

Reply to Li & Yang's comments on "Comparing the current and early 20th century warm periods in China"

Willie Wei-Hock Soon¹, Ronan Connolly*², Michael Connolly², Peter O'Neill³, Jingyun Zheng⁴,
Quansheng Ge⁴, Zhixin Hao⁴, and Hong Yan⁵

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*

Pre-print version

Published in Earth-Science Reviews.

Available online 5 September 2019.

<https://doi.org/10.1016/j.earscirev.2019.102950>

1. Introduction

From a cursory reading of Li & Yang's comments [Li & Yang, 2019, henceforth LY2019] on our recent review article, Soon et al. (2018) [henceforth S2018], some readers might think that LY2019 is somehow disputing our analysis and conclusions. Specifically, they claim to offer "*some comments on the arbitrary or deductive conclusions of Soon et al. (2018) as [sic.] the following five aspects...*" They then make some comments on the following five issues:

1. "On the representativeness of climate analysis"
2. "On the observational data from meteorological stations"
3. "On the comparisons of proxy data, model reanalysis data and instrumental observation data"
4. "On the mixture of homogenization process and urbanization"
5. "On the contribution of urbanization to the regional SAT series and its change"

If a reader had not read S2018, and only had read LY2018, they might mistakenly assume that LY2018 completely disagree with us. However, a careful comparison of LY2019 and S2018 reveals that Li and Yang apparently agree with **most** of our conclusions. That said, there are several points where LY2019 appear to disagree with us. Mostly this seems to be due to minor mistakes or misunderstandings by LY2019. However, in two cases, the reasons for the apparent disagreement are more complex, i.e.,

- (a) What is the relationship between the ensemble means of Global Climate Model (GCM) hindcasts of Surface Air Temperature (SAT) and observed SAT trends?
- (b) How does statistical homogenization lead to the blending problem described by us in S2018?

With that in mind, the rest of this reply will be divided into four parts:

- Section 2. Points where LY2019 agree with us
- Section 3. Minor mistakes and misunderstandings made by LY2019

- Section 4. Comparing climate model temperature hindcasts to observations
- Section 5. Comments on the blending problem of homogenization

2. Points where LY2019 agree with us

One of the goals of S2018 was to raise awareness of the many challenging, inter-related factors involved in trying to evaluate the available data on Chinese SAT trends since the late-19th century. LY2019 agree with us, and they join us in our call by reminding the scientific community that “[i]t is imperative that [researchers understand] the data sources, uncertainty, biases and other limitations of any data that they used.” In fact, they want to emphasise that as well as the many factors we discussed in detail, there are others which were beyond the scope of our review.

For instance, as we explained in the conclusions of S2018,

“In this review, we mostly focused on annual mean temperatures averaged over all of China. However, temperature trends often vary from season to season [...] Also, different regions within China often show different climatic trends [...] Therefore, it is also important to consider seasonality and regionality.”

LY2019 agree with our recommendation to also consider the effects of regionality. In particular, in their Section 1.1., they suggest that the trends of the western and eastern regions should be compared and contrasted.

LY2019 agree with us that there are multiple different instrumentally-derived estimates of Chinese temperature trends and that depending on “...the collection and processing methods of meteorological data, a certain degree of differences in the analysis results are inevitable”. In their Section 1.2., they briefly summarise some of these differences, but if the reader is interested in a more comprehensive review of these differences, we recommend re-reading Section 2 of S2018.

In their Section 1.3., they also agree with us that there are, “...large uncertainties in proxy series for exploring a warm early-20th century period and a warm recent period” and that, “...many more in-depth researches are needed...” (LY2019) to explain the differences between the various proxy estimates. For a more detailed discussion of this point, we refer the interested reader to Section 3.5 of S2018.

3. Minor mistakes and misunderstandings made by LY2019

LY2019 claim that we, “[...] concluded that both limits of long-term observations in rural areas and urbanization bias **mainly** led to the results that the recent warm period seemed much warmer than the earlier warm period [...]” [emphasis added in bold]. We did **not** make this specific claim, since each of the authors of S2018 has different opinions on the relative weight of each of the factors involved in this challenging topic. As we explained in S2018,

“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[...], some of us have argued that the recent

warm period is much warmer [...]. Therefore, we believe it is important to establish and assess the reasons for these differing views.”

In their Section 1.2, LY2019 claim that there are “*only about 200 stations totally in China*” in both version 3 and version 4 of the Global Historical Climatology Network (GHCN) dataset. We do not know where they got this estimate from since we explicitly noted that there are 417 Chinese stations in version 3 and 494 in version 4. For comparison, the CRUTEM dataset (which Li had specifically recommended to us when we shared an early draft of S2018 with him) only contained 160 stations in version 3, but now contains 703.

Speaking of the CRUTEM datasets, LY2019 also argue in their Section 1.2 that simply by processing the CRUTEM4 data in a different manner, you can obtain a very different estimate for China. They do not describe exactly how they processed the data (the process we used is described in Section 2 of S2018), but if they are correct then this adds further support to our recommendations.

In their Section 1.5, LY2019 propose a rather simplistic analysis to estimate “*the contribution of urbanization to the regional SAT series*”. They argue the magnitude of urbanization bias can be estimated from Table 1 of S2018 as follows:

1. Subtract the peak annual temperature of the early 20th century warm period (1946) from the peak annual temperature in the current warm period (1998 or 2007).
2. Divide this difference by either 2 or 3 depending on whether the recent peak occurred in 1998 or 2007 respectively.
3. According to LY2019, this is the “Estimated urbanization effects” in °C/10a.

LY2019 construct their own Table 1 calculating estimates from the values of 14 of the 16 series in our Table 1, dropping the CRUTEM3 series and replacing the CRUTEM4 series with their own version. For some reason, they seem to have also included the wrong values for two of the remaining 14 series, i.e., Tang & Ren (2005) and Wang et al. (2004).

We disagree with this analysis. Why do LY2019 assume that the **only** difference between the two peak years is urbanization bias? Also, urbanization bias is usually a persistent long-term multi-decadal phenomenon, so using a direct comparison of two individual years is simply inappropriate. LY2019 refer to their proposed analysis as a “*mistaken contribution of urbanization*”. We agree such an analysis would be mistaken.

4. Comparing climate model temperature hindcasts to observations

S2018 presented 14 different time series of Chinese SAT derived from meteorological observations. For comparison, we also presented two equivalent series which were based on GCM hindcasts (provided to us by Li). One series was the multi-model ensemble average of all 42 CMIP5 hindcasts for China. The CMIP5 hindcasts were the ones used for the IPCC 5th Assessment Report (AR5).

In their Section 1.3, LY2019 claim that directly comparing the CMIP5 ensemble averages to the 14 series “*...is unreasonable and insignificant statistically*”. However, they immediately contradict themselves and make a similar comparison only using just one of the 14 series. This time, where the comparison is more favourable to them, they insist that it “*enhance[s] the confidence level for SAT analysis from both CMIP5 ensemble and the observations*”.

The contradiction seems to arise from the inherent conflict between two different schools-of-thought within the scientific community. The existence of these two different camps is often unappreciated, and therefore in this section it may be helpful to elaborate on the rationales of the two camps (see also the discussion in Connolly et al., 2019). At the end of the section, we will show that whichever school-of-thought you favour, LY2019 were still wrong in their Section 1.3 to dismiss the significance of the comparison between observed trends and the CMIP5 multi-model ensemble average hindcast.

The two different schools-of-thought occur because with current GCM hindcasts, if the simulation run is adequately equilibrated and not majorly affected by drift, then the global temperatures for a given year are mostly determined by three factors:

1. External Radiative Forcing (RF) from “Anthropogenic” factors. Many of these factors are considered, but greenhouse gas and aerosol concentrations are the main two.
2. External Radiative Forcing from “Natural” factors. Currently, only two of these are considered, i.e., changes in Total Solar Irradiance (“solar”) and stratospheric aerosols from volcanic eruptions (“volcanic”).
3. Internal variability. This is the year-to-year random fluctuations in a given model run.

The GCMs are effectively only able to simulate multi-decadal trends using the first two factors. However, because the random year-to-year “internal variability” fluctuations vary between models and individual runs, if one plots all of the individual model runs on top of each other, e.g., as a “spaghetti plot”, the thicker “envelope” is more likely to encompass a time series of observations than a comparison with the ensemble mean.

One school-of-thought argues that this “internal variability” is essentially “noise”, and that by averaging together the results you can improve the signal-to-noise ratio, e.g., Douglass et al. (2007). However, the other school-of-thought disagrees and argues that this is a feature which can somehow approximate the “internal variability” of nature, e.g., Santer et al., (2008). Both camps agree that, because the random fluctuations are different for each model run, they tend to cancel each other out in ensemble averages.

We suggest that both schools-of-thought have some validity and should be considered. Despite LY2019’s claim that the ensemble average “...is unreasonable and insignificant statistically”, we argue that using the ensemble averages is better for describing the influence of the external radiative forcing factors, and that these are more relevant for studying multi-decadal trends.

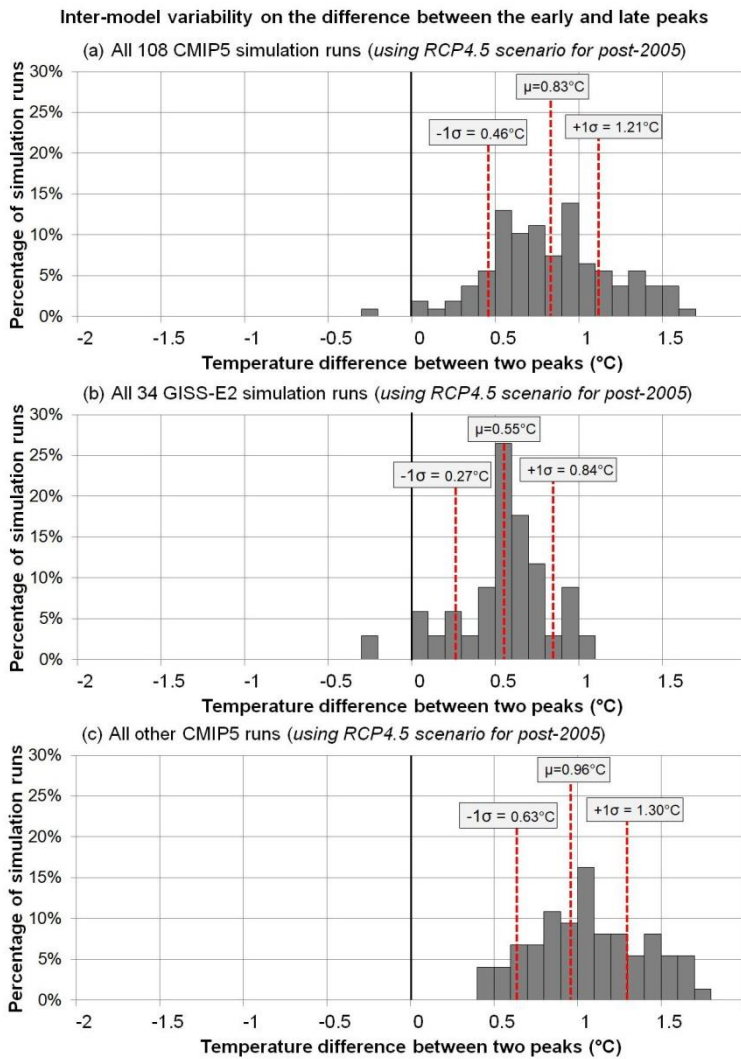


Figure 1. Histograms showing the difference between the early warm peak (maximum in the 1901-1950 period) and the recent warm peak (maximum in the 1951-2017 period) for all CMIP5 simulation runs (using RCP4.5 scenario for post-2005 projections).

Nonetheless, let us consider the inter-model variability. For S2018, we used the averages for each of the 42 models that submitted hindcasts to the CMIP5 project. These hindcasts covered the period 1861-2005. However, some of the modelling groups submitted multiple runs for each model (e.g., CSIRO-Mk-3-6-0 submitted 10 runs), and in those cases, our analysis (which used the same data as Li et al., 2017 – provided to us by Li) was based on the average of the multiple runs. Also, the modelling groups also projected these hindcasts into the future using a range of scenarios (RCP2.6, RCP4.5, RCP6.0 and RCP8.5). Therefore, we downloaded from KNMI’s ClimExp website (<https://climexp.knmi.nl/>) all 108 of the individual model runs and used the RCP4.5 projections to extend our analysis up to 2017. [We chose RCP4.5 as this was the most popular scenario submitted, but at any rate most of the differences between the four scenarios occur after 2017].

To sort the 108 model runs, we calculate the differences for each model run between the maximum temperature over the 1901-1950 period and the maximum temperature over the 1951-present (i.e., 2017) period. This crude, yet simple, metric allows comparison with part of the discussion in S2018. Figure 1(a) shows that the mean value of this metric was 0.83°C, i.e., the recent warm peak was

0.83°C warmer than the early-20th century peak. However, the exact value varied from model to model.

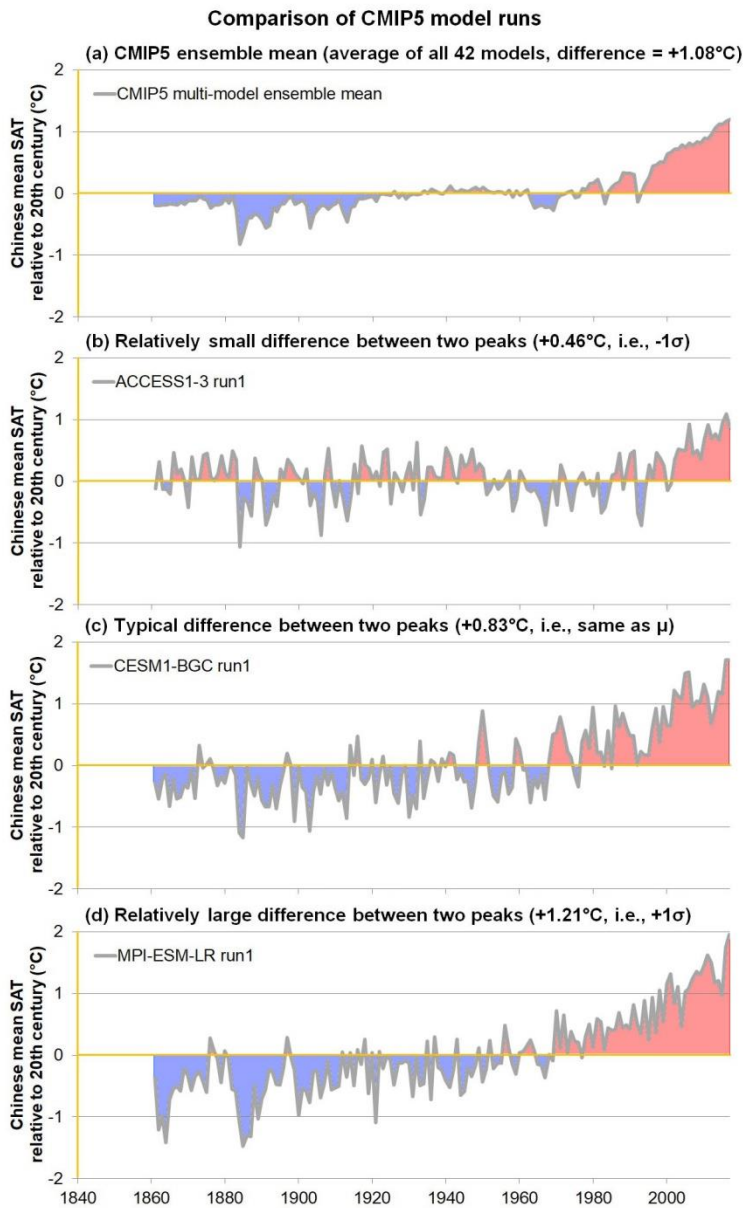


Figure 2. Comparison of (a) the CMIP5 multi-model ensemble mean to (b)-(d) three individual CMIP5 simulation runs.

As an aside, we briefly note that much of the spread in the histogram can be traced to one specific modelling group, NASA GISS, who submitted nearly 1/3 of the model runs. All of their model runs used a version of their GISS-E2 model. The differences between the 34 GISS-E2 model runs (Figure 1b) and the other 74 model runs (Figure 1c) are quite pronounced.

At any rate, in Figure 2, we compare the ensemble average (Figure 2a) to three of the 108 individual model runs (Figure 2b-d). The ensemble mean indeed removes most of the inter-annual variability of individual runs, making the average a lot “flatter” (this also makes the difference between the two peak years larger, i.e., +1.08°C). However, for runs where the difference between the two peaks is large, there is almost no “early 20th century warm period”, i.e., the same finding as for the ensemble mean. Zhou & Yu (2006) had noticed this for the earlier CMIP3 hindcasts for China. Meanwhile, for

runs where the difference between the peaks is small (e.g., Figure 2b), the recent warming is considerably muted, and not especially unusual.

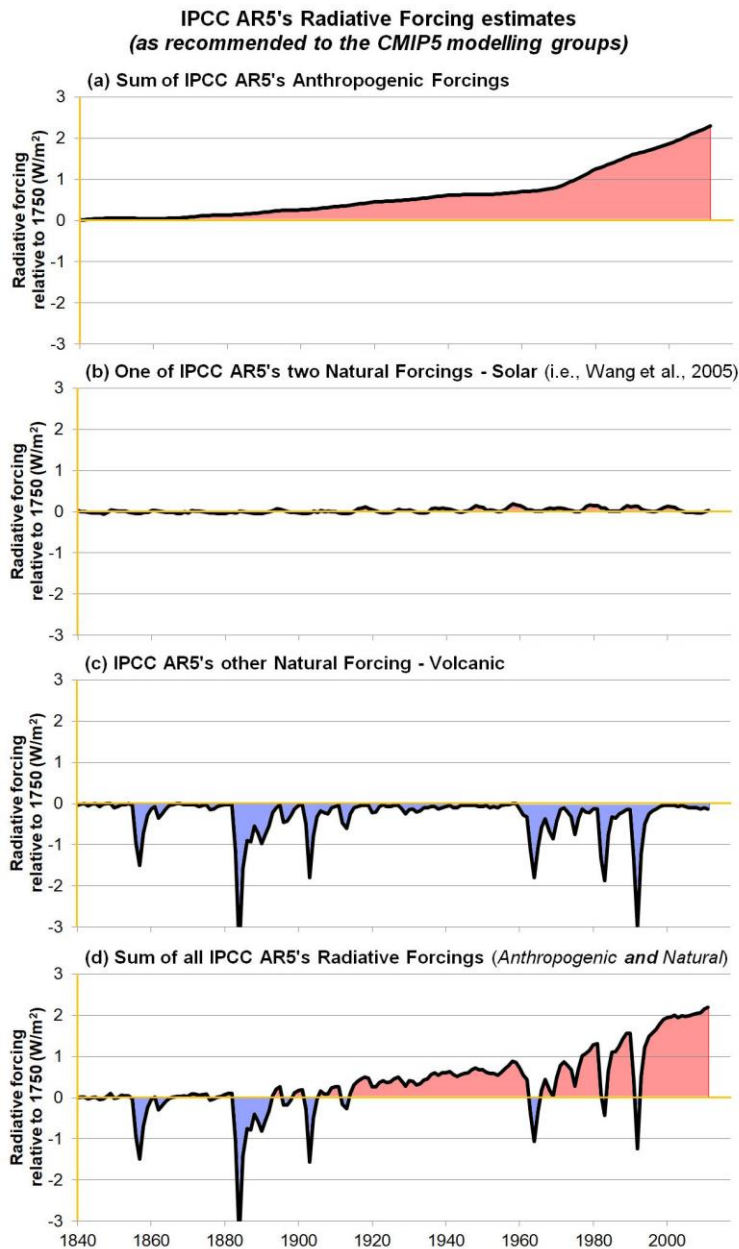


Figure 3. The various external "radiative forcings" estimates considered by the CMIP5 Global Climate Model hindcasts. Taken from IPCC AR5 Working Group 1 (Myhre et al., 2013).

Let us now consider the radiative forcing components used by the CMIP5 models – see Figure 3. Of the two “natural forcings” considered by the CMIP5 models, only the solar forcing could potentially introduce a multi-decadal warming trend, since the volcanic forcing (Figure 3c) only acts to introduce short 2-3 year cooling events. However, in Soon et al. (2015), some of us showed that the Wang et al. (2005) solar forcing dataset (or similar equivalents) used by the CMIP5 are very low variability estimates (Figure 3b).

This low variability is reduced even further in the IPCC AR5's dataset since they use a combined “albedo factor” of 0.55, i.e., they scale the original dataset by 0.70 (assuming an albedo of 30%) and

then by an additional 0.78 to account for “wavelength-albedo dependence” (Myhre et al., 2013; Section 8.4.1, p688).

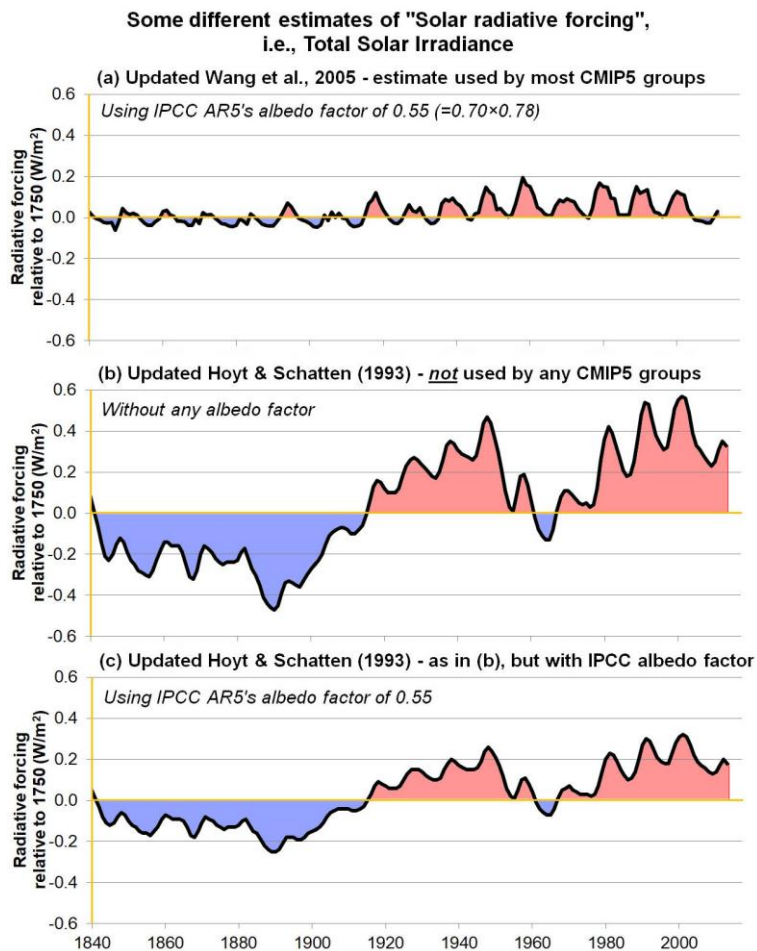


Figure 4. Comparison of the main "Solar radiative forcing" estimate used by the CMIP5 hindcasts (i.e., Wang et al., 2005) to an alternative estimate of solar radiative forcing (i.e., Hoyt & Schatten, 1993) that is at least as plausible.

Soon et al. (2015) argued that the CMIP5 modelling groups should have considered a range of the various available plausible solar variability datasets, rather than only considering the low variability ones. Figure 4 compares the Wang et al. (2005) dataset to another solar forcing dataset (Hoyt & Schatten, 1993) that is at least as plausible (Scafetta & Willson, 2014; Soon et al., 2015).

The choices made by GCM modelling groups as to which radiative forcing datasets directly influence the model output, as can be seen from Figure 5, in which different estimates of Chinese SAT are fitted to various combinations of “anthropogenic” and “natural” radiative forcing datasets, using a linear least squares best fit rescaling.

Figure 5a shows that the CMIP5 multi-model ensemble mean for China almost exactly overlaps with the combined “anthropogenic and natural forcings” dataset of Figure 3(d). In other words, the average Chinese temperature trends hindcasted by the CMIP5 models is determined almost entirely by the choice of forcing datasets used by the modelling groups.

This has profound implications for our understanding of the relative role of natural and anthropogenic factors in Chinese temperature trends since the 19th century, as well as in our assessment of the relative warmth of the various warm periods. For instance, while the Li et al. (2017) series (with a relatively modest 1940s warm period) is fairly well described using the IPCC's "anthropogenic forcings" (Figure 5b), the S2018 relatively rural series is fairly well described using the Hoyt & Schatten (1993) solar forcing dataset (Figure 5c). In other words, depending on which SAT series and which forcing datasets are used, you could come to completely different conclusions on whether Chinese temperatures since the 19th century were mostly determined by "anthropogenic factors" (Figure 5b) or "natural factors" (Figure 5c).

This builds on LY2019's point that "[u]nderstanding the long-term variation in surface air temperature related to climate warming is one of the important issues for understanding the regional and global climate change and its detection, attribution and impact". That is, we agree with LY2019 that it is important to understand "the data sources, uncertainty, biases and other limitations of [the various Chinese surface air temperature datasets]" – indeed, that was the primary motivation of S2018. But, further, we should also be aware of the considerable debate over the various radiative forcings datasets that have been used for the "attribution" of surface air temperature trends. We recommend Soon et al. (2015) for a comprehensive review of the ongoing debate over the solar radiative forcing datasets.

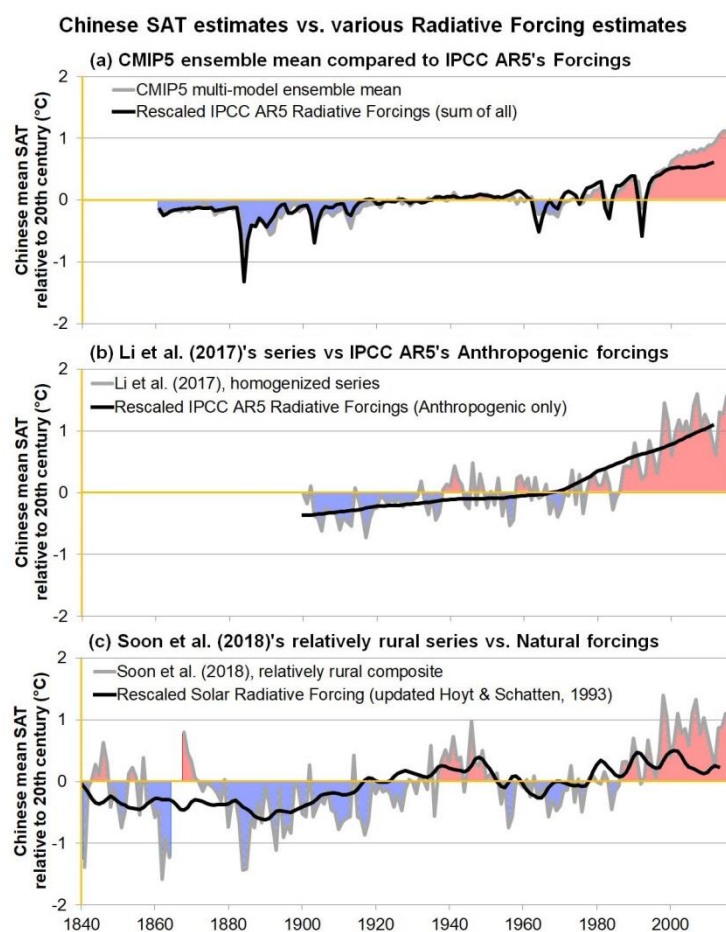


Figure 5. Comparison of three of the estimates of Chinese SAT considered in Soon et al. (2018) with various "radiative forcing" estimates.

5. Comments on the blending problem of homogenization

In their Section 1.4, LY2019 admit to being confused about the blending problem of the current homogenization processes. LY2019 briefly summarises the rationale of homogenization, but for those who want to gain a more comprehensive understanding of the process, we refer the interested reader to Section 3.2 of S2018 and the references cited therein. Unfortunately, LY2019 do not appear to have considered the blending problem associated with the current homogenization processes, which we described in Sections 3.2.3-3.2.5. We appreciate that the blending problem has been largely overlooked in the literature until S2018, and therefore it may be useful to provide some additional insights here. To understand the blending problem, it is important to distinguish between two separate stages of the homogenization process:

1. Identifying when a non-climatic step change bias occurred
2. Establishing (and adjusting for) the sign and magnitude of the bias

In the first stage, the main challenge is in minimising the number of false positives and false negatives, while maximising the number of true positives and true negatives. As LY2019 note, many of the current homogenization procedures perform quite well at this stage when tested with simulated and/or synthetic biases, e.g., Venema et al. (2012). We agree, but point out that those tests did not consider the urban blending problem (personal communication with Venema, 2017).

Often the homogenization process is carried out in the absence of any information about documented changes in the station location, instrumentation, etc., which the station observers record in accompanying station histories (sometimes called “station metadata”). This is the case with the widely-used GHCN dataset which we used for much of our analysis. However, while the CMA have access to such “station metadata” and used this in the Li et al. (2017) homogenization process, they have not yet provided public access to this important information, or to the homogenization adjustments they applied.

In October 2017, we invited Li to share this station metadata with us and/or collaborate with us to try and assess the accuracy of the GHCN homogenization adjustments with regards to this first stage. We repeat this invitation to Li and colleagues if they are interested in helping us to advance the scientific understanding on this important issue.

Nonetheless, the blending problem arises from the second stage and **not** from the first stage. All of LY2019’s comments on the homogenization process (in their Section 1.4) relate to the first stage. In S2018 we showed that the blending problem is a real statistical artefact. This is also confirmed by the statistical experiments of deGaetano (2006) and Pielke et al. (2007).

That said, quantifying the extent of the problem for a given region, such as China, is more challenging. In S2018 we showed that the problem was indeed substantial for a sample of 10 stations in the Beijing area which had been homogenized by He & Jia (2012) apparently using the same station history information endorsed by LY2019. However, the Beijing area is a highly urbanized region of China, and it is still unknown how large the problem is for the rest of China.

In Figure 6, we plot the net gridded mean of the homogenization adjustments applied by NOAA to the two most urbanized subsets (Figure 6a) and the other three subsets (Figure 6b). *Qualitatively*, the long-term trends are *consistent with* urban blending being a significant problem. That is,

homogenization partially reduces the warming trends of the most urban stations, but it introduces extra warming into the more rural stations.

However, when we consider the period before the 1950s (and to a lesser extent, after 1990), we can see that there is a considerable drop-off in station numbers for both the most urban (Figure 6c) and the least urban (Figure 6d) subsets. Moreover, the station numbers fluctuate from decade-to-decade (and even year-to-year). That is, the stations used for calculating the average Chinese temperature - as well as the stations used for homogenization - vary substantially over the years. As a result, we do not attempt to quantify here the extent of the blending problem for Chinese temperature estimates based on homogenized data. Rather, we merely note that the problem is real and insidious, and recommend further research to investigate its extent.

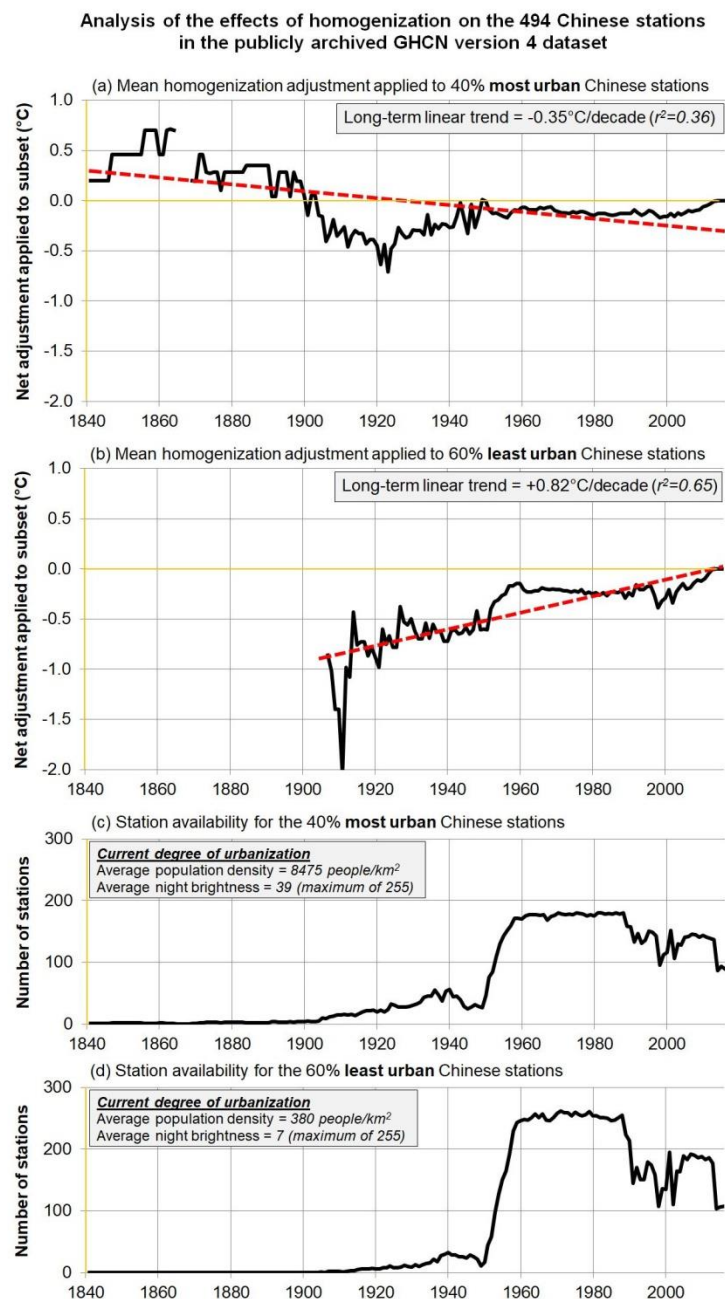


Figure 6. Gridded mean homogenization adjustments applied by NOAA to the 494 Chinese stations in version 4 of the Global Historical Climatology Network (GHCN) dataset.

Acknowledgement:

We are grateful to Prof. Geert Jan van Oldenborgh for creating and maintaining the KNMI Climate Explorer website, <https://climexp.knmi.nl/>, which we used to obtain the CMIP5 hindcasts/projections for China.

References

Connolly, R., M. Connolly, W. Soon, et al., 2019: Northern hemisphere snow-cover trends (1967-2018): A comparison between climate models and observations. *Geosci.*, **9**, 135, doi: [10.3390/geosciences9030135](https://doi.org/10.3390/geosciences9030135).

- 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](https://doi.org/10.1175/JCLI3662.1).
- Douglass, D.H., J. R. Christy, B. D. Pearson, and S. F. Singer, 2007: A comparison of tropical temperature trends with model predictions. *Int. J. Climatol.*, **28**, 1693-1701, doi: [10.1002/joc.1651](https://doi.org/10.1002/joc.1651).
- 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](https://doi.org/10.1002/joc.3370150403).
- 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](https://doi.org/10.1080/16742834.2012.11447029).
- Hoyt DV, Schatten KH (1993). A discussion of plausible solar irradiance variations, 1700-1992. *J Geophys Res* 98:18895-18906. doi:10.1029/93JA01944
- 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](https://doi.org/10.1175/BAMS-D-16-0092.1).
- Li.Q., Yang, Y., 2019: Comments on “Comparing the current and early 20th century warm periods in China” by Soon W., R. Connolly, M. Connolly et al. *Earth Sci. Rev.* [in press]
- 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](https://doi.org/10.1175/2008JCLI2263.1).
- Myhre G., D. Shindell, F.-M. Bréon, et al., 2013: Anthropogenic and natural radiative forcing. 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.
- 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](https://doi.org/10.1175/BAMS-88-6-913).
- Santer, B. D., P. W. Thorne, L. Haimberger, et al., 2008: Consistency of modeled and observed temperature trends in the tropical troposphere. *Int. J. Climatol.*, **28**, 1703-1722, doi: [10.1002/joc.1756](https://doi.org/10.1002/joc.1756).
- Scafetta N, Willson RC (2014) ACRIM total solar irradiance satellite composite validation versus TSI proxy models. *Astrophys Space Sci* 350:421-442. doi:10.1007/s10509-013-1775-9
- Soon W., R. Connolly, and M. Connolly, 2015: Re-evaluating the role of solar variability on Northern Hemisphere temperature trends since the 19th century. *Earth. Sci. Rev.*, **150**, 409-452, doi: [10.1016/j.earscirev.2015.08.010](https://doi.org/10.1016/j.earscirev.2015.08.010).
- Soon W., Connolly R., Connolly M., et al., 2018: Comparing the current and early 20th century warm periods in China. *Earth Sci. Rev.*, **185**, 80-101. doi: [10.1016/j.earscirev.2018.05.013](https://doi.org/10.1016/j.earscirev.2018.05.013)
- 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)
- Venema, V., O. Mestre, and the COST HOME Team, 2012: Benchmarking monthly homogenization algorithms. *Clim. Past*, **8**, 89-115, doi:[10.5194/cp-8-89-2012](https://doi.org/10.5194/cp-8-89-2012).
- 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](https://doi.org/10.1007/BF02915560).
- Wang Y.-M., Lean J. L., Sheeley N. R. Jr., 2005. Modeling the Sun's magnetic field and irradiance since 1713. *Astrophys. J.*, **625**, 522. doi:[10.1085/429689](https://doi.org/10.1085/429689)
- 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](https://doi.org/10.1002/jgrd.50791).
- 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](https://doi.org/10.1175/JCLI3952.1).