Comment on essd-2021-309

Anonymous Referee #1

- The idea and practice of applying (1) observation data like satellite remote sensing data, conventional ground observation data, (2) DEM data and (3) reanalysis products to build a new datasets through constructing near-surface air temperature model with a certain physical relationship should be encouraged. The researchers have finished a lot of work. However, part of the method in Figure 2 and Figure 3 are not rigorous for the large coverage and for the long time series so the model could be further studied according to some reasonable physical logics. And, The English scientific term usage level and grammar application in this paper needs to be improved further.

- Whether the temperature observation data from meteorological stations have been underwent homogeneous data processing and tested for homogeneity except the steps mentioned in line 207-214? It is very important.

- It is better to apply cloud mask product from the geostationary satellites to get clear/cloud detection for the complete diurnal variation observation, while the spatial resolution is limited. The clear/cloud detection from TERRA and AQUA polar satellites could meet the requirement well enough? Especially when: (1) the area is large enough in China in the scope of Figure 1, (2) the spatial resolution of the dataset is 0.1°. The swath of MODIS is only 2350 km, so **the spatial gap of missing observation is large**, even "same location or Euclidean distance < 0.3°". It is only a compromise to apply the remote sensing data from polar satellite in this research. The coverage is suspicious.

- The construction of near-surface air Temperature models in cloudy days or under complex weather conditions, including the determinations of Tmax and Tmin, are relatively difficult. The research object also involves the diurnal variation of T including Tmax and Tmin, but no detection for each location for 4 times a day or less. The single polar satellite flies over each area twice a day – day and night – **it is temporal missing**. The strategy under clear sky condition in this study is treated relatively to make sense, although some hypothesis may fail the test, like 4 clear sky conditions in the same location in a day, for the temperature peak/valley variation rule differ a lot due to many reasons. What's the portion of this situation? With TERRA as morning satellite and AQUA as afternoon satellite, the diurnal variation curves could not be captured well and the time and frequency of Tmax and Tmin in each pixel/location/grid could not be monitored even using the local sine function (Mao 2016, Jiang 2010) under clear sky days. How to process the temperature data in the same location/same

time with missing observation and reanalysis data? And how far is the closest location? Tave in (Mao 2016) is only limited in "daily mean temperature in various times (1:30, 10:30, 13:30, 22:30)" as clear shown in that paper from remote sensing data (it is feasible under the definition of Tm=(Tm 1:30+Tm 10:30+Tm 13:30+Tm 22:30)/4). The intervals between these 4 MODIS crossing time are 9 hours, 3 hours, 9 hours and 3 hours respectively. But the concept of Tave in this study should be the same with that from meteorological stations, it should be continuous in a day. Besides, the local sine function from (Mao 2016) was obtained for the global from MODIS data during 2001-2012, then whether the equation 5a-6b is statistically computed for China during 2002-2018 to get the coefficiencies? The new simulation method for the diurnal variation of temperature - sub-sine simulation in (Jiang 2010) was based on the hypothesis that the time of Tmax/Tmin is 12 hours later than the time of Tmin/Tmax and the variation from Tmax/Tmin to Tmin/Tmax changes sinusoidal but the temperature in China during 1979-2018/2002-2018 is also this case?

- Not match between "weather status" (like fog, rain, snow) and M*D11 QC fields. Please consider the description "weather status" in the paper. Weather status records observed by observers in meteorological station are not used in this study, as introduced in "3. Data".
- For EAR5 and CMFD datasets, the data themselves are relatively systematic. 2m air temperature from EAR5 reanalysis data is not suitable for evaluation separately like this as truth/reference values for it is the atmospheric reanalysis product, and the advantage, the main application are in isobaric pressure levels while the parameters near the surface is suggested to be analyzed systematically and integrally. The conclusion in Line 200 "the temperature data in the ERA5 data is selected to reconstruct the Ta dataset" may not be feasible. And, why the hourly ERA5 data is suitable to determine the daily weather status in each grid?
- Temperatures are comprehensively influenced by solar short-wave radiation, surface long-wave radiation of the earth, atmospheric water cycle, the weather process, etc. Comparing with other atmospheric elements, its variation extent is relatively small. Sometimes even research method is not perfect but the bias of the temperature is still in a small range. So, it is expected that the correlation analysis results are not poor.
- Why the different colors of scattered dots or anomaly bars used in Figure 6-13?
- Tiny problems.

(1) Line 80, it is suggested to consider the classification again or not mention there are 5 categories of estimation methods from remote sensing data.

(2) Figure 2, Quality control file exist, is it MODIS LST QC field exist? "meteorological station data", revise the description. The equations are too small to read in Figure 4.

(3) Some mistakes are necessary to revise, the following are for your reference. Line 57, "cold days and nights shorten", are they "number of days"? Line 68, time resolution may be temporal resolution. Line 69, which, is it "This way of detection"? Line 70, add "relatively". Line 102, the citation should be NCEP but not JRA-55. It is Kalnay but not Kobayashi. Line 109, forcing datasets, it is suggested not to use it directly. Such as, there are meteorological forcing dataset, atmospheric forcing dataset, and precipitation forcing dataset. Line 188, absorb, not suitable here, is it assimilate, or apply? Line 232, MODIS provides QC fields, the subject of the sentence is wrong, MODIS is a sensor. Line 374 and 711, the citation time of the paper is wrong for it is Ge 2014 but not Ge 2015. Line 640, add "data", add "to", the emphasis here should be "with geostationary satellite DATA, the variation could be monitored". 5.2 title, it is not calibration. The term is not suitable here.

(4) Some descriptions are difficult to understand. Line 177-178, it is suggested to revise the description, "Many studies and analyses show that the dataset's accuracy is high enough to meet the application requirements", what aspects of the accuracy and what

kind of application? Line 274, "Clouds and water vapor have a great influence on visible light and thermal infrared remote sensing", it is radiation but not light, and the emphasis is observation. The sentence should be rewritten. The description of the paragraph following Figure 3 is not clear, especially line 306-311. What's the distance from the closest adjacent stations used for missing gaps? If it is far, the method is not right. Line 391-392, the descriptions are not suitable, "we needed to consider the degree of influence of cloudy-sky weather phenomena. First, we performed effective value statistics on the MODIS data.", "When not all pixels of the MODIS data were valid".

10. https://doi.org/10.5281/zenodo.5513811 is the dataset but not Model code and software.