A global long-term (1981–2019) daily land surface radiation budget product from AVHRR satellite data using a residual convolutional neural network

Jianglei Xu[1], Shunlin Liang[2], Bo Jiang[3]

5   [1]School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China

[2]Corresponding author: Department of Geographical Sciences, University of Maryland, College Park, MD 20742, USA

[3]Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

Correspondence to: Shunlin Liang (slliang@whu.edu.cn)

Abstract. The surface radiation budget, also known as all-wave net radiation (R_n), is a key parameter for various land surface

10   processes including hydrological, ecological, agricultural and biogeochemical processes. Satellite data can be effectively used

to estimate R_n, but existing satellite products have coarse spatial resolutions and limited temporal coverage. In this study, a

point-surface matching estimation (PSME) method is proposed to estimate surface R_n using a residual convolutional neural

network (RCNN) integrating spatially adjacent information to improve the accuracy of retrievals. A global high-resolution

(0.05°) long-term (1981–2019) R_n product was subsequently generated from Advanced Very High-Resolution Radiometer

15   (AVHRR) data. Specifically, the RCNN was employed to establish a nonlinear relationship between globally distributed

ground measurements from 537 sites and AVHRR top of atmosphere (TOA) observations. Extended triplet collocation (ETC)

technology was applied to address the spatial scale mismatch issue resulting from the low spatial support of ground

measurements within the AVHRR footprint by selecting reliable sites for model training. The overall independent validation

results show that the generated AVHRR R_n product is highly accurate, with R^2, root-mean-square error (RMSE), and bias of

20   0.84, 26.66 Wm^{-2} (31.66%), and 1.59 Wm^{-2} (1.89%), respectively. Inter-comparisons with three other R_n products, i.e., the 5

km Global Land Surface Satellite (GLASS), the 1° Clouds and the Earth's Radiant Energy System (CERES), and the 0.5° ×

0.625° Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA2), illustrate that our AVHRR

R_n retrievals have the best accuracy under all of the considered surface and atmospheric conditions, especially thick cloud or

hazy conditions. The spatiotemporal analyses of these four R_n datasets indicate that the AVHRR R_n product reasonably

25   replicates the spatial pattern and temporal evolution trends of R_n observations. This dataset is freely available at

https://doi.org/10.5281/zenodo.5509854 for 1981-2019 (Xu et al., 2021).

1 Introduction

Net radiation (R_n), which characterizes the surface radiation budget, is the difference between downward and upward radiation

across the shortwave (0.3–3.0 μm) and longwave (3.0–100 μm) spectra. The surface radiation budget links the atmospheric

30   climate system to the land surface. R_n is thus a critical variable for studying Earth–atmosphere interactions, including

meteorological, hydrological, and biological processes, and is also responsible for the redistribution of surface available energy (Sellers et al., 1997). Accurate characterization and quantification of spatial-temporal variations in surface R_n are essential for both scientific and industrial applications, such as hydrological modeling and water resource management (Hao et al., 2019).

35 However, because the spatial and temporal dynamics of surface R_n are affected by multiple surface features (e.g., albedo, emissivity, and land surface temperature) and atmospheric parameters (e.g., clouds, aerosols, ozone, and water vapor) (Wang et al., 2015b), existing surface R_n data suffers from large uncertainties (Jia et al., 2016; Jia et al., 2018; Jiang et al., 2018; Yang and Cheng, 2020). Therefore, there is an urgent need for long-term, high-resolution surface R_n dataset to more properly understand its spatial pattern and temporal dynamics (i.e., seasonal and inter-annual variability).

Traditionally, historical R_n and surface radiative components have been measured at ground meteorological stations. These

40 ground-based measurements are widely used to study spatiotemporal variations in regional surface radiation and to evaluate gridded products (Jia et al., 2018; Zhang et al., 2020; Zhang et al., 2015). Nevertheless, the high cost of maintaining radiation radiometers means that stations are sparely distributed, severely hindering our ability to study and understand the spatiotemporal variability of surface R_n at global scale.

Alternatively, reanalysis products provide long-term global surface R_n information (Zhang et al., 2016). The greatest advantage

45 of reanalysis products is their global coverage over a long-term period; however, the large uncertainty and coarse spatial resolution of reanalysis products hinder their applications at regional spatial scale (Jia et al., 2018; Zhang et al., 2016; Zhang et al., 2020).

Retrieving R_n from satellite data is another effective method (Liang et al., 2010; Liang et al., 2019). Currently, satellite-based R_n retrieval methods can be broadly divided into two categories, physical methods based on radiative transfer (RT) and

50 empirical statistical methods. RT-based physical methods show a great generalization because they consider the physical processes of solar radiation from the top of the atmosphere to the Earth's surface (Tang et al., 2019). The look-up-table (LUT) and parameterization methods are two typical physical schemes that are widely used to estimate surface radiation from satellite data. To address the low computational efficiency of the radiative transfer model (RTM), the LUT method was proposed to estimate the surface radiation from satellite top of atmosphere (TOA) observations, which combines the advantages of RTM-

55 based simulations and statistical methods (Wang et al., 2015b, a; Wang et al., 2020b; Cheng and Liang, 2016; Cheng et al., 2017; Huang et al., 2011). This approach relies on several theoretical assumptions in the RTM simulation process, such as water vapor amounts, aerosol types, plane-parallel homogeneous clouds, horizontal homogeneity and directional properties of the surface (Jiang et al., 2019b), which results in errors in the final radiation estimates (Hao et al., 2018; Cheng et al., 2017; Jiao et al., 2015). The parameterization scheme is another typical physical method that uses various surface and atmospheric

60 parameter data to calculate surface radiation based on simplified RT (Huang et al., 2020; Qin et al., 2012). However, in the calculation of parameterized formulas, errors from each input variable accumulate in the final calculated surface radiation.

Conversely, different from physical methods, empirical statistical methods typically account for spatial-temporal variations in R_n by establishing statistical relationships between satellite measures or sensed variables, including surface and atmospheric variables or TOA observations, and surface radiation measurements (Tang et al., 2017; Huang et al., 2020; Jiang et al., 2015; Bisht and Bras, 2010; Bisht et al., 2005) using linear or non-linear models. Machine learning (ML) has played an important role in the development of empirical statistical methods owing to its strong nonlinear fitting ability (Jiang et al., 2014; Jiang et al., 2016; Chen et al., 2020; Xu et al., 2020). Although statistical methods incorporate very little physical knowledge and have limited ability to expand their coverage, they are still widely employed owing to their low computational cost and easy implementation.

Several well-known global R_n datasets have been generated from satellite data (Table 1), such as the Global Energy and Water Cycle experiment surface radiation budget (GEWEX-SRB) (Pinker and Laszlo, 1992), the Clouds and the Earth's Radiant Energy System (CERES) (Loeb et al., 2018), and the International Satellite Cloud Climatology Project (ISCCP) (Zhang et al., 2004). Although these products have been widely used in various fields, their coarse spatial resolution (≥ 100 km) cannot meet the requirements of high-resolution R_n data. A high-resolution (5 km) global R_n product was recently released (Jiang et al., 2018) from the Global Land Surface Satellite (GLASS) product suite (Liang et al., 2020). The GLASS R_n product has only been available since 2000 and was produced from Moderate Resolution Imaging Spectroradiometer (MODIS) data and reanalysis products. The Fluxcom initiative recently published a gridded product of surface R_n using multiple ML methods to merge energy flux measurements with remote sensing and meteorological data to estimate R_n retrievals, but the product only provides available data on areas covered by vegetation and the dataset only spans 15 years (Jung et al., 2019b). A long-term, high-resolution, and accurate surface R_n dataset is, therefore, still urgently needed to help more clearly understand the long-term spatiotemporal variation of global surface R_n.

In this study, deep learning ML methods were explored to produce a long-term, high-resolution surface R_n dataset from AVHRR data. Reviewing recent studies on surface R_n estimation using ML methods, many significant advancements have been made (Jiang et al., 2014; Husi et al., 2020; Wei et al., 2019; Wang et al., 2019); however, clear deviations still exist between satellite-derived estimations and ground-based measurements. Apart from the performance of the ML method itself, many of these discrepancies are attributed to two aspects; first, the spatial representation mismatch between satellite data and ground-based measurements and, second, the neglect of spatial adjacent effects on surface radiation estimation.

The spatial scale mismatch between surface radiation for domain averages and ground point measurements with insufficient spatial representativeness (Jiang et al., 2019b; Barker and Li, 1997) has attracted attention for a long time in the development of ML (Yuan et al., 2020a) and evaluation of gridded products (Huang et al., 2016; Yang, 2020; Román et al., 2009). However, many current studies still use matched point-surface sample datasets to train ML model regardless of the difference in spatial representativeness of matched point-surface data. The triple collocation (TC) technique (Stoffelen, 1998) was considered as an appropriate upscaling approach for the impact of random measurement error on ground-based measurements in comparison

to other complicated upscaling methods (e.g., the time stability approach and the block kriging algorithm (Crow et al., 2012))

95   (Yuan et al., 2020a). Furthermore, an extended triple collocation (ETC) method was proposed by Mccoll et al. (2014) and then

applied to the validation activities of the Soil Moisture Active Passive (SMAP) level-2 surface soil moisture (SSM) product

(Chen et al., 2017) and satellite surface albedo products (Wu et al., 2019). Therefore, the ETC technology is employed to limit

the effect of upscaling errors of ground measurements on the final surface R_n estimates at the satellite footprint scale.

Spatial adjacent effects should also be considered in the development of ML methods. With an increase in spatial resolution,

100   horizontal inhomogeneities in the atmosphere have become increasingly important and reduce accuracy of surface radiation

retrievals at higher spatial resolutions, especially in conjunction with high solar and viewing angles (Wyser et al., 2002); the

correlation between satellite TOA observations and surface radiation measurements weakens, and surface radiation cannot be

accessed directly from satellite TOA data for individual pixels. Convolutional neural networks (CNNs) were initially designed

to perform image recognition tasks; they can be readily used to extract various high-level, hierarchical, and abstract spatial

105   pattern features from original multispectral or hyperspectral satellite images (Yuan et al., 2020b; Ball et al., 2017). Using this

approach, multiple environmental parameters and their spatially adjacent effects can be accounted for in the estimation of

surface R_n (Jiang et al., 2019b). Therefore, CNN represents a promising method for integrating potential spatially adjacent

effects in surface radiation estimation (Jiang et al., 2020b; Jiang et al., 2019b).

Several studies have successfully employed CNNs and other deep neural networks to retrieve surface parameters, such as

110   global solar radiation (Jiang et al., 2019a), precipitation (Wu et al., 2020), and land surface temperature (Yin et al., 2020), with

varying success rates. However, no study has yet attempted to retrieve global surface R_n using a CNN model. In this study, a

residual CNN (RCNN) based point-surface matching estimation method (PSME) is proposed for estimating global land surface

R_n. Specifically, the RCNN model links ground-based R_n measurements with multiple image blocks of AVHRR TOA

observation data, including reflectance in visible channels and brightness temperature in thermal infrared channels, along with

115   other additional auxiliary variables. These auxiliary variables include angular information, i.e., solar zenith angle (SZA),

viewing zenith angle (VZA), and relative azimuth angle (RAA), and daily Modern-Era Retrospective analysis for Research

and Applications, Version 2 (MERRA2) R_n data (Jia et al., 2018; Zhang et al., 2016). Before training the RCNN, ETC

technology is applied to select reliable sites at a global level to generate representative training samples, making the established

statistical relationship more representative of surface R_n variation at the AVHRR footprint scale. After validation and

120   comparison, the best-trained model was subsequently implemented through the proposed PSME scheme to generate a global-

scale 39-year daily surface R_n dataset (1981–2019) with 0.05° resolution from the AVHRR data and MERRA2 reanalysis.

The remainder of this paper is organized as follows. Section 2 summarizes the characteristics of the data used for the

reconstruction of the ETC triplet and PSME method development. Section 3 describes the ETC method for the selection of

reliable sites and the PSME process for surface R_n estimation using the RCNN model. The results for selected sites based on

125   the ETC, the performance of the RCNN model and the long-term spatiotemporal variation of the R_n dataset are presented in

Sect. 4. The discussion is presented in Sect. 5. The data availability is described in Sect. 6. Finally, the conclusions are presented in Sect. 7.

Table 1: Summary of available R_n products.

Product name	Spatial resolution	Temporal resolution	Period	References
Reanalysis				
NCEP/CFSR	~38 km	6 hours	1979–2010	Saha et al. (2010)
NASA/MERRA2	0.5°×0.625°	1 hour	1979–present	Gelaro et al. (2017a)
ERA5	0.25°×0.25°	3 hours	1950–present	Hersbach et al. (2020)
JRA55	~55 km	3 hours	1958–present	Kobayashi et al. (2015)
NCEP-DOE R2	~200 km	6 hours	1979–present	Kanamitsu et al. (2002)
Satellite products				
CERES-SYN	1°×1°	1 hour	2000–present	Doelling et al. (2016)
GEWEX-SRB	1°×1°	3 hour	1983–2007	Stackhouse Jr et al. (2000)
ISCCP-FD	280 km	8 days	1983–2012	Rossow and Zhang (1995)
FLUXCOM	0.0833°×0.0833°	8 days	2001–2015	Jung et al. (2019a)
MODIS-TERRA	0.05°×0.05°	daily	2001–2009	Verma et al. (2016)
GLASS-MODIS	0.05°×0.05°	daily/daytime	2000–2019	Jiang et al. (2018)

130

2 Datasets

2.1 Ground measurements

Ground measurements of daily surface R_n were used for the RCNN model development. The in situ measurements used in this study cover the period from 2001 to 2019, and were obtained using various pyranometers (e.g., Kipp & ZonenCNR-1 and

135 Eppley) at 537 globally distributed stations from 15 observation networks/programs, as shown in Fig. 1. Detailed information about these observation networks/programs is listed in Table 2, and specific information about these sites is shown in Table S1. The operational thermoelectric pyranometers are known for their high-accuracy performance, with a spectral response of 0.3-3.0 μm, a sensitivity of 7-14μVW^{-1} m^2, a thermal effect of less than 5%, and an annual stability of 5% (Lu et al., 2011; Jiang et al., 2019b). These stations are maintained by multiple global and regional observation network organizations, such as

140 Global Fluxnet, the Greenland Climate Network (GCNET) and the Programme for Monitoring of the Greenland Ice Sheet (PROMICE) (Jiang et al., 2018). These stations vary in elevation from 4 to 5,063 m above sea level and are located in areas with different land covers, including forest, grassland, shrubland, wetland, cropland, ice/snow, and urban areas. The collective in situ measurements, therefore, represent an accurate and comprehensive dataset that is capable of accounting for surface R_n spatial-temporal variation on a global scale. To deal with equipment and operational errors, daily mean surface R_n

145 measurements were calculated based on several strict processing rules (Jia et al., 2018; Jiang et al., 2014; Chen et al., 2020; Jiang et al., 2018). As shown in Fig. 1, the surface R_n measurements from 460 stations were used to train the proposed RCNN

model (red circles), while the measurements from the remaining 77 stations (blue circles) were selected as independent test

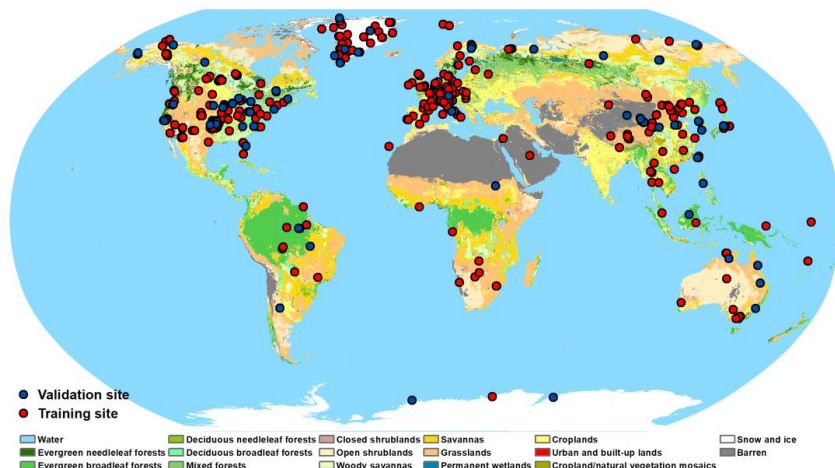
datasets to evaluate the model performance.

150

Figure 1: Geographic locations of the 537 ground stations used in this study. The 460 stations marked with red circles are used to train the RCNN model, while the 77 other stations marked with blue circles are used for the independent validation of the resulting trained model. Background colors indicate different land cover according to International Geosphere–Biosphere Programme (IGBP) classification system.

155

Table 2: Information about the observation networks. ARM: Atmospheric Radiation Measurement, BSRN: Baseline Surface Radiation Network, CEOP-Int: Coordinated Enhanced Observing Period, CEOP: Coordinated Enhanced Observation Network of China, EOL: Earth Observing Laboratory, GAME.ANN: GEWEX Asian Monsoon Experiment, GCNET: Greenland Climate Network, IMAU-Ktransect: Institute for Marine and Atmospheric Research Ice and Climate, PROMICE: Programme for

160 **Monitoring for the Greenland Ice Sheet, SURFRAD: Surface Radiation Budget Network.**

Network/Program	Sensors	Temporal Interval	Number of sites
ARM	Kipp&Zonen CNR-1	10 minutes	34
AsiaFlux	Kipp&Zonen CNR-1	30 minutes	31
BSRN	Kipp&Zonen CG4/Eppley. PIR	1 minutes	21
CEOP	\	30 minutes	16
CEOP-Int	Kipp&Zonen CG4/Eppley. PIR	30 minutes	8
ChinaFlux	Kipp&Zonen CNR-1	30/60 minutes	3
EOL	Kipp&Zonen pyrgeometers, Eppley	30/60 minutes	17
GCNEET	Li Cor Photodiode & REBS Q*7	60 minutes	18
GAME.ANN	EKO MS0202F	30 minutes	3
Global FluxNet	Kipp&Zonen CNR-1, etc	30 minutes	314
HiWATER	Kipp&Zonen CNR-1/CNR-4	10 minutes	19
IMAU-Ktransect	Kipp&Zonen CNR-1	60 minutes	4
LBA-ECO	Kipp&Zonen CG2/CNR-1	30 minutes	8
PROMICE	Kipp&Zonen CNR-1/CNR-4	10 minutes	24
SURFRAD	Eppley, Spectrosun	1/3 minutes	7

2.2 AVHRR data

AVHRR TOA observations at five spectral channels (a visible band (0.55–0.68 μm), a near-infrared band (0.75–1.1 μm), a middle-infrared band (3.55–3.93 μm), and two thermal bands (10.5–11.3 and 11.5–12.5 μm, respectively) were utilized for their comprehensive surface and atmospheric electromagnetic information. The National Aeronautics and Space Administration's (NASA) Land Long-term Data Record (LTDR) project produced a consistent long-term dataset by reprocessing Global Area Coverage (GAC) data, which were obtained from AVHRR sensors onboard the National Oceanic and Atmospheric Administration (NOAA) satellites (Pedelty et al., 2007). The primary reprocessing improvements included radiometric in-flight vicarious calibrations for the visible and near infrared channels, along with inverse navigation to relate a specific Earth location to each sensor's instantaneous field of view (Vermote and Kaufman, 1995; Pedelty et al., 2007). Multiple Climate Modeling Grid (CMG) data from AVHRR and MODIS instruments have been created for land climate studies (Xiao et al., 2017; Pedelty et al., 2007). In this study, we utilized a daily AVHRR TOA data product (AVH02C1) with a resolution of 0.05° from 1981 to 2019 to retrieve surface R_n estimates. Additionally, solar/viewing geometry data (i.e., SZA, VZA, and RAA) were also incorporated into the model as the amount of solar radiation incident on the Earth's surface varies greatly under different geometric observation conditions. A summary of these gridded products and their attributes is presented in Table 3.

Table 3: List of the satellite and reanalysis products used in this study.

Product names	Sensors	Spatial resolution	Temporal resolution	References
AVH02C1	AVHRR	0.05°×0.05°	daily	Pedelty et al. (2007); (Vermote and Kaufman, 1995)
GLASS07B11	MODIS	0.05°×0.05°	daily	Jiang et al. (2018); (Liang et al., 2020)
CERES-SYN	CERES/MODIS	1°×1°	1 hour	Doelling et al. (2016); (Doelling et al., 2013)
MERRA2	–	0.5°×0.625°	3 hours	Gelaro et al. (2017a)

2.3 GLASS product

The GLASS daily surface R_n product from MODIS data, one part of the GLASS product suite (Liang et al., 2020), was produced using two sets of algorithms. The main algorithm primarily uses the well-documented conversion relationships between downward shortwave radiation and all-wave R_n combinations (Wang and Liang, 2009; Jiang et al., 2015). It also incorporates a combination of other meteorological variables under various environmental conditions, such as different daytime lengths and land cover characteristics, which are designated based on the albedo and normalized difference vegetation index (NDVI). Multiple MARS learners were employed to establish efficient statistical relationships that allowed land surface R_n to be estimated from these inputs across most spatial domains (Jiang et al., 2016; Jiang et al., 2015). Conversely, when
surface solar radiation data were not available, the backup algorithm created a function that separately employed MODIS TOA observations to retrieve surface all-wave R_n using the length ratio of daytime (LRD) classification method, which was

190   accomplished by the genetic algorithm-artificial neural network (GA-ANN) (Chen et al., 2020). By using these two algorithms, the GLASS R_n product can provide seamless global land surface R_n estimates with a 0.05° resolution. Several studies have used in situ measurements to conduct evaluation studies, illustrating high accuracy performance as well as good application potential (Jiang et al., 2018; Guo et al., 2020). Thus, we used the GLASS daily R_n product covering 2000 to 2018 as a reference to help select reliable sites and validate the results from this study.

195   **2.4 CERES-SYN product**

The CERES instruments onboard the Terra, Aqua, and Suomi National Polar-Orbiting Partnership (Suomi NPP) satellites observe the TOA global energy budget by measuring shortwave reflected radiation, longwave Earth-emitted radiation and all wavelengths of radiation at a spatial resolution of 20 km at nadir (Wielicki et al., 1996). The CERES Synoptic (CERES-SYN) product combines CERES and MODIS observations with geostationary (GEO) data to provide hourly broadband TOA radiant

200   flux and cloud properties (Doelling et al., 2013). The CERES-SYN product also contains computed TOA and in-atmosphere and surface fluxes based on the Fu-Liou radiation transfer model. Aerosol and atmospheric data were included as inputs to calculate the radiation flux. CERES-SYN fluxes were provided as a 1° gridded product with an hourly temporal resolution. CERES-SYN surface R_n data have been evaluated in many studies, which indicate that the product has high accuracy, although systematic overestimation exists in the surface net radiation flux data (Jia et al., 2018; Jiang et al., 2016; Jia et al., 2016). Thus,

205   the CERES-SYN surface R_n obtained from four surface radiative components was used as a reference for comparison.

    2.5 MERRA2 reanalysis

MERRA2, produced by the NASA Global Modeling and Assimilation Office (GMAO), is the latest global atmospheric product and employs satellite observation data from 1980 to the present. The MERRA2 reanalysis assimilates space-based observations of aerosols and represents their interactions with other physical processes in the climate system. The goals of MERRA2 are to

210   provide a regularly gridded, homogeneous record of the global atmosphere, and to incorporate additional climatic variables and conditions, including trace gas constituents (stratospheric ozone), improved land surface representation, and cryospheric processes (Gelaro et al., 2017b). The MERRA2 products have a $0.5° \times 0.625°$ spatial resolution and hourly temporal resolution. Previous studies (Jiang et al., 2018; Jia et al., 2018; Delgado-Bonal et al., 2020) have confirmed that the MERRA2-calculated surface R_n and its radiative component provide outstanding accuracy and a reasonable spatial-temporal distribution compared

215   to other reanalysis data. Therefore, MERRA2 R_n data calculated from four surface radiative components were also used in this study to help retrieve accurate high-resolution surface R_n estimates by providing average information about the diurnal variation of daily surface R_n.

3 Methods

The entire workflow of the RCNN-based PSME method is shown in Fig. 2. First, the ETC technology was applied to the triplet

220 system constructed from ground-, satellite-, and model-based R_n data to identify reliable sites at which measurements can well

represent the dynamic variation of surface R_n at a 5 km scale. Then, AVHRR TOA reflectance, brightness temperature, angular

information and MERRA2 R_n were matched with the ground-based R_n measurements, both spatially and temporally.

Specifically, the site-measured R_n data were collocated with the 5 km AVHRR grid product covering the site. If one grid in

the AVH02C1 product covered multiple sites, the mean values from these sites' measurements were used to match the grid

225 data. Subsequently, the matched input-output training samples were fed into the RCNN to train the model. Reference R_n

measurements taken from reliable sites were used to evaluate the model's performance and, subsequently, identify the best

option to produce surface R_n by tuning the hyper-parameters of the RCNN. Finally, surface R_n retrievals were generated using

the best-trained model for the global scale, and CERES-SYN and GLASS R_n products were applied to perform inter-

comparisons to illustrate the accuracy and spatiotemporal variation of the surface R_n retrievals.

230

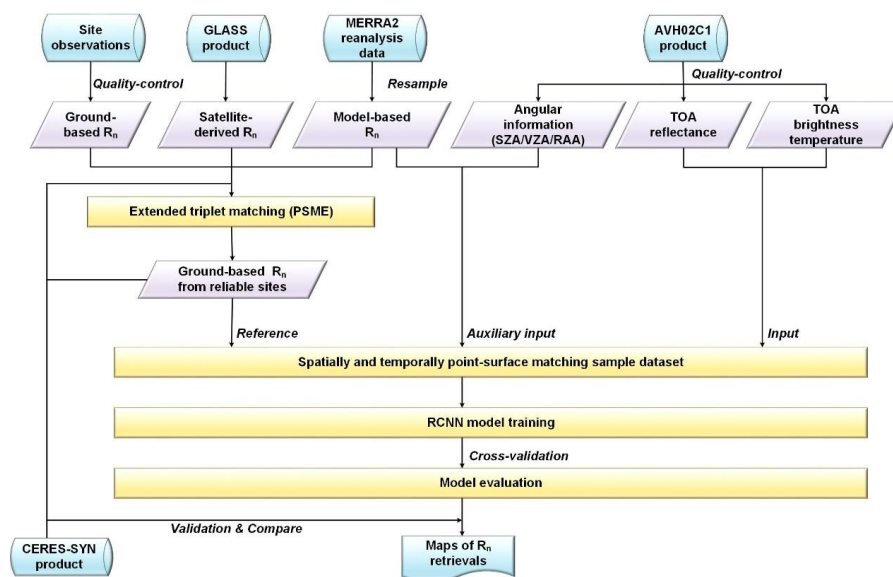



Figure 2: Workflow of RCNN-based point-surface matching estimation (PSME) method for surface R_n retrievals. TOA: Top of Atmosphere; RCNN: Residual Convolution Neural Network; GLASS: Global Land Surface Satellite; CRES-SYN: Clouds and the Earth's Radiant Energy System Synoptic; MERRA2: Modern-Era Retrospective analysis for Research and Applications, Version

235 **2; SZA: Solar Zenith Angle; VZA: Viewing Zenith Angle; RAA: Relative Azimuth Angle.**

3.1 Extended triple collocation (ETC)

To address the spatial scale mismatch issue owing to the small spatial support of sparse ground measurements in comparison

to gridded satellite data, which introduces large uncertainty into the collaborative inversion process, triplet collocation (TC)

technology was employed (Stoffelen, 1998; Yuan et al., 2020a). Specifically, based on the availability of three collocated,

240   independent measurement systems describing the same geophysical variable, TC was designed to estimate the unknown error

standard deviations (or RMSEs) of three mutually independent measurement systems, without treating any one system as

perfectly observed "truth" (Stoffelen, 1998; Gruber et al., 2016). To perform the TC, the following assumptions were made:

1) each of the triplet is related with the unknown truth of the geophysical variable in the linear form; 2) zero cross-correlation

across each of the triplet; 3) zero error cross-correlation between the triplet and the true signal state (T); and 4) the signal and

245   error statistics are stationary (Chen et al., 2017). Following the first assumption, the independent triplet systems (X_i, X_j, and

X_k) are related to the unknown true quantity in a linear error model:

$$X_i = \beta_i + \alpha_i T + \varepsilon_i \qquad (1)$$

where α_i and β_i are the additive and multiplicative bias terms, respectively; and ε_i is the mean-zero random error. Similar

calibration constants (α_j, β_j, α_k, and β_k) and random error terms (ε_j and ε_k) are also defined for X_j and X_k.

250   The objective of TC is to find a solution that individually estimates the variance of the random error term (ε_i) for each of

the triplet based on the listed assumptions (Stoffelen, 1998; Yuan et al., 2020a). However, to obtain the calibration constants,

one dataset is chosen from the three collocated measurement systems as the reference dataset, and the other two are rescaled

into the same reference data space. This results in a dependency of the error variance of the other two datasets on the

climatology of the scaling reference (Draper et al., 2013; Yuan et al., 2020a). To deal with this issue, ETC technology was

255   proposed by Mccoll et al. (2014), based on the same assumptions as TC, to estimate an additional evaluation metric

independent of the reference dataset, i.e., a correlation coefficient ($\rho_{(T,X_i)}$) of each measurement system with respect to the

unknown target variable as formulated below:

$$\rho(T, X_i) = sign(\pm)\sqrt{\frac{Cov(X_iX_j)Cov(X_iX_k)}{Cov(X_iX_i)Cov(X_jX_k)}} \qquad (2)$$

where $\rho_{(T,X_i)}$ is correct up to the sign ambiguity, as the measurement systems will almost always be positively correlated to

260   the unobserved truth.

Following the ideals of Chen et al. (2017) and Yuan et al. (2020a), ETC was applied to determine the reliable site measurements

over the AVHRR data footprint scale (i.e., 5 km grid). Specifically, the triplet dataset was first constructed using ground-based

measurements (i.e., site measurements), satellite-derived retrievals (GLASS R_n) and downscaled model-based simulations

(MERRA2 R_n) depending on the conversion ratio of GLASS R_n between original (0.05°) and aggregated (0.5°) spatial

265   resolutions, as they belong to different measurement systems and are not dependent on each other. Then, $\rho_{(T,X_i)}$ was

calculated using Eq. (2) for individual site, illustrating the fraction of 5 km satellite footprint-scale R_n dynamics captured by

point-scale ground-based measurements. An appropriate threshold of $\rho_{(T,X_i)}$ was determined to select reliable sites with the

greatest representativeness within a 5 km footprint to obtain sample datasets for the model training. After testing a series of

thresholds between 0.2 and 0.9 at intervals of 0.1, a threshold of 0.9 was selected, above which the sites were assigned as

270   'reliable' (also see Sect. 5).

3.2 RCNN-based PSME

3.2.1 RCNN

Loosely inspired by the human visual cortex, CNNs were originally applied to analyze common visual imagery using convolution instead of general matrix multiplication (Ball et al., 2017). A general CNN consists of multiple layers of operations,

275 such as convolution, pooling, normalization, and nonlinear activation functions. In the convolutional layers, a series of convolution (*Conv*) operations are performed using convolutional kernel weights and biases on the input images within the receptive field. The result of the locally weighted sum (feature map) is then passed through a nonlinear layer, such as a rectified linear unit (*ReLU*), which increases the nonlinear properties of the decision function and the overall network (Romanuke, 2017). The pooling layer, a form of non-linear down-sampling, merges semantically similar features into one, thereby reducing

280 the amount of computation in the network (Géron, 2019). Additionally, a batch normalization layer is placed between the convolutional layers and nonlinearities to speed up the training of the CNN and reduce the sensitivity to network initialization. By stacking two or three stages of convolution, nonlinearity, and pooling, followed by more fully connected layers, a typical CNN architecture is built.

To improve network performance in complicated tasks, a deeper CNN architecture is needed; however, a deeper neural network

285 is difficult to train well because a degradation problem occurs with deeper networks converge (He et al., 2016). Specifically, as the network depth increase, the training accuracy becomes saturated and then degrades rapidly. To address the degradation problem, He et al. (2016) proposed a residual block to improve the gradient flow through the network , which enables the training of deeper networks. Residual blocks were employed in our CNN architecture to construct the RCNN. The structure of the RCNN proposed in this study is shown schematically in Fig. 3. Table 4 lists the detailed configurations of the proposed

290 RCNN.

The input signals of the RCNN included AVHRR TOA reflectance and brightness temperature, angular information (SZA, VZA, and RAA), and daily MERRA2 R_n with a spatial size of 15×15 pixels (these specifications are further discussed in Sect. 5). To avoid introducing new errors, nearest-neighbor interpolation was used to resample MERRA2 R_n to 0.05° to match the spatial resolution of the other predictors. The output signal was ground-based R_n from the reliable sites. Essentially, the RCNN

295 model uses a convolution operation taken at stages of the feature map and residual learning block to recognize spatial patterns centered on a reliable site. Then, the multiple layer perceptron links abstract spatial patterns with ground-based measurements to construct a strong non-linear relationship to reproduce the spatial and temporal variation of surface R_n. This approach had been carried out in previous studies (Jiang et al., 2019b; Jiang et al., 2020a; Jiang et al., 2019a).

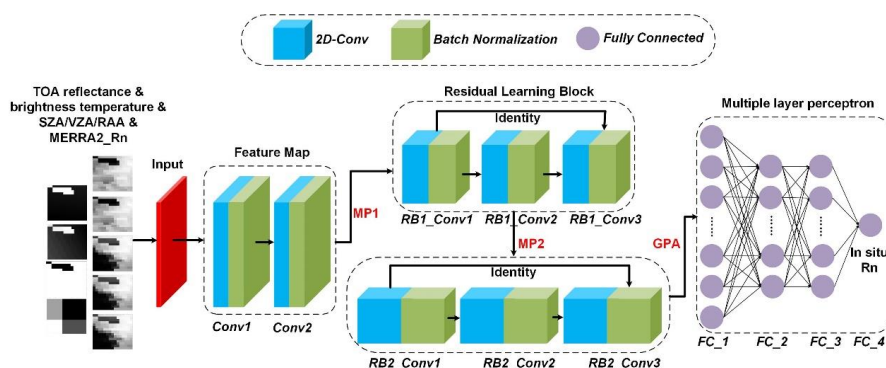



Figure 3: Depiction of the RCNN network. The model uses AVHRR TOA observations, angular information (SZA, VZA, and RAA), and MERRA-2 R$_n$ as inputs, which are used to calculate daily surface R$_n$ values as output. *Conv* **represents the convolution operation;** *MP* **and** *GAP* **are the max-pooling and global average-pooling operations, respectively;** *RB***: residual block;** *FC***: fully-connected layer.**

Table 4: Detailed configuration of the RCNN. *eLU***: exponential linear unit;** *DROP***: dropout layer.**

Module	Unit	Input size	Kernel num.	Kernel size	Stride	Activation function	Output size
	Input	$9 \times 15 \times 15$	—	—	—	—	$9 \times 15 \times 15$
Feature	*Conv1*	$9 \times 15 \times 15$	32	3×3	[1, 1]	*ReLU*	$32 \times 15 \times 15$
Mapping	*Conv2*	$32 \times 15 \times 15$	32	3×3	[1, 1]	*ReLU*	$32 \times 15 \times 15$
	MP1	$32 \times 15 \times 15$	—	2×2	[2, 2]	—	$32 \times 7 \times 7$
Residual	*RB1_Conv1*	$32 \times 7 \times 7$	64	3×3	[1, 1]	*ReLU*	$64 \times 7 \times 7$
Learning	*RB1_Conv2*	$64 \times 7 \times 7$	64	3×3	[1, 1]	*ReLU*	$64 \times 7 \times 7$
Block	*RB1_Conv3*	$64 \times 7 \times 7$	64	3×3	[1, 1]	*ReLU*	$64 \times 7 \times 7$
	MP2	$64 \times 7 \times 7$	—	2×2	[2, 2]	—	$64 \times 3 \times 3$
	RB2_Conv1	$64 \times 3 \times 3$	128	3×3	[1, 1]	*ReLU*	$128 \times 3 \times 3$
	RB2_Conv2	$128 \times 3 \times 3$	128	3×3	[1, 1]	*ReLU*	$128 \times 3 \times 3$
	RB2_Conv3	$128 \times 3 \times 3$	128	3×3	[1, 1]	*ReLU*	$128 \times 3 \times 3$
	GPA	$128 \times 3 \times 3$	—	—	[0, 0]	—	$128 \times 1 \times 1$
Multiple	*FC_1*	128	—	—	—	*eLU*	128
Layer	*DROP*	128	—	—	—	—	128
perceptron	*FC_2*	128	—	—	—	*eLU*	64
	FC_3	64	—	—	—	*eLU*	64
	FC4	64	—	—	—	*eLU*	1

3.2.2 RCNN model training and evaluation

Sample data from the reliable sites in the training site group (red circles as shown in Fig. 1) were used to train the RCNN model; datasets from reliable sites in the independent validation site group (blue circles shown in Fig. 1) served as test datasets to independently evaluate the model's performance. Specifically, in the training process, 10-fold cross-validation (CV) was used to test the model's predictive power. All of the sample datasets from the reliable training sites were randomly shuffled and divided into ten groups. One group of these data was them removed as a hold-out or validation dataset and the remaining nine groups of data were treated as the training datasets. The training datasets were used to fit the RCNN model, and the

validation datasets were applied to evaluate the trained model's performance to fine-tune the model's parameters. The process was repeated ten times to ensure that each group of data validated the model, and the remaining nine groups of data were trained. Finally, the evaluation results were presented by summarizing and averaging the ten evaluation scores. After determining the hyper-parameter settings using the CV, the model was trained again using datasets from all the reliable training sites, which was then independently evaluated using the test datasets from the reliable validation sites.

The following five evaluation metrics were used to evaluate the performance of the RCNN model and the R_n retrievals: bias, relative bias (rbias), RMSE, relative RMSE (rRMSE) and the coefficient of determination (R^2). Detailed information regarding the application of these metrics can be found in Yang et al. (2018).

4 Results

4.1 Identification of reliable sites

The reliable sites and unreliable sites identified a threshold of 0.9 for the ETC-derived correlation coefficient are summarized in Table 5 for each observation network. A total of 275 sites could be considered reliable, accounting for ~48% of the sites. Furthermore, no site was considered reliable for some observation networks/programs, namely ChinaFlux, GCNET, GAME.ANN, HiWATER, IMAU-Ktransect, and LBA-ECO. It is not surprising that sites from the ChinaFlux network were assigned as unreliable; several studies have revealed that the reliability of site measurements from China is questionable and should be examined carefully before use (Zhang et al., 2015; Tang et al., 2013; Tang et al., 2011). The GCNET network is located in inner Greenland, where systematic measurements errors are common due to difficulties in instrument maintenance and operation-related failures. Sites from the GAME.ANN network are located in the Tibetan Plateau (TP) region, where abnormal climate and complex terrain make it difficult to accurately measure variations in R_n. Similar issues also exist in in situ measurements from the IMAU-Ktransect, HiWATER and LBA-ECO networks. In contrast, some of the international observational networks, such as BSRN and FluxNet, provide many ground-based measurements with sufficient spatial representativeness for R_n at 5 km resolution. In addition, the ARM and SURFRAD networks were classified as containing reliable sites. In situ measurements from the SURFRAD network were well known in surface radiation budget studies because of their high data quality, and have been widely utilized as a result (Wang et al., 2015b; Hao et al., 2019; Wang et al., 2015a; Qin et al., 2012). Overall, compared to other networks, the sites from ARM, BSRN, SURFRAD, and FluxNet networks were mostly identified as reliable sites, illustrating the superiority of these observation networks.

The spatial and proportion distributions of the reliable training and validation sites for different surface types are presented in Fig. 4. The most reliable sites are distributed in the United States, Europe, and East Asia. In turn, many sites located in South America, Africa, Eurasia, and the Polar Regions were identified as unreliable. The reasons that these sites were classified as unreliable are closely related to the complex surface and atmospheric environment and poor instrument maintenance in their

corresponding regions. Most grassland and cropland sites were classified as reliable (~66% and ~62%, respectively), whereas the fewest reliable sites were classified in ice/snow-covered areas (~14%). In addition, sites neighboring the sea were mostly

345   identified as unreliability due to the presence of large water bodies within the satellite footprint. Thus, the processing of identifying reliable sites highlights the need to pay more attention to such areas for in surface radiation estimations.

Table 5: Summary of the selected reliable and unreliable sites based on ETC for each observation network

Network/Program	Number of reliable sites	Number of unreliable sites
ARM	34	0
AsiaFlux	12	19
BSRN	14	7
CEOP	5	11
CEOP-Int	4	4
ChinaFlux	0	3
EOL	2	15
GCNEET	0	18
GAME.ANN	0	3
Global FluxNet	189	125
HiWATER	0	19
IMAU-Ktransect	0	4
LBA-ECO	0	8
PROMICE	6	24
SURFRAD	7	0

350

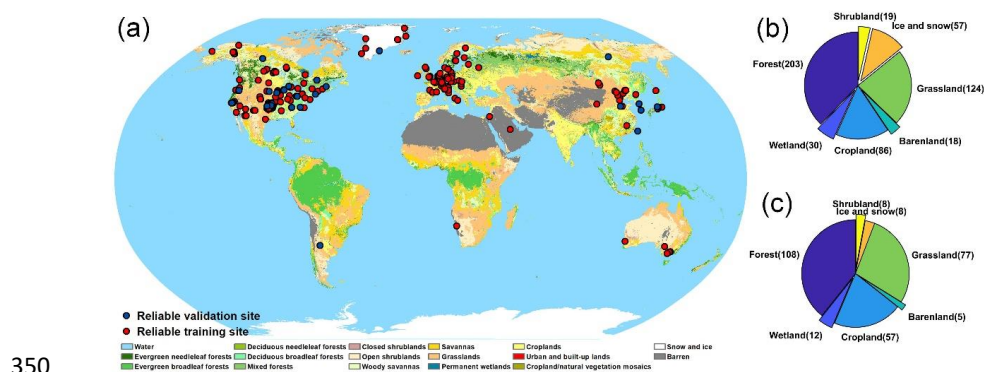



Figure 4: (a) Spatial distribution of reliable sites and the absolute numbers of (b) all sites and (c) reliable sites under different surface types.

4.2 Assessment of the RCNN model

Ten-CV was used to evaluate the performance of the RCNN model at reliable and all training sites, respectively, and the

355   evaluated performances are summarized in Table 6. Note that the model fitting result represents the model with the best fitting accuracy over the 10 CV rounds, while the cross-validation results are the averages of the 10-round combination. The RCNN model showed a high fitting accuracy at the reliable training sites with R^2, RMSE (rRMSE) and bias (rbias) values of 0.90, 20.67 Wm^{-2} (25.82%) and -0.18 Wm^{-2} (-0.23%), respectively. Compared to the model fitting accuracy across all sites, the

result for the reliable sites was improved, with R^2 values increased by 0.04 and rRMSE values reduced by 7.72%. The

360   implementation of ETC for the selection of reliable site ensures more consistently spatial representativeness of ground-based

measurements and AVHRR data, which improves the accuracy of R_n retrievals. Indeed, the CV-derived average accuracy is

extremely similar to the model fitting accuracy, illustrating that the trained RCNN model is highly robust. Additionally, an

unbiased estimation was achieved by the RCNN model with CV-derived biases close to zero.

Figure 5 shows the overall training accuracy and test accuracy for the RCNN model at reliable training and independent

365   validation sites. The over-training accuracy of the RCNN model is close to that of the CV-derived result. Between the model

training to the test phase, the R^2 score dropped by 0.06 and RMSE increased by 5.90 Wm^{-2} (5.6%), which indicates that slight

over-fitting by the proposed model. However, with the highest cross-validated R^2 of 0.90, and the lowest RMSE of 20.92

W/m^2, the RCNN model trained using ground-based R_n measurements obtained at the reliable sites is considered as the best-

trained model, which was selected for the subsequent analysis.

370

Table 6: Ten-fold cross-validation performance of the RCNN models.

Used sites	Model fitting					Cross-validation				
	R^2	RMSE	rRMSE	bias	rbias	R^2	RMSE	rRMSE	bias	rbias
All sites	0.86	25.64	33.74	-0.16	-0.21	0.86	25.77	33.85	-0.28	-0.36
Reliable sites	0.90	20.67	25.82	-0.18	-0.23	0.90	20.92	26.13	0.11	0.13

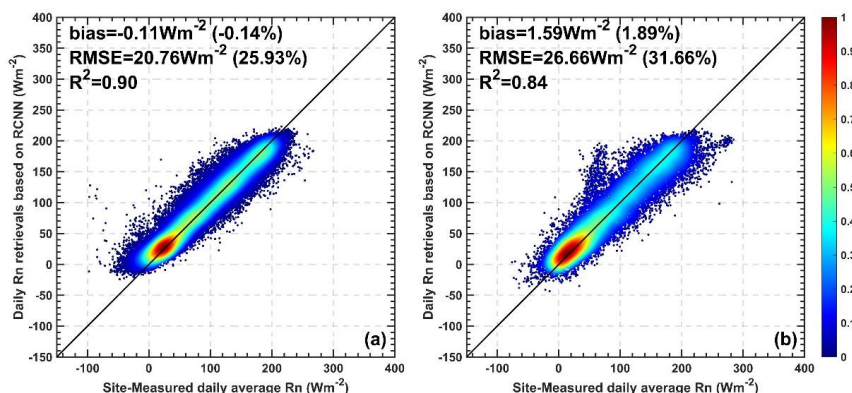



Figure 5: Scatterplots of (a) mode training (fitting) accuracy and (b) model test accuracy for the reliable training and independent

375   **validation sites.**

4.3 Evaluation of the RCNN-based AVHRR R_n retrievals

4.3.1 Inter-comparisons of R_n products

Figure 6 shows the validation results of the four datasets at the reliable sites, including the AVHRR, CERES-SYN, MERRA2,

380   and GLASS R_n estimates. Comparatively, the RCNN-derived AVHRR R_n retrievals show the best performance with R^2, RMSE,

and bias of 0.90, 20.99 Wm^{-2} (26.18%), -0.04 Wm^{-2} (-0.05%), respectively, followed by the GLASS R_n estimates with corresponding values of 0.89, 22.29 Wm^{-2} (27.81%) and -2.78 Wm^{-2} (-3.47%), respectively. The CERES R_n estimates show a notable overestimation at higher values against the in situ measurements with a bias of 7.25 Wm^{-2} (9.04%). In addition, greater uncertainty exists in the CERES R_n compared to the AVHRR and GLASS R_n estimates, with an RMSE of 25.11 Wm^{-2}. The

385   CERES-SYN cloud product, an input for the calculation of flux products, underestimated low-level clouds (by 11.8% and 20.9% for day and night, respectively) over the sun-glint ocean and polar regions both during daytime and nighttime (Xi et al., 2018; Xu et al., 2020). Additionally, when the aerosol optical depths (AODs) used to calculate CERES-SYN surface solar radiation are compared with the ground-based observations, the calculated shortwave radiation is 1–2% higher (Fillmore et al., 2021). Therefore, large uncertainties in these atmospheric input parameters may lead to serious overestimations of the CERES-

390   SYN R_n. In addition, the MERRA2 R_n shows the lowest accuracy, with an RMSE of 20.87 Wm^{-2}, reflecting the reanalysis model's inability to accurately describe the evolution of cloud properties (Betts et al., 2006). Comparatively, the AVHRR retrievals show a better accuracy than the other three products. In addition, our estimates have a higher spatial resolution compared to the CERES-SYN and MERRA2 data.

Compared to the validation results at the reliable sites, the accuracy evaluation at all sites shows the ability of the RCNN to

395   accurately capture R_n variation at a global scale, even though some measurements from unreliable sites added large uncertainties to the final evaluation. Figure S1 shows a comparison of results for the four datasets at all of the sites. Overall, the AVHRR and GLASS R_n retrievals were still better than those of CERES-SYN and MERRA2; however, the accuracy of AVHRR R_n decreases slightly, with R^2, RMSE, and bias values of 0.85, 26.51 Wm^{-2} (35.50%) and 1.32 Wm^{-2} (1.76%), which is comparable to the GLASS R_n retrievals, with values of 0.85, 26.56 Wm^{-2} (35.38%) and -0.83 Wm^{-2} (-1.10%), respectively.

400   Therefore, even if the RCNN model is trained using measurements from less reliable sites, it can accurately reproduce surface R_n distributions at global scale. In the following analysis, the GLASS R_n retrievals were used as the main comparison because of their high accuracy and reasonable spatiotemporal variation (Jia et al., 2018; Jiang et al., 2018).

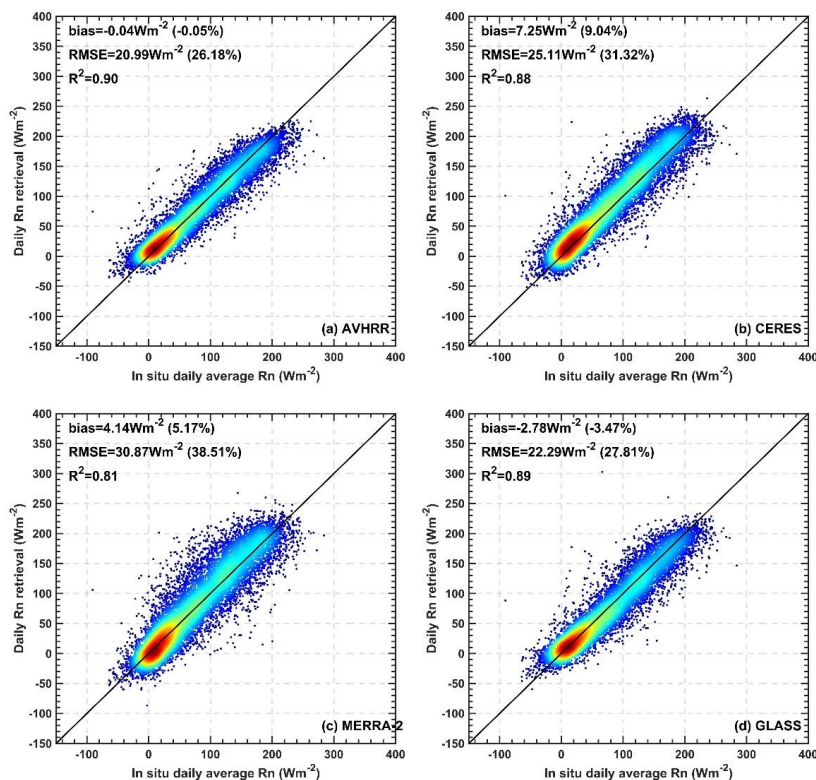



Figure 6: Scatterplots of product validation for (a) AVHRR, (b) CERES-SYN, (c) MERRA2, and (d) GLASS at the reliable sites.

405

Inter-comparison results for the AVHRR and GLASS R_n retrievals against the ground-based measurements over each network are displayed in Fig. 7. The AVHRR R_n retrievals performed better than the GLASS R_n retrievals in most of the observation networks. Specifically, the AVHRR R_n retrievals show lower RMSE values over seven networks, except for the EOL and PROMICE networks. However, the RMSE differences over the EOL and PROMICE networks are very small—only 1.4 and

410   0.3 Wm^{-2}, respectively. The EOL is a small regional network, and its measurements thus only reflect local-scale R_n variation. However, only two out of 17 sites were identified as reliable for the model training and, therefore, the RCNN model cannot capture specific R_n dynamics with such a small spatial extent. Similar reasons also account for the poor performance of the PROMICE network for most of the large sites in the GCNET and PROMICE networks identified as unreliable sites. Thus, the RCNN model has less knowledge of R_n dynamics under snow and ice surfaces. The most significant difference for RMSE was

415   observed over the ARM network, for which the mean RMSE value decreased by 2.1 Wm^{-2} for the AVHRR R_n retrievals relative to the GLASS R_n retrievals. Additionally, these two datasets showed very similar performance based on their R^2 values.

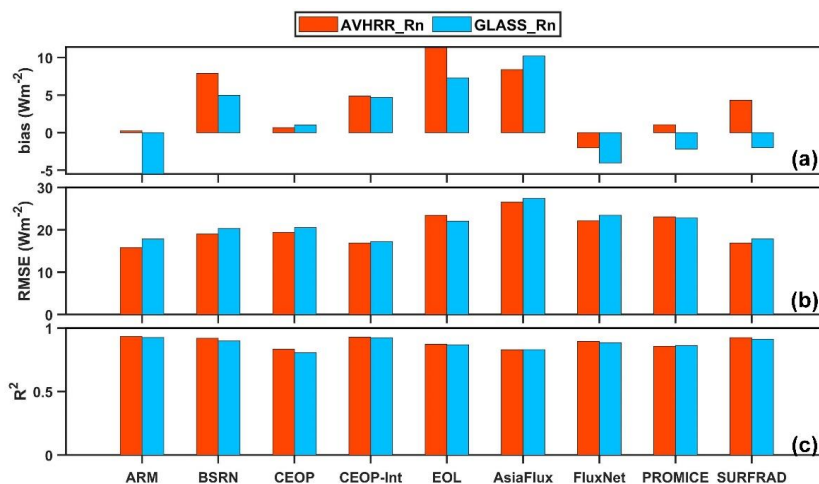



Figure 7: The average performance of the AVHRR and GLASS R_n retrievals against ground-based measurements at the reliable sites over each network.

420

To further improve understanding of the temporal variations of the AVHRR R_n retrievals, coincident time series from all the R_n datasets were inter-compared over seven sites representing different surface types, as shown in Fig. 8. Overall, all four datasets broadly captured the true dynamics of R_n under the different surface types. Comparatively, the AVHRR and GLASS R_n retrievals are more consistent with in situ measurements than the CERES-SYN and MERRA2 products. Specifically, the

425 MERRA2 and CERES-SYN R_n retrievals show higher values compared to the in situ measurements at the BSRA_DRA site, especially during 140–200 day period. In comparison, the AVHRR and GLASS R_n values closely match the ground-based measurements, and thereby better reflect the true temporal variation in R_n. At the Lath_UK-AMo site, four of the datasets slightly overestimated during the summer; however, the AVHRR and GLASS R_n retrievals still performed best. Moreover, large discrepancies occurred at the PM-SCO_U site under the snow/ice surface for the four datasets. Notably, MERRA2 R_n

430 values do not reflect true variations under snow and ice, especially during the 150–250 day period. Comparatively, the satellite-derived retrievals better capture R_n dynamics, although the CERES-SYN product still exhibits overestimation. Because less reliable sites were screened at the global scale to train the RCNN model under the snow/ice surfaces, the AVHRR R_n values only capture the general R_n trend in the snow/ice areas. The GLASS R_n retrievals are also most consistent with the in situ measurements among the four datasets, although small biases still exist.

435 The overall evaluation results of the AVHRR and GLASS R_n retrievals for different surface types are displayed in Table 7. Generally, both datasets achieved high accuracy, with RMSEs ranging from 20 to 25 Wm^{-2}. The AVHRR R_n retrievals show the better performance for most of the surface types, except for snow/ice, as previously discussed; however, the difference in the RMSEs between the AVHRR and GLASS R_n retrievals for the snow/ice cover type is small (0.5 W/m^{-2}). Together, these results further indicate that the RCNN model can generate accurate R_n estimates for different land cover types.

440

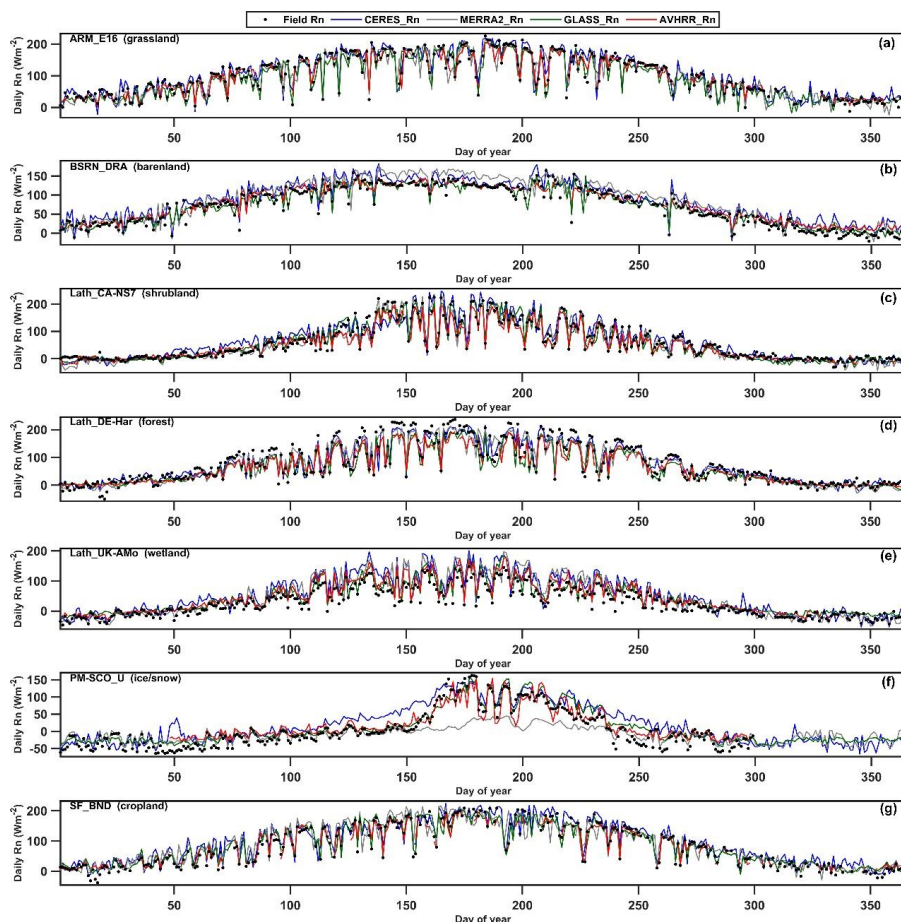



Figure 8: Coincident time series of the AVHRR, GLASS, MERRA2, CERES-SYN Rₙ retrievals and ground-based measurements over seven sites representing different surface cover types for (a) ARM_E06 (38.061°, -99.134°), (b) BSRN_DRA (36.626°, -116.018°), (c) Lath_CA-NS7 (56.635°, -99.948°), (d) Lath_DE-Har (47.934°, 7.601°), (e) Lath_UK-AMo (55.791°, -3.238°), (f) PM-SCO_U (72.393°, -27.233°) and (g) SF_BND (40.050°, -88.37°).

445

Table 7: Evaluation of the AVHRR and GLASS Rₙ retrievals for different surface cover types.

Surface types	AVHRR R_n					GLASS R_n				
	R^2	RMSE	rRMSE	bias	rbias	R^2	RMSE	rRMSE	bias	rbias
Forest	0.90	22.38	27.12	-4.18	-5.07	0.89	23.73	28.75	-5.16	-6.26
Crop	0.90	21.18	27.05	3.57	4.56	0.89	22.07	28.18	-0.87	-1.11
Grass	0.92	18.70	22.80	4.01	4.88	0.90	20.37	24.83	-0.57	-0.70
Shrub	0.90	18.52	21.93	-0.88	-1.04	0.89	20.08	23.78	-2.03	-2.40
Ice/snow	0.85	24.46	76.37	1.84	5.75	0.86	23.96	74.84	-0.59	-1.83
Barren	0.90	15.23	21.48	4.01	5.66	0.87	17.27	24.37	1.87	2.64
Wetland	0.91	23.15	27.72	-7.21	-8.63	0.90	23.96	28.69	-7.88	-9.44

4.3.2 Analysis of influencing factors

450   Variation in surface R_n is mainly affected by atmospheric conditions, but also, to a lesser degree, by surface characteristics (He et al., 2015). Under clear-sky conditions, AOD and column water vapor (CWV) are the main atmospheric constituents that modulate surface shortwave and longwave radiation, and further affect spatiotemporal variations of surface R_n. In contrast, clouds and CWV control surface R_n dynamics under cloud-sky conditions, especially clouds, which have significant impacts on surface shortwave and longwave radiation. Therefore, AOD, CWV, and cloud optical thickness (COT, as a surrogate for

455   cloud optical properties) were employed to analyze the sensitivity of the accuracy of the AVHRR and GLASS R_n retrievals to variations in these influencing factors. In addition, R_n retrieval performance at different elevations was also evaluated.

All the evaluation results are displayed in Figure 9. The AVHRR R_n retrievals were always better than the GLASS R_n retrievals under all conditions of the four influencing factors, except for elevations ranging 800–1000 and 1200–1500 m, which demonstrates the superiority of our algorithm. Specifically, as the COT increases (i.e., increasing cloud thickness), the AVHRR

460   and GLASS R_n RMSE values increase accordingly but still remain relatively low for both datasets (< 27 Wm^{-2}). Note that the differences in RMSE between the two datasets also increased with increasing COT (Fig, 9(a)). A small COT indicates relative clear-sky conditions, which results in surface total solar radiation dominated direct solar radiation. Therefore, the performance of the RCNN model and the MARS models used for the GLASS R_n product (Jiang et al., 2016) is comparable the accuracy of the R_n retrievals. However, when the absorption and scattering effects are enhanced for direct solar radiation from TOA,

465   depending on the individual pixel approximation (IPA), it is difficult to retrieve the total surface R_n accurately using the MARS model because the spatially adjacent effects (i.e., 3-D effects from clouds) are not considered. Although the RMSEs of the AVHRR retrievals also increase, the rate of increase is lower than that of the GLASS R_n retrievals. This is because the RCNN model recognizes spatial textural and contextual information and comprehensively considers atmospheric conditions within a certain area rather than an IPA, which to some extent addresses the spatially adjacent effect on the accuracy of AVHRR R_n

470   retrievals.

Aerosols also have absorption and scattering influences on solar radiation, and therefore, a similar conclusion can be drawn. For example, when AOD increases from 0.3 (a non-clean atmosphere), the difference in the performance of the two retrieved model becomes more pronounced. The AVHRR R_n retrievals maintain a stable level of uncertainty (RMSEs = 22–24 Wm^{-2}), while the errors GLASS R_n retrieval errors increase dramatically (up to 29 Wm^{-2}). This illustrates the importance of integrating

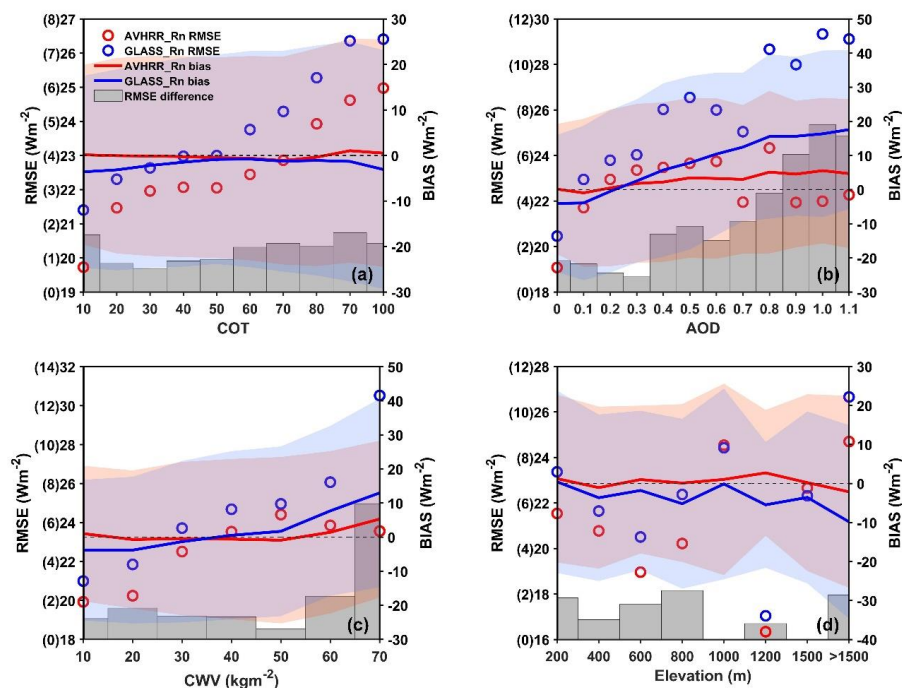
475   the spatial adjacent effect into the inversion model under hazy atmospheric conditions.

In the case of CWV, which has a strong influence on longwave radiation, when the condition is < 50 kgm^{-2}, the accuracy differences between the AVHRR and GLASS R_n values are small (< 1 Wm^{-2}). However, as CWV increases, the RMSEs of the GLASS R_n retrievals increase dramatically, while the AVHRR R_n estimates maintain a high level of accuracy (RMSEs = ~23 Wm^{-2}).

480   With respect to elevation, there was no notable difference between the two datasets; our estimates were better than GLASS R_n under different elevation ranges, except for the 800–1,000 and 1,200–1,500 m bins, although these differences were less than 1 Wm^{-2}. The lower accuracy of AVHRR R_n values for these two elevation ranges is attributable to the less reliable sites used for the RCNN training. In addition, the AVHRR R_n retrievals show steady and vary low (close to zero) biases under different conditions, while the biases of the GLASS R_n retrievals show a high degree of variation. This illustrates that the RCNN model

485   has a greater capability for unbiased surface R_n estimation.

    Overall, the RCNN-derived R_n retrievals show a high accuracy under different atmospheric and surface conditions relative to the GLASS R_n retrievals and especially for thick-cloud and hazy atmospheric conditions. In such cases, spatially adjacent information is important for accurately estimating the surface R_n retrievals. Although previous studies have proposed several methods for integrating spatial information to retrieve surface and atmospheric variables, such as PM$_{2.5}$ (Li et al., 2020b; Wang

490   et al., 2020a), ozone (Li and Cheng, 2021), and nitrogen dioxide (Li et al., 2020a), these methods only considered discrete surrounding points within a certain area to train the model using IPA. This artificially destroys the natural correlation between the target and the surroundings. Our RCNN model can automatically recognize complete spatial information centered at an interesting location and, thus, is a more reasonable and effective method.

495   **Figure 9: Accuracy changes in AVHRR and GLASS R_n retrievals under different conditions for (a) cloud optical thickness (COT), (b) aerosol optical depth (AOD), (c) column water vapor (CWV), and (d) elevation. The values in parentheses on the left-axis correspond to the RMSE differences denoted by bar charts. The shaded area show the variance ranges of the biases.**

4.3.3 Spatiotemporal analysis

500   The global-scale spatial distributions of mean AVHRR and GLASS surface R_n for January and July 2008 are displayed in Fig. 10. The missing values in the Polar Regions reflects the unavailability of valid data at the five bands in the case of the AVH02C1 product. The overall distribution of surface R_n for the two datasets is very similar, although slight differences exist in some regions, such as the TP region, the Sahara Desert, and Greenland. AVHRR R_n retrievals are notably larger than the GLASS R_n retrievals in the TP region. Based on the results shown in Fig. 9(d), greater confidence can be placed in the AVHRR R_n retrievals for high-elevation regions relative to GLASS R_n retrievals; however, the AVHRR R_n retrievals are relatively

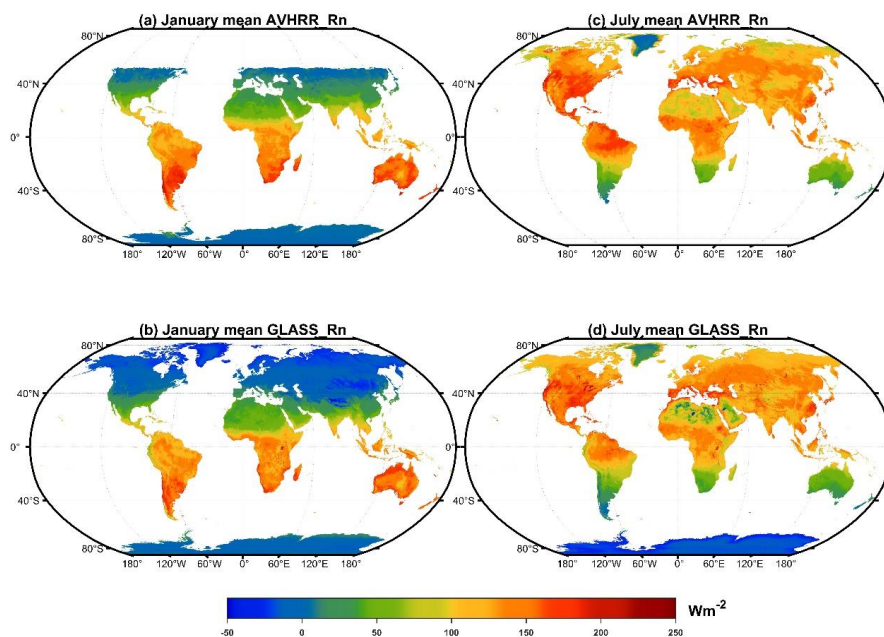
505   lower in Greenland. Because few sites from the GCNET and PROMICES networks were classified as reliable for the model training, the RCNN model has less knowledge about the spatiotemporal variations of R_n in Greenland compared to other regions. The validation results in Fig. 8 and Table 7 for the ice/snow surface cover type further confirm that GLASS R_n product may offer a better performance in Greenland region. Therefore, new algorithms and data are required for the Polar Regions to address this problem.

510

Figure 10: Spatial distribution of monthly mean AVHRR and GLASS R_n retrievals in January (a, b) and July (c, d) 2008.

The spatiotemporal consistency of the AVHRR and GLASS daily R_n retrievals against COT was examined at a global scale in January and July, respectively, as shown in Fig. 11. Overall, the spatial consistency is high for the two datasets. Specifically,

515   in January, as the COT increases, the mean R_n values and the absolute differences between the two datasets also increase. As shown in Fig. 9(a), when COT increases, the AVHRR R_n retrievals are more accurate. Thus we believe that the large

discrepancies under high COT conditions are mainly attributed to the GLASS R_n retrievals. In July, surface R_n remained

relatively stable under different COT conditions, and the absolute differences between the two datasets also remain steady,

with a mean absolute bias of 20 Wm^{-2}.

520 Based on our previous analysis, spatially adjacent information is important for surface R_n estimation when COTs values are

large; however, if the cover of the entire cloud layer is small compared to the scale of the AVHRR footprint, the spatial adjacent

effects will be significantly weakened in the inversion process, even if the corresponding COT is large. IPA-provided

information includes the properties of the entire cloud layer. Figure S2 shows the spatial distribution of the monthly mean

cloud cover fraction (CF) at the global scale in January and July, and the corresponding differences in CFs. In January, the

525 CFs are higher over most land regions except in northern Africa, southern America, northern Austria and southern Asia.

However, most regions had small CFs in July. The differences in CFs for the two months are also marked; the positive

differences demonstrate that more than 72% of the land pixels had a higher CFs in January than in July. The spatial adjacent

effects induced by clouds are more significant on surface R_n in January than in July. Therefore, when large and thick cloud

layers exist, such as in the Polar Regions, CNN is a better choice for surface R_n estimation, especially for downward longwave

530 radiation (DLR) because the temperature of cloud-base, which is an essential variable in the parameterized calculation of DLR,

is difficult to retrieve from multispectral remote sensing (Yang and Cheng, 2020).

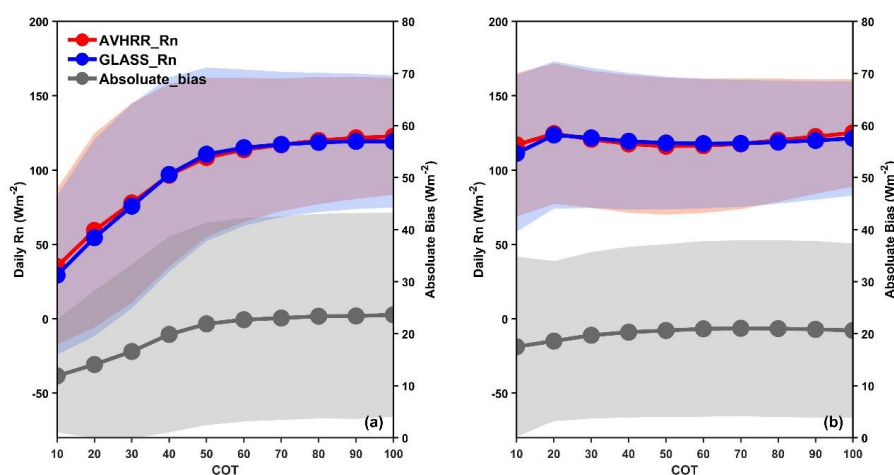



Figure 11: Variations in the spatial and temporal consistency of AVHRR and GLASS daily R_n retrievals against cloud optical thickness (COT) in (a) January and (b) July 2008.

535 **4.3.4 Temporal analysis**

To examine the temporal reliability of the generated AVHRR R_n dataset, a long-term analysis of surface R_n for the four datasets

was carried out, the results of which are shown in Fig. 12. In view of missing values for the Polar Regions, we focused on

surface R_n within the ± 60° latitudes region.

540 Overall, the AVHRR R_n retrievals are highly consistent with MERRA2 R_n values during the period of 1981 to 2019 as well as CERES and GLASS R_n retrievals after 2000. However, the MERRA2 and CERES R_n values are generally higher than AVHRR R_n retrievals. Inter-comparison results illustrated that the CERES and MERRA2 R_n values are overestimated against ground-based measurements. GLASS R_n temporal profile is more consistently correlated with the AVHRR R_n retrievals. Specifically, some large discrepancies occur, during some periods including 1994–1995, 1999–000, 2007–2008, and 2018–2019. These periods correspond to the alternative update times of the NOAA-series satellites. For example, NOAA-11 was successfully

545 produced by NOAA-14 from 1994 to 1995. Similarly, NOAA-16 replaced NOAA-14 in 2000 for monitoring of the Earth's surface and atmosphere. During these periods of satellite replacement, the corresponding AVHRR data contain large gaps. For example, no data are available for the winter of 1994, as shown in Fig. 12(a). Similarly, NOAA-20 was launched on November 18, 2017, yet the quality of the AVHRR TOA observations from this platform was poor during the second half of the year, with many missing values, especially in 2018 and 2019. This explains the abnormal temporal variations in the AVHRR R_n

550 profile in these years shown in Fig. 12(a). Therefore, effective multi-source data fusion algorithms and spatial gap-filling technology are urgently needed to improve the quality and coverage of the AVHRR R_n dataset.

The temporal variations in monthly R_n anomalies for the four datasets are shown in Fig. 12(b). High temporal consistency exists between AVHRR R_n anomalies and the other three datasets. Specifically, the correlation coefficient for the AVHRR and MERRA2 R_n anomalies for 1981–2019 is 0.952, and for the period after 2000, is 0.957 and 0.956 for the AVHRR and CERES,

555 and AVHRR and GLASS, R_n anomalies, respectively. Thus, the RCNN-derived AVHRR R_n dataset is temporally stable and reliable when the other three R_n datasets are used as a comparative baseline. In fact, the LTDR project has adapted a calibration method that can be consistently applied across the AVHRR instruments onboard various NOAA satellites to account for sensor degradation (Vermote and Kaufman, 1995), which enables a temporally reliable R_n dataset to be produced. Following the approach, overall, our AVHRR R_n dataset is more accurate and shows reasonable spatiotemporal variations compared to the

560 other three datasets. This dataset will play an important role in climate change study.

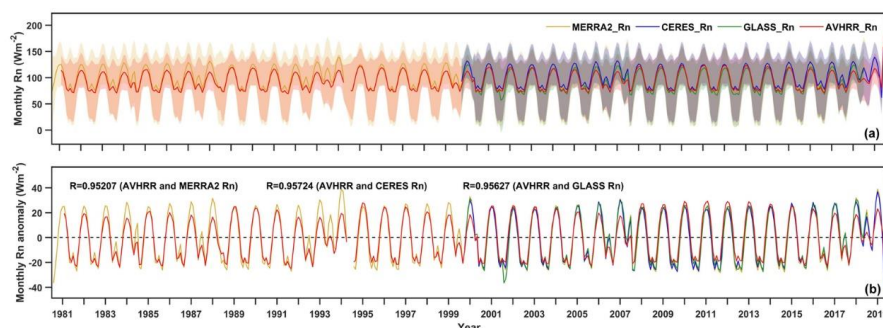



Figure 12: Long-term temporal variation of (a) monthly average R_n and (b) monthly R_n anomalies for the AVHRR, CERES, GALSS and MERRA2 datasets, respectively.

5 Discussion

5.1 Determination of an appropriate spatial scale

To provide appropriate AVHRR sub-image blocks containing sufficient information for the RCNN model to generate high-accuracy retrievals, the spatial adjacent effects on surface R_n under different valid spatial extents should be examined. For this, we used a simple multivariate linear regression (MLR) model (see supplementary data for further details). The spatial sizes of the sub-images denoted as B3 … B19 vary from 3×3 to 19×19, respectively, with an interval of 2 pixels. The true areas correspond to approximately 15×15 km^2 (B3) to 135×135 km^2 (B19) on the ground. The results are shown in Fig. 13. Overall, the average R increases from 0.61 to 7.08, and RMSE decreases from 50.12 to 46.17, respectively, for the MLR model. As the valid spatial extent increases, essential and complete spatial features are exposed and incorporated into the MLR model, which helps to continuously improve the model's retrieval accuracy. The spatial extent of B13 (approximately 65×65 km^2) is the smallest size that exhibits convergent R and RMSE values, and the spatial extent at the B15 reaches a more stable state for surface R_n estimations. This finding is in line with the results of Jiang et al. (2020b), showing that scale effects have a considerable impact on solar radiation retrieval accuracy, and distances of approximately 20 to 40 km from the central point (corresponding to areas of 40×40 km^2 to 80×80 km^2), are optimal spatial scale. In addition, previous studies (Hakuba et al., 2013; Huang et al., 2016) recommended a threshold distance of approximately 30 km, equal to a 13×13 grid region with a spatial resolution of 0.05°, for shortwave radiation estimation. Therefore, a 15×15 grid area was selected for the input sub-images to generate AVHRR R_n retrievals.

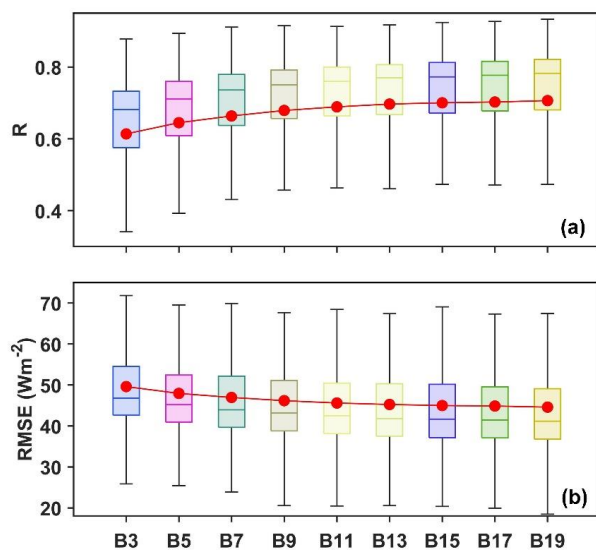



Figure 13: Variations of (a) R and (b) RMSE indices for each spatial scale in the MLR model. The red lines in the subplots are the average curves of indices at the different spatial scales.

5.2 Role of daily mean MERRA2 R_n

585

The RCNN model uses instantaneous satellite-sensed signals to directly estimate daily mean R_n retrievals. Though some previous studies (Chen et al., 2020; Wang et al., 2015a; Wang and Liang, 2017; Xu et al., 2020) directly estimated daily average surface radiation from instantaneous satellite observations, like MODIS, the idea is theoretically flawed because the AVHRR sensor only offers instantaneous "snapshots", which cannot capture daily mean information about the diurnal cycles of the atmosphere and clouds. King et al. (2013) acknowledged that the frequency of cloud variations is high at different times and locations based on twin MODIS cloud products. In view of the wide overpass times representing different instantaneous atmospheric conditions for different AVHRR sensors, daily mean MERRA2 R_n is incorporated into the input collection to provide daily mean information about the surface, atmosphere and clouds for the RCNN model.

590

Figure 14 shows the effect of the daily mean MERRA2 R_n on the final AVHRR R_n retrievals at different AVHRR overpass times. The improved effect is slightly more significant during the afternoon than in the morning when more over-land clouds are present (King et al., 2013). This improvement is also more pronounced during the night. The AVHRR R_n retrievals can only be obtained when solar radiation is available (Fig. 10) because of the missing values in the AVH01C1 product; therefore, the results during the night are based on the validation results for high latitudes, which demonstrate that daily mean information about the diurnal cycles of the atmosphere and clouds is more important for daily surface radiation estimation at high latitudes than that at middle and low latitudes. Study illustrated the cloud coverage in Arctic summer could be as high as 80% (Shupe et al., 2011).

595

600

Additionally, MERRA2 downward shortwave radiation (DSR) was used as a replacement for MERRA2 R_n to test its contribution to daily mean surface R_n estimations when using instantaneous satellite data. The results presented in Fig. S3 show that the improved effect of daily MERRA2 DSR is not comparable with that achieved using daily MERRA2 R_n, and the former is closer to the results obtained when only instantaneous AVHRR observations are used. Therefore, MERRA2 R_n is a meaningful input for the RCNN model. Moreover, the AVHRR R_n retrievals could also be further improved by using more accurate daily mean R_n data, such as GLASS R_n, or other parameters that accurately represent daily mean atmospheric and cloud variations.

605

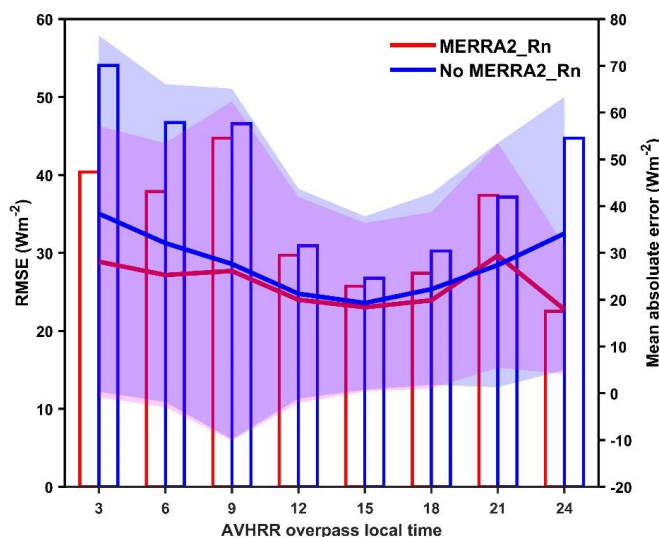



Figure 14: Effect of daily mean MERRA2 R_n on AVHRR R_n retrievals at different satellite overpass times. The bars indicate RMSE and lines indicates absolute biases.

5.3 Determination of a threshold for the ETC-derived correlation coefficient

The threshold for the ETC-derived correlation coefficient between in situ measurements and the unknown truth within the 5 km AVHRR grid in Eq. (2) affects the selection of reliable sites and the subsequent PSME process. A series of thresholds for the ETC-derived coefficients were considered, ranging from 0.2 to 0.9 with an interval of 0.1. In each case, the corresponding measurements from the selected reliable sites were fed into the RCNN model to train and subsequently generate AVHRR R_n retrievals. Then, the training and test accuracies of the RCNN were calculated over all of the reliable sites for comparison. Another important consideration is the representativeness of the RCNN for global R_n estimation, given that the number of reliable training sites decreases with higher thresholds. Thus, the trained RCNN model was again evaluated at all sites, including reliable and unreliable sites, to examine the global representativeness. The number of reliable sites (training and test accuracies) and the associated global accuracy are presented in Fig. 15. As the threshold increases, the number of reliable sites decreased. The training and test relative RMSEs of the RCNN model showed a general decreasing trend, especially above a threshold of 0.5, which illustrates that the selection of reliable sites and the measurements from these sites have better representativeness for the AVHRR footprint scale using ETC. This helps address the spatial scale mismatch issue and improve the accuracy of AVHRR retrievals at a 5 km resolution. In addition, a trade-off between the RCNN's fitting accuracy at the reliable sites and the global accuracy at all sites needs to be considered. Even when a threshold of 0.9 was used, the global accuracy of the RCNN was only slightly lower, which explains why this threshold was applied in Sect. 3.1.

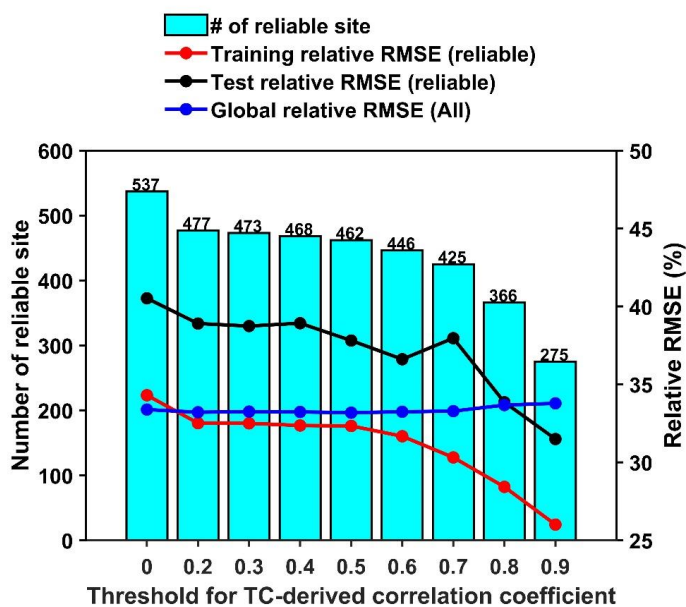
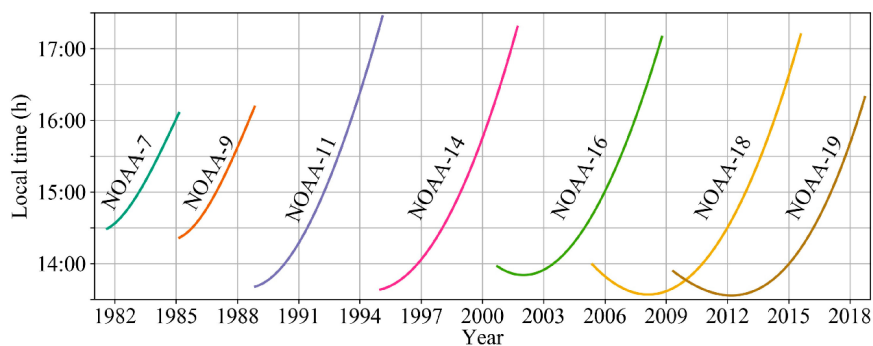



Figure 15: Effect of extended triplet collocation (ETC)-derived correlation coefficients on the number of reliable sites and the
corresponding RCNN's training, test, and global accuracies.

5.4 Orbital drift of the NOAA-series satellites

The orbital drift problem of the NOAA-series satellites has attracted the attention of many users applying AVHRR-derived high-level remote sensing products, such as land surface temperature (LST) (Ma et al., 2020; Liu et al., 2019) and TOA albedo (Song et al., 2018). As shown in Fig. 16, the orbital drift makes the true equatorial crossing time (ECT) of the NOAA-series afternoon satellites range from 13:00 to 17:30 in the solar time system. Previous geophysical variable retrievals based on AVHRR data are instantaneous values at satellite overpass times, which need to be corrected to a specific time, such as LST at 14:30 and albedo at noon local time. However, the RCNN model uses instantaneous AVHRR TOA observations at different satellite overpass times to directly retrieve daily surface R_n estimates, which differs from previous studies.

Additionally, daily MERRA2 R_n and instantaneous SZA values closely related to satellite transit times are taken as inputs; therefore, the RCNN model can automatically learn the relationships between instantaneous satellite data at different overpass times and corresponding daily surface R_n measurements. Moreover, the results of the long-term temporal analysis of the AVHRR R_n dataset provide more evidence to ensure that the quality of the long-term AVHRR daily R_n datasets is not affected by orbital-drift. As such, orbital drift does not affect our long-term AVHRR R_n dataset.

645   **Figure 16: Equatorial Crossing Time (ECT) for the National Oceanic and Atmospheric Administration (NOAA)-series afternoon Satellites (Liu et al., 2019).**

6 Data availability

Global surface R_n retrieved from NOAA/AVHRR data from 1981 to 2019 are freely available at https://doi.org/10.5281/zenodo.5509854 for 1981-2019 (Xu et al., 2021).

650   The AVH02C1 product data were downloaded from Level-1 and Atmosphere Archive & Distribution System Distributed Active Archive Center (https://ladsweb.modaps.eosdis.nasa.gov/, last access: 2 July 2021, NASA, 2021). CERES-SYN product was downloaded from CERES team (https://ceres.larc.nasa.gov/, last access: 2 July 2021, NASA, 2021). GLASS R_n product was provided by GLASS team at (http://www.glass.umd.edu/, last access: 2 July 2021, Beijing Normal University and University of Maryland, College Park, 2021). MERRA2 reanalysis was downloaded from the Global Modelling and

655   Assimilation Office (https://gmao.gsfc.nasa.gov/, last access: 2 July 2021, NASA, 2021). The download links of ground-based measurements from different observational networks were referenced to Jiang et al. (2018).

7 Conclusions and outlook

A long-term (1981–2019) global daily surface R_n product with spatial resolution of 0.05° was generated from historical NOAA-series AVHRR data using a RCNN-based PSME method. The specific steps employed were as follows: (1) selecting reliable

660   sites from all sites based on ETC to generate the sample dataset; (2) training and independent testing of the proposed RCNN model; (3) evaluating the AVHRR R_n retrievals against in situ measurements and performing inter-comparisons with three other R_n products (GLASS, CERES-SYN, and MERRA2); and (4) generating and evaluating the long-term AVHRR R_n product.

ETC was first applied to select reliable sites to prepare a sample dataset with better spatial representativeness at the AVHRR

665   footprint scale (i.e., 5 km). In total, 275 sites were classified as reliable sites from a total of 537 sites and used as a sample dataset for the RCNN model. The proportions of the selected reliable sites representing cropland and grassland surfaces were highest (~66% and ~62%, respectively), while those representing ice/snow surface were lowest (~14%). The sample dataset

from the 275 sites ensured that the trained RCNN model had both a good fitting accuracy for the reliable sites and global accuracy across all sites.

670   A simple MLR model was used to examine the spatial adjacent effects on surface R_n estimation, and a spatial extent of 15×15 pixels (75×75 km^2) was then determined as the input size of the RCNN to provide sufficient spatial information. Based on 10-fold CV, the trained RCNN model achieved an R^2 of 0.90, with an RMSE of 20.76 Wm^{-2} (25.93%), and a bias of -0.11 Wm^{-2} (-0.14%); the corresponding independent validation values were 0.84, 26.66 Wm^{-2} (31.66%), and 1.59 Wm^{-2} (1.89%) at the reliable sites, respectively. These results demonstrate the overall ability of the RCNN model to accurately predict surface

675   R_n.

The results of an inter-comparison between the AVHRR R_n retrievals and three other products illustrated that our retrievals show a better accuracy against in situ measurements, with an R^2 of 0.90, a RMSE of 20.99 Wm^{-2} (26.18%), and a bias of -0.04 W/m^2 (-0.05%) at the reliable sites, and an R^2 of 0.85, a RMSE of 26.51 Wm^{-2} (35.30%), and a bias of 1.32 Wm^{-2} (1.76%) across all sites. At the same time, the AVHRR R_n retrievals show better performance for different observational networks and

680   surface cover types, except for the snow/ice surface cover. Under different elevations and atmospheric conditions, the AVHRR R_n retrievals performed better than the GLASS R_n equivalent, especially in the presence of thick clouds and hazy atmospheric conditions because of the integration of spatially adjacent information into the inversion process in the RCNN model. In addition, the spatiotemporal variation of the AVHRR R_n retrievals is similar to that of the GLASS R_n values, demonstrating the ability of the RCNN model to generate a long-term global R_n product.

685   The long-term global R_n dataset generated by the RCNN model displays high accuracy and reasonable spatiotemporal variation at the global scale, which is suited to many applications including, for example, studies to understand the radiation budget and global climate change. However, further research is needed to solve some problems to further improve the data quality of the AVHRR R_n dataset. First, new algorithms and satellite data are needed to estimate surface R_n in the Polar Regions, such as MODIS data (Chen et al., 2020). Second, an effective data gap-filling method or multi-source data-fusion algorithm is required

690   to fill the data gaps over land, especially during periods of satellite replacement works. Third, coupled with spatially adjacent information, real-time temporal information, or historical information should be incorporated to further improve the accuracy of the R_n retrievals.

As a type of machine learning, deep learning involves using data-driven models to find potential relationships and patterns, and offers high adaptability to training data sample inputs. The predictive ability of a data-driven model completely depends

695   on the limitations of the training dataset and in the case of R_n, the ability of the model to accurately portray spatiotemporal dynamics in areas where the availability of training data is relatively poor, such as for AVHRR R_n retrievals for ice/snow-covered surface. To address this problem, more physical knowledge is needed to fully utilize data-driven modeling to estimate surface R_n under different atmospheric and surface conditions. In particular, more attention should be paid to understanding

inherent physical processes in addition to obtaining optimal estimation by coupling physical process models with the versatility

700 of data-driven machine learning (Reichstein et al., 2019).

Author contributions

JX and SL contributed to the design of this study and developed the overall methodology. JX and BJ collected and preprocessed the data. JX carried out the experiment and produced the product. JX wrote the first draft. All authors revised the manuscript.

Competing interests

705 All authors declare that they have no conflicts of interest.

Acknowledgements

The authors would like to thank the LTDR project for providing AVH02C1 product, CERES team for offering CERES-SYN flux product, NASA team for MERRA2 product and GLASS team for GLASS R_n product. We express our gratitude to each observation network project for providing ground-based radiation measurements, including ARM, BSRN, AsiaFlux, GCNET,

710 Global FluxNet, PROMICE and other networks/programs.

Financial support

This work was supported by the Chinese Grand Research Program on Climate Change and Response (project 2016YFA0600103) and the National Natural Science Foundation of China (grant 41971291).

References

715 Ball, J. E., Anderson, D. T., and Chan, C. S.: Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community, Journal of Applied Remote Sensing, 11, 042609, 2017.
Barker, H. W. and Li, Z.: Interpreting shortwave albedo-transmittance plots: True or apparent anomalous absorption?, Geophysical research letters, 24, 2023-2026, 1997.
Betts, A. K., Zhao, M., Dirmeyer, P., and Beljaars, A.: Comparison of ERA40 and NCEP/DOE near-surface data sets with
720 other ISLSCP-II data sets, Journal of Geophysical Research: Atmospheres, 111, 2006.
Bisht, G. and Bras, R. L.: Estimation of net radiation from the MODIS data under all sky conditions: Southern Great Plains case study, Remote Sensing of Environment, 114, 1522-1534, 2010.
Bisht, G., Venturini, V., Islam, S., and Jiang, L.: Estimation of the net radiation using MODIS (Moderate Resolution Imaging Spectroradiometer) data for clear sky days, Remote Sensing of Environment, 97, 52-67, 2005.
725 Chen, F., Crow, W. T., Colliander, A., Cosh, M. H., Jackson, T. J., Bindlish, R., Reichle, R. H., Chan, S. K., Bosch, D. D., Starks, P. J., Goodrich, D. C., and Seyfried, M. S.: Application of Triple Collocation in Ground-Based Validation of Soil Moisture Active/Passive (SMAP) Level 2 Data Products, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10, 489-502, 10.1109/JSTARS.2016.2569998, 2017.

730 Chen, J., He, T., Jiang, B., and Liang, S.: Estimation of all-sky all-wave daily net radiation at high latitudes from MODIS data, Remote Sensing of Environment, 245, 111842, 2020.

Cheng, J. and Liang, S.: Global estimates for high-spatial-resolution clear-sky land surface upwelling longwave radiation from MODIS data, IEEE Transactions on Geoscience and Remote Sensing, 54, 4115-4129, 2016.

Cheng, J., Liang, S., Wang, W., and Guo, Y.: An efficient hybrid method for estimating clear-sky surface downward longwave radiation from MODIS data, Journal of Geophysical Research: Atmospheres, 122, 2616-2630,

735 https://doi.org/10.1002/2016JD026250, 2017.

Crow, W. T., Berg, A. A., Cosh, M. H., Loew, A., Mohanty, B. P., Panciera, R., de Rosnay, P., Ryu, D., and Walker, J. P.: Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products, Reviews of Geophysics, 50, 2012.

Delgado-Bonal, A., Marshak, A., Yang, Y., and Holdaway, D.: Analyzing changes in the complexity of climate in the last

740 four decades using MERRA-2 radiation data, Scientific Reports, 10, 922, 10.1038/s41598-020-57917-8, 2020.

Doelling, D. R., Sun, M., Nguyen, L. T., Nordeen, M. L., Haney, C. O., Keyes, D. F., and Mlynczak, P. E.: Advances in geostationary-derived longwave fluxes for the CERES synoptic (SYN1deg) product, Journal of Atmospheric and Oceanic Technology, 33, 503-521, 2016.

Doelling, D. R., Loeb, N. G., Keyes, D. F., Nordeen, M. L., Morstad, D., Nguyen, C., Wielicki, B. A., Young, D. F., and

745 Sun, M.: Geostationary Enhanced Temporal Interpolation for CERES Flux Products, Journal of Atmospheric and Oceanic Technology, 30, 1072-1090, 10.1175/jtech-d-12-00136.1, 2013.

Draper, C., Reichle, R., de Jeu, R., Naeimi, V., Parinussa, R., and Wagner, W.: Estimating root mean square errors in remotely sensed soil moisture over continental scale domains, Remote Sensing of Environment, 137, 288-298, 2013.

Fillmore, D., Rutan, D., Kato, S., Rose, F., and Caldwell, T.: Evaluation of aerosol optical depths and clear-sky radiative

750 fluxes of the CERES Edition 4.1 SYN1deg data product, Atmos. Chem. Phys. Discuss., 2021, 1-50, 10.5194/acp-2021-283, 2021.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., and Reichle, R.: The modern-era retrospective analysis for research and applications, version 2 (MERRA-2), Journal of climate, 30, 5419-5454, 2017a.

755 Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), Journal of Climate, 30, 5419-5454, 10.1175/jcli-d-16-0758.1, 2017b.

760 Géron, A.: Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems, O'Reilly Media2019.

Gruber, A., Su, C.-H., Zwieback, S., Crow, W., Dorigo, W., and Wagner, W.: Recent advances in (soil moisture) triple collocation analysis, International Journal of Applied Earth Observation and Geoinformation, 45, 200-211, 2016.

Guo, X., Yao, Y., Zhang, Y., Lin, Y., and Bei, X.: Discrepancies in the Simulated Global Terrestrial Latent Heat Flux from

765 GLASS and MERRA-2 Surface Net Radiation Products, Remote Sensing, 12, 2763, 2020.

Hakuba, M. Z., Folini, D., Sanchez-Lorenzo, A., and Wild, M.: Spatial representativeness of ground-based solar radiation measurements, Journal of Geophysical Research: Atmospheres, 118, 8585-8597, 2013.

Hao, D., Asrar, G., Zeng, Y., Zhu, Q., Wen, J., Xiao, Q., and Chen, M.: Estimating hourly land surface downward shortwave and photosynthetically active radiation from DSCOVR/EPIC observations, Remote Sensing of Environment, 232, 111320,

770 2019.

Hao, D., Wen, J., Xiao, Q., Wu, S., Lin, X., Dou, B., You, D., and Tang, Y.: Simulation and analysis of the topographic effects on snow-free albedo over rugged terrain, Remote Sensing, 10, 278, 2018.

He, K., Zhang, X., Ren, S., and Sun, J.: Deep Residual Learning for Image Recognition, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778, 10.1109/CVPR.2016.90, 2016.

775 He, T., Liang, S., Wang, D., Shi, Q., and Goulden, M. L.: Estimation of high-resolution land surface net shortwave radiation from AVIRIS data: Algorithm development and preliminary results, Remote Sensing of Environment, 167, 20-30, 2015.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., and Schepers, D.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999-2049, 2020.

Huang, G., Li, X., Lu, N., Wang, X., and He, T.: A General Parameterization Scheme for the Estimation of Incident
780 Photosynthetically Active Radiation Under Cloudy Skies, IEEE Transactions on Geoscience and Remote Sensing, 2020.

Huang, G., Ma, M., Liang, S., Liu, S., and Li, X.: A LUT-based approach to estimate surface solar irradiance by combining MODIS and MTSAT data, Journal of Geophysical Research: Atmospheres, 116, 2011.

Huang, G., Li, X., Huang, C., Liu, S., Ma, Y., and Chen, H.: Representativeness errors of point-scale ground-based solar radiation measurements in the validation of remote sensing products, Remote sensing of environment, 181, 198-206, 2016.

785 Husi, LETU, Jiancheng, SHI, Ming, LI, Tianxing, WANG, Huazhe, and SHANG: A review of the estimation of downward surface shortwave radiation based on satellite data: Methods, progress and problems, ence China(Earth ences), v.63, 26-41, 2020.

Jia, A., Jiang, B., Liang, S., Zhang, X., and Ma, H.: Validation and spatiotemporal analysis of CERES surface net radiation product, Remote Sensing, 8, 90, 2016.

790 Jia, A., Liang, S., Jiang, B., Zhang, X., and Wang, G.: Comprehensive assessment of global surface net radiation products and uncertainty analysis, Journal of Geophysical Research: Atmospheres, 123, 1970-1989, 2018.

Jiang, B., Zhang, Y., Liang, S., Zhang, X., and Xiao, Z.: Surface Daytime Net Radiation Estimation Using Artificial Neural Networks, Remote Sensing, 6, 11031-11050, 2014.

Jiang, B., Liang, S., Jia, A., Xu, J., Zhang, X., Xiao, Z., Zhao, X., Jia, K., and Yao, Y.: Validation of the Surface Daytime
795 Net Radiation Product From Version 4.0 GLASS Product Suite, IEEE Geoscience and Remote Sensing Letters, 16, 509-513, 2018.

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

Jiang, B., Zhang, Y., Liang, S., Wohlfahrt, G., Arain, A., Cescatti, A., Georgiadis, T., Jia, K., Kiely, G., Lund, M.,
800 Montagnani, L., Magliulo, V., Ortiz, P. S., Oechel, W., Vaccari, F. P., Yao, Y., and Zhang, X.: Empirical estimation of daytime net radiation from shortwave radiation and ancillary information, Agricultural & Forest Meteorology, 211-212, 23-36, 2015.

Jiang, H., Lu, N., Qin, J., and Yao, L.: Surface global and diffuse solar radiation over China acquired from geostationary Multi-functional Transport Satellite data, Earth Syst. Sci. Data Discuss., 2019, 1-22, 10.5194/essd-2019-209, 2019a.

805 Jiang, H., Lu, N., Qin, J., and Yao, L.: Hourly 5-km surface total and diffuse solar radiation in China, 2007–2018, Scientific Data, 7, 311, 10.1038/s41597-020-00654-4, 2020a.

Jiang, H., Lu, N., Qin, J., Tang, W., and Yao, L.: A deep learning algorithm to estimate hourly global solar radiation from geostationary satellite data, Renewable & Sustainable Energy Reviews, 114, 109327, 2019b.

Jiang, H., Lu, N., Huang, G., Yao, L., Qin, J., and Liu, H.: Spatial scale effects on retrieval accuracy of surface solar
810 radiation using satellite data, Applied Energy, 270, 115178, 2020b.

Jiao, Z., Yan, G., Zhao, J., Wang, T., and Chen, L.: Estimation of surface upward longwave radiation from MODIS and VIIRS clear-sky data in the Tibetan Plateau, Remote Sensing of Environment, 162, 221-237, 2015.

Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., Papale, D., Schwalm, C., Tramontana, G., and Reichstein, M.: The FLUXCOM ensemble of global land-atmosphere energy fluxes, Scientific data, 6, 1-14, 2019a.

815 Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., Papale, D., Schwalm, C., Tramontana, G., and Reichstein, M.: The FLUXCOM ensemble of global land-atmosphere energy fluxes, Scientific Data, 6, 74, 10.1038/s41597-019-0076-8, 2019b.

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S.-K., Hnilo, J., Fiorino, M., and Potter, G.: Ncep–doe amip-ii reanalysis (r-2), Bulletin of the American Meteorological Society, 83, 1631-1644, 2002.

820   King, M. D., Platnick, S., Menzel, W. P., Ackerman, S. A., and Hubanks, P. A.: Spatial and temporal distribution of clouds observed by MODIS onboard the Terra and Aqua satellites, IEEE transactions on geoscience and remote sensing, 51, 3826-3852, 2013.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., and Endo,
H.: The JRA-55 reanalysis: General specifications and basic characteristics, Journal of the Meteorological Society of Japan.
825   Ser. II, 93, 5-48, 2015.

Li, T. and Cheng, X.: Estimating daily full-coverage surface ozone concentration using satellite observations and a spatiotemporally embedded deep learning approach, International Journal of Applied Earth Observation and Geoinformation, 101, 102356, 2021.

Li, T., Wang, Y., and Yuan, Q.: Remote Sensing Estimation of Regional NO2 via Space-Time Neural Networks, Remote
830   Sensing, 12, 2514, 2020a.

Li, T., Shen, H., Yuan, Q., and Zhang, L.: Geographically and temporally weighted neural networks for satellite-based mapping of ground-level PM 2.5, ISPRS Journal of Photogrammetry and Remote Sensing, 167, 178-188, 2020b.

Liang, S., Wang, D., He, T., and Yu, Y.: Remote sensing of earth's energy budget: synthesis and review, International Journal of Digital Earth, 12, 737-780, 2019.

835   Liang, S., Wang, K., Zhang, X., and Wild, M.: Review on estimation of land surface radiation and energy budgets from ground measurement, remote sensing and model simulations, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 3, 225-240, 2010.

Liang, S., Cheng, J., Jia, K., Jiang, B., Liu, Q., Xiao, Z., Yao, Y., Yuan, W., Zhang, X., and Zhao, X.: The Global LAnd Surface Satellite (GLASS) product suite, Bulletin of the American Meteorological Society, 1-37, 2020.

840   Liu, X., Tang, B.-H., Yan, G., Li, Z.-L., and Liang, S.: Retrieval of global orbit drift corrected land surface temperature from long-term AVHRR data, Remote Sensing, 11, 2843, 2019.

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

845   Lu, N., Qin, J., Yang, K., and Sun, J.: A simple and efficient algorithm to estimate daily global solar radiation from geostationary satellite data, Energy, 36, 3179-3188, 2011.

Ma, J., Zhou, J., Göttsche, F. M., Liang, S., Wang, S., and Li, M.: A global long-term (1981-2000) land surface temperature product for NOAA AVHRR, Earth Syst. Sci. Data Discuss., 2020, 1-43, 10.5194/essd-2020-143, 2020.

McColl, K. A., Vogelzang, J., Konings, A. G., Entekhabi, D., Piles, M., and Stoffelen, A.: Extended triple collocation:
850   Estimating errors and correlation coefficients with respect to an unknown target, Geophysical research letters, 41, 6229-6236, 2014.

Pedelty, J., Devadiga, S., Masuoka, E., Brown, M., Pinzon, J., Tucker, C., Vermote, E., Prince, S., Nagol, J., Justice, C.,
Roy, D., Junchang, J., Schaaf, C., Jicheng, L., Privette, J., and Pinheiro, A.: Generating a Long-term Land Data Record from the AVHRR and MODIS Instruments, 2007 IEEE International Geoscience and Remote Sensing Symposium, 23-28 July
855   2007, 1021-1025,   10.1109/IGARSS.2007.4422974,

Pinker, R. and Laszlo, I.: Modeling surface solar irradiance for satellite applications on a global scale, Journal of Applied Meteorology and Climatology, 31, 194-211, 1992.

Qin, J., Yang, K., Liang, S., and Tang, W.: Estimation of Daily Mean Photosynthetically Active Radiation under All-Sky Conditions Based on Relative Sunshine Data, Journal of Applied Meteorology & Climatology, 51, 150-160, 2012.

860   Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., and Carvalhais, N.: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195-204, 2019.

Román, M. O., Schaaf, C. B., Woodcock, C. E., Strahler, A. H., Yang, X., Braswell, R. H., Curtis, P. S., Davis, K. J., Dragoni, D., and Goulden, M. L.: The MODIS (Collection V005) BRDF/albedo product: Assessment of spatial representativeness over forested landscapes, Remote Sensing of Environment, 113, 2476-2498, 2009.

865   Romanuke, V.: Appropriate number and allocation of ReLUs in convolutional neural networks, Наукові вісті Національного технічного університету України Київський політехнічний інститут, 69-78, 2017.

Rossow, W. and Zhang, Y. C.: Calculation of surface and top of atmosphere radiative fluxes from physical quantities based on ISCCP data sets: 2. Validation and first results, Journal of Geophysical Research: Atmospheres, 100, 1167-1197, 1995.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., and Behringer, D.: The
870   NCEP climate forecast system reanalysis, Bulletin of the American Meteorological Society, 91, 1015-1058, 2010.

Sellers, P., Dickinson, R., Randall, D., Betts, A., Hall, F., Berry, J., Collatz, G., Denning, A., Mooney, H., and Nobre, C.: Modeling the exchanges of energy, water, and carbon between continents and the atmosphere, Science, 275, 502-509, 1997.

Shupe, M. D., Walden, V. P., Eloranta, E., Uttal, T., Campbell, J. R., Starkweather, S. M., and Shiobara, M.: Clouds at Arctic atmospheric observatories. Part I: Occurrence and macrophysical properties, Journal of Applied Meteorology and
875   Climatology, 50, 626-644, 2011.

Song, Z., Liang, S., Wang, D., Zhou, Y., and Jia, A.: Long-term record of top-of-atmosphere albedo over land generated from AVHRR data, Remote Sensing of Environment, 211, 71-88, 2018.

Stackhouse Jr, P., Gupta, S., Cox, S., Chiacchio, M., and Mikovitz, J.: The WCRP/GEWEX surface radiation budget project release 2: An assessment of surface fluxes at 1 degree resolution, 2000.

880   Stoffelen, A.: Toward the true near-surface wind speed: Error modeling and calibration using triple collocation, Journal of geophysical research: oceans, 103, 7755-7766, 1998.

Tang, W.-J., Yang, K., Qin, J., Cheng, C., and He, J.: Solar radiation trend across China in recent decades: a revisit with quality-controlled data, Atmospheric Chemistry and Physics, 11, 393-406, 2011.

Tang, W., Yang, K., Qin, J., and Min, M.: Development of a 50-year daily surface solar radiation dataset over China,
885   Science China Earth Sciences, 56, 1555-1565, 2013.

Tang, W., Yang, K., Qin, J., Li, X., and Niu, X.: A 16-year dataset (2000–2015) of high-resolution (3 h, 10 km) global surface solar radiation, Earth System ence Data, 11, 1905-1915, 2019.

Tang, W., Qin, J., Yang, K., Niu, X., Min, M., and Liang, S.: An efficient algorithm for calculating photosynthetically active radiation with MODIS products, Remote Sensing of Environment, 194, 146-154, 2017.

890   Verma, M., Fisher, J. B., Mallick, K., Ryu, Y., Kobayashi, H., Guillaume, A., Moore, G., Ramakrishnan, L., Hendrix, V., and Wolf, S.: Global surface net-radiation at 5 km from MODIS Terra, Remote Sensing, 8, 739, 2016.

Vermote, E. and Kaufman, Y.: Absolute calibration of AVHRR visible and near-infrared channels using ocean and cloud views, International Journal of Remote Sensing, 16, 2317-2340, 1995.

Wang, B., Yuan, Q., Yang, Q., Zhu, L., Li, T., and Zhang, L.: Estimate hourly PM2. 5 concentrations from Himawari-8 TOA
895   reflectance directly using Geo-intelligent long short-term memory network, Environmental Pollution, 116327, 2020a.

Wang, D. and Liang, S.: Estimating top-of-atmosphere daily reflected shortwave radiation flux over land from MODIS data, IEEE Transactions on Geoscience and Remote Sensing, 55, 4022-4031, 2017.

Wang, D., Liang, S., He, T., and Shi, Q.: Estimation of daily surface shortwave net radiation from the combined MODIS data, IEEE Transactions on Geoscience and Remote Sensing, 53, 5519-5529, 2015a.

900   Wang, D., Liang, S., He, T., and Shi, Q.: Estimating clear-sky all-wave net radiation from combined visible and shortwave infrared (VSWIR) and thermal infrared (TIR) remote sensing data, Remote Sensing of Environment, 167, 31-39, 2015b.

Wang, D., Liang, S., Zhang, Y., Gao, X., and Jia, A.: A New Set of MODIS Land Products (MCD18): Downward Shortwave Radiation and Photosynthetically Active Radiation, Remote Sensing, 12, 168, 2020b.

Wang, K. and Liang, S.: Estimation of daytime net radiation from shortwave radiation measurements and meteorological
905   observations, Journal of Applied Meteorology and Climatology, 48, 634-643, 2009.

Wang, Y., Jiang, B., Liang, S., Wang, D., He, T., Wang, Q., Zhao, X., and Xu, J.: Surface Shortwave net radiation estimation from Landsat TM/ETM+ data using four machine learning algorithms, Remote Sensing, 11, 2847, 2019.

Wei, Y., Zhang, X., Hou, N., Zhang, W., Jia, K., and Yao, Y.: Estimation of surface downward shortwave radiation over China from AVHRR data based on four machine learning methods, Solar Energy, 177, 32-46, 2019.

910   Wielicki, B. A., Barkstrom, B. R., Harrison, E. F., Lee, R. B., III, Smith, G. L., and Cooper, J. E.: Clouds and the Earth's Radiant Energy System (CERES): An Earth Observing System Experiment, Bulletin of the American Meteorological Society, 77, 853-868, 10.1175/1520-0477(1996)077<0853:catere>2.0.co;2, 1996.

Wu, H., Yang, Q., Liu, J., and Wang, G.: A spatiotemporal deep fusion model for merging satellite and gauge precipitation in China, Journal of Hydrology, 584, 124664, 2020.

915 Wu, X., Xiao, Q., Wen, J., and You, D.: Direct comparison and triple collocation: Which is more reliable in the validation of coarse-scale satellite surface albedo products, Journal of Geophysical Research: Atmospheres, 124, 5198-5213, 2019.

Wyser, K., O'Hirok, W., Gautier, C., and Jones, C.: Remote sensing of surface solar irradiance with corrections for 3-D cloud effects, Remote Sensing of Environment, 80, 272-284, https://doi.org/10.1016/S0034-4257(01)00309-1, 2002.

Xi, B., Qiu, S., Dong, X., Smith Jr, W. L., Minnis, P., and Sun-Mack, S.: Evaluation of CERES-MODIS Ed4 Cloud

920 Macrophysical Properties using 11-yr ARM Ground-based Observations over the Arctic, AGUFM, 2018, A11I-2348, 2018.

Xiao, Z., Liang, S., Tian, X., Jia, K., Yao, Y., and Jiang, B.: Reconstruction of long-term temporally continuous NDVI and surface reflectance from AVHRR data, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10, 5551-5568, 2017.

Xu, J., Liang, S., and Jiang, B.: Daily surface all-wave net radiation over global land (1981-2019) from AVHRR data [Data

925 set], https://doi.org/10.5281/zenodo.5509854, 2021.

Xu, J., Jiang, B., Liang, S., Li, X., Wang, Y., Peng, J., Chen, H., Liang, H., and Li, S.: Generating a High-Resolution Time-Series Ocean Surface Net Radiation Product by Downscaling J-OFURO3, IEEE Transactions on Geoscience and Remote Sensing, 2020.

Yang, D.: Quantifying the spatial scale mismatch between satellite-derived solar irradiance and in situ measurements: A case

930 study using CERES synoptic surface shortwave flux and the Oklahoma Mesonet, Journal of Renewable and Sustainable Energy, 12, 056104, 2020.

Yang, F. and Cheng, J.: A framework for estimating cloudy sky surface downward longwave radiation from the derived active and passive cloud property parameters, Remote Sensing of Environment, 248, 111972, 2020.

Yang, L., Zhang, X., Liang, S., Yao, Y., Jia, K., and Jia, A.: Estimating surface downward shortwave radiation over China

935 based on the gradient boosting decision tree method, Remote Sensing, 10, 185, 2018.

Yin, Z., Wu, P., Foody, G. M., Wu, Y., Liu, Z., Du, Y., and Ling, F.: Spatiotemporal Fusion of Land Surface Temperature Based on a Convolutional Neural Network, IEEE Transactions on Geoscience and Remote Sensing, 2020.

Yuan, Q., Xu, H., Li, T., Shen, H., and Zhang, L.: Estimating surface soil moisture from satellite observations using a generalized regression neural network trained on sparse ground-based measurements in the continental US, Journal of

940 Hydrology, 580, 124351, 2020a.

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., and Wang, J.: Deep learning in environmental remote sensing: Achievements and challenges, Remote Sensing of Environment, 241, 111716, 2020b.

Zhang, X., Liang, S., Wild, M., and Jiang, B.: Analysis of surface incident shortwave radiation from four satellite products, Remote Sensing of Environment, 165, 186-202, 2015.

945 Zhang, X., Lu, N., Jiang, H., and Yao, L.: Evaluation of Reanalysis Surface Incident Solar Radiation Data in China, Scientific Reports, 10, 2020.

Zhang, X., Liang, S., Wang, G., Yao, Y., Jiang, B., and Cheng, J.: Evaluation of the reanalysis surface incident shortwave radiation products from NCEP, ECMWF, GSFC, and JMA using satellite and surface observations, Remote Sensing, 8, 225, 2016.

950 Zhang, Y., Rossow, W. B., Lacis, A. A., Oinas, V., and Mishchenko, M. I.: Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global data sets: Refinements of the radiative transfer model and the input data, Journal of Geophysical Research: Atmospheres, 109, 2004.