

Review of the manuscript: “A daily highest air temperature estimation method and spatial-temporal changes analysis of high temperature in China from 1979 to 2018” by P. Wang et al.

General Comments

The authors present a novel method for obtaining daily maximum air temperatures (T_{\max}) over China. The proposed framework combines global reanalyses, station and satellite data, applying decision trees, regression methods and diurnal interpolation. The method is then used to construct an optimized T_{\max} estimation covering whole China for 40 years. As expected, the overall trend of T_{\max} is positive, but with regional and annual variations. Significant but small to moderate correlations to three Atmosphere and Ocean climate modal indices can be found regionally.

As the authors pointed out, their estimation method improves the accuracy of a traditional data set. Future global and regional studies on climate change will benefit from such an improvement of the China T_{\max} data set. Moreover, an adapted construction method may be suited for transfer to other world regions.

While the paper is written in a clearly structured way, I got a bit tired reading through all the detailed statistics in the results section. It would help the reader to get some more information on the geographical nature of the China regions in the introduction. One could do this with more visual information like climographs for selected stations of the six regions. There is also room for adding a table with basic climate information for the regions as well as the number of stations. In the discussion part, you may also elaborate on possible error sources in the trend analysis like changing data quality over time and influence of growing urban heat island effects. The Mann-Kendall test for abrupt change analysis is not well introduced and results are confusing. I'd propose to skip the respective sections (4.3 and lines 450-474 in 5.3.1). One could think of publishing this manuscript rather in a journal dedicated to climatic change, but the construction framework itself is a kind of model in geoscientific context which merits publication in GMD. Taking care of the above propositions, I generally recommend to publish the paper with minor corrections as stated out in the following.

Specific Comments

Lines 47-48: As pointed out by Yang et al. (2017), highest daily air temperatures have a significant influence on the end of plant growing season on the Tibetan Plateau. Evidence for an effect on the start time of the plant growth season on the Tibetan Plateau is given in Shen et al. (2016). However, these findings cannot be transferred easily to other climatic regions. Please remove association here or elaborate more detailed.

Lines 53-58: Station measurements have high accuracy for T_{\max} , both in its value and in the time of occurrence. Spatial representativeness is of course limited. For this paper it will be of interest to know how many station observations are used for your analysis. Please refer to Figure 1 for the spatial distribution.

Line 62: Replace “land surface temperature” by “retrieved land surface temperature”.

Line 86: Replace “extreme high temperature” by “extreme high temperature over China”.

Line 93: “almost no one” suggests that at least a few studies (of the above cited?) have constructed a high-resolution T_{\max} data set for China. If so, please refer to these papers here. Otherwise, you may state that to your best knowledge no such studies exist.

Lines 107-108: What do you want to say here? It seems obvious that wintertime temperatures get higher with more southerly regions. Does “opposite” mean that the northern part of region I is not affected by monsoon?

Lines 111-112: Better say something like: “The area exhibits little annual precipitation which decreases from east to west.”

Lines 129-137: Please elaborate shortly on how CMFD data has been derived.

Lines 140-141: Cite ERA5 with Hersbach et al. (2020).

Line 145: ERA5 data can be derived from the Copernicus Climate Data Store. Please change accordingly (also in the acknowledgements).

Lines 145-147: Please specify the spatial and temporal resolution of ERA5 as used in this study. You should also mention that multiple kinds of data assimilation have been performed to derive the reanalysis meteorological variables.

Lines 149-157: How many stations and observations are included in the original data set? How many of these are used in your study after performing checks and tests mentioned here? Certainly, not all stations have observational record for the whole time period. How do you deal with those stations?

Lines 175-178: You suppose here that your T_{\max} is constructed using station and reanalysis data only. In addition, for clear-sky condition the CMFD data is used. I argue that this data set is constructed with the help of satellite observations. Derivation of clear-sky conditions is also based on satellite retrievals?

Lines 179-180: How are the pixels mentioned here defined and which is the resolution? Is it based on satellite observations or on the era5 resolution of about 30 km? To better justify your approach, you could even give the total number of pixels and the percentage value of pixels without stations. Can you have several stations for one pixel?

Lines 217-226: I argue that the correction has been performed for all pixels, not only for pixels with poor quality (whatever this means). Please describe the correction process more clearly.

Lines 271-286: I propose to skip this section as stated above.

Lines 284-285: Rephrase sentence.

Lines 301-302: References can be omitted here.

Lines 316-317: How do you define your data as being very accurate? Please elaborate a more science-based description of Fig. 4.

Lines 372-373: How do you handle winter seasons 1978/79 and 2018/19?

Lines 387-388: Wintertime trends for region II are not significant. Omit.

Lines 440-441: Replace “... previous research results (Hong and Ying, 2018)” by “... previous research results by Hong and Ying (2018, not shown here)”

Lines 450-474: I propose to skip this section including Fig. 10.

Line 509: Replace “lead to” by “contribute to”. Similarly in line 518.

Lines 510-513: Please avoid to interpret the weak negative correlation with cause and effect. This could be easily attributed to other factors which have not been examined here.

Lines 602-603: You could refer to IPCC here.

Lines 619-620: Skip sentence referring to abrupt changes.

Lines 622-626: Mention that most of the correlations shown here are small.

Line 623: Replace “led to” by “coincides with”.

Lines 627-632: This section should be extended. What is the value of the new data set? How can it be used nationally and internationally? How can this study possibly interact with other international efforts for data construction?

Figure 4: Please add the total number of data points for each region.

References:

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, *Q. J. Roy. Meteor. Soc.*, 146, 1999–2049, <https://doi.org/10.1002/qj.3803>, 2020.

Shen, M., Piao, S., Chen, X., An, S., Fu, Y. H., Wang, S., Cong, N., Janssens, I. A.: Strong impacts of daily minimum temperature on the green-up date and summer greenness of the Tibetan Plateau. *Global Change Biology*, 22(9), 3057–3066, <https://doi.org/10.1111/gcb.13301>, 2020.