

Comment on **essd-2022-257**

Minyan Wang (Referee)

Referee comment on "HiTIC-Monthly: a monthly high spatial resolution (1 km) human thermal index collection over China during 2003–2020" by Hui Zhang et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2022-257-RC1>, 2022

Major comments.

The dataset is very important to study the climate change during this period over this region. The high spatial resolution and the monthly temporal resolution are basically reasonable, based on the previous datasets. The use of homogenous meteorological observation data is the right way to evaluate. This research is novel and belongs to the application of big data for large volume of satellite data used, and machine learning algorithm. The authors are from 8 universities/colleges. The results supply important reference for other research fields like the field of geoscience, biology, sociology, and medicine and engineering.

- Line 415, harm, it is suggested to change the word. All 11 indices are calculated based on SAT, so the accuracy of this dataset is up to the accuracy of the dataset (Zhang 2022b) and the algorithm. SAT has important influence to the other 11 indices. How to understand these?
- (Zhang 2022b) SAT is LST, but not Tair (1.5 meters above the surface). Why use LST but not Tair in Table 1 to compute the other 11 indices?
- Line 157-158 "Section 6 compares our products with two existing datasets, and the main findings of this paper are summarized in Section 7". Where is section 6 and section 7? They are data availability and conclusions.
- Why not keep Figures S1-S9, Tables S1-S5 as formal ones? Please consider again which to keep in the manuscript.
- How to get monthly data from MODIS daily mid-daytime 13:30 and mid-nighttime 01:30 LST? The monthly mean is different considering of the diurnal variation and satellite observations from ascending orbits and descending orbits.
- Line 190-191, how to compute and use the covariates?
- HiTIC? What is it? Is it HTI (human thermal index)? No "I" after "human" and before "thermal".

Minor comments.

- Add abbr. of LGBM when it is first used. Light Gradient Boosting Machine. Check others, please.
- Author contribution. It seems only H.Z is responsible for analysis and data processing, others are all involved in writing. For computations, are there anyone else in the author list?
- Line 439. The unit of the dataset is degree or 0.01 degree? Is 0.01 the scale factor?
- Please revise the followings.

- Line 247-248, as the figure displays
- Line 384, Figures S8j&m. "Figure S8", not figures, and add a blank before j.
- It is suggested to delete equations (1-2) and (3-6), add the reference.
- If necessary, please keep the same description of the time period for there are many kinds in the manuscript, like 2003~2020, from 2003 to 2020, during the year 2003~2020, (2003 to 2020), during 2003~2020, from January 2003 to December 2020, 2003-2020.
- Line 755. Delete the last sentence. Line 759, change : to . after 2003~2020.
- Add longitude and latitude (or the numerical scope signs) in Fig 1, 4-7, 9-12, 14.
- Figure 8. Prediction accuracies of 12 human thermal indices... Add 12. In individual years, change the description. Time series? Add the description of (a) (b) (c).
- Figure 10. Spatial distributions of 12 human thermal indices... Add 12.
- Line 780. Figure 11. National average, are you sure to write like this? Check it in other places in the manuscript. Just use average is fine. Delete "straight". Line 789, Figure 13, national, delete this. Add 12, 12 human thermal indices.
- Line 785, Figure 12, add 12. "the trends of annual mean", please consider how to express better. Inter-annual variation?
- Line 795, Figure 14. HITIC, is it "i"? in mainland China, is it over mainland of China? Please check the description of "mainland China" in the manuscript. In July 2018, move to the end of four major UAs. Delete i.e., change to :.