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

Point 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.

Response 1: Thank you for your good comments and guidance. The daily maximum and minimum temperatures, as well as their average values, are very important for research on climate change and regional energy balance. There are very few maximum-minimum-average air temperature datasets with high spatiotemporal resolution. In order to provide a complete dataset, we have done this work. This dataset is very popular with more than 43,000 downloads (https://zenodo.org/record/5502275#.YasWCLq-uUk). According to your suggestion, we have tried our best to modify and improve it.

Point 2： 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.

Response 2: Thank you for your guidance. The meteorological observation station data we obtained from the China Meteorological Administration have undergone homogeneity data processing and homogeneity testing. Moreover, we have done further checks and have not written very detail here. we have made supplementary explanations in the manuscript, thank you.

Point 3： 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.

Response 3: Thank you for your guidance. You have given us a very good suggestion, that is, in theory, geostationary satellites can indeed better perform the clear/cloud detection for all-sky change observations. However, due to some reasons, it is difficult for us to obtain continuous and stable geostationary meteorological satellite data in China for the past 20 years. MODIS data in TERRA and AQUA are stable and continuous, which are currently the best long-term satellite data. Although the swath of MODIS is only 2350 km, the satellites are sun-synchronized and there are very little missing data. In addition, MODIS data products have done strict data quality control, which is more assured to use, and they are free and publicly downloadable. Furthermore, when the data is missing, we still use the assimilation data ERA5 as the control condition. We didn't descript very detail here, we made a supplementary explanation. In short, your suggestion is very good. In the future, the datasets should try to use geostationary satellite data to judge weather conditions. Thanks again.

Point 4: 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? (0.3°) 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?

Response 4: Thank you for your guidance. You are right, and it is relatively more difficult under cloudy and complicated weather conditions. In fact, whether it is sunny or cloudy or other complicated conditions, the daily temperature changes partially satisfy the sine or cosine function changes. We already have data with a daily interval of 3 hours, so we can determine the time range for the maximum and minimum values. But in order to improve the accuracy, we further constructed the local cosine or sine function and used the daily hourly data of ERA5 as auxiliary data, which can already ensure that we can calculate the local maximum and minimum accurately. In order to calculate the maximum or minimum value, we only need the three locally adjacent maximum or minimum values to determine the coefficients of the function. In this research, we recalculate the coefficients for each pixel, because the maximum and minimum values of places with different latitude, longitude, and altitude are different and change with time. Mao et al. (2016) did a global temperature analysis, and they analyzed the global as a point (a value). In their research, there is no need to distinguish the weather state, so their coefficients can be unchanged. For our study, the weather is changed, so the coefficients of different pixels are changing, and even the same pixel is changing with time. We have made more supplementary explanations in the manuscript.

Point 5： 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”.

Response 5: Thank you for your guidance. We have made revisions to the manuscript. The weather conditions we are referring to here mainly distinguish between clear and non-clear sky. Because the data of weather records in many stations are inconsistent and some data are missing, and there are no weather stations in most areas. So here we mainly use mature MODIS products for judgment. In addition, in fact, we already have 3 hourly interval data every day and ERA5 low-resolution hourly data. These data are also the main basis for our judgment. We have made supplementary explanations in the manuscript.

Point 6： 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?

Response 6: Thank you for your guidance. EAR5 is only used for auxiliary data, which is used to determine the changing trend of temperature and help us determine the time range of the maximum and minimum temperature, and thus determine the coefficient of the function. It is not used to judge the weather status, and we have made revisions.

Point 7： 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.

Response 7: Thank you for your guidance. The trends are pretty good.

Point 8： Why the different colors of scattered dots or anomaly bars used in Figure 6-13?

Response 8: Thank you for your good suggestion. The scatter points of different colors in the figure are used to distinguish the accuracy values before and after correction. In Figure 9-11, the black scatter points are before correction, and orange, blue and green represent the corrected Tmax, Tmin, and Tavg respectively. It can be seen from the figure that the accuracy of the temperature (Tmax, Tmin, and Tavg) after the further correction has been greatly improved.

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

(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.
Response: Thank you for your good suggestion. It is the MODIS LST QC field, and we have made revisions.

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

(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?
Response: Thank you for your good suggestion, and we have made revisions. Some studies use CMFD temperature data as input parameters to construct a surface air temperature model, which shows that the correlation coefficient between CMFD temperature and measured data is greater than 0.99 and has high consistency, and grid data can reflect the temporal and spatial changes of regional air temperature (Zhang et al., 2019; Wang et al., 2017). The CMFD data as an input element to build a surface temperature model can also significantly reduce model deviation and improve model accuracy (Chen et al., 2011).

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.
Response: Thank you for your guidance. We have made revisions to the manuscript and redraw Figure 3. You are right. In fact, we deal with missing data in accordance with this principle. We mainly use 0.3° as a judgment condition. When the distance between stations is less than 0.3° and there are sites, we use site data. When the distance between stations is less than 0.3° and there is no site data, we use the interpolation method. When the distance between sites is greater than 0.3°, we downscale using ERA5 and CMFD data as well as MODIS data.

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”.

Response: Thank you for your guidance, and we have made revisions. Here we mainly want to make full use of the advantages of various data, especially with the help of high-resolution MODIS data to improve the accuracy of the dataset as much as possible. According to the QC field of MODIS data, when the quality of MODIS data is guaranteed, we use MODIS data with high spatio-temporal resolution to improve local accuracy.

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

Response: Thank you for your good suggestion. The model code is at the bottom of the link page, and the folder is named program.zip.