Article

Estimation of Surface Downward Shortwave Radiation over China from Himawari-8 AHI Data Based on Random Forest

Ning Hou [1,2], Xiaotong Zhang [1,2,*], Weiyu Zhang [1,2], Yu Wei [1,2], Kun Jia [1,2], Yunjun Yao [1,2], Bo Jiang [1,2] and Jie Cheng [1,2]

[1] State Key Laboratory of Remote Sensing Science, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China
[2] Beijing Engineering Research Center for Global Land Remote Sensing Products, Institute of Remote Sensing, Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China
* Correspondence: xtngzhang@bnu.edu.cn; Tel.: +86-10-5880-7712

check for updates

Abstract: Downward shortwave radiation (R_S) drives many processes related to atmosphere–surface interactions and has great influence on the earth's climate system. However, ground-measured R_S is still insufficient to represent the land surface, so it is still critical to generate high accuracy and spatially continuous R_S data. This study tries to apply the random forest (RF) method to estimate the R_S from the Himawari-8 Advanced Himawari Imager (AHI) data from February to May 2016 with a two-km spatial resolution and a one-day temporal resolution. The ground-measured R_S at 86 stations of the Climate Data Center of the Chinese Meteorological Administration (CDC/CMA) are collected to evaluate the estimated R_S data from the RF method. The evaluation results indicate that the RF method is capable of estimating the R_S well at both the daily and monthly time scales. For the daily time scale, the evaluation results based on validation data show an overall R value of 0.92, a root mean square error (RMSE) value of 35.38 (18.40%) Wm^{-2}, and a mean bias error (MBE) value of 0.01 (0.01%) Wm^{-2}. For the estimated monthly R_S, the overall R was 0.99, the RMSE was 7.74 (4.09%) Wm^{-2}, and the MBE was 0.03 (0.02%) Wm^{-2} at the selected stations. The comparison between the estimated R_S data over China and the Clouds and Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) R_S dataset was also conducted in this study. The comparison results indicate that the R_S estimates from the RF method have comparable accuracy with the CERES-EBAF R_S data over China but provide higher spatial and temporal resolution.

Keywords: downward shortwave radiation; Himawari-8 AHI; machine learning methods; Random Forest; multi-channel

1. Introduction

Downward shortwave radiation (R_S) incident at the earth's surface plays a vital role in the energy exchange between the land surface and atmosphere; R_S drives the significant ecological and biophysical processes on the earth [1–4]. The R_S information acts as an indicator of climate change since its availability on the earth depends on the atmospheric load and sky conditions, and also an essential variable of the earth's surface radiation budget [5–9]; therefore, it is critical to acquire accurate R_S estimates.

Ground measurements provide the most accurate R_S data; however, special maintenance is needed, and they are still not adequate to deduce the spatial distribution of R_S on account of the sparsely distributed stations [9,10]. Therefore, the spatial analysis of R_S based on the limited ground stations at regional and global scales are always inadequate [4,11]. The ground measurements over

the whole homogeneous flat area could be obtained by extending a single station if the atmosphere is horizontally homogeneous. Nevertheless, the extrapolation may be not valid if the land surface is heterogeneous or rugged [12]. Remote sensing is one of the alternative ways for estimating the R_S on the local, regional, and global scales [13–16], due to its extensive spatiotemporal coverage of the earth surface. Many studies have made great efforts for estimating the R_S using satellite observations based on various methods, which include the empirical statistical methods [14,17], parametrization methods [18–21], and retrieval methods based on radiative transfer models [3,22–30]. These developed approaches have both advantages and disadvantages. For example, one of the advantages of empirical methods is that they are easy to operate, but the proposed empirical methods are always site-dependent. Compared to empirical methods, the parametrization methods and retrieval approaches based on radiative transfer models have a clear physical basis. On the other hand, these two methods require multiple atmospheric products (e.g., cloud optical depth and aerosol optical depth [31–34]) as input variables. The errors in cloud and aerosol products may cause possible uncertainties in the estimated products [31,35]. Moreover, it is not easy to simultaneously balance the accuracy and efficiency for the retrieval methods based on the radiative transfer models.

In addition to the aforementioned methods, the machine learning method is another feasible way to estimate the R_S using satellite observations [3,36–45]. Ryu et al. [3] applied the artificial neural network (ANN) for computing the R_S with the MODIS product as inputs. The validation results indicated that the relative bias was −2.3% for the ANN model. Wang et al. [37] proposed an ANN-based approach to derive the R_S based on MODIS products. The validation results exhibited that the maximum root mean square error (RMSE) was less than 45 Wm^{-2}. Ghimire et al. [43] selected the support vector regression (SVR) method to estimate the R_S using MODIS data; the result showed that it was also a feasible way to apply the hybrid SVR model for obtaining R_S using satellite observations. Machine learning methods are weak in the physical basis; however, previous studies proved that machine learning methods are one effective way to estimate R_S using satellite observations [44,45]. Compared to the traditional approaches for estimating R_S, machine learning methods have the advantage of catching potential nonlinear relationships between the input variables and the R_S, and they can be applied to a variety of remote sensing variables [38]. The random forest (RF) method has been widely applied in the regression and classification analysis within the remote sensing research field owing to the high accuracy and computational efficiency [46,47]. However, it has rarely been used to estimate the R_S for the new satellite missions.

Himawari-8 is a new generation of geostationary meteorological satellite with the most advanced optical sensors, showing significant improvements over previously available satellites in the geostationary orbit [48,49]. It provides a scan of full-disk, a hemispheric region including the Pacific Ocean (with central coordinates of 0.0°N, 140.7°E) with 2-km spatial resolution and 10-min temporal resolution [50]. To make widespread application of the enhanced monitoring capabilities of Himawari-8, great efforts have been expended, and a few level-2 physical products (e.g., the sea surface temperature, aerosol properties, and R_S) have been officially released from March 2016. Some studies have been conducted to evaluate the performance of the R_S product from the Himawari-8, which was generated based on the algorithm proposed by Frouin and Murakami [51]. Lee et al. [29] retrieved the R_S at the top of the atmosphere using Himawari-8 Advanced Himawari Imager (AHI) data; the results showed the RMSE was 52.12 Wm^{-2} compared with Terra, Aqua, and S-NPP/CERES data. Shi et al. [52] evaluated the R_S product from the Himawari-8 using the Chinese Ecosystem Research Network (CERN) R_S data; the results indicated that the officially released daily R_S product had a mean bias error (MBE) value of 13.8 Wm^{-2} when compared to the CERN R_S measurements. Yu et al. [53] evaluated the Himawari-8 R_S products using ground measurements collected from five networks with an MBE value of 19.7 Wm^{-2}. Damiani et al. [49] evaluated the Himawari-8 R_S product using surface observations from four SKYNET stations in Japan and the Japan Meteorological Agency (JMA) surface network. The comparison results showed that the Himawari-8 R_S product was in good agreement with the ground-measured data with the MBE values ranging from 20 to 30 Wm^{-2}. Although many studies have

been conducted to estimate R_S using diverse methods, few studies have investigated the possibility of estimating R_S using machine learning methods over China, especially based on the RF method.

In this study, we try to estimate the R_S using the Himawari-8 Advanced Himawari Imager (AHI) data based on the RF method. The estimated R_S data at the daily time scale and monthly time scale are evaluated against ground-measured R_S at 86 Climate Data Center of the Chinese Meteorological Administration (CDC/CMA) stations in China. We also make a comparison between the RF method and the traditional ANN method in estimating the R_S. This paper is organized as follows: In Section 2, the datasets applied in this study are introduced. Section 3 introduces the RF machine learning method. Section 4 shows the results and a brief analysis, and Section 5 provides a discussion of the RF method. Finally, a short summary is presented in the last section of this paper.

2. Data

2.1. Himawari-8 AHI Data

Himawari-8 is a geostationary satellite, which is managed by the JMA. It is located at 0.0°N and 140.7°E, about 35,800 km over the land surface. The Himawari-8 has the most advanced optical sensors named AHI. The number of spectral bands, and the spectral and spatial resolution have been greatly promoted [48,53]. The Himawari-8 AHI has 16 bands (three for visible, three for near-infrared, and ten for infrared) in spatial resolutions of 0.5–2.0 km, which can provide abundant spectral information from visible to infrared [29]. The Himawari-8 AHI is capable of supplying a scan of the full-disk of the earth's surface and the target region at the temporal resolution of 10-min and 2.5-min, respectively. On account of the high temporal resolution, it is important for better understanding the R_S spatiotemporal variations in short time scales [54]. The information of Himawari-8 AHI bands is shown in Table 1 [55,56], and more detailed information about the Himawari-8 AHI is described in Bessho et al. [50]. The full-disk scan of Himawari-8 AHI data with 16 observational bands from February to May 2016 were used for analysis in this study.

Table 1. Specifications of Himawari-8 Advanced Himawari Imager (AHI) spectral bands.

Band	Descriptive Name	Central Wavelength (μm)	Spatial Resolution (km)	Primary Purpose
1	Blue	0.46	1.0	Daytime aerosol over land, coastal water mapping
2	Green	0.51	1.0	Green band-to produce color composite imagery
3	Red	0.65	0.5	Day time vegetation/burn scar and aerosols over water, winds
4	Vegetation	0.86	1.0	Daytime cirrus cloud
5	Snow/ice	1.61	2.0	Daytime cloud-top phase and particle size, snow
6	Cloud particle size	2.26	2.0	Daytime land/cloud properties, particle size, vegetation, snow
7	Shortwave window	3.85	2.0	Surface and cloud, fog at night, fire, and winds
8	Upper-level water vapor	6.25	2.0	High-level atmospheric water vapor, winds, and rainfall
9	Mid-level water vapor	6.95	2.0	Mid-level atmospheric water vapor, winds, and rainfall
10	Lower-level/Mid-level water vapor	7.35	2.0	Lower-level atmospheric water vapor, winds, and SO_2
11	Cloud-top phase	8.60	2.0	Total water for stability, cloud phase, dust, SO_2, and rainfall
12	O_3	9.63	2.0	Total ozone, turbulence, and winds

Band	Descriptive Name	Central Wavelength (μm)	Spatial Resolution (km)	Primary Purpose
13	Clean longwave window	10.45	2.0	Surface and cloud
14	Longwave window	11.20	2.0	Imagery, sea surface temperature, clouds, and rainfall
15	Dirty longwave window	12.35	2.0	Total water, ash, and sea surface temperature
16	CO_2	13.30	2.0	Air temperature, cloud heights and amounts

Clouds and aerosols are the two important parameters affecting the R_S [32–34]. The accuracy of the estimated R_S is dependent on the quality of clouds and aerosols data. Besides clouds and aerosols, water vapor is also a significant parameter affecting the R_S [4,14]. Water vapor provided by the Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2) at an hourly temporal resolution and 0.5×0.625-degree spatial resolution [57,58] was used as an ancillary variable. The length of daytime is also suggested to be used for estimating the R_S data [59,60]. The impact of solar zenith angle and surface elevation also should be considered in estimating the R_S [4,61]. In this study, we presented a direct estimation method to generate the R_S data from the Himawari-8 AHI based on the RF method. The Himawari-8 AHI top-of-atmosphere (TOA) radiance of 16 observational bands, water vapor, solar zenith angle, the length of daytime, and elevation data were used as input variables of the RF method to estimate the R_S in this study.

2.2. Ground Measurements

The ground measurements used to build R_S estimation models were obtained from the CDC/CMA. The spatial coverage of the collected Himawari-8 AHI data covered 86 CDC/CMA radiation stations in China. Ground-measured R_S of daily time scale at the corresponding 86 stations from February to May 2016 were collected from CDC/CMA. The R_S measurement by CDC/CMA began in 1957; since 1994, there have been only 96 stations with the records of solar radiation. Before releasing the solar radiation measurements, the CDC/CMA carries out quality control on the radiation data, which includes spatiotemporal consistency checks, and manual correction and adjustment. Previous studies [62,63] indicated that the quality of ground-measured solar radiation data from CDC/CMA should be examined more critically before using them. With the aim of guaranteeing the solar radiation data is reliable, a quality control procedure, proposed by Zhang et al. [10], was conducted in this study. The quality of R_S ground measurements from CDC/CMA is controlled based on the reconstructed daily and monthly integrated R_S data using daily routine meteorological data from the CDC/CMA [64,65]. More detailed information about the quality control procedure is described in Zhang et al. [10]. Figure 1 displays the spatial distribution of the 86 radiation stations over China.

2.3. CERES–EBAF R_S Data

The CERES–EBAF data from February to May 2016 were applied to compare with the estimated R_S data from the RF method in this study. The CERES–EBAF data [24,66] are specifically generated for the applications of evaluating and improving climate models and improving the understanding of the variability in the earth's energy budget. The CERES project offers surface irradiance and TOA irradiance in a variety of spatiotemporal scales [67]. Surface irradiance is calculated with satellite-derived data including cloud, aerosol, temperature, and humidity profiles [68]. The EBAF-surface dataset provides monthly downward shortwave irradiance, which is constrained by the TOA irradiance derived from CERES. The EBAF data are generated for addressing two disadvantages of the CERES level-3 dataset: the requirement of absolute accuracy in quantifying the earth's energy imbalance (EEI) [69], and the fact that CERES–EBAF R_S data combine more accurate cloud information from CERES instruments [24].

It has proven to be one of the most accurate R_S datasets [10,70]. The CERES–EBAF R_S is a gap-filled product with a temporal resolution of one month and a spatial resolution of one degree from March 2000 to February 2017 [71].
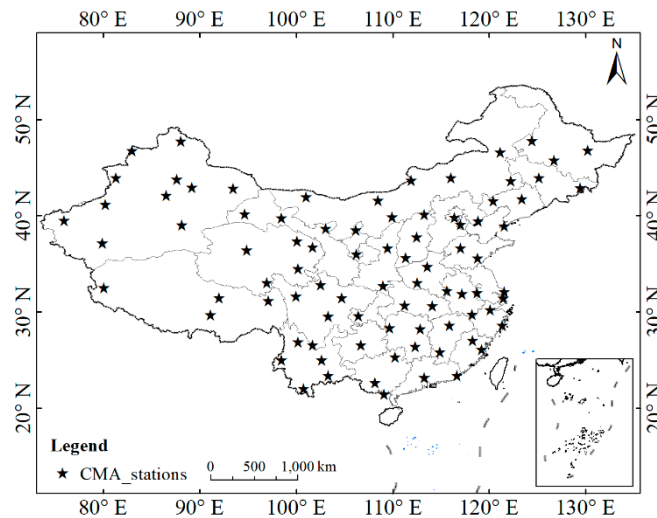

Figure 1. Spatial distribution of the 86 Climate Data Center of the Chinese Meteorological Administration (CDC/CMA) radiation stations.

3. Methodology

3.1. Random Forest

Random forest (RF), which contains abundant regression trees, is a powerful ensemble-learning algorithm for regression and classification studies [72]. The RF method is a modified method of the bagging regression tree, which belongs to the family of regression trees [73–75]. The binary regression tree method includes iteratively dividing the dataset into two separate sets based on some specific rules (e.g., thresholds) [74]; a group of rules on the decision about the predictors is established [76], so that the smaller sets can be divided from the data according to an individual predictor, and the binary split can be performed [75]. Each set is split into two sets, and the final set is named a leaf; then the predictive value can be obtained from the values of the leaf. Each tree firstly grows to the maximum size, and then prunes overfitting trees into optimal sizes using cross-validation and other techniques. A randomization step was added to bagging by Breiman [72]; in the random subgroup of the predictors, the segmentation of each bagged regression tree was built [74]. During the process, around one-third of the training samples that are not used in the bootstrap process is known as out-of-bag data (OBB) [72]. For each bootstrap sample, the optimal segmentation is decided by means of a randomly selected subset at each node according to the minimized Gini index. A sufficiently large number of trees are grown creating a random forest. During the prediction process, the random forest averages the predictors of all regression trees. The residual mean square (RMS) of OBB is used to evaluate the prediction accuracy [77]. The RF method was carried out using the scikit-learn toolbox in this study [78]. The structure of the method is displayed in Figure 2.

Figure 2. Structure of the random forest (RF) method. OBB is out-of-bag data.

3.2. Model Construction

This study estimated the R_S from Himawari-8 AHI data and other ancillary parameters using the RF method. The prediction of the R_S started with putting variables in the RF model. To construct the optimal model, we studied the various band intervals for their importance in determining R_S. The experiments were tested with different band intervals (1, 2, 3, and 4 bands). The validation results showed that the accuracy of the estimated R_S data decreased as the band intervals increased, which indicated that the estimated R_S data were the most accurate when all Himawari-8 AHI bands were used. The ground-measured daily R_S of 86 stations from February to May 2016 were selected as the object variables. Sixteen variables of the Himawari-8 AHI data, which were TOA radiance of bands 1–16, were the input variables in this study. Besides these sixteen variables, water vapor data from MERRA-2, elevation data, the length of daytime, and the solar zenith angle data were also used as the input variables. The solar zenith angle and the length of daytime were calculated according to the specific geographical location and time. The datasets were randomly divided into two groups to construct the models: 80% for training data and the remaining 20% for validation data. We then applied the k-fold cross validation for selecting the best parameters during the training process. Several main parameters of the RF method were used for adjustment to obtain the optimal RF model. The n-estimator parameter, which is the tree number of the forest, can increase the RF method accuracy but at a higher computational cost when it is larger. Additionally, the underfitting and overfitting may occur when the n-estimators are smaller or larger than the best number [79]. As for the max-features parameter, which is the number of random feature subgroups, it can reduce variance but increase MBE. To improve the performance of the RF method, we also modified the min-samples-leaf and

min-samples-split. Looping in each parameter threshold as shown in Table 2, these parameters were optimized using the k-fold cross validation approach according to the R, RMSE, and MBE values. It is the optimization that improved the RF method performance for the best fitting. The optimization results indicated that the RF method had the optimum performance when the n-estimators parameter was 400, the max features parameter was 10, the min-samples-split was 2, and the min-samples-leaf was 1. After the optimal parameters were determined, the RF model was applied for estimating the R_S.

Table 2. Parameters to set for determining the optimal RF model.

Parameters	Threshold	Intervals
n-estimators	50–400	50
max-features	2–19	1
min-samples-split	2–10	1
min-samples-leaf	1–10	1

3.3. Sensitivity Analysis and Scaling Issue

To further understand the RF method in estimating the R_S, a sensitivity analysis was conducted for investigating the influence of the RF method parameters on the accuracy of estimated R_S data. In order to conduct the sensitivity analysis of each parameter, we obtained the variations on R, RMSE, and MBE of the model for the validation data by keeping other parameters of the optimal RF model invariant.

As shown in Figure 3, the sensitivity analysis of the RF method parameters illustrated that R values were sensitive to the max-features. When the max-features increased from 2 to 10, the R values increased from 0.77 to 0.90 gradually. However, the R values decreased to 0.69 as the max-features increased from 10 to 19. The variations of n-estimators, min-samples-split, and min-sample-leaf had little impacts on R values. As shown in Figure 4, the sensitivity analysis of the RF method parameters illustrated that RMSE and MBE values were more sensitive to the max-features and min-samples-leaf parameters than the other two parameters according to the range of variations. The RMSE and MBE values increased as the min-sample-leaf parameter and min-samples-split parameter increased, while for the n-estimator, the RMSE and MBE values decreased as it increased. When the max-features increased from 2 to 10, the RMSE values decreased from 41.42 Wm^{-2} to 36.62 Wm^{-2}; when the max-features increased from 10 to 19, the RMSE values increased to 43.14 Wm^{-2}. The MBE values of the RF method were not sensitive to the min-sample-split parameter. Almost no variation of the MBE values was observed when the min-sample-split parameter varied from 2 to 10.

According to the sensitivity analysis, the accuracy of the estimated R_S (R, RMSE, and MBE) was the most sensitive to the max-features of the RF method of the four parameters. The min-sample-split and n-estimator parameters had low impacts on the accuracy of R_S data.

A problem of scaling in comparing R_S from satellite data and ground measurements was found in previous studies [80–82]. The impact of window size of spatial averaging on the consistency of the estimated R_S data from the RF method and ground measurements was studied. Figure 5 shows that the RMSE values of the estimated R_S data from the RF method decreased with increasing window size when the window size was less than 58 km, then the RMSE values increased with increasing window size. The smallest RMSE appeared when the window size was 58 km. For the monthly data, the optimal window size was 26 km.
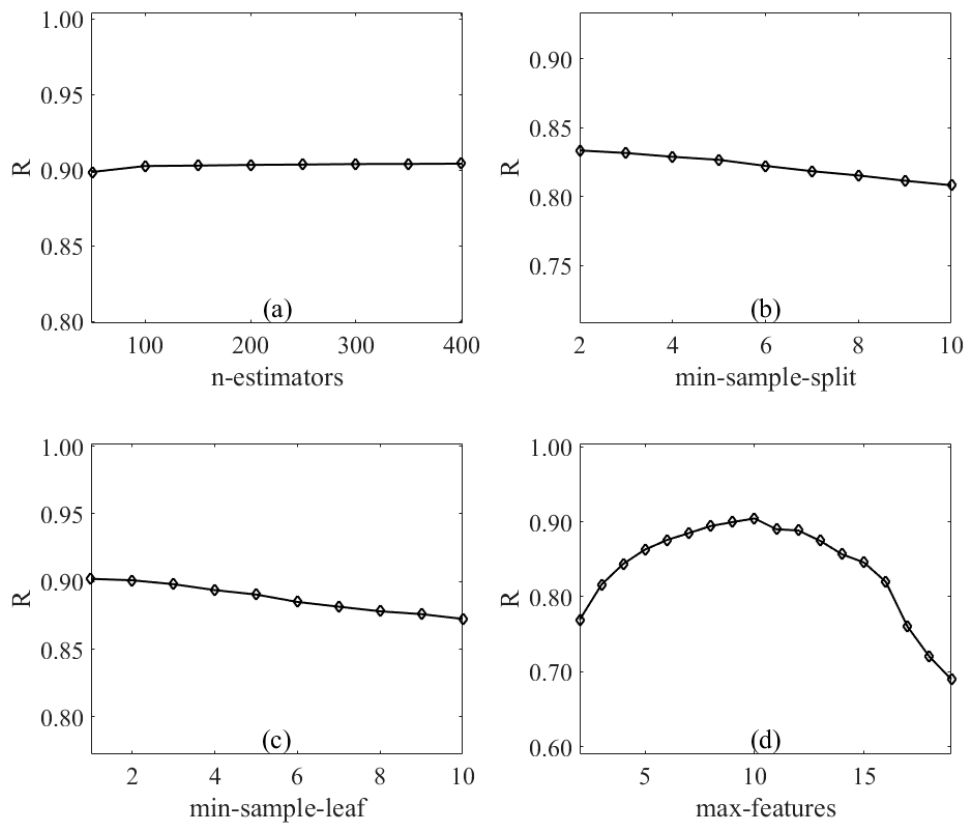
Figure 3. Sensitivity analysis of R values to the (**a**) n-estimators; (**b**) min-sample-split; (**c**) min-sample-leaf; (**d**) max-features of the RF method.
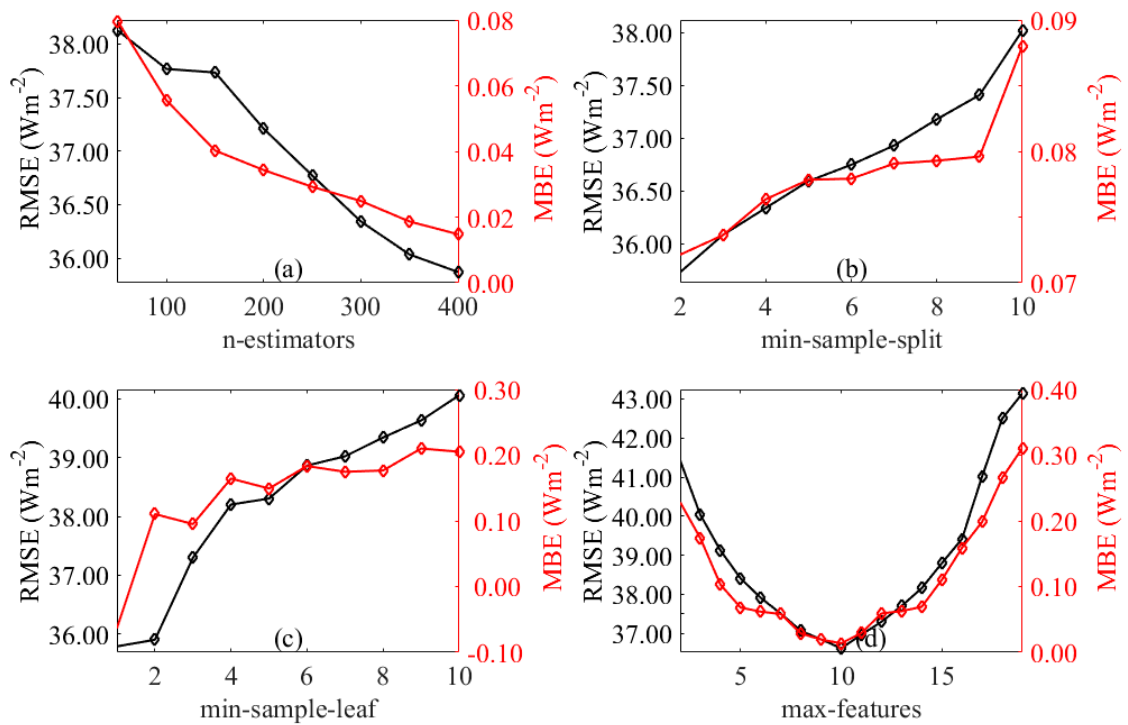

Figure 4. Sensitivity analysis of RMSE and MBE values to the (**a**) n-estimators; (**b**) min-sample-split; (**c**) min-sample-leaf; (**d**) max-features of the RF method.
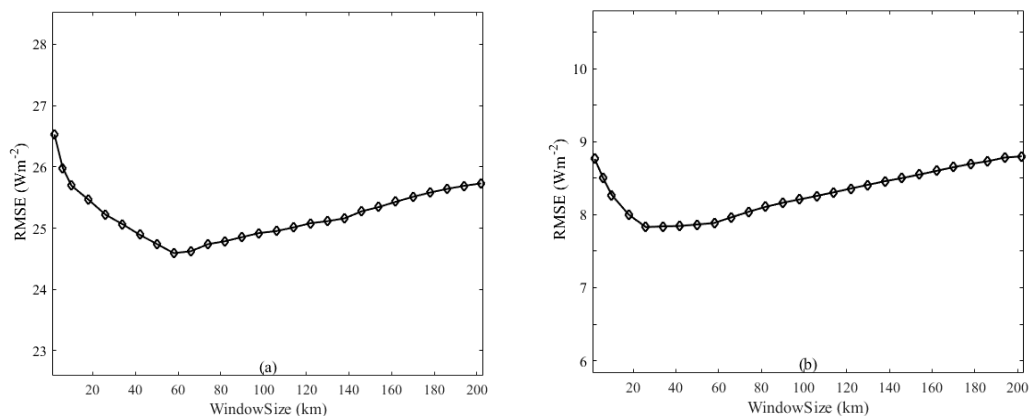
Figure 5. Impact of window size on the evaluation results of monthly mean estimated R_S data from the RF method against ground measurements from CDC/CMA from February to May 2016. (**a**) Daily R_S; (**b**) Monthly R_S.

4. Results and Analysis

The estimated R_S data based on the RF method were evaluated against ground R_S measurements of 86 CDC/CMA stations at the daily time scale, as well as the monthly time scale. The correlation coefficient (R), root mean square errors (RMSE), and MBE were computed for evaluating the accuracy of the estimated R_S data. Additionally, the comparison between the estimated R_S data and the CERES–EBAF R_S data was also conducted.

4.1. Validation Against Ground Measurements

4.1.1. Validation at a Daily Time Scale

The RF method performance for estimating daily R_S from the Himawari-8 AHI data is displayed in Figure 6. The daily estimated R_S data were averaged from a 58 km × 58 km window centered at the stations. It showed that the daily mean R_S estimates from the Himawari-8 AHI data using the RF method agreed well with ground measurements from CDC/CMA. For the training data, the RF method had better performance in estimating daily R_S; it showed that the overall R of the estimated R_S data based on the RF method was 0.99, the RMSE was 11.16 (5.83%) Wm^{-2}, and the MBE was −0.06 (−0.03%) Wm^{-2}. For the validation data, it showed that the overall R was 0.92, the RMSE was 35.38 (18.40%) Wm^{-2}, and the MBE was 0.01 (0.01%) Wm^{-2}. The results indicated that the RF method performed well in estimating the R_S with the Himawari-8 AHI data at the daily time scale.
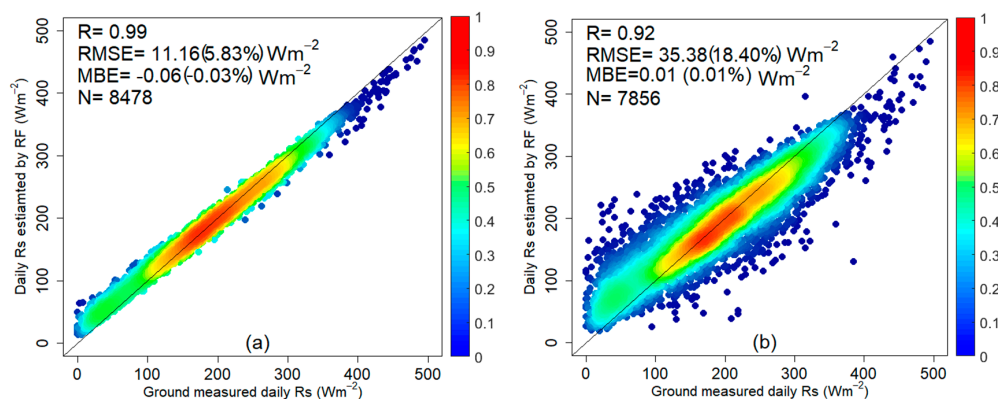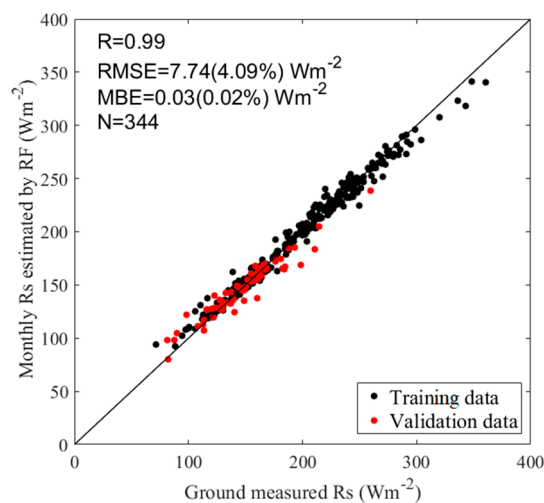

Figure 6. Evaluation results of daily estimated R_S data from Himawari-8 AHI data based on the RF method against ground measurements from CDC/CMA from February to May 2016: (**a**) for training data; (**b**) for validation data. Estimated R_S data are averaged from a 58 km × 58 km window centered at the stations. N is the number of data points.

4.1.2. Validation at a Monthly Time Scale

The monthly mean R_S were obtained from averaging all the daily R_S in a month including the training and validation data. The monthly R_S of a month was not be calculated if there were more than nine missing days in a month. The monthly mean estimated R_S data from the RF method from February to May were validated using the ground-measured R_S data from CDC/CMA. The monthly estimated R_S data were averaged from a 26 km × 26 km window centered at the stations. The performance of the RF method for estimating monthly R_S is displayed in Figure 7. The figure illustrates that the monthly mean estimated R_S data were in good agreement with the ground measurement from CDC/CMA. The overall R was 0.99, the RMSE was 7.74 (4.09%) Wm^{-2}, and the MBE was 0.03 (0.02%) Wm^{-2}. It can be concluded that the RF method could estimate the R_S accurately at the monthly time scale.

Figure 7. Evaluation results of monthly mean estimated R_S data from Himawari-8 AHI data based on the RF method against ground measurements from CDC/CMA from February to May 2016. Estimated R_S data are averaged from a 26 km × 26 km window centered at the stations. N is the number of data points.

4.2. Comparison with CERES–EBAF

4.2.1. Validation Against Ground Measurements

It has been demonstrated that the CERES–EBAF R_S is one of the most accurate global R_S datasets available to date [10,70]. The estimated R_S data from the Himawari-8 AHI data using the RF method were compared with the CERES–EBAF R_S data in this study.

The monthly mean R_S estimates from February to May 2016 at 86 CDC/CMA stations were compared with ground measurements. Figures 7 and 8 show the comparison of R_S estimates and R_S ground measurements. The estimated R_S data from the RF method had relatively better accuracy than those from the CERES–EBAF data at the selected stations over China with an overall R value 0.99, an RMSE value of 7.74 (4.09%) Wm^{-2}, and an MBE value of 0.03 (0.02%) Wm^{-2}. The CERES–EBAF R_S data had similar accuracy at the monthly time scale; the overall R was 0.89, the RMSE was 24.24 (12.79%) Wm^{-2}, and the MBE was −0.25 (−0.13%) Wm^{-2}. The validation results indicated that the estimated R_S data from the RF method had comparable accuracy compared to the CERES–EBAF R_S data at the selected stations over China, and the estimated R_S data from the RF method were in good agreement with ground measurements.

In addition, the estimated R_S data from the RF method and CERES–EBAF data were analyzed with the two-sample Kolmogorov–Smirnov (K–S) test. The two-sample K–S test is a non-parametric test that requires no assumptions on the distribution. It is designed to test whether two samples are drawn from the same distribution [83]. The null hypothesis was that these two distributions had no significant

change. The resulting H was 1 if the test rejected the null hypothesis at the 5% significance level, and 0 otherwise. As shown in Figure 9, the resulting H was 0, which confirmed the null hypothesis. It indicated that the estimated R_S data from the RF method and CERES–EBAF data had no observable difference at the 5% significance level. According to the validation results, we concluded that the estimated R_S data based on the RF method showed comparable precision to the CERES–EBAF R_S data but with higher resolution.
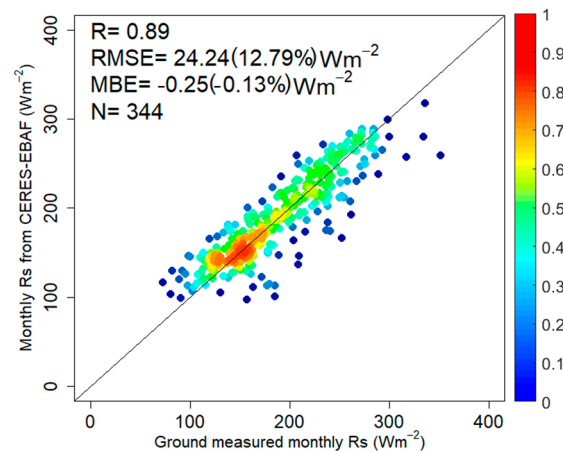
Figure 8. Evaluation results of monthly mean estimated R_S data from CERES–EBAF data against ground measurements from CDC/CMA from February to May 2016. N is the number of data points.
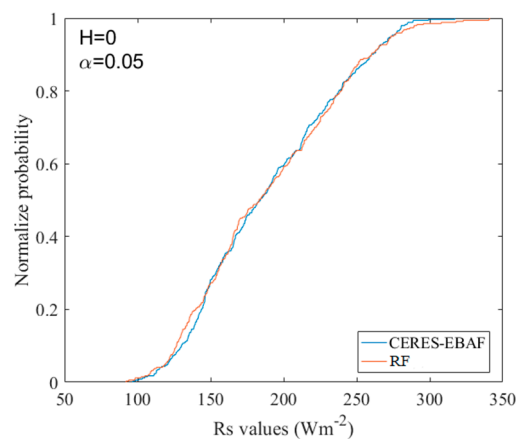
Figure 9. Cumulative density function of two-sample K–S test values for the estimated R_S data from the RF method and CERES–EBAF from February to May 2016.

4.2.2. Mapping R_S of China

Remote sensing is one of the feasible ways to generate the R_S products involving high temporal and spatial resolution from satellite observations. In this study, the Himawari-8 AHI data were applied to estimate R_S with one-hour temporal resolution and 2-km spatial resolution in mainland China from February to May 2016 based on the RF method. The spatial distribution over China of the monthly mean R_S generated from the CERES–EBAF data and R_S estimates from the Himawari-8 AHI data based on the RF method were compared, as displayed in Figure 10. It illustrates that the spatial distribution and temporal variation of monthly mean estimated R_S data from the RF method were in line with those of CERES–EBAF R_S data over China. It indicates that the estimated R_S data from the RF method were accurate but with higher spatial resolution and time resolution compared to the CERES–EBAF R_S data.
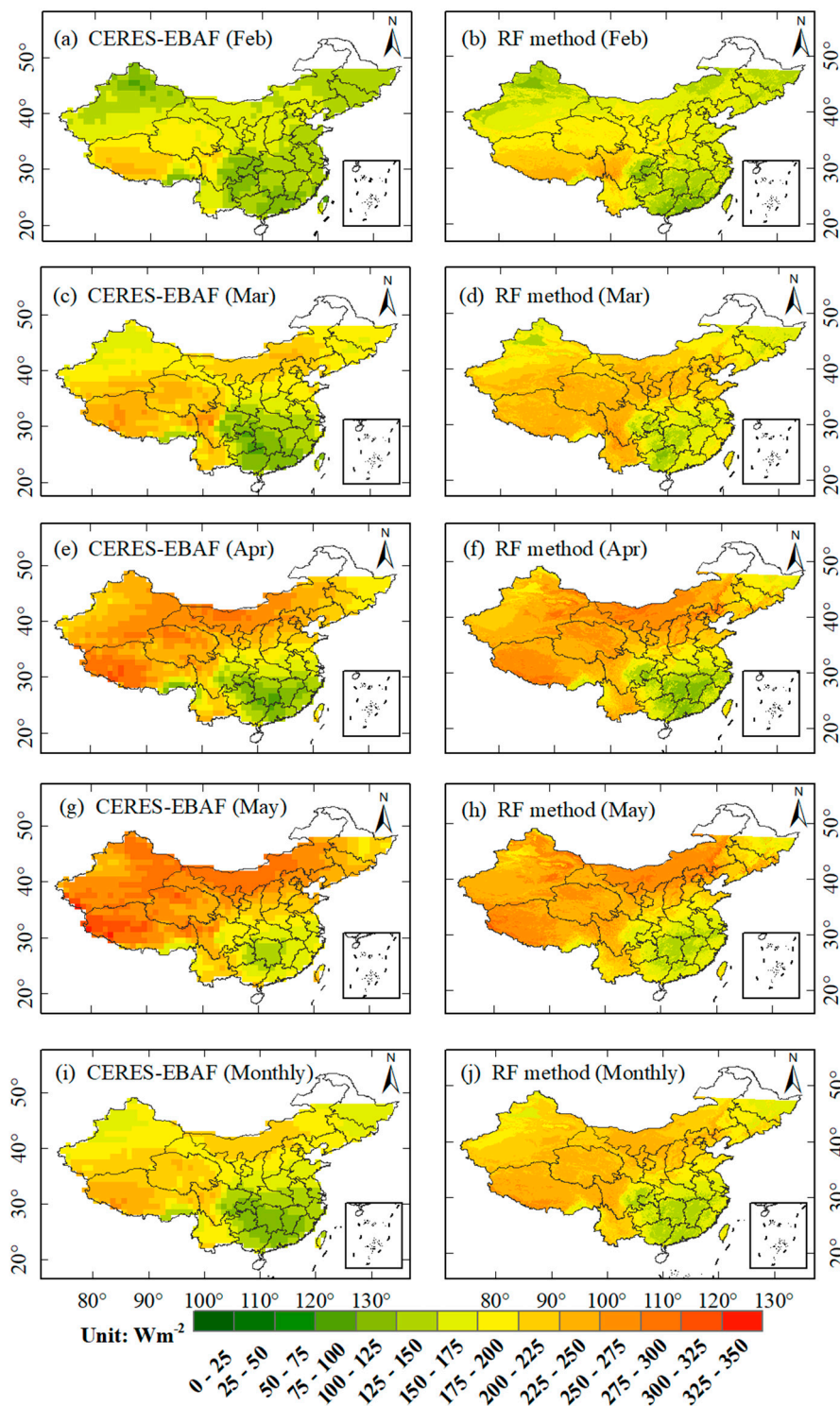
Figure 10. The spatial distribution of the monthly mean estimated R_S data over China of CERES–EBAF and the RF method from February to May 2016: panels (**a**–**h**) for each month; (**i,j**) for monthly mean R_S from February to May.

The CERES–EBAF R_S is one of the most accurate global R_S datasets, according to previous studies; however, some studies pointed out that it tends to overestimate the R_S in China [10,70]. To quantify the difference, the spatial resolution of the estimated R_S data from the RF method was aggregated into 1-degree steps for matching those of the CERES–EBAF R_S data, and the correlation between them was calculated (Figure 11). The estimated R_S data from the RF method agreed reasonably well

with CERES–EBAF data with an R value of 0.91, an RMSE of 22.00 (10.83%) Wm^{-2}, and an MBE of −2.93 (−1.44%) Wm^{-2}. The spatial distribution patterns of their discrepancies are shown in Figure 12. It shows that the differences between the estimated R_S data from the RF method and the CERES–EBAF were relatively large in February and March, especially in southern China. The differences between them varied from −20 Wm^{-2} to 20 Wm^{-2} in April and May. It is also found that the discrepancies between the estimated R_S and the CERES–EBAF gradually changed from negative to positive from February to May 2016. As for the monthly mean R_S from February to May, the differences of the multiple months mean R_S was not as obvious as these in each month. The mean differences values in China were relatively larger in May, and the mean difference values in February, March, and April gradually changed from −11.04 Wm^{-2} to 8.51 Wm^{-2}. Overall, the R_S estimates from Himawari-8 AHI data based on the RF method and from the CERES–EBAF had similar temporal variation and spatial distribution over China. The relatively large differences in southern China between the estimated R_S data from the RF method and the CERES–EBAF were probably related to the influences of aerosols, clouds, and their interactions on the R_S [32–34].
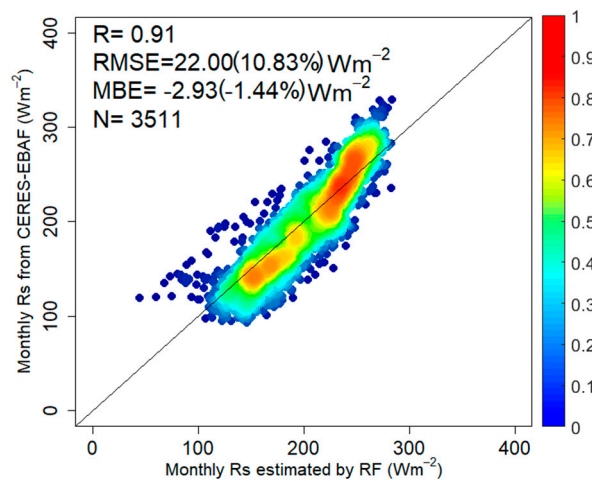

Figure 11. Evaluation results of the estimated monthly mean R_S based on the RF method using CERES–EBAF R_S data from February to May 2016. N is the number of data points.

4.3. Comparison with ANN

The RF method was applied for estimating the R_S using the Himawari-8 AHI data in this study. The ANN method, as the traditional machine learning method, is a non-linear modeling algorithm, which was developed based on the biological structure of the human neuron [84]. The ANN method does not require the knowledge of a specified problem, because the method can learn from examples; moreover, the method can provide flexible mathematical algorithms for different problems. The ANN method has been widely used in various applications including estimating the R_S [3,84]. In this study, the RF method was compared with the commonly used ANN method in estimating the R_S.

The ANN method performance in estimating the daily R_S using Himawari-8 AHI data is displayed in Figure 13. For the training data, the overall R of the estimated R_S data based on the ANN method was 0.90, the RMSE was 41.09 (21.49%) Wm^{-2}, and the MBE was 1.46 (0.76%) Wm^{-2}. For the validation data, the overall R was 0.86, the RMSE was 45.96 (23.90%) Wm^{-2}, and the MBE was 1.48 (0.77%) Wm^{-2}. For the monthly time scale, Figure 14 shows that the overall R of the estimated R_S data based on the ANN method was 0.93, the RMSE was 20.09 (10.62%) Wm^{-2}, and the MBE was 1.87 (0.99%) Wm^{-2}.

The comparison of the results (Table 3) indicated that the RF method was more accurate than the ANN method in estimating the R_S over China at the daily and monthly time scales, and the RF method could estimate the R_S with reasonable accuracy at the daily and monthly time scale.
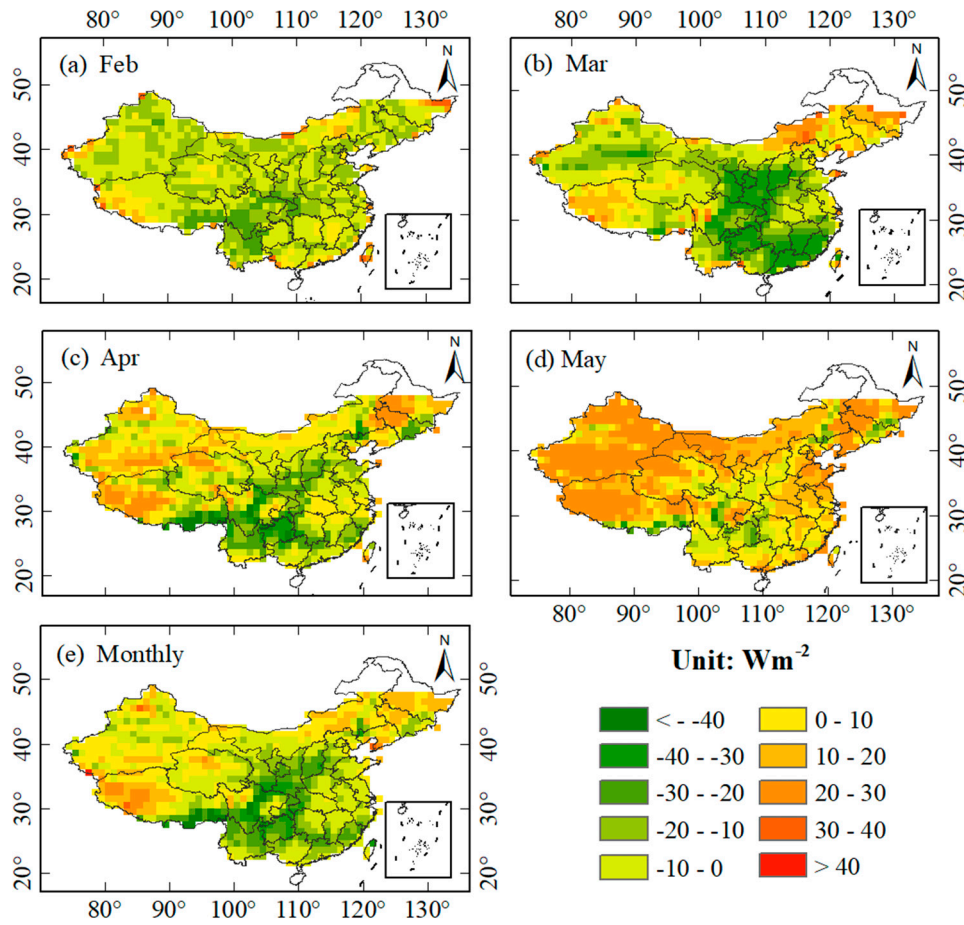
Figure 12. The differences of the monthly mean estimated R_S data between the CERES–EBAF and the RF method (i.e., the CERES–EBAF estimates minus the RF-based estimates) from February to May 2016: panels (**a–d**) for each month; (**e**) for the difference of monthly mean R_S from February to May.
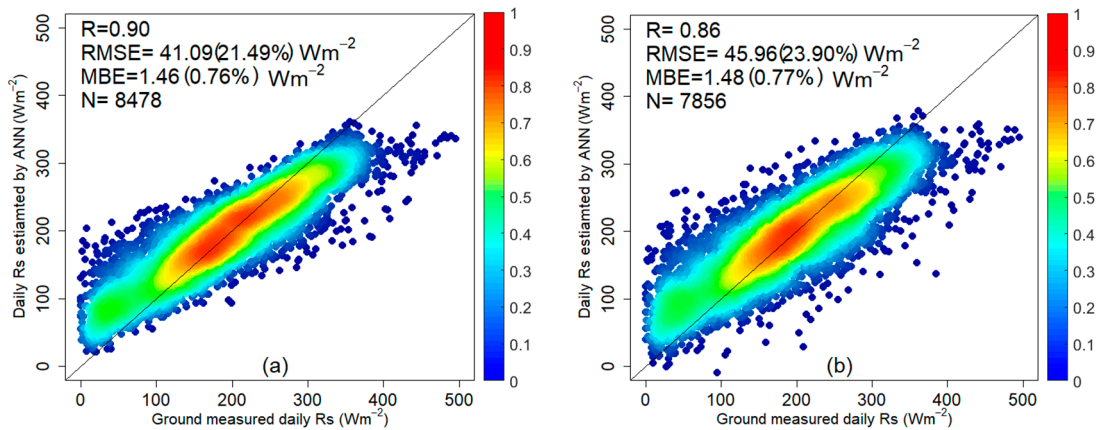

Figure 13. Evaluation results of daily estimated R_S data from Himawari-8 AHI data using the ANN method with ground-measured R_S from CDC/CMA from February to May 2016: (**a**) for training data; (**b**) for validation data. N is the number of data point.
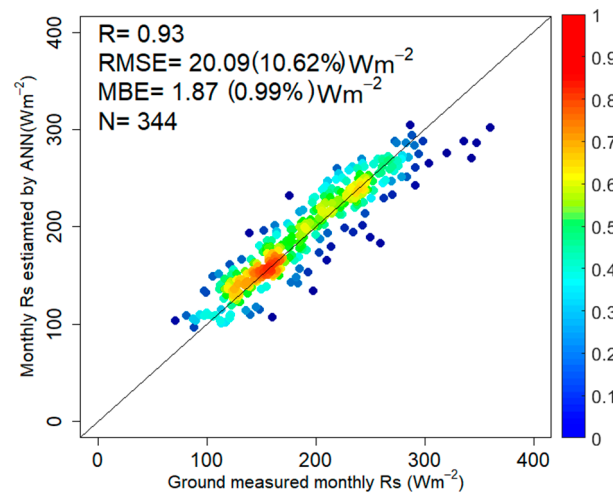
Figure 14. Results of evaluation for monthly mean estimated R_S data from Himawari-8 AHI data using the ANN method with ground-measured R_S from CDC/CMA from February to May 2016. N is the number of data point.

Table 3. Statistical parameters of training and validation data of two models at the daily and monthly time scale.

Time Scale	Data	Method	R	RMSE (Wm^{-2})	MBE (Wm^{-2})
Daily	Training Data	RF	0.99	11.16 (5.83%)	−0.06 (−0.03%)
		ANN	0.90	41.09 (21.49%)	1.46 (0.76%)
	Validation Data	RF	0.92	35.38 (18.40%)	0.01 (0.01%)
		ANN	0.86	45.96 (23.90%)	1.48 (0.77%)
Monthly	All Data	RF	0.99	7.74 (4.09%)	0.03 (0.02%)
		ANN	0.93	20.09 (10.62%)	1.81 (0.99%)

Figure 15 displays the accuracy of spatial distribution of daily mean R_S estimated from Himawari-8 AHI data based on the RF and ANN method at 86 CDC/CMA stations. According to Figure 15, R_S estimates from Himawari-8 AHI data using the RF method agreed well with ground measurements in the majority of the sites, with R values varying from 0.92 to 0.99, MBE values varying from −12.83 to 10.36 Wm^{-2}, and RMSE values varying from 12.04 to 29.21 Wm^{-2}. It is noted that R values were larger than 0.92 in 84 of 86 stations, and MBE values were less than 10 Wm^{-2} in 76 of 86 stations. For the ANN method, the R values varied from 0.70 to 0.96, the MBE values varied from −55.71 to 36.70 Wm^{-2}, and the RMSE values varied from 22.23 to 87.03 Wm^{-2}. This showed that the MBE and RMSE values of the ANN model were greater than those of the RF method at the daily time scale. From the results, one may conclude that the RF method performed better than the ANN method at most stations.
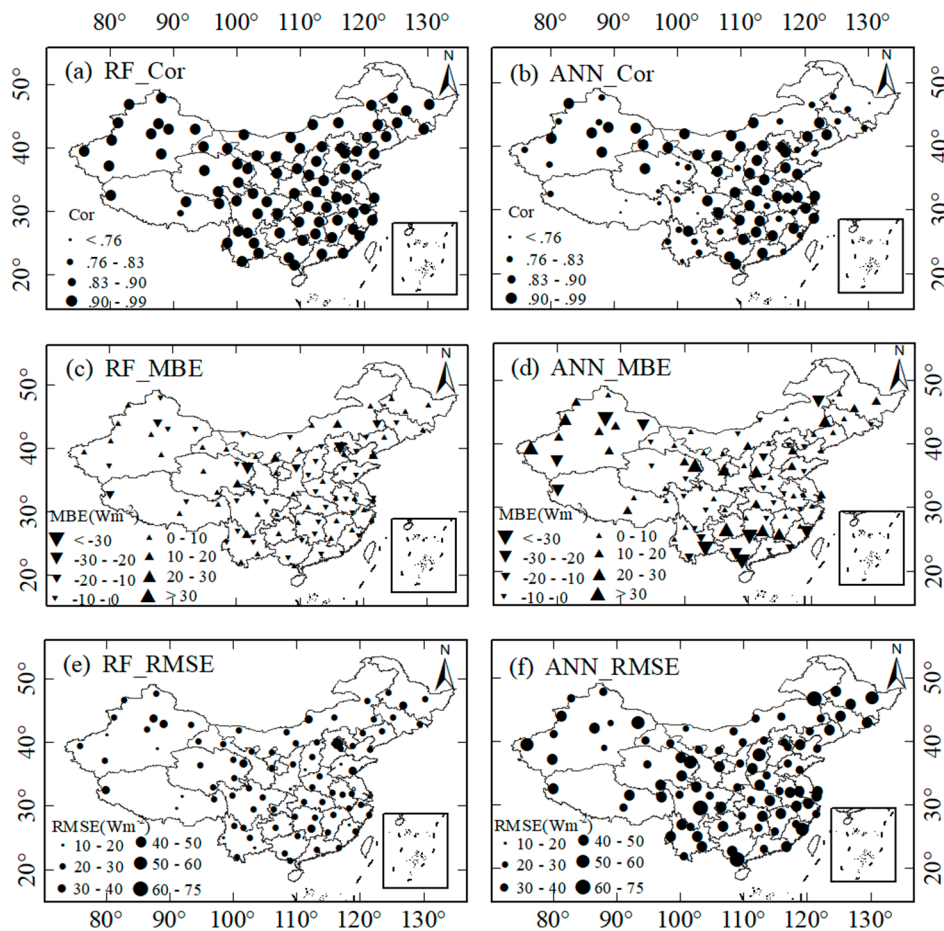
Figure 15. Spatial distribution of accuracy of R_S estimates from Himawari-8 AHI data based on two machine learning methods at a daily time scale at the CDC/CMA stations from February to May 2016: panels (**a**,**b**) are the R between ground measurements and estimates from the RF and ANN methods, respectively; panels (**c**,**d**) are the MBE and panels (**e**,**f**) are the RMSE for two methods, respectively.

5. Discussion

In this study, the RF method was applied to estimate the R_S data directly using the Himawari-8 AHI data. Compared to the ground measurements and CERES–EBAF R_S data, the validation results showed that the estimated R_S data had reasonable accuracy. The results also showed that the differences between the RF method estimates and two other data were relatively large in southern China. This might be related to aerosols and clouds, since previous studies reported that aerosols absorb and scatter the R_S and indirectly influence the R_S by modifying cloud properties [85–87]. Zhang et al. [32] pointed out that southern China includes some well-developed places where heavy pollution occurs. Li et al. [88] pointed out that the aerosol optical depth of southern China in the winter is relatively high. This is related to the monsoon, as well as anthropogenic aerosols, which are caused by pollution due to the rapid economic development in southern China. The aerosol optical depth and aerosol–cloud interaction might make the R_S estimation more difficult and cause relatively larger differences between the estimated R_S data from the RF method and CERES–EBAF in southern China. Since ground-measured clouds and aerosol data were not available at CDC/CMA stations, the influence of aerosols and clouds were not included in this study. The influence of aerosols, clouds, and their interactions on estimating the R_S should be analyzed in future studies. Moreover, urbanization and topographical effects might cause bias in ground measurements [10,89], and their influence on the R_S also should be quantified in future studies.

The RF method had been widely applied in many studies. The RF method has some advantages; for example, the nonlinear interaction can be automatically found by the RF method through regression

tree learning. There are also some disadvantages of the RF method; for example, it may be ineffective because regression trees are regarded as a black box [90]; moreover, it is not easily interpretable. Overfitting problems may occur in applications of the RF method because of complicated regression trees, particularly when noisy data are involved [91]. When the network of ANN grows too large, the ANN method may also have the overfitting issue [92].

6. Conclusions

The RF machine learning method is applied in this study to estimate R_S at the daily and monthly time scales and 2-km spatial resolution using the Himawari-8 AHI data and ancillary datasets. The sensitivity analysis of the RF method parameters shows that the max-features parameter has the most significant influence on the accuracy of estimated R_S data. The R_S estimates from the Himawari-8 AHI data using the RF method are evaluated against ground measurements at 86 CDC/CMA stations from February to May 2016 as a target population. At the daily time scale, the estimated R_S data from the RF method agree well with ground-measured R_S for the validation data with an overall R of 0.92, an RMSE of 35.38 (18.40%) Wm^{-2}, and an MBE of 0.01 (0.01%) Wm^{-2}, while at the monthly time scale, the estimated R_S data from the RF method agree better with ground-measured R_S. The results show that the overall R is 0.99, the RMSE is 7.74 (4.09%) Wm^{-2}, and the MBE is 0.03 (0.02%) Wm^{-2}.

The monthly R_S estimates from the RF method are used to make a comparison to the CERES–EBAF data, the results indicate that the estimated R_S data from the RF method agree well with the CERES–EBAF R_S data. Moreover, the estimated R_S data from the RF method are relatively more accurate than the CERES–EBAF R_S data, when compared with the ground-measured data at the selected stations over China, and they show similar temporal variation and spatial distribution. The differences between the CERES–EBAF R_S data and estimated R_S data from the RF method vary with time; relatively larger mean differences values occur in May, and the mean differences values in February, March, and April gradually change from −11.04 Wm^2 to 8.51 Wm^2. In addition, this study also compares the RF method to the ANN method in estimating the R_S. The results indicate that the RF method has higher accuracy in estimating the R_S from Himawari-8 AHI data at both the daily time scale and monthly time scale than the ANN method.

Overall, the results in this study indicate that the Himawari-8 AHI data based on the RF method can be used to estimate the R_S over China with reasonable accuracy. The estimated R_S data using the Himawari-8 AHI data based on the RF method can be used in other research, such as hydrological and climate change studies, and so on.

References

1. Gui, S.; Liang, S.; Wang, K.; Li, L.; Zhang, X. Assessment of three satellite-estimated land surface downwelling shortwave irradiance data sets. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 776–780. [CrossRef]
2. Liang, S.; Wang, K.; Zhang, X.; Wild, M. Review on estimation of land surface radiation and energy budgets from ground measurement, remote sensing and model simulations. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2010**, *3*, 225–240. [CrossRef]

3. Ryu, Y.; Jiang, C.; Kobayashi, H.; Detto, M. MODIS-derived global land products of shortwave radiation and diffuse and total photosynthetically active radiation at 5 km resolution from 2000. *Remote Sens. Environ.* **2018**, *204*, 812–825. [CrossRef]

4. Yang, K.; He, J.; Tang, W.; Qin, J.; Cheng, C.C.K. On downward shortwave and longwave radiations over high altitude regions: Observation and modeling in the Tibetan Plateau. *Agric. For. Meteorol.* **2010**, *150*, 38–46. [CrossRef]

5. Singh, J.; Kumar, M. Solar radiation over four cities of India: Trend analysis using Mann-Kendall test. *Int. J. Renew. Energy Res.* **2016**, *6*, 1385–1395.

6. Norris, J.R.; Wild, M. Trends in aerosol radiative effects over China and Japan inferred from observed cloud cover, solar "dimming," and solar "brightening". *J. Geophys. Res. Atmos.* **2009**, *114*. [CrossRef]

7. Kambezidis, H.D.; Kaskaoutis, D.G.; Kharol, S.K.; Moorthy, K.K.; Satheesh, S.K.; Kalapureddy, M.C.R.; Badarinath, K.V.S.; Sharma, A.R.; Wild, M. Multi-decadal variation of the net downward shortwave radiation over south Asia: The solar dimming effect. *Atmos. Environ.* **2012**, *50*, 360–372. [CrossRef]

8. Qian, Y.; Wang, W.; Leung, L.R.; Kaiser, D.P. Variability of solar radiation under cloud-free skies in China: The role of aerosols. *Geophys. Res. Lett.* **2007**, *34*. [CrossRef]

9. Liang, S.L.; Zheng, T.; Liu, R.G.; Fang, H.L.; Tsay, S.C.; Running, S. Estimation of incident photosynthetically active radiation from Moderate Resolution Imaging Spectrometer data. *J. Geophys. Res. Atmos.* **2006**, *111*. [CrossRef]

10. Zhang, X.; Liang, S.; Wild, M.; Jiang, B. Analysis of surface incident shortwave radiation from four satellite products. *Remote Sens. Environ.* **2015**, *165*, 186–202. [CrossRef]

11. Liu, X.; Mei, X.; Li, Y.; Zhang, Y.; Wang, Q.; Jensen, J.R.; Porter, J.R. Calibration of the Ångström—Prescott coefficients (a, b) under different time scales and their impacts in estimating global solar radiation in the Yellow River basin. *Agric. For. Meteorol.* **2009**, *149*, 697–710. [CrossRef]

12. Wang, J.; White, K.; Robinson, G.J. Estimating surface net solar radiation by use of Landsat-5 TM and digital elevation models. *Int. J. Remote Sens.* **2000**, *21*, 31–43. [CrossRef]

13. Harries, J.E.; Russell, J.E.; Hanafin, J.A.; Brindley, H.; Futyan, J.; Rufus, J.; Kellock, S.; Matthews, G.; Wrigley, R.; Last, A.; et al. The geostationary earth radiation budget project. *Bull. Am. Meteorol. Soc.* **2005**, *86*, 945–960. [CrossRef]

14. Shahi, N.R.; Agarwal, N.; Sharma, R.; Thapliyal, P.K.; Joshi, P.C.; Sarkar, A. Improved estimation of shortwave radiation over equatorial Indian Ocean using geostationary satellite data. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 563–566. [CrossRef]

15. Zhang, X.; Liang, S.; Zhou, G.; Wu, H.; Zhao, X. Generating Global LAnd Surface Satellite incident shortwave radiation and photosynthetically active radiation products from multiple satellite data. *Remote Sens. Environ.* **2014**, *152*, 318–332. [CrossRef]

16. Wang, K.; Zhou, X.; Liu, J.; Sparrow, M. Estimating surface solar radiation over complex terrain using moderate-resolution satellite sensor data. *Int. J. Remote Sens.* **2005**, *26*, 47–58. [CrossRef]

17. Cess, R.D.; Dutton, E.G.; Deluisi, J.J.; Jiang, F. Determining surface solar absorption from broadband satellite measurements for clear skies: Comparison with surface measurements. *J. Clim.* **1991**, *4*, 236–247. [CrossRef]

18. Tang, W.J.; Yang, K.; Sun, Z.; Qin, J.; Niu, X.L. Global performance of a fast parameterization scheme for estimating surface solar radiation from MODIS data. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3558–3571. [CrossRef]

19. Qin, J.; Tang, W.J.; Yang, K.; Lu, N.; Niu, X.L.; Liang, S.L. An efficient physically based parameterization to derive surface solar irradiance based on satellite atmospheric products. *J. Geophys. Res. Atmos.* **2015**, *120*, 4975–4988. [CrossRef]

20. Mueller, R.W.; Matsoukas, C.; Gratzki, A.; Behr, H.D.; Hollmann, R. The CM-SAF operational scheme for the satellite based retrieval of solar surface irradiance—A LUT based eigenvector hybrid approach. *Remote Sens. Environ.* **2009**, *113*, 1012–1024. [CrossRef]

21. Xia, S.; Mestas-Nuñez, A.; Xie, H.; Tang, J.; Vega, R. Characterizing variability of solar irradiance in San Antonio, Texas using satellite observations of cloudiness. *Remote Sens.* **2018**, *10*, 2016. [CrossRef]

22. Chen, L.; Yan, G.; Wang, T.; Ren, H.; Hu, R.; Chen, S.; Zhou, H. Spatial scale consideration for estimating all-sky surface shortwave radiation with a modified 1-D radiative transfer model. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 821–835. [CrossRef]

23. Zhang, Y.; He, T.; Liang, S.; Wang, D.; Yu, Y. Estimation of all-sky instantaneous surface incident shortwave radiation from Moderate Resolution Imaging Spectroradiometer data using optimization method. *Remote Sens. Environ.* **2018**, *209*, 468–479. [CrossRef]

24. Kato, S.; Loeb, N.G.; Rose, F.G.; Doelling, D.R.; Rutan, D.A.; Caldwell, T.E.; Yu, L.; Weller, R.A. Surface irradiances consistent with CERES-derived top-of-atmosphere shortwave and longwave irradiances. *J. Clim.* **2013**, *26*, 2719–2740. [CrossRef]

25. Kato, S.; Rose, F.G.; Sun-Mack, S.; Miller, W.F.; Chen, Y.; Rutan, D.A.; Stephens, G.L.; Loeb, N.G.; Minnis, P.; Wielicki, B.A.; et al. Improvements of top-of-atmosphere and surface irradiance computations with CALIPSO-, CloudSat-, and MODIS-derived cloud and aerosol properties. *J. Geophys. Res. Atmos.* **2011**, *116*. [CrossRef]

26. Romano, F.; Cimini, D.; Cersosimo, A.; Di Paola, F.; Gallucci, D.; Gentile, S.; Geraldi, E.; Larosa, S.; Nilo, S.T.; Ricciardelli, E.; et al. Improvement in surface solar irradiance estimation using HRV/MSG data. *Remote Sens.* **2018**, *10*, 1288. [CrossRef]

27. Riihelä, A.; Kallio, V.; Devraj, S.; Sharma, A.; Lindfors, A. Validation of the SARAH-E satellite-based surface solar radiation estimates over India. *Remote Sens.* **2018**, *10*, 392. [CrossRef]

28. Olpenda, A.; Stereńczak, K.; Będkowski, K. Modeling solar radiation in the forest using remote sensing data: A review of approaches and opportunities. *Remote Sens.* **2018**, *10*, 694. [CrossRef]

29. Lee, S.-H.; Kim, B.-Y.; Lee, K.-T.; Zo, I.-S.; Jung, H.-S.; Rim, S.-H. Retrieval of reflected shortwave radiation at the top of the atmosphere using Himawari-8/AHI data. *Remote Sens.* **2018**, *10*, 213. [CrossRef]

30. Ineichen, P. High turbidity solis clear sky model: Development and validation. *Remote Sens.* **2018**, *10*, 435. [CrossRef]

31. Kambezidis, H.D.; Psiloglou, B.E.; Karagiannis, D.; Dumka, U.C.; Kaskaoutis, D.G. Meteorological Radiation Model (MRM v6.1): Improvements in diffuse radiation estimates and a new approach for implementation of cloud products. *Renew. Sustain. Energy Rev.* **2017**, *74*, 616–637. [CrossRef]

32. Zhang, M.; Wang, Y.; Ma, Y.Y.; Wang, L.C.; Gong, W.; Liu, B.M. Spatial distribution and temporal variation of aerosol optical depth and radiative effect in South China and its adjacent area. *Atmos. Environ.* **2018**, *188*, 120–128. [CrossRef]

33. Feng, F.; Wang, K. Determining factors of monthly to decadal variability in surface solar radiation in China: Evidences from current reanalyses. *J. Geophys. Res. Atmos.* **2019**, *124*, 9161–9182. [CrossRef]

34. Deneke, H.M.; Feijt, A.J.; Roebeling, R.A. Estimating surface solar irradiance from METEOSAT SEVIRI-derived cloud properties. *Remote Sens. Environ.* **2008**, *112*, 3131–3141. [CrossRef]

35. Levy, R.C.; Remer, L.A.; Kleidman, R.G.; Mattoo, S.; Ichoku, C.; Kahn, R.; Eck, T.F. Global evaluation of the Collection 5 MODIS dark-target aerosol products over land. *Atmos. Chem. Phys.* **2010**, *10*, 10399–10420. [CrossRef]

36. Deo, R.C.; Şahin, M. Forecasting long-term global solar radiation with an ANN algorithm coupled with satellite-derived (MODIS) land surface temperature (LST) for regional locations in Queensland. *Renew. Sustain. Energy Rev.* **2017**, *72*, 828–848. [CrossRef]

37. Wang, T.; Yan, G.; Chen, L. Consistent retrieval methods to estimate land surface shortwave and longwave radiative flux components under clear-sky conditions. *Remote Sens. Environ.* **2012**, *124*, 61–71. [CrossRef]

38. Zhou, Q.T.; Flores, A.; Glenn, N.F.; Walters, R.; Hang, B.S. A machine learning approach to estimation of downward solar radiation from satellite-derived data products: An application over a semi-arid ecosystem in the US. *PLoS ONE* **2017**, *12*, e0180239. [CrossRef]

39. Jiang, B.; Liang, S.; Ma, H.; Zhang, X.; Xiao, Z.; Zhao, X.; Jia, K.; Yao, Y.; Jia, A. GLASS daytime all-wave net radiation product: Algorithm development and preliminary validation. *Remote Sens.* **2016**, *8*, 222. [CrossRef]

40. Yang, L.; Zhang, X.T.; Liang, S.L.; Yao, Y.J.; Jia, K.; Jia, A.L. Estimating surface downward shortwave radiation over China based on the gradient boosting decision tree method. *Remote Sens.* **2018**, *10*, 185. [CrossRef]

41. Tang, W.J.; Qin, J.; Yang, K.; Liu, S.M.; Lu, N.; Niu, X.L. Retrieving high-resolution surface solar radiation with cloud parameters derived by combining MODIS and MTSAT data. *Atmos. Chem. Phys.* **2016**, *16*, 2543–2557. [CrossRef]

42. Wei, Y.; Zhang, X.; Hou, N.; Zhang, W.; Jia, K.; Yao, Y. Estimation of surface downward shortwave radiation over China from AVHRR data based on four machine learning methods. *Sol. Energy* **2019**, *177*, 32–46. [CrossRef]

43. Ghimire, S.; Deo, R.C.; Raj, N.; Mi, J. Wavelet-based 3-phase hybrid SVR model trained with satellite-derived predictors, particle swarm optimization and maximum overlap discrete wavelet transform for solar radiation prediction. *Renew. Sustain. Energy Rev.* **2019**, *113*, 109247. [CrossRef]

44. Fan, J.; Wu, L.; Ma, X.; Zhou, H.; Zhang, F. Hybrid support vector machines with heuristic algorithms for prediction of daily diffuse solar radiation in air-polluted regions. *Renew. Energy* **2020**, *145*, 2034–2045. [CrossRef]

45. Chen, F.; Zhou, Z.; Lin, A.; Niu, J.; Qin, W.; Yang, Z. Evaluation of direct horizontal irradiance in China using a physically-based model and machine learning methods. *Energies* **2019**, *12*, 150. [CrossRef]

46. Hao, D.; Asrar, G.R.; Zeng, Y.; Zhu, Q.; Wen, J.; Xiao, Q.; Chen, M. Estimating hourly land surface downward shortwave and photosynthetically active radiation from DSCOVR/EPIC observations. *Remote Sens. Environ.* **2019**, *232*, 111320. [CrossRef]

47. 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]

48. Kim, B.-Y.; Lee, K.-T.; Jee, J.-B.; Zo, I.-S. Retrieval of outgoing longwave radiation at top-of-atmosphere using Himawari-8 AHI data. *Remote Sens. Environ.* **2018**, *204*, 498–508. [CrossRef]

49. Damiani, A.; Irie, H.; Horio, T.; Takamura, T.; Khatri, P.; Takenaka, H.; Nagao, T.; Nakajima, T.Y.; Cordero, R.R. Evaluation of Himawari-8 surface downwelling solar radiation by ground-based measurements. *Atmos. Meas. Tech.* **2018**, *11*, 2501–2521. [CrossRef]

50. Bessho, K.; Date, K.; Hayashi, M.; Ikeda, A.; Imai, T.; Inoue, H.; Kumagai, Y.; Miyakawa, T.; Murata, H.; Ohno, T.; et al. An introduction to Himawari-8/9-Japan's new-generation geostationary meteorological satellites. *J. Meteorol. Soc. Jpn.* **2016**, *94*, 151–183. [CrossRef]

51. Frouin, R.; Murakami, H. Estimating photosynthetically available radiation at the ocean surface from ADEOS-II Global Imager data. *J. Oceanogr.* **2007**, *63*, 493–503. [CrossRef]

52. Shi, H.; Li, W.; Fan, X.; Zhang, J.; Hu, B.; Husi, L.; Shang, H.; Han, X.; Song, Z.; Zhang, Y.; et al. First assessment of surface solar irradiance derived from Himawari-8 across China. *Sol. Energy* **2018**, *174*, 164–170. [CrossRef]

53. Yu, Y.; Shi, J.; Wang, T.; Letu, H.; Yuan, P.; Zhou, W.; Hu, L. Evaluation of the Himawari-8 Shortwave Downward Radiation (SWDR) product and its comparison with the CERES-SYN, MERRA-2, and ERA-interim datasets. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *12*, 519–532. [CrossRef]

54. Sawada, Y.; Okamoto, K.; Kunii, M.; Miyoshi, T. Assimilating every-10-minute Himawari-8 infrared radiances to improve convective predictability. *J. Geophys. Res. Atmos.* **2019**, *124*, 2546–2561. [CrossRef]

55. Schmit, T.J.; Griffith, P.; Gunshor, M.M.; Daniels, J.M.; Goodman, S.J.; Lebair, W.J. A closer look at the ABI on the GOES-R series. *Bull. Am. Meteorol. Soc.* **2016**, *98*, 681–698. [CrossRef]

56. Zhang, W.; Xu, H.; Zheng, F. Aerosol optical depth retrieval over East Asia using Himawari-8/AHI data. *Remote Sens.* **2018**, *10*, 137. [CrossRef]

57. Gelaro, R.; McCarty, W.; Suárez, M.J.; Todling, R.; Molod, A.; Takacs, L.; Randles, C.A.; Darmenov, A.; Bosilovich, M.G.; Reichle, R.; et al. The modern-era retrospective analysis for research and applications, Version 2 (MERRA-2). *J. Clim.* **2017**, *30*, 5419–5454. [CrossRef]

58. Hinkelman, L.M. The global radiative energy budget in MERRA and MERRA-2: Evaluation with respect to CERES EBAF data. *J. Clim.* **2019**, *32*, 1973–1994. [CrossRef]

59. Matuszko, D. Long-term variability in solar radiation in Krakow based on measurements of sunshine duration. *Int. J. Climatol.* **2014**, *34*, 228–234. [CrossRef]

60. He, Y.; Wang, K.; Zhou, C.; Wild, M. A revisit of global dimming and brightening based on the sunshine duration. *Geophys. Res. Lett.* **2018**, *45*, 4281–4289. [CrossRef]

61. Yang, K.; Koike, T.; Stackhouse, P.; Mikovitz, C.; Cox, S.J. An assessment of satellite surface radiation products for highlands with Tibet instrumental data. *Geophys. Res. Lett.* **2006**, *33*. [CrossRef]

62. Moradi, I. Quality control of global solar radiation using sunshine duration hours. *Energy* **2009**, *34*, 1–6. [CrossRef]

63. Tang, W.; Yang, K.; He, J.; Qin, J. Quality control and estimation of global solar radiation in China. *Sol. Energy* **2010**, *84*, 466–475. [CrossRef]

64. Tang, W.J.; Yang, K.; Qin, J.; Cheng, C.C.K.; He, J. Solar radiation trend across China in recent decades: A revisit with quality-controlled data. *Atmos. Chem. Phys.* **2011**, *11*, 393–406. [CrossRef]

65. Tang, W.J.; Qin, J.; Yang, K.; Niu, X.L.; Zhang, X.T.; Yu, Y.; Zhu, X.D. Reconstruction of daily photosynthetically active radiation and its trends over China. *J. Geophys. Res. Atmos.* **2013**, *118*, 13292–13302. [CrossRef]

66. Loeb, N.G.; Wielicki, B.A.; Doelling, D.R.; Smith, G.L.; Keyes, D.F.; Kato, S.; Manalo-Smith, N.; Wong, T. Toward optimal closure of the earth's top-of-atmosphere radiation budget. *J. Clim.* **2009**, *22*, 748–766. [CrossRef]

67. Kato, S.; Rose, F.G.; Rutan, D.A.; Thorsen, T.J.; Loeb, N.G.; Doelling, D.R.; Huang, X.; Smith, W.L.; Su, W.; Ham, S.-H. Surface irradiances of edition 4.0 clouds and the earth's radiant energy system (CERES) energy balanced and filled (EBAF) data product. *J. Clim.* **2018**, *31*, 4501–4527. [CrossRef]

68. Su, W.; Corbett, J.; Eitzen, Z.; Liang, L. Next-generation angular distribution models for top-of-atmosphere radiative flux calculation from CERES instruments: Methodology. *Atmos. Meas. Tech.* **2015**, *8*, 1467. [CrossRef]

69. Loeb, N.G.; Doelling, D.R.; Wang, H.; Su, W.; Nguyen, C.; Corbett, J.G.; Liang, L.; Mitrescu, C.; Rose, F.G.; Kato, S. Clouds and the earth's radiant energy system (CERES) energy balanced and filled (EBAF) top-of-atmosphere (TOA) edition-4.0 data product. *J. Clim.* **2018**, *31*, 895–918. [CrossRef]

70. Ma, Q.; Wang, K.; Wild, M. Impact of geolocations of validation data on the evaluation of surface incident shortwave radiation from Earth System Models. *J. Geophys. Res. Atmos.* **2015**, *120*, 6825–6844. [CrossRef]

71. Loeb, N. The Climate Data Guide: CERES EBAF: Clouds and Earth's Radiant Energy Systems (CERES) Energy Balanced and Filled (EBAF). Available online: https://climatedataguide.ucar.edu/climate-data/ceres-ebaf-clouds-and-earths-radiant-energy-systems-ceres-energybalanced-and-filled (accessed on 1 February 2019).

72. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

73. Yu, P.S.; Yang, T.C.; Chen, S.Y.; Kuo, C.M.; Tseng, H.W. Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting. *J. Hydrol.* **2017**, *552*, 92–104. [CrossRef]

74. Zamo, M.; Mestre, O.; Arbogast, P.; Pannekoucke, O. A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production. Part II: Probabilistic forecast of daily production. *Sol. Energy* **2014**, *105*, 804–816. [CrossRef]

75. Prasad, A.M.; Iverson, L.R.; Liaw, A. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems* **2006**, *9*, 181–199. [CrossRef]

76. Verbyla, D.L. Classification trees—A new discrimination tool. *Can. J. For. Res.* **1987**, *17*, 1150–1152. [CrossRef]

77. Gislason, P.O.; Benediktsson, J.A.; Sveinsson, J.R. Random Forests for land cover classification. *Pattern Recognit. Lett.* **2006**, *27*, 294–300. [CrossRef]

78. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2012**, *12*, 2825–2830.

79. Ibrahim, I.A.; Khatib, T. A novel hybrid model for hourly global solar radiation prediction using random forests technique and firefly algorithm. *Energy Convers. Manag.* **2017**, *138*, 413–425. [CrossRef]

80. Qu, Z.; Gschwind, B.; Lefevre, M.; Wald, L. Improving HelioClim-3 estimates of surface solar irradiance using the McClear clear-sky model and recent advances in atmosphere composition. *Atmos. Meas. Tech.* **2014**, *7*, 3927–3933. [CrossRef]

81. Psiloglou, B.E.; Kambezidis, H.D.; Kaskaoutis, D.G.; Karagiannis, D.; Polo, J.M. Comparison between MRM simulations, CAMS and PVGIS databases with measured solar radiation components at the Methoni station, Greece. *Renew. Energy* **2020**, *146*, 1372–1391. [CrossRef]

82. Linares-Rodriguez, A.; Ruiz-Arias, J.A.; Pozo-Vazquez, D.; Tovar-Pescador, J. An artificial neural network ensemble model for estimating global solar radiation from Meteosat satellite images. *Energy* **2013**, *61*, 636–645. [CrossRef]

83. Massey, F.J. The Kolmogorov-Smirnov test for goodness of fit. *J. Am. Stat. Assoc.* **1951**, *46*, 68–78. [CrossRef]

84. Meenal, R.; Selvakumar, A.I. Assessment of SVM, empirical and ANN based solar radiation prediction models with most influencing input parameters. *Renew. Energy* **2018**, *121*, 324–343. [CrossRef]

85. Albrecht, B.A. Aerosols, cloud microphysics, and fractional cloudiness. *Science* **1989**, *245*, 1227–1230. [CrossRef] [PubMed]

86. Hatzianastassiou, N.; Matsoukas, C.; Fotiadi, A.; Pavlakis, K.G.; Drakakis, E.; Hatzidimitriou, D.; Vardavas, I. Global distribution of Earth's surface shortwave radiation budget. *Atmos. Chem. Phys.* **2005**, *5*, 2847–2867. [CrossRef]

87.   Yu, R.C.; Yu, Y.Q.; Zhang, M.H. Comparing cloud radiative properties between the eastern China and the Indian monsoon region. *Adv. Atmos. Sci.* **2001**, *18*, 1090–1102.

88.   Li, Z.; Lau, W.K.M.; Ramanathan, V.; Wu, G.; Ding, Y.; Manoj, M.G.; Liu, J.; Qian, Y.; Li, J.; Zhou, T.; et al. Aerosol and monsoon climate interactions over Asia. *Rev. Geophys.* **2016**, *54*, 866–929. [CrossRef]

89.   Wang, K.; Ma, Q.; Wang, X.; Wild, M. Urban impacts on mean and trend of surface incident solar radiation. *Geophys. Res. Lett.* **2014**, *41*, 4664–4668. [CrossRef]

90.   Johnson, R.; Zhang, T. Learning nonlinear functions using regularized greedy forest. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 942–954. [CrossRef]

91.   Salles, T.; Goncalves, M.; Rodrigues, V.; Rocha, L. Improving random forests by neighborhood projection for effective text classification. *Inf. Syst.* **2018**, *77*, 1–21. [CrossRef]

92.   Hagiwara, K.; Fukumizu, K. Relation between weight size and degree of over-fitting in neural network regression. *Neural Netw.* **2008**, *21*, 48–58. [CrossRef] [PubMed]