Constraining the Nineteenth-Century Temperature Baseline for Global Warming

L. SCHNEIDER,[a,b] O. KONTER,[c] J. ESPER,[c,d] AND K. J. ANCHUKAITIS[e,f]

[a] Department of Geography, Justus Liebig University, Giessen, Germany
[b] Center for international Development and Environmental Research, Justus Liebig University, Giessen, Germany
[c] Department of Geography, Johannes Gutenberg University, Mainz, Germany
[d] Global Change Research Institute of the Czech Academy of Sciences (CzechGlobe), Brno, Czech Republic
[e] School of Geography, Development, and Environment, The University of Arizona, Tucson, Arizona
[f] Laboratory of Tree-Ring Research, The University of Arizona, Tucson, Arizona

(Manuscript received 26 October 2022, in final form 22 May 2023, accepted 23 May 2023)

ABSTRACT: Since the Paris Agreement, climate policy has focused on 1.5° and 2°C maximum global warming targets. However, the agreement lacks a formal definition of the nineteenth-century "pre-industrial" temperature baseline for these targets. If global warming is estimated with respect to the 1850–1900 mean, as in the latest IPCC reports, uncertainty in early instrumental temperatures affects the quantification of total warming. Here, we analyze gridded datasets of instrumental observations together with large-scale climate reconstructions from tree rings to evaluate nineteenth-century baseline temperatures. From 1851 to 1900 warm season temperatures of the Northern Hemisphere extratropical landmasses were 0.20°C cooler than the twentieth-century mean, with a range of 0.14°–0.26°C among three instrumental datasets. At the same time, proxy-based temperature reconstructions show on average 0.39°C colder conditions with a range of 0.19°–0.55°C among six records. We show that anomalously low reconstructed temperatures at high latitudes are underrepresented in the instrumental fields, likely due to the lack of station records in these remote regions. The nineteenth-century offset between warmer instrumental and colder reconstructed temperatures is reduced by one-third if spatial coverage is reduced to those grid cells that overlap between the different temperature fields. The instrumental dataset from Berkeley Earth shows the smallest offset to the reconstructions indicating that additional stations included in this product, due to more liberal data selection, lead to cooler baseline temperatures. The limited early instrumental records and comparison with reconstructions suggest an overestimation of nineteenth-century temperatures, which in turn further reduces the probability of achieving the Paris targets.

SIGNIFICANCE STATEMENT: The warming targets formulated in the Paris Agreement use a "pre-industrial" temperature baseline that is affected by significant uncertainty in the instrumental temperature record. During the second half of the nineteenth century, much of the continental landmasses were not yet covered by the observational station network and existing records were often subject to inhomogeneities and biases, thus resulting in uncertainty regarding the large-scale mean temperature estimate. By analyzing summer temperature reconstructions from tree-rings for the Northern Hemisphere extratropical land areas, we examine an independent climate archive with a typically broader and more continuous spatial extent during the "pre-industrial" period. Despite the additional uncertainty when using climate reconstructions instead of direct observations, there is evidence for an overestimation of land temperature during the summer season in early instrumental data. Colder early instrumental temperatures would reduce the probability of reaching the Paris targets.

KEYWORDS: Northern Hemisphere; Paleoclimate; Surface temperature; Automatic weather stations; Bias; Tree rings

1. Introduction

With the United Nations' decision to adopt the Paris Agreement at the 21st Conference of the Parties, climate change research was motivated to address the newly defined target of limiting global warming to less than 1.5°C above "pre-industrial levels" (Allen et al. 2019; Jehn et al. 2021; Knutti et al. 2016). However, the Paris Agreement does not define a particular temperature nor a time period that could serve as a reference (Hawkins et al. 2017; Schurer et al. 2017). The Fifth IPCC Assessment Report (AR5), published 2 years earlier, reports the likelihood for crossing the 1.5° and 2°C thresholds relative to the 1850–1900 period in the summary for policymakers (IPCC 2013) without using the term "pre-industrial" for referring to this temperature baseline (Kirtman et al. 2013). The IPCC's Special Report on Global Warming of 1.5°C and the Sixth IPCC Assessment Report (AR6) closes this gap by describing the 1850–1900 period as an "approximate" (Allen et al. 2019) and "pragmatic choice" (Chen et al. 2021) for global warming estimates from "pre-industrial" to modern times. However, the wording indicates that using this period is a compromise. There is evidence that global warming was occurring before the second half of the nineteenth century (Hegerl et al. 2007; Schurer et al. 2013; Abram et al. 2016), implying that a baseline from 1850 to 1900 cannot reflect

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-22-0806.s1.

Corresponding author: Lea Schneider, lea.schneider@geogr.uni-gießen.de

DOI: 10.1175/JCLI-D-22-0806.1

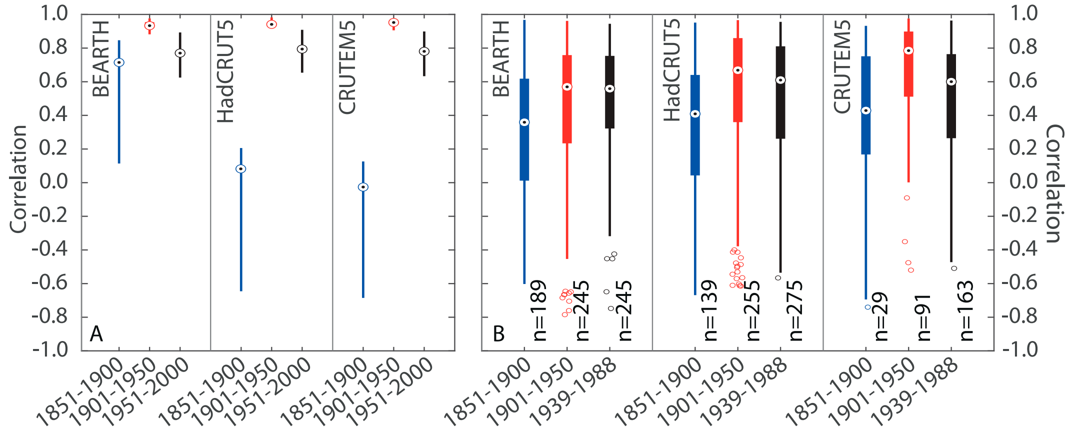

FIG. 3. Correlation between reconstructed and instrumental temperatures for different time windows after low-pass filtering with 11-yr running means. (a) Time series of six NH extratropical reconstructions from Sch15, Sto15, Wil16, Gui17, and Bün21 correlated with the large-scale averages of the three instrumental fields. Points denote the median of the six correlation values and the vertical lines are the minimum and maximum. (b) Correlations between single grid cells in Anc17 and the instrumental fields. Boxplots indicate the median, the 25th and 75th quantiles, and the extremes (maximum whisker length is 1.5 times the interquartile range). Note that the number of correlation values varies over time and between datasets.

4. Discussion

a. Uncertainty in the instrumental datasets

A difference in BT of 0.12°C between NH extratropical land and summer averages from BEARTH and CRUTEM5 reveals some of the uncertainty in large-scale warming estimates derived from these datasets. Expressed in the terminology of the IPCC reports, warming between 1851–1900 and 1986–2005 was 0.68°C for CRUTEM5 and 0.78°C for BEARTH using the seasonal and spatial limits of this study. Obviously, the 0.10°C difference largely results from the offset in BTs. Hawkins et al. (2017) also reported an 0.10°C difference between the warming calculated for global land and ocean (instead of NH land) temperatures over the whole year (instead of summer only): For the same time interval, temperatures in HadCRUT4 rose by 0.61°C and in BEARTH by 0.71°C. Like CRUTEM5, the original version of HadCRUT4 does not use infilling. AR6 reports 0.05°C less warming for global land areas if warming is calculated with the updated and infilled HadCRUT5 dataset instead of BEARTH (Gulev et al. 2021). This difference between HadCRUT5 and BEARTH is again close to the offset in BTs that we found within the spatial and seasonal limits of this study (BT_{diff} = 0.06°C). However, these analogies do not imply that our interpretation of the results can be upscaled to annual and global averages. There might be other processes and biases impacting temperature variability in other seasons and regions and on larger spatial scales. One of the most obvious examples is SSTs, which make up 70% of the world's surface. Their recording in ship logs follows a very different protocol from land stations, resulting in specific biases and statistical treatment. Reduced spatial coverage in high latitudes might be a feature similarly impacting temperature datasets of air temperature over land and SST (Kent et al. 2017).

BEARTH and HadCRUT5 differ in the underlying station network while both apply infilling. The resulting large-scale

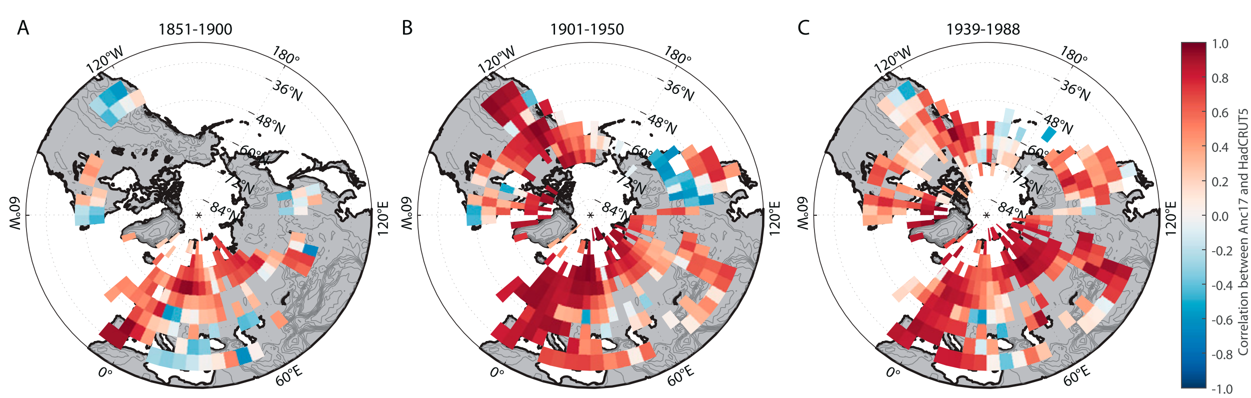

FIG. 4. Correlation between low-pass filtered reconstructed (Anc17) and instrumental (HadCRUT5) fields for different time intervals. Colors indicate Pearson correlations calculated from (a) 1851–1900, (b) 1901–50, and (c) 1939–88. The last 50-yr period terminates with the recent end of Anc17.

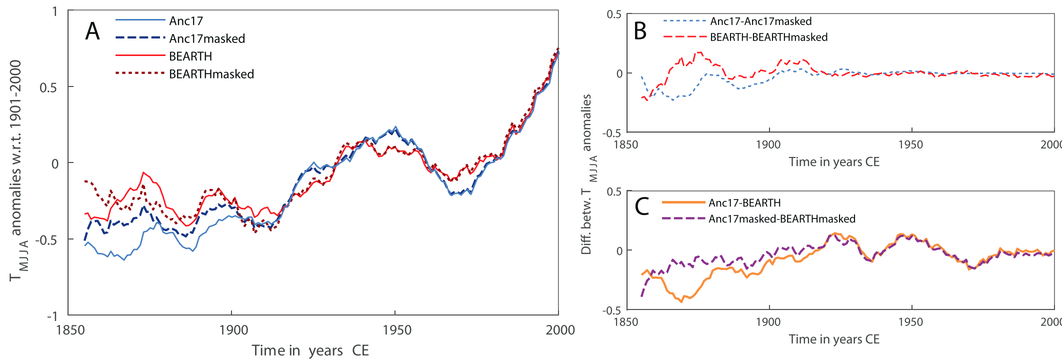

FIG. 5. Full and masked temperature averages for instrumental and reconstructed fields. Masked are those grid cells that do not overlap between Anc17 and CRUTEM5 (Fig. S2). (a) Large-scale temperature estimates smoothed with an 11-yr moving average. (b),(c) Differences between the unmasked and masked averages.

temperature estimates agree well after ~1870 and diverge strongly during the preceding decades. During this early period BEARTH covers a larger area of the NH (Fig. S1). But an average of BEARTH with only those grid cells available in HadCRUT5 results in almost identical large-scale temperature estimates as the full BEARTH average (not shown). In addition, the offset between BEARTH and the spatially more complete infilled HadCRUT4 dataset (Cowtan and Way 2014) is similar to the offset between BEARTH and HadCRUT5. As suggested by Rohde and Hausfather (2020), this proves that the divergence between these two datasets is a result of differences in the underlying station network rather than an artifact of the infilling intensity.

BEARTH and HadCRUT5/CRUTEM5 rely on different strategies for selecting and processing station records. The more liberal BEARTH approach results in a much larger number of input records, and as a consequence depends more strongly on successful nonclimatic noise or error cancellation during temperature interpolation. In a more conservative approach, HadCRUT5 and CRUTEM5 uses fewer and preselected station records in order to minimize the amount of nonclimatic noise from the very beginning. Both approaches are well justified and should be viewed as complementary (Menne et al. 2018).

b. *Agreement between instrumental and reconstructed BTs*

Large-scale reconstructions for summer temperatures add another perspective on BTs. In contrast to the instrumental data, the numbers of sites and trees entering the large-scale averages from 1851 to 1900 are less variable in the proxy networks and the quality of the reconstructions is presumably constant over the nineteenth and much of the twentieth century. Only since the 1980s the tree-ring chronologies become sparser, because much of the network was established during the late twentieth century (Briffa et al. 2001; Schweingruber et al. 1988). This is also one of the reasons why some reconstructions diverge from instrumental temperatures during this period (Fig. 1) (D'Arrigo et al. 2008). Tree-ring reconstructions estimate BTs to be lower than their instrumental counterparts. The only exception to this is Sto15, a reconstruction with an underestimated first-order autocorrelation (Esper et al. 2018) that targeted the representation of interannual variability versus decadal and longer trends (St. George and Esper 2019). The detrending method, used to remove age trends in the tree-ring measurements, can reduce the variability in the low-frequency domain. Such methodological choices affect not only BTs but also the temperature estimates during the late twentieth century and are another reason for the "divergence problem" (D'Arrigo et al. 2008; Esper and Frank 2009). Likewise, some of the spread between the other reconstructed BTs can be explained with different reconstruction targets (spatial domain or seasonal window) and analytical goals (Büntgen et al. 2021b). All of the reconstructions are associated with uncertainties arising from the spatial sampling, from calibration biases, and from biological memory affecting the spectral properties (Esper et al. 2018; Schneider et al. 2015; Wilson et al. 2016). There is no consensus for the best way to represent reconstruction uncertainties, particularly for spatial estimates, and as a consequence individual studies apply varying strategies. An estimate of the uncertainties can be derived by comparing the reconstructed temperatures to the instrumental counterparts in the twentieth century, when instrumental uncertainty is very small.

It should be noted that all temperature reconstructions are calibrated against instrumental temperatures using the gridded products discussed in this study or their previous versions, such as Anc17 with HadCRUT4 (Cowtan and Way 2014). Scaling and linear regression is used to transfer tree-ring indices or their principal components into temperature with the intrinsic assumption that the instrumental temperatures are the "true" target (Frank et al. 2007). If the instrumental target is erroneous, the error might propagate into the reconstruction. Due to uncertainties in the early portion of the instrumental fields, some reconstructions use exclusively the twentieth century for proxy calibration (Anchukaitis et al. 2017; Schneider et al. 2015). In this study, all reconstructions were scaled to a common instrumental target outside the BT period (see section 2) to avoid circularity.

The much warmer 1851–70 temperatures in HadCRUT5 and CRUTEM5 result in an overall decreasing trend over the

BT period. This cooling is observed neither in BEARTH nor in any of the index reconstructions for NH extratropical summers. Thus, correlation between the large-scale temperature averages from reconstructions and instrumental records is low for HadCRUT5 and CRUTEM5 in the low-frequency domain (Fig. 3). BEARTH, in contrast, correlates well with most of the index reconstructions. Correlations with Anc17 on the level of individual grid cells are on average slightly stronger for HadCRUT5 and CRUTEM5 than for BEARTH. Better agreement between HadCRUT5/CRUTEM5 and Anc17 during the twentieth century is presumably related to the fact that the 1901–88 period from the preceding HadCRUT4 dataset was used as the instrumental target for Anc17. Alternatively, it might imply that the more conservative approach of data screening in HadCRUT5 and CRUTEM5 results in more robust temperature estimates during times with plenty of station records that allow for a more rigorous screening. During the BT period higher correlations are likely the result of the reduced spatial coverage in HadCRUT5 and, particularly, CRUTEM5 compared to BEARTH. Most available HadCRUT5 and CRUTEM5 grid cells in the nineteenth-century cluster over western and northern Europe, a region that is relatively well sampled in both the instrumental and the proxy network. Spatial correlation fields do not reveal the reason for the correlation increase over time. Improving correlations are a result of both increasing correlation values over time (e.g., in Europe) and new, additional strongly correlating grid cells in the twentieth century (e.g., in Alaska).

To further disentangle the effect of spatial coverage on the offset between reconstructed and instrumental BTs, we harmonized coverage to a shared portion of grid cells in the reconstructed and instrumental fields. The masking experiments revealed that a BT difference of 0.10°C between Anc17 and BEARTH results from the unequal spatial coverage between these temperature fields. Northwestern North America and most of northwestern Asia, two regions with particularly cold estimates for BT are almost entirely masked out because CRUTEM5 provides no or few grid cells there. Despite the application of infilling, even HadCRUT5 and BEARTH show no data for many of these regions. But reconstructed and instrumental temperatures correlate well in northwestern North America and northwestern Asia during the twentieth century, indicating that the local reconstructions provide robust temperature estimates during the BT period, too. Instrumental temperatures simply miss this cooling signal in the Arctic, even though infilling helps to reduce the bias from poor spatial sampling (CRUTEM5 versus HadCRUT5). The remaining difference between time series of masked averages from Anc17 and BEARTH during the BT period (after 1860) is of a similar magnitude as differences in the twentieth century and thus is most likely ascribed to the uncertainty in the tree-ring-reconstructed temperature fields.

c. BT for NH land areas are likely to be colder during summer than currently understood

Using reconstructed temperature as an independent estimate for large-scale temperatures in the nineteenth century yields a clear result regarding the likelihood of BT estimates from the different instrumental datasets CRUTEM5, HadCRUT5 and BEARTH. Our findings show that the BEARTH temperature estimate is 0.12°C cooler than CRUTEM5 and is therefore closer to the tree-ring-reconstructed temperatures. HadCRUT5, which uses infilling together with the CRUTEM5 station network, yields a BT estimate in between the BEARTH and the CRUTEM5 value. Our numbers agree well with the offsets reported in Hawkins et al. (2017) although those authors looked at the whole globe and annual temperatures and not only at NH extratropical land temperatures during summer, as we have done here. While the most obvious difference between BEARTH and CRUTEM5 is the increased spatial coverage, this does not fully explain the significant offset between these datasets during the BT period. Apparently, the BEARTH BT estimate is not much affected by grid cells outside Europe and eastern North America. This was the result of masking BEARTH with the CRUTEM5 grid mask. Thus, the lower BT estimates in BEARTH must be strongly influenced by the additional station records included in the gridded product after the more liberal data selection process that does not require prior homogenization or quality control by National Meteorological Services. This approach could, however, potentially introduce more uncertainty. The two most relevant sources of error in early instrumental station records are the exposure bias and the effects of urbanization (Jones 2016). The urbanization effect refers here to the movement of weather stations to sites out of town and not to the additional warming from growing cities that is often discussed for the second half of the twentieth century (Jones et al. 2008). Both biases, exposure and urbanization, have the potential to alter the large-scale mean to more positive values (Brohan et al. 2006; Dienst et al. 2018; Parker 1994; Wickham et al. 2013). If the additional station records entering the BEARTH dataset would be significantly impacted by these biases, a positive deviation from CRUTEM5 in the masking experiment would be more likely. Similarly, additional random, uncorrelated noise, potentially introduced by station records of lower quality, would reduce the magnitude of the BEARTH BT estimate due to noise cancellation. Instead, the additional station records result in more negative values, likely indicating that they are not mainly introducing biases.

A lower BT estimate is supported by the spatial distribution of warmth and cold during the BT period. Some of the coldest reconstructed temperatures are found in regions poorly or not at all covered by the instrumental fields. This accounts mostly for the high latitudes of Asia and North America. Bekryaev et al. (2010) found a significant polar amplification after analyzing long instrumental station records from high latitudes. Although they found the effect to be weaker in summer, tree-ring-reconstructed temperatures seem to support a polar amplification of the nineteenth-century cooling. This underlines the importance of temperature estimates from high latitudes in contributing to large-scale averages. While we could show that the coverage bias explains some of the offset between BEARTH and Anc17 temperatures, it should be noted that other reconstructions might underestimate BT temperature, because their proxy networks are biased toward high-latitude tree line sites and often index reconstructions do not adequately account for the relative spatial representation of the underlying paleoclimate network (Anchukaitis et al. 2017).

In addition, proxy reconstructions are accompanied by uncertainty introduced during proxy selection, methodological choices, and the model calibration processes (Büntgen et al. 2021b; Anchukaitis and Smerdon 2022) explaining in part the large range of proxy-based BT estimates. Spectral biases introduced by biological memory in proxy reconstructions using tree-ring width data (Franke et al. 2013) can result in an overestimation of low-frequency variability and thus too low BT estimates. But even Sch15, a reconstruction built exclusively with the more robust temperature parameter of maximum latewood density (Schneider et al. 2015) yields lower BT estimates than the instrumental data. To reduce the range between reconstructed BT estimates, it is important to construct tree-ring reconstructions in regions that are not well covered in the proxy network (e.g., the eastern Mediterranean). Where centennial old trees are rare even shorter, reconstructions would be helpful in order to improve the robustness particularly during the past 200 years. In this study we did not use simulations from general circulation models to compare with the BT estimates. Other studies have already shown a multimodel ensemble range of approximately 0.5°C for warming estimates in global annual temperatures with HadCRUT5 in the upper middle of the model range (Hawkins et al. 2017). This range is of similar magnitude as the maximum and minimum BT found in this study indicating that confining BT estimates with model simulations is a more complex endeavor.

Although we find that BEARTH is in better agreement with proxy reconstructions, it is important to note that the BEARTH approach is somewhat riskier and could still result in better agreement for the wrong reasons if station records with a negative bias are not corrected adequately in the automated homogenization process. Ideally, the additional station records in the BEARTH dataset should undergo a thorough quality assessment and individual homogenization based on station metadata. This might reduce the offset—and thus the uncertainty—between different instrumental BT estimates. There is also still a lot of potential for data rescue initiatives that collect, digitize, and merge early instrumental observations (Brönnimann et al. 2018; Hawkins et al. 2019; Rennie et al. 2014) to have an impact on the large-scale average of early instrumental temperatures. While we found an infilled dataset to be closest to the reconstructed data, it depends on the research question of whether or not to prefer infilling. With the comparison of CRUTEM5 and HadCRUT5, we could show that infilling can impact the large-scale average to some degree. However, if regional characteristics are of interest, it can be more beneficial to reduce coverage to a common field (Cowtan et al. 2018) as in our masking experiments. Infilled instrumental datasets showed weaker correlation with the reconstructed data, likely due to larger errors at infilled grid cells.

5. Conclusions

The comparison of instrumental summer temperature fields from the NH extratropical landmass with proxy reconstructions suggests that instrumental temperatures overestimate BTs. The BEARTH dataset yields the estimate closest to the reconstructed values. Our analyses showed that in the first decades of the BT period a more liberal selection approach for station records is beneficial to reduce the offset between instrumental and reconstructed temperatures, although this introduces the risk of integrating records of lower quality. Infilling of the instrumental datasets cannot fully account for poor coverage during the BT period. Anomalously cold regions at high latitudes in the reconstructed field—potentially as a result of polar amplification—are not represented in the instrumental datasets, leading to an increased offset between the two independent temperature estimates. Despite considerable spread between different proxy-based temperature reconstructions, they all suggest BTs to be lower than estimated by HadCRUT5 and CRUTEM5. Cooler BTs lead to larger estimates for observed warming, which in turn reduces the probability of reaching the 1.5°C target set in the Paris Agreement. Closer agreement within instrumental data, within reconstructed data, and between instrumental and reconstructed data would reduce the uncertainty associated with early large-scale temperature estimates. We therefore emphasize the importance of recovery and integration of early instrumental data where such data are available. In addition, the tree-ring network should be extended in regions with sparse coverage, even if no trees older than 300+ years are available. Together, these approaches can reduce the gap between reconstructed and instrumental BTs and yield more robust forecasts for future warming rates and climate change impacts.

Acknowledgments. J. E. acknowledges support from SustES (CZ.02.1.01/0.0/0.0/16_019/0000797), ERC (AdG 882727), and the German Research Foundation (ES 161/12.1). K. J. A. is partially supported by a grant from the U.S. National Science Foundation (AGS-2102993).

Data availability statement. The data used in this study are available online or can be provided by the original authors upon request.

REFERENCES

Abram, N. J., and Coauthors, 2016: Early onset of industrial-era warming across the oceans and continents. *Nature*, **536**, 411–418, https://doi.org/10.1038/nature19082.

Allen, M. R., and Coauthors, 2019: Framing and context. *Global Warming of 1.5°C*, V. Masson-Delmotte et al., Eds., Cambridge University Press, 49–92, https://doi.org/10.1017/9781009157940.003.

Anchukaitis, K. J., and J. E. Smerdon, 2022: Progress and uncertainties in global and hemispheric temperature reconstructions of the Common Era. *Quat. Sci. Rev.*, **286**, 107537, https://doi.org/10.1016/j.quascirev.2022.107537.

——, and Coauthors, 2017: Last millennium Northern Hemisphere summer temperatures from tree rings: Part II, spatially resolved reconstructions. *Quat. Sci. Rev.*, **163**, 1–22, https://doi.org/10.1016/j.quascirev.2017.02.020.

Auer, I., R. Böhm, and W. Schöner, 2001: Austrian long-term climate 1767–2000: Multiple instrumental climate time series from central Europe. *Österreichische Beitr. Meteor. Geophys.*, **25**, 1–147, http://www.zamg.ac.at/histalp/download/abstract/Auer-etal-2001b-F.pdf.

Bekryaev, R. V., I. V. Polyakov, and V. A. Alexeev, 2010: Role of polar amplification in long-term surface air temperature variations and modern Arctic warming. *J. Climate*, **23**, 3888–3906, https://doi.org/10.1175/2010JCLI3297.1.

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

Briffa, K. R., T. J. Osborn, F. H. Schweingruber, I. C. Harris, P. D. Jones, S. G. Shiyatov, and E. A. Vaganov, 2001: Low-frequency temperature variations from a northern tree ring density network. *J. Geophys. Res.*, **106**, 2929–2941, https://doi.org/10.1029/2000JD900617.

Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones, 2006: Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *J. Geophys. Res.*, **111**, D12106, https://doi.org/10.1029/2005JD006548.

Brönnimann, S., and Coauthors, 2018: A roadmap to climate data rescue services. *Geosci. Data J.*, **5**, 28–39, https://doi.org/10.1002/gdj3.56.

Büntgen, U., L. Wacker, K. Nicolussi, M. Sigl, D. Güttler, W. Tegel, P. J. Krusic, and J. Esper, 2014: Extraterrestrial confirmation of tree-ring dating. *Nat. Climate Change*, **4**, 404–405, https://doi.org/10.1038/nclimate2240.

——, and Coauthors, 2021a: Recent European drought extremes beyond Common Era background variability. *Nat. Geosci.*, **14**, 190–196, https://doi.org/10.1038/s41561-021-00698-0.

——, and Coauthors, 2021b: The influence of decision-making in tree ring–based climate reconstructions. *Nat. Commun.*, **12**, 3411, https://doi.org/10.1038/s41467-021-23627-6.

Chen, D., and Coauthors, 2021: Framing, context, and methods. *Climate Change 2021: The Physical Science Basis*, V. Masson-Delmotte et al., Eds., Cambridge University Press, 147–286.

Cowtan, K., and R. G. Way, 2014: Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Quart. J. Roy. Meteor. Soc.*, **140**, 1935–1944, https://doi.org/10.1002/qj.2297.

——, P. Jacobs, P. Thorne, and R. Wilkinson, 2018: Statistical analysis of coverage error in simple global temperature estimators. *Dyn. Stat. Climate Syst.*, **3**, dzy003, https://doi.org/10.1093/climsys/dzy003.

Daniel, H., 1973: *One Hundred Years of International Co-Operation in Meteorology (1873–1973): A Historical Review*. Secretariat of the World Meteorological Organization, 53 pp.

D'Arrigo, R., R. Wilson, B. Liepert, and P. Cherubini, 2008: On the 'divergence problem' in northern forests: A review of the tree-ring evidence and possible causes. *Global Planet. Change*, **60**, 289–305, https://doi.org/10.1016/j.gloplacha.2007.03.004.

Dienst, M., J. Lindén, E. Engström, and J. Esper, 2017: Removing the relocation bias from the 155-year Haparanda temperature record in Northern Europe. *Int. J. Climatol.*, **37**, 4015–4026, https://doi.org/10.1002/joc.4981.

——, ——, and J. Esper, 2018: Determination of the urban heat island intensity in villages and its connection to land cover in three European climate zones. *Climate Res.*, **76** (1), 1–15, https://doi.org/10.3354/cr01522.

——, ——, Ò. Saladié, and J. Esper, 2019: Detection and elimination of UHI effects in long temperature records from villages—A case study from Tivissa, Spain. *Urban Climate*, **27**, 372–383, https://doi.org/10.1016/j.uclim.2018.12.012.

Edwards, P. N., 2004: "A vast machine": Standards as social technology. *Science*, **304**, 827–828, https://doi.org/10.1126/science.1099290.

Esper, J., and D. Frank, 2009: Divergence pitfalls in tree-ring research. *Climatic Change*, **94**, 261–266, https://doi.org/10.1007/s10584-009-9594-2.

——, and Coauthors, 2016: Ranking of tree-ring based temperature reconstructions of the past millennium. *Quat. Sci. Rev.*, **145**, 134–151, https://doi.org/10.1016/j.quascirev.2016.05.009.

——, and Coauthors, 2018: Large-scale, millennial-length temperature reconstructions from tree-rings. *Dendrochronologia*, **50**, 81–90, https://doi.org/10.1016/j.dendro.2018.06.001.

Frank, D., U. Büntgen, R. Böhm, M. Maugeri, and J. Esper, 2007: Warmer early instrumental measurements versus colder reconstructed temperatures: Shooting at a moving target. *Quat. Sci. Rev.*, **26**, 3298–3310, https://doi.org/10.1016/j.quascirev.2007.08.002.

Franke, J., D. Frank, C. C. Raible, J. Esper, and S. Brönnimann, 2013: Spectral biases in tree-ring climate proxies. *Nat. Climate Change*, **3**, 360–364, https://doi.org/10.1038/nclimate1816.

Guillet, S., and Coauthors, 2017: Climate response to the Samalas volcanic eruption in 1257 revealed by proxy records. *Nat. Geosci.*, **10**, 123–128, https://doi.org/10.1038/ngeo2875.

Gulev, S. K., and Coauthors, 2021: Changing state of the climate system. *Climate Change 2021: The Physical Science Basis*, V. Masson-Delmotte et al., Eds., Cambridge University Press, 287–422.

Hawkins, E., and Coauthors, 2017: Estimating changes in global temperature since the preindustrial period. *Bull. Amer. Meteor. Soc.*, **98**, 1841–1856, https://doi.org/10.1175/BAMS-D-16-0007.1.

——, S. Burt, P. Brohan, M. Lockwood, H. Richardson, M. Roy, and S. Thomas, 2019: Hourly weather observations from the Scottish Highlands (1883–1904) rescued by volunteer citizen scientists. *Geosci. Data J.*, **6**, 160–173, https://doi.org/10.1002/gdj3.79.

Hegerl, G. C., T. J. Crowley, M. Allen, W. T. Hyde, H. N. Pollack, J. Smerdon, and E. Zorita, 2007: Detection of human influence on a new, validated 1500-year temperature reconstruction. *J. Climate*, **20**, 650–666, https://doi.org/10.1175/JCLI4011.1.

IPCC, 2013: Summary for policymakers. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 1–29.

Jehn, F. U., M. Schneider, J. R. Wang, L. Kemp, and L. Breuer, 2021: Betting on the best case: Higher end warming is underrepresented in research. *Environ. Res. Lett.*, **16**, 084036, https://doi.org/10.1088/1748-9326/ac13ef.

Jones, P. D., 2016: The reliability of global and hemispheric surface temperature records. *Adv. Atmos. Sci.*, **33**, 269–282, https://doi.org/10.1007/s00376-015-5194-4.

——, D. H. Lister, and Q. Li, 2008: Urbanization effects in large-scale temperature records, with an emphasis on China. *J. Geophys. Res.*, **113**, D16122, https://doi.org/10.1029/2008JD009916.

——, ——, T. J. Osborn, C. Harpham, M. Salmon, and C. P. Morice, 2012: Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010. *J. Geophys. Res.*, **117**, D05127, https://doi.org/10.1029/2011JD017139.

Kadow, C., D. M. Hall, and U. Ulbrich, 2020: Artificial intelligence reconstructs missing climate information. *Nat. Geosci.*, **13**, 408–413, https://doi.org/10.1038/s41561-020-0582-5.

Kennedy, J. J., 2014: A review of uncertainty in in situ measurements and data sets of sea surface temperature. *Rev. Geophys.*, **52** (1), 1–32, https://doi.org/10.1002/2013RG000434.

Kent, E. C., and Coauthors, 2017: A call for new approaches to quantifying biases in observations of sea surface temperature.

Bull. Amer. Meteor. Soc., **98**, 1601–1616, https://doi.org/10.1175/BAMS-D-15-00251.1.

King, J. M., K. J. Anchukaitis, J. E. Tierney, G. J. Hakim, J. Emile-Geay, F. Zhu, and R. Wilson, 2021: A data assimilation approach to last millennium temperature field reconstruction using a limited high-sensitivity proxy network. *J. Climate*, **34**, 7091–7111, https://doi.org/10.1175/JCLI-D-20-0661.1.

Kirtman, B., and Coauthors, 2013: Near-term climate change: Projections and predictability. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 953–1028.

Knerr, I., M. Dienst, J. Lindén, P. Dobrovolný, J. Geletic, U. Büntgen, and J. Esper, 2019: Addressing the relocation bias in a long temperature record by means of land cover assessment. *Theor. Appl. Climatol.*, **137**, 2853–2863, https://doi.org/10.1007/s00704-019-02783-2.

Knutti, R., J. Rogelj, J. Sedláček, and E. M. Fischer, 2016: A scientific critique of the two-degree climate change target. *Nat. Geosci.*, **9**, 13–18, https://doi.org/10.1038/ngeo2595.

Ljungqvist, F. C., and Coauthors, 2020: Ranking of tree-ring based hydroclimate reconstructions of the past millennium. *Quat. Sci. Rev.*, **230**, 106074, https://doi.org/10.1016/j.quascirev.2019.106074.

Menne, M. J., C. N. Williams, B. E. Gleason, J. J. Rennie, and J. H. Lawrimore, 2018: The Global Historical Climatology Network monthly temperature dataset, version 4. *J. Climate*, **31**, 9835–9854, https://doi.org/10.1175/JCLI-D-18-0094.1.

Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones, 2012: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *J. Geophys. Res.*, **117**, D08101, https://doi.org/10.1029/2011JD017187.

——, and Coauthors, 2021: An updated assessment of near-surface temperature change from 1850: The HadCRUT5 data set. *J. Geophys. Res. Atmos.*, **126**, e2019JD032361, https://doi.org/10.1029/2019JD032361.

Oke, T. R., 1973: City size and the urban heat island. *Atmos. Environ.*, **7**, 769–779, https://doi.org/10.1016/0004-6981(73)90140-6.

Osborn, T. J., P. D. Jones, D. H. Lister, C. P. Morice, I. R. Simpson, J. P. Winn, E. Hogan, and I. C. Harris, 2021: Land surface air temperature variations across the globe updated to 2019: The CRUTEM5 data set. *J. Geophys. Res. Atmos.*, **126**, e2019JD032352, https://doi.org/10.1029/2019JD032352.

PAGES 2k Consortium, 2019: Consistent multidecadal variability in global temperature reconstructions and simulations over the Common Era. *Nat. Geosci.*, **12**, 643–649, https://doi.org/10.1038/s41561-019-0400-0.

Parker, D. E., 1994: Effects of changing exposure of thermometers at land stations. *Int. J. Climatol.*, **14** (1), 1–31, https://doi.org/10.1002/joc.3370140102.

Rennie, J. J., and Coauthors, 2014: The International Surface Temperature Initiative Global Land Surface Databank: Monthly temperature data release description and methods. *Geosci. Data J.*, **1**, 75–102, https://doi.org/10.1002/gdj3.8.

Rizwan, A. M., L. Y. C. Dennis, and C. Liu, 2008: A review on the generation, determination and mitigation of urban heat island. *J. Environ. Sci.*, **20**, 120–128, https://doi.org/10.1016/S1001-0742(08)60019-4.

Rohde, R. A., and Z. Hausfather, 2020: The Berkeley Earth land/ocean temperature record. *Earth Syst. Sci. Data*, **12**, 3469–3479, https://doi.org/10.5194/essd-12-3469-2020.

——, and Coauthors, 2013a: A new estimate of the average Earth surface land temperature spanning 1753 to 2011. *Geoinf. Geostat.*, **1** (1), 1–7, http://doi.org/10.4172/2327-4581.1000101.

——, and Coauthors, 2013b: Berkeley Earth temperature averaging process. *Geoinf. Geostat.*, **1**, 20–100, https://doi.org/10.4172/2327-4581.1000103.

Schneider, L., J. E. Smerdon, U. Büntgen, R. J. S. Wilson, V. S. Myglan, A. V. Kirdyanov, and J. Esper, 2015: Revising midlatitude summer temperatures back to AD 600 based on a wood density network. *Geophys. Res. Lett.*, **42**, 4556–4562, https://doi.org/10.1002/2015GL063956.

Schurer, A. P., G. C. Hegerl, M. E. Mann, S. F. B. Tett, and S. J. Phipps, 2013: Separating forced from chaotic climate variability over the past millennium. *J. Climate*, **26**, 6954–6973, https://doi.org/10.1175/JCLI-D-12-00826.1.

——, M. E. Mann, E. Hawkins, S. F. B. Tett, and G. C. Hegerl, 2017: Importance of the pre-industrial baseline for likelihood of exceeding Paris goals. *Nat. Climate Change*, **7**, 563–567, https://doi.org/10.1038/nclimate3345.

Schweingruber, F. H., T. Bartholin, E. Schaur, and K. R. Briffa, 1988: Radiodensitometric-dendroclimatological conifer chronologies from Lapland (Scandinavia) and the Alps (Switzerland). *Boreas*, **17**, 559–566, https://doi.org/10.1111/j.1502-3885.1988.tb00569.x.

Simmons, A. J., K. M. Willett, P. D. Jones, P. W. Thorne, and D. P. Dee, 2010: Low-frequency variations in surface atmospheric humidity, temperature, and precipitation: Inferences from reanalyses and monthly gridded observational data sets. *J. Geophys. Res.*, **115**, D01110, https://doi.org/10.1029/2009JD012442.

St. George, S., and J. Esper, 2019: Concord and discord among Northern Hemisphere paleotemperature reconstructions from tree rings. *Quat. Sci. Rev.*, **203**, 278–281, https://doi.org/10.1016/j.quascirev.2018.11.013.

Stoffel, M., and Coauthors, 2015: Estimates of volcanic-induced cooling in the Northern Hemisphere over the past 1500 years. *Nat. Geosci.*, **8**, 784–788, https://doi.org/10.1038/ngeo2526.

Trewin, B., 2010: Exposure, instrumentation, and observing practice effects on land temperature measurements. *Wiley Interdiscip. Rev.: Climate Change*, **1**, 490–506, https://doi.org/10.1002/wcc.46.

Vaccaro, A., J. Emile-Geay, D. Guillot, R. Verna, C. Morice, J. Kennedy, and B. Rajaratnam, 2021: Climate field completion via Markov random fields: Application to the HadCRUT4.6 temperature dataset. *J. Climate*, **34**, 4169–4188, https://doi.org/10.1175/JCLI-D-19-0814.1.

Vose, R. S., and Coauthors, 2021: Implementing full spatial coverage in NOAA's global temperature analysis. *Geophys. Res. Lett.*, **48**, e2020GL090873, https://doi.org/10.1029/2020GL090873.

Wickham, C., R. Rohde, R. Muller, J. Wurtele, and J. Mosher, 2013: Influence of urban heating on the global temperature land average using rural sites identified from MODIS classifications. *Geoinf. Geostat.*, **1** (2), http://doi.org/10.4172/2327-4581.1000104.

Wilson, R., and Coauthors, 2016: Last millennium Northern Hemisphere summer temperatures from tree rings: Part I: The long term context. *Quat. Sci. Rev.*, **134**, 1–18, https://doi.org/10.1016/j.quascirev.2015.12.005.

WMO, 1966: Climatic change. Tech. Note 79, WMO-195, TP 100, 80 pp.

Zhang, H., and Coauthors, 2018: East Asian warm season temperature variations over the past two millennia. *Sci. Rep.*, **8**, 7702, https://doi.org/10.1038/s41598-018-26038-8.