

Reply on RC2

Peng Si

Community comment on "Construction of homogenized daily surface air temperature for the city of Tianjin during 1887–2019" by Peng Si et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2020-343-CC3>, 2021

Thank you very much for the valuable time devoted to this paper by the referees and responsible editor of ESSD, as well as for our opportunity to reply to these comments. A point-by-point response to Anonymous Referee #2' comments is as follows.

Specific comments:

- Please provide the information about SAT stations (location, surroundings, rural, towns or city, as well as altitude).

Reply Thank you for your suggestions.

We have provided the information about SAT stations in the new Table 3, and added some description sentence on L279 in revised manuscript.

- Provide a correlation analysis and other measure to convince the use these data.

Reply Thank you for your suggestions.

It is difficult to find much observational data surrounding Tianjin before the 1950s, so in the establishing the reference series, we employed the station series or the interpolated temperature series using neighboring grid boxes from three global land surface temperature observation series (CRUTS4.03, GHCNV3 and Berkeley Earth). For monthly reference series, we established two types using a weighted average method. One was based on the combination of the interpolated temperature series from Berkeley Earth and CRUTS4.03 and station series from GHCNV3 data for Tianjin site and the other was based on the interpolated temperature series (they are the selected 9 stations) from Berkeley Earth data only. It is worth mentioning that the reference series established by the 3 datasets are very consistent, which reflects the consistency of air temperature variation in this region among different datasets. All the interpolated station series used in this current study have much higher correlations with each other, with the correlation coefficients all higher than 0.95.

- Justify why 300km is reasonable.

Reply Thank you for your comments.

In the process of data homogenization, the standard practice is to select 3-5 surrounding stations as the reference stations (the correlation coefficient between these stations and candidate stations is greater than 0.8, and the distance between them is 300-500km) for discontinuous point detection and adjustment, establish a relatively homogenous reference series that can represent the local climate change characteristics for the candidate stations, and detect the homogeneity of the difference series between the candidate and the reference series. However, this ideal depends on the density of the network. In the current study, we found 9 stations (including 54527 itself) within 300 km as the basis of reference series construction, which shows that the reference series establishment in this area is relatively reasonable, and using as many as possible stations can offset some inhomogeneity of a station that might affect the reasonableness of the reference series as far as possible.

- Please explore why use monthly and daily reference series respectively for the extreme temperature series? The extreme temperature series is monthly or daily series?

Reply □ **Thank you for your comments.**

Because daily temperatures vary on relatively small spatial scales and are influenced by local processes that are complex and nonlinear, homogenization detection for daily data is difficult. The monthly reference series were used to detect breakpoints for the extreme temperature series in the process of PMT, and the daily reference series was used to adjust breakpoints in the process of QM-adjustment. The extreme temperature series are daily series. What we have done in this current study is the same as in Xu et al (2013).

Ref:

Xu, W. Q., Li, Q. X., Wang, X. L., Yang, S., Cao, L. J., and Feng, Y.: Homogenization of Chinese daily surface air temperatures and analysis of trends in the extreme temperature indices, *J. Geophys. Res. Atmos.*, 118, [http:// doi.org/10.1002/jgrd.50791](http://doi.org/10.1002/jgrd.50791), 2013.

Minor comments:

- L248 For Table 2, why there are four data sources shown here, the Table 1 is puzzle?

There are three data sources in Table 2, which are CRUTS4.03, Berkeley Earth and GHCNV3, respectively.

The CRUTS4.03 and GHCNV3 data meet the requirements of this current study and both are on monthly timescales, but the Berkeley Earth data are on monthly and daily timescales. Table 1 only introduced the metadata information of Tianjin meteorological observation station.

We made some revisions to show this clearer in revised manuscript.

- L284 "Berkeley-daily data" should be "Berkeley Earth-monthly data".

It should be "BE-daily data". We have changed on L286 in revised manuscript.

Please also note the supplement to this comment:

<https://essd.copernicus.org/preprints/essd-2020-343/essd-2020-343-CC3-supplement.zip>