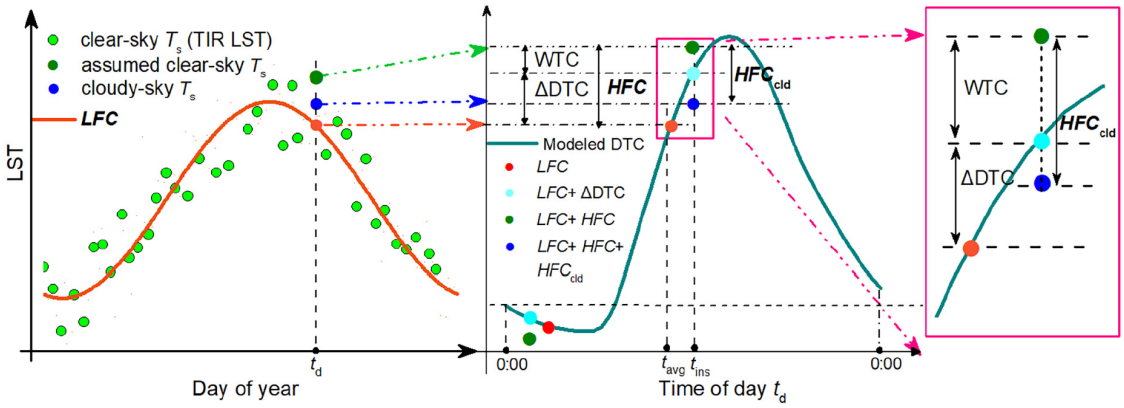
General Comment:

This proposed manuscript produces a daily (four records per day) 1-km all-weather land surface temperature (LST) dataset over China by integrating MODIS Terra, Aqua, and GLDAS reanalysis LSTs from 2000 to 2021. The product after 2002 is produced based on Zhang et al. (2021), and the key contribution of the proposed work is to fill the satellite operation gap from Jan to Feb 2000 for MODIS Terra using random forest (RF) - based extrapolation; the period, Jan 2000 to Jul 2002, for MODIS Aqua LST, is recovered referring to Terra LST in the same period. The core assumption of the proposed algorithm is that the trained RF model from Terra LST with meteorological factors can be directly used for Aqua LST.

Overall, such a contribution is expected to be of interest to the community, particularly compared to the published all-weather LST datasets that typically start from 2002. However, there are some concerns regarding the RF modeling approach, theoretical flaws in the core assumption, and lack of site validation for the proposed algorithm. While the assessment of the product generated after 2002 appears to be sound, there are reservations about the proposed method and the validation of this gap period, which do not appear to be convincing.

Major

In Module III (Section 3.3), it is assumed that the random forest model, trained on Terra LST (morning time), can be directly applied to Aqua LST at noon time to calculate both the LFC (Eq. 5) and HFC (Eq. 6). I have concerns about the validity of this assumption. Based on my understanding of Zhang et al. (2021), the HFC is used to adjust the initially reconstructed LSTs for each day from the averaged passing time (~10:30 am) in a year to the actual passing time (see Figure 2 of Zhang et al).



And such conversion utilized RF modeling, based on location and terrain information, NDVI, water vapor, and time difference for Terra LSTs. And the model was then directly applied for Aqua for the corresponding period, and I think it is not convincing because

1) around Terra's passing time, the warming rate should be high in a day, while at Aqua's passing time, the temperature reaches its peak at noon, which is relatively stable. Thus, LST in the morning is statistically more sensitive to the time difference, and the morning model should also be sensitive to ΔT while when it is used for Aqua data, things are different.

2) The model (eq 6) used in the morning does not take into account the dominant role of solar radiation in warming, which may affect its accuracy when applied to Aqua LST.

Additionally, 3) LST should have quite different sensitivities to the factors in Eq. 6 at both morning and noon times. Therefore, the direct application of the Terra-trained random forest model to Aqua LST may not be correct in theory. Even though some of the previous studies have assumptions to connect Terra and Aqua LSTs, they may not build models in the way as this paper.

In Eq. 8, cloud cover correction for Aqua LST is performed using an RF model that is trained on LST and meteorological factors at Terra time. However, I have concerns about the applicability of the morning model to noon time.

1) GLDAS temperatures, which are essential inputs for the model, exhibit significantly different bias magnitudes at different times of the day. Therefore, it is questionable whether the model trained in the morning can be used for noon time.

2) In addition, cloud conditions can change substantially throughout the day, including changes in cloud cover and cloud types. The statistical relationship used in the RF model may not adequately reflect such changes, and further validation is required to ensure its accuracy.

3) Moreover, the importance of different factors that affect LST may differ between morning and noon times. In the morning, LST warming is mainly controlled by solar radiation, while at noon time, when solar radiation is sufficient, the peak temperature may be highly affected by water vapor. Therefore, it is necessary to consider these factors when applying the cloud cover correction method.

(2) RF modeling details

The proposed method utilized RF modeling several times but in many different ways: Module II extrapolate the Terra LST to Jan 01 2000 by learning the relationship between LST and factors (e.g., DEM, NDVI, soil moisture). Module III reconstructs Aqua LST based on the relationship of Terra LST with environmental factors.

The first question is how the RF model is trained. Did you build one generic model that is applied to all pixels over China? Or several subset models for different climate zones or land cover types? I do not see any description of the RF model training introduction, how the training data is sampled, or how the model is tested, even though RF modeling has been used in every step.

In addition, one-year (2000.03 – 2001.02) data for training is not enough (in theory) as the drought condition may change considerably year by year due to the climate's internal variability.

Did you separate samples for daytime and nighttime? Factors (e.g. albedo) should play different roles in daytime and nighttime.

Driving factors from GLDAS is three-hourly, and how did such coarse time resolution impact your modeling accuracy? Which interpolation methods did you use?

(3) Independent site validation does not cover 2000-2002

This is also a serious issue, and only the spatial comparison with ERA5 and GLDAS is not enough for claiming the proposed algorithm is ready for production. Independent ground validation is still necessary. Suggest either collecting more site data for assessment or assuming 2003-04 are missing and recovering these two years by the same model you trained in modules II and III. And then the site validation and inter-comparison with Zhang et al (2021) can be done.

You may also try cross-validation of modules II and III to check which module is most robust. Essentially, both modules II and III can recover Aqua LST before 2002.07.

Other Concerns

(1) Table 1: site representativeness of fluxes should be highly related to the observing height and local terrain; however, comparing SDQ to D105 and GAZ, even D105 and GAZ have very low heights and high elevation, they have similar *u*REP with SDQ, can you explain that? What does the '*' mean?

(2) How did you get the albedo and NDVI before the Terra recording time for Module II?

(3) I double-checked the data while it is not standardized, suggest outputting it as NC or HDF file, scaling the matrix, and providing scale factor, and offset, just like MODIS product, which would be easier to be used for modeling input. Write the view time and LST into one file, data quality mark, and cloud mask are also necessary for users. Suggest not compressing the data, even if it is required by the platform, it is better to use ZIP rather than RAR, ZIP is more accessible worldwide.

Minor

(1) Line 11: many spatial missing -> many invalid pixels

(2) Line 16: suggest pointing out four times records per day, which is the strength of this work.

(3) Line 22: the "temporal gap" should be clarified otherwise it may mislead readers as the gaps caused by cloud cover rather than satellite operation time.

(4) The quotative analysis part in the abstract should divide the validation for the years after 2002 (previous work) and before 2002 (proposed work) because the results after 2002 should be good as it has been peer-reviewed. Further, the bias at Line 27 is huge honestly, suggest also giving the overall statistics rather than the range.

(5) Line 43: "Grain for Green Program" is the official term

(6) Line 71: "regions. LST (Martins et al., 2019)." wrong sentence

(7) Line 72: as 'a' physical method

(8) Line 75: MLST-AS should be spelled out, and double-check the whole context for abbr.

(9) Line 80: 'observe LST' -> 'retrieve LST'

(10) Line 97: suggest directly specifying the period rather than mentioning ‘outside the temporal gaps’

(11) Line 109: space missed in ‘product(’

(12) Line 116-120: as GLDAS is the key input of LST reconstruction here, its LST data should be introduced in detail.

(13) Table 1: adding a spatial map of sites and land cover types would be great

(14) Line 127: literature reference is necessary here for “3% - 10%”

(15) Line 129: which BBE data do you use for calculating site LST measurements

(16) Line 133: Brief introduction of uRep should be given.

(17) Line 175, Line 190: suggest replacing ‘T-i’ to ‘T,i’ as ‘–‘ means minus that will mislead readers, d-1 looks like the day before d.

(18) Line 184-185: what is the data source of the factors, how interpolated, why solar radiation is not included, why deep layer soil moisture is included, and is the model really sensitive to it?

(19) Line 197: what is the threshold for classification and any reference

(20) Line 199-200: hard to read

(21) Line 217: why DOY 3?

(22) Line 233: TI -> T1

(23) Eq. 7: what G means?

(24) Line 310: During the ‘temporal gap’, there are no Terra or Aqua passing, such statement is misleading and should clarify how the time in this period is determined.

(25) Line 312: clarify the interpolation method

(26) Line 328: ‘Due to limited space’, all statements like this should be deleted.

(27) Line 361: did the anomalies of site observations or products removed? And the grammar of the sentence is not correct.

(28) Table II: MXD11 has been suggested as having substantial bias at desert areas, and why in this table the MBE is 0.79, close to other land cover types?

(29) Line 392: MXD21 has been validated that performs better than MXD11 and why it is not used?

(30) Sect. 4.4 Would it be better to move to the intro to clarify the importance of the data?

(31) Line 442: spell TPDC, please

Reference

Zhang, X., Zhou, J., Liang, S., & Wang, D. (2021). A practical reanalysis data and thermal infrared remote sensing data merging (RTM) method for reconstruction of a 1-km all-weather land surface temperature. *Remote Sensing of Environment, 260*, 112437