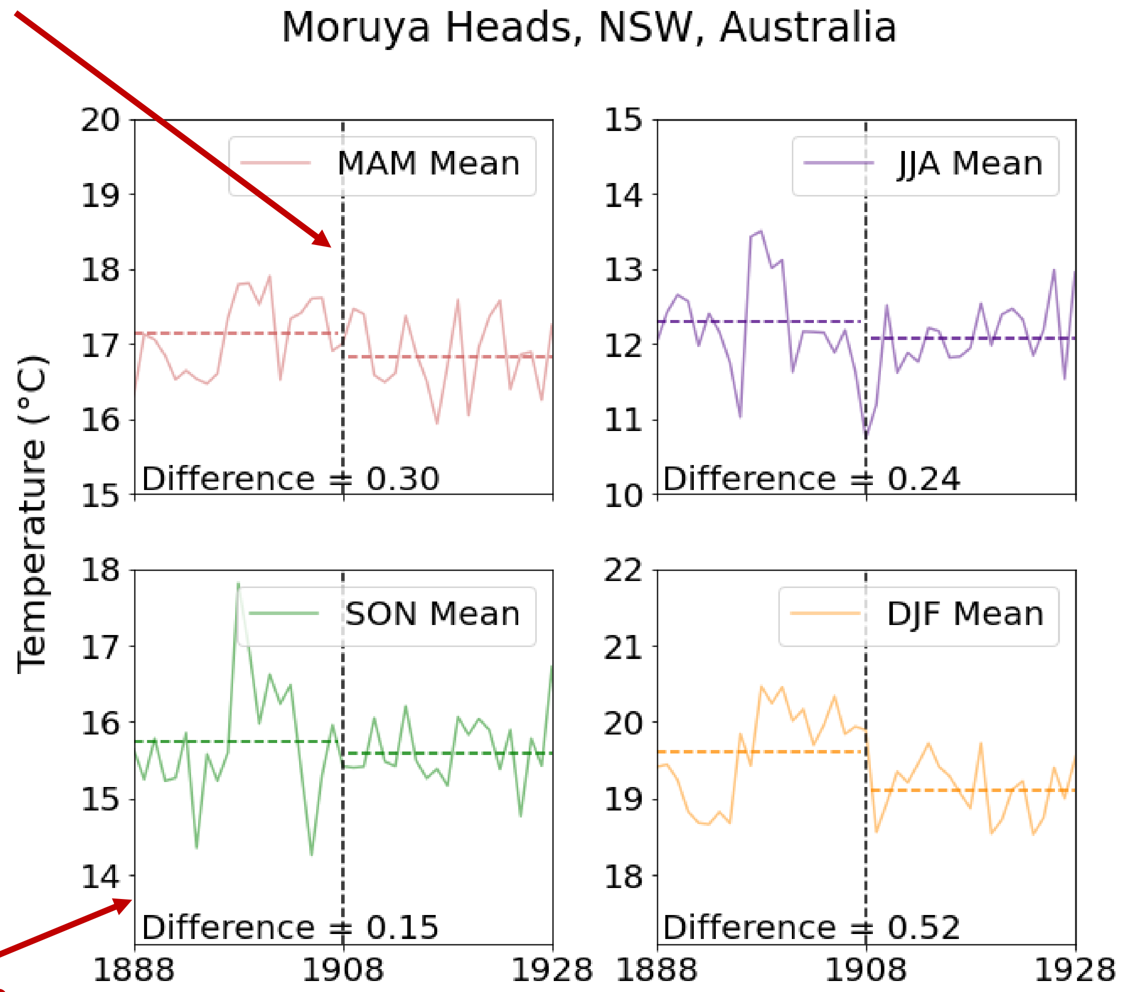
For example:

Stevenson Screen introduced

Mean level 20 years prior to change

-

Mean level 20 years post-change



The difference in the mean level may be representative of the exposure bias, assuming no change would have occurred otherwise

Data source: Linden Ashcroft.

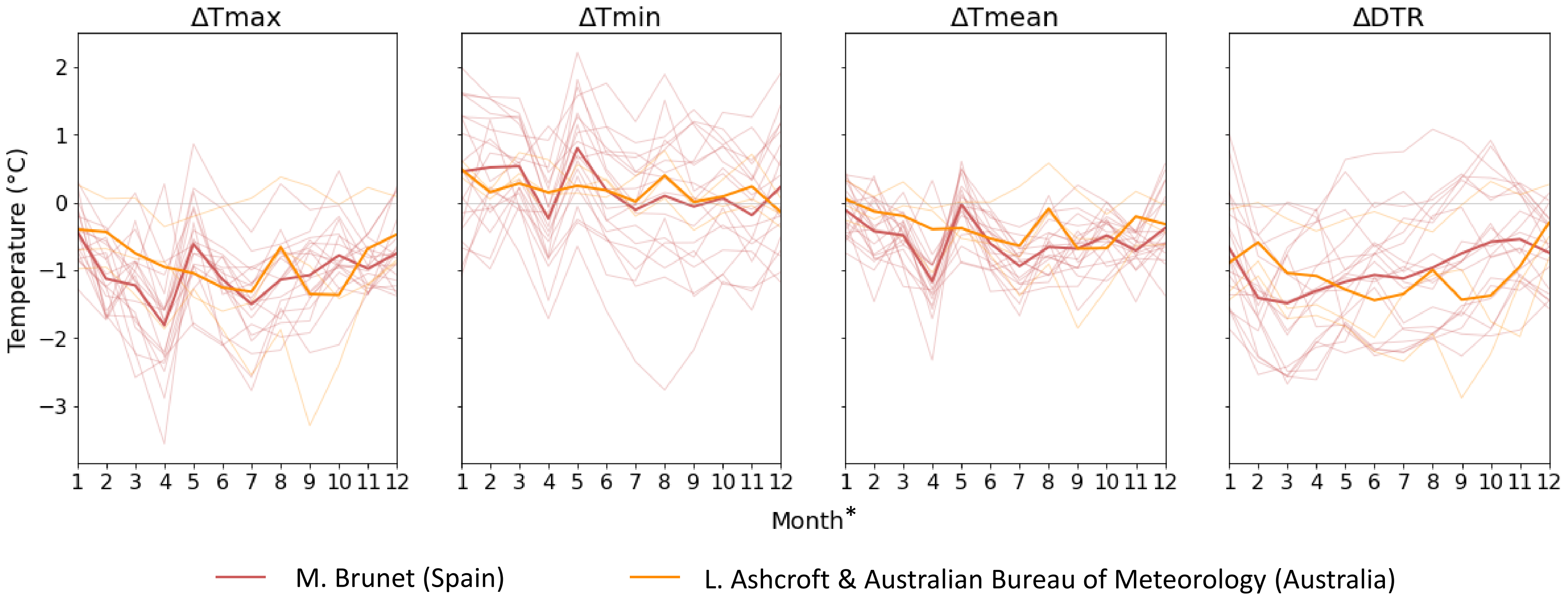
1. CHARACTERISE

Performing the same analyses as on the previous slide, across multiple stations, gives the data to the right. The data is noisy, but shows similar features to the parallel measurement studies:

On average, Stevenson screens tend to read cooler Tmax and warmer Tmin than "open" screens

Mean bias in monthly Tmean as large as 1°C in April

Mean level 20 years post-change (Stevenson screen) **−** Mean level 20 years prior to change ("open" screen)



*Monthly data for the Southern Hemisphere studies has been shifted 6 months so the seasons align

University of East Anglia

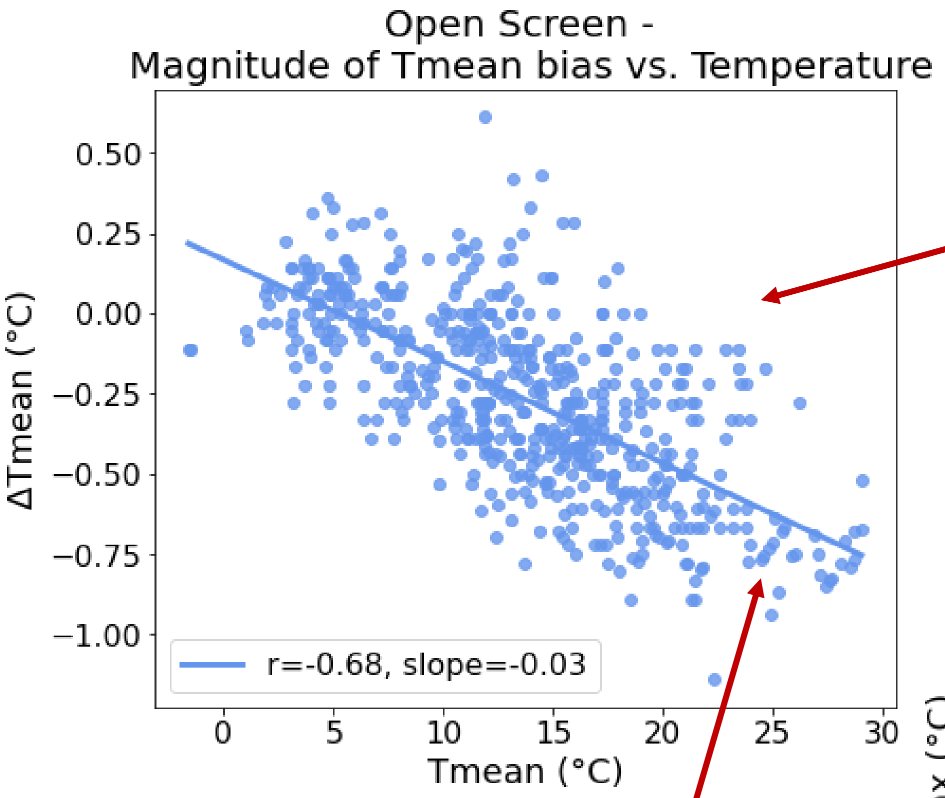
GloSAT Global Surface Air Temperature

1. CHARACTERISE

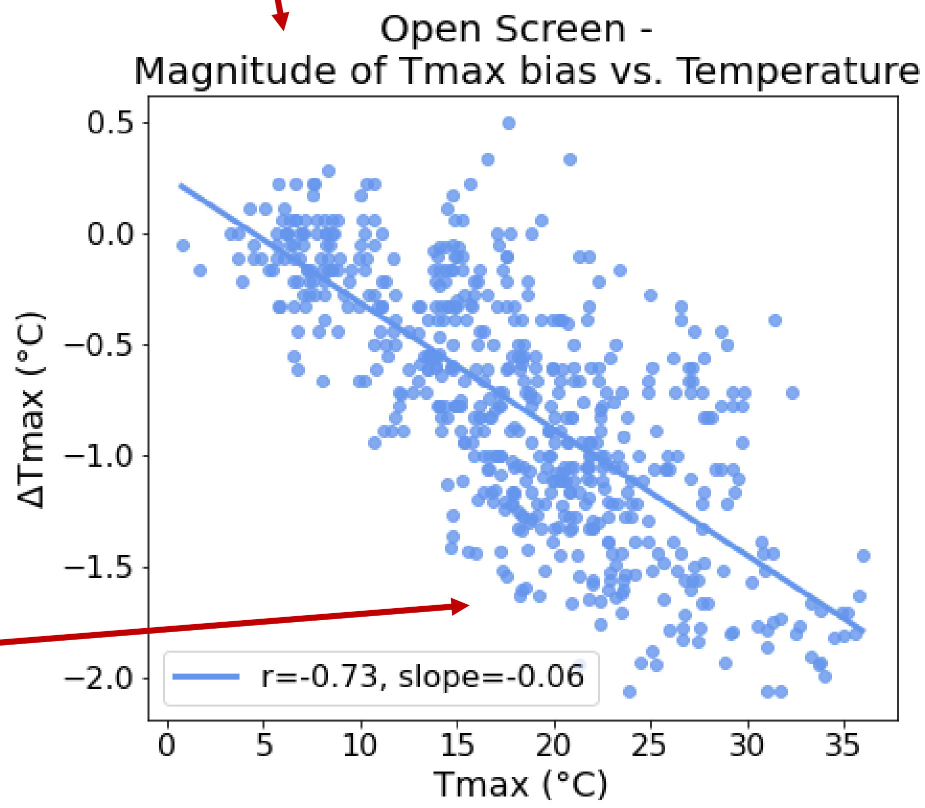
Is the magnitude of the bias influenced by temperature?

The plots on the right show a clear relationship between the temperature recorded (in the Stevenson screen) and the magnitude of the difference between the Stevenson screen and the "open" screen reading.

Warmer temperature = larger bias



Clear relationship between temperature and the magnitude of the bias in both Tmax and Tmean.

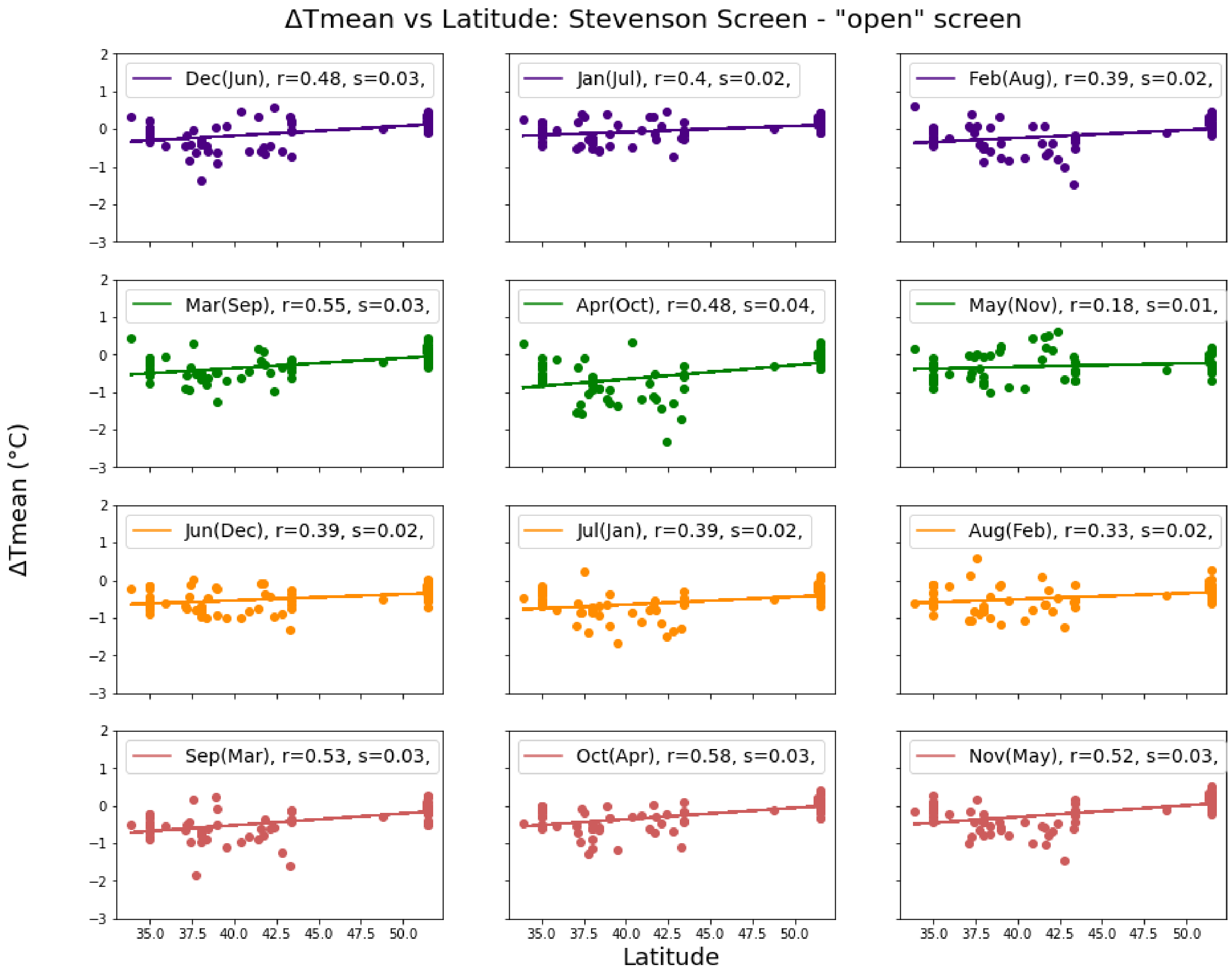


Relationship suggests warmer recorded temperatures lead to a greater magnitude of bias when comparing Stevenson and "open" screens.

1. CHARACTERISE

Is the magnitude of the bias influenced by location?

The plots on the right suggest latitude may have an influence on the magnitude of the (Tmean) bias - especially in spring and autumn - however the relationship is not strong enough to draw any firm conclusions.



University of East Anglia

GloSAT Global Surface Air Temperature

1. CHARACTERISE THE EXPOSURE BIAS

Summary: Stevenson screen vs. "open" stands

- Thermometers exposed in Stevenson screens tend to read cooler maximum temperatures (except in winter) and warmer minimum temperatures than open stands
- The bias is most evident in the maximum temperature and the diurnal temperature range (of the variables studied) but can bias the mean by as much as 1.1°C
- There is a clear seasonal cycle to the bias
- There is strong evidence that temperature influences the magnitude of the bias, but inconclusive evidence of a relationship between latitude and the magnitude of the bias

University of East Anglia

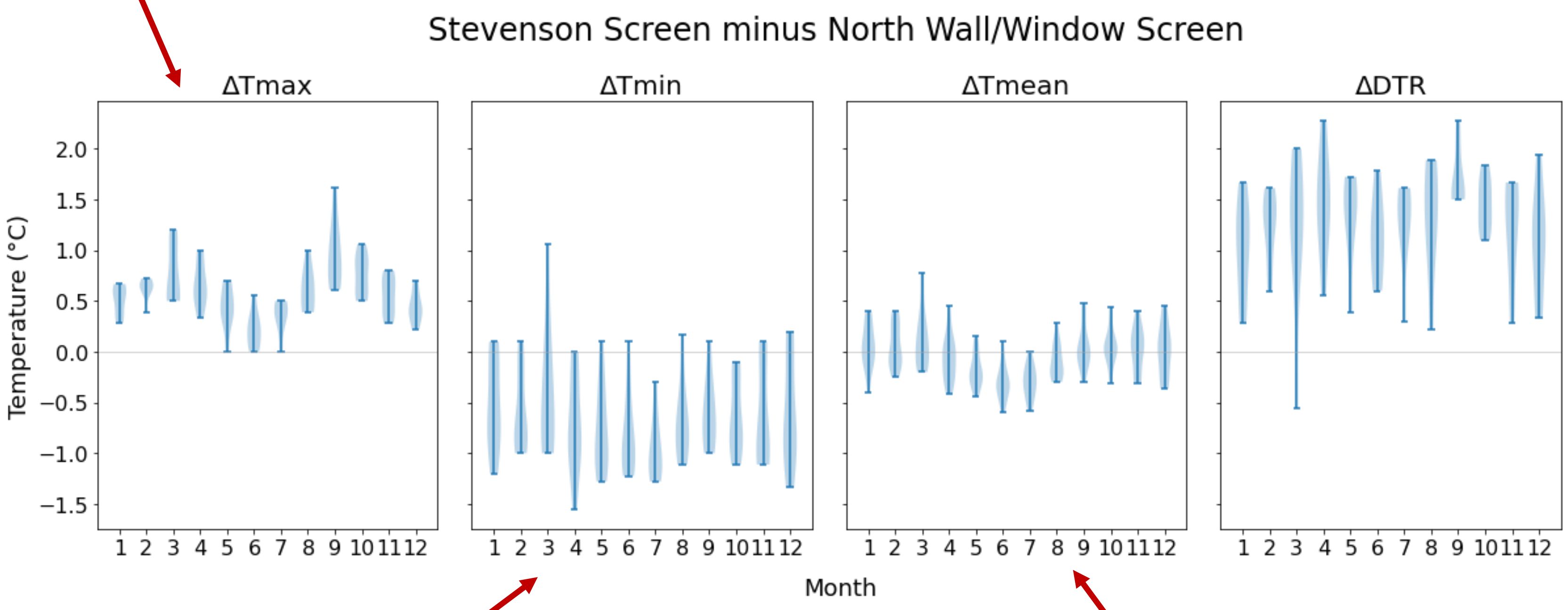
GloSAT Global Surface Air Temperature

1. CHARACTERISE

Stevenson screen minus North Wall / Window exposure

The previous slides focused on the comparison between Stevenson and "open" screens. The next two slides give a comparison between Stevenson screens and other common exposures, for information.

Stevenson screens tend to read warmer Tmax than a North Wall / Window exposure. This is likely due to the greater height of the thermometer in the latter exposure



Stevenson screen tends to read cooler Tmin than a North Wall / Window exposure. This is likely due to the thermal influence of the building keeping the thermometer warmer at night

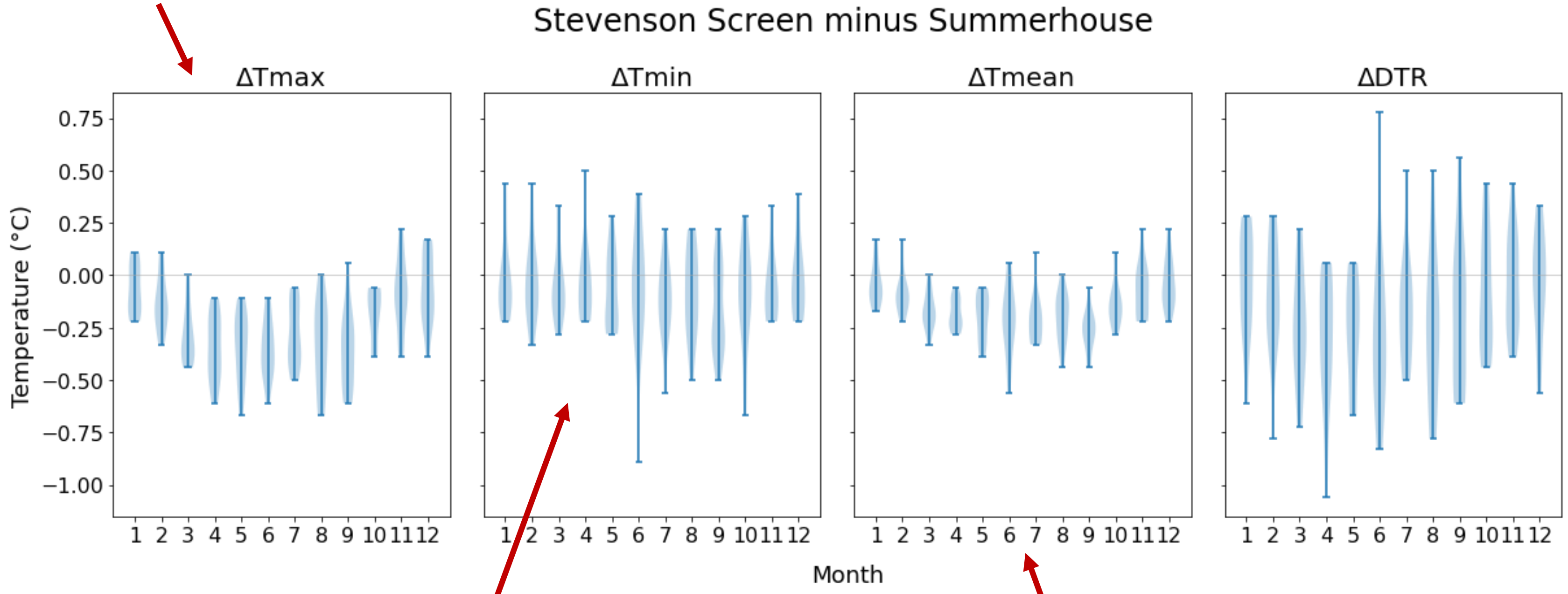
Although the bias in Tmax and Tmin are opposite to the previous comparison with "open" screens, the bias in Tmean is similar, and retains a seasonal cycle (when Tmean = ½(Tmax+Tmin))

Data sources: Butler, n.d.; Chandler, 1964; Chenoweth, 1992; Marriott, 1879; Veðráttan Journal

1. CHARACTERISE

Stevenson Screen minus Summerhouse

Stevenson screens tend to read cooler Tmax than thermometers exposed in Summerhouses - similar to "open" screens - although the magnitude of the bias is smaller



Variable difference in Tmin

As in the previous slides, Tmean is biased warmer in comparison to the Stevenson screen, especially in the summer months.

Data sources: Adelaide Observatory Yearbooks; Marriott, 1894

2. IDENTIFY THE EXPOSURE BIAS

Using the characteristics identified in Step 1, we next look at how we might identify possible instances of exposure bias in the instrumental data, using a combination of station metadata and comparator datasets, for a long station record in Germany.

CASE STUDY – BERLIN-DAHLEM STATION, GERMANY

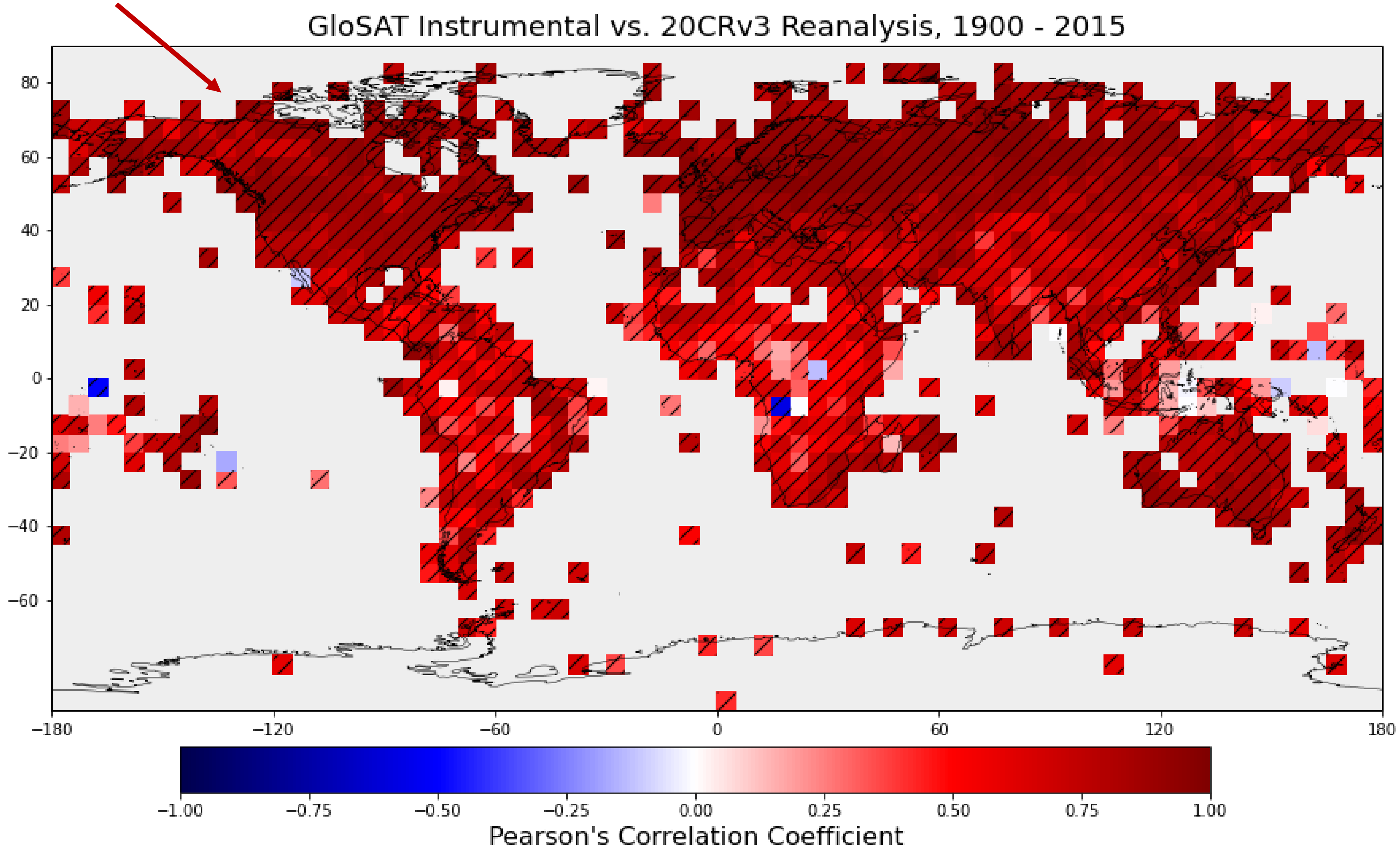
2. IDENTIFY

20CRv3 Reanalysis

20CRv3 Reanalysis is independent of observed land temperature data and can therefore be used as a comparator dataset (where/ when the data is considered to be representative).

The plot on the right shows generally good correlation between the 20CRv3 reanalysis ensemble mean and the instrumental data, meaning it may be useful as a comparator in this study.

Hash = significance (p<0.05).



20CRv3 Reanalysis: Slivinski et al., 2019

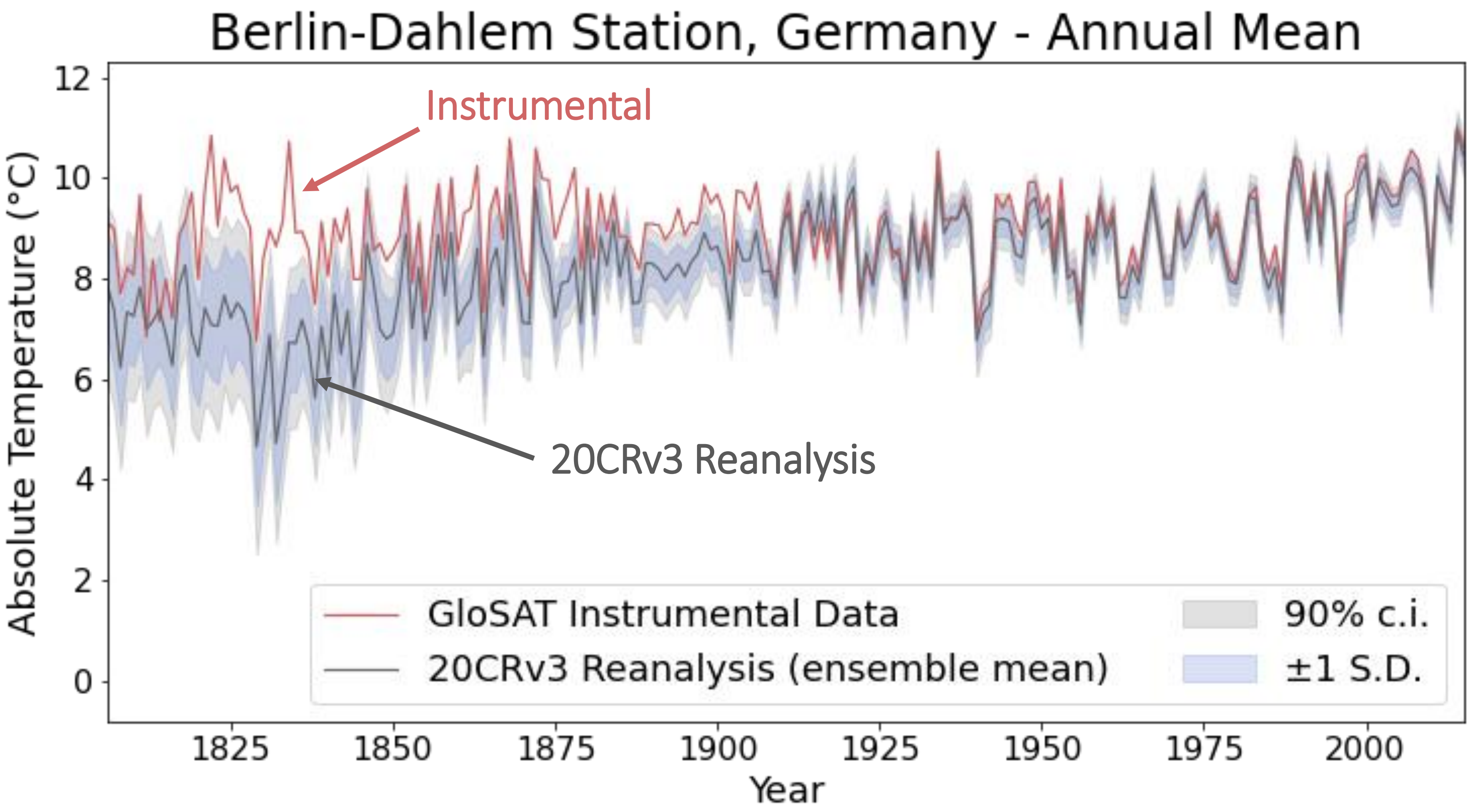
University of East Anglia

GloSAT Global Surface Air Temperature

2. IDENTIFY

Visually the 20CRv3 reanalysis ensemble mean compares well with the GloSAT instrumental data until c. 1840 and maintains a monthly correlation coefficient of >0.8 until 1860.

20CRv3 reanalysis is therefore considered a useful comparator dataset for this station.



University of East Anglia

GloSAT Global Surface Air Temperature

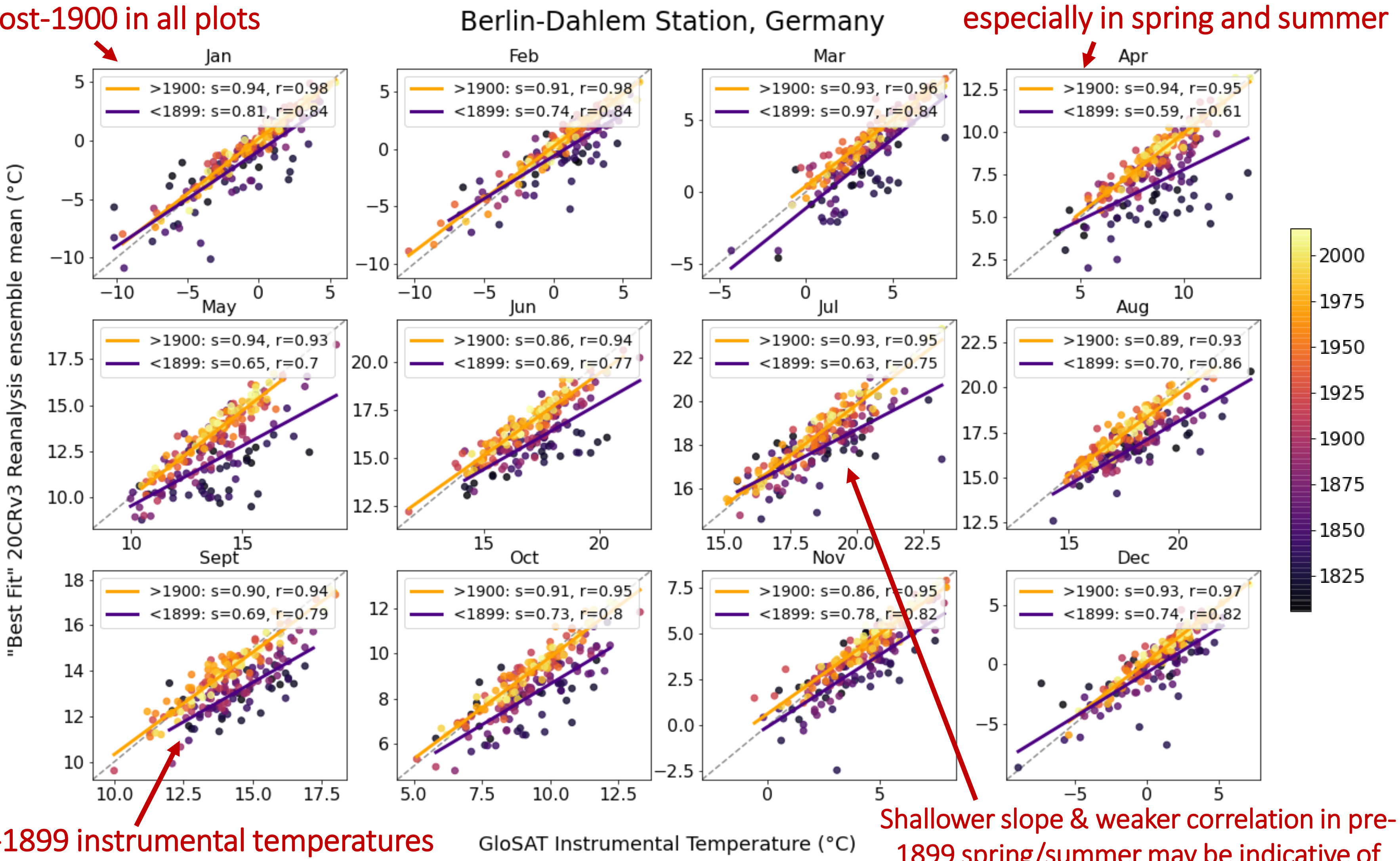
2. IDENTIFY

Linear regression (for 1950 – 2015) of the reanalysis onto the instrumental series is used to compensate for any biases in the reanalysis. This "Best Fit" version of the reanalysis is used for the remaining analyses.

Correlation between the "Best Fit" version of the reanalysis and the GloSAT instrumental data for Berlin-Dahlem station shows:

Very strong correlation post-1900 in all plots

Weaker correlation before 1899, especially in spring and summer



Pre-1899 instrumental temperatures tend to be warmer than the reanalysis, especially in non-winter months

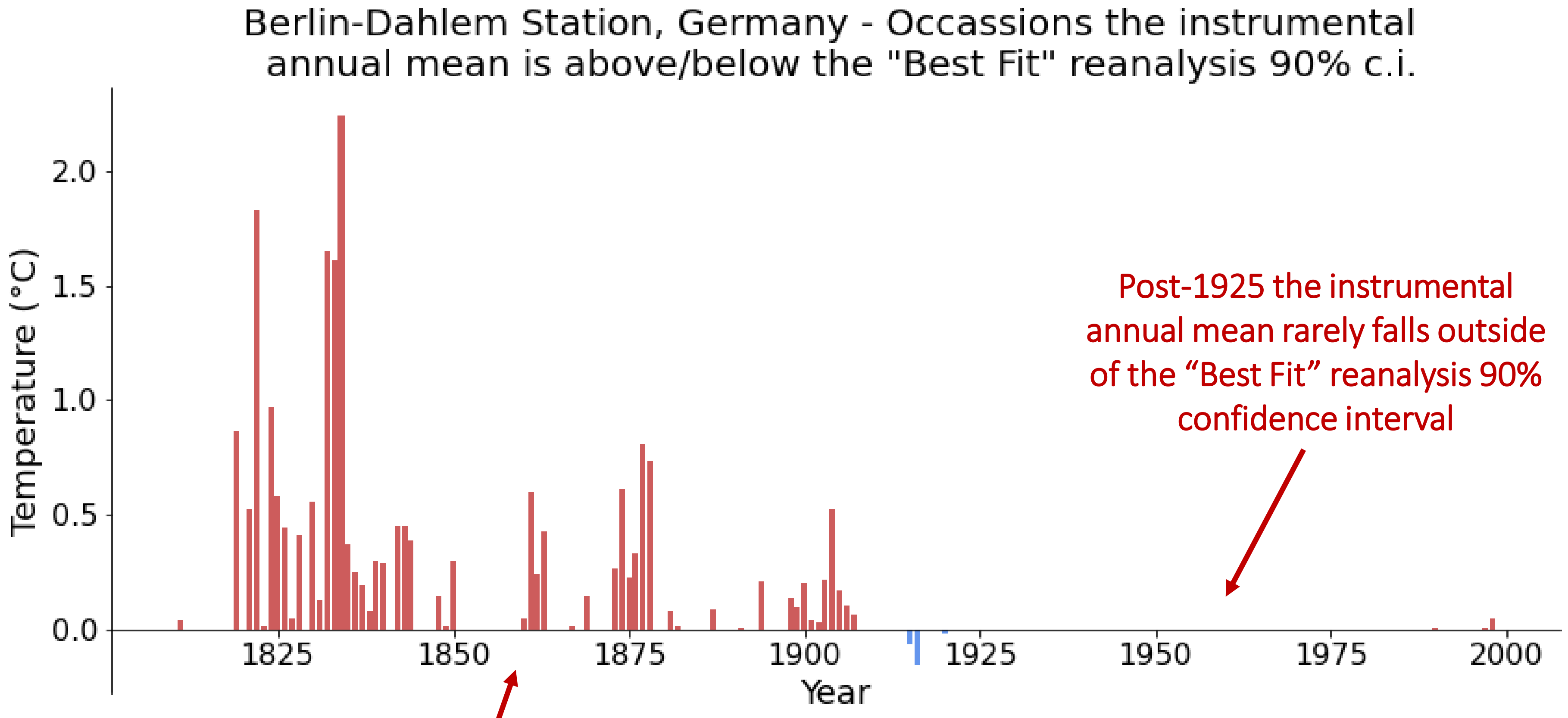
Shallower slope & weaker correlation in pre-1899 spring/summer may be indicative of exposure bias (Step 1 suggested summer season and warmer temperature = larger bias)

University of East Anglia

GloSAT Global Surface Air Temperature

2. IDENTIFY

To determine the extent of the difference between the two datasets we look at whether the GloSAT instrumental data for Berlin-Dahlem falls outside of the 90% confidence interval (5th – 95th percentile) at any time.



Post-1925 the instrumental annual mean rarely falls outside of the "Best Fit" reanalysis 90% confidence interval

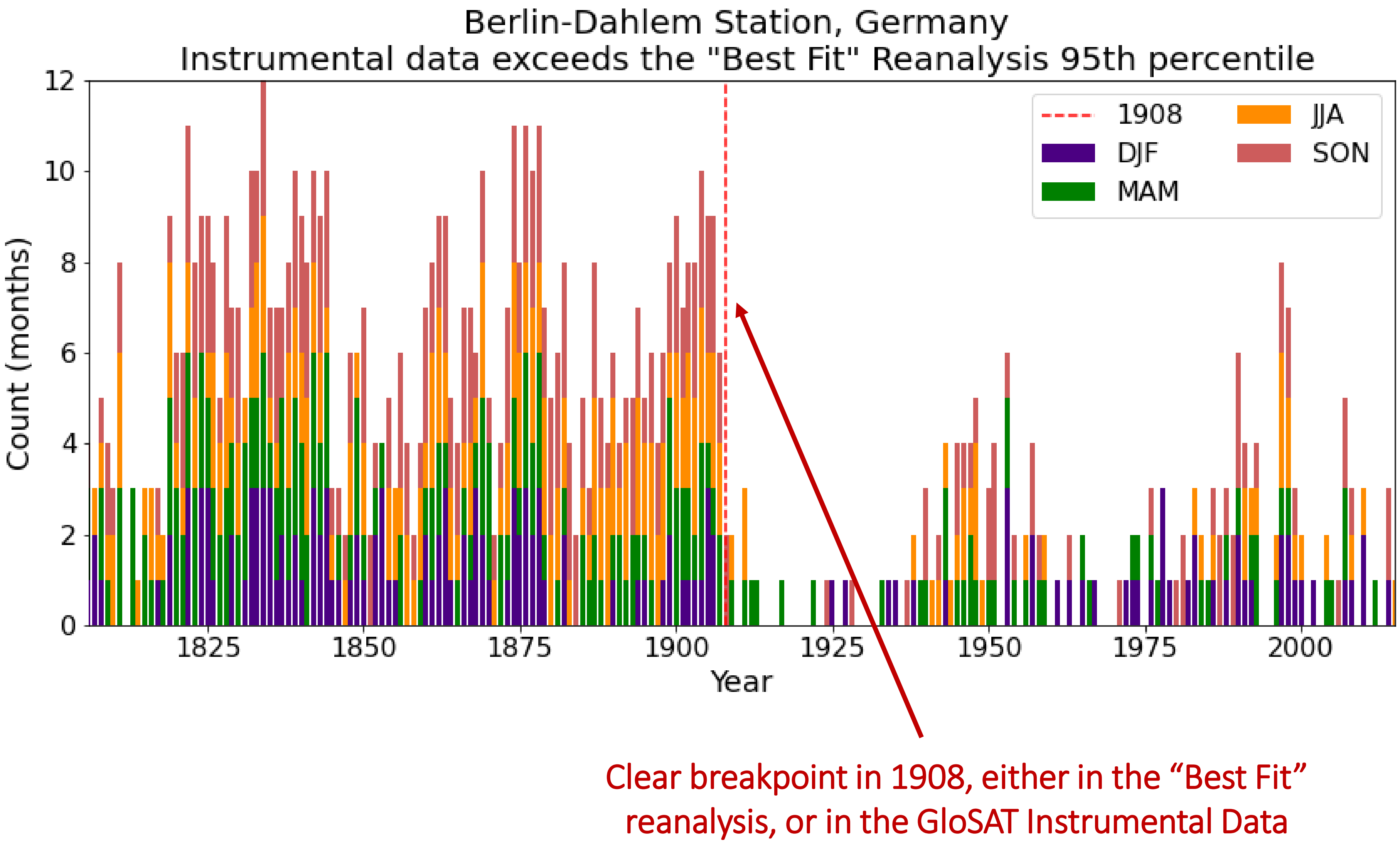
Pre-1910, the annual mean regularly exceeds the 90% confidence interval by up to 2.2°C (0.8°C in the more robust post-1860 period)

2. IDENTIFY

Looking in more detail at this, we can see that prior to 1908 the instrumental data exceeds the “Best Fit” reanalysis 95th percentile in at least 4 months in >80% of the years.

Post-1908 this occurs in only 12% of years, indicating a clear break-point in either dataset.

Exceedance appears to have a seasonal component – indicative of exposure bias?



University of East Anglia

GloSAT Global Surface Air Temperature

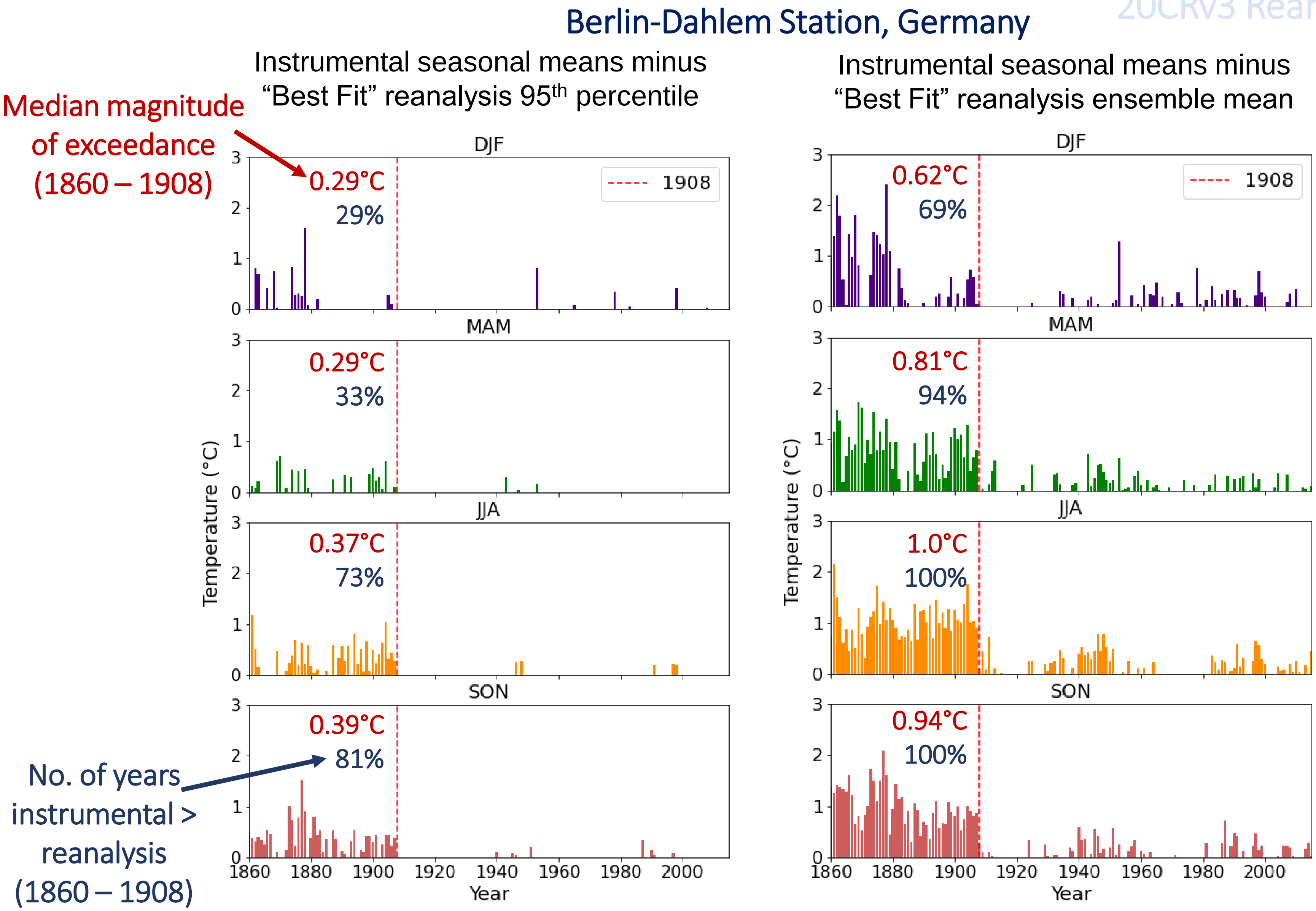
2. IDENTIFY

Looking at the more robust period of the reanalysis from 1860, it is clear there is a seasonal component to the incidence/magnitude of the divergence between the datasets.

The summer and autumn instrumental means most frequently exceed the reanalysis data pre-1908; they also have the greatest magnitude of exceedance.

Winter has lowest rate and magnitude of exceedance.

20CRv3 Reanalysis



University of East Anglia

GloSAT Global Surface Air Temperature

2. IDENTIFY THE EXPOSURE BIAS

Summary: 20CRv3 Reanalysis

- Preliminary findings suggest it is possible to use 20CRv3 to identify inhomogeneities/bias in instrumental land surface air temperature data (in selected locations)

Summary: What does 20CRv3 Reanalysis tell us about Berlin-Dahlem station?

- 20CRv3 Reanalysis is considered a useful comparator dataset for Berlin-Dahlem station
- Comparison between the two datasets suggests a breakpoint in 1908, with the instrumental temperature data diverging (warmer) before this date
- There is a seasonal component to the divergence, with the summer and autumn seasonal means exceeding the reanalysis 95th percentiles/ensemble means more frequently and to a greater extent than winter (and to a lesser extent) spring
- The magnitude of the divergence in each season is similar to the magnitude of the exposure bias identified in Step 1

2. IDENTIFY

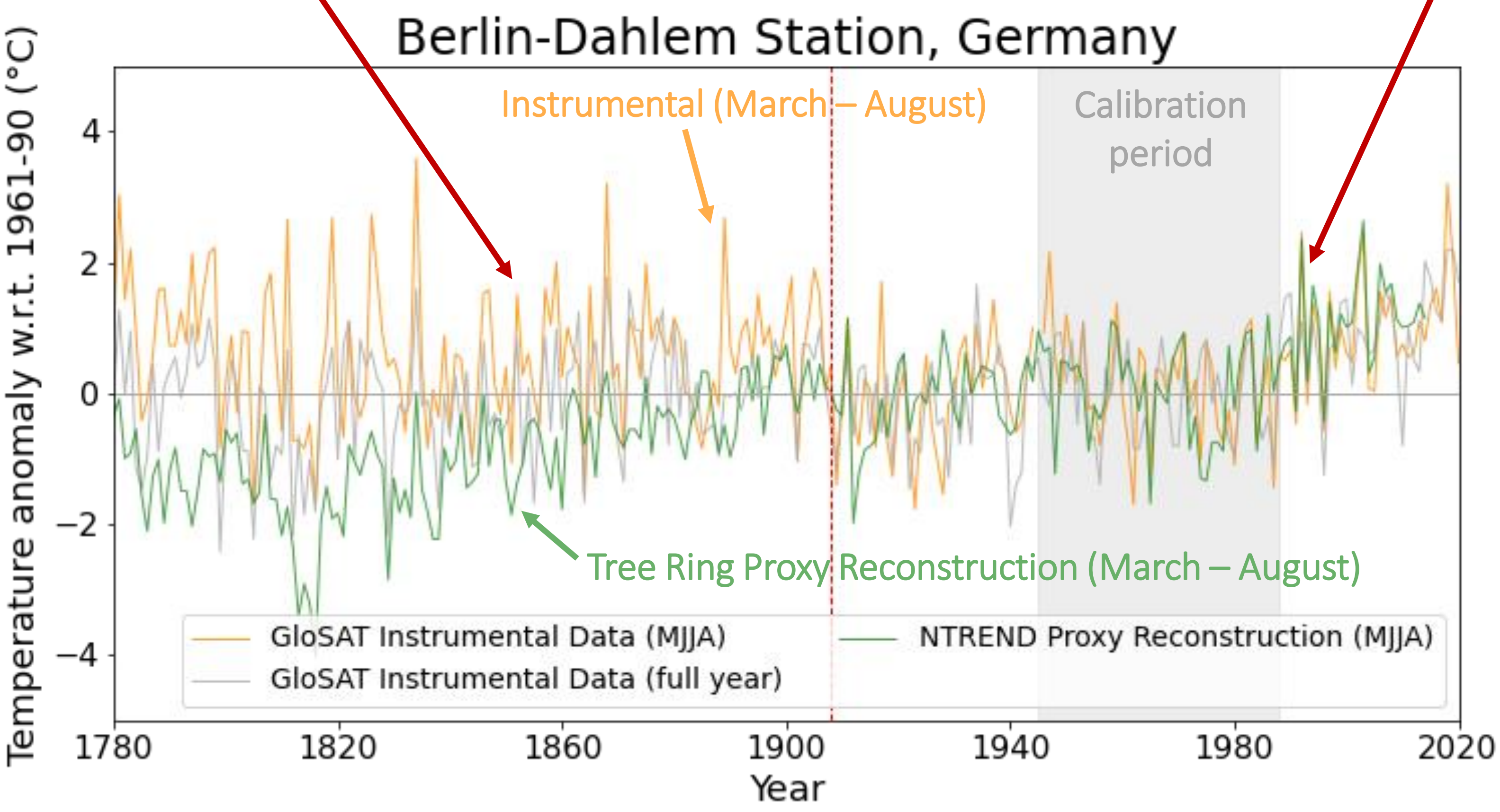
NTREND Proxy Reconstruction

Proxy reconstructions can also be used as comparator datasets, if robust, and if the calibration period/data does not overlap with the data/period of interest.

This figure compares the spatially-resolved NTREND tree ring reconstruction with the GloSAT instrumental data for Berlin-Dahlem station (MJJA mean).

Divergence pre-1900 (dashed line = 1908 for reference) with instrumental MJJA temperature anomalies warmer than the NTREND tree ring reconstruction

Visually, NTREND compares reasonably well with the instrumental data in the recent period

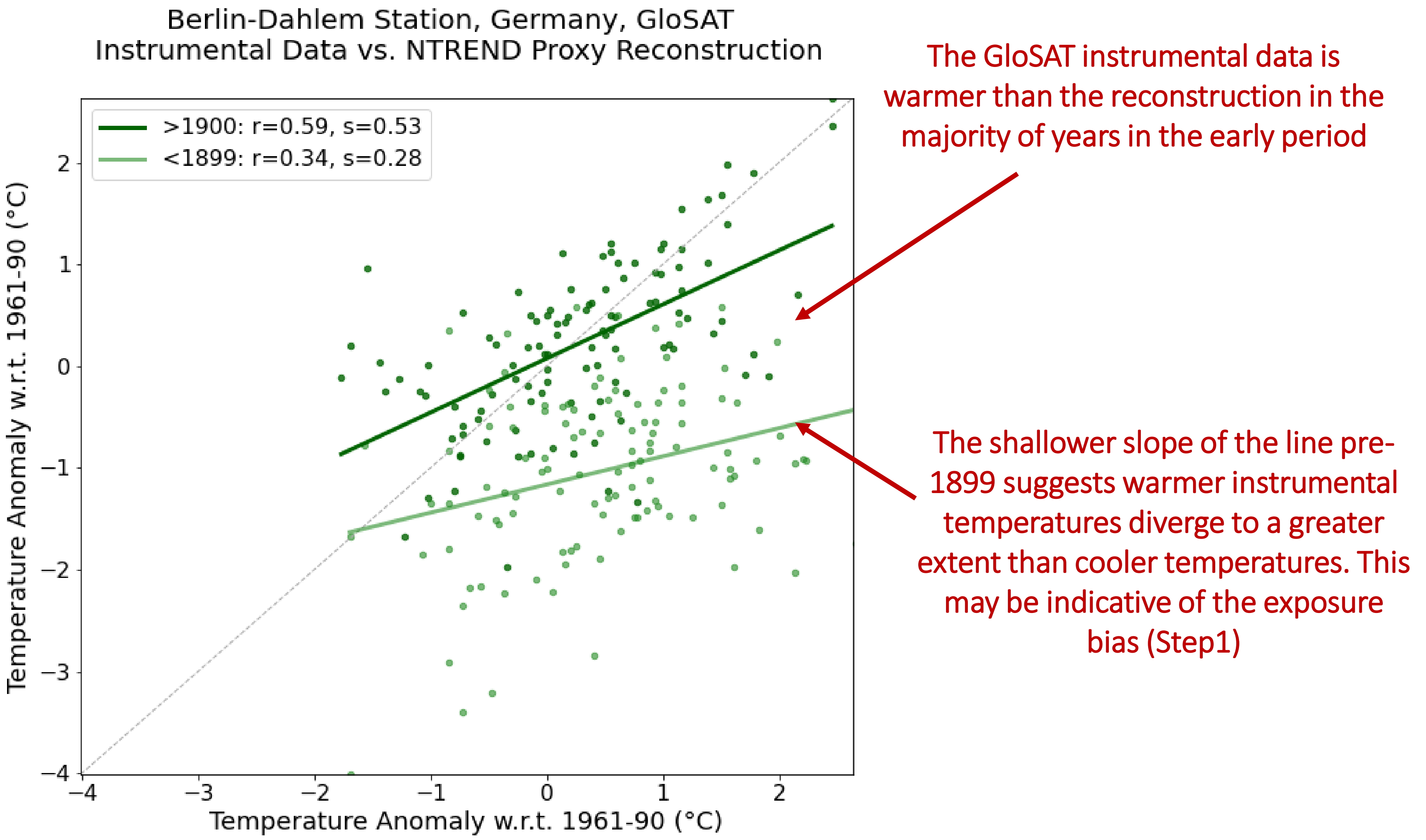


NTREND Reconstruction: Anchukaitis et al., 2017

2. IDENTIFY

The GloSAT instrumental data (MJJA) and NTREND reconstruction (MJJA) have a correlation coefficient of 0.59 for the period 1900 – 2010. This is a reasonable correlation for a proxy reconstruction.

This correlation weakens significantly in the early period (pre-1899) when the instrumental data is warmer than the reconstruction in the majority of years.



University of East Anglia

GlOSAT Global Surface Air Temperature

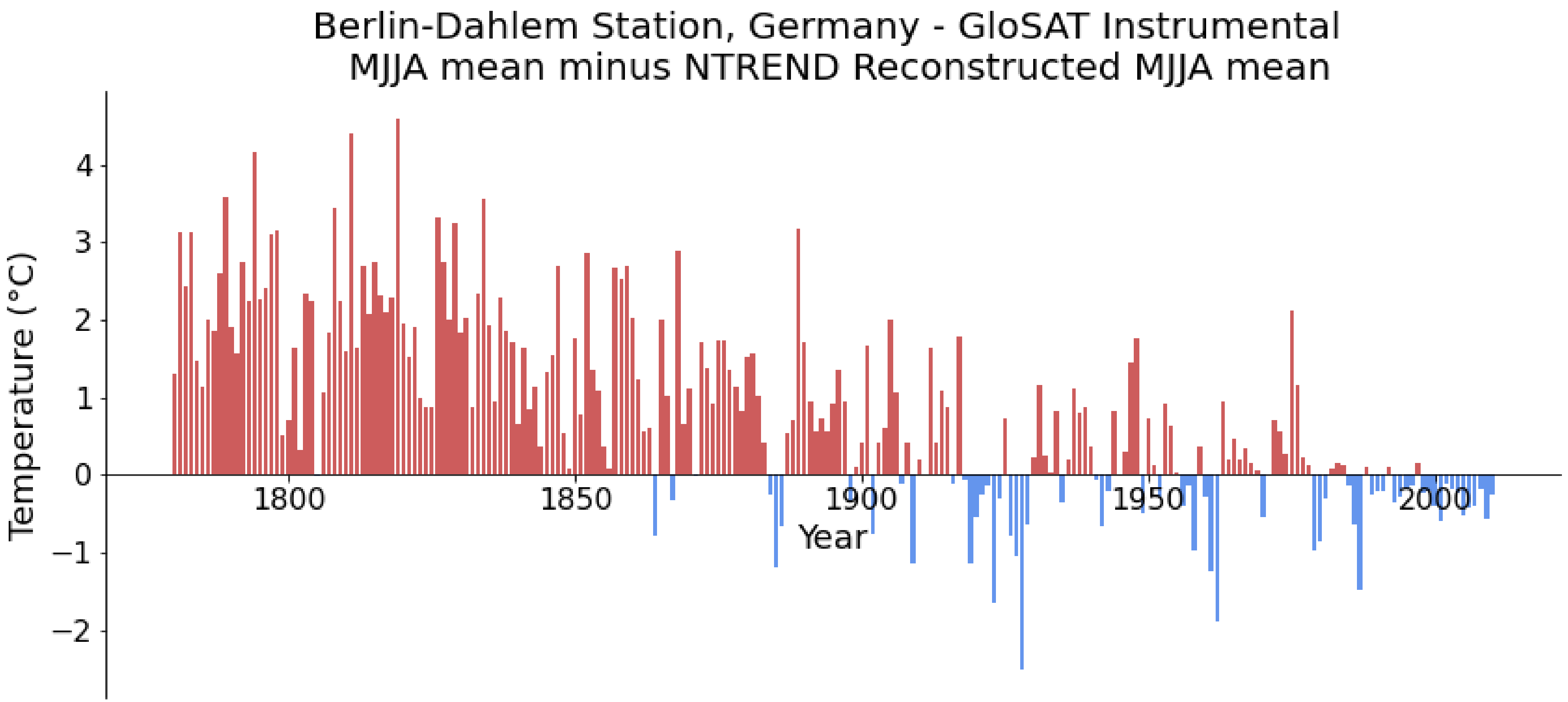
2. IDENTIFY

Although not as clearly as the reanalysis data, the proxy data also indicates a breakpoint around 1900.

Instrumental data > NTREND reconstruction in 92% of years pre-1900, but only 48% post-1900.

The magnitude of the exceedance is also greater pre-1900.

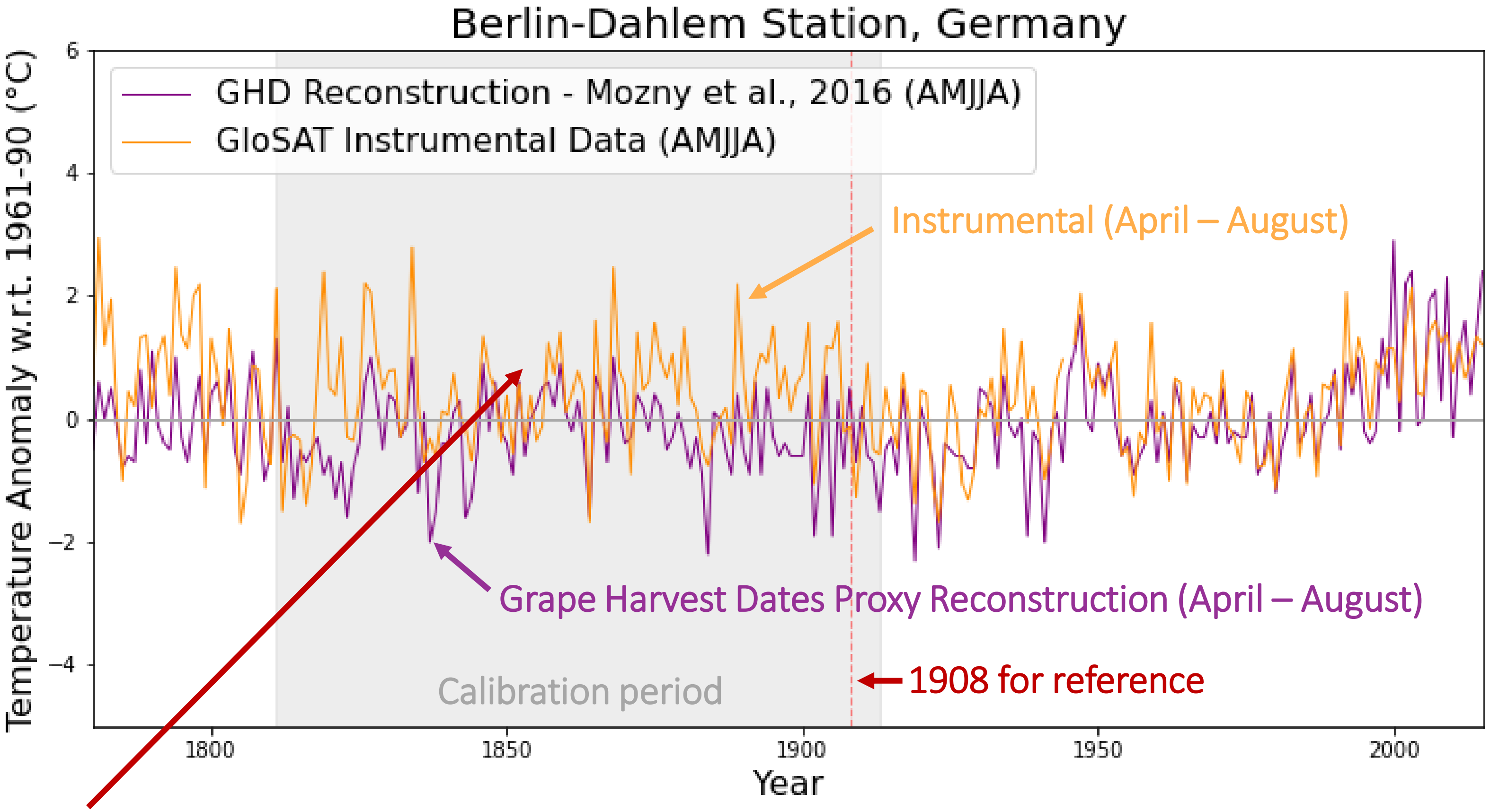
The median exceedance between 1860-1900 is 1.0°C, the same as in the 20CRv3 reanalysis.



2. IDENTIFY

To further illustrate the evidence for potentially too warm pre-1900 summers in the Berlin-Dahlem station record, we briefly share one further proxy reconstruction, developed from Grape Harvest Dates (GHD) nearby.

The series was calibrated over the interval of interest, but used data from the Czech region (Brazdil et al., 2012), independent of the Berlin-Dahlem series, so it can be used as a comparator.



As with the NTREND reconstruction, the Berlin-Dahlem instrumental data corresponds reasonably well with the GHD reconstruction post-1910, but shows some divergence pre-1910

2. IDENTIFY

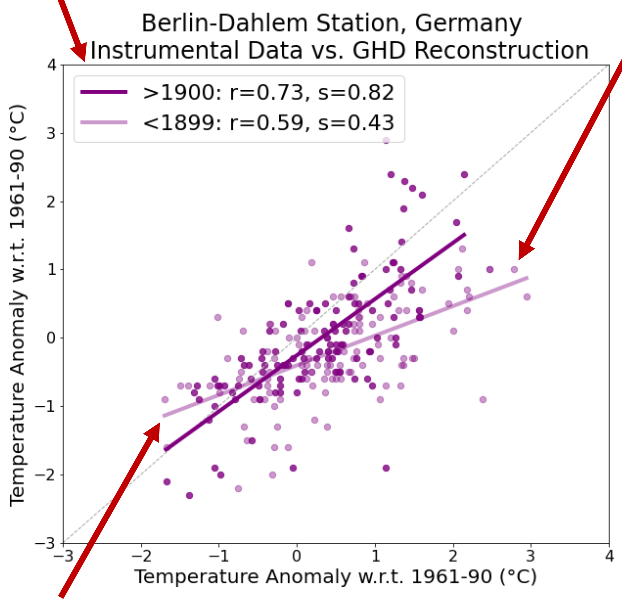
The instrumental data and GHD reconstruction have a robust correlation coefficient of 0.73 for the period 1900 – 2015. Pre-1900 the data shows the same pattern as the previous comparisons with a weaker correlation/shallower slope.

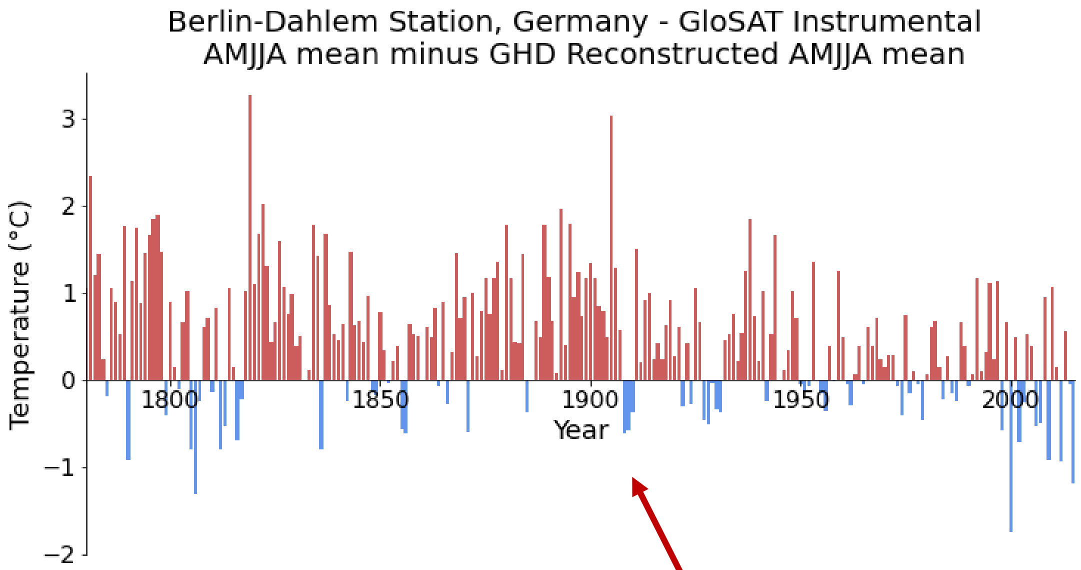
Instrumental data > GHD reconstruction in 81% of years pre-1900, 66% post.

No clear breakpoint but the magnitude of exceedance is greater pre-1900.

Robust correlation coefficient post-1900

Shallower slope pre-1900





The correlation weakens pre-1900, but remains fairly robust

No clear breakpoint, however the median magnitude of exceedance is 0.82°C between 1860 – 1900 and only 0.54°C post-1900

2. IDENTIFY THE EXPOSURE BIAS

Summary: Proxy Reconstructions

- Both proxy reconstructions, which use different proxies and calibration data, suggest the summer temperatures observed in Berlin-Dahlem are too warm during the pre-1910 period
- The magnitude of the divergence in the summer months corresponds with the 20CRv3 reanalysis data and is within the range we might expect as a result of the exposure bias (as identified in Step 1)
- The divergence between the proxy data and the Berlin-Dahlem instrumental data increases with warmer temperatures. This is also suggestive of the exposure bias (as identified in Step 1)

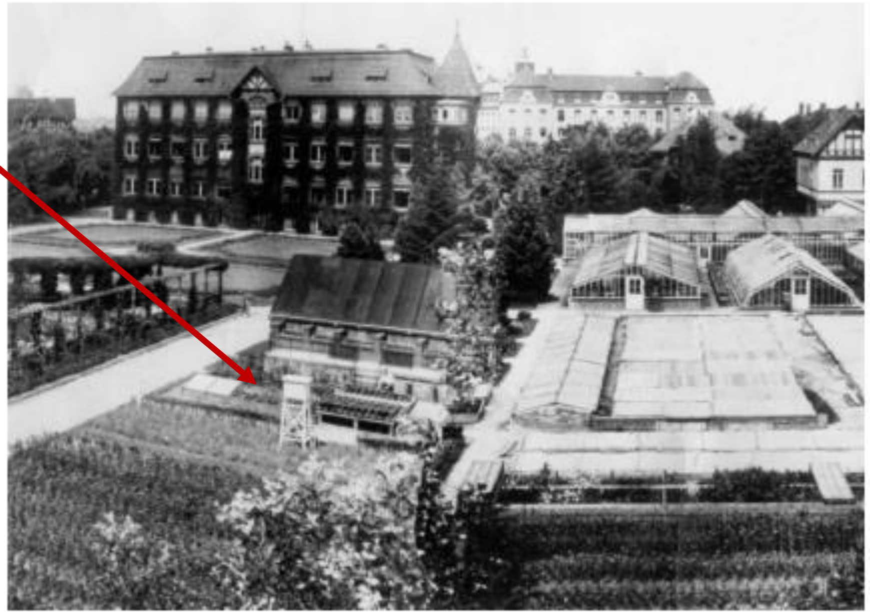
2. IDENTIFY

Station Metadata

The proxy reconstructions and 20CRv3 reanalysis data both indicate a breakpoint in the early 1900s in the Berlin-Dahlem station. The features of the bias are similar to the exposure bias, but does the metadata support this?

YES!

Stevenson-type screen in the garden of the Royal Prussian Gardening School.



Berlin-Dahlem Station metadata –

This dataset is a blend of temperature measurements made at a site in the city of Berlin (1701-1907) and (since 1908) at the Royal Prussian Gardening School in the Berlin suburb of Berlin-Dahlem. After the site change in 1908 a Stevenson-type screen was used, prior to this a wall-screen was used.

The metadata suggests the early data has been corrected to account for the urban environment, suggesting the remaining 'bias' may be a result of the change in exposure. This is supported by the characteristics of the bias and the divergence identified in steps 1 and 2 respectively.

Image source: Pelz, 2007; Data sources: Cubash and Kudow, 2011; Smithsonian Institution (WWR), 1927

University of East Anglia

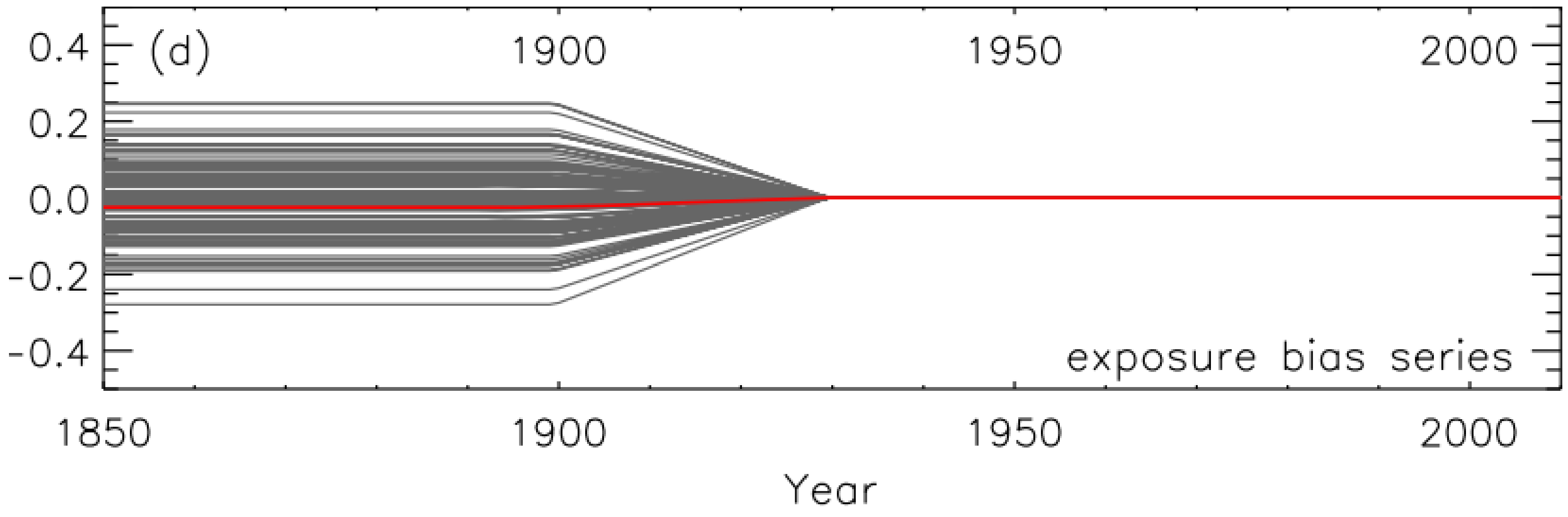
GloSAT Global Surface Air Temperature

3. ERROR MODEL

3. ERROR MODEL

The results of this study suggest there is a case for refining the representation of the exposure bias in the CRUTEM error model, which currently does not contain a seasonal cycle.

Work on this aspect of the project is ongoing.



Current representation of the exposure bias in CRUTEM5:

"For grid boxes in the latitude range of 20°S–20°N a 1σ uncertainty of 0.2°C is assumed prior to 1930. This then decreases linearly toward a value of zero in 1950. For stations that lie outside of 20°S– 20°N the exposure bias uncertainty takes a value of 0.1°C prior to 1900, decreasing linearly to zero by 1930" (Morice et al., 2012, p8)

Figure: Morice et al., 2012

SUMMARY

- Step 1: Changes in thermometer exposure can lead to significant differences in recorded temperatures. The magnitude of the bias differs according to season and temperature and can lead to a bias in the mean as large as 1°C
- Step 2: Preliminary work suggests it may be possible to identify the exposure bias using 20^{th} century reanalysis v3 and proxy reconstructions, supported by station metadata
- Step 3: Preliminary findings would support a revision of the error model to include a seasonal cycle, rather than a fixed annual bias

University of East Anglia

GloSAT Global Surface Air Temperature

NEXT STEPS

Characterise:

- Continue to collate and analyse parallel measurements and stations with a known date of Stevenson screen introduction to improve our knowledge of the characteristics of the exposure bias
- Explore how the bias differs diurnally and with method of mean calculation

Identify:

- Continue to use comparator datasets and metadata to identify possible instances of exposure bias
- Explore the possibility of using breakpoint detection algorithms to identify exposure bias

Error model / Impact:

- Propose refinements to the CRUTEM/GloSAT error model using the information gained in the characterisation stage of the project, as well as metadata regarding the type of exposure that was used previously in different regions and the likely date Stevenson screens were introduced in each region
- Explore the impact any identified biases have on long-term temperature trends, and our understanding of climate variability and change in the early instrumental period

REFERENCES

Adelaide Observatory Yearbooks – Government of South Australia (multiple years) 'Meteorological Observations made at the Adelaide Observatory and other places in South Australia and the Northern Territory' [online]. Available at: https://catalog.hathitrust.org/Record/012261912

Anchukaitis, K. J. et al. (2017) 'Last millennium Northern Hemisphere summer temperatures from tree rings: Part II, spatially resolved reconstructions', Quaternary Science Reviews, pp. 1–22. doi: 10.1016/j.quascirev.2017.02.020.

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

Ashcroft, L. (n.d.) 'Adjustments made to 1860–2009 southeastern Australian temperature, rainfall and mean sea level pressure data' [online]. Available at: https://lindenashcroft.com/research/

Böhm, R. *et al.* (2010) 'The early instrumental warm-bias: A solution for long central European temperature series 1760-2007', *Climatic Change*, 101(1), pp. 41–67. doi: 10.1007/s10584-009-9649-4.

Brázdil, R., Belínová, M., Dobrovolný, P., Mikšovský, J., Pišoft, P.,Reznícková, L., Štepánek, P., Valášek, H., and Zahradníˇcek, P. (2012) 'History of Weather and Climate in the Czech Lands IX: Temperature and Precipitation Fluctuations in the Czech Lands During the Instrumental Period', pp. 236. ISBN 978-80-210-6052-4.

Brunet, M. *et al.* (2006) 'The development of a new dataset of Spanish daily adjusted temperature series (SDATS) (1850-2003)', *International Journal of Climatology*, 26(13), pp. 1777–1802. doi: 10.1002/joc.1338.

Butler, C. J. *et al.* (no date) 'Meteorological Data recorded at Armagh Observatory: Vol 2 - Daily, Mean Monthly, Seasonal and Annual, Maximum and Minimum Temperatures, 1844-2004'.

Chandler, T. J. (1964) 'North-wall and stevenson screen temperatures at kew observatory', Quarterly Journal of the Royal Meteorological Society, 90(385), pp. 332–333. doi: 10.1002/qj.49709038514.

University of East Anglia

Global Surface Air Temperature

REFERENCES (cont'd)

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

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

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

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), pp. 240–249. doi: https://doi.org/10.1002/qj.4970178006.

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 11 to *Quarterly Weather Report for 1879*, Meteorological Office, London, pp. 13-39.

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), pp. 238–243. doi: 10.1002/qj.4970084404.

Greenwich Observatory Yearbooks – Board of Admiralty (multiple years) 'Results of the Magnetical and Meteorological Observations made at the Royal Observatory, Greenwich' [online]. Available at: http://www.geomag.bgs.ac.uk/data_service/data/yearbooks/yearbooks.html

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), pp. 209–226. doi: 10.1002/qj.49705021109.

Marriott, W. (1879) 'Thermometer exposure - Wall versus Stevenson Screens', *Quarterly Journal of the Royal Meteorological Society*.

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), pp. 164–168. doi: 10.1002/qj.4970209005.

REFERENCES (cont'd)

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

Morice, C. P. et al. (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), doi: 10.1029/2011JD017187.

Možný, M. et al. (2016) 'April–August temperatures in the Czech Lands, 1499–2015, reconstructed from grape-harvest dates', *Climate of the Past*, 12, pp. 1499–2015. doi: 10.5194/cp-12-1421-2016.

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

Pelz, J. (2007) 'Einhundert Jahre Wetteraufzeichnungen in (Berlin)-Dahlem', Beiträge des Instituts für Meteorologie der Freien Universität Berlin zur Berliner Wetterkarte [online] Available at:https://berliner-wetterkarte.de/Beilagen/2007/Pelz_29_07.pdf

Slivinski, L. C. et al. (2019) 'Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system', *Quarterly Journal of the Royal Meteorological Society*, 145(724), pp. 2876–2908. doi: 10.1002/qj.3598.

Smithsonian Institution (1927) 'World Weather Records, Smithsonian Miscellaneous Collections', Vol. 79, Washington, DC

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

Veðráttan Journal (1962) - 'Annual Report 1962', *Veðráttan* [online] Available at: https://timarit.is/page/3128806.