

# Estimating global downward shortwave radiation from VIIRS data using a transfer-learning neural network

Ruohan Li, Dongdong Wang<sup>\*</sup>, Shunlin Liang, Aolin Jia, Zhihao Wang

*Department of Geographical Sciences, University of Maryland, College Park, MD 20742, USA*



## ARTICLE INFO

**Keywords:**

Transfer learning  
Machine learning  
VIIRS  
Downward shortwave radiation  
Radiative transfer  
Solar energy

## ABSTRACT

In recent years, machine learning (ML) has been successfully used in estimating downward shortwave radiation (DSR). To achieve global estimations, traditional ML models need sufficient ground measurements covering various atmospheric and surface conditions globally, which is difficult to accomplish. Training on the simulated data of a radiative transfer model (RTM) is a possible solution, but widely used RTMs ignore some complex cloud conditions which brings bias to simulations. In this study, a neural network applied with the transfer-learning (TL) concept is introduced to utilize both radiative transfer simulations and ground measurement data, achieving global DSR estimation with only top-of-atmosphere and surface albedo at local solar noon as inputs. The proposed method estimates both instantaneous and daily DSR from Visible Infrared Imaging Radiometer Suite (VIIRS) data at 750-m resolution, and both the estimates are validated by 40 independent stations globally. The root mean-square error and relative root mean square error of instantaneous DSR validation over 25 Baseline Surface Radiation Network, seven Surface Radiation Network, and eight Greenland Climate Network stations in 2013 were 91.2 (16.1%), 106.3 (18.3%), 75.0 (24.2%)  $W/m^2$ , respectively, and the daily validation achieved 30.8 (15.5%), 33.5 (17.6%), and 31.3 (14.4)  $W/m^2$ , respectively. The proposed method presents significant high accuracy over polar regions and similar performances over other areas compared with traditional ML models, physics models (e.g., look-up tables and direct estimations), and existing DSR products. The algorithm is also applied to VIIRS swath data to test its global efficacy. Instantaneous mapping captures the spatial pattern of the cloud-mask product, and daily mapping shows spatial patterns similar to the Clouds and the Earth's Radiant Energy System Synoptic TOA and surface fluxes and clouds product, but with more detail. Further analysis indicates that model performance is less sensitive to the quantity of training data after TL has been incorporated. This study demonstrates the advantages of TL on boosting both the generality and accuracy of DSR estimation, which can potentially be applied to other variable retrievals.

## 1. Introduction

As the driving force of many ecological, hydrological, and biogeochemical processes at the Earth's surface and in the atmosphere, downward shortwave radiation (DSR) provides critical parameters for surface energy balance and has been the focus of many studies (Liang et al., 2019). Significant progress has been made regarding estimating DSR from satellite observations, and several global and regional DSR satellite products are currently available (Mueller et al., 2009; Rutan et al., 2015; Wang et al., 2020; Zhang et al., 2014).

Previous methods that estimate surface DSR from satellite data are largely based on the physical process of radiative transfer models (RTMs). These methods can be classified as either inverse or forward

algorithms (Wang et al., 2021). Inverse models estimate DSR directly from satellite observations of top-of-atmosphere (TOA) reflectance. A look-up-table (LUT) based approach is a typical inverse model. In this approach, simulated results from RTMs are stored, which consequently simplifies the radiative transfer process (Huang et al., 2019). By comparing the simulated TOA reflectance with satellite observations, the algorithm determines the atmospheric conditions and calculates surface flux using an offline LUT. Inversion methods such as LUT mostly use linear interpolation to sort atmospheric conditions within pre-calculated bins (Zhang et al., 2018). Parameterization algorithms are forward methods which provide parameterization schemes for different atmospheric conditions. These methods usually achieve more accurate estimations, but generate fewer valid retrievals as they require various

\* Corresponding author at: Department of Geographical Sciences, University of Maryland, 1127 LeFrak Hall, College Park, MD 20742, USA.  
E-mail address: [ddwang@umd.edu](mailto:ddwang@umd.edu) (D. Wang).

high-level surface and atmospheric products as inputs (Wang et al., 2021). This requirement also poses difficulties in applying parameterization method to generate high spatial and temporal resolution DSR products.

Machine-learning (ML) methods have been actively used to estimate DSR over the past few years, owing to their superior capability in capturing non-linear relationships with higher efficiency. Random-forest, gradient-boosting regression tree (GBRT), and neural-network (NN) models are widely used to estimate DSR (Babar et al., 2020; Chen et al., 2019; Hou et al., 2020; Yang et al., 2018). In recent years, some deep-learning models, such as convolutional NNs (CNN) (Jiang et al., 2019) and generalized regression NNs (Senkal, 2010) have been employed (Yuan et al., 2020). Several studies have compared these ML models using the same inputs and demonstrated that NN and GBRT presented the best performances (Brown et al., 2020; Wei et al., 2019).

However, the accuracy of ML methods greatly depends on the quality and quantity of training data (Zhang et al., 2018). Some ML models were trained on ground-measurement data (Babar et al., 2020; Brown et al., 2020; Hao et al., 2019; Peng et al., 2020; Wei et al., 2019; Yang et al., 2018). Such models generally achieve high accuracy when trained at regional scale or trained with data that are representative enough. The capabilities of these models when implemented at larger spatial scales or with more scarce conditions have not been fully demonstrated as sufficient representative training data is hard to obtain, especially for polar-orbiting sensors. Thus, the applications of MLs trained on ground-measurement data to estimate DSR globally is still limited. Other ML models are trained using radiative transfer simulations which cover sufficient instances under different atmospheric conditions and over different surface types (Bue et al., 2019; Chen et al., 2019; Ma et al., 2020; Ryu et al., 2018; Shi et al., 2019; Takenaka et al., 2011; Wang et al., 2012). The data size is usually large, which is preferred by NNs. However, owing to tradeoffs between the input numbers of driving parameters and RTM model efficiencies, gaps still exist between the RTM simulated data and the real-world data. For example, most simulations are still based on less computationally expensive 1-dimensional RTM, which consider cloud as horizontally homogeneous and infinite (Chen et al., 2019), meaning that the multilayer and 3-dimensional cloud effects are ignored. At regional scale, especially over areas with sufficient measurement data, simulation-based ML models will present less accuracies.

It would be ideal to combine both simulated and ground-measurement data in the training model which can be achieved by adapting the concept of transfer learning (TL). TL uses the model trained on source domain data as a starting point, and is fine-tuned in the process of developing a model for target domain (Pan and Yang, 2010). It is a powerful tool to transfer pre-known knowledge to a target model with improved generalization ability, even when the target training datasets are limited (Pan and Yang, 2010). Learning from simulations is a good practice of TL (Malmgren-Hansen et al., 2017), as simulations provide sufficient data covering comprehensive conditions and carries traceable physics knowledge which is valuable for initial model training. The gaps between the simulated data and real-world data can be filled when the model is further fine-tuned. TL has been applied in some studies of remote sensing, mostly in the area of land cover classification (Huang et al., 2018; Tong et al., 2020), where transferred CNN models were adapted to classify optical remote sensing images. Other areas include extending regional models (Wei et al., 2016) or extending one variable to similar variables estimation (Khaki et al., 2021). Transferring from simulated data has also conducted to classify synthetic aperture radar target (Zhou et al., 2020). However, to the best of our knowledge, TL has not been employed to improve DSR estimation.

This study proposes a robust NN method that employs TL to estimate instantaneous and daily DSR based on Visible Infrared Imaging Radiometer Suite (VIIRS) TOA reflectance and Moderate Resolution Imaging Spectroradiometer (MODIS) black-sky albedo (BSA) at local solar noon only. The proposed method leverages the advantages of both using

radiative transfer simulation data and ground measurement data for training, which enables accurate estimation of DSR under various climate regions, even over polar regions. The remainder of this paper is organized as follows: Section 2 introduces the input data and data separation method; Section 3 describes the methodology; Section 4 presents and discusses validation results; the main conclusions are drawn in Section 5.

## 2. Data

### 2.1. Real-world data, $D_{real}$

Let  $D_{real}$  be the real-world data serving as the target data domain in this study.  $D_{real}$  is composed of satellite data and in-situ measurements which are matched based on their locations.

#### 2.1.1. Satellite data

The TOA reflectance, geometry information, and surface BSA are extracted from satellite data: VIIRS/Suomi NPP Moderate Resolution Bands (VNP02MOD), VIIRS/Suomi NPP Moderate Resolution Terrain-Corrected Geolocation (VNP03MOD), and MODIS/Terra+Aqua Albedo (MCD43A3), respectively. VNP02MOD provides calibrated TOA spectral reflectance and pixel-level quality flags at 750-m resolution. Bands M1–M11 (except for M6) are used as inputs. VNP03MOD includes corresponding geographical coordinate locations and viewing geometry information. Three geometry angles are extracted. The surface BSA at local solar noon for MODIS bands 1 through 7 are retrieved from MCD43A3, which are produced daily at 500-m resolution. 16 days of Terra and Aqua data are weighted in MCD43, consequently containing more available data.

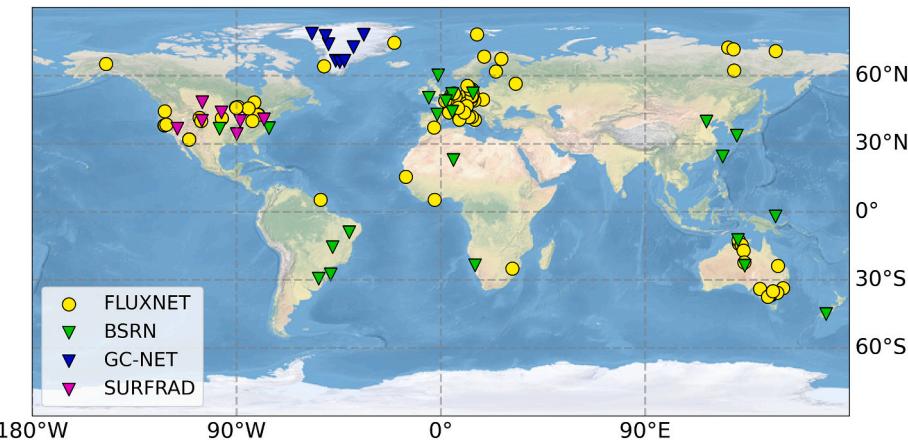
#### 2.1.2. In-situ observations

The in-situ observations of DSR are collected from four different networks serving as different purposes in this study. 96 sites from FLUXNET global network of micrometeorological tower sites in 2013 are used as training/fine-tuning datasets. The results are validated over 25 Baseline Surface Radiation Network (BSRN) stations, eight Greenland Climate Network (GC-NET) stations with data in 2013, and seven Surface Radiation Network (SURFRAD) stations with data in both 2013 and 2014. All the validation data are independent from the training data. Both SURFRAD and BSRN adopted high-level measurement standards set by the World Climate Research Program of the World Meteorological Organization (Augustine et al., 2000). The distribution of the in-situ sites is shown in Fig. 1. The instantaneous DSR is calculated by averaging the nearest half-hour data from the VIIRS overpass time, and the daily DSR is the mean of all instantaneous measurements in one day and is calculated as in Li et al. (2021). Measurements with a solar zenith angle (SZA) higher than 87 degrees are excluded.

### 2.2. Simulated data, $D_{sim}$

Let  $D_{sim}$  be the source domain of this study. MODTRAN5 (Berk et al., 2004) is employed to generate the simulated database. Following Wang et al. (2015), our radiative transfer model is operated for 294 cases of viewing geometry, including seven values of SZA, six values of viewing zenith angle (VZA), and seven values of relative azimuth angle (RAA). Different combinations of surface and atmospheric parameters are used as inputs for each view. The corresponding TOA reflectance of VIIRS bands and simulated instantaneous and daily DSR measures are also recorded.

To cover comprehensive surface types, a library of 245 surface albedo is used, which is obtained from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (Baldridge et al., 2009) and the U.S. Geological Survey (Clark et al., 2007) surface spectra databases. In terms of atmospheric inputs, different driving parameters are considered under clear and cloudy sky conditions. The aerosol types, aerosol optical



**Fig. 1.** Distribution of ground measurement stations. Training data are labeled as circles. Validating and testing data are labeled as triangles.

depths (AODs), and water-vapor loads are modeled under clear skies. For cloudy skies, cloud optical depths (CODs), types, heights, and water-vapor loads are set as input parameters. The scattering and absorption of atmospheric particulate matter are ignored when clouds are present.

Daily DSRs are averaged from instantaneous DSRs throughout the day, assuming that the atmospheric conditions remain the same. Under the same VZA, RAA, surface, and atmospheric conditions, instantaneous DSRs at any daytime,  $t$ , in a day are only related to SZA. Consequently, daily DSR simulations include two more parameters (i.e., latitude and solar declination angle) to calculate SZA( $t$ ). Detailed settings are listed in Table 1.

In summary, MODTRAN is operated with the combination of 294 cases of view geometry, 245 surface albedo, 576 cloudy-sky atmospheric conditions, and 210 clear-sky atmospheric conditions to simulate a comprehensive database. However, not all MODTRAN variables are used as NN inputs. Sensitivity analysis of previous studies have shown that SZA, albedo, and cloud fraction are the top-three factors influencing the DSR estimation under all sky conditions (Brown et al., 2020; Chen et al., 2019). Hence, only spectral TOA reflectance, surface albedo, and geometry information data are required for the proposed model. The other driving parameters for the simulation model are flattened. We then have our simulation database,  $D_{sim}$ .

### 2.3. Additional data

The proposed method is also compared with existing high resolution DSR products including the DSR product based on the Earth Polychromatic Imaging Camera (EPIC), the Breathing Earth System Simulator (BESS) shortwave radiation product, the Global Land Surface

**Table 1**  
Driving parameters used in MODTRAN.

Parameters	Values
	<b>View geometry</b>
Solar zenith angle (SZA)	0,10,20,30,45,60,75
View zenith angle (VZA)	0,10,20,30,45,60,75
Relative azimuth angle (RAA)	0,30,60,90,120,150,180
	<b>Clear sky</b>
Aerosols types	5 values
Aerosol optical depth (AOD)	0,0,0.05,0,10,0,2,0,4,0,6,1,0
Water vapor	0,5,1,0,1,5,3,0,5,0,7,0
	<b>Cloud sky</b>
Cloud types	stratus, cumulus, altostratus, nimbostratus
Cloud height	3 values
Cloud optical depth (COD)	0,25,1,2,3,5,10,20,30,60,120,180,240
Water vapor	1,5,3,0,7,0,12,0
	<b>Daily</b>
Latitude	0,10,20,30,40,50,60,70,80,90
Declination angle	-23,5,23,5,4,23,5

Satellite (GLASS) DSR product, and MODIS/Terra+Aqua Surface Radiation (MCD18). The EPIC uses the random forest method to generate DSR at 0.1-degree spatial resolution and 1-2 h temporal resolution. The BESS product is based on NN trained from RTM simulations. A series of MODIS atmosphere and land product data are included as the inputs of the RTM (Ryu et al., 2018). The product generates 5 km DSR globally. The GLASS also provides daily DSR at 5 km; it applies the direct estimation approach based on MODIS TOA spectral reflectance and GLASS broadband albedo (Zhang et al., 2019). The MCD18A1 Version 6.1 generates 3 h DSR at 1 km; it employs the LUT approach in which the MODIS TOA reflectance and albedo are used as the main inputs (Wang et al., 2020). It has been used as inputs of some surface variable retrieval methods (Jia et al., 2021; Wu et al., 2021). Clouds and the Earth's Radiant Energy System (CERES) for Synoptic TOA and surface fluxes and clouds (SYN) DSR data are employed in this study as a reference for daily estimation. It offers the accurate satellite DSR estimation at 1 degree worldwide (Li et al., 2021; Zhang et al., 2015). It uses Langley Fu-Liou radiative transfer codes and includes various satellite observations and reanalysis data as input (Rutan et al., 2015).

CLDMSK\_L2\_VIIRS\_SNPP data are also used. This product inherits the algorithms designed for MOD35 at a spatial resolution of 750 m. Both snow and cloud information used in this study are derived from the CLDMSK\_L2\_VIIRS\_SNPP product.

## 3. Methods

### 3.1. NN structure

Simple back-propagation NNs have been widely used to estimate DSR in various studies (Peng et al., 2020; Ryu et al., 2018; Wang et al., 2012). The theoretical base of the algorithm was described by Takenaka et al. (2011). The input layer of the proposed model has 20 neurons for the instantaneous estimation model, which includes 10 bands of TOA reflectance, seven bands of surface BSA, and three geometry angles. For daily models, two more neurons are added: latitude and declination angles. The sensitivity analysis of the input parameters is discussed in Section 4.1. To determine the hyper-parameters of the NN model, the models have been trained on different combinations of different hyper-parameters. SURFRAD data are used to validate and select the model parameters. To balance the model accuracies and the computational efficiencies, a four-layers (i.e., one input, two hidden, and one output) NN model is selected. The first hidden layer has 25 neurons followed by another hidden layer of five neurons with a rectified linear unit activation function (Nair and Hinton, 2010). The output layer has one neuron corresponding to the DSR value of linear activations. The weights of the model,  $w$ , are initialized following the Glorot uniform:

$$w \sim U \left[ -\frac{\sqrt{6}}{\sqrt{n_l + n_{l+1}}}, \frac{\sqrt{6}}{\sqrt{n_l + n_{l+1}}} \right], \text{ for } l = 0, \dots, L-1, \quad (1)$$

where  $L$  is the number of layers in the NN, and  $n_l$  is the number of neurons in the layers. Because the final outputs of the model are the numerical values of the DSR, the mean-square error (MSE) function is used as the error function:

$$E(w) = \frac{\sum_{i=1}^n [y(w, x_i) - h_i]^2}{n} \quad (2)$$

where  $h_i$  is the simulated/measured DSR for the  $i$ th sample. The error function is further minimized using RMSprop (Hinton et al., 2012), which is a type of stochastic gradient descent that enables a stable learning curve and boosts the convergence speed by calculating the differential square weighted average of the gradient.

### 3.2. Transfer learning model construction

The construction of the transfer learning model requires two rounds of training (Fig. 2). It is firstly trained from the source domain data, which is  $D_{sim}$  in this study. The model here is referred to as the base model (BM). Then the model is further fine-tuned by the target domain data,  $D_{real}$ . This is the final model used as DSR estimation, and is referred to as a transfer learning model (BM\_TL).

BM was constructed following the NN configuration, with training datasets consisting completely of simulation data. BM can be summarized as  $y(x, w)$ ,  $x \in D_{sim}$ . We randomly separated the simulation data into 80% training and 20% as training and validation data. Since the data size is significantly large, no cross-validation was applied. We also included SURFRAD measurements as additional validation data to

ensure that the selected BM model is not overfitted to the ground measurements. BM learned the knowledge from the simulation data. To adapt the model to real-world data, the model is fine-tuned by the matching data between satellite products and FLUXNET measurements. Previous studies revealed that the first layers of the NN models are not usually specific to a particular task but instead general to many related tasks, with the last few layers being task-specific (Yosinski et al., 2014). In this study, the weights of the first layer,  $w_1$ , were directly copied from BM, as shown in Fig. 2, meaning that they are initialized from  $\text{argmin } E(w, x)$ , for  $x \in D_{sim}$ . These weights were frozen and were not changed. The remaining weights from other layers were initialized by (1), and updated using RMSprop during the minimization process of (2), but with  $x \in D_{real}$ . Consequently, the BM\_TL can be summarized as  $y_{tl}(x, w_{tl})$ ,  $x \in D_{real}$ .

### 3.3. Evaluation

In this study, the proposed method was trained and evaluated on independent sites (i.e., no overlap sites exist between the training and testing data) as DSR estimation is related to surface properties. FLUXNET data served as training/tuning data, as they have comparatively sufficient measurements globally. SURFRAD sites were used for model selection and tuning hyper-parameters. The models were validated on BSRN and GCNET sites separately. FLUXNET, SURFRAD, and BSRN presented similar variable distributions while GCNET showed a different distribution with the peak of the visible bands of TOA and surface albedo located at larger values. GCNET provides unique measurements over high latitude areas where some existed DSR products degrade (Li et al., 2021). Hence, it helps to investigate the model's performance on under-representative conditions. The root mean square error (RMSE), relative root mean square error (rRMSE), mean bias difference (MBD) and R-square ( $R^2$ ) measures are the main assessment metrics of this study.

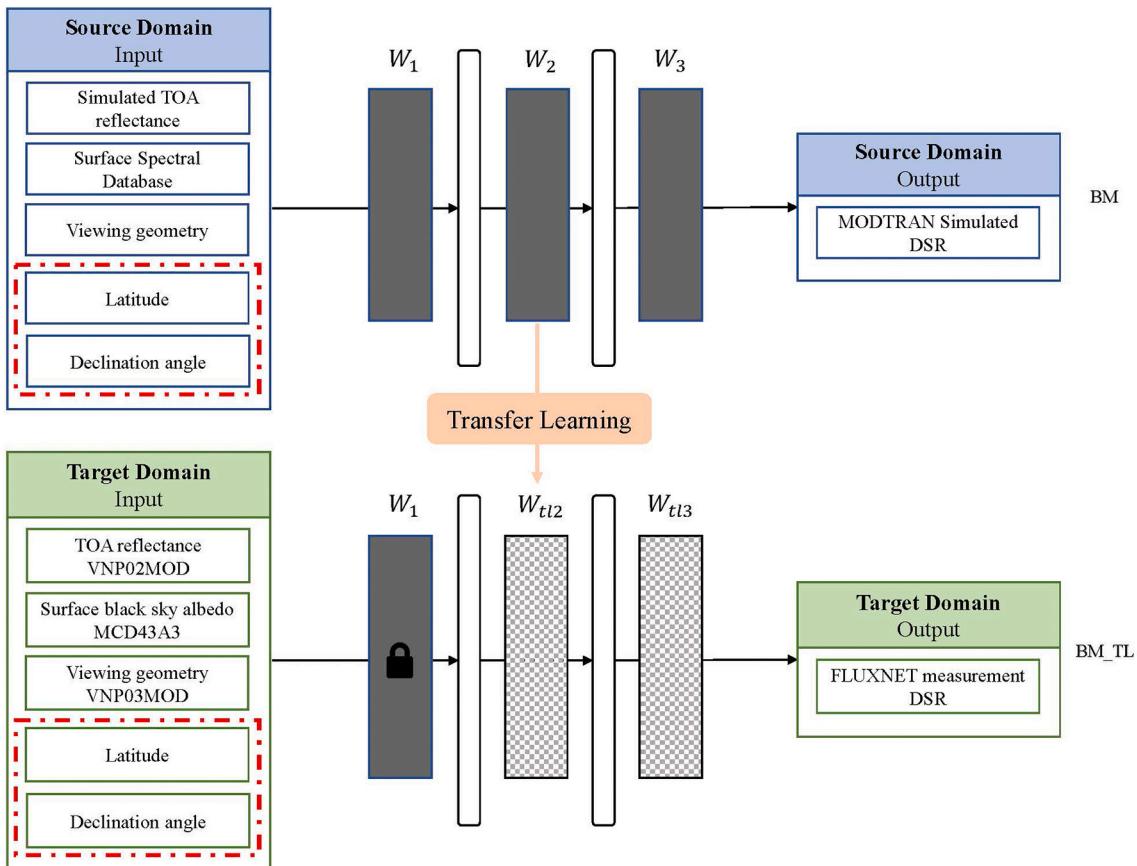


Fig. 2. Flowchart for constructing BM\_TL.

### 3.4. Other methods

To compare the traditional NN method with the proposed method, a baseline model, BM\_SA, is trained purely from in-situ measurements. BM\_SA can be summarized as  $y_{sa}(x, w_{sa})$ ,  $x \in D_{real}$ . The same NN configuration as BM\_TL is applied. FLUXNET data are used as the only training data and SURFRAD data are still used as validation datasets.

The narrow-band LUT approach was first proposed by (Liang et al., 2006), which includes a reflectance LUT to determine the atmospheric condition and a flux LUT to finalize surface DSR estimation. The LUT stores the pre-calculated results from the same RTM simulations. The method is further developed by Wang et al. (2020) to generate MCD18 products. It is noticeable that the original LUT method was applied to the MODIS TOA band. In this study, the VIIRS band is used instead. Divergence between VIIRS and MODIS bands may lead to the results being slightly different from the published ones (Wang et al., 2020).

Direct estimation utilizes the linear relationship between TOA reflectance and surface shortwave net radiation of narrow bands under cloudy and clear-sky conditions (Wang et al., 2021). The linear coefficients are also retrieved from the same RTM simulation data. In this case, we generate coefficients for each VIIRS band. The surface shortwave net radiation is then estimated, and the DSR is calculated by combining broadband albedos. This method employs cloud-mask information from the CLDMSK\_L2\_VIIRS\_SNPP products.

## 4. Results and discussion

### 4.1. Sensitivity analysis

Since using VIIRS TOA reflectance as inputs for DSR estimation is new, a sensitivity analysis was conducted to investigate the impact of the input variables for both instantaneous and daily DSR estimation. The Sobol analysis method (Sobol, 2001) from the SALib library in Python was employed. Both first-order indices (S1) and total-order indices (ST)

were calculated. S1 measures the contribution to the output variances caused by a single input alone. ST is S1 plus all higher-order interactions of this input. The results are shown in Fig. 3. For instantaneous DSR, the visible band 1–4 with the reflected range from 0.4–0.6 μm carry the most of the information needed for the DSR estimation. Among them, band3 and band4 account for more than 65% of the total variance. The near infrared to medium-wave infrared bands 5–11 also account for 12.5% in total, but most of the variance are from the interaction with other inputs. Unlike previous sensitivity analysis based on forward methods (Li et al., 2015; Chen et al., 2012), the SZA is not the first dominant factors but still plays an important role with 6.8%. Surface BSA band 3 and 4, combined, also influence the DSR estimation with 6.5%, which provides required surface conditions and multi-scattering information. The influence of other bands of surface BSA are negligible. The daily model has similar sensitivity analysis results as instantaneous one, but it is noticeable that more variances are from the interactions with other variables, as ST has a smaller portion of S1 compared with instantaneous model. Besides, SZA occupies smaller part. More percentages are given to latitude and declination angles.

### 4.2. Comparison with traditional NN methods

The proposed method is tested on the various in-situ networks in 2013 and the results are shown separately due to the different purposes of each network in this study. FLUXNET data serve as the training data of BM\_SA and BM\_TL model and hence represent the training accuracies of these two models. To demonstrate how TL improves the traditional ML, the proposed model (BM\_TL) is also compared with BM and BM\_SA. Fig. 4 and Fig. 5 show the validation results for instantaneous and daily estimations, respectively.

For instantaneous estimation, BM had a higher RMSE than BM\_SA over FLUXNET, SURFRAD and BSRN stations. The RMSE of the BM model over FLUXNET, SURFRAD and BSRN data scored 127.0, 126.2 and 104.5 W/m<sup>2</sup> respectively, whereas BM\_SA achieved an RMSE of

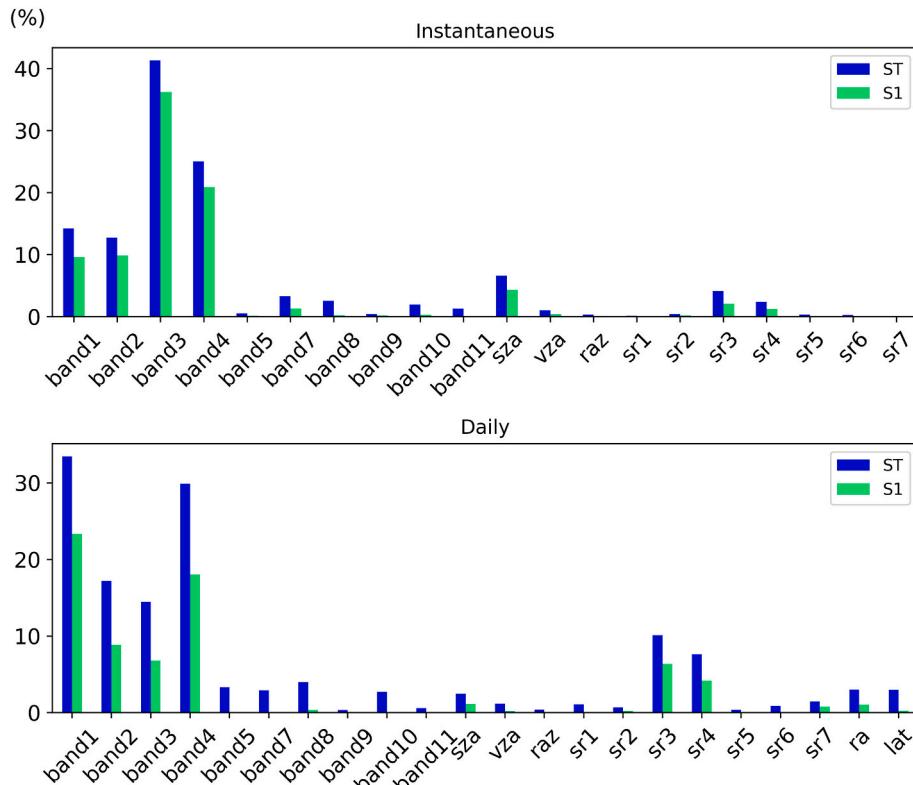
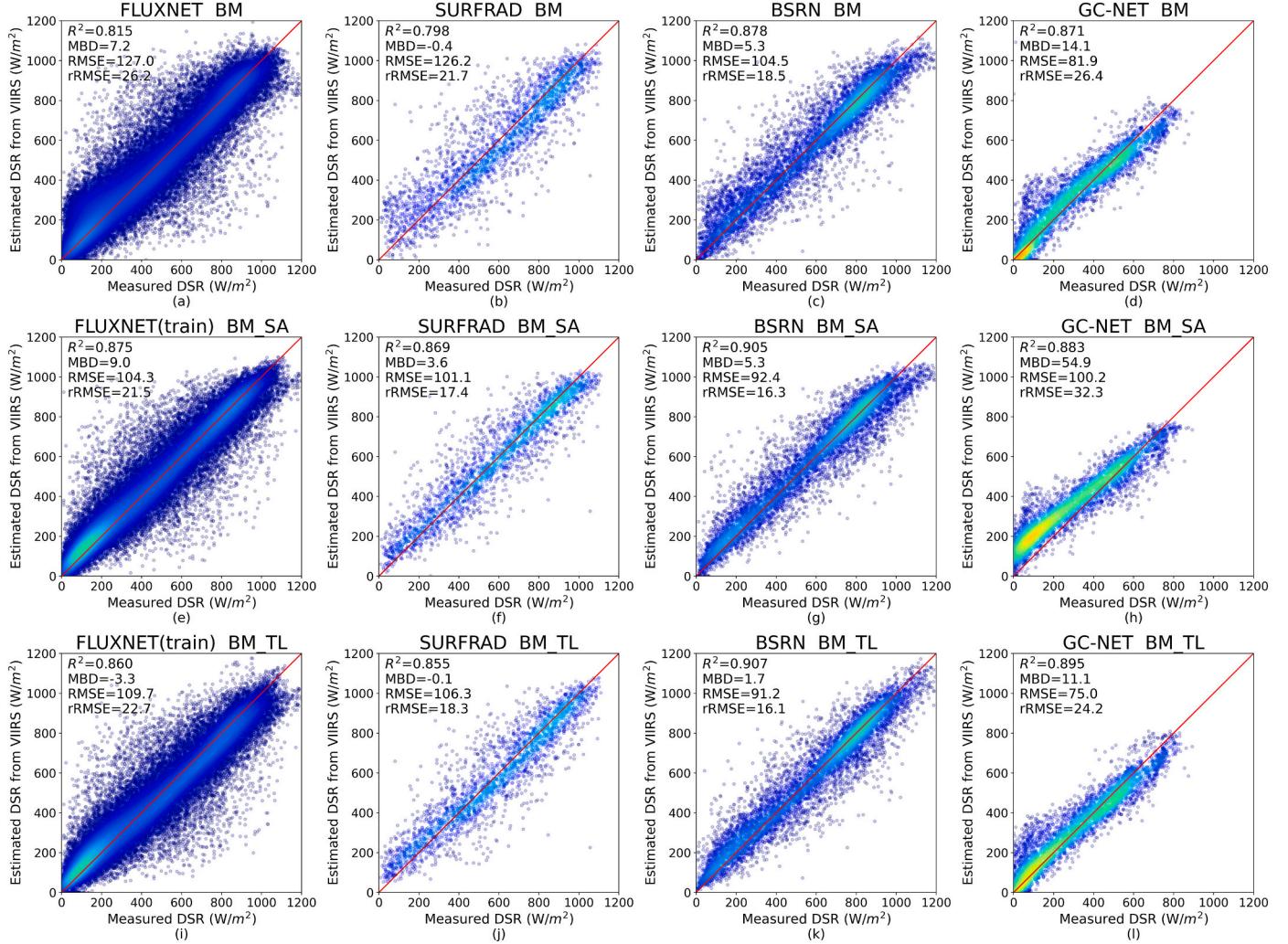


Fig. 3. The first-order indices (S1) and total-order indices (ST) for input variables of daily and instantaneous BM\_TL.



**Fig. 4.** Comparison among the estimated instantaneous DSR by BM, BM\_SA, and BM\_TL models over ground measurements at SURFRAD, BSRN, and GC-NET stations in 2013.

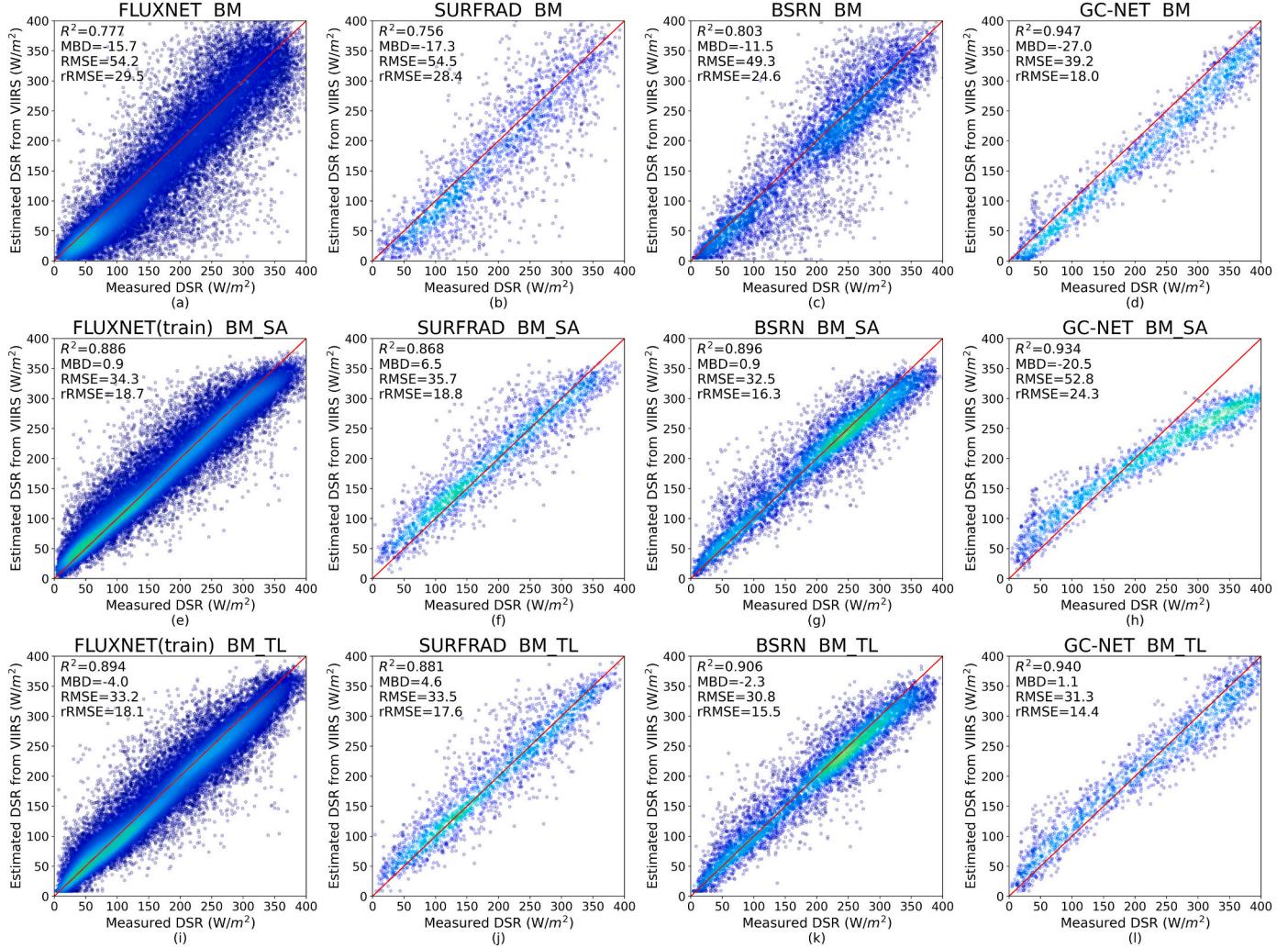
104.3, 101.1 and 92.4  $W/m^2$ . Over the GC-NET stations, BM had a lower RMSE of 81.9  $W/m^2$  compared with BM\_SA which had an RMSE of 100.2  $W/m^2$ . The RMSE of BM\_TL were 109.7 (22.7%), 106.3 (18.3%), 91.2 (16.1%) 75.0 (24.2%)  $W/m^2$  over FLUXNET, SURFRAD, BSRN, and GCNET respectively. Compared with BM\_SA, BM\_TL had a slightly higher RMSE over FLUXNET and SURFRAD, but lower over BSRN and GCNET stations. BM\_TL also had better accuracies than BM over all the sites.

The comparisons between the estimated daily DSR are shown in Fig. 5. Similar to the instantaneous estimation, BM\_SA achieved significantly better results than BM over the FLUXNET, SURFRAD and BSRN station data. The RMSE of the BM at FLUXNET, SURFRAD and BSRN stations were 54.2, 54.5, and 49.3  $W/m^2$ , respectively. The RMSE of BM\_SA at FLUXNET, SURFRAD and BSRN stations was 34.3, 35.7, and 32.5  $W/m^2$ . It is noticeable that the difference between BM and BM\_SA over these two networks was larger than the instantaneous estimations. This may indicate that the simulated daily samples have more systematic deviations from the truth than the instantaneous samples did due to the assumption of stable atmospheric conditions throughout a day. Owing to the insufficient training data over polar regions, the overfitting issues of the BM\_SA model persisted, which led to the highest RMSE of 52.8  $W/m^2$  over the GC-NET stations among the three models. The employment of BM\_TL presented significant improvements over GC-NET stations and slightly better results for the FLUXNET, SURFRAD and BSRN stations. The RMSEs of BM\_TL were 33.2 (18.1%), 33.5 (17.6%), 30.8 (15.5%),

and 31.3 (14.4)  $W/m^2$  over FLUXNET, SURFRAD, BSRN, and GC-NET, respectively, at raw resolutions for daily DSR estimations.

In summary, BM\_SA achieved a higher accuracy for the FLUXNET, SURFRAD and BSRN stations for both instantaneous and daily estimations, but it degraded dramatically for GC-NET stations. This corresponds to the overfitting issues of the ML model trained only on in-situ data only. The SURFRAD and BSRN stations have a similar sample distribution as with the training data from FLUXNET, which led to high accuracy. Over polar regions, no sufficient training samples were available; hence, BM\_SA fell short. On the other hand, BM displayed superior performance over high latitude areas. No obvious bias as observed, and the RMSE was comparable to previous methods (Zhang et al., 2018), demonstrating the strength of training using simulation data that cover sufficient atmospheric and surface cases. However, the accuracy of general cases was low, especially for daily estimations. BM\_TL successfully combined the advantages of both BM and BM\_SA, boosting its generality as it performed stably over all networks and resolved the issues of inaccurate daily simulated samples; it improves the accuracies as well. These results support the previous findings that TL can leverage both source and target data to better tune the model, and it demonstrates the feasibility of applying TL to DSR estimation.

To further investigate how BM\_TL improved the traditional ML, the BM, BM\_SA, and BM\_TL models were compared at different latitudes and SZAs, as shown in Fig. 6. The relative RMSE (rRMSE) of BM and BM\_SA were similar over low- to mid-latitudes. BM\_TL achieved comparable or



**Fig. 5.** Comparisons among the estimated daily DSR by BM, BM\_SA, and BM\_TL models over ground measurements at SURFRAD, BSRN, and GC-NET stations in 2013.

better results for those areas, but the improvements were limited. Over areas greater than 70°, the rRMSE of all three models dramatically increased, especially for BM\_SA. The difference between BM\_SA and BM was large in this case. BM\_TL still achieved better results than BM over high latitude areas, and the same conclusions were drawn from the comparisons of different SZA. BM\_SA, BM, and BM\_TL all showed similar performance when SZA was low. When SZA was greater than 60°, BM\_SA tended to fail as rRMSE increased. The rRMSE of BM increased as well, but the magnitude was small compared to BM\_SA. The rRMSE of BM\_TL drops in the middle of BM and BM\_SA for high SZA samples.

Fig. 7 shows the comparison of the three instantaneous models under different snow and cloud conditions. Both snow and cloud information are derived from CLDMSK\_L2\_VIIRS\_SNPP product. First, the performance of all three models varied with the testing data. For GCNET stations, regardless of conditions, BM\_SA had the highest rRMSE followed by BM. BM\_TL significantly improved this by more than 5% rRMSE under all conditions. For BSRN and SURFRAD stations, the three models had difference performances under different conditions. For non-snow samples, the rRMSE of all three models were similar. For snow-covered pixels, the rRMSE of all three models increased, but BM\_TL achieved the lowest rRMSE. For atmospheric conditions, the rRMSE of all models increase when it pertained to cloud over the BSRN and SURFRAD stations. When conditions were clear or probably clear, the three models present similar estimation accuracies. As it was probably cloudy or cloudy, BM showed the worst performance. This may

originate from the simulation data which did not consider the effect of AOD, multilayer cloud, or 3-dimensional cloud. This issue is addressed by BM\_TL through incorporating the ground measurements especially for completely cloudy conditions. However, BM\_TL still not function as well as BM\_SA for probably cloudy conditions. Overall, the results demonstrate that BM\_TL normally aligns well with the lower RMSE achieved by the two traditional models, leading to significant improvement under scarce atmospheric and surface conditions mostly over high latitudes and high SZAs areas compared with BM\_SA, and over cloudy cases compared with BM.

#### 4.3. Comparison with physical methods

In this study, the BM\_TL method was compared with two traditional physical methods, narrow-band LUT and direct estimation methods, using the same input data to ensure consistent comparisons. The detailed description of these two methods can be found in Wang et al. (2021). The coefficients of LUT and the direct estimation method were also retrieved from the same RTM simulated data. We kept all the available estimations for each method in the comparison. Fig. 8 and Fig. 9 show the validation results over SURFRAD, BSRN, and GC-NET stations for instantaneous and daily estimation, respectively. For instantaneous DSR estimation, the RMSE of LUT and the direct estimation method for SURFRAD stations were 140.0 and 143.2 W/m<sup>2</sup>, respectively, and for BSRN stations, they were 108.4 and 116.8 W/m<sup>2</sup>.

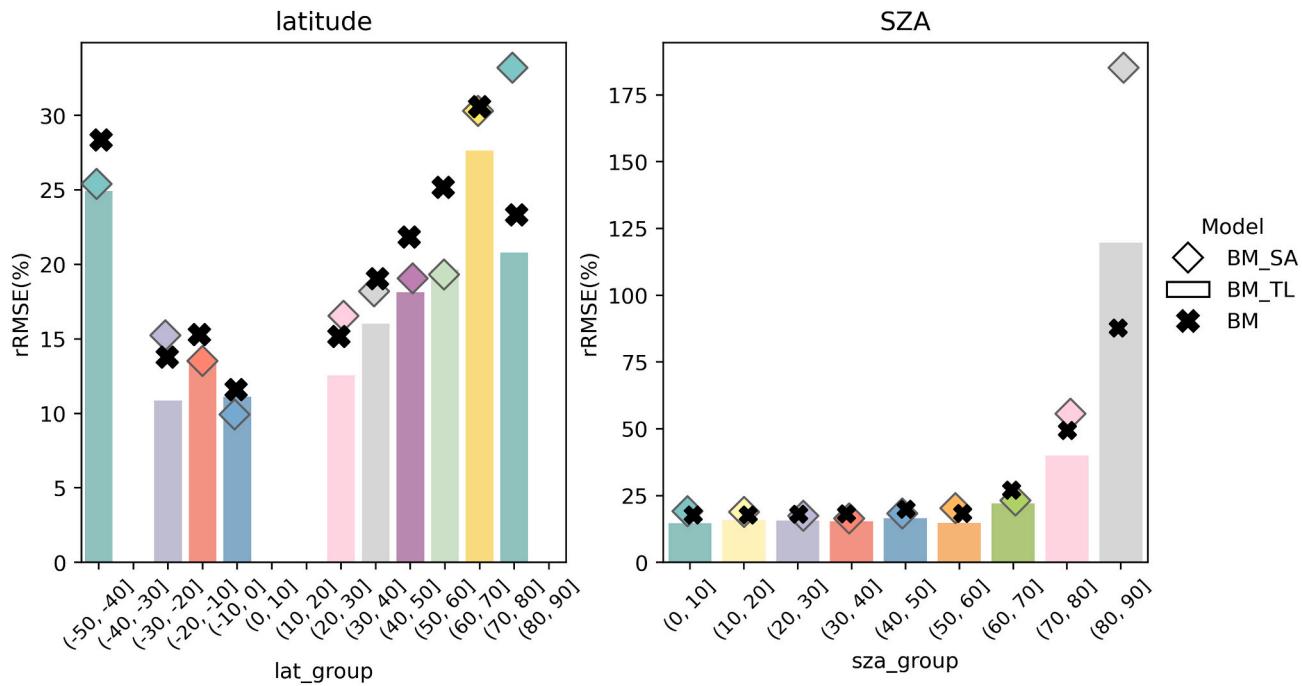


Fig. 6. Comparisons of the estimated instantaneous DSR by BM, BM\_SA, and BM\_TL models at different latitudes and SZA.

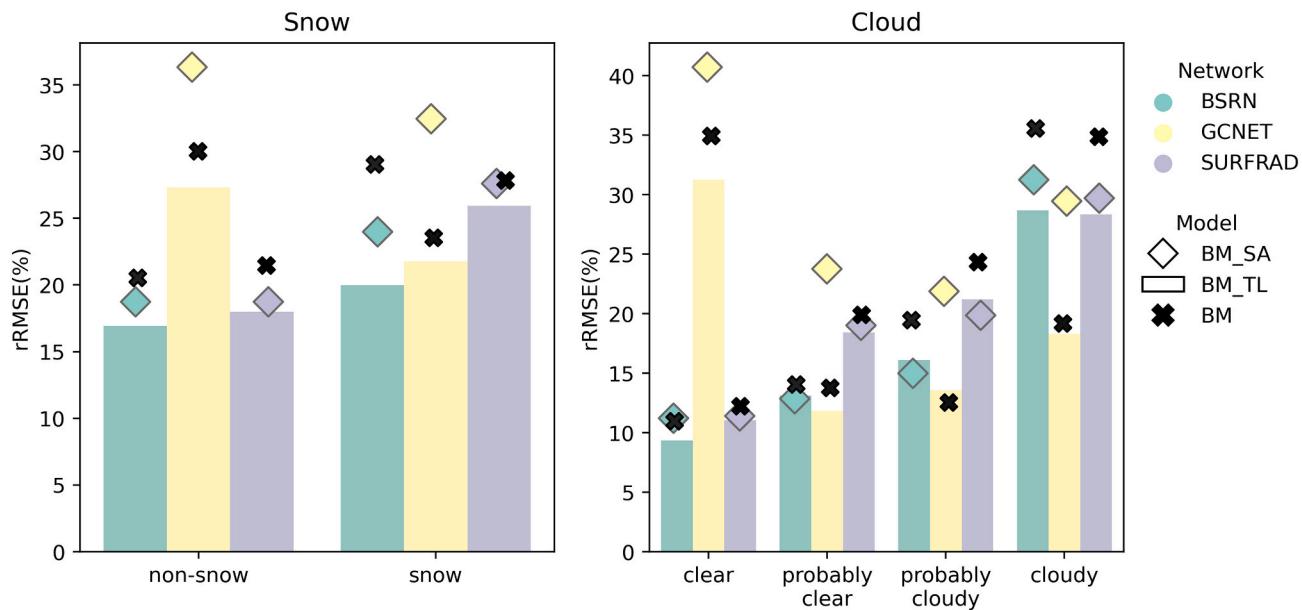
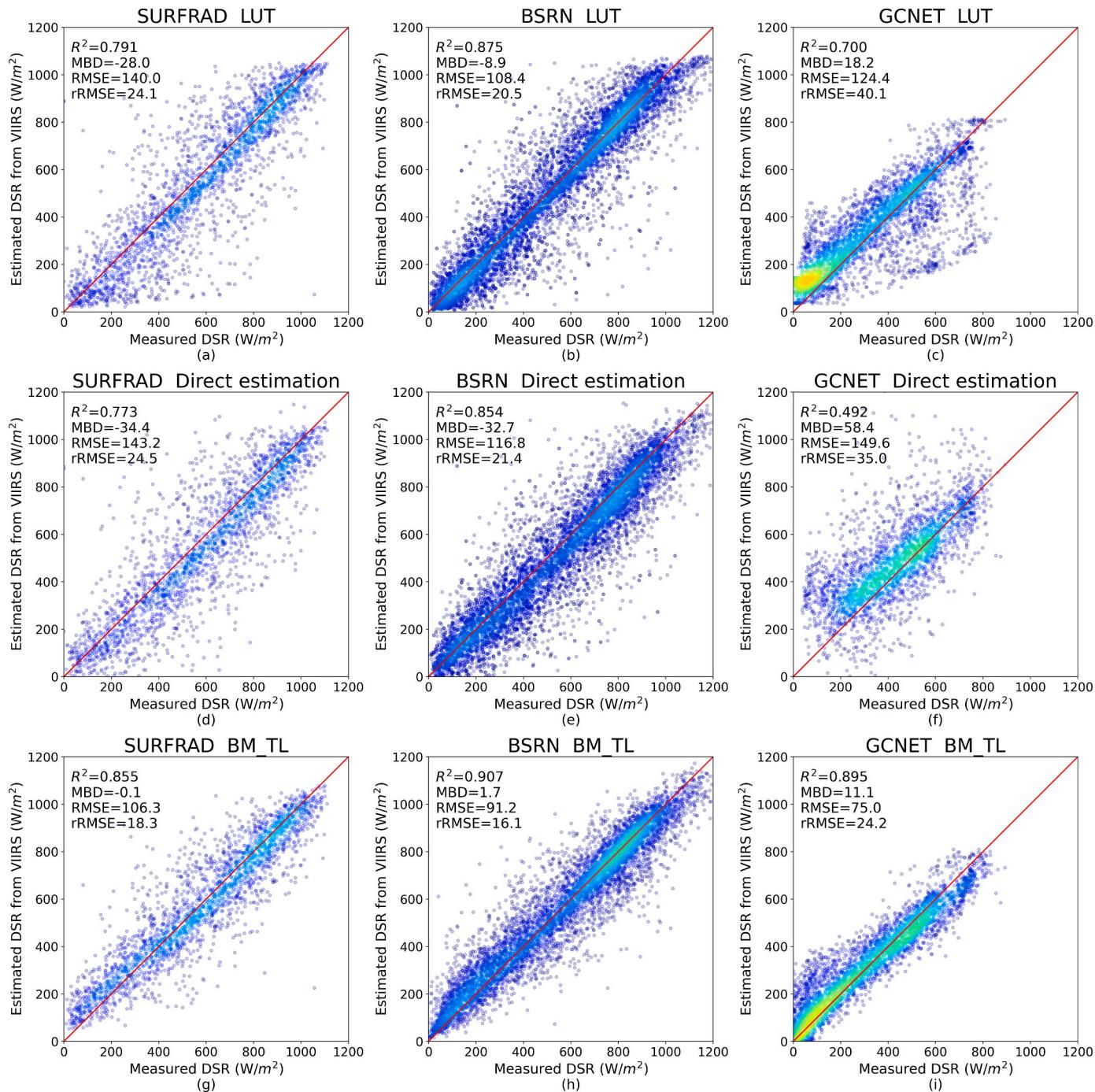


Fig. 7. Comparison between the estimated instantaneous DSR by BM, BM\_SA, and BM\_TL model at different snow and cloud conditions for each testing network.

respectively. The BM\_TL model achieved a much lower RMSE with 106.3 and 91.2  $W/m^2$  over SURFRAD and BSRNs. For daily DSR estimation, the RMSE of the LUT and the direct estimation method for SURFRAD stations were 48.6 and 44.9  $W/m^2$  respectively, and for BSRN stations, they were 43.5 and 42.9  $W/m^2$ , respectively. Again, the BM\_TL model achieved a lower RMSE with 30.8 and 31.3  $W/m^2$  over SURFRAD and BSRN, respectively. There are some possible reasons for the superior performance of BM\_TL. Though the three methods are based on the same simulation data, both LUT and the direct estimation method only use linear interpolation to calculate DSR, while BM\_TL can capture the nonlinear relationship. Furthermore, in addition to the simulation data, BM\_TL leverages information from the ground measurements, which can minimize the uncertainties that originate from the ignored conditions in

the simulation data.

The superiority of BM\_TL was also particular to the case of estimating over GC-NET stations. The RMSE of LUT and the direct estimation method instantaneous estimation were 124.4 and 149.6  $W/m^2$ , respectively. Invalid estimations of the direct estimation method exist at samples with SZA higher than 75 degrees which explains the higher RMSE while lower rRMSE compared with LUT. BM\_TL has a much lower RMSE with 75  $W/m^2$ (24.2%) and, notably,  $R^2$  was dramatically improved. For daily estimation, the RMSEs were 52.1 and 40.8  $W/m^2$ , respectively, for LUT and direct estimation methods, whereas BM\_TL achieved a RMSE of 38.0  $W/m^2$ . In addition to the above-mentioned reasons, the improvements of BM\_TL over LUT at GCNET stations are also due to the combination of multiple bands ranging from visible to



**Fig. 8.** Comparison between the estimated instantaneous DSR by LUT, Direct estimation, and BM\_TL model with ground measurements at SURFRAD, BSRN, GCNET stations.

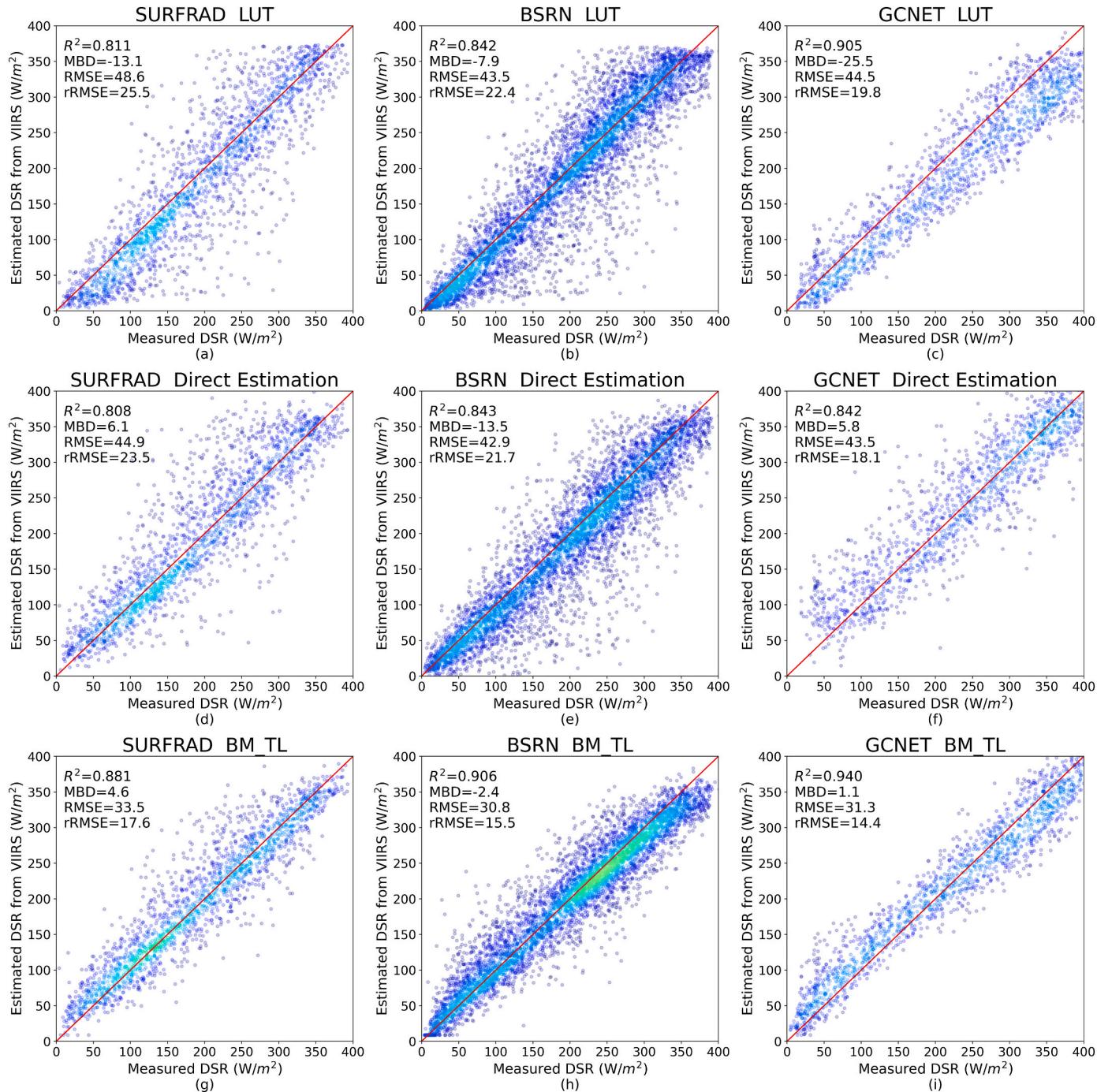
medium-wave infrared to better discriminate the snow from cloud. The validation results further reveal the dilemmas faced by traditional inversion methods can be improved by BM\_TL especially over high-latitude areas.

#### 4.4. Comparison with existed DSR products

The validation results of the proposed model were also compared with existing high resolution DSR products. Table 2 presents the comparison results with GLASS, BESS, and EPIC at an instantaneous time-scale. The statistics were collected from several recently published literatures (Hao et al., 2019; Ryu et al., 2018). The proposed method had a

significantly higher  $R^2$  and lower RSME and rRMSE than existed products. However, the results are not based on the same validation sites during the same time period, hence it only suggests the comparability between the products.

For the daily DSR estimation, we validated and compared BESS, GLASS, and MCD18 with our proposed method over the same sites from SURFRAD, BSRN and GCNET (Table 3). Invalid values were removed from the comparison. BM\_TL has a comparable RMSE and rRMSE over all stations, and is least biased compared to the other three products. Furthermore, it presents significantly better accuracies over the GCNET stations with the lowest RMSE and bias. Over the BSRN and SURFRAD sites, the advantages of the propose method are not pronounced. Some



**Fig. 9.** Comparison between the estimated daily DSR by LUT, Direct estimation, and BM\_TL model with ground measurements at SURFRAD, BSRN, GC-NET stations.

**Table 2**

Summary of comparisons between BM\_TL estimated instantaneous DSR with EPIC, GLASS, and BESS products.

Products	Time scale	R <sup>2</sup>	RMSE (W/m <sup>2</sup> )	Bias (W/m <sup>2</sup> )	rRMSE (%)	Reference
EPIC	Hourly	0.83	113.21	1.99	23.44	(Hao et al., 2019)
GLASS	Instantaneous	0.83	115	-6.5	-	(Ryu et al., 2018)
BESS	Instantaneous	0.84	123.3	-0.3	-	(Ryu et al., 2018)
BM_TL	Instantaneous	0.91	89.6	4.3	18.30	

possible reasons can be considered. Firstly, spatial resolution influences the DSR validation results (Li et al., 2021). The lower the spatial resolution is, the higher the validation accuracies are. BESS and GLASS are 5 km products and MCD18 are 1 km products, while the proposed method estimated daily DSR at 750 m. Secondly, the data source may influence the retrieval accuracies. BESS, GLASS, and MCD18 are all based on MODIS data while the proposed method is using newly VIIRS TOA data. BESS, GLASS, and MCD18 are mature products while the proposed method is still at the development stage. No adjustments procedures have been added for product generation.

**Table 3**

Summary of comparisons between BM\_TL estimated daily DSR with BESS, GLASS, and MCD18.

Network	Products	$R^2$	RMSE ( $W/m^2$ )	Bias ( $W/m^2$ )	rRMSE (%)
All	BESS	0.91	-3.2	32.0	15.4
	GLASS	0.83	-16.6	45.2	21.7
	MCD18	0.77	-19.4	53.9	25.9
	BM_TL	0.90	0.2	31.7	15.2
BSRN and SURFRAD	BESS	0.91	-1.9	30.6	15.2
	GLASS	0.93	-7.7	27.8	13.8
	MCD18	0.89	-7.8	35.1	17.4
	BM_TL	0.90	0.2	31.6	15.7
GCNET	BESS	0.89	-9.2	37.9	16.0
	GLASS	0.64	-57.1	88.4	37.2
	MCD18	0.55	-72.1	102.6	43.2
	BM_TL	0.92	-0.2	32.2	13.6

#### 4.5. Temporal analysis

Since the training data of the proposed models are completely from 2013, we also validated the models with measurements in 2014 over SURFRAD sites to test its capability for future inter-annual analysis. The results are shown in Fig. 10. The rRMSE are 17.5% and 17.1% respectively for instantaneous and daily estimation, and no pronounce bias are observed. It presents a similar accuracy for both instantaneous and daily estimation compared with the validation results over the same sites in 2013 which are 18.3% and 17.6% in terms of rRMSE (Fig. 4 and Fig. 5).

To evaluate the intra-annual stability of the proposed method, the rRMSE and rBias were calculated for each month over all validation sites. These results are shown in Fig. 11. Similar to previous analysis (Li et al., 2021), comparatively higher rRMSE were observed from November to March, and the lowest rRMSE was achieved from May to August. The difference between the highest and lowest rRMSE of each month was about 7%. In terms of rBias, the model tended to underestimate DSR from November to February, and overestimate DSR from August to October.

Previous analysis also suggests that the higher the latitude, the more obvious seasonal trend was in current DSR products (Li et al., 2021). To evaluate the temporal stability of the proposed method over high latitude areas, the daily variation of DSR over eight GCNET stations in 2013 are plotted in Fig. 12. Consistencies are observed between measured and estimated DSR. The variability and magnitude of DSR are captured in most cases. By adopting satellite observations, the proposed method can

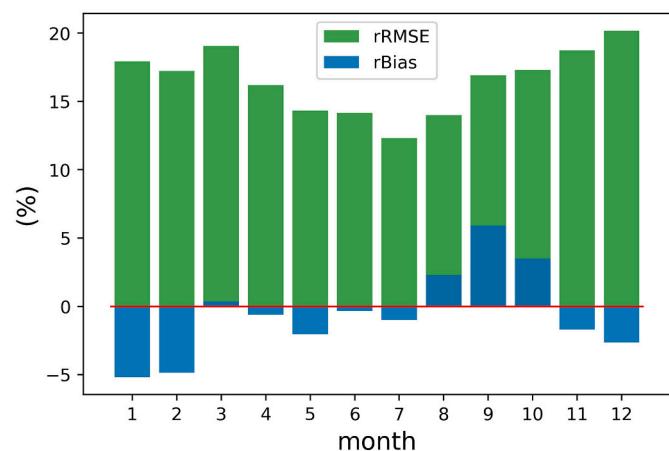


Fig. 11. The rRMSE and rBias of estimated DSR for each month.

also fill the gaps of ground measurements. For example, no measurements are available at Humboldt stations before May but continual estimations can be provided by the proposed method. The accurate retrieval of DSR over polar regions was challenging and limited even for ground measurements due to the severe weather; hence, the stable and accurate estimation of TL over high latitude areas will be of great importance in future study to understand polar climate.

#### 4.6. DSR mapping

BM\_TL was also applied to satellite imagery to demonstrate global effectiveness. Three VIIRS swaths over Greenland, Ural, and Middle East regions were chosen, where few stations are available to provide training data and where traditional DSR estimation methods degrade. A random day, June 6th, 2013, was chosen for both instantaneous and daily mapping.

Fig. 13 presents the estimated instantaneous DSR mapping. VIIRS red-green-blue (RGB) images are plotted as references for instantaneous estimation because the existence of clouds always corresponds to low DSR. To mask the cloud cover over snowy areas on RGB images, CLDMSK\_L2\_VIIRS\_SNPP data were also used. It shows that the BM\_TL based DSR estimations presented similar spatial patterns with cloud over all three regions. BM\_TL was also able to distinguish different cloud

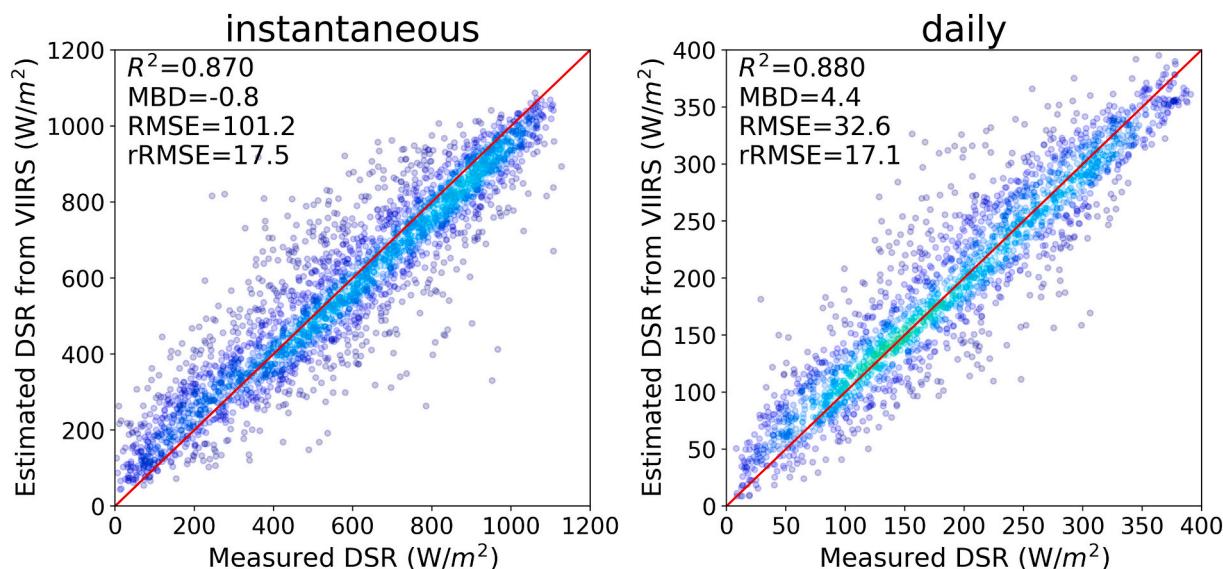
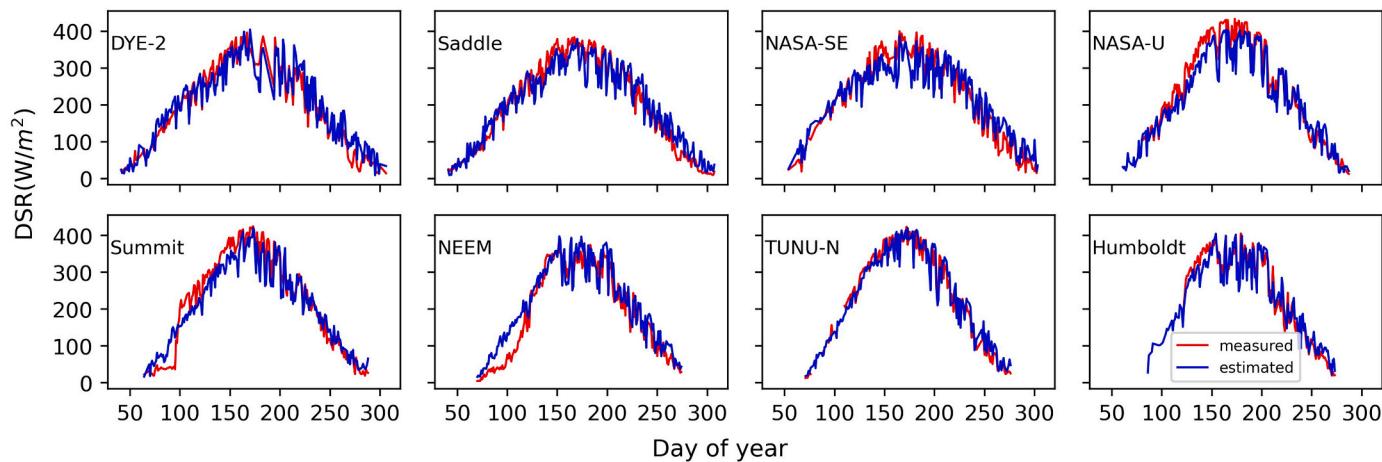
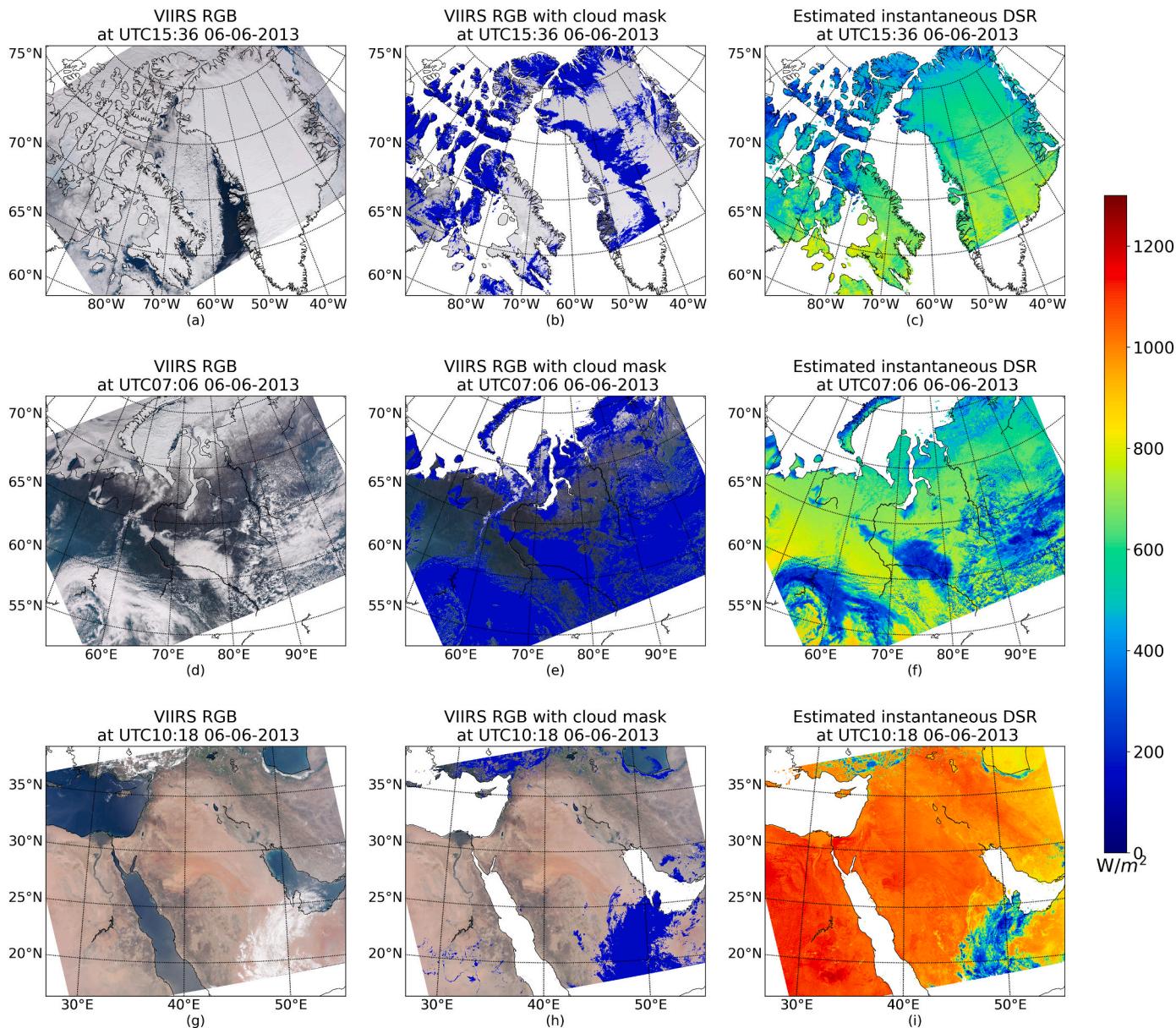


Fig. 10. Validation results of instantaneous and daily BM\_TL over SURFRAD stations in 2014.



**Fig. 12.** Daily variation of estimated and measured DSR over eight GCNET stations in 2013.



**Fig. 13.** Estimated instantaneous DSR and VIIRS RGB images and those with cloud-cover masks over Greenland, Ural, and Middle East areas on June 6th, 2013.

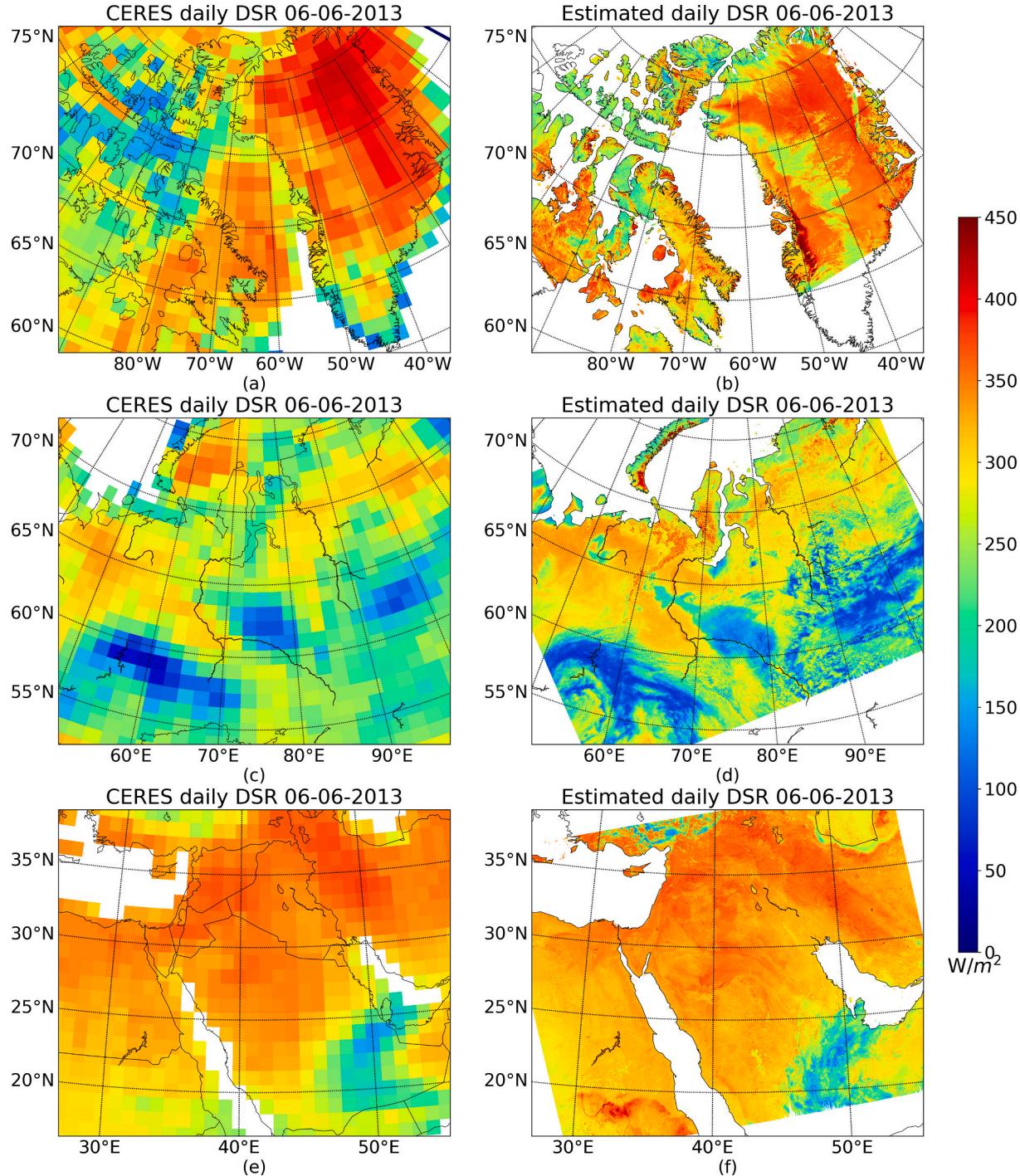
phase to estimate DSR values properly. For example, higher DSR were estimated over periphery areas than the middle area of the clouds in Middle East. The BM\_TL model also performed well over snow-covered regions. The cloud mask product revealed the presence of cloud in the middle of Greenland, which is hard to observe visually in the VIIRS RGB image. The proposed model successfully discriminated clouds from snow and estimated DSR properly. This agrees well with previous validation results over Greenland regions. Fig. 14 compares the estimated daily DSR mapping and corresponding CERES-SYN results over the same three areas on June 6, 2013. Similar spatial patterns and color scale were presented between the daily DSR estimated by BM\_TL and that retrieved from CERES-SYN. Moreover, owing to the high spatial resolution of

VIIRS, more detailed information could be observed from the BM\_TL maps. However, some differences persist, such as those over the west part of Greenland and those southwest of the Middle East.

#### 4.7. Impact of the training data size

As TL was first introduced to combine RTM simulated data and in-situ data, more analysis was performed to show how the size of training data influence the TL performance compared with traditional NN. We used instantaneous estimations as illustration here.

Six random subsets of the full data are employed by using a different number of stations (5, 10, 30, 50, 70, and 96) to train the BM\_SA and



**Fig. 14.** Estimated daily DSR and comparison with CERES-SYN daily DSR over Greenland, Ural, and Middle East areas on June 6th, 2013.

BM\_TL models. For each number, the stations were randomly chosen from all FLUXNET stations ten times to minimize the influence of training data representativeness. The models were tested over SURFRAD and GCNET stations. The same model structure and configuration were retained, except for the frozen layers and epoch number, which were justified separately for each training dataset. The results are presented in Fig. 15 and Table 4. Several conclusions can be drawn. First, a lower mean RMSE was observed for BM\_TL, especially when the training data size was smaller than 50 stations. For SURFRAD stations, the rRMSE of BM\_TL remains 17%–19% regardless of the training size while rRMSE of BM\_SA ranged from 19%–28%. The difference was more obvious when testing over GCNET stations; the rRMSE of BM\_SA reached 71.7% while BM\_TL achieved nearly half of BM\_SA's, at 36.6%, for GCNET sites when there was 5 training stations. Moreover, lower standard deviations (STDs) by 10 times and less outliers were observed for BM\_TL. Over both SURFRAD and GCNET stations, almost 0  $W/m^2$  STD was observed for BM\_TL when trained from all data. When training from small sample size, BM\_TL still has a low STD comparing with BM\_SA. The STD was lower than 6  $W/m^2$  for BM\_TL, while it reached 100  $W/m^2$  for BM\_SA over SURFRAD stations, and was lower than 34  $W/m^2$  for BM\_TL while it reached 62.6  $W/m^2$  for BM\_SA over GCNET stations. One possible explanation is that, unlike BM\_SA, BM\_TL will not be influenced by the random initial weights. Moreover, when randomly selecting data from the training data pool, it is possible that some stations are more representative and have better measurement quality. The results suggest that BM\_TL is less sensitive to the representativeness of training data. It is also noticeable that the difference between BM\_TL and BM\_SA diminishes as training data size increases, especially over SURFRAD stations. This is possible because, when including all of the training data from FLUXNET, the data has already covered sufficient conditions for the testing sites. The comparatively simple structure of the proposed method may also be another reason.

In summary, TL has superior results when facing limited training data compared with traditional NN. These superior performances include the improvement of the accuracy as well as the stability of the model. The advantages of TL are less pronounced when the training data size increases, but it still performs better over under-representative conditions, such as high latitude areas. These results indicate that with the knowledge learned from simulation data, TL can achieve increased performances even when the size of the training data is small. This characteristic will potentially help estimate more peculiar geographical variables when ground-measurement data are limited.

## 5. Conclusion

This study proposed an NN combined with TL to estimate global instantaneous and daily DSR from VIIRS TOA reflectance sources. A four-layer NN model was constructed and first trained on RTM simulated

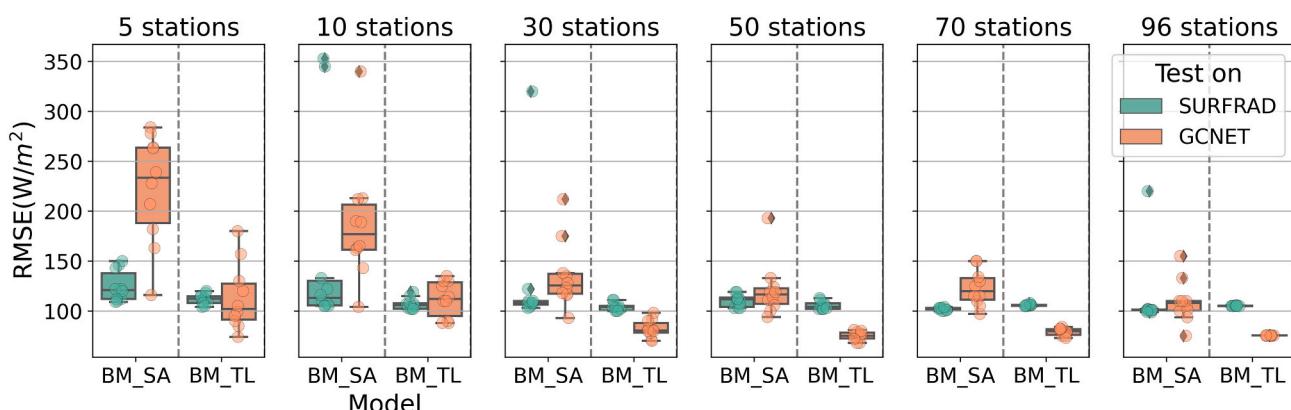
data. It transfers the knowledge learned from the simulated data to the real data by freezing one layer of the model and re-training it on FLUXNET measurements. The results were independently validated over seven SURFRAD stations, 25 BSRN stations, and eight GC-NET stations. For instantaneous estimation, the RMSEs were 106.3 (18.3%), 91.2 (16.1%), 75.0 (24.2%)  $W/m^2$ , respectively. For daily estimation, the RMSEs were 30.8 (15.5%), 33.5 (17.6%), and 31.3 (14.4)  $W/m^2$ , respectively. The results were also compared with traditional ML algorithms and physical algorithms, and presented significant improvements, especially for GC-NET data. The proposed method was also compared with existing DSR products and showed comparable accuracies over BSRN and SURFRAD stations and better results over GC-NET stations.

There are several advantages of the proposed algorithm. It only requires TOA and surface albedo as inputs, which avoids uncertainties inherently brought by other high-level products and enables retrievals at VIIRS raw resolutions (~ 750 m). TL is a bridge that combines the strengths of models trained by physical simulations and real-word data. Generality and accuracy were both improved, which enables a global application. The method was applied on three VIIRS swaths where no training data were collected. The instantaneous mapping captured the spatial pattern of the clouds, and the daily estimation corresponded well to the daily results of CERES-SYN. BM\_TL also presents high accuracies over polar regions where some existing methods degrade. Lastly, BM\_TL revealed the potential of training a model that mitigates the restrictions of training data size. BM\_TL significantly improved the accuracy when limited stations were used as training data. For geographical science studies where both physical models and ground measurements exist, TL revealed unique values when combining model simulations with real-world data.

The current algorithm still has room for further improvement. For example, the model did not consider topography as an explicit input which has been noted by previous studies that influences DSR estimation dramatically through shadowing and multi-scattering (Hao et al., 2018; Proy et al., 1989). Due to the data limitation, BSA at noon rather than the blue sky albedo is used as input in this study. Nevertheless, invalid estimations still exist in the proposed method due to the lack of different surface albedo bands as inputs. Future studies can incorporate blue sky albedo and climatology albedo data or train the model with less albedo bands to address the issues. In this study, a simple four-layers NN was employed with relatively sufficient training data. The potential of TL may not be fully leveraged with all available training data as discussed in Section 4.7. The combination of TL with a deeper and more sophisticated NN model can potentially deliver further improved performance.

## Author credit statement

D. Wang and R. Li conceptualized the research. R. Li and D. Wang



**Fig. 15.** Comparison between BM\_TL, BM\_SA models trained on 5, 10, 30, 50, 70, and all (96) FLUXNET stations.

**Table 4**

Evaluation of BM\_SA and BM\_TL trained on 10, 30, and 50 FLUXNET stations ten times over SURFRAD and GC-NET stations.

N	Model	SURFRAD RMSE (W/m <sup>2</sup> )	SURFRAD rRMSE (%)	SURFRAD STD (W/m <sup>2</sup> )	GCNET RMSE (W/m <sup>2</sup> )	GCNET rRMSE (%)	GCNET STD (W/m <sup>2</sup> )
5	BM_SA	124.6	21.5	15.8	222.4	71.7	55.1
	BM_TL	111.7	19.2	5.4	113.5	36.6	33.6
10	BM_SA	160.2	27.6	100.0	188.0	60.6	62.6
	BM_TL	107.0	18.4	5.8	112.0	36.1	18.2
30	BM_SA	129.9	22.4	67.0	135.5	43.7	34.1
	BM_TL	104.2	17.9	4.1	82.2	26.5	8.9
50	BM_SA	110.3	19.0	6.2	121.3	39.1	27.7
	BM_TL	105.5	18.2	4.1	74.9	24.1	4.6
70	BM_SA	101.9	17.5	1.4	123.1	39.7	18.0
	BM_TL	105.8	18.2	0.9	79.0	25.5	3.6
96	BM_SA	112.6	19.4	37.8	109.4	35.3	21.7
	BM_TL	105.1	18.1	0.3	75.4	24.3	0.2

processed and analyzed the data. R. Li, D. Wang and S. Liang performed the interpretation of the results. R. Li led the manuscript writing. All authors contributed to the manuscript revision.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgements

This study is supported by NASA (Grant ID: 80NSSC18K0620 and 80NSSC21K0701). We thank the four reviewers for their valuable comments and suggestions.

### References

- Augustine, J.A., DeLuise, J.J., Long, C.N., 2000. SURFRAD—A national surface radiation budget network for atmospheric research. *Bull. Am. Meteorol. Soc.* 81 (10), 2341–2358.
- Babar, B., Luppino, L.T., Boström, T., Anfinsen, S.N., 2020. Random forest regression for improved mapping of solar irradiance at high latitudes. *Sol. Energy* 198 (1), 81–92. <https://doi.org/10.1016/j.solener.2020.01.034>.
- Baldridge, A.M., Hook, S.J., Grove, C.I., Rivera, G., 2009. The aster spectral library version 2.0. *Remote Sens. Environ.* 113 (4), 711–715. <https://doi.org/10.1016/j.rse.2008.11.007>.
- Berk, A., Cooley, T.W., Anderson, G.P., Acharya, P.K., Bernstein, L.S., Muratov, L., Lewis, P.E., 2004. MODTRAN5: A Reformulated Atmospheric Band Model with Auxiliary Species and Practical Multiple Scattering Options. In: *Remote Sensing of Clouds and the Atmosphere IX*, vol. 5571. International Society for Optics and Photonics, pp. 78–85.
- Brown, M.G.L., Skakun, S., He, T., Liang, S., 2020. Intercomparison of machine-learning methods for estimating surface shortwave and photosynthetically active radiation. *Remote Sens.* 12 (3), 372. <https://doi.org/10.3390/rs12030372>.
- Bue, B.D., Thompson, D.R., Deshpande, S., Eastwood, M., Green, R.O., Natraj, V., Mullen, T., Parente, M., 2019. Neural network radiative transfer for imaging spectroscopy. *Atmo. Measur. Techn.* 12 (4), 2567–2578. <https://doi.org/10.5194/amt-12-2567-2019>.
- Chen, L., Yan, G., Wang, T., Ren, H., Calbó, J., Zhao, J., McKenzie, R., 2012. Estimation of surface shortwave radiation components under all sky conditions: modeling and sensitivity analysis. *Remote Sens. Environ.* 123, 457–469.
- Chen, L., Yan, G., Wang, T., Ren, H., Hu, R., Chen, S., Zhou, H., 2019. Spatial scale consideration for estimating all-sky surface shortwave radiation with a modified 1-d radiative transfer model. *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.* 12 (3), 821–835. <https://doi.org/10.1109/JSTARS.2019.2896644>.
- Clark, R.N., Swazyne, G.A., Wise, R.A., Livo, K.E., Hoefen, T.M., Kokaly, R.F., Sutley, S.J., 2007. USGS Digital Spectral Library splib06a (Data Series). US Geological Survey. <https://doi.org/10.3133/ds231>.
- Hao, D., Wen, J., Xiao, Q., Wu, S., Lin, X., You, D., Tang, Y., 2018. Impacts of DEM geolocation bias on downward surface shortwave radiation estimation over clear-sky rugged terrain: a case study in Dayekou Basin, China. *IEEE Geosci. Remote Sens. Lett.* 16 (1), 10–14.
- Hao, D., Asrar, G.R., Zeng, Y., Zhu, Q., Wen, J., Xiao, Q., Chen, M., 2019. Estimating hourly land surface downward shortwave and photosynthetically active radiation from DSCOVR/EPIC observations. *Remote Sens. Environ.* 232, 111320.
- Hinton, G., Srivastava, N., Swersky, K., 2012. Lecture 6a Overview of Mini-Batch Gradient Descent. Coursera Lecture Slides. [Online]. Available. <https://class.coursera.org/neuralnets-2012-001/lecture>.
- Hou, N., Zhang, X., Zhang, W., Wei, Y., Jia, K., Yao, Y., Jiang, B., Cheng, J., 2020. Estimation of surface downward shortwave radiation over China from himawari-8 ahi data based on random forest. *Remote Sens.* 12 (1), 181. <https://doi.org/10.3390/rs12010181>.
- Huang, B., Zhao, B., Song, Y., 2018. Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery. *Remote Sens. Environ.* 214, 73–86.
- Huang, G., Li, Z., Li, X., Liang, S., Yang, K., Wang, D., Zhang, Y., 2019. Estimating surface solar irradiance from satellites: past, present, and future perspectives. *Remote Sens. Environ.* 233, 111371. <https://doi.org/10.1016/j.rse.2019.111371>.
- Jia, A., Ma, H., Liang, S., Wang, D., 2021. Cloudy-sky land surface temperature from VIIRS and MODIS satellite data using a surface energy balance-based method. *Remote Sens. Environ.* 263, 112566.
- Jiang, H., Lu, N., Qin, J., Tang, W., Yao, L., 2019. A deep learning algorithm to estimate hourly global solar radiation from geostationary satellite data. *Renew. Sust. Energ. Rev.* 114, 109327. <https://doi.org/10.1016/j.rser.2019.109327>.
- Khaki, S., Pham, H., Wang, L., 2021. Simultaneous corn and soybean yield prediction from remote sensing data using deep transfer learning. *Sci. Rep.* 11 (1), 11132. <https://doi.org/10.1038/s41598-021-89779-z>.
- Li, L., Xin, X., Zhang, H., Yu, J., Liu, Q., Yu, S., Wen, J., 2015. A method for estimating hourly photosynthetically active radiation (PAR) in China by combining geostationary and polar-orbiting satellite data. *Remote Sens. Environ.* 165, 14–26.
- Li, R., Wang, D., Liang, S., 2021. Comprehensive assessment of five global daily downward shortwave radiation satellite products. *Sci. Remote Sens.* 4, 100028.
- Liang, S., Zheng, T., Liu, R., Fang, H., Tsay, S.-C., Running, S., 2006. Estimation of incident photosynthetically active radiation from moderate resolution imaging spectrometer data. *J. Geophys. Res.* 111 (D15), 121. <https://doi.org/10.1029/2005JD006730>.
- Liang, S., Wang, D., He, T., Yu, Y., 2019. Remote sensing of earth's energy budget: synthesis and review. *Int. J. Digital Earth* 12 (7), 737–780. <https://doi.org/10.1080/17538947.2019.1597189>.
- Ma, R., Letu, H., Yang, K., Wang, T., Shi, C., Xu, J., Shi, J., Shi, C., Chen, L., 2020. Estimation of surface shortwave radiation from himawari-8 satellite data based on a combination of radiative transfer and deep neural network. *IEEE Trans. Geosci. Remote Sens.* 58 (8), 5304–5316. <https://doi.org/10.1109/TGRS.2019.2963262>.
- Malmgren-Hansen, D., Kusk, A., Dall, J., Nielsen, A.A., Engholm, R., Skriver, H., 2017. Improving Sar automatic target recognition models with transfer learning from simulated data. *IEEE Geosci. Remote Sens. Lett.* 14 (9), 1484–1488. <https://doi.org/10.1109/LGRS.2017.2717486>.
- Mueller, R.W., Matsoukas, C., Gratzki, A., Behr, H.D., Hollmann, R., 2009. The cm-saf operational scheme for the satellite based retrieval of solar surface irradiance — a lut based eigenvector hybrid approach. *Remote Sens. Environ.* 113 (5), 1012–1024. <https://doi.org/10.1016/j.rse.2009.01.012>.
- Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted boltzmann machines. In: *Proceedings of the 27th International Conference on Machine Learning*, pp. 807–814.
- Pan, S.J., Yang, Q., 2010. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22 (10), 1345–1359. <https://doi.org/10.1109/TKDE.2009.191>.
- Peng, Z., Letu, H., Wang, T., Shi, C., Zhao, C., Tana, G., Zhao, N., Dai, T., Tang, R., Shang, H., Shi, J., Chen, L., 2020. Estimation of shortwave solar radiation using the artificial neural network from himawari-8 satellite imagery over China. *J. Quant. Spectrosc. Radiat. Transf.* 240, 106672. <https://doi.org/10.1016/j.jqsrt.2019.106672>.
- Proy, C., Tanre, D., Deschamps, P.Y., 1989. Evaluation of topographic effects in remotely sensed data. *Remote Sens. Environ.* 30 (1), 21–32.
- Rutan, D.A., Kato, S., Doelling, D.R., Rose, F.G., Le Nguyen, T., Caldwell, T.E., Loeb, N. G., 2015. Ceres synoptic product: methodology and validation of surface radiant flux. *J. Atmos. Ocean. Technol.* 32 (6), 1121–1143. <https://doi.org/10.1175/JTECH-D-14-00165.1>.
- Ryu, Y., Jiang, C., Kobayashi, H., Detto, M., 2018. Modis-derived global land products of shortwave radiation and diffuse and total photosynthetically active radiation at 5 km resolution from 2000. *Remote Sens. Environ.* 204 (D09104), 812–825. <https://doi.org/10.1016/j.rse.2017.09.021>.

- Şenkal, O., 2010. Modeling of solar radiation using remote sensing and artificial neural network in Turkey. *Energy* 35 (12), 4795–4801. <https://doi.org/10.1016/j.energy.2010.09.009>.
- Shi, H., Xiao, Z., Tian, X., 2019. Exploration of machine learning techniques in emulating a coupled soil–canopy–atmosphere radiative transfer model for multi-parameter estimation from satellite observations. *IEEE Trans. Geosci. Remote Sens.* 57 (11), 8522–8533. <https://doi.org/10.1109/TGRS.2019.2921392>.
- Sobol, I.M., 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simul.* 55 (1–3), 271–280.
- Takenaka, H., Nakajima, T.Y., Higurashi, A., Higuchi, A., Takamura, T., Pinker, R.T., Nakajima, T., 2011. Estimation of solar radiation using a neural network based on radiative transfer. *J. Geophys. Res.* 116 (D8), 2003. <https://doi.org/10.1029/2009JD013337>.
- Tong, X.Y., Xia, G.S., Lu, Q., Shen, H., Li, S., You, S., Zhang, L., 2020. Land-cover classification with high-resolution remote sensing images using transferable deep models. *Remote Sens. Environ.* 237, 111322.
- Wang, T., Yan, G., Chen, L., 2012. Consistent retrieval methods to estimate land surface shortwave and longwave radiative flux components under clear-sky conditions. *Remote Sens. Environ.* 124, 61–71. <https://doi.org/10.1016/j.rse.2012.04.026>.
- Wang, D., Liang, S., He, T., Shi, Q., 2015. Estimation of daily surface shortwave net radiation from the combined modis data. *IEEE Trans. Geosci. Remote Sens.* 53 (10), 5519–5529. <https://doi.org/10.1109/TGRS.2015.2424716>.
- Wang, D., Liang, S., Zhang, Y., Gao, X., Brown, M.G.L., Jia, A., 2020. A new set of modis land products (mcd18): downward shortwave radiation and photosynthetically active radiation. *Remote Sens.* 12 (1), 168. <https://doi.org/10.3390/rs12010168>.
- Wang, D., Liang, S., Li, R., Jia, A., 2021. A synergic study on estimating surface downward shortwave radiation from satellite data. *Remote Sens. Environ.* 264, 112639 <https://doi.org/10.1016/j.rse.2021.112639>.
- Wei, Y., Zheng, Y., Yang, Q., 2016. Transfer knowledge between cities. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1905–1914.
- Wei, Y., Zhang, X., Hou, N., Zhang, W., Jia, K., Yao, Y., 2019. Estimation of surface downward shortwave radiation over China from avhrr data based on four machine learning methods. *Sol. Energy* 177 (10), 32–46. <https://doi.org/10.1016/j.solener.2018.11.008>.
- Wu, W., Epstein, H.E., Guo, H., Li, X., Gong, C., 2021. A pigment ratio index based on remotely sensed reflectance provides the potential for universal gross primary production estimation. *Environ. Res. Lett.* 16 (5), 054065.
- Yang, L., Zhang, X., Liang, S., Yao, Y., Jia, K., Jia, A., 2018. Estimating surface downward shortwave radiation over China based on the gradient boosting decision tree method. *Remote Sens.* 10 (2), 185. <https://doi.org/10.3390/rs10020185>.
- Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How Transferable Are Features in Deep Neural Networks?
- Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., Zhang, L., 2020. Deep learning in environmental remote sensing: achievements and challenges. *Remote Sens. Environ.* 241 (11), 111716 <https://doi.org/10.1016/j.rse.2020.111716>.
- Zhang, X., Liang, S., Zhou, G., Wu, H., Zhao, X., 2014. Generating global land surface satellite incident shortwave radiation and photosynthetically active radiation products from multiple satellite data. *Remote Sens. Environ.* 152, 318–332. <https://doi.org/10.1016/j.rse.2014.07.003>.
- Zhang, X., Liang, S., Wild, M., Jiang, B., 2015. Analysis of surface incident shortwave radiation from four satellite products. *Remote Sens. Environ.* 165 (Suppl. 1), 186–202. <https://doi.org/10.1016/j.rse.2015.05.015>.
- Zhang, Y., He, T., Liang, S., Wang, D., Yu, Y., 2018. Estimation of all-sky instantaneous surface incident shortwave radiation from moderate resolution imaging spectroradiometer data using optimization method. *Remote Sens. Environ.* 209 (24), 468–479. <https://doi.org/10.1016/j.rse.2018.02.052>.
- Zhang, X., Zhao, X., Li, W., Liang, S., Wang, D., Liu, Q., Yao, Y., Jia, K., He, T., Jiang, B., Wei, Y., Ma, H., 2019. An operational approach for generating the global land surface downward shortwave radiation product from MODIS data. *IEEE Trans. Geosci. Remote Sens.* 57 (7), 4636–4650. <https://doi.org/10.1109/TGRS.2019.2891945>.
- Zhou, C., Zhang, J., Liu, J., Zhang, C., Shi, G., Hu, J., 2020. Bayesian transfer learning for object detection in optical remote sensing images. *IEEE Trans. Geosci. Remote Sens.* 58 (11), 7705–7719. <https://doi.org/10.1109/TGRS.2020.2983201>.