Comparing the current and early 20th century warm periods in China

Willie Wei-Hock Soon[1], Ronan Connolly*[1], Michael Connolly[2], Peter O'Neill[3], Jingyun Zheng[4], Quansheng Ge[4], Zhixin Hao[4], and Hong Yan[5]

[1]*Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA*

[2]*Independent scientists, Dublin, Ireland*

[3]*School of Mechanical and Materials Engineering (retired), University College Dublin, Ireland*

[4]*Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China*

[5]*State Key Laboratory of Loess and Quaternary Geology, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an 710061, China*

ABSTRACT

Most estimates of Chinese regional Surface Air Temperatures since the late-19th century have identified two relatively warm periods – 1920s-40s and 1990s-present. However, there is considerable debate over how the two periods compare to each other. Some argue the current warm period is much warmer than the earlier warm period. Others argue the earlier warm period was comparable to the present. In this collaborative paper, including authors from both camps, the reasons for this ongoing debate are discussed. Several different estimates of Chinese temperature trends, both new and previously published, are considered. A study of the effects of urbanization bias on Chinese temperature trends was carried out using the new updated version of the Global Historical Climatology Network (GHCN) – version 4 (currently in beta production). It is shown that there are relatively few rural stations with long records, but urbanization bias artificially makes the early warm period seem colder and the recent warm period seem warmer. However, current homogenization approaches (which attempt to reduce non-climatic biases) also tend to have similar effects, making it unclear whether reducing or increasing the relative warmth of each period is most appropriate. A sample of 17 Chinese temperature proxy series (12 regional and 5 national) is compared and contrasted specifically for the period since the 19th century. Most proxy series imply a warm early-20th century period and a warm recent period, but the relative warmth of these two periods differs between proxies. Also, with some proxies, one or other of the warm periods is absent.

Key words: China regional temperatures • Global warming • Urbanization bias • Global Climate Models • Temperature proxies • Early 20th Century Warm Period.

Authors draft version
Accepted for publication in Earth-Science Reviews: 18 May 2018
Journal version available at https://doi.org/10.1016/j.earscirev.2018.05.013

---

*Corresponding author : Dr. Ronan Connolly
Email: ronanconnolly@yahoo.ie

Contents

1. Introduction
2. Data sources and methodology
2.1. Previous regional Chinese temperature reconstructions using meteorological records
2.2. Data for new Chinese temperature reconstructions
2.3. CRUTEM3 and CRUTEM4 estimates
2.4. CMIP5 Global Climate Model hindcasts for Chinese region
2.5. 20-CR hindcast for Chinese region
2.6. Temperature proxy reconstructions for Chinese region
3. Results and discussion
3.1. The representativeness problem
3.1.1. Spatial sampling challenges
3.1.2. Comparing thermometer data to other estimates, i.e., model output and proxy data
3.1.3. Annual mean surface air temperatures vs. other metrics
3.2. The problem of non-climatic biases
3.2.1. The urbanization bias debate
3.2.2. The debate over homogenization approaches
3.2.3. The blending problem of current homogenization techniques
3.2.4. Demonstration of urban blending during the homogenization of 10 stations near Beijing
3.2.5. Some approaches to minimizing the blending problem of homogenization
3.3. Construction of new relatively-rural estimate of Chinese temperature trends
3.4. Comparison of different estimates of Chinese temperature trends
3.5. Comparison of temperature proxy reconstructions for Chinese region
4. Conclusions

1. Introduction

Over the last two decades, there have been several different attempts to estimate regional Surface Air Temperature (SAT) trends for China since the start of the 20th century (or earlier) (Wang S. et al., 2001, 2004; Tang & Ren, 2005; Tang et al., 2010; Ren et al., 2012, 2017; Cao et al., 2013; Ding et al., 2014; Wang J. F. et al., 2014; Soon et al., 2015; Li et al., 2017). All of these estimates identify two relatively warm periods (1920s-40s and 1990s-present) separated by a relatively cool period (1950s-70s). However, there has been considerable debate over how these two warm periods compare to each other.

Recently, Li et al. (2017) found that the early 20th century warm period (1920s-40s) only involved a relatively modest warming and that the current warm period is the hottest on record by a substantial amount. This agreed with the earlier studies by Cao et al. (2013) and Wang J. F. et al. (2014), as well as some temperature proxy-based studies, e.g., (Ding et al., 2016; Liu et al., 2017; Zheng et al., 2017).

On the other hand, several studies have suggested that both warm periods have been comparable, albeit with the current warm period being warmer on average (Wang S. et al., 2001, 2004; Tang & Ren, 2005; Tang et al., 2010; Ren et al., 2012, 2017; Ding et al., 2014). Meanwhile, Soon et al. (2015) came to the opposite conclusion of Li et al. (2017) and found that the early 20th century warm period was the hottest on record for China.

In this collaborative paper, each of us has different views on this contentious issue. Specifically, while some of us have argued that the early 20th century warm period was comparable to the recent warm period for China (e.g., Soon et al., 2011; Soon et al., 2015), some of us have argued that the recent warm period is much warmer (e.g., Ding et al., 2016; Liu et al., 2017; Zheng et al., 2017). Therefore, we believe it is important to establish and assess the reasons for these differing views.

Several challenging, inter-related factors seem to be involved:

1. There are relatively few Chinese stations with temperature records beginning before 1954, i.e., the period during which the early 20th century warm period occurred.
2. Moreover, the methods by which daily temperatures were estimated are especially poorly documented for the earlier periods, e.g., instruments used, times-of-observation. Therefore, it is plausible that changes in these methods may have introduced non-climatic biases into the estimates of the warmth of the earlier period.
3. On the other hand, there is considerable evidence that, in recent decades, many instrumental records in China have been affected by warming biases caused by urbanization. So, urbanization bias may have artificially inflated the apparent warmth of the recent period. This would also have the effect of artificially decreasing the relative warmth of the early period. However, determining the magnitude of this effect has been quite contentious. Some argue that it has only had a very small or negligible effect (Li et al., 2004, 2010a, 2017; Jones et al., 1990; Yan et al., 2016; Wang J. & Yan, 2016; Wang J. et al., 2017). Others argued it has

had a substantial effect (Soon et al., 2015; Ren, 2015; Ren et al., 2015, 2017; Sun et al., 2016).

4. The main homogenization approaches currently applied in an attempt to reduce the effects of non-climatic biases have a tendency to reduce the warmth of the early period and increase the warmth of the recent period. This has led several groups to conclude that the apparent warmth of the early period is mostly due to non-climatic biases, e.g., Li et al. (2017). On the other hand, Soon et al. (2015) note that the current homogenization approaches lead to "urban blending" when applied to a highly urbanized station network. That is, the homogenization process "aliases" (deGaetano, 2006; Pielke et al., 2007a) a fraction of the urbanization bias of urban neighbours onto the records of less urbanized station. This blending problem would have a tendency to artificially increase the warmth of the recent period and decrease the warmth of the early period.
5. A somewhat independent approach to estimating the relative warmth of the two periods could be to use temperature proxies, such as tree rings and ice cores (Wang S. et al., 2001, 2004; Ding et al., 2016; Liu et al., 2017; Zheng et al., 2017). Most temperature proxies for China agree that the 20th and 21st centuries are relatively warm compared with the 18th and 19th centuries, and usually identify both a warm early-20th century period and a warm recent period. However, the relative warmth of these two periods differs between proxies.

In this paper, we provide a review of the above challenges. We also compare and contrast the early and recent warm periods for each of the different Chinese temperature trend estimates. Additionally, we present new estimates derived from the updated Global Historical Climatology Network temperature dataset (Lawrimore et al., 2011). The latest updates to this dataset (version 4, currently in beta production) have substantially increased the amount of publicly archived temperature records for China for both the recent warm period and the early 20th century warm period. The first assessments of the magnitude of urbanization bias for China in this new dataset are also carried out. Finally, we present a new compilation of various different proxy reconstructions representative of three regions in China (central China, northeast China and the Tibetan Plateau) as well as five multiproxy reconstructions for all of China. Previous assessments of the temperature proxy data have tended to primarily focus on the points of agreement between individual series. However, given the complexity of the debate described in the rest of the paper, we believe it is important to both compare **and** contrast the similarities and differences of the proxy data for the post-19th century period.

2. Data sources and methodology

2.1. Previous regional Chinese temperature reconstructions using meteorological records

In this paper, we will consider several different previously-published estimates of Chinese temperature trends, derived from meteorological records. Many of these estimates are constructed from similar data sources, and hence, they are not entirely independent. However, each of the estimates has taken a slightly different approach to using this data, and there are also some differences in the data sources.

A brief overview of each of these estimates is given below:

- **Wang et al. (2004)**. Wang S. et al. (2001, 2004) divided the country into ten different regions. They then used a combination of raw station records, annual regional climatologies and some temperature proxy records to construct temperature series for each of the ten regions. These regional series were then averaged to yield a single time series for all of China covering the period 1880-2002. Recently, Ren et al. (2017) updated this series to 2015, and we obtained the updated series by digitizing Figure 1 of Ren et al. (2017). Wang et al. (2001, 2007) have also extended this series back in time using temperature proxies – see Section 2.5.
- **Tang & Ren (2005)**. Tang & Ren (2005) noted that there was no unified observation time for China before 1950, and that inconsistencies in observation times could have introduced non-climatic biases into station records. In an attempt to reduce the magnitude of these Time of Observation Biases (TOB), they averaged together the records for monthly maximum temperatures and monthly minimum temperatures from 616 stations, instead of using the monthly mean averages. Their original series spanned the period 1905-2001, but this series has since been updated by both Tang et al. (2010) and more recently Ren et al. (2017). We obtained this series by digitizing Figure 1 of Ren et al. (2017).
- **China Climate Change Monitoring Bulletin, 2014**. In 2014, the China Meteorological Agency published an estimate of Chinese temperature trends for 1901-2013.

We obtained this series by digitizing Figure 1 of Ding et al. (2014).

- **Soon et al. (2015) "mostly rural composite"**. Soon et al. (2015) analysed all 417 Chinese stations in version 3 of the Global Historical Climatology Network (GHCN) dataset. Soon et al. noted that only 30% of these stations are still rural, and that these rural stations had relatively short and incomplete station records. During the 1951-1990 period (when all stations had relatively complete station records), a strong urbanization bias was apparent in the non-rural station records in that the temperature trend increased from +0.025°C/decade for the rural subset to +0.088°C/decade for the moderately urbanized subset +0.119°C/decade for the most heavily urbanized subset. This suggested that ~86% of the warming trend of the full dataset (+0.109°C/decade) for this period was urbanization bias. Similar results were obtained whether using the raw GHCN dataset or the homogenized dataset, indicating that homogenization had failed to remove this urbanization bias. Moreover, outside this period, they found indications that the (automated) homogenization process had introduced "urban blending" into the homogenized dataset, i.e., warming trends had been artificially inserted into the (relatively rare) rural station records so that their trends better matched those of their more numerous urban neighbours. For these reasons, Soon et al. intentionally used the non-homogenized dataset and preferentially used rural station records to minimise the effects of urbanization bias. For the 1951-1990 period, their reconstruction only used rural stations. However, because very few of the rural stations had data outside this period, for the 1907-1950 and 1991-2014 periods, they were only able to remove the most heavily urbanized subset and therefore included some stations that were moderately urbanized. All of the stations with data before 1907 were in the most heavily urbanized subset, but it was argued that most of the urbanization bias in these records has probably occurred in more recent decades. Therefore, to extend the length of their reconstruction, for the 1841-1906 period they included all available stations, regardless of how urbanized they were.
- **Li & Xu (adjusted)**. Li et al. (2017) reached different conclusions from Soon et al. (2015) and therefore used a different approach for constructing their series. Li et al. estimated that the average magnitude of urbanization bias in their homogenized dataset is quite small (Li et al., 2004, 2010a; Jones et al., 2008; Xu et al., 2017). They therefore did not attempt to explicitly remove any urbanization bias. They also argued that the homogenization process should have reduced the magnitude of other non-climatic biases, e.g., station moves, changes in instrumentation. They also applied Tang & Ren (2005)'s approach of averaging together the separate maximum and minimum monthly series to try to reduce Time of Observation Biases. The Li & Xu (adjusted) series therefore is the homogenized regional series of Li et al. (2010) and Xu et al. (2013), updated by Li et al. (2017) to cover the period 1900-2015.
- **Li & Xu (raw)**. For comparison, Li et al. (2017) also repeated their analysis using the non-homogenized version of their dataset. Both of the Li & Xu series were provided by personal communication with Li.

Many of the time series studied in this paper were calculated using different baseline periods, e.g., 1961-1990, 1971-2000, etc. Therefore, to allow direct inter-comparisons between time series, for the purposes of this paper, all of the above series were rescaled into temperature anomalies relative to the same baseline period, 1901-2000.

2.2. Data for new Chinese temperature reconstructions

As discussed above, Soon et al. (2015) developed a "mostly rural composite" of Chinese temperatures since 1841 using version 3 of the Global Historical Climatology Network (GHCN) dataset. One limitation of this dataset was that more than 80% of the 417 Chinese station records finished in 1990 (particularly rural stations). However, in recent years, there have been considerable efforts to recover, digitize and compile more temperature records, especially for the early 20th century, e.g., Williamson et al. (2017). In particular, the International Surface Temperature Initiative (ISTI) project (Rennie et al., 2014) and the updates to the Global Historical Climatology Network daily dataset (Menne et al., 2012) have substantially increased the amount of publicly archived temperature data – both globally and specifically for China.

At any rate, version 4 of the GHCN dataset (in beta version, at the time of writing) has updated many of these stations to include post-1990 data and also includes some additional stations. This new dataset is mostly based on the recent International Surface Temperature Initiative (ISTI) dataset – see Rennie et al. (2014), but also incorporates new data from the updated "daily" version of the GHCN (Menne et al., 2012) and various other sources. In total there are 494

Chinese stations, and nearly 75% of these stations have at least some post-1990 data.

Therefore, it is worth repeating the Soon et al. (2015) analysis using version 4 of the GHCN dataset. However, while version 3 of the GHCN data provides two different estimates of how urbanized each of the stations is (one based on population and the other based on night light brightness), there is no such information for version 4.

With that in mind, we have estimated how urbanized each of the stations is by comparing the station co-ordinates to two different maps estimating urbanization:

1. The average population density associated with the station location
2. The average night brightness associated with the station location

To estimate the average population densities, we used the "Gridded Population of the World" (GPW) version 4 dataset[2]. To estimate the average night brightnesses, we used the "Global Radiance Calibrated Nighttime Lights for 2006" dataset[3]. The average values for each station location were determined from the mean average of the nine pixel values centred at the station location, of the appropriate datasets for both population density and night brightness.

Establishing which thresholds to use to define whether a station is "urban" or not can be somewhat arbitrary. Therefore, we simply ranked all the stations from "most urban" to "least urban" according to each of the two metrics. As can be seen from Figure 1, stations that have a particular ranking according to one metric tend to have a similar ranking according to the other metric. Therefore, the overall ranking for each station is simply the average of the two ranks. All of the Chinese stations are then grouped into 5 equal subsets according to this overall ranking, with Subset 1 corresponding to the 20% most urbanized stations and Subset 5 corresponding to the 20% least urbanized stations – see Figure 1.

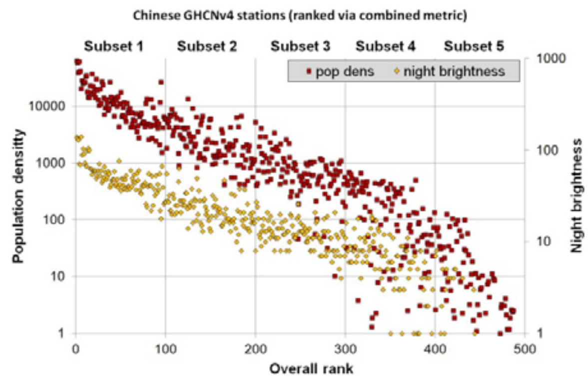

Figure 1. The night brightness and population densities associated with all Chinese stations in the GHCN version 4 dataset, ranked according to urbanization. The y-axes are both shown using log-scale.

We note that this approach to grouping the stations is explicitly based on the degree of urbanization of the current location of the stations. However, many of the stations have been moved over time, and in particular there has been a tendency in China to occasionally move stations that have become heavily urbanized to more rural locations (Li et al., 2017; Ren et al., 2007, 2017; Yan Z.W. et al., 2016; Wang J. & Yan, 2016; Yang et al., 2013; Shi T. et al., 2015; Zhang L. et al., 2014; Zhang Y. & Ren, 2014). Therefore, this approach may identify some stations as being less urban than they originally were. A more comprehensive approach would be to use station histories (a.k.a. "station metadata") to account for these station moves. The Chinese Meteorological Administration (CMA) have internal access to station histories for many of their stations, and as will be discussed in Section 3.2, they incorporate this information into their homogenization procedure. However, such information was not available to us, at the time of writing.

In Section 3, we will use these subsets to study the influence of urbanization on estimates of Chinese temperature trends for version 4 of the GHCN dataset, and also to construct a new "relatively rural composite" analogous to that from Soon et al. (2015).

Because version 4 of the GHCN has increased the number of available Chinese stations, and also updated many of the station records which had previously finished in 1990, it is important to separately investigate what effects (if any) the transition between versions 3 and 4 of the datasets has on estimates of Chinese temperature trends.

[2] Gridded Population of the World, Version 4 (GPWv4): Population Density Adjusted to Match 2015 Revision of UN WPP, Year of Estimate: 2015. Downloaded on December 5th, 2016 as "gpw-v4-population-density-adjusted-to-2015-unwpp-country-totals-2015.zip" from http://sedac.ciesin.org/data/set/gpw-v4-population-density-adjusted-to-2015-unwpp-country-totals/data-download (NASA EARTHDATA login required)

[3] Global Radiance Calibrated Nighttime Lights for 2006 dataset was downloaded on March 19th, 2015, from http://www.ngdc.noaa.gov/dmsp/download_radcal.html as "F162006.v4.tar". Note that as of August 4th 2017, the previous link no longer works, but a different version of this dataset appears at a moved website, https://www.ngdc.noaa.gov/eog/dmsp/download_radcal.html.

Also, both version 3 and version 4 of the GHCN provide two different datasets. The first dataset contains the unadjusted raw station records (with some minor quality control adjustments), while the second dataset has been homogenized using the Menne & Williams (2009) automated homogenization algorithm. As will be discussed in Section 3.2, there is ongoing debate over whether such homogenization procedures improve (Li et al., 2017; Menne & Williams, 2009) or reduce (Soon et al., 2015) the reliability of the data. There is also debate over whether homogenization increases (Ren et al., 2007; Yang et al., 2013; Shi T. et al., 2015) or decreases (Yan et al., 2016; Wang J. & Yan, 2016; Yan Z. W. et al., 2010; Wang J. et al., 2013) the apparent magnitude of urbanization biases.

Therefore, in this paper, to study the effects of (a) homogenization vs non-homogenization and (b) switching between versions 3 and 4 of the GHCN datasets, we will also generate an additional four time series:

1. GHCN version 3 (raw, i.e., non-homogenized)
2. GHCN version 3 (homogenized)
3. GHCN version 4 (raw)
4. GHCN version 4 (homogenized)

For these four time series, we will use all available Chinese stations in the respective datasets, whether rural or urban. For comparison with Soon et al. (2015) [9], we use the same datasets for version 3 which were downloaded from ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/ on 9th January 2015. The version 4 datasets were downloaded from the same website on 16th February 2017.

To generate each of the time series described in this section, we adopt a similar approach to that in Soon et al. (2015), i.e.,

- All station records are converted into temperature anomaly records relative to their 1961-1990 average, and then assigned into 5° × 5° grid boxes. Stations require at least 15 years of complete data for this 1961-1990 period to be included in the analysis.
- For each year, the temperature anomaly for each grid box is the mean anomaly of all stations with 12 months of data for that year, in that grid box.
- The temperature anomaly for China for that year is calculated from the area-weighted average of all grid boxes with data. Area-weighting is approximated by taking the cosine of the mid-latitude of each grid box.

Although Soon et al. (2015) did not present error bars with their analysis, we include them here as twice the standard error of the means. Also, for comparison with the other time series, after generating each series, we rescale them into temperature anomalies relative to the 1901-2000 average.

2.3. CRUTEM3 and CRUTEM4 estimates

Another widely-used temperature dataset is that of the UK-based Climate Research Unit (CRU). Version 3 of this dataset (CRUTEM3) was described by Brohan et al. (2006) [18], and version 4 (the latest version, CRUTEM4) was described by Jones et al. (2012) [19].

CRUTEM3 has about 160 Chinese stations (for some of the stations, the CRUTEM3 inventory file did not include country codes, but we manually identified some extra stations in the Chinese region using the provided station co-ordinates), but only 102 of these stations had at least 15 complete years of data in the 1961-1990 anomaly period. Therefore, for our CRUTEM3 analysis, we reduced the required number of years in the anomaly period to 5. This provided us with 151 stations.

CRUTEM4 has a much larger number of stations for China (703) than CRUTEM3, and 667 of these stations have at least 15 complete years of data in the 1961-1990 anomaly period. Therefore, for our CRUTEM4, we used the same requirement for at least 15 years in the anomaly period, as we did for the GHCN dataset.

We downloaded the CRUTEM3 station data from the CRU's website: https://crudata.uea.ac.uk/cru/data/crutem3/station-data/ [Last accessed 25/10/2017]. We note that the station co-ordinates for some non-Chinese stations have been updated since the station data was first published in 2010: https://www.metoffice.gov.uk/hadobs/crutem3/jan_2010_update.html. We downloaded the CRUTEM4 station data from the UK Met Office's website: https://www.metoffice.gov.uk/hadobs/crutem4/data/download.html [Last accessed 25/10/2017].

2.4. CMIP5 Global Climate Model hindcasts for Chinese region

In preparation for the IPCC's 5th Assessment Report (Bindoff et al., 2013), climate modelling groups from around the world were invited to submit the results of their Global Climate Model (GCM) simulations through the 5th phase of the Coupled Model Intercomparison Project (CMIP5) (Taylor et al., 2012). This was an update to the earlier CMIP3 phase which had been used for the IPCC's 4th Assessment Report in 2007. One component of the CMIP5 submissions involved hindcasts of 20th century climate trends from 1900 to 2005.

As part of their analysis, Li et al. (2017) calculated the regional temperature trends for China from these hindcasts for all 41 of the CMIP5 models. They then calculated the multimodel average of these 41 hindcasts. These regional hindcasts were provided by personal communication with Li. We use the multimodel average as representative of the CMIP5 hindcasts (rescaled relative to its 1901-2000 average).

2.5. 20-CR hindcast for Chinese region

The 20th Century Reanalysis (20-CR) dataset is an alternative 20th century hindcast that was generated by running the atmosphere-land component of a Global Climate Model using similar CO_2, solar and volcanic forcings to those of the CMIP5 hindcasts. The sea surface temperatures and sea ice estimates of the HadISST temperature dataset were used as boundary conditions for the oceanic component. However, rather than being an entirely model-generated hindcast, after each 6-hourly time step, the model outputs were adjusted so that the *modelled surface pressures* better matched the *observed surface pressures* at weather stations for that date (Compo et al., 2011).

Since the 20-CR dataset incorporates some observational data into its output, its hindcasted temperature estimates should better reflect observations than the estimates from a pure GCM hindcast. Hence, it provides a useful semi-empirical intermediary between the GCM hindcasts and the various observational datasets (Compo et al., 2011, 2013; Li et al., 2017). Therefore, we include the 20-CR surface temperature estimates for the Chinese region, as calculated by Li et al. (2017) which was provided by personal communication with Li. This data from the 20-CR dataset was originally provided by NOAA/OAR/ESRL PSD, Boulder, Colorado, USA (accessible from their website at http://www.esrl.noaa.gov/psd/). As before, we have rescaled this time series relative to its 1901-2000 average, for comparison with the other series.

2.6. Temperature proxy reconstructions for Chinese region

A large number of temperature proxy reconstructions have been made for various parts of China, e.g., see (Wang S. et al., 2001, 2007; Shi F. et al., 2012; Yang B. et al., 2002; Ge et al., 2013b, 2017) and references therein. A detailed analysis of all of these proxy series is beyond the scope of this paper. Instead, we will confine ourselves to analysing typical examples of published temperature proxies from different parts of China. Specifically, we have chosen three regions with a relatively large number of temperature proxy series: i) central China; ii) northeast China; and iii) the Tibetan Plateau. We have then (somewhat arbitrarily) selected four temperature proxy series for each of these regions. We have tried to ensure that these series include a mixture of proxy type (e.g., tree rings, ice cores, speleothems) and seasons for each region.

The temperature proxy series we used were taken from the following studies:

Central China

- Chen et al. (2014)
- Ge et al. (2003)
- Tan et al. (2013)
- Yi et al. (2012)

Northeast China:

- Chu et al. (2011)
- Wiles et al. (2014) - note technically this is a northeast Asia proxy.
- Zhu et al. (2015)
- Zhu et al. (2016)

Tibetan Plateau:

- Fan et al. (2010)
- Shi et al. (2016)
- Wang et al. (2015)
- Yang et al. (2008)

We also compiled five multiproxy reconstructions for all of China:

- Shi et al. (2012)
- Wang et al. (2007)
- Yang et al. (2002)
- Ge et al. (2017) – Principal Component Regression (PCR) reconstruction
- Ge et al. (2017) – Partial Least Squares (PLS) reconstruction

Since the focus of this paper is mainly on temperature trends since the 19th century, we only consider the post-1800 portions of these reconstructions. Again, we have rescaled each of these series relative to their 1901-2000 average values.

3. Results and discussion

3.1. The representativeness problem

3.1.1. Spatial sampling challenges

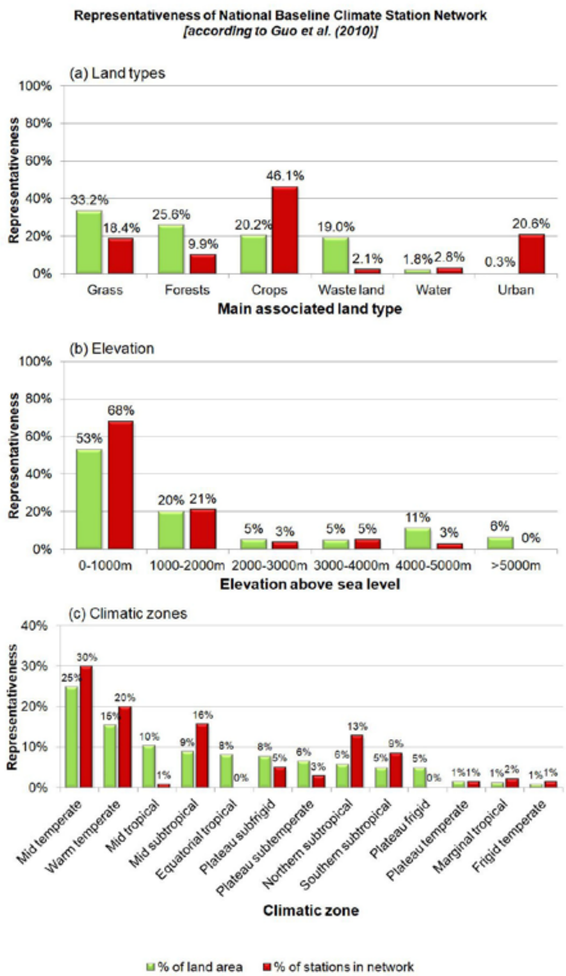

Figure 2. Analysis of the spatial representativeness of the 143 stations in the National Baseline Climate Station Network dataset for China, as determined by Guo et al. (2010). Data was digitized or transcribed from Guo et al., (2010). For each category, the left (green) columns correspond to the actual % of China's land surface, while the right (red) columns correspond to the % of stations in the network.

Figure 2 presents an analysis of the spatial representativeness of the 143 stations in the National Baseline Climate Station Network dataset for China, as determined by Guo et al. (2010). Although most of the different climatic regimes are represented by stations in this dataset, several of the regimes are significantly under-represented (or even have no stations). For instance, 59% of the Chinese land area is currently mostly grass land or forestry, but only 28% of the stations in the dataset are from these types of land. Similarly, 27% of the land area is at an elevation of greater than 2000m, but only 11% of the stations are from those elevations.

Other regimes are significantly over-represented, e.g., 67% of the stations in the dataset are associated with either crop land or urban areas, yet this only represents 20% of the actual land area. In particular, urban areas represent less than 1% of the land area (0.3%) yet 20.6% of the stations in the dataset. As will be discussed in Section 3.2, urban stations often report more warming than rural stations due to urbanization bias, so this is especially problematic.

In Soon et al. (2015), three of us [WS, RC & MC] attempted to address the overrepresentation of urban stations by constructing a Chinese time series using only stations that are rural or mostly rural. We repeat this approach in this paper (Section 3.3). However, while urban areas currently represent only a small fraction of China, this fraction is increasing over time. Also, the local trends due to expanding urban heat islands are genuine climatic changes, and the majority of the population currently lives in urban areas.

Therefore, another method is to include urban stations, but reduce their weighting accordingly. This is the approach taken by one of us [QG] in a series of studies to account for the over-representation of urban stations in thermometer-based estimates (Wang F. & Ge, 2012; Ge et al., 2013; Wang F. et al., 2015). The weighting given by Ge et al. (2013; Wang F. et al., 2015) to the urban areas is very small - although slightly higher at 0.7% than Guo et al. (2010)'s estimate of 0.3%. Therefore, it is effectively equivalent to Soon et al. (2015)'s approach of simply removing urban stations. However, if the urban area continues to expand, perhaps this is a more rigorous and systematic approach to take going forward.

This reweighting approach offers a reasonably straightforward method for reducing the influence of stations which are over-represented. However, if some climatic regimes are under-sampled or even have no stations, then the data for these regimes may be too limited, or simply absent. For this reason, Guo et al. (2010) recommended urgently setting up new climate stations in the under-represented regimes. However, while we echo this recommendation, it does not resolve the problems with establishing past trends.

Ren et al. have put considerable effort into developing a reference network for China that is more spatially representative than the standard networks (Ren et al., 2015, 2017; Zhang et al., 2010; Ren & Ren, 2011). However, they found that the available data was very limited before about 1960, and so their dataset only provides data for the period from 1960 onwards.

3.1.2. Comparing thermometer data to other estimates, i.e., model output and proxy data

We saw from Figure 2 that the distribution of weather stations is not ideal, with some climatic regimes over-represented (e.g., urban areas and croplands) and others under-represented (e.g., high altitude forested regions). Similarly, the distribution of temperature proxies is also uneven. Ironically, in many cases the over- and under-represented regions are almost reversed. For instance, both tree ring-based and ice core-based temperature proxies are mostly taken from high altitude, isolated locations that are often far from populated areas (e.g., Zheng et al., 2015).

Wang S. et al. (2001, 2004, 2007) actually used this point to improve the spatial distribution of their estimates by using temperature proxy series to improve the coverage in climatic regions which were under-represented by thermometer data. However, in general, this mismatch in spatial distributions poses a challenge when directly comparing thermometer-based estimates with proxy-based estimates for specific regions.

On the other hand, the output from Global Climate Models hindcasts and reanalyses such as the 20th Century Reanalysis (20CR) are perfectly evenly distributed. That is, the model output is typically generated in terms of evenly distributed grid boxes, e.g., 1°×1° (Li et al., 2017).

Therefore, when we are comparing model output, thermometer-based and proxy-based estimates of regional temperature trends (as we do in this paper), we should be wary of the fact that they may each be sampling different climatic regimes.

Having said that, if the thermometer-based estimates and proxy-based estimates both contain data from a sufficiently representative range of climatic regimes then it may be reasonable to compare both types of estimates to each other and to model output – even if the specific spatial distributions are different. We saw from Figure 2 that, although the spatial distribution of weather stations was not perfect, most of the main climatic regimes have at least some data. However, as Guo et al. (2010) noted, it would be desirable to improve the spatial representativeness of the weather station data further. The same applies to the temperature proxy data.

One way to improve the spatial representativeness of the available temperature proxies is to construct proxies from historical documents, e.g., non-thermometer based weather records, since these documents were often written in regions close to present day weather stations. This is an area of research which several of us have already been working on (e.g., Ge et al., 2003; Ding et al., 2015; Zheng et al., 2017; Liu et al., 2017), as historical records can act as very useful temperature proxies, and we encourage further research along these lines (Ge et al., 2016). Another way to improve the representativeness could be to weight the data according to the relative area occupied by the climatic regime being sampled, as discussed in the previous section.

3.1.3. Annual mean surface air temperatures vs. other metrics

In this review, we will mostly be focusing on the annually-averaged mean surface air temperatures for China. Even for this relatively straightforward temperature metric, we have seen that there is considerable debate. For this reason, we will mostly limit the scope of this review to this metric. However, the annually-averaged mean trends do not always capture all of the socially- and/or scientifically-relevant aspects of temperature changes. Therefore, it is worth briefly reviewing the relevance of other temperature metrics.

Surface air temperature trends are typically described in terms of the mean daily temperature (“T_{mean}”). However, the temperature trends are often different for the daily minima (“T_{min}”) values (usually night-time) and the daily maxima (“T_{max}”) values (usually afternoon). For instance, in a recent simulation of the urban heat island associated with Beijing, Liu X. et al. (2018) calculated that the urban warming was most pronounced for the night-time minima (i.e., T_{min}) and less pronounced for the daily maxima (i.e., T_{max}). This is consistent with many studies of urbanization bias in China which typically seem to find the bias is greatest for T_{min}, e.g., Zhou et al. (2004); Zhang et al. (2005); Hua et al. (2008); Yang Y.-J. et al. (2013); Wang J. et al. (2013); Zhang et al. (2014); Ren (2015). Fall et al. (2011) also find that siting biases which occur when weather stations are poorly sited (and thereby strongly influenced by local microclimates) are most pronounced for T_{min}. On the other hand, McNider et al. (2012) argue that changes in atmospheric temperatures, e.g., from increased greenhouse gas concentrations, are more likely to be captured by changes in T_{max} than T_{min}.

For these reasons, depending on the aspects of temperature variability that are being studied, it may be more relevant to study T_{max} or T_{min} instead of T_{mean}. Another related metric which can be of interest is the so-called “Diurnal Temperature Range” (DTR), i.e., the difference in temperature between T_{max} and T_{min}, e.g., Zhang et al.

(2005); Fall et al. (2011); Wang J. et al. (2013); Li et al. (2015); Ren (2015).

From a societal perspective, it can often be more relevant to study changes in the frequencies of days above/below certain thresholds, e.g., changes in the growing season length (for agricultural purposes), or in the occurrence of heat waves or extreme cold days (e.g., Pielke et al., 2002; Ren et al., 2012; Li et al., 2015).

On the other hand, if you are assessing trends in the atmospheric heat content of the Earth system, then studying changes in surface air temperatures may not be sufficient. For instance, Pielke et al. (2004) note that changes in the heat content of surface air can also manifest themselves as changes in absolute humidity instead of just temperature (see also Liu X. et al., 2018). Therefore, Pielke et al. (2004) recommend combining humidity measurements with temperature records when studying trends in surface heat content. Also, Lin et al. (2015) note that the temperature trends when measured at a height of 2m (the typical height of most temperature stations) can be different from the trends at different heights.

Often temperature trends vary with season, e.g., Pielke et al. (2002); Hua et al. (2008); Ren et al. (2012); Li et al. (2015); Yan et al. (2015); Sun et al. (2017b); Kawakubo et al. (2017); Liu Z. et al. (2018). For this reason, it can be desirable to separate the annual temperature trends into their seasonal components – winter, spring, summer and autumn. This is especially relevant when comparing temperature proxies to instrumental records since the behavior of a temperature proxy is often most influenced by a particular season, e.g., tree ring growth may be most influenced by temperatures during the summer. However, although we will briefly revisit the issue of seasonality in Section 3.5, for this review, we will confine our discussion to the annually-averaged T_{mean} surface air temperature trends.

3.2. The problem of non-climatic biases

It is well-known that the multi-decadal temperature records of weather stations are frequently affected by various non-climatic biases (e.g., Mitchell, 1953; Oke, 1973; Karl & Williams, 1987; Easterling & Peterson, 1995; Pielke et al., 2007a; Pielke et al., 2007b; Menne & Williams, 2009; Ren et al., 2015; Soon et al., 2015; Li et al., 2017). Therefore, when using weather station records to estimate long-term regional (or global) climatic temperatures trends, it is important to correct or account for these non-climatic biases.

Some events which could introduce a non-climatic bias into a station's temperature record include: station moves (Karl & Williams, 1987; Butler et al., 2005; Menne & Williams, 2009; Soon et al., 2015); changes in the time of observation (Karl et al., 1986; Tang & Ren, 2005; Li et al., 2017); changes in the types of thermometer used and/or the instrument shelter used to house the thermometer (Mitchell, 1953; Butler et al., 2005); changes in the method by which daily temperatures are calculated (Mitchell, 1953; Butler et al., 2005); changes in the immediate surroundings of the thermometer, such as the construction of new buildings, car parks, etc. (Pielke et al., 2007a; Pielke et al., 2007b; Menne et al., 2010; Fall et al., 2011); urbanization bias (e.g., Mitchell, 1953; Oke, 1973; Karl et al., 1988; Ge et al., 2013; Wang F. et al., 2015; Soon et al., 2015; Ren et al., 2017).

3.2.1. The urbanization bias debate

There has been considerable ongoing debate since the 1990s over the extent to which urbanization bias has affected estimates of Chinese temperature trends. Until now, the discussion of the urbanization bias problem for China has mostly focused on its relevance for recent temperature trends, i.e., the rate of warming in recent decades. This is understandable since the rate of urban development in China has been most pronounced in recent decades (especially since the early 1980s). As a result, those studies which have found a substantial urbanization bias have found that it has increased the apparent warmth of the recent warm period, but not had as much influence on temperatures in the earlier decades. This seems to have led many to the mistaken assumption that it has no relevance for the apparent warmth of the early 20th century. It is true that the urbanization bias in China is most pronounced for the recent warm period. However, when urbanization bias substantially increases the apparent warmth of the recent warm period, it also reduces the apparent relative warmth of the early warm period. Specifically, if urbanization bias makes the recent warm period seem warmer, then it also makes the earlier warm period seem cooler (in comparison). Therefore, in this section, we will review the reasons for the debate over the extent to which urbanization bias has affected Chinese temperature estimates.

Most studies agree that individual station records from highly urbanized cities may have been significantly affected by urbanization bias, e.g., Beijing (Ren et al., 2007; Zhang L. et al., 2014; Zhang Y. & Ren, 2014; Yan et al., 2010; Wang et al., 2013). However, calculating the net effects of urbanization bias on regional and national temperature trend estimates has been more challenging and contentious.

For instance, Jones et al. (1990) argued that there had been no significant urbanization bias for eastern China. Yet, in a separate study involving two co-authors of Jones et al. (1990) and using the same dataset, Wang et al. (1990) found that the urban stations in that dataset showed an average urbanization bias of +0.04°C/decade over the 1954-83 study period. This suggested that ~33% of the +0.12°C/decade warming trend of the urban stations was due to urbanization bias. They also noted that most of the "rural" stations in the dataset were probably quite urbanized themselves. Portman (1993) carried out a more detailed analysis for one of the regions in the previous studies ("Northern plains") and found similar biases (between 0.05°C/decade and 0.09°C/decade) for the same period. Moreover, he found a slight cooling trend for the rural stations (-0.01°C/decade), meaning that all of the warming observed for that region was due to urbanization bias.

However, after homogenizing the Chinese station data, Li et al. (2004) found that removing the 30% most urban stations (according to associated population) made almost no difference to their estimates of Chinese temperature trends over the 1951-2000 period. They concluded that the net effect of urbanization bias on regional trends was almost negligible (at least after homogenization). Zhou et al. (2004) did find some evidence of urbanization bias for winter temperatures in their study of southeast China for the 1979-1998 period, but calculated that the net contribution to regional trends was only 11% (+0.05°C/decade out of the observed +0.45°C/decade warming trend). On the other hand, when Zhang et al. (2005) extended the analysis of Zhou et al. to cover all of eastern China for 1960-1999, they found that urbanization bias (and other changes in land use) could explain up to +0.12°C/decade of the observed warming, i.e., 18% of the observed +0.66°C/decade warming trend.

In 2005, Ren et al. published a series of analyses (Zhou & Ren, 2005; Chu & Ren, 2005; Chen et al., 2005; Zhang & Ren, 2005; Ren et al., 2005) each demonstrating a significant effect from urbanization bias to regional temperature trends for: North China, the Shandong and Hubei provinces and the Beijing area. Ren et al. (2005) also noted that balloon measurements only showed a very modest rate of warming (+0.05°C/decade) for the lower-to-mid troposphere (400-850hPa) relative to surface temperatures (+0.30°C/decade) for China over the 1961-2004 period. This was consistent with the surface temperature estimates being significantly biased by urbanization.

Since then, there have been numerous different estimates of the magnitude of the urbanization bias problem, with several studies claiming it is considerable (Soon et al., 2015; Ren, 2015; Ren et al., 2008, 2015, 2017; Sun et al., 2016; Hua et al., 2008; Zhang et al., 2010; Ren & Ren, 2011; He & Jia, 2012; He et al., 2013; Ge et al., 2013a; Li et al., 2013), others claiming it is relatively modest (Jones et al., 2008; Wu & Yang, 2013; Zhao et al., 2014), and others claiming it is almost negligible (Li et al., 2010a, 2017; Yan et al., 2016; Wang J. & Yan, 2016; Wang et al., 2017; Wang F. et al., 2015).

One factor for the debate over the magnitude of the urbanization bias is that the rural and urban stations are often in climatically different regions of China, e.g., many of the rural stations are in mountainous regions, while most of the urban stations are located in the plains (Ren et al., 2015; Wang J. et al., 2013). For instance, Wang J. et al. (2013) analysed 20 stations in the Greater Beijing area for the period 1978-2008. They identified 7 of these stations as "urban", 4 as "suburban" and the remaining 9 as being rural. However, because 6 of these rural stations were from mountainous regions, they excluded them from their "rural" subset and only used the 3 non-mountainous rural stations for their assessment of urbanization bias. They found that the 7 urban stations showed an extra +0.148°C/decade warming relative to the 6 rural mountain stations (i.e., 24.5% of the +0.604°C/decade warming trend of the urban stations). However, relative to the 3 rural plains stations, the extra warming of the urban stations was only 0.066°C/decade (i.e., only 10.9% of the +0.604°C/decade warming trend of the urban stations).

Another factor is the methods used for classifying stations into "urban" and "rural" subsets. Early studies often simply divided stations into those with an associated population above ("urban") or below ("rural") a certain threshold value, e.g., (Jones et al., 1990; Wang et al., 1990; Portman, 1993; Li et al., 2004). However, it is now recognised that such simple classifications are often inadequate for estimating the true magnitude of urbanization bias effects (Stewart & Oke, 2012). First, the growth of urban heat islands is a continual process. It is **not** a simple situation where there is a single degree of urbanization below which there is no urban heat island and above which there is a constant urban heat island. Instead, the magnitude of the urban heat island tends to gradually increase as the area becomes more urbanized. So, the use of a single "urban/rural" threshold could mistakenly include urban stations into the "rural" subset and vice versa. Hence, several studies have started classifying stations into multiple subsets/categories, e.g., (Soon et al., 2015; Stewart & Oke, 2012; Wang F. & Ge, 2012; Li et al., 2013; Yang X. C. et al., 2011).

Also, while heavily urbanized areas tend to have a much higher population than rural areas, the use of a single population value is a very crude metric for estimating the rate of urbanization of an individual station. For instance, in some large old cities, much of the population growth may have occurred before the station was set up (Jones et al., 2008), while the actual urbanization experienced by a station in a smaller, but faster growing urban area might be greater (Stewart & Oke, 2012). Also, the rate of urbanization experienced by a station could be different in the urban centre compared to the outskirts of a city (Yang et al., 2013; Shi T. et al., 2015; Zhang et al., 2014; Zhang & Ren, 2014).

For these reasons, several studies have tried to develop more sophisticated methods for classifying the urbanization of stations using combinations of different urbanization metrics, such as satellite-derived land surface temperatures (i.e., thermal environment); satellite-measured nighttime light imagery; estimates of land use changes; as well as population-based metrics, e.g., (Soon et al., 2015; Ren & Ren, 2011; Wang Y. et al., 2011; Stewart & Oke, 2012; Wang F. & Ge, 2012; He & Jia, 2012; He et al., 2013; Ge et al., 2013a; Li et al., 2013; Yang et al., 2013; Shi T. et al., 2015; Zhang et al., 2014; Zhang & Ren, 2014; Yang et al., 2011; Li et al., 2015). For instance, using satellite imagery, Zhang et al. (2017) found an increase of 0.62°C in the average land surface temperature for every 10% increase in impervious surface area (a measure of urbanization). Ideally, these methods should be able to quantify the changes in urbanization over time as well as the degree of urbanization, so that the rates of urbanization can be studied, e.g., (Stewart & Oke, 2012; Wang F. & Ge, 2012; He & Jia, 2012; He et al., 2013; Ge et al., 2013a; Li et al., 2013; Yang et al., 2011).

A major problem for estimating the magnitude of the urbanization bias is that there are other non-climatic biases which can affect station records, and it is very challenging to separate these biases from each other. For instance, while most Chinese weather stations currently use the same observation times for recording daily temperatures, before the 1950s, observation times may have varied from station to station and over time (Tang & Ren, 2005; Tang et al., 2010). Also, individual station records may have non-climatic biases due to changes in instrumentation and/or changes in the local environment.

In particular, several groups have noted that in China, it is common practice to occasionally relocate weather stations that have become highly urbanized into less built-up suburban locations. However, over time these suburban locations can themselves become urbanized, resulting in a complex mixture of non-climatic biases (gradual urban warming followed by abrupt cooling station moves). Some groups have argued that correcting for these station moves could reduce the magnitude of the apparent urbanization bias (Yan et al., 2016; Wang J. & Yan, 2016; Yan Z. W. et al., 2010; Wang J. et al., 2013). Other groups have argued that correcting for these station moves increases the apparent urbanization bias (Ren et al., 2007; Yang et al., 2013; Shi T. et al., 2015).

3.2.2. The debate over homogenization approaches

In an attempt to simultaneously correct for all of these non-climatic biases, several statistics-based, "homogenization" procedures have been developed, e.g., Karl & Williams (1987); Easterling & Peterson (1995); Li et al. (2004; 2010); Menne & Williams (2009); Xu et al., 2013; Lakatos et al., 2013. These homogenization algorithms statistically compare each station record to the records of its neighbours to try and identify anomalous step changes in a given record which might be due to a non-climatic bias, such as a station move. The algorithms then estimate the magnitude of this proposed bias and apply an equivalent (but opposite) adjustment to the raw station record. The use of these homogenization algorithms can be combined with station history (or "metadata") information (if available) and/or the manual inspection of station records, e.g., Karl & Williams (1987); Li et al. (2004; 2010; 2017); Ren et al. (2017). Alternatively, it may be an entirely automated process which does not consider the station histories, e.g., Easterling & Peterson (1995); Menne & Williams (2009).

Some studies have suggested that using homogenized data reduces the apparent urbanization bias problem, e.g., (Xu et al., 2017; Yan et al., 2016; Wang J. & Yan, 2016). Two different arguments seem to be involved. Wang J. et al. (Wang J. & Yan, 2016; Yan et al., 2016) suggest that rural sites might be more affected than urban sites by non-climatic cooling biases, and that this leads to an overestimation of the urbanization bias when using non-homogenized data. On the other hand, Xu et al. (2017) suggest that the homogenization process itself could remove some of the urbanization biases in the data. They argue that "*in many cases, urban influences on temperatures at a station will be manifested as a step change [...] and will be adjusted for as part of the general homogenization process*". However, many of the studies which found evidence of urbanization bias in Chinese temperature estimates were using homogenized data, e.g., (Ren et al., 2008, 2015, 2017; Ren & Ren, 2011; He & Jia, 2012; He et al., 2013; Ge et al., 2013a; Yang et al., 2013; Yang et al., 2011). Also, as mentioned earlier, Soon et al. (2015) noted that applying a popular homogenization algorithm (i.e., Menne & Williams, 2009) failed to remove or even identify any of the urbanization bias for the Chinese data when there was a relatively high fraction of rural data (i.e., 1951-1990).

Moreover, Soon et al. (2015) found evidence that, whenever the rural stations were rarer, the homogenization process often led to "urban blending". When the homogenization process is estimating the sign and magnitude of a potential non-climatic bias, the target record is compared to that of its neighbours. However, if a significant fraction of the neighbours are affected by urbanization bias, then the calculated sign and magnitude of the identified "non-climatic bias" will be inadvertently biased by the urbanization bias of the neighbours. That is, some of the urbanization bias of the neighbours will be "aliased" into the homogenized record – see deGaetano (2006) and Pielke et al. (2007a). This means that applying homogenization with an urbanized network will have a tendency to underestimate the magnitude of any biases which introduce a "warming" trend, while overestimating the magnitude of any biases which introduce a "cooling" trend. The net tendency of this urban blending would be to artificially cool the earlier period and warm the recent period. Additionally, while the trends of the most urban stations would tend to be partially reduced, the trends of the rural stations would also tend to be increased to match those of their urban neighbours. This mixing process certainly would tend to make the station data more "homogeneous", in that all stations would have fairly similar trends. But, it would not actually remove the non-climatic biases. Instead, homogenization would distribute the biases more evenly amongst all stations.

Although the "blending" or "aliasing" problem of the current homogenization processes has been recognized for more than a decade (deGaetano, 2006; Pielke et al., 2007a; Soon et al., 2015), its effects on homogenized temperature data seem to have been largely overlooked, e.g., there is no discussion of the problem in any of the following studies which describe homogenized Chinese datasets: Li Q. et al. (2004; 2010; 2017); Li Z. et al (2010; 2015); Ren et al. (2012; 2017); Xu et al. (2013); Cao et al. (2013); Ding et al. (2014); Wang J.F. et al. (2014). Menne & Williams (2009) and Hausfather et al. (2013) briefly considered the possibility that "aliasing" might be a concern, but only carried out a preliminary assessment. Therefore, it is worth addressing this issue in a separate subsection.

3.2.3. The blending problem of current homogenization techniques

All of the current homogenization procedures essentially work by comparing each station's record to those of its neighbours. For instance, the Easterling & Peterson (1995) approach involves the construction of a "reference series" for each of the stations. This reference series is the average of five of the neighbouring station records. The reference series is then subtracted from the record of the station being tested ("the target series") and the "difference series" is then statistically analysed for any unusual step changes (known as "breakpoints"). Whenever a breakpoint passes a given statistical significance test, it is assumed to be a non-climatic bias. Li et al. (2004)'s first homogenized dataset for China was largely based on Easterling & Peterson (1995), but station history information was also used to assess if the statistically-identified breakpoints corresponded to documented station changes.

Other algorithms use variations of this basic approach, e.g., rather than using a single reference series, the Menne & Williams (2009) algorithm directly compares each target series with 40 different neighbouring stations (one-at-a-time). The "MASH" algorithm (Lakatos et al., 2013) used by Li Z. et al. (Li Z. & Yan, 2010; Li Z. et al., 2015) is similar. However, in all cases, for our discussion here, the basic principles are essentially the same.

The next step of the algorithm involves estimating the magnitude and sign of this apparent bias. Typically, to do this, the algorithm calculates the mean temperature (for some given length of time) of the target station before and after the break-point. It then compares this difference to the equivalent differences of the neighbouring stations. The "RHtests" algorithm (Vincent et al., 2012) currently used by Li Q. et al. (2010; 2017; Xu et al., 2013) takes a slightly more complex approach and compares the statistical distribution of the temperatures before and after the break-point in terms of quantiles instead of just means (Vincent et al., 2012; Xu et al., 2013). However, for our discussion, both approaches are equivalent. It is this step which inadvertently leads to the blending problem.

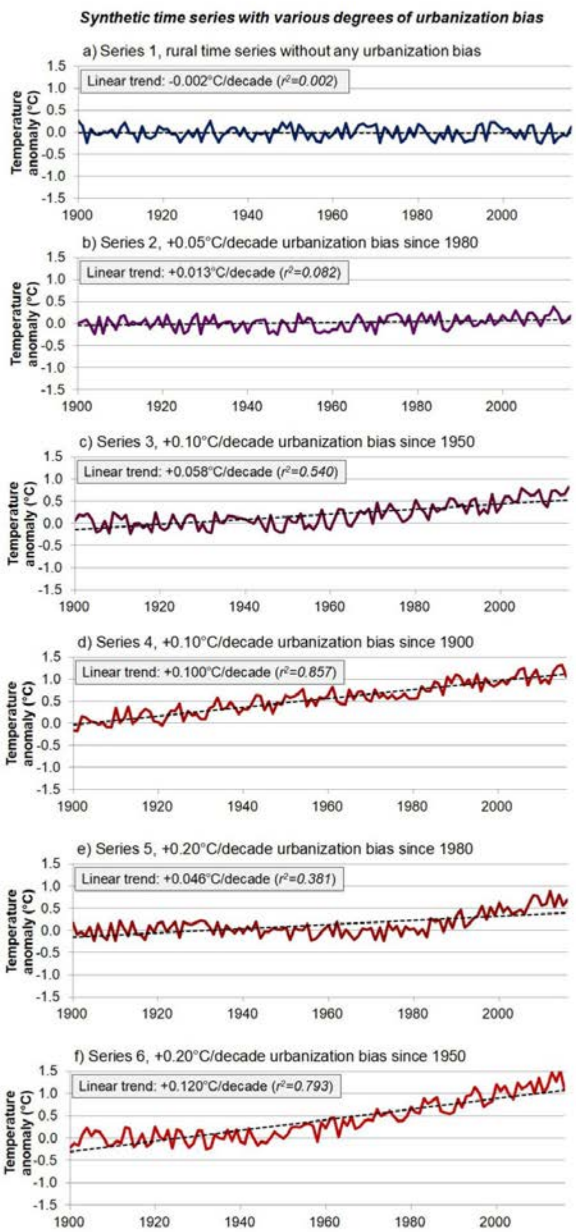

Figure 3. The six synthetic time series of "temperature anomalies" with various degrees of urbanization bias added, which we use to demonstrate the theoretical basis for the urban blending problem of current homogenization techniques. All six series initially consisted of random values between -0.25°C and +0.25°C, and therefore have no long-term secular trend. However, for series 2-6 [(b) to (f)], different linear "ramps" have been added to mimic "urbanization bias": b) +0.05°C/decade since 1980; c) +0.1°C/decade since 1950; d) +0.1°C/decade since 1900; e) +0.2°C/decade since 1980; f) +0.2°C/decade since 1950.

As a thought experiment, let us suppose that global temperatures have been constant since 1900. Now, let us suppose we have 6 neighbouring stations with complete station records (1900-2016). Let us also suppose that each of these stations have experienced different degrees of urbanization over the length of their record. In Figure 3, we plot 6 synthetic time series. To mimic the local variability, for each of the series, the underlying temperature anomaly for each year is a random value between -0.25°C and + 0.25°C. To mimic urbanization bias, we have then added various linear trend ramps to each series. Series 1 we have left unaffected by urbanization bias, and is representative of a rural station for our thought experiment. Of the rest of the series, Series 2 has had the smallest bias added to it with a relatively small trend of +0.05°C/decade from 1980 onwards. Meanwhile Series 6 has had the largest urbanization bias added to it, with a linear trend of +0.2°C/decade from 1950 onwards.

As we discussed above, there are various approaches to using reference series for homogenization. For simplicity, let us take the Easterling & Peterson (1995) approach. Therefore, the "reference series" for homogenizing Series 1 (a rural station) will be the mean of Series 2-6 and the "reference series" for homogenizing Series 6 (a heavily urbanized station) will be the mean of Series 1-5. Both reference series will be affected by urbanization bias, because either four or five of the five neighbours contain at least some urbanization bias.

Now, let us suppose that Series 1 (rural) actually has experienced several station moves, and that all of these station moves have been documented. In Figure 4, we have arbitrarily added four step change biases each at 25 year intervals, with each coincidentally cancelling out the previous bias: +0.5°C, -0.5°C, -0.5°C, +0.5°C. What would happen when you homogenize this record?

In our thought experiment, because the station moves are documented, the timings of the station moves are known. However, since none of the other series have experienced station moves (this is of course unlikely) and the step changes are quite obvious, most of the standard automated homogenization algorithms would probably also correctly identify the **timings** of the four station moves. Instead, the problem occurs in identifying the magnitudes (and possibly sign) of the associated non-climatic bias.

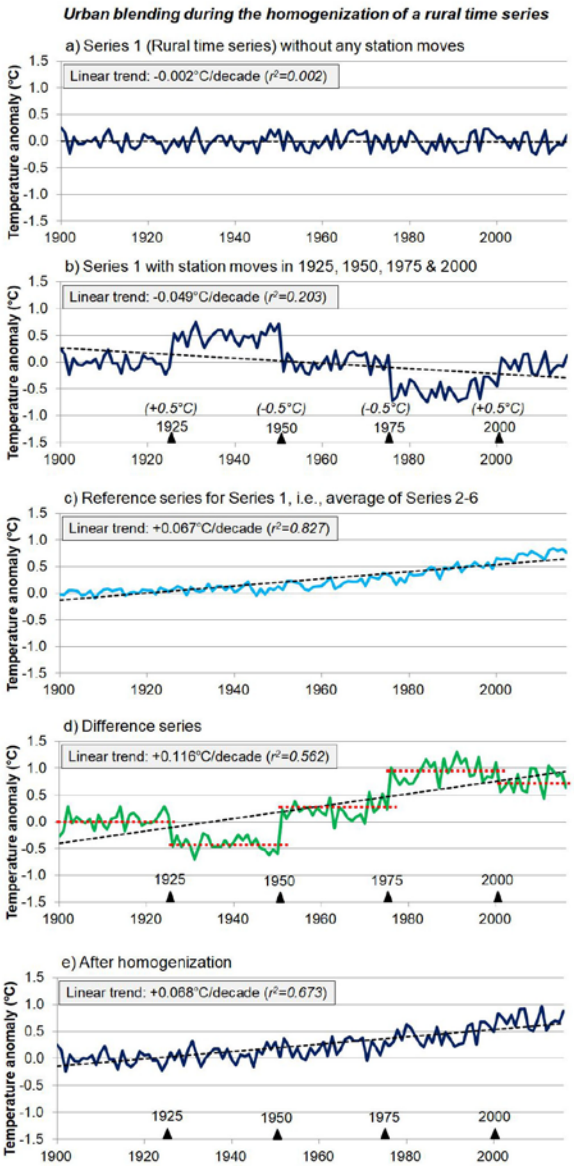

Figure 4. Illustration of how the homogenization of a rural time series can lead to urban blending. a) Series 1 from Figure 3; b) as for a) but with four station moves leading to a +0.5°C step bias in 1925, a -0.5°C step bias in 1950, a -0.5°C bias in 1975 and a +0.5°C bias in 2000; c) the reference series, i.e., the mean of the other five time series; d) the difference between b) and c); e) the "rural" time series after homogenization which now has a linear trend comparable to the urbanized reference series.

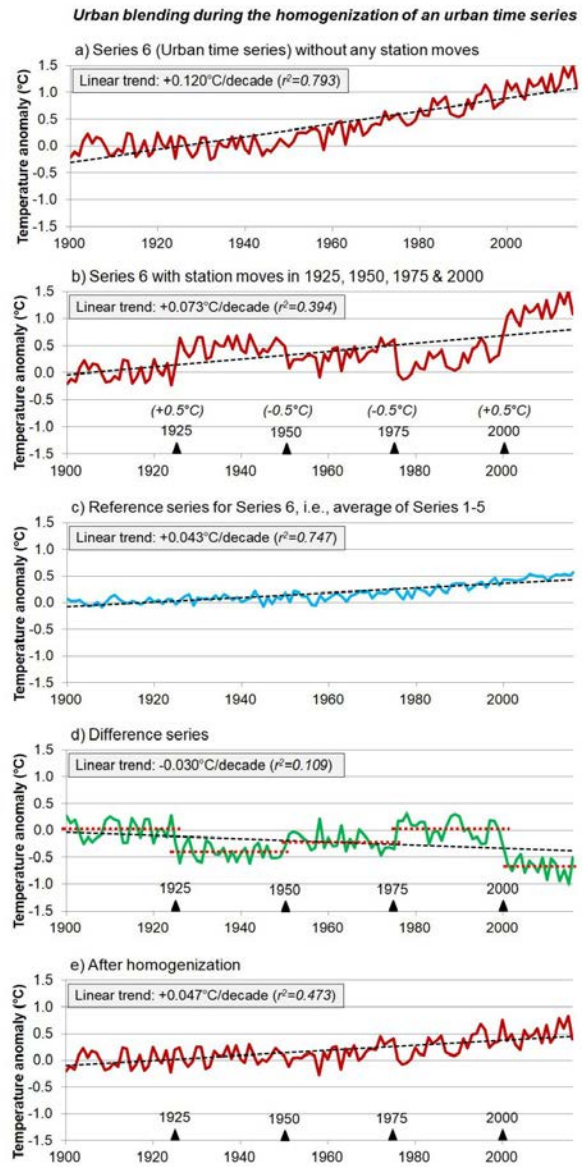

Figure 5. Illustration of how the homogenization of an urban time series can also lead to urban blending. a) Series 1 from Figure 3; b) as for a) but with four station moves leading to a +0.5°C step bias in 1925, a -0.5°C step bias in 1950, a -0.5°C bias in 1975 and a +0.5°C bias in 2000; c) the reference series, i.e., the mean of the other five time series; d) the difference between b) and c); e) the "rural" time series after homogenization which now has a linear trend comparable to the (still-urbanized) reference series.

Because the reference series has a long-term warming urbanization bias, on average the means of the difference series will be warmer after the station move than before. Since most of the algorithms estimate the magnitudes of each non-climatic bias by (in some way) comparing the means of the difference series before and after each of the station moves, this means that homogenizing Series 1 will artificially transfer some of the urbanization bias of the neighbour average into the rural station record. Indeed, we can see by comparing Figure 4(c) and Figure 4(e) that the linear trend of the homogenized series is almost exactly the same as the reference series (+0.068°C/decade and +0.067°C/decade respectively). The homogenization process has arguably been successful in removing the abrupt step changes in the original Figure 4(b) series. However, in the process, the warming trend of the reference series has been "aliased" (deGaetano, 2006; Pielke et al., 2007a) onto the target series. In our thought experiment, we know (by design) that this warming trend is urbanization bias. In other words, through "urban blending", the homogenized "rural" time series is now affected by a similar amount of "urbanization bias" as its neighbours.

Now, let us repeat the same thought process for Series 6 (highly urbanized). In Figure 5, we have applied our four (arbitrary) step change biases to Series 6. Again, most of the standard homogenization algorithms should easily identify the **timings** of the four documented station moves. And, in this case, since the urbanization bias in Series 6 is greater than for the reference series, the blending process should actually partially reduce the urbanization bias in Series 6 as well as partially reducing the magnitude of the step biases. However, because the reference series itself is also affected by urbanization bias, the blending process only **partially** reduces the magnitude of the urbanization bias: from +0.073°C/decade in Figure 5(b) to +0.047°C/decade in the homogenized Figure 5(e) series.

Note that this blending problem occurs regardless of whether the actual station moves (or other step change) led to cooling or warming. The problem relates to the trends of the reference series. In general, the more "non-climatic biases" (i.e., breakpoints) the homogenization procedure identifies and the more urban neighbours there are, the more urban blending will occur. The more rigorous the homogenization procedure is (i.e., the more breakpoints are identified) the more blending can occur.

The blending problem means that homogenization will tend to blend (or smooth or homogenize) the non-climatic biases amongst all stations. Therefore, after homogenization, most stations will have fairly similar trends. Ironically, because the "rural" stations now have some urbanization bias and the most heavily urbanized stations will have had their urbanization bias partially reduced, comparing the "rural" and "urban" trends of homogenized data will tend to underestimate the extent of urbanization bias. This seems to have led several groups to conclude that the homogenization process has reduced the urbanization bias problem, e.g., Menne & Williams (2009); Yan Z. W. et al. (2010); Wang J. et al. (2013); Hausfather et al. (2013); Yan et al. (2016); Wang J. & Yan (2016). However, the above thought experiment demonstrates that the homogenized trends will also include the average of the non-climatic biases that are common to the neighbours, e.g., urbanization bias.

3.2.4. Demonstration of urban blending during the homogenization of 10 stations near Beijing

The above thought experiment demonstrates that – in principle – current homogenization algorithms can inadvertently lead to urban blending. Indeed, a similar phenomenon has already been noted by deGaetano (2006) and Pielke et al. (2007a) who demonstrated using various tests of common homogenization algorithms that homogenization often "aliases" the trend biases of reference series onto that of the target series. However, does the phenomenon occur in real world situations, and in particular, how relevant is it for the Chinese data?

He & Jia (2012) studied the effects of a popular homogenization procedure called "Multiple Analysis of Series for Homogenization" or "MASH" (Lakatos et al., 2013) on the temperature trends for 10 stations near Beijing, China over the 1978-2008 period, using the dataset developed by Li Z. et al. (2010). This homogenization process apparently used the station histories provided by the CMA similarly to the process used for homogenizing the CMA's own datasets (Li et al., 2004; 2010; 2017).

During this period, the region experienced considerable urbanization, and He & Jia ranked the 10 stations according to how much urbanization each station had experienced. They then compared the temperature trends for each station over the 1978-2008 period before and after homogenization. Before homogenization, 68% of the apparent warming seemed to be urbanization bias. However, after homogenization, the difference in warming between the stations was reduced to ~20%.

Initially, the results from He & Jia (2012) seem to be consistent with those studies which suggest that homogenization reduces urbanization bias e.g., (Xu et al., 2017; Yan et al., 2016; Wang J. & Yan, 2016). However, a close inspection of the results reveals that the homogenization process has actually led to urban blending. In Figure 6, we have digitized and re-plotted the results from He & Jia (2012). We can see that the homogenization process has indeed substantially reduced the apparent differences between all of the stations. However, rather than the warming trends of all of the urban stations being reduced to match those of the least urbanized stations, all of the trends appear to have been adjusted to match those of the average stations – which happen to be intermediately urbanized. That is, the trends of the most urban stations are partially reduced, but the trends of the most rural stations are also increased.

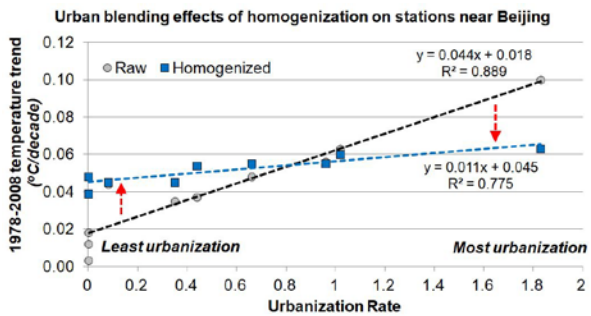

Figure 6. Illustration of the urban blending problem based on He & Jia (2012)'s study of the effects of "MASH" homogenization on temperature trends for 10 stations near Beijing, China over the 1978-2008 period. The data for this figure was digitized from He & Jia (2012)'s Figure 3. Values to the right correspond to the stations which have experienced the most urbanization over the 1978-2008 period.

3.2.5. Some approaches to minimizing the blending problem of homogenization

We have shown above that the application of the current temperature homogenization processes to weather station records can inadvertently lead to a blending of non-climatic biases among all of the stations. In these cases, the homogenized data will still be affected by non-climatic biases. Moreover, those stations which had previously been unaffected by those biases will have themselves become biased in the process. This leads to the counterintuitive result that the homogenization of temperature records (in an attempt to improve their reliability), in many cases, may mistakenly be making the data less reliable.

We have limited our discussion of the blending problem to that for urbanization bias, i.e., "urban blending". However, the same phenomenon could also occur for other non-climatic biases, e.g., siting biases arising from changes in station exposure (Pielke et al., 2007a; Pielke et al., 2007b; Menne et al., 2010; Fall et al., 2011). That is, if a given type of non-climatic bias has similarly affected multiple stations in a given region, then the current homogenization approaches may lead to a blending of those biases among all stations. If there are significant differences between the trends of a subset known to be unaffected by a particular bias and those of a subset affected by the bias in the non-homogenized data, but these differences are apparently reduced by the homogenization process this is actually a strong indication that blending of biases may have occurred. With this in mind, it is worth noting that both Menne et al. (2010) and Fall et al. (2011) have noted that the apparent differences in trends between well-sited stations and poorly-sited stations in the U.S. are almost entirely removed after homogenization. Previously this has been assumed to indicate that homogenization has reduced the siting biases (e.g., Menne et al., 2010; Fall et al., 2011). However, it should now be apparent that homogenization may be leading to a **blending** of siting biases instead of their removal. One way to avoid (or at least considerably reduce) the urban blending problem would be to only use definitively rural stations for homogenization. Some groups seem to have put considerable effort into doing this when homogenizing urban stations, e.g., Ren et al. (2007, 2015; Zhang et al., 2014; Zhang & Ren, 2014; Ren & Ren, 2011) and Yang Y. J. et al. (Yang et al., 2013; Li Y. B. et al., 2015; Shi T. et al., 2015). It is interesting to note that in these cases, homogenization tended to increase the difference between rural and urban station records, i.e., the opposite of blending, thereby making urbanization bias easier to quantify. This could be a factor in why those studies reached different conclusions from (Yan et al., 2010, 2016; Wang J. & Yan, 2016; Wang J. et al., 2013).

Ren et al. have constructed a relatively large network of (mostly rural) reference stations for China for the post-1960 period, and they have been able to use this for homogenizing the Chinese temperature data from 1961-onwards (Ren et al., 2015, 2017; Zhang et al., 2010; Ren & Ren, 2011). They confirmed that there has been a warming trend since the 1970s for China, but that urbanization bias had exaggerated that trend by at least 27%. However, while using rural stations for homogenizing is feasible when there is a relatively high density of rural stations, there are very few rural stations with data for the early 20th century, i.e., the period including the 1920s-40s warm period. Therefore, Ren et al.'s homogenized reference network cannot be used for directly assessing the early 20th century period.

Similarly, Karl et al. (1988) put a considerable effort into constructing the U.S. Historical Climatology Network for the United States. However, while this should reduce the blending problems of homogenization, it might not completely remove them. For instance, Fall et al. (2011) and Soon et al. (2015) have shown that urbanization bias and siting biases are still present in even the high quality U.S. Historical Climatology Network.

So, while it is clearly desirable to remove any non-climatic biases from the station records before estimating Chinese temperature trends, we should also be cautious that the current homogenization procedures may not actually be removing all of these biases, and may even be introducing new biases. Hence, while there is widespread agreement that the raw, unadjusted station records contain non-climatic biases, there is an open debate over whether the current homogenized datasets are more reliable or not.

3.3. Construction of new relatively-rural estimate of Chinese temperature trends

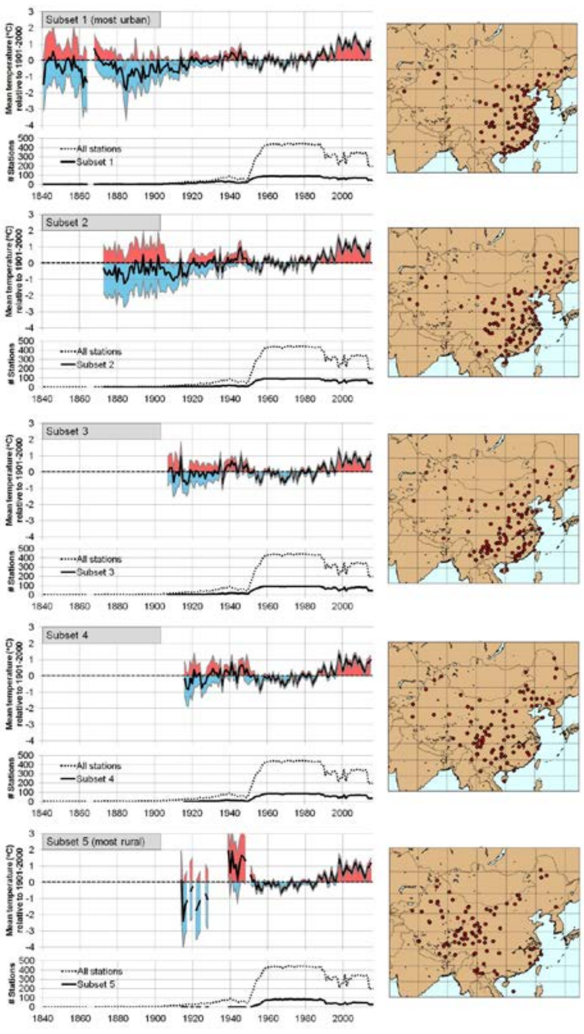

Figure 7. Chinese temperature trends according to each of the five GHCN version 4 subsets. Error bars are indicated with gray curves, and correspond to twice the standard error. All reconstructions are shown relative to their 20th century mean (1901-2000). Values above and below that mean are highlighted with red and blue shading, respectively. The station locations for each subset are shown in the maps on the right hand side.

Figure 7 shows the different estimates of Chinese temperature trends which are obtained when using the five different subsets from Figure 1 of the Chinese stations in version 4 of the GHCN (non-homogenized) dataset. Unlike the analysis in Soon et al. (2015) which only considered three subsets, the differences between each subset can be relatively subtle, e.g., Subsets 3 and 4 are quite similar to each other while Subsets 2 and 3 are also quite similar to each other. However, in general, the more rural the subset is, the warmer the early 20th century warm period becomes and the cooler the current warm period becomes. This is certainly consistent with the more urban subsets being affected by urbanization bias.

On the other hand, it can be seen from the subset maps that the spatial distribution of stations also varies with each subset. As discussed in the previous section, most of the urban stations tend to be in eastern China (especially near the eastern coast), while most of the rural stations tend to be in western China (especially in the more mountainous regions). Therefore, it is possible that some of the differences between the subsets could be related to their different spatial distributions and geographical settings.

Another problem (also discussed in the previous section) is that the longest and most complete station records tend to be from the more urban subsets. This can be seen more clearly in Figure 8, which shows the percentages of each subset with data for a given year. Although each subset contains exactly 20% of the stations in the GHCN dataset, it can be seen that the more rural subsets tend to have much shorter records and the relative breakdown of the subsets changes quite dramatically over time.

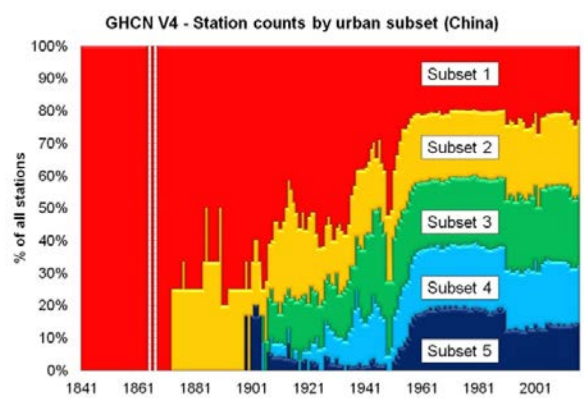

Figure 8. Breakdown of the Chinese urban subsets with data available for each year in the GHCN version 4 dataset.

For instance, the only stations with data from 1841-1873 are all from Subset 1 (the most urban subset), and for the rest of the 19th century (aside from one year, 1899), all of the station data comes from Subsets 1 and 2 (the two most urban subsets). While the data for the 1960-1990 period is relatively evenly distributed among the subsets (and to a lesser extent, the post-1990 period), for the early 20th century, the most rural subsets have relatively little data. Unfortunately, this coincides with one of the two periods we are focusing on in this paper, i.e., 1920s-1940s. So, while there is a reasonable amount of data for comparing the two warm periods for the urban subsets, the data is much more limited for the rural subsets.

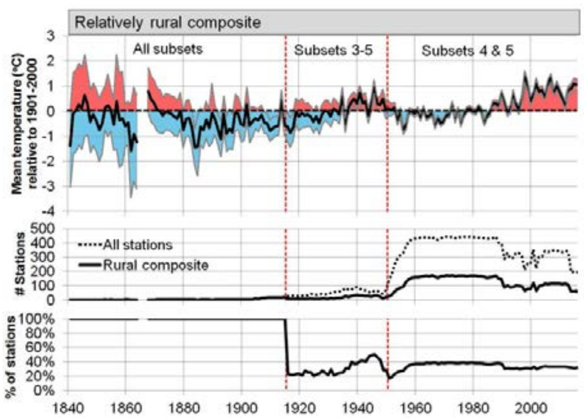

Figure 9. Relatively rural composite Chinese temperature trends since the 19th century, relative to the 20th century mean (1901-2000). Error bars are indicated with gray curves, and correspond to twice the standard error. Values above and below that mean are highlighted with red and blue shading, respectively. The bottom panels indicate the fraction of available Chinese GHCN version 4 stations used for each year.

Hence, rather than using just the most rural subset (Subset 5) for comparing the two periods, we adopt a similar approach to Soon et al. (2015) and develop a "relatively rural composite" by including some of the data from the urbanized subsets for the earlier periods to increase the amount of available data. This composite is plotted in Figure 9 along with details on its composition.

For the 1951-2016 period, all of the subsets have a relatively large number of stations with data. Therefore, we only use the two most rural subsets (Subsets 4 & 5) for our composite. However, as can be seen from Figure 8, neither of these subsets have a lot of data before 1951. Therefore, for the 1916-1950 period, we also include stations from the moderately urbanized Subset 3. Before 1916, there is almost no available data from the rural subsets. Therefore, for the 1841-1915 period, we use all available stations, regardless of how urbanized they are. Because our composite, by necessity, includes some data from urban stations (particularly for the early years), it is **not** a completely rural composite. Therefore, we refer to it as a "relatively rural composite". It should be **less** affected by urbanization bias than estimates constructed from all stations, but it may still contain some urbanization bias.

The error bars are greatest when the numbers of available stations are low. Therefore, the uncertainties associated with our composite increase the further back in time we consider. Because there are very few stations with data before the 1950s, and almost none for the 19th century, the error bars are fairly wide for the early 20th century and very wide for the 19th century. Nonetheless, the composite suggests that the 1940s were relatively warm, if slightly cooler than the recent warm period. It is also possible that there were some relatively warm periods during the 19th century, but given the large error bars for this period, this possibility should be considered cautiously.

3.4. Comparison of different estimates of Chinese temperature trends

As a simple metric for comparing the various estimates of Chinese temperature trends, we calculated the hottest year for each of the estimates for the 1901-1950 period, and separately for the 1951-present period. In Table 1, we have ranked all of the estimates according to the differences (ΔT) between these two peak years. That is, according to the estimate at the start of the table - Soon et al. (2015) – the recent peak year (2007) was *colder* than the early 20th century peak year (1946) by 0.45°C, while according to the estimate at the end of the table – Li et al. (2017) – the recent peak year (also 2007) was *hotter* than the earlier peak year (also 1946) by 1.12°C. The full data series are also plotted in **Figure 10** using this same ranking.

Although this metric for comparing the various estimates is rather crude, it nonetheless provides some useful insights. For instance, we note that – aside from the two computer model-based hindcasts (i.e., the 20th Century Reanalysis and CMIP5 multimodel mean) – all of the estimates identify the same year (1946) as being the hottest year in the 1901-1950 period. Similarly, all of the estimates except for the two hindcasts identify one of two years (1998 or 2007) as being the hottest for the 1951-present period. This suggests that, despite the differences between each of the estimates, they are closely related. It also highlights that the computer model-based hindcasts are not very good at simulating the observed peaks.

Table 1. Linear trends for the early (1901-1950) and late (1951-2000) 20th century, along with the hottest years during the 1901-1950 and 1951-present periods for each of the sixteen reconstructions in Figure 5, ranked according to the differences between the peak years. Temperatures (°C) are relative to the 1901-2000 annual means.

				1901-2000	1901-1950			1951-2000	1951-end of series		Differences	
#	Study	Reconstruction details	Period	Trend (°C/10y)	Trend (°C/10y)	Peak T (°C)	Peak Year	Trend (°C/10y)	Peak T (°C)	Peak Year	Δ Trend (°C/10y)	ΔT (°C)
1	Soon et al., 2015	GHCN v3 (raw), mostly rural composite	1841-2014	+0.00	+0.25	1.30	1946	+0.10	0.85	2007	-0.15	-0.45
2	This study	GHCN v3 (raw), all stations	1841-2014	+0.05	+0.23	1.05	1946	+0.18	1.21	1998	-0.05	0.16
3	This study	CRUTEM v3 (homogenized)	1841-2010	+0.07	+0.21	1.01	1946	+0.19	1.34	1998	-0.02	0.33
4	Ding et al., 2014	China Climate Change Monitoring Bulletin	1901-2013	+0.06	+0.22	1.03	1946	+0.16	1.44	2007	-0.06	0.41
~5	This study	GHCN v4 (raw), relatively rural composite	1841-2016	+0.06	+0.20	0.97	1946	+0.16	1.39	1998	-0.04	0.42
~5	This study	GHCN v3 (MW09-adjusted), all stations	1841-2014	+0.08	+0.19	0.91	1946	+0.19	1.33	1998	0.00	0.42
7	Tang & Ren, 2005	Max/Min (*updated by Ren et al., 2017*)	1901-2016	+0.06	+0.23	1.03	1946	+0.17	1.49	2007	-0.06	0.46
8	Li et al., 2017	20th Century Reanalysis hindcast	1900-2005	+0.11	+0.09	0.25	1944	+0.13	0.73	1994	+0.04	0.48
9	Li et al., 2017	Li & Xu records (raw)	1900-2015	+0.07	+0.18	0.99	1946	+0.18	1.54	2007	0.00	0.55
10	This study	GHCN v4 (raw), all stations	1841-2016	+0.08	+0.20	0.85	1946	+0.18	1.45	1998	-0.02	0.60
11	This study	CRUTEM v4 (homogenized)	1841-2016	+0.08	+0.14	1.03	1946	+0.20	1.66	2007	+0.06	0.63
12	Wang et al., 2004	Raw + proxies (*updated by Ren et al., 2017*)	1880-2015	+0.04	+0.24	0.79	1946	+0.17	1.48	2007	-0.07	0.69
13	Li et al., 2017	CMIP5 regional China hindcast mean	1900-2005	+0.06	+0.08	0.09	1948	+0.12	0.83	2004	+0.04	0.74
14	This study	GHCN v4 (MW09-adjusted), all stations	1841-2016	+0.11	+0.20	0.88	1946	+0.19	1.63	1998	-0.01	0.75
15	This study	GHCN v4 (raw), most urban subset	1841-2016	+0.08	+0.18	0.78	1946	+0.18	1.69	2007	0.00	0.91
16	Li et al., 2017	Li & Xu records (Li10-adjusted)	1900-2015	+0.09	+0.10	0.48	1946	+0.18	1.60	2007	+0.08	1.12
		Average values		***+0.07***	***+0.18***	***0.84***	***1946***	***+0.17***	***1.35***	***1998/2007***	***-0.02***	***0.51***

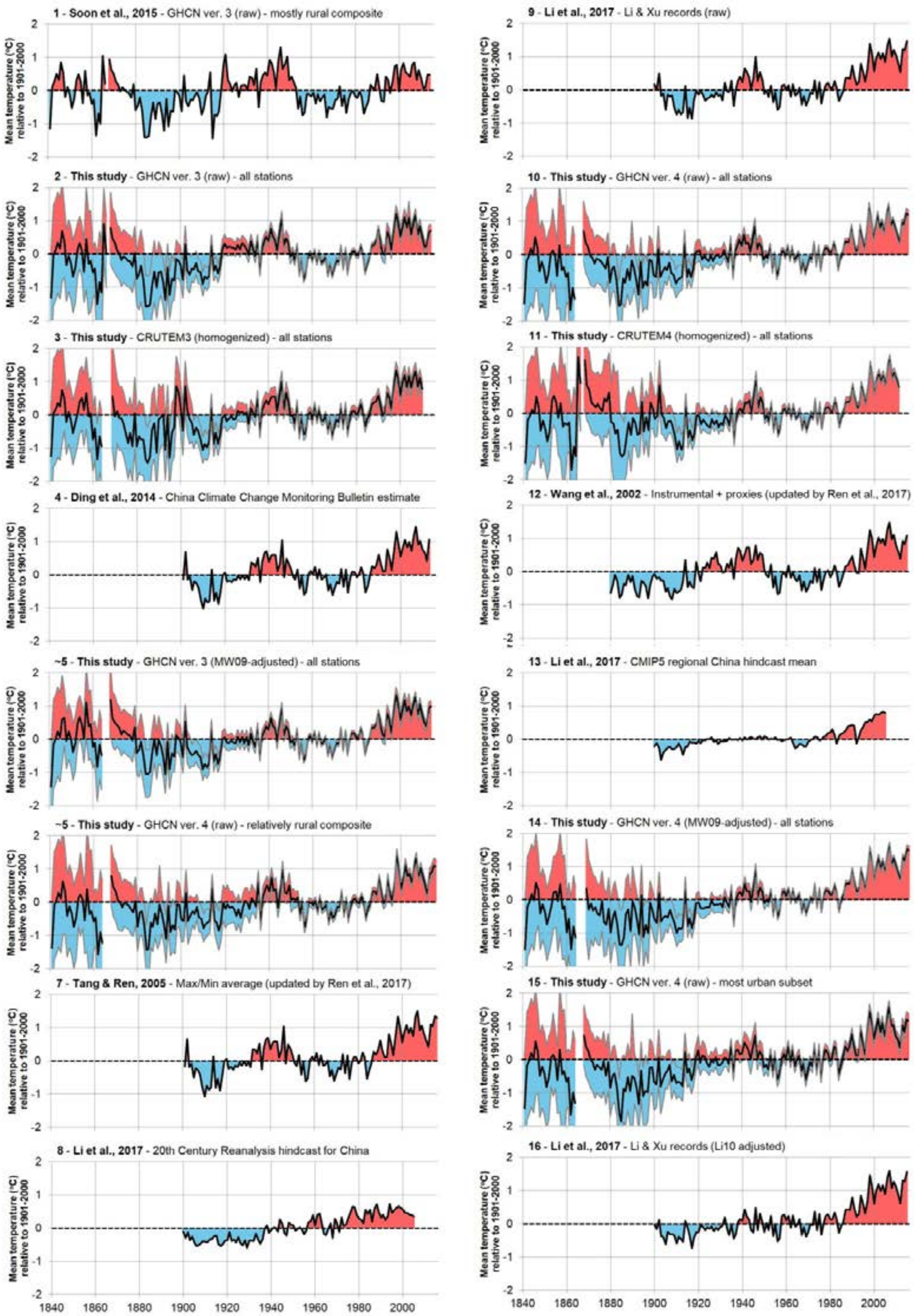

Figure 10. Various instrumental-based annual temperature reconstructions (and climate model hindcasts) for China. When error bars were provided, these are indicated with gray curves. All reconstructions are shown relative to their 20th century mean (1901-2000). Values above and below that mean are highlighted with red and blue shading, respectively.

The fact that many of the estimates imply 1998 is the hottest year in the 1951-present period is consistent with the suggestion that recent decades have involved a "warming hiatus" for Chinese temperatures, e.g., (Li et al., 2015; Chen & Zhai, 2017; Sun et al., 2017b). However, we note that several studies have suggested that this hiatus did not occur for some regions within China, e.g., the Tibetan Plateau (Yan & Liu, 2014; Duan & Xiao, 2015; An et al., 2017).

In Table 1, we also present the linear trends for each of the series over three periods: 1901-1950, 1951-2000 and 1901-2000. Interestingly, nine of the 16 series suggest a greater warming rate for the 1901-1950 period than for the 1951-2000 period. Moreover, only four of the series suggest that the warming rate increased for the 1951-2000 period and two of these series were the model-based hindcasts.

From Table 1 and **Figure 10** we can see several factors which appear to influence the relative magnitudes of the peaks in the estimates:

- Reducing the contribution from urbanized stations increases the apparent warmth of the early 20th century warm period and decreases the apparent warmth of the current warm period.
- The two computer model-based hindcasts simulated almost no early 20th century warm period.
- The transition from GHCN version 3 to version 4 reduces the apparent warmth of the early 20th century warm period and increased the apparent warmth of the current warm period. The Li & Xu datasets seem more comparable to version 4 of the GHCN than version 3. The transition from CRUTEM3 to CRUTEM4 also seems to have had similar effects.
- Using the homogenized versions of the datasets tends to reduce the apparent warmth of the early 20th century warm period and increase the apparent warmth of the current warm period – compare the "raw" and "adjusted" versions of "GHCN v3, all", "GHCN v4, all" and "Li & Xu records".

As Li et al. (2017) have already noted, the use of homogenized datasets brings the estimates more in line with the computer model-based hindcasts. If the hindcasts are reliable then this would suggest that the homogenized datasets are more reliable. On the other hand, the use of homogenized datasets also brings the estimates more in line with the urbanized subsets than the rural subsets. This is consistent with Soon et al. (2015)'s argument that the current homogenization approach leads to urban blending for the Chinese temperature series.

At any rate, although in general homogenization appears to reduce the relative warmth of the early period and increase the relative warmth of the recent period, other factors seem to be involved in determining the relative warmth of the two periods, e.g., reducing the number of urban stations seems to increase the relative warmth of the early 20th century. For instance, if we consider the relative warmth of the early period, the 3rd warmest of the 16 series was a homogenized series (i.e., the CRUTEM3). On the other hand, the 2nd coldest was a non-homogenized series (i.e., the most urban GHCN v4 subset).

It is unclear why the change from version 3 to version 4 of the GHCN dataset has reduced the apparent warmth of the early 20th century warm period and increased the apparent warmth of the current warm period. However, it is worth noting that most of the changes introduced by version 4 seem to be for the post-1951 period (and particularly post-1990). So, while version 4 has significantly increased the amount of available data for the current warm period, the data available to assess the early 20th century warm period is still relatively limited.

Tang & Ren (2005) noted that before 1950, there was no unified time of observation for China, and this might have introduced non-climatic "time of observation biases" into the raw data for the early 20th century. So, it is worth considering whether this could have led to an overestimation or underestimation of the warmth of the early 20th century warm period. Tang & Ren attempted to reduce the magnitude of any such time of observation biases in the raw data by using the average of maximum and minimum temperature series instead of using the standard mean temperature series. Others have suggested that using homogenized temperature records should have reduced these biases, e.g., Li et al. (2017).

As noted earlier, the effects of homogenization on Chinese temperature estimates is generally to reduce the relative warmth of the early period and increase the relative warmth of the recent period. This has led some groups to conclude that the net effect of the various non-climatic biases (e.g., time of observation bias) on the early 20th century was to exaggerate the warmth of the early warm period, e.g., Li et al. (2017). However, if Soon et al. (2015) are correct and the homogenization has been affected by urban blending, then this would provide an alternative explanation.

In theory, we might get some idea of the potential impact of time of observation biases on the early 20th century warm period by comparing the homogenized series and Tang & Ren series to the raw series. Comparing the

Tang & Ren series to the Wang et al. (2004) series which was based on raw data, we see that the Tang & Ren approach has apparently reduced the warmth of the 1920s, but both series have a relatively warm 1940s period. Indeed, the peak year (1946) is slightly warmer in the Tang & Ren series. Moreover, the peak warmth of the 1940s period for the Wang et al. (2004) series is actually ranked quite low in our list (12 out of 16), and several homogenized series have a warmer 1940s peak. Therefore, while it is possible that some of the apparent warmth of the 1920s in the raw series may be due to time of observation biases, this does not seem to explain the 1940s warm period.

3.5. Comparison of temperature proxy reconstructions for Chinese regions

Temperature proxies are essential for studying Chinese temperature trends on multi-decadal and centennial timescales, since instrumental records are relatively short. For this reason, in the last decade or so, there have been considerable advances within the field of paleoclimate within China, e.g., see Ge et al. (2016) for a recent review. Using temperature proxies, several of us have suggested elsewhere that Chinese temperatures may have been comparable to the recent warm period approximately 1,000 years ago, during the "Medieval Warm Period", e.g., Ge et al. (2013b, 2017), Yan et al. (2015). However, in this paper, we will confine our analysis of Chinese temperature proxy series to temperature trends since the 19th century.

Figure 11 compares different temperature proxy series for three regions in China (central China, northeast China and the Tibetan Plateau) as well as five multiproxy reconstructions for all of China. As discussed in Section 2.6, different types of proxies are presented for each of the three regions, and proxies for different seasons are included.

All of the proxy series suggest there have been considerable changes in regional temperatures since the 19th century. However, on these relatively short time-scales, there seem to be considerable inconsistencies between the temperature fluctuations implied by each of the proxy series – even for the same regions. In particular, there are notable differences between individual series on the timing, length and magnitude of different warm and cold periods. Although most of the proxy series suggest the existence of relatively warm periods for both the recent period and the early 20th century, the relative warmth of the two periods varies between proxies. Also, in some cases, one of other of the warm periods seems to be absent. Unfortunately, this makes it challenging to use temperature proxies for directly comparing the early 20th century and current warm periods.

Part of the reason for these inconsistencies could be due to genuine differences between the climatic trends for different regions. However, inconsistencies also exist within individual regions. For instance, for central China, while Yi et al. (2012) implies that the early 20th century warm period was hotter than the current warm period, Tan et al. (2013) implies the opposite result, and Chen et al. (2014) suggests that neither warm period was particularly noteworthy. Similarly, for the Tibetan Plateau, both Fan et al. (2010) and Shi et al. (2016) imply that the early 20th century warm period was warmer, while Wang et al. (2015) implies a slightly warmer current period and Yang et al. (2008) implies that the 1920s-40s included a lot of relatively cold years.

Another factor which could be involved is the differences between seasons. In our analysis of the instrumentally-based series in the previous section, we were considering the annual mean temperature trends. However, most temperature proxies are typically for one particular season, e.g., summer or winter, and climatic trends can sometimes be different for different seasons, e.g., (Li et al., 2015; Sun et al., 2017b; Yan et al., 2015; Kawakubo et al., 2017). On the other hand, we can see that, even for the same region and season, different proxies can imply different trends. For example, three of the four Tibetan Plateau proxy series in Figure 11 are summer temperature proxies.

An additional factor which can influence the relative warmth of the different periods is the reconstruction method applied. Zheng et al. (2015) have applied two alternative reconstruction methods to the same multi-proxy series for a region in northwest China (Xinjiang). Using a fairly standard reconstruction method yielded a reconstruction that agreed there had been warming since the 1970s, but suggested the 1930s-40s were a relatively cold period, and that it was about as warm in the mid-19th century as at the end of the 20th century. On the other hand, when they applied an alternative reconstruction method which separated the low-frequency and high-frequency components of the proxy data, their reconstruction suggested a relatively warm 1930s-40s period, and also reduced the warmth of the mid-19th century, while increasing the warmth of the recent warm period.

Differences between alternative reconstruction methods are also apparent (although less dramatically so) by comparing the Ge et al. (2017) "PLS" and "PCR" reconstructions for all-China in Figure 11. Both reconstructions used the same proxy dataset, but the PCR method implies a slightly warmer and longer early-20th century warm period, as well as a warmer late 19th century than the PLS method.

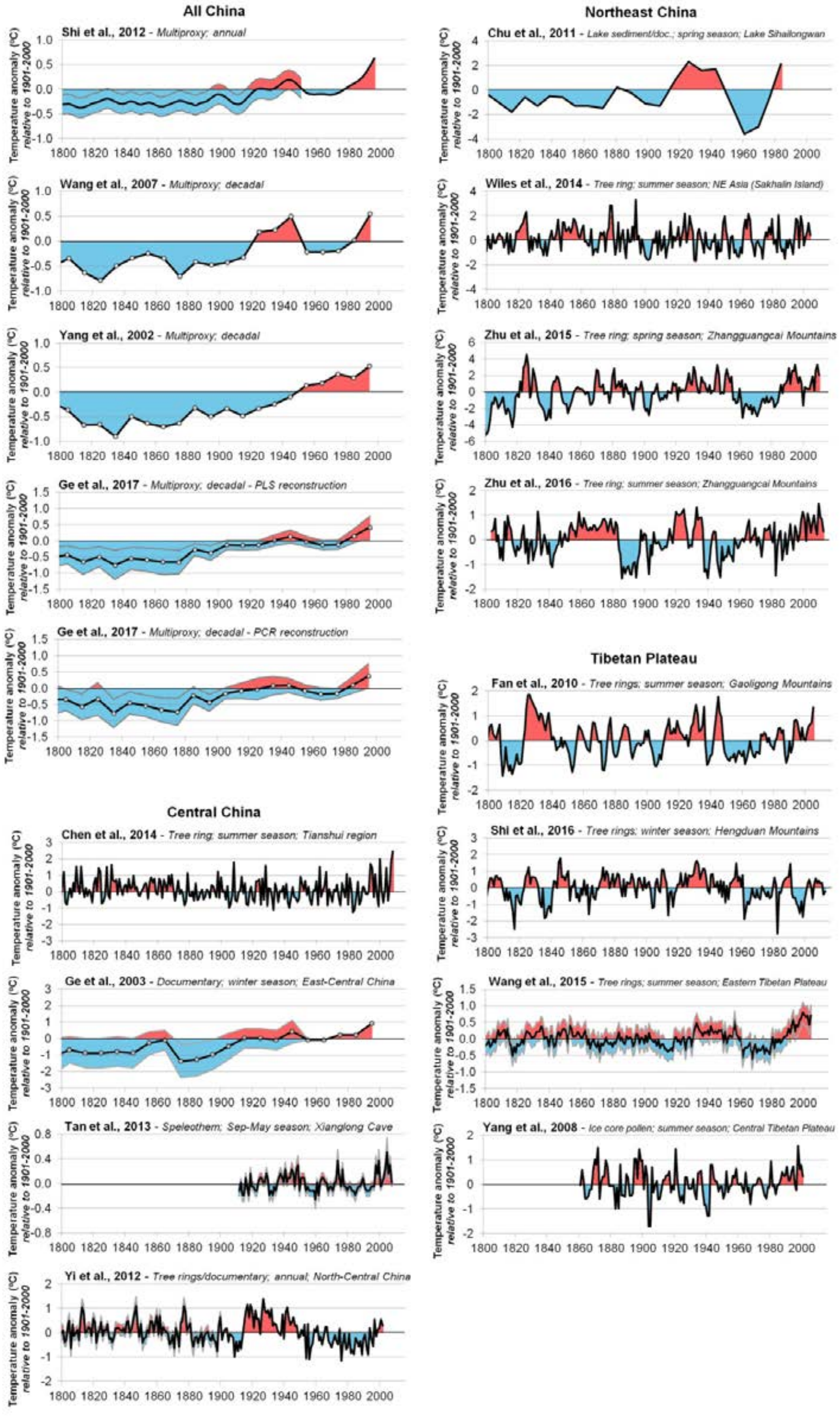

Figure 11. Examples of various temperature proxy reconstructions for three regions in China (Central, Northeastern and Tibetan Plateau) and for all China. When error bars were provided, these are indicated with gray curves. All reconstructions are shown

relative to their 20th century mean (1901-2000). Values above and below that mean are highlighted with red and blue shading, respectively.

So before temperature proxies can be satisfactorily used to address this specific issue, more research is probably required to understand the reasons for these differences. Ge et al. (2016) have highlighted several issues in which further research is needed, as well as summarising the many advances which have been made in the last 10 years.

4. Conclusions

In this collaborative paper, we analysed and reviewed some of the main reasons why there is such ongoing debate over how the early 20th century warm period (1920s-1940s) compared to the current warm period (1990s-present) in China.

Most Global Climate Model (GCM) hindcasts are currently unable to simulate the early 20th century warm period (Li et al., 2017; Zhou & Yu, 2006; Guo et al., 2013). This has led some to suggest that its apparent warmth has been overestimated (perhaps due to non-climatic biases), e.g., Li et al. (2017). On the other hand, Soon et al. (2011, 2015) have argued that both warm periods and the intervening cool period can be explained almost entirely in terms of natural climate change.

There is general agreement that there are individual stations in China from highly urbanized areas such as Beijing (Ren et al., 2007; Zhang L. et al., 2014; Zhang Y. & Ren, 2014; Yan et al., 2010; Wang et al., 2013) which are significantly affected by urbanization bias, and that this has introduced a warming trend bias into their station records. However, there is considerable ongoing debate over the exact extent to which urbanization bias has affected estimates of national temperature trends.

In this study, we assessed the extent of urbanization bias among the Chinese component of the new, substantially revised, version of the Global Historical Climatology Network (version 4, currently in beta production). We found that urbanization bias substantially altered the apparent relative warmth of the two periods. Specifically, the more that urban stations were removed from the analysis, the warmer the early 20th century warm period became, and the cooler the recent warm period became.

Some groups have argued that the application of statistically-based homogenization procedures such as Menne & Williams (2009) has substantially reduced the magnitude of urbanization biases and other non-climatic biases in the data, e.g., (Xu et al., 2017; Yan et al., 2016; Wang J. & Yan, 2016). Others have argued that urbanization biases are still a problem after homogenization, e.g., (Ren et al., 2008, 2015, 2017; Ren & Ren, 2011; He & Jia, 2012; He et al., 2013; Ge et al., 2013a; Yang et al., 2013; Yang et al., 2011; Soon et al., 2015). Meanwhile, Soon et al. (2015) have argued that these homogenization procedures are inadvertently introducing substantial warming biases into the Chinese records through urban blending. We provided a theoretical description of why this blending problem occurs (Sec. 3.2.3); a case study of the problem for the Beijing area (Sec. 3.2.4); and some suggestions of how to reduce the problem (Sec. 3.2.5) in Section 3.2.

Although version 4 of the GHCN dataset has increased the amount of data for both the early and recent warm periods, the available data before about 1954 is still quite limited. This is an even bigger problem for rural stations. For this reason, the ACRE China project which aims to recover and digitize more of this early 20th century data for China is particularly important (Williamson et al., 2017).

Another possible approach to overcoming this shortage would be to supplement the limited instrumental records with temperature proxy series, for instance, as had been done by Wang S. et al. (2001, 2004). Additionally, these proxy series could then be used for extending our estimates of Chinese temperature trends back into the pre-instrumental era, e.g., Wang S. et al. (2007). Unfortunately, an analysis of a sample of 12 temperature proxy series (taken from three separate regions in China) reveals that there are still a lot of inconsistencies between individual proxy series. Therefore, more research into resolving the reasons for these inconsistencies is recommended.

In this review, we mostly focused on annual mean temperatures averaged over all of China. However, temperature trends often vary from season to season, e.g., (Li et al., 2015; Sun et al., 2017b). Also, different regions within China often show different climatic trends, e.g., Sun et al. (2017b) found a very pronounced "hiatus" for northeast China, but Yan & Liu (2014) found none for the Tibetan Plateau. Therefore, it is also important to consider seasonality and regionality.

We have identified several key questions with regards to the current and early 20th century warm periods in China which have still not been satisfactorily resolved:

- Is the apparent warmth of the 1920s-40s in China merely an artefact of non-climatic biases in the observational data, as suggested by Li et al. (2017)?
- If not, and the 1920s-40s warmth was genuine, then why are most of the current GCMs unable to reproduce it? Zhou & Yu (2006) and Soon et al. (2011, 2015) have suggested that the current GCMs may be significantly underestimating natural climate changes.
- Similarly, for the current warm period, has there been a "warming hiatus" (Li et al., 2015; Chen & Zhai, 2017; Sun et al., 2017b; An et al., 2017; Yan & Liu, 2014; Duan & Xiao, 2015) in recent decades, and if so, why?
- Are these apparent discrepancies between observations and modelled Chinese temperatures only a regional phenomenon for China? If so, what are the explanations for these regional variations? Zhou & Yu (2006), Soon et al. (2011) and Li et al. (2015) amongst others have suggested some possible explanations which might be worthy of further research.
- Alternatively, are these phenomena global (or at least hemispheric) in nature? Both Sun et al. (2017a) and Soon et al. (2015) have noted that the global temperature data before ~1940s is unfortunately quite limited, and might be substantially affected by non-climatic biases such as urbanization bias. Moreover, Soon et al. (2015) have argued that, after accounting for urbanization biases, similar 1920s-40s warm periods were observed across the entire northern hemisphere.

Acknowledgments We thank Dr. Jinbao Li and Prof. Bao Yang for providing some of the temperature proxy series, and Prof. Qingxiang Li for providing us with the data for the four Li et al. (2017) time series which we plotted in Figure 5. We also thank Prof. Guoyu Ren, Prof. Qingxiang Li, Prof. Bao Yang and Prof. Panmao Zhai for several useful comments and suggestions on an early draft of this manuscript. Finally, we thank the two anonymous reviewers whose comments led to significant improvements in the paper. All of the research in this collaborative paper was carried out during the authors' free time and at their own expense, and none of the authors received any funding for this research.

References

An W. L., S. G. Hou, Y. Y. Hu, and S. Y. Wu, 2017: Delayed warming hiatus over the Tibetan Plateau. *Earth. Space Sci.*, **4**, 128-137, doi: 10.1002/2016EA000179.

Bindoff N. L., P. A. Stott, K. M. AchutaRao, et al., 2013: Detection and attribution of climate change: from global to regional. *Climate change 2013: the physical science basis. Contribution of Working Group 1 to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, [Stocker T. F. et al. (eds)]. Cambridge University Press, Cambridge, UK and New York, NY, USA.

Brohan P., J.J. Kennedy, I. Harris, et al., 2006: Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *J. Geophys. Res.*, **111**, D12106, doi: 10.1029/2005JD006548.

Butler C. J., A. M. García Suárez, A. D. S. Coughlin, and C. Morrell, 2005: Air temperatures at Armagh Observatory, Northern Ireland, from 1796 to 2002. *Int. J. Climatol.*, **25**, 1055-1079, doi: 10.1002/joc.1148.

Cao L. J., P. Zhao, Z. W. Yan, et al., 2013: Instrumental temperature series in eastern and central China back to the nineteenth century. *J. Geophys. Res.*, **118**, 8197-8207, doi: 10.1002/jgrd.50615.

Chen F., and Y. Yuan, 2014: May-June maximum temperature reconstruction from mean earlywood density in north central China and its linkages to the summer monsoon activities. *PLos ONE*, **9**, e107501, doi: 10.1371/journal.pone.0107501.

Chen Y., and P. M. Zhai, 2017: Persisting and strong warming hiatus over eastern China during the past two decades. *Environ. Res. Lett.*, in press, doi: 10.1088/1748-9326/aa822b.

Chen Z. H., H. J. Wang, G. Y. Ren, et al., 2005: Change of urban heat island intensity and its effect on regional temperature series: A case study in Hubei Province. *Clim. Environ. Res.* (in Chinese), **10**, 771-779.

Chu G., Q. Sun, X. Wang, et al., 2011: Seasonal temperature variability during the past 1600 years recorded in historical documents and varved lake sediment profiles from northeastern China. *Holocene*, **22**, 785-792, doi: 10.1177/0959683611430413.

Chu Z. Y., and G. Y. Ren, 2005: Change in urban heat island magnitude and its effect on mean air temperature record in Beijing region. *Acta Meteor. Sinica* (in Chinese), **63**, 534-540.

Compo G. P., J. S. Whitaker, P.D. Sardeshmukh, et al., 2011: The Twentieth Century Reanalysis project. *Q. J. R. Meteorol. Soc.*, **137**, 1-28, doi: 10.1002/qj.776.

Compo G. P., P. D. Sardeshmuk, J. S. Whitaker, et al., 2013: Independent confirmation of global land warming without the use of station temperatures. *Geophys., Res., Lett.*, **40**, 3170-3174, doi: 10.1002/grl.50425.

deGaetano, A. T. 2006. Attributes of several methods for detecting discontinuities in mean temperature series. *J. Clim.*, **19**, 838-853. doi:10.1175/JCLI3662.1.

Ding L. L., Q. S. Ge, J. Y. Zheng, and Z. X. Hao, 2016: Variations in annual winter mean temperature in South China since 1736. *Boreas*, **45**, 252-259, doi: 10.1111/bor.12144.

Ding Y. H., Y. J. Liu, S. J. Liang, et al., 2014: Interdecadal variability of the East Asian Winter Monsoon and its possible links to global climate change. *J. Meteor. Res.*, **28**, 693-713, doi: 10.1007/s13351-014-4046-y.

Duan A. M., and Z. X. Xiao, 2015: Does the climate warming hiatus exist over the Tibetan Plateau? *Sci. Rep.*, **5**, 13711, doi: 10.1038/srep13711.

Easterling D. R., and T. C. Peterson, 1995: A new method for detecting undocumented discontinuities in climatological time series. *Int. J. Climatol.*, **15**, 369-377, doi: 10.1002/joc.3370150403.

Fall S., A. Watts, J. Nielsen-Gammon et al., 2011: Analysis of the impacts of station exposure on the U.S. Historical Climatology Network temperatures and temperature trends. *J. Geophys. Res.*, **116**, D14120, doi: 10.1029/2010JD015146.

Fan Z.-X., A. Bräuning, Q.-H. Tian, et al., 2010: Tree ring recorded May-August temperature variations since A.D. 1585 in the Gaoligong Mountains, southeastern Tibetan Plateau. *Palaeogeog. Palaeoclim. Palaeoecol.*, **296**, 94-102, doi: 10.1016/j.palaeo.2010.06.017.

Ge Q., J. Zheng, X. Fang, et al., 2003: Winter half-year temperature reconstruction for the middle and lower reaches of the Yellow River and Yangtze River, China, during the past 2000 years. *Holocene*, **13**, 933-940, doi: 10.1191/0959683603hl680rr.

Ge Q. S., F. Wang, and J. Luterbacher, 2013a: Improved estimation of average warming trend of China from 1951-2010 based on satellite observed land-use data. *Clim. Change*, **121**, 365-379, doi: 10.1007/s10584-013-0867-4.

Ge Q., Z. Han, J. Zheng, and X. Shao, 2013b: Temperature changes over the past 2000 yr in China and comparison with the Northern Hemisphere. *Clim. Past*, **9**, 1153-1160, doi: 10.5194/cp-9-1153-2013.

Ge Q. S., J. Y. Zheng, Z. X. Hao, et al., 2016: Recent advances on reconstruction of climate and extreme events in China for the past 2000 years. *J. Geogr. Sci.* **26**, 827-854, doi: 10.1007/s11442-016-1301-4.

Ge Q. S., H. L. Liu, X. Ma, et al., 2017: Characteristics of temperature change in China over the last 2000 years and spatial patterns of dryness/wetness during cold and warm periods. *Adv. Atmos. Sci.*, **34**, 941-951, doi: 10.1007/s00376-017-6238-8.

Guo J. X., L. Chen, H. H. Liang, and X. Lin, 2010: Representativeness evaluation of China's national baseline climate station network. TECO-2010 – WMO Technical Conference on Meteorological and Environmental Instruments and Methods of Observation. Helsinki, Finland, 30 August – 1 September 2010. url: http://www.wmo.int/pages/prog/www/IMOP/publications/IOM-104_TECO-2010/1_1_Guo_China.pdf [Accessed: 23/04/2018]

Guo Y., W.-J. Dong, F.-M. Ren, et al., 2013: Surface air temperature simulations over China with CMIP5 and CMIP3. *Adv. Clim. Change Res.*, **4**, 145-152, doi: 10.3724/SP.J.1248.2013.145.

Hartmann D. L., A. M. G. Klein Tank, M. Rusticucci, et al., 2013: Observations: atmosphere and surface. Climate change 2013: the physical science basis. Contribution of Working Group 1 to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker T. F. et al. (eds)]. Cambridge University Press, Cambridge, UK and New York, NY, USA.

Hausfather, Z., M. J. Menne, C. N. Jr. Williams, et al., 2013: Quantifying the effect of urbanization on U.S. Historical Climatology Network temperature records. *J. Geophys. Res.*, **118**, 481-494, doi:10.1029/2012JD018509.

He Y. T., and G. S. Jia, 2012: A dynamic method for quantifying natural warming in urban areas. *Atmos. Oceanic Sci. Lett.*, **5**, 408-413, doi: 10.1080/16742834.2012.11447029.

He Y. T., G. S. Jia, Y. H. Hu, and Z. J. Zhou, 2013: Detecting urban warming signals in climate records. *Adv. Atmos. Sci.*, **30**, 1143-1153. doi: 10.1007/s00376-012-2135-3.

Hua L. J., Z. G. Ma, and W. D. Guo, 2008: The impact of urbanization on air temperature across China. *Theor. Appl. Climatol.*, **93**, 179-194, doi: 10.1007/s00704-007-0339-8.

Jones P. D., P. Y. Groisman, M. Coughlan, et al., 1990: Assessment of urbanization effects in time series of surface air temperature over land. *Nature*, **347**, 169-172, doi: 10.1038/347169a0.

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

Jones P. D., D. H. Lister, T. J. Osborn, et al., 2012: Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010. *J. Geophys. Res.*, **117**, D05127, doi: 10.1029/2011JD017139.

Karl T. R., C. N. Williams, P. J. Young, and W. M. Wendland, 1986: A model to estimate the time of observation bias associated with monthly mean maximum, minimum and mean temperatures for the United States. *J. Climate Appl. Meteor.*, **25**, 145-160, doi: 10.1175/1520-0450(1986)025<0145:AMTETT>2.0.CO;2.

Karl T. R., and C. N. Jr. Williams, 1987: An approach to adjusting climatological time series for discontinuous inhomogeneities. *J. Climate Appl. Meteorol.*, **26**, 1744-1762, doi: 10.1175/1520-0450(1987)026<1744:AATACT>2.0.CO;2.

Karl, T. R., H. F. Diaz, and G. Kukla, 1988: Urbanization: its detection and effect in the United States climate record. *J. Clim.*, **1**, 1099-1123, doi: 10.1175/1520-0442(1988)001<1099:UIDAEI>2.0.CO;2.

Kawakubo Y., C. Alibert, and Y. Yokoyama, 2017: A reconstruction of subtropical western North Pacific SST variability back to 1578, based on a *Porites* Coral Sr/Ca record from the Northern Ryukyus, Japan. *Paleoocean.*, **32**, 1352-1370, doi: 10.1002/2017PA003203.

Lakatos, M., T. Szentimrey, Z. Bihari, and S. Szalai. 2013. "Creation of a homogenized climate database for the Carpathian region by applying the MASH procedure and the preliminary analysis of the data". *Quarterly Journal of the Hungarian Meteorological Service*. 117:143-158.

Lawrimore J. H., M. J. Menne, B. E. Gleason, et al., 2011: An overview of the Global Historical Climatology Network monthly mean temperature data set, version 3. *J. Geophys. Res. Atmos.*, **116**, D19121, doi: 10.1029/2011JD016187.

Li Q., H. Zhang, X. Liu, et al., 2004: Urban heat island effect on annual mean temperature during the last 50 years in China. *Theor. Appl. Climatol.*, **79**, 165-174, doi: 10.1007/s00704-004-0065-4.

Li Q. X., W. Li, P. Si, et al., 2010a: Assessment of surface air warming in northeast China, with emphasis on the impacts of urbanization. *Theor. Appl. Climatol.*, **99**, 469-478, doi: 10.1007/s00704-009-0155-4.

Li Q. X., W. J. Dong, W. Li, et al., 2010b: Assessment of the uncertainties in temperature change in China during the last century. *Chin. Sci. Bull.*, **55**, 1974-1982, doi: 10.1007/s11434-010-3209-1.

Li Q. X., S. Yang, W. H. Xu, et al., 2015: China experiencing the recent warming hiatus. *Geophys. Res. Lett.*, **42**, 889-898. doi: 10.1002/2014GL062773.

Li Q., L. Zhang, W. Xu, et al., 2017: Comparisons of time series of annual mean surface air temperature for China since the 1900s: Observations, model simulations and extended reanalysis. *Bull. Amer. Meteor. Soc.*, **98**, 699-711, doi: 10.1175/BAMS-D-16-0092.1.

Li Y., L. J. Zhu, X. Y. Zhao, et al., 2013: Urbanization impact on temperature change in China with emphasis on land cover change and human activity. *J. Clim.*, **26**, 8765-8780, doi: 10.1175/JCLI-D-12-00698.1.

Li Y. B., T. Shi, Y. J. Yang, et al., 2015: Satellite-based investigation and evaluation of the observational environment of meteorological stations in Anhui province, China. *Pure Appl. Geophys.*, **172**, 1735-1749, doi: 10.1007/s00024-014-1011-8.

Li Z., and Z. W. Yan, 2010: Application of multiple analysis of series for homogenization to Beijing daily temperature series (1960-2006). *Adv. Atmos. Sci.*, **27**, 777-787, doi: 10.1007/s00376-009-9052-0.

Li Z., Z. W. Yan, and W. Hong, 2015: Updated homogenized Chinese temperature series with physical consistency. *Atmos. Oceanic Sci. Lett.*, **8**, 17-22, doi: 10.3878/AOSL20140062.

Lin X., R. A. Sr. Pielke, R. Mahmood, et al., 2015: Observational evidence of temperature trends at two levels in the surface layer. *Atmos. Chem. Phys.*, **16**, 827-841, doi: 10.5194/acp-16-827-2016.

Liu X. J., G. J. Tian, J. M. Feng, et al., 2018: Modeling the warming impact of urban land expansion on hot weather using the Weather Research and Forecasting model: A case study of Beijing, China. *Adv. Atmos. Sci.*, **35**, 723-736, doi: 10.1007/s00376-017-7137-8.

Liu Y., J. Y. Zheng, Z. X. Hao, and X. Z. Zhang, 2017: Unprecedented warming revealed from multi-proxy reconstruction of temperature in southern China for the past 160 years. *Adv. Atmos. Sci.*, **34**, 977-982, doi: 10.1007/s00376-017-6228-x.

Liu Z., Y. Liu, S. Wang, et al., 2018: Evaluation of spatial and temporal performances of ERA-Interim precipitation and temperature in mainland China. *J. Clim.*, **31**, 4347-4365, doi: 10.1175/JCLI-D-17-0212.1.

McNider R. T., G. J. Steeneveld, A. A. M. Holtslag, et al., 2012: Response and sensitivity of the nocturnal boundary layer over land to added longwave radiative forcing. *J. Geophys. Res.*, **117**, D14106, doi: 10.1029/2012JD017578.

Menne M. J., and C. N. Jr. Williams, 2009: Homogenization of temperature series via pairwise comparisons. *J. Clim.*, **22**, 1700-1717, doi: 10.1175/2008JCLI2263.1.

Menne, M. J., C. N. Jr. Williams, and M. A. Palecki. 2010: On the reliability of the U.S. surface temperature record. *J. Geophys. Res.*, **115**, D11108, doi:10.1029/2009JD013094.

Menne M. J., I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston, 2012: An overview of the Global Historical Climatology Network-Daily dataset. *J. Atm. Oceanic Tech.*, **29**, 897-910. doi:10.1175/JTECH-D-11-00103.1.

Mitchell, J. M. Jr. 1953: On the causes of instrumentally observed secular temperature trends". *J. Meteorol.*, **10**, 244-261, doi:10.1175/1520-0469(1953)010<0244:OTCOIO>2.0.CO;2.

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

Pielke R. A. Sr., T. Stohlgren, L. Schell, et al., 2002: Problems in evaluating regional and local trends in temperature: an example from eastern Colorado, USA. *Int. J. Climatol.*, **22**, 421-434, doi: 10.1002/joc.706.

Pielke R. A. Sr., C. Davey, and J. Morgan, 2004: Assessing "global warming" with surface heat content. *Eos*, **85**, 210-211, doi: 10.1029/2004EO210004.

Pielke, R. A. Sr., J. Nielsen-Gammon, C. Davey, et al., 2007a: Documentation of uncertainties and biases associated with surface temperature measurement sites for climate change assessment. *Bull. Amer. Meteorol. Soc.*, **88**, 913-928, doi: 10.1175/BAMS-88-6-913.

Pielke R. A. Sr., C. Davey, D. Niyogi, et al., 2007b: Unresolved issues with the assessment of multi-decadal global land surface temperature trends. *J. Geophys. Res.*, **112**, D24S08, doi: 10.1029/2006JD008229.Portman D., 1993: Identifying and correcting urban bias in regional time series: surface temperature in China's northern plain. *J. Clim.*, **6**, 2298-2308, doi: 10.1175/1520-0442(1993)006<2298:IACUBI>2.0.CO;2.

Ren G. Y., 2015: Urbanization as a major driver of urban climate change. *Adv. Clim. Change Res.*, **6**, 1-6, doi: 10.1016/j.accre.2015.08.003.

Ren G. Y., Z. Y. Chu, Y. Q. Zhou, et al., 2005: Recent progresses in studies of regional temperature changes in China. *Clim. Environ. Res.* (in Chinese), **10**, 701-716.

Ren G. Y., Z. Y. Chu, Z. H. Chen, and Y. Y. Ren, 2007: Implications of temporal change in urban heat island intensity observed at Beijing and Wuhan stations. *Geophys. Res. Lett.*, **34**, L05711, doi: 10.1029/2006GL027927.

Ren G. Y., Y. Q. Zhou, Z. Y. Chu, et al., 2008: Urbanization effects on observed surface air temperature trends in north China. *J. Clim.*, **21**, 1333-1348, doi: 10.1175/2007JCLI1348.1.

Ren Y. Y., and G. Y. Ren, 2011: A remote-sensing method of selecting reference stations for evaluating urbanization effect on surface air temperature trends. *J. Clim.*, **24**, 3179-3189, doi: 10.1175/2010JCLI3658.1.

Ren G. Y., Y. H. Ding, Z. C. Zhao, et al., 2012: Recent progress in studies of climate change in China. *Adv. Atmos. Sci.*, **29**, 958-977, doi: 10.1007/s00376-012-1200-2.

Ren G. Y., J. Li, Y. Y. Ren, et al., 2015: An integrated procedure to determine a reference station network for evaluating and adjusting urban bias in surface air temperature data. *J. Appl. Meteor. Climatol.*, **54**, 1248-1266, doi: 10.1175/JAMC-D-14-0295.1.

Ren G. Y., Y. H. Ding, and G. L. Tang, 2017: An overview of mainland China temperature change research. *J. Meteor. Res.*, **31**, 3-16, doi: 10.1007/s13351-017-6195-2.

Rennie J. J., J. H. Lawrimore, B. E. Gleason, et al., 2014: The international surface temperature initiative global land surface databank: monthly temperature data release description and methods. *Geosci. Data J.*, **1**, 75-102, doi: 10.1002/gdj3.8.

Shi F., B. Yang, and L. Von Gunten, 2012: Preliminary multiproxy surface air temperature field reconstruction for China over the past millennium. *Sci. China Earth Sci.*, **55**, 2058-2067, doi: 10.1007/s11430-012-4374-7.

Shi S., J. Li, J. Shi, et al., 2016: Three centuries of winter temperature change on the southeastern Tibetan Plateau and its relationship with the Atlantic Multidecadal Oscillation. *Clim. Dyn.*, in press, doi: 10.1007/s00382-016-3381-3.

Shi T., Y. Huang, W. Hong, et al., 2015: Influence of urbanization on the thermal environment of meteorological station: Satellite-observed evidence. *Adv. Clim. Change Res.*, **6**, 7-15, doi: 10.1016/j.accre.2015.07.001.

Soon W., K. Dutta, D. R. Legates, et al., 2011: Variation in surface air temperature of China during the 20^{th} century. *J. Atmos. Sol. Terr. Phys.*, **73**, 2331-2344, doi: 10.1016/j.jastp.2011.07.007.

Soon W., R. Connolly, and M. Connolly, 2015: Re-evaluating the role of solar variability on Northern Hemisphere temperature trends since the 19^{th} century. *Earth. Sci. Rev.*, **150**, 409-452, doi: 10.1016/j.earscirev.2015.08.010.

Stewart I. D., and T. R. Oke, 2012: Local climate zones for urban temperature studies. *Bull. Amer. Meteor. Soc.*, **93**, 1879-1900, doi: 10.1111/j.1541-0064.2000.tb.00709.x.

Sun Y., X. Zhang, G. Ren, et al., 2016: Contribution of urbanization to warming in China. *Nature Clim. Change.*, **6**, 706-709, doi: 10.1038/nclimate2956.

Sun X. B., G. Y. Ren, W. H. Xu, et al., 2017b: Global land-surface air temperature change based on the new CMA GLSAT data set. *Sci. Bull.*, **62**, 236-238, doi 10.1016/j.scib.2017.01.017.

Sun X. B., G. Y. Ren, Y. Y. Ren, et al., 2017b: A remarkable climate warming hiatus over Northeast China since 1998. *Theor. Appl. Climatol.*, in press, doi: 10.1007/s00704-017-2205-7.

Tan L., L. Yi, Y. Cai, et al., 2013: Quantitative temperature reconstruction based on growth rate of annually-layered stalagmite: a case study from central China. *Quat. Sci. Rev.*, **72**, 137-145, doi: 10.1016/j.quascirev.2013.04.022.

Tang G. L., and G. Y. Ren GY, 2005: Reanalysis of surface air temperature change of the last 100 years over China. *Clim. Environ. Res.*, **10**, 791-798 (in Chinese)

Tang G. L., Y. H. Ding, S. W. Wang, et al., 2010: Comparative analysis of China surface air temperature series for the past 100 years. *Adv. Clim. Change Res.*, **1**, 11-19, doi: 10.3724/SP.J.1248.2010.00011.

Taylor K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experimental design. *Bull. Amer. Meteor. Soc.*, **93**, 485-498, doi: 10.1175/BAMS-D-11-00094.1.

Vincent L. A., X. L. Wang, E. J. Milewska, et al., 2012: A second generation of homogenized Canadian monthly surface air temperature for climate trend analysis. *J. Geophys. Res.*, **117**, D18110, doi: 10.1029/2012JD017859.

Wang F., and Q. S. Ge, 2012: Estimation of urbanization bias in observed surface temperature change in China from 1980 to 2009 using satellite land-use data. *Chin. Sci. Bull.*, **57**, 1708-1715, doi: 10.1007/s11434-012-4999-0.

Wang F., Q. S. Ge, S. W. Wang, et al., 2015: A new estimation of urbanization's contribution to the warming trend in China. *J. Clim.*, **28**, 8923-8938, doi :10.1175/JCLI-D-14-00427.1.

Wang J., Z. W. Yan, L. Zhen, et al., 2013: Impact of urbanization on changes in temperature extremes in Beijing during 1978-2008. *Chin. Sci. Bull.*, **58**, 4679-4686, doi: 10.1007/s11434-013-5976-y.

Wang J. F., C. D. Xu, M. G. Hu, et al., 2014: A new estimate of the China temperature anomaly series and uncertainty assessment in 1900-2006. *J. Geophys. Res. Atmos.*, **119**, 1-9, doi: 10.1002/2013JD020542.

Wang J., B. Yang, and F. C. Ljungqvist, 2015: A millennial summer temperature reconstruction for the eastern Tibetan Plateau from tree-ring width. *J. Clim.*, **28**, 5289-5304, doi: 10.1175/JCLI-D-14-00738.1.

Wang J., and Z. W. Yan, 2016: Urbanization-related warming in local temperature records: a review. *Atmos. Oceanic Sci. Lett.*, **9**, 129-138, doi: 10.1080/16742834.2016.1141658.

Wang J., S. F. B. Tett, Z. Yan, 2017: Correcting urban bias in large-scale temperature records in China, 1980-2009. *Geophys. Res. Lett.*, **44**, 401-408, doi: 10.1002/2016GL071524.

Wang, S. W., D. Y. Gong, and J. H. Zhu, 2001: Twentieth-century climatic warming in China in the context of the Holocene. *Holocene*, **11**, 313-321, doi: 10.1191/095968301673172698.

Wang, S. W., J. H. Zhu, and J. N. Cai, 2004: Interdecadal variability of temperature and precipitation in China since 1880. *Adv. Atmos. Sci.*, **21**, 307-313, doi: 10.1007/BF02915560.

Wang S., X. Wen, Y. Luo, et al., 2007: Reconstruction of temperature series of China for the last 1000 years. *Chin. Sci. Bull.*, **52**, 3272-3280, doi: 10.1007/s11434-007-0425-4.

Wang W. C., Z. M. Zeng, and T. R. Karl, 1990: Urban heat islands in China. *Geophys. Res. Lett.*, **17**, 2377-2380, doi: 10.1029/GL017i013p02377.

Wang Y. Y., G. Li, Y. Zhang, 2011: Regional representativeness analysis of national reference climatological stations based on MODIS/LST product. *J. Appl. Meteorol. Sci.*, **22**, 214-220 (in Chinese). url: http://html.rhhz.net/yyqxxb/html/20110210.htm

Wiles G. C., O. Solomina, R. D'Arrigo, et al., 2015: Reconstructed summer temperatures over the last 400 years based on larch ring widths: Sakhalin Island, Russian Far East. *Clim. Dyn.*, **45**, 397, doi: 10.1007/s00382-014-2209-2.

Williams C. N., M. J. Menne, and P. W. Thorne, 2012: Benchmarking the performance of pairwise homogenization of surface temperatures in the United States. *J. Geophys. Res.*, **117**, D05116, doi: 10.1029/2011JD016761.

Williamson F., G. Y. Ren, and R. Allan, 2017: The Atmospheric Circulation Reconstructions over the Earth (ACRE) initiative ACRE China workshop. *Earth Space Sci.*, **4**, 40-43, doi: 10.1002/2016EA000215.

Wu K., and X. Q. Yang, 2013: Urbanization and heterogeneous surface warming in eastern China. *Chin. Sci. Bull.*, **58**, 1363-1373, doi:10.1007/s11434-012-5627-8.

Xu W., Q. Li, X. L. Wang, et al., 2013: Homogenization of Chinese daily surface air temperatures and analysis of trends in the extreme temperature indices. *J. Geophys. Res.*, **118**, 9708-9720, doi: 10.1002/jgrd.50791.

Xu W. H., Q. X. Li, P. Jones, et al., 2017: A new integrated and homogenized global monthly land surface air temperature dataset for the period since 1900. *Clim. Dyn.*, in press, doi: 10.1007/s00382-017-3755-1.

Yan H., W. Soon, and Y. H. Wang, 2015: A composite sea surface temperature record of the northern South China Sea for the past 2500 years: A unique look into seasonality and seasonal climate changes during warm and cold periods. *Earth Sci. Rev.*, **141**, 122-135, doi: 10.1016/j.earscirev.2014.12.003.

Yan L. B., and X. D. Liu, 2014: Has climatic warming over the Tibetan Plateau paused or continued in recent years? *J. Earth Ocean Atmos. Sci.*, **1**, 13-28, url: http://jeoas.uscip.us/Publishedissues.aspx [Last accessed: 7/11/2017]

Yan Z. W., Z. Li, Q. X. Li, and P. Jones, 2010: Effects of site change and urbanisation in the Beijing temperature series 1977-2006. *Int. J. Climatol.*, **30**, 1226-1234, doi: 10.1002/joc.1971.

Yan Z.-W., J. Wang, J.-J. Xia, et al., 2016: Review of recent studies of the climatic effects of urbanization in China. *Adv. Clim. Change Res.*, **7**, 154-168, doi: 10.1016/j.accre.2016.09.003.

Yang B., A. Braeuning, K. R. Johnson, and S. Yafeng, 2002: General characteristics of temperature variation in China during the last two millennia. *Geophys. Res. Lett.*, **29**, 1324, doi: 10.1029/2001GL014485.

Yang B., L. Tang, A. Bräuning, et al., 2008: Summer temperature reconstruction on the central Tibetan Plateau during 1860-2002 derived from annually resolved ice core pollen. *J. Geophys. Res.*, **113**, D24102, doi: 10.1029/2008JD010142.

Yang X. C., Y. L. Hou, and B. D. Chen, 2011: Observed surface warming induced by urbanization in east China. *J. Geophys. Res.*, **116**, D14113, doi: 10.1029/2010JD015452.

Yang Y. J., B. W. Wu, C. E. Shi, et al., 2013: Impacts of urbanization and station-relocation on surface air temperature series in Anhui province, China. *Pure Appl. Geophys.*, **170**, 1969-1983, doi: 10.1007/s0024-012-0619-9.

Yi L., H. Yu, J. Ge, et al., 2012: Reconstructions of annual summer precipitation and temperature in north-central China since 1470 AD based on drought/flood index and tree-ring records. *Clim. Change*, **110**, 469-498, doi: 10.1007/s10584-011-0052-6.

Zhang A. Y., and G. Y. Ren, 2005: Urban heat island effect on change of regional mean temperature over Shandong Province, China. *Clim. Environ. Res.* (in Chinese), **10**, 754-762.

Zhang A. Y., G. Y. Ren, J. X. Zhou, et al., 2010: On the urbanization effect on surface air temperature trends over China. *Acta. Meteor. Sinica.* (in Chinese), **68**, 957-966.

Zhang J. Y., W. J. Dong, L. Y. Wu, et al., 2005: Impact of land use changes on surface warming in China. *Adv. Atmos. Sci.*, **22**, 343-348, doi: 10.1007/BF02918748.

Zhang L., G. Y. Ren, Y. Y. Ren, et al., 2014: Effect of data homogenization on estimate of temperature trend: a case of Huairou station in Beijing municipality. *Theor. Appl. Climatol.*, **115**, 365-373, doi: 10.1007/s00704-013-0894-0.

Zhang X. X., Y. H. Hu, G. S. Jia, et al., 2017: Land surface temperature shaped by urban fractions in megacity region. *Theor. Appl. Climatol.*, **127**, 965-975, doi :10.1007/s00704-015-1683-8.

Zhang Y., and G. Y. Ren, 2014: Correcting urban bias for surface air temperature series of Beijing station over time period 1915-2012. *Chinese J. Geophy.* (in Chinese), **57**, 2197-2207

Zhao P., P. Jones, L. J. Cao, et al., 2014: Trend of surface air temperature in eastern China and associated large-scale climate variability over the last 100 years. *J. Clim.*, **27**, 4693-4703, doi: 10.1175/JCLI-D-13-00397.1.

Zheng J. Y., Y. Liu, and Z. X. Hao, 2015: Annual temperature reconstruction by signal decomposition and synthesis from multi-proxies in Xinjiang, China, from 1850 to 2001. *PLoS ONE*, **10**, e0144210, doi: 10.1371/journal.pone.0144210.

Zheng J. Y., Y. Liu, Z. X. Hao, et al., 2017: Winter temperatures of southern China reconstructed from phonological cold/warm events recorded in historical documents over the past 500 years. *Quat. Int.*, in press, doi: 10.1016/j.quaint.2017.08.033.

Zhou L. M., R. E. Dickinson, Y. H. Tian, et al., 2004: Evidence for a significant urbanization effect on climate in China. *Proc. Natl. Acad. Sci.*, **101**, 9540-9544, doi: 10.1073/pnas.0400357101.

Zhou T. J., and R. C. Yu, 2006: Twentieth-century surface air temperature over China and the globe simulated by coupled climate models. *J. Clim.*, **19**, 5843-5858, doi: 10.1175/JCLI3952.1.

Zhou Y. Q., and G. Y. Ren, 2005: Identifying and correcting urban bias for regional surface air temperature series of north China over period of 1961-2004. *Clim. Environ. Res.* (in Chinese), **10**, 743-753.

Zhu L., S. Li, and X. Wang, 2015: Tree-ring reconstruction of February-March mean minimum temperature back to 1790A.D. in Yichun, China. *Quat. Sci.* (in Chinese), **35**, 1175-1184.

Zhu L., Z. Li, Y. Zhang, et al., 2016: A 211-year growing season temperature reconstruction using tree-ring width in Zhangguangcai Mountains, northeast China: linkages to the Pacific and Atlantic Oceans. *Int. J. Climatol.*, **37**, 3145-3153, doi: 10.1002/joc.4906.