Supplement of Earth Syst. Sci. Data, 18, 97–116, 2026
https://doi.org/10.5194/essd-18-97-2026-supplement

Open Access Earth System Science Data

Supplement of

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

Xuetong Wang et al.

Correspondence to: Liang He (heliang_hello@163.com), Hao Shi (haoshi@rcees.ac.cn), and Qiang Yu (yuq@nwafu.edu.cn)

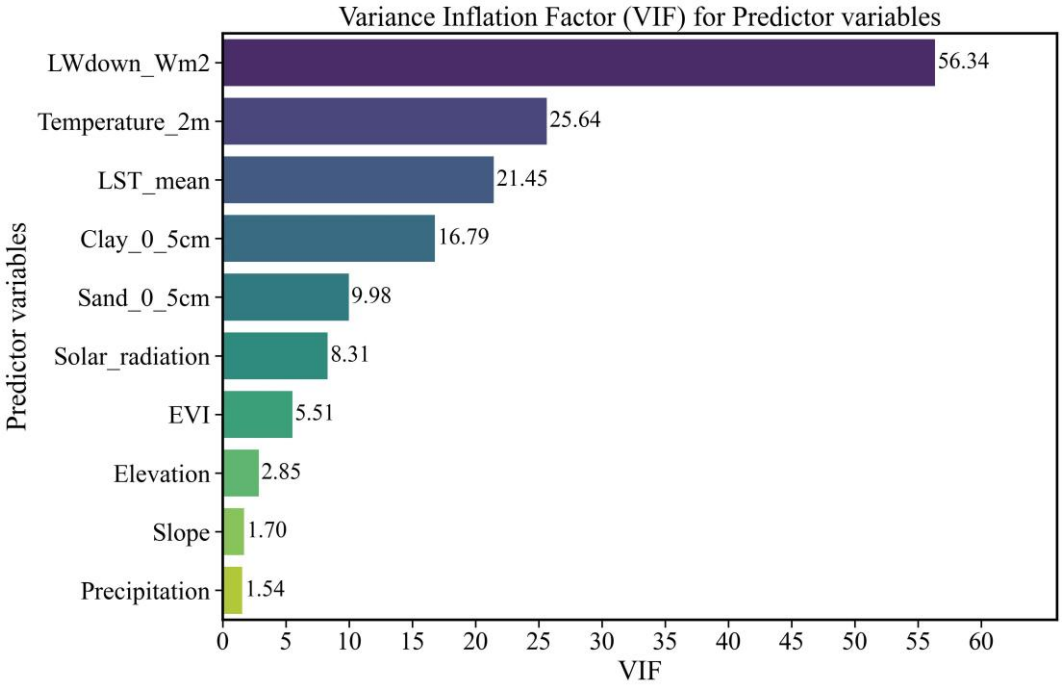

Figure S1. Variance Inflation Factor (VIF) of predictor variables (with LWD)

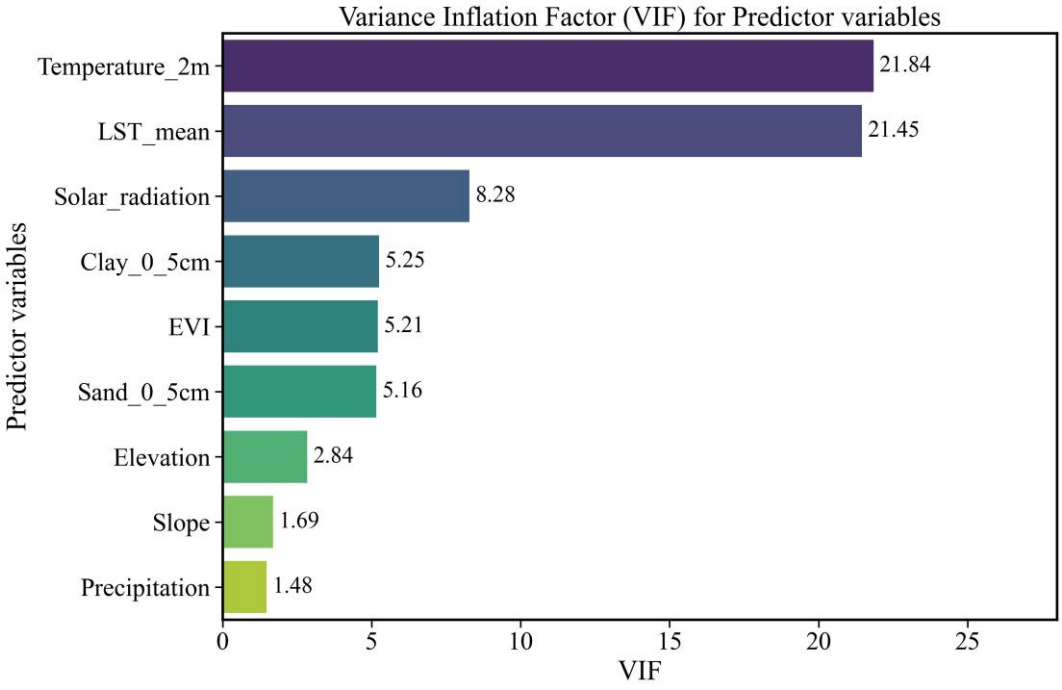

Figure S2. Variance Inflation Factor (VIF) of predictor variables

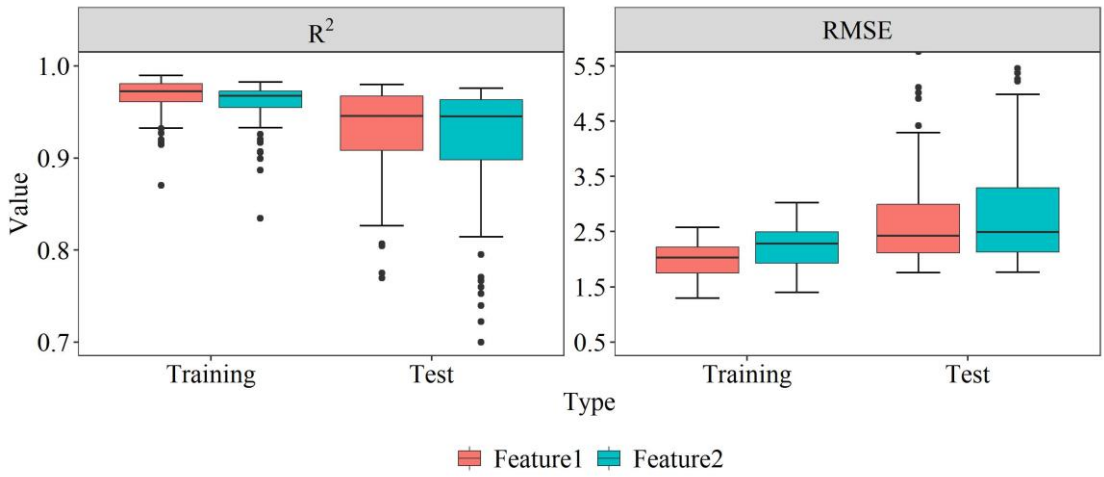

Figure S3. Comparison of Modeling Accuracy with Different Feature Variables (Feature1 represents using both air temperature and LST together with other feature variables, while Feature 2 represents using only air temperature together with other feature variables)

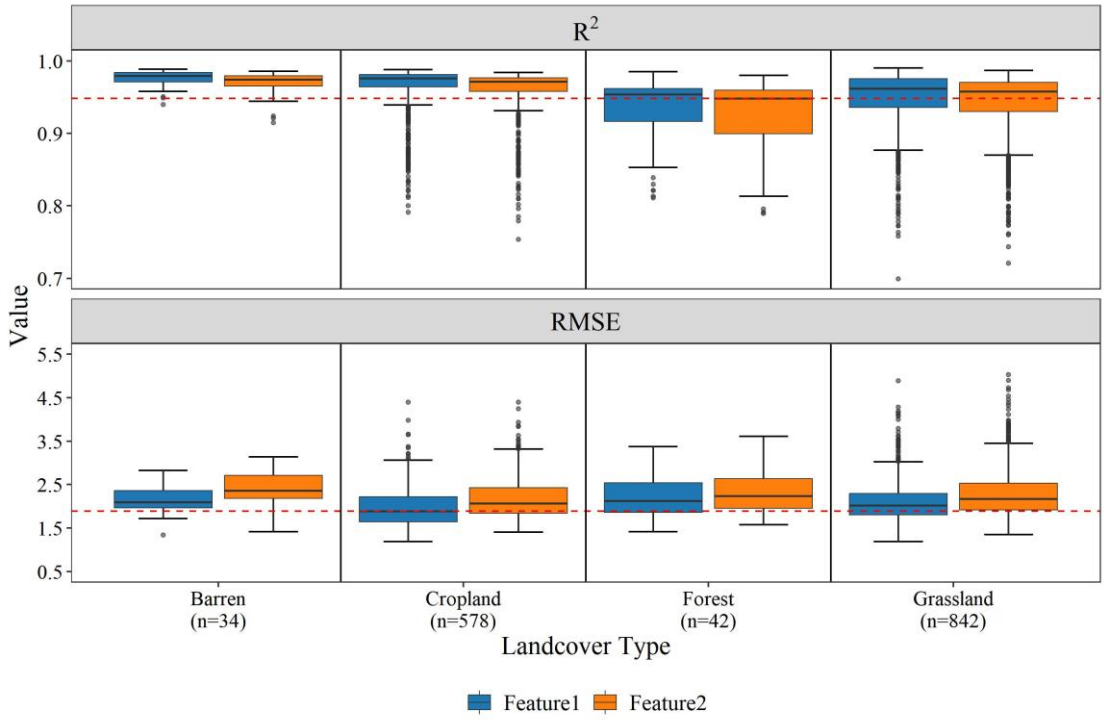

Figure S4. Differences in model accuracy across land cover types under different feature variable combinations. (Feature1 represents using both air temperature and LST together with other feature variables, while Feature 2 represents using only air temperature together with other feature variables)

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.

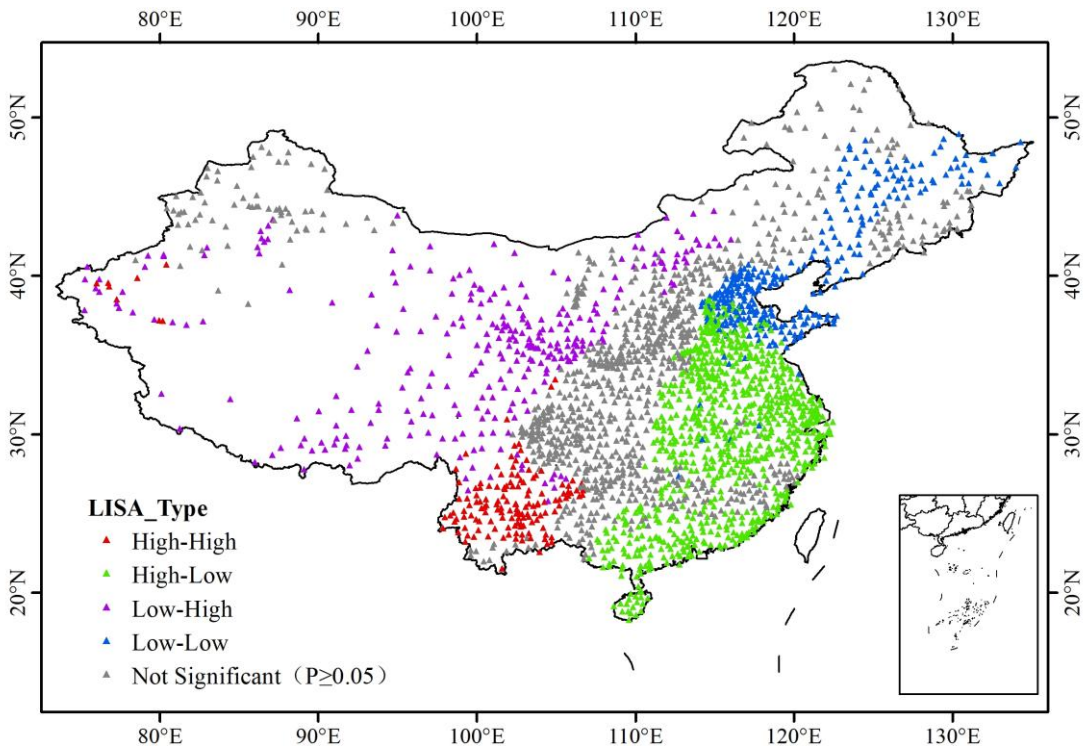

Figure S5. Spatial patterns of the bivariate Local Moran's I between GST_Avg and elevation at meteorological stations across China.

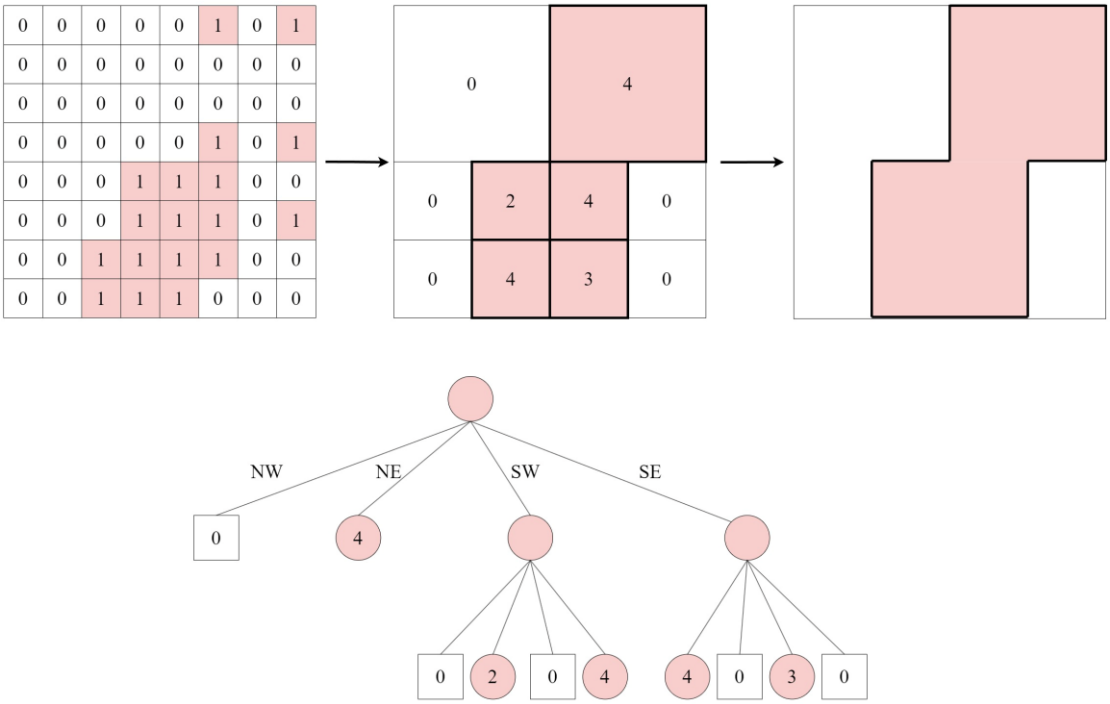

Figure S6. Quadtree spatial index principle. (Northwest (NW), Northeast (NE), Southwest (SW), Southeast (SE) refer to the four quadrants into which the quadtree divides the two-dimensional space.)

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 S7 (a–c), we systematically evaluated quadtree performance under a series of point-count thresholds (10, 30, 50, 70, 90): Figure S7a 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 S7b 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 S7c 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.

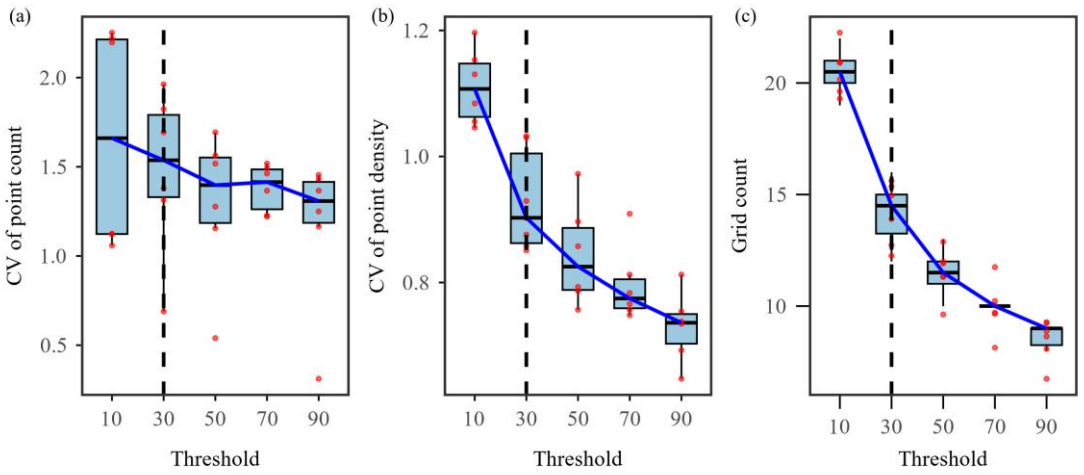

Figure S7. Performance evaluation of quadtree partitioning under different point-count thresholds. (a) Coefficient of variation (CV) of point count across spatial units. (b) CV of point density (point count per unit area). (c) Total number of generated grid cells. Dashed vertical line indicates the selected threshold of 30.

Table.S1 Candidate values of hyperparameters in XGBoost.

hyperparameters	candidate value		
	Start	End	Step
n_estimators	60	200	20
max_depth	5	15	1
learning rate	0.1	1	0.1

Table.S2 Metadata of flux tower soil temperature observations used for validation

Site name	Ecosystem type	Depth (cm)	Time series
Baotianman Forest Station	Forest	0,5,20	2010-2014
Changling Rice Paddy Station	Cropland	5,10,20	2018-2020
Daan Cropland Station	Cropland	0,5,10,15,20	2017-2020
Damao Grassland Station	Grassland	0,5,10,15,20,40	2017-2020
Danzhou Rubber Plantation Station	Forest	5,10,20	2010
Haibei Alpine Meadow Station	Grassland	5,10,15,20,40	2015-2020
Haibei Shrubland Station	Grassland	0,5,20,40	2016-2018
Huzhong Boreal Forest Station	Forest	5,10,20	2014-2018
Jinzhou Cropland Station	Cropland	5,10,15,20,40	2011-2014
Lijiang Alpine Meadow Station	Grassland	5,10,15,20,40	2013-2020
Maoershan Forest Station	Forest	5	2016-2018
Panjin Reed Wetland Station	Wetland	10,20,40	2018-2020
Qianyanzhou Plantation Forest Station	Forest	5,10,20	2011-2015

Ruoergai Alpine Wetland Station	Wetland	0,5,10,20	2013-2020
Sanjiangyuan Alpine Grassland Station	Grassland	0,5,15	2013-2015
Taoyuan Cropland Station	Cropland	5,10,15,20,40	2010-2014
Xishuangbanna Rubber Plantation Station	Forest	0,5,20	2010-2014
Yuanjiang Dry-Hot Valley Savanna Station	Grassland	5,10,20,40	2013-2015

95

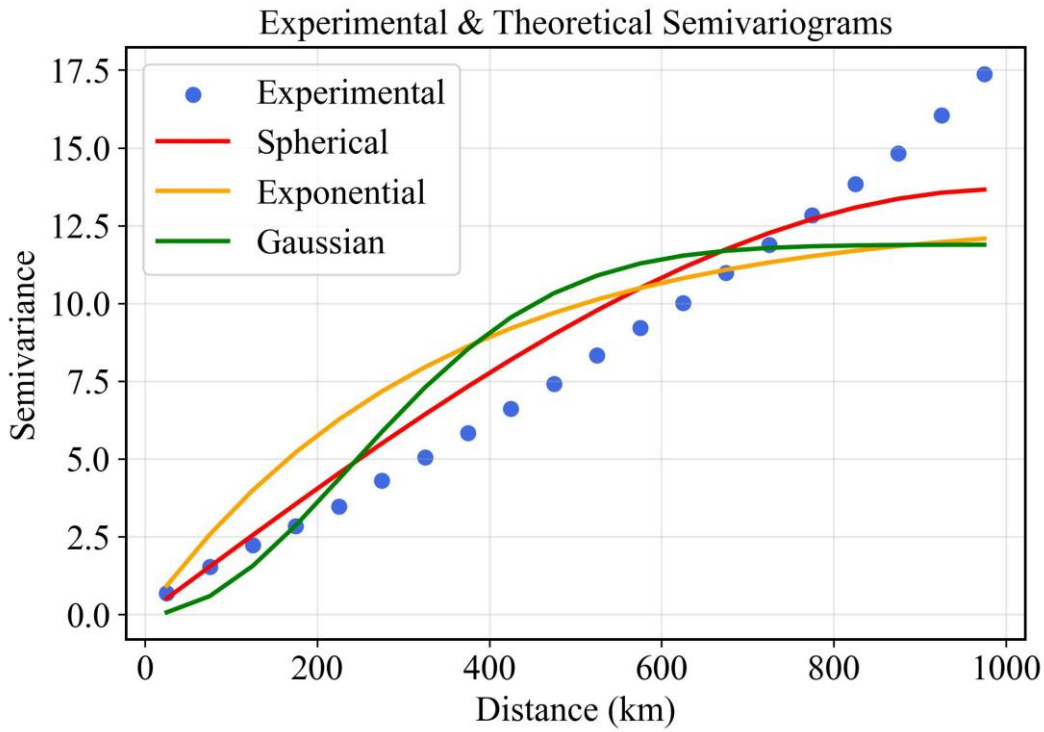

Figure S8. Experimental and theoretical semivariograms of annual mean T_s in 2020 at 0 cm (The spherical, exponential, and Gaussian models are fitted for comparison).

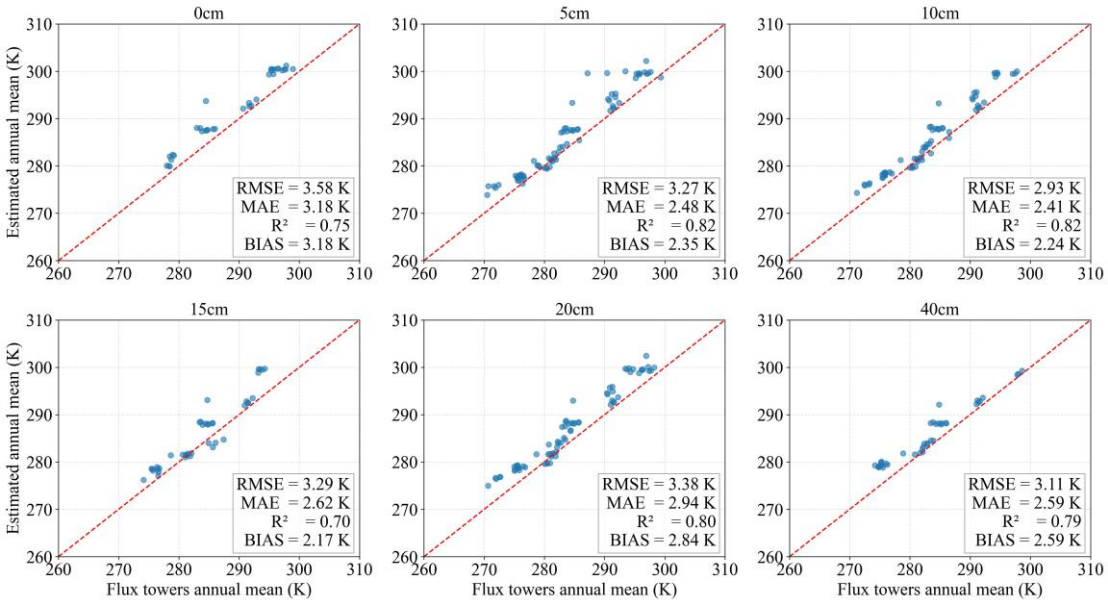

Figure S9. Comparison between estimated and FLUX towers annual mean T_s across six depths (0~40 cm)

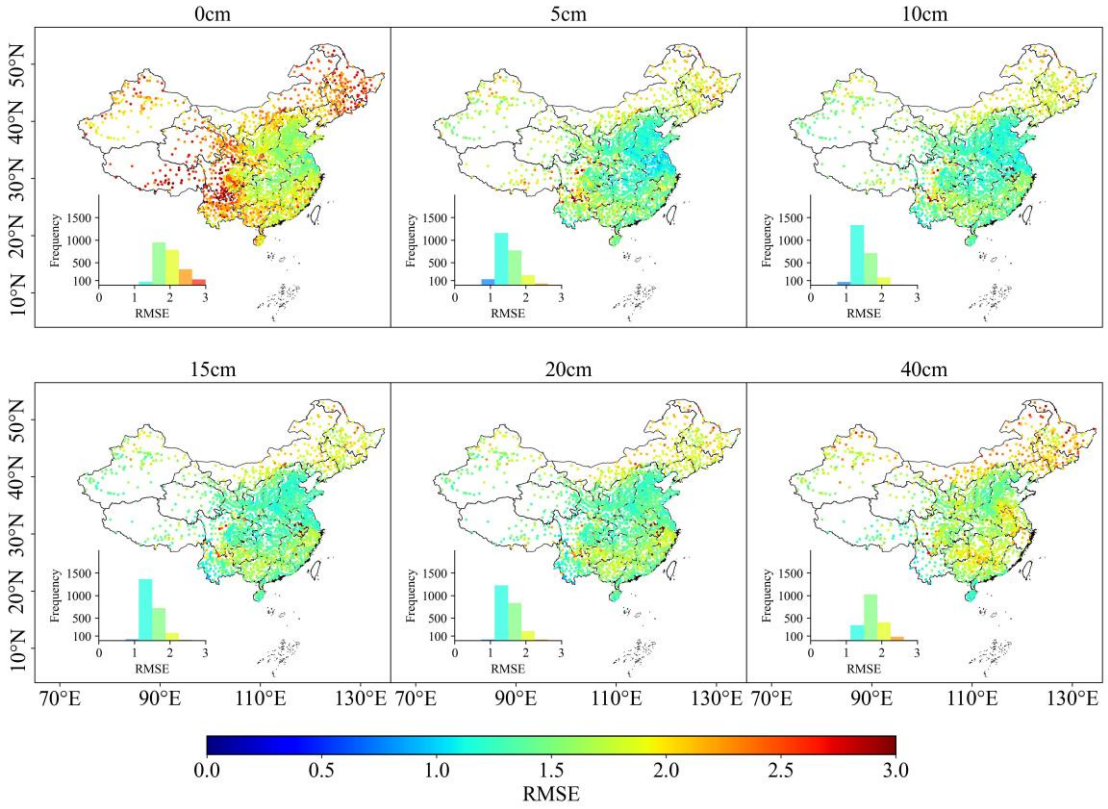

Figure S10. Goodness of RMSE across China estimated during the model testing phase. Perfrmance metrics are calculated between predicted T_s and in-situ T_s data sets.

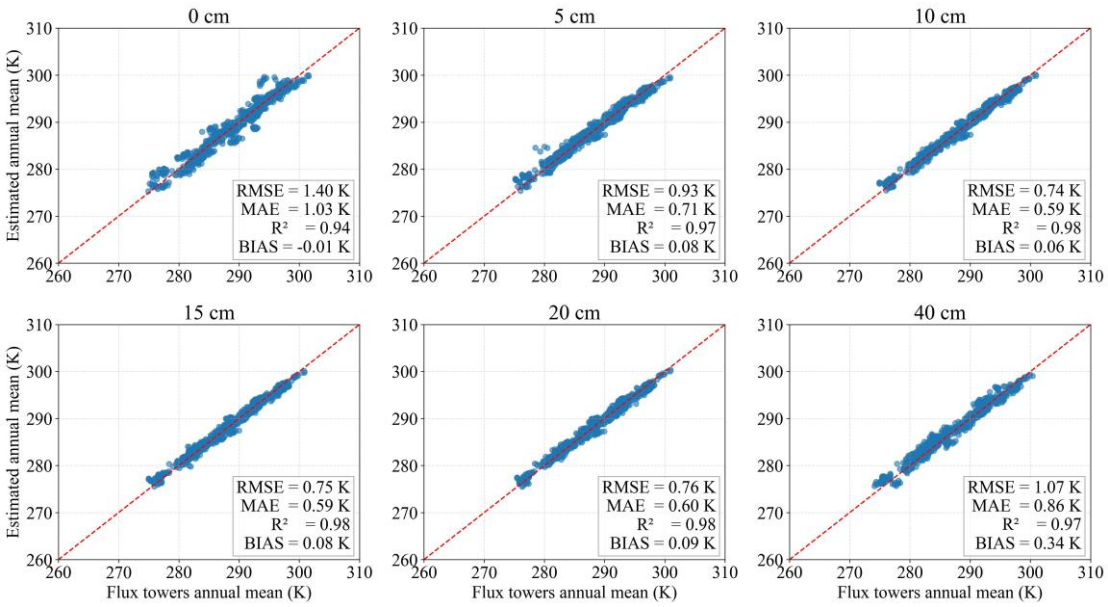

Figure S11. Comparison between estimated and observed annual mean T_s across six soil depths (0~40 cm) on the test set

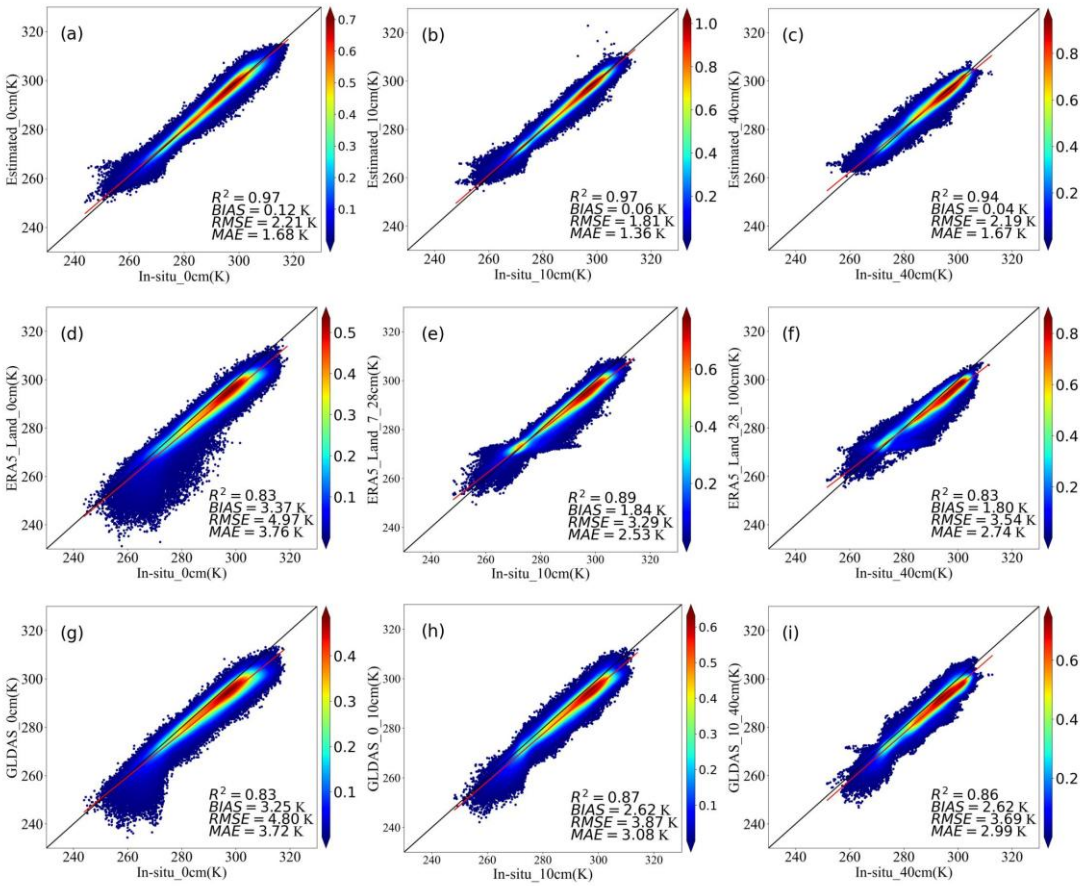

Figure S12. Scatter density plot comparing the accuracy of different products (e.g., 0, 10, and 40 cm)

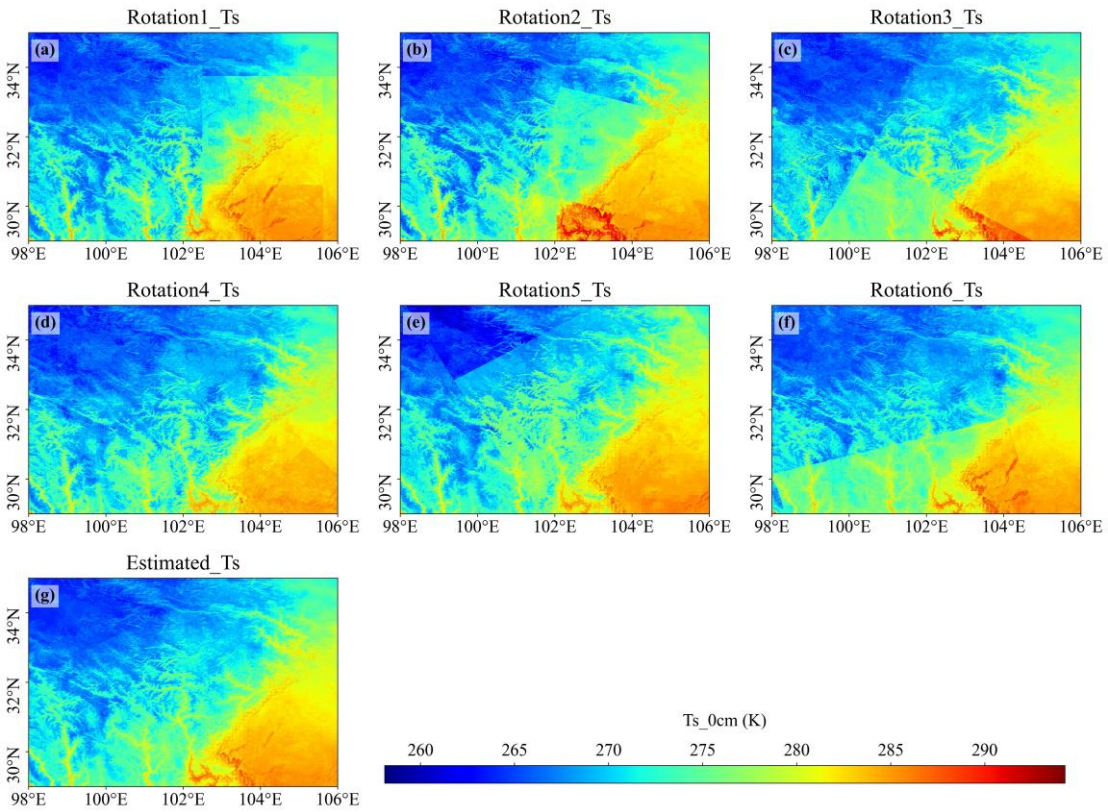

Figure S13. The difference in spatial performance after modeling different rotations separately (Rotation1-Rotation 6) and the average value of all different rotations (Estimated_T_s) within the same sub-region (29°N–35°N, 98°E–106°E)

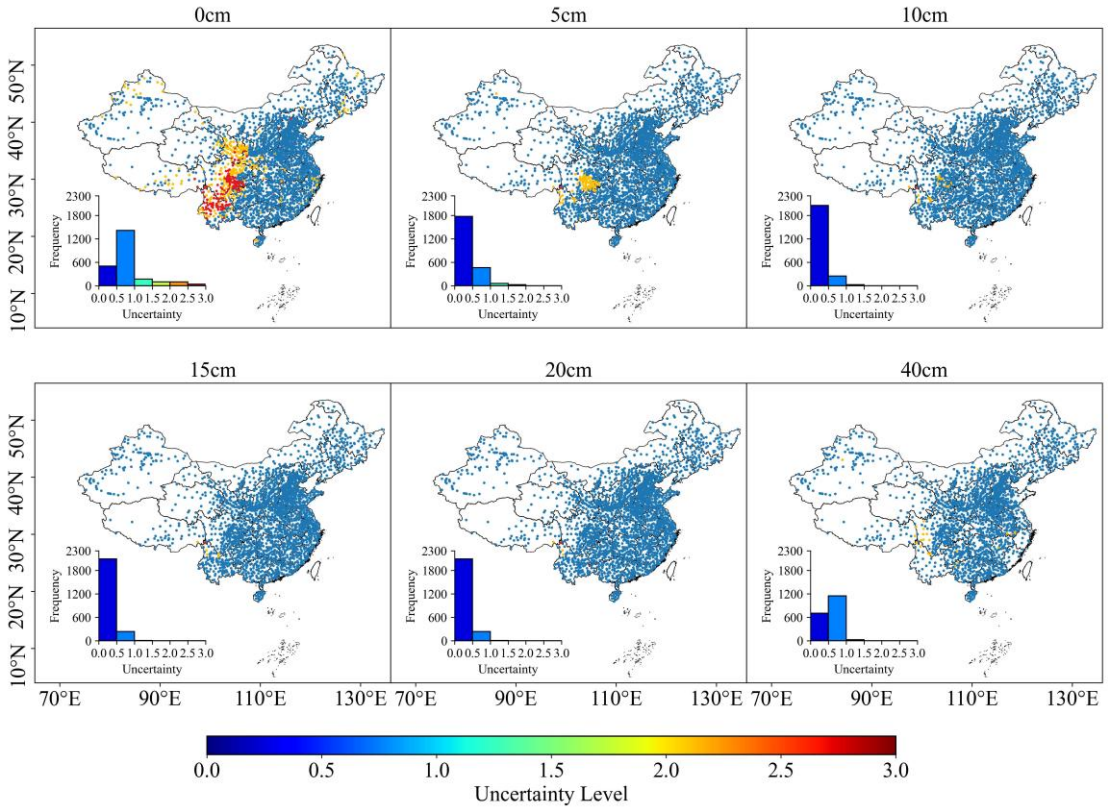

Figure S14. Spatial patterns of prediction uncertainty at six depths (0~40 cm) based on the rotated-quadtree ensemble. Note:The uncertainty metric shown here represents the variability induced by the spatial partitioning scheme rather than the total predictive uncertainty. Colored points represent site-level uncertainty values, with warmer colors indicating higher uncertainty. Insets show the frequency distribution of histograms uncertainty levels at each depth.