# Responses to RC2

*Dear Reviewer #2:*

*Thanks very much for your time on reviewing our manuscript. We sincerely thank the reviewer for your efforts on the reviewing of our manuscript. We deeply appreciate your valuable comments on our manuscript, and we have carefully revised the manuscript according to the comments. The point-by-point responses to your comments are provided in this document.*

*Best regards,*
*Zhenwei Zhang*
*Nanjing University of Information Science & Technology*

The manuscript essd-2024-548 has been reviewed. It presents the first global hourly surface air temperature dataset (GHRSAT) from 2011 to 2023. The manuscript demonstrates clear organization, rigorous logical flow, and natural transitions between sections. While the overall quality of the work is commendable, several critical issues require attention, as outlined in the following comments:

**Comments #1.** The core method of this manuscript is the RF-KR method, that is, the RF model is used to build the site-scale SAT estimation model, and then the residual is interpolated by the Kriging interpolation method to obtain pixel-by-pixel residual data. Logically, it is possible, but the validation is not sufficient, that is, whether the cross-validation used in the test process can be validated with independent data to explain its accuracy better.

**Responses #1**: Thanks for your valuable comments. To offer a further validation of our models, we have additionally performed site-based cross-validation (CV) for all models developed in our study. In the site-based cross-validation, the sites were first randomly divided into ten sets, and samples from the sites in each set are treated as one fold of samples (see Lines 289-294 in Sec. 3.3). For site-based CV, models are trained using samples from nine sets of ground station, and the models are then validated by samples from one remaining set of stations. Thus, the models are assessed by completely independent sites. The validation results for sited-based CV have been discussed in our revised manuscript (Lines 26-29 in the Abstract, Lines 332-359 in Sec. 4.1, and Lines 694-699 in Sec. 6). Fig. 4 has been revised to contain the overall validation results for our models under both sample-based and site-based cross-validation (Lines 384-389), and some additional figures have been added in our revised supplement file (Fig. S4, S5, S6).

**Comments #2.** In the process of model construction, the GHA-LST dataset is used as the main input, but it is recommended to discuss how its uncertainty will affect SAT estimation.

**Responses #2**: Thanks for valuable comments. We agree with you that uncertainty associated with model inputs, especially the GHA-LST reconstructed dataset, is important for the SAT models based on GHR-LST. But there is no available uncertainty information at the pixel level for GHR-LST, it is hard to design modeling experiments to quantitively analyze the impacts of GHR-LST reconstruction uncertainty on SAT estimation models. The SAT models use the ground measurements of SAT as the target variable, which is measured at high accuracy. In general, the errors and uncertainty of reconstructed LST and other inputs will be reflected by the predictive performance of the SAT models. We have briefly discussed this issue in our revised manuscript (Lines 411-413).

**Comments #3.** How to consider the spatial representativeness of the air temperature observed at the station on the 5km scale.

**Responses #3**: Thanks for pointing out this important question. The scale-mismatch between ground point-level station measurements and areal (pixel-level) observations from satellites is an innate and challenging issue not only for SAT estimation studies, but for a wide range of research fields using remote sensing data. SAT estimation models are trained using matched samples from ground stations. The scale-mismatch issue will **impact the spatial representativeness of the matched samples (or sampling errors), which further influences the predictive performance (errors) of the SAT models**. Better predictive performance for estimating SAT can be achieved by building models with more representative samples processed from high-density ground stations. However, it is very difficult to completely resolve the issue regarding spatial representativeness for ground sampling, given current status and developments of ground-based observational networks. To obtain more spatially representative samples for modeling SAT at the 5-km scale, very-high density of ground stations should be available for a study area. For example, the HiWATER research program (Li et al. 2013) established one network of high-density ground observation sensors in Heihe for matching a footprint of MODIS observations. But it is generally impossible to operationally maintain a very-high density of networks for large-scale areas under the constraints of financial supports.

*Li, X., Cheng, G., Liu, S., Xiao, Q., Ma, M., Jin, R., Che, T., Liu, Q., Wang, W., Qi, Y., Wen, J., Li, H., Zhu, G., Guo, J., Ran, Y., Wang, S., Zhu, Z., Zhou, J., Hu, X., and Xu, Z.: Heihe Watershed Allied Telemetry Experimental Research (HiWATER), 16, 2013.*

**Comments #4.** Figure 2, which Kriging method was used for TR-4?

**Responses #4**: Thanks for the comment. We used the FRK method to model the site residuals from RF models constructed for TR-4. We have modified Figure 2 (Line 195).

**Comments #5.** Figure 4, why are RMSE and MAE so large in TR-1 and TR-6?

**Responses #5**: Thanks for your important comments. In general, SAT estimation models developed for regions with complex geographical environments and climatic dynamics such as polar regions (Nielsen et al., 2023; Meyer et al., 2016; Kilibarda et al., 2014) exhibit larger predictive errors and uncertainties. There are very limited ground stations in the polar regions (TR-1 and TR-6) and sampling representativeness by the scarce stations will deteriorate the predictive performance of SAT estimation models trained using samples from the stations in the regions. In our revised manuscript (Lines 360-371), we have discussed issue of scarcity of stations for SAT model building, and state that developing models based on transfer learning (Wang W. et al., 2025) offers a promising and important approach for estimating SAT in regions with the scarcity of stations (Lines 370-371).

*Kilibarda, M., Hengl, T., Heuvelink, G.B.M., Gräler, B., Pebesma, E., Perčec Tadić, M., Bajat, B., 2014. Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. J. Geophys. Res.: Atmos. 119, 2294–2313. https://doi.org/10.1002/2013JD020803*

*Nielsen, E. B., Katurji, M., Zawar-Reza, P., and Meyer, H.: Antarctic daily mesoscale air temperature dataset derived from MODIS land and ice surface temperature, Sci Data, 10, 833, https://doi.org/10.1038/s41597-023-02720-z, 2023.*

*Meyer, H., Katurji, M., Appelhans, T., Müller, M., Nauss, T., Roudier, P., and Zawar-Reza, P.: Mapping Daily Air Temperature for Antarctica Based on MODIS LST, Remote Sens., 8, 732, https://doi.org/10.3390/rs8090732, 2016.*

*Wang, W., Brönnimann, S., Zhou, J., Li, S., and Wang, Z.: Near-surface air temperature estimation for areas with sparse observations based on transfer learning, ISPRS Journal of Photogrammetry and Remote Sensing, 220, 712–727, https://doi.org/10.1016/j.isprsjprs.2025.01.021, 2025.*

**Comments #6.** Figure 5, why does the RMSE of the model have such strong periodicity?

**Responses #6**: Thanks for your important comments. The changing of surface air temperature (SAT) has very strong seasonality and periodicity. For the northern hemisphere, SAT reaches two extremes for summer months and winter months. In general, it is harder to build models for extreme SAT (summer or wither months) than for moderate conditions of SAT (spring and autumn months). Therefore, the RMSE for models developed across different months exhibit periodicity. The periodical variability of RMSE for SAT models has also been reported in previous studies (Wang M. et al., 2024; Yao R. et al, 2023). We have discussed the issue in Line 390-410.

*Yao, R., Wang, L., Huang, X., Cao, Q., Wei, J., He, P., Wang, S., and Wang, L.: Global seamless and high-resolution temperature dataset (GSHTD), 2001–2020, Remote Sens. Environ., 286, 113422, https://doi.org/10.1016/j.rse.2022.113422, 2023.*

*Wang, M., Wei, J., Wang, X., Luan, Q., and Xu, X.: Reconstruction of all-sky daily air temperature datasets with high accuracy in China from 2003 to 2022, Sci. Data, 11, 1133, https://doi.org/10.1038/s41597-024-03980-z, 2024.*

**Comments #7.** Line 211, what is RF-KR?

**Responses #7**: Thanks for your concerns. We developed hybrid estimation models that integrate RF and residual kriging for reconstructing our dataset. Two types of kriging including OK and FRK have been utilized for constructing the hybrid models. In our manuscript, we use the abbreviation RF-KR to generally refer the hybrid models. We have reorganized and rewritten some contents of section 3.2 (Lines 215-223, and Lines 244-251) to clearly describe all model abbreviations used in our manuscript.

**Comments #8.** Line 226, Should the formula number be Eq. (3)?

**Responses #8**: Thank you for pointing out this mistake. We have corrected it (Line 261).

**Comments #9.** Is the time label of the dataset local time or UTC? This is critical for users.

**Responses #9**: Thanks for this important comment. The time standard for our GHRSAT product is UTC. We have added contents for the time standard in sec. 3.1 (Lines 199-200) and sec. 5 (Line 676).

**Comments #10.** For the air temperature estimate in the case of sparse sites, please ref er to these two articles: https://doi.org/10.1016/j.isprsjprs.2025.01.021; https://doi.org/10.1 109/JSTARS.2022.3161800

**Responses #10**: Thanks for your valuable comments. The two articles are very pertaining to our study, and we have referred the two articles in our revised manuscript to discuss the significance of building models based on reconstructed LST (Wang et al. 2022, in Line 85) and state the significance of developing SAT model based on transfer learning for estimating SAT in regions with limited stations (Wang et al., 2025, in Line 371).