

Research Article

Comparison of Machine-Learning Algorithms for Near-Surface Air-Temperature Estimation from FY-4A AGRI Data

Ke Zhou,¹ Hailei Liu ,^{1,2} Xiaobo Deng,^{1,3} Hao Wang ,¹ and Shenglan Zhang¹

¹Key Laboratory of Atmospheric Sounding, Chengdu University of Information Technology, Chengdu 610225, China

²National Satellite Meteorology Center, China Meteorological Administration, Beijing 100081, China

³Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science & Technology, Nanjing 210044, China

Correspondence should be addressed to Hailei Liu; liuhailei@cuit.edu.cn

Received 21 April 2020; Revised 11 August 2020; Accepted 22 September 2020; Published 6 October 2020

Academic Editor: Stefania Bonafoni

Copyright © 2020 Ke Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Six machine-learning approaches, including multivariate linear regression (MLR), gradient boosting decision tree, k -nearest neighbors, random forest, extreme gradient boosting (XGB), and deep neural network (DNN), were compared for near-surface air-temperature (T_{air}) estimation from the new generation of Chinese geostationary meteorological satellite Fengyun-4A (FY-4A) observations. The brightness temperatures in split-window channels from the Advanced Geostationary Radiation Imager (AGRI) of FY-4A and numerical weather prediction data from the global forecast system were used as the predictor variables for T_{air} estimation. The performance of each model and the temporal and spatial distribution of the estimated T_{air} errors were analyzed. The results showed that the XGB model had better overall performance, with R^2 of 0.902, bias of -0.087°C , and root-mean-square error of 1.946°C . The spatial variation characteristics of the T_{air} error of the XGB method were less obvious than those of the other methods. The XGB model can provide more stable and high-precision T_{air} for a large-scale T_{air} estimation over China and can serve as a reference for T_{air} estimation based on machine-learning models.

1. Introduction

Air temperature (T_{air}) is one of the basic meteorological observation parameters [1–3] and is of great concern in scientific disciplines like hydrology, meteorology, and environmental science. Furthermore, it influences most land-surface processes, such as photosynthesis and land-surface evapotranspiration [4]. Obtaining high-resolution T_{air} data can reduce human health risks and promote urban heat island research, so high-resolution T_{air} information is quite crucial [5, 6]. The summer T_{air} value in China is generally above 20°C , except in the high-altitude regions (e.g., Qinghai-Tibet Plateau). Summer heat waves have a major impact on agricultural food production, as well as the use of water and electricity [7]. This study focuses on the issue of summer T_{air} estimation in China using Advanced Geostationary Radiation Imager (AGRI) data.

Large-scale T_{air} data are mainly obtained by interpolation from the data collected by surface meteorological stations.

However, the distribution of meteorological stations is usually uneven due to geographical factors, and some sparsely populated areas even have no meteorological observation [8]. Therefore, the accuracy of the interpolated T_{air} data is limited, and researchers are unable to obtain high-spatial-resolution T_{air} information [9].

Meteorological satellites such as low-Earth-orbit (LEO) satellites and geostationary-Earth-orbit (GEO) satellites can provide continuous surface (i.e., land-surface temperature (LST)) and atmospheric observations with a wide spatial coverage at global and regional scales [10–12]. In the last several decades, LEO and GEO observations have been gradually applied to T_{air} estimation with the development of meteorological satellite technology. LEO satellites can only acquire data once or twice a day for one place. In addition, cloud contamination will reduce the effective data for T_{air} estimation [13–15]. Unlike LEO satellites, GEO meteorological satellites can continuously provide data every 15 or 30 min on

one-third of the Earth's surface [16–20]. Therefore, GEO satellites comprise an effective method of obtaining high-spatial- and high-temporal-resolution T_{air} data in a fixed area and have the potential to facilitate the study on the daily change of T_{air} [20, 21].

At present, the methods for T_{air} estimation from satellite brightness temperatures (BTs) and land-surface temperature (LST) product data can be divided into simple linear, multivariate linear, and nonlinear approaches [21, 22]. Previous studies [7, 23, 24] have shown that machine-learning algorithms can obtain higher-accuracy T_{air} values than those in other methods. For example, a machine-learning model (e.g., a neural network model (NN)) has higher accuracy, and the root-mean-square error (RMSE) is reduced by 1.29°C compared with linear models [7].

The AGRI aboard Fengyun-4A (FY-4A) has 14 spectral bands [18, 20, 25, 26]—six visible/near-infrared (VIS/NIR), six infrared (IR), and two water vapor bands—with a temporal resolution of 15 min for the full disk and a spatial resolution of 4 km at IR bands. It provides an unprecedented opportunity for obtaining high-precision T_{air} data over China and surrounding areas.

Machine-learning methods are used to estimate T_{air} based on moderate-resolution imaging spectroradiometer (MODIS) data in several studies [27–29]. However, there is currently a lack of relevant studies on T_{air} estimation based on FY-4A. The use of FY-4A data to estimate high-resolution T_{air} is of great significance to the study of human health and high-temporal- and high-spatial-resolution T_{air} in East Asia. In addition, there is a need for timely and high-resolution T_{air} data for the sustainable planning and management of climate-resilient cities [3].

This study aims to develop the machine-learning approaches for T_{air} estimation using FY-4A data and compares the performances of different machine-learning models [i.e., multivariate linear regression (MLR), gradient boosting decision tree (GBTD), k -nearest neighbors (KNN), random forest (RF), extreme gradient boosting (XGB), and deep neural network (DNN)] in T_{air} estimation, which, to the best of our knowledge, has never been done before. By comparing different machine-learning algorithms, a machine-learning algorithm with good applicability for estimating T_{air} is selected. The algorithm is widely applicable to meteorological satellites without surface-temperature products.

The remainder of this paper is organized as follows. In Section 2, the study area and data used for model development are introduced, and the construction of the above-listed six machine-learning models for T_{air} estimation is described. Variable importance analysis, validation results, and discussion are described in Section 3. Conclusions are presented in Section 4.

2. Materials and Methods

2.1. Study Area. The study area is located in China, and Figure 1 shows the spatial distribution of 1,812 meteorological stations used in this study. There is a higher altitude in the West over China than in the East, and even the Qinghai-Tibet Plateau has an average elevation of over

4,000 m [30]. There are more stations in the East areas than in the West ones due to the uneven distribution of population and economic development in China (Figure 1).

2.2. Data. The data used in this study mainly include FY-4A/AGRI brightness temperature (BT) and L2 cloud mask data, global forecast system (GFS) 3 h forecast data, meteorological data of 1,812 stations in China, and other auxiliary data (longitude, latitude, and Julian day).

2.2.1. Satellite Data. FY-4A, the new generation of Chinese geostationary meteorological satellites, was launched on December 11, 2016. It was fixed at a position of 99.5°E above the equator. As thermal infrared split-window channels, the 12 and 13 bands of AGRI (BT₁₂ and BT₁₃, respectively) are mainly used for studies of cloud, aerosol, and T_{air} estimation. Their central wavelengths are 10.8 and 12.0 μm [31].

BT₁₂, BT₁₃, and L2 cloud mask products during Summer 2018 (i.e., June, July, and August) were used. The AGRI data were selected at 3 h intervals (i.e., 00, 03, 06, 09, 12, 15, 18, and 21 UTC) per day. The data were downloaded from the China National Satellite Meteorological Center (<http://satellite.nsmc.org.cn/PortalSite/Data/Satellite.aspx>).

2.2.2. Meteorological Data. This study selected meteorological data at 3 h intervals from 1,812 observation stations in China during summer 2018. The meteorological variables used in this study include T_{air} and the digital elevation model (DEM). T_{air} in summer 2018 ranges from -5°C to 40°C , and the DEM of the station was between 0 and 5000 m. These data were obtained from the China Meteorological Data Service Center (CMDC) (<http://data.cma.cn/>).

2.2.3. Numerical Weather Prediction Data and Auxiliary Data. Previous studies showed that the relationship between BTs (or LST) and T_{air} is easily affected by surface characteristics and atmospheric conditions [7, 31]. Therefore, the accuracy of T_{air} estimation was effectively improved by adding several auxiliary parameters [32]. In this study, GFS 3 h precipitable water vapor (GFS PWV) and relative humidity (GFS RH) forecast fields data were used. The forecast length of the GFS data (GFS PWV and GFS RH) used was 3 h per day, and there were eight periods of data per day (i.e., 00, 03, 06, 09, 12, 15, 18, and 21 UTC). The GFS data were interpolated according to the location and time information of the AGRI pixels. GFS data were obtained through the U.S. National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Prediction (<http://www.nco.ncep.noaa.gov/pmb/products/gfs>). Table 1 presents the temporal and spatial resolution information of the data used in this study.

2.3. Methods

2.3.1. Preparation of Training Dataset. The BT₁₂, BT₁₃, GFS PWV, GFS RH, and auxiliary data were used as the input variables, T_{air} was used as the response variable of the

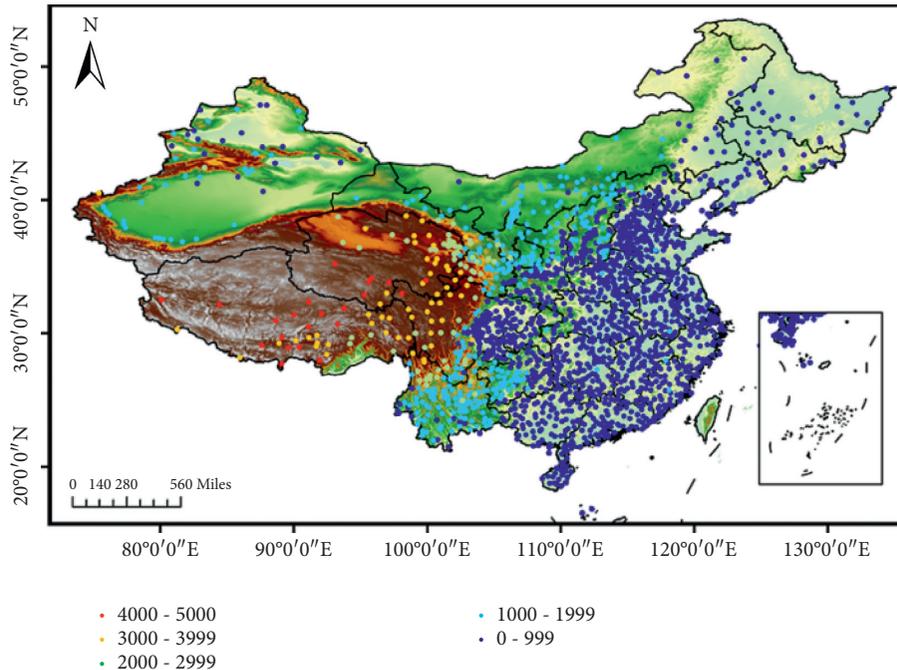


FIGURE 1: Elevation of study area and meteorological sites in China. Differently colored dots represent sites located at different elevations.

TABLE 1: Temporal and spatial information of the primary data used in this study.

| Abbreviation | Units | Spatial resolution | Temporal resolution | Source |
|------------------|-------|--------------------|---------------------|------------|
| AGRI BTs | K | 4 km | 15 min | FY-4A AGRI |
| GFS PWV | mm | 0.5° | 3 ^h | GFS |
| GFS RH | % | 0.5° | 3 ^h | GFS |
| DEM | m | Site | — | FY-4A AGRI |
| Longitude | — | Site | — | CMDC/AGRI |
| Latitude | — | Site | — | CMDC/AGRI |
| T_{air} | °C | Site | 1 ^h | CMDC |
| Julian day | — | — | — | — |

machine-learning models (Table 1), and all data points (across space and time) were included in one model (i.e., the XGB model) [33]. The construction of the representative training data was crucial to develop successful retrieval models using machine learning. Thus, data from June to August—except the 1st, 10th, 20th, and 30th of each month—were collected as the original dataset, and the original dataset was randomly divided into a training dataset (80%, 97,1773 samples) and a test dataset (20%, 24,2944 samples) with the same number of pieces of data for each bin (i.e., 1.0°C in temperature) as shown in Figure 2. For the validation, the data that were not used for training were selected from June to August 1st, 10th, 20th, and 30th.

2.3.2. Machine-Learning Algorithm. Machine-learning methods have been widely used in classification and regression in the field of remote sensing [34–41]. In this study, six machine-learning approaches, that is, MLR, GBTD, KNN, RF, XGB, and DNN, were used for constructing T_{air} estimation models. The flowchart of T_{air} estimation based on machine-learning approaches is shown in Figure 3. L2 cloud

mask products were used to detect cloud. If the data were cloudless, FY-4A data matched both the GFS data and meteorological station data (same space and time), and then T_{air} was estimated through the machine-learning models.

As a simple machine-learning algorithm, MLR has usually been the basic tool for the estimation of meteorological parameters [42, 43]. Similarly, as a local non-linear algorithm, the prediction process of KNN is generally divided into two steps. First, when the KNN algorithm predicts a point, it searches for the k -nearest neighbors closest to the point in the training dataset. Second, the mean of the target variable of the k -nearest neighbors is computed [44, 45]. In this study, the hyperparameters of MLR and KNN were set to default values. Unlike MLR and KNN, RF is an ensemble to a decision-tree-based approach for improving the prediction accuracy, such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest [34, 43, 46–50]. The Scikit-learn library was used for hyperparameter tuning named GridSearchCV from Python to filter the hyperparameters including number of

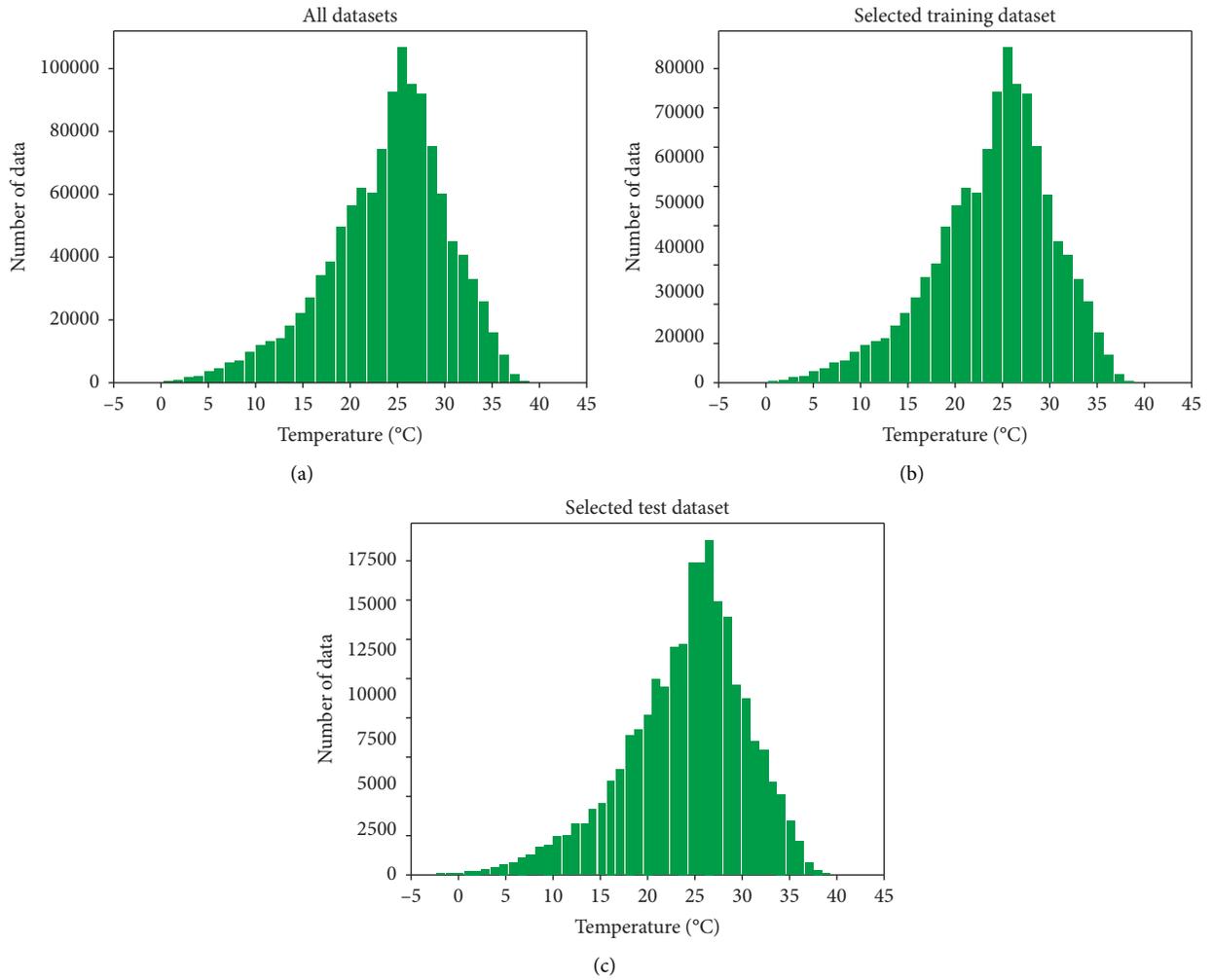


FIGURE 2: Histograms of training data for machine learning. (a) All datasets, (b) training dataset, and (c) test dataset.

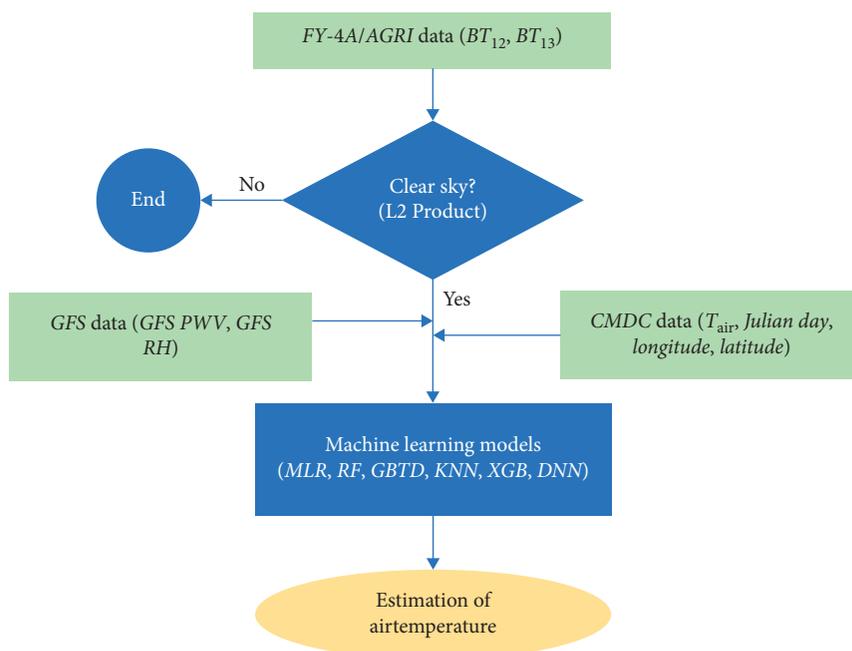


FIGURE 3: Flowchart of the T_{air} estimation model in this study.

trees (`n_estimators`), minimum number of samples (`min_samples_leaf`), and maximum depth of a tree (`max_depth`). The result of parameter selection is `n_estimators = 200`, `min_samples_leaf = 50`, and `max_depth = 3`.

The principle of GBTD is to sequentially apply a classification algorithm to the weighted version of the training data [51, 52], descending along the gradient direction of the model loss function previously established, and then perform a weighted majority vote on the resulting classifier sequence. As an improved algorithm of GBTD, XGB uses all data in each iteration, which is similar to RF [53, 54]. Therefore, XGB reduces the complexity of the model and makes the learned model simpler [35, 54–58]. In this study, four hyperparameters in GBTD and XGB models (i.e., `n_estimators`, `max_depth`, `learning_rate` (`lr`), and minimum loss reduction) required to make a further partition on a leaf node of the tree (`gamma`) were empirically tuned based on RMSE. The optimum `n_estimators`, `gamma`, `max_depth`, and `lr` in the two models were 500, 0.2, 5, and 0.1, respectively.

An artificial neural network (ANN) is a biologically inspired machine-learning method [59]. Here, DNN, a subset of ANN with multiple hidden layers, uses a fully connected structure, which has the ability to learn time and space relationships [60, 61]. It adjusts the connection strength through back-propagation and minimizes the prediction error by iterating between neurons [62–64]. Each hidden layer was tested in the DNN model at one to five hidden layers and 5–200 neurons in five intervals. In addition, some widely used optimizers (i.e., stochastic gradient descent, RMSProp, and Adam) were tested by comparing the calculated results. In this study, the hyperparameters of the DNN were set as follows: `batch_size`, 128; `dropout_rate`, 0.1; `stop_steps`, 20 (if the validation-set loss function was not improved within 20, training will be terminated); and `learning_rate`, 0.001. The optimizer chose Adam, the number of hidden layers was three, and the number of hidden neurons was 256.

2.4. Error Analyses. Four statistical factors—determination coefficient (R^2), RMSE, MSE, and mean bias (bias)—were used to evaluate the accuracy of T_{air} estimation model as follows:

$$R^2 = \frac{[\sum (T_{ea} - \bar{T}_{ea})(T_{oa} - \bar{T}_{oa})]^2}{\sum (T_{ea} - \bar{T}_{ea})^2 \sum (T_{oa} - \bar{T}_{oa})^2}, \quad (1)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (T_{ea} - T_{oa})^2}{N}}, \quad (2)$$

$$\text{bias} = \frac{\sum_{i=1}^N (T_{ea} - T_{oa})}{N}, \quad (3)$$

$$\text{MSE} = \frac{\sum_{i=1}^N (T_{ea} - T_{oa})^2}{N}, \quad (4)$$

where T_{ea} is the estimated T_{air} , T_{oa} is the observed T_{air} at the meteorological stations, and N is the sample size.

3. Results and Discussion

In this section, the results of variable importance were presented, and the performance of the six machine-learning models was verified. The spatial distribution characteristics of the T_{air} errors of each model were also analyzed.

3.1. Variable Importance Results. Correlation analysis was performed to analyze the linear relationship between T_{air} and BT_{12} , BT_{13} , GFS PWV, GFS RH, DEM, longitude (LONG), latitude (LAT), and Julian day (JD). Table 2 shows the correlation coefficient matrix of these variables.

As described in Figure 4(a), GFS PWV, DEM, BT_{12} , and BT_{13} had a better correlation with T_{air} than other variables, and the R values of the four variables were 0.635, -0.596 , 0.459, and 0.413, respectively. This indicated that these variables played more important roles in the linear T_{air} estimation models. However, the Pearson correlation coefficient only described the linear correlation between two variables; it could not identify the nonlinear relationship between two variables. Therefore, the variable importance of the RF algorithm was also analyzed (Figure 4(b)). The RF algorithm modeled the nonlinear relationship well. The GFS PWV was identified as the most important variable for T_{air} estimation in the RF model, while the GFS RH and BT_{12} also played important roles than other predictors. Therefore, PWV and RH were used as inputs to effectively improve the accuracy of T_{air} estimation, which was consistent with the previous study [65].

3.2. Model Performance Results. For evaluating the overall performance of each model, a 10-fold cross-validation method was used. K -fold cross-validation was used for model configuration selection. When a particular value of K was selected (where K was 10), the datasets were randomly and equally distributed among K groups. One group was folded for test, and the $K - 1$ group was folded for training. In a total of k validations, the model performance was calculated using different test folds for each validation [35]. Finally, the average validation results were used to evaluate the overall performance of each model.

Figure 5 illustrates the six models with different statistical parameters, including RMSE, Bias, MSE, and R^2 . The MLR model had the lowest performance of the six models. The variation range of RMSE, Bias, MSE, and R^2 in the MLR model was quite wide; even the range of RMSE was 1.602°C – 4.487°C , while the DNN model used in this study had better overall performance and higher efficiency than the other five models. The DNN model showed the highest accuracy, with an average RMSE of 1.736°C . The

TABLE 2: Pearson correlation matrix for variables considered in the T_{air} estimation model.

| | T_{air} | BT ₁₂ | BT ₁₃ | GFS PWV | GFS RH | DEM | LONG | LAT | JD |
|------------------|------------------|------------------|------------------|---------|--------|--------|--------|--------|--------|
| T_{air} | 1.000 | 0.459 | 0.413 | 0.635 | -0.182 | -0.596 | 0.303 | -0.288 | -0.383 |
| BT ₁₂ | 0.459 | 1.000 | 0.995 | 0.047 | -0.256 | -0.287 | 0.199 | -0.022 | -0.099 |
| BT ₁₃ | 0.413 | 0.995 | 1.000 | 0.010 | -0.249 | -0.268 | 0.190 | 0.002 | -0.077 |
| GFS PWV | 0.635 | 0.047 | 0.010 | 1.000 | 0.383 | -0.585 | 0.366 | -0.463 | -0.310 |
| GFS RH | -0.182 | -0.256 | -0.249 | 0.383 | 1.000 | -0.020 | 0.189 | -0.355 | -0.046 |
| DEM | -0.596 | -0.287 | -0.268 | -0.585 | -0.020 | 1.000 | -0.663 | 0.057 | 0.003 |
| LONG | 0.303 | 0.199 | 0.190 | 0.366 | 0.189 | -0.663 | 1.000 | 0.134 | -0.004 |
| LAT | -0.288 | -0.022 | 0.002 | -0.463 | -0.355 | 0.057 | 0.134 | 1.000 | -0.006 |
| JD | -0.383 | -0.099 | -0.077 | -0.310 | -0.046 | 0.003 | -0.004 | -0.006 | 1.000 |

Data represent the correlation coefficient between different variables.

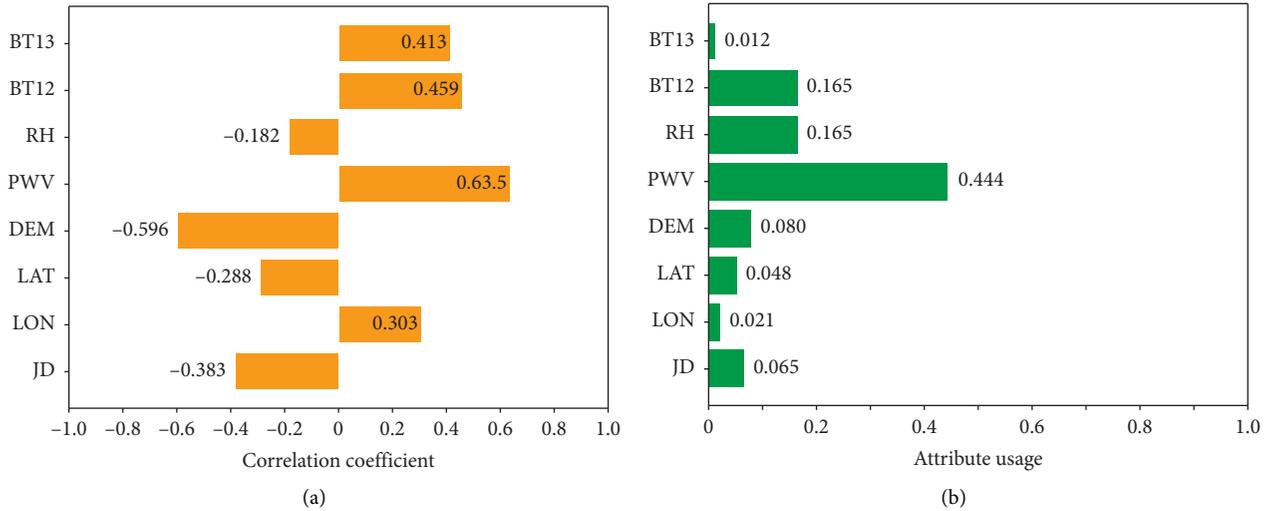


FIGURE 4: (a) Relative variable importance identified by Pearson correlation coefficient and (b) RF variable importance.

range of RMSE in the DNN model was 0.852°C – 2.584°C , showing good concentration and stability, as presented in Figure 5(a). In addition, the overall performances of the XGB and GBTD models of the remaining models were equivalent, which were better than those of the MLR, KNN, and RF models.

3.3. Validation Results. Model performance was used as an indicator to internally validate each model. The model accuracy must be evaluated with a dataset that was not used for training or testing. To validate the developed MLR, RF, KNN, GBTD, XGB, and DNN models, the observed data not used for both training and testing were utilized (validation dataset in Section 2.3.1). Figure 6 illustrates the quantitative validation results of the estimated T_{air} during the validation time (the 1st, 10th, 20th, and 30th of June–August 2018). Compared with the results in the test dataset, the overall accuracy of the six models on the validation dataset decreased. For example, for the DNN model, the RMSE of T_{air} using the test dataset was 1.736°C , while that of the validation results was 2.006°C . This difference may be caused by overfitting due to the fact that the best model was not selected based on the final validation results [35].

The biases of the MLR, RF, DNN, GBTD, and XGB models were within $\pm 0.2^{\circ}\text{C}$, indicating no obvious overestimation or underestimation. In contrast, the KNN model showed a larger negative bias of -0.492°C . The reason that the KNN model had a larger negative bias may be that it had poor robustness. Robustness mainly depended on the dataset, and poor robustness made the model difficult to directly apply to other cases, so the KNN model had a low bias on the test dataset and a high bias on the validation dataset.

The XGB model had excellent modeling performance with R^2 of 0.902. The R^2 values of the GBTD and DNN models were 0.898 and 0.890, respectively, and the R^2 value of the remaining three models was less than 0.89. Moreover, compared with the other models, the XGB and GBTD models can repeatedly learn to generate a weighted average of the weak learners. Therefore, the XGB and GBTD models showed a relatively better performance in the validation dataset in most sites. In general, the XGB model showed a higher overall performance than the other five models on the validation dataset.

The T_{air} estimation models based on satellite and numerical forecast data are susceptible to factors such as altitude and surface roughness. To further evaluate the

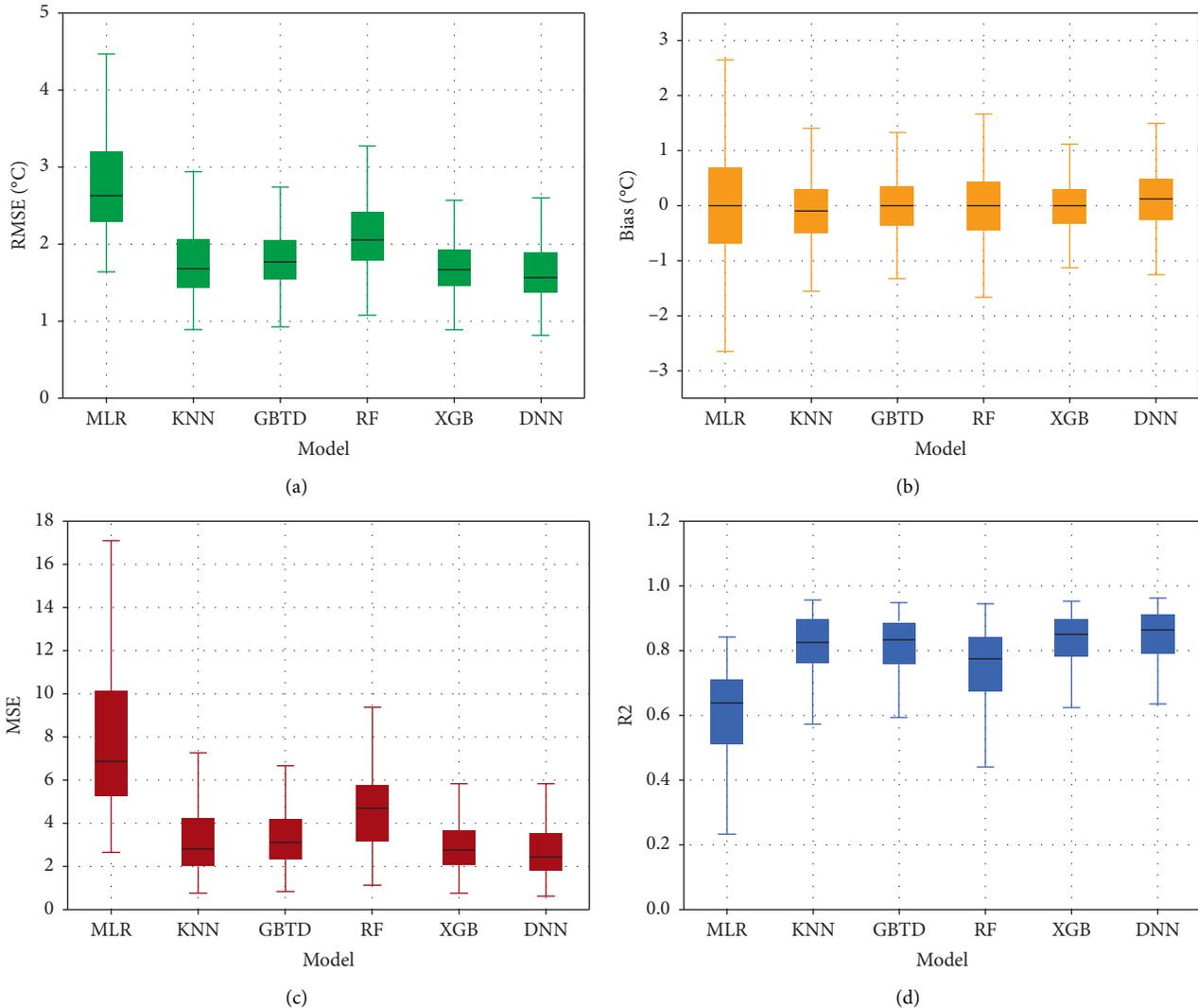


FIGURE 5: Boxplots of performance evaluation for the six models (MLR, KNN, SVM, RF, XGB, and DNN) using the test dataset in terms of (a) RMSE, (b) Bias, (c) MSE, and (d) R^2 .

applicability of these models, the spatial distribution of each meteorological observation was evaluated (Figures 7–9).

It can be seen that the T_{air} estimation errors of all models showed obvious spatial distribution characteristics (Figure 7). Generally, the RMSE is relatively low in the eastern regions (e.g., Guangdong Province) and high in the northwestern regions for each model (e.g., Xinjiang Province). For example, the RMSE in Guangdong Province of the XGB model was approximately 1.2°C – 1.8°C , while that in Xinjiang Province was about 2.0°C – 3.2°C . Because the northwestern regions have relatively wide T_{air} changes during day and night, high altitude, and few meteorological observations, the accuracy difference between northwestern and eastern China is obvious. Moreover, the RMSE of the KNN, DNN, GBTD, and XGB models was relatively low in the eastern and southern regions. However, the MLR, RF, KNN, and DNN models had a higher RMSE in northwestern China. In contrast, the GBTD and XGB models had a relatively smaller RMSE in northwestern China because the

GBTD and XGB models can generate repeated weighted averages to adjust the applicability of different regions through repeated learning of numerous data.

Furthermore, Gong's study (2015) [66] illustrated that the RMSE of GFS T_{air} in most eastern regions reaches 1.5°C – 3.0°C and was above 3.5°C in the northwestern regions. By contrast, the results showed that the RMSE of T_{air} estimated by the DNN, XGB, and GBTD models was obviously lower than that of GFS data. In the present study, the RMSE of the XGB model was 1.0°C – 2.0°C in most eastern regions, and it was below 3.5°C in the northwestern regions. In addition, $\text{RMSE} < 2.0^{\circ}\text{C}$ accounted for 48.2% and $\text{RMSE} < 2.5^{\circ}\text{C}$ accounted for 87.6% in the XGB model.

The six models showed the same distribution trend as shown in Figure 8, with R^2 being higher in the eastern regions, but R^2 gradually became lower as it got closer to the southwestern regions. Compared with the central regions (e.g., Henan Province), the viewing zenith angle (VZA) of ARG1 over the western China is larger. The larger the VZA

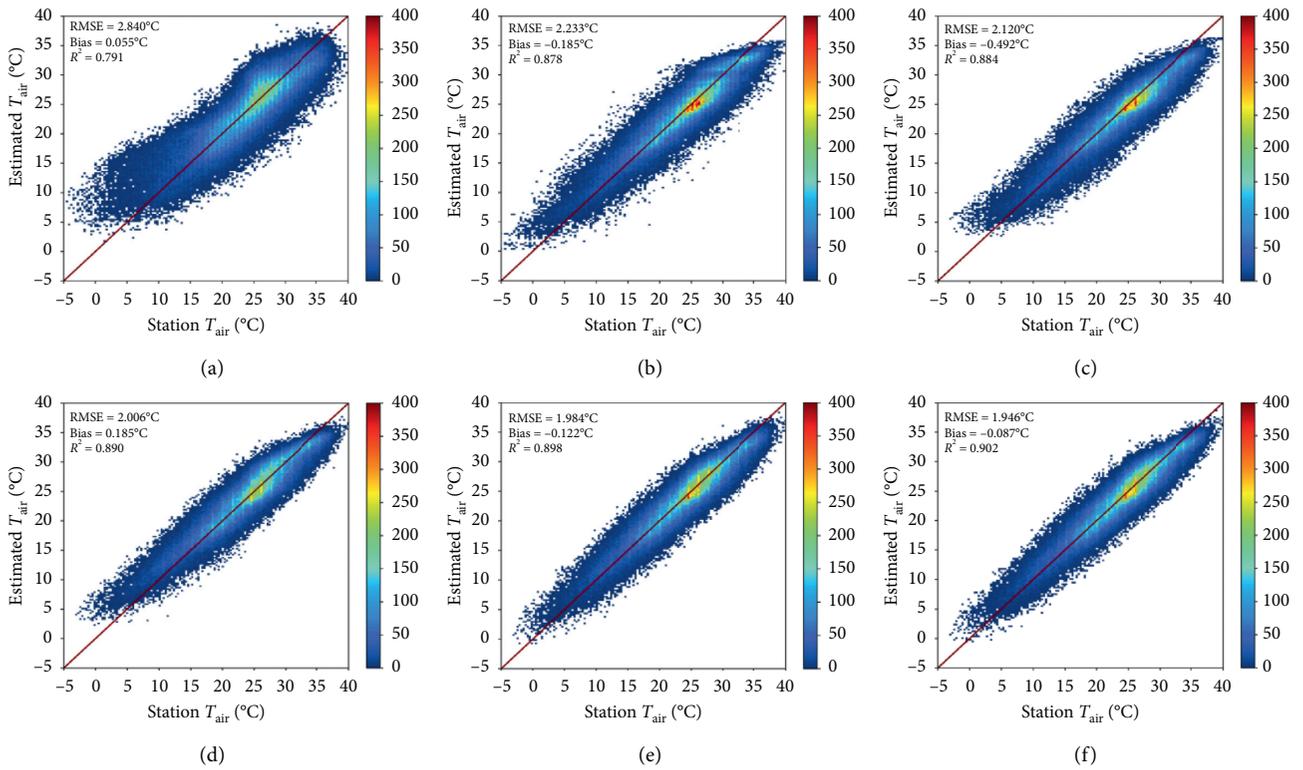


FIGURE 6: Two-dimensional histogram of predicted T_{air} data and meteorological observed T_{air} data based on six machine-learning models. (a) MLR model. (b) RF model. (c) KNN model. (d) DNN model. (e) GBTD model. (f) XGB model.

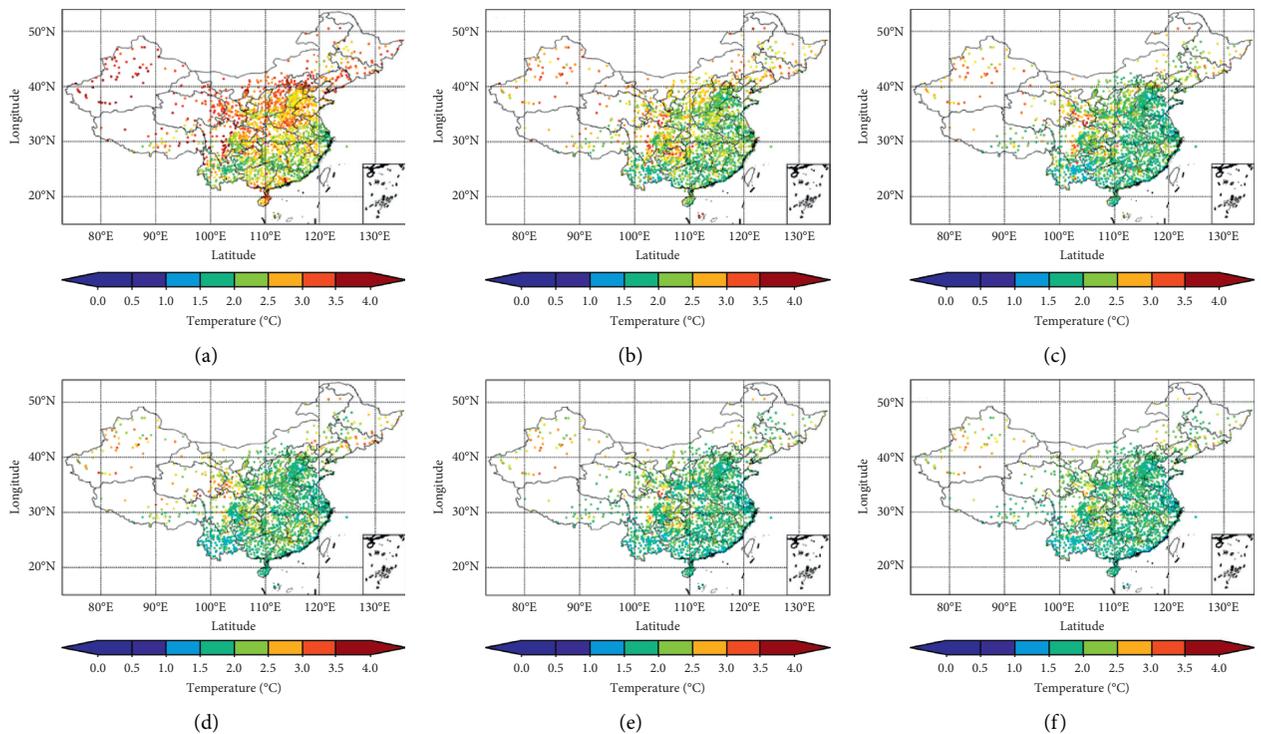


FIGURE 7: Spatial distribution of RMSE for six machine-learning models. (a) MLR-RMSE. (b) RF-RMSE. (c) KNN-RMSE. (d) DNN-RMSE. (e) GBTD-RMSE. (f) XGB-RMSE.

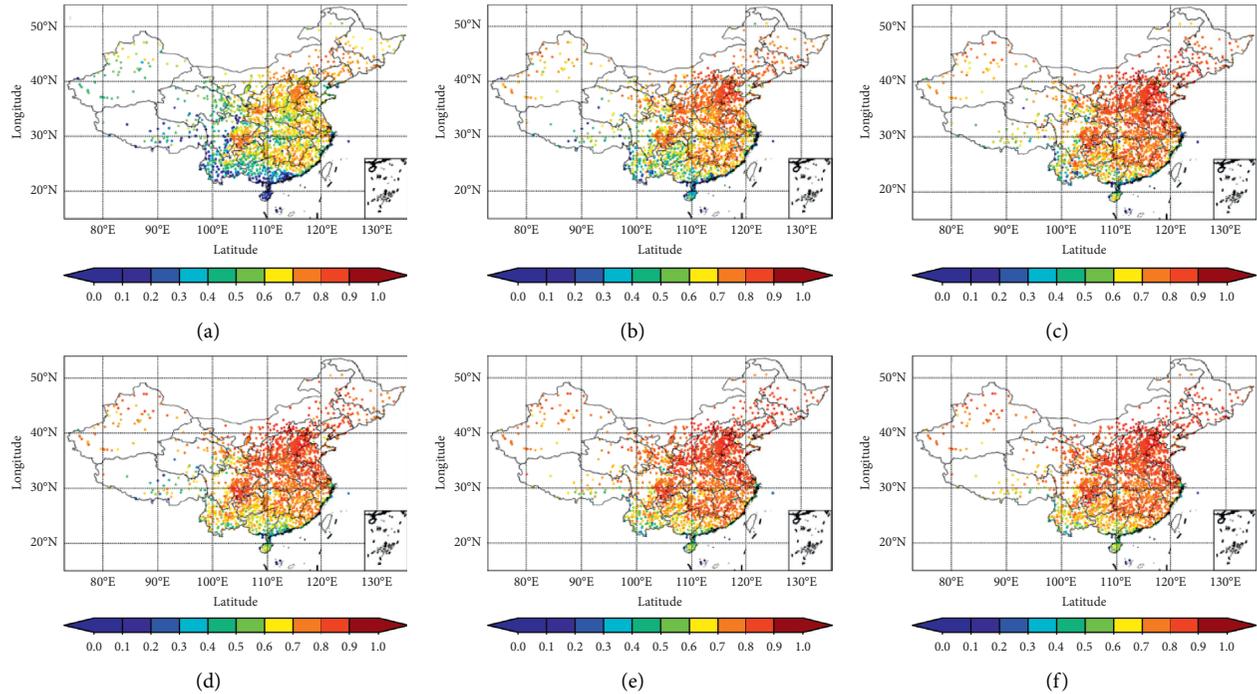


FIGURE 8: Spatial distribution of R^2 for six machine-learning models. (a) MLR- R^2 . (b) RF- R^2 . (c) KNN- R^2 . (d) DNN- R^2 . (e) GBTD- R^2 . (f) XGB- R^2 .

is, the more the radiation reaching the sensor will be highly affected by the atmosphere, which may cause differences in R^2 of the estimated T_{air} value between the southwestern and central regions.

For the MLR model, the bias for all of China was large. For the RF and KNN models, relatively high negative bias existed in southwestern China (e.g., Yunnan-Guizhou Plateau), as shown in Figure 9. This may be the relatively simple structure of the three models mentioned above, which cannot well simulate the complex T_{air} changes in China, resulting in underfitting. Besides, T_{air} estimated by the DNN model was overestimated in northwestern China, which was the reason that the RMSE in the DNN model was also high in these regions. In contrast, the GBTD and XGB models had relatively low bias in northwestern China, where the absolute bias ranges from 2.0°C to 3.0°C . In conclusion, the bias is lower in the coastal areas and higher in northwestern areas, which is mainly related to the characteristics of Summer T_{air} change.

Figure 10 shows the time series of RMSE for the six models during the validation period. The RMSE of the MLR model was significantly higher than other models, with the RMSE ranging from 2.5°C to 4.3°C . In contrast, the RMSE of the GBTD and XGB models showed a relatively lower RMSE (i.e., 1.8°C – 2.2°C) than that in the RF, KNN, and DNN models.

Based on the above analysis, it is expected that the XGB model can provide a more reliable and accurate T_{air} estimation than other models. For purposes of evaluating the contribution of predictive factors in the XGB model

to T_{air} estimation, BTs data (BT₁₂ and BT₁₃) and GFS data (GFS PWV and RH) were successively introduced (Table 3). As shown in Table 3, DEM, longitude, latitude, and Julian day were used as input variables, and the RMSE of the XGB model was 3.003°C . The accuracy of T_{air} estimation was obviously improved when BT₁₂ and BT₁₃ were included in the model. Moreover, when GFS PWV and RH were added to the input variables, the RMSE of the XGB model was decreased to 2.164°C , indicating important influences of GFS PWV and RH on the T_{air} estimation. These results are understandable due to the fact that PWV and RH are the main parameters needed for atmospheric correction and LST retrieval. The RMSE of XGB model was improved by 0.228°C compared with just GFS data which were introduced when both AGRI BTs and GFS data were introduced to the input variables. This indicates that both GFS data and satellite observation data have an important role in improving the T_{air} estimation model. The RMSE of T_{air} estimation model was less than 2.0°C when both satellite BTs and GFS data were introduced, which was considered to be the precision level of “accurate” [67].

The relationship of XGB model errors with altitude, observed T_{air} , and VZA was analyzed. Figure 11 demonstrates the scatter plot of the estimated T_{air} error with DEM, T_{air} , and VZA. It can be seen that the T_{air} error mainly ranges from -3°C to 3°C . The results showed positive deviation at high-altitude areas, which produced a larger RMSE than low-altitude areas. The model showed a positive deviation when T_{air} was low while exhibiting a

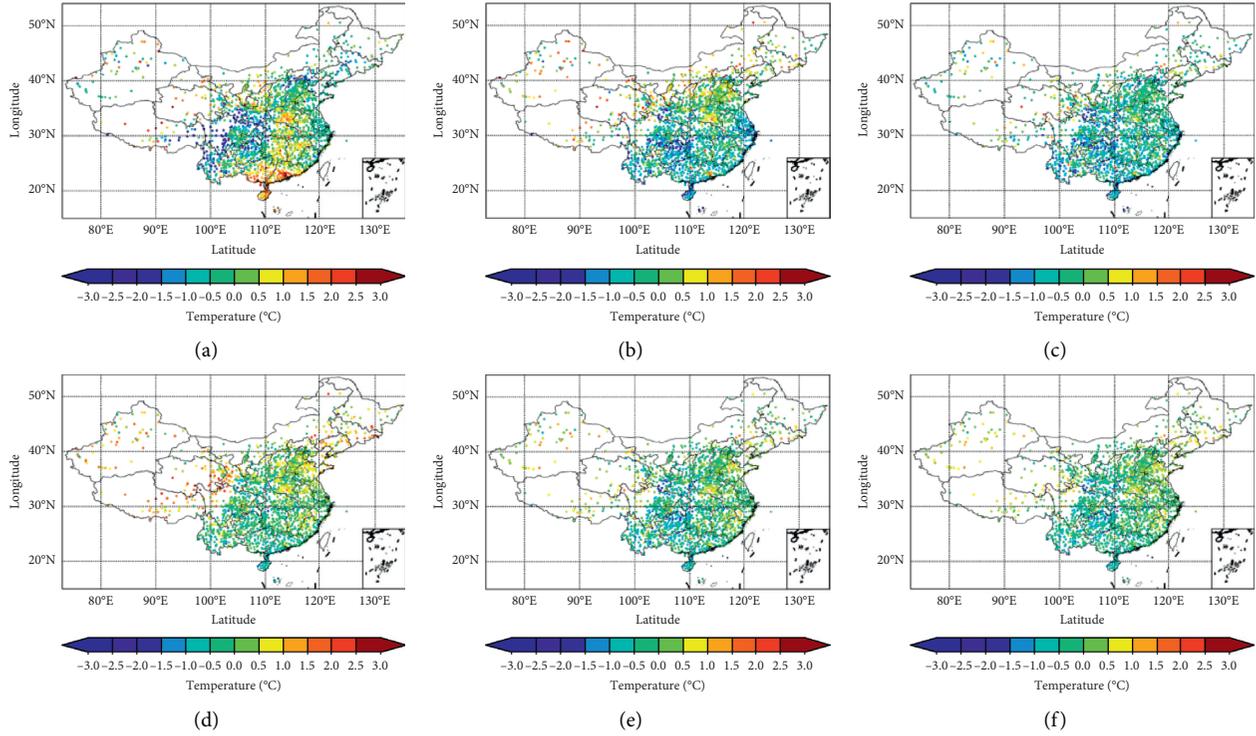


FIGURE 9: Spatial distribution of bias for six machine-learning models. (a) MLR-Bias. (b) RF-Bias. (c) KNN-Bias. (d) DNN-Bias. (e) GBTD-Bias. (f) XGB-Bias.

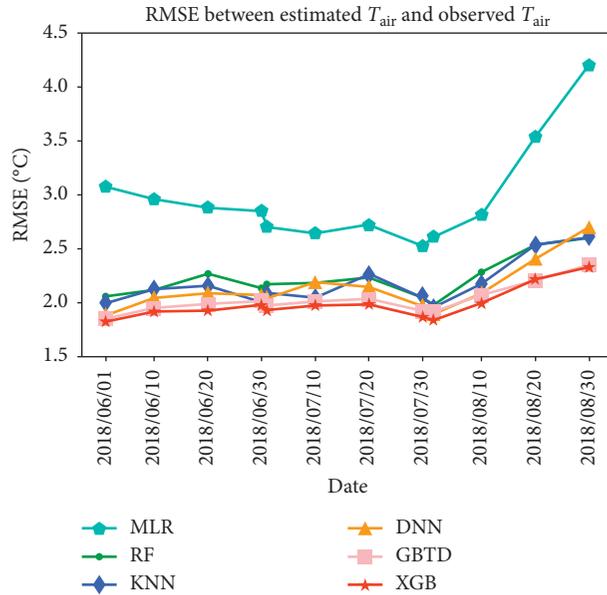


FIGURE 10: Time series (June to August 1st, 10th, 20th, and 30th) of RMSE of estimated T_{air} for six machine-learning models.

TABLE 3: The contribution of AGRI BTs and GFS data to the XGB T_{air} estimation model.

| Predictive factors | XGB model RMSE (°C) |
|---|---------------------|
| DEM, longitude, latitude, Julian day | 3.003°C |
| DEM, longitude, latitude, Julian day, BT_{12} , BT_{13} | 2.376°C |
| DEM, longitude, latitude, Julian day, GFS PWV, GFS RH | 2.164°C |
| DEM, longitude, latitude, Julian day, BT_{12} , BT_{13} , GFS PWV, GFS RH | 1.946°C |

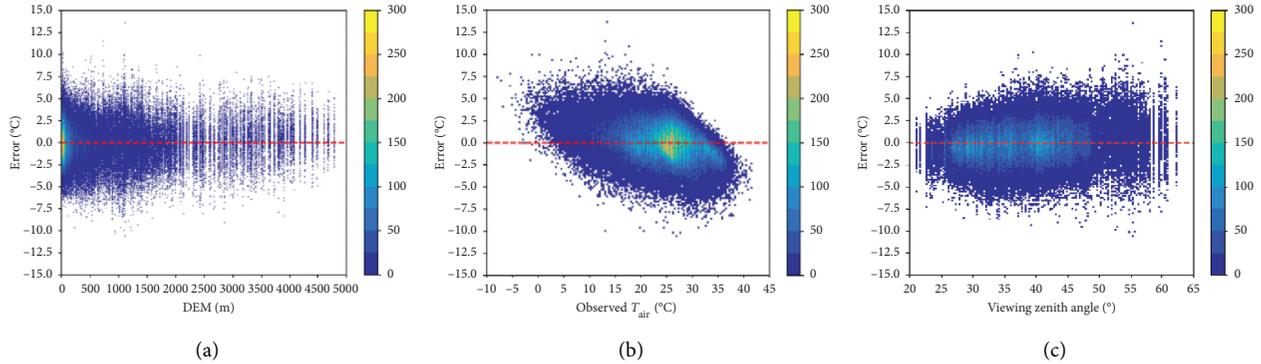


FIGURE 11: Scatter plot of T_{air} error with (a) DEM, (b) observed T_{air} , and (c) VZA for the XGB model.

negative bias for the high-air-temperature condition. Therefore, the model showed a larger RMSE in the lower- and higher-air-temperature conditions due to underestimation and overestimation. This is similar to the results of previous studies [38]. Furthermore, the uneven distribution of stations makes the applicability of the model in high-altitude areas poor. It is worth mentioning that the effect of VZA on model performance is negligible as shown in Figure 11(c).

4. Conclusions

In this study, six machine-learning approaches (MLR, RF, KNN, DNN, GBTD, and XGB) for T_{air} estimation from FY-4A AGRI data in China were compared and analyzed in terms of the spatial and temporal characteristics of their performance. The validation results highlighted the high potential of T_{air} estimation approaches using machine learning and showed that the accuracy of the XGB model was better than that of the MLR, RF, KNN, GBTD, and DNN models at most sites for T_{air} estimation over China. The validation was performed using spatially and temporally independent data, and hence the model performance was considered to be quite reliable.

This study improves on previous studies in the following key areas. First, T_{air} estimation models were constructed based on FY-4A AGRI data and other auxiliary data. The results showed that high-temporal- and high-spatial-resolution T_{air} values (RMSE $< 2.0^{\circ}\text{C}$) can be obtained based on FY-4A data. According to the study of Vazquez [67], the level of precision generally accepted as “accurate” for remote-sensing-based T_{air} estimation is between 1°C and 2°C . Second, the accuracy and performance of the six machine-learning models (MLR, RF, KNN, XGB, GBTD, and DNN) were compared and analyzed. The results showed that the XGB model can provide more stable and high-precision T_{air} estimation, which provides a reference for T_{air} estimation based on machine-learning models. Finally, the accuracy of T_{air} estimation based on satellite data can be effectively improved by adding a numerical model of T_{air} . The experimental results showed that only satellite data were used for large-scale T_{air} estimation in China, and the RMSE of the XGB model was 2.376°C , but the RMSE using satellite data combined with numerically modeled T_{air} data reached 1.946°C .

However, aside from the novelties of this study, the limitation of the dataset used is the restriction to clear-sky conditions. Similarly, machine-learning algorithms cannot infer beyond the range of observed T_{air} value. If the T_{air} value increases beyond the range that cannot be observed within the current training period, the model must be retrained. Moreover, future research may explore whether adding other predictors, such as distance-to-coast and vegetation information (normalized difference vegetation index, etc.), can improve the accuracy of the T_{air} estimation models.

Data Availability

FY-4A AGRI data were downloaded from the China National Satellite Meteorological Center (NSMC) (<http://satellite.nsmc.org.cn/PortalSite/Data/Satellite.aspx>). The GFS data were obtained from the National Centers for Environmental Prediction (<https://www.nco.ncep.noaa.gov/pmb/products/gfs/>). The meteorological station data were accessed at the China Meteorological Data Service Center (CMDC) (<http://data.cma.cn/>).

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (41527806) and the National Key Research and Development Program of China (2016YFA0600101). The authors would like to thank National Satellite Meteorological Center (NSMC) for providing FY-4A data, China Meteorological Data Service Center (CMDC) for the meteorological data, and National Centers for Environmental Prediction (NCEP) for the GFS data.

References

- [1] M. P. Cresswell, A. P. Morse, M. C. Thomson, and S. J. Connor, “Estimating surface air temperatures, from Meteosat land surface temperatures, using an empirical solar zenith angle model,” *International Journal of Remote Sensing*, vol. 20, no. 6, pp. 1125–1132, 1999.

- [2] L. Prihodko and S. N. Goward, "Estimation of air temperature from remotely sensed surface observations," *Remote Sensing of Environment*, vol. 60, no. 3, pp. 335–346, 1997.
- [3] Z. S. Venter, O. Brousse, I. Esau, and F. Meier, "Hyperlocal mapping of urban air temperature using remote sensing and crowdsourced weather data," *Remote Sensing of Environment*, vol. 242, no. 1, Article ID 111791, 2020.
- [4] H. C. Ho, A. Knudby, Y. Xu, M. Hodul, and M. Aminipouri, "A comparison of urban heat islands mapped using skin temperature, air temperature, and apparent temperature (Humidex), for the greater Vancouver area," *Science of the Total Environment*, vol. 544, pp. 929–938, 2016.
- [5] T. R. Lookingbill and D. L. Urban, "Spatial estimation of air temperature differences for landscape-scale studies in montane environments," *Agricultural & Forest Meteorology*, vol. 114, no. 3-4, pp. 141–151, 2003.
- [6] J. W. Hurrell and K. E. Trenberth, "Satellite versus surface estimates of air temperature since 1979," *Journal of Climate*, vol. 9, no. 9, pp. 2222–2232, 1996.
- [7] H. Liu, Q. Zhou, S. Zhang, and X. Deng, "Estimation of summer air temperature over China using himawari-8 AHI and numerical weather prediction data," *Advances in Meteorology*, vol. 2019, pp. 1–10, Article ID 2385310, 2019.
- [8] M. Konda, N. Imasato, and A. Shibata, "A new method to determine near-sea surface air temperature by using satellite data," *Journal of Geophysical Research Oceans*, vol. 101, no. C6, pp. 14349–14360, 1996.
- [9] U. Marcel, "Comparison of satellite-derived land surface temperature and air temperature from meteorological stations on the pan-arctic scale," *Remote Sensing*, vol. 5, no. 5, pp. 2348–2367, 2013.
- [10] W. Wagner, V. Naeimi, K. Scipal, R. de Jeu, and J. Martínez-Fernández, "Soil moisture from operational meteorological satellites," *Hydrogeology Journal*, vol. 15, no. 1, pp. 121–131, 2007.
- [11] C. Oppenheimer, "Review article: volcanological applications of meteorological satellites," *International Journal of Remote Sensing*, vol. 19, no. 15, pp. 2829–2864, 1998.
- [12] H. W. Yates, J. D. Tarpley, S. R. Schneider, D. F. McGinnis, and R. A. Scofield, "The role of meteorological satellites in agricultural remote sensing," *Remote sensing of environment*, vol. 14, no. 1–3, pp. 219–233, 1984.
- [13] C. O. Justice, "The moderate resolution imaging spectroradiometer (MODIS): land remote sensing for global change research," *IEEE Transactions on Geoscience & Remote Sensing*, vol. 36, no. 4, pp. 1228–1249, 2002.
- [14] D. A. Chu, "Remote sensing of smoke from MODIS airborne simulator during the SCAR-B experiment," *Journal of Geophysical Research Atmospheres*, vol. 103, no. D24, Article ID 31979, 1998.
- [15] L. Mohammadi, N. Molanian, and A. Heidari, "Determination of the best coverage area for receiver stations of LEO remote sensing satellites," in *Proceedings of the 3rd International Conference on Information and Communication Technologies: From Theory to Applications*, Damascus, Syria, 2008.
- [16] K. Yumimoto, T.M. Nagao, M. Kikuchi et al., "Aerosol data assimilation using data from Himawari-8, a next-generation geostationary meteorological satellite," *Geophysical Research Letters*, vol. 43, no. 11, pp. 5886–5894, 2016.
- [17] K. Bessho, K. Date, M. Hayashi et al., "An introduction to himawari-8/9— Japan's new-generation geostationary meteorological satellites," *Journal of the Meteorological Society of Japan. Ser. II*, vol. 94, no. 2, pp. 151–183, 2016.
- [18] J. Yang, Z. Zhang, C. Wei, F. Lu, and Q. Guo, "Introducing the new generation of Chinese geostationary weather satellites, Fengyun-4," *Bulletin of the American Meteorological Society*, vol. 98, no. 8, pp. 1637–1658, 2017.
- [19] W. Hui, F. Huang, and R. Liu, "Characteristics of lightning signals over the Tibetan Plateau and the capability of FY-4A LMI lightning detection in the Plateau," *International Journal of Remote Sensing*, vol. 41, no. 12, pp. 4605–4625, 2020.
- [20] S. S. Chu, L. Zhu, H. F. Sun et al., "Automated volcanic hot-spot detection based on FY-4A/AGRI infrared data," *International Journal of Remote Sensing*, vol. 41, no. 1, pp. 2410–2438, 2020.
- [21] X. Zhang and W. Jiao, "Estimation of land surface temperature using geostationary meteorological satellite data," *Remote Sensing Technology & Application*, vol. 28, no. 1, 2013.
- [22] J. Xu, "Estimation of near-surface air temperature from HJ-1B satellite data in Northwest China," *Nongye Gongcheng Xuebao/transactions of the Chinese Society of Agricultural Engineering*, vol. 29, no. 22, pp. 145–153, 2013.
- [23] X. Zhu, Q. Zhang, C.-Y. Xu, P. Sun, and P. Hu, "Reconstruction of high spatial resolution surface air temperature data across China: A new geo-intelligent multisource data-based machine learning technique," *Science of the Total Environment*, vol. 665, pp. 300–313, 2019.
- [24] J. Hrisko, P. Ramamurthy, Y. Yu, P. Yu, and D. Melecio-Vázquez, "Urban air temperature model using GOES-16 LST and a diurnal regressive neural network algorithm," *Remote Sensing of Environment*, vol. 237, Article ID 111495, 2020.
- [25] Q. Zhang, Y. Yu, W. Zhang, T. Luo, and X. Wang, "Cloud detection from FY-4A's geostationary interferometric infrared sounder using machine learning approaches," *Remote Sensing*, vol. 11, no. 24, p. 3035, 2019.
- [26] Y. Chen, G. Chen, C. Cui et al., "Retrieval of the vertical evolution of the cloud effective radius from the Chinese FY-4 (Feng Yun 4) next-generation geostationary satellites," *Atmospheric Chemistry and Physics*, vol. 20, no. 2, pp. 1131–1145, 2020.
- [27] I. Kloog, F. Nordio, B. A. Coull, and J. Schwartz, "Predicting spatiotemporal mean air temperature using MODIS satellite surface temperature measurements across the Northeastern USA," *Remote Sensing of Environment*, vol. 150, pp. 132–139, 2014.
- [28] N. Janatian, M. Sadeghi, S. Hossein Sanaeiejad et al., "A statistical framework for estimating air temperature using MODIS land surface temperature data," *International Journal of Climatology*, vol. 37, no. 3, 2016.
- [29] R. Huang, J.-x. Huang, C. Zhang et al., "Soil temperature estimation at different depths, using remotely-sensed data," *Journal of Integrative Agriculture*, vol. 19, no. 1, pp. 277–290, 2020.
- [30] J. Y. Fang and K. Yoda, "Climate and vegetation in China (I). Changes in the altitudinal lapse rate of temperature and distribution of sea level temperature," *Ecological Research*, vol. 3, no. 1, pp. 37–51, 1988.
- [31] C. Du, R. Huazhong, Q. Qin, S. Zhao, and J. Meng, "Research of split-window algorithm for retrieval of land surface temperature from Landsat 8 data," *Journal of Geomatics*, vol. 39, pp. 73–77, 2014.
- [32] F. Chen, Y. Liu, Q. Liu, and F. Qin, "A statistical method based on remote sensing for the estimation of air temperature in China," *International Journal of Climatology*, vol. 35, 2014.
- [33] P. S. G. de Mattos Neto, G. D. C. Cavalcanti, P. R. A. Firmino, E. G. Silva, and S. R. P. Vila Nova Filho, "A temporal-window

- framework for modelling and forecasting time series,” *Knowledge-Based Systems*, vol. 193, Article ID 105476, 2020.
- [34] R. Bycroft, J. X. Leon, and D. Schoeman, “Comparing random forests and convoluted neural networks for mapping ghost crab burrows using imagery from an unmanned aerial vehicle,” *Estuarine, Coastal and Shelf Science*, vol. 224, pp. 84–93, 2019.
- [35] Y. Lee, D. Han, M.-H. Ahn, J. Im, and S. J. Lee, “Retrieval of total precipitable water from himawari-8 AHI data: A comparison of random forest, extreme gradient boosting, and deep neural network,” *Remote Sensing*, vol. 11, no. 15, p. 1741, 2019.
- [36] J. García-Gutiérrez, F. Martínez-Álvarez, A. Troncoso, and J. C. Riquelme, “A comparison of machine learning regression techniques for LiDAR-derived estimation of forest variables,” *Neurocomputing*, vol. 167, no. 1, pp. 24–31, 2015.
- [37] D. Upreti, W. Huang, W. Kong et al., “A comparison of hybrid machine learning algorithms for the retrieval of wheat biophysical variables from sentinel-2,” *Remote Sensing*, vol. 11, no. 5, p. 481, 2019.
- [38] R. Li, L. Cui, H. Fu, Y. Meng, J. Li, and J. Guo, “Estimating high-resolution PM1 concentration from Himawari-8 combining extreme gradient boosting-geographically and temporally weighted regression (XGBoost-GTWR),” *Atmospheric Environment*, vol. 229, Article ID 117434, 2020.
- [39] G. Papacharalampous and H. Tyralis, “Hydrological time series forecasting using simple combinations: Big data testing and investigations on one-year ahead river flow predictability,” *Journal of Hydrology*, vol. 590, Article ID 125205, 2020.
- [40] R. Pérez-Chacón, G. Asencio-Cortés, F. Martínez-Álvarez, and A. Troncoso, “Big data time series forecasting based on pattern sequence similarity and its application to the electricity demand,” *Information Sciences*, vol. 540, pp. 160–174, 2020.
- [41] H. Abdollahi, “A novel hybrid model for forecasting crude oil price based on time series decomposition,” *Applied Energy*, vol. 267, Article ID 115035, 2020.
- [42] R. S. dos Santos, “Estimating spatio-temporal air temperature in London (UK) using machine learning and earth observation satellite data,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 88, Article ID 102066, 2020.
- [43] H. J. Richardson, D. J. Hill, D. R. Denesiuk, and L. H. Fraser, “A comparison of geographic datasets and field measurements to model soil carbon using random forests and stepwise regressions (British Columbia, Canada),” *GI Science & Remote Sensing*, vol. 54, no. 4, pp. 573–591, 2017.
- [44] H. Franco-Lopez, A. R. Ek, and M. E. Bauer, “Estimation and mapping of forest stand density, volume, and cover type using the k -nearest neighbors method,” *Remote Sensing of Environment*, vol. 77, no. 3, pp. 251–274, 2001.
- [45] R. Haapanen, A. R. Ek, M. E. Bauer, and A. O. Finley, “Delineation of forest/nonforest land use classes using nearest neighbor methods,” *Remote Sensing of Environment*, vol. 89, no. 3, pp. 265–271, 2004.
- [46] M. Belgiu and L. Drăguț, “Random forest in remote sensing: a review of applications and future directions,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 114, no. 114, pp. 24–31, 2016.
- [47] M. J. Cracknell and A. M. Reading, “The upside of uncertainty: Identification of lithology contact zones from airborne geophysics and satellite data using random forests and support vector machines,” *Geophysics*, vol. 78, no. 3, 2013.
- [48] X. Ye, X. Yang, X. Xiong, Y. Shen, M. Hao, and R. Gu, “A quality control method based on an improved random forest algorithm for surface air temperature observations,” *Advances in Meteorology*, vol. 2017, pp. 1–15, Article ID 8601296, 2017.
- [49] B. Babar, L. T. Luppino, T. Boström, and S. N. Anfinsen, “Random forest regression for improved mapping of solar irradiance at high latitudes,” *Solar Energy*, vol. 198, pp. 81–92, 2020.
- [50] L. V. Utkin, M. S. Kovalev, and F. P. A. Coolen, “Imprecise weighted extensions of random forests for classification and regression,” *Applied Soft Computing*, vol. 92, Article ID 106324, 2020.
- [51] L. Liu, M. Ji, and M. Buchroithner, “Combining partial least squares and the gradient-boosting method for soil property retrieval using visible near-infrared shortwave infrared spectra,” *Remote Sensing*, vol. 9, no. 12, p. 1299, 2017.
- [52] J. Son, I. Jung, K. Park, and B. Han, “Tracking-by-Segmentation with online gradient boosting decision tree,” in *Proceedings of the International Conference on Computer Vision, Las Condes, Chile, December 2015*.
- [53] H. Mo, H. Sun, J. Liu, and S. Wei, “Developing window behavior models for residential buildings using XGBoost algorithm,” *Energy and Buildings*, vol. 205, no. 15, Article ID 109564, 2019.
- [54] M. H. D. M. Ribeiro and L. dos Santos Coelho, “Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series,” *Applied Soft Computing*, vol. 86, Article ID 105837, 2020.
- [55] I. B. Mustapha and F. Saeed, “Bioactive molecule prediction using extreme gradient boosting,” *Molecules*, vol. 21, no. 8, p. 983, 2016.
- [56] C. Li, “Power load forecasting based on the combined model of LSTM and XGBoost,” in *Proceedings of the International Conference on Pattern Recognition, Wenzhou, China, June 2019*.
- [57] Z.-Y. Chen, T.-H. Zhang, R. Zhang et al., “Extreme gradient boosting model to estimate PM2.5 concentrations with missing-filled satellite data in China,” *Atmospheric Environment*, vol. 202, pp. 180–189, 2019.
- [58] S. Zhao, D. Zeng, W. Wang et al., “Mutation grey wolf elite PSO balanced XGBoost for radar emitter individual identification based on measured signals,” *Measurement*, vol. 159, Article ID 107777, 2020.
- [59] S. J. Lee, M.-H. Ahn, and Y. Lee, “Application of an artificial neural network for a direct estimation of atmospheric instability from a next-generation imager,” *Advances in Atmospheric Sciences*, vol. 33, no. 2, pp. 221–232, 2016.
- [60] J. Tang, C. Deng, G.-B. Huang, and B. Zhao, “Compressed-domain ship detection on spaceborne optical image using deep neural network and extreme learning machine,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 3, pp. 1174–1185, 2015.
- [61] G. T. Ribeiro, V. C. Mariani, and L. Coelho, “Enhanced ensemble structures using wavelet neural networks applied to short-term load forecasting,” *Engineering Applications of Artificial Intelligence*, vol. 82, pp. 272–281, 2019.
- [62] W. Wang, X. Sun, R. Zhang, Z. Li, Z. Zhu, and H. Su, “Multi-layer perceptron neural network based algorithm for estimating precipitable water vapour from MODIS NIR data,” *International Journal of Remote Sensing*, vol. 27, no. 3, pp. 617–621, 2006.
- [63] K. I. Chronopoulos, I. X. Tsiros, I. F. Dimopoulos, and N. Alvertos, “An application of artificial neural network models to estimate air temperature data in areas with sparse

- network of meteorological stations,” *Journal of Environmental Science and Health, Part A*, vol. 43, no. 14, pp. 1752–1757, 2008.
- [64] S. Rodrigues Moreno, R. Gomes da Silva, V. Cocco Mariani, and L. dos Santos Coelho, “Multi-step wind speed forecasting based on hybrid multi-stage decomposition model and long short-term memory neural network,” *Energy Conversion and Management*, vol. 213, Article ID 112869, 2020.
- [65] A. Bayat and S. Mashhadizadeh Maleki, “Comparison of precipitable water vapor derived from AIRS and SPM measurements and its correlation with surface temperature of 29 synoptic stations over Iran,” *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 178, pp. 24–31, 2018.
- [66] W. W. Gong, “Evaluation of surface meteorological elements from several numerical models in China,” *Climatic and Environmental Research*, vol. 20, no. 1, pp. 53–62, 2015.
- [67] D. Pozo Vázquez, F. J. Olmo Reyes, and L. Alados Arboledas, “A comparative study of algorithms for estimating land surface temperature from AVHRR Data,” *Remote Sensing of Environment*, vol. 62, no. 3, pp. 215–222, 1997.