Estimation of Long-Term Surface Downward Longwave Radiation over the Global Land from 2000 to 2018

Chunjie Feng, Xiaotong Zhang *, Yu Wei, Weiyu Zhang, Ning Hou, Jiawen Xu, Shuyue Yang, Xianhong Xie and Bo Jiang

Abstract: It is of great importance for climate change studies to construct a worldwide, long-term surface downward longwave radiation ($L_d$, 4–100 µm) dataset. Although a number of global $L_d$ datasets are available, their low accuracies and coarse spatial resolutions limit their applications. This study generated a daily $L_d$ dataset with a 5-km spatial resolution over the global land surface from 2000 to 2018 using atmospheric parameters, which include 2-m air temperature (Ta), relative humidity (RH) at 1000 hPa, total column water vapor (TCWV), surface downward shortwave radiation ($S_d$), and elevation, based on the gradient boosting regression tree (GBRT) method. The generated $L_d$ dataset was evaluated using ground measurements collected from AmeriFlux, AsiaFlux, baseline surface radiation network (BSRN), surface radiation budget network (SURFRAD), and FLUXNET networks. The validation results showed that the root mean square error (RMSE), mean bias error (MBE), and correlation coefficient (R) values of the generated daily $L_d$ dataset were 17.78 W m$^{-2}$, 0.99 W m$^{-2}$, and 0.96 ($p < 0.01$). Comparisons with other global land surface radiation products indicated that the generated $L_d$ dataset performed better than the clouds and earth’s radiant energy system synoptic (CERES-SYN) edition 4.1 dataset and ERA5 reanalysis product at the selected sites. In addition, the analysis of the spatiotemporal characteristics for the generated $L_d$ dataset showed an increasing trend of 1.8 W m$^{-2}$ per decade ($p < 0.01$) from 2003 to 2018, which was closely related to Ta and water vapor pressure. In general, the generated $L_d$ dataset has a higher spatial resolution and accuracy, which can contribute to perfecting the existing radiation products.

Keywords: surface downward longwave radiation; air temperature; relative humidity; surface downward shortwave radiation; total column water vapor; gradient boosting regression tree

1. Introduction

The surface downward longwave radiation ($L_d$, 4–100 µm) is an indispensable component needed to study the Earth’s surface radiation budget and energy balance [1]. Currently, there are four main ways of obtaining $L_d$: ground measurement data, reanalysis retrieval methods, general circulation model (GCM) simulations and satellite products. However, $L_d$ is not always treated as a conventional observation as other common meteorological parameters are, such as air temperature (Ta), relative humidity (RH), etc. Moreover, its observation stations are sparsely distributed and even entirely absent in certain areas due to a high cost, a difficult calibration process, and a required quality control step [2–5]. In addition, there are uncertainties and biases in GCM simulations [6–8], reanalysis retrievals [9,10], and satellite products [11]. Therefore, establishing a more accurate long-term global $L_d$ dataset is not only useful for improving the knowledge of the surface radiation balance but is also helpful for perfecting the existing $L_d$ products.
Under clear-sky conditions, $L_d$ is primarily influenced by temperature profiles and water vapor in the lower atmosphere. Zeppetello et al. [12] found that $L_d$ is tightly coupled to surface temperature, and changes in surface temperature cause at least 63% of the clear-sky $L_d$ response in greenhouse forcing. Water vapor is the most crucial atmospheric gas contributing to thermal radiation which can absorb and emit longwave radiation, thereby resulting in $L_d$ estimates with great uncertainty [13]. RH, which is closely related to water vapor pressure, is the percentage of water vapor pressure in the atmosphere to the saturated vapor pressure at a given temperature. Numerous studies [5,14–20] have estimated $L_d$ on the basis of traditional methods using $T_a$, water vapor, RH, and other basic variables derived from meteorological observations. These methods mainly include empirical, physics-based, and hybrid methods. Among them, empirical models, including the representative Brunt [14] and Brutsaert [15] equations, establish the regression relationship between various meteorological parameters and $L_d$ observations, with an accuracy that is mainly limited by ground measurements and actual geographical environments, such as climate and terrain. Although this method is relatively simple, it is difficult to apply to $L_d$ estimation on a large regional scope. Compared with empirical methods, physics-based methods containing the LOWTRAN and MODTRAN models can not only estimate $L_d$ with a high accuracy but also describe the atmospheric radiative transfer process in detail [21–23]. Due to the intricacy of the model and the difficulty associated with obtaining an input dataset, however, this approach is only used for research and is difficult to apply to business products [4]. Hybrid methods [24–28] establish the relationship between $L_d$ and the top-of-atmosphere radiance on the basis of physical radiative transfer processes. In contrast, this method, with its higher simulation accuracy and greater general applicability, can be applied on a global scale, which has become an effective method for $L_d$ retrieval. For example, Wang et al. [27] developed a hybrid method to estimate instantaneous land clear $L_d$ on the basis of extensive radiative transfer simulation and statistical analysis, obtaining root mean squared error (RMSE) values of 17.60 W m$^{-2}$ (Terra) and 16.17 W m$^{-2}$ (Aqua) for the nonlinear models.

Under cloudy conditions, the influence of clouds on $L_d$ is also nonnegligible. Clouds are visible polymers of tiny water droplets or ice crystals formed by the condensation of water vapor in the atmosphere, which can absorb heat from the ground and radiate it back to the surface to enhance $L_d$ [29,30]. The cloud cover fraction is mostly utilized to quantify the effects of clouds on $L_d$ and is an essential parameter for $L_d$ estimation under cloudy conditions, which can be obtained from ground measurements and satellite cloud detection products [31–33]. However, the effects of clouds cannot be corrected when cloud cover fraction observations are not available. Crawford et al. [34] first proposed that the cloud cover fraction under cloudy-sky conditions can be estimated from the proportion of the observed surface downward shortwave radiation ($S_d$) to the theoretical clear-sky $S_d$ under the same conditions. They evaluated the performance of estimating $L_d$ using $S_d$, barometric pressure, vapor pressure, and temperature datasets. The evaluation results showed that the RMSEs and mean bias errors (MBE) of the monthly $L_d$ estimates ranged from 11 to 22 W m$^{-2}$ and −9 to 4 W m$^{-2}$ compared to ground observations over a one-year time period, respectively, which indicated that it is reliable to use $S_d$ to represent the impact of clouds on $L_d$. It is easier to obtain $S_d$ data compared with cloud cover fractions, so an increasing number of studies have utilized $S_d$ to estimate $L_d$ under cloudy conditions [5,13,35–37]. Choi et al. [35] estimated the daily $L_d$ using 2-m air temperature, 2-m RH, and $S_d$ observations in Florida from 2004 to 2005, obtaining RMSEs of less than 13 W m$^{-2}$ and squared correlation coefficients ($R^2$) of more than 0.9 relative to the ground measurements collected at 11 stations. Lhomme et al. [13] demonstrated that the cloud correction function of the Crawford et al. [34] model also performed relatively credibly for estimating $L_d$ in high elevation regions between 3700 and 4100 m above sea level. The presence of clouds makes it impossible for satellites to accurately observe surface information. It is also difficult to model the properties of clouds due to the uncertainty...
associated with their distribution and variability. The ready availability of \( S_d \) data makes the \( L_d \) estimation model more readily applicable under cloudy conditions.

In addition, \( S_d \) and \( L_d \) both show a strong dependence on altitude. Zeng et al. [38] evaluated the global land surface satellite (GLASS) \( L_d \) product using the ground observations collected from 141 stations in six networks at different surface elevations. The RMSE values are 22.09, 23.31, 26.94, and 26.99 W m\(^{-2}\) at elevations of <500, 500–1000, 1000–3000, and >3000 m, respectively. The bias values are −3.19, −4.73, −2.26, and 15.34 W m\(^{-2}\) at the four elevation intervals, respectively. The validation results showed that the performance of \( L_d \) degraded as the surface elevation increased. This may be due to special environmental conditions present at high altitudes with lower air pressures, smaller water vapor densities, and fewer clouds, leading to a greater uncertainty in \( S_d \) and \( L_d \) data at high elevations [37,39–43]. In addition, some studies have also quantitatively measured the effect of elevation on \( L_d \) and attempted to correct its deviation [37,39–42]. Yang et al. [42] reported that the MBE of GEWEX-SRB V2.5 \( L_d \) can be reduced by 7–10 W m\(^{-2}\) after an altitudinal correction of 2.8 W m\(^{-2}\) per hundred meters in the Tibet Plateau. It can be concluded that the influence of elevation cannot be ignored in addition to the abovementioned influencing factors including temperature profiles, water vapor, and clouds. Although the importance of elevation has been verified by previous studies, few studies have taken elevation as an important variable to predict \( L_d \). This paper used elevation as the input variable of the model, hoping to reduce the errors caused by elevation.

Based on the above summary, it is clear that \( L_d \) is closely related to \( T_a \), RH, water vapor, \( S_d \), and elevation. Therefore, this study utilized the gradient boosting regression tree (GBRT) method with the daily mean \( T_a \) of 2 m, RH at 1000 hPa, total column water vapor, \( S_d \), and elevation to estimate daily \( L_d \) over global land surface from 2000 to 2018. In contrast to prior methods, this machine learning method can automatically establish the relationship between the input data and target variable, and has a strong predictive ability [44,45], which has been widely employed to retrieve radiation [46–49]. Yang et al. [46] applied the GBRT method to estimate daily \( S_d \) with a spatial resolution of 5 km in China using ground observations and satellite retrievals with good results. The RMSE and R between the ground measurements and daily \( L_d \) estimates were 27.71 W m\(^{-2}\) and 0.91, respectively, under cloudy conditions; these values were 42.97 W m\(^{-2}\) and 0.80, respectively, under clear conditions. To date, few studies have used this method to predict \( L_d \) over the globe based on ground observations. We demonstrated that it can be reasonably and reliably used for \( L_d \) estimation by building the relationship between \( L_d \) observations and its influencing factors based on the GBRT method [49,50]. Therefore, the objective of this study is to use the GBRT model to generate a 5-km \( L_d \) dataset over the global land surface with a daily time scale from 2000 to 2018.

The structure of this paper is as follows: Section 2 introduces the data used, including the ground measurements, ERA5 reanalysis data, GLASS \( S_d \), global multi-resolution terrain elevation data 2010, and existing \( L_d \) products. The detailed model construction process is displayed and described in Section 3. Section 4 provides the evaluation results and analyzes the spatiotemporal distribution of \( L_d \). Finally, the discussion and conclusion are presented in Sections 5 and 6, respectively.

2. Data

2.1. Ground Measurements

The ground measurements of surface downward longwave radiation (\( L_d \)) used in this study from 2000 to 2018 were collected from the AmeriFlux network (175 sites), AsiaFlux network (26 sites), baseline surface radiation network (BSRN, 57 sites), surface radiation budget network (SURFRAD, 7 sites), and FLUXNET (84 sites). The observation sites were randomly divided into 90% (314 sites) and 10% (35 sites) datasets, as shown in Figure 1. After removing the outliers, the \( L_d \) observations collected at 314 sites were used as target variables to build and train the model. The remaining \( L_d \) observations collected at 35 sites were used to evaluate the generated global land daily \( L_d \). The spatial distribution of the
observation sites used to build the model and validate it is shown in Figure 1. The detailed information of ground sites is listed in the Appendix A Table A1.

![Geographical distribution of observation sites used to model (314 sites in total, green) and validate (35 sites in total, red) the $L_d$ dataset in this study collected at AmeriFlux (squares) with 159 and 16 sites, AsiaFlux (pentagrams) with 23 and 3 sites, BSRN networks (circles) with 51 and 6 sites, FLUXNET (inverted triangle) with 75 and 9 sites, and SURFRAD (positive triangle) with 6 and 1 sites, respectively.](image)

**Figure 1.** Geographical distribution of observation sites used to model (314 sites in total, green) and validate (35 sites in total, red) the $L_d$ dataset in this study collected at AmeriFlux (squares) with 159 and 16 sites, AsiaFlux (pentagrams) with 23 and 3 sites, BSRN networks (circles) with 51 and 6 sites, FLUXNET (inverted triangle) with 75 and 9 sites, and SURFRAD (positive triangle) with 6 and 1 sites, respectively.

Critical quality control procedures were implemented to calculate the daily $L_d$ because the selected networks only provided instantaneous $L_d$ values, except for FLUXNET. The daily mean $L_d$ was integrated from the instantaneous values if the portion of missing instantaneous values was less than 20% in one day. The monthly mean values used for validation were obtained by averaging the effective daily values if the missing daily data reached less than 10 days in one month.

2.1.1. AmeriFlux, AsiaFlux, and FLUXNET Data

FLUXNET [51,52] is a joint regional network that provides continuous measurements of various ecological parameters at five temporal resolutions, including carbon dioxide, water, meteorological data, and radiation data. The FLUXNET2015 dataset contains 1532 site-years of data from 1996 to 2014, of which daily $L_d$ observations are used to build and evaluate $L_d$ estimates over global land surface in this study. The AmeriFlux network [53–55] includes 151 sites with more than 100 active sites as of 2012, providing half-hourly or hourly $L_d$ data spanning from 1996 to present. Flux tower sites of the AsiaFlux network [56,57] are spread across various representative climate zones (from humid to arid climates) and land cover types (forest, grass, cropland, and urban area), of which $L_d$ observations have half-hourly or hourly temporal resolutions from 1998 to 2018.

To reduce systematic measurement errors, the data QA/QC checks proposed by Pastorello et al. [58], including single-variable, multi-variable, and specialized checks, are implemented at each site within the three networks. Single-variable checks are aimed at exploring the consistency of one variable in the long and short time series trends. Multi-variable checks focus on the relationship among correlation variables to ascertain discrepant periods. Specialized checks look at common issues in eddy covariance (EC) and meteorological data, such as timestamp shifts or sensor deterioration patterns. The last step for data QA/QC is automatic checks that use specific variable de-spiking routines adapted from Papale et al. [59] to set a range for each variable.
2.1.2. BSRN Data

The baseline surface radiation network (BSRN) was initiated by the world climate research program (WCPR) and aimed to provide accurate observations for validation of satellite radiometry and climate models [60]. The BSRN project has established more than 60 stations globally since January 1992 spanning latitudes ranging from 80°N to 90°S, providing continuous meteorological and radiation data on a minute time scale. By improving its calibration process, the difference between \( L_d \) observations from different pyrgeometers only reached 10 W m\(^{-2}\) in 1995 [61]. Only 6.5% of the \( L_d \) data are missing, which indicates that the pyrgeometers within the BSRN maintain high standards [62]. Moreover, the missing data have less influence on \( L_d \) because \( L_d \) has a small diurnal cycle. Overall, the BSRN \( L_d \) observations are relatively accurate and reliable.

2.1.3. SURFRAD Data

The surface radiation budget network (SURFRAD) has provided meteorological and radiation data used for evaluating satellite products and researching climate changes in the United States since 1995. Currently, it is composed of seven stations representing diverse climates with elevations ranging from 98 to 1689 m. It provides long-term and continuous surface radiation measurements with 3 min and 1 min time intervals before and after 2009, respectively. The \( L_d \) measured by SURFRAD, with an uncertainty of ±9 W m\(^{-2}\), covers a wavelength spanning from 4 to 50 µm [63]. The time period of \( L_d \) measurements used ranges from 2000 to 2018 in this study.

2.2. Input Data

2.2.1. ERA5 Reanalysis Dataset

ERA5 [64], produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), is the fifth generation reanalysis dataset and a successor of ERA-Interim. It provides complete and consistent hourly temperature, relative humidity, and radiation datasets, in addition to many other atmospheric parameter datasets, with a 25-km spatial resolution from 1979 to near real time. Compared with ERA-Interim [65], ERA5 applied the updated integrated forecast system (IFS) “Cy41r2” 4D-var and produced many new parameters, such as a 100-m wind vector [66]. Many studies have also compared the accuracy of ERA5 and used it to analyze climate change. For example, Wang et al. [66] found that the warm bias of ERA5 2-m air temperature (\( T_a \)) is smaller in the warm season and larger in the cold season in relation to the buoy observations over Arctic sea ice. Zhen et al. [67] indicated that the mean relative humidity (RH) of ERA5 displayed a sharp decreasing jump for China during the early 2000s. In this study, the parameters of the ERA5 hourly reanalysis dataset, including the 2-m \( T_a \) (°C), the RH at 1000 hPa (%), and the total column water vapor (TCWV, kg m\(^{-2}\)) from 2000 to 2018, were consolidated into a daily temporal resolution as input data to construct global land \( L_d \) (W m\(^{-2}\)) dataset based on the GBRT method.

2.2.2. GLASS Surface Downward Shortwave Radiation Product

The global land surface satellite (GLASS) daily surface downward shortwave radiation (\( S_{bd} \), W m\(^{-2}\)) product with a 5-km spatial resolution from 2000 to 2018 was produced from the moderate resolution imaging spectroradiometer top-of-atmosphere (TOA) spectral reflectance on the basis of a direct estimation method [68,69]. First, the TOA reflectance was retrieved using atmospheric radiation transfer simulations under different solar or view geometries. Then, surface shortwave net radiation (\( S_n \)) was estimated from the TOA reflectance on the basis of a linear regression relationship between them under different atmospheric conditions and surface properties. Finally, the GLASS daily \( S_d \) was produced using daily \( S_n \) estimates and surface broadband albedo values. The GLASS daily \( S_d \) values obtained an overall RMSE and bias of 32.84 and 3.72 W m\(^{-2}\), respectively, compared to the ground observations at 525 sites from 2003 to 2005 [68].
2.2.3. Global Multi-Resolution Terrain Elevation Data 2010

The 2010 Global Multi-resolution Terrain Elevation dataset (GMTED2010DEM) [70] is a global continent-wide elevation dataset generated by the U.S. Geological Survey (USGS) and the National Geospatial-Intelligence Agency (NGA). This product contains three spatial resolutions (approximately 250, 500, and 1000 m) aimed at providing generic products for different applications. Carabajal et al. [71] indicated that the GMTED2010DEM products exhibited a great improvement relative to previous elevation data at comparable resolutions. Compared to the global set of the ice, cloud, and land elevation satellite (ICESat) geodetic ground control points, it obtained a positive bias of approximately 3 m. In this study, GMTED2010DEM data with a spatial resolution of approximately 250 m were resampled to a 5-km resolution as input data for estimating $L_d$ to match the generated $L_d$ dataset.

2.3. Exiting Surface Downward Longwave Radiation Datasets

The $L_d$ products used for validation and comparison with the generated $L_d$ dataset contain the clouds and earth’s radiant energy system synoptic (CERES-SYN) edition 4.1 and ERA5 reanalysis datasets. The CERES-SYN product with a 100-km spatial resolution, generated on the basis of the Langley Fu-Liou radiation transfer model [72], provides flux estimates at the TOA and surface, as well as four atmospheric pressure levels (70, 200, 500, and 850 hPa) from 2000 to 2020. Compared with CERES-SYN Edition 3A, the $L_d$ of Edition 4A has been improved due to the improvement of nighttime retrieved cloud properties [73,74]. The ERA5 hourly $L_d$ with a 25-km spatial resolution from 1979 to near real time used the more complicated method proposed by Morcrette [9] to replace the old $L_d$ parametrization [75]. Silber et al. [10] demonstrated that ERA5 underestimated $L_d$ compared with the ground measurements collected from the ARM West Antarctic radiation experiment (AWARE) campaign at McMurdo Station and the West Antarctic Ice Sheet (WAIS) divide. In this study, the daily ERA5 $L_d$ dataset consolidated from the hourly dataset and the CERES-SYN product were compared and used to evaluate the generated $L_d$ dataset from 2000 to 2018.

3. Method
3.1. Gradient Boosting Regression Tree

The gradient boosting regression tree (GBRT) is an ensemble approach that enhances the accuracy of the model by aggregating multiple weak forms of regression and decision trees first proposed by Friedman [76]. The GBRT method is capable of predicting and solving overfitting problems [77]. The core idea of this model is to select the appropriate decision tree function based on the current model and fitting function in order to minimize the loss function. The model produces a strong predictive model by constructing an $M$ amount of different weak classifiers through multiple iterations in order to obtain an accurate prediction rule. Each iteration is to improve the previous results by reducing the residuals of the previous model and establish a new combined model in the gradient direction of the reduced residual [46]. Supposing $(x_i, y_i)_{i=1}^N$ is the training dataset, where $x$ represents the predictor variables, $y$ represents the target variable, and $N$ is the number of the training dataset. The GBRT model constructs $M$ different individual decision trees, expressed as $(h(x, a_i))_{i=1}^M$, which can be used to calculate the approximation function of the target variable $f(x)$ as follows:

$$
\begin{align*}
\left\{ \begin{array}{l}
  f(x) = \sum_{m=1}^M f_m(x) = \sum_{m=1}^M \beta_m h(x; a_m) \\
  h(x; a_m) = \sum_{j=1}^J \gamma_{jm} I(x \in R_{jm}), \text{ where } I = 1 \text{ if } x \in R_{jm}, I = 0, \text{ otherwise}
\end{array} \right.
\end{align*}
$$

(1)

where $\beta_m$ and $a_m$ are the weight and classifier parameter of each decision tree, respectively. A loss function $L(y, f(x))$ is introduced to describe the accuracy of the model. Each tree partitions the input space into $J$ regions $R_{1m}, R_{2m}, \cdots, R_{Jm}$ and each $R_{jm}$ corresponds to
a predicted value $\gamma_{jm}$. The general process of the GBRT method is shown in Appendix A, Algorithm A1. More details about the GBRT method can be found in Hastie et al. [78] and Ridgeway [79].

The accuracy of the GBRT model which is implemented in the scikit-learn toolbox is mainly affected by its n-estimator, learning rate, max-depth, and subsample parameters. The n-estimator parameter is the maximum number of iterations completed by a weak learner. Larger n-estimators are more likely to lead to overfitting due to a poorer prediction ability with an increasing model complexity. The learning rate parameter is the weight reduction factor of each weak learner, which is usually used together with the n-estimator parameter to determine the fitting effect of the algorithm. The max-depth parameter is the maximum depth of each regression tree, which limits the number of nodes in the tree. The subsample parameter is the proportion of samples used for fitting the base decision tree. Selecting a subsample less than 1 can reduce overfitting but increase the deviation of sample fitting. In this study, the root mean square error (RMSE), mean bias error (MBE), and correlation coefficient (R) between the $L_d$ observations and estimates are used to evaluate the accuracy of the model.

3.2. Model Construction

The daily 2-m air temperature ($T_a$), relative humidity (RH) at 1000 hPa, total column water vapor (TCWV), surface downward shortwave radiation ($S_{sd}$), and elevation datasets are selected as predictor variables to estimate the daily surface downward longwave radiation ($L_d$). The target variable is the daily $L_d$ observations collected at AmeriFlux, AsiaFlux, BSRN, FLUXNET, and SURFRAD from 2000 to 2018. First, the predictor variables were extracted from global datasets corresponding to the ground stations. Then, the dataset of 314 sites was divided into two portions at random: 80% for the training dataset and the remaining 20% for the test dataset. To select the optimal model, 5-fold cross-validation was applied during the training process. The main steps are as follows:

(1) Calculating daily $L_d$ observations. The daily mean $L_d$ was integrated from the instantaneous values if the missing instantaneous values were less than 20% in one day because the AmeriFlux, AsiaFlux, BSRN, and SURFRAD networks only provide instantaneous $L_d$ values;

(2) Data preprocessing. After resampling to a 5-km resolution, the ERA5 $T_a$, ERA5 RH, ERA5 TCWV, GLASS $S_{sd}$, and GMTED2010DEM elevation datasets were extracted according to the latitude, longitude, and time corresponding to the ground stations;

(3) Training the GBRT model. By circulating within the range of each parameter displayed in Table 1, the GBRT model where the n-estimator parameter is set to 50, the learning rate is set to 0.1, the max-depth is set to 6, and the subsample parameter of 0.8 was selected as the optimal model to estimate global land $L_d$, achieving the lowest RMSE and MBE values on the test dataset;

(4) Implementing the model. The global land $L_d$ was produced on the basis of the trained model using the daily ERA5 $T_a$, ERA5 RH, ERA5 TCWV, GLASS $S_{sd}$, and GMTED2010DEM elevation datasets;

(5) Evaluation of the generated global land $L_d$ dataset. Daily $L_d$ values collected at 35 observation sites were used to validate the generated global land $L_d$ dataset and compare it with the existing $L_d$ datasets. The main flowchart in this study is shown in Figure 2.

### Table 1. Parameter settings to determine the optimal parameters for the GBRT method.

| Parameters   | Threshold | Intervals |
|--------------|-----------|-----------|
| n-estimator  | 50–300    | 50        |
| learning rate| 0.1–0.9   | 0.1       |
| max-depth    | 4–9       | 1         |
| subsample    | 0.2–1     | 0.1       |
Evaluation of the generated global land \( L_d \) dataset. Daily \( L_d \) values collected at 35 observation sites were used to validate the generated global land \( L_d \) dataset and compare it with the existing \( L_d \) datasets. The main flowchart in this study is shown in Figure 2.

Table 1. Parameter settings to determine the optimal parameters for the GBRT method.

| Parameter             | Thresholds       | Values   |
|-----------------------|------------------|----------|
| \( n_{\text{estimator}} \) | 50–300           | 50       |
| \( \text{learning rate} \) | 0.1–0.9          | 0.1      |
| \( \text{max-depth} \)  | 4–9              | 1        |
| \( \text{subsample} \)  | 0.2–1            | 0.1      |

Ground Measurements (Daily \( L_d \))

![Diagram showing the main flowchart in this study.](remote-sens-13-1848-g002)

Figure 2. The main flowchart in this study.

In order to investigate the impact of the predictor variables used in the GBRT model on the \( L_d \) estimation, the feature importance measures provided by the GBRT method was conducted. As shown in Table 2, the importance of the predictor variables of the GBRT model was in the order of the total column water vapor (TCWV), 2-m air temperature (\( T_a \)), relative humidity at 1000hPa (RH), surface downward shortwave radiation (\( S_d \)), and elevation. The \( L_d \) estimates are shown to be more sensitive to the TCWV and \( T_a \) than to most of other variables, thus highlighting the importance of taking TCWV and \( T_a \) as inputs.

Table 2. Importance rankings of all predictor variables for \( L_d \) estimation.

| Predictor Variables | Importance |
|---------------------|------------|
| Total column water vapor (TCWV) | 0.78       |
| 2-m air temperature (\( T_a \)) | 0.19       |
| Relative humidity at 1000 hPa (RH) | 0.01       |
| Surface downward shortwave radiation (\( S_d \)) | 0.01       |
| Elevation           | 0.01       |

4. Results

4.1. Validation against Ground Measurements

4.1.1. Performance of the Model

After confirming the optimal parameters, 80% and 20% of the extracted dataset collected at 314 stations were used as the training and test datasets, respectively, to train the GBRT model and evaluate the \( L_d \) estimates. Figure 3 displays the evaluation results of daily \( L_d \) estimates for the training and test datasets against the ground measurements collected at the AmeriFlux, AsiaFlux, BSRN, FLUXNET, and SURFRAD networks from 2000 to 2018.

For the training dataset, the root mean square error (RMSE), mean bias error (MBE), and correlation coefficient (R) are 16.73 W m\(^{-2}\), 0 W m\(^{-2}\), and 0.96 (\( p < 0.01 \)), respectively, between the ground observations and \( L_d \) estimates on the basis of the GBRT model from 2000 to 2018. Those values are 16.75 W m\(^{-2}\), 0.05 W m\(^{-2}\), and 0.96 (\( p < 0.01 \)) for the test dataset, respectively, which shows a tendency to slightly overestimate \( L_d \). As a whole, the performance of the GBRT model on the test dataset is satisfactory and reliable with an MBE close to zero.
4. Results

4.1. Validation against Ground Measurements

Figure 3. Evaluation results of daily $L_d$ estimates on the basis of the GBRT model for (a) the training dataset and (b) the test dataset against the ground measurements from March 2000 to December 2018.

4.1.2. Validation of the Generated $L_d$ Dataset

The $L_d$ observations of 35 sites were used to evaluate the generated $L_d$ dataset collected at the AmeriFlux, AsiaFlux, BSRN, FLUXNET, and SURFRAD networks from March 2000 to December 2018. As shown in Figure 4, the RMSE, MBE, and R values on the daily time scale are 17.78 W m$^{-2}$, 0.99 W m$^{-2}$, and 0.96 ($p < 0.01$), respectively, between the ground observations and $L_d$ estimates obtained by the GBRT model. On the monthly time scale, those values are 11.53 W m$^{-2}$, 0.68 W m$^{-2}$, and 0.98 ($p < 0.01$), respectively. To further evaluate the performance of the generated $L_d$ dataset, the RMSE, MBE, and R values of the daily $L_d$ estimates at each site were calculated from 2000 to 2018. The minimum and maximum RMSE of the 35 sites are 11.26 and 37.82 W m$^{-2}$, respectively. As shown in Figure 5, 24 out of the 35 sites had RMSEs less than 20 W m$^{-2}$, and only two sites had RMSEs greater than 30 W m$^{-2}$. Overall, 35 sites had absolute MBE values varying from 0.12 to 36.83 W m$^{-2}$, and 23 sites had MBEs between −10 and 10 W m$^{-2}$. The number of stations with MBE less than −10 W m$^{-2}$ and greater than 10 W m$^{-2}$ are both six.

Figure 4. Evaluation results of $L_d$ estimates with 5-km resolution based on the GBRT model on the (a) daily and (b) monthly time scales against the ground measurements from March 2000 to December 2018.
Figure 5. (a) RMSE and (b) MBE histograms of daily \( L_d \) estimates with 5-km resolution based on the GBRT model against the ground measurements from March 2000 to December 2018.

4.2. Comparison with Existing \( L_d \) Products

To better evaluate the accuracy of the generated \( L_d \) dataset, the evaluation result against the 35 sites from 2000 to 2018 was compared with the CERES-SYN and ERA5 products. The generated \( L_d \) and ERA5 products were resampled to a 100-km resolution using the nearest neighbor interpolation method to match the CERES-SYN product. As shown in Figure 6, the RMSE and MBE are 17.94 and 0.25 W m\(^{-2}\), 18.81 and 1.76 W m\(^{-2}\), 18.52 and \(-2.09\) W m\(^{-2}\), respectively, for the daily generated, CERES-SYN, and ERA5 \( L_d \) datasets. CERES-SYN and ERA5 \( L_d \) show overestimated and underestimated trends on the daily time scale, respectively. Relatively speaking, the overestimated trend of the generated \( L_d \) dataset with an MBE of 0.25 W m\(^{-2}\) is slight. In addition, the RMSE of the ERA5 daily \( L_d \) dataset is less than that of the CERES-SYN product, and it can be concluded that the ERA5 \( L_d \) product over land is more accurate than that of the CERES-SYN on the daily time scale. This is consistent with the conclusion of Tang et al. [11] that the ERA5 \( L_d \) product over land surface has a higher accuracy on average than the CERES-SYN on the hourly, daily, and monthly time scales but has a worse accuracy than the CERES-SYN dataset over ocean surface. On the monthly time scale, the RMSE and MBE are 11.75 and 0.18 W m\(^{-2}\), respectively, for the generated, CERES-SYN, and ERA5 \( L_d \) datasets. It can be concluded that the generated \( L_d \) dataset based on the GBRT model performed best on both daily and monthly scales. To further compare the performance of the three daily \( L_d \) datasets, the RMSE, MBE, and R values at each site were calculated from 2000 to 2018. The RMSE of the 35 sites varied from 11.21 to 31.90 W m\(^{-2}\), 9.09 to 41.99 W m\(^{-2}\), 8.68 to 35.51 W m\(^{-2}\), respectively, for the daily generated, CERES-SYN, and ERA5 \( L_d \) datasets. As shown in Figure 7, there are 28, 28, and 25 sites with RMSEs less than 25 W m\(^{-2}\) for the daily generated, CERES-SYN, and ERA5 \( L_d \) datasets. Only 3, 3, and 2 out of 35 sites had RMSEs greater than 30 W m\(^{-2}\) for the three daily \( L_d \) datasets, respectively. These three daily \( L_d \) datasets have 23, 19, and 24 sites with MBEs between \(-10\) and 10 W m\(^{-2}\), respectively. However, the daily CERES-SYN \( L_d \) product obtained 10 sites with MBEs greater than 10 W m\(^{-2}\), compared with 6 for the generated \( L_d \) dataset and 3 for the ERA5 \( L_d \) retrieval.
Figure 6. Evaluation results of the daily and monthly (a,d) $L_d$ estimates based on the GBRT model, (b,e) CERES-SYN $L_d$ product, and (c,f) ERA5 $L_d$ retrieval with a 100-km resolution against the ground measurements from March 2000 to December 2018.

Figure 7. (a) RMSE and (b) MBE histograms of the daily $L_d$ estimates based on the GBRT model, CERES-SYN $L_d$ product, and ERA5 $L_d$ retrieval with a 100-km resolution against the ground measurements from March 2000 to December 2018.

4.3. Spatial and Temporal Analysis of $L_d$

4.3.1. Spatial Distribution

The multiyear seasonal and annual mean values of the generated $L_d$ dataset from 2003 to 2018 (i.e., not from 2000 to 2018) were calculated due to the absence of daily $L_d$ values from 2000 to 2002. The CERES-SYN and ERA5 $L_d$ products were resampled to a 5-km resolution by the bilinear interpolation method for comparison with the generated $L_d$ dataset. The spatial distributions of the multiyear seasonal and annual mean $L_d$ estimations over the global land surface from 2003 to 2018 are displayed in Figures 8 and 9, respectively.
The highest multiyear seasonal mean $L_d$ value is 333.21 W m$^{-2}$ in Northern hemisphere summer (June, July, and August), followed by 311.94 W m$^{-2}$ in Northern hemisphere autumn (September, October, and November), and the lowest value is 286.09 W m$^{-2}$ in Northern hemisphere winter (December, January, and February). The seasonal variation in $L_d$ is closely related to the annual solar zenith cycle and the maximum sunshine duration. After the winter solstice, the direct sun point moves northward from the Tropic of Capricorn, causing changes in the global heat distribution, which increases the overall $L_d$ value in the Northern hemisphere. Overall, the multiyear annual mean value of the generated $L_d$ dataset is 308.76 W m$^{-2}$, which is greater than the ERA5 value of 306.92 W m$^{-2}$ and less than the CERES-SYN value of 313.83 W m$^{-2}$ from 2003 to 2018. The spatial distribution of $L_d$ not only shows significant latitudinal dependencies in which the mean $L_d$ value decreases with increasing latitude but also relates to the surface elevation and regional climate. The mean $L_d$ values estimated over the Andes and Tibetan Plateau are comparatively and obviously low due to their high elevation with a low cloud coverage, thin air and readily lost heat. The mean $L_d$ values of Antarctica and Greenland are always lowest owing to the perennial snow cover and frigid climate. Apparently, the generated $L_d$ value is lower than the CERES-SYN value and higher than the ERA5 value. The lowest and highest differences between the generated $L_d$ and the CERES-SYN product are $-81.44$ and $60.56$ W m$^{-2}$, respectively; and the values between the generated $L_d$ and the ERA5 product are $-46.17$ and $58.83$ W m$^{-2}$, respectively. The generated $L_d$ value is significantly lower than the CERES-SYN in the Tibetan Plateau, Andes Mountains, and Antarctica, and is significantly higher than it in a small area of the northern Amazon Rainforest and eastern Indonesia. Compared with the CERES-SYN dataset, the difference between the generated $L_d$ dataset and the ERA5 product is evenly distributed with no obvious high and low values over the global land surface.

**Figure 8.** The spatial distribution of the multiyear seasonal mean value of the generated $L_d$ dataset in Northern hemisphere (a) spring (March, April, and May), (b) summer (June, July, and August), (c) autumn (September, October, and November), and (d) winter (December, January, and February) over the global land surface from 2003 to 2018.
The multiyear annual mean $L_d$ values of the generated dataset, ERA5 retrieval, and CERES-SYN product are consistent with the evaluation results against the ground measurements that ERA5 and CERES-SYN tend to underestimate and overestimate $L_d$ value, respectively. However, there is still much debate about the specific multiyear annual mean $L_d$ value over the global land surface. The uncertainty of the global land mean $L_d$ estimation is difficult to quantify, and different periods may influence the estimated values. Ma et al. [80] summarized that multiyear annual mean $L_d$ values over the global land surface varied between 287.35 and 316.62 W m$^{-2}$ per decade. Therefore, it is reliable for the generated and ERA5 datasets, except for 2005 and 2010, which im-
nediate $L_d$ values for this period are below the multiyear average over 1970–2005, and its median value is 1.86 W m$^{-2}$.

The multiyear annual mean $L_d$ values of the generated dataset, (a) generated $L_d$ dataset, (b) generated $L_d$ minus CERES-SYN, and (c) generated $L_d$ minus ERA5 over the global land surface from 2003 to 2018.

Figure 9. The spatial distribution of the multiyear annual mean value of the (a) generated $L_d$ dataset, (b) generated $L_d$ minus CERES-SYN, and (c) generated $L_d$ minus ERA5 over the global land surface from 2003 to 2018. To study the temporal variations of the generated $L_d$ dataset, we calculated the monthly and annual mean $L_d$ from 2003 to 2018, as shown in Figure 10. We analyzed whether the interannual changes in the generated $L_d$ dataset were reliable by comparing with the ERA5 and CERES-SYN $L_d$ datasets. Overall, ERA5 has a relatively lower value, and CERES-SYN shows a larger value for the multiyear monthly mean $L_d$. The multiyear monthly mean $L_d$ of the ERA5, generated dataset, and CERES-SYN are all lowest in January, with values of 280.02, 284.22, and 284.49 W m$^{-2}$, respectively; they are all largest in July, with values of 336.12, 336.24, 343.94 W m$^{-2}$, respectively. The multiyear monthly mean $L_d$ values of the three datasets all increase from January to July and decrease from July to December in connection with the revolution of the earth around the sun resulting in more total solar radiation in Northern hemisphere summer than in Northern hemisphere winter. Compared with ERA5 and CERES-SYN, the boxed part of the box-plot (Figure 10a) of the generated $L_d$
dataset is relatively compact, indicating that its monthly mean $L_d$ values in different years are concentrated. Similar to the monthly mean $L_d$, the annual mean $L_d$ values of ERA5 and CERES-SYN are lower and higher than the generated $L_d$ dataset, respectively, in the same year. The annual mean $L_d$ values ranged from 304.93 to 309.92 W m$^{-2}$, 306.94 to 311.99 W m$^{-2}$, and 311.88 to 316.29 W m$^{-2}$ from 2003 to 2018, respectively, for the ERA5 retrieval, the generated $L_d$ dataset, the CERES-SYN product. The three datasets all obtained the lowest and largest annual mean $L_d$ values in 2008 and 2016, respectively.

As displayed in Figure 10c, before 2015, the anomalies of the annual mean $L_d$ values are negative for the generated and ERA5 $L_d$ datasets, except for 2005 and 2010, which implies that the annual mean $L_d$ values for this period are below the multiyear average over 16 years. In addition, the anomalies of annual mean $L_d$ values are also more than zero for the CERES-SYN product in 2003. Moreover, the CERES-SYN $L_d$ product had a smaller growth trend of 0.8 W m$^{-2}$ per decade ($p = 0.20$) from 2003 to 2018, but the growth trend was not significant. Overall, the temporal variation and trend of the generated $L_d$
dataset are more consistent with the ERA5 product, and the annual mean $L_d$ values display a gradual increasing trend from 2003 to 2018. Ma et al. [80] concluded that the trend of the annual mean $L_d$ over the global land surface for 44 CMIP5 GCMs varied from 0.69 to 2.86 W m$^{-2}$ per decade ($p < 0.01$) during the time period of 1970–2005, and its median value is 1.86 W m$^{-2}$ per decade. Therefore, it is reliable for the generated and ERA5 $L_d$ datasets, with trends of 1.8 ($p < 0.01$) and 1.9 ($p < 0.01$) W m$^{-2}$ per decade over the global land surface, respectively.

4.3.3. Relationships between the Long-Term $L_d$ and the Key Factors

Previous studies indicated that the accuracy of $L_d$ estimation mainly depends on the reliability of the air temperature, precipitable water vapor, cloud, and elevation data retrieved from the reanalysis and satellite products. In view of the small variations in elevation and cloud cover over the long time series, the trend of $L_d$ estimation is mainly influenced by air temperature and water vapor pressure. Therefore, we calculated the anomalies of the 2-m air temperature (Ta, °C) and water vapor pressure ($e$, hPa) datasets based on the ERA5 hourly products from 2003 to 2018. $e$ can be calculated with Ta using the following equations based on the Ta and relative humidity at 1000 hPa (RH) derived from the ERA5 products.

\[ \text{RH} = \frac{e}{e_s} \times 100\% \quad (2) \]

\[ e_s = 6.11 \exp \left( \frac{L_d}{273.15} \left( \frac{1}{273.15} - \frac{1}{\text{Ta} + 273.15} \right) \right) \quad (3) \]

where $e$ and $e_s$ are the water vapor pressure and saturation water vapor pressure, respectively.

Figure 11 presents the temporal variation in the annual mean anomalies for the generated $L_d$ estimation, Ta and $e$ from 2003 to 2018. The annual mean values ranged from 9.12 to 10.22 °C, and 7.82 to 8.26 hPa from 2003 to 2018 for Ta and $e$, respectively. Before 2015, the annual mean anomalies were negative for Ta and $e$, excluding 2005 and 2010, which was similar to the $L_d$ estimation. In addition, the anomalies of annual mean Ta values were also greater than zero in 2007. The $L_d$ increases with the increase in Ta and $e$. The increasing rates are 0.3 °C per decade ($p < 0.01$), 0.1 hPa per decade ($p < 0.05$), and 1.8 W m$^{-2}$ per decade ($p < 0.01$), respectively, for Ta, $e$, and $L_d$ from 2003 to 2018. Overall, Ta and $e$ positively influence $L_d$ with correlation coefficients of 0.96 ($p < 0.01$) and 0.97 ($p < 0.01$), respectively. The strong absorption and re-emission of radiation by water vapor molecules result in a high correlation between $e$ and $L_d$. However, the influence of temperature on $L_d$ relies on the dependence of the outgoing longwave radiation on the absolute temperature of the Earth. In addition, the spatial distributions of the annual mean values of Ta and $e$ from 2003 to 2018 are shown in Figure 12. The minimum and maximum annual mean values are −53.35 and 34.12 °C, and 0.05 and 32.85 hPa, respectively, for Ta and $e$. Their distribution characteristics are similar to that of the generated $L_d$ dataset that its spatial distribution not only shows notable latitudinal dependencies as the annual mean values decrease with increasing latitudes but also relates to the surface elevation and regional climate. The annual mean Ta and $e$ values on the Andes and Tibetan Plateau are comparatively and obviously low due to their high elevations. The annual mean Ta values of Antarctica and Greenland are always the lowest due to their perennial snow coverage and frigid climates. The annual mean $e$ values are relatively low, less than 21.72 hPa at middle to high latitudes. The spatial distribution of R between the generated $L_d$ estimation and Ta and $e$ from 2003 to 2018 is also drawn, as shown in Figure 13. Only significant pixels where $p$ values are less than 0.05 appeared. The R values ranged from 0.50 and −0.67 to 1 for Ta and $e$, respectively. There was a positive correlation between the generated $L_d$ and Ta in the region where the R passed the significant test. For the $e$, there are few pixels with R value less than 0, which even cannot be shown up on the map. Except for values less than 0, the minimum value of R between the generated $L_d$ and $e$ is also 0.50. The R values between annual mean $L_d$ estimates and Ta and $e$ failed the significant test mainly occurred in the Andes Mountains, Brazilian Plateau, Tibet Plateau, Australia, Southern Africa, and
southern North America. This may be due to the influences of clouds, elevation controls, and carbon dioxide emissions [29,30,41,82], that play a dominant role in these regions. The possible reasons need to be further explored.

![Image](image.png)

**Figure 11.** The trend of the annual mean anomalies of the generated $L_d$ estimation, ERA5 2-m air temperature, and water vapor pressure from 2003 to 2018.

![Image](image.png)

**Figure 12.** The spatial distribution of annual mean values for the (a) ERA5 2-m air temperature and (b) water vapor pressure from 2003 to 2018.

![Image](image.png)

**Figure 13.** The spatial distribution of the correlation coefficient between the generated $L_d$ estimation and (a) ERA5 2-m air temperature and (b) water vapor pressure from 2003 to 2018. Only significant pixels where $p$ values are less than 0.05 appeared.
5. Discussion

5.1. Shortcomings of the GBRT Model

The gradient boosting regression tree (GBRT) method has advantages in forecasting and solving overfitting problems [76]. The evaluation results demonstrated that the generated $L_d$ dataset based on the GBRT method performed better at selected stations than the ERA5 and CERES-SYN products on daily and monthly time scales. However, there are still some disadvantages in the machine learning methods for radiation estimation represented in the GBRT method [45,77,83]. Alizamir et al. [77] utilized six different machine learning models to estimate solar radiation from two stations of two different locations, and found that the six models all tend to overestimate $L_d$ for low values of it and to underestimate $L_d$ for high values of it. Fan et al. [83] also exposed similar problem in using support vector machine and extreme gradient boosting methods to predict daily solar radiation in China. With reference to Figures 3, 4 and 6, it is evident that the GBRT method for $L_d$ estimation also overestimate $L_d$ at low values and underestimate $L_d$ at high values. Since the departures of slope from 1 and intercept from 0 for fitting linear regression equations can measure the degree of deviation, the linear regression equations were fitted between the ground measurements of 35 stations and the generated, ERA5, and CERES-SYN $L_d$ datasets on daily and monthly time scales, as listed in Table 3. Compared to those two $L_d$ products, the fitted linear regression equations of the generated $L_d$ has a smaller slope and a greater intercept on both daily and monthly time scales, which indicates that the fitted line deviates more from the 1:1 line and that the GBRT method underestimates $L_d$ for high values and overestimates it for low values. Other machine learning methods, including support vector regressions, multivariate adaptive regression splines, and artificial neural networks used to estimate $L_d$, also show the same problem [49]. On the other hand, the GBRT method makes predictions by learning rules from many sample data, so it has higher requirements for the accuracy and quantity of its training datasets. However, the ground measurements of $L_d$ used as the target variable exhibit missing values and deviations, although the obviously incorrect data have been removed, which may be limit the accuracy of the GBRT method. Finally, the learning and training process of the GBRT method is a black box whose processes are not known and may not be effective [45].

Table 3. The fitted linear regression equations for the generated, ERA5, and CERES-SYN $L_d$ datasets on both daily and monthly time scales. Where the x and y represent the ground measurements of $L_d$ and the $L_d$ estimates, respectively.

| Time Scale          | Dataset         | Fitted Linear Regression Equation |
|---------------------|-----------------|-----------------------------------|
| Daily time scale    | $L_d$ estimation| $y = 0.91 \times x + 27.99 *$    |
|                     | ERA5 $L_d$      | $y = 0.97 \times x + 6.39 *$     |
|                     | CERES-SYN $L_d$| $y = 0.96 \times x + 13.38 *$    |
| Monthly time scale  | $L_d$ estimation| $y = 0.94 \times x + 19.06 *$    |
|                     | ERA5 $L_d$      | $y = 0.99 \times x + 1.64 *$     |
|                     | CERES-SYN $L_d$| $y = x + 0.59 *$                  |

* The coefficient of the fitted linear regression equation passed the significance test ($p < 0.01$).

5.2. Accuracy and Completeness of Input Datasets and Ground Measurements

Because $L_d$ was estimated based on the relationships between its ground observations and input variables, including the 2-m air temperature (Ta), relative humidity (RH) at 1000 hPa, total column water vapor (TCWV), surface downward shortwave radiation ($S_d$), and elevation, the accuracy and completeness of the input datasets and ground measurements are vital. However, ground observations exist measurement errors and problem of spatial representativeness, which are potential sources of errors in $L_d$ estimation. A larger part of measurements errors is caused by systematic deviations and calibration process differences. Ohmura et al. [60] demonstrated the accuracy of $L_d$ observations in the baseline surface radiation network improved from 30 W m$^{-2}$ in 1999 to 10 W m$^{-2}$ in 1995.
due to improvement of the calibration process. Currently, although pyrgeometers for \( L_d \) measurement are regularly maintained and calibrated, there is still a lack of recognized world reference calibration standard \([60,84]\). The different calibration methods of different observation networks can lead to inconsistencies of \( L_d \) measurements at the close positions, which also brings uncertainties of ground measurements \([81]\). The spatial representativeness plays an important role in the surface radiation retrieval and validation \([85–88]\). Jiang et al. \([87]\) indicated the accuracy of \( S_d \) retrieval can be improved, that maximum improvement of root mean square error is up to 9%, after considering the scale information. In this study, we only compared the accuracies of the generated, ERA5, and CERES-SYN \( L_d \) datasets at 100-km spatial resolution but did not examine the representativeness of surface observation points, which maybe lead to the uncertainty of compared result.

On the other hand, the completeness of input datasets limits the continuity of the generated \( L_d \) dataset. For example, the generated daily \( L_d \) dataset was discontinuous before 2003 due to the data missing of the global land surface satellite (GLASS) daily \( S_d \) product which was produced by using moderate resolution imaging spectroradiometer (MODIS) top of atmosphere (TOA) reflectance data. Compared with the previous methods of \( S_d \) retrieval, however, GLASS \( S_d \) had a higher spatial resolution of 5 km and was directly estimated using TOA reflectance without the need for cloud and aerosol data, which contributes to a better ability to demonstrate temporal variations in \( S_d \) over a long time period. Moreover, the ERA5 and elevation datasets were resampled to a 5-km spatial resolution matching with \( S_d \), which can also introduce uncertainty into the data. In summary, the \( L_d \) estimates will be more accurate if the accuracy and completeness of the input datasets and ground measurements are improved.

6. Conclusions

It is of great importance for studying the Earth’s surface radiation budget and energy balance to construct a long-term surface downward longwave radiation (\( L_d \), 4–100 \( \mu \)m) dataset worldwide. This study generated a daily \( L_d \) dataset with a 5-km spatial resolution over the global land surface utilizing the gradient boosting regression tree (GBRT) method with 2-m air temperature (\( T_a \)), relative humidity (\( R_H \)) at 1000 hPa, total column water vapor (TCWV), surface downward shortwave radiation (\( S_d \)), and elevation datasets from 2000 to 2018. The \( L_d \) observations of 349 stations collected at the AmeriFlux, AsiaFlux, baseline surface radiation network (BSRN), surface radiation budget network (SURFRAD), and FLUXNET networks were randomly divided into 90% (314 sites), as the target variable to build the model, and 10% (35 sites), as the evaluation dataset to independently validate the \( L_d \) estimates. First, the predictor variables were extracted from the global datasets according to the latitude, longitude, and time corresponding to the ground stations. Then, the dataset of 314 sites was further divided into two portions at random to train the GBRT model: 80% for the training dataset and the remaining 20% for the test dataset. Then, the daily \( L_d \) observations collected at 35 stations were used to validate the generated global land \( L_d \) dataset.

The evaluation results showed that the root mean square error (RMSE), mean bias error (MBE), and correlation coefficient (R) values on the daily time scale were 17.78 W m\(^{-2}\), 0.99 W m\(^{-2}\), and 0.96 (\( p < 0.01 \)), respectively, between the \( L_d \) estimates with a 5-km spatial resolution and the ground measurements. On the monthly time scale, those values are 11.53 W m\(^{-2}\), 0.68 W m\(^{-2}\), and 0.98 (\( p < 0.01 \)), respectively. At a 100-km spatial resolution, the performance of the generated \( L_d \) dataset is better than that of ERA5 and CERES-SYN. On the daily time scale, the RMSE and MBE are 17.94 and 0.25 W m\(^{-2}\), 18.81 and 1.76 W m\(^{-2}\), 18.52 and −2.09 W m\(^{-2}\), respectively, for the generated, CERES-SYN, and ERA5 \( L_d \) datasets. The multiyear seasonal and annual mean values of the generated \( L_d \) dataset from 2003 to 2018 were calculated due to the absence of daily \( L_d \) from 2000 to 2002. In terms of their temporal variation, the multiyear monthly mean \( L_d \) values of the three datasets increase from January to July and decrease from July to December in connection with the revolution of the earth around the sun resulting in more total solar radiation in Northern hemisphere.
summer than in Northern hemisphere winter. Overall, the temporal variation and trend of the generated $L_d$ dataset are more consistent with the ERA5 product that the annual mean $L_d$ values display a gradual increasing trend from 2003 to 2018. The spatial distribution of $L_d$ not only shows a notable latitudinal dependency in which the mean $L_d$ value decreases with increasing latitudes but also relates to the surface elevation and regional climate. In addition, $L_d$ is positively affected by the 2-m air temperature and water vapor pressure with R values of 0.96 ($p < 0.01$) and 0.97 ($p < 0.01$), respectively.

Overall, the generated $L_d$ dataset has a higher spatial resolution and accuracy, contributing to knowledge of the surface radiation budget and energy balance of the Earth.

**Author Contributions:** Conceptualization, X.Z.; methodology, C.F.; software, C.F.; data curation, C.F.; writing—original draft preparation, C.F.; writing—review and editing, X.Z., C.F., Y.W., W.Z., N.H., J.X., S.Y., X.X., B.J.; supervision, X.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was supported by the National Natural Science Foundation of China under Grant 42090012 and 41571340.

**Data Availability Statement:** The generated $L_d$ dataset will be publicly available at https://doi.org/10.5281/zenodo.4704019 and https://doi.org/10.5281/zenodo.4739724 (accessed on 28 March 2021). The ground measurements collected from the AmeriFlux, AsiaFlux, FLUXNET, BSRN, and SURFRAD stations are available at https://ameriflux.lbl.gov, http://www.asiaflux.net/, https://fluxnet.org/, https://dataportals.pangaea.de/bsrn, and https://www.esrl.noaa.gov/gmd/grad/surfrad/ (accessed on 28 March 2021), respectively. The ERA5 and CERES-SYN data are available at https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5 and https://ceres.larc.nasa.gov/ (accessed on 28 March 2021), respectively.

**Acknowledgments:** We sincerely thank the institutions and researchers who provided the data used in this study and made them available to the public. We also sincerely thank the anonymous reviews for their constructive suggestions which are greatly helpful for improving the article.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Appendix A

| Number | Site Code | Site Name                          | Latitude (deg) | Longitude (deg) | Elevation (m) | Time Period   |
|--------|-----------|------------------------------------|----------------|----------------|---------------|---------------|
| 1      | BR-Npw    | Northern Pantanal Wetland           | −16.50         | −56.41          | 120           | 2013–2017     |
| 2      | BR-Sa3    | Santarem-Km83-Logged Forest        | −3.02          | −54.97          | 100           | 2001–2004     |
| 3      | CA-ARF    | Attawapiskat River Fen              | 52.70          | −83.96          | 88            | 2011–2015     |
| 4      | CA-Ca1    | British Columbia–1949 Douglas-fir stand | 49.87      | −125.33         | 300           | 2000–2010     |
| 5      | CA-Ca3    | British Columbia–Pole sapling Douglas-fir stand | 49.53      | −124.90         | \             | 2002–2016     |
| 6      | CA-Cbo    | Ontario–Mixed Deciduous, Borden Forest Site | 44.32       | −79.93          | 120           | 2005–2018     |
| 7      | CA-DBB    | Delta Burns Bog                    | 49.13          | −122.98         | 4             | 2016–2018     |
| 8      | CA-Gro    | Ontario–Groundhog River, Boreal Mixedwood Forest New Brunswick–1967 Balsam Fir–Nashwaak Lake Site 01 (Mature balsam fir forest) | 48.22       | −82.16          | 340           | 2003–2014     |
| 9      | CA-Na1    | Saskatchewan–Western Boreal, Mature Aspen | 46.47      | −67.10          | 341           | 2003–2005     |
| 10     | CA-Oas    | Saskatchewan–Western Boreal, Mature Black Spruce | 53.63       | −106.20         | 530           | 2000–2010     |
| 11     | CA-Obs    | Saskatchewan–Western Boreal, Mature Jack Pine | 53.99       | −105.12         | 628.94        | 2000–2010     |
| 12     | CA-Ojp    | Saskatchewan–Western Boreal, Mature Jack Pine | 53.92       | −104.69         | 579           | 2000–2010     |
| Number | Site Code | Site Name | Latitude (deg) | Longitude (deg) | Elevation (m) | Time Period |
|--------|-----------|-----------|----------------|----------------|---------------|-------------|
| 13     | CA-Qcu    | Quebec–Eastern Boreal, Black Spruce/Jack Pine Cutover | 49.27 | -74.04 | 392.3 | 2004–2010 |
| 14     | CA-Qfo    | Quebec–Eastern Boreal, Mature Black Spruce | 49.69 | -74.34 | 382 | 2003–2010 |
| 15     | CA-SCB    | Scotty Creek Bog | 61.31 | -121.30 | 280 | 2014–2017 |
| 16     | CA-SCC    | Scotty Creek Landscape | 61.31 | -121.30 | 285 | 2013–2016 |
| 17     | CA-SF1    | Saskatchewan–Western Boreal, forest burned in 1977 | 54.49 | -105.82 | 536 | 2003–2006 |
| 18     | CA-SF2    | Saskatchewan–Western Boreal, forest burned in 1989 | 54.25 | -105.88 | 520 | 2002–2005 |
| 19     | CA-SF3    | Saskatchewan–Western Boreal, forest burned in 1998 | 54.09 | -106.01 | 540 | 2002–2006 |
| 20     | CA-SJ1    | Saskatchewan–Western Boreal, Jack Pine forest harvested in 1994 | 53.91 | -104.66 | 580 | 2001–2010 |
| 21     | CA-SJ2    | Saskatchewan–Western Boreal, Jack Pine forest harvested in 2002 | 53.95 | -104.65 | 580 | 2003–2010 |
| 22     | CA-SJ3    | Saskatchewan–Western Boreal, Jack Pine forest harvested in 1975 (BOREAS Young Jack Pine) | 53.88 | -104.65 | 580 | 2004–2010 |
| 23     | CA-TP4    | Ontario–Turkey Point 1939 Plantation White Pine | 42.71 | -80.36 | 184 | 2003–2017 |
| 24     | CA-TPD    | Ontario–Turkey Point Mature Deciduous | 42.64 | -80.56 | 260 | 2012–2017 |
| 25     | CA-WP1    | Alberta–Western Peatland–LaBiche River, Black Spruce/Larch Fen | 54.95 | -112.47 | 540 | 2003–2009 |
| 26     | US-A03    | ARM-AMF3-Oliktok | 70.50 | -149.88 | 5 | 2014–2018 |
| 27     | US-A10    | ARM-NSA-Barrow | 71.32 | -156.61 | 4 | 2011–2018 |
| 28     | US-A32    | ARM SGP Medford hay pasture | 36.82 | -97.82 | 335 | 2015–2017 |
| 29     | US-A74    | ARM SGP milo field | 36.81 | -97.55 | 337 | 2016–2017 |
| 30     | US-AR1    | ARM USDA UNL OSU Woodward Switchgrass 1 | 36.43 | -99.42 | 611 | 2009–2012 |
| 31     | US-AR2    | ARM USDA UNL OSU Woodward Switchgrass 2 | 36.64 | -99.60 | 646 | 2009–2012 |
| 32     | US-ARM    | ARM Southern Great Plains site-Lamont | 36.61 | -97.49 | 314 | 2003–2018 |
| 33     | US-An1    | Anaktuvuk River Severe Burn | 68.99 | -150.28 | 600 | 2008–2009 |
| 34     | US-An2    | Anaktuvuk River Moderate Burn | 68.95 | -150.21 | 600 | 2008–2019 |
| 35     | US-An3    | Anaktuvuk River Unburned | 68.93 | -150.27 | 600 | 2008–2010 |
| 36     | US-Bi1    | Boulid Island Alfalfa | 38.10 | -121.50 | -2.7 | 2016–2018 |
| 37     | US-Bi2    | Boulid Island corn | 38.11 | -121.54 | -5 | 2018–2017 |
| 38     | US-Bkg    | Brookings | 44.35 | -96.84 | 510 | 2004–2010 |
| 39     | US-Blk    | Black Hills | 44.16 | -103.65 | 1718 | 2004–2008 |
| 40     | US-Bo1    | Bondville | 40.01 | -88.29 | 219 | 2000–2008 |
| 41     | US-Br1    | Brooks Field Site 10- Ames | 41.97 | -93.69 | 313 | 2005–2011 |
| 42     | US-Br3    | Brooks Field Site 11- Ames | 41.97 | -93.69 | 313 | 2005–2011 |
| 43     | US-CPk    | Chimney Park | 41.07 | -106.12 | 2750 | 2009–2013 |
| 44     | US-ChR    | Chestnut Ridge | 35.93 | -84.33 | 286 | 2005–2010 |
| 45     | US-Ctn    | Cottonwood | 43.95 | -101.85 | 744 | 2006–2009 |
| 46     | US-Dia    | Diablo | 37.68 | -121.53 | 323 | 2010–2012 |
| 47     | US-Dk1    | Duke Forest-open field | 35.97 | -79.09 | 168 | 2004–2008 |
| 48     | US-Dk2    | Duke Forest-hardwoods | 35.97 | -79.10 | 168 | 2004–2008 |
| 49     | US-Dk3    | Duke Forest-loblolly pine | 35.98 | -79.09 | 163 | 2004–2008 |
| 50     | US-EDN    | Eden Landing Ecological Reserve | 37.62 | -122.11 | | 2018 |
| 51     | US-EML    | Eight Mile Lake Permafrost thaw gradient, Healy Alaska. | 63.88 | -149.25 | 700 | 2011–2018 |
| 52     | US-FPe    | Fort Peck | 48.31 | -105.10 | 634 | 2000–2008 |
| 53     | US-FR2    | Freeman Ranch- Mesquite Juniper | 29.95 | -98.00 | 271.9 | 2008 |
| 54     | US-FR3    | Freeman Ranch-Woodland | 29.94 | -97.99 | 232 | 2008–2012 |
Table A1. Cont.

| Number | Site Code | Site Name | Latitude (deg) | Longitude (deg) | Elevation (m) | Time Period |
|--------|-----------|-----------|----------------|----------------|--------------|-------------|
| 55     | US-Fmf    | Flagstaff–Managed Forest | 35.14 | -111.73 | 2160 | 2005–2010 |
| 56     | US-Fuf    | Flagstaff–Unmanaged Forest | 35.09 | -111.76 | 2180 | 2005–2010 |
| 57     | US-Fw1    | Flagstaff–Wildfire | 35.45 | -111.77 | 2270 | 2005–2010 |
| 58     | US-GLE    | GLEES | 41.37 | -106.24 | 3197 | 2004–2018 |
| 59     | US-Goo    | Goodwin Creek | 34.25 | -89.87 | 87 | 2002–2008 |
| 60     | US-HBK    | Hubbard Brook Experimental Forest | 43.94 | -71.72 | 367 | 2017–2018 |
| 61     | US-HRA    | Humnoke Farm Rice Field–Field A | 34.59 | -91.75 | 367 | 2016–2017 |
| 62     | US-HRC    | Humnoke Farm Rice Field–Field C | 34.59 | -91.75 | 367 | 2016–2017 |
| 63     | US-Ha2    | Harvard Forest Hemlock Site | 42.54 | -72.18 | 360 | 2014–2018 |
| 64     | US-Hn3    | Hobcaw Barony Longleaf Pine Restoration | 46.69 | -119.46 | 1209 | 2017–2018 |
| 65     | US-Ho1    | Howland Forest (main tower) | 45.20 | -68.74 | 60 | 2007–2018 |
| 66     | US-Ho2    | Howland Forest (west tower) | 45.21 | -68.75 | 61 | 2007–2009 |
| 67     | US-Ho3    | Howland Forest (harvest site) | 45.21 | -68.73 | 61 | 2007–2009 |
| 68     | US-Ivo    | Ivotuk | 68.49 | -155.75 | 568 | 2003–2006 |
| 69     | US-KFS    | Kansas Field Station | 39.06 | -95.19 | 310 | 2008–2018 |
| 70     | US-KLS    | Kansas Land Institute | 38.77 | -97.57 | 373 | 2012–2017 |
| 71     | US-KM4    | KBS Marshall Farms Smooth Brome Grass (Ref) | 42.44 | -85.33 | 288 | 2010–2018 |
| 72     | US-KS3    | Kennedy Space Center (salt marsh) | 28.71 | -90.74 | 0 | 2018 |
| 73     | US-KUT    | KUOM Turfgrass Field | 44.99 | -93.19 | 301 | 2006–2009 |
| 74     | US-Kon    | Konza Prairie LTER (KNZ) | 39.08 | -96.56 | 417 | 2006–2018 |
| 75     | US-los    | Lost Creek | 46.08 | -89.98 | 480 | 2014–2018 |
| 76     | US-MMS    | Morgan Monroe State Forest | 39.32 | -86.41 | 275 | 2000–2018 |
| 77     | US-MOz    | Missouri Ozark Site | 38.74 | -92.20 | 219.4 | 2004–2017 |
| 78     | US-MRF    | Mary’s River (Fir) site | 44.65 | -123.35 | 263 | 2007–2011 |
| 79     | US-MSR    | Montana Sun River winter wheat | 47.48 | -111.72 | 1110 | 2016 |
| 80     | US-Me2    | Metolius mature ponderosa pine | 44.45 | -121.56 | 1253 | 2005–2018 |
| 81     | US-Me3    | Metolius-second young aged pine | 44.32 | -121.61 | 1005 | 2009 |
| 82     | US-Me6    | Metolius Young Pine Burn | 44.32 | -121.61 | 998 | 2010–2018 |
| 83     | US-Men    | Lake Mendota, Center for Limnology Site | 43.08 | -89.40 | 260 | 2012–2018 |
| 84     | US-Mpj    | Mountainair Pinyon-Juniper Woodland | 34.44 | -106.24 | 2196 | 2008–2018 |
| 85     | US-MTb    | Mt Bigelow | 32.42 | -110.73 | 2573 | 2008–2018 |
| 86     | US-Nc1    | Mt Bigelow | 35.81 | -76.71 | 5 | 2005–2012 |
| 87     | US-Nc2    | NC_Lobliolgy Plantation | 35.80 | -76.67 | 5 | 2005–2018 |
| 88     | US-Nc3    | NC_Clearcut#3 | 35.80 | -76.66 | 5 | 2013–2018 |
| 89     | US-Nc4    | NC_AlligatorRiver | 35.79 | -75.90 | 1 | 2015–2018 |
| 90     | US-NGB    | NGEE Arctic Barrow | 71.28 | -156.61 | 5273 | 2012–2018 |
| 91     | US-NGC    | NGEE Arctic Council | 64.86 | -163.70 | 35 | 2017–2018 |
| 92     | US-NR1    | Niwot Ridge Forest (LTER NWT1) | 40.03 | -105.55 | 3050 | 2000–2018 |
| 93     | US-Ne1    | Mead–irrigated continuous maize site | 41.17 | -96.48 | 361 | 2001–2018 |
| 94     | US-Ne2    | Mead–irrigated maize-soybean rotation site | 41.16 | -96.47 | 362 | 2001–2018 |
| 95     | US-Ne3    | Mead–rainfed maize-soybean rotation site | 41.18 | -96.44 | 363 | 2001–2018 |
| 96     | US-Orv    | Olentangy River Wetland Research Park | 40.02 | -83.02 | 221 | 2011–2016 |
| 97     | US-Ohio   | Oak Openings | 41.55 | -83.84 | 230 | 2004–2013 |
| 98     | US-PHm    | Plum Island High Marsh | 42.74 | -70.83 | 1.4 | 2013–2018 |
| 99     | US-PnP    | Lake Mendota, Picnic Point Site | 43.09 | -89.42 | 260 | 2016–2018 |
| 100    | US-Prr    | Poker Flat Research Range Black Spruce Forest | 65.12 | -147.49 | 210 | 2010–2016 |
| 101    | US-Rls    | RCEW Low Sagebrush | 43.14 | -116.74 | 1608 | 2014–2018 |
| 102    | US-Rms    | RCEW Mountain Big Sagebrush | 43.06 | -116.75 | 2111 | 2014–2018 |
| 103    | US-Ro1    | Rosemount- G21 | 44.71 | -93.09 | 260 | 2004–2016 |
| 104    | US-Ro2    | Rosemount- C7 | 44.73 | -93.09 | 292 | 2015–2016 |
Table A1. Cont.

| Number | Site Code | Site Name | Latitude (deg) | Longitude (deg) | Elevation (m) | Time Period |
|--------|-----------|-----------|----------------|-----------------|--------------|-------------|
| 105    | US-Ro4    | Rosemount Prairie | 44.68          | -93.07          | 274          | 2015–2018   |
| 106    | US-Ro5    | Rosemount I18_South | 44.69          | -93.06          | 283          | 2017–2018   |
| 107    | US-Ro6    | Rosemount I18_North | 44.69          | -93.06          | 282          | 2017–2018   |
| 108    | US-Rpf    | Succession from fire scar to deciduous forest | 65.12         | -147.43         | 497          | 2013–2018   |
| 109    | US-Rwe    | RCEW Reynolds Mountain East | 43.07       | -116.76          | 2098         | 2005–2007   |
| 110    | US-Rwf    | RCEW Upper Sheep Prescribed Fire | 43.12       | -116.72          | 1878         | 2014–2018   |
| 111    | US-Rws    | Reynolds Creek Wyoming big sagebrush | 43.17        | -116.71          | 1425         | 2014–2018   |
| 112    | US-SFP    | Sioux Falls Portable | 43.24         | -96.90           | 386          | 2007–2009   |
| 113    | US-SRC    | Santa Rita Creosote | 31.91         | -110.84          | 950          | 2008–2014   |
| 114    | US-SRG    | Santa Rita Grassland | 31.79         | -110.83          | 1291         | 2008–2014   |
| 115    | US-SRM    | Santa Rita Mesquite | 31.82         | -110.87          | 1120         | 2004–2018   |
| 116    | US-Seg    | Sevilleta grassland | 34.36         | -106.70          | 1622         | 2007–2018   |
| 117    | US-Ses    | Sevilleta shrubland | 34.33         | -106.74          | 1604         | 2007–2018   |
| 118    | US-Skr    | Shark River Slough (Tower SRS-6 Everglades) | 25.36        | -81.08           | 0            | 2004–2011   |
| 119    | US-Slt    | Silas Litle- New Jersey | 39.91        | -74.60           | 30           | 2007–2012   |
| 120    | US-Sne    | Sherman Island Restored Wetland | 38.04       | -121.75          | -5           | 2016–2018   |
| 121    | US-Snf    | Sherman Barn | 38.04         | -121.73          | -4           | 2018        |
| 122    | US-Srr    | Suisun marsh–Rush Ranch | 38.20        | -122.03          | 8            | 2014–2017   |
| 123    | US-Ton    | Ton Ranch | 38.43         | -120.97          | 177          | 2014–2018   |
| 124    | US-Tw1    | Twitchell Wetland West Pond | 38.11        | -121.65          | -5           | 2011–2018   |
| 125    | US-Tw2    | Twitchell Corn | 38.10         | -121.64          | -5           | 2012–2013   |
| 126    | US-Tw3    | Twitchell Alfalfa | 38.12         | -121.65          | -4           | 2013–2018   |
| 127    | US-Tw4    | Twitchell East End Wetland | 38.10        | -121.64          | -5           | 2013–2018   |
| 128    | US-Tw5    | East Pond Wetland | 38.11         | -121.64          | -5           | 2018        |
| 129    | US-UM3    | Douglas Lake | 45.57         | -84.67           | 234          | 2013–2014   |
| 130    | US-UMB    | Univ. of Mich. Biological Station | 45.56      | -84.71           | 234          | 2007–2018   |
| 131    | US-UMd    | UMBS Disturbance | 45.56         | -84.70           | 239          | 2008–2018   |
| 132    | US-Uaf    | University of Alaska, Fairbanks | 64.07       | -147.86          | 155          | 2009–2018   |
| 133    | US-Ula    | University of Illinois Switchgrass | 40.06      | -88.20           | 224          | 2015        |
| 134    | US-Var    | Vaira Ranch- Lane | 38.41         | -120.95          | 129          | 2004–2018   |
| 135    | US-Vcm    | Valles Caldera Mixed Conifer | 35.89       | -106.83          | 3003         | 2009–2018   |
| 136    | US-Vcp    | Valles Caldera Ponderosa Pine | 35.86       | -106.60          | 2542         | 2007–2018   |
| 137    | US-Vcs    | Valles Caldera Sulphur Springs Mixed Conifer | 35.92      | -106.61          | 2752         | 2016–2018   |
| 138    | US-WBW    | Walker Branch Watershed | 35.96        | -84.29           | 283          | 2001–2007   |
| 139    | US-WCr    | Willow Creek | 45.81         | -90.08           | 520          | 2000–2018   |
| 140    | US-WPT    | Winous Point North Marsh | 41.46       | -83.00           | 175          | 2011–2013   |
| 141    | US-Wdn    | Walden | 40.78         | -106.26          | 2469         | 2006–2008   |
| 142    | US-Wgr    | Willamette Grass | 45.11         | -122.66          | 52           | 2015        |
| 143    | US-Whs    | Walnut Gulch Lucky Hills Shrub | 31.74        | -110.05          | 1370         | 2009–2018   |
| 144    | US-Wjs    | Willard Juniper Savannah | 34.43       | -105.86          | 1931         | 2007–2018   |
| 145    | US-Wkg    | Walnut Gulch Kendall Grasslands | 31.74       | -109.94          | 1531         | 2004–2018   |
| 146    | US-Wpp    | Willamette Poplar | 44.14         | -123.18          | 111          | 2015        |
| 147    | US-Wrc    | Wind River Crane Site | 45.82        | -121.95          | 371          | 2000–2015   |
| 148    | US-xAB    | NEON Abby Road (ABBY) | 45.76        | -122.33          | 363          | 2017–2018   |
| 149    | US-xBN    | NEON Caribou Creek–Poker Flats Watershed (BONA) | 65.15       | -147.50          | 263          | 2018        |
| 150    | US-xBR    | NEON Bartlett Experimental Forest (BART) | 44.06       | -71.29           | 232          | 2017–2018   |
| 151    | US-xCP    | NEON Central Plains Experimental Range (CPER) | 40.82       | -104.75          | 1654         | 2017–2018   |
| 152    | US-xDC    | NEON Dakota Coteau Field School (DCFS) | 47.16        | -99.11           | 559          | 2017–2018   |
| 153    | US-xDJ    | NEON Delta Junction (DEJU) | 63.88        | -145.75          | 529          | 2017–2018   |
Table A1. Cont.

| Number | Site Code | Site Name | Latitude (deg) | Longitude (deg) | Elevation (m) | Time Period |
|--------|-----------|-----------|----------------|----------------|--------------|-------------|
| 154    | US-xDL    | NEON Dead Lake (DELA) | 32.54 | −87.80 | 22 | 2017–2018 |
| 155    | US-xGR    | NEON Great Smoky Mountains National Park, Twin Creeks (GRSM) | 35.69 | −83.50 | 579 | 2018 |
| 156    | US-xHA    | NEON Harvard Forest (HARV) | 42.54 | −72.17 | 351 | 2017–2018 |
| 157    | US-xHE    | NEON Healy (HEAL) | 63.88 | −149.21 | 705 | 2017–2018 |
| 158    | US-xJE    | NEON Jones Ecological Research Center (JERC) | 31.19 | −84.47 | 44 | 2017–2018 |
| 159    | US-xJR    | NEON Jornada LTER (JORN) | 32.59 | −106.84 | 1329 | 2017–2018 |
| 160    | US-xKA    | NEON Konza Prairie Biological Station–Relocatable (KONA) | 39.11 | −96.61 | 1329 | 2017–2018 |
| 161    | US-xKZ    | NEON Konza Prairie Biological Station (KONZ) | 39.10 | −96.56 | 381 | 2017–2018 |
| 162    | US-xNG    | NEON Northern Great Plains Research Laboratory (NOGP) | 46.77 | −100.92 | 578 | 2017–2018 |
| 163    | US-xNQ    | NEON Onaqui-Ault (ONAQ) | 40.18 | −112.45 | 1685 | 2017–2018 |
| 164    | US-xRM    | NEON Rocky Mountain National Park, CASTNET (RMNP) | 40.28 | −105.55 | 2743 | 2017–2018 |
| 165    | US-xSE    | NEON Smithsonian Environmental Research Center (SERC) | 38.89 | −76.56 | 15 | 2017–2018 |
| 166    | US-xSL    | NEON North Sterling, CO (STER) | 40.46 | −103.03 | 1364 | 2017–2018 |
| 167    | US-xSP    | NEON Soaproot Saddle (SOAP) | 37.03 | −119.26 | 1160 | 2017–2018 |
| 168    | US-xSR    | NEON Santa Rita Experimental Range (SRER) | 31.91 | −110.84 | 983 | 2017–2018 |
| 169    | US-xST    | NEON Steigerwaldt Land Services (STEI) | 45.51 | −89.59 | 481 | 2017–2018 |
| 170    | US-xTE    | NEON Lower Teakettle (TEAK) | 37.01 | −119.01 | 2147 | 2018 |
| 171    | US-xTR    | NEON Treehaven (TREE) | 45.49 | −89.59 | 472 | 2017–2018 |
| 172    | US-xUK    | NEON The University of Kansas Field Station (UKFS) | 39.04 | −95.19 | 335 | 2017–2018 |
| 173    | US-xUN    | NEON University of Notre Dame Environmental Research Center (UNDE) | 46.23 | −89.54 | 518 | 2017–2018 |
| 174    | US-xWD    | NEON Woodworth (WOOD) | 47.13 | −99.24 | 579 | 2017–2018 |
| 175    | US-xWR    | NEON Wind River Experimental Forest (WREF) | 45.82 | −121.95 | 407 | 2018 |
| 176    | MSE       | Mase paddy flux site | 36.05 | 140.03 | 13 | 2001 |
| 177    | PSO       | Pasoh Forest Reserve | 2.97 | 102.31 | 75–150 | 2003–2009 |
| 178    | BKS       | Bukit Soeharto | −0.86 | 117.04 | 20 | 2001–2002 |
| 179    | CBS       | Changbaishan Site | 41.40 | 128.10 | 731 | 2003–2005 |
| 180    | FHK       | Fuji Hokuuroku Flux Observation Site | 35.44 | 138.76 | 1050–1150 | 2006–2012 |
| 181    | GCK       | Gwangjeung Coniferous forest | 37.75 | 127.16 | 132 | 2007–2008 |
| 182    | HBG       | Haibe Potentilla fruticosa bosk Site | 37.48 | 101.20 | 756 | 2003–2004 |
| 183    | HFK       | Haenam Farmland | 34.55 | 127.57 | 12 | 2008 |
| 184    | IRI       | IRRI Flux Research Site | 14.14 | 121.27 | 21 | 2009–2014 |
| 185    | KBU       | Kherlenbayan Ulaan | 47.21 | 108.74 | 1235 | 2003–2009 |
| 186    | LSH       | Loashan | 45.28 | 127.58 | 340 | 2002 |
| 187    | MBF       | Moshiri Birch Forest Site | 44.38 | 142.32 | 585 | 2003–2005 |
| 188    | MKL       | Mae Klong | 14.59 | 98.84 | 585 | 2003–2004 |
| 189    | MMF       | Moshiri Mixd Forest Site | 44.32 | 142.26 | 340 | 2003–2005 |
| 190    | PDF       | Palangkaraya drained forest | −2.35 | 114.04 | 30 | 2002–2005 |
| 191    | QYZ       | Qianyanzhou Site | 26.73 | 115.07 | 100 | 2003–2004 |
| 192    | SKR       | Sakera | 14.49 | 101.92 | 543 | 2001–2003 |
| 193    | SKT       | Southern Khentii Taiga | 48.35 | 108.65 | 1630 | 2003–2006 |
| 194    | SMF       | Seto Mixed Forest Site | 35.26 | 137.08 | 205 | 2002–2015 |
| 195    | SWL       | Suwa Lake Site | 36.05 | 138.11 | 799 | 2015–2018 |
| 196    | TKC       | Takayama evergreen coniferous forest site | 36.14 | 137.37 | 800 | 2007 |
| 197    | TMK       | Tomakomai Flux Research Site | 42.74 | 141.51 | 140 | 2001–2003 |
| Number | Site Code | Site Name | Latitude (deg) | Longitude (deg) | Elevation (m) | Time Period |
|--------|-----------|-----------|----------------|----------------|--------------|-------------|
| 198    | TSE       | CC-LaG Teshio Experimental Forest | 45.06          | 142.11         | 70           | 2001–2005   |
| 199    | YCS       | Yuchen Site | 36.83          | 116.57         | 28           | 2003–2005   |
| 200    | YLF       | Yakutsk Spasskaya Pad larch  | 62.26          | 129.17         | 220          | 2003–2007   |
| 201    | YPF       | Yakutsk Pine | 62.24          | 129.65         | 220          | 2004–2007   |
| 202    | APE       | Alert      | 82.49          | −62.42         | 127          | 2004–2014   |
| 203    | ASP       | Alice Springs | −23.30        | 135.89         | 547          | 2000–2018   |
| 204    | BAR       | Barrow     | 71.32          | −156.61        | 8            | 2000–2017   |
| 205    | BIL       | Billings   | 36.61          | −97.52         | 317          | 2000–2017   |
| 206    | BON       | Bondville  | 40.07          | −88.37         | 213          | 2009–2018   |
| 207    | BOS       | Boulder    | 40.13          | −105.24        | 1689         | 2009–2018   |
| 208    | BOU       | Boulder    | 40.05          | −105.01        | 1577         | 2000–2016   |
| 209    | BRB       | Brasilia   | −15.60         | −47.71         | 1023         | 2008–2018   |
| 210    | CAB       | Cabauw     | 51.97          | 4.93           | 0            | 2005–2018   |
| 211    | CAM       | Camborne   | 50.22          | −5.32          | 88           | 2001–2017   |
| 212    | CAR       | Carpentras | 44.08          | 5.06           | 100          | 2000–2018   |
| 213    | CNR       | Cener      | 42.82          | −1.60          | 471          | 2009–2018   |
| 214    | COC       | Cocos Island | −12.19        | 96.84          | 6            | 2004–2018   |
| 215    | DAA       | De Aar     | −30.67         | 23.99          | 1287         | 2000–2018   |
| 216    | DAR       | Darwin     | −12.43         | 130.89         | 30           | 2002–2015   |
| 217    | DOM       | Concordia Station, Dome C | −75.10        | 123.38         | 3233         | 2006–2018   |
| 218    | DRA       | Desert Rock | 36.63          | −116.02        | 1007         | 2009–2018   |
| 219    | DWN       | Darwin Met Office | −12.42        | 130.89         | 32           | 2008–2018   |
| 220    | E13       | Southern Great Plains | 36.61          | −97.49         | 318          | 2000–2017   |
| 221    | ENA       | Eastern North Atlantic | 39.09          | −28.03         | 15.2         | 2013–2015   |
| 222    | EUR       | Eureka     | 79.99          | −85.94         | 85           | 2007–2011   |
| 223    | FLO       | Florianopolis | −27.60         | −48.52         | 11           | 2013–2018   |
| 224    | FPE       | Fort Peck   | 48.32          | −105.10        | 634          | 2009–2018   |
| 225    | FUA       | Fukuoka    | 33.58          | 130.38         | 3            | 2010–2018   |
| 226    | GAN       | Gandhinagar | 23.11          | 72.63          | 65           | 2014–2015   |
| 227    | GCR       | Goodwin Creek | 34.25          | −89.87         | 98           | 2009–2018   |
| 228    | GOB       | Gobabeb    | −23.56         | 15.04          | 407          | 2012–2018   |
| 229    | GR R      | Gurgaon    | 28.42          | 77.16          | 259          | 2014–2018   |
| 230    | GVN       | Georg von Neumayer | −70.65        | −8.25          | 42           | 2000–2008   |
| 231    | HOW       | Howrah     | 22.55          | 88.31          | 51           | 2014–2018   |
| 232    | ISH       | Ishigakijima | 24.34          | 124.16         | 5.7           | 2010–2018   |
| 233    | LAU       | Lauder     | −45.05         | 169.69         | 350          | 2000–2008   |
| 234    | LER       | Lerwick    | 60.14          | −1.18          | 80           | 2001–2017   |
| 235    | LIN       | Lindenberg | 52.21          | 14.12          | 125          | 2000–2017   |
| 236    | LRC       | Langley Research Center | 37.10          | −76.39         | 3            | 2014–2018   |
| 237    | LYU       | Lanyu Station | 22.04          | 121.56         | 324          | 2018        |
| 238    | MAN       | Momote     | −2.06          | 147.43         | 6            | 2000–2013   |
| 239    | NAU       | Nauru Island | −0.52          | 166.92         | 7            | 2000–2013   |
| 240    | NEW       | Newcastle  | −32.88         | 151.73         | 18.5         | 2017–2018   |
| 241    | NYA       | Ny-Ålesund | 78.93          | 11.93          | 11           | 2000–2018   |
| 242    | PAL       | Palaiseau, SIRTA Observatory | 48.71          | 2.21           | 156          | 2003–2018   |
| 243    | PAY       | Payerne    | 46.82          | 6.94           | 491          | 2000–2018   |
| 244    | PSU       | Rock Springs | 40.72          | −77.93         | 376          | 2009–2018   |
| 245    | PTR       | Petrolina  | −9.07          | −40.32         | 387          | 2008–2018   |
| 246    | REG       | Regina     | 50.21          | −104.71        | 578          | 2000–2011   |
| 247    | SAP       | Sapporo    | 43.06          | 141.33         | 17.2         | 2010–2018   |
| 248    | SBO       | Sede Boquer | 30.86          | 34.78          | 500          | 2003–2012   |
| 249    | SMS       | São Martinho da Serra | −29.44        | −53.82         | 489          | 2008–2017   |
| 250    | SON       | Sonnblick  | 47.05          | 12.96          | 3108.9       | 2013–2018   |
| 251    | SOV       | Solar Village | 24.91          | 46.41          | 650          | 2000–2002   |
| 252    | SXF       | Sioux Falls | 43.73          | −96.62         | 473          | 2009–2018   |
| 253    | SYO       | Syowa      | −69.01         | 39.59          | 18           | 2000–2018   |
| 254    | TAM       | Tamanrasset | 22.79          | 5.53           | 1385         | 2000–2018   |
| 255    | TAT       | Tateno     | 36.06          | 140.13         | 25           | 2000–2018   |
Table A1. Cont.

| Number | Site Code | Site Name | Latitude (deg) | Longitude (deg) | Elevation (m) | Time Period |
|--------|-----------|-----------|----------------|----------------|--------------|-------------|
| 256    | TIR       | Tiruvallur | 13.09          | 79.97          | 36           | 2014–2018   |
| 257    | TOR       | Toravere  | 58.25          | 26.46          | 70           | 2003–2018   |
| 258    | XIA       | Xianghe   | 39.75          | 116.96         | 32           | 2005–2015   |
| 259    | AT-Neu    | Neustift   | 47.12          | 11.32          | 970          | 2005–2012   |
| 260    | AU-ASM    | Alice Springs | −22.28      | 133.25        |   | 2010–2014   |
| 261    | AU-Ade    | Adelaide River | −13.08      | 131.12        |   | 2007–2009   |
| 262    | AU-Cpr    | Calperum   | −34.00         | 140.59         |   | 2010–2014   |
| 263    | AU-Cum    | Cumberland | −33.62         | 150.72         |   | 2012–2014   |
| 264    | AU-DaP    | Daly River Savanna | −14.06     | 131.32        |   | 2007–2013   |
| 265    | AU-DaS    | Daly River Cleared | −14.16    | 131.39        |   | 2008–2014   |
| 266    | AU-Dry    | Dry River  | −15.26         | 132.37         |   | 2008–2014   |
| 267    | AU-Emr    | Emerald    | −23.86         | 148.47         |   | 2011–2013   |
| 268    | AU-Fog    | Fogg Dam   | −12.55         | 131.31         |   | 2006–2008   |
| 269    | AU-GWW    | Great Western Woodlands, Western Australia, Australia | −30.19     | 120.65        |   | 2013–2014   |
| 270    | AU-Gin    | Gingin     | −31.38         | 115.71         |   | 2011–2014   |
| 271    | AU-Lox    | Loxton     | −34.47         | 140.66         |   | 2008–2009   |
| 272    | AU-RDF    | Red Dirt Melon Farm, Northern Territory | −14.56     | 132.48         |   | 2011–2013   |
| 273    | AU-Rig    | Riggs Creek | −36.65         | 145.58         |   | 2011–2014   |
| 274    | AU-Rob    | Robson Creek, Queensland, Australia | −17.12     | 145.63         |   | 2014         |
| 275    | AU-Stp    | Sturt Plains | −17.15        | 133.35         |   | 2008–2014   |
| 276    | AU-TTE    | Ti Tree East | −22.29        | 133.64         |   | 2012–2014   |
| 277    | AU-Tum    | Tumburumba | −35.66         | 148.15         | 1200 | 2007–2014   |
| 278    | AU-Whr    | Whroo       | −36.67         | 145.03         |   | 2011–2014   |
| 279    | AU-Wom    | Wombat      | −37.42         | 144.09         | 705  | 2010–2014   |
| 280    | AU-Ync    | Jaxa        | −34.99         | 146.29         |   | 2012–2014   |
| 281    | BE-Bra    | Brasschaat  | 51.31          | 4.52           | 16  | 2007–2014   |
| 282    | BE-Lon    | Lonzee      | 50.55          | 4.75           | 167 | 2005–2014   |
| 283    | CH-Cha    | Chamau      | 47.21          | 8.41           | 393 | 2005–2014   |
| 284    | CH-Dav    | Davos       | 46.82          | 9.86           | 1639 | 2006–2014   |
| 285    | CH-Fru    | Fruebitzel  | 47.12          | 8.54           | 982 | 2005–2014   |
| 286    | CH-Lae    | Laegern     | 47.48          | 8.36           | 689  | 2005–2014   |
| 287    | CH-Oe1    | Oensinger grassland | 47.29     | 7.73           | 450 | 2003–2008   |
| 288    | CH-Oe2    | Oensinger crop | 47.29       | 7.73           | 452  | 2004–2014   |
| 289    | CN-Cha    | Changbaishan | 42.40         | 128.10         |   | 2003–2005   |
| 290    | CN-Cng    | Changling   | 44.59          | 123.51         |   | 2007–2010   |
| 291    | CN-Dan    | Dangxiong  | 30.50          | 91.07          |   | 2004–2005   |
| 292    | CN-Din    | Dinghusan  | 23.17          | 112.54         |   | 2003–2005   |
| 293    | CN-Ha2    | Haibe Shrubland | 37.61       | 101.33         |   | 2003–2005   |
| 294    | CN-Qia    | Qianyanzhou | 26.74          | 115.06         |   | 2003–2005   |
| 295    | CZ-wet    | Trebon (CZECHWET) | 49.02     | 14.77           | 426  | 2006–2014   |
| 296    | DE-Akm    | Anklam      | 53.87          | 13.68          | −1  | 2009–2014   |
| 297    | DE-Geb    | Gebesee     | 51.10          | 10.91          | 161.5 | 2001–2014   |
| 298    | DE-Gri    | Grillenburg | 50.95          | 13.51          | 385  | 2006–2014   |
| 299    | DE-Hai    | Hainich     | 51.08          | 10.45          | 430  | 2002–2012   |
| 300    | DE-Kli    | Klingenber | 50.89          | 13.52          | 478  | 2004–2014   |
| 301    | DE-Lkb    | Lackenberg  | 49.10          | 13.30          | 1308 | 2009–2013   |
| 302    | DE-Lnf    | Leinefelde  | 51.33          | 10.37          | 451  | 2002–2012   |
| 303    | DE-Oba    | Oberbarenburg | 50.79       | 13.72          | 734  | 2008–2014   |
| 304    | DE-RuR    | Rollesbroich | 50.62         | 6.30           | 514.7 | 2011–2014   |
| 305    | DE-RuS    | Selhausen Juelich | 50.87       | 6.45           | 102.755 | 2011–2014 |
| 306    | DE-SfN    | Schechenfilz Nord | 47.81       | 11.33           | 590  | 2012–2014   |
| 307    | DE-Spw    | Spreewald   | 51.89          | 14.03          | 61   | 2010–2014   |
| 308    | DE-Thu    | Tharanntd   | 50.96          | 13.57          | 385  | 2004–2014   |
| 309    | DE-Zrk    | Zarnekoew   | 53.88          | 12.89          | 0    | 2013–2014   |
| 310    | FI-Hyy    | Hyytiala    | 61.85          | 24.29          | 181  | 2009–2014   |
| 311    | FI-Lom    | Lompolojanka | 68.00          | 24.21          | 274  | 2007–2009   |
Table A1. Cont.

| Number | Site Code | Site Name                  | Latitude (deg) | Longitude (deg) | Elevation (m) | Time Period    |
|--------|-----------|----------------------------|----------------|----------------|---------------|----------------|
| 312    | FR-Gri    | Grignon                    | 48.84          | 1.95           | 125           | 2004–2014      |
| 313    | FR-LBr    | Le Bray                    | 44.72          | −0.77          | 61            | 2003–2008      |
| 314    | FR-Pue    | Puechabon                  | 43.74          | 3.60           | 270           | 2005–2014      |
| 315    | IT-BCi    | Borgo Cioffi               | 40.52          | 14.96          | 26            | 2006–2011      |
| 316    | IT-CA1    | Castel d’Asso1             | 42.38          | 12.03          | 200           | 2011–2014      |
| 317    | IT-CA2    | Castel d’Asso2             | 42.38          | 12.03          | 200           | 2011–2014      |
| 318    | IT-CA3    | Castel d’Asso3             | 42.38          | 12.02          | 197           | 2011–2014      |
| 319    | IT-Col    | Collelongo                 | 41.85          | 13.59          | 1560          | 2004–2014      |
| 320    | IT-Isp    | Ispra ABC-IS               | 45.81          | 8.63           | 210           | 2013–2014      |
| 321    | IT-La2    | Lavarone2                  | 45.95          | 11.29          | 1350          | 2000–2002      |
| 322    | IT-Lav    | Lavarone                   | 45.96          | 11.28          | 1353          | 2003–2004      |
| 323    | IT-MBo    | Monte Bondone              | 46.01          | 11.05          | 1550          | 2003–2013      |
| 324    | IT-Noe    | Arca di Noe–Le Prigionette | 40.61          | 8.15           | 25            | 2004–2014      |
| 325    | IT-Ren    | Renon                      | 46.59          | 11.43          | 1730          | 2003–2013      |
| 326    | IT-Ro2    | Roccarespampioni 2         | 42.39          | 11.92          | 160           | 2010–2012      |
| 327    | IT-SR2    | San Rossore 2              | 43.73          | 10.29          | 4             | 2013–2014      |
| 328    | IT-SRo    | San Rossore                | 43.73          | 10.28          | 6             | 2004–2008      |
| 329    | IT-Tor    | Torgnon                    | 45.84          | 7.58           | 2160          | 2008–2014      |
| 330    | JP-MBF    | Moshiri Birch Forest Site  | 44.39          | 142.32         |               | 2003–2005      |
| 331    | NL-Hor    | Horstermeer                | 52.24          | 5.07           | 265           | 2004–2011      |
| 332    | NL-Loo    | Loobos                     | 52.17          | 5.74           | 25            | 2000–2014      |
| 333    | RU-Che    | Cherski                    | 68.61          | 161.34         | 6             | 2002–2005      |
| 334    | RU-Fyo    | Fyodorovskoye              | 56.46          | 32.92          | 263           | 2000–2014      |
| 335    | SE-St1    | Stordalen grassland        | 48.35          | 19.05          | 351           | 2012–2014      |
| 336    | SJ-Blv    | Bayelva, Spitsbergen       | 78.92          | 11.83          | 25            | 2008–2009      |
| 337    | US-CRT    | Curtice Walter-Berger cropland | 41.63       | −83.35         | 180           | 2011–2013      |
| 338    | US-GBT    | GLEES Brooklyn Tower       | 41.37          | −106.24        | 3191          | 2000–2006      |
| 339    | US-Syv    | Sylvania Wilderness Area   | 46.24          | −89.35         | 540           | 2010–2014      |
| 340    | US-Tw4    | Twitchell East End Wetland | 38.10          | −121.64        | −5            | 2013–2014      |
| 341    | ZA-Kru    | Skukuza                    | −25.02         | 31.50          | 359           | 2000–2003      |
| 342    | ZM-Mon    | Mongu                      | −15.44         | 23.25          | 1053          | 2000–2009      |
| 343    | BND       | Bondville                  | 40.05          | −88.37         | 230           | 2000–2018      |
| 344    | DRA       | Desert Rock                | 36.62          | −116.02        | 1007          | 2000–2018      |
| 345    | FPK       | Fort Peck                  | 48.31          | −105.10        | 634           | 2000–2018      |
| 346    | GWN       | Goodwin Creek              | 34.25          | −89.87         | 98            | 2000–2018      |
| 347    | PSU       | Penn State                 | 40.72          | −77.93         | 376           | 2000–2018      |
| 348    | SXF       | Sioux Falls                | 43.73          | −96.62         | 473           | 2003–2018      |
| 349    | TBL       | Table Mountain             | 40.13          | −105.24        | 1689          | 2000–2018      |

The first 175 stations are the AmeriFlux sites, followed by 26 AsiaFlux sites (beginning with site code named “MSE”), 57 BSRN sites (beginning with site code named “ALE”), 84 FLUXNET sites (beginning with site code named “AT-Neu”), and 7 SURFRAD sites (beginning with site code named “BND”).

Algorithm A1. The Gradient Boosting Regression Tree Algorithm.

Initialize $f_0(x) = \arg \min_p \sum_{i=1}^{N} L(y_i, p)$

For $m = 1$ to $M$

For $i = 1$ to $N$

Compute the negative gradient $\tilde{y}_{im} = -\left[ \frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right] f(x) = f_{m-1}(x) - f_{m-1}(x)$

End

Fit a regression tree $h(x; \alpha_m)$ to predict the targets $\tilde{y}_{im}$ from covariates $x_i$ for all training dataset

Compute a gradient descent step size as $\rho_m = \arg \min_p \sum_{i=1}^{N} L(y_i, f_{m-1}(x_i) + \rho h(x; \alpha_m))$

Update the model as $f_m(x) = f_{m-1}(x) + \rho_m h(x; \alpha_m)$

End

Output the final model $f_M(x)$
87. Jiang, H.; Lu, N.; Huang, G.H.; Yao, L.; Qin, J.; Liu, H.Z. Spatial scale effects on retrieval accuracy of surface solar radiation using satellite data. *Appl. Energy.* 2020, 270. [CrossRef]

88. Tang, W.J.; Yang, K.; Qin, J.; Li, X.; Niu, X.L. A 16-year dataset (2000-2015) of high-resolution (3 h, 10 km) global surface solar radiation. *Earth Syst. Sci. Data* 2019, 11, 1905–1915. [CrossRef]