Atmospheric Research 270 (2022) 106056

Contents lists available at ScienceDirect

Atmospheric Research

journal homepage: www.elsevier.com/locate/atmosres

# Assessment of surface downward longwave radiation in CMIP6 with comparison to observations and CMIP5

Jiawen Xu [a,b], Xiaotong Zhang [a,b,*], Weiyu Zhang [a,b], Ning Hou [a,b], Chunjie Feng [a,b], Shuyue Yang [a,b], Kun Jia [a,b], Yunjun Yao [a,b], Xianhong Xie [a,b], Bo Jiang [a,b], Jie Cheng [a,b], Xiang Zhao [a,b], Shunlin Liang [c]

[a] State Key Laboratory of Remote Sensing Science, Jointly Sponsored by Beijing Normal University and Institute of Remote Sensing and Digital Earth of Chinese Academy of Sciences, Beijing 100875, China

[b] Beijing Engineering Research Center for Global Land Remote Sensing Products, Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

[c] Department of Geographical Sciences, University of Maryland, College Park, MD 20742, USA

ARTICLE INFO

*Keywords:*
Surface downward longwave radiation (SDLR)
CMIP6
CMIP5
GCMs
Bayesian model averaging
Multimodel ensemble

ABSTRACT

Surface downward longwave radiation (SDLR) plays an important role in understanding the greenhouse effect and global warming. The simulated SDLR from 47 coupled models in the Coupled Model Intercomparison Project (CMIP6) general circulation models (GCMs) was evaluated by comparing them with ground measurements and CMIP5 results. The estimated SDLR using all CMIP6 GCMs based on the multimodel ensemble (MME) methods was validated as well. The bias values of the SDLR simulations from individual CMIP6 GCMs averaged over the selected 183 sites around the world varied from −10 to 10 W m$^{-2}$, while the root mean squared error (RMSE) values ranged from 20 to 26 W m$^{-2}$. Compared to CMIP5 models, the CMIP6 GCMs did not show a significant tendency to underestimate SDLR. However, the SDLR from CMIP6 GCMs exhibited the relatively better precision at low altitude and low latitude sites compared to that at high altitude and high latitude sites. Moreover, the Bayesian model averaging (BMA) method increased the correlation coefficient (R) by approximately 0.02 and reduced the RMSE by approximately 5 W m$^{-2}$ on average compared to the individual CMIP6 GCMs. The trend in SDLR was also investigated in this study, which has been related to the changes in air temperature (SAT), and water vapor pressure (WVP).

## 1. Introduction

The surface radiation budget plays a vital role in atmospheric and oceanic general circulations and fundamentally influences radiation fluxes within the Earth's climate system and its internal distributions (Cheng et al., 2019; Stephens et al., 2012a; Wild, 2020). It is a main component of the radiative energy exchange between the land/ocean surface and atmosphere and consists of upward and downward fluxes of longwave and shortwave radiation (Gupta et al., 1999; Qin et al., 2020a). Surface downward longwave radiation (SDLR) is not only an essential variable in energy balances, meteorological and climatic studies (Ahmed et al., 2020; Cheng et al., 2020; Guo et al., 2019) but also affects the carbon, water and energy cycles. Additionally, SDLR has been related to global warming and greenhouse effects (Ma et al., 2014; Philipona et al., 2004).

SDLR can be obtained from ground measurements (Augustine et al., 2000; Delia Garcia et al., 2019; Ohmura et al., 1998), reanalysis data (Flynn et al., 2019; Hinkelman, 2019; Wang and Dickinson, 2013), remote sensing (Loeb et al., 2013; Wang et al., 2020; Zhou et al., 2019), and general circulation models (GCMs) (Ma et al., 2014; Wild, 2020; Wild et al., 1998). Among them, GCMs, which are tools primary used to investigate past and future climate changes (Kim et al., 2020; Perez et al., 2014), have the advantage of producing long-term global or regional energy budget components (Zhang et al., 2019b). To support an activity proposed through the World Climate Research Programme (WCRP), the Working Group on Coupled Modeling (WGCM) organized the 5th phase of the Coupled Model Intercomparison Project (CMIP5) as common protocols to be applied in experiments with the same

* Corresponding author at: State Key Laboratory of Remote Sensing Science, Jointly Sponsored by Beijing Normal University and Institute of Remote Sensing and Digital Earth of Chinese Academy of Sciences, Beijing 100875, China.
*E-mail address:* xtngzhang@bnu.edu.cn (X. Zhang).

https://doi.org/10.1016/j.atmosres.2022.106056
Received 2 April 2021; Received in revised form 24 January 2022; Accepted 26 January 2022
Available online 9 February 2022

parameters by different GCMs (Guilyardi et al., 2013). A series of experiments completed during CMIP5 contributed to the Intergovernmental Panel on Climate Change (IPCC) 5th Assessment Report (AR5) as the basis for researching the responses of the climate to external forcings (Kusunoki and Arakawa, 2015; Taylor et al., 2012). Currently, the latest generation of GCMs conducted for the 6th phase of the Coupled Model Intercomparison Project (CMIP6) experiments (Eyring et al., 2016) has become available with improvements in its dynamical processes and higher spatial resolution (Kim et al., 2020; Marotzke et al., 2017). It focuses on examining changes under climate extremes and understanding the associated physical processes (Kim et al., 2020) and provides opportunities to evaluate how well SDLR is simulated by CMIP6 GCMs compares to that of accessible ground observations.

Previous studies have comprehensively validated the abilities of GCMs to simulate SDLR (Li et al., 2016; Ma et al., 2014; Wild, 2008; Wild et al., 2013; Wild et al., 2019). For instance, Bodas-Salcedo et al. (2008) reported that the Hadley Centre Global Environmental Model version 1 (HadGEM1) underestimated SDLR at Baseline Surface Radiation Network (BSRN) sites by 6 W m$^{-2}$. It was also found that the monthly SDLR simulations from 44 CMIP5 GCMs showed negative average biases of 1.2 W m$^{-2}$ and 5.4 W m$^{-2}$ by comparing ground measurements from the BSRN and buoy, respectively (Ma et al., 2014). Wild et al. (2015) validated 43 CMIP5 GCMs in simulating annual SDLR at 41 BSRN sites with a negative average bias of 3 W m$^{-2}$. In comparison with the Clouds and the Earth's Radiant Energy System, Energy Balanced and Filled (CERES EBAF), the SDLR simulations from the CMIP5 GCMs were significantly underestimated in most land regions, especially during the summer season (JJA) (Li et al., 2016). Negative biases of the Coupled Earth System Model version 1 with the Coupled Atmosphere Model version 5 (CESM1-CAM5) SDLR against CERES EBAF within a range of 58–70° S and 60–90° N over the ocean for each calendar month were found by Li et al. (2017) and Li et al. (2019). Based on these studies, it is obvious that the GCMs showed a tendency to underestimate SDLR, which has been a long-standing issue over several decades and evolutions of GCMs (Bodas-Salcedo et al., 2008; Wild, 2020; Wild et al., 2015). The Earth System Grid Federation (ESGF) has recently released CMIP6 GCMs using common formats and metadata (Eyring et al., 2016). Wild (2020) demonstrated that a global mean SDLR of 343.8 W m$^{-2}$ estimated by 38 CMIP6 GCMs from 2000 to 2014, was higher than that obtained based on 22 CMIP5 GCMs (342.3 W m$^{-2}$) (Wild et al., 2013) and 43 CMIP5 GCMs (341.5 W m$^{-2}$) (Wild et al., 2015) from 2000 to 2004. These results indicated that CMIP6 GCMs have substantially improved the long-standing underestimation of SDLR. However, the SDLR simulations from the CMIP6 GCMs have not yet been evaluated with ground measurements.

SDLR began being observed in the early 1990s at far fewer BSRN sites. Worldwide ground-measured SDLR has been used to detect long-term SDLR variability (Nyeki et al., 2019; Prata, 2010; Stephens et al., 2012b; Wacker et al., 2011; Wang and Liang, 2009). For example, Wacker et al. (2011) reported that the clear-sky SDLR showed an increasing trend of 3.5 W m$^{-2}$ per decade from 1996 to 2007 at four sites from the Swiss Alpine Surface Radiation Budget Network (ASRB). Stephens et al. (2012b) demonstrated that the clear-sky SDLR over the global oceans increased by 1.8 W m$^{-2}$ per decade using observations of sea surface temperature and oceanic-wide precipitable water collecting during 1988–2005. In addition to the ground measurements, considerable effort has been made in tracking long-term SDLR variabilities using SDLR simulations from GCMs (Kim et al., 2013; Ma et al., 2014; Wild, 2016; Wild et al., 2008). Wild et al. (2008) illustrated that the SDLR increased at a rate of 2.6 W m$^{-2}$ per decade at the 12 earliest sites collected from the BSRN during 1992–2000. This was consistent with the respective change (+2.4 W m$^{-2}$ per decade) based on a transient GCM experiment. In a later study, an increasing trend in SDLR of 2.0 W m$^{-2}$ per decade at 25 BSRN sites since the early 1990s was identified, which agreed with that obtained from the SDLR simulations from CMIP5 GCMs (Wild, 2016). An increase in the global mean SDLR of 1.5 W m$^{-2}$ per decade over the period of 1979–2005 was found using 44 CMIP5 GCMs by Ma et al. (2014). Although considerable effort has been made on identifying long-term variabilities in SDLR, it is still of great importance to assess the trends of SDLR in CMIP6 GCMs.

Individual GCMs have merits and shortcomings in simulating SDLR, which suggests that the uncertainty of SDLR estimations may be ignored and underestimated when only considering and implementing a single model (Bhat et al., 2011; Zhang et al., 2019a). The multimodel ensemble (MME) method, which is a promising approach that utilizes advantage of different GCMs, has been effectively applied in hydrologic and climatic variable estimations (Duan et al., 2007; Yao et al., 2016; Zhang et al., 2019a). Many studies have indicated that even a simple MME method is more consistent with ground measurements than single model estimations, such as the simple model averaging (SMA) method (Duan et al., 2007; Wu et al., 2012; Yao et al., 2016; Zhang et al., 2019b). In addition to the SMA method, more complicated MME methods have been developed that assign weights based on the ability of the models during a training period (Bhat et al., 2011). For example, the Bayesian model averaging (BMA) method proposed by Raftery et al. (2005) is a postprocessing approach to estimate climatic variables using simulations from multiple models (Jia et al., 2020). It has been widely used in a series of scientific studies, such as energy budget estimation (Wu et al., 2012; Zhang et al., 2019b), climate changes (Bhat et al., 2011; Min et al., 2007; Smith et al., 2009) and hydrological simulation (Duan et al., 2007). Those studies demonstrated that the BMA method was excellent for incorporating the strengths of different models (Duan et al., 2007; Fang and Li, 2016; Jia et al., 2020).

Therefore, the object of the present research is to evaluate the performance and applicability of the MME methods to estimate SDLR with simulations from the CMIP6 GCMs. This study first validated the SDLR simulations from 47 CMIP6 GCMs with ground measurements and investigated how well the SMA and BMA methods perform in the estimation of SDLR using CMIP6 GCMs. Second, the ability of CMIP6 GCMs to simulate SDLR was compared with that of CMIP5 GCMs with respect to SDLR observations. Finally, we explored the temporal evolutions of both ground-measured SDLR and CMIP6 GCM SDLR simulations.

## 2. Data

### 2.1. CMIP5 and CMIP6 GCMs

The available simulations on monthly SDLR from 93 GCMs, which include 46 GCMs from CMIP5 and 47 GCMs from CMIP6, were used in this study. Tables A1 and A2 summarize the names of the models as well as the institutions to which they are associated, their time period and their spatial resolution.The SDLR simulations in the 46 CMIP5 GCMs under ensemble 'r1i1p1' during 1861–2005 are from "historical" experiments (https://esgf-node.llnl.gov/search/cmip5/). For the 47 CMIP6 GCMs, "historical" experiments under ensemble 'r1i1p1f1' employed in this study cover the period from 1850 to 2014 (https://esgf-node.llnl.gov/search/cmip6/). The "historical" experiments mainly considered anthropogenic and natural forcings, such as solar radiation, aerosols, greenhouse gases, and land use (Eyring et al., 2016). Both CMIP5 and CMIP6 GCMs were resampled to 1° × 1° with bilinear interpolation before evaluation and comparison.

### 2.2. Ground-measured data

Ground-measured SDLR at 183 sites across the world were collected in this study, which included 61 BSRN sites, 7 Surface Radiation Budget Network (SURFRAD) sites, and 115 FLUXNET2015 (FLUXNET) sites. These stations are globally located in different climatic zones, with latitudes ranging from 82.490° N to 89.983° S and longitudes ranging from 156.607° W to 169.689° E. The sites encompass an altitude ranging from −9 m to 4319 m and include cropland, grassland, forest, bare land, desert and other land types. A list of the sites utilized in the present research is given in Table A3. The locations of all 183 sites are exhibited in Fig. 1.

BSRN is a global observation network established by the WCRP in 1992 to measure surface radiation with a high temporal resolution and a high accuracy (Ohmura et al., 1998; Wild, 2017; Zhang et al., 2016). These samples, with a channel frequency of 1 Hz, are recorded as 1-min values (Ohmura et al., 1998). The uncertainty of monthly SDLR measurements was within 10 W $m^{-2}$ after the BSRN standard was established in the early 1990s (Ma et al., 2014). At present, there are 68 BSRN stations in various climate zones (Hatzianastassiou et al., 2020) (https://bsrn.awi.de/), wherein the SDLR observations at 61 BSRN sites from 1992 to 2014 were used to evaluate the performance of CMIP6 GCMs to simulate SDLR.

SURFRAD has been maintained by the National Oceanic and Atmospheric Administration (NOAA) since 1993 to supply long-term and continuous radiation measurements (Cheng et al., 2020; Guo et al., 2020), which has contributed to American climate research (Liang et al., 2010). Twenty years (1995–2014) of SDLR measurements at seven operational SURFRAD sites (Sekertekin et al., 2020) were downloaded and applied in this study (https://www.esrl.noaa.gov/gmd/grad/surfrad/). These SDLR measurements, with frequencies of 1 Hz for all instruments, were recorded every 3 min (Augustine et al., 2000; Verma et al., 2016) and have been aggregated to 1-min averages since 2009 (Qin et al., 2020a). The measurement uncertainty is approximately ±9 W $m^{-2}$ (Augustine et al., 2000).

FLUXNET is a worldwide network that provides micrometeorological measurements (Winter et al., 2009), which is composed of AmeriFlux, AsiaFlux, CarboEuropeIP, ChinaFlux and other regional networks (Liang et al., 2010). The FLUXNET tower sites measure the surface radiation budget with frequencies ranging from 10 to 20 Hz using eddy covariance methods and record samples every 30 min (Carrer et al., 2012; Pastorello et al., 2020; Verma et al., 2016). The SDLR accuracy exceeds 10 W $m^{-2}$ for certain FLUXNET stations (Liang et al., 2010). FLUXNET2015, which is the third-generation dataset (Pastorello et al., 2020) with 166 sites in the CC-BY-4.0 dataset, was created as a global eddy covariance dataset to improve consistency and comparability across sites (https://fluxnet.fluxdata.org/data/fluxnet2015-dataset/). The measured SDLR used in this study was collected from 115 CC-BY-4.0 sites during 1998–2014.

The SDLR measurements provided by BSRN and SURFRAD are instantaneous values and need to be processed critically to obtain accurate monthly SDLR values. First, the missing values were filled using linear interpolation when there were at least 80% effective values collected in a day. Then, the daily mean SDLR was estimated from the instantaneous values through integration. Finally, the monthly mean SDLR was obtained using the daily mean SDLR values collected over a month. Only the daily mean SDLR values were missing within 10 days in a month, the values for that month were used for research.

### 2.3. ERA5

ERA5, which was developed by the European Centre for Medium-Range Weather Forecasts (ECMWF) in 2017, is the 5th generation reanalysis (Graham et al., 2019). It was conducted for the global climate to replace ERA-Interim with data available from 1979 to present (Hamal et al., 2020). ERA5 assimilates ground measurements using the 4D-Var method (Chen et al., 2020). The monthly mean SDLR data on single levels with a spatial resolution of 0.25° × 0.25° (Flynn et al., 2019) during 1979–2014 from ERA5 were used in this study (https://cds.climate.copernicus.eu/), which were resampled to the same spatial resolution as that of GCMs at 1° × 1°.

### 2.4. CERES EBAF

CERES EBAF is produced by the National Aeronautics and Space Administration (NASA) for evaluating climate models and estimating the surface energy budget, with data ranging from 2000 to the present (Hinkelman, 2019; Zhang et al., 2016). The data is obtained from instruments on the Aqua, Terra, Suomi National Polar-orbiting Partnership (S-NPP) (Loeb et al., 2018; Smith et al., 2014) and Joint Polar Satellite System 1 (JPSS-1) satellites (Smith et al., 2018). The estimates of SDLR are improved by the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) and CloudSat (Verma et al., 2016). The SDLR data applied in this research included the monthly estimates provided by the CERES EBAF Ed4.1 dataset with a spatial resolution of 1° × 1° (Loeb et al., 2013) from 2000 to 2014 (https://ceres.larc.nasa.gov/).

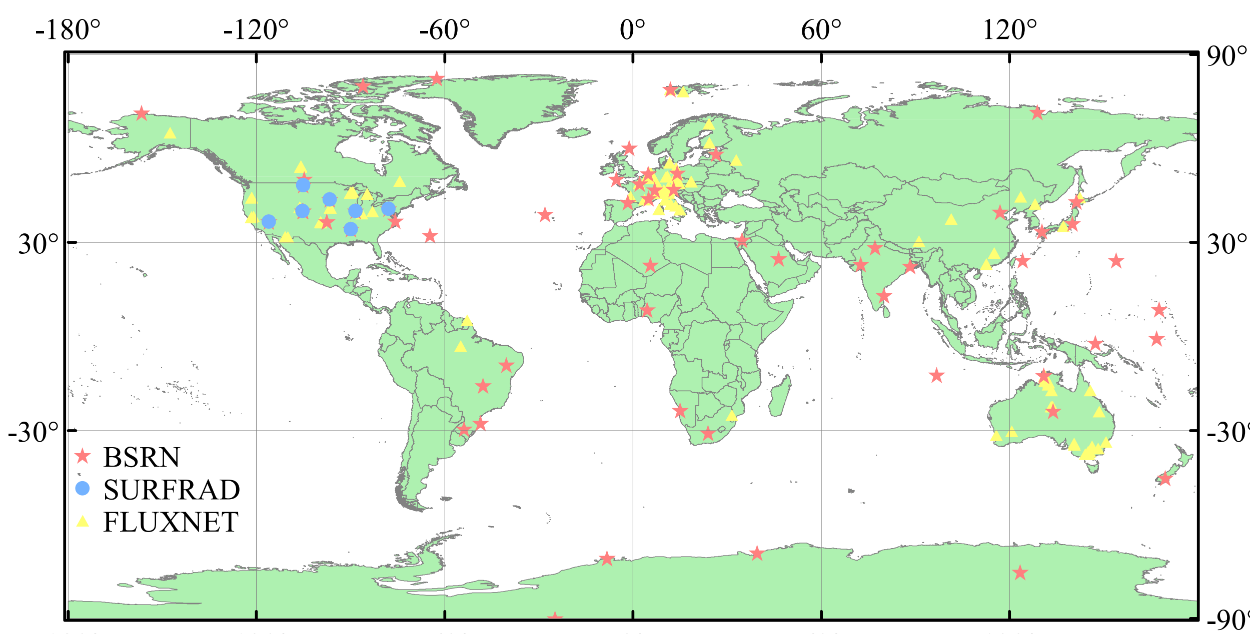


**Fig. 1.** Geographical distributions of the observation sites.

## 3. Methods

### 3.1. The Multimodel ensemble (MME) methods

In this study, two common MME methods were evaluated for SDLR estimation using simulations from GCMs based on the following equations. For the SMA method, the weights of the different GCMs are equal and set to 1/K, wherein K denotes the number of GCMs. The SDLR estimated by the SMA method is calculated by weighted averaging the individual models. The BMA method is applied to calculate the relative weights of individual models and the deterministic result based on the Bayesian theory (Fang and Li, 2016). The weights are the posterior probability of each participating model, which are nonnegative values that add up to one (Jia et al., 2020). They can be acquired through a maximum likelihood function from the training data (Guo et al., 2019; Medina and Tian, 2020). To ensure the numerical stability and simplification of the training procedure, Raftery et al. (2005) introduced the expectation maximization (EM) algorithm to maximize the log-likelihood function rather than the likelihood function itself. The deterministic result is estimated by weighted averaging the simulations from multiple models, which are obtained from the bias-correction process (Fang and Li, 2016). Raftery et al. (2005) comprehensively introduced more descriptions of the BMA method. In the present research, these SDLR measurements during the period of 1992–2014 at 183 sites spread across the world were applied for the BMA analysis.

### 3.2. Statistical metrics

Five statistical metrics were utilized in this study: the root mean square error (RMSE), relative root mean square error (RRMSE), mean bias error (bias), relative mean bias error (Rbias) and correlation coefficient (R) values. These selected statistical metrics are calculated using the following equations:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(e_i - o_i)^2} \tag{1}$$

$$Bias = \frac{1}{n}\sum_{i=1}^{n}(e_i - o_i) \tag{2}$$

$$RRMSE = \frac{100}{\overline{o}} \times \sqrt{\frac{1}{n}\sum_{i=1}^{n}(e_i - o_i)^2} \tag{3}$$

$$RBias = \frac{100}{\overline{o}} \times \frac{1}{n}\sum_{i=1}^{n}(e_i - o_i) \tag{4}$$

$$R = \frac{n\sum_{i=1}^{n} e_i o_i - \sum_{i=1}^{n} e_i \times \sum_{i=1}^{n} o_i}{\sqrt{\left[n\sum_{i=1}^{n} {e_i}^2 - \left(\sum_{i=1}^{n} e_i\right)^2\right] \times \left[n\sum_{i=1}^{n} {o_i}^2 - \left(\sum_{i=1}^{n} o_i\right)^2\right]}} \tag{5}$$

where $n$ denotes the number of data; $e_i$ and $o_i$ are the simulated and observed SDLR values, respectively; and $\overline{o}$ denotes the average observed SDLR values.

The accuracy of the simulated SDLR values was analyzed using five statistical metrics; however, individual metrics could not indicate the overall accuracy. Thus, the global performance indicator (GPI) (Qin et al., 2020b) was applied in this study to validate the overall accuracy of the individual GCMs, which can be calculated as follows:

$$GPI_k = \sum_{j=1}^{m} X_j\left(\widetilde{Y}_j - Y_{kj}\right) \tag{6}$$

where $\tilde{Y}_j$ and $Y_{kj}$ denote the median of the absolute values of metric $j$ and the absolute value of metric $j$ for model $k$, respectively, and $m$ is the number of metrics. $X_j$ is −1 for R and 1 for other metrics. For the GPI, a higher value suggests a better performance of model $k$.

## 4. Results and analysis

This study used the ground-measured SDLR as a major reference dataset to validate the SDLR simulations from CMIP6 and CMIP5 GCMs. We compared the GCM gridded SDLR simulations with the SDLR measurements within the grids. Since both CMIP5 and CMIP6 GCMs applied in this study had different spatial resolutions, which ranged from 3.75° × 3.75° to 0.56° × 0.56° and from 2.81° × 2.81° to 0.70° × 0.70°, respectively, all GCMs were resampled to 1° × 1° with bilinear interpolation before evaluation and comparison.

### 4.1. Evaluation with ground measurements

#### 4.1.1. CMIP6 GCMs SDLR evaluation

SDLR ground measurements (17,625 samples) collected at 183 sites during 1992–2014 were applied to evaluate the performance of monthly SDLR simulations from 47 CMIP6 GCMs. Fig. 2 indicate the evaluation results of the SDLR simulations at the sites from BSRN, SURFRAD and FLUXNET. The CMIP6 GCMs reported RMSE (RRMSE) values in the range of 17–26 W m$^{-2}$ (5.5%–8.2%) with respect to the seven SURFRAD sites within twenty years (1995–2014) of SDLR records, with an average RMSE (RRMSE) value of 20 W m$^{-2}$ (6.5%). The RMSE (RRMSE) values for over half (27 GCMs) of the CMIP6 GCMs were less than 20 W m$^{-2}$ (6.5%). The biases (Rbias) of the individual GCMs ranged from −16 to 7 W m$^{-2}$ (−5.0% to 2.3%). It was found that SDLR was underestimated at 37 out of the 47 GCMs with the average bias (Rbias) amounting to −5 W m$^{-2}$ (−1.5%). The R values at the seven SURFRAD sites for the individual GCMs varied from 0.89 to 0.95, with an average R value of 0.93. No individual GCM had an R value greater than 0.95. The GPI values at the SURFRAD sites were higher than −5 for 31 out of the 47 GCMs, and the average GPI value was −3. Among the 47 CMIP6 GCMs, BCC-ESM1 had the best performance with an RMSE of 18.01 W m$^{-2}$, a bias of 0.81 W m$^{-2}$, an R of 0.926 and a maximum GPI value of 6.563 for the seven SURFRAD sites, followed by NorESM2-LM, FIO-ESM-2-0, MIROC6 and AWI-ESM-1-1-LR. NorCPM1 demonstrated the poorest ability to simulate SDLR at the SURFRAD sites, which had an RMSE of 24.22 W m$^{-2}$, a bias of 15.19 W m$^{-2}$, an R of 0.928 and a minimum GPI value of −20.679, followed by EC-Earth3-Veg-LR and CAMS-CSM1–0.

We also validated the SDLR simulations from CMIP6 GCMs using 61 sites from the BSRN, which has a worldwide distribution and is considered one of the most qualified ground measurements. The RMSE (RRMSE) values averaged over 61 sites from the BSRN for the SDLR simulations ranged from 18 to 25 W m$^{-2}$ (5.9% to 8.1%), and the average RMSE (RRMSE) value was 21 W m$^{-2}$ (6.7%). 15 out of the 47 GCMs had RMSE (RRMSE) values less than 20 W m$^{-2}$ (6.5%). The biases (Rbias) at the BSRN sites for the individual GCMs varied from −13 to 9 W m$^{-2}$ (−4.1% to 2.7%). The SDLR simulations were underestimated by only 27 GCMs, and they agreed well with ground measurements with a lower negative average bias (Rbias) of 2 W m$^{-2}$ (0.6%). The SDLR simulations showed R values ranging from 0.95 to 0.98 at the BSRN sites, which were typically higher than those at the SURFRAD sites. The R values exceeded 0.95 at all 47 GCMs, with a higher average R value of 0.97. With respect to the BSRN sites, the majority of the GCMs (37 GCMs) had GPI values above −5, and the average GPI value was −1. AWI-ESM-1-1-LR, which also performed better at the seven sites from SURFRAD, was the best model for 61 BSRN sites compared to other individual GCMs, with an RMSE of 19.56 W m$^{-2}$, a bias of 0.80 W m$^{-2}$, an R of 0.967 and a maximum GPI value of 4.475, followed by FIO-ESM-2-0 and MPI-ESM1–2-HR. Consistent with the evaluation results for SURFRAD sites, the NorCPM1 SDLR simulations showed the poorest performance at the BSRN sites, which had an RMSE of 24.89 W m$^{-2}$, a bias of 12.66 W m$^{-2}$, an R of 0.964 and a minimum GPI value of

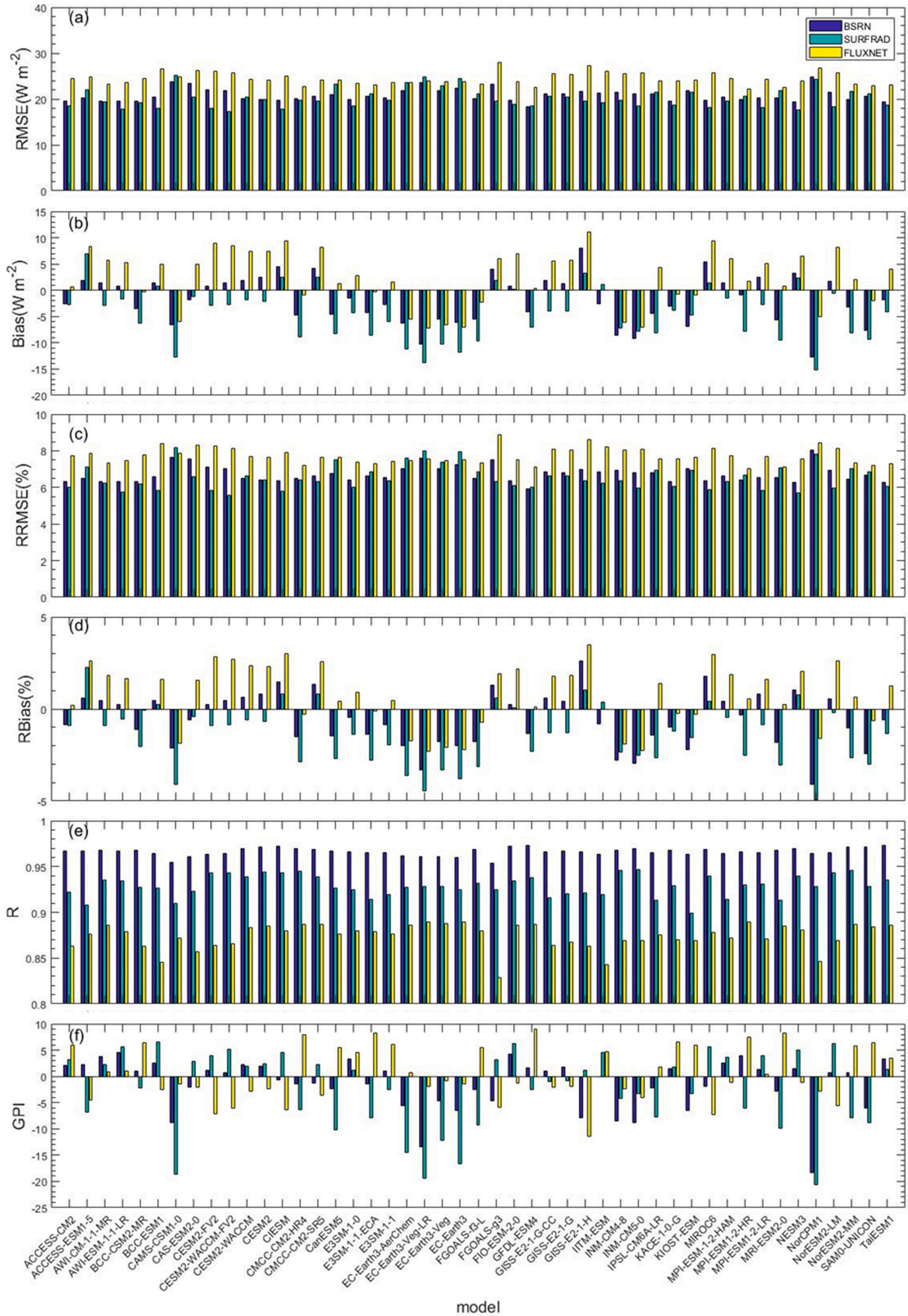


**Fig. 2.** Comparison of RMSE (a), Bias (b), RRMSE (c), RBias (d), R (e) and GPI (f) histograms of three observation networks for monthly downward longwave radiation (SDLR) simulations from the 47 CMIP6 GCMs.

−18.256, followed by EC-Earth3-Veg-LR and CAMS-CSM1–0.

To further evaluate the influence of site selection and record quality on the GCMs in SDLR, the above analysis was also repeated at 115 FLUXNET sites. For the selected sites, the RMSE (RRMSE) values varied between 22 and 29 W $m^{-2}$ (7.0% and 8.9%). No individual GCM had an RMSE (RRMSE) value within 20 W $m^{-2}$ (6.5%). The biases (Rbias) of the SDLR simulations at the 115 sites ranged from −8 to 12 W $m^{-2}$ (−2.3% to 3.5%), which were positive for most GCMs (31 GCMs). The CMIP6 GCMs showed an average RMSE (RRMSE) of 24 W $m^{-2}$ (7.7%) and absolute average bias (Rbias) of 2 W $m^{-2}$ (0.7%) at the 115 FLUXNET sites, which were higher than those at the 61 BSRN sites. This may be because FLUXNET has a lower temporal resolution of 30 min, while the SDLR observations are recorded as 1-min values at the BSRN sites. The R values at the FLUXENT sites were obviously lower, in the range of 0.82–0.89, with an average R value of 0.87. There was no individual GCM with an R value greater than 0.95. The GPI values for the majority of the GCMs (40 GCMs) at the FLUXNET sites were greater than −5, with an average GPI value of 0.5. Among all of the CMIP6 GCMs, GFDL-ESM4 agreed best with the FLUXNET measurements, which showed an RMSE of 22.53 W $m^{-2}$, a bias of 0.30 W $m^{-2}$, an R of 0.887 and a maximum GPI value of 8.982, followed by MRI-ESM2–0, E3SM-1-1-ECA, CMCC-CM2-HR4 and MPI-ESM1–2-HR. The performance of MPI-ESM1–2-HR was better for both the BSRN and FLUXNET sites. GISS-E2–1-H performed the worst in simulating SDLR at the FLUXNET sites, with an RMSE of 27.23 W $m^{-2}$, a bias of 11.05 W $m^{-2}$, an R of 0.863 and a minimum GPI value of −9.791.

Fig. 3 show the statistical metrics representing the model performance at all 183 sites from SURFRAD, BSRN, and FLUXNET. The results illustrated that the RMSE (RRMSE) values of the individual GCMs with respect to the 183 sites varied between 20 and 26 W $m^{-2}$ (6.4% and 8.3%), with an average RMSE (RRMSE) value of 22 W $m^{-2}$ (7.2%). There was no individual GCM with an RMSE (RRMSE) value within 20 W $m^{-2}$ (6.5%). The biases (Rbias) averaged over 183 sites for the various GCMs varied from −10 to 10 W $m^{-2}$ (−3.1% to 2.9%). Only 24 out of the 47 CMIP6 GCMs underestimated SDLR. Overall, the CMIP6 GCMs did not exhibit a significant tendency to underestimate SDLR at the selected 183 sites, which had a lower positive average bias (Rbias) of 0.18 W $m^{-2}$ (0.06%). The CMIP6 GCMs had R values between 0.92 and 0.95 at all 183 sites, and the average R value amounted to 0.94. No individual GCM had an R value above 0.95. 39 out of the 47 GCMs displayed GPI values over −5, with an average GPI value of −1. Among the 47 CMIP6 GCMs, the lowest simulation deviation was found in MPI-ESM1–2-HR which reported an RMSE of 21.02 W $m^{-2}$, a bias of 0.33 W $m^{-2}$, an R of 0.944 and a maximum GPI value of 5.797 for all 183 sites, followed by TaiESM1, E3SM-1-0 and GFDL-ESM4, which had higher spatial resolutions. At all 183 sites, the largest model error was found in NorCPM1, with an RMSE of 25.67 W $m^{-2}$, a bias of 9.47 W $m^{-2}$, an R of 0.931 and a minimum GPI value of −12.412, which had a lower spatial resolution, followed by GISS-E2–1-H and EC-Earth3-Veg-LR. However, the CMIP6 GCMs with higher spatial resolutions did not always show greater GPI values than those that have lower spatial resolutions, such as EC-Earth3, EC-Earth3-Veg and EC-Earth3-AerChem.

We also used the SMA and BMA methods to estimate SDLR by merging the 47 CMIP6 GCMs. The weights of single GCMs acquired by the BMA method, shown in Fig. 4, ranged from 0.019 to 0.024, and 39 out of the 47 GCMs showed weights between 0.020 and 0.022. EC-Earth3-Veg, with a maximum weight of 0.0231, which was approximately 8% higher than the mean value (0.0213), made a greater contribution to the SDLR ensemble, followed by EC-Earth3-AerChem and EC-Earth3, with weights of 0.0227 and 0.0226, respectively. The weight of NorCPM1 was 0.0198, which was approximately 7% lower than the mean value, followed by FGOALS-g3 with a weight of 0.0200. The RMSE exhibited a negative correlation with the weight, with a correlation coefficient of −0.51. R was positively correlated with weight, with a correlation coefficient of 0.64.

Taylor diagrams (Taylor, 2001), which are particularly useful to evaluate the relative performance of numerous models, were applied to assess the abilities of the 47 CMIP6 GCMs and the MME methods to simulate SDLR with respect to the ground measurements. Fig. 5 illustrates the comparison results between the monthly SDLR estimates and the ground-measured SDLR for the BSRN, SURFRAD, FLUXNET and all sites. According to Fig. 5, the MME methods always showed a better performance than the individual GCMs. The evaluation results of the estimated SDLR by the SMA and BMA methods at all 183 sites are also given in Fig. 3. It was obvious that the SDLR obtained by the SMA and BMA methods exhibited lower RMSE (RRMSE) and bias (Rbias) and higher R values with respect to the individual GCMs at all 183 sites; specifically, the RMSE, bias, RRMSE, Rbias and R values reported by the SMA results were 17.35 W $m^{-2}$, −0.18 W $m^{-2}$, 5.55%, −0.06% and 0.96, respectively, while these values were 17.33 W $m^{-2}$, 0 W $m^{-2}$, 5.54%, 0% and 0.96, respectively, for the BMA results. The statistical metrics derived from the BMA results were close to those of the SMA results. The MME methods reduced the RMSE (RRMSE) by approximately 5 W $m^{-2}$ (1.7%) and increased the R by approximately 0.02 on average compared to the individual GCMs. The uncertainty of the CMIP6 GCMs was reduced by the MME methods by incorporating multiple models.

#### 4.1.2. Comparison with CMIP5

The majority of the SDLR simulations from CMIP5 and CMIP6 GCMs cover the periods 1861–2005 and 1850–2014, respectively, while the SDLR measurements begin in 1992. Therefore, SDLR observations (6097 samples) from 1992 to 2005 at 101 sites were used to compare the performance of the 47 CMIP6 GCMs to simulate SDLR with that of the 46 CMIP5 GCMs. Table 1 illustrates a comparison of statistical metrics between CMIP6 and CMIP5 GCMs at 42 BSRN sites, 7 SURFRAD sites, 52 FLUXNET sites and all 101 sites. For the seven SURFRAD sites, the CMIP5 GCMs showed RMSE (RRMSE) values in the range of 16–27 W $m^{-2}$ (5.4%–8.7%), with an average RMSE (RRMSE) value of 21 W $m^{-2}$ (6.7%) and a median RMSE (RRMSE) value of 21 W $m^{-2}$ (6.7%). The RMSE (RRMSE) values for 20 out of the 46 CMIP5 GCMs were less than 20 W $m^{-2}$ (6.5%). The biases (Rbias) in the individual CMIP5 GCMs ranged from −20 to 5 W $m^{-2}$ (−6.3% to 1.5%). SDLR was underestimated at 42 out of the 46 CMIP5 GCMs with an average bias (Rbias) of −8 W $m^{-2}$ (−2.5%) and a median bias (Rbias) of −6 W $m^{-2}$ (−2.1%). The R values for the individual CMIP5 GCMs varied from 0.89 to 0.95, with an average R value of 0.93 and a median R value of 0.93. There was no individual CMIP5 GCM with an R value greater than 0.95. The RMSE (RRMSE) values for the SDLR simulations from individual CMIP6 GCMs ranged from 16 to 27 W $m^{-2}$ (5.3% to 8.5%). The average RMSE (RRMSE) value was 20 W $m^{-2}$ (6.6%), and the median RMSE (RRMSE) value amounted to 20 W $m^{-2}$ (6.4%), which were slightly lower than those of CMIP5 GCMs. 27 out of the 47 CMIP6 GCMs had RMSE (RRMSE) values less than 20 W $m^{-2}$ (6.5%). The biases (Rbias) for the individual CMIP6 GCMs varied from −18 to 7 W $m^{-2}$ (−5.7% to 2.2%). The SDLR simulations were underestimated by 40 out of the 47 CMIP6 GCMs, with a negative average bias (Rbias) of 6 W $m^{-2}$ (1.9%) and a negative median bias (Rbias) of 5 W $m^{-2}$ (1.6%), which were typically smaller than those of the CMIP5 GCMs. The SDLR simulations from the CMIP6 GCMs showed R values ranging from 0.90 to 0.95, which were somewhat higher than those of the CMIP5 GCMs. The R values exceeded 0.95 at one out of the 47 CMIP6 GCMs, with an average R value of 0.93 and a median R value of 0.93.

For 42 sites in the BSRN, the RMSE (RRMSE) values of the individual CMIP5 GCMs varied between 18 and 29 W $m^{-2}$ (6.0% and 9.4%). Three out of the 46 CMIP5 GCMs showed RMSE (RRMSE) values within 20 W $m^{-2}$ (6.5%), with an average RMSE (RRMSE) value of 22 W $m^{-2}$ (7.1%) and a median RMSE (RRMSE) value of 21 W $m^{-2}$ (6.9%). The biases (Rbias) for the various CMIP5 GCMs varied from −21 to 4 W $m^{-2}$ (−6.5% to 1.3%). The majority of the CMIP5 GCMs (41 GCMs) underestimated SDLR, with an average bias (Rbias) of −6 W $m^{-2}$ (−1.9%) and a median bias (Rbias) of −5 W $m^{-2}$ (−1.6%). The CMIP5 GCMs had R values between 0.95 and 0.97, which were above 0.95 for all 46 CMIP5

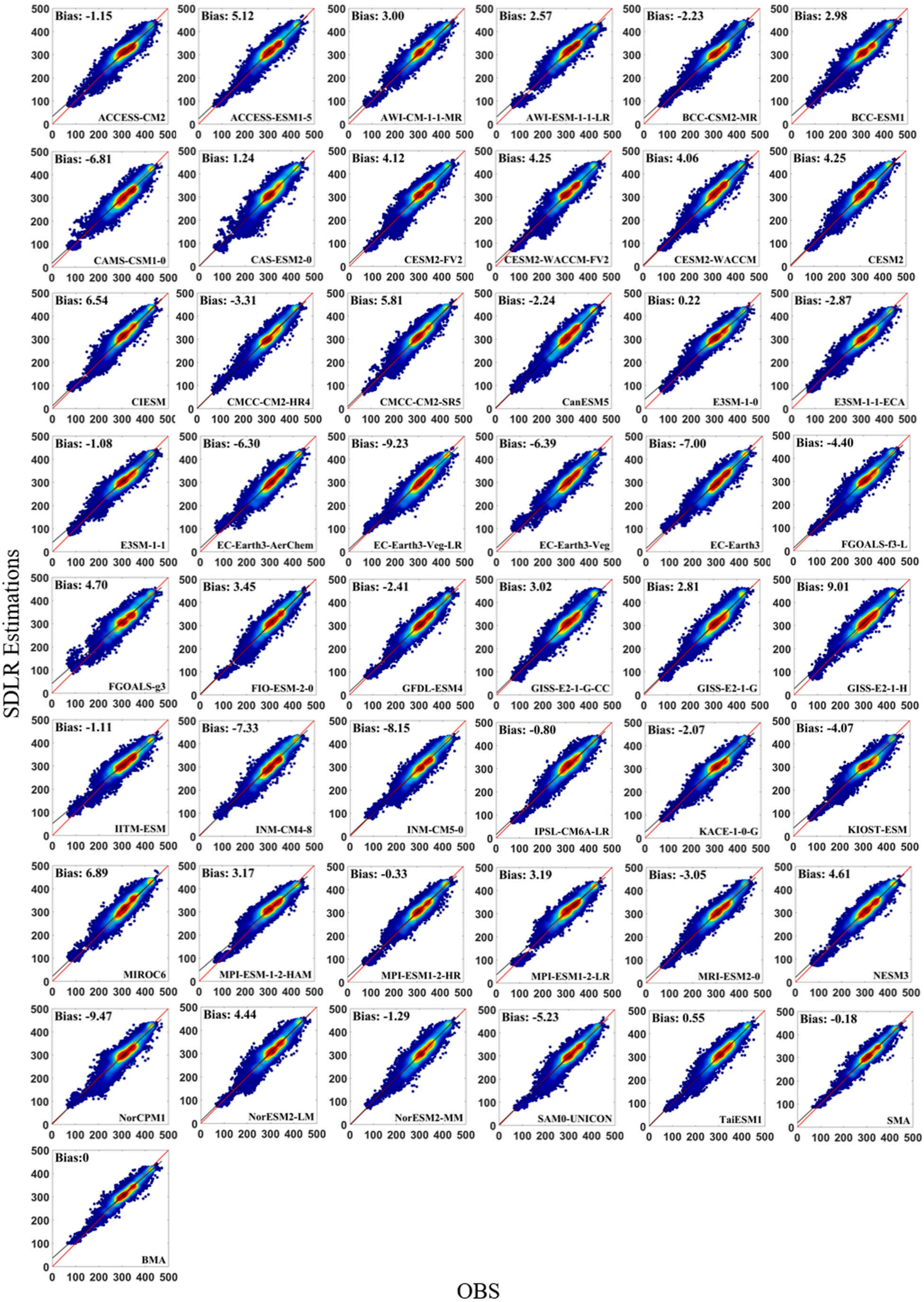


**Fig. 3.** Scatterplots of the monthly SDLR observations from 183 sites and the corresponding SDLR simulations from the 47 CMIP6 GCMs and the multimodel ensemble (MME) methods (in units of W $m^{-2}$).

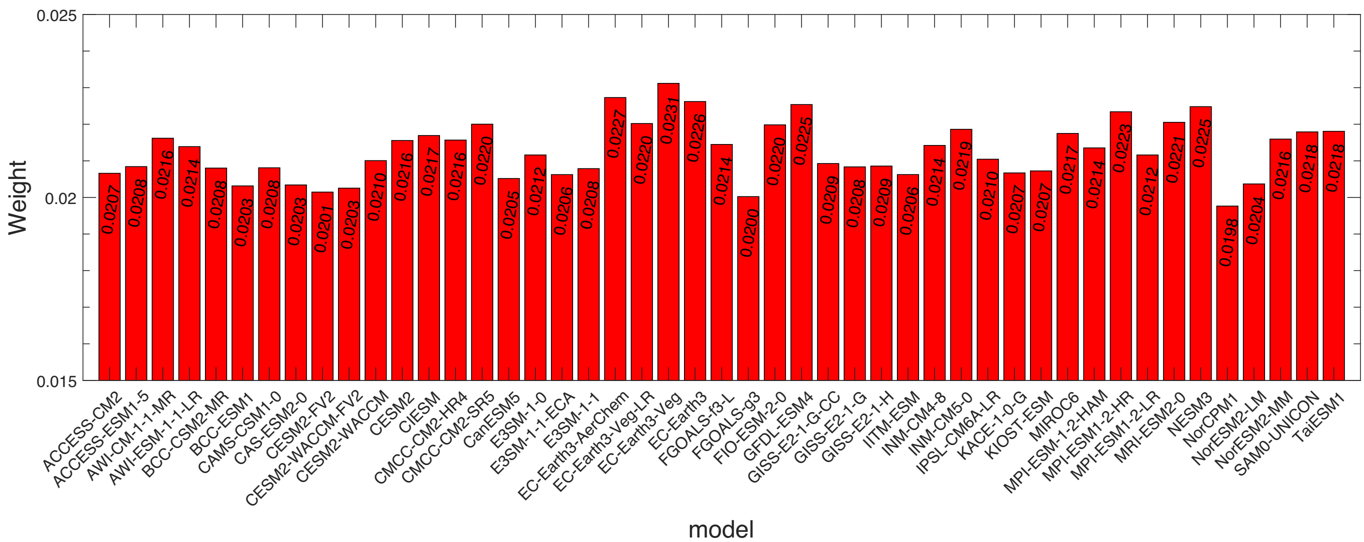


**Fig. 4.** Relative weights of the 47 CMIP6 GCMs acquired by the Bayesian model averaging (BMA) method.

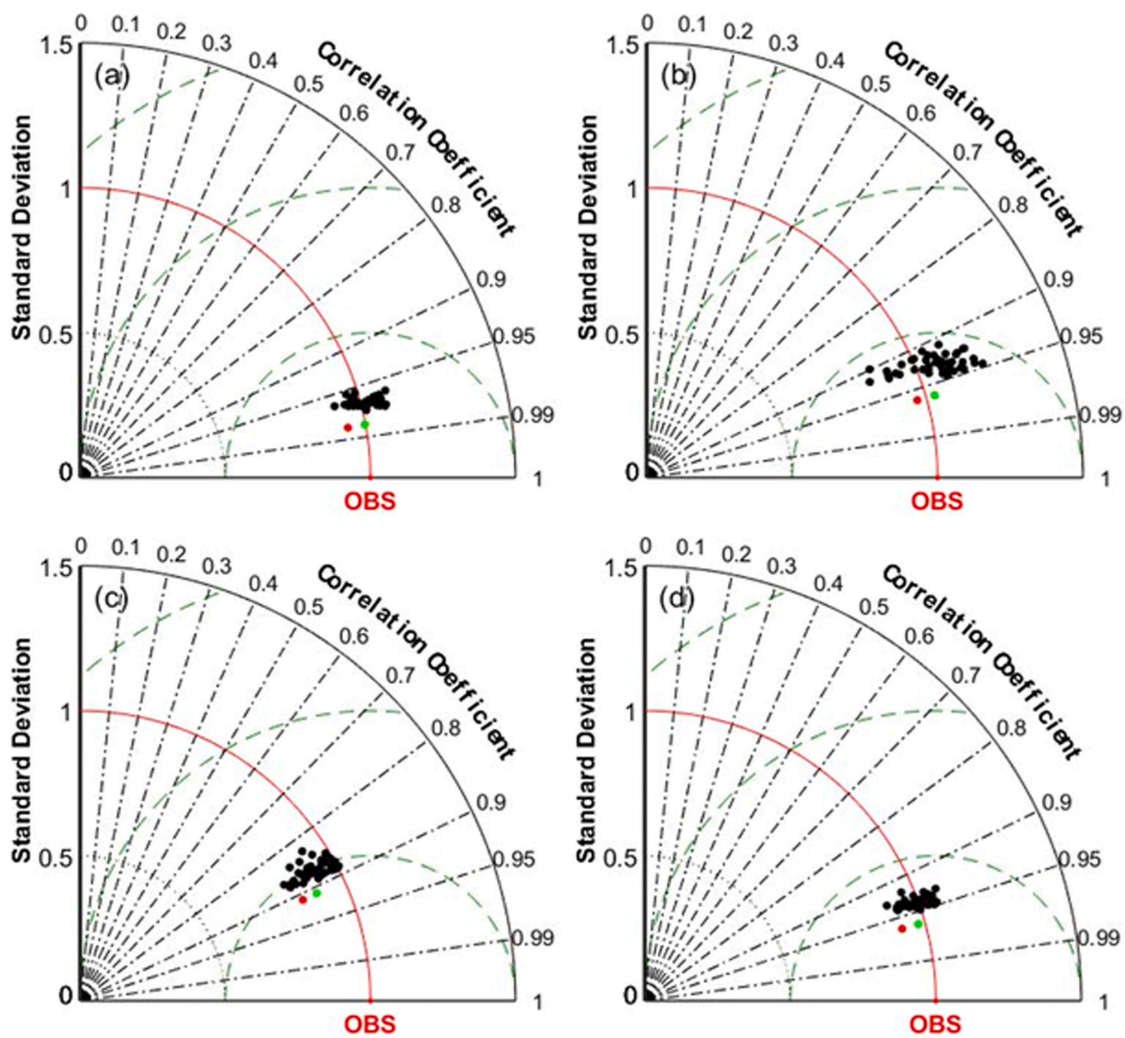


**Fig. 5.** Taylor diagrams of the monthly SDLR measurements and MME estimates for BSRN (a), SURFRAD (b), FLUXNET (c) and all 183 sites (d). The 47 CMIP6 GCMs, the simple model averaging (SMA) method and the BMA method are represented by black, green and red dots, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 1**
Comparison of statistical metrics for monthly SDLR simulations from the CMIP6 and CMIP5 GCMs.

| | | BSRN | | SURFRAD | | FLUXNET | | All Sites | |
|---|---|---|---|---|---|---|---|---|---|
| | | CMIP6 | CMIP5 | CMIP6 | CMIP5 | CMIP6 | CMIP5 | CMIP6 | CMIP5 |
| RMSE | Mean | **21.37** | 21.93 | **20.37** | 20.71 | **24.24** | 24.53 | **22.08** | 22.55 |
| | Median | **21.01** | 21.46 | **19.82** | 20.63 | **23.71** | 24.34 | **21.92** | 22.25 |
| | Min | 19.13 | **18.72** | **16.54** | 16.96 | **21.57** | 21.80 | 20.00 | **19.72** |
| | Max | **26.37** | 28.88 | **26.10** | 26.84 | **28.07** | 30.52 | **26.69** | 28.86 |
| *Bias* | *Mean* | ***−2.94*** | *−5.95* | ***−5.80*** | *−7.74* | *3.03* | ***0.55*** | ***−1.68*** | *−4.43* |
| | *Median* | ***−3.15*** | *−4.98* | ***−4.92*** | *−6.36* | *3.07* | ***2.36*** | ***−1.46*** | *−2.99* |
| | *Min* | ***−14.31*** | *−20.08* | ***−17.56*** | *−19.25* | ***−7.39*** | *−13.94* | ***−12.85*** | *−18.16* |
| | *Max* | *6.84* | ***3.73*** | *6.71* | ***4.58*** | *12.66* | ***9.21*** | *7.79* | ***5.29*** |
| R | Mean | 0.96 | 0.96 | 0.93 | 0.93 | **0.91** | 0.90 | 0.95 | 0.95 |
| | Median | 0.96 | **0.97** | 0.93 | 0.93 | **0.91** | 0.90 | 0.95 | 0.95 |
| | Min | 0.95 | 0.95 | **0.90** | 0.89 | 0.87 | 0.87 | 0.94 | 0.94 |
| | Max | 0.97 | 0.97 | 0.95 | 0.95 | 0.92 | **0.93** | 0.96 | 0.96 |

GCMs. The average R value amounted to 0.96, and the median R value was 0.97. The RMSE (RRMSE) values of the various CMIP6 GCMs ranged from 19 to 27 W m$^{-2}$ (6.1% to 8.6%). Six out of the 47 CMIP6 GCMs had RMSE (RRMSE) values within 20 W m$^{-2}$ (6.5%), with an average RMSE (RRMSE) value of 21 W m$^{-2}$ (6.9%) and a median RMSE (RRMSE) value of 21 W m$^{-2}$ (6.8%), which were lower than those of the CMIP5 GCMs. The biases (Rbias) of the SDLR simulations from the individual CMIP6 GCMs varied between −15 and 7 W m$^{-2}$ (−4.7% and 2.3%), which were negative for 31 out of the 47 CMIP6 GCMs. The SDLR from the CMIP6 GCMs had a lower negative average bias (Rbias) of 3 W m$^{-2}$ (1.0%) and a lower negative median bias (Rbias) of 3 W m$^{-2}$ (1.0%). It was clear that the CMIP6 GCMs have obviously improved the underestimation of SDLR in comparison with the CMIP5 GCMs. The R values for the CMIP6 GCMs were close to those for the CMIP5 GCMs, with an average R value of 0.96 and a median R value of 0.96, which were in the range of 0.95–0.97, and all 47 CMIP6 GCMs had R values greater than 0.95.

For the 52 FLUXNET sites, the RMSE (RRMSE) values for the individual CMIP5 GCMs ranged between 21 and 31 W m$^{-2}$ (7.0% and 9.9%), with an average RMSE (RRMSE) value of 25 W m$^{-2}$ (7.9%) and a median RMSE (RRMSE) value of 24 W m$^{-2}$ (7.8%). There was no individual GCM with an RMSE (RRMSE) value within 20 W m$^{-2}$ (6.5%). The CMIP5 GCMs showed biases (Rbias) ranging from −14 to 10 W m$^{-2}$ (−4.5% to 3.0%), and 28 out of the 46 CMIP5 GCMs overestimated SDLR. The average bias (Rbias) was 1 W m$^{-2}$ (0.2%), and the median bias (Rbias) amounted to 2 W m$^{-2}$ (0.8%). The R values of the various CMIP5 GCMs varied between 0.87 and 0.93, with an average R value of 0.90 and a median R value of 0.90. No individual CMIP5 GCM had an R value greater than 0.95. Compared with the CMIP5 GCMs, the CMIP6 GCMs exhibited lower RMSE (RRMSE) values, which ranged from 21 to 29 W m$^{-2}$ (6.9% to 9.1%), with an average RMSE (RRMSE) value of 24 W m$^{-2}$ (7.8%) and a median RMSE (RRMSE) value of 24 W m$^{-2}$ (7.6%). No individual CMIP6 GCM reported an RMSE (RRMSE) value within 20 W m$^{-2}$ (6.5%). The SDLR simulations from the various CMIP6 GCMs produced biases (Rbias) in the range of −8-13 W m$^{-2}$ (−2.4%–4.1%), which were positive at most CMIP6 GCMs (33 GCMs), with an average bias (Rbias) of 3 W m$^{-2}$ (1.0%) and a median bias (Rbias) of 3 W m$^{-2}$ (1.0%). The R values of the individual CMIP6 GCMs varied from 0.87 to 0.92. There was no individual CMIP6 GCM with an R value above 0.95. The SDLR simulations from the CMIP6 GCMs correlated well with the SDLR observations with a higher average R value of 0.91 and a higher median R value of 0.91 than the CMIP5 GCMs.

A comparison of statistical metrics between CMIP6 and CMIP5 GCMs at all 101 sites is also shown in Table 1. The CMIP5 GCMs had an average RMSE (RRMSE) of 23 W m$^{-2}$ (7.3%), a negative average bias (Rbias) of 4 W m$^{-2}$ (1.4%), and an average R of 0.95. The SDLR simulations from the CMIP6 GCMs agreed well with the SDLR observations with an average RMSE (RRMSE) of 22 W m$^{-2}$ (7.1%) and a negative average bias (Rbias) of 2 W m$^{-2}$ (0.5%), which were lower than those of the CMIP5 GCMs. Similar to the CMIP5 GCMs, the average R for the CMIP6 GCMs amounted to 0.95. The histograms of the statistical metrics for the 47 CMIP6 GCMs and the 46 CMIP5 GCMs at all 101 sites are shown in Fig. 6. The RMSE values for the SDLR simulations from the individual CMIP5 GCMs ranged from 19 to 29 W m$^{-2}$. 25 out of the 46 CMIP5 GCMs had RMSE values in the range of 21–23 W m$^{-2}$. The RMSE values were less than 21 W m$^{-2}$ at six CMIP5 GCMs. 15 CMIP5 GCMs showed RMSE values greater than 23 W m$^{-2}$. The CMIP6 GCMs had RMSE values between 19 and 27 W m$^{-2}$. The RMSE values for 31 out of the 47 CMIP6 GCMs varied between 21 and 23 W m$^{-2}$. Eight CMIP6 GCMs showed RMSE values within 21 W m$^{-2}$. The RMSE values were above 23 W m$^{-2}$ at only eight CMIP6 GCMs, and no individual CMIP6 GCM had an RMSE value greater than 27 W m$^{-2}$. The biases for the individual CMIP5 GCMs varied from −20 to 10 W m$^{-2}$ and were negative for 38 out of the 46 CMIP5 GCMs. 28 CMIP5 GCMs reported absolute biases within 5 W m$^{-2}$. The biases of 17 CMIP5 GCMs were less than −5 W m$^{-2}$. Only one CMIP5 GCM had a bias above 5 W m$^{-2}$. The biases of the individual CMIP6 GCMs ranged from −15 to 10 W m$^{-2}$, and the SDLR simulations were underestimated by only 27 out of the 47 CMIP6 GCMs. The absolute biases of 32 CMIP6 GCMs were within 5 W m$^{-2}$. 12 CMIP6 GCMs showed biases less than −5 W m$^{-2}$, and there was no individual CMIP6 GCM with a bias less than −15 W m$^{-2}$. The biases of the three CMIP6 GCMs were above 5 W m$^{-2}$. Overall, the CMIP6 GCMs exhibited better performance in simulating SDLR than the CMIP5 GCMs.

### 4.2. *Spatial distribution*

The monthly mean SDLR records derived from ERA5 and CERES EBAF begin 1979 and 2000, respectively, while the majority of the SDLR simulations from the CMIP5 GCMs end in 2005. Thus, the BMA method, which showed better performance in estimating SDLR, was applied to produce a gridded global SDLR dataset (1° × 1°) during 2000–2005 using the CMIP6 and CMIP5 GCMs. Fig. 7 (a) indicate the spatial variation in the SDLR estimations of the CMIP6 GCMs using the BMA method in 2000–2005 worldwide. The variation pattern of SDLR shows a higher correlation with the variation in surface air temperature (Bodas-Salcedo et al., 2008). Thus, lower SDLR values were found in Antarctica, amounting to approximately 80–210 W m$^{-2}$ in the local winter season (JJA) and 120–280 W m$^{-2}$ in the local summer season (DJF). Lower SDLR values also occurred in the Arctic, which were equal to approximately 140–190 W m$^{-2}$ in the local winter season (DJF) and 220–320 W m$^{-2}$ in the local summer season (JJA). Tropical regions have always been areas with greater SDLR. In addition to the ocean surface regions, some land surface areas showed greater SDLR values, such as Indonesia, Malaysia, the Philippines, the Congo Basin and the Amazon rainforest, which was consistent with other studies (Bodas-Salcedo et al., 2008; Wang and Dickinson, 2013; Wild et al., 2015). Overall, there was a gradual SDLR decrease from low latitude areas towards polar areas.

The spatial variation of the different values in the global annual mean SDLR values obtained by the BMA method between the CMIP6 and

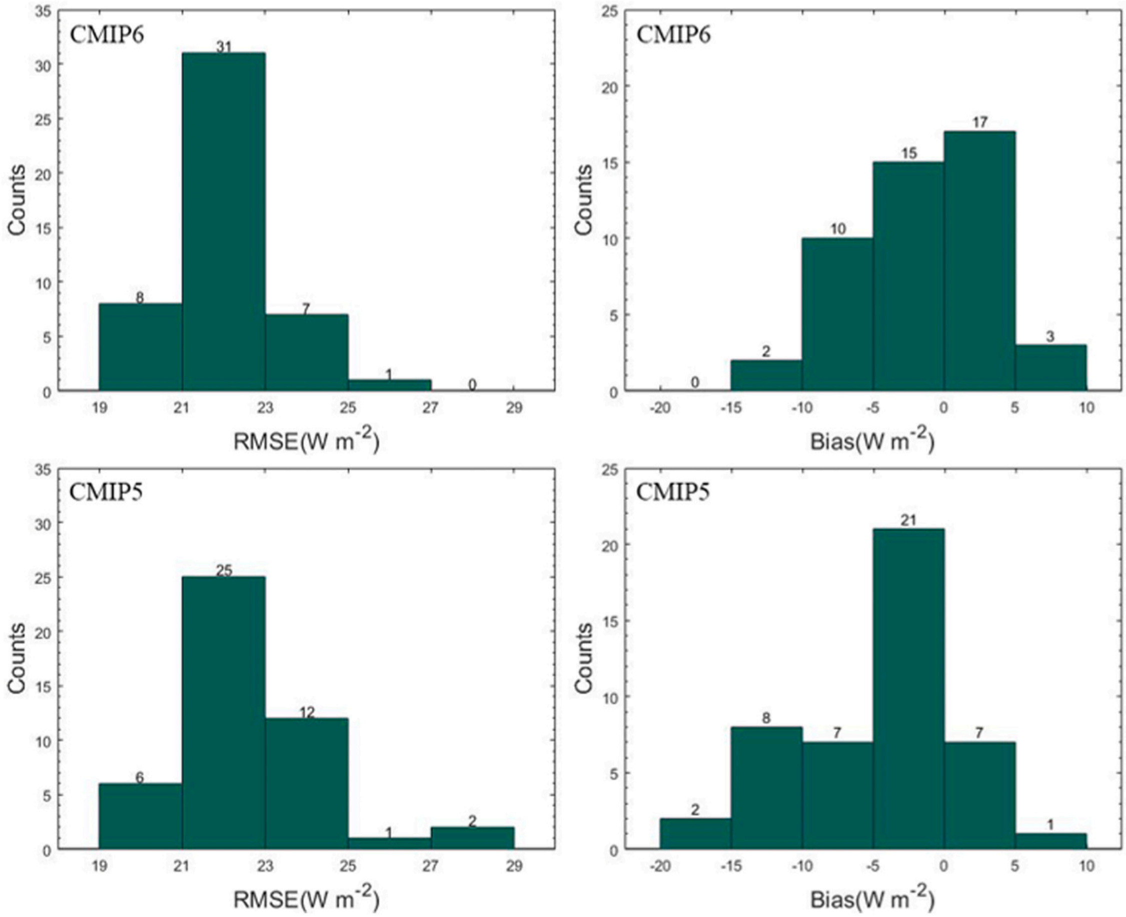


**Fig. 6.** RMSE and Bias histograms for monthly SDLR simulations from the 47 CMIP6 GCMs and the 46 CMIP5 GCMs.

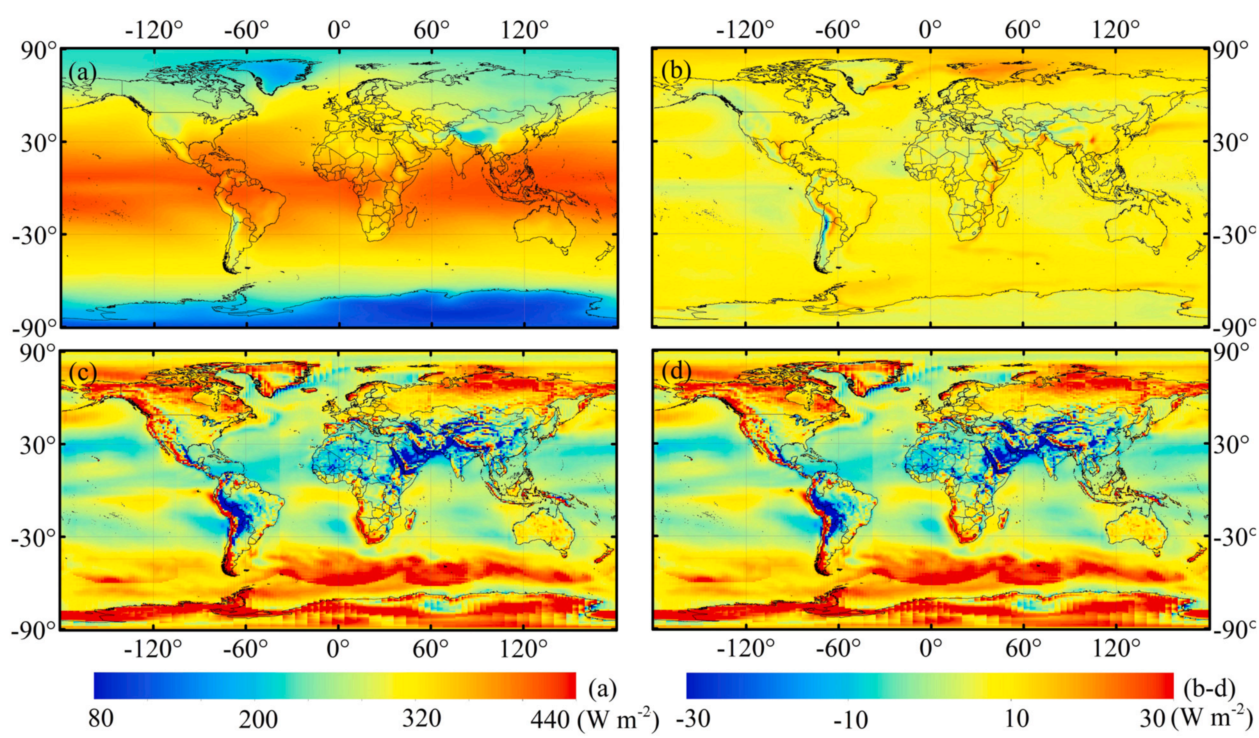


**Fig. 7.** Spatial variation of the CMIP6 GCM SDLR estimations using the BMA method (a), and the biases between CMIP6 and CMIP5 GCMs (b), between CMIP6 GCMs and ERA5 (c), and between CMIP6 GCMs and CERES EBAF (d) from 2000 to 2005 (in units of W $m^{-2}$).

CMIP5 GCMs, which ranged from −22 to 11 W m$^{-2}$, are shown in Fig. 7 (b). The annual mean SDLR values estimated by the CMIP6 GCMs using the BMA method were significantly higher in the southern Himalayas, eastern Andes, north Asia and the oceans in the high latitude regions than those from the CMIP5 GCMs, which had a maximum positive bias of 11.11 W m$^{-2}$. The BMA estimations from the CMIP6 GCMs had lower annual mean SDLR values in the western Andes, with a maximum negative bias of −22.04 W m$^{-2}$. Overall, the BMA estimations from the CMIP6 GCMs showed higher global annual mean SDLR compared to CMIP5 GCMs, which indicated that the CMIP6 GCMs have improved the underestimation problem in SDLR in comparison with CMIP5 GCMs. The spatial differences between the BMA estimations of CMIP6 GCMs and ERA5 are exhibited in Fig. 7 (c) and varied between −10 and 20 W m$^{-2}$ in most areas. The BMA estimations of CMIP6 GCMs showed negative biases in the southern Himalayas and the eastern Andes, and the largest negative bias was −36.67 W m$^{-2}$. Positive biases were found over much of the Earth's surface, especially in the oceans near Antarctica in the high latitude regions, northern Himalayas, Kunlun Mountains, Altun Mountains, Qilian Mountains, Cascade Mountains, Sierra Nevada and western Andes, and the maximum positive bias amounted to 59.83 W m$^{-2}$. Fig. 7 (d) illustrates the spatial dissimilarities between the BMA estimations from the CMIP6 GCMs and the CERES EBAF, which varied from −20 to 10 W m$^{-2}$ in most areas. Negative biases were found in the southern Himalayas, Kunlun Mountains, Altun Mountains, Qilian Mountains, eastern Andes and other high altitude areas, where the maximum negative bias was approximately −89.70 W m$^{-2}$. The BMA estimations from the CMIP6 GCMs had positive biases in the oceans near Antarctica in the high latitude regions, northern Himalayas, Cascade Mountains, Sierra Nevada and western Andes, with a maximum positive bias of approximately 65.92 W m$^{-2}$.

### 4.3. Annual mean and long-term trend

There are 37 sites (including 27 BSRN sites, 6 SURFRAD sites and 4 FLUXNET sites) with long-term SDLR measurements from 1995 to 2014, while only 10 sites record ground-measured SDLR from 1992 to 1994. Therefore, to analyze the temporal evolutions of both ground-measured SDLR and SDLR simulations from GCMs, we calculated the annual mean SDLR of 37 sites and the corresponding MME results of the CMIP6 GCMs from 1995 to 2014, as shown in Fig. 8. The mean SDLR of 37 sites increased at a rate of 2.3 W m$^{-2}$ per decade ($P > 0.05$) during 1995–2014. It was also found that SDLR showed an increasing trend from 1995 to 2014 under the BMA (3.9 W m$^{-2}$ per decade, $P < 0.05$) and SMA results of the CMIP6 GCMs (4.1 W m$^{-2}$ per decade, P < 0.05). Both the ground-measured SDLR and the MME SDLR estimates showed a similar increasing trend in SDLR from 1995 to 2014 but with different magnitudes. In general, the annual mean SDLR trend based on the BMA method were closer to those from the ground measurements compared to the SMA results.

Therefore, the BMA method was used to esitimate the global annual mean SDLR of the CMIP6 GCMs, as shown in Table 2. The SDLR values obtained by the BMA method using the CMIP6 GCMs were similar to the estimates found in other studies. For instance, Ma et al. (2014) estimated a global annual mean SDLR of 341 W m$^{-2}$ based on 44 CMIP5 GCMs from 1992 to 2005, supporting a value consistent with the estimation (340.5 W m$^{-2}$) derived by the BMA method from the 47 CMIP6 GCMs used in the present research. The estimate in Wild et al. (2013) of 342 W m$^{-2}$ acquired by 22 CMIP5 GCMs in 2000–2004 were in line with the value (341.6 W m$^{-2}$) estimated by the BMA method using the 47 CMIP6 GCMs here. L'Ecuyer et al. (2015) determined the global annual mean SDLR from satellite observations at 341 W m$^{-2}$ during 2000–2009, which is close to the estimate (341.9 W m$^{-2}$) made from the 47 CMIP6 GCMs using the BMA method in this study. The estimate reported by Wang et al. (2013) was 342 W m$^{-2}$ in 2003–2010, which was in agreement with that (342.1 W m$^{-2}$) calculated by the BMA method with the 47 CMIP6 GCMs. The global annual mean SDLR found by Wild (2020) of 344 W m$^{-2}$ from 38 CMIP6 GCMs during 2000–2014, was close to that (342.2 W m$^{-2}$) obtained by the BMA method using the 47 CMIP6 GCMs here. Overall, the BMA method performed better in the estimation of global annual mean SDLR using the CMIP6 GCMs. For the global annual mean SDLR, the best estimation of 342 W m$^{-2}$ in 2000–2014 obtained by the BMA method using the CMIP6 GCMs was applied in this study.

## 5. Discussion

### 5.1. Uncertainties of evaluation

Validating and merging the SDLR simulations from the CMIP6 GCMs with only ground measurements from the BSRN, SURFRAD and FLUXNET cause large uncertainties, which are discussed as follows. First, the uncertainty of SDLR measurements impacts the evaluation and fusion of SDLR simulations from the CMIP6 GCMs. Although SDLR measurements are relatively accurate, there are still missing values and measured errors, within approximately 10 W m$^{-2}$ (Augustine et al., 2000; Liang et al., 2010; Ma et al., 2014). Linear interpolation was applied to supply

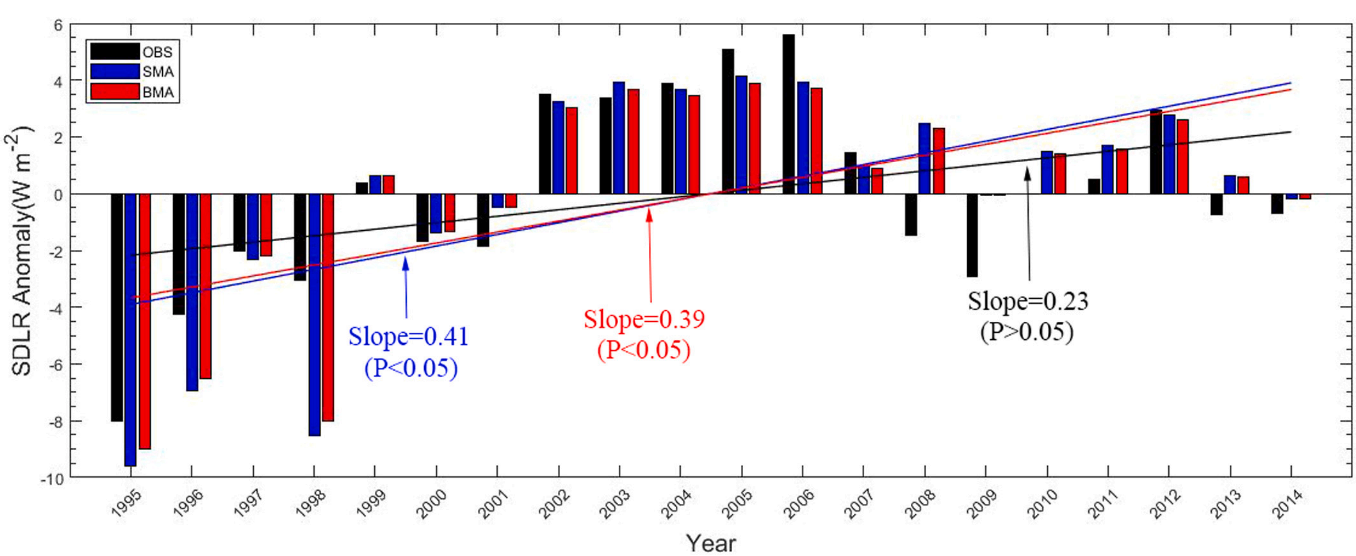


**Fig. 8.** The annual variations in anomalous mean SDLR of the ground measurements and the MME results of the CMIP6 GCMs.

**Table 2**
The averaged global annual mean SDLR of the CMIP6 GCMs, CMIP5 GCMs, ERA5, CERES EBAF and reference estimates (in units of W m$^{-2}$).

| Period | CMIP6 GCMs | | CMIP5 GCMs | | ERA5 | CERES EBAF | Reference Estimates |
|---|---|---|---|---|---|---|---|
| | SMA | BMA | SMA | BMA | | | |
| 1992–2005 | 342.0 | 340.5 | 339.0 | 341.1 | 338.6 | – | 341 (Ma et al., 2014) |
| 2000–2004 | 343.2 | 341.6 | 340.0 | 342.1 | 339.0 | 345.5 | 342 (Wild et al., 2013) |
| 2000–2009 | 343.5 | 341.9 | – | – | 339.3 | 345.2 | 341 (L'Ecuyer et al., 2015) |
| 2003–2010 | 343.8 | 342.1 | – | – | 339.6 | 345.2 | 342 (Wang and Dickinson, 2013) |
| 2000–2014 | 343.9 | 342.2 | – | – | 339.5 | 345.2 | 344 (Wild, 2020) |

the missing data in the present study, but these interpolations still increased the uncertainty of the SDLR measurements. In addition, different evaluation datasets might show a systematic bias and result in different validation results. This results in substantial degree of uncertainty in the evaluation and fusion of the CMIP6 GCMs.

Second, the observation site footprints are not consistent with the CMIP6 GCM gridded footprints at the spatial scale. Generally, the footprints of the observation sites cover an area of several hundred meters, whereas the CMIP6 GCMs have a spatial resolution exceeding 100 km. Therefore, the CMIP6 GCM gridded SDLR cannot be represented by the SDLR values of the observation stations. Nevertheless, the SDLR measurements were applied to denote the "true" value to validate and merge CMIP6 GCMs in the present research. This inaccurate denotation will cause large discrepancies between the SDLR measurements and the SDLR simulations from the CMIP6 GCMs.

Third, it is noteworthy that the deviations of the SDLR simulations from the CMIP6 GCMs as well as biases produced by the resampling procedure also result in the uncertainty of SDLR estimation using the BMA method. For example, the deviations of SDLR simulations from the CMIP6 GCMs lead to the input errors of the SDLR estimated by the BMA method. Moreover, in terms of various CMIP6 GCMs with different spatial resolutions, the biases of the merged SDLR that propagated through the resampling procedure are highlighted. It was found that the RMSE values for multiple gridded SDLR datasets were variable at various scales. Although it shows small impacts by improving the spatial resolution of the GCMs, it still impacts the SDLR fusion (Ma et al., 2014; Yao et al., 2016). This resampling procedure could increase the uncertainty of the SDLR obtained by the BMA method.

### 5.2. Altitudinal and latitudinal dependency of SDLR estimation

The altitude of the selected 183 sites during 1992–2014 varies from −9 m to 4319 m, which allowed us to validate the performance of the CMIP6 GCMs in simulating SDLR at different altitudes. Significant differences between the CMIP6 GCMs and ERA5, CERES EBAF, and CMIP5 GCMs were found in high altitude areas, such as the Himalayas, Kunlun Mountains, Altun Mountains, Qilian Mountains and Andes. Therefore, the SDLR ground measurements were grouped into 154 sites with altitudes below 1000 m and 29 sites with altitudes above 1000 m to evaluate the altitudinal dependency of SDLR estimations. The evaluation results indicated that the RMSE values of the SDLR estimations for all 47 CMIP6 GCMs increased with increasing elevation. 29 out of the 47 CMIP6 GCMs had higher absolute biases at sites with altitudes above 1000 m than at sites with altitudes below 1000 m. Among the 47 CMIP6 GCMs, FGOALS-g3 demonstrated the poorest ability to simulate SDLR at sites with altitudes above 1000 m, which had a maximum RMSE of 34.52 W m$^{-2}$ and an absolute bias of 18.06 W m$^{-2}$. It was also found that the RMSE and bias values for the MME methods increased with increasing elevation. To further validate the performance of SDLR simulations from the CMIP6 GCMs at different altitudes, the SDLR observations were also divided into 500 m altitudinal zones. Fig. 9 shows the evaluation results of the SDLR simulations from the CMIP6 GCMs and the MME results for different altitudinal zones. It was obvious that the RMSE values at the sites with altitudes above 1500 m were higher than those at the sites with altitudes below 1500 m. 33 out of the 47 CMIP6 GCMs tended to show a significant altitudinal dependency of the RMSE, which had positive slope values ($p < 0.05$). The SDLR at sites with altitudes above 2000 m were in worse agreement with higher biases, especially in FGOALS-g3, BCC-ESM1, IITM-ESM and MIROC6. Overall, the CMIP6 GCMs performed worse in simulating SDLR in high altitude areas.

Since water vapor is a key greenhouse gas that influences SDLR (Li et al., 2020; Shi and Liang, 2013; Vaquero-Martínez et al., 2020), we tried to investigate whether the altitudinal dependency of the SDLR estimates are caused by total column water vapor (TCWV). However, the TCWV are not provided by the 183 sites selected in this study. Therefore, the TCWV simulations from the CMIP6 GCMs were assessed with the TCWV observations at 169 sites collected from Suominet (Ware et al., 2000) in different altitudinal zones around the world. The elevation of the selected Suominet sites ranges from 3 m to 3657 m. The validation

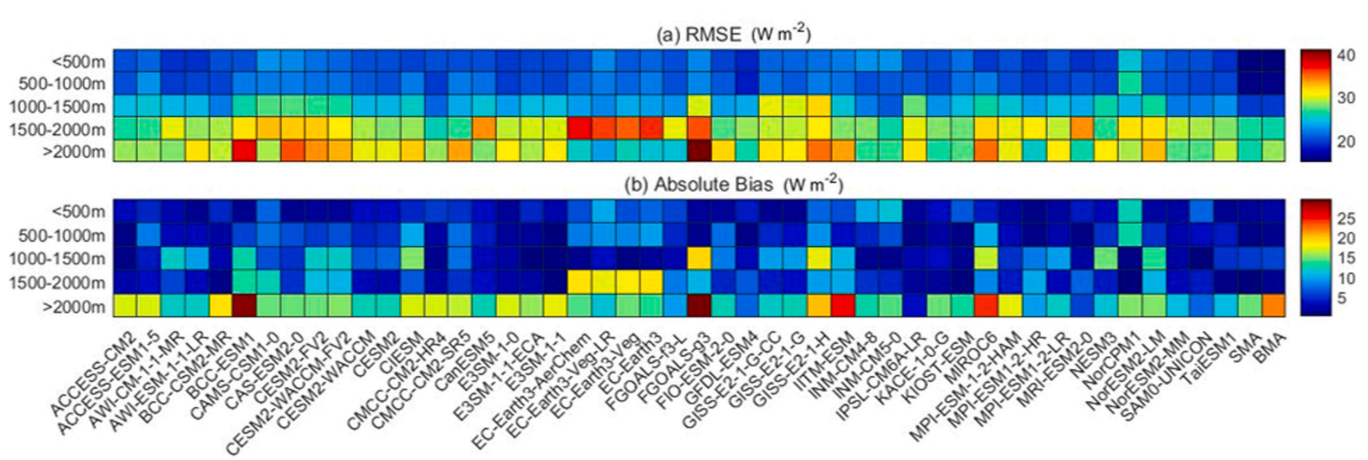


**Fig. 9.** Comparison of the RMSE (a) and absolute Bias (b) values at the sites in different altitudinal zones for the monthly SDLR simulations from the 47 CMIP6 GCMs and the MME results.

results of the TCWV simulations from the CMIP6 GCMs are showed in Fig. 10. Similar to the SDLR simulations, the TCWV simulations from CMIP6 GCMs showed higher RMSE and absolute biases at sites with the altitudes above 1500 m than that with the altitudes below 1500 m. Therefore, the large uncertainties existed in CMIP6 GCM SDLR simulations at high altitudinal sites may be partially related to the lower TCWV precision at those sites.

The ability of the CMIP6 GCMs to simulate SDLR at different latitudes was also evaluated in this study. The ground-measured SDLR were grouped into 30° latitudinal bands. The comparison results of the SDLR simulations from the CMIP6 GCMs and the MME results for different latitudinal bands are shown in Fig. 11. Overall, the majority of the CMIP6 GCMs showed lower RMSE values at the sites located in the relatively lower latitudinal bands. The RMSE values at the sites located in the higher latitudinal bands were higher than the lower latitudinal bands for most CMIP6 GCMs models, especially in FGOALS-g3, NorCPM1 and CESM2-FV2. In general, the CMIP6 GCMs showed poor ability to simulate SDLR in the high latitudinal bands. Most of the CMIP6 GCM SDLR simulations exhibited a latitudinal dependency.

### 5.3. Relations between SDLR, SDSR, SAT, RH, and WVP

Both the ground-measured SDLR and the MME SDLR estimates showed an increasing trend in SDLR during the period of 1995–2014 (Fig. 8). To fully explore the causes of the variables influencing SDLR, the correlation coefficient between the trends in SDLR and the corresponding trends in downward shortwave radiation (SDSR), air temperature (SAT), relative humidity (RH) and water vapor pressure (WVP) in GCM simulations were investigated (Fig. 12). An obvious positive correlation was observed between the trends in SDLR and SAT, with an R value of 0.64. As it is known to all, SAT is an important parameter in estimating SDLR (Swinbank, 1963). The trend in RH was negatively correlated with the trend in SAT, with an R value of −0.19. WVP is also considered a main contributor to SDLR because water vapor is a key greenhouse gas that influences SDLR (Li et al., 2020; Shi and Liang, 2013). Figure 12 shows that the increases in SAT and atmospheric WVP were the most important factors controlling the long-term variation of SDLR. The trend in SDSR was negatively correlated with the trend in SDLR. This might be explained by the fact that the existence of clouds impedes SDSR from reaching the Earth's surface, whereas clouds are vital radiators of SDLR (Shi and Liang, 2013). The relationship between SDLR, SDSR, SAT, RH, and WVP was analyzed with only the GCM simulations, which may result in errors, it is necessary to use ground-based measurements for a further analysis in the future.

## 6. Conclusions

This study validated SDLR simulations from 47 CMIP6 GCMs in comparison to globally distributed ground measurements and then investigated how well the SMA and BMA methods performed in the estimation of SDLR using CMIP6 GCMs. The evaluation data sets consist of 183 sites (17,625 samples) from BSRN, SURFRAD and FLUXNET over a period covering 1992–2014. A comparison between the SDLR simulations from the CMIP6 GCMs versus the ground measurements indicated large differences in the ability of the individual CMIP6 GCMs to simulate SDLR. Individual GCMs with higher spatial resolutions did not always perform better than those with lower spatial resolutions. The RMSE (RRMSE), bias (Rbias), R and GPI values from the SDLR simulations from the individual CMIP6 GCMs averaged over 183 sites varied from 20 to 26 W m$^{-2}$ (6.4% to 8.3%), −10 to 10 W m$^{-2}$ (−3.1% to 2.9%), 0.92 to 0.95 and − 13 to 6, respectively. The CMIP6 GCMs did not show a significant tendency to underestimate SDLR. Only 24 out of the 47 CMIP6 GCMs underestimated SDLR and showed negative biases for all 183 sites. Among the CMIP6 GCMs, NorCPM1 had the worst performance with the lowest GPI value and weight at 183 sites, and its SDLR simulations exhibited a poor predictive ability for both the BSRN and SURFRAD sites. It was also found that the CMIP6 GCMs performed worse in simulating SDLR in high altitude and high latitude areas. The BMA method always had a better performance in estimating SDLR than that of the individual CMIP6 GCMs, with an RMSE (RRMSE) of 17.33 W m$^{-2}$ (5.54%), a bias (Rbias) of 0 W m$^{-2}$ (0%), and an R of 0.96 with respect to the 183 sites. It reduced the RMSE (RRMSE) by approximately 5 W m$^{-2}$ (1.7%) and increased the R by approximately 0.02 on average compared to the individual GCMs. The ability of the CMIP6 GCMs to simulate SDLR was compared with that of the CMIP5 GCMs with respect to the SDLR observations collected at 101 sites (6097 samples) from 1992 to 2005. The results demonstrated that the CMIP6 GCMs exhibited a better ability to simulate SDLR with an average RMSE (RRMSE) of 22 W m$^{-2}$ (7.1%) and a negative average bias (Rbias) of 2 W m$^{-2}$ (0.5%) for the 101 sites, which were lower than those of the CMIP5 GCMs. This indicated that the CMIP6 GCMs obviously improved the underestimation of SDLR in comparison with the CMIP5 GCMs.

We applied the BMA method to produce a gridded global SDLR dataset (1° × 1°) during 2000–2005 using the CMIP6 GCMs. The spatial distributions in SDLR values worldwide were discussed using the SDLR dataset produced in this research. Overall, there was a gradual decrease in SDLR from low latitude areas to the polar areas. Significant differences between the CMIP6 GCMs and ERA5, CERES EBAF, and CMIP5 GCMs were found in high altitude areas, such as the Himalayas, Kunlun

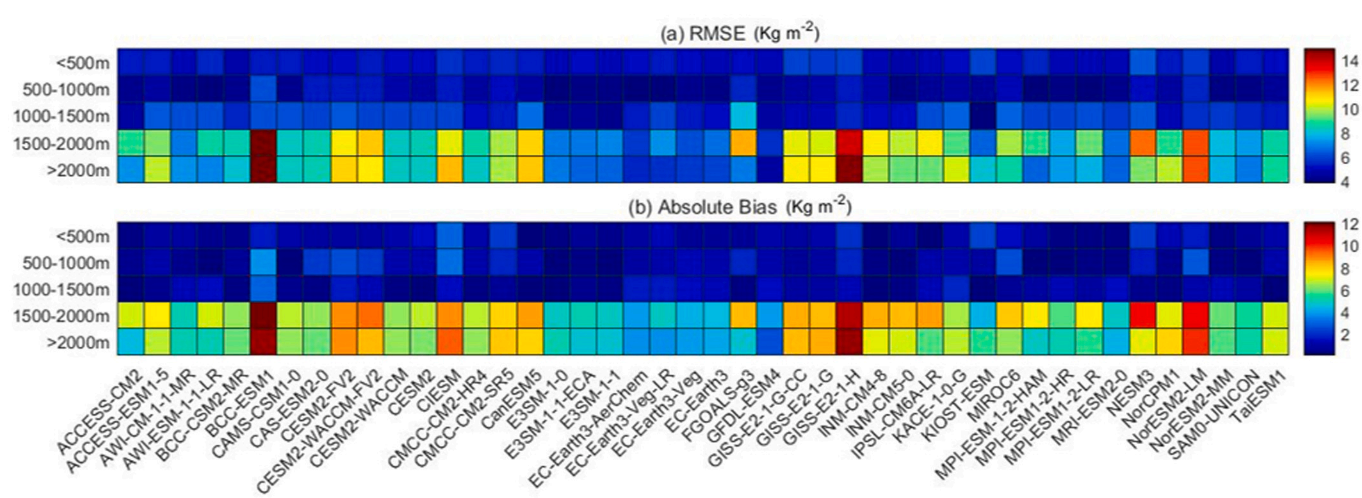


**Fig. 10.** Comparison of the RMSE (a) and absolute Bias (b) values at the sites in different altitudinal zones for the monthly total column water vapor (TCWV) simulations from the 44 CMIP6 GCMs.

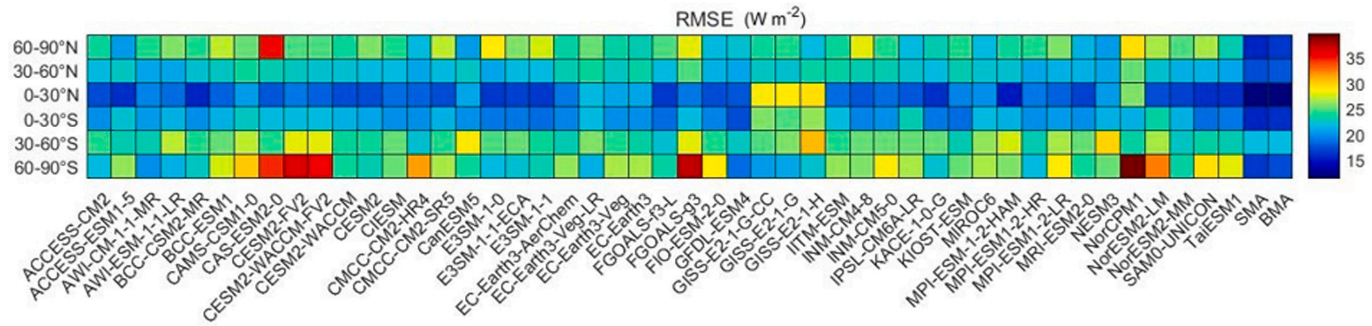


**Fig. 11.** Comparison of the RMSE values at the sites in different latitudinal bands for the monthly SDLR simulations from the 47 CMIP6 GCMs and the MME results.

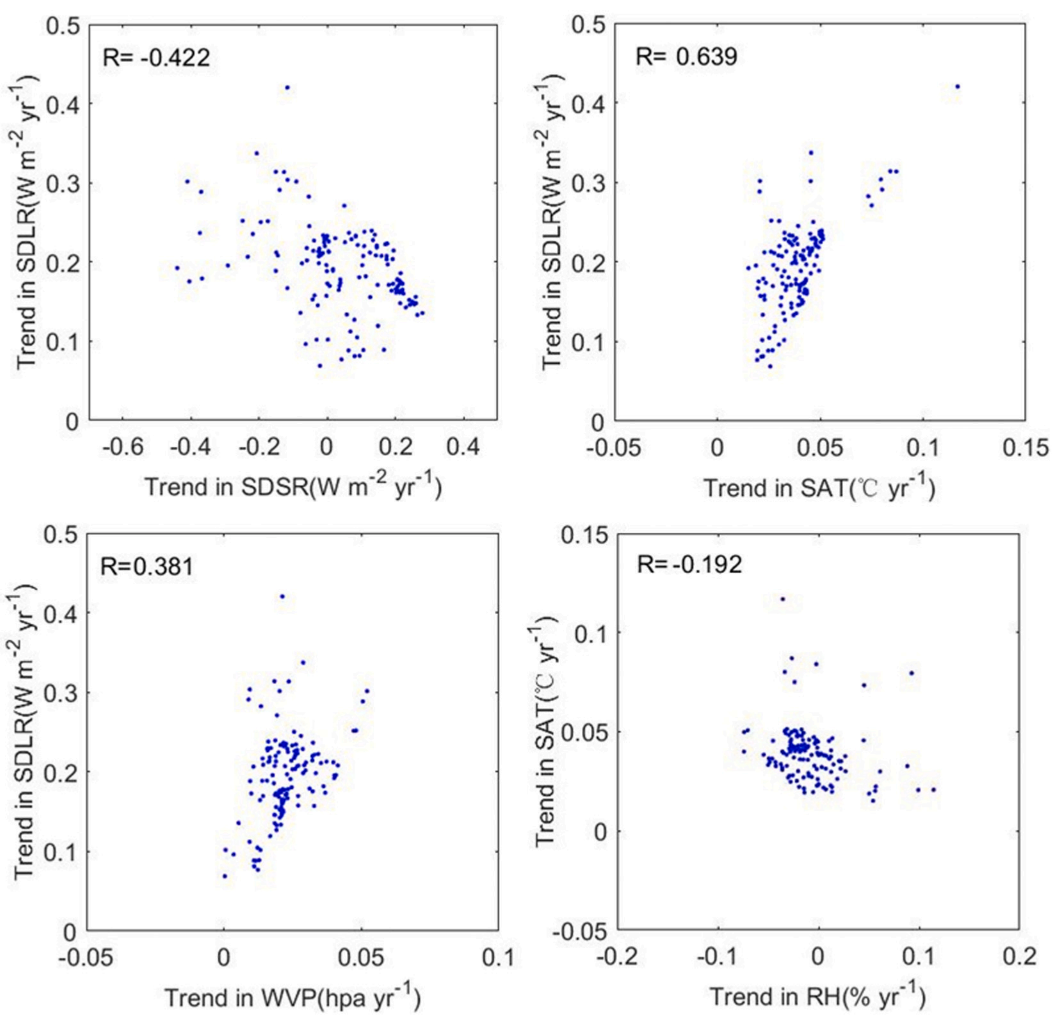


**Fig. 12.** Scatterplots of the trends in SDLR (in units of W m$^{-2}$ per year) and the corresponding trends in downward shortwave radiation (SDSR, in units of W m$^{-2}$ per year), air temperature (SAT, in units of K per year), relative humidity (RH, in units of % per year), and water vapor pressure (WVP, in units of hPa per year) in GCM simulations.

Mountains, Altun Mountains, Qilian Mountains and Andes. The temporal evolutions of both ground-measured SDLR and SDLR simulations from GCMs were also analyzed in this study. In general, the annual mean SDLR based on the BMA method were closer to those from the ground measurements compared to the SMA results. In terms of the global annual mean SDLR, the SDLR estimated by the BMA method based on the 47 CMIP6 GCMs agreed with the results of other studies. The best estimation at 342 W m$^{-2}$ in 2000–2014 obtained by the BMA method using the CMIP6 GCMs was applied in this study. SDLR showed an increasing trend during the period of 1995–2014.This increase may be caused by changes in SAT and WVP related to global warming.

## Data availability

The SDLR simulations from CMIP5 and CMIP6 GCMs were available at https://esgf-node.llnl.gov/search/cmip5/ and https://esgf-node.llnl.gov/search/cmip6/. The ground-measured data of surface downward longwave radiation was downloaded from BSRN (https://bsrn.awi.de/), SURFRAD (https://www.esrl.noaa.gov/gmd/grad/surfrad/), and FLUXNET (https://fluxnet.fluxdata.org/data/fluxnet2015-dataset/). The TCWV observations from the Suominet were available at https://www.unidata.ucar.edu/data/suominet/. The ERA5 data utilized in the present research was obtained from https://cds.climate.copernicus.eu/. The CERES EBAF data was downloaded from the NASA Langley Research Center (https://ceres.larc.nasa.gov/).

## CRediT authorship contribution statement

**Jiawen Xu:** Methodology, Validation, Data curation, Writing – original draft, Writing – review & editing. **Xiaotong Zhang:** Conceptualization, Writing – review & editing, Supervision. **Weiyu Zhang:** Writing – review & editing. **Ning Hou:** Writing – review & editing. **Chunjie Feng:** Writing – review & editing. **Shuyue Yang:** Writing – review & editing. **Kun Jia:** Writing – review & editing. **Yunjun Yao:** Writing – review & editing. **Xianhong Xie:** Writing – review & editing. **Bo Jiang:** Writing – review & editing. **Jie Cheng:** Writing – review & editing. **Xiang Zhao:** Writing – review & editing. **Shunlin Liang:** Writing – review & editing.

## Declaration of Competing Interest

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

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants 42171320, in part by the National Key Research and Development Program of China under Grant 2016YFA0600103.

## Appendix

**Table A1**
Detailed information on the 5th phase of the Coupled Model Intercomparison Project (CMIP5) general circulation models (GCMs) used in this study.

| ID | Model Name | Institute ID | Time | Resolution |
|---|---|---|---|---|
| 1 | ACCESS1–0 | CSIRO-BOM | 185001–200512 | 1.88° × 1.24° |
| 2 | ACCESS1–3 | CSIRO-BOM | 185001–200512 | 1.88° × 1.24° |
| 3 | BNU-ESM | GCESS | 185001–200512 | 2.81° × 2.81° |
| 4 | CCSM4 | NCAR | 185001–200512 | 1.25° × 0.94° |
| 5 | CESM1-BGC | NSF-DOE-NCAR | 185001–200512 | 1.25° × 0.94° |
| 6 | CESM1-CAM5 | NSF-DOE-NCAR | 185001–200512 | 1.25° × 0.94° |
| 7 | CESM1-FASTCHEM | NSF-DOE-NCAR | 185001–200512 | 1.25° × 0.94° |
| 8 | CESM1-WACCM | NSF-DOE-NCAR | 185001–200512 | 2.50° × 1.88° |
| 9 | CMCC-CESM | CMCC | 185001–200512 | 3.75° × 3.75° |
| 10 | CMCC-CMS | CMCC | 185001–200512 | 1.88° × 1.88° |
| 11 | CMCC-CM | CMCC | 185001–200512 | 0.75° × 0.75° |
| 12 | CNRM-CM5–2 | CNRM-CERFACS | 185001–200512 | 1.41° × 1.41° |
| 13 | CNRM-CM5 | CNRM-CERFACS | 185001–200512 | 1.41° × 1.41° |
| 14 | CSIRO-Mk3–6-0 | CSIRO-QCCCE | 185001–200512 | 1.88° × 1.88° |
| 15 | CanCM4 | CCCMA | 196101–200512 | 2.81° × 2.81° |
| 16 | CanESM2 | CCCMA | 185001–200512 | 2.81° × 2.81° |
| 17 | FGOALS-g2 | LASG-CESS | 185001–200512 | 2.81° × 3.00° |
| 18 | GFDL-CM2p1 | NOAA GFDL | 186101–200512 | 2.50° × 2.00° |
| 19 | GFDL-CM3 | NOAA GFDL | 186001–200512 | 2.50° × 2.00° |
| 20 | GFDL-ESM2G | NOAA GFDL | 186101–200512 | 2.50° × 2.00° |
| 21 | GFDL-ESM2M | NOAA GFDL | 186101–200512 | 2.50° × 2.00° |
| 22 | GISS-E2-H-CC | NOAA GISS | 185001–201012 | 2.50° × 2.00° |
| 23 | GISS-E2-H | NOAA GISS | 185001–200512 | 2.50° × 2.00° |
| 24 | GISS-E2-R-CC | NOAA GISS | 185001–201012 | 2.50° × 2.00° |
| 25 | GISS-E2-R | NOAA GISS | 185001–200512 | 2.50° × 2.00° |
| 26 | HadCM3 | MOHC | 185912–200512 | 3.75° × 3.47° |
| 27 | HadGEM2-AO | NIMR/KMA | 186001–200512 | 1.88° × 1.24° |
| 28 | HadGEM2-CC | MOHC | 185912–200511 | 1.88° × 1.24° |
| 29 | HadGEM2-ES | MOHC | 185912–200511 | 1.88° × 1.24° |
| 30 | IPSL-CM5A-LR | IPSL | 185001–200512 | 3.75° × 1.88° |
| 31 | IPSL-CM5A-MR | IPSL | 185001–200512 | 2.50° × 1.26° |
| 32 | IPSL-CM5B-LR | IPSL | 185001–200512 | 3.75° × 1.88° |
| 33 | MIROC-ESM-CHEM | MIROC | 185001–200512 | 2.81° × 2.81° |
| 34 | MIROC-ESM | MIROC | 185001–200512 | 2.81° × 2.81° |
| 35 | MIROC4h | MIROC | 195001–200512 | 0.56° × 0.56° |
| 36 | MIROC5 | MIROC | 185001–201212 | 1.41° × 1.41° |
| 37 | MPI-ESM-LR | MPI-M | 185001–200512 | 1.88° × 1.88° |
| 38 | MPI-ESM-MR | MPI-M | 185001–200512 | 1.88° × 1.88° |
| 39 | MPI-ESM-P | MPI-M | 185001–200512 | 1.88° × 1.88° |
| 40 | MRI-CGCM3 | MRI | 185001–200512 | 1.13° × 1.13° |
| 41 | MRI-ESM1 | NCC | 185101–200512 | 1.13° × 1.13° |
| 42 | NorESM1-ME | NCC | 185001–200512 | 2.50° × 1.88° |
| 43 | NorESM1-M | NCC | 185001–200512 | 2.50° × 1.88° |
| 44 | bcc-csm1–1-m | BCC | 185001–201212 | 1.13° × 1.13° |
| 45 | bcc-csm1–1 | BCC | 185001–201212 | 1.13° × 1.13° |
| 46 | inmcm4 | UNM | 185001–200512 | 2.00° × 1.50° |

**Table A2**
Detailed information on the 6th phase of the Coupled Model Intercomparison Project (CMIP6) general circulation models (GCMs) used in this study.

| ID | Model Name | Institute ID | Time | Resolution |
|---|---|---|---|---|
| 1 | ACCESS-CM2 | CSIRO-ARCCSS | 185001–201412 | 1.88° × 1.25° |
| 2 | ACCESS-ESM1–5 | CSIRO | 185001–201412 | 1.88° × 1.24° |
| 3 | AWI-CM-1-1-MR | AWI | 185001–201412 | 0.94° × 0.94° |
| 4 | AWI-ESM-1-1-LR | AWI | 185001–201412 | 1.88° × 1.88° |
| 5 | BCC-CSM2-MR | BCC | 185,001–201,412 | 1.13° × 1.13° |
| 6 | BCC-ESM1 | BCC | 185001–201412 | 2.81° × 2.81° |
| 7 | CAMS-CSM1–0 | CAMS | 185001–201412 | 1.13° × 1.13° |
| 8 | CAS-ESM2–0 | CAS | 185001–201412 | 1.41° × 1.41° |
| 9 | CESM2-FV2 | NCAR | 185001–201412 | 2.50° × 1.88° |
| 10 | CESM2-WACCM-FV2 | NCAR | 185001–201412 | 2.50° × 1.88° |
| 11 | CESM2-WACCM | NCAR | 185001–201412 | 1.25° × 0.94° |
| 12 | CESM2 | NCAR | 185001–201412 | 1.25° × 0.94° |
| 13 | CIESM | THU | 185001–201412 | 1.25° × 0.94° |
| 14 | CMCC-CM2-HR4 | CMCC | 185001–201412 | 1.25° × 0.94° |
| 15 | CMCC-CM2-SR5 | CMCC | 185001–201412 | 1.25° × 0.94° |
| 16 | CanESM5 | CCCma | 185001–201412 | 2.81° × 2.81° |
| 17 | E3SM-1-0 | E3SM-Project | 185001–201412 | 1.00° × 1.00° |
| 18 | E3SM-1-1-ECA | E3SM-Project | 185001–201412 | 1.00° × 1.00° |
| 19 | E3SM-1-1 | E3SM-Project | 185001–201, | 1.00° × 1.00° |
| 20 | EC-Earth3-AerChem | EC-Earth-Consortium | 185001–201412 | 0.70° × 0.70° |
| 21 | EC-Earth3-Veg-LR | EC-Earth-Consortium | 185001–201412 | 1.13° × 1.13° |
| 22 | EC-Earth3-Veg | EC-Earth-Consortium | 185001–201412 | 0.70° × 0.70° |
| 23 | EC-Earth3 | EC-Earth-Consortium | 185001–201412 | 0.70° × 0.70° |
| 24 | FGOALS-f3-L | CAS | 185001–201412 | 1.25° × 1.00° |
| 25 | FGOALS-g3 | CAS | 185001–201612 | 2.00° × 2.25° |
| 26 | FIO-ESM-2-0 | FIO-QLNM | 185001–201412 | 1.25° × 0.94° |
| 27 | GFDL-ESM4 | NOAA-GFDL | 185001–201412 | 1.25° × 1.00° |
| 28 | GISS-E2–1-G-CC | NASA-GISS | 185001–201412 | 2.50° × 2.00° |
| 29 | GISS-E2–1-G | NASA-GISS | 185001–201412 | 2.50° × 2.00° |
| 30 | GISS-E2–1-H | NASA-GISS | 185001–201412 | 2.50° × 2.00° |
| 31 | IITM-ESM | CCCR-IITM | 185001–201412 | 1.88° × 1.91° |
| 32 | INM-CM4–8 | INM | 185001–201412 | 2.00° × 1.50° |
| 33 | INM-CM5–0 | INM | 185001–201,412 | 2.00° × 1.50° |
| 34 | IPSL-CM6A-LR | IPSL | 185,001–201412 | 2.50° × 1.26° |
| 35 | KACE-1-0-G | NIMS-KMA | 185001–201412 | 1.88° × 1.25° |
| 36 | KIOST-ESM | KIOST | 185001–201412 | 1.88° × 1.88° |
| 37 | MIROC6 | MIROC | 185001–201412 | 1.41° × 1.41° |
| 38 | MPI-ESM-1-2-HAM | HAMMOZ-Consortium | 185001–201412 | 1.88° × 1.88° |
| 39 | MPI-ESM1–2-HR | MPI-M | 185001–201412 | 0.94° × 0.94° |
| 40 | MPI-ESM1–2-LR | MPI-M | 185001–201412 | 1.88° × 1.88° |
| 41 | MRI-ESM2–0 | MRI | 185001–201412 | 1.13° × 1.13° |
| 42 | NESM3 | NUIST | 185001–201412 | 1.88° × 1.88° |
| 43 | NorCPM1 | NCC | 185001–202912 | 2.50° × 1.88° |
| 44 | NorESM2-LM | NCC | 185001–201412 | 2.50° × 1.88° |
| 45 | NorESM2-MM | NCC | 185001–201412 | 1.25° × 0.94° |
| 46 | SAM0-UNICON | SNU | 185001–201412 | 1.25° × 0.94° |
| 47 | TaiESM1 | AS-RCEC | 185001–201412 | 1.25° × 0.94° |

**Table A3**
Information of the sites from BSRN, SURFRAD, and FLUXNET used in this study.

| Network | Site ID | Latitude | Longitude | Network | Site ID | Latitude | Longitude |
|---|---|---|---|---|---|---|---|
| BSRN | ALE | 82.49° N | 62.42° W | FLUXNET | BE-Bra | 51.31° N | 4.52° E |
| BSRN | ASP | 23.80° S | 133.89° E | FLUXNET | BE-Lon | 50.55° N | 4.75° E |
| BSRN | BAR | 71.32° N | 156.61° W | FLUXNET | BR-Sa3 | 3.02° S | 54.97° W |
| BSRN | BER | 32.27° N | 64.67° W | FLUXNET | CA-Qfo | 49.69° N | 74.34° W |
| BSRN | BIL | 36.61° N | 97.52° W | FLUXNET | CA-SF1 | 54.49° N | 105.82° W |
| BSRN | BON | 40.07° N | 88.37° W | FLUXNET | CA-SF2 | 54.25° N | 105.88° W |
| BSRN | BOS | 40.13° N | 105.24° W | FLUXNET | CA-SF3 | 54.09° N | 106.01° W |
| BSRN | BOU | 40.05° N | 105.01° W | FLUXNET | CH-Cha | 47.21° N | 8.41° E |
| BSRN | BRB | 15.60° S | 47.71° W | FLUXNET | CH-Dav | 46.82° N | 9.86° E |
| BSRN | CAB | 51.97° N | 4.93° E | FLUXNET | CH-Fru | 47.12° N | 8.54° E |
| BSRN | CAM | 50.22° N | 5.32° W | FLUXNET | CH-Lae | 47.48° N | 8.37° E |
| BSRN | CAR | 44.08° N | 5.06° E | FLUXNET | CH-Oe1 | 47.29° N | 7.73° E |
| BSRN | CLH | 36.91° N | 75.71° W | FLUXNET | CH-Oe2 | 47.29° N | 7.73° E |
| BSRN | CNR | 42.82° N | 1.60° W | FLUXNET | CN-Cha | 42.40° N | 128.10° E |
| BSRN | COC | 12.19° S | 96.84° E | FLUXNET | CN-Cng | 44.59° N | 123.51° E |
| BSRN | DAA | 30.67° S | 23.99° E | FLUXNET | CN-Dan | 30.50° N | 91.07° E |
| BSRN | DAR | 12.43° S | 130.89° E | FLUXNET | CN-Din | 23.17° N | 112.54° E |
| BSRN | DOM | 75.10° S | 123.38° E | FLUXNET | CN-Ha2 | 37.61° N | 101.33° E |
| BSRN | DRA | 36.63° N | 116.02° W | FLUXNET | CN-Qia | 26.74° N | 115.06° E |

(*continued on next page*)

**Table A3** (*continued*)

| Network | Site ID | Latitude | Longitude | Network | Site ID | Latitude | Longitude |
|---|---|---|---|---|---|---|---|
| BSRN | DWN | 12.42° S | 130.89° E | FLUXNET | CZ-BK1 | 49.50° N | 18.54° E |
| BSRN | E13 | 36.61° N | 97.49° W | FLUXNET | CZ-BK2 | 49.49° N | 18.54° E |
| BSRN | ENA | 39.09° N | 28.03° W | FLUXNET | CZ-wet | 49.02° N | 14.77° E |
| BSRN | EUR | 79.99° N | 85.94° W | FLUXNET | DE-Akm | 53.87° N | 13.68° E |
| BSRN | FLO | 27.60° S | 48.52° W | FLUXNET | DE-Geb | 51.10° N | 10.91° E |
| BSRN | FPE | 48.32° N | 105.10° W | FLUXNET | DE-Gri | 50.95° N | 13.51° E |
| BSRN | FUA | 33.58° N | 130.38° E | FLUXNET | DE-Hai | 51.08° N | 10.45° E |
| BSRN | GAN | 23.11° N | 72.63° E | FLUXNET | DE-Kli | 50.89° N | 13.52° E |
| BSRN | GCR | 34.25° N | 89.87° W | FLUXNET | DE-Lkb | 49.10° N | 13.30° E |
| BSRN | GOB | 23.56° S | 15.04° E | FLUXNET | DE-Obe | 50.79° N | 13.72° E |
| BSRN | GUR | 28.42° N | 77.16° E | FLUXNET | DE-RuR | 50.62° N | 6.30° E |
| BSRN | GVN | 70.65° S | 8.25° W | FLUXNET | DE-RuS | 50.87° N | 6.45° E |
| BSRN | HOW | 22.55° N | 88.31° E | FLUXNET | DE-SfN | 47.81° N | 11.33° E |
| BSRN | ILO | 8.53° N | 4.57° E | FLUXNET | DE-Spw | 51.89° N | 14.03° E |
| BSRN | ISH | 24.34° N | 124.16° E | FLUXNET | DE-Tha | 50.96° N | 13.57° E |
| BSRN | KWA | 8.72° N | 167.73° E | FLUXNET | DK-Sor | 55.49° N | 11.64° E |
| BSRN | LAU | 45.05° S | 169.69° E | FLUXNET | FI-Hyy | 61.85° N | 24.29° E |
| BSRN | LER | 60.14° N | 1.18° W | FLUXNET | FI-Lom | 68.00° N | 24.21° E |
| BSRN | LIN | 52.21° N | 14.12° E | FLUXNET | FR-Gri | 48.84° N | 1.95° E |
| BSRN | MAN | 2.06° S | 147.43° E | FLUXNET | FR-LBr | 44.72° N | 0.77° W |
| BSRN | MNM | 24.29° N | 153.98° E | FLUXNET | FR-Pue | 43.74° N | 3.60° E |
| BSRN | NAU | 0.52° S | 166.92° E | FLUXNET | GF-Guy | 5.28° N | 52.92° W |
| BSRN | NYA | 78.93° N | 11.93° E | FLUXNET | IT-BCi | 40.52° N | 14.96° E |
| BSRN | PAL | 48.71° N | 2.21° E | FLUXNET | IT-CA1 | 42.38° N | 12.03° E |
| BSRN | PAY | 46.82° N | 6.94° E | FLUXNET | IT-CA2 | 42.38° N | 12.03° E |
| BSRN | PSU | 40.72° N | 77.93° W | FLUXNET | IT-CA3 | 42.38° N | 12.02° E |
| BSRN | PTR | 9.07° S | 40.32° W | FLUXNET | IT-Col | 41.85° N | 13.59° E |
| BSRN | REG | 50.21° N | 104.71° W | FLUXNET | IT-Isp | 45.81° N | 8.63° E |
| BSRN | SAP | 43.06° N | 141.33° E | FLUXNET | IT-La2 | 45.95° N | 11.29° E |
| BSRN | SBO | 30.86° N | 34.78° E | FLUXNET | IT-Lav | 45.96° N | 11.28° E |
| BSRN | SMS | 29.44° S | 53.82° W | FLUXNET | IT-MBo | 46.01° N | 11.05° E |
| BSRN | SON | 47.05° N | 12.96° E | FLUXNET | IT-NOE | 40.61° N | 8.15° E |
| BSRN | SOV | 24.91° N | 46.41° E | FLUXNET | IT-Ren | 46.59° N | 11.43° E |
| BSRN | SPO | 89.98° S | 24.80° W | FLUXNET | IT-Ro1 | 42.41° N | 11.93° E |
| BSRN | SXF | 43.73° N | 96.62° W | FLUXNET | IT-Ro2 | 42.39° N | 11.92° E |
| BSRN | SYO | 69.01° S | 39.59° E | FLUXNET | IT-SR2 | 43.73° N | 10.29° E |
| BSRN | TAM | 22.79° N | 5.53° E | FLUXNET | IT-SRo | 43.73° N | 10.28° E |
| BSRN | TAT | 36.06° N | 140.13° E | FLUXNET | IT-Tor | 45.84° N | 7.58° E |
| BSRN | TIK | 71.59° N | 128.92° E | FLUXNET | JP-MBF | 44.39° N | 142.32° E |
| BSRN | TIR | 13.09° N | 79.97° E | FLUXNET | JP-SMF | 35.26° N | 137.08° E |
| BSRN | TOR | 58.25° N | 26.46° E | FLUXNET | NL-Hor | 52.24° N | 5.07° E |
| BSRN | XIA | 39.75° N | 116.96° E | FLUXNET | NL-Loo | 52.17° N | 5.74° E |
| SURFRAD | BND | 40.05° N | 88.37° W | FLUXNET | RU-Fyo | 56.46° N | 32.92° E |
| SURFRAD | TBL | 40.12° N | 105.24° W | FLUXNET | SJ-Adv | 78.19° N | 15.92° E |
| SURFRAD | DRA | 36.62° N | 116.02° W | FLUXNET | SJ-Blv | 78.92° N | 11.83° E |
| SURFRAD | FPK | 48.31° N | 105.10° W | FLUXNET | US-AR1 | 36.43° N | 99.42° W |
| SURFRAD | GWN | 34.25° N | 89.87° W | FLUXNET | US-AR2 | 36.64° N | 99.60° W |
| SURFRAD | PSU | 40.72° N | 77.93° W | FLUXNET | US-GBT | 41.37° N | 106.24° W |
| SURFRAD | SXF | 43.73° N | 96.62° W | FLUXNET | US-GLE | 41.37° N | 106.24° W |
| FLUXNET | AT-Neu | 47.12° N | 11.32° E | FLUXNET | US-Los | 46.08° N | 89.98° W |
| FLUXNET | AU-Ade | 13.08° S | 131.12° E | FLUXNET | US-Me2 | 44.45° N | 121.56° W |
| FLUXNET | AU-ASM | 22.28° S | 133.25° E | FLUXNET | US-Me6 | 44.32° N | 121.61° W |
| FLUXNET | AU-Cpr | 34.00° S | 140.59° E | FLUXNET | US-MMS | 39.32° N | 86.41° W |
| FLUXNET | AU-Cum | 33.62° S | 150.72° E | FLUXNET | US-Ne1 | 41.17° N | 96.48° W |
| FLUXNET | AU-DaP | 14.06° S | 131.32° E | FLUXNET | US-Ne2 | 41.16° N | 96.47° W |
| FLUXNET | AU-DaS | 14.16° S | 131.39° E | FLUXNET | US-Ne3 | 41.18° N | 96.44° W |
| FLUXNET | AU-Dry | 15.26° S | 132.37° E | FLUXNET | US-ORv | 40.02° N | 83.02° W |
| FLUXNET | AU-Emr | 23.86° S | 148.47° E | FLUXNET | US-Prr | 65.12° N | 147.49° W |
| FLUXNET | AU-Fog | 12.55° S | 131.31° E | FLUXNET | US-SRG | 31.79° N | 110.83° W |
| FLUXNET | AU-Gin | 31.38° S | 115.71° E | FLUXNET | US-SRM | 31.82 N | 110.87° W |
| FLUXNET | AU-GWW | 30.19° S | 120.65° E | FLUXNET | US-Syv | 46.24° N | 89.35° W |
| FLUXNET | AU-How | 12.49° S | 131.15° E | FLUXNET | US-Tw1 | 38.11° N | 121.65° W |
| FLUXNET | AU-Lox | 34.47° S | 140.66° E | FLUXNET | US-Tw2 | 38.10° N | 121.64° W |
| FLUXNET | AU-RDF | 14.56° S | 132.48° E | FLUXNET | US-Tw3 | 38.12° N | 121.65° W |
| FLUXNET | AU-Rig | 36.65° S | 145.58° E | FLUXNET | US-Tw4 | 38.10° N | 121.64° W |
| FLUXNET | AU-Rob | 17.12° S | 145.63° E | FLUXNET | US-UMB | 45.56° N | 84.71° W |
| FLUXNET | AU-Stp | 17.15° S | 133.35° E | FLUXNET | US-UMd | 45.56° N | 84.70° W |
| FLUXNET | AU-TTE | 22.29° S | 133.64° E | FLUXNET | US-Var | 38.41° N | 120.95° W |
| FLUXNET | AU-Tum | 35.66° S | 148.15° E | FLUXNET | US-WCr | 45.81° N | 90.08° W |
| FLUXNET | AU-Wac | 37.43° S | 145.19° E | FLUXNET | US-Whs | 31.74° N | 110.05° W |
| FLUXNET | AU-Whr | 36.67° S | 145.03° E | FLUXNET | US-Wkg | 31.74° N | 109.94° W |
| FLUXNET | AU-Wom | 37.42° S | 144.09° E | FLUXNET | ZA-Kru | 25.02° S | 31.50° E |
| FLUXNET | AU-Ync | 34.99° S | 146.29° E | | | | |

## References


Ahmed, H.A., Yu-xin, T., Qi-chang, Y., Al-Helal, I.M., Shady, M.R., Abdel-Ghany, A.M., 2020. Estimation of sky thermal irradiance in arid climate under clear sky conditions. Int. J. Thermophys. 41 (6), 76. https://doi.org/10.1007/s10765-020-02656-1.

Augustine, J., DeLuisi, J., Long, C., 2000. SURFRAD-A national surface radiation budget network for atmospheric research. Bull. Am. Meteorol. Soc. 81 (10), 2341–2357. https://doi.org/10.1175/1520-0477(2000)081<2341:SANSRB>2.3.CO;2.

Bhat, K.S., Haran, M., Terando, A., Keller, K., 2011. Climate projections using Bayesian model averaging and space–time dependence. J. Agric. Biol. Environ. Stat. 16 (4), 606–628. https://doi.org/10.1007/s13253-011-0069-3.

Bodas-Salcedo, A., Ringer, M.A., Jones, A., 2008. Evaluation of the surface radiation budget in the atmospheric component of the Hadley Centre Global Environmental Model (HadGEM1). J. Clim. 21 (18), 4723–4748. https://doi.org/10.1175/2008jcli2097.1.

Carrer, D., Lafont, S., Roujean, J.-L., Calvet, J.-C., Meurey, C., Le Moigne, P., Trigo, I.F., 2012. Incoming solar and infrared radiation derived from METEOSAT: impact on the modeled land water and energy budget over France. J. Hydrometeorol. 13 (2), 504–520. https://doi.org/10.1175/JHM-D-11-059.1.

Chen, X., Pauluis, O.M., Leung, L.R., Zhang, F., 2020. Significant contribution of mesoscale overturning to tropical mass and energy transport revealed by the ERA5 reanalysis. Geophys. Res. Lett. 47 (1) https://doi.org/10.1029/2019GL085333 e2019GL085333.

Cheng, J., Yang, F., Guo, Y., 2019. A comparative study of bulk parameterization schemes for estimating cloudy-sky surface downward longwave radiation. Remote Sens. 11 (5), 528. https://doi.org/10.3390/rs11050528.

Cheng, J., Liang, S., Shi, J., 2020. Impact of air temperature inversion on the clear-sky surface downward longwave radiation estimation. IEEE Trans. Geosci. Remote Sens. 58 (7), 4796–4802. https://doi.org/10.1109/TGRS.2020.2967432.

Delia Garcia, R., Cuevas, E., Ramos, R., Eugenia Cachorro, V., Redondas, A., Moreno-Ruiz, J.A., 2019. Description of the Baseline Surface Radiation Network (BSRN) station at the Izaña Observatory (2009–2017): measurements and quality control/assurance procedures. Geosci. Instrument. Methods Data Syst. 8 (1), 77–96. https://doi.org/10.5194/gi-8-77-2019.

Duan, Q., Ajami, N.K., Gao, X., Sorooshian, S., 2007. Multi-model ensemble hydrologic prediction using Bayesian model averaging. Adv. Water Resour. 30 (5), 1371–1386. https://doi.org/10.1016/j.advwatres.2006.11.014.

Eyring, V., Bony, S., Meehl, G.A., Senior, C.A., Stevens, B., Stouffer, R.J., Taylor, K.E., 2016. Overview of the coupled Model Intercomparison Project phase 6 (CMIP6) experimental design and organization. Geosci. Model Dev. 9 (5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016.

Fang, M., Li, X., 2016. Application of Bayesian model averaging in the reconstruction of past climate change using PMIP3/CMIP5 Multimodel ensemble Simulations. J. Clim. 29 (1), 175–189. https://doi.org/10.1175/jcli-d-14-00752.1.

Flynn, C.J., Eloranta, E.W., Cadeddu, M., Fridlind, A.M., Bromwich, D.H., Wang, S.-H., Verlinde, J., Silber, I., 2019. Cloud influence on ERA5 and AMPS surface downwelling longwave radiation biases in West Antarctica. J. Clim. 32 (22), 7935–7949. https://doi.org/10.1175/jcli-d-19-0149.1.

Graham, R.M., Cohen, L., Ritzhaupt, N., Segger, B., Graversen, R.G., Rinke, A., Walden, V.P., Granskog, M.A., Hudson, S.R., 2019. Evaluation of six atmospheric reanalyses over Arctic Sea ice from winter to early summer. J. Clim. 32 (14), 4121–4143. https://doi.org/10.1175/JCLI-D-18-0643.1.

Guilyardi, E., Balaji, V., Lawrence, B., Callaghan, S., Deluca, C., Denvil, S., Lautenschlager, M., Morgan, M., Murphy, S., Taylor, K.E., 2013. Documenting climate models and their simulations. Bull. Am. Meteorol. Soc. 94 (5), 623–629. https://doi.org/10.1175/BAMS-D-11-00035.1.

Guo, Y., Cheng, J., Liang, S., 2019. Comprehensive assessment of parameterization methods for estimating clear-sky surface downward longwave radiation. Theor. Appl. Climatol. 135 (3–4), 1045–1058. https://doi.org/10.1007/s00704-018-2423-7.

Guo, J., Ren, H., Zheng, Y., Lu, S., Dong, J., 2020. Evaluation of land surface temperature retrieval from Landsat 8/TIRS images before and after stray light correction using the SURFRAD dataset. Remote Sens. 12 (6), 1023. https://doi.org/10.3390/rs12061023.

Gupta, S., Ritchey, N., Wilber, A., Whitlock, C., Gibson, G., Stackhouse, P., 1999. A climatology of surface radiation budget derived from satellite data. J. Clim. 12 (8), 2691–2710. https://doi.org/10.1175/1520-0442(1999)012<2691:ACOSRB>2.0.CO;2.

Hamal, K., Sharma, S., Baniya, B., Khadka, N., Zhou, X., 2020. Inter-annual variability of winter precipitation over Nepal coupled with Ocean-Atmospheric patterns during 1987–2015. Front. Earth Sci. 8, 161. https://doi.org/10.3389/feart.2020.00161.

Hatzianastassiou, N., Ioannidis, E., Korras-Carraca, M.-B., Gavrouzou, M., Papadimas, C. D., Matsoukas, C., Benas, N., Fotiadi, A., Wild, M., Vardavas, I., 2020. Global dimming and brightening features during the first decade of the 21st Century. Atmosphere 11 (3), 308. https://doi.org/10.3390/atmos11030308.

Hinkelman, L.M., 2019. The global radiative energy budget in MERRA and MERRA-2: evaluation with respect to CERES EBAF data. J. Clim. 32 (6), 1973–1994. https://doi.org/10.1175/JCLI-D-18-0445.1.

Jia, B., Cai, X., Zhao, F., Liu, J., Chen, S., Luo, X., Xie, Z., Xu, J., 2020. Potential future changes of terrestrial water storage based on climate projections by ensemble model simulations. Adv. Water Resour. 142, 103635 https://doi.org/10.1016/j.advwatres.2020.103635.

Kim, K.-Y., Na, H., Lee, H.-G., 2013. Energy budget change in the tropics according to the SRES A1B scenario in the IPCC AR4 models. J. Geophys. Res.-Atmos. 118 (6), 2521–2534. https://doi.org/10.1002/jgrd.50240.

Kim, Y.-H., Min, S.-K., Zhang, X., Sillmann, J., Sandstad, M., 2020. Evaluation of the CMIP6 multi-model ensemble for climate extreme indices. Weather and Climate Extremes 29, 100269. https://doi.org/10.1016/j.wace.2020.100269.

Kusunoki, S., Arakawa, O., 2015. Are CMIP5 models better than CMIP3 models in simulating precipitation over East Asia? J. Clim. 28 (14), 5601–5621. https://doi.org/10.1175/jcli-d-14-00585.1.

L'Ecuyer, T.S., et al., 2015. The observed state of the energy budget in the early twenty-first Century. J. Clim. 28 (21), 8319–8346. https://doi.org/10.1175/jcli-d-14-00556.1.

Li, J.L.F., Lee, W.-L., Yu, J.-Y., Hulley, G., Fetzer, E., Chen, Y.-C., Wang, Y.-H., 2016. The impacts of precipitating hydrometeors radiative effects on land surface temperature in contemporary GCMs using satellite observations. J. Geophys. Res.-Atmos. 121 (1), 67–79. https://doi.org/10.1002/2015jd023776.

Li, J.L.F., Richardson, M., Hong, Y., Lee, W.-L., Wang, Y.-H., Yu, J.-Y., Fetzer, E., Stephens, G., Liu, Y., 2017. Improved simulation of Antarctic Sea ice due to the radiative effects of falling snow. Environ. Res. Lett. 12 (8), 084010 https://doi.org/10.1088/1748-9326/aa7a17.

Li, J.-L.F., Richardson, M., Lee, W.-L., Fetzer, E., Stephens, G., Jiang, J., Hong, Y., Wang, Y.-H., Yu, J.-Y., Liu, Y., 2019. Potential faster Arctic Sea ice retreat triggered by snowflakes' greenhouse effect. Cryosphere 13 (3), 969–980. https://doi.org/10.5194/tc-13-969-2019.

Li, J., Zhou, M., Lenschow, D.H., Cheng, Z., Dou, Y., 2020. Observed relationships between the urban heat island, urban pollution island, and downward longwave radiation in the Beijing area. Earth and Space. Science 7 (6). https://doi.org/10.1029/2020ea001100.

Liang, S., Wang, K., Zhang, X., Wild, M., 2010. Review on estimation of land surface radiation and energy budgets from ground measurement, remote sensing and model simulations. IEEE J. Selected Top. Appl. Earth Observat. Remote Sensing 3 (3), 225–240. https://doi.org/10.1109/jstars.2010.2048556.

Loeb, N.G., Kato, S., Rose, F.G., Doelling, D.R., Rutan, D.A., Caldwell, T.E., Yu, L., Weller, R.A., 2013. Surface irradiances consistent with CERES-Derived Top-of-Atmosphere shortwave and longwave irradiances. J. Clim. 26 (9), 2719–2740. https://doi.org/10.1175/jcli-d-12-00436.1.

Loeb, N.G., Doelling, D.R., Wang, H., Su, W., Nguyen, C., Corbett, J.G., Liang, L., Mitrescu, C., Rose, F.G., Kato, S., 2018. Clouds and the Earth's Radiant Energy System (CERES) energy Balanced and Filled (EBAF) Top-of-Atmosphere (TOA) Edition-4.0 data product. J. Clim. 31 (2), 895–918. https://doi.org/10.1175/JCLI-D-17-0208.1.

Ma, Q., Wang, K., Wild, M., 2014. Evaluations of atmospheric downward longwave radiation from 44 coupled general circulation models of CMIP5. J. Geophys. Res.-Atmos. 119 (8), 4486–4497. https://doi.org/10.1002/2013jd021427.

Marotzke, J., et al., 2017. Climate research must sharpen its view. Nat. Clim. Chang. 7 (2), 89–91. https://doi.org/10.1038/nclimate3206.

Medina, H., Tian, D., 2020. Comparison of probabilistic post-processing approaches for improving numerical weather prediction-based daily and weekly reference evapotranspiration forecasts. Hydrol. Earth Syst. Sci. 24 (2), 1011–1030. https://doi.org/10.5194/hess-24-1011-2020.

Min, S.K., Simonis, D., Hense, A., 2007. Probabilistic climate change predictions applying Bayesian model averaging. Philosophical Transact. Royal Society A-Math. Phys. Eng. Sci. 365 (1857), 2103–2116. https://doi.org/10.1098/rsta.2007.2070.

Nyeki, S., Wacker, S., Aebi, C., Gröbner, J., Martucci, G., Vuilleumier, L., 2019. Trends in surface radiation and cloud radiative effect at four Swiss sites for the 1996–2015 period. Atmos. Chem. Phys. 19 (20), 13227–13241. https://doi.org/10.5194/acp-19-13227-2019.

Ohmura, A., et al., 1998. Baseline Surface Radiation Network (BSRN/WCRP): new precision radiometry for climate research. Bull. Am. Meteorol. Soc. 79 (10), 2115–2136. https://doi.org/10.1175/1520-0477(1998)079<2115:BSRNBW>2.0.CO;2.

Pastorello, G., Trotta, C., Canfora, E., e. a., 2020. The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. Sci. Data 7 (1), 225. https://doi.org/10.1038/s41597-020-0534-3.

Perez, J., Menendez, M., Mendez, F.J., Losada, I.J., 2014. Evaluating the performance of CMIP3 and CMIP5 global climate models over the north-East Atlantic region. Clim. Dyn. 43 (9–10), 2663–2680. https://doi.org/10.1007/s00382-014-2078-8.

Philipona, R., Durr, B., Marty, C., Ohmura, A., Wild, M., 2004. Radiative forcing - measured at Earth's surface - corroborate the increasing greenhouse effect. Geophys. Res. Lett. 31 (3) https://doi.org/10.1029/2003gl018765.

Prata, F., 2010. The climatological record of clear-sky longwave radiation at the Earth's surface: evidence for water vapour feedback? Int. J. Remote Sens. 29 (17–18), 5247–5263. https://doi.org/10.1080/01431160802036508.

Qin, B., Cao, B., Li, H., Bian, Z., Hu, T., Du, Y., Yang, Y., Xiao, Q., Liu, Q., 2020a. Evaluation of six high-spatial resolution clear-sky surface upward longwave radiation estimation methods with MODIS. Remote Sens. 12 (11), 1834. https://doi.org/10.3390/rs12111834.

Qin, W., Wang, L., Gueymard, C.A., Bilal, M., Lin, A., Wei, J., Zhang, M., Yang, X., 2020b. Constructing a gridded direct normal irradiance dataset in China during 1981–2014. Renew. Sust. Energ. Rev. 131, 110004 https://doi.org/10.1016/j.rser.2020.110004.

Raftery, A.E., Gneiting, T., Balabdaoui, F., Polakowski, M., 2005. Using Bayesian model averaging to calibrate forecast ensembles. Mon. Weather Rev. 133 (5), 1155–1174. https://doi.org/10.1175/MWR2906.1.

Sekertekin, A., Arslan, N., Bilgili, M., 2020. Modeling diurnal land surface temperature on a local scale of an arid environment using Artificial Neural Network (ANN) and time series of Landsat-8 derived spectral indexes. J. Atmos. Sol. Terr. Phys. 206, 105328 https://doi.org/10.1016/j.jastp.2020.105328.

Shi, Q., Liang, S., 2013. Characterizing the surface radiation budget over the Tibetan Plateau with ground-measured, reanalysis, and remote sensing data sets: 2.

Spatiotemporal analysis. J. Geophys. Res.-Atmos. 118 (16), 8921–8934. https://doi.org/10.1002/jgrd.50719.
Smith, R.L., Tebaldi, C., Nychka, D., Mearns, L.O., 2009. Bayesian modeling of uncertainty in ensembles of climate models. J. Am. Stat. Assoc. 104 (485), 97–116. https://doi.org/10.1198/jasa.2009.0007.
Smith, N.P., Thomas, S., Shankar, M., Szewczyk, Z.P., Wilson, R.S., Walikainen, D.R., Daniels, J.L., Hess, P.C., Priestley, K.J., 2014. Assessment of the Clouds and the Earth's Radiant Energy System (CERES) Flight Model 5 (FM5) Instrument Performance and Stability. In: Butler, J., Xiong, X., Gu, X. (Eds.), Conference on Earth Observing Systems XIX. https://doi.org/10.1117/12.2061833. UNSP 921802, San Diego.
Smith, N.P., Wilson, R., Szewczyk, Z.P., Thomas, S., Priestley, K.J., 2018. Early trends on the Clouds and the Earth's Radiant Energy System (CERES) flight Model 6 (FM6) instrument's performance. In: Butler, J., Xiong, X., Gu, X. (Eds.), Conference on Earth observing Systems XXIII. https://doi.org/10.1117/12.2322712. UNSP 107640R San Diego.
Stephens, G.L., Li, J., Wild, M., Clayson, C.A., Loeb, N., Kato, S., L'Ecuyer, T., Stackhouse, P.W., Lebsock, M., Andrews, T., 2012a. An update on Earth's energy balance in light of the latest global observations. Nat. Geosci. 5 (10), 691–696. https://doi.org/10.1038/ngeo1580.
Stephens, G.L., Wild, M., Stackhouse, P.W., L'Ecuyer, T., Kato, S., Henderson, D.S., 2012b. The global character of the flux of downward longwave radiation. J. Clim. 25 (7), 2329–2340. https://doi.org/10.1175/jcli-d-11-00262.1.
Swinbank, W., 1963. Long-wave radiation from clear skies. Q. J. R. Meteorol. Soc. 89 (381), 339–348. https://doi.org/10.1002/qj.49708938105.
Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res.-Atmos. 106 (D7), 7183–7192. https://doi.org/10.1029/2000jd900719.
Taylor, K.E., Stouffer, R.J., Meehl, G.A., 2012. An overview of CMIP5 and the experiment design. Bull. Am. Meteorol. Soc. 93 (4), 485–498. https://doi.org/10.1175/bams-d-11-00094.1.
Vaquero-Martínez, J., Antón, M., Sanchez-Lorenzo, A., Cachorro, V.E., 2020. Evaluation of Water Vapor Radiative Effects using GPS Data Series over Southwestern Europe. Remote Sens. 12 (8), 1307. https://doi.org/10.3390/rs12081307.
Verma, M., et al., 2016. Global surface net-radiation at 5 km from MODIS Terra. Remote Sens. 8 (9), 739. https://doi.org/10.3390/rs8090739.
Wacker, S., Gröbner, J., Hocke, K., Kämpfer, N., Vuilleumier, L., 2011. Trend analysis of surface cloud-free downwelling long-wave radiation from four Swiss sites. J. Geophys. Res. 116(D10). https://doi.org/10.1029/2010jd015343.
Wang, K., Dickinson, R.E., 2013. Global atmospheric downward longwave radiation at the surface from ground-based observations, satellite retrievals, and reanalyses. Rev. Geophys. 51 (2), 150–185. https://doi.org/10.1002/rog.20009.
Wang, K., Liang, S., 2009. Global atmospheric downward longwave radiation over land surface under all-sky conditions. from 1973 to 2008 J. Geophys. Res. 114(D19). https://doi.org/10.1029/2009jd011800.
Wang, T., Shi, J., Ma, Y., Letub, H., Li, X., 2020. All-sky longwave downward radiation from satellite measurements: General parameterizations based on LST, column water vapor and cloud top temperature. ISPRS J. Photogramm. Remote Sens. 161 https://doi.org/10.1016/j.isprsjprs.2020.01.011.
Ware, R., Fulker, D., Stein, S., Anderson, D., Avery, S., Clark, R., Droegemeier, K., Kuettner, J., Minster, J., Sorooshian, S., 2000. SuomiNet: a real-time national GPS network for atmospheric research and education. Bull. Am. Meteorol. Soc. 81 (4), 677–694. https://doi.org/10.1175/1520-0477(2000)081<0677:SARNGN>2.3.CO;2.
Wild, M., 2008. Short-wave and long-wave surface radiation budgets in GCMs: a review based on the IPCC-AR4/CMIP3 models. Tellus A: Dynamic Meteorol. Oceanogr. 60 (5), 932–945. https://doi.org/10.1111/j.1600-0870.2008.00342.x.
Wild, M., 2016. Decadal changes in radiative fluxes at land and ocean surfaces and their relevance for global warming. Wiley Interdiscip. Rev. Clim. Chang. 7 (1), 91–107. https://doi.org/10.1002/wcc.372.
Wild, M., 2017. In: Davies, R., Egli, L., Schmutz, W. (Eds.), Changes in Shortwave and Longwave Radiative Fluxes as Observed at BSRN Sites and Simulated with CMIP5 Models, in *Radiation Processes in the Atmosphere and Ocean*. Univ Auckland, Auckland, NEW ZEALAND. https://doi.org/10.1063/1.4975554.
Wild, M., 2020. The global energy balance as represented in CMIP6 climate models. Clim. Dyn. 55 (3–4), 553–577. https://doi.org/10.1007/s00382-020-05282-7.
Wild, M., Ohmura, A., Gilgen, H., Roeckner, E., Giorgetta, M., Morcrette, J.J., 1998. The disposition of radiative energy in the global climate system: GCM-calculated versus observational estimates. Clim. Dyn. 14 (12), 853–869. https://doi.org/10.1007/s003820050260.
Wild, M., Grieser, J., Schär, C., 2008. Combined surface solar brightening and increasing greenhouse effect support recent intensification of the global land-based hydrological cycle. Geophys. Res. Lett. 35 (17) https://doi.org/10.1029/2008gl034842.
Wild, M., Folini, D., Schär, C., Loeb, N., Dutton, E.G., König-Langlo, G., 2013. The global energy balance from a surface perspective. Clim. Dyn. 40 (11–12), 3107–3134. https://doi.org/10.1007/s00382-012-1569-8.
Wild, M., Folini, D., Hakuba, M.Z., Schär, C., Seneviratne, S.I., Kato, S., Rutan, D., Ammann, C., Wood, E.F., König-Langlo, G., 2015. The energy balance over land and oceans: an assessment based on direct observations and CMIP5 climate models. Clim. Dyn. 44 (11−12), 3393–3429. https://doi.org/10.1007/s00382-014-2430-z.
Wild, M., Hakuba, Maria Z., Folini, D., Dorig-Ott, P., Schar, C., Kato, S., Long, C.N., 2019. The cloud-free global energy balance and inferred cloud radiative effects: an assessment based on direct observations and climate models. Clim. Dyn. 52 (7–8), 4787–4812. https://doi.org/10.1007/s00382-018-4413-y.
Winter, J.M., Pal, J.S., Eltahir, E.A.B., 2009. Coupling of Integrated Biosphere Simulator to Regional climate Model Version 3. J. Clim. 22 (10), 2743–2757. https://doi.org/10.1175/2008jcli2541.1.
Wu, H., Zhang, X., Liang, S., Yang, H., Zhou, G., 2012. Estimation of clear-sky land surface longwave radiation from MODIS data products by merging multiple models. J. Geophys. Res. 117, D22107. https://doi.org/10.1029/2012JD017567.
Yao, Y., et al., 2016. Assessment and simulation of global terrestrial latent heat flux by synthesis of CMIP5 climate models and surface eddy covariance observations. Agric. For. Meteorol. 223, 151–167. https://doi.org/10.1016/j.agrformet.2016.03.016.
Zhang, X., Liang, S., Wang, G., Yao, Y., Jiang, B., Cheng, J., 2016. Evaluation of the reanalysis surface incident shortwave radiation products from NCEP, ECMWF, GSFC, and JMA using satellite and surface observations. Remote Sens. 8 (3), 225. https://doi.org/10.3390/rs8030225.
Zhang, H., Chu, P.-S., He, L., Unger, D., 2019a. Improving the CPC's ENSO forecasts using Bayesian model averaging. Clim. Dyn. 53 (5–6), 3373–3385. https://doi.org/10.1007/s00382-019-04710-7.
Zhang, W., Zhang, X., Li, W., Hou, N., Wei, Y., Jia, K., Yao, Y., Cheng, J., 2019b. Evaluation of Bayesian multimodel estimation in surface incident shortwave radiation simulation over high latitude areas. Remote Sens. 11 (15), 1776. https://doi.org/10.3390/rs11151776.
Zhou, W., Shi, J.C., Wang, T.X., Peng, B., Husi, L., Yu, Y.C., Zhao, R., 2019. New methods for deriving clear-sky surface longwave downward radiation based on remotely sensed data and ground measurements. Earth and Space Sci. 6 (11), 2071–2086. https://doi.org/10.1029/2019ea000754.