Revisions of Manuscript: ESSD-2025-192

Title: Spatially adaptive estimation of multi-layer soil temperature at a daily time-step across China during 2010-2020

Author(s): Xuetong Wang, Liang He, Peng Li, Jiageng Ma, Yu Shi, Qi Tian, Gang Zhao, Jianqiang He, Hao Feng, Hao Shi, Qiang Yu

Dear Reviewer,

We sincerely thank you for your thoughtful comments and constructive suggestions on our manuscript. We have carefully revised the manuscript in response to your feedback, with all changes clearly marked using track changes. In the revised manuscript and accompanying supplementary materials, modifications are highlighted in blue for ease of reference.

Below, we provide a detailed, point-by-point response to each of your comments. For clarity, your original remarks are shown in *italics*, followed by our corresponding replies. We have made every effort to address all concerns comprehensively and to improve the scientific rigor, clarity, and overall quality of the manuscript.

We sincerely appreciate the time and effort you invested in reviewing our work.

Response to Reviewer3_Comments

Reviewer Comment 1:

The effectiveness of spatial block CV is dependent on the block size being large enough to account for the spatial autocorrelation range of the data. The authors have used 1° (~100 km), which needs to be justified. Some previous studies (e.g., Ploton, et al. 2020 Nature Communications 11 (1): 4540) showed that typical climate variables can exhibit significant spatial correlation up to 500 km. Therefore, the authors should provide a justification for their choice of 1°, for instance by using semivariograms of the observed Ts to determine the distance at which spatial autocorrelation becomes negligible. Otherwise, the data leakage problem is only reduced, not solved.

Response to Reviewer Comment 1:

Thank you very much for your valuable comment. We recognize that the effectiveness of spatial block cross-validation (CV) largely depends on the size of the blocks, which must be sufficiently large to capture the spatial autocorrelation structure inherent in the data. In addition, the spatial block partitioning strategy previously adopted in our study was limited by an inadequately defined threshold and, more importantly, by the fact that the boundaries of adjacent blocks were often very close to each other. As a result, stations located near block edges could still be separated by only short distances and thus remain spatially correlated, making it difficult to fully eliminate potential data leakage under the original partitioning scheme.

Following your suggestion, we performed a semivariogram analysis to determine the optimal distance required to effectively reduce spatial autocorrelation in the T_s data. As shown in Figure S8, the semivariogram of the T_s data reaches a plateau at a distance of approximately 400–500 km, with only minor variations beyond this range. This indicates that spatial autocorrelation in T_s declines substantially around 500 km and becomes negligible at greater distances. This finding is consistent with Ploton et al., (2020), who also reported significant spatial dependence in climate variables over scales of 250~500 km.

Based on these results, we revised our sampling strategy to more effectively mitigate spatial autocorrelation in the T_s station data and to reduce the risk of data leakage between the training and testing subsets arising from spatial dependence. The revised method has been updated in Section 2.3.3 of the manuscript. We believe that these improvements better address the reviewer's concerns regarding spatial autocorrelation and further enhance the reliability of our model's generalization assessment. We sincerely appreciate your insightful comment, which has substantially strengthened the methodological rigor of our study.

Revised Text in Section 2.3.3 (L354-L375):

Significant spatial autocorrelation commonly exists among nearby T_s observation sites.

To prevent potential data leakage caused by randomly splitting the training and testing subsets, we conducted the partitioning at the station level and constructed a buffer zone around the selected test station. All other stations located within this buffer were removed, and only stations outside the buffer were retained as the training set. This strategy effectively ensures that samples within the same sub-grid do not appear simultaneously in both the training and testing subsets due to spatial autocorrelation, thereby allowing a more robust and unbiased assessment of the model's generalization performance.

Specifically, considering the availability of sufficient training samples, one station was randomly selected as the test sample within each sub-grid. A 500 km buffer was subsequently created around the test station, with the radius determined based on the effective distance for reducing spatial autocorrelation among stations as shown in Appendix Figure S8. All stations within the buffer were excluded, and only those outside the buffer were used for model training. Subsequently, five-fold cross-validation was performed at the station level, and GridSearchCV was used to optimize three key hyperparameters: the number of trees (n_estimators), maximum tree depth (max_depth), and learning rate (learning_rate). The search ranges for these parameters are provided in Appendix Table S1. The optimal hyperparameter combination was identified by minimizing the mean validation error. Finally, the model was retrained on the full training subset using the optimized parameters and evaluated on the spatially independent test sample to rigorously assess its generalization capability.

Here are the revisions, supplemented in the Appendix (L95-L100):

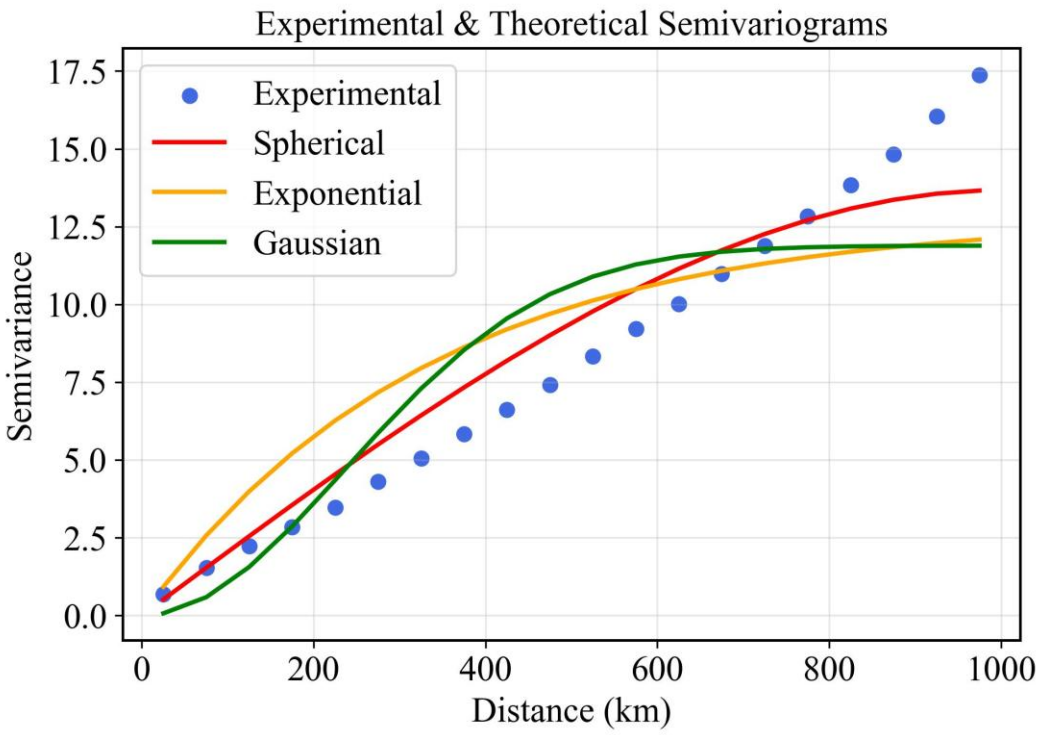

Figure S8. Experimental and theoretical semivariograms of annual mean T_s at 0 cm (The spherical, exponential, and Gaussian models are fitted for comparison).

Reference

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S., and Pélissier, R.: Spatial validation reveals poor predictive performance of large-scale ecological mapping models, Nat. Commun., 11, 4540, https://doi.org/10.1038/s41467-020-18321-y, 2020.

Reviewer Comment 2:

In their response, the authors present a new figure (labeled Figure 1 in the response letter) that validates the model's ability to capture the spatial distribution of annual mean Ts. However, the number of points in this figure looks far greater than the ~200 sites that would constitute a 10% test set, suggesting that all stations are included in this validation. For a validation to be a true test of generalization, it must be performed exclusively on the held-out test blocks.

Furthermore, I suggest an enhancement: color-code the points in the plot by their parent quadtree grid. This would provide a visual assessment of the performance of the different localized models.

Response to Reviewer Comment 2:

Thank you for your valuable suggestions, which have helped us further enhance the rigor and clarity of our spatial validation and visualization. In the previous response letter, Figure 1 indeed included all stations rather than only the independent test-block stations. Therefore, in the revised manuscript, we validate the model's ability to capture the spatial distribution of annual mean T_s using only the test set.

Regarding your suggestion to color-code the points based on their quadtree grids, our quadtree system is rotated at six different angles. Under different rotations, both the number of sub-grids and their identifiers vary, and the selected test set is not fixed across rotations. As a result, it is not feasible to apply a consistent color-coding scheme to represent grid membership within a single figure.

Nevertheless, the revised Figure S12 clearly demonstrates that across all rotation angles and corresponding grid structures, the localized models consistently and accurately capture the spatial distribution of annual mean T_s when evaluated on the independent test set. This further confirms the robustness of the rotated-quadtree modeling framework.

Here are the revisions, supplemented in the Appendix (L118-L121):

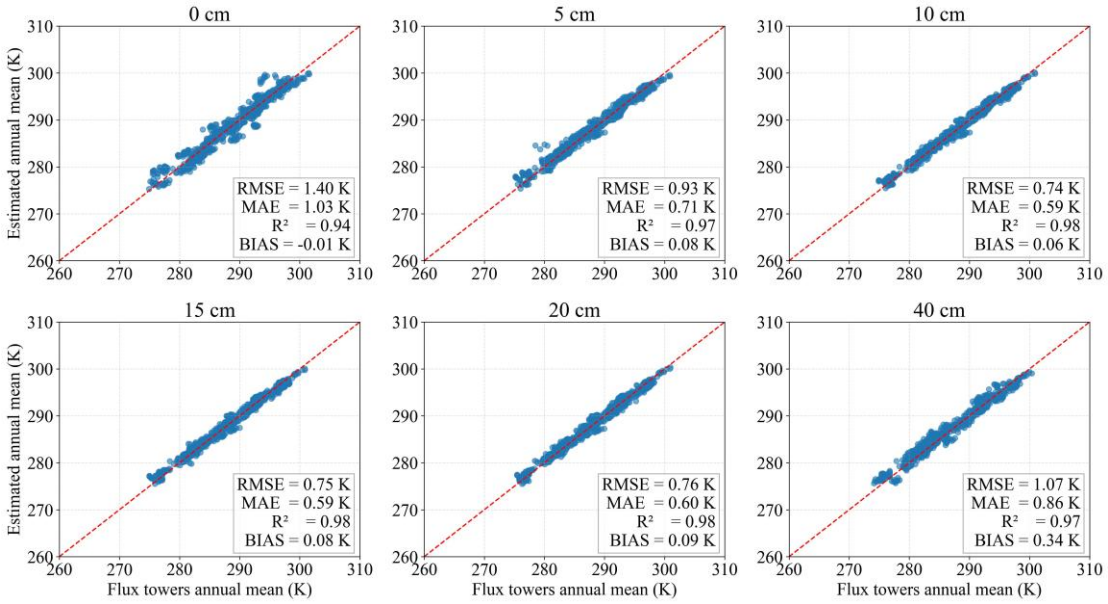

Figure S12. Comparison between estimated and observed annual mean T_s across six depths (0~40 cm)

Reviewer Comment 3:

The addition of an independent validation using 18 flux tower sites is a great improvement. The authors present the daily time-series comparison in the new Figure 5. To maintain consistency with the most robust validation practices, I suggest that the authors also present the annual mean Ts (spatial-only validation) comparison between model and observation.

Response to Reviewer Comment 3:

We appreciate the reviewer's insightful suggestion. Following your recommendation, we have now included an additional spatial-only validation based on the annual mean T_s derived from the 18 independent flux tower sites. This new analysis directly assesses the model's ability to reproduce the spatial distribution of annual mean T_s without relying on temporal information.

The results are presented in the newly added Figure S13, which compares the observed and estimated annual mean T_s across six soil depths (0~40 cm). Despite the small number of independent sites and the strong climatic heterogeneity among tower locations, the model achieves reasonable agreement with observations, demonstrating its capacity to capture the spatial variability of annual mean T_s under independent conditions. This addition ensures consistency with best practices in model evaluation and provides a more comprehensive and rigorous validation of the model's spatial performance.

Here are the revisions, supplemented in the Appendix (L122-L126):

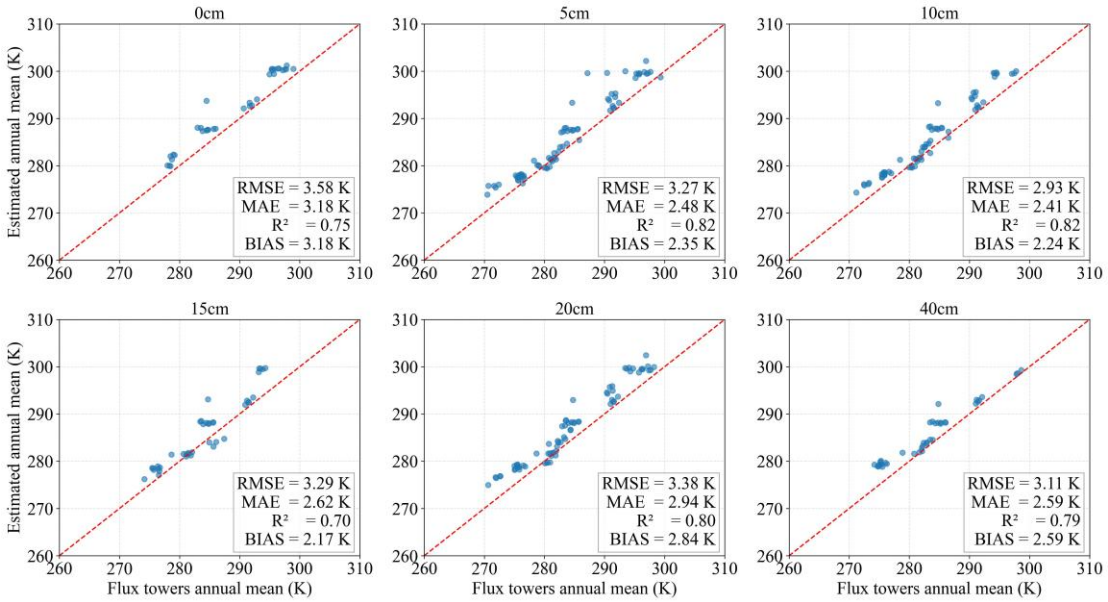

Figure S13. Comparison between estimated and FLUX towers annual mean T_s across six depths (0~40 cm)

Response to Reviewer4_Comments

Specific comments:

Reviewer Comment 1:

Discussion on time-series is missing. How has the Ts changed during this decade?

Response to Reviewer Comment 1:

Thank you for this helpful comment. We agree that our current manuscript does not include a full discussion of long-term T_s changes over the entire decade (2010–2020). In our study, interannual variability of T_s during this period is relatively small, and our primary objective was to evaluate the model's ability to reproduce T_s dynamics rather than to analyze long-term climate trends. Therefore, instead of presenting a full decadal trend analysis, we selected four representative stations and examined in detail their daily variations of air temperature, LST and T_s for the period 2018–2019 as illustrative examples.

Revised Text (L524-L541):

To further assess the temporal performance of T_s estimation, Fig. 11 presents the time series of estimated T_s alongside in-situ measurements at four randomly selected stations (e.g., Station 56748, 99.18°E, 25.12°N) from January 2018 to January 2020. The figure displays T_s at two depths (0 cm and 40 cm), including estimated T_s (Estimated_0cm, Estimated_40cm), in-situ T_s (In-situ_0cm, In-situ_40cm), daily mean land surface temperature (Daily_mean_LST), and 2-meter air temperature (Temperature_2m). The air temperature shows distinct seasonal cycles, while T_s exhibits smoother temporal variations. In general, T_s reaches its peak during summer and its minimum in winter, though its temporal dynamics vary with soil depth. Specifically, T_s at 0 cm responds rapidly to air temperature changes and exhibits larger amplitude variations, while T_s at 40 cm shows slower responses and a noticeable lag, reflecting the damping effect of vertical heat conduction. Site-level accuracy was evaluated using RMSE, which ranged from 1.24 K to 2.05 K across both depths, indicating strong agreement between predicted and observed values. Overall, the time series analysis confirms the robustness and reliability of the model in estimating T_s across varying depths, offering valuable insights into regional soil thermal dynamics.

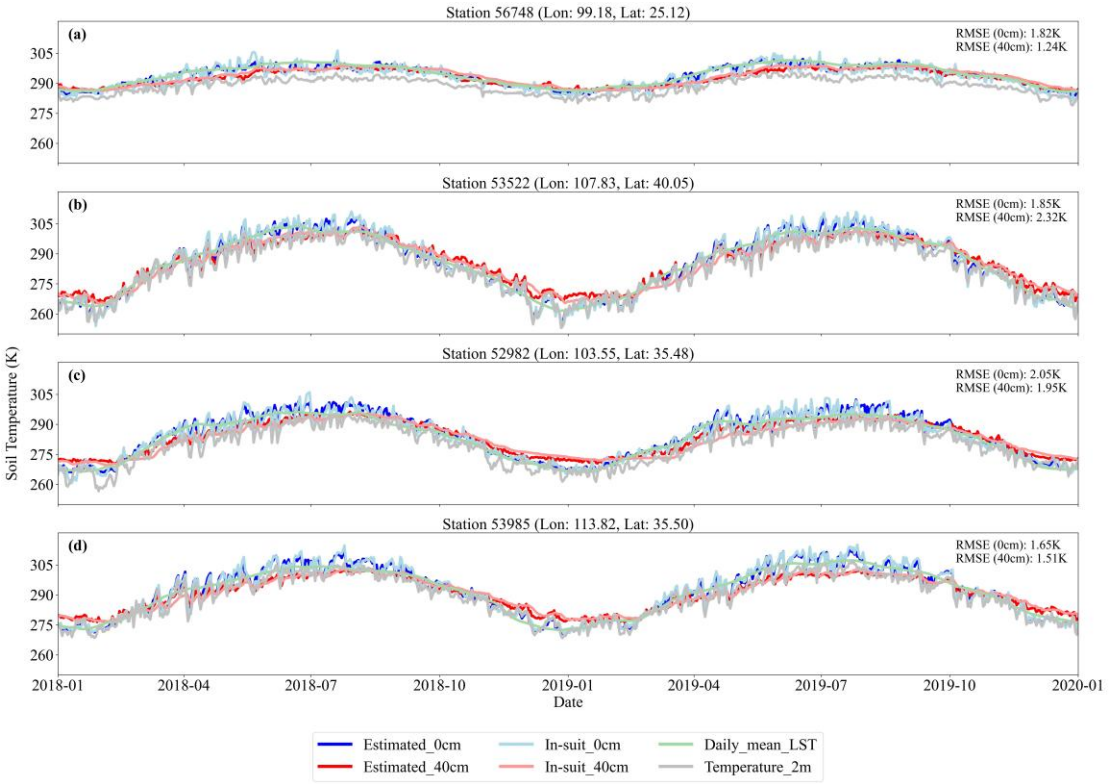

Figure 11. Time series of the Estimated_0cm, Estimated_40cm, Daily_mean_LST, and Temperature_2m at four sites from different regions between 2018-2019.

Reviewer Comment 2:

Too many figures in the main text. Suggest to move some into supplemental material.

Response to Reviewer Comment 2:

Thank you for your helpful suggestion. Following your recommendation, we have streamlined the presentation of figures in the main text. Specifically, we have merged the original Sections 3.1 and 3.2 into a new Section 3.1 to reduce redundancy and improve the clarity of the Results. In addition, Figure 7 from the former Section 3.2 has been moved to the Supplementary Material. These adjustments help keep the main manuscript focused on essential results while ensuring that all supporting visualizations remain accessible in the appendix.

Reviewer Comment 3:

Adaptive scaling can lead to confusion when interpreting the spatial resolution of a dataset. I would recommend authors to report the "effective spatial resolution" information of each grid, or store it as a separated supporting dataset. Otherwise, it is hard to be used for comparison against other regional/global products with fixed spatial resolution.

Response to Reviewer Comment 3:

Thank you very much for your thoughtful comment. We fully understand the concern that adaptive spatial partitioning may cause confusion regarding the spatial resolution of the final dataset. Here we would like to clarify that, although our modeling framework uses a rotated-quadtree structure to adaptively partition the training space

based on the density of in-situ observations, the final T_s product is always generated at a fixed spatial resolution of 1 km.

The adaptive quadtrees are used only during the modeling stage to train localized XGBoost models under multiple rotation angles. After model training, the predictions from all rotated-quadtree configurations are aggregated and mapped onto a uniform 1-km grid across China. Therefore, the effective spatial resolution of the final product does not vary across space, and all output grids represent the same 1-km resolution regardless of quadtree size during training.

We hope this clarification addresses your concern, and we appreciate your attention to dataset usability and transparency.

Reviewer Comment 4:
The dataset is claimed to be freely accessible, but I cannot access the data through FTP. Authors need to verify and check the link, or upload a copy to an open access repository, otherwise it does not meet the requirement of ESSD.
Response to Reviewer Comment 4:
Thank you for pointing out this issue. We sincerely apologize for the inconvenience caused by the temporary inaccessibility of the FTP link. After receiving your comment, we immediately contacted the data center and confirmed that the FTP remote server itself had not experienced any malfunction. To ensure that the dataset fully meets ESSD's data accessibility requirements, we have taken the following steps:
(1) The complete T_s dataset has been re-uploaded to the National Tibetan Plateau Data Center (TPDC), and the DOI download links have been fully verified and are functioning properly. Two stable access links are now provided for users:
https://doi.org/10.11888/Terre.tpdc.302333
https://cstr.cn/18406.11.Terre.tpdc.302333
(2) If access issues persist, we kindly recommend users to switch to another network environment (e.g., non-campus network or networks without firewall restrictions), as some institutional networks may block external data repository access.
At present, the dataset is fully accessible through the verified DOI links and fully satisfies the data availability requirements of ESSD. We sincerely appreciate your valuable feedback.

Technical corrections:

Reviewer Comment 1:
Line 108 - 116: This part can be moved to objectives. When describing each objective, state the uniqueness/improvement from your study.
Response to Reviewer Comment 1:
Thank you very much for your valuable suggestion. Following your recommendation, we have revised the corresponding part of the Introduction. The updated content has now been incorporated into the revised manuscript.

Revised Text (L105-L135):

Recent advances in spatially adaptive modeling have increasingly emphasized the importance of addressing spatial heterogeneity and uneven sampling density in environmental datasets. Classical quadtree structures and related hierarchical spatial data models provide the theoretical foundation for constructing adaptive, variable-sized spatial partitions, enabling efficient organization of multiscale spatial information through recursive subdivision (Samet, 1984). Building on this foundation, Lagonigro et al., (2020) developed the AQuadtree R package, which provides an adaptive spatial partitioning framework capable of generating variable-sized grid cells according to the spatial distribution of observations. This adaptive partitioning produces finer grids in data-dense regions and coarser grids where observations are sparse, ensuring a spatial structure that better reflects sampling heterogeneity and improves the model's capacity to capture localized spatial variability. Extending this idea, we develop a rotated-quadtree strategy that applies multiple orientation angles during the quadtree subdivision process. This enhancement allows the model to capture spatial heterogeneity from multiple directional perspectives, and averaging predictions across rotation angles substantially reduces the boundary artifacts that may arise from single-angle grid partitioning, ultimately improving the robustness of local modeling under complex environmental gradients.

To address the irregular station distribution, and non-stationarity commonly encountered in large-scale T_s estimation, we construct a spatially adaptive modeling framework based on the rotated quadtree approach. Within each grid cell, multi-source environmental predictors are integrated with in situ station records, and T_s is estimated using XGBoost models. Based on this framework, the objectives of this study are to: (1) construct a spatially adaptive modeling system; (2) generate a multi-layer T_s dataset at a daily time-step and one kilometer resolution in China from 2010-2020; and (3) evaluate the dataset through independent validation with flux tower observations and benchmarking against widely used T_s products. The proposed methodology could directly address the scaling challenges induced by spatial heterogeneity and uneven data distribution. The generated products would provide a robust foundation for high-resolution environmental modeling, precision agriculture and climate impact assessments.

Reference

Lagonigro, R., Oller, R., Martori, J.C., 2020. AQuadtree: An R package for quadtree anonymization of point data.

Samet, H., 1984. The quadtree and related hierarchical data structures. ACM Comput. Surv. CSUR 16, 187–260.

Reviewer Comment 2:

Line 110: I would be curious about how to use your data product, since most of the observed/modeled data products have uniform spatial resolution, and your varying resolution product will be hard for intercomparison against them.

Response to Reviewer Comment 2:

Thank you for raising this important question. We would like to clarify that although our modeling framework employs an adaptive, rotated-quadtree structure during the training stage, the final T_s product is always generated on a uniform 1-km grid across China. The adaptive partitioning only determines how local XGBoost models are trained according to site density, and it does not affect the spatial resolution of the final gridded product. To avoid potential misunderstanding, we provide the following clarification here:

(1) The adaptive grid system is used solely for local model training;

(2) The final predicted T_s fields are mapped to a fixed 1-km grid;

Thus, users can apply our T_s dataset in the same manner as any other 1-km gridded environmental product, without concerns related to heterogeneous spatial resolution.

Reviewer Comment 3:

Line 252: How did you treat auxiliary predictors differently from main predictors?

Response to Reviewer Comment 3:

Thank you for raising this point. In addition to the distinction between main and auxiliary predictors, we would like to further clarify how multicollinearity was handled and why both air temperature and LST were retained in the final model.

Before modeling, we conducted a comprehensive variance inflation factor (VIF) analysis for all predictor variables to remove those exhibiting strong multicollinearity. The results are shown in the Supplementary Figures. As expected, air temperature and satellite-derived LST showed high collinearity. To determine whether both variables should be retained, we performed a comparative modeling experiment using two predictor combinations:

(1) Air temperature + other predictors

(2) Air temperature + LST + other predictors

As shown in Supplementary Figure S3 and Figure S4, the second combination (air temperature + LST + other predictors) consistently produced the best predictive performance across depths. This indicates that although air temperature and LST are correlated, they contain complementary thermal information—air temperature captures large-scale atmospheric conditions, whereas LST provides fine-resolution surface radiometric temperature signals.

Therefore, despite their statistical collinearity, both variables were retained in the final model to maximize predictive accuracy. This rationale has been added to Section 2.3.1 of the revised manuscript.

Here are the revisions, supplemented in the Appendix (L5-L20):

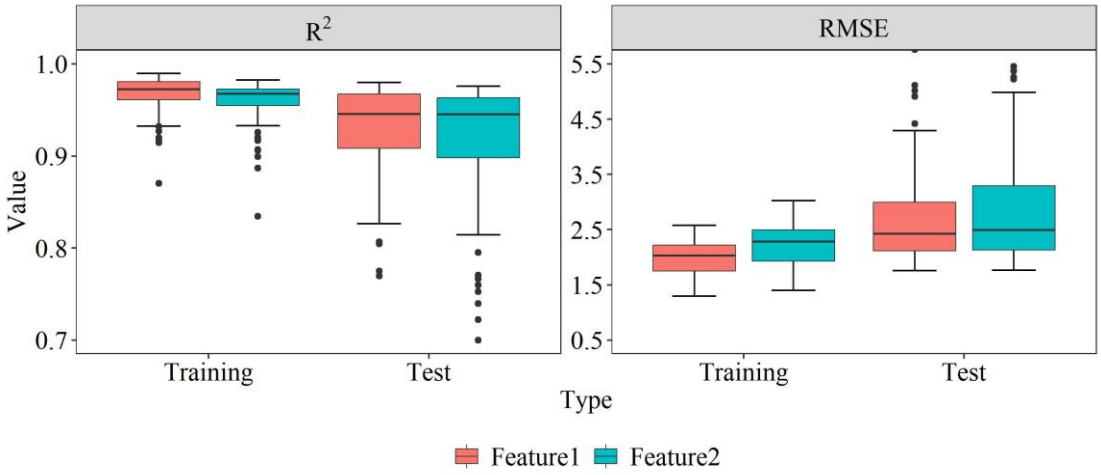

Figure S3. Comparison of Modeling Accuracy with Different Feature Variables (Feature1 represents using both air temperature and LST together with other feature variables, while Feature 2 represents using only air temperature together with other feature variables)

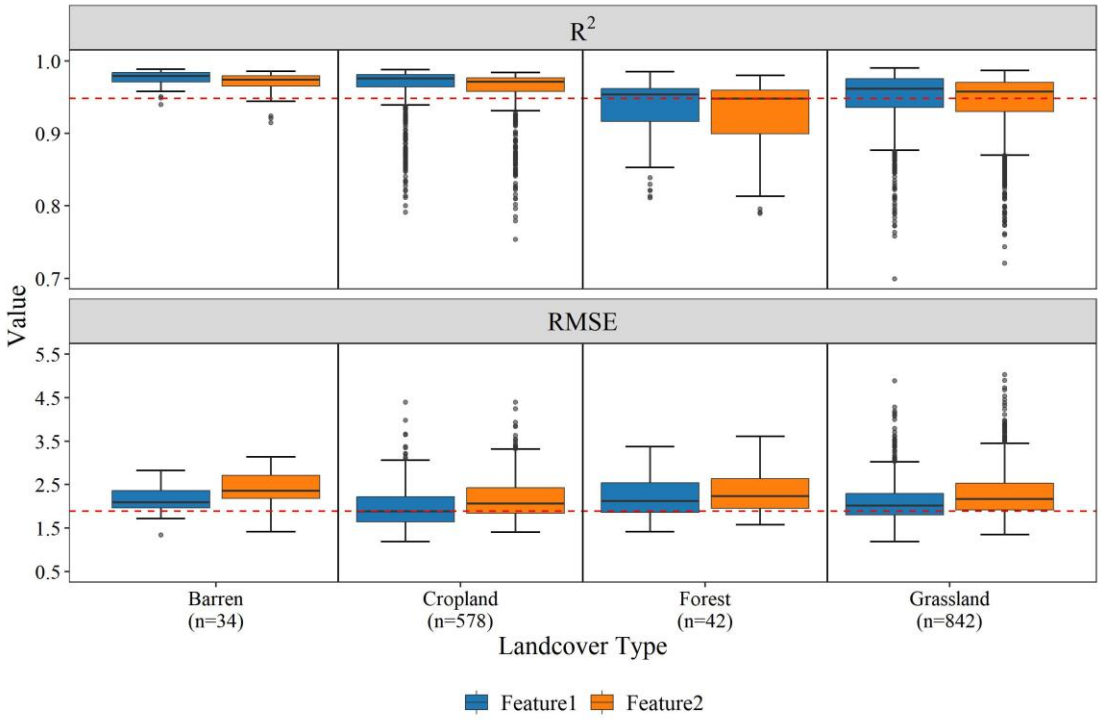

Figure S4. Differences in model accuracy across land cover types under different feature variable combinations. (Feature1 represents using both air temperature and LST together with other feature variables, while Feature 2 represents using only air temperature together with other feature variables)

Reviewer Comment 4:

Line 268: Change "fig. 3" to "fig. 2"

Response to Reviewer Comment 4:

Thank you for your correction. We have updated the figure citation accordingly in the revised manuscript.

Reviewer Comment 5:

Line 300: "Such differences are particularly important in complex ecosystems such as forests, where canopy structure and biological processes substantially influence thermal dynamics (Liu et al., 2025)." I'm not quite understanding this sentence and its connection to the context, can you explain a bit or rephrase?

Response to Reviewer Comment 5:

Thank you for pointing out the lack of clarity in this sentence. We agree that the original description did not clearly convey its intended meaning nor its connection to the context. The purpose of this sentence was to justify why both daily mean LST and air temperature were retained despite their high VIF values, by emphasizing their physical differences and complementary thermal information. To improve clarity, we have rewritten the sentence in the revised manuscript.

Revised Text (L285-L297):

Although the daily mean LST (LST_mean) and air temperature exhibit high collinearity (VIF > 10; Fig. S2), we chose to retain both variables because they represent different thermal information. LST_mean captures high-resolution surface radiative temperature signals, whereas air temperature reflects broader-scale atmospheric thermal conditions. In ecosystems with complex canopy structures, such as forests, the canopy can alter radiative transfer processes and cause LST to deviate from the true subsurface thermal environment(Liu et al., 2025). Therefore, the two variables provide complementary thermal information that helps better characterize soil thermal dynamics.In addition, we compared the model performance under different combinations of predictor variables (Fig. S3 and Fig. S4). The results show that the combination of air temperature + LST + other predictors achieved the best modeling accuracy at the surface soil layers. Therefore, retaining both air temperature and LST in the final model is reasonable and necessary.

Reference

Liu X., Li Z.-L., Duan S.-B., Leng P., Si M., 2025. Retrieval of global surface soil and vegetation temperatures based on multisource data fusion. Remote Sens. Environ. 318, 114564. https://doi.org/10.1016/j.rse.2024.114564

Reviewer Comment 6:

Line 440: Should MAE equation be the difference between prediction and observations divided by total sample number? Is "2" a typo?

Response to Reviewer Comment 6:

Thank you for pointing this out. You are correct — the denominator should represent the total number of samples (N). The “2” in the MAE equation was a typographical

error. The correct formulation of the Mean Absolute Error (MAE) is:

$$MAE = \frac{\sum_{i=1}^{N} | x_i - y_i |}{N} \quad (1.1)$$

where x_i and y_i denote the observed and predicted values, respectively. We have corrected this in the revised manuscript.

Reviewer Comment 7:
Fig. 6, 7: There are narrow regions with quite dark color over the eastern coast. Are these regions with low R-square and RMSE or just the boundary? Please clarify.
Response to Reviewer Comment 7:
Thank you for your observation. The narrow dark regions in Figures 6 and 7 along the eastern coast correspond to islands and boundary areas, rather than regions with low R^2 or high RMSE values. We apologize for any confusion this may have caused.

Reviewer Comment 8:
Line 579: Did you upscale your fine resolution results from XGBoost before comparing it to other data products?
Response to Reviewer Comment 8:
Thank you for your question. Before comparing with the ERA5_Land and GLDAS products, all three datasets were resampled to match the spatial resolution of ERA5_Land. This ensured consistency in spatial scale and allowed for a fair and accurate evaluation.

Reviewer Comment 9:
Section 3.1 and 3.2 both are evaluated at site level, so I suggest merging them.
Response to Reviewer Comment 9:
As you correctly pointed out, both Sections 3.1 and 3.2 focused on model evaluation at the site level. Accordingly, we have merged the original Sections 3.1 and 3.2 into a single new Section 3.1 based on your suggestion. In addition, Figure 7 from the former Section 3.2 has been moved to the Appendix.

Reviewer Comment 10:
section 3.3: Fine resolution product shall be able to reflect the response of Ts to elevation. I did not see any discussion on this point.
Response to Reviewer Comment 10:
Thank you for this valuable comment. In the revised manuscript, we have expanded our discussion on how the fine-resolution T_s product captures the response of T_s to elevation. In this section, we compare our product with other existing datasets and highlight the advantages of our 1-km T_s estimates in representing spatial variations across different topographic conditions. Please refer to Section 3.3 of the revised manuscript for the detailed discussion.

Revised Text (L470-L479):

Figure 9 presents a comparison of the T_s products at the 0 cm depth with the ERA5-Land and GLDAS 2.1 reanalysis datasets, including both national-scale patterns (Fig. 9a–c) and zoomed-in regional details (Fig. 9d–f). Compared with the two reanalysis products, our generated T_s dataset exhibits substantially finer spatial resolution, enabling a clearer representation of localized spatial heterogeneity. As illustrated in the zoomed-in panels of Figure 9, our T_s product accurately captures terrain- and elevation-driven temperature gradients in regions with strong topographic variability, such as the transition zone from the Sichuan Basin to the margins of the QTP. In contrast, the coarse spatial resolution of ERA5-Land and GLDAS 2.1 tends to smooth out these fine-scale topographic effects, resulting in a loss of spatial detail.

Reviewer Comment 11:

Fig. 12: Authors shall improve the figure quality. Lines are too thin and resolution seems to be low.

Response to Reviewer Comment 11:

Thank you for your helpful comment regarding the figure quality. We have improved the visualization by increasing the line thickness and rendering resolution. The figure has been regenerated at higher quality and the updated version has been incorporated into the revised manuscript. The original Figure 12 has now been updated and is presented as Figure 11 in the revised manuscript.

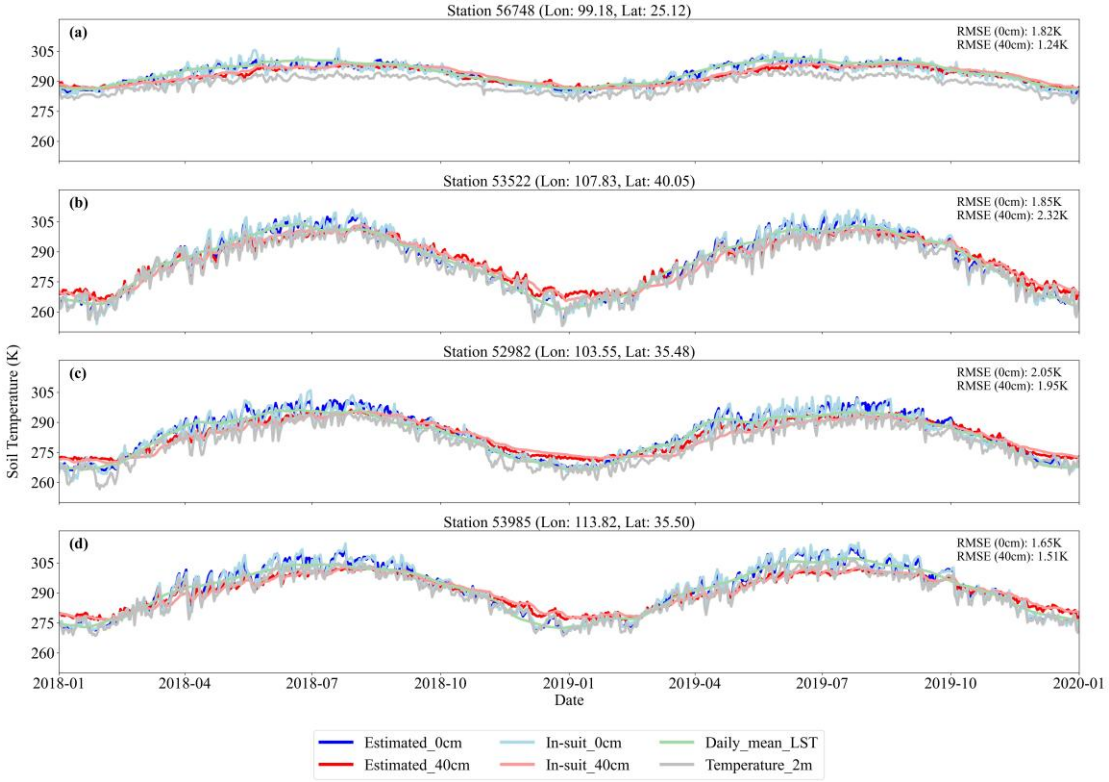

Figure 11. Time series of the Estimated_0cm, Estimated_40cm, Daily_mean_LST, and Temperature_2m at four sites from different regions between 2018-2019.

Reviewer Comment 12:

Line 822: Missing soil moisture can also bring substantial error in capturing daily or sub-daily variations of soil temperature. Authors shall discuss it as well.

Response to Reviewer Comment 12:

Thank you for this insightful comment. We agree that soil moisture plays an important role in regulating daily and sub-daily soil temperature dynamics, and the absence of soil moisture information can introduce additional uncertainty in T_s estimation. In the revised manuscript, we have added a dedicated discussion on this issue, highlighting how the lack of soil moisture data may affect the model's performance and the potential pathways for future improvements. The detailed discussion can be found in Section 4.3.

Revised Text (L674-L682):

Short-term changes in soil moisture alter fundamental soil thermal properties, including heat capacity, thermal conductivity, and thermal diffusivity, which in turn control heat transfer processes and sub-daily T_s dynamics. (Abu-Hamdeh, 2003; Subin et al., 2013). Consequently, the absence of soil moisture information may introduce additional uncertainty when modeling daily and sub-daily T_s dynamics, especially at deeper layers. Incorporating high-resolution soil moisture datasets in future work would improve the representation of soil hydrothermal interactions and further enhance T_s estimation accuracy.

Reference

Abu-Hamdeh, N.H., 2003. Thermal properties of soils as affected by density and water content. Biosyst. Eng. 86, 97–102.

Subin, Z.M., Koven, C.D., Riley, W.J., Torn, M.S., Lawrence, D.M., Swenson, S.C., 2013. Effects of soil moisture on the responses of soil temperatures to climate change in cold regions. J. Clim. 26, 3139–3158.

Reviewer Comment 13:

Line 826: "explained by the physical characteristics of soil temperature". Should this be "the physical characteristics of soil texture profile"?

Response to Reviewer Comment 13:

Thank you for this helpful clarification. We agree that the original expression was imprecise. The depth-dependent behavior of T_s is indeed more closely related to the physical characteristics of the soil profile, including soil texture, bulk density, and thermal properties, rather than to the "characteristics of soil temperature" itself. We have revised the manuscript accordingly to replace the phrase with "the physical characteristics of the soil profile." This improves both accuracy and clarity.

Revised Text (L663-L664):

This depth-dependent pattern can be explained by the physical characteristics of the soil profile.

Reviewer Comment 14:

Line 830: "dampen high-frequency fluctuations and stabilize the relationship between predictors and Ts". This conclusion is a bit misleading. In relatively deeper soil depth, "dampen high-frequency fluctuations" is true, but this does not stabilize the predictor -

Ts relationship, but reduces the signal of the predictor. Following these reasons, "thereby improving performance at these depths" is not true. The lower RMSE/R2 is a consequence of low variability of deeper soil temperature itself, not reflecting a better performance for deeper soil temperature. Please revise.

Response to Reviewer Comment 14:

We appreciate this insightful comment. We agree that the original wording may have been misleading. The reduced RMSE and increased R^2 at intermediate depths do not necessarily imply that the predictor–T_s relationships are stronger at those depths. Instead, the improved metrics primarily reflect the dampened variability of T_s caused by thermal buffering and increased heat capacity in mid-soil layers. In the revised manuscript, we have rewritten the sentence to clarify that the performance metrics at middle depths are largely a consequence of reduced temporal and spatial fluctuations in T_s, rather than inherently better model performance or stronger predictor–response relationships.

Revised Text (L667-L670):

In contrast, intermediate soil layers buffer high-frequency temperature fluctuations through thermal diffusion and higher heat capacity. As a result, T_s becomes more stable with lower natural variability at these depths, leading to lower RMSE and higher R^2 values.

Reviewer Comment 15:

Line 832: "At greater depths, however, surface-level errors propagate downward through the cascading framework, resulting in reduced accuracy—particularly during summer and winter." Can this be a consequence from uncertain soil texture profile input? As soil goes deeper, the uncertainty from soil texture will accumulate and becomes higher,

Response to Reviewer Comment 15:

Thank you for this valuable suggestion. We agree that uncertainty in soil texture inputs can accumulate with depth and may contribute to the reduced accuracy observed at deeper soil layers. Although our modeling framework incorporates multi-layer soil texture information, uncertainties in deep soil texture may propagate through the cascading prediction structure. We have added this point to Section 4.3 of the Discussion, acknowledging that soil texture uncertainty is an additional factor influencing deep-layer T_s errors.

Revised Text (L671-L674):

At deeper layers, prediction accuracy decreases because surface-level errors propagate downward through the hierarchical modeling framework, and uncertainties in soil texture inputs gradually accumulate with depth; during periods such as summer and winter, these combined uncertainties may be further amplified.

Reviewer Comment 16:

Line 836: Another reason is that the LST product reflects the temperature that the remote sensor observed. So it can be either soil surface temperature, snow temperature

or temperature at canopy top. Please address how this uncertainty source can impact your results.

Response to Reviewer Comment 16:

We greatly appreciate this important comment. We agree that the LST product represents the radiometric temperature observed by the satellite sensor, and the specific temperature it reflects may vary depending on land cover and seasonal conditions. As a result, LST may correspond to surface T_s, snow surface temperature, or canopy-top temperature. This inherent ambiguity indeed introduces an additional source of uncertainty in the estimation of T_s. We have added a discussion of this issue in Section 4.3 of the revised manuscript.

Revised Text (L683-L703):

Seasonal variations and differences in land cover also contribute to the spatiotemporal differences in model performance. As shown in Figures 7 and 8, the model performs better in spring and autumn, whereas its accuracy declines in summer and winter. In summer, vigorous vegetation growth and canopy closure alter surface–atmosphere energy exchange processes and weaken the relationship between canopy temperature and subsurface T_s, thereby reducing the effectiveness of LST as a proxy for near-surface T_s (Kropp et al., 2020; Cui et al., 2022). Moreover, because satellite sensors measure radiometric temperature, LST in densely vegetated regions often represents canopy-top temperature rather than the surface T_s, introducing an additional source of uncertainty. In winter, snow cover further increases complexity: the high albedo of snow reduces net radiation (Loranty et al., 2014; Li et al., 2018), and its insulating effect weakens the soil's response to cold-air fluctuations (Zhang, 2005; Myers-Smith et al., 2015). Meanwhile, Meanwhile, freezing of soil water alters soil thermal conductivity and heat capacity, and frequent freeze–thaw cycles introduce nonlinear dynamics into T_s, increasing modeling uncertainty (Li et al., 2023a; Imanian et al., 2024). Although our multi-source adaptive modeling framework demonstrates robust performance across varying depths and environmental conditions, it does not explicitly represent the physical mechanisms governing vertical heat transfer. Future research could incorporate deep learning models capable of learning complex spatiotemporal dependencies to enhance the physical interpretability of T_s variations across time, space, and depth.

Reference

Cui, X., Xu, G., He, X., Luo, D., 2022. Influences of seasonal soil moisture and temperature on vegetation phenology in the Qilian Mountains. Remote Sens. 14, 3645. https://doi.org/10.3390/rs14153645

Imanian, H., Mohammadian, A., Farhangmehr, V., Payeur, P., Goodarzi, D., Hiedra Cobo, J., Shirkhani, H., 2024. A comparative analysis of deep learning models for soil temperature prediction in cold climates. Theor. Appl. Climatol. 155, 2571–2587. https://doi.org/10.1007/s00704-023-04781-x

Kropp, H., Loranty, M.M., Natali, S.M., Kholodov, A.L., Rocha, A.V., Myers-Smith, I., Abbot, B.W., Abermann, J., Blanc-Betes, E., Blok, D., Blume-Werry, G., Boike,

J., Breen, A.L., Cahoon, S.M.P., Christiansen, C.T., Douglas, T.A., Epstein, H.E., Frost, G.V., Goeckede, M., Høye, T.T., Mamet, S.D., O'Donnell, J.A., Olefeldt, D., Phoenix, G.K., Salmon, V.G., Sannel, A.B.K., Smith, S.L., Sonnentag, O., Vaughn, L.S., Williams, M., Elberling, B., Gough, L., Hjort, J., Lafleur, P.M., Euskirchen, E.S., Heijmans, M.M., Humphreys, E.R., Iwata, H., Jones, B.M., Jorgenson, M.T., Grünberg, I., Kim, Y., Laundre, J., Mauritz, M., Michelsen, A., Schaepman-Strub, G., Tape, K.D., Ueyama, M., Lee, B.-Y., Langley, K., Lund, M., 2020. Shallow soils are warmer under trees and tall shrubs across arctic and boreal ecosystems. Environ. Res. Lett. 16, 015001. https://doi.org/10.1088/1748-9326/abc994

Li, Q., Ma, M., Wu, X., Yang, H., 2018. Snow cover and vegetation-induced decrease in global albedo from 2002 to 2016. J. Geophys. Res. Atmospheres 123, 124–138. https://doi.org/10.1002/2017JD027010

Li, X., Zhu, Y., Li, Q., Zhao, H., Zhu, J., Zhang, C., 2023. Interpretable spatio-temporal modeling for soil temperature prediction. Front. For. Glob. Change 6, 1295731. https://doi.org/10.3389/ffgc.2023.1295731

Loranty, M.M., Berner, L.T., Goetz, S.J., Jin, Y., Randerson, J.T., 2014. Vegetation controls on northern high latitude snow-albedo feedback: Observations and CMIP 5 model simulations. Glob. Change Biol. 20, 594–606. https://doi.org/10.1111/gcb.12391

Myers-Smith, I.H., Elmendorf, S.C., Beck, P.S.A., Wilmking, M., Hallinger, M., Blok, D., Tape, K.D., Rayback, S.A., Macias-Fauria, M., Forbes, B.C., Speed, J.D.M., Boulanger-Lapointe, N., Rixen, C., Lévesque, E., Schmidt, N.M., Baittinger, C., Trant, A.J., Hermanutz, L., Collier, L.S., Dawes, M.A., Lantz, T.C., Weijers, S., Jørgensen, R.H., Buchwal, A., Buras, A., Naito, A.T., Ravolainen, V., Schaepman-Strub, G., Wheeler, J.A., Wipf, S., Guay, K.C., Hik, D.S., Vellend, M., 2015. Climate sensitivity of shrub growth across the tundra biome. Nat. Clim. Change 5, 887–891. https://doi.org/10.1038/NCLIMATE2697

Zhang, T., 2005. Influence of the seasonal snow cover on the ground thermal regime: An overview. Rev. Geophys. 43. https://doi.org/10.1029/2004RG000157

Response to Reviewer5_Comments

Reviewer Comment 1:

The Introduction would benefit from incorporating recent advances in spatial adaptive modeling using quadtree recursive retrieval, with explicit discussion of the comparative advantages of the proposed methodology.

Response to Reviewer Comment 1:

Thank you very much for this constructive suggestion. We agree that a more systematic discussion of recent advances in spatial adaptive modeling based on quadtree recursive partitioning would strengthen the Introduction. In the revised manuscript, we have expanded the Introduction to incorporate relevant developments in quadtree-based spatial modeling and have further clarified the advantages of our proposed rotated-quadtree framework.

Revised Text (L105-L135):

Recent advances in spatially adaptive modeling have increasingly emphasized the importance of addressing spatial heterogeneity and uneven sampling density in environmental datasets. Classical quadtree structures and related hierarchical spatial data models provide the theoretical foundation for constructing adaptive, variable-sized spatial partitions, enabling efficient organization of multiscale spatial information through recursive subdivision (Samet, 1984). Building on this foundation, Lagonigro et al., (2020) developed the AQuadtree R package, which provides an adaptive spatial partitioning framework capable of generating variable-sized grid cells according to the spatial distribution of observations. This adaptive partitioning produces finer grids in data-dense regions and coarser grids where observations are sparse, ensuring a spatial structure that better reflects sampling heterogeneity and improves the model's capacity to capture localized spatial variability. Extending this idea, we develop a rotated-quadtree strategy that applies multiple orientation angles during the quadtree subdivision process. This enhancement allows the model to capture spatial heterogeneity from multiple directional perspectives, and averaging predictions across rotation angles substantially reduces the boundary artifacts that may arise from single-angle grid partitioning, ultimately improving the robustness of local modeling under complex environmental gradients.

To address the irregular station distribution, and non-stationarity commonly encountered in large-scale T_s estimation, we construct a spatially adaptive modeling framework based on the rotated quadtree approach. Within each grid cell, multi-source environmental predictors are integrated with in situ station records, and T_s is estimated using XGBoost models. Based on this framework, the objectives of this study are to: (1) construct a spatially adaptive modeling system; (2) generate a multi-layer T_s dataset at a daily time-step and one kilometer resolution in China from 2010-2020; and (3) evaluate the dataset through independent validation with flux tower observations and benchmarking against widely used T_s products. The proposed methodology could directly address the scaling challenges induced by spatial heterogeneity and uneven

data distribution. The generated products would provide a robust foundation for high-resolution environmental modeling, precision agriculture and climate impact assessments.

Reference

Lagonigro, R., Oller, R., Martori, J.C., 2020. AQuadtree: An R package for quadtree anonymization of point data.

Samet, H., 1984. The quadtree and related hierarchical data structures. ACM Comput. Surv. CSUR 16, 187–260.

Reviewer Comment 2:

Additional clarification regarding auxiliary variable selection criteria is warranted. Beyond literature review, principles such as correlation analysis should be elaborated. It is recommended to first validate the existence of spatially non-stationary relationships using bivariate local Moran's I, then proceed with spatial partitioning and predictive modeling via quadtree recursive retrieval.

Response to Reviewer Comment 2:

Thank you very much for your valuable suggestion. We agree that it is necessary to further clarify the criteria for selecting auxiliary variables and the evaluation of spatial non-stationarity. In the revised manuscript, we focus on the local form of spatial association analysis and provide a more detailed explanation of the analytical procedures used in this study.

Here are the revisions, supplemented in the Appendix (L23-L50):

To examine whether the relationships between T_s (GST_Avg) and the auxiliary variables exhibit spatial non-stationarity, we employed the Local Bivariate Moran's I, a local statistic within the Local Indicators of Spatial Association (LISA) framework. This method allows us to reveal localized spatial associations and spatially varying relationships between the target variable (X) and the spatially lagged auxiliary variable (Wy). First, we constructed a spatial weights matrix using the K-nearest neighbors method (K = 8). This configuration is suitable for the irregular spatial distribution of meteorological stations across China and ensures that each station has a comparable number of spatial neighbors.

Based on this spatial weights structure, we calculated the Local Bivariate Moran's I between GST_Avg (X) and elevation (Y), and obtained permutation-based p-values.We then computed the spatially lagged auxiliary variable (Wy) and classified each station into one of four significant LISA cluster types ($p < 0.05$): High–High (red), High–Low (green), Low–High (purple), and Low–Low (blue). Stations with non-significant local associations ($p \geq 0.05$) are shown in gray. As illustrated in Figure S5, approximately 64% of the stations exhibit statistically significant local spatial associations, and all four cluster types occur across different regions of China. These spatially heterogeneous local association patterns clearly indicate pronounced spatial non-stationarity in the T_s–elevation relationship.

These findings further demonstrate the necessity of adopting a spatially adaptive modeling framework. Accordingly, the rotated quadtree model developed in this study is well justified, as it can effectively capture localized variations in predictor–response relationships.

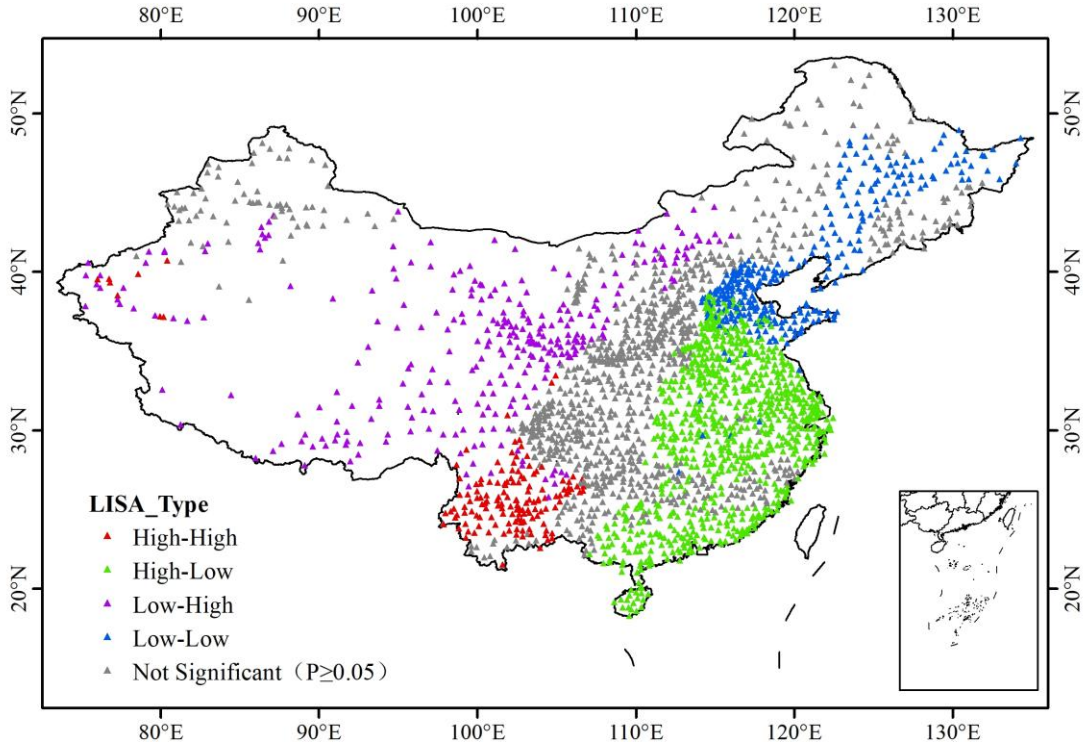

Figure S5. Spatial patterns of the bivariate Local Moran's I between GST_Avg and elevation at meteorological stations across China.

Revised Text (L300-L305):

We applied the Local Bivariate Moran's I analysis to assess the local spatial relationship between surface T_s (GST_Avg) and elevation as an illustrative example (Fig. S5). The results reveal significant spatial variations in their local association ($p < 0.05$), indicating pronounced spatial non-stationarity in the T_s–elevation relationship. These findings justify the need for a spatially adaptive modeling strategy capable of capturing localized heterogeneity.

Reviewer Comment 3:

Methodological details concerning quadtree rotation require elaboration: Can this approach achieve complete coverage of the study area? Is 360-degree rotation necessary to cover all prediction grids followed by averaging for final predictions? Supplementary materials illustrating the detailed procedures of the proposed model would be valuable.

Response to Reviewer Comment 3:

First, the grid cells generated by a single quadtree subdivision may not fully cover the entire study area and may omit stations located near the domain boundaries. To address this limitation, we employ six rotated quadtree configurations at different orientation angles, which collectively ensure complete spatial coverage and prevent potential loss of edge-area observations caused by a single subdivision. Second, a full 360° rotation

is unnecessary. We selected six representative angles—0°, 15°, 30°, 45°, 60°, and 75°—which sufficiently cover different directional alignments; additional angles would only introduce redundancy without improving performance. Third, we average the predictions obtained from the six rotation angles, which allows the model to capture spatial heterogeneity from multiple directional perspectives while effectively mitigating boundary artifacts induced by any single quadtree partition. This ensemble approach markedly enhances the stability and robustness of the final soil temperature estimates. Finally, following your recommendation, we have added detailed workflow diagrams and supplementary materials that illustrate the complete rotated-quadtree modeling framework, including grid rotation, spatial subdivision, model training, and prediction integration.

Revised Text (L306-L336):

A quadtree is a hierarchical spatial data structure that recursively subdivides a two-dimensional space into four quadrants, enabling efficient spatial indexing and localized data organization. In this study, we adopted a bottom-up, rotated quadtree–based spatial partitioning strategy that adaptively generates finer grids in regions with dense observations and coarser grids in sparsely sampled areas. Compared with global modeling or static grid partitioning, this adaptive approach improves regional modeling fidelity while maintaining computational efficiency. The procedure consists of the following steps:

(1) Initialization of Minimum Units

The entire study area was first divided into uniform minimum-sized units (leaf nodes), each representing a basic spatial element that may contain zero or more soil temperature observations. This initialization provides the base spatial resolution for subsequent hierarchical construction. An illustration of the quadtree structure and spatial indexing principles is provided in Fig. S2.

(2) Bottom-up Hierarchical Merging

Starting from the leaf nodes, groups of four adjacent quadrants were recursively merged into parent nodes if each contained fewer than 30 observation sites (threshold selection detailed in Fig. S4). The merging process continued upward until no further groups met the threshold. This approach ensures that each node has sufficient sample size while achieving spatially adaptive partitioning across the study area. Each subregion is then assigned a localized T_s prediction model.

(3) Rotation at Multiple Angles

To reduce potential edge effects introduced by static grid boundaries, we implemented a rotated quadtree partitioning strategy. The quadtree structure was rotated at six angles (0°, 15°, 30°, 45°, 60°, and 75°), producing distinct sets of spatial partitions for each orientation (Fig. 3). Independent models were trained for each rotated configuration, and the final T_s estimates were obtained by averaging the outputs from all six models. This rotation-based ensemble method improves spatial smoothness and minimizes discontinuities at partition boundaries.

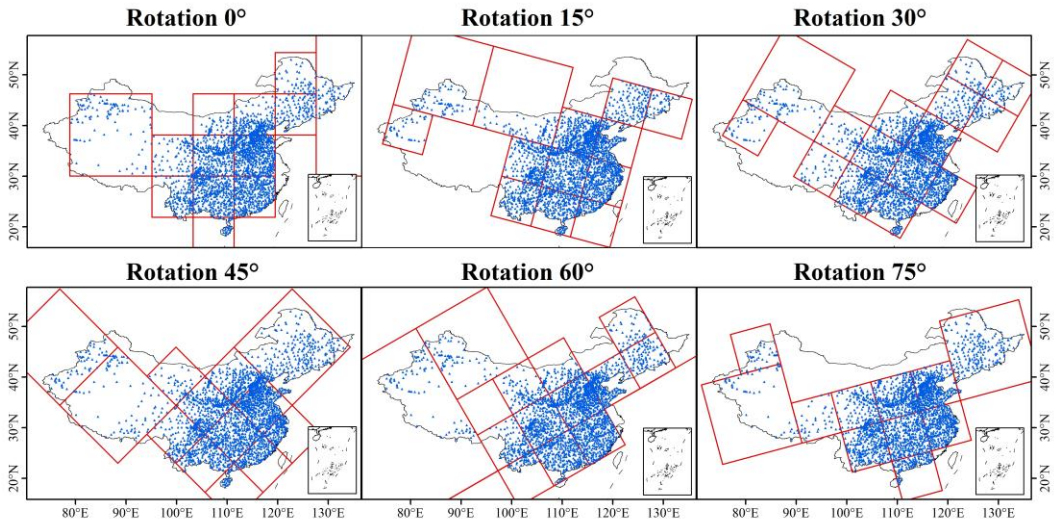

Figure 3. Multi-angle adaptive quadtree partitioning of site observations (0°, 15°, 30°, 45°, 60°, 75°)

Reviewer Comment 4:

It is recommended to add a section of pseudocode to illustrate the computational process of spatial adaptive partition method.

Response to Reviewer Comment 4:

Thank you very much for your suggestion regarding improving the transparency and reproducibility of our method. In response, we have made the complete R implementation of the rotated-quadtree spatial adaptive partitioning algorithm publicly available on GitHub. The repository includes all scripts used to construct the six rotated quadtree partitions, generate the spatial blocks, and export the polygon shapefiles. The code is openly accessible at: https://github.com/wangxt1314/Rotated-quadtree

This repository also provides detailed documentation and example files, enabling users to fully reproduce the quadtree construction and subsequent modeling workflow. We believe that making the full code publicly accessible will substantially enhance the reproducibility and transparency of our study.

Revised Text (L730-L732):

7. Code availability

The R scripts used to implement the rotated-quadtree spatial adaptive partitioning are publicly available at: https://github.com/wangxt1314/Rotated-quadtree

Reviewer Comment 5:

Given the simultaneous inclusion of elevation and slope (often derived from elevation) as auxiliary variables, potential multi-collinearity concerns should be addressed. Please discuss whether this correlation might affect model predictive performance.

Response to Reviewer Comment 5:

Thank you very much for raising this important point. As slope is derived from the digital elevation model (DEM), it is indeed correlated with elevation. To assess whether

this relationship may introduce multicollinearity issues in our modeling framework, we conducted a Variance Inflation Factor (VIF) analysis for all auxiliary variables. The results are presented in Figure S2. Variance Inflation Factor (VIF) of predictor variables in the Supplementary Materials. The VIF values for both elevation and slope are well below commonly accepted thresholds (VIF < 10), indicating that their correlation is not strong enough to compromise model stability. Furthermore, our modeling framework is based on a tree-based algorithm (XGBoost), which learns through recursive partitioning driven by information gain. Such models are inherently robust to correlations among predictor variables and are far less susceptible to multicollinearity issues than linear regression models, where parameter estimates can become unstable under collinearity.

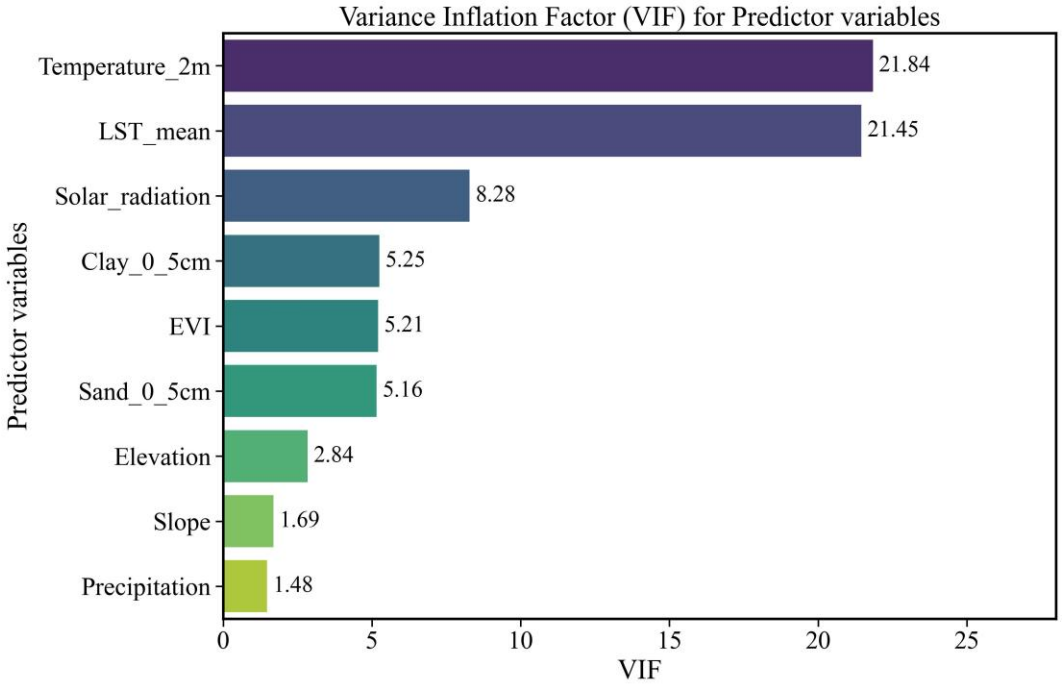

Figure S2. Variance Inflation Factor (VIF) of predictor variables

Reviewer Comment 6:

As model predictions and uncertainty typically coexist, and given the high accuracy demonstrated in validation results, provision of corresponding prediction uncertainty estimates would strengthen the methodological rigor.

Response to Reviewer Comment 6:

Thank you very much for your valuable comment. We agree that incorporating a prediction uncertainty assessment further strengthens the scientific rigor of our methodology. In response to your suggestion, we have added an uncertainty analysis based on an ensemble of six quadtree models constructed under different rotation angles.

Specifically, each observation station is predicted six times using quadtree partitioning structures generated at six different rotation angles. We quantify prediction uncertainty as the standard deviation of these six predicted T_s values, which reflects the stability

and sensitivity of the model predictions to changes in spatial partitioning orientation. A larger standard deviation indicates substantial divergence among predictions from different rotations, suggesting higher structural uncertainty caused by spatial heterogeneity, partition boundary effects, or variation in local sample density. Conversely, when the six predictions remain highly consistent, the standard deviation is small, indicating that the model is stable across different partitioning orientations and exhibits lower uncertainty.

To derive a more robust annual uncertainty estimate, we further average the daily uncertainty values for each station within each year, resulting in a station-level annual uncertainty index (Fig.S14). This uncertainty metric serves as a useful complement to the prediction results by identifying areas where the model exhibits stronger structural variability or weaker observational support, thereby enhancing the transparency and credibility of the modeling results.

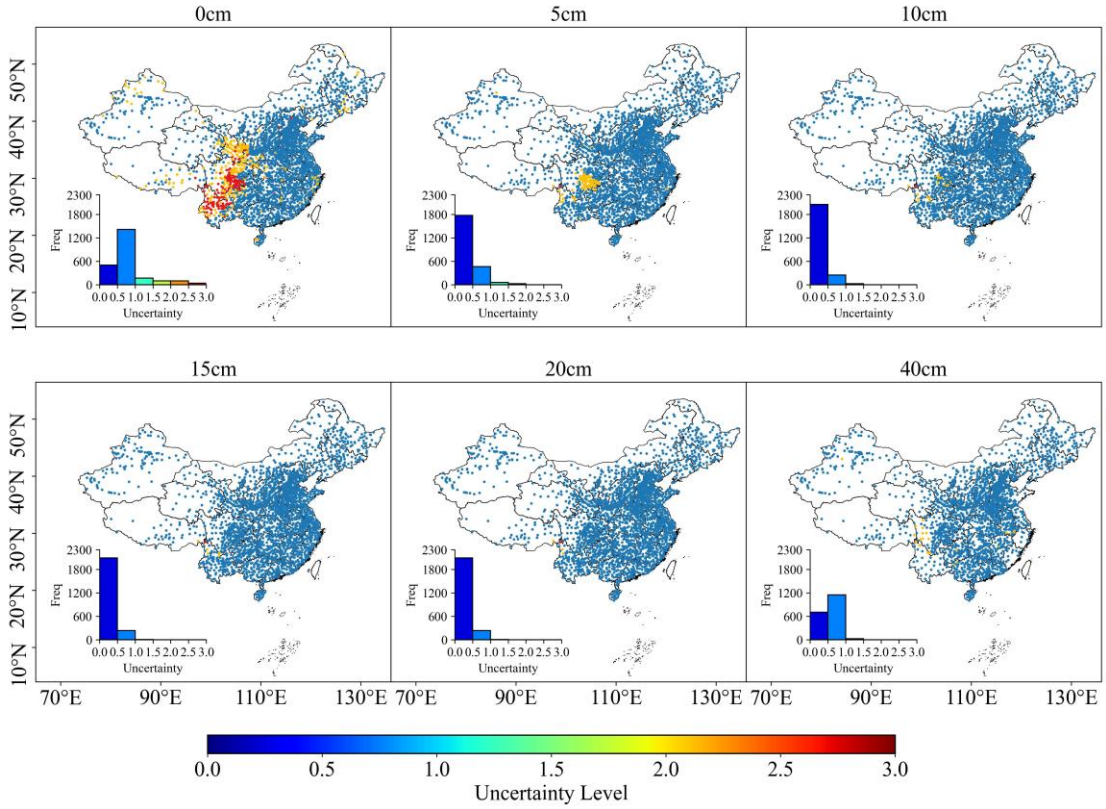

Figure S14. Spatial patterns of prediction uncertainty at six soil depths based on the rotated-quadtree ensemble.

The station-based uncertainty analysis shows that at the 0 cm depth, a substantially larger proportion of stations exhibit high uncertainty compared with other depths. The stations with higher uncertainty are mainly concentrated in the Sichuan Basin, the Yunnan–Guizhou Plateau, and the Qinghai–Tibet Plateau, which are characterized by complex geological and geomorphological environments. In contrast, the overall uncertainty levels at the remaining depths are considerably lower and spatially more stable. We believe that incorporating this improvement will further strengthen the

methodological rigor, enhance the reliability of the results, and provide valuable guidance for future users of the dataset.