atmosphere

MDPI

Article

Comparative Evaluation of Multi-Source Geospatial Data and Machine Learning Models for Hourly Near-Surface Air Temperature Mapping

Zexiang Yan [1,2,3], Yixu Chen [1,2,3], Ruoxue Li [1,2,3] and Meiling Gao [1,2,4,*]

[1] State Key Laboratory of Loess Science, Chang'an University, Xi'an 710054, China
[2] College of Geological Engineering and Geomatics, Chang'an University, Xi'an 710054, China
[3] Big Data Center for Geosciences and Satellites (BDCGS), Xi'an 710054, China
[4] Key Laboratory of Ecological Geology and Disaster Prevention, Ministry of Natural Resources, Xi'an 710054, China
* Correspondence: gaomeiling@chd.edu.cn

Abstract

Accurate estimation of hourly near-surface air temperature (NSAT) is critical for climate analysis, environmental monitoring, and urban thermal studies. However, existing temperature datasets remain constrained by coarse spatial resolution and limited hourly accuracy. This study systematically evaluates four widely used land surface temperature (LST) datasets—MODIS, ERA5-Land, FY-2F, and CGLS—and five machine learning models (RF, MDN, DNN, XGBoost, and GTNNWR) for NSAT estimation across two contrasting regions in Shaanxi, China: a complex-terrain region in southwestern Shaanxi and the urban area of Xi'an. Results demonstrate that single-source LST inputs outperform multi-source LST stacking, largely due to compounded systematic biases across heterogeneous datasets. MODIS provides the best performance in the mountainous region, while CGLS excels in the urban environment. Among all models, GTNNWR—which explicitly captures spatiotemporal non-stationarity—consistently achieves the highest accuracy, reducing *RMSE* by 44.8% and 44.2% relative to the second-best model in the two study areas, respectively, whereas the remaining four models exhibit broadly comparable performance. This work identifies effective data–model configurations for generating high-resolution hourly NSAT products and provides methodological insights for climate and environmental applications in regions with complex terrain or strong urban heterogeneity.

Keywords: hourly near-surface air temperature; land surface temperature; machine learning; MODIS; CGLS; GTNNWR

Academic Editor: Christian Natale Gencarelli

Received: 1 December 2025
Revised: 3 January 2026
Accepted: 8 January 2026
Published: 9 January 2026

1. Introduction

Near-surface air temperature (hereafter referred to as NSAT) is one of the most fundamental and essential variables in earth systems [1,2]. As a key parameter characterizing the interaction between the land surface energy balance and the atmospheric boundary layer, air temperature not only reflects the long-term trend and short-term fluctuations of climate change but also plays a vital role in environmental quality monitoring, agricultural and ecological research, and public heat related health risk warning [3–7]. High accuracy air temperature datasets are indispensable for understanding the energy cycling of terrestrial ecosystems, optimizing urban planning, and improving human thermal comfort [8,9]. In the context of global warming and increasing frequency of extreme temperature events,

there is an urgent need to obtain air temperature data with high spatiotemporal resolution and high accuracy for climate dynamics analysis, regional environmental management, and meteorological disaster prevention [10,11].

Traditionally, NSAT has been obtained primarily from ground-based meteorological stations [12]. While such observations provide high temporal resolution (typically hourly or sub-hourly level), their spatial distribution is sparse and uneven, especially in regions with complex topography such as plateaus, mountains, and high-latitude areas. This sparse coverage limits the ability to capture fine-scale spatial heterogeneity for regional analysis [12,13]. Meanwhile, satellite-based remote sensing has become an essential complement, enabling continuous spatial estimation of near-surface air temperature through empirical or physically based relationships with land surface temperature (LST) [14–16].

Numerous studies have confirmed the strong statistical relationship between LST and NSAT [14,17], and various LST products have been utilized to estimate air temperature across different spatial and temporal scales. Among satellite-derived datasets, the Moderate Resolution Imaging Spectroradiometer (MODIS) LST has been one of the most widely used. For example, previous studies have employed MODIS LST together with ancillary variables such as the Enhanced Vegetation Index (EVI), surface albedo, and digital elevation model to generate spatially continuous NSAT maps for regions including the Yellow River Basin [18], the Loess Plateau [19], the Jingjinji Area [20], the Portugal [15], and China [21]. Although MODIS LST provides relatively high spatial resolution at 1 km, its temporal coverage is constrained by overpass frequency and cloud contamination, making it unsuitable for continuous diurnal or hourly analyses.

In recent years, geostationary satellite systems have emerged as valuable platforms for LST retrieval. Instruments such as the Himawari-8 and Fengyun-4 (FY-4) series provide LST data at 10 to 15 min intervals, enabling detailed monitoring of diurnal thermal dynamics. Therefore, geostationary remotely sensed LST has been widely used for hourly air temperature estimation due to its superior temporal sampling compared with polar-orbiting satellites [22,23]. However, the spatial resolution of these products is typically coarse.

Reanalysis datasets represent another important source for air temperature estimation [14,24]. By integrating multi-source observations into numerical models, reanalysis products offer spatially continuous and temporally consistent data suitable for large-scale analyses [25]. For instance, the ERA5-Land reanalysis dataset provides hourly LST and air temperature at a spatial resolution of approximately 10 km [25], effectively capturing regional temperature patterns but often failing to represent local variations driven by microclimates, urban morphology, or vegetation heterogeneity. Therefore, fine-scale hourly NSAT estimation remains challenging due to the limitations of single LST sources.

Parallel to data advancements, machine learning (ML) has greatly improved air temperature estimation and has become one of the most widely used approaches [14,19,23,26]. ML methods are particularly effective in capturing complex nonlinear relationships and integrating multi-source datasets. Various traditional ML models, such as Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM), as well as several deep learning models, including Deep Neural Networks (DNN) and Deep Belief Networks (DBN), have been applied to air temperature estimation [17,18,20,23,27,28]. Among these learning models, most conventional ML approaches are trained in a point-wise manner and often fail to account for the spatial and temporal dependencies among grids. To overcome this limitation, some new models have been developed. For example, spatiotemporal models such as the Geographical–Temporal Neural Network Weighted Regression (GTNNWR) have been proposed [29] to estimate spatiotemporal non-stationary relationships. However, the comparative performance of such models under different data source conditions for NSAT estimation remains insufficiently investigated.

Overall, for the estimation of NSAT with high spatiotemporal resolution (1 km, hourly), the effective utilization of multi-source remote sensing LST data and the selection of appropriate machine learning algorithms still require further exploration. To address this gap, this study employs four types of LST data from different sources (polar-orbiting satellite product (MODIS LST), reanalysis product (ERA5-Land LST), geostationary satellite product (FY-2F LST), and multi-satellite composite product (CGLS LST)) as well as five representative machine learning models, including RF, XGBoost, DNN, MDN (Mixture Density Network), and GTNNWR. Based on the Google Earth Engine (GEE) cloud platform and the Python computing environment (Python 3.9), we developed a comparative framework integrating these multi-source LST datasets and machine learning algorithms to estimate hourly NSAT. Using in situ observations as reference data, this study systematically evaluates the accuracy of different "data source–model" combinations. The results aim to provide both methodological guidance and empirical evidence for selecting the optimal temperature estimation strategy under various application scenarios.

2. Materials and Methods

All data acquisition, preprocessing, and variable integration were conducted using the Google Earth Engine (GEE) cloud platform (Google Earth Engine, Google LLC, Mountain View, CA, USA; accessed in 2025) and the Python computing environment (Python 3.9, Python Software Foundation, Wilmington, DE, USA).

2.1. The Study Region

Because the analysis was conducted on the GEE platform and subject to free-user memory limitations, it was not feasible to include the entire province within the study domain. Therefore, two representative regions in Shaanxi Province, China (Figure 1a), were selected for analysis. As shown in Figure 1b,c, the first study area is located in the southwestern part of Shaanxi Province (Figure 1b). This region lies within the Qinba Mountains, where the terrain is dominated by mountains and hills, with elevations generally decreasing from south to north and showing considerable variation [30]. Vegetation is mainly composed of forests and grasslands. The area has a northern subtropical humid monsoon climate. Owing to the barrier effect of the Qinling Mountains, it possesses abundant thermal and moisture resources. Land cover is primarily forest and cropland, and the natural environment exhibits strong spatial heterogeneity. Precipitation generally decreases from south to north and shows distinct seasonality, being concentrated in summer and autumn with pronounced interannual variability. Spanning both the Yellow River and Yangtze River basins, this area represents a significant climatic and ecological transition zone in Shaanxi Province.

To assess the regional generalizability of the optimal data sources and ML models, the second study area, corresponding to the Xi'an metropolitan region, was selected (Figure 1c). Xi'an is situated in the central Guanzhong Plain at an average elevation of approximately 400 m. The terrain is flat and open, and the region is highly urbanized, with built-up and agricultural lands dominating the underlying surface. The area experiences a warm temperate, semi-humid continental monsoon climate [31]. In terms of topography, land-surface structure, and human activity intensity, Xi'an forms a marked contrast with the mountainous southwestern part of Shaanxi Province.

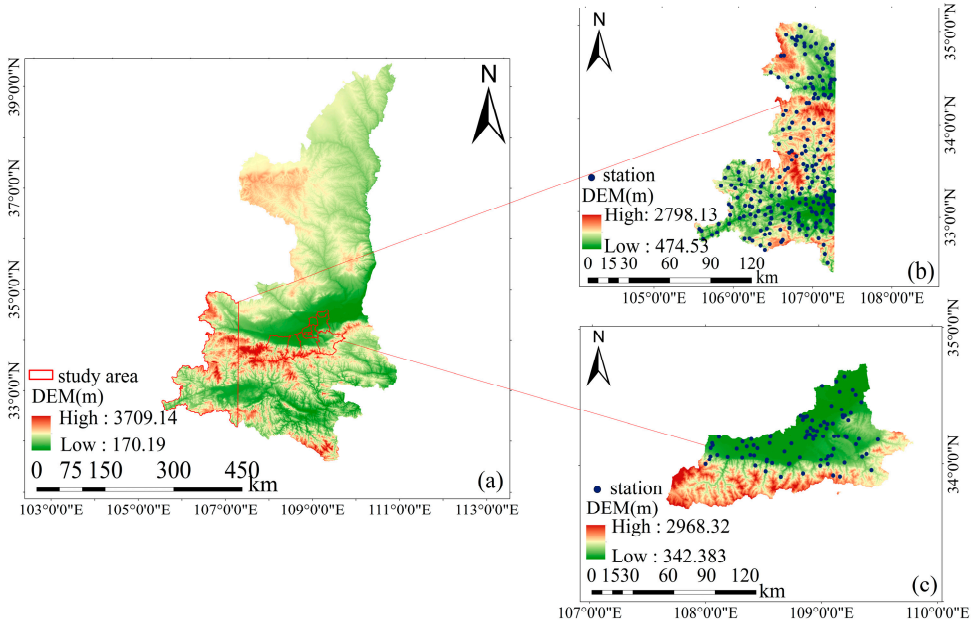

Figure 1. The study region: (**a**) location of Shaanxi Province; (**b**) location of the first study area; (**c**) location of the second study area. The points in subfigures (**b**,**c**) indicate the locations of in situ stations.

2.2. The Data Sources

This study primarily utilized four types of LST datasets, hourly near-surface air temperature observations from meteorological stations, and a series of auxiliary datasets to support the hourly estimation of NSAT.

2.2.1. LST Data Sources

(1) MODIS LST Data

Given that this study focuses on high-accuracy estimation of NSAT at the hourly scale, which is highly sensitive to the diurnal variation of LST, the MOD11A1 daily product (from the Terra satellite) was selected. This product provides two observations per day at approximately 02:30 and 14:30 UTC [32,33]. Although the daily product is more susceptible to instantaneous cloud contamination compared with the 8-day composite product, its advantage lies in its temporal resolution, which is crucial for capturing the diurnal variation patterns of air temperature. Before performing NSAT estimation, a sinusoidal function [34,35] was used to interpolate MODIS LST to an hourly resolution.

(2) ERA5-Land LST Data

ERA5-Land is a global dataset released by the European Centre for Medium-Range Weather Forecasts (ECMWF). It provides data at a spatial resolution of 0.1° and a temporal resolution of 1 h, covering the period from 1950 to the present. Based on the ERA5 reanalysis, ERA5-Land incorporates specific enhancements for land surface hydrological and energy exchange processes, thereby offering improved representations of near-surface variables such as air temperature and soil moisture [25]. In this study, ERA5-Land is used as one source of LST. Additionally, hourly meteorological variables, including land surface temperature, downward radiation, 10 m wind speed, precipitation, and specific humidity, were extracted from ERA5-Land as auxiliary inputs for the air temperature estimation model.

(3) FY-2F LST Data

The Fengyun-2F (FY-2F) satellite is a geostationary meteorological satellite independently developed by China, capable of providing high-frequency observations over fixed regions (https://satellite.nsmc.org.cn/DataPortal/cn/home/index.html, accessed on 27 April 2025). FY-2F currently offers a variety of remote sensing–derived products, including those related to land surface, atmosphere, ocean, and radiation. The LST product from FY-2F has a spatial resolution of approximately 5 km and temporal coverage from 2012 to 2022. This product provides hourly observations that fully capture the diurnal variation in land surface temperature, thereby supplying high–temporal-resolution inputs for the hourly air temperature estimation model.

(4) CGLS LST Data

The Copernicus Global Land Service (CGLS) is a core component of the European Union's Copernicus program, providing a range of land surface datasets, including LST products. To maintain temporal consistency with the FY-2F dataset, this study utilized the "Land Surface Temperature 2010–2021 (raster 5 km), global, hourly—v1" dataset (https://doi.org/10.2909/90ca3e33-7926-4f71-857b-7336818cbd23, accessed on 20 May 2025) from CGLS. This product integrates observations from multiple geostationary meteorological satellites [36], offering a spatial resolution of approximately 5 km and continuous hourly global coverage from 2010 to 2021. It provides crucial temporal dynamics for constructing the hourly NSAT estimation model.

Although additional LST products are available, this study deliberately selected MODIS, ERA5-Land, FY-2F, and CGLS to ensure representativeness, complementarity, and suitability for hourly NSAT estimation. These four datasets cover different LST generation mechanisms and observation platforms, including polar-orbiting satellite retrievals (MODIS), geostationary satellite observations (FY-2F and CGLS), and land surface reanalysis products (ERA5-Land).

MODIS provides a widely validated benchmark LST product with high accuracy, ERA5-Land offers physically consistent land–surface thermal information, FY-2F supplies high-frequency regional observations, and CGLS integrates multi-satellite geostationary data to achieve continuous hourly global coverage. Together, these datasets represent the main operational LST sources currently used in regional and large-scale air temperature estimation studies, enabling a systematic and comparable assessment under a unified modeling framework.

2.2.2. Auxiliary Variables

The selection of auxiliary variables in this study follows established practices in previous air temperature estimation studies, where topographic, meteorological, and anthropogenic factors have been widely demonstrated to be effective predictors of NSAT [17,18,20,23]. The auxiliary variables were primarily used to support the estimation of NSAT based on LST. Most of these data were obtained from multi-source remote sensing and reanalysis datasets available on the GEE platform. The employed auxiliary variables can be categorized into four groups:

(1) Meteorological Variables:

These were derived from the ECMWF ERA5-Land hourly reanalysis dataset and include downward longwave radiation, downward shortwave radiation, wind speed, precipitation rate, and specific humidity.

(2) Land Surface Feature Variables:

The Enhanced Vegetation Index (EVI) from the MOD13A2 (Collection 6) product was employed, which has an original spatial resolution of 1 km and a temporal resolution of 16 days. Additionally, elevation and slope were derived from the SRTM digital elevation model, and population density was obtained from the GPWv4.11 dataset.

(3) Geographical Variables:

In this study, longitude and latitude were included as spatial coordinate variables to characterize the spatial distribution of air temperature.

(4) Temporal Identifier Variables:

These include year, month, day, and hour, which are used to capture the temporal variation patterns of air temperature.

For the objectives of this study, multiple datasets, including ERA5-Land reanalysis data, MODIS LST products, SRTM elevation data, and population density data, were efficiently acquired and processed using the GEE platform, avoiding the challenges associated with offline data processing and storage. All variables were spatially and temporally matched within GEE, and prior to use in the estimation models, all data were preprocessed to a uniform spatial resolution of 1 km and temporal resolution of 1 h. Specifically, MODIS LST data were interpolated to an hourly resolution using the sinusoidal function model.

In addition, EVI data were linearly interpolated to a daily resolution, and for hourly modeling, the same daily EVI value was assigned to all hours within the corresponding day. The ERA5-Land data were resampled to 1 km spatial resolution using bilinear interpolation, while FY-2F and CGLS data were linearly downscaled to 1 km spatial resolution. The data processing period covered 3–5 August and 7–10 August, as these periods were available for all data types used in this study.

2.2.3. Meteorological Station Data

The hourly near-surface air temperature data used in this study were obtained from the China Meteorological Administration. This dataset primarily includes variables such as wind speed, wind direction, precipitation, air temperature, relative humidity, and air pressure. For the study, station observations of NSAT were selected within the southwestern region of Shaanxi Province and Xi'an City (as shown in Figure 1), corresponding to the temporal coverage of the available LST datasets described above.

Due to partial data gaps in the FY-2F land surface temperature product caused by factors such as cloud cover and satellite malfunctions, the analysis period was adjusted to ensure temporal consistency across multiple datasets and to maintain data completeness. The final study period was set from 05:00 UTC on 3 August to 11:00 UTC on 5 August and from 13:00 UTC on 7 August to 23:00 UTC on 10 August. After spatiotemporal filtering and quality control, a total of 30,269 and 11,753 high-quality hourly air temperature observations were obtained for study areas 1 and 2, respectively, and were used as ground truth data for model training and accuracy validation.

2.3. NSAT Estimation Models

Five representative machine learning models were selected in this study for air temperature estimation. The basic principles of each model, the rationale for their selection, and the corresponding model configurations are described below.

2.3.1. Mixture Density Network

The MDN is a framework that combines deep neural networks with probabilistic modeling [37]. In this model, the neural network is used to learn the parameters of a

conditional probability distribution, such as a mixture of Gaussian distributions, making it suitable for regression problems with multimodal or uncertain outputs. The MDN provides probabilistic predictions, allowing the quantification of prediction uncertainty, and the mean of the predicted distribution is taken as the estimated value.

In this study, the MDN consisted of two fully connected hidden layers, each containing 256 neurons with ReLU activation functions. The output layer predicted the parameters of a Gaussian mixture model, including mixture weights, means, and variances. The number of Gaussian components was set to three. The model was trained using the Adam optimizer, and the negative log-likelihood of the mixture distribution was used as the loss function.

2.3.2. Random Forest

RF is an ensemble learning algorithm widely used for both remotely sensed regression and classification tasks [38,39]. The model constructs multiple decision trees and aggregates their predictions, using the mean for regression tasks, to effectively reduce the risk of overfitting. Previous studies have demonstrated that RF performs robustly when handling high-dimensional features and complex nonlinear relationships, making it well suited for air temperature estimation involving multi-source data [18,20].

In this study, the RF model was implemented using 1500 decision trees. The maximum depth of each tree was limited to 20 to avoid overfitting, and the minimum number of samples required at leaf nodes was set to two. The square root of the total number of features was used for feature selection at each split.

2.3.3. Deep Neural Network

DNN is a classical deep learning model that extracts high-level features from data through multiple layers of nonlinear transformations and is widely used for regression and classification tasks [40,41]. Studies have shown that DNNs can effectively capture the complex nonlinear relationships between meteorological variables and air temperature [18]. However, their performance is highly dependent on the network architecture, hyperparameter tuning, and the size of the training dataset.

The DNN architecture adopted in this study consisted of three hidden layers with 256, 128, and 64 neurons, respectively, all using ReLU activation functions. To enhance model generalization, L2 regularization was applied to the weights of the first two hidden layers, and dropout layers with a dropout rate of 0.2 were incorporated. The network was trained using the Adam optimizer with a learning rate of 0.001 and mean squared error as the loss function.

2.3.4. Extreme Gradient Boosting

XGBoost is an efficient gradient boosting decision tree algorithm [42]. The model constructs an ensemble of weak decision trees (i.e., CART regression trees) in a sequential manner, where each subsequent tree is trained to fit the residuals of the previous tree, thereby progressively reducing prediction errors. XGBoost incorporates a regularization term in its objective function to control model complexity, effectively preventing overfitting. It also supports parallel processing, offering excellent computational efficiency and superior predictive performance on large datasets. Studies have shown that XGBoost performs robustly when handling structured data and complex feature interactions, and it has been widely applied with notable success in data science competitions and environmental modeling tasks, including air temperature estimation [28,43].

In this study, the XGBoost model was configured with 1500 boosting trees and a learning rate of 0.03. The maximum tree depth was set to eight, and both row subsampling and column subsampling ratios were set to 0.8. L1 and L2 regularization terms were applied to further improve model robustness.

2.3.5. Geospatial-Temporal Neural Network Weighted Regression

GTNNWR is a hybrid model that integrates neural networks with geographically weighted regression to better capture the spatiotemporal non-stationarity among variables [29]. In GTNNWR, a spatiotemporal proximity neural network adaptively learns sample weights, replacing the fixed kernel function used in traditional geographically weighted regression, thereby allowing more flexible modeling of spatiotemporal dependencies. In this study, GTNNWR was employed to better represent the complex relationships between meteorological drivers and air temperature across different locations and time periods, improving the accuracy of hourly air temperature estimation.

In this study, the GTNNWR model was implemented using a neural-network-based spatiotemporal weighting framework. Specifically, the spatiotemporal weighting network consisted of one hidden layer with three neurons, while the main regression network adopted a three-layer fully connected architecture with 256, 128, and 64 neurons, respectively. Dropout regularization with a rate of 0.3 was applied to enhance model generalization. Model training was conducted using the Adadelta optimizer combined with a multi-step learning rate scheduler, where the initial learning rate was set to 0.01 and decayed at predefined training milestones.

2.4. Accuracy Assessment of NSAT Estimates

To comprehensively evaluate the estimation performance of the different models, this study employed the coefficient of determination (R^2), root mean square error (*RMSE*), and mean absolute error (*MAE*) as evaluation metrics. The formulas for each metric are as Equations (1)–(3).

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(\widehat{y_i} - y_i)^2}{\sum_{i=1}^{m}(\overline{y_i} - y_i)^2}, \quad (1)$$

$$RSME = \sqrt{(\frac{1}{m}\sum_{i=1}^{m}(y_i - \widehat{y_i})^2)}, \quad (2)$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|y_i - \widehat{y_i}| \quad (3)$$

where y_i denotes the observed value of the i-th sample, $\hat{y}_i$ represents the model-predicted value for the i-th sample, $\overline{y_i}$ is the mean of all observed values, and m is the total number of samples.

The R^2 measures the strength of the linear relationship between the predicted and observed values. Both the *RMSE* and *MAE* quantify the magnitude of prediction errors, where smaller values indicate higher model accuracy. *RMSE* is more sensitive to large errors and thus better captures extreme deviations in predictions, whereas *MAE* reflects the average absolute level of errors. To ensure the robustness and generalization of the evaluation, all metrics are calculated based on five-fold cross-validation results.

2.5. Experiment Design

The experimental design of this study consists of two main parts. To examine the regional applicability of the optimal scheme, experiments were conducted in both the southwestern region of Shaanxi and the Xi'an City. The overall workflow for these two objectives is illustrated in Figure 2.

(1) Comparison of LST combination schemes

This part aims to identify the optimal combination of input LST features. The auxiliary variables introduced in Section 2.2.2 were systematically combined with four types of LST data (MODIS, ERA5-Land, CGLS, and FY-2F) to construct a total of 15 variable combination schemes (as shown in Table 1). Considering computational efficiency, four representative

models (MDN, RF, DNN, and XGBoost) were employed to evaluate these combinations. Specifically, the mean metrics of the four models were used as indicators to assess the performance of different LST sources.

Table 1. Experimental design for testing the impact of LST sources for NSAT estimation.

Case Number	Sources of Input LST Data	Case Name
1	CGLS	C_1km
2	ERA-Land	E_1km
3	FY-2F	F_1km
4	MODIS	M_1km
5	CGLS + ERA-Land	CE_1km
6	CGLS + FY-2F	CF_1km
7	CGLS + MODIS	CM_1km
8	ERA-Land + FY-2F	EF_1km
9	ERA-Land + MODIS	EM_1km
10	FY-2F + MODIS	FM_1km
11	CGLS + ERA-Land + FY-2F	CEF_1km
12	CGLS + ERA-Land + MODIS	CEM_1km
13	CGLS + FY-2F + MODIS	CFM_1km
14	ERA-Land + FY-2F + MODIS	EFM_1km
15	CGLS + ERA-Land + FY-2F + MODIS	CEFM_1km

In addition, five-fold cross-validation was adopted, and the feature combination scheme with the highest accuracy was selected as the input set for subsequent model comparisons. All valid samples were randomly shuffled and divided into five subsets at the sample level, without explicitly separating spatial locations or temporal periods. In each fold, four subsets were used for model training and the remaining subset was used for validation. This random cross-validation strategy was adopted to evaluate the overall predictive performance of different LST combination schemes under mixed spatiotemporal conditions.

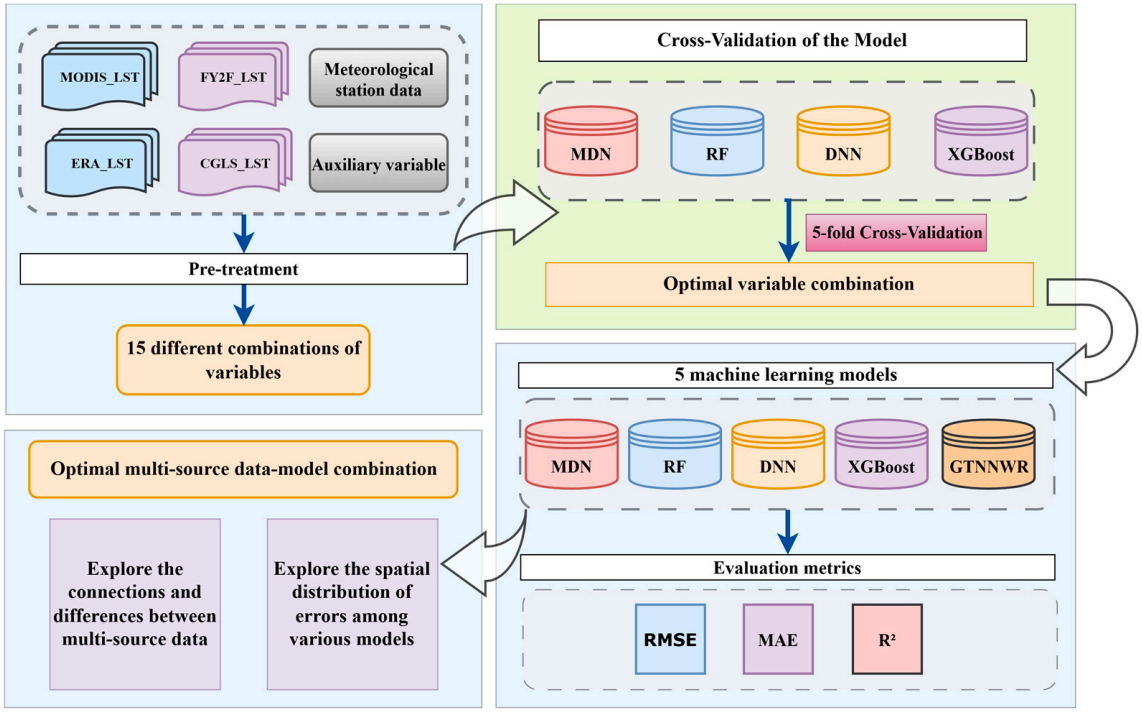

Figure 2. Flowchart of optimal "data source–model" combination selection.

(2) Comparison of model performance

Based on the optimal input LST scheme identified in the previous step, this part further compares the estimation performance of five machine learning models—MDN, RF, DNN, XGBoost, and GTNNWR. All models were trained and tested using the same dataset, and their performance was comprehensively evaluated using R^2, *RMSE*, and *MAE*. This analysis aims to identify the strengths, weaknesses, and applicability of each model in hourly NSAT estimation, as well as to assess the robustness of the optimal model across different regions.

3. Results and Analysis

3.1. Comparison of the Accuracy of NSAT Estimation Using Different LST Datasets

Table 2 presents the mean accuracy metrics (*RMSE*, *MAE*, and R^2) for the four estimation algorithms (RF, DNN, XGBoost, and MDN) in the southwestern region of Shaanxi Province (study area 1), providing a comprehensive assessment of how LST from different data sources affects hourly air temperature estimation. Table 3 reports the accuracy evaluation of LST data sources in Xi'an City, Shaanxi Province (study area 2).

Table 2. Estimated NSAT Accuracy Using Different LST Sources and Their Combinations in the Study area 1.

Case	*RMSE* (°C)	*MAE* (°C)	R^2
C_1km	0.9287	0.6806	0.9569
E_1km	0.9718	0.7127	0.9519
F_1km	0.9640	0.7033	0.9534
M_1km	0.9166	0.6700	0.9579
CE_1km	0.9445	0.6902	0.9553
CF_1km	0.9625	0.7049	0.9536
CM_1km	0.9417	0.6893	0.9556
EF_1km	0.9752	0.7106	0.9522
EM_1km	1.0031	0.7303	0.9490
FM_1km	0.9522	0.6964	0.9546
CEF_1km	1.0260	0.7514	0.9466
CEM_1km	0.9596	0.7017	0.9538
CFM_1km	0.9897	0.72205	0.9508
EFM_1km	1.0024	0.7315	0.9493
CEFM_1km	0.9654	0.7058	0.9532

Table 3. Estimated NSAT Accuracy Using Different LST Sources and Their Combinations in the Study area 2.

Case	*RMSE* (°C)	*MAE* (°C)	R^2
C_1km	0.7880	0.5756	0.9748
E_1km	0.7946	0.5777	0.9743
F_1km	0.7936	0.5763	0.9744
M_1km	0.8030	0.5831	0.9738
CE_1km	0.8040	0.5845	0.9737
CF_1km	0.8101	0.5910	0.9733
CM_1km	0.8132	0.5933	0.9732
EF_1km	0.8118	0.5899	0.9732
EM_1km	0.8058	0.5873	0.9736
FM_1km	0.8103	0.5903	0.9733
CEF_1km	0.8183	0.5966	0.9727

Table 3. *Cont.*

Case	*RMSE* (°C)	*MAE* (°C)	R^2
CEM_1km	0.8189	0.5975	0.9727
CFM_1km	0.8226	0.6015	0.9725
EFM_1km	0.8187	0.5972	0.9727
CEFM_1km	0.8322	0.6094	0.9718

It should be noted that the R^2 values exhibit only limited variation across different LST datasets and their combinations in both study areas. This is because R^2 primarily reflects the overall explanatory power of the full predictor set, which is dominated by predictors common to all LST schemes. In contrast, differences among LST datasets are more sensitively captured by error-based metrics such as *RMSE* and *MAE*, which better reflect the influence of LST quality and systematic biases on estimation accuracy.

3.1.1. Comparison of Prediction Performance Using LST from a Single Source

By evaluating the performance of the four LST datasets used independently for NSAT estimation (Tables 2 and 3), notable differences were observed across data sources. In the southwestern region of Shaanxi Province (Table 2), MODIS (M_1km case) performed the best, yielding the lowest *RMSE* (0.9166 °C) and *MAE* (0.6700 °C), along with the highest R^2 (0.9579). CGLS (C_1km) and FY-2F (F_1km) followed, while ERA5-Land (E_1km) showed the largest estimation error, with an *RMSE* of 0.9718 °C and the weakest model fit (R^2 is 0.9519). Accordingly, the performance ranking of individual LST data sources in southwestern Shaanxi is as follows: MODIS performed best, followed by CGLS and FY-2F, whereas ERA5-Land exhibited the lowest accuracy.

However, this ranking exhibits clear regional dependence. In Xi'an City (Table 3), CGLS (C_1km case) emerges as the best single-source dataset, performing slightly better than MODIS in this area. This finding indicates that the optimal choice of LST data source must account for local geographic and environmental conditions.

3.1.2. Comparison of Prediction Performance Using LST from Multi-Source Combinations

The comparison between using a single LST source (cases 1–4) and multi-source LST combinations (cases 5–15) shows that, in both the southwestern region of Shaanxi Province (Table 2) and Xi'an City (Table 3), incorporating additional LST variables into the best single-source dataset (M_1km for study area 1 and C_1km for study area 2) did not improve model estimation accuracy. Instead, it generally resulted in decreased performance.

In the southwestern region of Shaanxi Province (study area 1), all combination schemes exhibited higher *RMSE* and *MAE* values than the best single-source dataset (M_1km), while R^2 values were consistently lower (Table 2). Specifically, the combination of MODIS and ERA5-Land (EM_1km) increased *RMSE* by approximately 8.62%, from 0.9166 °C to 1.0031 °C. The combination of MODIS and CGLS (CM_1km) increased *RMSE* by about 9.44%, reaching 0.9417 °C. The combination of MODIS and FY-2F (FM_1km) increased *RMSE* by roughly 3.88%, reaching 0.9522 °C. The combination of CGLS, ERA5-Land and FY-2F (CEF_1km) showed the poorest performance, with an *RMSE* of 1.0260 °C.

A similar pattern was observed in Xi'an City (study area 2). Using the best single-source dataset (C_1km), as a baseline, all combination schemes resulted in reduced accuracy (Table 3). For example, the combination of CGLS and MODIS (CM_1km) increased *RMSE* from 0.7880 °C to 0.8132 °C. The full combination of all sources (CEFM_1km) reached an *RMSE* of 0.8322 °C, which was the worst performance among all schemes.

Overall, the results from both regions indicate that simple multi-source LST inputs do not provide additional effective information. Instead, they may introduce redundancy or conflicting systematic biases, which ultimately degrade model performance.

3.1.3. Determination of the Optimal LST Scheme

Based on the above analysis, selecting a high-quality single data source for a specific region represents the optimal strategy in this study. In the southwestern region of Shaanxi Province, MODIS (M_1km) serves as the best single-source dataset, with *RMSE*, *MAE*, and R^2 values of 0.9166 °C, 0.6700 °C, and 0.9579, respectively. Its superior performance is attributed to the relatively high spatial resolution of MODIS, which enables more accurate characterization of the complex land surface thermal conditions in this region. In contrast, for Xi'an City, CGLS (C_1km) exhibits better potential as a single data source, achieving an *RMSE* of 0.7880 °C and slightly outperforming MODIS in this area (*RMSE* of 0.8030 °C). These findings emphasize the regional dependence of optimal data source selection.

Results from both study areas indicate that naive feature stacking may introduce systematic biases or redundant information, ultimately reducing model performance. Therefore, selecting a high-quality single data source tailored to the specific region is more effective than directly combining multiple LST datasets. It should be noted that this conclusion is based on the datasets used in the present study. If LST datasets of higher quality with truly complementary characteristics become available, the results may differ.

3.2. Performance Comparison of Different Machine Learning Models for Hourly NSAT Estimation

To account for regional applicability, the final assessment used single-source MODIS LST for the southwestern region of Shaanxi and single-source CGLS LST for Xi'an City as model inputs, allowing for a comparative analysis of the estimation accuracy across the five machine learning models.

3.2.1. Performance Comparison of Different Machine Learning Models

In this study, key hyperparameters for each model were optimized using the grid search method until model performance stabilized. The optimized models were then used to systematically compare their predictive performance. A five-fold cross-validation was employed to evaluate the performance of all ML models.

Figure 3 presents the scatter plots and density distributions of estimated versus observed NSAT for the five models in the southwestern region of Shaanxi Province (study area 1), with the red line representing the 1:1 reference. The results reveal clear differences in predictive performance among the models. The MDN (Figure 3a) exhibits relatively dispersed predictions, with many points deviating from the 1:1 line, resulting in comparatively large errors (*RMSE* of 1.0560 °C, *MAE* of 0.7514 °C, and R^2 of 0.9453). Predictions from the RF model (Figure 3b) are more concentrated and generally closer to the observed values, indicating a better fit, while the DNN model (Figure 3c) shows similarly concentrated predictions with slightly improved accuracy (*RMSE* of 0.9326 °C, *MAE* of 0.6927 °C, and R^2 of 0.9573 for DNN, compared to *RMSE* of 1.0082 °C, *MAE* of 0.7476 °C, and R^2 of 0.9501 for RF). XGBoost (Figure 3d) further enhances the prediction accuracy, with predicted points distributed more tightly along the 1:1 line and fewer outliers, achieving *RMSE* of 0.6991 °C, *MAE* of 0.5022 °C, and R^2 of 0.9760. Among all models, GTNNWR (Figure 3e) provides the best performance, with predictions almost entirely aligned with the 1:1 line and the highest point density, achieving an *RMSE* of 0.3861 °C, an *MAE* of 0.2477 °C, and an R^2 of 0.9927, which highlights its clear advantage in capturing the nonlinear relationships among multiple factors for air temperature estimation.

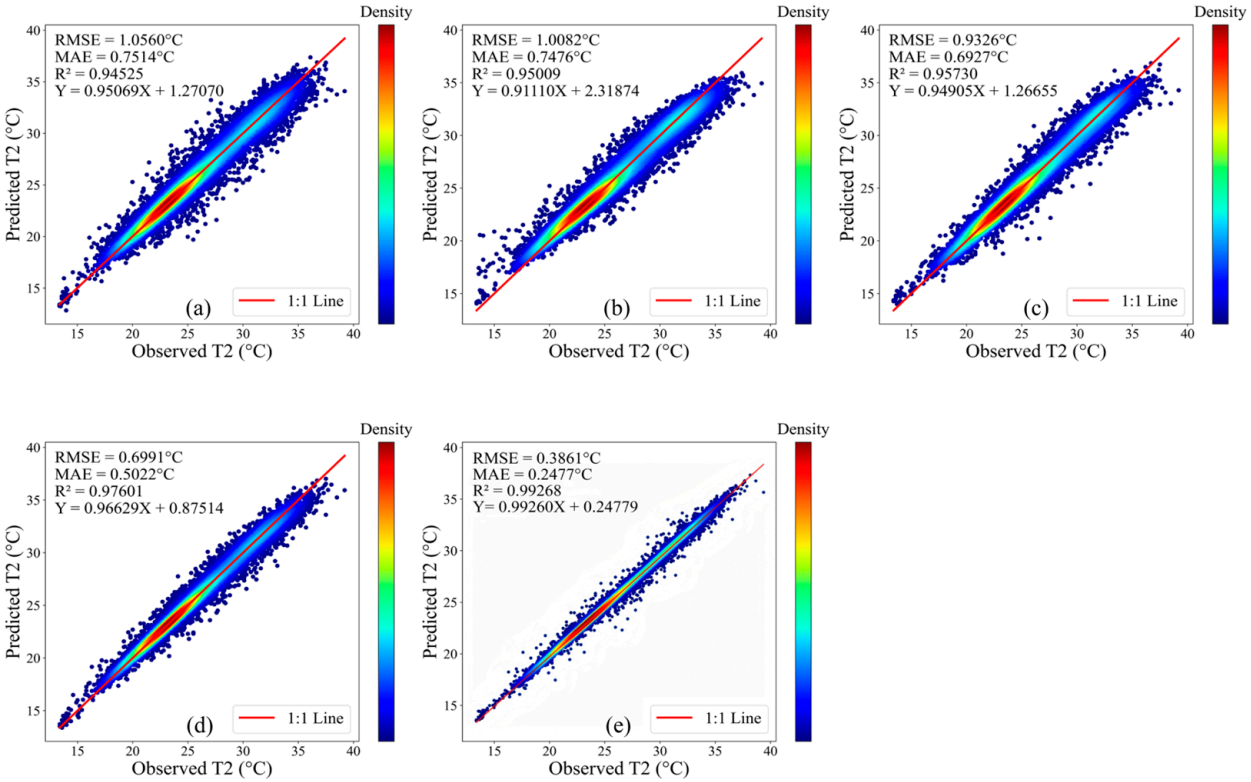

Figure 3. NSAT accuracy of five models based on single MODIS LST in Study Area 1. Subfigures (**a**–**e**) correspond to MDN, RF, DNN, XGBoost, and GTNNWR, respectively.

As shown in Figure 4, significant differences in predictive performance are also observed among the models for Xi'an City. The MDN model (Figure 4a) produces relatively concentrated predictions that are generally close to the observed values, with an *RMSE* of 0.8468 °C, an *MAE* of 0.6029 °C, and an R^2 of 0.9718. In contrast, the RF model (Figure 4b) predictions are more dispersed, with a larger number of points deviating from the 1:1 line, resulting in greater errors, with an *RMSE* of 1.0158 °C, an *MAE* of 0.7708 °C, and an R^2 of 0.9594. The DNN model (Figure 4c) generates concentrated predictions that fit the observations well, achieving an *RMSE* of 0.7768 °C, an *MAE* of 0.5849 °C, and an R^2 of 0.9763. XGBoost (Figure 4d) further improves prediction accuracy, with predicted points distributed more tightly along the 1:1 line and fewer outliers, resulting in an *RMSE* of 0.6426 °C, an *MAE* of 0.4626 °C, and an R^2 of 0.9838. Among all models, GTNNWR (Figure 4e) demonstrates the best performance, with predictions nearly perfectly aligned with the 1:1 line and the highest point density, achieving the lowest *RMSE* and *MAE* (0.3588 °C and 0.2032 °C, respectively) and an R^2 of 0.9949. These results are consistent with those observed in study area 1, confirming the superior capability of GTNNWR in capturing the nonlinear relationships among multiple factors for NSAT estimation.

Overall, the GTNNWR model demonstrates the best performance, with *RMSE* and *MAE* values substantially lower than those of the other models, as well as the highest coefficient of determination. This superior performance is likely due to GTNNWR's ability to simultaneously capture spatiotemporal heterogeneity and nonlinear relationships. XGBoost ranks second in predictive accuracy, outperforming the remaining three models. The DNN shows slightly lower regression performance compared with GTNNWR and XGBoost. RF offers advantages in handling nonlinear relationships and assessing variable importance, and it is relatively simple to implement; however, its predictive accuracy for hourly air temperature in this study is lower than that of the aforementioned models. The MDN performs the least well overall, as it does not achieve regression accuracy compa-

rable to the other models. In summary, the overall ranking of the five models in terms of predictive accuracy in this study is as follows: GTNNWR exhibits the highest accuracy, followed by XGBoost, DNN, RF, and MDN. Nevertheless, all models achieve relatively high performance, with mean absolute errors for hourly air temperature predictions remaining below 1 °C and coefficients of determination exceeding 0.9. Apart from GTNNWR, which shows a clear advantage, the differences in estimation accuracy among the other four models are relatively minor.

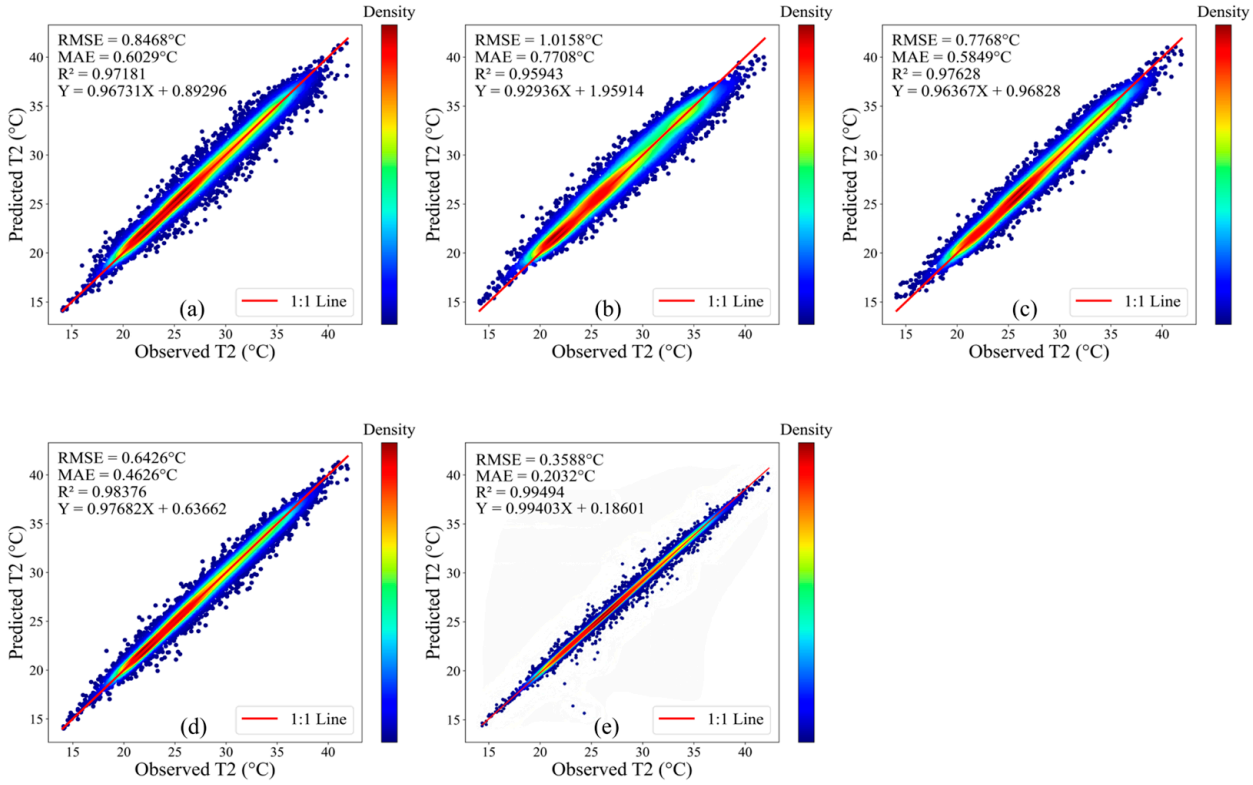

Figure 4. NSAT accuracy of five models based on single MODIS LST in Study Area 2. Subfigures (**a**–**e**) correspond to MDN, RF, DNN, XGBoost, and GTNNWR, respectively.

3.2.2. Spatial Distribution of *RMSE* for NSAT Estimates

Figures 5 and 6 present the spatial distribution of estimation errors for NSAT in the two study areas. In the southwestern region of Shaanxi Province (Figure 5), the GTNNWR model (Figure 5e) demonstrates the best performance, with approximately 100% of stations exhibiting *RMSE* values below 0.73 °C, indicating a clear advantage over the other models. XGBoost (Figure 5d) ranks second, with 88.20% of stations showing *RMSE* between 0.5 and 1.1 °C, and 11.80% of stations falling in the range of 0.38–0.5 °C. The DNN model (Figure 5c) shows similar performance to XGBoost, although 30.77% of stations exhibit *RMSE* values above 1.0 °C, with larger errors mainly occurring in the western and central high-altitude regions. For the RF model (Figure 5b), 69.23% of stations exhibit *RMSE* between 0.48 and 1.0 °C, while 30.77% exceed 1.0 °C, indicating a more dispersed error distribution compared to DNN. The MDN model (Figure 5a) shows an error distribution similar to RF, but the proportion of stations with *RMSE* above 1.0 °C is the highest, reaching 34.84%, reflecting weaker fitting capability in complex terrain.

In Xi'an City (Figure 6), GTNNWR remains the best-performing model, with *RMSE* values at all stations below 0.77 °C (Figure 6e), outperforming the other models. XGBoost (Figure 6d) and DNN (Figure 6c) follow, while RF (Figure 6b) and MDN (Figure 6a) exhibit slightly larger errors.

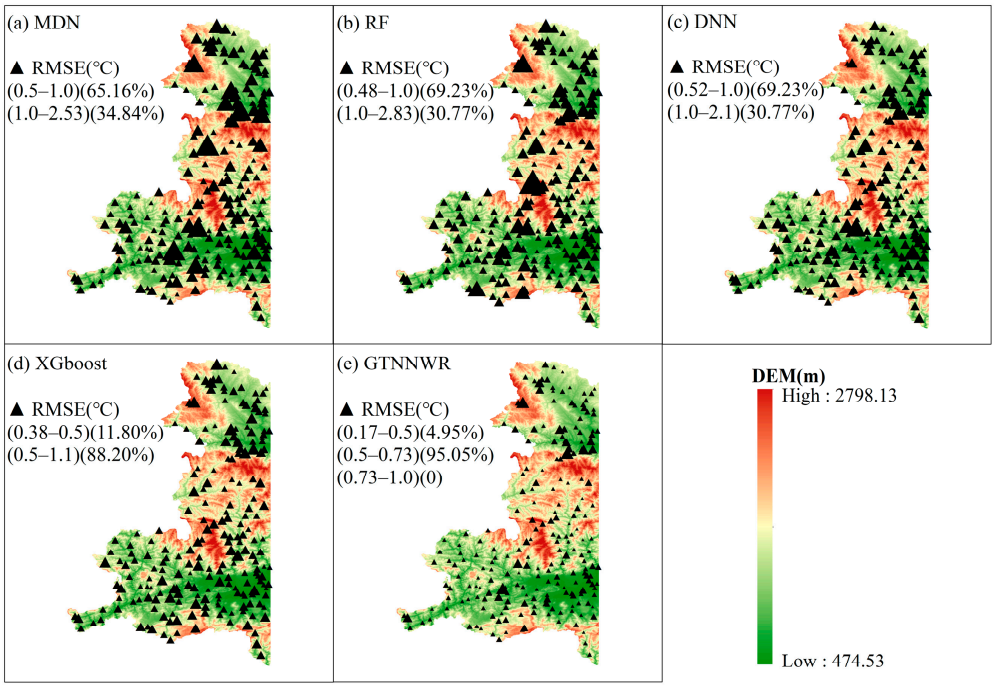

Figure 5. Spatial distribution of *RMSE* for NSAT estimates in Study area 1. Subfigures (**a**–**e**) correspond to MDN, RF, DNN, XGBoost, and GTNNWR, respectively. Black triangular markers indicate station-based *RMSE*, and larger triangles represent higher *RMSE* values.

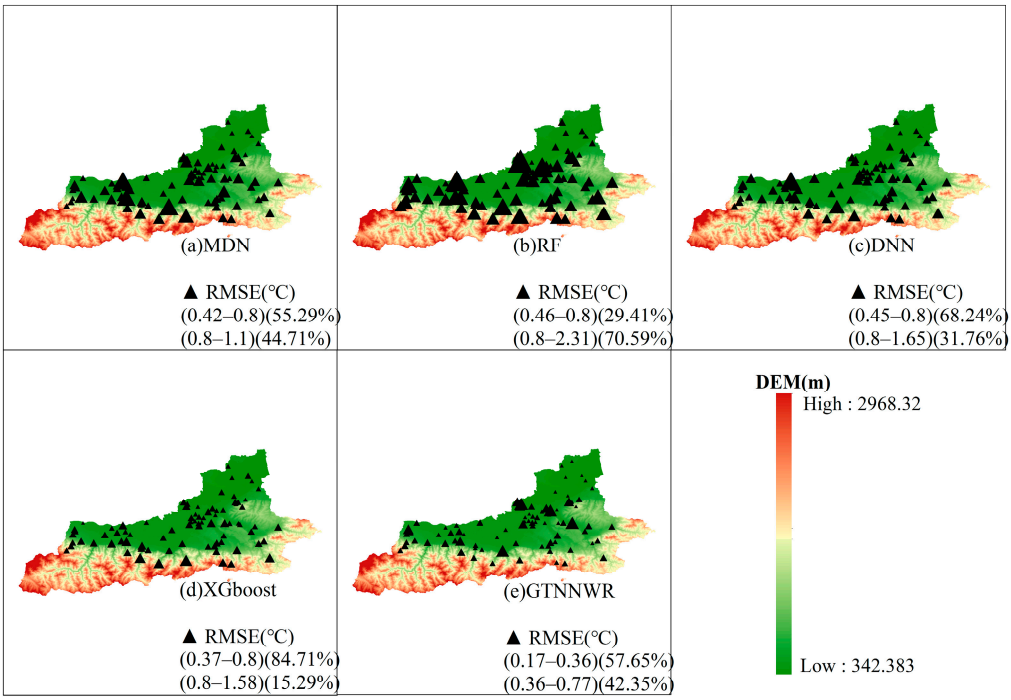

Figure 6. Spatial distribution of *RMSE* for NSAT estimates in Study area 2. Subfigures (**a**–**e**) correspond to MDN, RF, DNN, XGBoost, and GTNNWR, respectively. Black triangular markers indicate station-based *RMSE*, and larger triangles represent higher *RMSE* values.

Overall, the GTNNWR model consistently maintains high estimation accuracy across regions with different terrain and climatic conditions, demonstrating superior spatial generalization ability and stability, and significantly outperforming XGBoost, DNN, RF, and MDN.

4. Discussion

4.1. Impact of Multi-Source LST Stacking on NSAT Estimation

To further investigate the underlying cause of the observed phenomenon that a single high-quality data source outperforms multi-source LST stacking, we analyzed the relationships among different LST products (Figure 7). The results show that there are substantial systematic biases among these LST datasets. The discrepancies between any two LST products (e.g., MODIS, FY-2F, CGLS, ERA5-Land) are considerable, with pairwise *RMSE* values ranging from 4.79 °C to 8.64 °C. Although some pairs exhibit moderate linear correlations (e.g., FY-2F and CGLS with R of 0.63), most combinations yield negative coefficients of determination (R^2) and consistently high *RMSE* values. These results indicate that the different data sources are not providing complementary information but instead contain prominent and mutually inconsistent systematic deviations. This finding is consistent with previous studies showing that LST products derived from different satellite sensors frequently exhibit substantial systematic differences due to variations in overpass time, spectral response, retrieval algorithms, and viewing geometry [44,45]. When such biased datasets are jointly used as model inputs, they introduce additional errors. Therefore, using a single high-quality LST product (e.g., MODIS or CGLS) avoids the interference of systematic biases, offers a more reliable foundation for model training, and ultimately leads to higher air temperature estimation accuracy.

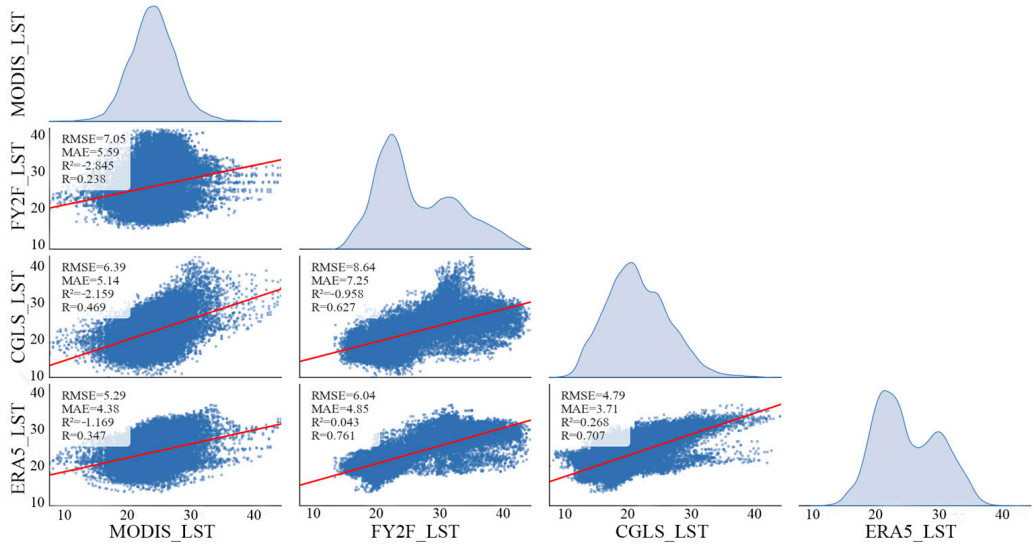

Figure 7. Analysis of Systematic Biases Among Different LST Data Sources.

4.2. Impact of Different LST Data Sources on Model Performance

The results presented in Tables 2 and 3 represent the averages across four machine learning models. To examine how different LST data sources (CGLS, ERA, FY, MODIS) affect the performance of individual models, we compared the *RMSE* values achieved by each model under different data sources (Figure 8).

In the southwestern Shaanxi region (Figure 8a), the *RMSE* differences across data sources are more pronounced for the MDN and RF models: ERA5-Land and FY-2F generally yield higher *RMSEs* than CGLS and MODIS, indicating that these models are more sensitive to data source resolution and retrieval errors. In contrast, the DNN and XGBoost models exhibit relatively small *RMSE* differences across data sources, demonstrating higher tolerance to variations in input features and maintaining stable predictive accuracy under multi-source data conditions.

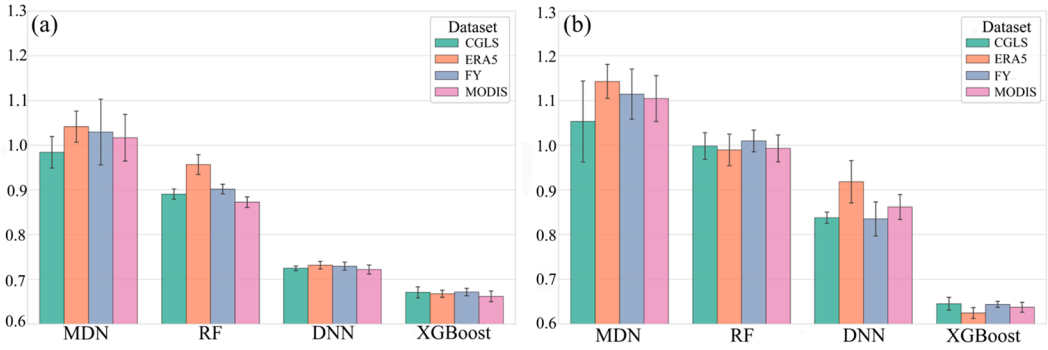

Figure 8. Accuracy errors of different models using various LST data sources, with (**a**) Study Area 1 and (**b**) Study Area 2.

In the Xi'an region (Figure 8b), the discrepancies among data sources become even more pronounced. For the DNN model, ERA5-Land produces significantly higher *RMSE* values than CGLS and FY-2F, implying that the lower spatial resolution and reanalysis characteristics of ERA5-Land may limit model accuracy in areas with complex land surface characteristics and pronounced urban thermal heterogeneity. The XGBoost model shows the smallest *RMSE* variation across the four data sources, further confirming its strong adaptability and generalization capability under heterogeneous data conditions.

Overall, LST data sources exert a noticeable yet model-dependent influence on performance. (1) For MDN and RF, model accuracy is strongly affected by data source quality and spatial resolution, leading to substantial performance fluctuations across data sources. (2) For XGBoost, data source differences have minimal impact, highlighting its strong stability across heterogeneous inputs. (3) For DNN, sensitivity to data sources is low in Region 1 (southwestern Shaanxi) but increases in Region 2 (Xi'an). Furthermore, models in the urban region (Xi'an) show higher sensitivity to data source selection, whereas those in the mountainous region (southwestern Shaanxi) exhibit relative consistency. This phenomenon may be attributed to the inherently complex urban thermal environment, in which both LST and NSAT tend to exhibit greater spatial and temporal variability [46,47].

5. Summary

This study systematically evaluated four land surface temperature datasets (MODIS, CGLS, FY-2F, and ERA5-Land) and five machine learning models (RF, MDN, DNN, XGBoost, and GTNNWR) for hourly near-surface air temperature estimation. The results show that, regarding data source selection, a single high-quality LST dataset outperforms multi-source LST stacking. MODIS performs best in the southwestern Shaanxi region, whereas CGLS shows clear advantages in the Xi'an metropolitan area, indicating that the optimal LST source is region-dependent. Multi-source LST combinations reduce model performance due to conflicting systematic biases among datasets. Regarding model performance, GTNNWR significantly outperforms the other models. GTNNWR achieves a reduction in *RMSE* of 44.8% and 44.2% in Study Area 1 (using MODIS) and Study Area 2 (using CGLS), respectively, compared with the second-best model. XGBoost consistently ranks second, followed by DNN and RF, while MDN performs comparatively weaker. GTNNWR maintains the highest accuracy across all data sources and both study regions, demonstrating strong adaptability and robustness. Although the remaining four models differ in accuracy, the differences are relatively small.

This study can provide practical guidance for selecting suitable LST sources and models for hourly NSAT estimation, thereby supporting the production and application of high spatiotemporal resolution temperature datasets. However, several limitations of this study should be acknowledged. In terms of computational efficiency, GTNNWR requires

substantially longer training and prediction time than the other four models. GTNNWR explicitly models spatiotemporal heterogeneity by constructing local spatiotemporal weight matrices and performing location-specific parameter estimation. Such operations involve intensive distance calculations and matrix computations, resulting in higher computational complexity. Although computationally demanding, GTNNWR consistently achieves the highest estimation accuracy across different LST sources and study regions.

Due to the memory limitations of GEE, this study selected two representative study areas—a mountainous region and an urban region—to characterize contrasting environmental conditions within Shaanxi Province. These two regions capture typical differences in topography, land cover, and land surface thermal characteristics that are common across the province. Nevertheless, both study areas are located in northwestern China. Therefore, the findings primarily reflect the model performance and data-source sensitivity under contrasting terrain and urbanization conditions and caution should be exercised when extrapolating the results to other climatic regions.

In addition, due to data availability constraints, the analysis was conducted using data from specific periods in the summer season only. Therefore, the generalization of the proposed strategy to other seasons remains to be further examined. Future work should extend the analysis to broader spatial domains and longer temporal periods to comprehensively assess the robustness and applicability of the proposed approach.

Author Contributions: Conceptualization, Z.Y. and M.G.; methodology, Z.Y. and M.G.; software, Z.Y.; validation, Z.Y., Y.C., R.L. and M.G.; formal analysis, Z.Y.; investigation, Z.Y. and M.G.; resources, Z.Y.; data curation, Z.Y.; writing—original draft preparation, Z.Y. and M.G.; writing—review and editing, Z.Y. and M.G.; visualization, Z.Y., Y.C. and R.L.; supervision, M.G.; project administration, M.G.; funding acquisition, M.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (No. 42471392), the Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China (No. JYB2025XDXM10), and the Fundamental Research Funds for the Central Universities, CHD (No. 300102265203).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Canadell, J.G.; Monteiro, P.M.; Costa, M.H.; Cotrim da Cunha, L.; Cox, P.M.; Eliseev, A.V.; Henson, S.; Ishii, M.; Jaccard, S.; Koven, C. Intergovernmental Panel on Climate Change (IPCC). Global Carbon and Other Biogeochemical Cycles and Feedbacks. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*; Cambridge University Press: Cambridge, UK, 2023; pp. 673–816. ISBN 1-009-15789-2.
2. Pinkerton, K.E.; Rom, W.N. (Eds.) *Climate Change and Global Public Health*; Respiratory Medicine; Springer International Publishing: Cham, Switzerland, 2021; ISBN 978-3-030-54745-5.
3. Girgibo, N.; Lü, X.; Hiltunen, E.; Peura, P.; Dai, Z. The Air Temperature Change Effect on Water Quality in the Kvarken Archipelago Area. *Sci. Total Environ.* **2023**, *874*, 162599. [CrossRef]
4. Yadav, N.; Rajendra, K.; Awasthi, A.; Singh, C.; Bhushan, B. Systematic Exploration of Heat Wave Impact on Mortality and Urban Heat Island: A Review from 2000 to 2022. *Urban Clim.* **2023**, *51*, 101622. [CrossRef]
5. Wang, B.; Gao, M.; Li, Y.; Xu, H.; Li, Z.; Peng, J. Spatiotemporal Trends of Extreme Temperature Events Along the Qinghai-Tibet Plateau Transportation Corridor From 1981 to 2019 Based on Estimated Near-Surface Air Temperature. *JGR Atmos.* **2023**, *128*, e2023JD039040. [CrossRef]

6. Han, J.; Fang, S.; Wang, X.; Zhuo, W.; Yu, Y.; Peng, X.; Zhang, Y. The Impact of Intra-Annual Temperature Fluctuations on Agricultural Temperature Extreme Events and Attribution Analysis in Mainland China. *Sci. Total Environ.* **2024**, *949*, 174904. [CrossRef] [PubMed]
7. Wang, D.; Zhang, P.; Chen, S.; Zhang, N. Adaptation to Temperature Extremes in Chinese Agriculture, 1981 to 2010. *J. Dev. Econ.* **2024**, *166*, 103196. [CrossRef]
8. Hrisko, J.; Ramamurthy, P.; Yu, Y.; Yu, P.; Melecio-Vázquez, D. Urban Air Temperature Model Using GOES-16 LST and a Diurnal Regressive Neural Network Algorithm. *Remote Sens. Environ.* **2020**, *237*, 111495. [CrossRef]
9. Lüthi, S.; Fairless, C.; Fischer, E.M.; Scovronick, N.; Armstrong, B.; Coelho, M.D.S.Z.S.; Guo, Y.L.; Guo, Y.; Honda, Y.; Huber, V.; et al. Rapid Increase in the Risk of Heat-Related Mortality. *Nat. Commun.* **2023**, *14*, 4894. [CrossRef]
10. Box, J.E. Key Indicators of Arctic Climate Change: 1971–2017. *Environ. Res. Lett.* **2019**, *14*, 045010. [CrossRef]
11. Shen, Z.; Shi, C.; Shen, R.; Tie, R.; Ge, L. Spatial Downscaling of Near-Surface Air Temperature Based on Deep Learning Cross-Attention Mechanism. *Remote Sens.* **2023**, *15*, 5084. [CrossRef]
12. Liu, S.; Su, H.; Tian, J.; Wang, W. An Analysis of Spatial Representativeness of Air Temperature Monitoring Stations. *Theor. Appl. Clim.* **2018**, *132*, 857–865. [CrossRef]
13. Wang, W.; Brönnimann, S.; Zhou, J.; Li, S.; Wang, Z. Near-Surface Air Temperature Estimation for Areas with Sparse Observations Based on Transfer Learning. *ISPRS J. Photogramm. Remote Sens.* **2025**, *220*, 712–727. [CrossRef]
14. Gao, M.; Xu, H.; Tan, Z.; Li, Z.; Yang, G. A 40-Year 1-Km Daily Seamless Near-Surface Air Temperature Product Over Yellow River Basin of China. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 7433–7446. [CrossRef]
15. Benali, A.; Carvalho, A.C.; Nunes, J.P.; Carvalhais, N.; Santos, A. Estimating Air Surface Temperature in Portugal Using MODIS LST Data. *Remote Sens. Environ.* **2012**, *124*, 108–121. [CrossRef]
16. He, Q.; Wang, M.; Liu, K.; Li, K.; Jiang, Z. GPRChinaTemp1km: A High-Resolution Monthly Air Temperature Data Set for China (1951–2020) Based on Machine Learning. *Earth Syst. Sci. Data* **2022**, *14*, 3273–3292. [CrossRef]
17. Shen, H.; Jiang, Y.; Li, T.; Cheng, Q.; Zeng, C.; Zhang, L. Deep Learning-Based Air Temperature Mapping by Fusing Remote Sensing, Station, Simulation and Socioeconomic Data. *Remote Sens. Environ.* **2020**, *240*, 111692. [CrossRef]
18. Gao, M.; Li, Z.; Tan, Z.; Li, H.; Peng, J. Use of Google Earth Engine to Generate a 20-Year 1 Km × 1 Km Monthly Air Temperature Product Over Yellow River Basin. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 10079–10090. [CrossRef]
19. He, T.; Liu, F.; Wang, A.; Fei, Z. Estimating Monthly Surface Air Temperature Using MODIS LST Data and an Artificial Neural Network in the Loess Plateau, China. *Chin. Geogr. Sci.* **2023**, *33*, 751–763. [CrossRef]
20. Wang, C.; Bi, X.; Luan, Q.; Li, Z. Estimation of Daily and Instantaneous Near-Surface Air Temperature from MODIS Data Using Machine Learning Methods in the Jingjinji Area of China. *Remote Sens.* **2022**, *14*, 1916. [CrossRef]
21. Wang, M.; Wei, J.; Wang, X.; Luan, Q.; Xu, X. Reconstruction of All-Sky Daily Air Temperature Datasets with High Accuracy in China from 2003 to 2022. *Sci. Data* **2024**, *11*, 1133. [CrossRef]
22. Su, Q.; Wang, Y.; Yang, Y.; Zhou, Y.; Wan, B.; Zong, L.; Li, T.; Zhong, T.; Lu, Z.; Xie, Z.; et al. Space-Time Deep Hybrid Boosting Learning for Investigating Day-Night Hourly Seamless Air Temperature Distribution from FY-4A over China. *J. Hydrol.* **2025**, *663*, 134245. [CrossRef]
23. Zhou, S.; Wang, Y.; Yuan, Q. Estimation of Hourly Air Temperature in China Based on LightGBM and Himawari-8. In Proceedings of the IGARSS 2022—2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17 July 2022; IEEE: Washington, DC, USA; pp. 6558–6561.
24. Fang, S.; Mao, K.; Xia, X.; Wang, P.; Shi, J.; Bateni, S.M.; Xu, T.; Cao, M.; Heggy, E.; Qin, Z. Dataset of Daily Near-Surface Air Temperature in China from 1979 to 2018. *Earth Syst. Sci. Data* **2022**, *14*, 1413–1432. [CrossRef]
25. Muñoz-Sabater, J.; Dutra, E.; Agustí-Panareda, A.; Albergel, C.; Arduini, G.; Balsamo, G.; Boussetta, S.; Choulga, M.; Harrigan, S.; Hersbach, H.; et al. ERA5-Land: A State-of-the-Art Global Reanalysis Dataset for Land Applications. *Earth Syst. Sci. Data* **2021**, *13*, 4349–4383. [CrossRef]
26. Maionchi, D.D.O.; Silva, J.G.D.; Balista, F.A.; Junior, W.A.M.; Paulo, S.R.D.; Paulo, I.J.C.D.; Biudes, M.S. Estimating Hourly Air Temperature in an Amazon-Cerrado Transitional Forest in Brazil Using Machine Learning Regression Models. *Theor. Appl. Clim.* **2024**, *155*, 7827–7843. [CrossRef]
27. Li, B.; Liang, S.; Ma, H.; Liu, X.; He, T.; Zhang, Y. Generation of Global 1 Km All-Weather Instantaneous and Daily Mean Land Surface Temperature from MODIS Data; ESSD—Global/Energy and Emissions. *Earth Syst. Sci. Data* **2024**, *16*, 3795–3819. [CrossRef]
28. Sebbar, B.; Khabba, S.; Merlin, O.; Simonneaux, V.; Hachimi, C.E.; Kharrou, M.H.; Chehbouni, A. Machine-Learning-Based Downscaling of Hourly ERA5-Land Air Temperature over Mountainous Regions. *Atmosphere* **2023**, *14*, 610. [CrossRef]
29. Wu, S.; Wang, Z.; Du, Z.; Huang, B.; Zhang, F.; Liu, R. Geographically and Temporally Neural Network Weighted Regression for Modeling Spatiotemporal Non-Stationary Relationships. *Int. J. Geogr. Inf. Sci.* **2021**, *35*, 582–608. [CrossRef]
30. Gao, M.; Ma, J.; Li, Z.; Yang, G. Evolution of Vegetation Phenology and Its Response to Urbanisation in Shaanxi Province, China: A Fine-Scale Perspective. *J. Geogr. Sci.* **2025**, *35*, 1024–1048. [CrossRef]

31. Zhang, X.; Steeneveld, G.-J.; Zhou, D.; Ronda, R.J.; Duan, C.; Koopmans, S.; Holtslag, A.A.M. Modelling Urban Meteorology with Increasing Refinements for the Complex Morphology of a Typical Chinese City (Xi'an). *Build. Environ.* **2020**, *182*, 107109. [CrossRef]
32. Duan, S.-B.; Li, Z.-L.; Li, H.; Göttsche, F.-M.; Wu, H.; Zhao, W.; Leng, P.; Zhang, X.; Coll, C. Validation of Collection 6 MODIS Land Surface Temperature Product Using in Situ Measurements. *Remote Sens. Environ.* **2019**, *225*, 16–29. [CrossRef]
33. Wan, Z. *MODIS Land Surface Temperature Products Users' Guide*; Institute for Computational Earth System Science, University of California: Santa Barbara, CA, USA, 2006; Volume 805, p. 26.
34. Mu, Q.; Jones, L.A.; Kimball, J.S.; McDonald, K.C.; Running, S.W. Satellite Assessment of Land Surface Evapotranspiration for the pan-Arctic Domain. *Water Resour. Res.* **2009**, *45*, W09420. [CrossRef]
35. Spano, D.; Duce, P.; Snyder, R.L.; Cesaraccio, C. An Improved Model for Determining Degree-Day Values from Daily Temperature Data. *Int. J. Biometeorol.* **2001**, *45*, 161–169. [CrossRef] [PubMed]
36. Freitas, S.C.; Trigo, I.F.; Macedo, J.; Barroso, C.; Silva, R.; Perdigão, R. Land Surface Temperature from Multiple Geostationary Satellites. *Int. J. Remote Sens.* **2013**, *34*, 3051–3068. [CrossRef]
37. Bishop, C.M. Mixture Density Networks. In *Neural Computing Research Group Report NCRG/94/004*; Aston University: Birmingham, UK, 1994.
38. Belgiu, M.; Drăguţ, L. Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
39. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
40. Samek, W.; Montavon, G.; Lapuschkin, S.; Anders, C.J.; Müller, K.-R. Explaining Deep Neural Networks and beyond: A Review of Methods and Applications. *Proc. IEEE* **2021**, *109*, 247–278. [CrossRef]
41. Sze, V.; Chen, Y.-H.; Yang, T.-J.; Emer, J.S. Efficient Processing of Deep Neural Networks: A Tutorial and Survey. *Proc. IEEE* **2017**, *105*, 2295–2329. [CrossRef]
42. Chen, T.; Guestrin, C. Xgboost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
43. Carrión, D.; Arfer, K.B.; Rush, J.; Dorman, M.; Rowland, S.T.; Kioumourtzoglou, M.-A.; Kloog, I.; Just, A.C. A 1-Km Hourly Air-Temperature Model for 13 Northeastern U.S. States Using Remotely Sensed and Ground-Based Measurements. *Environ. Res.* **2021**, *200*, 111477. [CrossRef]
44. Adeniran, I.A.; Zhu, R.; Yang, J.; Zhu, X.; Wong, M.S. Cross-Comparison between Sun-Synchronized and Geostationary Satellite-Derived Land Surface Temperature: A Case Study in Hong Kong. *Remote Sens.* **2022**, *14*, 4444. [CrossRef]
45. Guillevic, P.; Göttsche, F.; Nickeson, J.; Hulley, G.; Ghent, D.; Yu, Y.; Trigo, I.; Hook, S.; Sobrino, J.A.; Remedios, J. Land Surface Temperature Product Validation Best Practice Protocol. Version 1.1. *Best Practice for Satellite-Derived Land Product Validation*. 2018, 60p. Available online: https://scholar.google.com.hk/scholar?hl=zh-TW&newwindow=1&as_sdt=0%2C5&q=Land+Surface+Temperature+Product+Validation+Best+Practice+Protocol.+&btnG= (accessed on 7 January 2026).
46. Xu, Z.; Yi, Z.; Wang, Y.; Wang, D.; Zhang, L.; Huo, H. Estimating Near-Surface Air Temperature in Urban Functional Zones in China Using Spatial-Temporal Attention. *Build. Environ.* **2025**, *276*, 112860. [CrossRef]
47. Mitraka, Z.; Chrysoulakis, N.; Doxani, G.; Del Frate, F.; Berger, M. Urban Surface Temperature Time Series Estimation at the Local Scale by Spatial-Spectral Unmixing of Satellite Observations. *Remote Sens.* **2015**, *7*, 4139–4156. [CrossRef]