

Comparison between deep learning architectures for the 1 km, 10/15-min estimation of downward shortwave radiation from AHI and ABI



Ruohan Li^a, Dongdong Wang^{a,*}, Shunlin Liang^b

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

^b Department of Geography, University of Hong Kong, Hong Kong 999077, China

ARTICLE INFO

Edited by Jing M. Chen

Keywords:
 Deep learning
 Shortwave radiation
 Solar energy
 Solar radiation
 Geostationary
 ABI
 AHI
 CNN
 GRU
 DeNET

ABSTRACT

The retrieval of downward shortwave radiation (DSR) with high spatiotemporal resolution and short latency is critical. It is the fundamental driving force of surface energy, carbon, and hydrological circulations, and a key energy source for photovoltaic electricity. However, existing methods face significant challenges owing to cloud heterogeneity and their reliance on other satellite-derived products, which hinder the retrieval of accurate and timely DSR with high spatiotemporal resolution. In addition to the spectral features used in traditional approaches, deep learning (DL) can incorporate the spatial and temporal features of satellite data. This study developed and compared three DL methods, namely the DenseNet, the bidirectional gated recurrent unit without surface albedo as inputs ($\text{BiGRU}_{\text{nor}}$), and the convolutional neural network with gated recurrent unit without surface albedo as inputs ($\text{CNNGRU}_{\text{nor}}$). These methods were used to estimate DSR at 1 km and 10/15 min resolutions directly from top-of-atmosphere reflectance over the Advanced Himawari Imager (AHI) onboard Himawari-8 and the Advanced Baseline Imager (ABI) onboard GOES-16 coverage, achieving high accuracies. The instantaneous root mean square error (RMSE) and relative RMSE for the three models were 68.4 (16.1%), 69.4 (16.3%), and 67.1 (15.7%) W/m^2 , respectively, which are lower than the baseline machine learning method, the multilayer perceptron model (MLP), with RMSE at 76.8 W/m^2 (18.0%). Hourly accuracies for the three DL methods were 58.6 (14.1%), 57.8 (14.0%), and 57.3 (13.8%) W/m^2 , which are within the DSR RMSEs that we estimated for existing datasets of the Earth's Radiant Energy System (CERES) (88.8 W/m^2 , 21.4%) and GeoNEX (77.8 W/m^2 , 18.8%). The study illustrates that DL models that incorporate temporal information can eliminate the need for surface albedo as an input, which is crucial for timely monitoring and nowcasting of DSR. Incorporating spatial information can enhance retrieval accuracy in overcast conditions, and incorporating infrared bands can further improve the accuracy of DSR estimation.

1. Introduction

Recent years have witnessed increasing demand for high spatiotemporal surface downward shortwave radiation (DSR) with high accuracies (Liang et al., 2010; Liang et al., 2019; Stephens et al., 2012; Liang et al., 2019). As DSR is the main input for surface ecological, hydrological, and biogeochemical models, high spatiotemporal resolution is a prerequisite for estimating many surface variables at a fine scale, such as hourly ground-level ozone (Wei et al., 2022), land surface temperature (Jia et al., 2022b; Jia et al., 2023), and evapotranspiration (Huang et al., 2019). DSR with high spatiotemporal resolution is also critical for monitoring photovoltaic (PV) output and the economic and stable operation of the solar grid (Buster et al., 2021). Short-term

fluctuations in the DSR can adversely affect power grid quality (Gandhi et al., 2020). Power system models, such as the Global Solar Energy Estimator (Pfenninger and Staffell, 2016) and the Integrated Grid Modeling System (Palmintier, 2015), require hourly DSR data as input. Analysis of storage requirements for large grid-connected PV plants requires sub-hour or even minute-by-minute DSR data (Marcos et al., 2014; Nguyen et al., 2016). The World Meteorological Organization (WMO) suggests that a 10-min temporal resolution would be a breakthrough for DSR nowcasting and very short-range forecasting (Huang et al., 2019). In addition, coarse spatial resolution DSR cannot accommodate the current trend of extensive small-distributed PV deployment, particularly at the residential scale (Jain et al., 2017). Reportedly, issues of spatial heterogeneity for DSR exist on an hourly

* Corresponding author at: Department of Geographical Sciences, University of Maryland, 4600 River Road, College Park, MD 20737, USA.
 E-mail address: ddwang@umd.edu (D. Wang).

Table 1

The Summary of input variables and corresponding datasets.

Variables	Dataset	Dimension	Spatial resolution	Temporal resolution	References
TOA reflectance	GOES16/ HIMAWARI8 Level 1G	15 (bands)	1 km	10 min or 15 min	(Wang et al., 2020b)
Geometry angles	GOES16/ HIMAWARI8 Level 1G	3 (angles)	1 km	10 min or 15 min	(Wang et al., 2020b)
Elevation	GTOPO30	1	30 arc seconds	Static	(Earth Resources Observation And Science (EROS), 2017)
Surface albedo	MCD43A3	7 (bands)	500 m	Daily	(Schaaf and Wang, 2015)
Surface albedo (climatology)	Blue-sky Actual and Snow-free Albedo Climatology	7 (bands)	1 km	Daily	(Jia et al., 2022a)

scale and are more pronounced in mountainous regions as the spatial scale increases (Li et al., 2023). High accuracy is also demanded by DSR nowcasting (Huang et al., 2019). However, achieving accurate retrievals at high temporal and spatial resolutions remains a challenging topic.

Accurately estimating DSR with high spatiotemporal resolution currently presents various challenges and limitations. Existing DSR retrieval methods rely on high-level satellite-derived products that require complex processing algorithms for data extraction from raw satellite observations (Zhang et al., 2020). Although parameterization of atmospheric radiative transfer (RT) is an accurate and stable method, its performance still heavily relies on the choice of satellite-derived products as inputs, limiting its adaptability to high spatial and temporal resolution sensors (Van Laake and Sanchez-Azofeifa, 2004; Huang et al., 2018; Tang et al., 2019; Wang et al., 2021). Lookup tables (LUTs) are also physics-based methods that invert the DSR from satellite top-of-atmosphere (TOA) reflectance (Pinker et al., 1995; Liang et al., 2006; Wang et al., 2020a). Owing to their simplicity, they have been widely applied to estimate DSR from geostationary sensors; however, even LUTs pose challenges owing to their reliance on various satellite-derived products. Zhang et al. (2021) employed Moderate Resolution Imaging Spectroradiometer (MODIS) aerosol optical depth (AOD), albedo, and GOES-16 Clear Sky Mask products to the LUT to estimate hourly photosynthetically activated radiation (PAR). Li et al. (2015) used MODIS AOD, water vapor, ozone, surface albedo, and snow cover products to estimate PAR over China. Letu et al. (2022) also required a series of Himawari8 atmosphere products and ERA5 reanalysis data when estimating surface radiation from the Advanced Himawari Imager (AHI). The production of GeoNEX is comparatively less dependent on the high-level products, but still requires MODIS daily surface albedo product (MCD43) as inputs (Li et al., 2023). Some machine learning (ML) methods also estimate DSR forwardly similar to parameterization method and have been applied to generate DSR products. Hao et al. (2019) included MCD43 and products from a series of the Earth Poly-chromatic Imaging Camera (EPIC) onboard the Deep Space Climate Observatory (DSCOVR) to generate global hourly DSR and PAR products using random forest. Ryu et al. (2018) produced Breathing Earth System Simulator DSR using a neural network based on a series of MODIS products. The resolutions of high-level products restrict the retrieval of DSR with high spatiotemporal resolutions and prolong the latency for in time DSR estimation and nowcasting. With the increasing analyses on current solar energy supply and demand (Tong et al., 2021), storage (Gandhi et al., 2020), and solar power prediction (Bansal et al., 2022), developing methods that rely less on input data and provide more timely information on DSR is critical.

Cloud complexity has been identified as another factor constraining the enhancement of DSR estimation accuracy (Ham et al., 2014; Huang et al., 2019; Wyser et al., 2005). Current products and retrieval algorithms have primarily been developed from one-dimensional RT theory,

which is based on the assumption of independent pixel approximation (IPA) owing to its relatively high efficiency. It ignores the three-dimensional (3D) effect of clouds including the nonlocal cloud shadows and photon diffusion from neighboring columns, which works well for large pixel retrieval but is inadequate for high-resolution products (Chen et al., 2019). Wang et al. (2017) demonstrated that shadowing effect brought by the solar-cloud-satellite geometry may lead to up to 80% errors in DSR estimation. The errors are more prominent at high temporal resolutions, such as instantaneous estimation, and indirectly influence current global daily high-spatial resolution products extrapolated from instantaneous estimations. With the increasing demand for DSR with high spatial and temporal resolution, the complex effect of clouds on DSR retrieval can no longer be ignored.

Deep learning (DL) can extract spatial and temporal features on top of spectral information from satellite data by incorporating specific structures, such as convolutional layers or forget gates, presenting a potential solution to address the challenges posed by cloud heterogeneity and the reliance on high-level products when estimating DSR. The use of DL in DSR is still at its infancy. Few studies have utilized DL for estimating DSR, particularly with both spatial and temporal structures, while most have focused on a single DL model for accuracy improvement. Jiang et al. (2019) employed residual convolutional neural network (CNN) models to estimate DSR over mainland China. Zhang et al. (2022) used a spectral-wise CNN to extract the relationship between different bands for DSR estimation from the Visible Infrared Imager-Radiometer Suite. Although certain studies have applied recurrent neural network (RNN) to solar energy forecasting (Sønderby et al., 2020; Bansal et al., 2022), their estimation capability has not been investigated.

The current study developed and compared three DL methods that accurately estimate 1 km DSR from AHI onboard Himawari-8 with 10-min frequency and the Advanced Baseline Imager (ABI) onboard GOES-16 with 15-min frequency. The models in this study were designed with minimal inputs. The DenseNET model requires both TOA reflectance and surface albedo as inputs, whereas the bidirectional gated recurrent unit (BiGRU) and CNN with gated recurrent unit (CNNGRU) models only require TOA reflectance, henceforth referred as BiGRU_{nor} and CNNGRU_{nor} models, respectively. These three DL models with spatial, temporal, or combined structures were compared and analyzed for their accuracy under different temporal resolutions, model efficiency, and ability to address issues related to dependency on high-level products and cloud effects. The remainder of this manuscript is organized as follows: Section 2 describes the input and ground measurement data; Section 3 elaborates on the data preprocessing methods, model structures, and validation methods; Section 4 discusses the overall validation results and analysis; Section 5 summarizes the study conclusions.

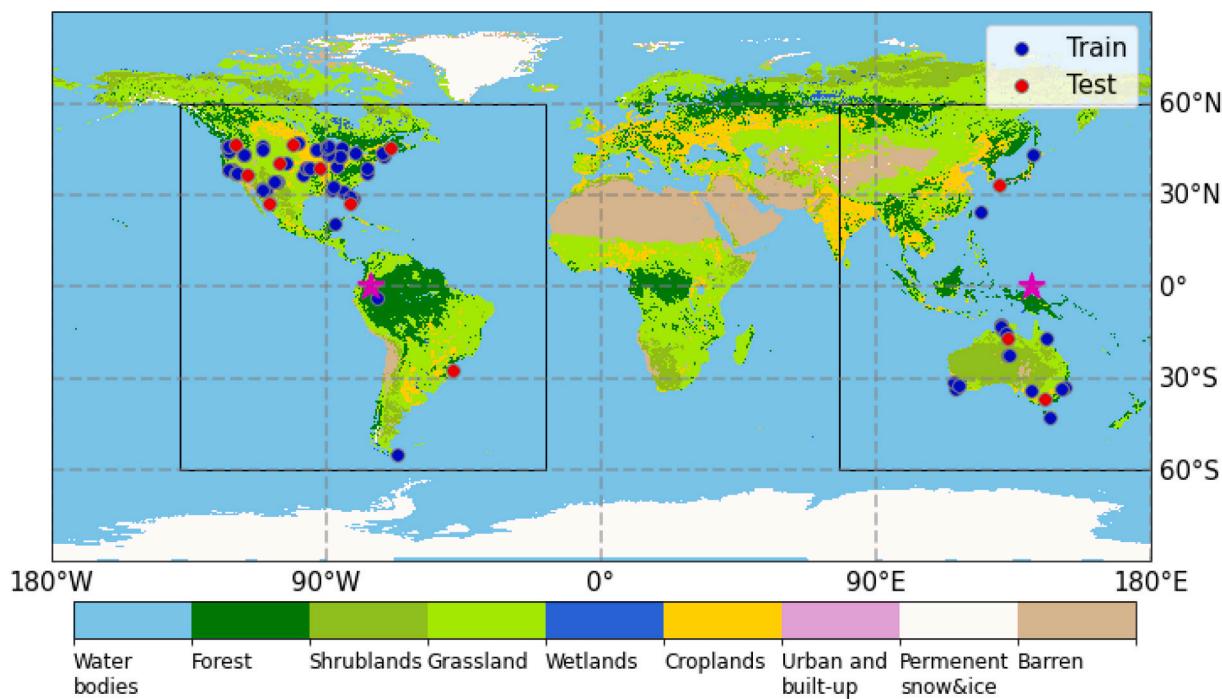


Fig. 1. Distribution of ground measurement stations. Training sites are labeled in blue and testing sites are labeled in red. The MCD12 International Geosphere-Biosphere Programme (IGBP) maps are used as the background. The pink stars indicate the location of ABI and AHI sensors. The black rectangles outline the coverage areas of ABI and AHI. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2. Data

2.1. Input data

The input variables and the corresponding data source are listed in Table 1. The gridded TOA reflectance was retrieved from GOES16/HIMAWARI8 Level 1G archived in the NASA GeoNEX data portal (Wang et al., 2020b). The 15 bands shared by ABI onboard GOES-16 and AHI onboard Himawari-8, ranging from visible to infrared, were included in this study. ABI and AHI sensors cover areas ranging from 60°N, 138°W to 60°S, 18°W and 60°N, 78°E to 60°S, 198°E, respectively (Fig. 1). The study also includes the geometric angles of each scan, namely solar zenith angles (SZA), viewing zenith angles (VZA), and relative azimuth angles (RAA), as inputs. The Global 30 Arc-Second Elevation (GTOPO30) dataset provided the static surface elevation data (Earth Resources Observation And Science (EROS), 2017). The surface albedo information used in this study is carried by the black sky albedo (BSA) of MODIS bands 1–7 (MCD43). The blanks left by MCD43 are filled with the surface albedo climatology data which are generated by Jia et al. (2022a) through averaging the available surface albedo from 2001 to 2020 for the existing MODIS products. Surface albedo and BSA are used synonymously in this study.

2.2. Ground measurement data

Ground measurement data cover 110 sites measured by four networks: AMERIFLUX, BSRN, FLUXNET, and SURFRAD, in 2018 (Fig. 1). All data used as training and testing samples first passed the BSRN quality check (Roesch et al., 2011). Data with SZA larger than 85° were excluded. The data were subjected to a quality check following the procedure described by Li et al. (2021). BSRN and SURFRAD provide measurements at 1 min intervals, whereas FLUXNET and AMERIFLUX measure DSR at 30 min intervals. To derive FLUXNET and AMERIFLUX measurements at every 10/15 min, we first generated the per minute measurements by downscaling through forward filling. For example, the 8:01 to 8:30 am data were all filled with 8:30 am measurements. The

final in-situ measurements of all four networks used as the true value of training or validating were averaged from the 15 min measurements centered at the satellite passing time to eliminate the representative errors of the point-scale DSR measurement (Huang et al., 2016). The hourly measurements were directly averaged, and the daily aggregation DSR were calculated as per procedure stated by Li et al. (2021).

2.3. Ancillary data sets

In this study, we compare our DSR estimations with hourly CERES SYN1deg product (CERES). CERES estimates the radiation fluxes at TOA, surface, and different atmospheric layers with 1° spatial resolution. The product is generated using the Langley Fu-Liou RT code from the observations of CERES instruments onboard the Terra, Aqua, and other polar-orbiting satellites' high-level products. It also includes a geostationary satellite imager to derive high-temporal-resolution data (Rutan et al., 2015). Owing to its high accuracy and global coverage, the validation accuracy of CERES has been used as the baseline result in many studies (Riihela et al., 2017; Zhang et al., 2020; Li et al., 2022). GeoNEX DSR is a newly developed product with the highest spatiotemporal resolution among the existing DSR products, with spatial and temporal resolutions of 1 h and 1 km (Li et al., 2023). It employs the LUT approach on the ABI and AHI TOA bands and aggregates the instantaneous results into an hourly scale. It demonstrates greater accuracy than CERES at both the hourly and daily scales and is regarded to have the highest accuracy in retrieving DSR from satellites (Li et al., 2023).

In addition, the NOAA GOES-R Series Level 2 clear sky mask and cloud top height product were used to analyze the performance of DSR retrievals under different cloud conditions. Full disk versions of the two products were generated at the same temporal resolution as the ABI observations. The clear sky mask was retrieved at a horizontal spatial resolution of 2 km, and cloud top height of 10 km. Both were matched with the ground measurement after conversion to the geographic coordinate system. The products were only available for ABI coverage.

Table 2

The Inputs and outputs of the models compared in this study; the models-of-choice are highlighted in gray.

Model	Input data dimension	Input data description	Feature description	Output data dimension	Output data description
MLP	(26)	(Feature)	15 TOA bands, 7 BSA bands, 3 angles, elevation	(1)	(DSR)
MLP _{nor}	(19)	(Feature)	15 TOA bands, 3 angles, elevation	(1)	(DSR)
MLP _{sw}	(16)	(Feature)	5 TOA shortwave bands, 7 BSA bands, 3 angles, elevation	(1)	(DSR)
BiGRU	(60,26)	(Timestep, Feature)	15 TOA bands, 7 BSA bands, 3 angles, elevation	(60,1)	(Timestep, DSR)
BiGRU _{nor}	(60,19)	(Timestep, Feature)	15 TOA bands, 3 angles, elevation	(60,1)	(Timestep, DSR)
DenseNET	(26,26,26)	(Spatial width, Spatial height, Feature)	15 TOA bands, 7 BSA bands, 3 angles, elevation	(1)	(DSR)
DenseNET _{nor}	(26,26,19)	(Spatial width, Spatial height, Feature)	15 TOA bands, 3 angles, elevation	(1)	(DSR)
CNNGRU	(60,26,26,26)	(Timestep, Spatial width, Spatial height, Feature)	15 TOA bands, 7 BSA bands, 3 angles, elevation	(60,1)	(Timestep, DSR)
CNNGRU _{nor}	(60,26,26,19)	(Timestep, Spatial width, Spatial height, Feature)	15 TOA bands, 3 angles, elevation	(60,1)	(Timestep, DSR)

3. Methods

3.1. Data preprocessing and evaluation methods

The inputs and outputs for each model are listed in [Table 2](#). All of the input variables were matched with the geographic locations of the in-situ sites to generate all the samples. Measurements and satellite observations in 2018 for each 10/15 min over 110 sites were used in this study. The total sample size was approximately 1.2 million. The extensive ground measurement data used in this study also ensured that samples covered as many different land cover types as possible ([Fig. 1](#)). The broadband BSA ranged from 0 to 0.8, and the elevation of the sites ranged from 0 to 3000 m.

Carefully splitting training and testing data are critical for testing and ensuring the model generality. Samples from geospatially independent sites were used as training and testing data since surface-reflecting properties play a critical role in estimating DSR. We manually selected 12 sites as testing sites that were geospatially independent from the training sites and covered different surface types ([Fig. 1](#)). Prior to feeding the data into any of the models, the input variables were normalized by the mean and standard deviation calculated over the entire training sites. The models were trained on 80% of the training sites data and validated on the remaining 20% to determine their configurations.

Two evaluation methods based on sites were employed in this study. The models were firstly tested on the 12 sites that were completely left out during the model tuning process. The second evaluation metric comprised 10-fold cross-validation across sites. All the sites were randomly divided into ten groups. The selected group of sites served as validation data, and the model was trained on the remaining data. This process was repeated for each of the ten groups. The mean of the ten iteration scores was used as the final accuracy.

The evaluation metrics for the validation and comparison of the models includes R-square (R^2), mean bias difference (MBD), root mean square error (RMSE), relative root-mean-square error (rRMSE), and ratio of variance (ROV) ([Tramontana et al., 2016](#)). Among them, RMSE and rRMSE are the most widely used metrics of DSR estimation studies; hence, they were primarily used to compare the models among each other and with existing products. The standard deviation of RMSE among the 10-fold validation (STD) was also included to quantify the uncertainties of each model. The evaluation metrics were calculated using following equations:

$$R^2 = \frac{\sigma_{R_p R_g}}{\sigma R_p \sigma R_g} \quad (1)$$

$$MBD = \frac{1}{N} \sum_{i=1}^N (R_p^i - R_g^i) \quad (2)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (R_p^i - R_g^i)^2} \quad (3)$$

$$rRMSE = RMSE / \bar{R}_g \quad (4)$$

$$ROV = \frac{(\sigma R_g)^2}{(\sigma R_p)^2} \quad (5)$$

where σR_p represents the standard deviation of the predicted values for the testing samples, σR_g represents the standard deviation of the ground measurements for the testing samples, $\sigma_{R_p R_g}$ represents covariance of the predicted values and ground measurements for the testing samples, R_p^i represents the predicted value for observation i , R_g^i represents the ground measurement for observation i , \bar{R}_g represents the mean of the ground measurements for the testing samples, and N is the total number of the testing samples.

3.2. Model construction

The study used an NVIDIA Tesla V100 GPU and the TensorFlow DL framework in Python for implementation. Detail configuration of each model can be found at <https://github.com/Ruohan-Li/DeepLearningDSREstimation.git> (doi: <https://doi.org/10.5281/zenodo.7926773>).

3.2.1. MLP

MLP denotes a fully connected feedforward neural network (NN), one of the most extensively used NN models. Over the last few years, it has been used to estimate DSR ([Takenaka et al., 2011](#); [Ryu et al., 2018](#); [Brown et al., 2020](#); [Peng et al., 2020](#)). A four-layer MLP model was developed as the baseline model based on previous studies where it achieved high accuracy ([Li et al., 2022](#)). The input layer included 26 neurons. The first hidden layer had 25 neurons, followed by another hidden layer of five neurons. One neuron in the output layer corresponded to the DSR value of linear activation. All layers were activated by the rectified linear unit (ReLU) activation function ([Nair and Hinton, 2010](#)). The model used Root Mean Square Propagation (RMSprop) optimizer and mean squared error (MSE) loss function.

3.2.2. DenseNET

The structure of the convolutional and pooling layers in CNN enables learning the spatial relationships from the input data with fewer trainable parameters. CNNs have been used for cloud classification from TOA, demonstrating their ability to detect cloud spatial structure, which is critical for improving DSR estimation ([Chai et al., 2019](#); [Jeppesen et al., 2019](#)). Many deep CNN models such as LeNet-5 ([LeCun et al., 1998](#)), AlexNet ([Krizhevsky et al., 2017](#)), and VGG-16 ([Simonyan and Zisserman, 2014](#)) have been developed and tested for high efficiency in image classification. As the input features include the dominant

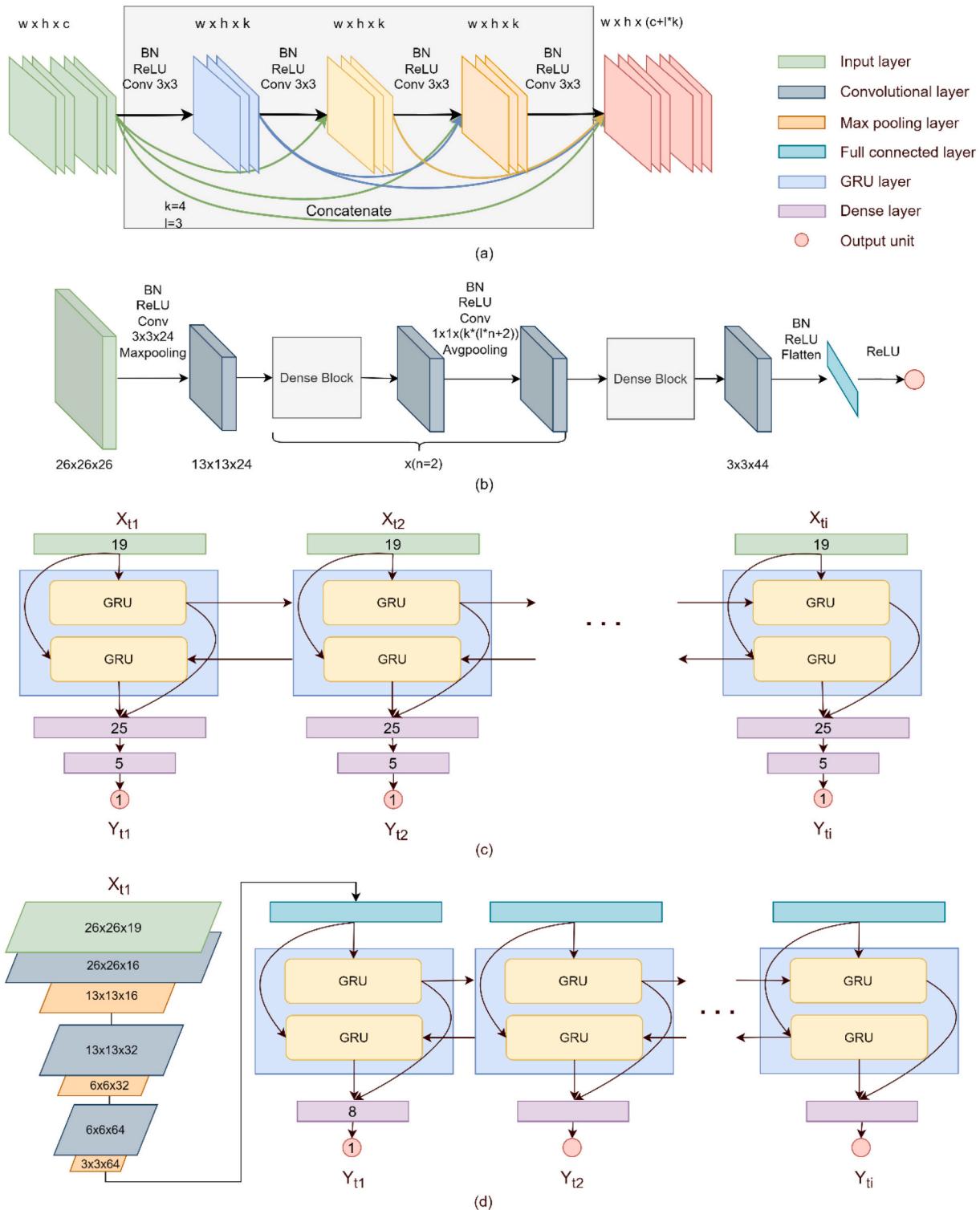


Fig. 2. (a) Structure of one DenseNET dense block. c , w and h are the channel number, width and height of the input sample; k is the growth rate; l is the number of layers within a dense block; n is the number of dense blocks. (b) Structure of the DenseNET. (c) Structure of the BiGRU_{nor}; i represents the time steps within one day. (d) Structure of CNNGRU_{nor}. The size of inputs, convolutional filter, max pooling layer and the number of neurons are labeled on these layers.

information of DSR estimation (Wang et al., 2021) and the spatial information are the complementary knowledge, a dense convolutional network (DenseNET) (Huang et al., 2017) was employed for our specific task to address the issue of vanishing or losing information (He et al., 2016). This allowed constantly concatenation of new features from all preceding features as new input layers, and ensured that the information from the inputs or the first few layers is passed to the end of the

structure. These concatenation operations were encapsulated in dense blocks (Fig. 3a). The developed model has the potential to achieve deep feature extraction with fewer variables and has a regularizing effect, which mitigates overfitting issues.

The structure of the DenseNET is shown in Fig. 2b. The growth rate k was set to a small number, $k = 4$, owing to the large input features. The DenseNET comprised three dense blocks, each with three layers. The

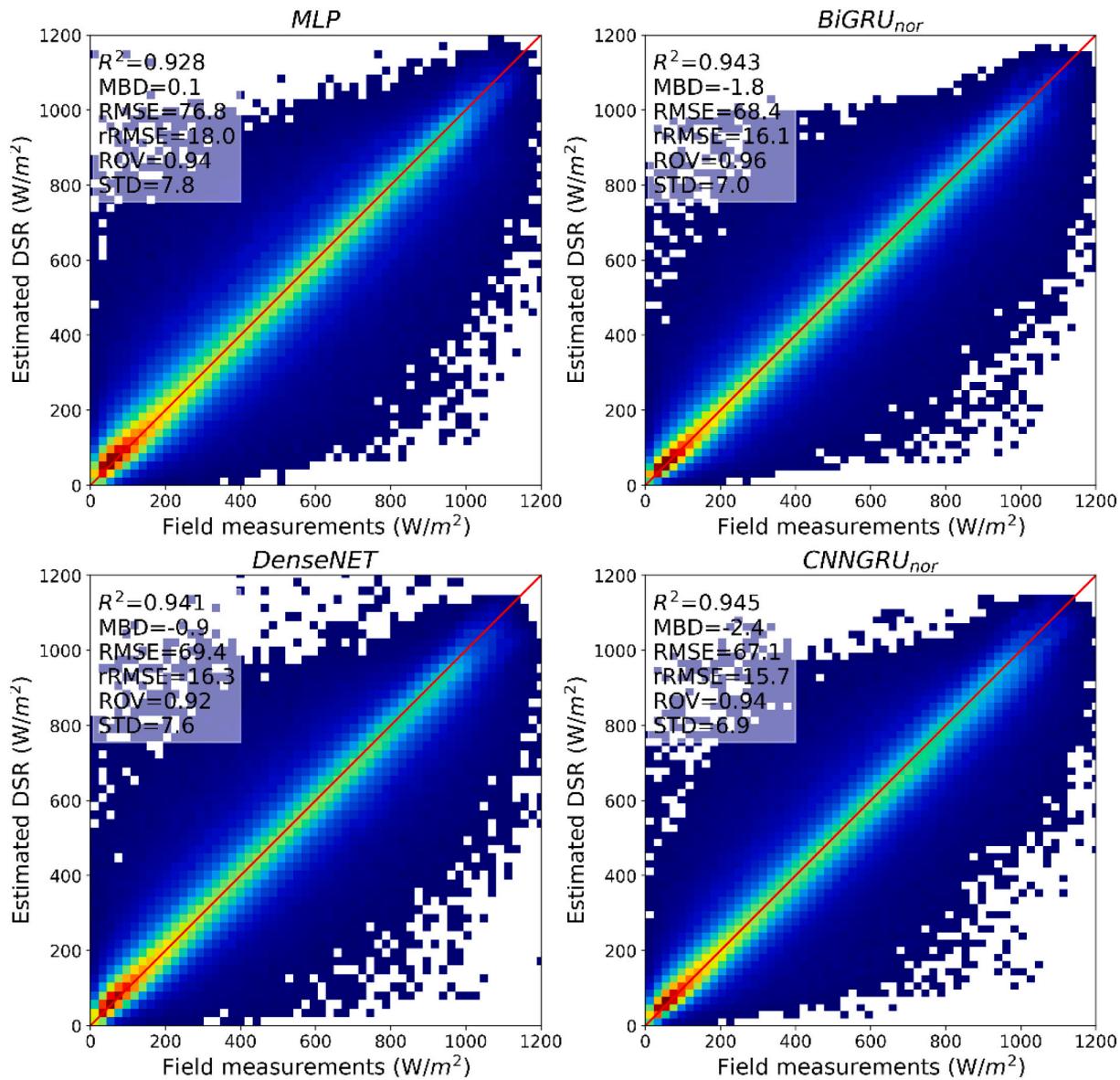


Fig. 3. The 10-fold cross validation results of MLP, BiGRU_{nor}, DenseNET, and CNNGRU_{nor} models.

transition layers consisting of a 1×1 convolutional layer were followed by an average pooling layer employed to connect the dense blocks to reduce the dimensions. The outputs of the last pooling layer were flattened to form a fully connected layer; the output layer with one node estimated the final estimated DSR value. The RMSprop with learning rate of 0.001 and MSE loss function were used to train the model. Previous studies have suggested that the 3D cloud effect should be considered when the spatial scale is smaller than approximately 24 km (Chen et al., 2019). The highest accuracy of DSR estimation for current 1 km sensors is normally achieved when averaging over window size of 25 km (Li et al., 2021); hence, the input sample size of this study was set at 26×26 pixels, corresponding to $26 \text{ km} \times 26 \text{ km}$ in reality.

3.2.3. BiGRU_{nor}

The RNNs, which study the relationships between time-series data in DL, are adapted to eliminate the need for BSA as additional inputs by leveraging multiple observations made within a single day of geostationary sensors. Various RNNs have been developed that use temporal information from time-series data (Sherstinsky, 2020), including the simple RNN (Hopfield, 1982), long short-term memory network

(Hochreiter and Schmidhuber, 1997), gated recurrent unit (GRU) (Cho et al., 2014), and RNNs equipped with bidirectional flows (Schuster and Paliwal, 1997). After comparing the above mentioned RNNs, bidirectional GRU (BiGRU) produced the best results with the highest efficiencies; hence, BiGRU was used in this study. Given that the overall variation in DSR in a day is asymmetrical, the BiGRU has the ability to take advantage of the former and the latter input values to adjust current DSR estimation. The BiGRU includes two GRU layers that are chained in opposite directions, enabling learning information flow from both sides (Fig. 2c).

In this study, we applied a many-to-many BiGRU structure that takes observations from an entire day as input and simultaneously generates a DSR estimate at each time step in a day. The inputs of the BiGRU_{nor} model do not include BSA (Table 2); the structure of this model is shown in Fig. 2c. All observations of one day from 6 am to 8 pm were reconstructed as one sample and were aligned by their local time with a 10/15 min interval that corresponds to a time step. Invalid measurements or observations were filled with -1. The outputs of one sample correspond to the DSR estimations at each time step within a day. The model is not designed to be that deep because the diurnal variation of DSR are still

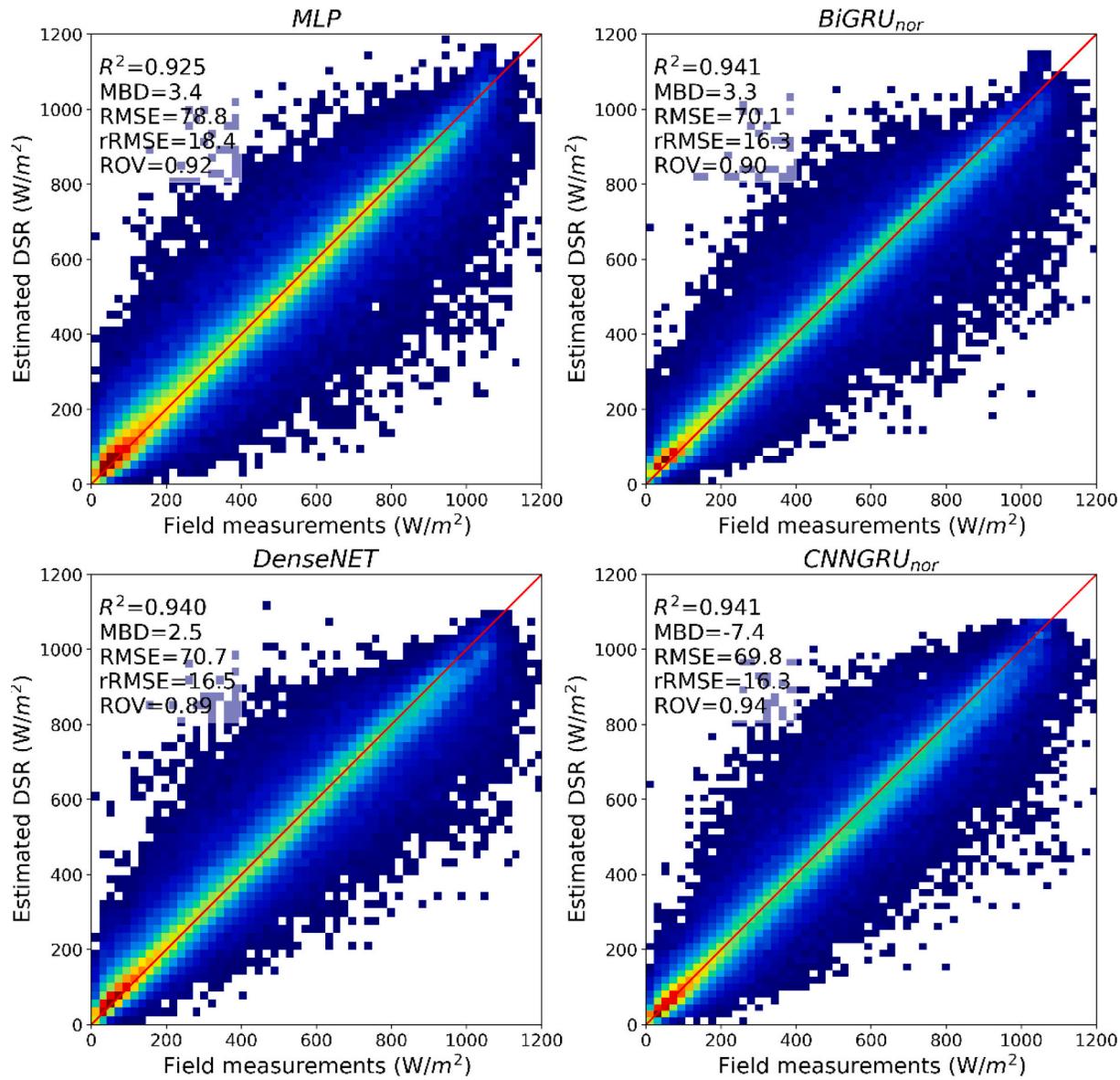


Fig. 4. Validation results of MLP, BiGRU_{nor}, DenseNET, and CNNGRU_{nor} models over 12 independent sites.

heavily influenced by diurnal cloud conditions and not by the temporal correlation within one day. The model includes one BiGRU layer with 32 units each, two dense layers with 25 and 5 nodes each, and one output layer. The adaptive moment estimation (Adam) with learning rate of 0.001 optimization function and MSE loss function were employed for this model.

3.2.4. CNNGRU_{nor}

To investigate whether the combination of temporal and spatial information can further improve the estimation accuracy, a combination of CNN and RNN models, namely the CNNGRU structure, was developed as part of this study owing to the unique temporal and spatial characteristics brought by satellite observations on DSR estimations (Fig. 2d). As the model did not include BSA as inputs, it is referred as CNNGRU_{nor}. The inputs of this model are listed in Table 2. It used a CNN for spatial feature extraction and a GRU for time sequence analyses. This type of model is currently in the early stages of predicting solar energy or precipitation (Sønderby et al., 2020; Bansal et al., 2022); however, to the best of our knowledge, it has not been used to estimate DSR directly. The input data were first fed into the CNN structure via three convolutional layers, followed by batch normalization (BN) and pooling layers. The

flattened nodes were then fed into BiGRU structure, similar to the ones described above but comprising one dense layer including eight neurons. Finally, the model was made to produce the diurnal DSR simultaneously after digesting the spatial and temporal information of all spatial satellite images for one day. The Adam with learning rate of 0.001 optimization function and MSE loss function were employed for this model.

4. Results

4.1. Comparison among the models and with the existing products

The 10-fold cross-validation results have been shown in Fig. 3. All the models were compared over the same samples. CNNGRU_{nor} achieved the highest accuracies with an RMSE of 67.1 W/m² and an rRMSE of 15.7%, followed by BiGRU_{nor} with an RMSE of 68.4 (16.1%) W/m² and the DenseNET with an RMSE of 69.4 (16.3%) W/m². MLP shows the worst RMSE at 76.8 (18%) W/m² but the lowest MBD at 0.1 W/m². The ROV for all four models were higher than 0.9, indicating that the estimations have a similar level of variance as the ground truth. The STD of

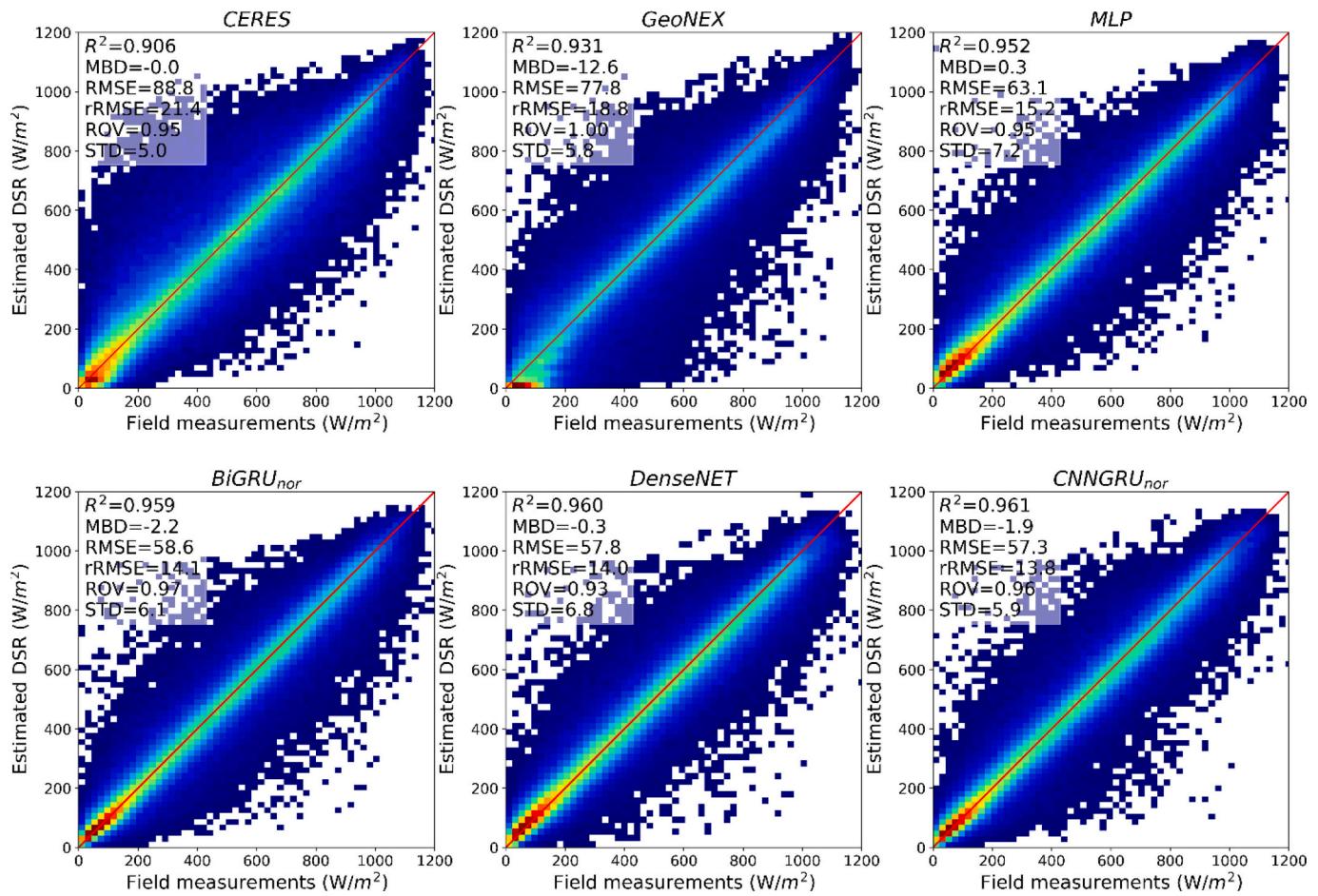


Fig. 5. Comparison of the 10-fold validation results for MLP, BiGRU_{nor}, DenseNET, and CNNGRU_{nor} with CERES and GeoNEX at an hourly scale.

10-fold validation results of the four models were similar, ranging from 6.9 to 7.8 W/m². Fig. 4 shows the validation results on 12 independent sites. The overall accuracy was found to be similar to that of the 10-fold validation, which further demonstrated the generality of the proposed models. CNNGRU_{nor} still achieved the highest accuracy with an RMSE of 69.8 W/m² and an rRMSE of 16.3%. However, it presented the largest MBD at -7.4 W/m². BiGRU_{nor} presented an RMSE of 70.1 (16.3%) W/m² and the DenseNET model showed an RMSE of 70.7 (16.5%) W/m². The MLP method showed the lowest accuracy, with an RMSE of 78.8 (18.4%) W/m². Compared to the baseline MLP model, the results demonstrated the ability of DL methods to achieve higher accuracy in retrieving DSR at 1 km and 10/15 min with no overfitting issues. It is noticeable that both the BiGRU_{nor} and CNNGRU_{nor} models did not include surface BSA as inputs but achieved comparable accuracies.

Hourly validation results were compared with CERES and GeoNEX. CERES provides hourly DSR estimations with global coverage. Notably, the spatial resolution of CERES is 1° which is different from that of other products and methods included in this comparison (Jiang et al., 2020). GeoNEX is a newly developed product based on AHI and ABI with 1 km and 1 h spatiotemporal resolution. These two products present greater accuracy than existing DSR hourly products and hence were compared in this study (Li et al., 2023).

As shown in Fig. 5, the 10/15 min retrievals of the proposed DL models were aggregated into hourly scales for comparison. CERES showed the highest RMSE at 88.8 (21.4%) W/m² but with no bias. GeoNEX presents better accuracies than CERES, with an RMSE of 77.8 (18.8%) W/m² which aligned well with existing validations (Li et al., 2022). MLP shows higher accuracies than existing two products, with an RMSE of 63.1 (15.2%) W/m². The reasons for the high accuracy of the

simple MLP method could be the sufficient training data and comparably small coverage areas in this specific geostationary sensor-based DSR retrieval. The other DL methods showed high accuracies with RMSE values of 58.6 (14.1%), 57.8 (14.0%), and 57.3 (13.9%) W/m² for BiGRU_{nor}, DenseNET, and CNNGRU_{nor}, respectively. To the best of our knowledge, these accuracies are higher than those of most of the existing hourly products or operational results (Letu et al., 2022; Chen et al., 2021; Zhang et al., 2020). Notably, when compared with the instantaneous estimation shown in Fig. 3, the hourly accuracies improved by approximately 10 W/m² for all ML models. Among them, BiGRU_{nor} and CNNGRU_{nor} showed slightly smaller improvements, suggesting that the instantaneous estimation of the GRU-related model incorporated some information from the temporal averaging results. As DSR normally exhibits pronounced diurnal cycles, we analyzed the intra-hour anomalies by employing the methodology described in Tramontana et al. (2016). These anomalies are determined as the deviation of an instantaneous value from the climatological averaged hourly value. To calculate the climatological DSR values for this study, we utilized the CERES 2017–2019 hourly data, extracting the mean hourly cycle for each site, each day, and each hour. Fig. 6 demonstrates the successful reproduction of the intra-hour variation in DSR by all four models. Notably, the DL models exhibit higher R^2 values, approximately 0.8, surpassing that of the MLP model.

We further aggregated hourly DSR into daily scale for comparison (Fig. 7). For DSR retrieval at daily scale, the superiority of DL was reduced, particularly the model with temporal structure. DenseNET showed the lowest RMSE at 15.1 W/m² with no bias. MLP, BiGRU_{nor}, and CNNGRU_{nor} show similar performance with RMSE values of 17.0 (8.2%), 16.1 (7.0%), and 16.0 (7.7%) W/m² respectively. Nevertheless,

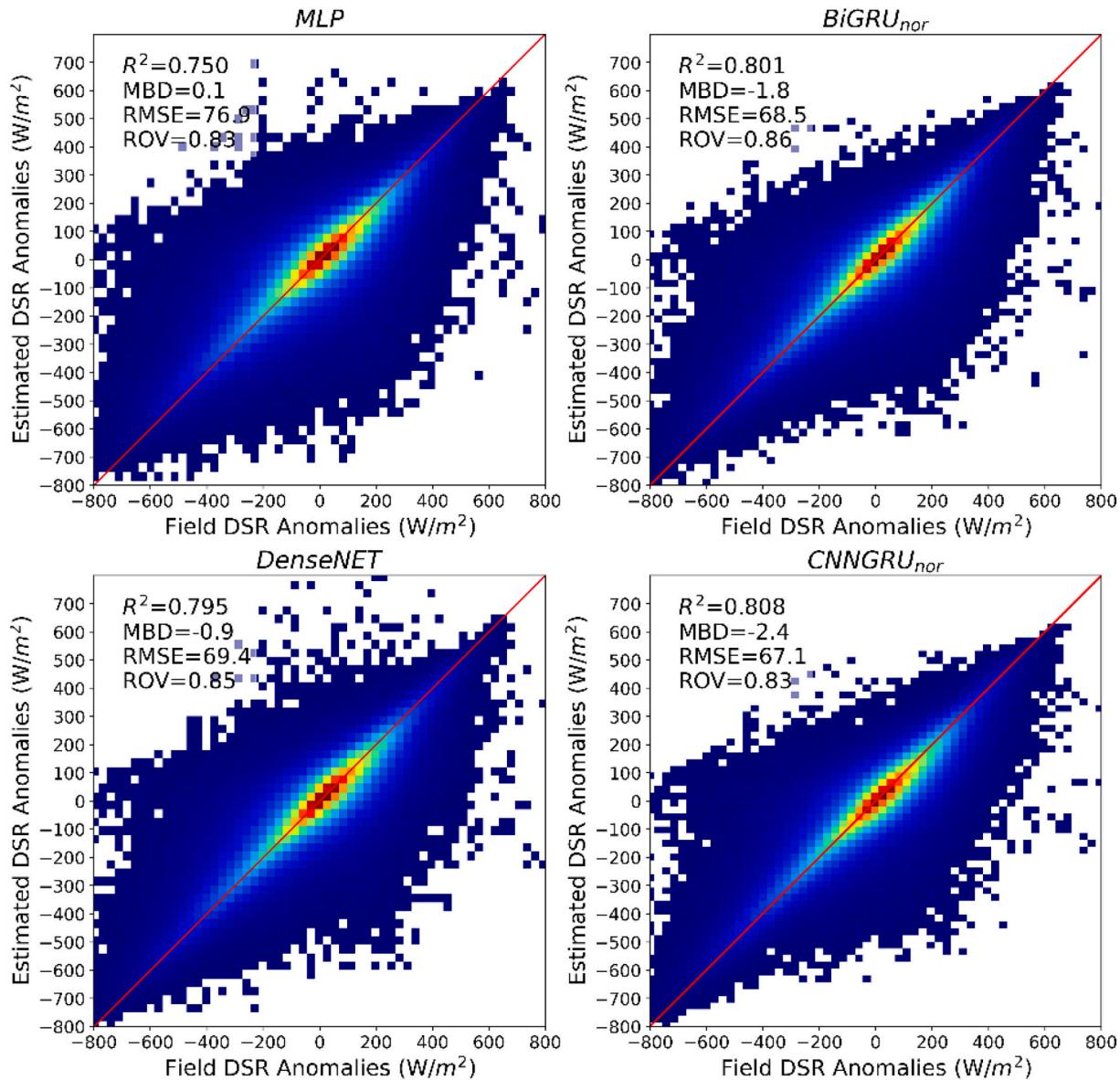


Fig. 6. The anomalies of the 10-fold validation results for MLP, BiGRU_{nor}, DenseNET, and CNNGRU_{nor} at an hourly scale.

they presented lower RMSE than CERES at 24.0 (11.5%) W/m² with no bias and GeONEX at 19.6 (9.4%) W/m².

Fig. 8 shows the diurnal variations in the proposed methods at an hourly scale for comparison with the GeoNEX and CERES products. The results were plotted from day of year (DOY) 190 to 210 across six sites covering different land surface types. The results show that both GeoNEX and the proposed methods can capture the diurnal patterns in DSR variation, while CERES is more prone to overestimation under cloudy conditions. GeoNEX displayed more variations than in-situ measurements such as DOY 196 over US-Pnp sites and DOY 209 over US-SRG sites, which is also partly observed in MLP methods, likely owing to single pixel retrieval. DL-based methods performed better in these cases of overcast conditions and barren surfaces. Nevertheless, even with spatial information, DL methods lost some sudden changes in DSR like DOY 209 over Collie sites. Even when DL methods could capture the change, they tend to smooth the change.

Fig. 9 shows the spatial distribution of the 10-fold validation rRMSE of the proposed methods, CERES and GeoNEX, over all sites. In general, the rRMSE of all sites improved with the proposed ML models. The results further demonstrated that similar patterns observed for both the

proposed methods and existing products eliminated the overfitting issues. DL methods showed a more evenly distributed rRMSE than MLP. High rRMSE values in South America and over coastal regions were observed for all the models and products. One possible reason is that more diverse cloud conditions over coastal regions and less accurate or representative ground measurements may contribute to the low accuracy of all models and products. The VZA-related radial degradation that resulted in a high rRMSE at sites distant from AHI and ABI owing to high VZA in GeoNEX was also less pronounced in the DL-based methods.

To conduct a comprehensive comparison of the models, this study included a list of parameter numbers and floating point operations (FLOPs) for each model in Table 3. BiGRU_{nor} and CNNGRU_{nor} both simultaneously estimated 60 values in a day; hence the average of each estimation was calculated and compared. MLP shows the lowest parameter numbers and FLOPs, followed by BiGRU_{nor} for each single estimation. DenseNET presented the most complex structure.

Table 3 compares the model accuracies at different temporal scales, providing a useful reference for deciding which model to use under different conditions. CNNGRU_{nor} with both temporal and spatial structures showed the highest accuracy at temporal resolutions higher than

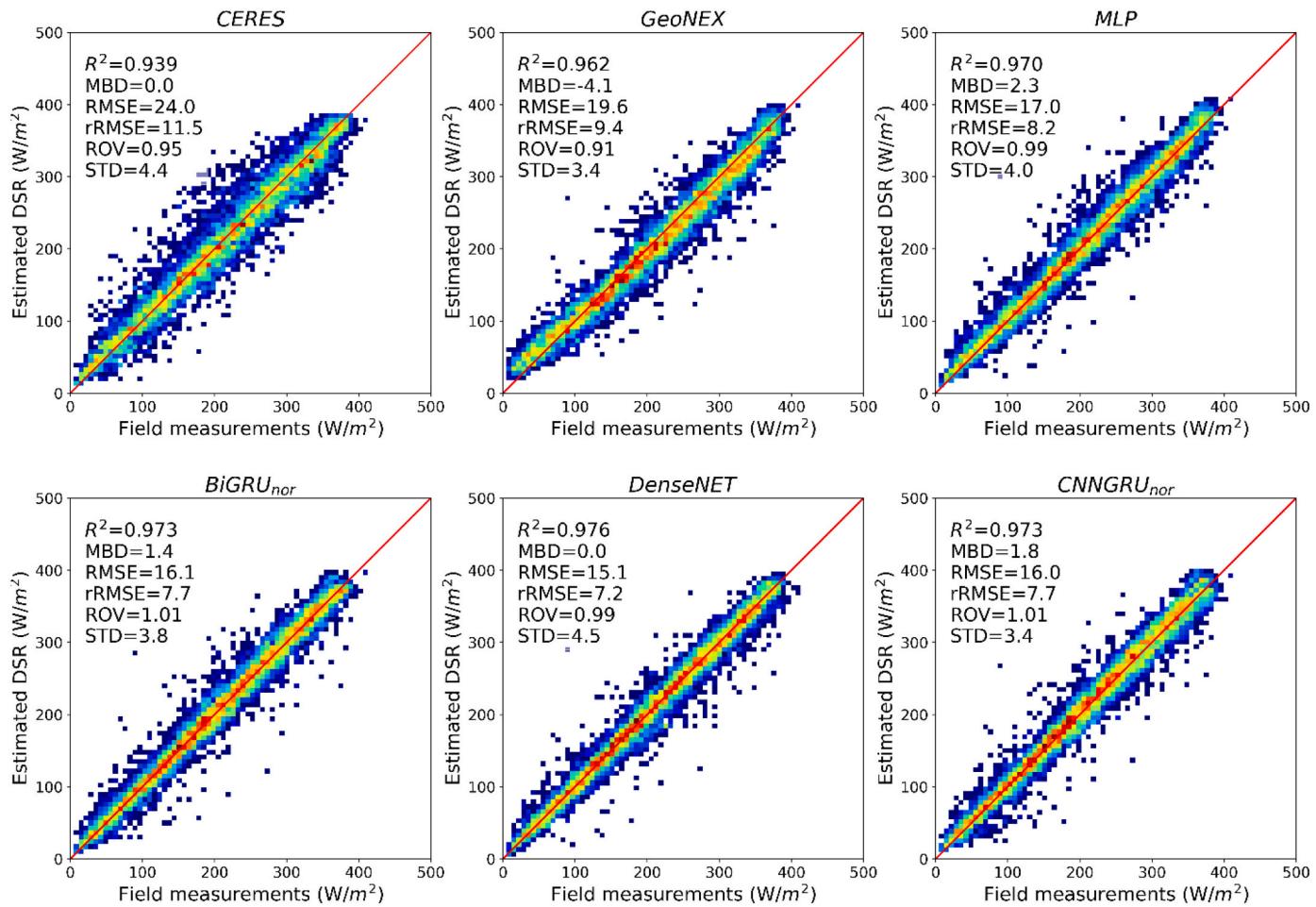


Fig. 7. Comparison of the 10-fold validation results of MLP, BiGRU_{nor}, DenseNET, and CNNGRU_{nor} with CERES and GeoNEX at the daily scale.

an hour. The superiority of GRU-related temporal structures disappeared at daily scales, as DenseNET surpassed both BiGRU_{nor} and CNNGRU_{nor}. However, both CNNGRU_{nor} and DenseNET have high numbers of parameters and FLOPs. Considering the speed and computing resources, BiGRU_{nor} would be the top choice for DSR estimation under all three temporal scales.

4.2. Impact of surface albedo

Surface albedo is a crucial variable in DSR estimation and currently accounts for a minimal input in existing DSR products (Wang et al., 2020; Wang et al., 2021). However, the retrieval of surface albedo degrades in the presence of clouds, snow, and other bright surfaces (Schaaf and Wang, 2015; He et al., 2018). Data gaps exist in current surface albedo data, even in the widely used MCD43, and using surface albedo as an explicit variable may introduce further uncertainties (Schaaf et al., 2002).

The proposed BiGRU_{nor} and CNNGRU_{nor} models successfully achieved stable and accurate DSR estimations without using surface BSA as inputs. To further testify whether the time-sequence structure helps eliminate the dependencies on surface albedo, we trained and compared the proposed DL models with and without BSA as inputs (Fig. 10). The inputs of the models compared in this section are summarized in Table 2. We also compared the DL models with the LUT method (Wang et al., 2021). When BSA was high, LUT produced extremely poor results and increased the interquartile range (IQR) among 10-fold sites, primarily because both the unreliable sources of surface albedo and limited bands which brought difficulties in distinguishing snow and clouds contribute

to the degradation of physical methods over bright surfaces (Karlsson et al., 2017; Wang et al., 2020a, 2020b). Compared to LUT, ML improved the accuracy for all BSA intervals. ML models present a similar accuracy when the surface BSA was lower than 0.2, comparable to most cases in reality. As the surface BSA increased, the difference between inputs with and without surface BSA became visible. For both the MLP and DenseNET models, the absence of surface BSA as input, led to a higher rRMSE than that of the same model structure with surface BSA, particularly when the surface albedo was higher than 0.4. The difference in such cases can exceed 10% when the surface BSA ranges from between 0.6 and 0.8 for both models. The models without surface BSA as inputs either tend to have larger IQR. However, the same conclusion was not observed in the GRU-related models. BiGRU_{nor} achieved the highest accuracy than all the models without surface BSA as inputs at all intervals. Additionally, the differences between BiGRU and BiGRU_{nor} in terms of median rRMSE and IQR were found to be small in all cases and substantially smaller than those of other methods when the surface BSA was large. Compared with all models without surface BSA as inputs, the BiGRU structure has better generalities as it is less affected by surface albedo. This is likely owing to the incorporation of time series information in one day, which provides more reliable information on the stable surface. The CNNGRU_{nor} model displayed satisfactory results when the surface BSA is between 0.4 and 0.6. However, it degrades when the surface BSA was higher than 0.6. One possible reason for this was the depth of the models. Some critical temporal information was lost in the CNN layers, which could not be passed on to the GRU layers.

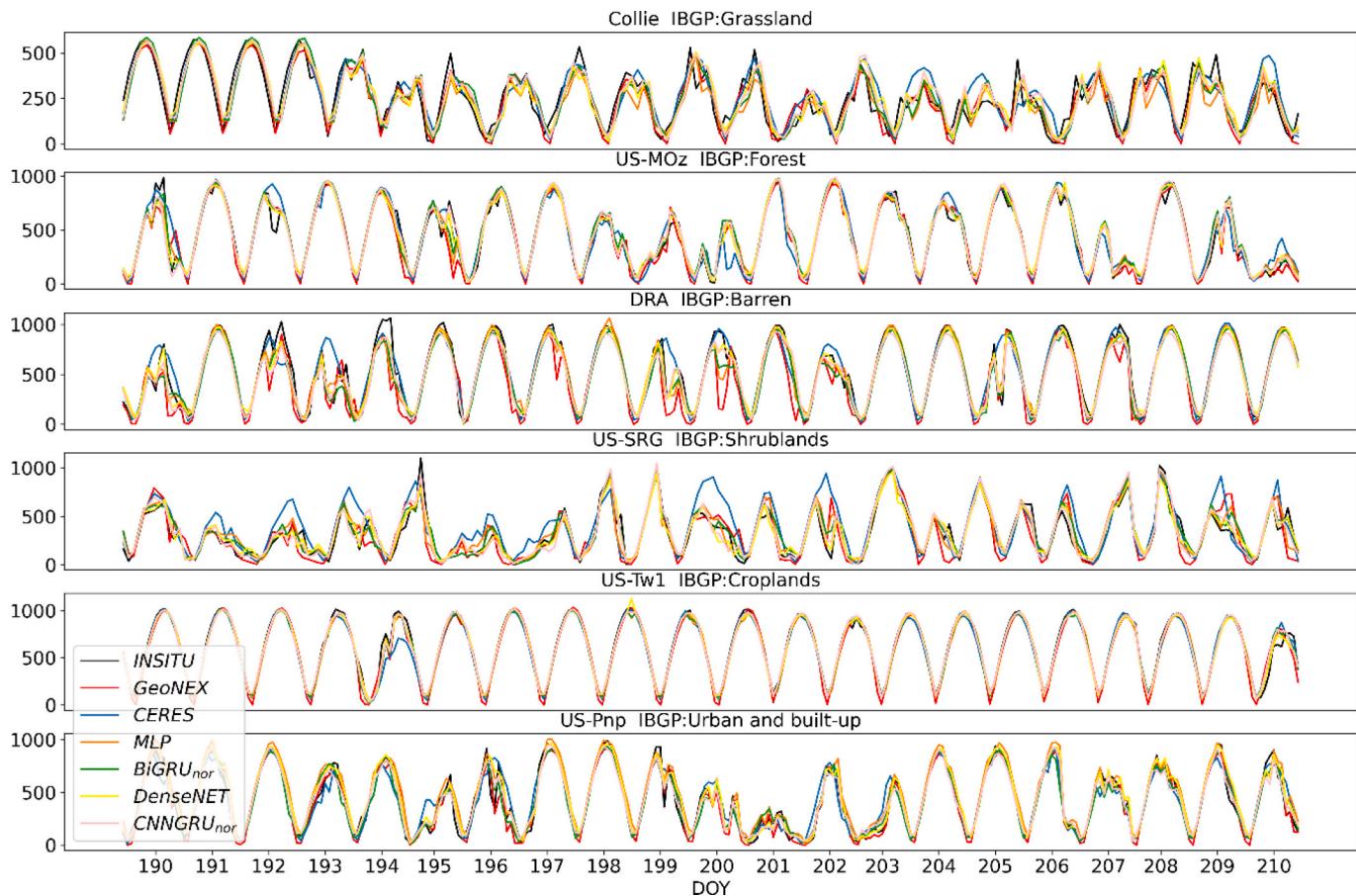


Fig. 8. Diurnal variation in DSR of MLP, BiGRU_{nor}, DenseNET, and CNNGRU_{nor} with CERES and GeoNEX from day of year (DOY) 190 to 210 across six sites.

4.3. Impacts of cloud conditions

The study also compared all models with LUT under clear and cloudy conditions (Fig. 11a). For all models, accuracies under clear-sky conditions were significantly higher than those under cloud-sky conditions. Under clear sky conditions, the difference between different neural networks was trivial, but varied across different models under cloudy sky conditions. The LUT methods presented the highest rRMSE (37.2%) under cloudy conditions, which corresponds well with previous findings (Li et al., 2023). MLP presented the second-highest rRMSE (25.4%) and the highest IQR. DL models were shown to improve the DSR estimation accuracies under cloudy conditions. Single-pixel-based BiGRU_{nor} showed high accuracy, with a rRMSE of 22.3%. One possible reason is that the temporal information before and after the current moment contributes useful information in terms of cloud type and cloud positions. DenseNET and CNNGRU_{nor} had lower rRMSE under cloudy skies of 22.1% and 21.5%, respectively. These two models incorporate spatial information into their input, which aided in recognizing different cloud conditions. All DL models showed less IQR than MLP which demonstrate their high generality under cloudy condition.

Furthermore, we examined the performance of different models for different cloud top heights (Fig. 11b). ML models presented similar accuracy when the cloud height was lower than 2000 m. With the increase in cloud height, attributed to the solar-cloud-satellite geometry effects are more prone to occur (Wang et al., 2017), the DL-based models particularly the models with the spatial structure showed high accuracies, indicating that the convolutional filter may help models distinguish the cloud structure and hence improve the DSR estimations.

4.4. Influence of infrared bands

Existing DSR retrieval methods, particularly physics-based methods, primarily use shortwave bands of TOA reflectance. With flexibility in the inputs of DL models, this study tested the importance of infrared (IR) bands in retrieving DSR. Fig. 12a shows the 10-fold validation results of MLP trained on all bands ranging from 0.47 to 13.3 μm shared by AHI and ABI sensors, and Fig. 12b shows the validation results of the same model but trained on shortwave bands only (0.47–3.9 μm) (MLP_{sw}). The inputs difference between MLP and MLP_{sw} are also summarized in Table 2. The total RMSE of MLP_{sw} is 4 W/m² lower than MLP. We further analyzed the impacts of infrared bands under different surface conditions as illustrated in Fig. 12c. Our findings indicate that when the surface albedo is less than 0.4, the differences between MLP and MLP_{sw} are negligible. However, for surface albedo greater than 0.4, MLP_{sw} exhibited significantly higher accuracy. This can be attributed to the complementary function of IR bands to the visible bands, enabling the distinction of clouds from high albedo surface types. To investigate the relative contributions of each channel, a sensitivity analysis was applied to all normalized inputs via the Sobol analysis method (Sobol, 2001) based on the MLP method (Fig. 12d). First-order indices (S1) measure the contribution of a single input to the output variances, while total-order indices (ST) are the total of all single and higher-order interactions of this input. This shows that the blue band (0.47 μm) contributed the most to the surface shortwave radiation estimation, which is consistent with previous findings (Wang et al., 2020a, 2020b). IR bands especially between 12.3 and 13.3 μm, also showed great contributions in estimating DSR. Both bands were used to generate products such as the cloud mask and the cloud-top height (Menzel and Strabala, 1997; Chang and Xiong, 2019) owing to their ability to delineate the

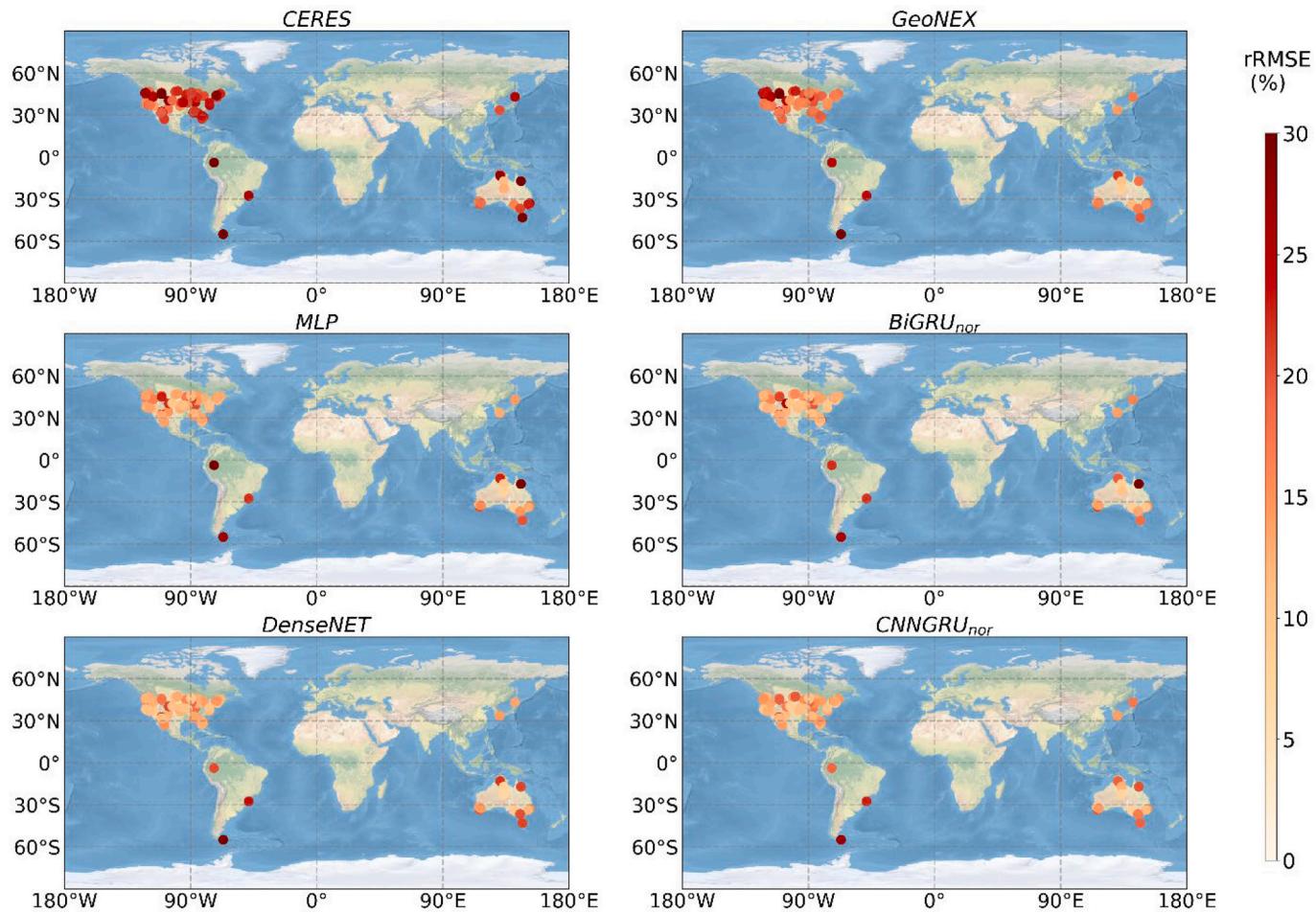


Fig. 9. Spatial distribution of the rRMSE for MLP, BiGRU_{nor}, DenseNET, and CNNGRU_{nor} with CERES and GeoNEX.

Table 3

Summary of models' parameter numbers, FLOPs, and accuracy at 10/15 min, hourly, and daily scale; the number in brackets represent the average for each time step in GRU-related model.

Model	Parameters number	FLOPs	10/15 min RMSE (W/m ²)	Hourly RMSE (W/m ²)	Daily RMSE (W/m ²)
MLP	811	795	76.8	63.1	17.0
BiGRU _{nor}	18,013 (300)	115,610 (1927)	68.4	58.6	16.1
DenseNET	18,913	4,614,984	69.4	57.8	15.1
CNNGRU _{nor}	143,761 (2396)	198,746,606 (3,312,443)	67.1	57.3	16.0

upper level of the cloud (Dupont et al., 2009). Previous studies have reported similar results for IR bands in estimating the diffuse portion of the DSR (Chen et al., 2022).

5. Conclusions

Recently, there has been an increasing need for the retrieval of DSR with high spatiotemporal resolution and less latency. However, the issues of cloud heterogeneity and the dependence of existing methods on high-level products hamper the development of fine-resolution DSR retrieval. DL, which has been widely used in different areas, provides a promising way by incorporating spatial and temporal information in addition to the spectral information used by conventional approaches.

This study developed and compared three DL models for retrieving DSR with high spatiotemporal resolution from both ABI and AHI TOA reflectance. DenseNET used convolutional layers to extract spatial information and employed densely connected layers to pass the critical band and angle information to the end of the model. BiGRU_{nor} employed hidden states to transmit information from previous or future time series, as well as two gates to determine the amount of information to be retained. CNNGRU_{nor} combined spatial and temporal structures. Both BiGRU_{nor} and CNNGRU_{nor} did not include surface albedo as input. The ground measurement data covered 110 stations over AHI and ABI coverage areas, with approximately 1.2 million samples in total. To avoid data overfitting, the models were evaluated over 12 independent sites that covered sufficient land types as well as a 10-fold site-based validation. Both independent site validation and 10-fold site-based validation presented similar accuracies, which demonstrated that the proposed methods do not have overfitting issues.

DL methods show high accuracy for estimating high spatial- and temporal-resolution DSR. BiGRU_{nor}, DenseNET, and CNNGRU_{nor} achieved an instantaneous RMSE of 68.4 (16.1%), 69.4 (16.3%), 67.1 (15.7%) W/m², respectively. The hourly results were compared with those of CERES-SYN and GeoNEX. BiGRU_{nor}, DenseNET, and CNNGRU_{nor} achieve 58.6 (14.1%), 57.8 (14.0%), and 57.3 (13.9%) respectively, which is higher than CERES with 88.8 (21.4%) and GeoNEX (18.8%) with 77.8 (18.8%). Different DL models exhibit different strengths under different conditions. CNNGRU_{nor} presented the highest accuracy at hourly or sub-hourly resolution. DenseNET showed the highest accuracy at daily scale. BiGRU_{nor} was considered to be the best choice for DSR estimation at all temporal resolution considering the model efficiency.

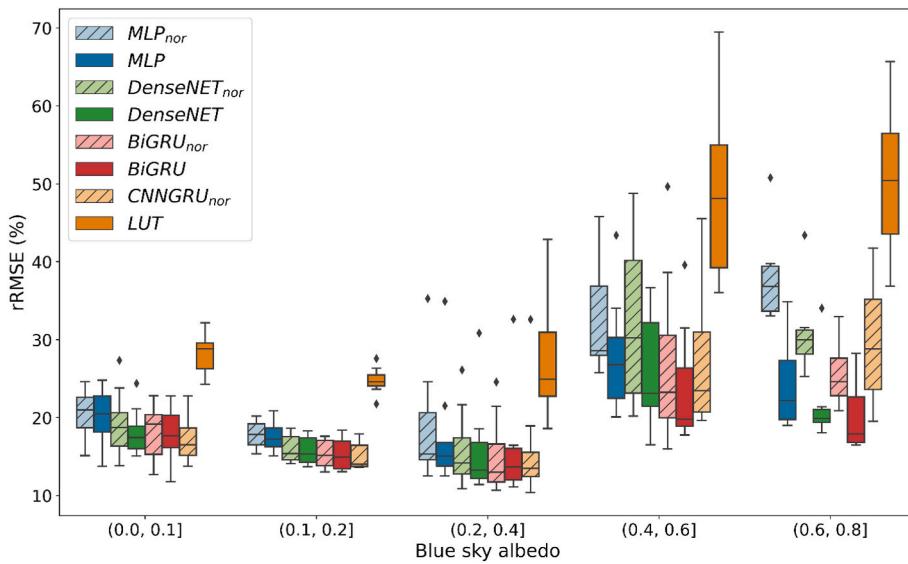


Fig. 10. The boxplots label the median, quartiles, and outliers of rRMSE with and without surface BSA as inputs during the 10-fold validation; the whiskers represent the maximum and minimum values within 1.5 IQR beyond the first and third quartiles; the model without surface BSA as inputs include *nor* as subscript and highlighted with the shadow.

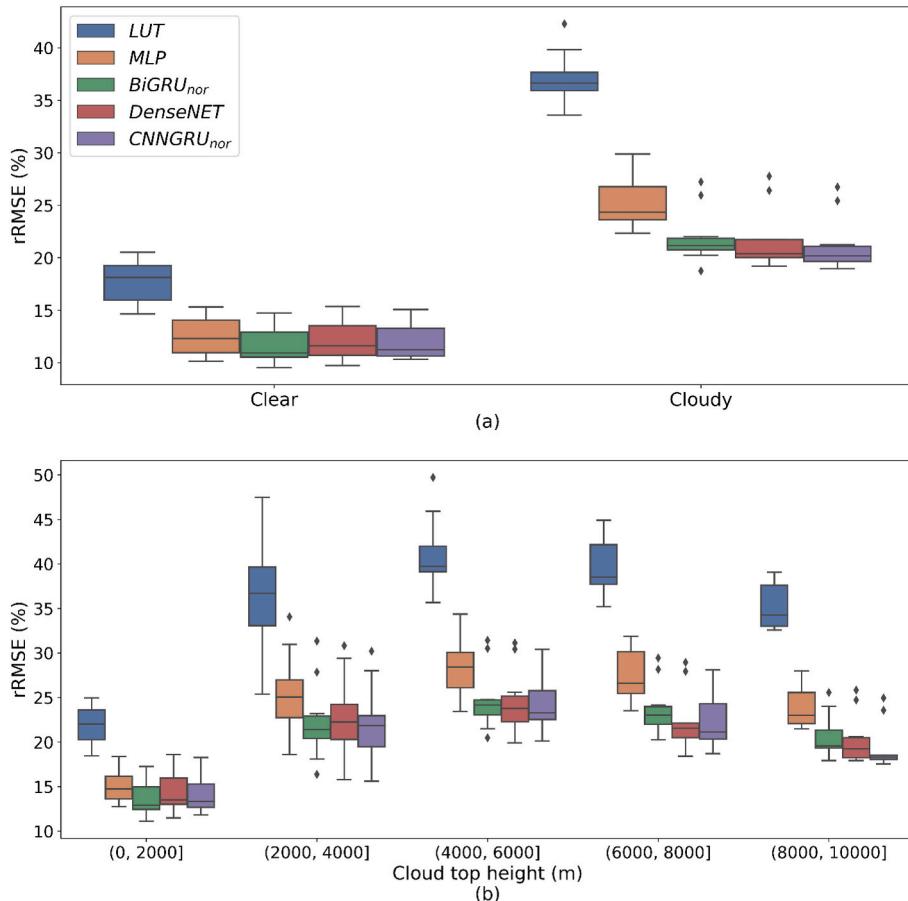


Fig. 11. The boxplots label the median, quartiles, and outliers of rRMSE (a) under clear and cloudy sky conditions and (b) under different cloud top heights during the 10-fold validation; the whiskers represent the maximum and minimum values within 1.5 IQR beyond the first and third quartiles.

Models, such as DenseNET and CNNGRU_{nor}, incorporating spatial structures, presented higher accuracies under cloudy and particularly high cloud conditions. The use of temporal structure in DL models, such as BiGRU_{nor}, not only allows for similar accuracies without surface al-

bedo but also presents higher accuracies when surface albedo is high. The incorporation of IR bands can increase estimation accuracy, particularly over high albedo surface.

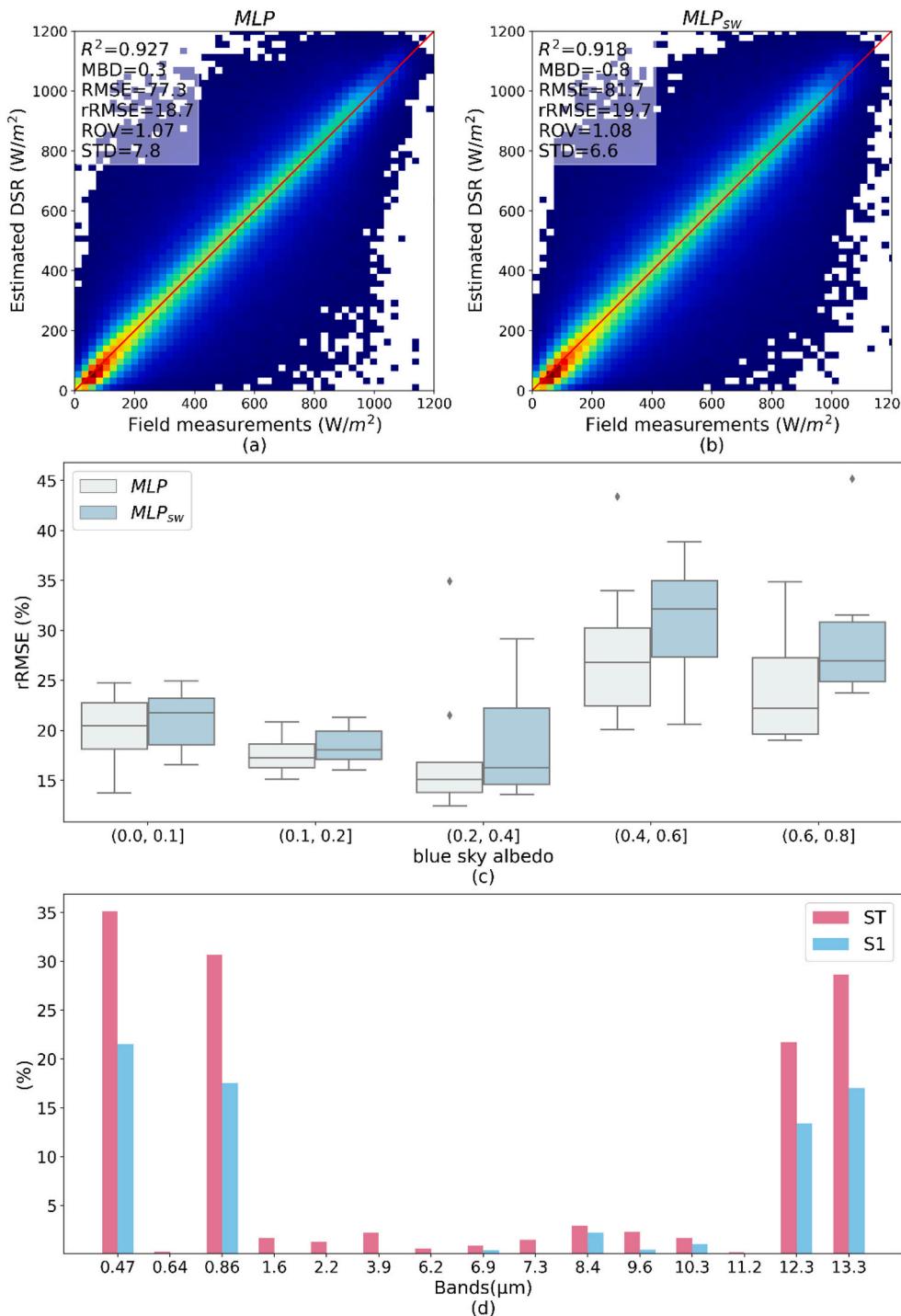


Fig. 12. (a) The scatterplots of MLP model. (b) The scatterplots of MLP_{sw} model. (c) The boxplots label the median, quartiles, and outliers of rRMSE over different surface BSA; the whiskers represent the maximum and minimum values within 1.5 IQR beyond the first and third quartiles. (d) Sensitivity analysis of all bands in MLP model.

Author credit statement

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

Data availability

Data will be made available on request.

References

- Bansal, A.S., Bansal, T., Irwin, D., 2022. A moment in the sun: solar nowcasting from multispectral satellite data using self-supervised learning. In: Proceedings of the Thirteenth ACM International Conference on Future Energy Systems, pp. 251–262.
- 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>.
- Buster, G., Rossol, M., Maclaurin, G., Xie, Y., Sengupta, M., 2021. A physical downscaling algorithm for the generation of high-resolution spatiotemporal solar irradiance data. *Sol. Energy* 216, 508–517.
- Chang, T., Xiong, X., 2019. GOES-16/ABI thermal emissive band assessments using GEO-LEO-GEO double difference. *Earth Space Sci.* 6 (12), 2303–2316.
- Chai, D., Newsam, S., Zhang, H.K., Qiu, Y., Huang, J., 2019. Cloud and cloud shadow detection in landsat imagery based on deep convolutional neural networks. *Remote Sens. Environ.* 225, 307–316.
- Chen, J., Zhu, W., Yu, Q., 2021. Estimating half-hourly solar radiation over the continental United States using GOES-16 data with iterative random forest. *Renew. Energy* 178, 916–929.
- Chen, J., Zhu, W., Yu, Q., 2022. High-spatiotemporal-resolution estimation of solar energy component in the United States using a new satellite-based model. *J. Environ. Manag.* 302, 114077.
- 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. Sel. Top. Appl. Earth Obs. Remote Sens.* 12 (3), 821–835.
- Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation arXiv preprint arXiv:1406.1078.
- Dupont, J.C., Haefelin, M., Long, C.N., 2009. Cirrus cloud radiative effect on surface-level shortwave and longwave irradiances at regional and global scale. *Atmos. Chem. Phys. Discuss.* 9 (6), 26777–26832.
- Earth Resources Observation And Science (EROS), 2017. Global 30 Arc-Second Elevation (GTOPO30), U.S. 465 Geological Survey [Data set]. <https://doi.org/10.5066/F7DF6PQ5>.
- Gandhi, O., Kumar, D.S., Rodríguez-Gallegos, C.D., Srinivasan, D., 2020. Review of power system impacts at high PV penetration part I: factors limiting PV penetration. *Sol. Energy* 210, 181–201.
- Ham, S.H., Kato, S., Barker, H.W., Rose, F.G., Sun-Mack, S., 2014. Effects of 3-D clouds on atmospheric transmission of solar radiation: cloud type dependencies inferred from A-train satellite data. *J. Geophys. Res. Atmos.* 119 (2), 943–963.
- 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 DISCOVR/EPIC observations. *Remote Sens. Environ.* 232, 111320.
- He, T., Liang, S., Wang, D., Cao, Y., Gao, F., Yu, Y., Feng, M., 2018. Evaluating land surface albedo estimation from landsat MSS, TM, ETM+, and OLI data based on the unified direct estimation approach. *Remote Sens. Environ.* 204, 181–196.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.
- Hopfield, J.J., 1982. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences* 79 (8), 2554–2558.
- Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural Comput.* 9, 1735–1780.
- Huang, G., Li, X., Huang, C., Liu, S., Ma, Y., Chen, H., 2016. Representativeness errors of point-scale ground-based solar radiation measurements in the validation of remote sensing products. *Remote Sens. Environ.* 181, 198–206.
- Huang, G., Liang, S., Lu, N., Ma, M., Wang, D., 2018. Toward a broadband parameterization scheme for estimating surface solar irradiance: development and preliminary results on MODIS products. *J. Geophys. Res.-Atmos.* 123, 12180–12193.
- Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708.
- 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.
- Jain, R.K., Qin, J., Rajagopal, R., 2017. Data-driven planning of distributed energy resources amidst socio-technical complexities. *Nature Energy* 2 (8), 1–11.
- Jeppesen, J.H., Jacobsen, R.H., Inceoglu, F., Toftegaard, T.S., 2019. A cloud detection algorithm for satellite imagery based on deep learning. *Remote Sens. Environ.* 229, 247–259.
- Jia, A., Wang, D., Liang, S., Peng, J., Yu, Y., 2022a. Global daily actual and snow-free blue-sky land surface albedo climatology from 20-year MODIS products. *J. Geophys. Res. Atmos.* 127 (8), e2021JD035987.
- Jia, A., Liang, S., Wang, D., 2022b. Generating a 2-km, all-sky, hourly land surface temperature product from advanced baseline imager data. *Remote Sens. Environ.* 278, 113105.
- Jia, A., Liang, S., Wang, D., Ma, L., Wang, Z., Xu, S., 2023. Global hourly, 5 km, all-sky land surface temperature data from 2011 to 2021 based on integrating geostationary and polar-orbiting satellite data. *Earth Syst. Sci. Data* 15 (2), 869–895.
- 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.
- Jiang, H., Lu, N., Huang, G., Yao, L., Qin, J., Liu, H., 2020. Spatial scale effects on retrieval accuracy of surface solar radiation using satellite data. *Appl. Energy* 270, 115178.
- Karlsson, K.G., Anttila, K., Trentmann, J., Stengel, M., Fokke Meirink, J., Devasthale, A., Hollmann, R., 2017. CLARA-A2: the second edition of the CM SAF cloud and radiation data record from 34 years of global AVHRR data. *Atmos. Chem. Phys.* 17 (9), 5809–5828.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. Imagenet classification with deep convolutional neural networks. *Commun. ACM* 60 (6), 84–90.
- Letu, H., Nakajima, T.Y., Wang, T., Shang, H., Ma, R., Yang, K., Shi, J., 2022. A new benchmark for surface radiation products over the East Asia-Pacific region retrieved from the Himawari-8/AHI next-generation geostationary satellite. *Bull. Am. Meteorol. Soc.* 103 (3), E873–E888.
- LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. *Proc. IEEE* 86 (11), 2278–2324.
- 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.
- Li, R., Wang, D., Liang, S., Jia, A., Wang, Z., 2022. Estimating global downward shortwave radiation from VIIRS data using a transfer-learning neural network. *Remote Sens. Environ.* 274, 112999.
- Li, R., Wang, D., Wang, W., Nemani, R., 2023. A GeoNEX-based high-spatiotemporal-resolution product of land surface downward shortwave radiation and photosynthetically active radiation. *Earth Syst. Sci. Data* 15 (3), 1419–1436.
- Liang, S., Wang, D., He, T., Yu, Y., 2019. Remote sensing of earth's energy budget: synthesis and review. *Int. J. Digital Earth* 12, 737–780.
- Liang, S., Wang, K., Zhang, X., Wild, M., 2010. Review on estimation of land surface radiation and energy budgets from ground measurement, remote sensing and model simulations. *IEEE J. Sel. Top. in Appl. Earth Obs. Remote Sens.* 3 (3), 225–240. <https://doi.org/10.1109/JSTARS.2010.2048556>.
- 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.-Atmos.* 111.
- Marcos, J., Storkel, O., Marroyo, L., Garcia, M., Lorenzo, E., 2014. Storage requirements for PV power ramp-rate control. *Sol. Energy* 99, 28–35.
- Menzel, W.P., Strabala, K., 1997. Cloud top properties and cloud phase algorithm theoretical basis document. University of Wisconsin–Madison.
- Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted boltzmann machines ICML-10, 807–814.
- Nguyen, A., Velyai, M., Schoene, J., Zheglov, V., Kurtz, B., Murray, K., Kleissl, J., 2016. High PV penetration impacts on five local distribution networks using high resolution solar resource assessment with sky imager and quasi-steady state distribution system simulations. *Sol. Energy* 132, 221–235.
- Palaintier, B., 2015. Integrated Grid Modeling System (IGMS) for Combined Transmission and Distribution Simulation (No. NREL/PR-5D00-65552). National Renewable Energy Lab.(NREL), Golden, CO (United States).
- 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>.
- Pfenninger, S., Staffell, I., 2016. Long-term patterns of european PV output using 30 years of validated hourly reanalysis and satellite data. *Energy* 114, 1251–1265.
- Pinker, Frouin, Li, 1995. A review of satellite methods to derive surface shortwave irradiance. *Remote Sens. Environ.* 51 (1), 108–124.
- Riihela, A., Key, J.R., Meirink, J.F., Kuipers Munneke, P., Palo, T., Karlsson, K.-G., 2017. An intercomparison and validation of satellite-based surface radiative energy flux estimates over the Arctic. *J. Geophys. Res. Atmos.* 122 (9), 4829–4848. <https://doi.org/10.1002/2016JD026443>.
- Roesch, A., Wild, M., Ohmura, A., Dutton, E.G., Long, C.N., Zhang, T., 2011. Assessment of BSRN radiation records for the computation of monthly means: BSRN. *Atmos. Meas. Tech.* 4 (2), 339–354. <https://doi.org/10.5194/amt-4-339-2011>.
- Rutan, D.A., Kato, S., Doelling, D.R., Rose, F.G., Nguyen, L.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.
- 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>.
- Sobol, I.M., 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simul.* 55 (1–3), 271–280.
- Sönderby, C.K., Espeholt, L., Heek, J., Dehghani, M., Oliver, A., Salimans, T., Kalchbrenner, N., 2020. Metnet: A neural weather model for precipitation forecasting arXiv preprint arXiv:2003.12140.
- Schaaf, C.B., Gao, F., Strahler, A.H., Lucht, W., Li, X.W., Tsang, T., Strugnell, N.C., Zhang, X.Y., Jin, Y.F., Muller, J.P., Lewis, P., Barnsley, M., Hobson, P., Disney, M., Roberts, G., Dunderdale, M., Doll, C., d'Entremont, R.P., Hu, B.X., Liang, S.L., Privette, J.L., Roy, D., 2002. First operational BRDF, albedo nadir reflectance products from MODIS. *Remote Sens. Environ.* 83, 135–148.
- Schaaf, C., Wang, Z., 2015. MCD4A3 MODIS/Terra+Aqua BRDF/Albedo Daily L3 Global -500 m V006, NASA 550 EOSDIS Land Processes DAAC [Data Set]. <https://doi.org/10.5067/MODIS/MCD4A3.006>.
- Schuster, M., Paliwal, K.K., 1997. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* 45 (11), 2673–2681.

- Stephens, G.L., Li, J., Wild, M., Clayson, C.A., Loeb, N., Kato, S., L'Ecuyer, T., Stackhouse, P.W., Lebsack, M., Andrews, T., 2012. An update on Earth's energy balance in light of the latest global observations. *Nat. Geosci.* 5 (10), 691–696. <https://doi.org/10.1038/ngeo1580>.
- Sherstinsky, A., 2020. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Phys. D: Nonlinear Phenom.* 404, 132306.
- Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition *arXiv preprint arXiv:1409.1556*.
- 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>.
- Tang, W.J., Yang, K., Qin, J., Li, X., Niu, X.L., 2019. A 16-year dataset (2000–2015) of high-resolution (3 h, 10 km) global surface solar radiation. *Earth Syst. Sci. Data* 11, 1905–1915.
- Tramontana, G., Jung, M., Schwalm, C.R., Ichii, K., Camps-Valls, G., Ráduly, B., Papale, D., 2016. Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. *Biogeosciences* 13 (14), 4291–4313.
- Tong, D., Farnham, D.J., Duan, L., Zhang, Q., Lewis, N.S., Caldeira, K., Davis, S.J., 2021. Geophysical constraints on the reliability of solar and wind power worldwide. *Nat. Commun.* 12 (1), 1–12.
- Van Laake, P.E., Sanchez-Azofeifa, G.A., 2004. Simplified atmospheric radiative transfer modelling for estimating incident PAR using MODIS atmosphere products. *Remote Sens. Environ.* 91, 98–113.
- Wang, D., Liang, S., Zhang, Y., Gao, X., Brown, M.G., Jia, A., 2020. A new set of MODIS land products (MCD18): downward shortwave radiation and photosynthetically active radiation. *Remote Sens.* 12 (1), 168.
- 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.
- Wang, T., Shi, J., Husi, L., Zhao, T., Ji, D., Xiong, C., Gao, B., 2017. Effect of solar-cloud-satellite geometry on land surface shortwave radiation derived from remotely sensed data. *Remote Sens.* 9 (7), 690.
- Wang, W., Li, S., Hashimoto, H., Takenaka, H., Higuchi, A., Kalluri, S., Nemani, R., 2020. An introduction to the geostationary-NASA earth exchange (GeoNEX) products: 1. Top-of-atmosphere reflectance and brightness temperature. *Remote Sens.* 12 (8), 1267.
- Wei, J., Li, Z., Li, K., Dickerson, R.R., Pinker, R.T., Wang, J., Cribb, M., 2022. Full-coverage mapping and spatiotemporal variations of ground-level ozone (O₃) pollution from 2013 to 2020 across China. *Remote Sens. Environ.* 270, 112775.
- Wyser, K., O'Hirok, W., Gautier, C., 2005. A simple method for removing 3-D radiative effects in satellite retrievals of surface irradiance. *Remote Sens. Environ.* 94 (3), 335–342.
- Zhang, H., Dong, X., Xi, B., Xin, X., Liu, Q., He, H., Yu, S., 2021. Retrieving high-resolution surface photosynthetically active radiation from the MODIS and GOES-16 ABI data. *Remote Sens. Environ.* 260, 112436.
- Zhang, Y., Liang, S., He, T., 2022. Estimation of Land Surface Downward Shortwave Radiation Using Spectral-Based Convolutional Neural Network Methods: A Case Study From the Visible Infrared Imaging Radiometer Suite Images. *IEEE Trans. Geosci. Remote Sens.* 60, 1–15.
- Zhang, Y., Liang, S., He, T., Wang, D., Yu, Y., Ma, H., 2020. Estimation of land surface incident shortwave radiation from geostationary advanced himawari imager and advanced baseline imager observations using an optimization method. *IEEE Trans. Geosci. Remote Sens.* 60, 1–11.