A dense station-based long-term and high-accuracy dataset of daily surface solar radiation in China

Wenjun Tang[1,*], Junmei He[1,2], Jingwen Qi[1], Kun Yang[3,1,*]

1. National Tibetan Plateau Data Center (TPDC), State Key Laboratory of Tibetan Plateau Earth System, Environment and Resources (TPESER), Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China.

2. Faculty of Geography, Yunnan Normal University, Kunming 650500, China.

3. Department of Earth System Science, Ministry of Education Key Laboratory for Earth System Modeling, Institute for Global Change Studies, Tsinghua University, Beijing 100084, China.

Corresponding authors: Prof. Wenjun Tang (tangwj@itpcas.ac.cn); Prof. Kun Yang (yangk@tsinghua.edu.cn)

Abstract: The lack of long-term and high-quality solar radiation data has been an obstacle for scientific and industrial fields. In this study, a dense station-based long-term and high-accuracy dataset of daily surface solar radiation was developed using two surface radiation models. One is the model developed by Yang et al. (2006) for global radiation estimation, and the other is the model developed by Tang et al. (2018) for direct radiation estimation. The main inputs for the development of the dataset are surface pressure, air temperature, relative humidity, horizontal visibility, and sunshine duration, which are the routine meteorological variables observed at the 2743 China Meteorological Administration (CMA) weather stations. Validation against in-situ observations and comparisons with two satellite-based radiation products show that our station-based radiation dataset clearly outperforms the satellite-based radiation products at both daily and monthly scales. In addition, our dataset is available for more than 60 years and includes three radiation components of global, direct, and diffuse radiation, which is not possible with satellite products. This station-based radiation dataset will contribute to the climate change research and solar energy engineering applications in the future. The station-based dataset is now available at https://doi.org/10.11888/Atmos.tpdc.300461 and https://data.tpdc.ac.cn/en/disallow/55fc9768-ea6a-4d16-9207-28f6aed4900b (Tang, 2023).

Keywords: Global radiation; Direct radiation; Diffuse radiation; High-accuracy; Long-term; Dataset.

1. Introduction

Solar radiation provides energy to everything on Earth, drives the water, energy and carbon cycles of the Earth’s climate system, and largely determines the climatic conditions of human habitats (Wild et al., 2009). Therefore, solar radiation information at the Earth's surface is crucial in research of agriculture, hydrology, ecology, climate change, and simulations of land surface processes (Wang et al., 2012). Solar radiation reaching the horizontal surface is called as total solar radiation or global radiation (R_g), and is composed of direct radiation (R_{dir}) and diffuse radiation (R_{dif}). Global radiation is an important component of the surface energy budgets (Wild et al., 2015), and accurate global radiation data will contribute to the simulation of land surface-related processes, such as ecological, hydrological, agricultural, and glacial simulations (Tang et al., 2019a). In addition, both direct and diffuse radiation provide energy for plant photosynthesis and transpiration, and are essential for hydrological and agricultural studies (Lee et al., 2017). Due to its multidirectional nature, diffuse radiation penetrates in the vegetation canopy more than direct radiation, and more diffuse radiation can increase light energy use efficiency of the canopy (Mercado et al., 2009, Yang et al., 2019). For example, under the same global radiation condition, increasing the proportion of diffuse radiation can increase the light energy use efficiency of different vegetation types by 6-180% (Gu et al., 2002; Alton et al., 2007).

In addition to basic scientific research, the distribution and intensity of surface solar radiation is urgently needed for solar energy applications. All three components of solar radiation, i.e., global radiation, direct radiation, and diffuse radiation, are prerequisites for the siting, design, evaluation, and optimization of different solar energy systems (Karakoti et al., 2011; Mellit et al., 2010). For example, photovoltaic (PV) power systems rely on global radiation to generate electricity, and global radiation on arbitrarily oriented PV panels can be calculated by direct and diffuse radiation, while concentrated solar power systems use only direct radiation to generate electricity (Boland et al., 2013; Tang et al., 2018). In addition, recently developed bifacial PV panels also use the backside of the PV panel to generate electricity, and the source of solar radiation on the backside is diffuse radiation (Rodríguez-Gallegos et al., 2018; Pelaez et al., 2019).

In-situ measurements are considered the most effective and direct means of obtaining surface solar radiation data. However, the number of radiation observation stations is very low due to high maintenance and calibration costs. For example, radiation fluxes measured at about 2500 stations worldwide are stored in the Global Energy Balance Archive (GEBA), a database that stores the different components of the surface energy budget from different data sources, such as national meteorological services, various experimental observation networks, and project reports (Wild et al., 2017). Among the GEBA, there are only about 100 radiation stations (Jiang et al. 2020a), which are provided by the China Meteorological Administration (CMA). Another issue is that the radiation observation stations are very unevenly distributed, with most of radiation stations being located in flat and densely populated areas, and with a very small number of stations being located in the complex terrain areas and sparsely populated areas. In addition, almost all radiation stations include global radiation observations, but most of them do not include direct radiation and diffuse radiation observations. In addition, radiation measurements are considered to be more prone to problems and unreliable data than those of other meteorological variables. For example, erroneous or spurious data were frequently found in the CMA radiation observations (Shi et al., 2008; Tang et al., 2011). Therefore, the lack of long-term and dense solar radiation observations has become a tough challenge.

Satellites can provide continuous spatiotemporal observations, and retrieval based on satellite remote sensing is considered one of the most effective and commonly used means to fill the gap in ground-based radiation measurements (Lu et al., 2011; Zhang et al., 2014; Huang et al., 2019; Wang et al., 2022; Letu et al., 2021). In the past, satellite-based retrieval algorithms and corresponding radiation products have emerged, and the retrieval accuracy of the corresponding products has improved (Huang et al., 2019). In particular, retrievals applied to the new generation of geostationary satellites have improved in temporal resolution, spatial resolution, and accuracy to varying degrees (Tang et al., 2019b; Letu et al., 2020). Especially, Li et al. (2023) produced a high-spatiotemporal-resolution radiation product based on the new generation of geostationary satellites from the United States and Japan, with accuracy higher than other existing satellite products. In general, the accuracy of the satellite-based radiation product is higher than that of the reanalysis product (Jiang et al., 2020b). In addition, Hao et al. (2020) developed a global radiation product based on the unique Deep Space Climate Observatory (DSCOVR) satellite, whose orbit is at the Lagrange point. Despite the obvious advantages, there are still several problems with satellite-based radiation products. First, the time series of satellite radiation products are generally not long enough to meet the demand for long time series data, such as for climate change studies. Second, the updating of satellite sensors and the fusion of multi-source satellites would lead to inconsistent data quality and introduce large uncertainties into the analysis of long-term variations (Feng and Wang, 2021a; Shao et al., 2022). Third, most satellite radiation products only provide global radiation, but do not include direct and diffuse radiation, because the uncertainty in algorithms for estimating direct and diffuse radiation is much larger than the uncertainty in global radiation.

Alternatively, estimation based on meteorological variables observed at routine weather stations is another effective solution that can overcome the scarcity of radiation data, since the number of routine weather stations is much denser than that of radiation stations (Feng and Wang, 2021b). For example, the number of radiation stations maintained by the CMA is only about 100, but the number of routine weather stations with long-term observations is much denser, exceeding 2400 stations. Empirical models for estimating global radiation using surface meteorological variables were usually found in the hydrological and agricultural fields (Wang et al., 2016), and these models can be broadly classified into three types: air temperature-based models, sunshine duration-based models, and cloud cover-based models (Liu et al., 2009; Ehnberg & Bollen, 2005). In general, a well-calibrated sunshine duration-based model was considered to perform better than the other two types of models (Pohlert, 2004). Although the above empirical models, especially the Ångström-Prescott relationships (Ångström, 1924; Prescott, 1940), have been widely used and have achieved great success, these models require ground radiation observations for calibration and the calibrated parameters are usually site-dependent, which poses significant risks and challenges when applied in areas with sparse or no radiation observations.

Yang et al. (2001) developed a hybrid model to estimate daily global radiation by combining pure physical processes under clear skies and a parameterization formula for cloud transmission under cloudy skies. The hybrid model has been shown to work well without local calibration (Yang et al., 2006; Yang et al., 2010). Tang et al. (2013) used this hybrid model to construct a dataset of daily global radiation at 716 CMA weather stations during 1961-2010, and also found that its accuracy was generally higher than that of locally calibrated empirical models and satellite-based retrievals. However, few studies have focused on the development of empirical models for direct and diffuse radiation estimates. Fortunately, Tang et al. (2018) also developed a similar hybrid model to estimate daily direct radiation by adopting the strategy of Yang et al. (2001) to estimate global radiation. Therefore, these two hybrid models provide us with an opportunity to construct daily surface solar radiation at routine weather stations, once the observations of meteorological variables are available. The constructed dataset will contribute to simulations of land surface processes, climate change analysis, and solar energy applications.

In this study, based on our previous study by Tang et al. (2013) for global radiation estimation, we expanded the number of stations, radiation elements and time length, and finally developed a dense station-based long-term and high-accuracy daily surface solar radiation dataset in China, which includes three elements: global radiation, direct radiation, and diffuse radiation. This long-term dataset will contribute to the analysis of long-term variations in surface process simulations and solar energy applications, such as the assessment of solar energy potential, the determination of the optimal angle for solar PV panels and their long-term variation analysis, as well as the assessment of historical extreme events on solar energy systems. The rest of the paper is organized as follows. The methods used to estimate daily global, direct, and diffuse radiation at weather stations are presented in Section 2, and Section 3 describes the input data used to drive the station-based models, the in-situ data used to validate the accuracy of the developed dataset, and two satellite-based radiation products used for comparison with our products. The performance of our dataset and two satellite-based radiation products against the in-situ data is evaluated in Section 4, and the information on data availability is described in Section 5. Finally, Section 6 presents the summary of this study.

2. Methods

In this study, daily global radiation and daily direct radiation were estimated using the hybrid model of Yang et al. (2006) and the method developed by Tang et al. (2018), respectively. Finally, daily diffuse radiation was calculated by subtracting direct radiation from global radiation. The methods for estimating daily global, direct, and diffuse radiation can be simply described by the following five mathematical formulas:

$$R_{\mathrm{g}} = (R_{\mathrm{b,clr}} + R_{\mathrm{d,clr}})\tau_{\mathrm{c,g}}, \tag{1}$$

$$R_{\mathrm{b}} = R_{\mathrm{b,clr}}\tau_{\mathrm{c,b}}, \tag{2}$$

$$R_{\mathrm{d}} = R_{\mathrm{g}} - R_{\mathrm{b}}, \tag{3}$$

$$\tau_{c,g} = 0.2505 + 1.1468\left(\frac{n}{N}\right) - 0.3974\left(\frac{n}{N}\right)^2, \tag{4}$$

$$\tau_{c,b} = 0.4868\left(\frac{n}{N}\right) + 0.5132\left(\frac{n}{N}\right)^2, \tag{5}$$

where R_g, R_b, and R_d [W m^{-2}] are the daily all-sky global, direct, and diffuse radiation at the horizontal surface, respectively. $R_\mathrm{b,clr}$ and $R_\mathrm{b,clr}$ [W m^{-2}] are the daily clear-sky global and direct radiation at the horizontal surface, respectively. $\tau_\mathrm{c,g}$ and $\tau_\mathrm{c,b}$ are the cloud transmittance for the daily global and direct radiation, respectively. n and N are the actual sunshine duration and the maximum possible sunshine duration, respectively. The daily clear-sky global and direct radiation ($R_\mathrm{b,clr}$ and $R_\mathrm{b,clr}$) at the horizontal surface are calculated by the following equations:

$$R_\mathrm{b,clr} = \frac{1}{24\mathrm{hours}}\int_{t1}^{t2} I_0\,\bar{\tau}_\mathrm{b}\mathrm{d}t, \tag{6}$$

$$R_\mathrm{d,clr} = \frac{1}{24\mathrm{hours}}\int_{t1}^{t2} I_0\,\bar{\tau}_\mathrm{d}\mathrm{d}t, \tag{7}$$

$$\bar{\tau}_\mathrm{b} \approx \max\left(0, \bar{\tau}_\mathrm{o}\bar{\tau}_\mathrm{w}\bar{\tau}_\mathrm{g}\bar{\tau}_\mathrm{r}\bar{\tau}_\mathrm{a}\right), \tag{8}$$

$$\bar{\tau}_\mathrm{d} \approx \max\left(0, 0.5\bar{\tau}_\mathrm{o}\bar{\tau}_\mathrm{w}\bar{\tau}_\mathrm{g}(1 - \bar{\tau}_\mathrm{r}\bar{\tau}_\mathrm{a})\right), \tag{9}$$

$$\bar{\tau}_\mathrm{o} = \exp\left[-0.0365(ml)^{0.7136}\right], \tag{10}$$

$$\bar{\tau}_\mathrm{g} = \exp\left[-0.0117(m_\mathrm{c})^{0.3139}\right], \tag{11}$$

$$\bar{\tau}_\mathrm{w} = \min\left[1.0, 0.909\ln(mw)\right], \tag{12}$$

$$\bar{\tau}_\mathrm{r} = \exp\left[-0.008735 m_\mathrm{c}(0.547 + 0.014 m_\mathrm{c} - 0.00038 m_\mathrm{c}^2 + 4.6\times10^{-6} m_\mathrm{c}^3)^{-4.08}\right], \tag{13}$$

$$\bar{\tau}_\mathrm{a} = \exp\left\{-m\beta\left[0.6777 + 0.1464(m\beta) - 0.00626(m\beta)^2\right]^{-1.3}\right\}, \tag{14}$$

$$m = 1/\left[\sin(h) + 0.15(57.296h + 3.885)^{-1.253}\right], \tag{15}$$

$$m_\mathrm{c} = m p_\mathrm{s}/p_0, \tag{16}$$

where I_0 [W m^{-2}] is the horizontal solar radiation at the top of the atmosphere (TOA), and $t1$ [hour] and $t2$ [hour] are the times of sunrise and sunset, respectively. $\bar{\tau}_{\mathrm{b}}$ and $\bar{\tau}_{\mathrm{d}}$ are the transmittances for the daily direct and diffuse radiation under clear skies, respectively. $\bar{\tau}_{\mathrm{o}}$, $\bar{\tau}_{\mathrm{g}}$, $\bar{\tau}_{\mathrm{w}}$, $\bar{\tau}_{\mathrm{r}}$, and $\bar{\tau}_{\mathrm{a}}$ are transmittances due to ozone absorption, permanent gas absorption, water vapor absorption, Rayleigh scattering, and aerosol absorption and scattering in the atmospheric layer, respectively. m is the air mass, and l (cm) is ozone layer thickness. m_{c} is the pressure-corrected air mass, and w (cm), is the precipitable water. β and h [radian] are the Ångström turbidity coefficient and the solar elevation angle, respectively. p_0 [Pa] is the standard atmospheric pressure, and p_{s} [Pa] is the surface pressure.

The overall flowchart for the estimation of the station-based radiation products is shown in Figure 1, which consists mainly of input, model and output sections. The inputs are surface pressure, air temperature, relative humidity, ozone amount, aerosol data and sunshine duration. Air temperature and relative humidity are used to estimate precipitable water. The aerosol Ångström turbidity is mainly converted from horizontal visibility observations measured at weather stations using the method of Tang et al. (2017a). The ozone amount is obtained from the zonal means of the Total Ozone Mapping Spectrometer (TOMS). The outputs are daily global, direct and diffuse radiation, respectively. For more detailed information on the above methods, we can refer to the articles by Yang et al. (2006) and Tang et al. (2018).

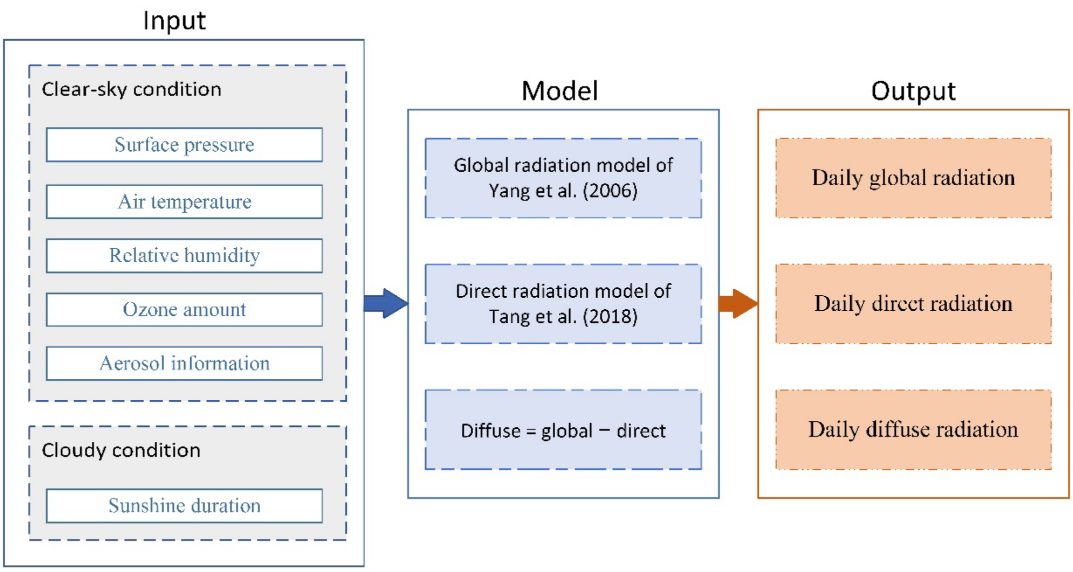

Figure 1 Flowchart of data production, including input, model, and output sections.

3 Data

3.1 Input data

The inputs to the above methods for estimating daily global, direct, and diffuse radiation are mainly surface pressure, air temperature, relative humidity, horizontal visibility, sunshine duration, and ozone amount. Except for ozone, all input data are observed at the 2473 CMA routine meteorological stations, where the estimation methods can work well. Here, the ozone amount was used from the climatological data obtained from the TOMS zonal means. Figure 2 shows the spatial distribution of these routine meteorological stations, indicated by the small blue circles, with characteristics of a dense distribution in the eastern and southern regions of China and a sparse distribution in the western and northern regions of China.

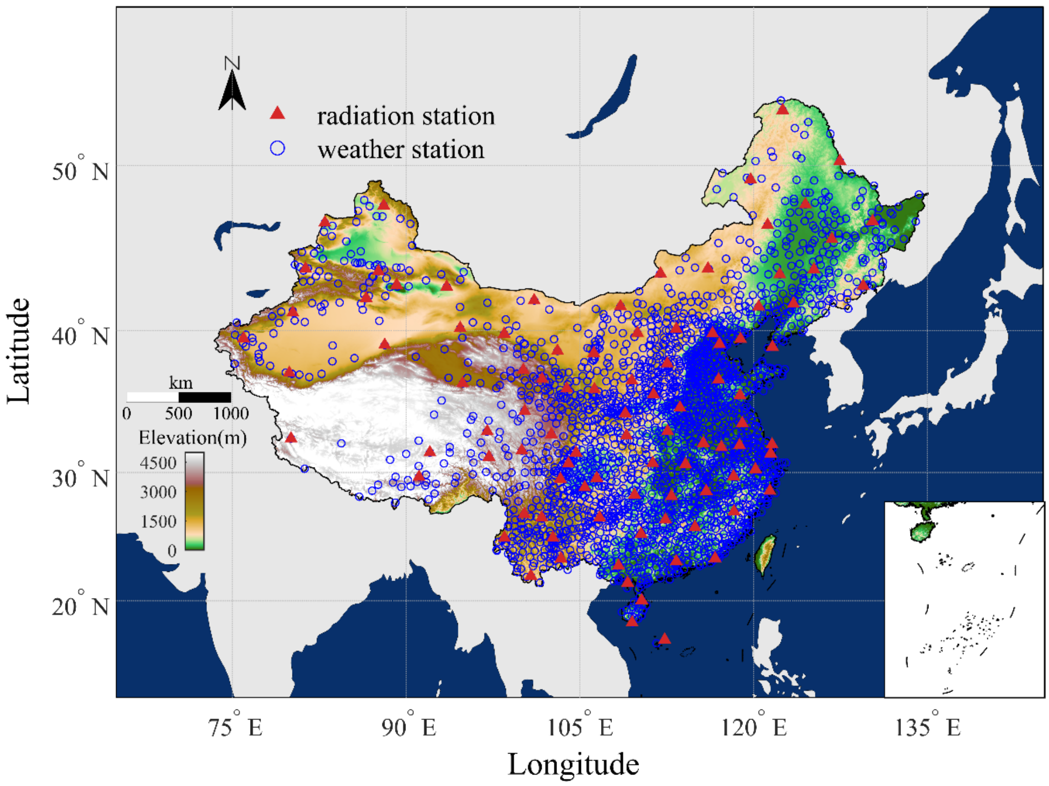

Figure 2 Spatial distribution of China Meteorological Administration weather station and radiation station. Red triangles represent radiation stations, and blue circles denote weather stations without radiation observation.

Finally, in this study, a dataset of more than 60 years of daily global, direct and diffuse radiation was constructed using the methods described above at the 2473 CMA routine meteorological stations from the 1950s to 2021, with most stations covering the period of 1961-2021.

3.2 In-situ radiation data

Due to the extensive renewal and replacement of radiation instruments in 1993, and because more erroneous observations were found before 1993, radiation observations during the period 1993-2010 were used to evaluate the performance of our developed station-based radiation dataset and the other two satellite-based radiation products. Among the 2473 CMA routine meteorological stations, there are about 96 radiation stations (denoted by the upper red triangle in Figure 2) where radiation observations have been carried out in addition to routine meteorological observations since 1993. Among the 96 CMA radiation stations, there are only 19 stations where direct and diffuse radiation observations are carried out in addition to global radiation observations.

One issue to note is that we did not use the direct radiation observations as validation basis, but the value obtained by subtracting the diffuse radiation observations from the global radiation observations, because the quality of the direct radiation observations is significantly lower than those of the global and diffuse radiation observations (Tang et al., 2018). Another issue that we need to be aware of is that obviously erroneous or false values have usually been found in the CMA radiation data (Shi et al., 2008), although a preliminary quality check of the raw radiation observations was carried out before release. Therefore, we further applied the quality control procedures developed by Tang et al. (2010) to the raw radiation observations and filtered out the corresponding spurious and erroneous observations. More detailed information on the quality control scheme can be found in the article by Tang et al. (2010).

3.3 Satellite-based radiation products

In this study, two satellite-based radiation products (Jiang et al., 2020a; Tang et al., 2019a) were used for comparison with our station-based radiation dataset estimated at 2743 CMA routine meteorological stations.

One is the product of Jiang et al. (2020a), which provides hourly global radiation and diffuse radiation in China with a spatial resolution of ~5 km and a time span from 2007 to 2018. The product was generated using a deep learning algorithm developed by Jiang et al. (2019) to retrieve global radiation, and a transfer learning approach to retrieve diffuse radiation from Multi-functional Transport Satellites (MTSAT) imagery. The algorithm successfully tackled spatial adjacency effects induced by photon transport through convolutional neural networks, resulting in excellent performance in instantaneous radiation retrieval, especially for diffuse radiation (Jiang et al., 2020c).

The other is the high-resolution global product of global radiation developed by Tang et al. (2019a) based on the latest cloud products of the International Satellite Cloud Climatology Project H-series pixel-level global (ISCCP-HXG), routine meteorological variables of the ERA5 reanalysis data, aerosol optical thickness of the MERRA-2 reanalysis data, and albedo of the MODIS and CM-SAF products with the physical algorithm of Tang et al. (2017b). This global product has a spatial resolution of 10 km, a temporal resolution of 3 hours, and spans the period of 1983.7-2018.12. Global comparative validation with observations shows that the accuracy of this global radiation product is generally better than several global satellite radiation products, such as the Earth's Radiant Energy System (CERES; Kato et al., 2013), the Global Energy and Water Cycle Experiment surface radiation budget (GEWEX-SRB; Pinker and Laszlo, 1992), and the ISCCP flux dataset (ISCCP-FD; Zhang et al., 2004).

In this study, these two satellite-based radiation products were first averaged to daily and monthly means, and then validated against observations collected at the CMA radiation stations, and finally compared with our station-based radiation dataset in China.

4 Results and Discussion

The station-based radiation dataset was first validated against the CMA radiation observations at daily and monthly scales, respectively, then compared with two other satellite-based radiation products, and finally its spatial distribution characteristics were further analyzed. These are described in detail in the following four subsections. In this study, we used the statistical metrics of mean bias error (MBE), relative MBE (rMBE), root mean square error (RMSE), relative RMSE (rRMSE) and correlation coefficient (R) to measure the accuracies of our station-based dataset and the other two satellite-based radiation products.

4.1 Validation at daily scale

Figure 3 shows the validation results for the daily global radiation estimates against observations from 96 CMA radiation stations over the period of 1993-2010. Overall, our station-based estimates over China perform well, with an MBE of 2.5 W m^{-2}, a RMSE of 23.2 W m^{-2}, and an R of 0.96 (Figure 3 a). This indicates that the accuracy of our station-based estimates is significantly higher than that of almost all of the satellite products (such as, ISCCP-FD, GEWEX-SRB, CERES, GLASS and ISCCP-HXG) and reanalysis data (such as, ERA5 and MERRA-2), as well as other regional satellite-based radiation products from Tang et al. (2016), Jiang et al. (2020a), and Letu et al. (2021). Therefore, we would expect our station-based estimates to be more accurate than the five global radiation products mentioned by Li et al. (2021), as CERES generally performs best among them. Of course, this speculation needs to be further verified with in-situ measurements collected in China in the future. In terms of MBE, there is a slight positive deviation, but the relative deviation is close to 2%, which is within the tolerance range of the PV potential assessment.

In addition, we also calculated the statistical metrics for each radiation station and their boxplots and spatial distributions are shown in Figure 3 (b)-(d). It can be seen that most of the stations had MBE between −4 and 8 W m^{-2}, RMSE less than 25 W m^{-2}, and R greater than 0.95. The stations with relatively large errors were mainly found in the southern and eastern regions of China. This is largely due to the fact that these regions have more cloud cover and overcast skies, making it difficult to use sunshine duration to accurately parameterize cloud transmission. In addition, the uncertainty in the aerosol data would also partly contribute to the large errors.

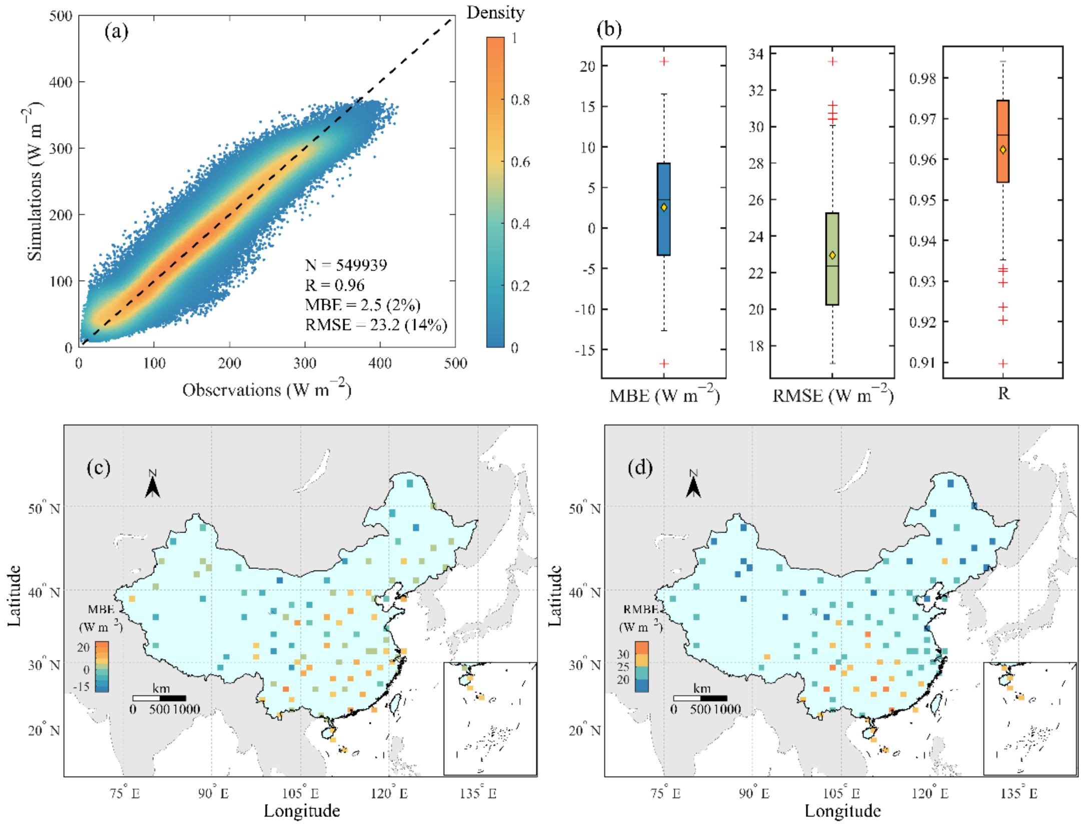

Figure 3 Validation results for our station-based daily global radiation dataset against observations collected at 96 CMA radiation stations during 1993-2010. (a) Comparison of daily global radiation between our dataset and observations, (b) Boxplots of three statistical error metrics (MBE, RMSE and R), (c) spatial distribution of MBE for individual radiation station, and (d) spatial distribution of RMSE for individual radiation station.

Similar evaluations for the daily direct and diffuse radiation estimates are presented in Figure 4 and Figure 5, respectively. Averaged over 19 CMA radiation stations, our estimates for daily direct radiation produces an MBE of 7.4 W m^{-2}, a RMSE of 27.2 W m^{-2} and an R of 0.92, respectively; while our estimates for daily diffuse radiation produces an MBE of -3.3 W m^{-2}, a RMSE of 19.2 W m^{-2} and an R of

0.83, respectively. Both accuracies for direct and diffuse radiation are lower than for global radiation, as can be seen from their rRMSE and rMBE. The absolute rMBE for global radiation is about 2%, while those for direct and diffuse radiation are about 9% and 4%, respectively. The rRMSE for global radiation is about 14%, while those for direct and diffuse radiation are about 33% and 22%, respectively. We found that there is a slight overestimation for direct radiation, with an MBE of about 7.4 W m^{-2}. This may be due to the presence of high clouds, in which case the cloud transmission for direct radiation cannot be well parameterized by the sunshine duration (Tang et al. 2018). For direct radiation, the RMSE at most stations is less than 29 W m^{-2} and the R at most stations is greater than 0.91; for diffuse radiation, the RMSE at most stations is less than 21 W m^{-2} and the R at most stations is greater than 0.81. The largest RMSE for both direct and diffuse radiation is found at the Lhasa station, where popcorn clouds were common, posing a major challenge to the simulation of cloud transmittance with sunshine duration.

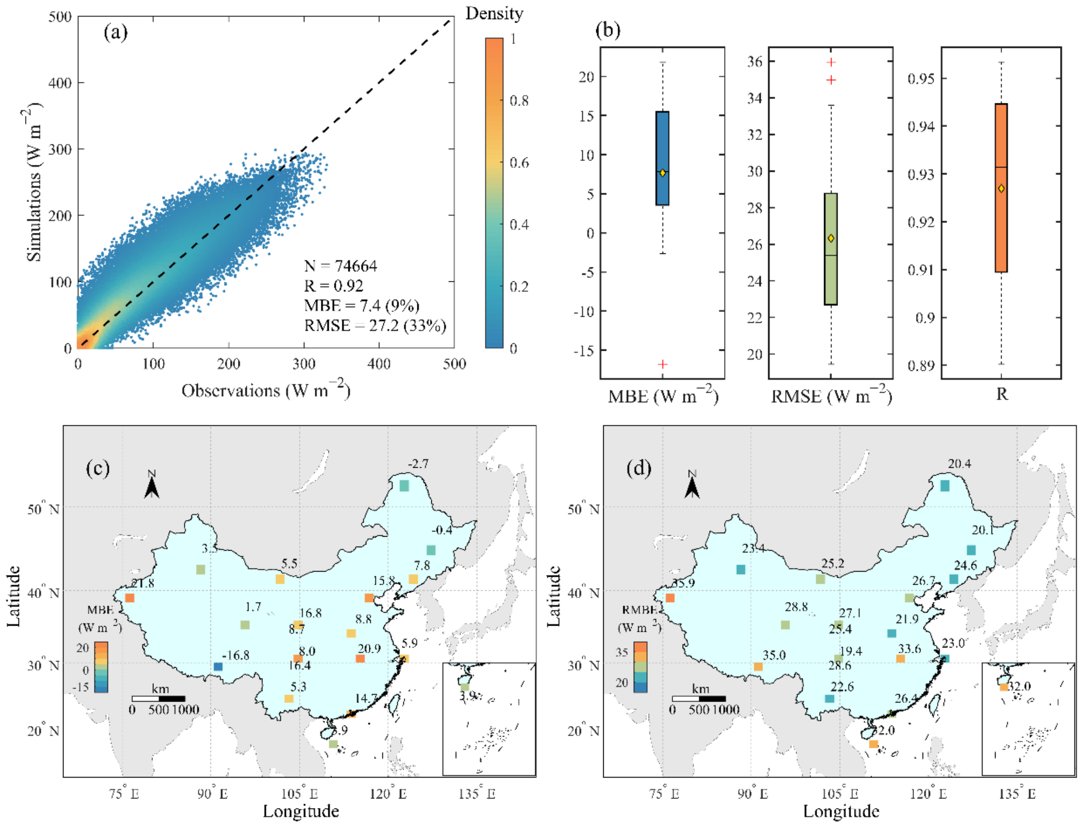

Figure 4 Same as **Figure 3**, but for daily direct radiation.

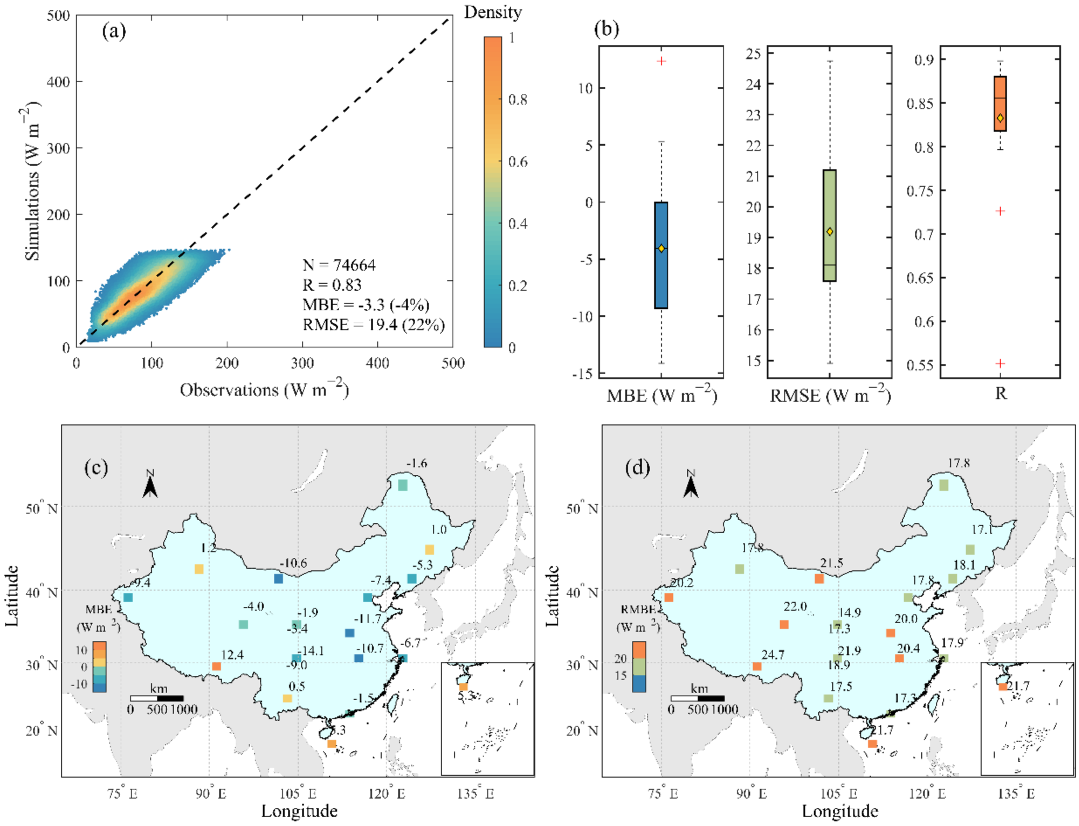

Figure 5 Same as **Figure 3**, but for daily diffuse radiation.

4.2 Validation at monthly scale

Based on the daily estimates, we also calculated their monthly averages and validated them against observations. Figure 6 shows the validations for the estimates of monthly global radiation against observations at the CMA radiation stations. Averaged over all stations, the monthly global radiation estimates give an MBE of 2.4 W m^{-2}, a RMSE of 13.1 W m^{-2} and an R of 0.98. The rRMSE is about 8%. From the boxplots of the error indicators, R was greater than 0.98 and RMSE was less than 15 W m^{-2} at most stations. The spatial distribution features of MBE and RMSE are similar to those for daily global radiation, with relatively larger errors in the southern and eastern regions of China, and the largest MBE and RMSE are both found at Panzhihua station (with longitude of 101.717° and latitude of 26.583°).

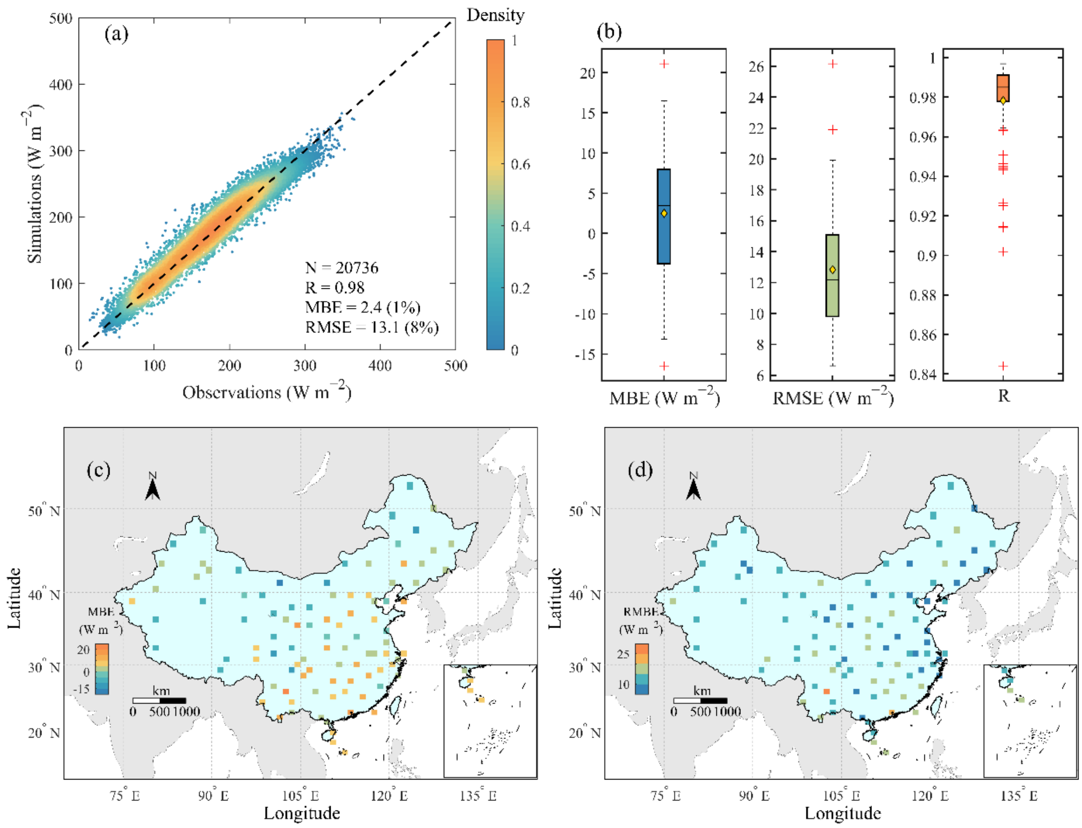

Figure 6 Same as **Figure 3**, but for monthly global radiation.

Figure 7 and Figure 8 also show the validation results for monthly direct and diffuse radiation against 19 CMA radiation stations. The rRMSE values for monthly direct and diffuse radiation are 20% and 12%, respectively. The MBE, RMSE and R for monthly direct radiation are 6.6 W m^{-2}, 16.3 W m^{-2} and 0.94, respectively, while the MBE, RMSE and R for monthly diffuse radiation are -3.3 W m^{-2}, 10.5 W m^{-2} and 0.94, respectively. Similar to the case of the daily scale, a slight overestimation is also found for the monthly direct radiation. The RMSE for monthly direct and diffuse radiation is less than 18 W m^{-2} and 13 W m^{-2}, respectively, at most stations, while R for monthly direct and diffuse radiation is greater than 0.91 and 0.95, respectively, at most stations. The spatial distribution characteristics of MBE and RMSE are similar to those of daily conditions for both direct and diffuse radiation.

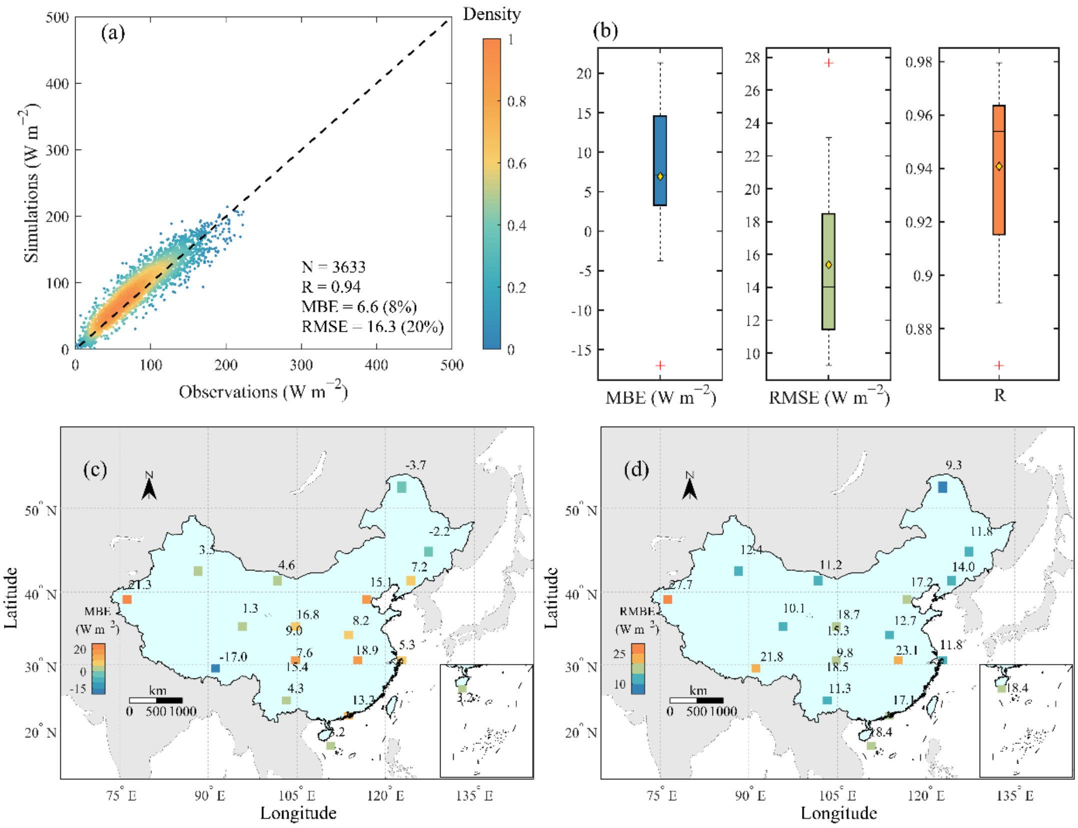

Figure 7 Same as **Figure 3**, but for monthly direct radiation.

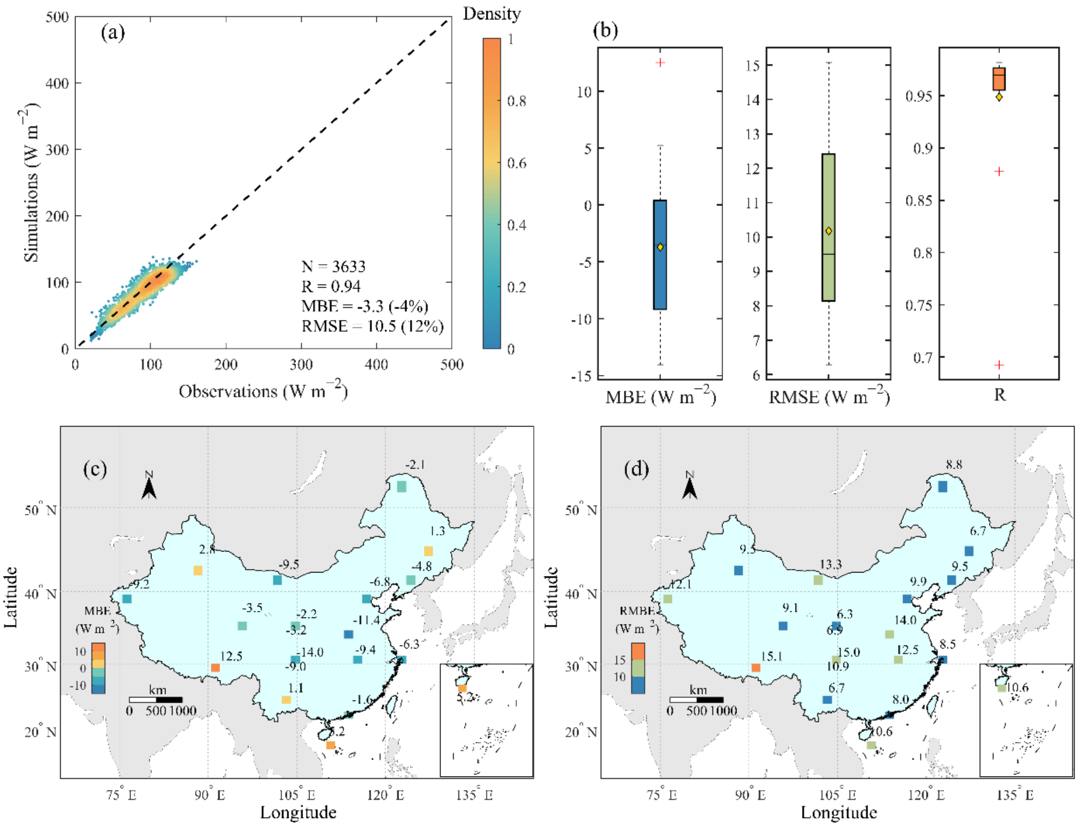

Figure 8 Same as **Figure 3**, but for monthly diffuse radiation.

4.3 Comparison with satellite-based products

To demonstrate the superiority of the station-based radiation dataset developed in this study, we compared the station-based dataset with two widely used satellite-based radiation products. One is the product developed with deep learning and trained with surface radiation observations (Jiang et al., 2020a), and the other is the long-term global product of global radiation (Tang et al., 2019a), which was mainly developed based on the latest ISCCP-HXG cloud products with the improved physical algorithm (Tang et al., 2017b). The former can provide hourly global, direct and diffuse radiation, while the latter can only provide the 3-hourly global radiation. Here, we use a 3 × 3 spatial window to smooth the raw radiation product from Tang et al. (2019a), as its accuracy is clearly improved when upscaled to 30 km. As we only have access to the CMA radiation observations up to 2010, the time period chosen for comparison is 2000-2010. Table 1 and Table 2 present the comparison of the evaluation results among the three radiation products against the observations measured at all CMA radiation stations, for daily and monthly conditions, respectively.

For daily data comparisons, our station-based estimate for daily global radiation obviously outperforms the other two satellite-based radiation products, with MBE of 2.7 W m^{-2}, RMSE of 23.2 W m^{-2} and R of 0.96, followed by the product of Tang et al. (2019a) (with MBE of 6.7 W m^{-2}, RMSE of 26.8 W m^{-2} and R of 0.95) and the product of Jiang et al. (2020a) (with MBE of 4.4 W m^{-2}, RMSE of 33.6 W m^{-2} and R of 0.92). Our station-based estimates for daily direct radiation are also apparently more accurate than the satellite-based product of Jiang et al. (2020a), with the former producing a lower RMSE and a higher R. The RMSE of our estimates is about 9 W m^{-2} lower than that of the product of Jiang et al. (2020a).

For the daily diffuse radiation, our estimate is comparable to the diffuse radiation product of Jiang et al. (2020a). It should be noted that the CMA radiation observations were used to train the deep learning model of Jiang et al. (2020a), while no observations were used in our estimates. Overall, these results indicate that the station-based estimates of global, direct, and diffuse radiation generally outperform those from satellite retrievals.

Table 1. Accuracy comparisons of our estimates in this study with other two satellite-based products at daily scale. The units of MBE and RMSE are both W m^{-2}. Observations of global radiation measured at 96 CMA radiation stations, and observations of direct radiation and diffuse radiation measured at 19 CMA radiation stations during 2000-2010 are used.

	This study			Jiang et al. (2020a)			Tang et al. (2019a)
	global	direct	diffuse	global	direct	diffuse	global (30 km)
MBE	2.7	8.6	-3.8	4.4	8.6	0.1	6.7
RMSE	23.2	27.6	19.6	33.6	36.6	20.1	26.8
R	0.96	0.92	0.83	0.92	0.85	0.84	0.95
N	328977	43630	43630	328977	43630	43630	328977

For the monthly comparisons, we also found that our estimate of monthly global radiation is clearly more accurate than the two satellite products. The MBE and RMSE of our estimate are 2.6 W m^{-2} and 13.4 W m^{-2}, respectively, which are lower than those of the two satellite products, with MBE and RMSE values of 4.6 W m^{-2} and 18.5 W m^{-2} for the Jiang et al. (2020a) product and 6.7 W m^{-2} and 16.3 W m^{-2} for the Tang et al. (2019a) product. The R of our estimate is 0.98, which is higher than those of the two satellite products. Our direct radiation estimate also outperforms the satellite product of Jiang et al. (2020a), with the former having lower RMSE and MBE and higher R. Similar to the daily comparison, the two monthly diffuse products are comparable to each other.

Table 2. Accuracy comparisons of our estimates in this study with other two satellite-based products at monthly scale. The units of MBE and RMSE are both W m^{-2}. Observations of global radiation measured at 96 CMA radiation stations, and observations of direct radiation and diffuse radiation measured at 19 CMA radiation stations during 2000-2010 are used.

	This study			Jiang et al. (2020a)			Tang et al. (2019a)
	global	direct	diffuse	global	direct	diffuse	global (30 km)
MBE	2.6	7.9	-3.7	4.6	9.0	-0.5	6.7
RMSE	13.4	16.5	10.9	18.5	21.3	12.1	16.3
R	0.98	0.94	0.94	0.96	0.89	0.93	0.97
N	12059	2088	2088	12059	2088	2088	12059

4.4 Spatial distribution features of solar radiation in China

Based on the developed station-based solar radiation dataset, Figure 9 present the spatial distributions of the multi-year average global, direct, and diffuse radiation in China during 1961 -2021. The spatial distribution of direct radiation is similar to that of global radiation, with the highest value over the Tibetan Plateau and the lowest value over the Sichuan Basin. However, the spatial distribution of diffuse radiation is very different from that of global radiation, with the highest value being found at the southernmost tip of China, and the lowest value being found in northeastern China. In general, clear-sky days correspond to more direct radiation and cloudy days correspond to more diffuse radiation, mainly because the scattering effect of aerosols is much smaller than the scattering effect of clouds. Therefore, high direct radiation is usually found in areas with frequent sunny days, such as Northwest China, North China, and Inner Mongolia, while low direct radiation is usually found in areas with frequent cloud cover, such as East China and South China. The opposite is true for diffuse radiation.

It should be noted that our station-based products are spatially discontinuous, especially in northwestern China, which may introduce significant uncertainty when applied to the assessment of solar power system potential. However, the uncertainty caused by spatial discontinuity in flat areas would be relatively small, as the spatial representation of a station on flat ground is generally larger than 25 km (Hakuba et al., 2013). Fortunately, most solar power systems are built on land with slopes of less than 3%. In contrast, applications over complex terrain will introduce large uncertainties. Combining station-based data with satellite products will be a good solution in the future to improve the accuracy of solar energy potential assessment.

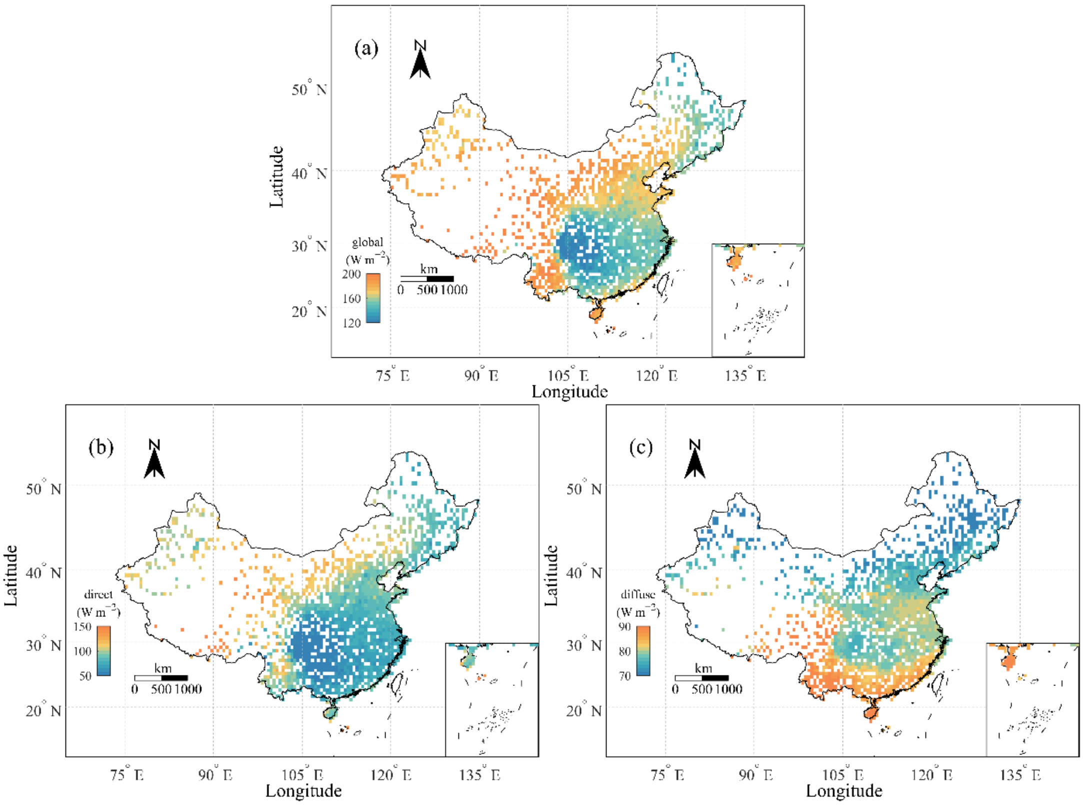

Figure 9 Spatial distribution of the multi-year mean (a) global radiation, (b) direct radiation, and (c) diffuse radiation from the station-based radiation dataset in China during 1961-2021.

The spatial distribution of the multi-year average ratio of direct radiation to global radiation in China during the period of 1961-2021 is also shown in Figure 10. Direct radiation is mainly greater than diffuse radiation (ratio < 50%) in northern China, northeastern China, northwestern China, and the Tibetan Plateau, while diffuse radiation dominates in eastern and southern China. This information can guide the planning of solar PV power and concentrating solar power (CSP) in China. For example, regions such as Xinjiang, Inner Mongolia, Gansu, and the Tibetan Plateau with a high proportion of direct radiation are suitable for the construction of CSP projects. Regions such as Hainan, Yunnan, Guangxi, and Guangdong with relatively high global and diffuse radiation are suitable for the use of bifacial PV panels, as both sides of the panels can be used to generate electricity and the main source on the back is from diffuse radiation. Conversely, in areas with low diffuse radiation, it may not be necessary to use bifacial PV panels.

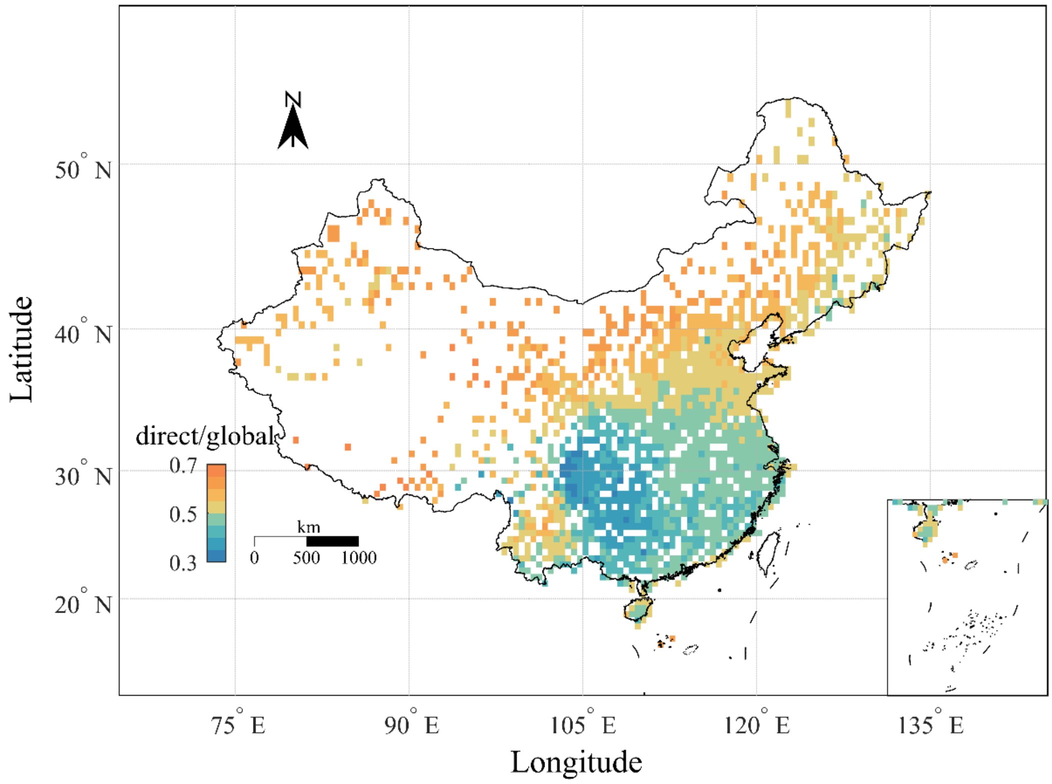

Figure 10 Spatial distribution of the multi-year mean ratio of direct radiation to global radiation from the station-based radiation dataset in China during 1961-2021.

5 Data availability

The dense station-based long-term dataset of daily global, direct and diffuse radiation product with high-accuracy is stored at the National Tibetan Plateau Data Center (https://data.tpdc.ac.cn/en/disallow/55fc9768-ea6a-4d16-9207-28f6aed4900b and https://doi.org/10.11888/Atmos.tpdc.300461, Tang, 2023), Institute of Tibetan Plateau Research, Chinese Academy of Sciences.

6 Summary

In this study, we have developed a dense station-based long-term dataset of daily surface solar radiation in China with high-accuracy. The dataset consists of estimates of three radiation components (global, direct and diffuse radiation) at the 2473 CMA meteorological stations during the period from the 1950s to 2021, with most stations covering the period of 1961-2021. The methods used to develop the dataset are the global radiation estimation model of Yang et al. (2006) and the direct radiation estimation model of Tang et al. (2018), and the main inputs are the five meteorological variables of surface pressure, air temperature, relative humidity, horizontal visibility, and sunshine duration measured at the meteorological stations. The developed dataset was evaluated against in-situ measurement collected at 96 CMA radiation stations, and further compared with other two satellite-based radiation products.

Averaged over all radiation stations and the time period of 1993-2010, the total RMSE values for daily global, direct and diffuse radiation are 23.2, 27.6 and 19.6 W m^{-2}, respectively. At the monthly mean scale, our estimates give RMSE values of about 13.4, 16.5 and 10.9 W m^{-2}, respectively, for monthly global, direct and diffuse radiation.

These error indicators on both daily and monthly scales are generally lower than those of the satellite radiation products, especially for global and direct radiation. Comparisons with the satellite-based radiation products indicate that our station-based estimates have a clear advantage in terms of accuracy and length of time series. However, our dataset does not provide radiation data beyond the weather stations. Merging the station-based estimates with the satellite-based retrievals will have good potential to improve the accuracy of the radiation products in the future. We expect that our station-based radiation dataset to contribute significantly to relevant basic research, engineering applications and fusion with satellite-based retrievals in the future.

Author contributions. All authors discussed the results and contributed to the manuscript. WT calculated the dataset, analyzed the results, and drafted the manuscript. JH drew the figures.

Competing interests. The authors declare that they have no conflicts of interest.

Acknowledgments. The ISCCP-HXG global product of global radiation was provided by the National Tibetan Plateau Data Center (TPDC). The radiation product of Jiang et al. (2016) was available from the Pangaea at https://doi.org/10.1594/PANGAEA.904136, and the CMA routine meteorological variables and radiation data were obtained from the CMA Meteorological Information Center. The authors would like to thank the staff of the data management and production

organizations for their valuable work.

Financial support. This work was supported by the National Key Research and Development Program of China (Grant No. 2022YFB4202104), and the National Natural Science Foundation of China (Grant No. 42171360).

References

Alton, P. B., North, P. R., Los, S. O.: The impact of diffuse sunlight on canopy light-use efficiency, gross photosynthetic product and net ecosystem exchange in three forest biomes. Global Change Biology, 13: 776-787, 2007.

Ångström, A.: Solar and terrestrial radiation. Q.J.R. Meteorol. Soc.50, 121-125, 1924.

Boland, J., Huang, J., & Ridley, B.: Decomposing global solar radiation into its direct and diffuse components. Renewable and Sustainable Energy Reviews, 28, 749-756. https://doi.org/https://doi.org/10.1016/j.rser.2013.08.023, 2013.

Ehnberg, J. S. G., & Bollen, M. H. J.: Simulation of global solar radiation based on cloud observations. Solar Energy, 78(2), 157–162. https://doi.org/10.1016/j.solener.2004.08.016, 2005.

Feng, F., Wang, K.: Merging high-resolution satellite surface radiation data with meteorological sunshine duration observations over China from 1983 to 2017. Remote Sensing. 13(4), 602, 2021a.

Feng, F., and Wang, K.: Merging ground-based sunshine duration observations with satellite cloud and aerosol retrievals to produce high-resolution long-term surface solar radiation over China, Earth System Science Data, 13(3): 907-922, 2021b.

Gu, L. H., Baldocchi, D., Verma, S. B., et al.: Advantages of diffuse radiation for terrestrial ecosystem productivity. Journal of Geophysical Research-Atmospheres, 107: 4050, 2002.

Hakuba, M.Z., Folini, D., Sanchez-Lorenzo, A., & Wild, M.: Spatial representativeness of ground-based solar radiation measurements. Journal of Geophysical Research: Atmospheres, 118, 8585-8597, https://doi.org/10.1002/2017JD027261, 2013.

Hao, D., Asrar, G. R., Zeng, Y., Zhu, Q., Wen, J., Xiao, Q., & Chen, M.: DSCOVR/EPIC-derived global hourly and daily downward shortwave and photosynthetically active radiation data at 0.1°× 0.1° resolution. Earth System Science Data, 12(3), 2209-2221, 2020.

Huang, G., Li, Z., Li, X., Liang, S., Yang, K., Wang, D., and Zhang, Y.: Estimating surface solar irradiance from satellites: Past, present, and future perspectives, Remote Sens. Environ., 233, 111371, https://doi.org/10.1016/j.rse.2019.111371, 2019.

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

Jiang, H., Lu, N., Qin, J., & Yao, L.: Hourly 5-km surface total and diffuse solar radiation in China, 2007–2018. Scientific Data, 7(1). https://doi.org/10.1038/s41597-020-00654-4, 2020a.

Jiang, H., Yang, Y., Bai, Y. & Wang, H.: Evaluation of the total, direct, and diffuse solar radiations from the ERA5 reanalysis data in China. IEEE Geosci. Remote S. 17, 47–51, 2020b.

Jiang, H., Yang, Y., Wang, H., Bai, Y. & Bai, Y.: Surface diffuse solar radiation determined by reanalysis and satellite over East Asia: evaluation and comparison. Remote Sensing, 12, 1387, 2020c.

Karakoti, I., Pande, B., & Pandey, K.: Evaluation of different diffuse radiation models for Indian stations and predicting the best fit model. Renewable and Sustainable Energy Reviews, 15(5), 2378-2384. https://doi.org/https://doi.org/10.1016/j.rser.2011.02.020, 2011.

Kato, S., Loeb, N. G., Rose, F. G., Doelling, D. R., Rutan, D. A., Caldwell, T. E., Yu, L., and Weller, R. A.: Surface irradiances consistent with CERES-derived top-of-atmosphere shortwave and longwave irradiances, J. Climate, 26, 2719–2740, 2013.

Lee, M. et al.: Model-based analysis of the impact of diffuse radiation on CO_2 exchange in a temperate deciduous forest. Agr. Forest Meteorol. 249, 377–389, 2017.

Letu, H., Yang, K., Nakajima, T. Y., Ishimoto, H., Nagao, T. M., Riedi, J., Baran, A. J., Ma, R., Wang, T., Shang, H., Khatri, P., Chen, L., Shi, C., Shi, J.: High-resolution retrieval of cloud microphysical properties and surface solar radiation using Himawari-8/AHI next-generation geostationary satellite. Remote Sens. Environ., 239, 111583, http://dx.doi.org/10.1016/j.rse.2019.111583, 2020.

Letu, H., Nakajima, T.Y., Wang, T.X., Shang, H.Z., Ma, R., Yang, K., Baran, A.J., Riedi, J., Ishimoto, H., Yoshida, M., Shi, C., Khatri, P., Du, Y.H., Chen, L.F., & Shi, J.C.: A new benchmark for surface radiation products over the East Asia-Pacific region retrieved from the Himawari-8/AHI next-generation geostationary satellite. Bulletin of the American Meteorological Society, 2021, E873–E888, https://journals.ametsoc.org/view/journals/bams/103/3/BAMS-D-20-0148.1.xml, 2021.

Li, R., Wang, D., & Liang, S.: Comprehensive assessment of five global daily downward shortwave radiation satellite products. Science of Remote Sensing, 4, 100028, 2021.

Li, R., Wang, D., Wang, W., and Nemani, R.: A GeoNEX-based high-spatiotemporal-resolution product of land surface downward shortwave radiation and photosynthetically active radiation, Earth Syst. Sci. Data, 15, 1419–1436, https://doi.org/10.5194/essd-15-1419-2023, 2023.

Liu, X. Y., Mei, X. R., Li, Y. Z., Wang, Q., Jensen, J. R., Zhang, Y., & Porter, J. R.: Evaluation of temperature-based global solar radiation models in China. Agricultural and Forest Meteorology, 149, 1433–1446, 2009.

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, https://doi.org/10.1016/j.energy.2011.03.007, 2011.

Mellit, A., Eleuch, H., Benghanem, M., Elaoun, C., & Pavan, A. M.: An adaptive model for predicting of global, direct and diffuse hourly solar irradiance. Energy Conversion and Management, 51(4), 771-782. https://doi.org/https://doi.org/10.1016/j.enconman.2009.10.034, 2010.

Mercado L. M., Bellouin N., Sitch S., et al.: Impact of changes in diffuse radiation on the global land carbon sink. Nature, 458: 1014-1017, 2009.

Pelaez, S. A., Deline, C., Macalpine, S. M., Marion, B., Stein, J. S., & Kostuk, R. K.: Comparison of Bifacial Solar Irradiance Model Predictions With Field Validation. IEEE Journal of Photovoltaics, 9(1), 82-88. https://doi.org/10.1109/jphotov.2018.2877000, 2019.

Pinker, R. T. and Laszlo, I.: Modeling surface solar irradiance for satellite application on a global scale, J. Appl. Meteorol., 31, 194–211, https://doi.org/10.1175/1520-0450(1992)031<0194:MSSIFS>2.0.CO;2, 1992.

Pohlert, T.: Use of empirical global radiation models for maize growth simulation. Agric. Forest Meteorol. 126, 47–58, 2004.

Prescott, J.A.: Evaporation from a water surface in relation to solar radiation. Trans. Roy. Soc. Austr. 641, 114-125, 1940.

Rodríguez-Gallegos, C. D., Bieri, M., Gandhi, O., Singh, J. P., Reindl, T., & Panda, S. K.: Monofacial vs bifacial Si-based PV modules: Which one is more cost-effective? Solar Energy, 176, 412-438. https://doi.org/https://doi.org/10.1016/j.solener.2018.10.012, 2018.

Shao, C., Yang, K., Tang, W., et al.: Convolutional neural network-based homogenization for constructing a long-term global surface solar radiation dataset. Renewable and Sustainable Energy Reviews. 169, 112952, 2022.

Shi, G.Y., Hayasaka, T., Ohmura, A., Chen, Z.H., Wang, B., Zhao, J.Q., Che, H.Z., Xu, L.: Data quality assessment and the long-term trend of ground solar radiation in China. J. Appl. Meteor. Climatology, 47, 1006–1016, 2008.

Tang, W., Yang, K., He, J., & Qin, J.: Quality control and estimation of global solar radiation in China. Solar Energy, 84(3), 466-475. https://doi.org/10.1016/j.solener.2010.01.006, 2010.

Tang, W. J., Yang, K., Qin, J., Cheng, C. C. K., & He, J.: Solar radiation trend across China in recent decades: a revisit with quality-controlled data. Atmospheric Chemistry and Physics, 11(1), 393-406. https://doi.org/10.5194/acp-11-393-2011, 2011.

Tang, W., Yang, K., Qin, J., & Min, M.: Development of a 50-year daily surface solar radiation dataset over China. Science China Earth Sciences, 56(9), 1555–1565. https://doi.org/10.1007/s11430-012-4542-9, 2013.

Tang, W., Qin, J., Yang, K., Liu, S., Lu, N., and Niu, X.: Retrieving high-resolution surface solar radiation with cloud parameters derived by combining MODIS and MTSAT data, Atmos. Chem. Phys., 16, 2543–2557, https://doi.org/10.5194/acp-16-2543-2016, 2016.

Tang W., Yang, K., Qin, J., Niu, X., Lin, C., Jing, X.: A revisit to decadal change of aerosol optical depth and its impact on global radiation over China. Atmospheric Environment, 150: 106-115, 2017a.

Tang, W., Yang, K., Sun, Z., Qin, J., and Niu, X.: Global Performance of a Fast Parameterization Scheme for Estimating Surface Solar Radiation From MODIS Data, IEEE T. Geosci. Remote Sens, 55, 3558–3571, https://doi.org/10.1109/TGRS.2017.2676164, 2017b.

Tang, W., Yang, K., Qin, J., Min, M., & Niu, X.: First Effort for Constructing a Direct

Solar Radiation Data Set in China for Solar Energy Applications. Journal of Geophysical Research: Atmospheres, 123(3), 1724-1734. https://doi.org/10.1002/2017jd028005, 2018.

Tang, W., Yang, K., Qin, J., Li, X., & Niu, X.: A 16-year dataset (2000–2015) of high-resolution (3 h, 10 km) global surface solar radiation. Earth System Science Data, 11(4), 1905-1915. https://doi.org/10.5194/essd-11-1905-2019, 2019a.

Tang, W., Li, J., Yang, K., Qin, J., Zhang, G., Wang, Y.: Dependence of remote sensing accuracy of global horizontal irradiance at different scales on satellite sampling frequency, Solar Energy, 193: 597–603, 2019b.

Tang, W. (2023). A dense station-based long-term and high-accuracy dataset of daily surface solar radiation in China. National Tibetan Plateau/Third Pole Environment Data Center, https://doi.org/10.11888/Atmos.tpdc.300461.

Wang, D., Liang, S., Zhang, Y., Gao, X., Brown, M. G., & Jia, A.: A new set of MODIS land products (MCD18): Downward shortwave radiation and photosynthetically active radiation. Remote Sensing, 12(1), 168, 2020.

Wang, K. C., Dickinson, R. E., Wild, M., and Liang, S.: Atmospheric impacts on climatic variability of surface incident solar radiation, Atmos. Chem. Phys., 12, 9581–9592, https://doi.org/10.5194/acp-12-9581-2012, 2012.

Wang, L., Kisi, O., Zounemat-Kermani, M., Salazar, G., Zhu, Z., & Gong, W.: Solar radiation prediction using different techniques: Model evaluation and comparison. Renewable and Sustainable Energy Reviews, 61, 384–397. https://doi.org/10.1016/j.rser.2016.04.024, 2016.

Wild, M.: Global dimming and brightening: A review, J. Geophys.Res.-Atmos., 888 114, D00D16, 2009.

Wild, M., Folini, D., Hakuba, M.Z., et al.: The energy balance over land and oceans: an assessment based on direct observations and CMIP5 climate models. Climate Dynamics. 44, 3393-3429, 2015.

Wild, M., Ohmura, A., Schär, C., et al.: The Global Energy Balance Archive (GEBA) version 2017: A database for worldwide measured surface energy fluxes. Earth System Science Data. 9(2), 601-613, 2017.

Yang, X., Li, J., Yu, Q., et al.: Impacts of diffuse radiation fraction on light use efficiency and gross primary production of winter wheat in the North China Plain. Agricultural and Forest Meteorology, 275: 233-242, 2019.

Yang, K., Huang, G.-W., & Tamai, N.: A hybrid model for estimating global solar radiation. Solar Energy, 70(1), 13–22. https://doi.org/10.1016/S0038-092X(00)00121-3, 2001.

Yang, K., Koike, T., & Ye, B.: Improving estimation of hourly, daily, and monthly downward shortwave radiation by importing global data sets. Agricultural and Forest Meteorology, 137(1-2), 43–55. https://doi.org/10.1016/j.agrformet.2006.02.001, 2006.

Yang, K., He, J., Tang, W., Qin, J., Cheng, C. C. K.: On downward shortwave and longwave radiations over high altitude regions: Observation and modeling in the Tibetan Plateau, Agricultural and Forest Meteorology, 150(1): 38-46, 2010.

Zhang, X., Liang, S., Zhou, G., Wu, H., and Zhao, X.: Generating Global LAnd Surface

Satellite incident shortwave radiation and photosynthetically active radiation products from multiple satellite data, Remote Sens. Environ., 152, 318–332, 2014.

Zhang, Y. C., Rossow, W. B., Lacis, A. L., Valdar, O., and Michael, I. M.: 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, J. Geophys. Res., 109, D19105, https://doi.org/10.1029/2003JD004457, 2004.