



## Comment on **essd-2021-267**

Anonymous Referee #2

---

Referee comment on "GPRChinaTemp1km: a high-resolution monthly air temperature dataset for China (1951–2020) based on machine learning" by Qian He et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2021-267-RC2>, 2021

---

This study conducted by Qian He et.al produced a high-resolution air temperature dataset using three types of machine learning methods. The dataset is timely, and fits well the scope of the journal, which could be valuable and interested to the readers and community. The language and the methods of the work is overall good and I enjoyed reading it. I would really like to see this dataset published.

However, there are some points/aspects not clearly enough or needed to be clarified further. I have a number of general comments and suggestions listed below:

- 1) Generating high precision long time series of temperature data in China can effectively meet the needs of scientific research, but there are already high precision temperature data with 1km resolution in China have been released (Zhu X et al, 2019; Peng S et al, 2019), what are the innovative and different points of your data/methods?
- 2) The selection of the characteristic factors: The authors chose three spatially invariant variables, lon, lat and elevation, to predict the dynamic changes of temperature. Whereas these three static factors do not really reflect the changes of temperature and the real spatial distribution characteristics of temperature. Have you ever considered factors such as NDVI vegetation index, land use change, surface temperature, and temporal and spatial correlations, month changes, etc.
- 3) The accuracy of machine learning depends on the adjustment and calibration of hyperparameters. Here, 840 models are used in this study, are these 840 models using the same set of parameters or are each set of parameters different?

4) How do you conduct the accuracy verification of the raster products? The authors used a limited number of 613 sites to generate 1km raster data products. However, as far as I know, the climate modelling sites are too sparse for the Qinghai-Tibet Plateau and Northwest China.