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.

Reviewer Comment 1:

In the introduction (lines 92–104), the authors clearly outline two key challenges in current research: first, the significant heterogeneity of T_s leads to unclear relationships between variables; second, modeling is hindered by data scarcity and uneven distribution. However, in lines 106–113, when introducing the objectives and scope of this study, the authors do not explain how the study addresses these two challenges. It is also unclear what specific methods are used to overcome them, and why these methods are effective. It is recommended that the authors restructure this section by focusing on the core problems, rather than simply listing the research contents. This would improve the clarity and logical flow of the introduction.

Response to Reviewer Comment 1:

We greatly appreciate your insightful comments and constructive feedback. We agree that the introduction should more explicitly link the identified challenges with the study's objectives and methodology. In response, we have revised and reorganized the relevant section to clarify how our approach directly addresses the two key challenges currently facing T_s prediction.

Revised Text (L105-L121):

To address the above challenges, this study proposes a spatially adaptive methodology based on quadtrees. This approach dynamically partitions the study area into grids of varying sizes, with smaller grids in densely observed regions and larger grids in sparsely sampled areas, thereby enabling localized modeling that better captures spatial heterogeneity across complex environmental gradients. In addition, multi-source environmental predictors are integrated, and XGBoost models are applied within each grid cell to capture the nonlinear relationships between T_s and its driving factors. Importantly, we employ a spatial block cross-validation strategy to evaluate the model's generalization ability in unseen regions. 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.

Reviewer Comment 2:

In Section 2.1, the authors describe the use of CMA T_s observational data. However, it is unclear how these data were processed. Were the observations directly provided as daily averages, or were they aggregated from hourly data? Was any quality control applied? How were missing data handled, both in the vertical profile and in the time series? Were any filtering or screening steps performed, and if so, what were the specific criteria?

Response to Reviewer Comment 2:

We appreciate the reviewer's thoughtful comment. The multi-layer T_s data were obtained from the national CMA weather station network, where measurements were automatically recorded every 10 minutes and used to compute daily means at each depth. Data preprocessing steps are described in Section 2.1.

Revised Text (L125-L133):

In this study, in-situ T_s *observations* was measured at six depths: at the surface (0 m), and at subsurface levels of 0.05, 0.10, 0.15, 0.20, and 0.40 meters. Data were collected through the national weather station network operated by the China Meteorological Administration (CMA), in accordance with standardized measurement protocols. At each site, T_s was recorded every 10 minutes and automatically uploaded to a central server. Daily mean values at each depth were calculated from these high-frequency records. We then assessed data completeness for the period 2010–2020 and excluded stations with more than 20% missing daily records at any depth. After quality control, 2,093 stations were retained for model development.

Reviewer Comment 3:

In lines 186–190, as well as in Section 4.3, the authors provide a brief discussion of the study's limitations. However, it is concerning that the missing land surface temperature (LST) data caused by cloud cover were filled using a simple linear interpolation method. This approach may be questionable, as the interpolated values represent a theoretical cloud-free state, while cloud presence can significantly influence radiative transfer and thus impact LST. There are existing interpolation methods that take into account energy transfer and energy balance. It is recommended that the authors investigate these alternatives and consider adopting a more reliable method.

Response to Reviewer Comment 3:

We appreciate your comments on the interpolation method used to address LST gaps resulting from cloud contamination. Indeed, cloud cover presents a major challenge in remote sensing–based LST reconstruction, as it significantly alters surface radiative fluxes and interferes with the physical basis of thermal observations. As the reviewer correctly noted, linear interpolation does not explicitly account for the thermal effects of clouds and may produce overly idealized estimates under cloud-free assumptions.

In this study, we employed a spatiotemporal linear interpolation method primarily due to its computational efficiency, simplicity, and suitability for large-scale reconstruction of missing data. To further reduce short-term fluctuations and noise introduced during interpolation, we applied a Savitzky–Golay filter during the preprocessing stage to smooth the time series (Kong et al., 2019; Chen et al., 2021). Notably, this method can be readily implemented on the Google Earth Engine (GEE) platform, enabling efficient global processing of MODIS LST products and the rapid generation of daily gap-free land surface temperature composites. This facilitates scalable model training and T_s estimation.

Nevertheless, we fully acknowledge the limitations of this method in cases of prolonged cloud cover. We concur that incorporating physically based interpolation methods could enhance the reliability of the reconstructed data. In future work, we plan to explore energy balance–based reconstruction techniques, such as incorporating surface energy balance system models and diurnal temperature cycle models (Hong et al., 2022; Firozjaei et al., 2024; Wang et al., 2024).

Moving forward, we aim to explore hybrid approaches that combine physically based models with machine learning algorithms to better capture the effects of cloud cover, land surface heterogeneity, and seasonal variability on T_s reconstruction. Additionally, we intend to incorporate passive microwave–based land surface temperature products, which are less affected by cloud contamination, as supplementary information for gap-filling. We believe these advancements will help reduce uncertainties in LST reconstruction and further enhance the accuracy and robustness of the resulting T_s dataset.

Reference

Chen, Y., Cao, R., Chen, J., Liu, L., and Matsushita, B.: A practical approach to reconstruct high-quality Landsat NDVI time-series data by gap filling and the Savitzky–Golay filter, ISPRS J. Photogramm. Remote Sens., 180, 174–190, https://doi.org/10.1016/j.isprsjprs.2021.08.015, 2021.

Firozjaei, M. K., Mijani, N., Kiavarz, M., Duan, S.-B., Atkinson, P. M., and Alavipanah, S. K.: A novel surface energy balance-based approach to land surface temperature downscaling, Remote Sens. Environ., 305, 114087, https://doi.org/10.1016/j.rse.2024.114087, 2024.

Hong, F., Zhan, W., Göttsche, F.-M., Liu, Z., Dong, P., Fu, H., Huang, F., and Zhang, X.: A global dataset of spatiotemporally seamless daily mean land surface temperatures: Generation, validation, and analysis, Earth Syst. Sci. Data, 14, 3091–3113, https://doi.org/10.5194/essd-14-3091-2022, 2022.

Kong, D., Zhang, Y., Gu, X., and Wang, D.: A robust method for reconstructing global MODIS EVI time series on the Google Earth Engine, ISPRS J. Photogramm. Remote Sens., 155, 13–24, https://doi.org/10.1016/j.isprsjprs.2019.06.014, 2019.

Wang, Q., Tang, Y., Tong, X., and Atkinson, P. M.: Filling gaps in cloudy landsat LST product by spatial-temporal fusion of multi-scale data, Remote Sens. Environ., 306, 114142, https://doi.org/10.1016/j.rse.2024.114142, 2024.

Reviewer Comment 4:

In Section 2.3.1, it is suggested to provide further explanation of the Variance Inflation Factor (VIF). Specifically, what is its purpose, how is it calculated, and if possible, a formula should be included to make the description more complete.

Response to Reviewer Comment 4:

We appreciate the reviewer's valuable suggestion. In the revised manuscript, we have added a detailed explanation of the purpose, calculation, and interpretation of the Variance Inflation Factor (VIF) in Section 2.3.1. The updated text now includes the VIF formula and clarifies its role in diagnosing multicollinearity among predictors.

Here are the revisions (L244-L254):

Multicollinearity among multiple source variables may affect the robustness of the models. Therefore, we rigorously evaluated the multicollinearity among the independent variables using the variance inflation factor (VIF) before modeling to remove highly correlated variables. The VIF is a diagnostic statistic used to quantify the degree of multicollinearity by measuring how much the variance of a regression coefficient is inflated due to correlations with other predictors (Akinwande et al., 2015). It is calculated as:

$$VIF_i = \frac{1}{1-R_i^2} \tag{1.1}$$

where R_i^2 is the coefficient of determination obtained by regressing the i-th predictor against all other predictors. Variables with VIF exceeding 10 are generally considered severely multicollinear and should be removed.

Reference

Akinwande, M. O., Dikko, H. G., and Samson, A.: Variance inflation factor: As a condition for the inclusion of suppressor variable(s) in regression analysis, Open J. Stat., 5, 754–767, https://doi.org/10.4236/ojs.2015.57075, 2015.

Reviewer Comment 5:

In Section 2.3.2, a substantial portion is devoted to the spatial partitioning strategy based on a rotated quadtree. I have several questions regarding this part. First, why was the quadtree data structure chosen? The manuscript does not clearly explain this. Is it intended to address the issue of uneven distribution of observation sites? If so, why is the quadtree suitable for this purpose? Second, what was achieved by using the quadtree? Was there an effort to ensure that each node contains a roughly equal number of sites, for example around 30? Why was 30 selected as the threshold, and what is the basis for this value? Lastly, a minor suggestion (optional for consideration): if the goal is to achieve a more balanced spatial distribution of stations, a top-down data structure such as the K-D tree (with K = 2 in this study) may be more effective than the bottom-up quadtree. A K-D tree can ensure the difference in the number of points between leaf nodes does not exceed one, and can also support rotation operations.

Response to Reviewer Comment 5:

We thank the reviewer for the insightful and detailed questions. Below we provide point-by-point clarifications regarding:

(1) the rationale for choosing the rotated quadtree;

(2) the threshold of 30 observation sites; and

(3) a comparison with the suggested K-D tree approach.

1. Rationale for Choosing the Rotated Quadtree

As noted in the revised manuscript, our study faced a significant challenge of spatially uneven distribution of observation stations. The objective of using a quadtree-based partitioning strategy was not to ensure that each grid cell contains an equal number of samples, but rather to enable spatial adaptivity. The quadtree recursively subdivides space from the bottom up based on a point-count threshold, thereby generating finer grids in densely sampled regions and retaining coarser units in sparse areas. This design allows the model to accommodate spatial variability in data density, thereby improving both its adaptability and predictive accuracy.

Moreover, since local models are trained separately for each spatial unit, boundary effects between neighboring grids may arise due to discontinuities. To mitigate such effects, we implemented a rotated quadtree ensemble approach, in which multiple quadtree configurations are generated under different rotation angles. Averaging predictions across rotated quadtree configurations helps mitigate boundary-related artifacts and improves the spatial smoothness and robustness of the final outputs. This spatial ensemble strategy is visually illustrated in Figure S4. These methodological details and justifications have been incorporated into the revised manuscript in Section 2.3.2 (L269-L298) and further discussed in Section 4.1 (L516-L564).

2. Justification for Using a Threshold of 30 Sites

We sincerely thank the reviewer for the insightful comments regarding the design rationale of the quadtree-based partitioning strategy. To justify our choice of threshold = 30 as the final splitting criterion, we conducted a systematic evaluation of the partitioning performance under different thresholds using three key metrics. The supporting analysis and figures are included in the Appendix.

Here are the revisions, supplemented in the Appendix (L14-L44):

We conducted a systematic evaluation of the partitioning performance under different thresholds using three key metrics: the coefficient of variation (CV) of point count, the CV of point density, and the total number of grid cells. The CV of point count was used to evaluate the balance of sample distribution across spatial units under different thresholds. Point density was defined as the number of observation stations within a grid cell divided by its area. A lower CV of point density indicates that the partitioning effectively adjusted grid size according to local station density—i.e., producing smaller grids in dense regions and larger grids in sparse areas—thus reflecting a more adaptive spatial division. Conversely, a higher CV suggests that the partitioning failed to capture the spatial heterogeneity of station density. Therefore, the CV of point density serves as a key indicator of the spatial adaptivity of the quadtree partitioning.

The total number of grids corresponds to the number of local models to be trained, and thus indirectly reflects the computational and time cost associated with model training. As shown in Figure S4 (a–c), we systematically evaluated quadtree performance under a series of point-count thresholds (10, 30, 50, 70, 90): Figure S4a shows that the CV of point count drops rapidly with increasing threshold, indicating improved balance in sample allocation across grids. However, this trend levels off beyond threshold = 30, suggesting diminishing returns. Thus, threshold 30 marks an optimal trade-off. Figure S4b shows a notable inflection point in the CV of point density near threshold = 30. Although not the global minimum, this point represents an optimal trade-off where grid subdivision sufficiently reflects sample density variation without causing over- or under-segmentation—thereby capturing spatial adaptivity effectively. Figure S4c shows that the number of grid cells decreases rapidly as the threshold increases, leading to substantial computational savings. However, the rate of reduction slows considerably beyond threshold = 30, indicating limited additional benefit from further increases.

In summary, threshold = 30 achieves a favorable balance among sample distribution equity, spatial adaptivity, and computational efficiency, and was therefore selected as the final splitting threshold in this study. The detailed results of this threshold evaluation, including figures and metric comparisons, have been added to the revised manuscript as supplementary material (Appendix, Lines 9–41) to support transparency.

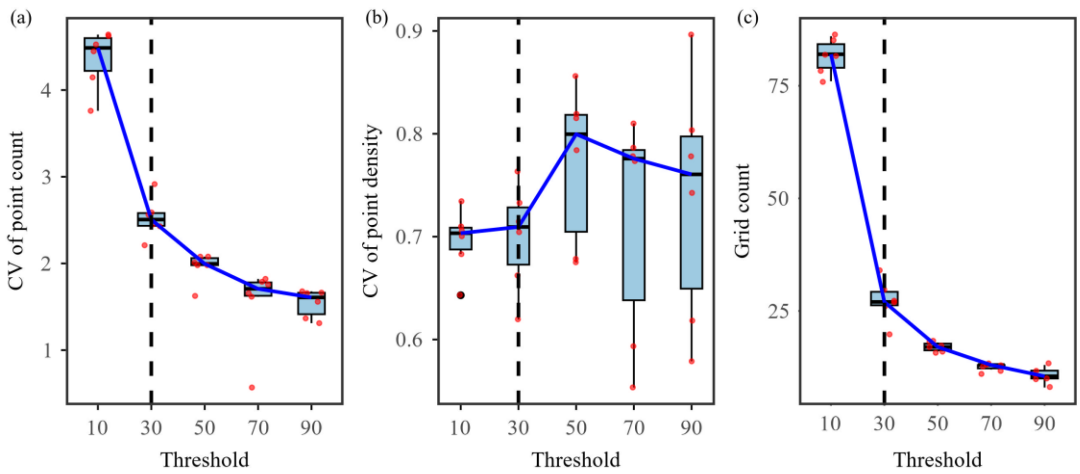

Figure S4. Performance evaluation of quadtree partitioning under different point-count thresholds. (a) Coefficient of variation (CV) of point count across spatial units. (b) CV of point density (point count per unit area). (c) Total number of generated grid cells. Dashed vertical line indicates the selected threshold of 30.

3. Comparison with the K-D Tree Approach

We appreciate the reviewer's thoughtful suggestion regarding the use of K-D trees for achieving a balanced spatial distribution of stations. We agree that K-D trees offer precise control over sample counts in each partition and can be advantageous when strict sample balance is the primary objective. However, the core objective of our study is not to enforce equal sample sizes in each spatial unit, but rather to enhance the adaptability and predictive performance of local modeling under spatially heterogeneous station distributions. To this end, we adopted a bottom-up quadtree-based strategy, which recursively subdivides space based on a point-count threshold. This enables the generation of finer grids in data-rich areas and larger cells in sparse regions, allowing the model structure to adapt to local data density and environmental variability. Compared to top-down methods like K-D trees, the quadtree is better suited for capturing spatial adaptivity than enforcing uniform sample counts. That said, we acknowledge the merits of K-D trees and agree that they represent a promising alternative for future work, particularly in applications where sample balance is more critical than spatial adaptivity.

Reviewer Comment 6:

In Section 2.3.3 (lines 285–295), the authors introduce XGBoost as the core machine learning algorithm used in the study. They present its advantages and compare it with other methods such as SVM, RF, and neural networks. However, the stated advantages are not sufficient to demonstrate that XGBoost is superior to the other listed methods. Machine learning models differ in structure, number of parameters, optimization strategy, and suitability for different tasks. Therefore, the current explanation is not enough to justify the model choice. Considering that the algorithm is not the main focus of this paper, it is suggested to either include a brief comparative experiment to support the claimed superiority or rephrase the section to emphasize the strengths of XGBoost without direct comparison to other models.

Response to Reviewer Comment 6:

We appreciate the reviewer's constructive comments regarding the justification of our model choice. As suggested, to clarify our reasoning, we have elaborated on the key considerations below.

Revised Text (L303-L317):

We adopted the XGBoost (Extreme Gradient Boosting) algorithm as the core regression model for T_s estimation due to its strong predictive performance, computational efficiency, and scalability across large environmental datasets. XGBoost builds an ensemble of regression trees in a stage-wise boosting process, where each tree is trained to minimize the residuals from the previous iteration, leading to a robust and optimized model (Chen and Guestrin, 2016). A key strength of XGBoost is its ability to handle heterogeneous and high-dimensional predictor sets, which are common in geoscience applications involving complex terrain, land cover variability, and climatic gradients. Recent studies have demonstrated its effectiveness in similar domains, including land surface temperature reconstruction (Li et al., 2024), multi-layer soil moisture estimation (Karthikeyan and Mishra, 2021), drought event attribution (Wang et al., 2025), and crop yield prediction (Li et al., 2023b). Given these proven strengths and the spatially nonstationary characteristics of T_s in our study area, XGBoost was selected to train localized prediction models within spatial subregions.

Reference

Chen, T. and Guestrin, C.: Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 785–794, https://doi.org/10.1145/2939672.2939785, 2016.

Karthikeyan, L. and Mishra, A. K.: Multi-layer high-resolution soil moisture estimation using machine learning over the United States, Remote Sens. Environ., 266, 112706, https://doi.org/10.1016/j.rse.2021.112706, 2021.

Li, B., Liang, S., Ma, H., Dong, G., Liu, X., He, T., and Zhang, Y.: Generation of global 1 km all-weather instantaneous and daily mean land surface temperatures from MODIS data, Earth Syst. Sci. Data, 16, 3795–3819, https://doi.org/10.5194/essd-16-3795-2024, 2024.

Li, Y., Zeng, H., Zhang, M., Wu, B., Zhao, Y., Yao, X., Cheng, T., Qin, X., and Wu, F.: A county-level soybean yield prediction framework coupled with XGBoost and multidimensional feature engineering, Int. J. Appl. Earth Obs. Geoinformation, 118, 103269, https://doi.org/10.1016/j.jag.2023.103269, 2023.

Wang, M., Wang, Y., Liu, X., Hou, W., Wang, J., Li, S., Zhao, L., and Hu, Z.: Vapor pressure deficit dominates vegetation productivity during compound drought and heatwave events in China's arid and semi-arid regions: Evidence from multiple vegetation parameters, Ecol. Inform., 88, 103144, https://doi.org/10.1016/j.ecoinf.2025.103144, 2025.

Reviewer Comment 7:

In lines 296–297, the validation set is twice the size of the test set. Is this split reasonable, and can it effectively evaluate the generalization performance of the model? Why not adopt more common ratios such as 8:1:1 or 6:2:2? In addition, the manuscript later mentions that five-fold cross-validation was used for evaluation. In this context, what are the roles of the two validation sets? Are they used for model selection, parameter tuning, or testing? It is recommended that the authors provide a clearer explanation. It is also suggested to report the specific sample sizes for each dataset.

Response to Reviewer Comment 7:

We thank the reviewer for this valuable comment. In the revised manuscript, we have refined the data partitioning strategy and provided a clearer explanation of the roles of each dataset.

Specifically, to rigorously evaluate the spatial generalization performance of the model and avoid potential data leakage, we employed spatial block cross-validation combined with GridSearchCV during localized modeling. In this method, observation sites were first grouped into spatial blocks based on their geographic locations, and cross-validation was then conducted across blocks rather than through random splitting at the individual site level. This approach ensured that geographically adjacent sites were not simultaneously included in both the training and testing subsets, thereby enabling a stricter and more realistic assessment of the model's generalization ability to new regions. Based on this revised scheme, we retrained and re-evaluated the XGBoost models. The updated results and methodological details are now presented in the revised manuscript (L318–336).

As this study involves multiple soil depths and spatial subregions, the exact sample sizes vary across cases and are therefore not reported individually in the main text.

However, we have clearly specified the data partitioning ratios and their purposes to ensure methodological transparency and reproducibility. We believe that this revised scheme not only aligns with common practice but also provides a stricter and more realistic evaluation of the model's generalization performance.

Revised Text (L318-336):

To rigorously account for the strong spatial autocorrelation of T_s and avoid potential data leakage between training and testing subsets, we employed a spatial block cross-validation scheme rather than random splitting. Specifically, within each rotated quadtree grid, observation sites were grouped into spatial blocks based on their geographic coordinates: station latitude and longitude were each divided by 1° and floored to integer values, and stations sharing the same index were assigned to the same block. This ensured that samples within the same spatial block were not simultaneously assigned to both the training and testing subsets, thereby avoiding data leakage due to spatial autocorrelation and enabling a more reliable evaluation of the model's generalization capability.

Within each spatial grid, the data were partitioned into training (90%) and testing (10%) subsets at the block level. The training subset was further subjected to 10-fold spatial block cross-validation using GridSearchCV to optimize three key hyperparameters: the number of trees (n_estimators), maximum tree depth (max_depth), and learning rate (learning_rate). Detailed parameter settings are provided in Appendix Table S1. The hyperparameter set that yielded the lowest average validation error across the ten folds was selected as optimal. The final model was retrained on the full training set with the optimized parameters and evaluated on the held-out testing set to assess generalization.

Reviewer Comment 8:

The authors produced data at a 1-kilometer resolution for China. How did the authors account for the spatial scale difference between point observations of soil temperature and the 1-kilometer resolution results? How was it ensured that the dataset constructed through point observation training could represent results at the 1-kilometer spatial scale? Additionally, regarding the dataset production, I am very interested in the subsequent maintenance and updates of the dataset over time. Can the authors' method be extended to produce datasets for subsequent years?

Response to Reviewer Comment 8:

We thank the reviewer for this important and thoughtful comment. It involves two critical aspects:

(1) the scale consistency between point-based observations and gridded predictions at a 1 km resolution;

(2) the potential for dataset maintenance and future updates. We address both issues below.

1. Addressing the Scale Difference Between Point Observations and 1 km Predictions

To reconcile the spatial scale mismatch between point-level T_s observations and the 1 km gridded outputs, we implemented a multi-pronged modeling strategy designed to ensure scale compatibility and representativeness:

(1) Predictor Resolution consistency:

All input variables used for model training (e.g., MODIS, ERA5-Land, and soil texture data) were uniformly resampled to a spatial resolution of 1 kilometer, thereby ensuring that the spatial scale of the predictors is consistent with that of the target output.

(2) Rotated Quadtree-Based Local Modeling:

As detailed in the revised Section 2.3.2, we employed a spatially adaptive modeling strategy based on rotated quadtree partitioning. This approach automatically divides the study area into spatial units of varying sizes according to the density of observation stations—finer grids in densely sampled areas and coarser grids in sparsely observed regions. Within each unit, a localized XGBoost model was trained using in-situ observations and 1 km-resolution environmental predictors. To mitigate edge effects and directional bias introduced by fixed partition boundaries, we constructed quadtree structures under six different rotation angles (0° to 75°). For each soil depth layer, the predictions from these rotated models were averaged, thereby reducing boundary artifacts and enhancing the spatial continuity and robustness of the final results.

(3) Robust Evaluation Framework:

A two-tier validation framework was established to comprehensively assess model performance. First, we applied spatial block cross-validation within each rotated quadtree grid. In this scheme, observation sites were partitioned into training (90%) and testing (10%) subsets at the block level, ensuring that geographically adjacent sites were not simultaneously included in both subsets. The training subset was further subjected to 10-fold cross-validation for parameter tuning, while the testing subset was used to rigorously evaluate spatial generalization. This approach effectively reduced the risk of data leakage caused by spatial autocorrelation and enhanced the robustness of the evaluation. Second, independent external validation was performed using daily T_s observations from 18 flux tower sites of the ChinaFLUX network. The results (Section 3.1, Figure 5) show that the dataset maintains high accuracy at these independent sites, further confirming the reliability and robustness of the evaluation framework.

(4) Established Precedents:

The use of point-based observations to train models for gridded prediction has been widely applied in related environmental studies, such as land surface temperature and soil moisture estimation (Karthikeyan and Mishra, 2021; Song et al., 2022; Yu et al., 2024). Our method builds on these established practices by incorporating spatial adaptivity and ensemble averaging, further enhancing consistency and robustness.

2. Potential for Dataset Extension and Future Updates

We greatly appreciate the reviewer's interest in the extensibility and long-term value of the dataset. As elaborated in the revised discussion section, the proposed spatially adaptive modeling framework is designed to be modular and scalable, making it readily applicable to future years. Given access to updated in-situ station observations and corresponding environmental predictors (e.g., MODIS and ERA5-Land), the same modeling pipeline can be re-applied to retrain the models and generate new products. This allows for filling historical data gaps and extending T_s estimates into future periods. In addition, we are currently generating T_s estimates for the period 2001–2010, which will soon be released through the National Tibetan Plateau Data Center (https://data.tpdc.ac.cn). Beyond this, the dataset will be continuously maintained and updated, with all future versions openly released on the same platform to ensure free and unrestricted access for the global scientific community. We believe these ongoing efforts will provide long-term benefits for environmental monitoring, climate research, and ecosystem modeling.

Reference

Karthikeyan, L. and Mishra, A. K.: Multi-layer high-resolution soil moisture estimation using machine learning over the United States, Remote Sens. Environ., 266, 112706, https://doi.org/10.1016/j.rse.2021.112706, 2021.

Song, P., Zhang, Y., Guo, J., Shi, J., Zhao, T., and Tong, B.: A 1 km daily surface soil moisture dataset of enhanced coverage under all-weather conditions over China in 2003–2019, Earth Syst. Sci. Data, 14, 2613–2637, https://doi.org/10.5194/essd-14-2613-2022, 2022.

Yu, Y., Fang, S., Zhuo, W., and Han, J.: A Fast and Easy Way to Produce a 1-Km All-Weather Land Surface Temperature Dataset for China Utilizing More Ground-Based Data, IEEE Trans. Geosci. Remote Sens., 62, 1–16, https://doi.org/10.1109/TGRS.2024.3368707, 2024.

Reviewer Comment 9:

How did the authors account for the impact of uneven spatial distribution of observation data points on the development of the national soil temperature dataset? How do factors such as topography, landform, and vegetation cover types influence the results and uncertainties of this dataset?

Response to Reviewer Comment 9:

1. Addressing the Impact of Uneven Spatial Distribution of Observations

We sincerely thank the reviewer for raising the important issue of uneven spatial distribution of T_s observation sites and its implications for national-scale dataset development. As noted, T_s stations in China are concentrated in eastern lowland regions, with sparse coverage across the western and high-altitude areas. This spatial imbalance poses a major challenge to constructing a robust and spatially representative T_s dataset.

To address this, we adopted a spatially adaptive modeling framework based on rotated quadtree partitioning approach. This method improves the dataset construction in two primary ways. First, it dynamically subdivides the study area into spatial units based on observation density: finer grids are assigned to densely sampled regions to improve local precision and avoid overfitting, while coarser grids are used in sparsely sampled areas to maintain model stability and statistical representativeness. Within each grid cell, a localized XGBoost model is trained to capture nonlinear relationships between T_s and relevant environmental drivers, including topography, landforms, climate, and vegetation. This strategy mitigates structural biases associated with training a single global model on unevenly distributed data. Second, to reduce boundary artifacts caused by fixed grid divisions, we generated quadtree structures under multiple rotation angles and averaged their predictions. This ensemble strategy enhanced the spatial coherence and robustness of the final T_s dataset (see revised Section 2.3 and Section 4.1 for detailed explanations).

Revised Text (L270-L298):
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 samples and coarser grids in sparse regions. Compared to global modeling or static grid partitioning, this adaptive approach offers improved regional modeling fidelity while significantly enhancing computational efficiency. The procedure consists of the following steps:

(1) Initialization of Minimum Units
The entire spatial domain was first divided into uniform, minimum-sized units (leaf nodes), each representing a fundamental spatial element. These units may contain zero or more in-situ observations. This initial step provides the base resolution for subsequent hierarchical construction. The structure and principle of quadtree spatial indexing are illustrated in Fig. S2.

(2) 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. S3). 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 different 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 (see Fig. 2). 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.

Revised Text (L516-L564):

4.1 The advantages of the spatially adaptive model

Previous studies have explored various approaches for constructing T_s datasets. For instance, Wang et al., (2023) created a daily multi-layer T_s dataset for China (1980-2010) at 0.25° resolution, employing interpolation techniques including the thin-plain spline and the angular distance weight interpolation methods with over 2,000 in-situ observations. A persistent challenge in building national-scale T_s datasets, however, lies in the highly uneven spatial distribution of observation stations—densely clustered in eastern lowlands while remaining sparse in western and high-altitude regions. Global modeling approaches, which train a single unified function across the entire domain, are inherently limited in capturing the nonlinear and non-stationary relationships between T_s and its predictors in such heterogeneous landscapes. Specifically, in sparsely sampled regions, global models lack sufficient data to learn effectively, resulting in low prediction accuracy. In contrast, in densely sampled areas, the model tends to overfit, and the training process becomes disproportionately influenced by those regions. This imbalance introduces systematic biases and limits model generalizability.

Reanalysis datasets, which synergize data assimilation systems with numerical weather prediction and land surface modeling frameworks, provide valuable representations of land-atmosphere interactions and subsurface heat transfer processes. These products are particularly advantageous for large-scale climate simulations and long-term environmental assessments. Yang and Zhang (2018) assessed the T_s accuracy of four reanalysis datasets (ERA-Interim/Land, MERRA-2, CFSR, and GLDAS-2.0) in China using in-situ monthly mean T_s observations. The results showed that all reanalysis datasets consistently underestimated T_s across the country. More recently, the ERA5-Land and GLDAS 2.1 T_s dataset offers high temporal resolution (hourly/3-hour), but it is limited by a spatial resolution of 0.1 or 0.25 degrees. Beyond reanalysis datasets, some efforts have focused on constructing empirical T_s products using ML approaches. For example, the Global Soil Bioclimatic Variables dataset (Lembrechts et al., 2022), derived from Random Forest modeling with 8,519 global sensors, provides only long-term climatological means, rather than high-resolution daily estimates.

In contrast, the methodological framework proposed in this study addresses both accuracy and resolution limitations. The spatially adaptive modeling strategy offers significant advantages over traditional interpolation and globally trained ML models. Its core strength lies in localized modeling, which accounts for regional variability in topography, soil properties, and climate conditions. As shown in Fig. S5, the rotated quadtree strategy partitions space at six orientations (0°–75°), enabling a more nuanced representation of spatial heterogeneity. By averaging predictions across these rotated configurations, the method reduces boundary artifacts often associated with static grids, resulting in smoother and more continuous spatial outputs. Moreover, the fine spatial resolution (1 km) enables the model to resolve localized thermal patterns that are critical for understanding vegetation dynamics and soil biogeochemistry. We also assessed the contribution of satellite-derived LST to model performance. As illustrated in Fig. S6, incorporating LST significantly improves spatial accuracy—especially in sparsely vegetated areas—compared to air temperature inputs, with notable enhancements in northwestern China. This highlights the importance of multi-source data fusion in boosting the performance of spatially adaptive models under data-scarce conditions. In summary, our spatially adaptive local modeling approach offers a more robust and scalable solution for large-scale T_s estimation under heterogeneous station distributions and complex environmental conditions.

2. Influence of Topography, Climate, and Vegetation on Model Performance and Uncertainty

We also thank the reviewer for pointing out the potential influence of environmental factors on model uncertainty. As shown in Sections 3.2 and 3.3 of the revised manuscript, although the overall accuracy of the dataset is satisfactory, the estimation performance exhibits clear spatial and seasonal heterogeneity. To address this, we expanded the discussion in Section 4.3 to systematically examine how factors such as topography, climate conditions, land cover types, and remote sensing variables may affect the stability and accuracy of T_s estimates across different regions and seasons. We also proposed future directions for improving model adaptability under complex environmental conditions. These revisions aim to clarify how our methodology accounts for spatial sampling bias and environmental complexity, and we hope they address the reviewer's concerns comprehensively.

Revised Text (L602-L662):
Despite the strong performance of our spatially adaptive T_s estimation framework, several limitations warrant acknowledgment. As shown in Figures 6 and 7, model validation at station level reveals spatial heterogeneity in prediction accuracy, with relatively lower performance observed in the YGP and the QTP regions. On the one hand, as evidenced by Figure 10, our multi-source modeling framework captures T_s variations across different elevations and geomorphic conditions more effectively than existing datasets. However, the QTP and YGP are characterized by complex terrain and high altitudes, coupled with rapidly changing climatic conditions, which significantly complicate T_s estimation. These findings align with previous studies showing that high elevations intensify the disconnect between air temperature and LST, thereby increasing the uncertainty in thermal modeling (Mo et al., 2025).

MODIS LST serves as a critical input to our modeling framework. However, as an optical remote sensing product, it is highly susceptible to cloud contamination, often resulting in data gaps. Despite the use of spatiotemporal interpolation and SG filtering, residual uncertainties persist in the reconstructed LST data. Future improvements in T_s reconstruction can be pursued along two main directions. First, more physically grounded LST reconstruction methods can be adopted, such as incorporating surface energy balance models and diurnal temperature cycle models (Hong et al., 2022; Firozjaei et al., 2024; Wang et al., 2024). These methods apply energy conservation principles to estimate T_s during periods of missing or unreliable observations, thereby providing more realistic estimates of land surface thermal conditions during periods of cloud cover. Second, integrating higher temporal resolution remote sensing observations may help overcome the limitations of MODIS. For instance, passive microwave satellite data provide all-weather observations and are less sensitive to cloud interference (Duan et al., 2017; Wu et al., 2022). In addition, next-generation geostationary satellites such as Himawari-8 offer observations at 10-minute intervals, substantially enhancing the temporal continuity and quality of surface temperature estimates (Yamamoto et al., 2022; You et al., 2024). These enhancements are expected to significantly improve the accuracy and temporal continuity of soil temperature monitoring.

Our results (Figures 8 and 9) show that model accuracy varies across different soil depths, with additional influences from season and land use. Accuracy is relatively lower at the surface (0 cm), improves at intermediate depths (5–10 cm), and then declines again at greater depths (20–40 cm). This depth-dependent pattern can be explained by the physical characteristics of soil temperature. Surface soil temperature is highly sensitive to short-term meteorological fluctuations such as radiation, precipitation, and evapotranspiration, leading to greater spatiotemporal variability and larger prediction errors. In contrast, intermediate soil layers benefit from the buffering effects of thermal diffusion and soil heat capacity, which dampen high-frequency fluctuations and stabilize the relationship between predictors and T_s, thereby improving performance at these depths. At greater depths, however, surface-level errors propagate downward through the cascading framework, resulting in reduced accuracy—particularly during summer and winter.

Seasonal changes and variations in land cover further contribute to differences in estimation accuracy. As shown in Figures 8 and 9, the model exhibits higher accuracy in spring and autumn, whereas its performance tends to decline during summer and winter. During summer, dense vegetation growth and canopy closure reduce the influence of surface–atmosphere energy exchanges on T_s, weakening the correlation between canopy temperature and subsurface T_s (Kropp et al., 2020; Cui et al., 2022). In winter, snow cover introduces a suite of confounding effects: high surface albedo reduces net radiation (Loranty et al., 2014; Li et al., 2018), while snow acts as an insulator, limiting the soil's response to cold air incursions (Zhang, 2005; Myers-Smith et al., 2015). Additionally, low temperatures lead to soil water freezing, which alters the soil's thermal conductivity and heat storage capacity. These factors, together with frequent freeze–thaw cycles, introduce complex nonlinear dynamics in T_s that increase modeling uncertainty (Li et al., 2023a; Imanian et al., 2024). While our multi-source adaptive modeling framework performs well across depths, it does not explicitly account for the physical mechanisms of vertical heat transfer. Future research could explore deep learning models that are capable of learning complex spatiotemporal features and improving the physical interpretability of T_s variations across time, space, and depth.

Reference

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

Duan, S.-B., Li, Z.-L., and Leng, P.: A framework for the retrieval of all-weather land surface temperature at a high spatial resolution from polar-orbiting thermal infrared and passive microwave data, Remote Sens. Environ., 195, 107–117, https://doi.org/10.1016/j.rse.2017.04.008, 2017.

Firozjaei, M. K., Mijani, N., Kiavarz, M., Duan, S.-B., Atkinson, P. M., and Alavipanah, S. K.: A novel surface energy balance-based approach to land surface temperature downscaling, Remote Sens. Environ., 305, 114087, https://doi.org/10.1016/j.rse.2024.114087, 2024.

Hong, F., Zhan, W., Göttsche, F.-M., Liu, Z., Dong, P., Fu, H., Huang, F., and Zhang, X.: A global dataset of spatiotemporally seamless daily mean land surface temperatures: Generation, validation, and analysis, Earth Syst. Sci. Data, 14, 3091–3113, https://doi.org/10.5194/essd-14-3091-2022, 2022.

Imanian, H., Mohammadian, A., Farhangmehr, V., Payeur, P., Goodarzi, D., Hiedra Cobo, J., and Shirkhani, H.: 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, 2024.

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., and Lund, M.: 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, 2020.

Li, Q., Ma, M., Wu, X., and Yang, H.: 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, 2018.

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

Loranty, M. M., Berner, L. T., Goetz, S. J., Jin, Y., and Randerson, J. T.: 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, 2014.

Mo, Y., Pepin, N., and Lovell, H.: Understanding temperature variations in mountainous regions: The relationship between satellite-derived land surface temperature and in situ near-surface air temperature, Remote Sens. Environ., 318, 114574, https://doi.org/10.1016/j.rse.2024.114574, 2025.

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., and Vellend, M.: Climate sensitivity of shrub growth across the tundra biome, Nat. Clim. Change, 5, 887–891, https://doi.org/10.1038/NCLIMATE2697, 2015.

Wang, Q., Tang, Y., Tong, X., and Atkinson, P. M.: Filling gaps in cloudy landsat LST product by spatial-temporal fusion of multi-scale data, Remote Sens. Environ., 306, 114142, https://doi.org/10.1016/j.rse.2024.114142, 2024.

Wu, P., Su, Y., Duan, S., Li, X., Yang, H., Zeng, C., Ma, X., Wu, Y., and Shen, H.: A two-step deep learning framework for mapping gapless all-weather land surface temperature using thermal infrared and passive microwave data, Remote Sens. Environ., 277, 113070, https://doi.org/10.1016/j.rse.2022.113070, 2022.

Yamamoto, Y., Ichii, K., Ryu, Y., Kang, M., and Murayama, S.: Uncertainty quantification in land surface temperature retrieved from Himawari-8/AHI data by operational algorithms, ISPRS J. Photogramm. Remote Sens., 191, 171–187, https://doi.org/10.1016/j.isprsjprs.2022.07.008, 2022.

You, W., Huang, C., Hou, J., Zhang, Y., Dou, P., and Han, W.: Reconstruction of MODIS LST Under Cloudy Conditions by Integrating Himawari-8 and AMSR-2 Data Through Deep Forest Method, IEEE Trans. Geosci. Remote Sens., 62, 1–17, https://doi.org/10.1109/TGRS.2024.3388409, 2024.

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

Reviewer Comment 10:

There are also some minor issues that should be addressed. For example, in Figures 6 and 7, it is recommended to include a color bar legend. As it stands, it is difficult to interpret the exact values represented by the orange points. In Equation (2), the variables x and y lack subscripts i. In Equation (4), the variable i used for summation is not defined. In the references, line 667 and 763 include "others" among the authors—what does this mean? It is suggested to carefully check the manuscript for such details, including grammar, figures, and reference formatting.

Response to Reviewer Comment 10:

We thank the reviewer for the careful reading and helpful suggestions. In response:

1. Figure revisions

We have redrawn the portion of Figure 1 related to dataset division in the revised manuscript to present it more clearly to the readers. Additionally, we have added color bar legends to both Figure 6 and Figure 7. This addition clarifies the exact values represented by the orange points and enhances the interpretability of the figures.

Revised Text (L342-L344):

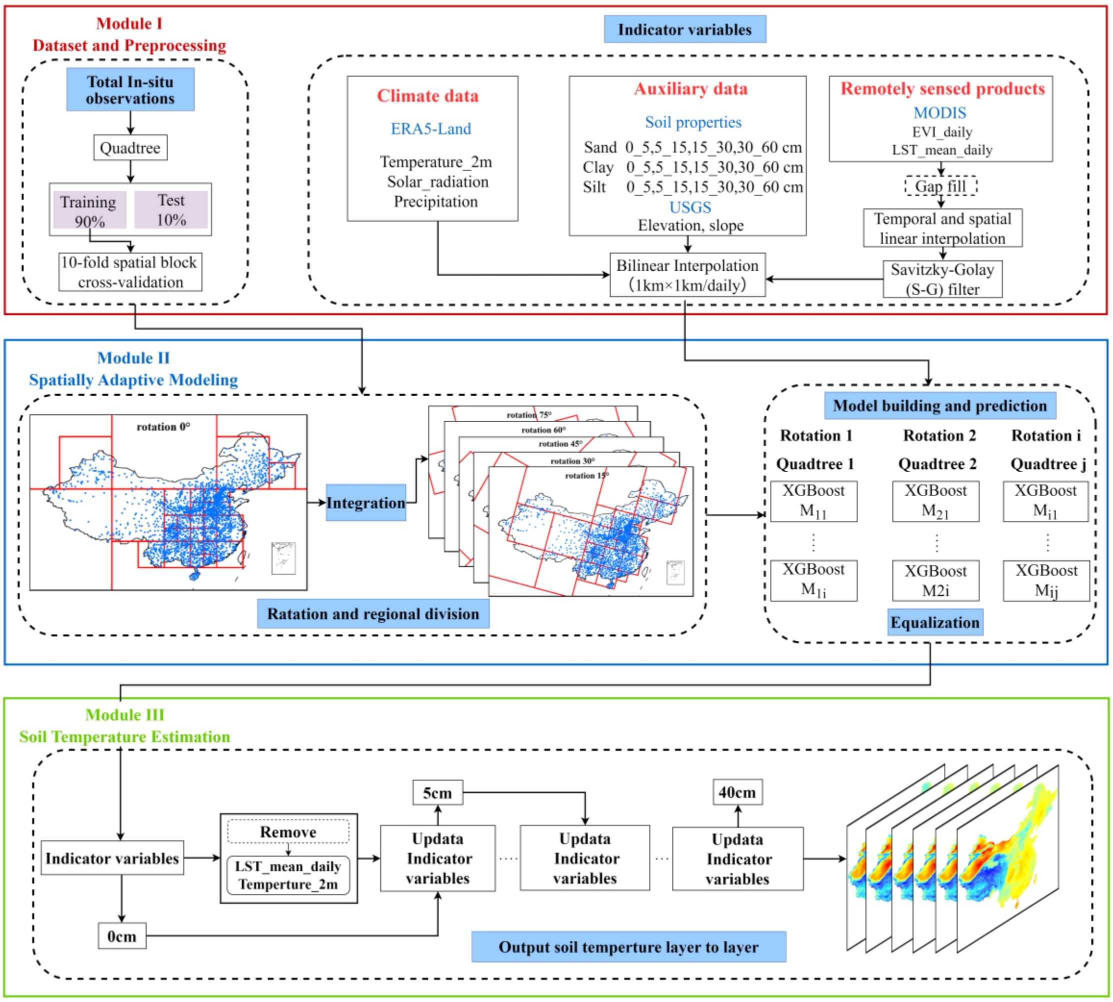

Figure 3. Workflow of the proposed method to obtain multi-layer T_s over the China.

Revised Text (L402-L408):

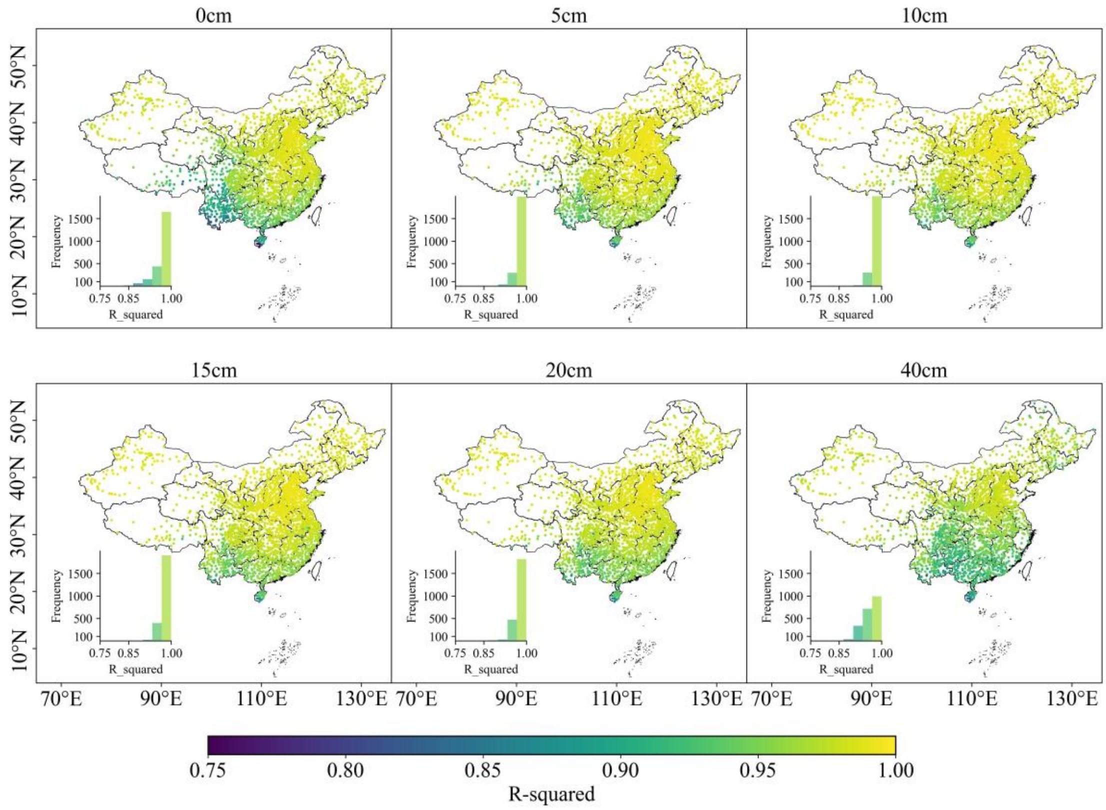

Figure 6. Goodness of R^2 across China estimated during the model testing phase. Performance metrics are calculated between predicted_T_s and in-situ T_s data sets.

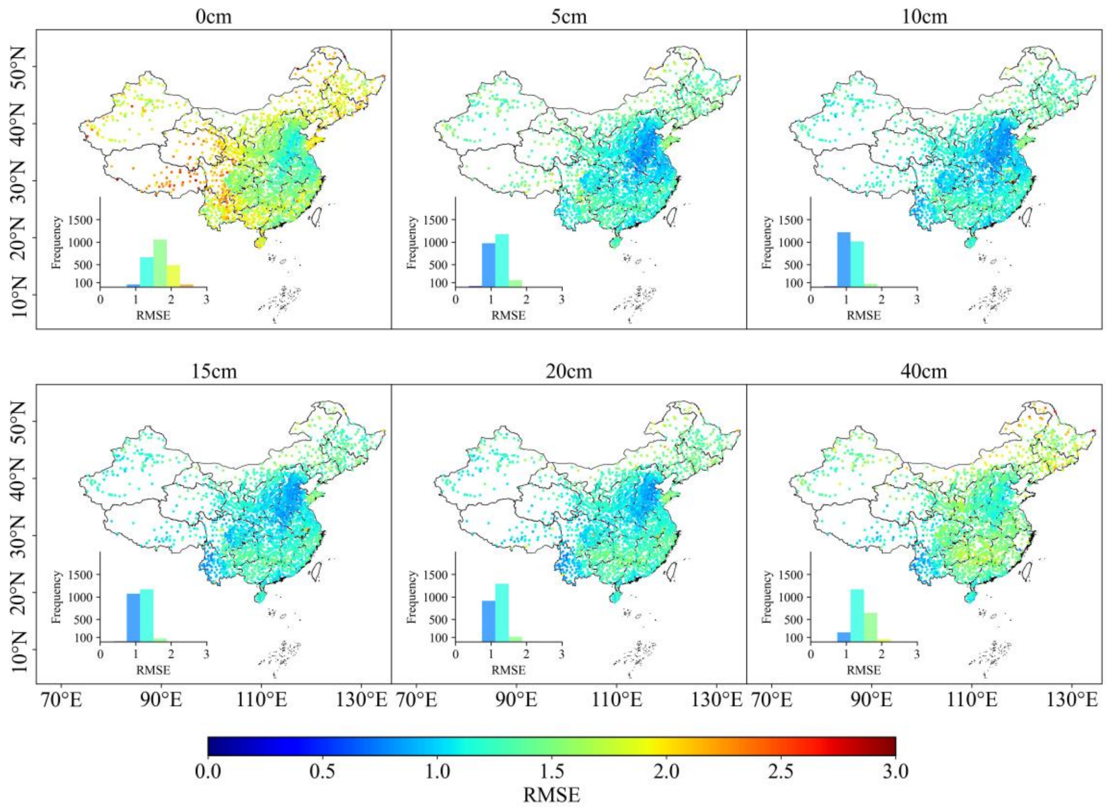

Figure 7. Goodness of RMSE across China estimated during the model testing phase. Performance metrics are calculated between predicted T_s and in-situ T_s data sets.

2. Equation corrections

Revised Text (L352-L359):

Equation (2) has been corrected to include subscripts i for both x and y, to clearly indicate that the RMSE is calculated over paired observations.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}[(x_i - \overline{X}) - (y_i - \overline{Y})]^2}{N}} \tag{1.2}$$

Equation (4) has been reformulated to explicitly define the summation index i and to reflect the mean bias across all samples.

$$Bias = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i) \tag{1.3}$$

3. Reference formatting

In accordance with the reviewer's suggestion, we have carefully reviewed and revised the entire reference list to ensure formatting accuracy and consistency, fully complying with the journal's citation requirements.

4. Additional Edits

We carefully reviewed the manuscript to address minor issues in grammar, figure annotations, and reference formatting. We are grateful for the reviewer's attention to these important details, which helped us further improve the overall clarity and quality of the manuscript.