Dear referees,

Thank you for your valuable comments on our manuscript. First, we would like to express our sincere appreciation for your professional and insightful remarks on our paper. These comments are all valuable and have helped us to improve the quality of our paper. We have studied each comment and have made revisions that we hope will meet with approval. Please find our detailed responses below. For convenience, we also attach a version of the manuscript with changes incorporated. Thanks again.

With our best regards,
Shu Fang and co-authors

Response to referees

Authors presented a dataset of land surface air temperature over China using in situ station data that is interpolated to 0.1º resolution by downscaling methods using ERA5, CMFD, and CMA data. They found the data set has a reasonable accuracy quantified by RMSE, MAE, and R2, and by cross-validation method. The days of the 90th/10th percentile temperature are increasing/decreasing, which is consistent with overall warming climate. My major comments are (a) the cross-validation part (Table 1) could be extended while the evaluation against in situ data (Figures 7-9, 10-12) can be shortened, and (b) it might be helpful if the data set can be compared against other independent datasets. A major revision is recommended.

Response: Thank you for your guidance, all the comments have been carefully and individually addressed. We have made revisions and enclosed below are our point-to-point responses to these comments.

1. L30-31, "and we further improve data accuracy through building correction equations for different regions" => and the data accuracy is improved through building correction equations for different regions

Response: Thank you for your good suggestion, and we have made revisions.

2. L41, "reliable accuracy", which needs explanations. Can you justify that RMSE of 0.86-1.78ºC is a reliable accuracy?

Response: Thank you for your guidance, and we have made revisions. We are mainly based on current application requirements and accuracy comparison with other datasets.

3. L57-58, L58-59, The statements of "the extremely cold days and nights gradually shorten" and "intensity and duration of extreme weather events are also increasing" do not appear consistent. Are the increasing extreme events all hot events?

Response: Thank you for your guidance, and we have made revisions. The number of extremely cold days and cold nights gradually shortens and the frequency of extreme weather events is also increasing.

4. L66-68, “monitoring estimating/obtaining/ Ta” => Ta monitored/estimated/obtained
Response: Thank you, and we have made revisions.

5. L88, delete “energy”
Response: Thank you, and we have made revisions.

6. L135-136, “cumulative temperature is between 2500°C and 4000°C.”, a citation is needed. Why the cumulative temperature is used instead of average temperature? How the temperature is accumulated? If so, the unit should be ºC*year.
Response: Thank you for your guidance. There are two main methods for calculating the accumulated temperature. Here we mainly use the accumulation of the daily average temperature ⩾10 °C. We have added references here, and the unit is also in accordance with the commonly used unit. Here is mainly to introduce the different conditions of six subregions, there is no special reason for the choice.

7. L145, “800°C” should be a typo.
Response: Thank you, and we have made revisions.

8. L146-151, (V), what temperature is this region?
Response: Thank you, and we have made revisions.

9. Section 2, How are the boundaries of these regions determined?
Response: We divide China into six regions mainly based on climatic conditions, such as temperature and rainfall, and topographical conditions such as elevation.

10. L200 and L185, it is not clear whether one or both are used to reconstruct Ta. How are these data used to reconstruct Ta, as a training data or validation data?
Response: Here we mainly want to introduce that ERA5 is an upgraded version with higher accuracy. In the non-clear sky condition, we used the ERA5 dataset for downscaling where there is no in situ data. Part of the data is used as training data, and part is used as validation data.

11. L208-210, reference is needed such as Du et al. 2020
Response: Thank you, and we have made revisions.

12. L244 “In currently” delete “in”
Response: Thank you. We have made revisions.

13. Figure 3, in “2002-2018” branch, are there any in situ data used? If not, are the final output purely from MODIS observation?
Response: Thank you for your guidance. During 2002-2018, we mainly used MODIS data as the judgment condition, but when MODIS data is missing, in situ

data and other data are also used at the same time. The manuscript didn't introduce it in detail, we revised it.

14. L305, “was less than 0.3°,” is questionable because the data resolution is 0.1º.
Response: The spatial resolution of ERA5 data is 0.3°. In order to ensure the consistency of our data, we use 0.3° as the condition for judging the distance between observation stations. When there is a station at the pixel location or the station distance is less than 0.3°, we use the in situ data. When the distance between stations is greater than 0.3°, we use ERA5 data for downscaling.

15. L306-315, Is there any elevation consideration between stations when filling from adjacent stations?
Response: Yes, we have considered the influence of elevation. In order to improve the accuracy of the interpolation, we first correct the data of the observation site to a uniform sea level and then perform further calculations based on the elevation of the interpolation point to obtain the corresponding temperature.

16. L323-326, again, any elevation consideration?
Response: Yes, we have considered the influence of elevation and have made revisions in the manuscript.

17. Equations (1)-(2), a reference is needed. Why do the indices I and j start from 0?
Response: Thank you for your guidance. This formula is derived by ourselves and does not require references, and we have made revisions.

18. In Figure 4, the 3rd column, the downscaling should use the surrounding data, not from all from the corner (The figure is merely visible due to low quality/resolution)
Response: Thank you for your guidance. We have made revisions to the manuscript and redraw Figure 4.

19. L369-370, How are At and Bt “obtained”, by least square method in the next sentence?
Response: Thank you for your guidance. A_t and B_t are obtained by the least square method. We didn't introduce it in detail, and we have modified it in the manuscript.

20. L403-404, Should “use … as” be “average” or “to calculate daily average temperature”?
Response: Thank you for your guidance, and we have made revisions.

21. L427. “Sect. 0,”, please check.
Response: Thank you, and we have made revisions.

22. L451-453, Any reason for “we selected areas with uniform surface types and flat terrain under clear skies as the comparative study area and compared this product with

the existing datasets”?

Response: Thank you for your guidance. It is very difficult to validate large-scale data. First of all, it is necessary to ensure that the reference dataset is relatively accurate. We mainly consider the representativeness of the in situ data from the ground station, so we choose as much as possible the area with a single surface type and flat terrain for comparison. Currently, there is no better method for large-scale calibration.

23. L455, Since ERA5 and CMA data have been used in the downscaling processes, they are not independent anymore and it is questionable to be used for validation/evaluation.

Response: Only ERA5 and CMFD data during non-clear sky are used for downscaling and filling. When doing validation, we do cross-comparison after upscaling all data to the same resolution, including ERA5 and CMFD data that did not participate in the calculation. The most important thing is that we made a cross-comparison with the CMA data that has not been involved in the calculation, and used the site data that did not participate in the calculation for validation.

24. L466, Should TXx/TNn be better TXn/TMn? Otherwise, what are the meaning of “x” and “N”?

Response: Here we mainly use the definition of World Meteorological Organization (WMO) [1][2]. WMO mainly defines 27 core indices, and we quote them in their original form here. TXx/TNn respectively represent the maximum value of maximum temperatures and the minimum value of minimum temperatures. The x in TXx represents max, and the n in TNn represents min in the same way.

1.Karl, T.R., N. Nicholls, and A. Ghazi, 1999: CLIVAR/GCOS/WMO workshop on indices and indicators for climate extremes: Workshop summary. Climatic Change, 42, 3-7.

2.Peterson, T.C., and Coauthors: Report on the Activities of the Working Group on Climate Change Detection and Related Rapporteurs 1998-2001. WMO, Rep. WCDMP-47, WMO-TD 1071, Geneve, Switzerland, 143pp.

25. Figures 6-8, Since the “dataset” have been using “in situ data” at pixels where in situ data exist, I doubt these comparisons are meaningful. A better way is to compare independent data/station, or compare with other available datasets created independently.

Response: Maybe it was not clearly written in the manuscript, and we revised it. In order to ensure the independence of the data, we use the jackknife method to divide the in situ data into 80% and 20%. Of these, 80% is used to build the model, and 20% is used to validate accuracy. Figures 6-8 and Figures 9-11 are accuracy validation under the premise of 20% validation data.

26. Figures 9-11, some comments as for Figure 6-8.

Response: Thank you for your guidance, and we have made revisions.

27. Table 1, Glad to see the cross validation. I would suggest replot Figures 7-9 using independent data and put original figure to appendix or supplementary materials. The question is how the cross-validation is arranged, which may need explanations in text.

Response: Thank you for your guidance. We used independent data and in situ data for cross-validation and validation. We left part of the data and did not participate in the calculation, so the validation is no problem. The expression may not be clear enough, we have revised it in the manuscript.

28. L473, "with more than 90% (less than 10%) correlation with the number of days in each year" may need to rephrase or rewritten. It is hard to understand. How do these percentile is determined, based on a certain time period?

Response: Thank you, and we have made revisions. We select TX90p and TN10p as extreme climate indices defined by the WMO organization [1][2]. The 90% (10%) corresponding value in the time series is used as the threshold for judging warm days (cold nights).

1.Karl, T.R., N. Nicholls, and A. Ghazi, 1999: CLIVAR/GCOS/WMO workshop on indices and indicators for climate extremes: Workshop summary. Climatic Change, 42, 3-7.

2.Peterson, T.C., and Coauthors: Report on the Activities of the Working Group on Climate Change Detection and Related Rapporteurs 1998-2001. WMO, Rep. WCDMP-47, WMO-TD 1071, Geneve, Switzerland, 143pp.

29. Figure 14c', check x-axis, any ***?

Response: Thank you, and we have made revisions and redrawn Figure 14.