Reviewer #2

Tmax and Tmin are both very important of ecosystem. High quality of long period and high resolution of air temperature products estimated from satellites are still lacking. This paper provided a set of such grided data with elaborated accuracy evaluations. However, there are some key questions needed to answer.

Response: Thank you for your comments and suggestions. Below please find our responses to your comments in detail.

Comment #2-1: I suggest you change the title because you olny produced Tmax and Tmin. 'Daily' means too much. Besides,'global' is more than what you did.

Response: Thank you for your suggestion. We have changed the title as 'A global dataset of daily maximum and minimum near-surface air temperature at 1-km resolution over land (2003-2020)'.

Comment #2-2: Line 77-78: I don't think you improved accuracies of spatial resolutions and temporal coverge. Many current studies have achieved 1km resolution products and the scale of daily even hourly.

Response: Thank you for your comment. We are unaware of any existing products of daily air temperature at the 1-km resolution with a global coverage. We believe our dataset is the first of this kind. We have clarified relevant description in the revised manuscript.

"Our dataset aims to provide the first ever 1-km resolution daily maximum and minimum Ta dataset with a global coverage" (lines 79-80)

Comment #2-3: line 103-104: As you trained models for each day of the period 2003-2020 as well as for each of the five regions, you would got the same number of results from 10-fold cross validation with trained models. Because different models will produce different validation results, how did you got only one result for each region, each landtype, each climate type, each year.... Possibly, you validate the products using all the records, thus, the validation is not independent. Or you caculated the average of RMSE of all models' 10-fold cross validation, such as in one region in one day, thus the RMSE through the paper is debatable.

Response: Thank you for your comment. Our cross-validation is a very rigorous and widely used method for evaluating the accuracy. The uncertainty estimates are based on independent data (not used in the model training). Moreover, our RMSE is a conservative estimate of the true uncertainties of our data because we use all available data, more than those (90%) in the cross-validation. In order to evaluate the model performance for estimating the gridded Ta, we used the 10-fold cross-validation for each day in each region. That is, we equally and randomly divided the data into 10 groups. Nine groups were used as training set and the rest one group was used as the testing set. This process was implemented for 10 times until all the groups were used as the testing set. Each test of the 10-fold cross-validation can generate a RMSE and the average RMSE of the 10 tests was used as the final RMSE for a specific day in a specific region. Therefore, the accuracy assessment was implemented based on independent validation data and can provide a reliable evaluation of the accuracy using the 10-fold cross-validation. We compared daily RMSEs based on 10-fold cross-validation and validation with 30% randomly selected testing data (Fig. R1). We found that the two validation methods show similar results in accuracy. The distributions of RMSEs using the 10-fold cross-validation method is more concentrated with higher maximum densities than those of the 30%-random-validation-based method, especially in Africa and South America due to the low number of validation records, indicating the superiority of the 10-fold cross-validation method in stability.

For each station, we can calculate RMSE based on the time series of estimated and validation Ta in the 10-fold cross-validation. Accordingly, we can calculate mean RMSE and corresponding standard deviation in each land cover type, climate type, and elevation range. We have improved relevant descriptions in the revised manuscript.

"The model performance for estimating gridded Ta was assessed based on root mean square error (RMSE) and mean square error (MAE) using the 10-fold cross-validation in these regions in each day. Taking the RMSE as an example, a RMSE was generated in each test of the 10-fold cross-validation and all RMSEs from the 10 tests were averaged as the final RMSE in a specific day in a specific region. This accuracy assessment using the 10-fold cross-validation was implemented based on independent validation data and can provide a reliable evaluation of the accuracy. For each station, we can also calculate RMSE based on the time series of estimated and validation Ta from the 10-fold cross-validation. Accordingly, we can calculate mean RMSE and corresponding standard deviation in each land cover type, climate type, and elevation range. Specifically, this accuracy assessment represents conservative estimates of the uncertainties of our data because when producing the final results, we used all the available data, more than those in the 10-fold cross-validation" (lines 134-143)

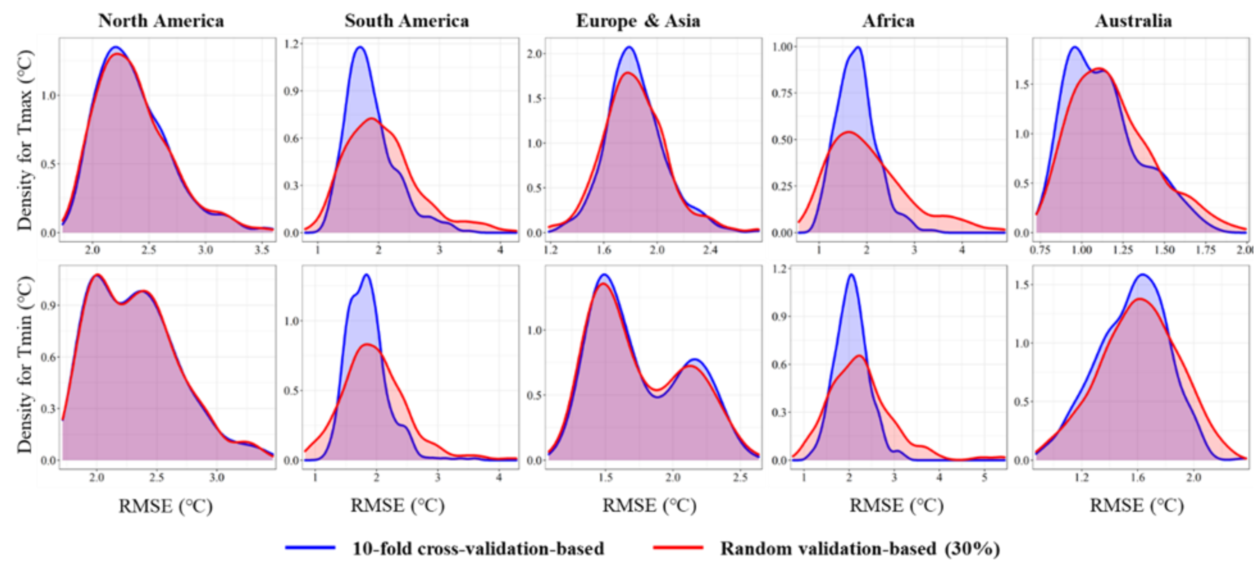

Figure R1: Density of daily RMSEs from 30%-random-validation (red) and mean RMSEs from 10-fold cross-validation (blue) in five regions in year 2010.

Comment #2-4: Please describe how many records you filled because of missing values of weather station observations, and what percent of filled records to all records. Please note it's 'records', not number of sations with valid observations. Figure S1 and S3 would confuse readers because Y axis and title are inconsistent. Besides, how do you define valid station number? For example, Latin America in 2020 has no more than 150 stations before filling missing data. After filling, it is over 150. If one station had only one valid record in 2020, it didn't make sense.

Response: Thank you for your suggestion. The number of filled records for Tmax and Tmin in Africa is 335,900 and 430,652, respectively, accounting for 17.7% and 21.4% of the total records. In Latin America, they are 469,637 and 348,554 for Tmax and Tmin, respectively, accounting for 29.4% and 19.4% of the total records. Based on the daily number of valid records, we drew a boxplot in each year. We have clarified it in the revised supplement.

"Figure S1: Number of valid records (Y-axis) in five regions. Each box is based on the daily number of valid records in a specific year. Each point represents the number of valid records in a specific day."

"Figure S3: Number of valid records (Y-axis) after filling missing values in Africa and Latin America. Each box is based on the daily number of valid records in a specific year. Each point represents the number of valid records in a specific day. The number of filled records for Tmax and Tmin in Africa is 335,900 and 430,652, respectively, accounting for 17.7% and 21.4% of the total records. In Latin America, they are 469,637 and 348,554 for Tmax and Tmin, respectively, accounting for 29.4% and 19.4% of the total records."

Comment #2-5: In equation (1), there is only one LST, however, as both Terra and Aqua observes in one daytime and nightime, there are two LST values. Did you use average value? or lower value? or higher value?

Response: Thank you for your questions. Observations from both Terra and Aqua satellites were used to build gap-filled LST by using a spatiotemporal gap-filling algorithm. More details can be found in Zhang et al. (2022a). We have clarified it in the revised manuscript.

Zhang, T., Zhou, Y., Zhu, Z., Li, X. and Asrar, G. R.: A global seamless 1 km resolution daily land surface temperature dataset (2003–2020), Earth Syst. Sci. Data, 14(2), 651–664, doi:https://doi.org/10.5194/essd-14-651-2022, 2022a.

"The LST dataset is a global seamless 1-km resolution LST dataset at a daily (mid-daytime and mid-nighttime) frequency from 2003 to 2020, which was gap-filled from the MODIS LST products (Zhang et al., 2022a)." (lines 94-96)

re available over large areas." (lines 290-291)