

Earth Syst. Sci. Data Discuss., referee comment RC2
<https://doi.org/10.5194/essd-2021-309-RC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on **essd-2021-309**

Anonymous Referee #2

Referee comment on "Dataset of daily near-surface air temperature in China from 1979 to 2018" by Shu Fang et al., Earth Syst. Sci. Data Discuss.,
<https://doi.org/10.5194/essd-2021-309-RC2>, 2021

Authors presented a dataset of land surface air temperature over China using in situ station data that is interpolated to 0.1° resolution by downscaling methods using ERA5, CMFD, and CMA data. They found the data set has a reasonable accuracy quantified by RMSE, MAE, and R^2 , 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.

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

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

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?

L66-68, "monitoring estimating/obtaining/ Ta" => Ta monitored/estimated/obtained

L88, delete "energy"

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 °C*year.

L145, "800°C" should be a typo.

L146-151, (V), what temperature is this region?

Section 2, How are the boundaries of these regions determined?

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?

L208-210, reference is needed such as Du et al. 2020

L244 "In currently" delete "in"

Figure 3, in "2002-2018" branch, are there any in situ data used? If not, are the final output purely from MODIS observation?

L305, "was less than 0.3°," is questionable because the data resolution is 0.1°.

L306-315, Is there any elevation consideration between stations when filling from adjacent stations?

L323-326, again, any elevation consideration?

Equations (1)-(2), a reference is needed. Why do the indices I and j start from 0?

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)

L369-370, How are At and Bt "obtained", by least square method in the next sentence?

L403-404, Should "use ... as" be "average" or "to calculate daily average temperature"?

L427. "Sect. 0,", please check.

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"?

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.

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

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.

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

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.

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?

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

