

Article

# Estimating Land Surface Temperature from Satellite Passive Microwave Observations with the Traditional Neural Network, Deep Belief Network, and Convolutional Neural Network

Shaofei Wang <sup>1</sup>, Ji Zhou <sup>1,\*</sup>, Tianjie Lei <sup>2</sup>, Hua Wu <sup>3</sup>, Xiaodong Zhang <sup>1</sup>, Jin Ma <sup>1</sup> and Hailing Zhong <sup>1</sup>

<sup>1</sup> School of Resources and Environment, Center for Information Geoscience, University of Electronic Science and Technology of China, Chengdu 611731, China; 201822070311@std.uestc.edu.cn (S.W.); 201611180105@std.uestc.edu.cn (X.Z.); majin@std.uestc.edu.cn (J.M.); 201721180114@std.uestc.edu.cn (H.Z.)

<sup>2</sup> China Institute of Water Resources and Hydropower Research (IWHR), Beijing 100038, China; leitj@iwhr.com

<sup>3</sup> State Key Laboratory of Resources and Environment Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; wuhua@igsnrr.ac.cn

\* Correspondence: jzhou233@uestc.edu.cn

Received: 16 July 2020; Accepted: 17 August 2020; Published: 20 August 2020



**Abstract:** Neural networks, especially the latest deep learning, have exhibited good ability in estimating surface parameters from satellite remote sensing. However, thorough examinations of neural networks in the estimation of land surface temperature (LST) from satellite passive microwave (MW) observations are still lacking. Here, we examined the performances of the traditional neural network (NN), deep belief network (DBN), and convolutional neural network (CNN) in estimating LST from the AMSR-E and AMSR2 data over the Chinese landmass. The examinations were based on the same training set, validation set, and test set extracted from 2003, 2004, and 2009, respectively, for AMSR-E with a spatial resolution of  $0.25^\circ$ . For AMSR2, the three sets were extracted from 2013, 2014, and 2016 with a spatial resolution of  $0.1^\circ$ , respectively. MODIS LST played the role of “ground truth” in the training, validation, and testing. The examination results show that CNN is better than NN and DBN by 0.1–0.4 K. Different combinations of input parameters were examined to get the best combinations for the daytime and nighttime conditions. The best combinations are the brightness temperatures (BTs), NDVI, air temperature, and day of the year (DOY) for the daytime and BTs and air temperature for the nighttime. By adding three and one easily obtained parameters on the basis of BTs, the accuracies of LST estimates can be improved by 0.8 K and 0.3 K for the daytime and nighttime conditions, respectively. Compared with the MODIS LST, the CNN LST estimates yielded root-mean-square differences (RMSDs) of 2.19–3.58 K for the daytime and 1.43–2.14 K for the nighttime for diverse land cover types for AMSR-E. Validation against the in-situ LSTs showed that the CNN LSTs yielded root-mean-square errors of 2.10–4.72 K for forest and cropland sites. Further intercomparison indicated that ~50% of the CNN LSTs were closer to the MODIS LSTs than ESA’s GlobTemperature AMSR-E LSTs, and the average RMSDs of the CNN LSTs were less than 3 K over dense vegetation compared to NASA’s global land parameter data record air temperatures. This study helps better the understanding of the use of neural networks for estimating LST from satellite MW observations.

**Keywords:** land surface temperature; neural network; deep belief network; convolutional neural network; passive microwave

## 1. Introduction

Land surface temperature (LST) is a key factor in earth–atmosphere interactions and an important indicator for monitoring environmental changes and energy balance on Earth’s surface [1–3]. Compared with ground-based LST measurements, satellite remote sensing is more effective at regional and global scales [4]. Thermal infrared (TIR) remote sensing is the main approach for estimating LST from satellite remote sensing data. In recent decades, many algorithms have been proposed [5–7], and the TIR LST datasets have become relatively mature [8]. TIR remote sensing has the advantages of high accuracy and fine resolution. However, it can only obtain valid observations under clear-sky conditions.

On average, 60% of the land surface is covered by clouds, which limits the practical application of TIR LSTs [9–11]. An alternative method is to use satellite passive microwave (MW) measurements to estimate LSTs. MW has the advantage of the ability to penetrate clouds, which can cover the deficiencies of TIR. Although MW observations have coarser spatial resolution than TIR, LST estimated from MW observations can be an important complement to TIR LST [12,13]. Several methods have been proposed to estimate MW LST, and these methods can be grouped into three categories: empirical algorithms [9,12,14,15], semi-empirical algorithms [16–20], and physically-based algorithms [13,21]. Currently, there are three main problems in estimating accurate MW LST. The first problem is the difficulty of determining MW emissivity, which depends greatly on surface conditions (e.g., vegetation cover and snow cover) [22,23]. The complex relationships between these influencing factors and surface emissivity make it difficult to express MW emissivity using an analytical expression. The second problem is the atmospheric correction. Although MW remote sensing is less affected by the atmosphere than TIR remote sensing, the removal of atmospheric effects is still necessary for higher frequencies (>10 GHz). Finally, the temperature retrieved from MW measurements is a bulk temperature, which is evidently different from the skin temperature obtained by TIR. Thus, there exists a physical difference between MW LST and TIR LST. The difference is caused by different thermal sampling depths (TSDs) of MW and TIR [3,24].

Neural networks have strong nonlinear fitting ability and have been applied in the retrieval of MW LST. For example, Aires et al. [25] established a neural network model for the simultaneous retrieval of atmospheric water vapor, the cloud liquid water path, LST, and emissivities from special sensor microwave/imager (SSM/I) data with a spatial resolution of  $0.25^\circ$ ; the theoretical root-mean-square error (RMSE) of the LST over the globe is 1.3 K under clear-sky and 1.6 K under cloudy conditions. Jiménez et al. [26] and Ermida et al. [27] used neural networks for LST retrieval from the Advanced Microwave Scanning Radiometer for EOS (AMSR-E) data and generated the European Space Agency’s (ESA) GlobTemperature AMSR-E LST product for the period from 2008–2010 with a spatial resolution of  $14 \times 8$  km; to reduce the dependence on auxiliary data, monthly AMSR-E emissivities [22,28] were taken as the input parameters except for the brightness temperatures; the evaluation results showed that after masking out snow and deserts, the LST differences between AMSR-E and MODIS were  $3.0 \pm 5.1$  K and  $1.4 \pm 3.9$  K for the daytime and nighttime conditions, respectively. These studies have confirmed that neural networks can be applied to the estimation of MW LST. In fact, the neural network utilized by these studies was the traditional neural network (hereafter termed NN).

With the development of neural networks, there have been some other networks with different architectures named deep learning (e.g., deep belief network—DBN, and convolutional neural network—CNN). The DBN was proposed by Hinton et al. [29] to solve the optimization problem of deep neural networks, while CNN has been well recognized by the scientific communities for use in image processing [30]. Recently, DBN and CNN have been gradually applied to the estimation of remote sensing parameters. For example, Li et al. [31] compared the performance of DBN with other methods for the process of upscaling evapotranspiration, and the results show that the accuracy of DBN is lower than that of the other methods. Shen et al. [32] employed the DBN for air temperature estimation by fusing MODIS LST, station, simulation, and socioeconomic data, and found that the RMSE was  $1.996^\circ\text{C}$  at the national scale. Ge et al. [33] compared the performance of the deep convolutional neural

network (DCNN) with NN for estimating soil moisture and found that DCNN performs slightly better than NN. Tan et al. [34] constructed the LST retrieval model for the Advanced Microwave Scanning Radiometer 2 (AMSR2) data with CNN and found that its frequencies except 6.9 GHz were able to obtain the most accurate CNN model, which has an RMSE of 2.69 K. Sadeghi et al. [35] examined the performance of CNN in the estimation of precipitation and confirmed the good ability of CNN.

Although there have been some studies using these new networks for retrieving surface parameters from satellite remote sensing data, examinations on the performances of the neural networks in the estimation of LST from MW observations are still lacking. In addition, the CNN model constructed by Tan et al. [34] only takes AMSR2 brightness temperatures (BTs) as inputs. However, whether adding more input parameters that can be easily obtained would improve the accuracy of the LST estimate is a question that is important for the scientific communities. Under this context, the first objective of this study was to examine the performances of NN, DBN, and CNN based on multiple input parameters, including BTs, surface parameters, atmospheric related parameters, and day of the year (DOY), and the second objective was to estimate LST from MW observations based on the best performing neural network with the best input parameter combinations. In this study, China was selected as the study area for two main reasons. The first reason is that China has diverse land cover types, high elevation variations, and different climate types. The second reason is that China, especially South China, suffers from frequent cloud coverage. These two reasons make it a good area for examining the performances of these networks in MW LST estimation. The time period under examination was 2003–2010 for AMSR-E and 2013–2016 for AMSR2.

## 2. Datasets

### 2.1. AMSR-E and AMSR2 Data

AMSR-E is a twelve-channel and six-frequency passive microwave radiometer that measures the BTs of the Earth's surface at 6.925, 10.65, 18.7, 23.8, 36.5, and 89.0 GHz. Vertically and horizontally polarized measurements are taken at all frequencies. AMSR-E was onboard the Aqua satellite from 2002 to 2011. The native spatial resolution is ~5 km at 89.0 GHz and 60 km at 6.9 GHz. The AMSR-E BTs used here are from the AMSR-E level 3 product with a spatial resolution of 0.25°. This product is the daily average of BT in level 1B product and projected by equi-rectangular. As the successor of AMSR-E, AMSR2 is onboard the Global Change Observation Mission 1st-Water (GCOM-W1) satellite and has seven frequencies (i.e., 6.925, 7.3, 10.65, 18.7, 23.8, 36.5, and 89.0 GHz) in both horizontal and vertical polarizations. Compared with AMSR-E, AMSR2 has higher spatial resolutions. The native spatial resolution of AMSR2 is 3 × 5 km at 89.0 GHz and 35 × 62 km at 6.925 GHz. The AMSR2 BTs used here are from the AMSR2 level 3 product with a spatial resolution of 0.1°. This product is also daily average data of level 1B data and projected by equi-rectangular. The overpass time of AMSR-E and AMSR2 are both ~13:30 (ascending) and ~01:30 (descending) local solar time. The level 3 products of AMSR-E and AMSR2 were downloaded at <https://gportal.jaxa.jp/gpr/>.

### 2.2. MODIS Land Surface Products

Three MODIS land surface products were used to parameterize MW emissivity, including MODIS/Aqua Snow Cover Daily L3 (MYD10C1), MODIS/Terra+Aqua Land Cover Type Yearly L3 (MCD12C1), and MODIS/Aqua Vegetation Indices 16-Day L3 (MYD13C1). Spatially averaged MYD11C1 LST was used as the target LST of the samples. All products were in version 6. MODIS LST is currently the most accurate and widely used TIR satellite LST product and has been used as the target LST to establish the MW LST retrieval model [12,17,20,34] and to retrieve MW LSE [28,36,37]. Validation results over homogeneous surfaces showed that MODIS LST products generally have higher accuracies than 1.0 K [38,39]. The aforementioned MYD\*\*C1 products have a spatial resolution of 0.05° and were derived from Aqua observations. Thus, these products achieved better spatial and temporal matching with the AMSR-E and AMSR2 pixels. The snow cover (SC, from MYD10C1),

NDVI (from MYD13C1), land cover type percent (LCTP, from MCD12C1, i.e., the percent cover of 17 IGBP classes at each  $0.05^{\circ}$  pixel), and LST (from MYD11C1) were upscaled to the spatial resolution of AMSR-E and AMSR2 by spatial averaging. In addition, the latest MODIS LST and emissivity product MYD21A1 with a spatial resolution of 1 km was used to (i) derive the MODIS channel emissivities for the calculation of the broadband emissivity (BBE) of the ground stations and (ii) to quantify the thermal heterogeneity within the MW pixels. These MODIS land products were downloaded from EARTHDATA (<https://search.earthdata.nasa.gov>).

### 2.3. Reanalysis and Assimilation Datasets

To quantify the atmospheric conditions for the corresponding MW observations, we collected the 2-meter air temperature (AT-2m) and total precipitable water vapor (TPWV) from the second Modern-Era Retrospective analysis for Research and Applications (MERRA-2) from <https://disc.gsfc.nasa.gov>. In addition, the surface skin temperature was also extracted to synchronize the Aqua MODIS with AMSR-2 observations. The dataset used in this study was tavg1\_2d\_slv\_Nx, which has a grid size of  $0.5^{\circ} \times 0.625^{\circ}$  and temporal resolution of 1 hour from 00:30 UTC.

The soil moisture (SM) values in four layers (i.e., 0–10 cm, 10–40 cm, 40–100 cm, and 100–200 cm) were extracted from the Global Land Data Assimilation System (GLDAS) NOAH025\_3H (in version 2.1) with a grid size of  $0.25^{\circ} \times 0.25^{\circ}$  and temporal resolution of 3 hours. The GLDAS dataset was also downloaded from <https://disc.gsfc.nasa.gov>. The 1-hourly MERRA-2 product and 3-hourly GLDAS product were interpolated to the overpass time of the MW sensors and the spatial resolution of the MW BT products.

### 2.4. In-Situ Measurements

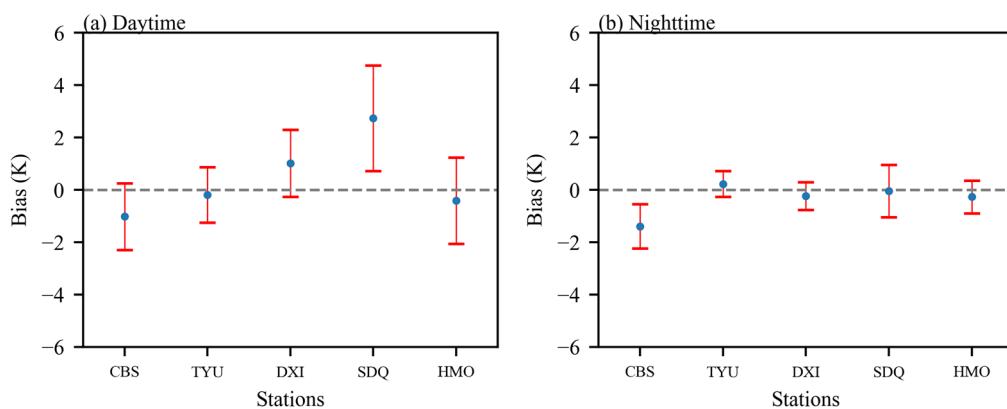
The in-situ measurements of five ground stations (CBS, TYU, DXI, SDQ, and HMO) with ground longwave radiation were collected to validate the LST estimates. Details about the five stations are shown in Table 1. Among these five stations, the measurements of CBS were from ChinaFlux (<http://www.chinaflux.org/>); the measurements of TYU were from the CEOP Asia-Australia Monsoon Project (CAMP) [40]; the measurements of DXI were from the Haihe experiments in the Hai River Basin, China [41]; and the measurements of SDQ and HMO were from the Heihe Watershed Allied Telemetry Experimental Research (HiWATER) program [42–44]. The elevations of these stations range from 20 m to 1050 m. The instruments for measuring the outgoing and incoming longwave radiation are the Kipp and Zonen CNR1 (spectral range: 5–50  $\mu\text{m}$ ; uncertainty in daily total: 10%) at CBS, DXI, and HMO; Kipp and Zonen CG4 pyrgeometer (spectral range: 4.5–42  $\mu\text{m}$ ; uncertainty in daily total: 3%) at TYU; and Kipp and Zonen CNR4 (spectral range: 4.5–42  $\mu\text{m}$ ; uncertainty in daily total: <10%) at SDQ. Although CNR1, CG4, and CNR4 have different uncertainties, the average errors of incoming and outgoing longwave radiation are  $\sim 6$  and  $3 \text{ W/m}^2$  [45,46], respectively, leading to an in-situ LST uncertainty of  $\sim 0.6 \text{ K}$  [3]. According to Stefan–Boltzmann’s law, in-situ LST can be calculated from the outgoing and incoming longwave radiation, and the BBE was calculated from the emissivities of MODIS channels 29, 31, and 32 [47]. Liang et al. [47] stated a theoretical residual standard error of 0.0289 for the BBE, leading to in-situ LST uncertainties of  $\sim 0.7 \text{ K}$  for CBS, TYU, DXI, and SDQ, and  $\sim 0.9 \text{ K}$  for HMO. The in-situ LST total uncertainties caused by measurements and BBE are  $\sim 0.9 \text{ K}$  for CBS, TYU, DXI, and SDQ and  $\sim 1.1 \text{ K}$  for HMO. Additionally, the  $3\sigma$  (standard deviations, STDs) criterion was used to filter the matched-up CNN LST and in-situ LST and remove the outliers [45,48,49].

**Table 1.** Details of the selected five ground stations.

Station	Longitude	Latitude	Instrument		Elevation (m)	Height (m)	Diameter of FOV (m)	Surface Type at Station	IGBP Class Percentin MW Pixel *	Period of Measurement	Interval of Measurement (min)
			Model								
CBS	128.10°E	42.40°N	Kipp & Zonen CNR1		736	6	44.78	Deciduous broadleaf forest	Deciduous broadleafforest: 62.1% Mixed Forests: 29.3% Woody Savannas: 4.4% Grasslands: 0.8% Croplands: 1.1% Urban and Built-up Lands: 2.3%	January 2003–December 2005	30
TYU	122.87°E	44.42°N	Kipp & Zonen CG4		184	3	22.39	Cropland	Grasslands: 91.6% Croplands: 8.4%	January 2003–December 2004	30
DXI	116.43°E	39.62°N	Kipp & Zonen CNR1		20	28	208.99	Cropland	Grasslands: 0.8% Croplands: 67.9% Urban and Built-up Lands: 31.3%	January 2009–December 2010	10
SDQ	101.14°E	42.00°N	Kipp & Zonen CNR4		873	10	74.63	Tamarix	Grasslands: 66.8% Barren: 33.2%	January 2015–December 2016	10
HMO	100.99°E	42.11°N	Kipp & Zonen CNR1		1054	6	44.78	Desert	Barren: 100%	May 2015–December 2016	10

Note: \* AMSR-E pixel for CBS, TYU, and DXI; AMSR2 pixel for SDQ and HMO.

As shown in Table 1, the mounting heights of these instruments vary from 3 m to 28 m above the ground surface, resulting in field-of-view (FOV) diameters ranging from 22.39 m to 208.99 m. However, the FOV is very small when compared with the spatial resolutions of AMSR-E and AMSR2. To assess the homogeneities of the MW pixels containing the stations, the percentage of the IGBP classes within the MW pixels were determined (Table 1). Furthermore, the LST differences between the MW pixels (reprojected by spatially averaging from MYD21A1 LST) and the 1-km MODIS pixels containing the ground stations were determined to evaluate thermal heterogeneities within the MW pixels (Figure 1). The statistical results were based on the MYD21A1 LST data of the time period of the in-situ measurements. For the daytime conditions, the LST of the MW spatial resolution has an underestimation for CBS, overestimation for DXI and SDQ, and no obvious systemic bias for TYU and HMO when compared with the 1-km MODIS LST. The underestimation of CBS may be induced by the elevation distribution, and the overestimations of DXI and SDQ are induced by over 30% of built-up surfaces and barren within the MW pixels, respectively. As for the nighttime conditions, only CBS has an underestimation induced by the elevation distribution. In addition, larger STDs can be seen during the daytime.



**Figure 1.** The LST differences between the MW pixels and the 1-km MODIS pixels containing the ground stations for the daytime (a) and nighttime (b). For CBS, TYU, and DXI, the spatial resolutions of the MW pixels are  $0.25^\circ$ ; for SDQ and HMO the spatial resolutions of the MW pixels are  $0.1^\circ$ . The symbol and error bar represent the mean value and STD of the LST difference, respectively.

## 2.5. Other Datasets

In addition to the aforementioned datasets, the ESA's GlobTemperature AMSR-E LST ( $14 \times 8$  km) [26,27] and the NASA's global land parameter data record (LPDR) air temperatures (ATs) (25 km) [50–53] are used for intercomparison. DEM data with a spatial resolution of 90 m were downloaded from the Geospatial Data Cloud (<http://www.gscloud.cn/>). The DEM data were used to filter the valid samples rather than directly as input parameters. More details can be found in Section 3.3.

## 3. Methodology

### 3.1. Neural Networks

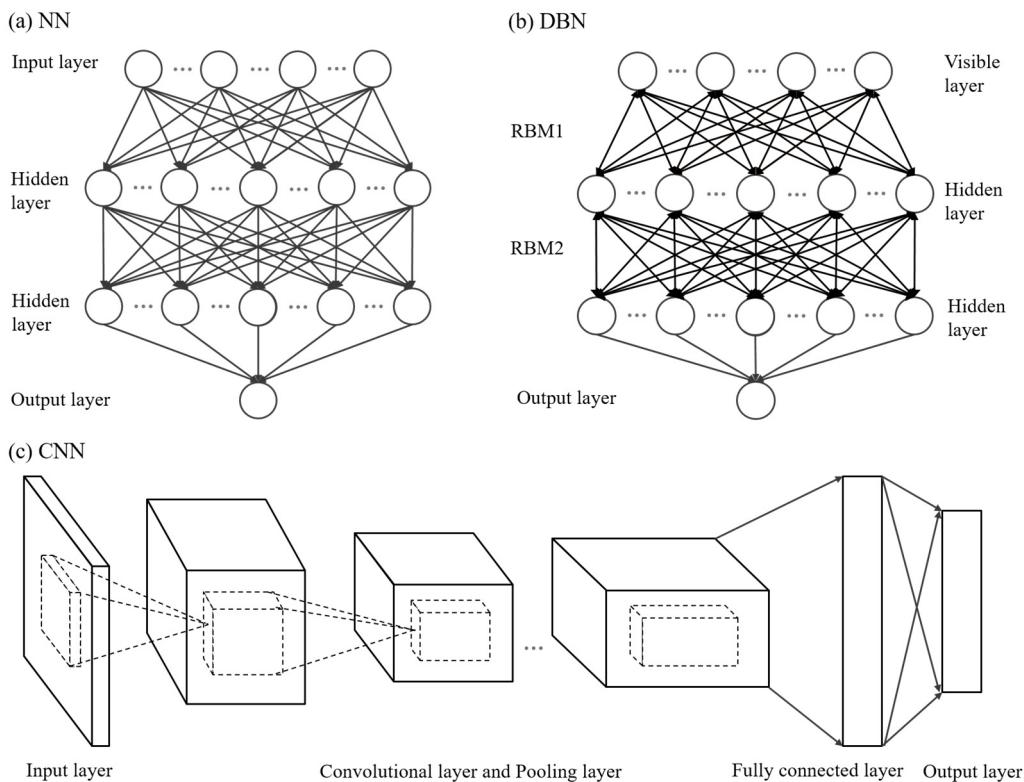
Three different neural networks, i.e., NN, DBN, and CNN, were selected for assessments of their performances in the estimation of LST from the aforementioned AMSR-E and AMSR2 observations. Their architectures are shown in Figure 2. NN consists of three layers, including an input layer, hidden layer, and output layer (see Figure 2a). In this study, NN was designed with 2 hidden layers, and the

numbers of hidden neurons were set as 64 and 48, respectively. Discussions on this point can be found in Section 5. The output of the hidden layer  $h$  and the output layer  $\hat{y}$  can be written as follows:

$$\begin{aligned} h &= \sigma(W_1x + b_1) \\ \hat{y} &= W_2h + b_2 \end{aligned} \quad (1)$$

where  $x$  is the input vector;  $W_1$  is the weight matrix between the input vector and output vector of the hidden layer;  $b_1$  is the bias matrix of the hidden layer;  $h$  is the output vector of the hidden layer;  $W_2$  is the weight matrix between the hidden layer and the output layer;  $b_2$  is the bias matrix of output layer; and  $\sigma$  is the activation function. The activation function is used to build a nonlinear model, and the activation function used in this study was the rectified linear unit (ReLU), which has the following formula:

$$f(x) = \max(x, 0) \quad (2)$$



**Figure 2.** The architectures of NN (a), DBN (b), and CNN (c).

DBN (see Figure 2b) is a generative model that generates training data according to the maximum probability by training the weights between its neurons. DBN is a stack of layers of restricted Boltzmann machine (RBM), which has only two layers of neurons, including the visible layer and hidden layer [29]. The characteristics of RBM are as follows: given the state of the visible units, the activation conditions of each hidden unit are independent; when determining the state of the hidden layer units, the activation conditions of the visible layer units are also independent. The significance of training an RBM is adjusting the parameters of the model to fit the given input data and ultimately making the probability distribution of the visible units consistent with the input data. In an RBM, the activation probabilities of hidden unit  $h_j$  and visible unit  $v_i$  can be written as follows:

$$\begin{aligned} P(h_j|v) &= \sigma(b_j + \sum_i W_{i,j}x_i) \\ P(v_i|h) &= \sigma(c_i + \sum_j W_{i,j}h_j) \end{aligned} \quad (3)$$

where  $x$  is the input data of RBM;  $W$  is the RBM weight matrix;  $b$  is the bias vector for the visible units;  $c$  is the bias vector for the hidden units; and  $\sigma$  is the activation function.

In contrast with NN, DBN uses a layer-by-layer training method to update the network parameters. The first RBM is trained, and then the output of the previous RBM is used as the input of the next RBM. After the training of the RBMs, a fully connected layer is set to obtain the final output. In this study, the number of RBMs was 1 and the number of hidden units in the RBM was set as 112 (see discussion).

For CNN, the neurons between two adjacent layers are partly connected (see Figure 2c), which is different from NN. A CNN consists of five parts, including an input layer, convolutional layer, pooling layer, fully connected layer, and output layer. Convolutional layer is used to extract local features, and the pooling layer is used to reduce the number of parameters in the network. However, the pooling layer has little effect on the accuracy of CNN [54]. The fully connected layer is used to integrate the local features extracted by the previous convolution layer. In this study, CNN was designed with one convolutional layer and one fully connected layer. The kernel size was set as  $1 \times 7$ , the number of convolution kernels was set as 96, and the number of neurons in the fully connected layer was set as 96 (see discussion). The output of the convolutional layer  $y^l$  and the output layer  $\hat{y}$  can be written as follows:

$$\begin{aligned} y^l &= \sigma(W^l x^l + b^l) \\ \hat{y} &= W^2(\sigma(W^1 y^l + b^1)) + b^2 \end{aligned} \quad (4)$$

where  $W^l$  is the convolution kernel weight;  $x^l$  is the input of the convolutional layer;  $b^l$  is the bias of the convolutional layer;  $W^1$  and  $W^2$  are the weights of the fully connected layer and output layer, respectively;  $b^1$  and  $b^2$  are the biases of the fully connected layer and output layer, respectively; and  $\sigma$  is the activation function.

### 3.2. Determination of Inputs for the Networks

The MW brightness temperatures measured by satellite sensors can be written as follows [20]:

$$T_p = \tau \varepsilon_p T_s + (1 - \varepsilon_p) \tau T_a^\downarrow + T_a^\uparrow \quad (5)$$

where the subscript  $p$  is the polarization ( $p \in \{v, h\}$ );  $T_p$ ,  $T_s$ ,  $T_a^\uparrow$ , and  $T_a^\downarrow$  are the MW brightness temperature, LST, atmospheric upwelling brightness temperature, and atmospheric downwelling brightness temperature in K, respectively;  $\tau$  is the atmospheric transmissivity; and  $\varepsilon_p$  is the MW land surface emissivity (LSE).

Equation (5) demonstrates that LSE and three atmospheric parameters are important factors in the estimation of the MW LST. However, these four parameters are difficult to obtain directly. Fortunately, the LSE of MW is closely related to soil moisture, snow, and vegetation [23,55]. Therefore, here, we used surface parameters including NDVI, SC, LCTP, and SM to implicitly present the LSE in the subsequent networks. In addition, atmospheric parameters rely on some easily available meteorological variables, such as the TPWV [20]. Thus, we used the TPWV and AT-2m to implicitly parameterize  $T_a^\uparrow$ ,  $T_a^\downarrow$ , and  $\tau$ . The reason for including AT-2m is that it is better related to atmospheric radiances in conditions with high air temperature but low atmospheric humidity [38].

Soil temperatures at different depths are important parameters for solving the problem of TSD [23,28,56,57]. Currently, soil temperatures mainly come from reanalysis and assimilation datasets. For GLDAS, the first layer depth of soil temperature is 0–10 cm, which means that the soil temperature of the first layer may be very close to the LST. The accuracy of the LST retrieval model may greatly depend on the accuracy of the soil temperature, and microwave BTs will have small contributions when the soil temperature is used as the input parameter. Considering that MW BT products are more reliable than soil temperature products from reanalysis or assimilation datasets, we did not take soil temperatures as input parameters. In addition to the aforementioned parameters, DOY is used to characterize the changes in LST at the annual scale. Finally, the input parameter set includes BTs (all frequencies and all polarizations), NDVI, SC, LCTP, SM, AT-2m, TPWV, and DOY.

### 3.3. Extraction of the Samples

The overpass time of AMSR2 is not the same as that of Aqua MODIS. This difference in the overpass time prevents the MYD11C1 LST from representing the LST at the overpass time of AMSR2. The daytime and nighttime LST at the overpass time of AMSR2 and Aqua MODIS is close to the maximum and minimum LST of the day, respectively. Furthermore, we found that over 90% of the overpass time differences between Aqua MODIS and AMSR2 over China are less than 20 minutes. Therefore, it is reasonable to assume that MERRA-2 skin temperature changes linearly with time within the hour that is nearest to the overpass time of AMSR2. First, the MERRA-2 surface skin temperatures were interpolated over time to obtain the skin temperature at the overpass times of MODIS ( $T_{\text{MERRA-2}}(t_{\text{MODIS}})$ ) and AMSR2 ( $T_{\text{MERRA-2}}(t_{\text{AMSR2}})$ ). Then, the difference  $\Delta T(t_{\text{MODIS}})$  between  $T_{\text{MODIS}}(t_{\text{MODIS}})$  and  $T_{\text{MERRA-2}}(t_{\text{MODIS}})$  at the overpass time of MODIS can be obtained by:

$$\Delta T(t_{\text{MODIS}}) = T_{\text{MODIS}}(t_{\text{MODIS}}) - T_{\text{MERRA-2}}(t_{\text{MODIS}}) \quad (6)$$

Assuming that  $\Delta T(t_{\text{AMSR2}})$  is equal to  $\Delta T(t_{\text{MODIS}})$ ,  $T_{\text{MODIS}}(t_{\text{AMSR2}})$  (the LST at the AMSR2 overpass time) can be obtained by:

$$T_{\text{MODIS}}(t_{\text{AMSR2}}) = T_{\text{MERRA-2}}(t_{\text{AMSR2}}) + \Delta T(t_{\text{AMSR2}}) \quad (7)$$

The accuracy of neural networks relies on the validity of the training samples. To ensure the validity of the samples, we used two criteria to filter the valid samples. The first criterion is that the pixel quality is flagged as “good” (LST error < 1 K) and the view zenith angle (VZA) is less than 40° for all MYD11C1 pixels within the corresponding MW pixel. The second criterion is that the STD of the elevation in MW pixel is less than the threshold (i.e., 100 m for AMSR-E with a spatial resolution of 0.25° and 10 m for AMSR2 with a spatial resolution of 0.1°). Compared with AMSR-E, AMSR2 was filtered more strictly because it has a higher resolution (which means more samples) and lower training sample accuracy (affected by synchronization of MODIS and AMSR2).

The extracted samples were categorized into 4 groups, including daytime AMSR-E (Group I), nighttime AMSR-E (Group II), daytime AMSR2 (Group III), and nighttime AMSR2 (Group IV). Each group contains a training set, a validation set, and a test set. For AMSR-E, the training set, validation set, and test set were the samples from 2003, 2004, and 2009, respectively; for AMSR2, the training set, validation set, and test set were the samples from 2013, 2014, and 2016, respectively. Details for the extracted four groups of samples are shown in Table 2.

**Table 2.** Details for the extracted four groups of samples.

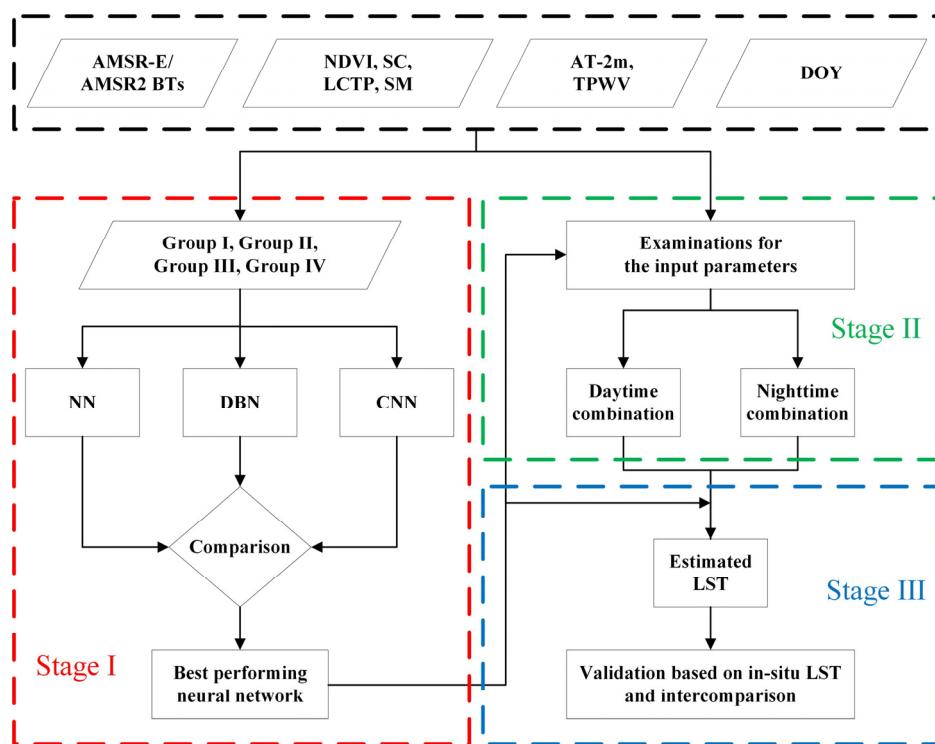
Group	Sensor	Condition	Sample Size		
			Training Set	Validation Set	Test Set
Group I	AMSR-E	Daytime	399,351	450,171	420,655
Group II	AMSR-E	Nighttime	170,144	262,803	315,708
Group III	AMSR2	Daytime	1,057,257	589,205	832,588
Group IV	AMSR2	Nighttime	570,848	400,695	448,753

Note: Sample size represents the number of valid pixels.

### 3.4. Implementation of the Networks

In this study, the three networks performed based on Tensorflow using a GPU with a computation capability of 5.0. Figure 3 shows the flowchart of this study. The process can be divided into three stages. Stage I was to examine the performances of NN, DBN, and CNN and to determine the best performing neural network. The accuracy indices in the comparison included the mean bias deviation (MBD), root-mean-square differences (RMSD), and coefficient of determination ( $R^2$ ) based on the training set, validation set, and test set. For neural networks, more relevant input parameters may

lead to better accuracy. However, some parameters may have small contributions and are difficult to obtain directly. Thus, in stage II, different input parameter combinations were examined to get the best combination for the daytime and nighttime. In stage III, the LST estimated from the MW observations with the determined best network and the best combination were validated based on the in-situ LST. Additionally, the latest MYD21A1 LST product, ESA's GlobTemperature AMSR-E LST, and NASA's LPDR ATs were also investigated for intercomparison purposes. Note that the criteria for filtering valid pixels are only applicable to the extraction of samples with the purpose of obtaining an accurate BT-LST conversion relationship. During the comparison in stage III, CNN LST estimates were from all available samples. Before intercomparison, GT LST, LPDR ATs, and MODIS LST were reprojected to the MW spatial resolutions (namely, 0.25° for AMSR-E and 0.1° for AMSR2).



**Figure 3.** The flow chart of the proposed method.

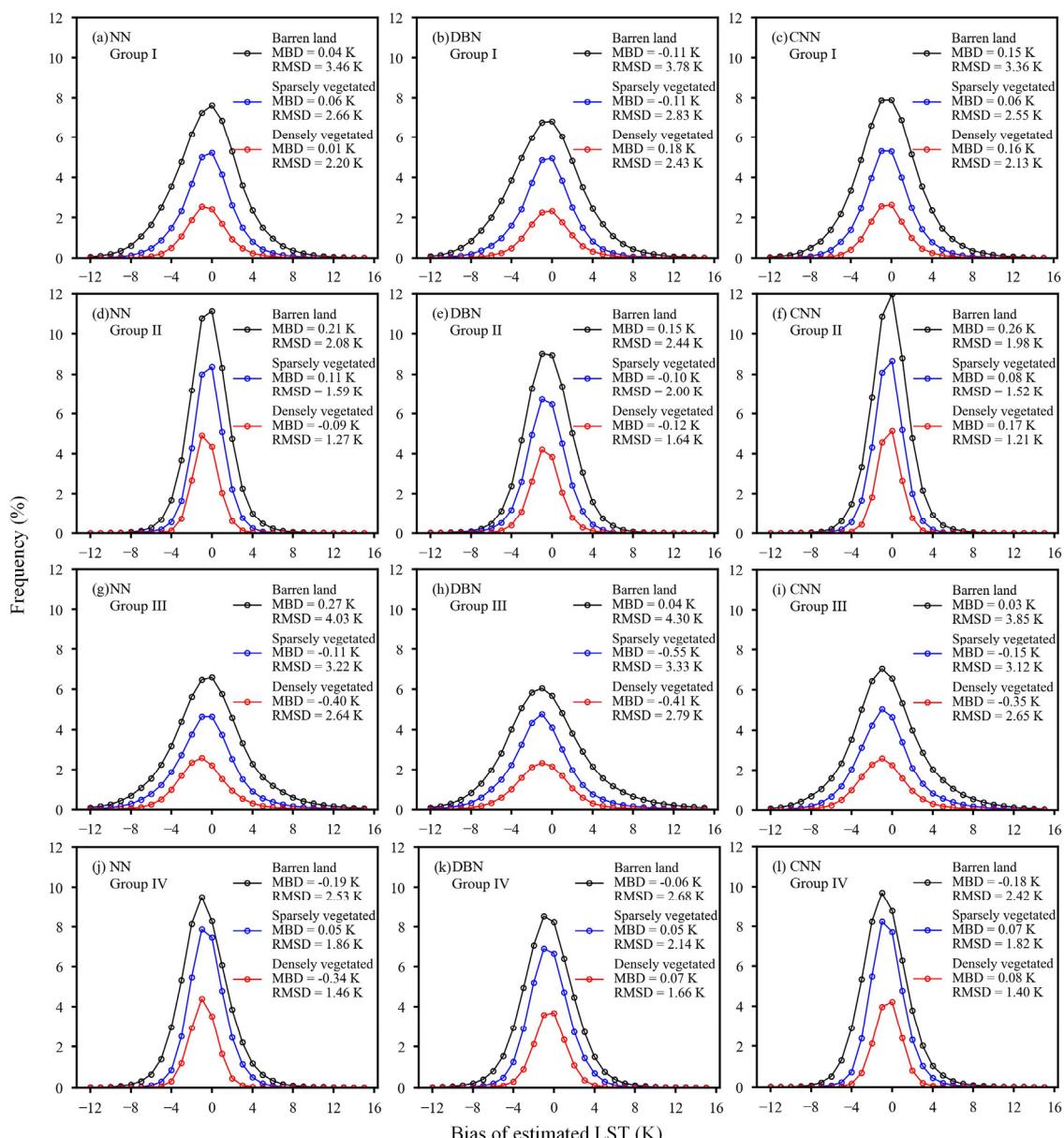
## 4. Results

### 4.1. Performances of the Different Neural Networks

The statistical results of the three networks on the training set, validation set, and test set are shown in Table 3. Table 3 shows that the LSTs estimated by the three networks have no obvious systematic deviation when compared to the MYD11C1 LST. The differences between the validation RMSDs and test RMSDs were lower than 0.30 K (lower than 0.10 K for Groups I, II, and IV), indicating the stability of the three network models. However, CNN had the lowest RMSDs: the RMSDs on the validation set and test set for CNN were lower by ~0.1 K and ~0.3 K than NN and DBN, respectively, for the four groups. In addition, we found that the accuracies of NN and CNN were extremely close. Therefore, a statistical test was used to evaluate the differences between the NN estimates and CNN estimates. Here, equivalence testing was chosen to reduce the risk of Type I inferential error [58,59]. The null hypothesis  $H_0$  was  $|T_1 - T_2| \geq \Delta T$ , and the alternative hypothesis  $H_1$  was  $|T_1 - T_2| < \Delta T$ . Note that the  $H_0$  would be rejected when both one-sided components would be rejected [58]. In this study,  $\Delta T$  was set as 0.1 K and the test results are shown in Table 4. It can be seen that the CNN estimates were different (null hypothesis is accepted) from the NN estimates for both AMSR-E and AMSR2 during the

daytime. In contrast, they were equivalent (null hypothesis is rejected) on the validation and test sets for AMSR-E and the training and validation sets for AMSR2 during the nighttime.

Then, NN, DBN, and CNN were further compared for different NDVI-based land cover types and different seasons. We classified the microwave pixels of the test sets into three types based on NDVI [3]: (i) barren land ( $NDVI < 0.2$ ); (ii) sparsely vegetated ( $0.2 \leq NDVI \leq 0.5$ ); and (iii) densely vegetated ( $NDVI > 0.5$ ). For the convenience of statistics, the spring represents March, April, and May; the summer represents June, July, and August; the autumn represents September, October, and November; and the winter represents December, January, and February. The accuracies of the three networks for different NDVI-based land cover types and different seasons are shown in Figures 4 and 5, respectively. It can be seen that CNN had the lowest RMSDs in almost all conditions and the RMSDs of NN and DBN were higher than those of CNN by 0.1 K–0.5 K.



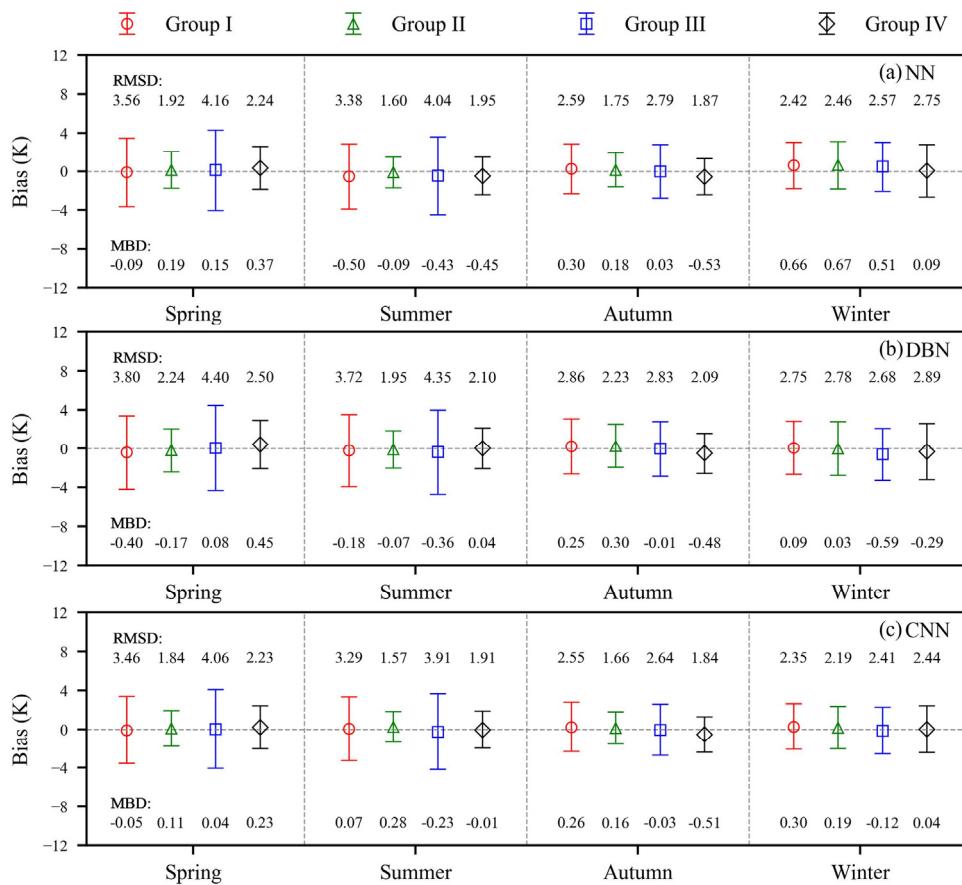
**Figure 4.** Histograms of the differences between the LSTs estimated from neural networks and the MYD11C1 LSTs on the test sets for NN (a,d,g,j), DBN (b,e,h,k), and CNN (c,f,i,l).

**Table 3.** The performances of NN, DBN, and CNN on the training sets, validation sets, and test sets.

Group	Set	NN			DBN			CNN		
		MBD (K)	RMSD (K)	$R^2$	MBD (K)	RMSD (K)	$R^2$	MBD (K)	RMSD (K)	$R^2$
Group I (Daytime AMSR-E)	Training	−0.04	2.77	0.97	−0.06	3.28	0.96	0.04	2.50	0.98
	Validation	−0.03	2.99	0.97	−0.03	3.34	0.96	0.08	2.88	0.97
	Test	0.04	3.11	0.97	−0.07	3.38	0.96	0.13	3.00	0.97
Group II (Nighttime AMSR-E)	Training	−0.09	1.56	0.98	−0.05	2.03	0.96	−0.01	1.32	0.98
	Validation	0.08	1.81	0.97	0.01	2.10	0.96	0.12	1.66	0.98
	Test	0.13	1.83	0.97	0.03	2.20	0.96	0.19	1.74	0.97
Group III (Daytime AMSR2)	Training	0.12	3.12	0.97	−0.03	3.46	0.96	0.01	2.90	0.97
	Validation	−0.03	3.32	0.96	−0.14	3.55	0.95	−0.12	3.22	0.96
	Test	0.05	3.62	0.96	−0.21	3.83	0.95	−0.08	3.48	0.96
Group IV (Nighttime AMSR2)	Training	−0.06	1.85	0.98	0.01	2.12	0.98	−0.07	1.70	0.98
	Validation	0.23	2.12	0.97	0.37	2.34	0.96	0.22	2.02	0.97
	Test	−0.13	2.19	0.97	−0.01	2.38	0.96	−0.06	2.10	0.97

**Table 4.** The probability values (*p*-values) of the equivalence testing for the NN LST and the CNN LST on the training sets, validation sets, and test sets. Note that the  $H_0$  would be rejected when both one-sided components would be rejected. The *p*-values of the sets on which the CNN estimates and the NN estimates are equivalent are highlighted by underline.

Set	Group I		Group II		Group III		Group IV	
	$T_1 - T_2 \leq -0.1 \text{ K}$	$T_1 - T_2 \geq 0.1 \text{ K}$	$T_1 - T_2 \leq -0.1 \text{ K}$	$T_1 - T_2 \geq 0.1 \text{ K}$	$T_1 - T_2 \leq -0.1 \text{ K}$	$T_1 - T_2 \geq 0.1 \text{ K}$	$T_1 - T_2 \leq -0.1 \text{ K}$	$T_1 - T_2 \geq 0.1 \text{ K}$
Training set	0.35	<0.01	0.31	<0.01	<0.01	0.86	<u>&lt;0.01</u>	<u>&lt;0.01</u>
Validation set	0.58	<0.01	<u>0.03</u>	<u>&lt;0.01</u>	<0.01	0.41	<u>&lt;0.01</u>	<u>&lt;0.01</u>
Test set	0.35	<0.01	<u>0.04</u>	<u>&lt;0.01</u>	<0.01	0.90	0.16	<0.01



**Figure 5.** Seasonal statistics of the difference between the LSTs estimated from neural networks and the MYD11C1 LSTs on the test sets for NN (a), DBN (b), and CNN (c). The bars are centered in the centroid of the symbol. The centroid of the symbol represents the mean value of the difference. The bar represents the STD of the difference. In addition, the RMSDs of the three networks compared to MYD11C1 LST are also annotated in the figure.

From the previous analysis, it is clear that CNN outperforms NN and DBN. The training times of NN, DBN, and CNN were related to the number of training samples and were 6, 8, and 12 min for Group I, respectively. Although CNN had the highest time cost among these three networks, it was still acceptable. Therefore, only CNN was employed in further analysis, including the analysis of input parameters, validation based on in-situ LST, and intercomparison with other products.

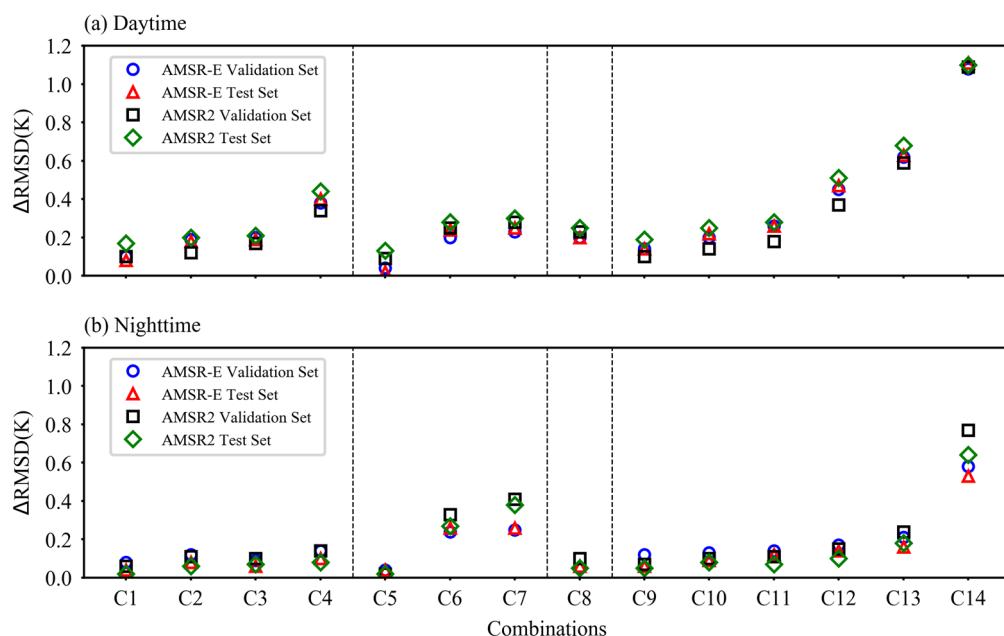
#### 4.2. Determination of the Best Input Parameter Combinations

Considering that BTs of all frequencies are acquired at the same time, we only examined the following input parameters: NDVI, SC, LCTP, SM, TPWV, AT-2m, and DOY. Due to space limitations, only some representative input parameter combinations and their total numbers of parameters are listed in Table 5. In these combinations, combination C0 contains all parameters in the input parameter set; C1–C4 were used for examining the surface parameters; C5–C7 were used for examining AT-2m and TPWV; C8 is for DOY; and C9–C14 were designed based on the results from C1–C8 to determine the best combinations for the daytime and nighttime conditions. The RMSD differences ( $\Delta$ RMSDs) between combination C1–C14 and C0 for the validation sets and test sets are shown in Figure 6. The best input parameter combinations were selected based on (i)  $\Delta$ RMSD less than 0.3 K; (ii) fewer input parameters.

**Table 5.** Some of the input parameter combinations and the corresponding number of parameters. ✓ represents the parameter contained in the input parameters, and × is the opposite. The numbers outside (within) the brackets represent the total number of input parameters for AMSR-E (AMSR2).

Combination	BTs	Surface Parameters			Atmospheric Related Parameters		DOY	Number of Parameters
		NDVI	LCTP	SM* and SC	AT-2m	TPWV		
C0	✓	✓	✓	✓	✓	✓	✓	38 (40)
C1	✓	✓	✓	×	✓	✓	✓	33 (35)
C2	✓	✓	×	×	✓	✓	✓	16 (18)
C3	✓	×	✓	×	✓	✓	✓	32 (34)
C4	✓	×	×	×	✓	✓	✓	15 (17)
C5	✓	✓	✓	✓	✓	×	✓	37 (39)
C6	✓	✓	✓	✓	×	✓	✓	37 (39)
C7	✓	✓	✓	✓	×	×	✓	36 (38)
C8	✓	✓	✓	✓	✓	✓	×	37 (39)
C9	✓	✓	✓	×	✓	×	✓	32 (34)
C10	✓	✓	×	×	✓	×	✓	15 (17)
C11	✓	×	✓	×	✓	×	✓	31 (33)
C12	✓	×	×	×	✓	×	✓	14 (16)
C13	✓	×	×	×	✓	×	×	13 (15)
C14	✓	×	×	×	×	×	×	12 (14)

Note: SM\* is soil moisture values in four layers. LCTP contains the percent of the 17 IGBP classes.



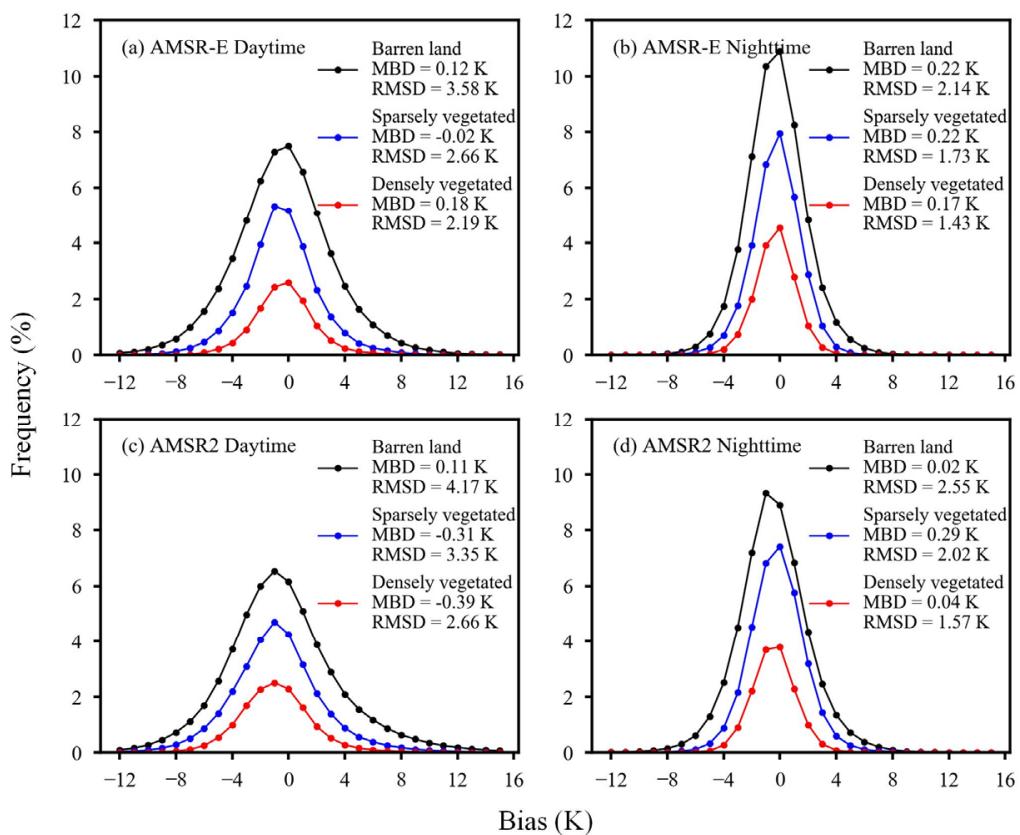
**Figure 6.**  $\Delta\text{RMSD}$ s between combination C1–C14 and C0 for the validation sets and test sets during the daytime (a) and nighttime (b).

For the daytime condition, SM and SC have small contributions to the improvement of model accuracy in surface parameters. When the input parameters did not contain SM and SC (i.e., combination C1), the  $\Delta\text{RMSD}$ s were below 0.15 K for both the validation sets and the test sets. If NDVI and LCTP were not removed at the same time, the increase in  $\Delta\text{RMSD}$  was small. For example, on the basis of C1, only LCTP (NDVI) was removed for C2 (C3), and the increases in  $\Delta\text{RMSD}$  were below 0.10 K; however, the increases in  $\Delta\text{RMSD}$  were above 0.25 K when NDVI and LCTP were removed at the same time (C4). AT-2m had a higher impact than TPWV. When AT-2m was not used as an input parameter (i.e., combination C6), the  $\Delta\text{RMSD}$  was higher than 0.20 K. In addition, DOY (i.e., combination C8) can help improve the accuracies of the daytime models. In these combinations,  $\Delta\text{RMSEs}$  of C10 and C11 were both below 0.30 K. Considering that NDVI is a quantitative parameter and C10 has

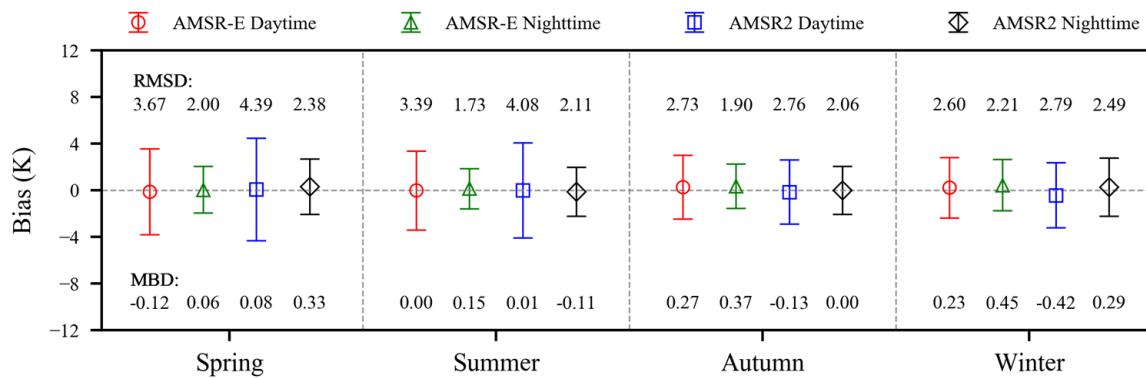
fewer input parameters, we finally selected C10 as the best combination of input parameters for the daytime models.

For the nighttime conditions, SM, SC, and TPWV have small contributions, and AT-2m has a large contribution. This finding is similar to that under daytime conditions. However, the nighttime NDVI, LCTP, and DOY values have fewer contributions than those in the daytime. For example, the  $\Delta$ RMSDs of C13 (only BTs and AT-2m) were all below 0.30 K. In these combinations, only C14 had fewer parameters than C13. However, it can be observed that C14 had the RMSDs of 0.5–0.8 K. Therefore, C13 was finally selected as the best combination of input parameters for the nighttime models.

The statistical results of the best input parameter combinations for different NDVI-based land cover types and seasons are shown in Figures 7 and 8. For the three NDVI-based land cover types, the highest accuracies occurred in the densely vegetated pixels and lowest accuracies occurred in the barren land pixels. During the daytime, the RMSDs of the best combinations were 2.19 K, 2.66 K, and 3.58 K for densely vegetated pixels, sparsely vegetated pixels, and barren land pixels for AMSR-E, respectively. The corresponding RMSDs were 1.43 K, 1.73 K, and 2.14 K during the nighttime. For AMSR2, there were higher RMSDs on all land cover types compared to AMSR-E (~0.6 K and ~0.3 K for the daytime and nighttime conditions, respectively). The lowest accuracies for the barren land may be concluded by the difficulty in determining the LSE of barren land. Figure 8 shows that the RMSDs of spring and summer are higher than those of autumn and winter for the daytime conditions. Spring and summer are the seasons of vegetation growth, increasing the heterogeneity within the pixels, and, thus, the upscaling method of spatial averaging introduces more uncertainty than other seasons. In contrast, the RMSDs of all seasons are below 2.20 K and 2.80 K for AMSR-E and AMSR2 at night, respectively.



**Figure 7.** Histograms of the difference between the LSTs estimated from CNN with the best combinations of input parameters and the MYD11C1 LSTs on the test sets for AMSR-E (**a,b**) and AMSR2 (**c,d**).



**Figure 8.** Similar to Figure 5, but only showing the difference between the LSTs estimated from CNN with the best combinations of input parameters and the MYD11C1 LSTs on the test sets.

#### 4.3. Validation Based on In-Situ LST

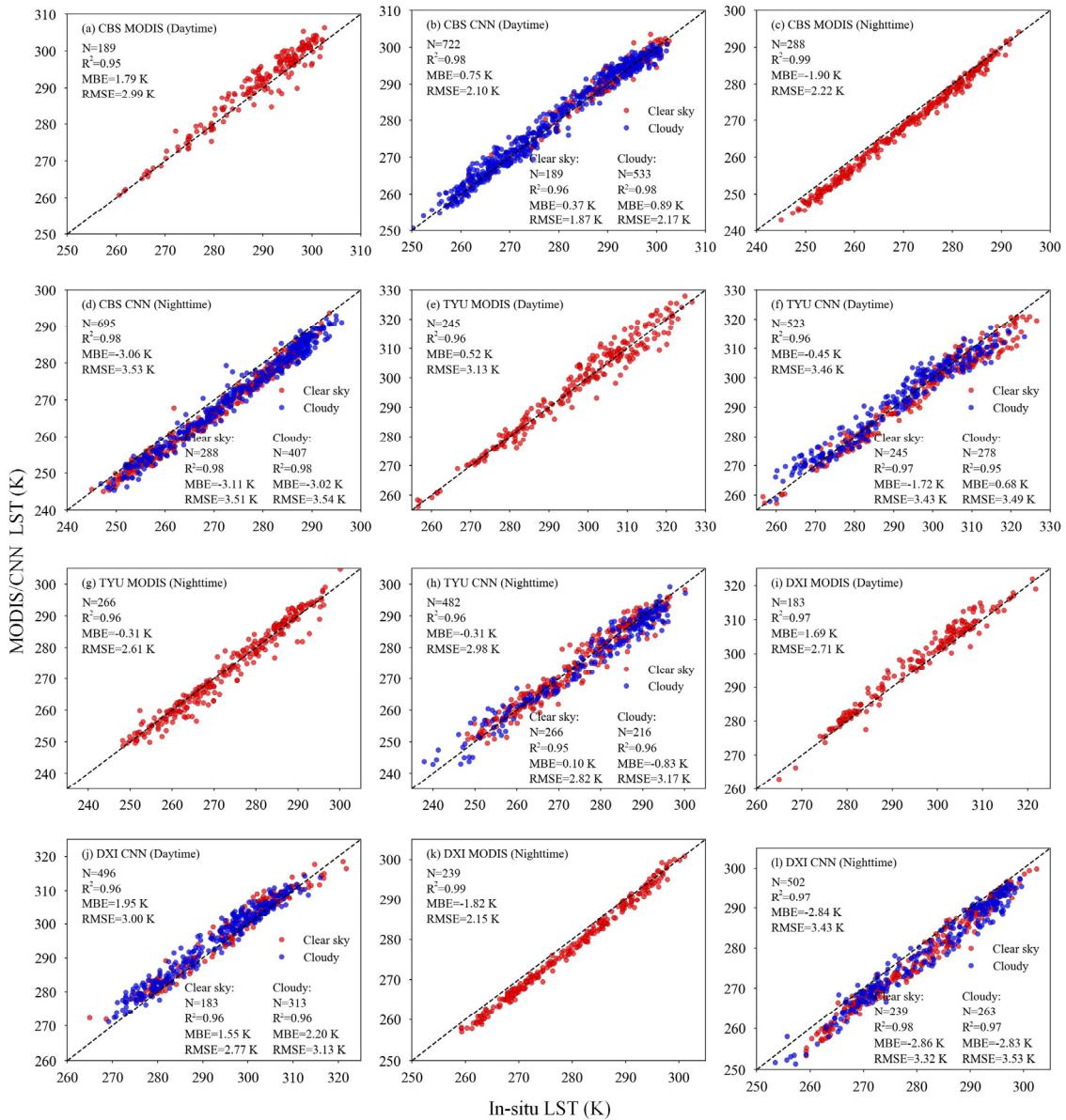
Based on the best combinations of input parameters determined in Section 4.2, the CNN LST was validated against the in-situ LST. In addition, the MODIS LST derived from the MYD21A1 products was also compared. Since the obtained in-situ measurements were from different years, the validations of the retrieved LST from AMSR-E were based on CBS, TYU, and DXI; the validations of the retrieved LST from the AMSR2 observations were based on SDQ and HMO. The results are shown in Figures 9 and 10.

For CBS, the MODIS LST had a slight overestimation and underestimation in the daytime and nighttime values, respectively. This could be related to the built-up surfaces around the station. Although only ~2% of MW pixels were covered by built-up surfaces, it can be seen from Google Earth that the built-up surfaces were concentrated near the station. The CNN LST is close to the in-situ LST during the daytime and had an evident underestimation during the nighttime: the MBE values were 0.75 K and −3.06 K during the daytime and nighttime, respectively; the corresponding RMSE values were 2.10 K and 3.53 K. The underestimations of the nighttime values are induced by the higher elevation (the highest elevation in MW pixel is 1210 m) at the south of the MW pixel.

For TYU, the MODIS LSTs and the CNN LSTs had no obvious systemic biases in either the daytime or nighttime. The absolute values of the MBEs were all below 0.50 K. The RMSEs of CNN LST were below 3.5 K and 3 K for the daytime and nighttime, respectively. Although the AMSR-E pixel containing the TYU station was not a pure pixel, the main land cover types, i.e., croplands and grasslands, had very similar thermal properties and LSTs. Thus, it is understandable that CNN LST was close to the in-situ LST and only a slight RMSD difference existed between the CNN LST and the MODIS LST.

For DXI, an overestimation of the daytime values and an underestimation of the nighttime values were observed for both the MODIS LST and the CNN LST. This may be induced by the built-up surfaces. Table 1 shows that over 30% of the AMSR-E pixel was covered by built-up surfaces. The MBE values were 1.95 K and −2.84 K for the CNN LST during the daytime and nighttime, respectively. The corresponding RMSE values were 3 K and 3.43 K.

For SDQ, the CNN LST had an RMSE of 4.72 K (MBE is 3.64 K) for the daytime conditions. Over 30% of the AMSR2 pixels were covered by barren, resulting in an overestimation of above 3 K. For the nighttime values, the MODIS LST had good performance: the MBE and RMSE were −0.56 K and 1.32 K, respectively. In contrast, the corresponding values were 1.84 K and 4.08 K for the CNN LST. From Figure 10d, evident overestimation can be observed when LST is less than 270 K. Based on the discriminant function algorithm (DFA) proposed by Wang et al. [60], we found that this overestimation is due to the frozen soil. Frozen soil has higher emissivity than unfrozen soil. Thus, higher BTs can be observed when soil freezing occurs [61], resulting in an overestimation of the CNN LST.

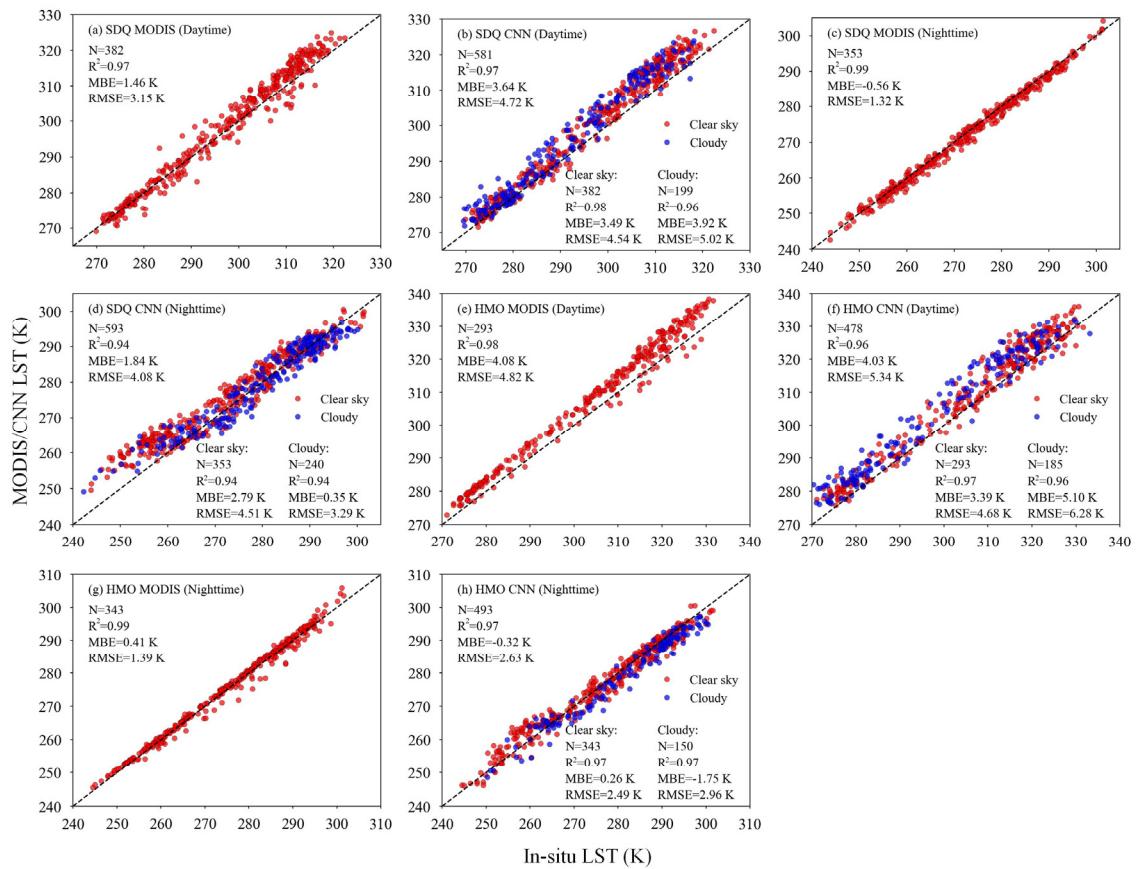


**Figure 9.** Scatter plots between (1) the MODIS LST and (2) the CNN LST estimates from AMSR-E data and the in-situ LST at CBS (**a–d**), TYU (**e–h**), and DXI (**i–l**).

For HMO, both the MODIS LST and the CNN LST have high systematic overestimation (above 4 K) during the daytime and have better accuracy at nighttime (RMSE was 1.39 K for the MODIS LST and 2.63 K for the CNN LST). The significant overestimation of the CNN LST during the daytime may be induced by the overestimation of the MODIS LST, which was the basis for the training of the CNN models. In addition, an overestimation also occurred when the LST was less than 270 K at night, which is similar to that at SDQ.

The validation results of the CNN LST under clear-sky conditions and cloudy conditions are also shown in Figures 9 and 10. It can be seen that the RMSE differences of the clear-sky conditions and the cloudy conditions were lower than 0.5 K in most cases. Exceptions occurred during the daytime of HMO and the nighttime of SDQ. The reason was the uneven distribution of the clear-sky and the cloudy samples. However, the CNN LSTs under cloudy conditions do not show significant outliers compared with the CNN LSTs under clear-sky conditions in all cases, indicating that CNN models established with clear-sky samples have good ability when extended to cloudy conditions. Overall, the CNN LST agrees well with the in-situ LST, with  $R^2$  values ranging from 0.94 to 0.98. Although the RMSE of CNN

LST was above 4 K in a few cases, the RMSE was mainly due to the different land cover types and elevation changes within the MW pixels.



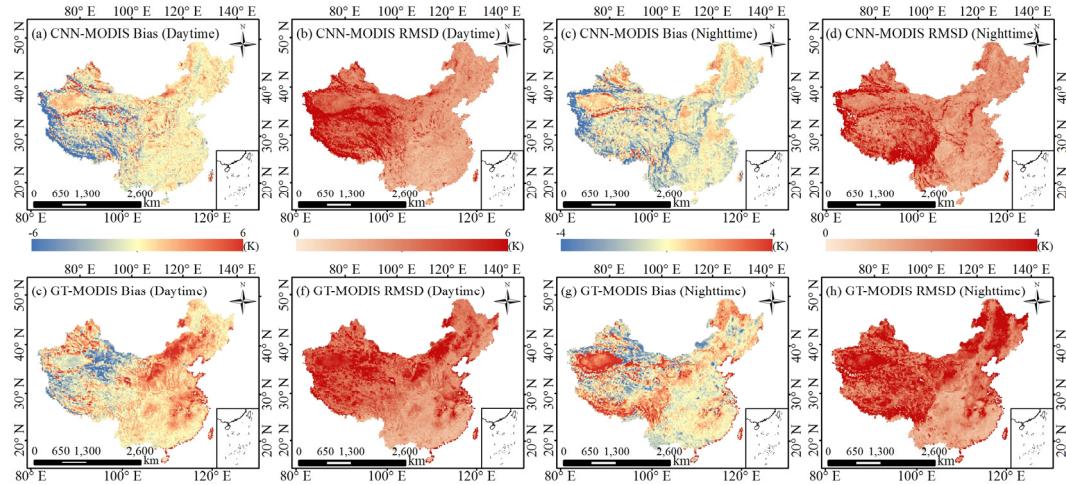
**Figure 10.** Scatter plots between (1) the MODIS LST and (2) the CNN LST estimates from AMSR2 data and the in-situ LST at SDQ (**a–d**) and HMO (**e–h**).

#### 4.4. Intercomparison with GlobTemperature AMSR-E LST and LPDR Air Temperature

After the validations with the in-situ LST, the CNN LST was intercompared with the AMSR-E LST provided by the GlobTemperature (hereafter termed GT LST) [26,27]. Since GT LST was only available from 2008 to 2010, the CNN LST, GT LST, and MODIS LST were intercompared based on the data of 2009 for AMSR-E (Figure 11). The pixel percentages of the RMSD difference between CNN LST and GT LST in different ranges over China are shown in Table 6. The statistical results show that ~50% of the pixels of the CNN LST have smaller RMSDs and only ~20% of the pixels of the CNN LST have larger RMSDs than GT LST. It can be observed that the pixels with significant underestimation for CNN LST are concentrated on the boundaries of the Tibetan Plateau. This phenomenon could be partly due to snow. Based on MYD10C1 product, we found that these pixels are frequently covered with snow, and snow pixels tend to have larger bias [27]. By analyzing the BTs of these pixels, we found an evident underestimation for the CNN LST, especially at high frequency (i.e., 36.5 and 89.0 GHz), resulting in the underestimation of the CNN LST for snow pixels. This corresponds to the strong volume scattering of snow [27,62].

In addition, the overpass times of AMSR-E and AMSR2 were both ~13:30 and 01:30 local solar time, which are close to the time when daily maximum and minimum AT appear [12]. Fily et al. [16] showed LST was close to AT at surface level over dense vegetation. Thus, a comparison between the CNN LST and the LPDR ATs was also performed. Figure 12 shows maps of RMSDs and  $R^2$  between the CNN LST and the LPDR ATs. It can be observed that the CNN LST agrees well with the LPDR ATs for most pixels (over 80% of pixel  $R^2$  values are above 0.8) and there are small RMSDs for pixels with

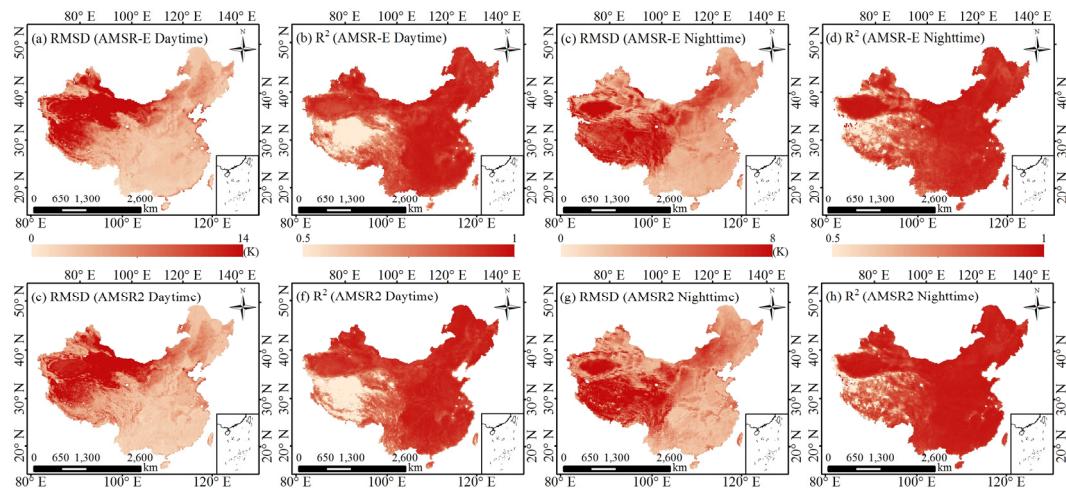
high vegetation coverage (e.g., the average RMSDs of South China and Northeast China are less than 3 K). This phenomenon further confirms that the CNN LST and AT have similar annual variations and are close over dense vegetation. Therefore, it can be concluded that CNN LST estimates have high accuracy.



**Figure 11.** Spatial distribution of the bias (a,c,e,g) of the difference and RMSD (b,d,f,h) between (1) the CNN LST and (2) the GT LST and the MODIS LST during the daytime and nighttime for AMSR-E. The statistics are based on the data of 2009.

**Table 6.** The pixel percentages of the RMSD difference between the CNN LST and the GT LST in different ranges: (i)  $\Delta\text{RMSD} < -0.5 \text{ K}$  represents CNN LST has better accuracy; (ii)  $-0.5 \text{ K} \leq \Delta\text{RMSD} \leq 0.5 \text{ K}$  represents CNN LST has similar accuracy to GT LST; (iii)  $\Delta\text{RMSD} > 0.5 \text{ K}$  represents GT LST has better accuracy.

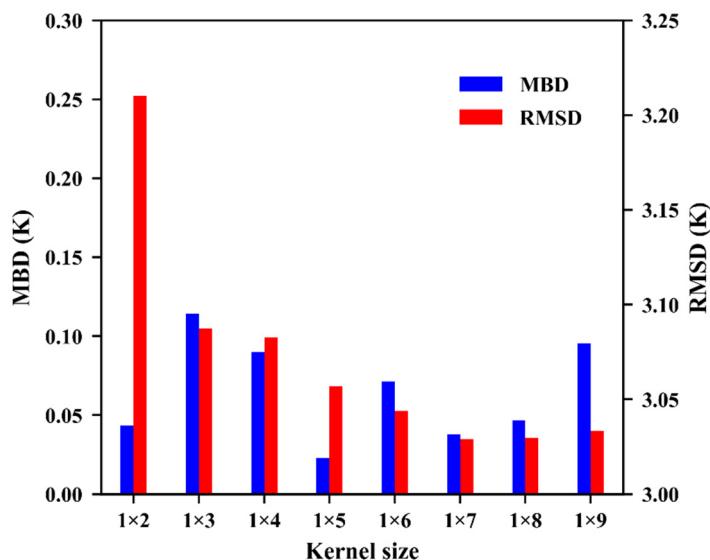
	$\Delta\text{RMSD} < -0.5 \text{ K}$	$-0.5 \text{ K} \leq \Delta\text{RMSD} \leq 0.5 \text{ K}$	$\Delta\text{RMSD} > 0.5 \text{ K}$
Daytime	47.7%	29.4%	22.9%
Nighttime	52.3%	29.1%	18.6%



**Figure 12.** Spatial distribution of the RMSD (a,c,e,g) and  $R^2$  (b,d,f,h) between the CNN LST and the LPDR ATs for AMSR-E (the top panels) and AMSR2 (the bottom panels). The statistics are based on the data of 2009 and 2016 for AMSR-E and AMSR2, respectively.

## 5. Discussion

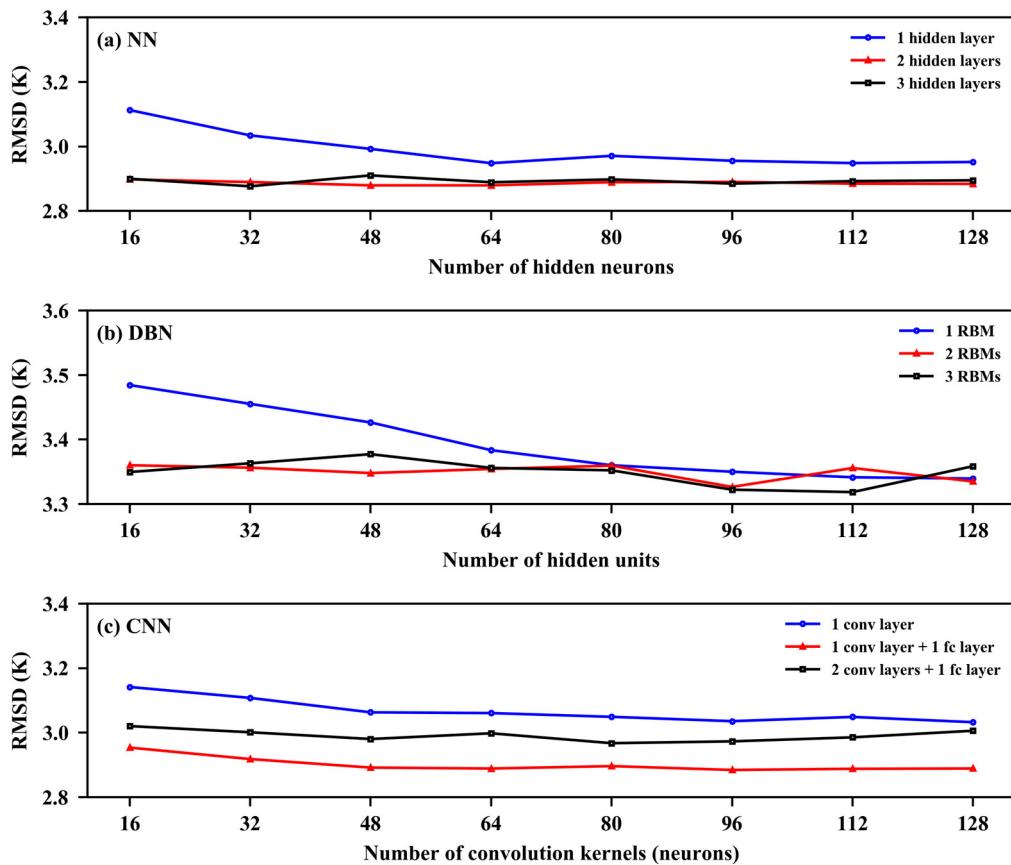
Considering that the comparison of the three networks should be performed when they are optimal respectively, we examined the structural parameters of NN, DBN, and CNN. The structural parameters were examined with the samples from Group I. The kernel size of the CNN was first examined, and the examination results based on the validation set are shown in Figure 13. As the size of the convolution kernel increased, the RMSD first decreased significantly; when the kernel size was greater than  $1 \times 7$ , the RMSD remained relatively stable. The MBDs did not exhibit a clear pattern but were lower than 0.15 K. Therefore,  $1 \times 7$  was designated as the eventual kernel size of the CNN model.



**Figure 13.** The mean bias deviation (MBD) and RMSD of different convolutional kernel sizes for CNN based on the validation set of Group I.

The examination results for layers are shown in Figure 14. The DBN models with two or three RBMs had similar accuracies to one RBM with more than 112 hidden units. For NN and CNN, the accuracies of the two-layer models had an increase of more than 0.10 K compared to the one-layer models. However, the accuracies of three-layer models were almost the same as those of two-layer models and were sometimes even lower. This situation shows that two hidden layers for NN, one RBM for DBN, and one convolutional layer and one fully connected layer for CNN are sufficient to describe the nonlinear relationship between the LST and the input parameters. For NN, the numbers of hidden neurons were designated as 64 and 48 for the first hidden layer and the second hidden layer, respectively. For DBN, the number of hidden units was designated as 112 for the RBM. For CNN, the number of convolution kernels was designated as 112, and the number of neurons in the fully connected layer was designated as 64.

In general, the accuracies for AMSR2 were lower than those for AMSR-E. The possible reasons are: (i) AMSR2 observations had higher spatial resolution than AMSR-E, resulting in greater uncertainty for AMSR2 samples during the process of spatial matching with MERRA-2 product and GLDAS product, and (ii) there still existed deviation between the synchronized MYD11C1 LST and “true” LST at the overpass time of AMSR2. The nighttime models had better accuracies than the daytime models, and the nighttime RMSDs were more than 1 K lower than those of the daytime RMSDs. The reason is that the daytime data are noisier than the nighttime data.



**Figure 14.** The RMSDs of NN (a), DBN (b), and CNN (c) with different layers based on the validation set. The “conv layer” and “fc layer” mean the convolutional layer and fully connected layer, respectively.

The best combinations of input parameters are BTs, NDVI, AT-2m, and DOY for the daytime conditions and BTs and AT-2m for the nighttime conditions. The small contributions of SM, SC, and TPWV can be explained by the fact that these parameters can be expressed by the MW BTs [51,62]. NDVI can implicitly characterize the proportion of different land cover types within a pixel, and it is also the reason why NDVI and LCTP have similar contributions. Thus, taking NDVI as an input can improve the accuracies of the daytime models. In contrast, the different land cover types within the pixels have similar thermal properties and LSTs at nighttime. Hence, it is understandable that NDVI has a small contribution to the nighttime models. As for DOY, it is interesting that we found DOY can improve the accuracies of the daytime models. However, DOY is irrelevant to Equation (5). The relationship between the MODIS LST ( $T_{MODIS}$ ) and the CNN LST ( $f_i(x_i)$ ) can be written as:

$$T_{MODIS} = f_i(x_i) + \delta_i \quad (8)$$

where  $x_i$  is the input parameters of combination  $C_i$ ;  $\delta_i$  is the residual corresponding to  $C_i$  and denotes the tiny fraction that cannot be interpreted by input parameters; and  $f_i$  is the conversion relationship established by CNN based on  $C_i$ . The purpose of inputting multiple parameters is to decompose the residuals as much as possible. Generally, the component temperature differences between barren land and vegetation within the pixels are larger in summer compared to those in winter. Therefore, it is understandable that the component temperature differences can be expressed as a function with DOY as the independent variable. In other words, DOY helps to characterize the component temperature differences within the pixels. This information has potential value for the LST retrieval, and, thus, can help decompose the residuals. Another possible reason is that the LST of a whole year can be divided into three temporal components (i.e., the annual temperature component-ATC, the diurnal temperature component-DTC, and the weather-change temperature

component-WTC) [3,56], and ATC is a function of DOY. In contrast, the small contributions of DOY to the nighttime models can be concluded to the smaller residuals of the models themselves.

It is noticeable that this study was not the first time that CNN was applied to estimate the MW LST. As described in Section 1, Tan et al. [34] constructed the CNN model. However, only BTs were used as inputs (i.e., C14 in this study). Figure 6 shows that the accuracy of C14 is lower by 0.8 K during the daytime and 0.3 K during the nighttime than the best combination. Thus, it can be concluded that adding some auxiliary data (e.g., AT-2m) on the basis of BTs can significantly improve the accuracies of the retrieval models.

In this study, the ground station FOVs were greatly different from the spatial resolutions of the MW pixels. To quantify the representativeness errors introduced by the scale mismatching between the ground stations and the MW pixels, we calculated the errors based on the biases and STDs from Figure 1 and the validation results of the MODIS LST. Similar to Huang et al. [63], by assuming that MODIS LST is the “true LST” on the 1-km scales, the representativeness error for each station is given by:

$$\begin{aligned} \text{MBE}_{\text{GtoMW}} &= \text{MBE}_{\text{Gto1KM}} + \text{MBE}_{\text{1KMtoMW}} \\ \text{STD}_{\text{GtoMW}}^2 &= \text{STD}_{\text{Gto1KM}}^2 + \text{STD}_{\text{1KMtoMW}}^2 \end{aligned} \quad (9)$$

where  $\text{MBE}_{\text{GtoMW}}$ ,  $\text{MBE}_{\text{Gto1KM}}$ , and  $\text{MBE}_{\text{1KMtoMW}}$  are systematic errors introduced by the scale mismatching between the ground station and the MW pixel, between the ground station and the 1-km MODIS pixel, and between the 1-km MODIS pixel and the MW pixel, respectively;  $\text{STD}_{\text{GtoMW}}$ ,  $\text{STD}_{\text{Gto1KM}}$ , and  $\text{STD}_{\text{1KMtoMW}}$  are the corresponding STD values. The statistical results of the representativeness errors of the five stations and the corresponding validation results of the CNN LST are shown in Table 7. It can be seen that in most cases, the  $\text{MBE}_{\text{CNN}}$  and  $\text{STD}_{\text{CNN}}$  were close to the  $\text{MBE}_{\text{GtoMW}}$  and  $\text{STD}_{\text{GtoMW}}$  (MBE differences and STD differences were both lower than 1 K), respectively. Exceptions occurred in the nighttime of SDQ and HMO, and this phenomenon was caused by frozen soil (see Section 4.3). Therefore, Table 7 further confirms that the main reason for the errors of the CNN LST against in-situ LST is the scale mismatching between the ground stations and the MW pixels.

**Table 7.** The representativeness errors of the five stations and the corresponding validation results of the CNN LST based on the in-situ LST. All MBE and STD values are in K.

Station	Daytime				Nighttime			
	$\text{MBE}_{\text{GtoMW}}$	$\text{MBE}_{\text{CNN}}$	$\text{STD}_{\text{GtoMW}}$	$\text{STD}_{\text{CNN}}$	$\text{MBE}_{\text{GtoMW}}$	$\text{MBE}_{\text{CNN}}$	$\text{STD}_{\text{GtoMW}}$	$\text{STD}_{\text{CNN}}$
CBS	0.77	0.75	2.72	1.96	-3.29	-3.06	1.43	1.76
TYU	0.32	-0.45	3.27	3.43	-0.09	-0.31	2.64	2.96
DXI	2.70	1.95	2.47	2.28	-2.06	-2.84	1.27	1.91
SDQ	4.19	3.64	3.45	2.99	-0.61	1.84	1.55	3.64
HMO	3.67	4.03	3.05	3.50	0.14	-0.32	1.47	2.62

For CNN, we only tried to convolve on the input vector of a single pixel. Another method to build the CNN model is to use the parameter images (e.g., BT images and NDVI images) of the entire study area as inputs. In this situation, the spatial relationships between the adjacent pixels may be helpful in improving model accuracy. Nevertheless, one should keep in mind that it is difficult to apply this method: MW images have coarse spatial resolutions, resulting in weak spatial connections between the adjacent pixels (which means a small contribution to the LSTs of the surrounding pixels). In addition, accurate and cloud-free LST images for large study areas are usually not available, which limits the extraction of LST for the training samples. Further studies to incorporate the spatial relationships in the estimation of LST from MW observations are needed.

A well-recognized feature of neural networks is the ‘black box’ problem, which means that it is difficult to understand the physical mechanisms for obtaining the output based on the inputs. The opacity of neural network makes it hard to further improve the accuracy of the neural network.

For the estimation of LST from MW observations, three approaches may contribute to the improvement of the accuracy. The first approach is to increase the sample size. Larger sample size would be helpful to reduce overfitting and, thus, improve the accuracy of the output. In this study, the upscaling method for the MODIS LST was spatial averaging, which may introduce uncertainty to the LST of the training samples. Therefore, the second approach is to improve the accuracy of the training samples. The third approach is to incorporate the physical models into the neural networks by using the physical models as the boundaries of the networks, to provide initial guesses, and/or to construct new input features for the networks. For example, the thermal sampling depth correction model developed by Zhou et al. [57] may be incorporated to help neural networks better estimate the LST over barren land.

Finally, there still exist some issues in this study. The first one is the training of the neural network was performed with the MODIS LST product. Although the MODIS LST has high accuracy, the uncertainty of MODIS LST will inevitably increase when reprojected to match with the MW pixels. Therefore, an accurate upscaling method related to land cover types and elevation distribution within the MW pixels is critical for reducing the uncertainty in the spatial matching process. The second one is the scale mismatch between the ground stations and the MW pixels. This issue is critical for the validation of low-resolution LST. Thus, it is essential to find an evaluation method of spatial representativeness of the ground station to realize the conversion of in-situ LST from station FOV to coarser resolution.

## 6. Conclusions

This study examined the performances of NN (i.e., the traditional neural network) and two deep learning methods (i.e., DBN and CNN) in the estimation of LST from satellite MW observations. The input parameter set for these networks included BTs, soil moistures, NDVI, snow cover, land cover type percent, air temperature at 2 m above the ground surface, total precipitable water vapor, and DOY. Based on the training, validation, and test sets derived from the MODIS LST, microwave BTs, and the aforementioned input parameters, the results demonstrate that CNN outperformed NN and DBN. The LSTs estimated by CNN from the AMSR-E and AMSR2 data were closer to MODIS LST: the RMSD values were 3 K during the daytime and 1.74 K at night for AMSR-E; the corresponding RMSD values were 3.48 K and 2.10 K for AMSR2. In contrast, the RMSDs of NN and DBN were higher by 0.1 K and 0.4 K, respectively. Additionally, CNN was more prominent than NN and DBN for different land cover types and seasons. Therefore, it should be concluded that CNN performs better than the NN and DBN in the estimation of LST from satellite MW observations.

More details for the impacts of the input parameters on the performances of CNN were obtained. Among the surface parameters used to implicitly parameterize the LSE, the NDVI and land cover type percent have greater impacts than the other parameters during the daytime, and NDVI and land cover type percent cannot be removed at the same time; nevertheless, their impacts decrease at nighttime. For the two parameters used to implicitly quantify the atmospheric effects, the air temperature plays a more important role than the total precipitable water vapor. In addition, DOY is helpful in improving the accuracy of the CNN model in the daytime. Therefore, the best combinations of input parameters are the BTs, NDVI, air temperature, and DOY for the daytime and BTs and air temperature for the nighttime. The CNN LST estimate from AMSR-E with the best combinations yields RMSDs of 2.19–3.58 K for daytime and 1.43–2.14 K for nighttime for diverse land cover types.

Thorough validation of the LST estimates of CNN was conducted based on the in-situ LST. The RMSEs range from 2.10 K to 5.34 K during the daytime and 2.63 K to 4.08 K during the nighttime. The CNN LST agrees well with the in-situ LST, with  $R^2$  values ranging from 0.94 to 0.98. Although the accuracy of the CNN LST is not as satisfactory as the TIR LST in a few cases, the differences between the CNN LSTs and the in-situ LSTs are mainly due to the different land cover types and elevation distributions within the MW pixels containing the stations. Further intercomparison indicates that the ~50% CNN LST estimates are closer to MODIS LSTs than ESA's GlobTemperature AMSR-E LST and the average RMSDs are less than 3 K over dense vegetation compared to NASA's LPDR ATs. Findings

from this study will be beneficial for an in-depth understanding of the use of neural networks for estimating LST from satellite observations.

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

**Funding:** This work was supported in part by the National Key Research and Development Program of China (grant number: 2017YFB0503903), by the National Natural Science Foundation of China (grant number: 41871241), by the National Key Research and Development Program of China (grant number: 2018YFC1505205), by the Fundamental Research Funds for the Central Universities of China (grant number: ZYGX2019J069), and by the ESA-MOST Dragon 5 Cooperation Programme (grant number: 59318).

**Acknowledgments:** The authors would like to thank the three reviewers for improving the clarity and relevance of this work. The authors would also like to thank JAXA for providing AMSR-E and AMSR2 data, the EARTH DATA for providing the MODIS data, GES DISC for providing MERRA-2 and GLDAS data, GlobTemperature Data Portal for providing AMSR-E LST data, and NASA for providing the LPDR dataset.

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

## References

1. Su, Z. The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes. *Hydrol. Earth Syst. Sci.* **2002**, *6*, 85–100. [[CrossRef](#)]
2. Li, Z.L.; Tang, B.H.; Wu, H.; Ren, H.; Yan, G.; Wan, Z.; Trigo, I.F.; Sobrino, J.A. Satellite-derived land surface temperature: Current status and perspectives. *Remote Sens. Environ.* **2013**, *131*, 14–37. [[CrossRef](#)]
3. Zhang, X.; Zhou, J.; Gottsche, F.M.; Zhan, W.; Liu, S.; Cao, R. A Method Based on Temporal Component Decomposition for Estimating 1-km All-Weather Land Surface Temperature by Merging Satellite Thermal Infrared and Passive Microwave Observations. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 4670–4691. [[CrossRef](#)]
4. Chen, Y.; Zhan, W.; Quan, J.; Zhou, J.; Zhu, X.; Sun, H. Disaggregation of Remotely Sensed Land Surface Temperature: A Generalized Paradigm. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 5952–5965. [[CrossRef](#)]
5. Wan, Z.; Dozier, J. A generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Trans. Geosci. Remote Sens.* **1996**, *34*, 892–905. [[CrossRef](#)]
6. Gillespie, A.R.; Rokugawa, S.; Hook, S.J.; Matsunaga, T.; Kahle, A.B. *Temperature/Emissivity Separation Algorithm Theoretical Basis Document, Version 2.4*; ATBD Contract NAS5-31372; NASA: Washington, DC, USA, 1999.
7. Qin, Z.; Karnieli, A.; Berliner, P. A mono-window algorithm for retrieving land surface temperature from Landsat TM data and its application to the Israel-Egypt border region. *Int. J. Remote Sens.* **2001**, *22*, 3719–3746. [[CrossRef](#)]
8. Martin, M.; Ghent, D.; Pires, A.; Götsche, F.M.; Cermak, J.; Remedios, J. Comprehensive In Situ Validation of Five Satellite Land Surface Temperature Data Sets over Multiple Stations and Years. *Remote Sens.* **2019**, *11*, 479. [[CrossRef](#)]
9. Chen, S.; Chen, X.; Chen, W.; Su, Y.; Li, D. A simple retrieval method of land surface temperature from AMSR-E passive microwave data—A case study over Southern China during the strong snow disaster of 2008. *Int. J. Appl. Earth Obs. Geoinf.* **2011**, *13*, 140–151. [[CrossRef](#)]
10. Prigent, C.; Jimenez, C.; Aires, F. Toward “all weather,” long record, and real-time land surface temperature retrievals from microwave satellite observations: Microwave land surface temperature. *J. Geophys. Res. Atmos.* **2016**, *121*, 5699–5717. [[CrossRef](#)]
11. Martins, J.P.A.; Trigo, I.F.; Ghilain, N.; Jimenez, C.; Götsche, F.M.; Ermida, S.L.; Olesen, F.S.; Gellens-Meulenberghs, F.; Arboleda, A. An All-Weather Land Surface Temperature Product Based on MSG/SEVIRI Observations. *Remote Sens.* **2019**, *11*, 3044. [[CrossRef](#)]
12. Zhou, J.; Dai, F.; Zhang, X.; Zhao, S.; Li, M. Developing a temporally land cover-based look-up table (TL-LUT) method for estimating land surface temperature based on AMSR-E data over the Chinese landmass. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *34*, 35–50. [[CrossRef](#)]

13. Huang, C.; Duan, S.B.; Jiang, X.G.; Han, X.J.; Leng, P.; Gao, M.F.; Li, Z.L. A physically based algorithm for retrieving land surface temperature under cloudy conditions from AMSR2 passive microwave measurements. *Int. J. Remote Sens.* **2019**, *40*, 1828–1843. [[CrossRef](#)]
14. McFarland, M.J.; Miller, R.L.; Neale, C.M.U. Land surface temperature derived from the SSM/I passive microwave brightness temperatures. *IEEE Trans. Geosci. Remote Sens.* **1990**, *28*, 839–845. [[CrossRef](#)]
15. Holmes, T.R.H.; De Jeu, R.A.M.; Owe, M.; Dolman, A.J. Land surface temperature from Ka band (37 GHz) passive microwave observations. *J. Geophys. Res.* **2009**, *114*, D04113. [[CrossRef](#)]
16. Fily, M. A simple retrieval method for land surface temperature and fraction of water surface determination from satellite microwave brightness temperatures in sub-arctic areas. *Remote Sens. Environ.* **2003**, *85*, 328–338. [[CrossRef](#)]
17. Gao, H.; Fu, R.; Dickinson, R.E.; Juarez, R.I.N. A Practical Method for Retrieving Land Surface Temperature from AMSR-E Over the Amazon Forest. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 193–199. [[CrossRef](#)]
18. Royer, A.; Poirier, S. Surface temperature spatial and temporal variations in North America from homogenized satellite SMMR-SSM/I microwave measurements and reanalysis for 1979–2008. *J. Geophys. Res. Atmos.* **2010**, *115*. [[CrossRef](#)]
19. Zhao, T.J.; Zhang, L.X.; Shi, J.C.; Jiang, L.M. A physically based statistical methodology for surface soil moisture retrieval in the Tibet Plateau using microwave vegetation indices. *J. Geophys. Res.* **2011**, *116*, D08116. [[CrossRef](#)]
20. André, C.; Ottlé, C.; Royer, A.; Maignan, F. Land surface temperature retrieval over circumpolar Arctic using SSM/I-SSMIS and MODIS data. *Remote Sens. Environ.* **2015**, *162*, 1–10. [[CrossRef](#)]
21. Weng, F.; Grody, N.C. Physical retrieval of land surface temperature using the special sensor microwave imager. *J. Geophys. Res. Atmos.* **1998**, *103*, 8839–8848. [[CrossRef](#)]
22. Aires, F.; Prigent, C.; Bernardo, F.; Jiménez, C.; Saunders, R.; Brunel, P. A Tool to Estimate Land-Surface Emissivities at Microwave frequencies (TELSEM) for use in numerical weather prediction. *Q. J. R. Meteorol. Soc.* **2011**, *137*, 690–699. [[CrossRef](#)]
23. Galantowicz, J.F.; Moncet, J.L.; Liang, P.; Lipton, A.E.; Uymin, G.; Prigent, C.; Grassotti, C. Subsurface emission effects in AMSR-E measurements: Implications for land surface microwave emissivity retrieval. *J. Geophys. Res.* **2011**, *116*, D17105. [[CrossRef](#)]
24. Prigent, C.; Rossow, W.B.; Matthews, E.; Marticorena, B. Microwave radiometric signatures of different surface types in deserts. *J. Geophys. Res. Atmos.* **1999**, *104*, 12147–12158. [[CrossRef](#)]
25. Aires, F.; Prigent, C.; Rossow, W.B.; Rothstein, M. A new neural network approach including first guess for retrieval of atmospheric water vapor, cloud liquid water path, surface temperature, and emissivities over land from satellite microwave observations. *J. Geophys. Res. Atmos.* **2001**, *106*, 14887–14907. [[CrossRef](#)]
26. Jiménez, C.; Prigent, C.; Ermida, S.L.; Moncet, J.L. Inversion of AMSR-E observations for land surface temperature estimation: 1. Methodology and evaluation with station temperature: AMSR-E LAND SURFACE TEMPERATURE. *J. Geophys. Res. Atmos.* **2017**, *122*, 3330–3347. [[CrossRef](#)]
27. Ermida, S.L.; Jiménez, C.; Prigent, C.; Trigo, I.F.; DaCamara, C.C. Inversion of AMSR-E observations for land surface temperature estimation: 2. Global comparison with infrared satellite temperature: AMSR-E LAND SURFACE TEMPERATURE. *J. Geophys. Res. Atmos.* **2017**, *122*, 3348–3360. [[CrossRef](#)]
28. Moncet, J.L.; Liang, P.; Galantowicz, J.F.; Lipton, A.E.; Uymin, G.; Prigent, C.; Grassotti, C. Land Surface Microwave Emissivities Derived from AMSR-E and MODIS Measurements with Advanced Quality Control. Available online: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010JD015429> (accessed on 1 July 2019).
29. Hinton, G.E.; Osindero, S.; Teh, Y.W. A Fast Learning Algorithm for Deep Belief Nets. *Neural Comput.* **2006**, *18*, 1527–1554. [[CrossRef](#)]
30. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
31. Li, X.; Liu, S.; Li, H.; Ma, Y.; Wang, J.; Zhang, Y.; Xu, Z.; Xu, T.; Song, L.; Yang, X.; et al. Intercomparison of Six Upscaling Evapotranspiration Methods: From Site to the Satellite Pixel. *J. Geophys. Res. Atmos.* **2018**, *123*, 6777–6803. [[CrossRef](#)]
32. Shen, H.; Jiang, Y.; Li, T.; Cheng, Q.; Zeng, C.; Zhang, L. Deep learning-based air temperature mapping by fusing remote sensing, station, simulation and socioeconomic data. *Remote Sens. Environ.* **2020**, *240*, 111692. [[CrossRef](#)]

33. Ge, L.; Hang, R.; Liu, Y.; Liu, Q. Comparing the Performance of Neural Network and Deep Convolutional Neural Network in Estimating Soil Moisture from Satellite Observations. *Remote Sens.* **2018**, *10*, 1327. [CrossRef]
34. Tan, J.; NourEldeen, N.; Mao, K.; Shi, J.; Li, Z.; Xu, T.; Yuan, Z. Deep Learning Convolutional Neural Network for the Retrieval of Land Surface Temperature from AMSR2 Data in China. *Sensors* **2019**, *19*, 2987. [CrossRef] [PubMed]
35. Sadeghi, M.; Asanjan, A.A.; Faridzad, M.; Nguyen, P.; Hsu, K.; Sorooshian, S.; Braithwaite, D. PERSIANN-CNN: Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks—Convolutional Neural Networks. *J. Hydrometeorol.* **2019**, *20*, 2273–2289. [CrossRef]
36. Prakash, S.; Norouzi, H.; Azarderakhsh, M.; Blake, R.; Tesfagiorgis, K. Global Land Surface Emissivity Estimation from AMSR2 Observations. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1270–1274. [CrossRef]
37. Prakash, S.; Norouzi, H.; Azarderakhsh, M.; Blake, R.; Prigent, C.; Khanbilvardi, R. Estimation of Consistent Global Microwave Land Surface Emissivity from AMSR-E and AMSR2 Observations. *J. Appl. Meteorol. Climatol.* **2018**, *57*, 907–919. [CrossRef]
38. Zhou, J.; Liang, S.; Cheng, J.; Wang, Y.; Ma, J. The GLASS Land Surface Temperature Product. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 493–507. [CrossRef]
39. Wan, Z. New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product. *Remote Sens. Environ.* **2014**, *140*, 36–45. [CrossRef]
40. Dong, W. EOL Data Archive—CAMP: Tongyu (Inner Mongolia) Surface Meteorology and Radiation Data Set. Available online: <https://data.eol.ucar.edu/dataset/76.141> (accessed on 13 May 2020).
41. Liu, S.M.; Xu, Z.W.; Zhu, Z.L.; Jia, Z.Z.; Zhu, M.J. Measurements of evapotranspiration from eddy-covariance systems and large aperture scintillometers in the Hai River Basin, China. *J. Hydrol.* **2013**, *487*, 24–38. [CrossRef]
42. Liu, S.M.; Xu, Z.W.; Wang, W.Z.; Jia, Z.Z.; Zhu, M.J.; Bai, J.; Wang, J.M. A comparison of eddy-covariance and large aperture scintillometer measurements with respect to the energy balance closure problem. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 1291–1306. [CrossRef]
43. Li, X.; Cheng, G.; Liu, S.; Xiao, Q.; Ma, M.; Jin, R.; Che, T.; Liu, Q.; Wang, W.; Qi, Y.; et al. Heihe Watershed Allied Telemetry Experimental Research (Hiwater). *Bull. Am. Meteorol. Soc.* **2013**, *94*, 1145–1160. [CrossRef]
44. Liu, S.; Li, X.; Xu, Z.; Che, T.; Xiao, Q.; Ma, M.; Liu, Q.; Jin, R.; Guo, J.; Wang, L.; et al. The Heihe Integrated Observatory Network: A Basin-Scale Land Surface Processes Observatory in China. *Vadose Zone J.* **2018**, *17*. [CrossRef]
45. Yang, J.; Zhou, J.; Götsche, F.M.; Long, Z.; Ma, J.; Luo, R. Investigation and validation of algorithms for estimating land surface temperature from Sentinel-3 SLSTR data. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *91*, 102136. [CrossRef]
46. Xu, Z.; Liu, S.; Li, X.; Shi, S.; Wang, J.; Zhu, Z.; Xu, T.; Wang, W.; Ma, M. Intercomparison of surface energy flux measurement systems used during the HiWATER-MUSOEXE. *J. Geophys. Res. Atmos.* **2013**, *118*, 13140–13157. [CrossRef]
47. Liang, S.L. *Quantitative Remote Sensing of Land Surfaces*; John Wiley & Sons: Hoboken, NJ, USA, 2004; p. 388.
48. Pearson, R.K. Outliers in process modeling and identification. *IEEE Trans. Control Syst. Technol.* **2002**, *10*, 55–63. [CrossRef]
49. Götsche, F.M.; Olesen, F.S.; Trigo, I.F.; Bork-Unkelbach, A.; Martin, M.A. Long term validation of land surface temperature retrieved from MSG/SEVIRI with continuous in-situ measurements in Africa. *Remote Sens.* **2016**, *8*, 410. [CrossRef]
50. Du, J.; Jones, L.A.; Kimball, J.S. *Daily Global Land Surface Parameters Derived from AMSR-E and AMSR2, Version 2*; NASA National Snow and Ice Data Center Distributed Active Archive Center: Boulder, CO, USA, 2017. [CrossRef]
51. Du, J.; Kimball, J.S.; Jones, L.A.; Kim, Y.; Glassy, J.; Watts, J.D. A global satellite environmental data record derived from AMSR-E and AMSR2 microwave Earth observations. *Earth Syst. Sci. Data* **2017**, *9*, 791–808. [CrossRef]
52. Jones, L.A.; Ferguson, C.R.; Kimball, J.S.; Zhang, K.; Chan, S.T.K.; McDonald, K.C.; Njoku, E.G.; Wood, E.F. Satellite Microwave Remote Sensing of Daily Land Surface Air Temperature Minima and Maxima From AMSR-E. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2010**, *3*, 111–123. [CrossRef]

53. Jones, L.A.; Kimball, J.S.; McDonald, K.C.; Chan, S.T.K.; Njoku, E.G.; Oechel, W.C. Satellite Microwave Remote Sensing of Boreal and Arctic Soil Temperatures From AMSR-E. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 2004–2018. [[CrossRef](#)]
54. Springenberg, J.T.; Dosovitskiy, A.; Brox, T.; Riedmiller, M. Striving for Simplicity: The All Convolutional Net. *arXiv* **2015**, arXiv:14126806 Cs.
55. Zheng, D.; Wang, X.; van der Velde, R.; Ferrazzoli, P.; Wen, J.; Wang, Z.; Schwank, M.; Colliander, A.; Bindlish, R.; Su, Z. Impact of surface roughness, vegetation opacity and soil permittivity on L-band microwave emission and soil moisture retrieval in the third pole environment. *Remote Sens. Environ.* **2018**, *209*, 633–647. [[CrossRef](#)]
56. Zhan, W.; Zhou, J.; Ju, W.; Li, M.; Sandholt, I.; Voogt, J.; Yu, C. Remotely sensed soil temperatures beneath snow-free skin-surface using thermal observations from tandem polar-orbiting satellites: An analytical three-time-scale model. *Remote Sens. Environ.* **2014**, *143*, 1–14. [[CrossRef](#)]
57. Zhou, J.; Zhang, X.; Zhan, W.; Götsche, F.M.; Liu, S.; Olesen, F.S.; Hu, W.; Dai, F. A thermal sampling depth correction method for land surface temperature estimation from satellite passive microwave observation over barren land. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 4743–4756. [[CrossRef](#)]
58. Foody, G.M. Classification accuracy comparison: Hypothesis tests and the use of confidence intervals in evaluations of difference, equivalence and non-inferiority. *Remote Sens. Environ.* **2009**, *113*, 1658–1663. [[CrossRef](#)]
59. de Beurs, K.M.; Henebry, G.M.; Owsley, B.C.; Sokolik, I. Using multiple remote sensing perspectives to identify and attribute land surface dynamics in Central Asia 2001–2013. *Remote Sens. Environ.* **2015**, *170*, 48–61. [[CrossRef](#)]
60. Wang, P.; Zhao, T.; Shi, J.; Hu, T.; Roy, A.; Qiu, Y.; Lu, H. Parameterization of the freeze/thaw discriminant function algorithm using dense *in-situ* observation network data. *Int. J. Digit. Earth* **2019**, *12*, 980–994. [[CrossRef](#)]
61. Zhao, S.; Zhang, L.; Zhang, Y.; Jiang, L. Microwave emission of soil freezing and thawing observed by a truck-mounted microwave radiometer. *Int. J. Remote Sens.* **2012**, *33*, 860–871. [[CrossRef](#)]
62. Cordisco, E.; Prigent, C.; Aires, F. Snow characterization at a global scale with passive microwave satellite observations. *J. Geophys. Res.* **2006**, *111*, D19102. [[CrossRef](#)]
63. Huang, G.; Li, X.; Huang, C.; Liu, S.; Ma, Y.; Chen, H. Representativeness errors of point-scale ground-based solar radiation measurements in the validation of remote sensing products. *Remote Sens. Environ.* **2016**, *181*, 198–206. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).