



## Article

# Evaluation of Surface Upward Longwave Radiation in the CMIP6 Models with Ground and Satellite Observations

Jiawen Xu, Xiaotong Zhang <sup>\*</sup>, Chunjie Feng , Shuyue Yang, Shikang Guan, Kun Jia , Yunjun Yao , Xianhong Xie , Bo Jiang , Jie Cheng and Xiang Zhao

State Key Laboratory of Remote Sensing Science, Jointly Sponsored by Beijing Normal University and Aerospace Information Research Institute of Chinese Academy of Sciences, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China; xujw@mail.bnu.edu.cn (J.X.); fcj20190901@mail.bnu.edu.cn (C.F.); ysy\_0128@mail.bnu.edu.cn (S.Y.); shikang\_guan@mail.bnu.edu.cn (S.G.); jiakun@bnu.edu.cn (K.J.); yaoyunjun@bnu.edu.cn (Y.Y.); xianhong@bnu.edu.cn (X.X.); bojiang@bnu.edu.cn (B.J.); Jie\_Cheng@bnu.edu.cn (J.C.); zhaoxiang@bnu.edu.cn (X.Z.)

\* Correspondence: xtngzhang@bnu.edu.cn

**Abstract:** Surface upward longwave radiation (SULR) is an indicator of thermal conditions over the Earth's surface. In this study, we validated the simulated SULR from 51 Coupled Model Inter-comparison Project (CMIP6) general circulation models (GCMs) through a comparison with ground measurements and satellite-retrieved SULR from the Clouds and the Earth's Radiant Energy System, Energy Balanced and Filled (CERES EBAF). Moreover, we improved the SULR estimations by a fusion of multiple CMIP6 GCMs using multimodel ensemble (MME) methods. Large variations were found in the monthly mean SULR among the 51 CMIP6 GCMs; the bias and root mean squared error (RMSE) of the individual CMIP6 GCMs at 133 sites ranged from  $-3$  to  $24 \text{ W m}^{-2}$  and  $22$  to  $38 \text{ W m}^{-2}$ , respectively, which were higher than those found between the CERES EBAF and GCMs. The CMIP6 GCMs did not improve the overestimation of SULR compared to the CMIP5 GCMs. The Bayesian model averaging (BMA) method showed better performance in simulating SULR than the individual GCMs and simple model averaging (SMA) method, with a bias of  $0 \text{ W m}^{-2}$  and an RMSE of  $19.29 \text{ W m}^{-2}$  for the 133 sites. In terms of the global annual mean SULR, our best estimation for the CMIP6 GCMs using the BMA method was  $392 \text{ W m}^{-2}$  during 2000–2014. We found that the SULR varied between  $386$  and  $393 \text{ W m}^{-2}$  from 1850 to 2014, exhibiting an increasing tendency of  $0.2 \text{ W m}^{-2}$  per decade ( $p < 0.05$ ).

**Keywords:** surface upward longwave radiation (SULR); CMIP6; CMIP5; GCMs; Bayesian model averaging; multimodel ensemble



**Citation:** Xu, J.; Zhang, X.; Feng, C.; Yang, S.; Guan, S.; Jia, K.; Yao, Y.; Xie, X.; Jiang, B.; Cheng, J.; et al. Evaluation of Surface Upward Longwave Radiation in the CMIP6 Models with Ground and Satellite Observations. *Remote Sens.* **2021**, *13*, 4464. <https://doi.org/10.3390/rs13214464>

Academic Editor: Bo-Hui Tang

Received: 11 October 2021

Accepted: 3 November 2021

Published: 6 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The surface energy budget is an important component in the energy exchange between the Earth's surface and atmosphere, which affects the fluxes of sensible and latent heat [1,2]. It is not only related to climate change [3] but also influences general oceanic and atmospheric circulations [4,5]. Surface upward longwave radiation (SULR), a surface radiation budget component [6–8], has been previously applied to determine ecologic and hydrologic characteristics [9–11]. SULR indicates the thermal conditions over the Earth's surface and is the main component during the nighttime and during the majority of the year over polar regions [12–14].

SULR can be acquired from ground measurements [15–17], general circulation models (GCMs) [18–20], reanalysis data [21,22], and remote sensing [7,23,24]. Among these sources, GCMs are one of the major tools applied to examine past and future climate changes on Earth and provide opportunities to investigate long-term SULR variability [25,26]. It has been organized by the World Climate Research Programme (WCRP) in the last several decades for the Coupled Model Intercomparison Project (CMIP) in various phases

and contributed to the Intergovernmental Panel on Climate Change (IPCC) Assessment Report (AR) [27–29]. The latest GCM version developed for CMIP6 was made available in 2016 [27,30–32] and includes improvements in the existing parameterizations, new physical processes, and higher resolution in comparison with CMIP5 [20,25,26,33–37]. Evaluating how well CMIP6 GCMs simulate SULR worldwide is vital for determining whether the new models have improved their performance.

Considerable efforts have been made to evaluate surface downward longwave radiation (SDLR) simulations from GCMs with ground measurements [35,38–40], while few studies have assessed the performance of GCMs to simulate SULR. In earlier analyses, Garratt [18] validated four GCMs in simulating annual SULR at 22 inland sites with an average bias of  $-11 \text{ W m}^{-2}$ . In addition to the ground observations, satellite estimates are also used to validate the GCMs. Although comparisons of surface downward shortwave radiation (SDSR) in GCMs against satellite observations have been performed globally and in specific regions [41–43], evaluations of SULR simulations from GCMs using satellite-derived SULR products are rare. Positive biases of the CMIP5 GCMs SULR against the Clouds and the Earth's Radiant Energy System, Energy Balanced and Filled (CERES EBAF) within Northern Eurasia and North America during the winter season (DJF) were found by Li et al. [44]. With respect to CERES EBAF, the SULR simulations from the Coupled Earth System Model version 1 with the Coupled Atmosphere Model version 5 (CESM1–CAM5) were underestimated for each calendar month in a range of 58–70°S over the ocean [45]. Overall, whether the SULR simulated by CMIP6 GCMs agrees well with ground and satellite observations has not been investigated comprehensively.

Due to the complexity of the Earth's climate system as well as the lack of a full understanding of it, our limited computation abilities, and often sparse observational sites, SULR simulations obtained from individual GCMs are subject to large uncertainties [46]. To reduce these structural uncertainties, the multimodel ensemble (MME) method was developed; this method merges separate GCMs, such as the Bayesian model averaging (BMA) method and simple model averaging (SMA) method [46,47]. It has been widely utilized in climatic and hydrologic variable estimations [48–50], and previous studies indicated that the uncertainties associated with individual GCMs can be decreased by using the MME method [48,50–53]. For example, Zhang et al. [51] reported that the estimated SDSR from 48 CMIP5 GCMs using the BMA method exhibited lower root mean squared error (RMSE) and bias values than the SMA method and individual GCMs. Yao et al. [50] applied the BMA method to merge latent heat flux simulations of CMIP5 GCMs, and found that the RMSE of the BMA results decreased by more than  $3 \text{ W m}^{-2}$  compared to the SMA results and individual GCMs.

Therefore, the purpose of this study is to assess the ability and applicability of MME methods in the estimation of SULR based on CMIP6 GCMs. This research first evaluated the performance of 51 CMIP6 GCMs in simulating SULR with ground measurements and examined how well the MME methods performed in estimating SULR with CMIP6 GCMs. Second, the SULR simulations from CMIP6 GCMs were compared to their counterparts from CMIP5 GCMs and CERES EBAF. Finally, we detected the long-term global mean SULR variability based on CMIP6 GCMs.

## 2. Data

### 2.1. CMIP5 and CMIP6 GCMs

Monthly SULR simulations from 45 CMIP5 GCMs and 51 CMIP6 GCMs, which are available at the Earth System Grid Federation (ESGF), were applied in this research. Detailed descriptions of the GCMs are summarized in Tables 1 and 2. The SULR simulations over the period of 1861–2005 from 45 CMIP5 GCMs under ensemble “r1i1p1” employed “historical” experiments (<https://esgf-node.llnl.gov/search/cmip5/>, accessed on 11 October 2021). As of March 2021, historical simulations on monthly SULR performed by 51 GCMs under ensemble “r1i1p1f1” during 1850–2014 were available from CMIP6 (<https://esgf-node.llnl.gov/search/cmip6/>, accessed on 11 October 2021). The “histori-

cal" experiments considered major anthropogenic and natural forcings, such as aerosol loadings, greenhouse gases, solar radiation, and land use [29]. All 96 GCMs from CMIP6 and CMIP5 were interpolated to  $1^\circ \times 1^\circ$  grid resolution with a bilinear interpolation algorithm before validation and comparison. GCMs have been interpolated to  $1^\circ \times 1^\circ$  resolution for comparison in many studies [26,31,50,51,54]. For example, Yao et al. [50] resampled the CMIP5 GCMs to  $1^\circ \times 1^\circ$  by the bilinear interpolation method before evaluating and merging GCMs with varying spatial resolutions. Zhang et al. [51] interpolated the CMIP5 GCMs to the same resolution as the CERES EBAF of  $1^\circ \times 1^\circ$  using bilinear interpolation. In a recent study, CMIP6 GCMs were interpolated to a resolution of  $1^\circ \times 1^\circ$  with bilinear interpolation before comparison [31]. Generally, bilinear interpolation is a commonly used method for regridding data from lower resolutions to higher resolutions without impacting climate signals [27].

**Table 1.** Detailed information on the 5th phase of the Coupled Model Intercomparison Project (CMIP5) general circulation models (GCMs) used in this study.

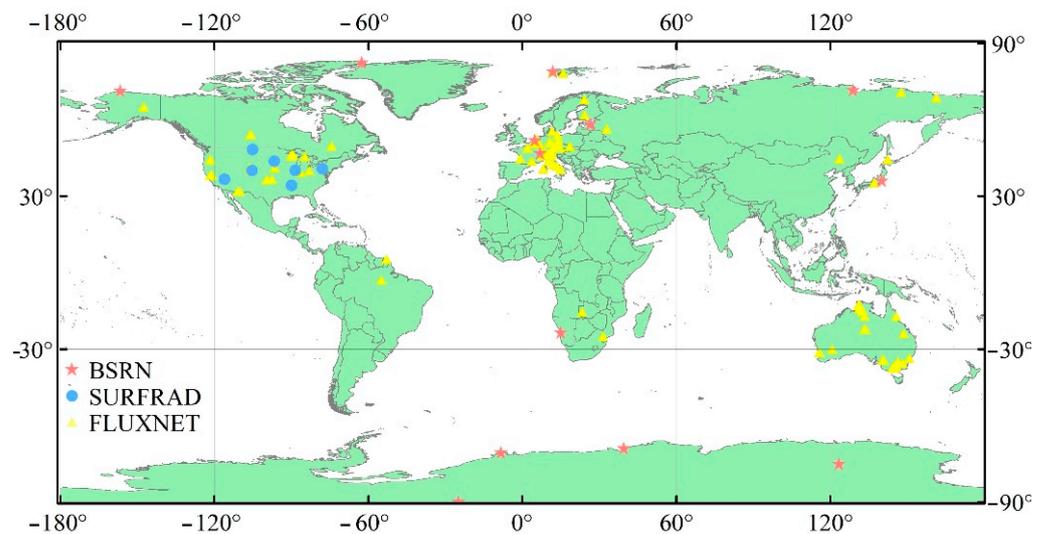
ID	Model Name	Institute ID	Time	Resolution
1	ACCESS1-0	CSIRO-BOM	185001–200512	$1.88^\circ \times 1.24^\circ$
2	ACCESS1-3	CSIRO-BOM	185001–200512	$1.88^\circ \times 1.24^\circ$
3	BNU-ESM	GCESS	185001–200512	$2.81^\circ \times 2.81^\circ$
4	CCSM4	NCAR	185001–200512	$1.25^\circ \times 0.94^\circ$
5	CESM1-BGC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

**Table 2.** Detailed information on the CMIP6 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	185001–201412	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	CMCC-ESM2	CMCC	185001–201412	1.25° × 0.94°
17	CanESM5	CCCma	185001–201412	2.81° × 2.81°
18	E3SM-1-0	E3SM-Project	185001–201412	1.00° × 1.00°
19	E3SM-1-1-ECA	E3SM-Project	185001–201412	1.00° × 1.00°
20	E3SM-1-1	E3SM-Project	185001–201412	1.00° × 1.00°
21	EC-Earth3-AerChem	EC-Earth-Consortium	185001–201412	0.70° × 0.70°
22	EC-Earth3-CC	EC-Earth-Consortium	185001–201412	0.70° × 0.70°
23	EC-Earth3-Veg-LR	EC-Earth-Consortium	185001–201412	1.13° × 1.13°
24	EC-Earth3-Veg	EC-Earth-Consortium	185001–201412	0.70° × 0.70°
25	EC-Earth3	EC-Earth-Consortium	185001–201412	0.70° × 0.70°
26	FGOALS-f3-L	CAS	185001–201412	1.25° × 1.00°
27	FGOALS-g3	CAS	185001–201612	2.00° × 2.25°
28	FIO-ESM-2-0	FIO-QLNM	185001–201412	1.25° × 0.94°
29	GFDL-ESM4	NOAA-GFDL	185001–201412	1.25° × 1.00°
30	GISS-E2-1-G-CC	NASA-GISS	185001–201412	2.50° × 2.00°
31	GISS-E2-1-G	NASA-GISS	185001–201412	2.50° × 2.00°
32	GISS-E2-1-H	NASA-GISS	185001–201412	2.50° × 2.00°
33	IITM-ESM	CCCR-IITM	185001–201412	1.88° × 1.91°
34	INM-CM4-8	INM	185001–201412	2.00° × 1.50°
35	INM-CM5-0	INM	185001–201412	2.00° × 1.50°
36	IPSL-CM5A2-INCA	IPSL	185001–201412	3.75° × 1.88°
37	IPSL-CM6A-LR-INCA	IPSL	185001–201412	2.50° × 1.26°
38	IPSL-CM6A-LR	IPSL	185001–201412	2.50° × 1.26°
39	KACE-1-0-G	NIMS-KMA	185001–201412	1.88° × 1.25°
40	KIOST-ESM	KIOST	185001–201412	1.88° × 1.88°
41	MIROC6	MIROC	185001–201412	1.41° × 1.41°
42	MPI-ESM-1-2-HAM	HAMMOZ-Consortium	185001–201412	1.88° × 1.88°
43	MPI-ESM1-2-HR	MPI-M	185001–201412	0.94° × 0.94°
44	MPI-ESM1-2-LR	MPI-M	185001–201412	1.88° × 1.88°
45	MRI-ESM2-0	MRI	185001–201412	1.13° × 1.13°
46	NESM3	NUIST	185001–201412	1.88° × 1.88°
47	NorCPM1	NCC	185001–202912	2.50° × 1.88°
48	NorESM2-LM	NCC	185001–201412	2.50° × 1.88°
49	NorESM2-MM	NCC	185001–201412	1.25° × 0.94°
50	SAM0-UNICON	SNU	185001–201412	1.25° × 0.94°
51	TaiESM1	AS-RCEC	185001–201412	1.25° × 0.94°

## 2.2. Ground-Measured Data

The ground-measured SULR data used in the present research were obtained from 133 sites across the world, including 7 sites from the Surface Radiation Budget Network (SURFRAD) [17], 13 sites from the Baseline Surface Radiation Network (BSRN) [16] and 113 sites from FLUXNET2015 (FLUXNET) [55]. The BSRN, SURFRAD, and FLUXNET sites record SULR measurements every 1, 3, and 30 min, respectively [16,56]. These stations cover a wide latitudinal range between 82.490°N and 89.983°S and a longitudinal range between 156.607°W and 161.341°E, with elevations in the range of −9–3233 m. The sites are spread across separate climatic zones and encompass cropland, forest, grassland, desert, bare land, and other land types. The sites applied in this study, which cover the period from 1992 to 2014, are exhibited in Appendix A Table A1. Figure 1 shows the locations of all 133 sites.



**Figure 1.** Geographical distribution of the observation sites.

## 2.3. ERA5

ERA5, which was released by the European Centre for Medium-Range Weather Forecasts (ECMWF) in 2017, is the newest version of the reanalysis for global climate [57]. It was produced to replace ERA-Interim, with data ranging from 1979 to the present [58]. In addition to the improved spatial resolution in comparison to ERA-Interim, ERA5 includes many more ground measurements utilized for assimilation [30]. The SULR data applied in the present research included monthly estimations at single levels with a spatial resolution of  $0.25^\circ \times 0.25^\circ$  [59] from 1979 to 2014 (<https://cds.climate.copernicus.eu/>, accessed on 11 October 2021), which were resampled to a  $1^\circ \times 1^\circ$  spatial resolution before the comparison between GCMs and ERA5.

## 2.4. CERES EBAF

CERES EBAF is conducted by the National Aeronautics and Space Administration (NASA) for application in radiation budget estimation and climate model assessment, with data available from 2000 to the present [22,60]. It utilizes data from the Aqua, Terra, Suomi National Polar-orbiting Partnership (S-NPP) [61,62] and Joint Polar Satellite System 1 (JPSS-1) satellites [63] and merges the satellite data using the radiative transfer model to estimate the surface energy budget [64]. The CERES EBAF uses more accurate cloud information and is considered one of the most qualified gridded products [60]. Fifteen years (2000–2014) of monthly mean SULR data of the CERES EBAF Ed4.1 dataset with a spatial resolution of  $1^\circ \times 1^\circ$  [23] were applied in this research (<https://ceres.larc.nasa.gov/>, accessed on 11 October 2021).

### 3. Methods

#### 3.1. Multimodel Ensemble (MME) Methods

In the present research, two MME approaches were utilized to merge individual CMIP6 GCM SULR simulations. In the SMA method, the weights of all GCMs were equal and set to  $1/K$ , where  $K$  was the number of the models. This method provides SULR estimations through the weighted averaging of individual GCMs. The BMA method is a postprocessing approach that estimates SULR from multiple models [65]. The weights represent the models' predictive abilities during the training period, which are determined by the posterior probabilities of individual models [66]. These probabilities can be calculated by maximizing the log-likelihood function through the expectation-maximization (EM) algorithm [66,67]. The posterior probabilities of all participating models are nonnegative values that sum to one [52]. The SULR values estimated based on the BMA method are obtained through the weighted averaging of the multiple bias-corrected models, which can be computed by the following equation. Detailed descriptions of the BMA method were introduced by Raftery et al. [47]. In the present research, SULR measurements from 1992 to 2014 at 133 sites globally distributed in the world were utilized for BMA analysis.

$$Y = \sum_{i=1}^K w_i(a_i + b_i y_i) \quad (1)$$

where  $y_i$  and  $Y$  are the simulated SULR values from individual GCMs and estimated SULR values using the BMA method, respectively.  $w_i$  denotes the BMA weights for the individual models.  $a_i$  and  $b_i$  are coefficients calculated through linear regression of observed and simulated SULR values in the bias correction procedure.

#### 3.2. Validation Metrics

The statistical metrics applied in this research include the mean bias error (bias), root mean square error (RMSE), relative mean bias error (Rbias), relative root mean square error (RRMSE), and correlation coefficient ( $R$ ) values, which are expressed as the following equations:

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

$$Bias = \frac{1}{n} \sum_{i=1}^n (e_i - o_i) \quad (3)$$

$$RRMSE = \frac{100}{\bar{o}} \times \sqrt{\frac{1}{n} \sum_{i=1}^n (e_i - o_i)^2} \quad (4)$$

$$RBias = \frac{100}{\bar{o}} \times \frac{1}{n} \sum_{i=1}^n (e_i - o_i) \quad (5)$$

$$R = \frac{n \sum_{i=1}^n e_i o_i - \sum_{i=1}^n e_i \times \sum_{i=1}^n o_i}{\sqrt{\left[ n \sum_{i=1}^n e_i^2 - \left( \sum_{i=1}^n e_i \right)^2 \right] \times \left[ n \sum_{i=1}^n o_i^2 - \left( \sum_{i=1}^n o_i \right)^2 \right]}} \quad (6)$$

where  $e_i$  and  $o_i$  denote the simulated and observed SULR values, respectively,  $\bar{o}$  represents the average observed SULR values, and  $n$  indicates the amount of data.

Although five statistical metrics were used to validate the SULR simulations from CMIP5 and CMIP6 GCMs, no individual metric can denote the overall accuracy. Thus, the

global performance indicator (GPI) was utilized in the present research to evaluate the overall model accuracy, which is calculated as follows [68]:

$$GPI_k = \sum_{j=1}^m X_j (\tilde{Y}_j - Y_{kj}) \quad (7)$$

where  $\tilde{Y}_j$  represents the median of the absolute values for metric  $j$ ,  $Y_{kj}$  represents the absolute value for metric  $j$  of model  $k$ , and  $m$  denotes the number of metrics.  $X_j$  is set to  $-1$  for R and set to 1 for other metrics. A higher GPI value indicates a better ability of model  $k$ .

#### 4. Results and Analysis

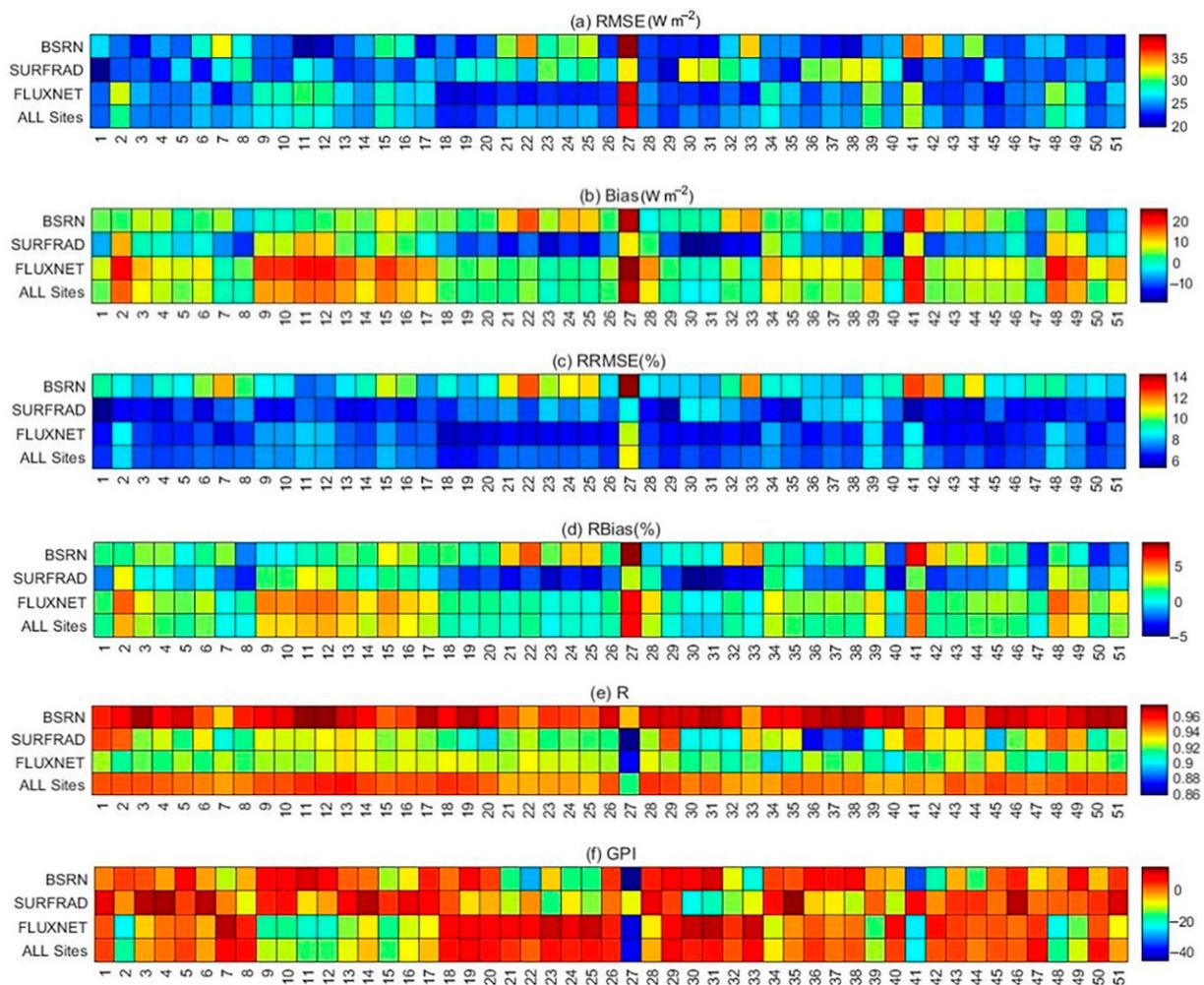
This study evaluated CMIP6 and CMIP5 model performances with ground-measured SULR as reference data. The GCM gridded SULR simulations were compared with respect to the SULR measurements within the grids. Since both the CMIP6 and CMIP5 GCMs utilized in the present research had different spatial resolutions, we regridded all GCMs to  $1^\circ \times 1^\circ$  using the bilinear interpolation method for a fair comparison.

##### 4.1. Evaluation with Ground Measurements

###### 4.1.1. CMIP6 GCMs SULR Evaluation

SULR observations (10,598 samples) collected at 133 sites from 1992 to 2014 were utilized to assess the ability of 51 CMIP6 GCMs to simulate SULR. Figure 2 illustrates the statistical metrics calculated for all 51 CMIP6 GCMs at the sites from SURFRAD, BSRN, and FLUXNET. The RMSE (RRMSE) values at seven sites from the SURFRAD for the SULR simulations varied from 19 to 33  $W m^{-2}$  (5.1% to 8.6%), and the average RMSE (RRMSE) value amounted to 26  $W m^{-2}$  (6.9%). Twenty-three out of the 51 GCMs showed RMSE (RRMSE) values within 25  $W m^{-2}$  (6.6%). The biases (Rbias) at seven SURFRAD sites for the individual GCMs ranged from  $-20$  to 13  $W m^{-2}$  ( $-5.2\%$  to  $3.3\%$ ). The SULR simulations were underestimated by 36 GCMs, with an average bias (Rbias) of  $-4 W m^{-2}$  ( $-1.1\%$ ). The SULR simulations had R values ranging between 0.85 and 0.96 at the SURFRAD sites, and the average R value was 0.92. The R values exceeded 0.93 at 15 GCMs. With respect to the SURFRAD sites, the majority of the GCMs (35 GCMs) showed GPI values above  $-5$ , with an average GPI value of  $-0.5$ . Among the 51 CMIP6 GCMs, INM-CM5-0 was the best model at the SURFRAD sites in comparison with other individual GCMs, with a bias of  $-0.19 W m^{-2}$ , an RMSE of 22.14  $W m^{-2}$ , an R of 0.927, and a maximum GPI value of 14.526. The GISS-E2-1-G-CC SULR simulations exhibited the poorest ability at seven SURFRAD sites, showing a bias of  $-19.60 W m^{-2}$ , an RMSE of 32.14  $W m^{-2}$ , an R of 0.901, and a minimum GPI value of  $-22.703$ .

We also evaluated the SULR simulations from 51 CMIP6 GCMs using 13 BSRN sites within twenty-three years (1992–2014) of SULR records. The CMIP6 GCMs had RMSE (RRMSE) values varying from 19 to 40  $W m^{-2}$  (7.0% to 14.2%) with respect to 13 BSRN sites, with an average RMSE (RRMSE) value of 26  $W m^{-2}$  (9.2%). The RMSE (RRMSE) values for 28 out of the 51 GCMs were less than 25  $W m^{-2}$  (8.9%). The biases (Rbias) for the individual GCMs ranged between  $-10$  and 24  $W m^{-2}$  ( $-3.3\%$  and  $8.5\%$ ). SULR was overestimated by 41 out of the 51 GCMs, and the average bias (Rbias) was 4  $W m^{-2}$  (1.5%). The R values of the individual GCMs at 13 BSRN sites ranged from 0.93 to 0.98, with an average R value of 0.96; this value was higher than that obtained at the SURFRAD sites. All 51 GCMs had R values greater than 0.93. The GPI values at 13 sites from the BSRN were higher than  $-5$  at 36 out of the 51 GCMs, and the average GPI value amounted to  $-3$ . CESM2-WACCM showed the best ability, with a bias of 1.70  $W m^{-2}$ , an RMSE of 19.94  $W m^{-2}$ , an R of 0.974, and a maximum GPI value of 10.571 for the 13 BSRN sites. FGOALS-g3 demonstrated the poorest performance in simulating SULR at the BSRN sites, exhibiting a bias of 23.71  $W m^{-2}$ , an RMSE of 39.97  $W m^{-2}$ , an R of 0.936, and a minimum GPI value of  $-46.414$ , followed by MIROC6.



**Figure 2.** Diagrams of the statistical metrics (root mean squared error (RMSE) (a), mean bias error (Bias) (b), relative root mean squared error (RRMSE) (c), relative mean bias error (RBias) (d), correlation coefficient (R) (e), and global performance indicator (GPI) (f)) of the comparison between the surface upward longwave radiation (SULR) simulations from the 51 CMIP6 GCMs and the SULR measurements for the Baseline Surface Radiation Network (BSRN), Surface Radiation Budget Network (SURFRAD), FLUXNET2015 (FLUXNET), and all 133 sites.

To further validate the effects of the site selection and observation quality on the GCMs in SULR, the above research was also repeated at 120 sites from FLUXNET. The RMSE (RRMSE) values of the SULR simulations at the 120 sites varied between 21 and 39  $\text{W m}^{-2}$  (5.7% and 10.4%), and the average RMSE (RRMSE) value was 25  $\text{W m}^{-2}$  (6.8%). Thirty out of the 51 GCMs had RMSE (RRMSE) values within 25  $\text{W m}^{-2}$  (6.8%). The biases (RBias) averaged over 120 FLUXNET sites ranged from  $-1$  to 26  $\text{W m}^{-2}$  ( $-0.2\%$  to 7.0%). The SULR simulations were overestimated by 48 GCMs, with an absolute average bias (RBias) of 9  $\text{W m}^{-2}$  (2.5%); this bias value was higher than those obtained at the sites for BSRN and SURFRAD. This may be due to the lower temporal resolution of SULR measurements at the FLUXNET sites [55]. The R values were in the range of 0.87–0.94 at the FLUXNET sites, with an average R value of 0.92. Only three GCMs showed R values greater than 0.93. The GPI values for 32 out of the 51 GCMs were greater than  $-5$  at the FLUXNET sites, with an average GPI value of  $-3$ . CAMS-CSM1-0 agreed best with the FLUXNET measurements, exhibiting a bias of 0.15  $\text{W m}^{-2}$ , an RMSE of 21.92  $\text{W m}^{-2}$ , an R of 0.924, and a maximum GPI value of 12.588. In line with the validation results for BSRN sites, FGOALS-g3 performed the worst to simulate SULR at the FLUXNET sites, with a bias of 25.78  $\text{W m}^{-2}$ , an RMSE of 38.47  $\text{W m}^{-2}$ , an R of 0.87, and a minimum GPI value of  $-41.022$ , followed by MIROC6.

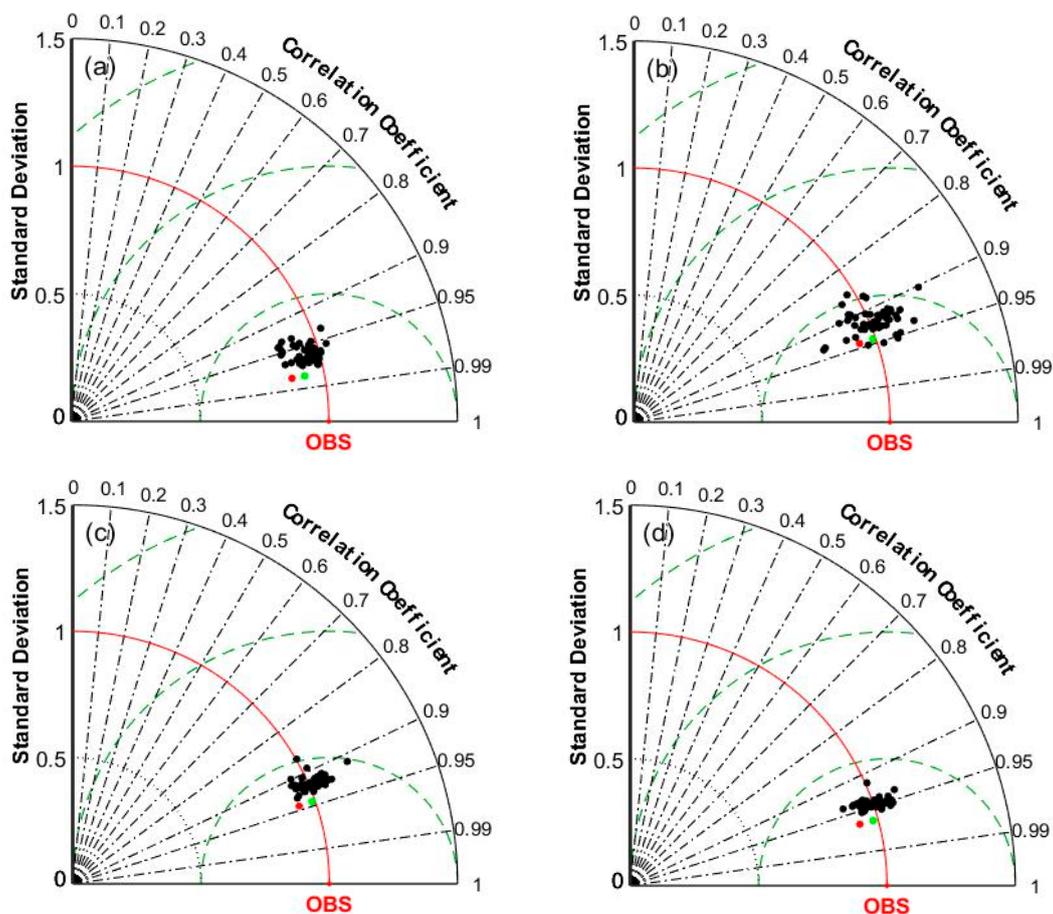
Figures 2 and 3 show the ability of the CMIP6 GCMs to simulate SULR at all 133 sites from SURFRAD, BSRN, and FLUXNET. The results indicated that the RMSE (RRMSE) values with respect to the 133 sites for the individual GCMs varied between 22 and 38  $W m^{-2}$  (6.3% and 10.8%), with an average RMSE (RRMSE) value of 26  $W m^{-2}$  (7.2%). Twenty-one out of the 51 GCMs displayed RMSE (RRMSE) values within 25  $W m^{-2}$  (7.1%). The biases (Rbias) of the various GCMs at 133 sites varied from  $-3$  to 24  $W m^{-2}$  ( $-0.8\%$  to 6.6%). Forty-six out of the 51 CMIP6 GCMs overestimated the SULR. Overall, the CMIP6 GCMs showed an obvious tendency to overestimate the SULR at the selected 133 sites, which had an average bias (Rbias) of 6  $W m^{-2}$  (1.8%). The CMIP6 GCMs exhibited R values between 0.91 and 0.96 at all 133 sites, and the average R value amounted to 0.95. Most GCMs (50 GCMs) had R values above 0.93. Thirty-five out of the 51 GCMs exhibited GPI values over  $-5$ , with an average GPI value of  $-3$ . Among the 51 CMIP6 GCMs, the maximum GPI value was found in E3SM-1-1-ECA at values of 8.383, which had a higher spatial resolution. FGOALS-g3, with a lower spatial resolution, exhibited the minimum GPI value of  $-39.443$ , followed by MIROC6. However, the CMIP6 GCMs, which had higher spatial resolutions, did not always show greater GPI values than those that had lower spatial resolutions, such as CESM2, CESM2-WACCM, and CMCC-CM2-SR5.

We also applied the MME method to estimate SULR by fusing the 51 CMIP6 GCMs. The weights obtained by the BMA method for individual GCMs, exhibited in Appendix B Figure A1, varied between 0.017 and 0.022, and 35 out of the 51 GCMs displayed weights in the range of 0.018–0.020. MRI-ESM2-0, with a maximum weight of 0.0215, which was approximately 10% higher than the mean value (0.0196), greatly contributed to the SULR ensemble. The weight of KACE-1-0-G was 0.0173, which was approximately 12% lower than the mean value.

A Taylor diagram [69], which is suitable for evaluating the relative abilities of multiple models, was utilized to validate the performances of the 51 CMIP6 GCMs and the MME methods in simulating SULR compared to the ground-measured SULR. The comparison results between the SULR estimations and the SULR measurements at the BSRN, SURFRAD, FLUXNET, and all sites are shown in Figure 4, indicating that the MME methods performed better than the individual GCMs in general. The validation results of the SULR estimated by the MME methods at all 133 sites are also displayed in Figure 3. Notably, the SULR estimations obtained based on the MME methods showed higher R, lower bias (Rbias), and lower RMSE (RRMSE) values in comparison with the individual GCMs at all 133 sites. The MME methods reduced the RMSE (RRMSE) by approximately 5–6  $W m^{-2}$  (1.5–1.8%) and increased the R by approximately 0.02 on average compared to the individual GCMs. The BMA method showed a better ability to simulate SULR than the individual GCMs and the SMA method, with a bias of 0  $W m^{-2}$ , an RMSE of 19.29  $W m^{-2}$ , and an R of 0.97 for the 133 sites, owing to the bias correction before weighting the multiple models to closely match the SULR measurements.



**Figure 3.** Scatterplots of the monthly SULR observations from 133 sites and the corresponding SULR simulations from the 51 CMIP6 GCMs and the multimodel ensemble (MME) methods (in units of  $W m^{-2}$ ).

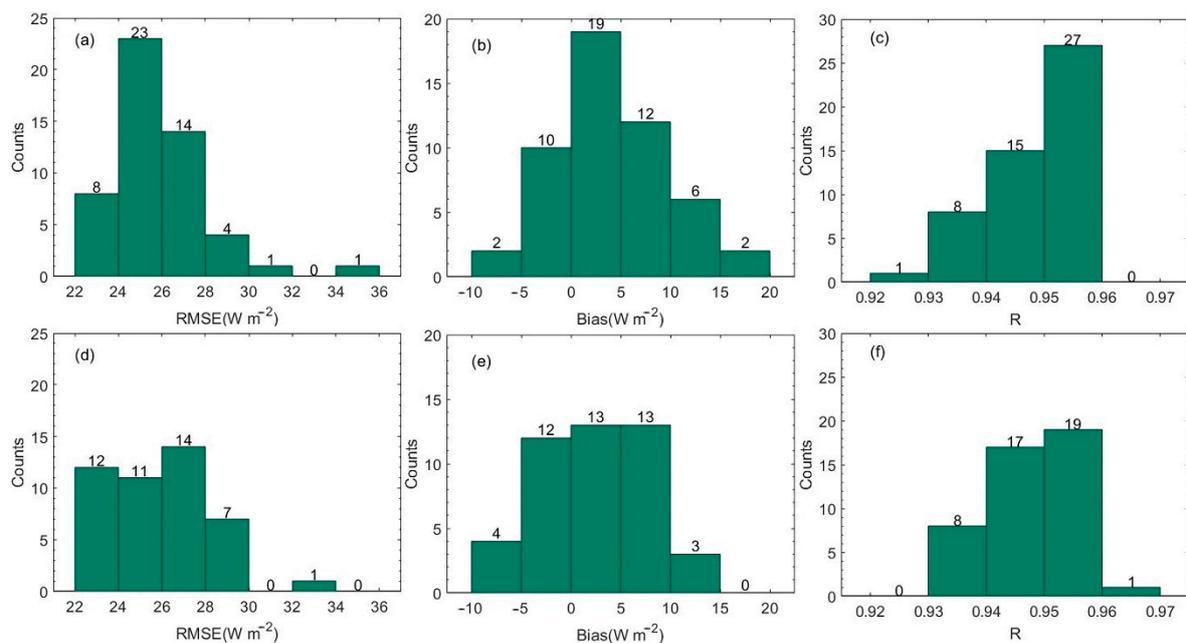


**Figure 4.** Taylor diagrams of the monthly SULR measurements and MME estimates for BSRN (a), SURFRAD (b), FLUXNET (c), and all 133 sites (d). The 51 CMIP6 GCMs, the simple model averaging (SMA) method, and the BMA method are represented by black, green, and red dots, respectively.

#### 4.1.2. Comparison with CMIP5

Most simulated SULR from CMIP6 and CMIP5 GCMs cover the periods 1850–2014 and 1861–2005, respectively, whereas the ground-measured SULR begins in 1992. There are only 61 sites with ground-measured SULR from 1992 to 2005. Therefore, SULR measurements (3052 samples) during 1992–2005 at 61 sites were applied to compare the ability of the 51 CMIP6 GCMs to simulate SULR with that of the 45 CMIP5 GCMs. The results illustrated that the CMIP5 GCMs showed an average bias ( $R_{bias}$ ) of  $2 \text{ W m}^{-2}$  (0.7%), an average RMSE (RRMSE) of  $26 \text{ W m}^{-2}$  (7.6%), and an average  $R$  of 0.95. For the CMIP6 GCMs, these values were  $4 \text{ W m}^{-2}$  (1.2%),  $26 \text{ W m}^{-2}$  (7.6%), and 0.95, respectively, and did not show better performance than the CMIP5 GCMs. Figure 5 shows the histogram of the statistical metrics calculated for the 45 CMIP5 GCMs and the 51 CMIP6 GCMs at all 61 sites. The CMIP5 GCMs exhibited RMSE values from 22 to  $34 \text{ W m}^{-2}$ . The RMSE values for 11 out of the 45 CMIP5 GCMs varied from 24 to  $26 \text{ W m}^{-2}$ . Twelve CMIP5 GCMs had RMSE values less than  $24 \text{ W m}^{-2}$ . The RMSE values were greater than  $26 \text{ W m}^{-2}$  at 22 CMIP5 GCMs. The RMSE values for the individual CMIP6 GCMs ranged from 22 to  $36 \text{ W m}^{-2}$ . Twenty-three out of the 51 CMIP6 GCMs had RMSE values between 24 and  $26 \text{ W m}^{-2}$ . The RMSE values were within  $24 \text{ W m}^{-2}$  at eight CMIP6 GCMs. A total of 20 CMIP6 GCMs showed RMSE values above  $26 \text{ W m}^{-2}$ . The biases for the individual CMIP5 GCMs ranged between  $-10$  and  $15 \text{ W m}^{-2}$ , and the SULR simulations were overestimated by 29 out of the 45 CMIP5 GCMs. The absolute biases of 25 CMIP5 GCMs were within  $5 \text{ W m}^{-2}$ . Sixteen CMIP5 GCMs reported biases above  $5 \text{ W m}^{-2}$ . The biases for the four CMIP5 GCMs were within  $-5 \text{ W m}^{-2}$ . The biases for the individual CMIP6 GCMs varied from

−10 to 20  $W m^{-2}$  and were positive for 39 out of the 51 CMIP6 GCMs. Twenty-nine CMIP6 GCMs showed absolute biases of less than 5  $W m^{-2}$ . The biases for 20 CMIP6 GCMs were above 5  $W m^{-2}$ . Only two CMIP6 GCMs had biases within −5  $W m^{-2}$ . The R values of the various CMIP5 GCMs varied from 0.93 to 0.97. Nineteen out of the 45 CMIP5 GCMs showed R values between 0.95 and 0.96. The R values were less than 0.95 at 25 CMIP5 GCMs. Only one CMIP5 GCM exhibited an R value above 0.96. The CMIP6 GCMs showed R values in the range of 0.92–0.96. The R values for 27 out of the 51 CMIP6 GCMs ranged from 0.95 to 0.96. Twenty-four CMIP6 GCMs had R values of less than 0.95, and there were no individual CMIP6 GCMs with an R value greater than 0.96. Overall, the CMIP6 GCMs did not perform better in simulating SULR than the CMIP5 GCMs.



**Figure 5.** RMSE, bias, and R histograms for monthly SULR simulations from the 51 CMIP6 GCMs (a–c) and the 45 CMIP5 GCMs (d–f).

The SMA and BMA methods were also utilized to combine 45 CMIP5 GCMs to estimate SULR. The validation result of the SMA method based on the CMIP5 GCMs had a higher RMSE of 25.87  $W m^{-2}$  and lower R of 0.94 than the SMA result based on the CMIP6 GCMs, while the bias of the SULR estimated by the SMA method using the CMIP5 GCMs was approximately −1.91  $W m^{-2}$ , which was lower than that using the CMIP6 GCMs. The estimated SULR of CMIP5 GCMs using the BMA method showed a bias of 0, an RMSE of 19.24, and an R of 0.97; this result was close to that of the CMIP6 GCMs obtained by the BMA method. Compared to the SMA result of the CMIP5 GCMs, the BMA method still exhibited better performance to estimate SULR by merging CMIP5 GCMs.

#### 4.2. Evaluation with CERES EBAF

The SULR simulations from the CMIP6 GCMs were validated with 133 sites from three observation networks across the world in Section 4.1.1; however, many oceanic and land surface areas still lack representativeness. The CERES EBAF, which is considered one of the most qualified gridded products [60], was utilized in this study. Compared to the CMIP6 GCMs, it agreed better with the SULR measurements at the 133 sites during 2000–2014, with a bias of 0.04  $W m^{-2}$ , RMSE of 16.77  $W m^{-2}$ , R of 0.972, and GPI of 17.733. Therefore, we validated the monthly SULR simulations from the 51 CMIP6 GCMs using the CERES EBAF SULR from 2000 to 2014 at  $1^\circ \times 1^\circ$  grid scale, as shown in Table 3. The start year of 2000 was chosen in the present research since the SULR data of the CERES EBAF only began in 2000. The RMSE (RRMSE) values for the various GCMs in comparison

with the CERES EBAF ranged between 14 and 31  $W m^{-2}$  (3.9% and 8.5%), with an average RMSE (RRMSE) value of 17  $W m^{-2}$  (4.8%). Fifty out of the 51 GCMs exhibited RMSE (RRMSE) values of less than 25  $W m^{-2}$  (6.1%), except for FGOALS-g3. The biases (Rbias) for the individual GCMs varied from  $-9$  to 11  $W m^{-2}$  ( $-2.4\%$  to 3.1%). The majority of the CMIP6 GCMs (39 GCMs) overestimated the monthly SULR, with an average bias (Rbias) of 2  $W m^{-2}$  (0.6%). The RMSE and biases for the GCMs with respect to the CERES EBAF were lower than those between the GCMs and ground measurements. The CMIP6 GCMs showed R values in the range of 0.96–0.99, with an average R value of up to 0.99. In addition to FGOALS-g3, almost all of the GCMs (50 GCMs) exhibited R values greater than 0.97. Forty out of the 51 GCMs had GPI values above  $-5$ , and the average GPI amounted to  $-2$ . Among the 51 CMIP6 GCMs, FGOALS-g3 showed the poorest ability to simulate SULR, with a bias of 10.74  $W m^{-2}$ , an RMSE of 30.17  $W m^{-2}$ , an R of 0.959, and a minimum GPI value of  $-28.348$  in comparison with the CERES EBAF. This result was similar to the evaluation result exhibited in Section 4.1.1.

**Table 3.** Evaluation of global monthly SULR simulations from the 51 CMIP6 GCMs using the Clouds and the Earth’s Radiant Energy System, Energy Balanced and Filled (CERES EBAF) data from 2000 to 2014 at  $1^\circ \times 1^\circ$  grid scale (in units of  $W m^{-2}$  for RMSE and Bias, % for RRMSE and RBias, respectively).

ID	Model Name	RMSE	Bias	RRMSE	RBias	R	GPI
1	ACCESS-CM2	15.39	0.48	4.33	0.14	0.988	3.76
2	ACCESS-ESM1-5	15.60	4.18	4.39	1.17	0.989	$-1.24$
3	AWI-CM-1-1-MR	14.18	2.32	3.99	0.65	0.990	2.96
4	AWI-ESM-1-1-LR	15.64	$-0.50$	4.40	$-0.14$	0.988	3.41
5	BCC-CSM2-MR	16.04	2.96	4.51	0.83	0.987	$-0.26$
6	BCC-ESM1	18.11	3.13	5.09	0.88	0.984	$-3.13$
7	CAMS-CSM1-0	19.35	1.91	5.44	0.54	0.981	$-3.16$
8	CAS-ESM2-0	20.32	$-0.53$	5.71	$-0.15$	0.980	$-2.63$
9	CESM2-FV2	16.26	2.54	4.57	0.71	0.987	0.01
10	CESM2-WACCM-FV2	15.86	2.89	4.46	0.81	0.988	0.07
11	CESM2-WACCM	14.99	3.10	4.22	0.87	0.989	0.91
12	CESM2	15.30	4.02	4.30	1.13	0.989	$-0.66$
13	CIesm	17.17	6.40	4.83	1.80	0.987	$-6.11$
14	CMCC-CM2-HR4	14.80	2.96	4.16	0.83	0.989	1.35
15	CMCC-CM2-SR5	18.94	6.71	5.33	1.89	0.984	$-8.77$
16	CMCC-ESM2	17.57	5.15	4.94	1.45	0.986	$-5.01$
17	CanESM5	16.51	0.11	4.64	0.03	0.986	2.80
18	E3SM-1-0	16.38	2.94	4.61	0.83	0.987	$-0.66$
19	E3SM-1-1-ECA	16.36	1.07	4.60	0.30	0.987	1.75
20	E3SM-1-1	16.33	1.83	4.59	0.52	0.987	0.83
21	EC-Earth3-AerChem	17.41	5.15	4.90	1.45	0.986	$-4.82$
22	EC-Earth3-CC	19.49	8.75	5.48	2.46	0.985	$-12.09$
23	EC-Earth3-Veg-LR	18.12	2.81	5.09	0.79	0.984	$-2.72$
24	EC-Earth3-Veg	17.75	5.58	4.99	1.57	0.986	$-5.80$
25	EC-Earth3	18.10	5.53	5.09	1.56	0.985	$-6.19$
26	FGOALS-f3-L	16.49	$-1.24$	4.64	$-0.35$	0.986	1.38
27	FGOALS-g3	30.17	10.74	8.48	3.02	0.959	$-28.35$
28	FIO-ESM-2-0	14.27	$-1.31$	4.01	$-0.37$	0.990	4.13
29	GFDL-ESM4	14.08	0.85	3.96	0.24	0.990	4.96
30	GISS-E2-1-G-CC	16.99	1.04	4.78	0.29	0.986	1.00
31	GISS-E2-1-G	17.07	0.32	4.80	0.09	0.985	1.80
32	GISS-E2-1-H	17.82	5.33	5.01	1.50	0.985	$-5.57$
33	IITM-ESM	21.48	3.41	6.04	0.96	0.978	$-7.80$
34	INM-CM4-8	17.39	$-0.88$	4.89	$-0.25$	0.985	0.68
35	INM-CM5-0	15.32	$-1.64$	4.31	$-0.46$	0.988	2.36
36	IPSL-CM5A2-INCA	17.37	$-1.25$	4.89	$-0.35$	0.985	0.23

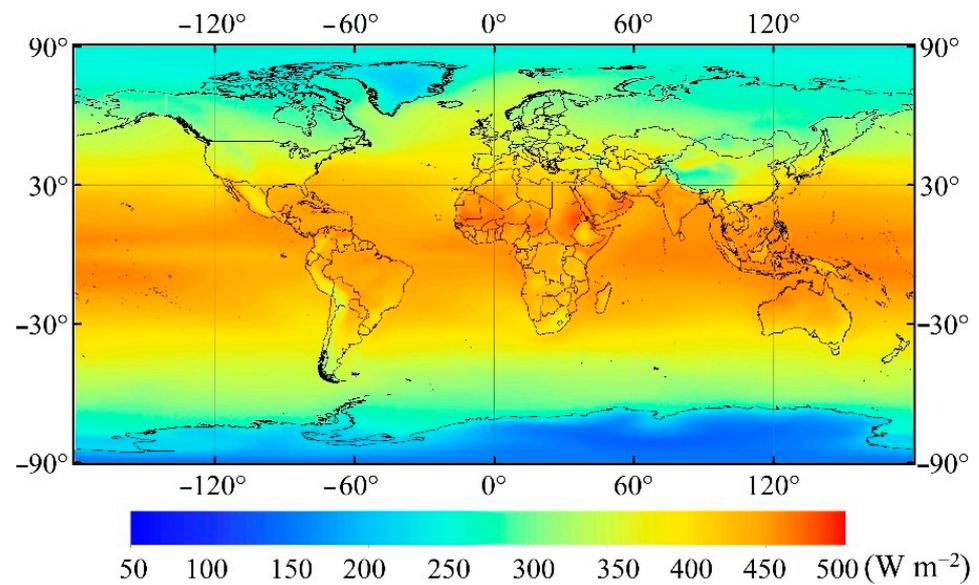
Table 3. Cont.

ID	Model Name	RMSE	Bias	RRMSE	RBias	R	GPI
37	IPSL-CM6A-LR-INCA	15.62	0.70	4.39	0.20	0.988	3.18
38	IPSL-CM6A-LR	15.29	−0.37	4.30	−0.10	0.988	4.04
39	KACE-1-0-G	16.05	2.32	4.51	0.65	0.987	0.56
40	KIOST-ESM	18.03	−1.40	5.07	−0.39	0.984	−0.80
41	MIROC6	21.07	7.62	5.92	2.14	0.981	−12.67
42	MPI-ESM1-2-HAM	17.59	0.30	4.95	0.08	0.985	1.17
43	MPI-ESM1-2-HR	14.64	2.31	4.12	0.65	0.989	2.38
44	MPI-ESM1-2-LR	16.13	1.59	4.54	0.45	0.987	1.38
45	MRI-ESM2-0	14.75	1.84	4.15	0.52	0.989	2.83
46	NESM3	16.95	1.11	4.77	0.31	0.985	0.95
47	NorCPM1	20.64	−8.39	5.80	−2.36	0.983	−13.11
48	NorESM2-LM	16.54	4.25	4.65	1.20	0.987	−2.55
49	NorESM2-MM	14.91	0.76	4.19	0.21	0.989	4.03
50	SAM0-UNICON	16.25	−5.23	4.57	−1.47	0.989	−3.43
51	TaiESM1	15.15	−1.26	4.26	−0.36	0.990	3.07

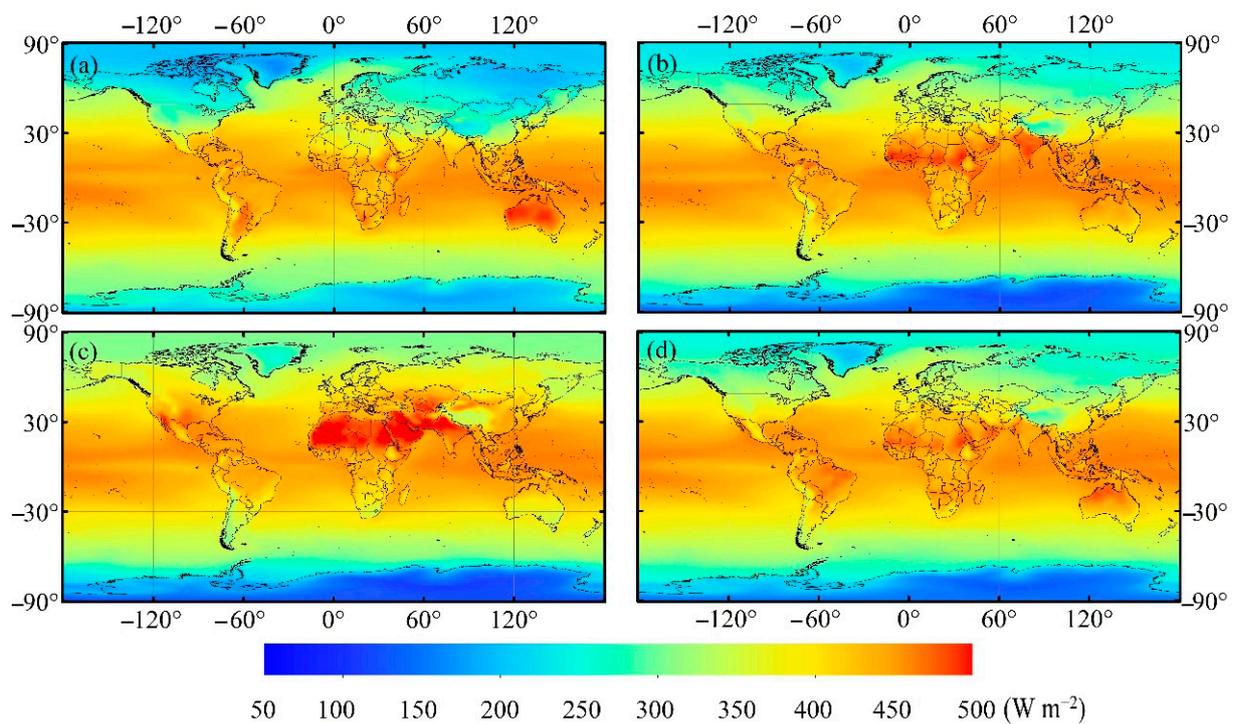
#### 4.3. Spatial Distribution and Seasonal Variations

For ERA5 and CERES EBAF, the monthly mean SULR estimations start in 1979 and 2000, respectively; in CMIP5 GCMs, most SULR simulations end in 2005. Thus, we applied the BMA method, which performed better in estimating SULR, to construct a gridded global SULR dataset ( $1^\circ \times 1^\circ$ ) from 2000 to 2005 based on CMIP5 and CMIP6 GCMs. Figures 6 and 7 exhibit the spatial variation in the SULR estimations from the CMIP6 GCMs based on the BMA method during 2000–2005 worldwide. The Arctic has always been an area with lower SULR values, equal to approximately  $160\text{--}220 \text{ W m}^{-2}$  in the local winter season (DJF) and  $250\text{--}320 \text{ W m}^{-2}$  in the local summer season (JJA). Antarctica was another area with lower SULR values, with values amounting to approximately  $100\text{--}210 \text{ W m}^{-2}$  in the local winter season (JJA) and  $180\text{--}280 \text{ W m}^{-2}$  in the local summer season (DJF). In addition to the ocean surface regions, SULR was higher in some land surface areas over tropical regions ( $440\text{--}490 \text{ W m}^{-2}$ ), such as the Amazon rainforest, Central Africa, Arabian Peninsula, Indian Peninsula, and North Australia (Figure 6); these findings were in line with other studies [9,20]. Generally, SULR gradually decreased from tropical regions to polar regions, ranging from  $130$  to  $490 \text{ W m}^{-2}$ . At the same latitude, SULR was lower in high-elevation areas than in low-elevation areas. For instance, SULR values as low as  $360\text{--}410 \text{ W m}^{-2}$  occurred in the Andes in the local summer season (DJF), while values as low as  $330\text{--}400 \text{ W m}^{-2}$  were observed in the local winter season (JJA). Corresponding features, with SULR values of  $330\text{--}400 \text{ W m}^{-2}$  in the local summer season (JJA) and  $220\text{--}290 \text{ W m}^{-2}$  in the local winter season (DJF), were found in the Qinghai–Tibet Plateau.

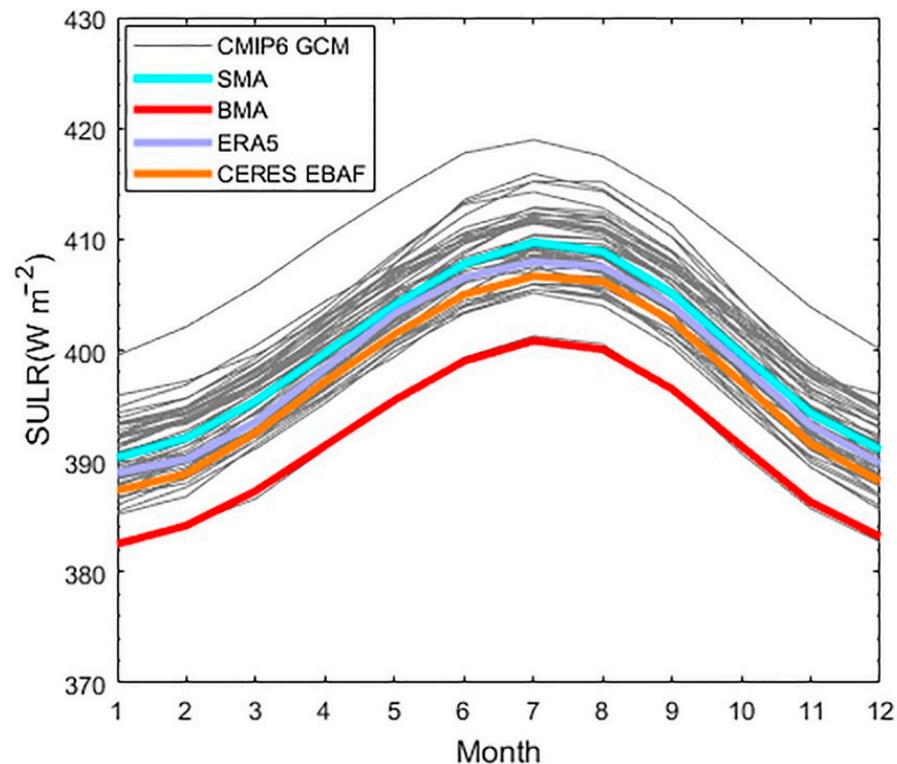
Figure 8 indicates the monthly variations in SULR from the CERES EBAF, ERA5, 51 CMIP6 GCMs, and MME methods during the period of 2000–2005 worldwide. The results exhibited the monthly mean SULR throughout the world ranging between  $382$  and  $419 \text{ W m}^{-2}$ , where the maximum SULR occurred in July and August ( $400\text{--}419 \text{ W m}^{-2}$ ), while the minimum SULR appeared in January and December ( $382\text{--}400 \text{ W m}^{-2}$ ). FGOALS-g3 had the highest global monthly mean SULR values, which varied from  $399$  to  $419 \text{ W m}^{-2}$ . The lowest SULR values, in the range of  $382\text{--}402 \text{ W m}^{-2}$ , were observed in NorCPM1. Large seasonal variations were found in MIROC6, with SULR values between  $392$  and  $416 \text{ W m}^{-2}$ , while GISS-E2-1-H showed small seasonal variations between  $395$  and  $412 \text{ W m}^{-2}$ . The maximum SULR in both the BMA and SMA methods was observed in July at values of  $401$  and  $410 \text{ W m}^{-2}$ , respectively, and the minimum SULR occurred in January at values of  $383$  and  $390 \text{ W m}^{-2}$ , respectively. Although the global monthly mean SULR obtained by the MME methods exhibited similar monthly variability to the SULR values estimated by the CERES EBAF and ERA5, the BMA results showed lower SULR values in comparison with the CERES EBAF, while the SMA results displayed higher values than the CERES EBAF.



**Figure 6.** Spatial variation in the SULR estimations from the CMIP6 GCMs using the BMA method from 2000 to 2005 (in units of  $\text{W m}^{-2}$ ).

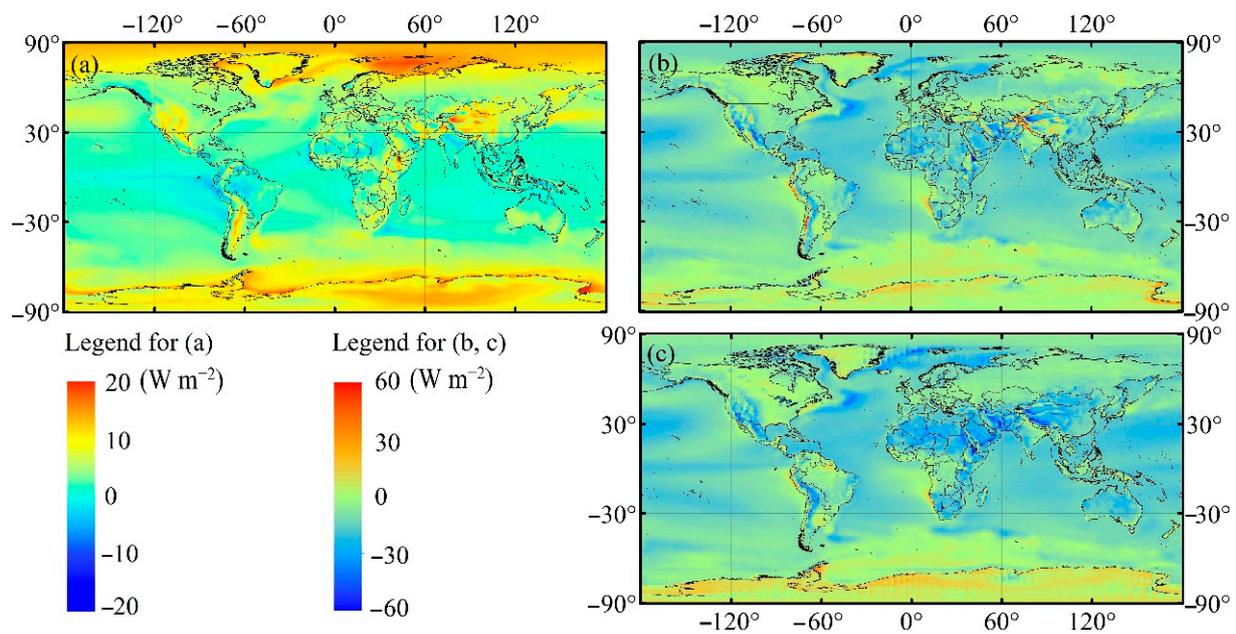


**Figure 7.** Spatial variations in the SULR estimations from the CMIP6 GCMs using the BMA method in the DJF (December, January, and February) (a), MAM (March, April, and May) (b), JJA (June, July, and August) (c), and SON (September, October, and November) seasons (d) from 2000 to 2005 (in units of  $\text{W m}^{-2}$ ).



**Figure 8.** Temporal variations in the SULR estimations of the 51 CMIP6 GCMs, the SMA method, the BMA method, the ERA5 and the CERES EBAF from 2000 to 2005 (in units of  $W m^{-2}$ ).

The spatial pattern of the biases between the CMIP6 and CMIP5 GCMs for the annual mean SULR estimated by the BMA method is illustrated in Figure 9a; these biases varied between  $-16$  and  $16 W m^{-2}$ . Positive biases occurred in the Arctic, Antarctica, Qinghai–Tibet Plateau, Sierra Nevada, Rocky Mountains, and Eastern Andes, while significant negative biases were found in the Western Andes. The spatial dissimilarities of the CMIP6 GCMs against ERA5 are shown in Figure 9b and ranged from  $-48$  to  $52 W m^{-2}$ . The SULR values of CMIP6 GCMs obtained by the BMA method were higher than ERA5 in the Himalayas, Hengduan Mountains, Qilian Mountains, Altun Mountains, Kunlun Mountains, Tianshan Mountains, and Western Andes, with a difference of more than  $10 W m^{-2}$ . In the Sierra Nevada and the Eastern Andes, the annual SULR of CMIP6 GCMs was obviously lower than that of ERA5, and the differences exceeded  $20 W m^{-2}$ . Figure 9c exhibits the spatial differences in the CMIP6 GCMs minus the CERES EBAF, which ranged between  $-59$  and  $45 W m^{-2}$ . The CMIP6 GCMs showed positive biases in Antarctica, Greenland, the Southern Himalayas, and the Western Andes, where the biases exceeded  $5 W m^{-2}$ . Clear negative biases above  $25 W m^{-2}$  were found in the Northern Himalayas, Sierra Nevada, and the Eastern Andes.



**Figure 9.** The biases of the SULR estimations between CMIP6 GCMs and CMIP5 GCMs (a), between CMIP6 GCMs and ERA5 (b), and between CMIP6 GCMs and CERES EBAF (c) from 2000 to 2005 (the SULR estimations from CMIP6 and CMIP5 GCMs were obtained by the BMA method, in units of  $\text{W m}^{-2}$ ).

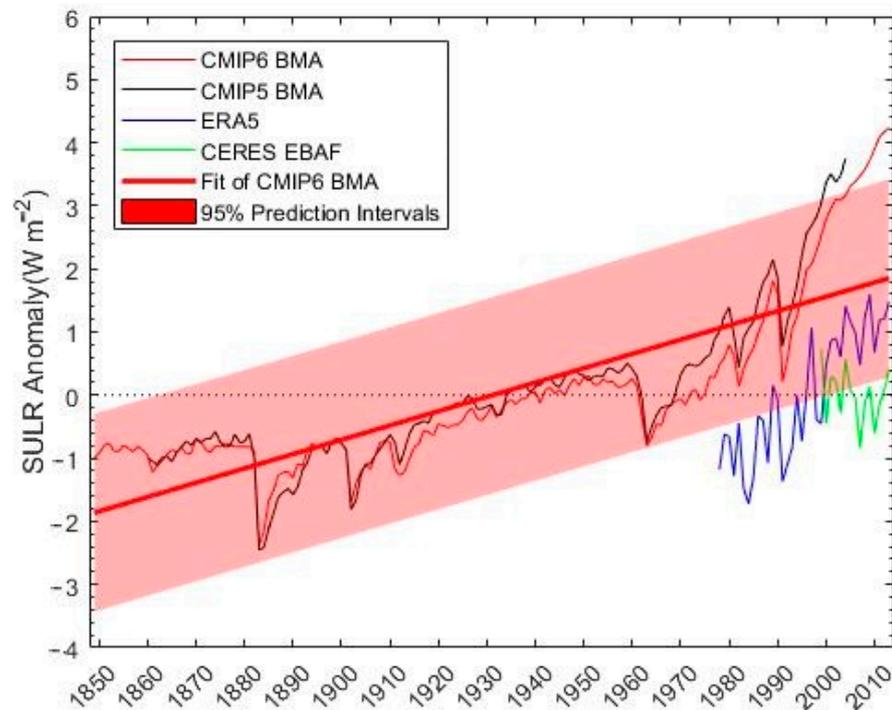
#### 4.4. Annual Mean and Long-Term Variabilities

Table 4 lists the global annual mean SULR for the CERES EBAF, ERA5, and the CMIP5 and CMIP6 GCMs based on the MME methods during different periods. The results showed that the global annual mean SULR for the CMIP6 GCMs based on the BMA method varied between  $386$  and  $393 \text{ W m}^{-2}$  from 1850 to 2014. The average SULR of  $389 \text{ W m}^{-2}$  and median SULR of  $388 \text{ W m}^{-2}$  were lower than the SMA results of the CMIP6 GCMs, owing to the overestimation of SULR in the CMIP6 GCMs. The CMIP5 GCMs had an annual mean SULR estimated by the BMA method between  $389$  and  $396 \text{ W m}^{-2}$  in 1861–2005. Both the average SULR value and the median SULR value were  $392 \text{ W m}^{-2}$ , which were higher than those from CMIP6 GCMs using the BMA method, and the differences were approximately  $4 \text{ W m}^{-2}$ . The SULR values from ERA5 varied from  $394$  to  $398 \text{ W m}^{-2}$  during 1979–2014, where the average SULR value and the median SULR value amounted to  $396 \text{ W m}^{-2}$ . Higher SULR values, varying between  $397$  and  $400 \text{ W m}^{-2}$  in 2000–2014, were found in the CERES EBAF, with an average SULR of  $399 \text{ W m}^{-2}$  and a median SULR of  $399 \text{ W m}^{-2}$ . The SULR values in both ERA5 and CERES EBAF were higher than those in the BMA results from CMIP6 GCMs and lower than those in the SMA results from CMIP6 GCMs; this may have been because the overestimation of SULR in the CMIP6 GCMs was significantly improved by the BMA method. Overall, the best global annual mean SULR estimation obtained by the CMIP6 GCMs using the BMA method was  $392 \text{ W m}^{-2}$  during 2000–2014.

**Table 4.** Comparison of the global annual mean SULR of the CMIP6 GCMs, CMIP5 GCMs, ERA5, and CERES EBAF over different periods (in units of  $W m^{-2}$ ).

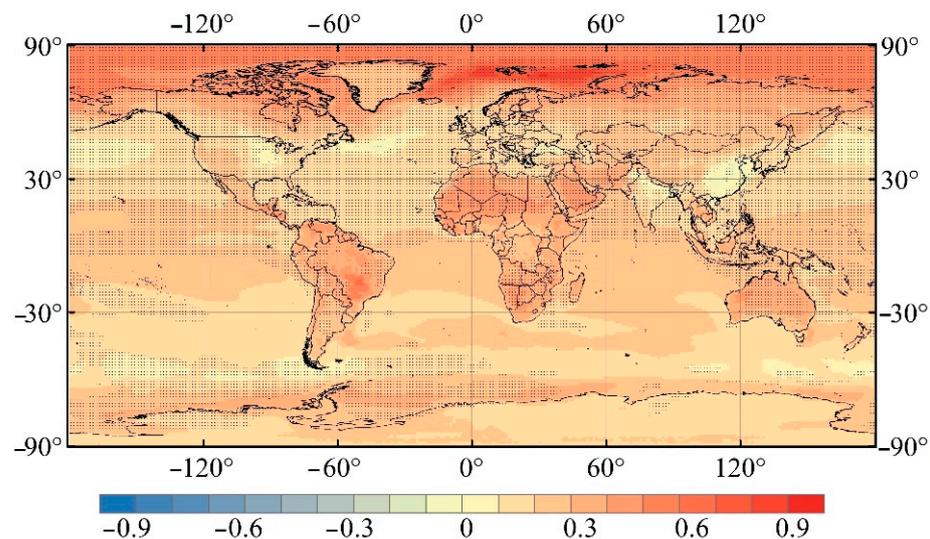
		CMIP6 GCMs		CMIP5 GCMs		ERA5	CERES EBAF
		SMA	BMA	SMA	BMA		
1850–2014	Mean	396.7	388.6	—	—	—	—
	Median	396.4	388.2	—	—	—	—
	Min	394.2	386.2	—	—	—	—
	Max	401.2	392.8	—	—	—	—
1861–2005	Mean	396.6	388.4	394.9	392.0	—	—
	Median	396.4	388.2	394.7	391.8	—	—
	Min	394.2	386.2	392.3	389.5	—	—
	Max	400.1	391.7	398.7	395.7	—	—
1979–2014	Mean	398.9	390.6	—	—	396.4	—
	Median	398.8	390.5	—	—	396.4	—
	Min	396.9	388.7	—	—	394.7	—
	Max	401.2	392.8	—	—	398.0	—
2000–2014	Mean	400.4	392.0	—	—	397.4	398.6
	Median	400.3	392.0	—	—	397.4	398.7
	Min	399.4	391.1	—	—	396.0	397.8
	Max	401.2	392.8	—	—	398.0	399.3

Figure 10 exhibits the variations in the global annual mean SULR from the CMIP6 and CMIP5 GCMs based on the BMA method. The annual mean SULR from CMIP6 GCMs exhibited an increasing tendency of  $0.2 W m^{-2}$  per decade ( $p < 0.05$ ) in 1850–2014. The SULR decreased insignificantly ( $-0.05 W m^{-2}$  per decade,  $p > 0.05$ ) from 1850 to 1899, but during the period of 1900–1939, it showed an increasing tendency ( $0.3 W m^{-2}$  per decade,  $p < 0.05$ ). A slight decrease appeared in 1940–1979 at values of  $-0.04 W m^{-2}$  per decade ( $p > 0.05$ ). However, the SULR increased sharply by  $1.2 W m^{-2}$  per decade ( $p < 0.05$ ) after 1980. Consistent with the CMIP6 GCMs, the SULR increased at values of  $0.2 W m^{-2}$  per decade ( $p < 0.05$ ) under the CMIP5 GCMs during 1861–2005. The annual variations in SULR for the ERA5 and CERES EBAF are also displayed in Figure 10. The global mean SULR showed an increasing trend of  $0.8 W m^{-2}$  per decade ( $p < 0.05$ ) under the ERA5 during the period 1979–2014, while under the BMA results of CMIP6 GCMs, the global mean SULR increased significantly at a rate of  $1.2 W m^{-2}$  per decade ( $p < 0.05$ ) from 1979 to 2014. It can also be found that both the BMA results from CMIP6 GCMs ( $1.1 W m^{-2}$  per decade,  $p < 0.05$ ) and the ERA5 ( $0.7 W m^{-2}$  per decade,  $p < 0.05$ ) showed an increasing trend in the global mean SULR over the period 2000 to 2014. However, the global mean SULR of the CERES EBAF decreased by  $0.2 W m^{-2}$  per decade ( $p > 0.05$ ) from 2000 to 2014. This may be because the SULR data of the CERES EBAF only began in March 2000, and the lack of data may have impacted the estimation of the annual mean SULR in 2000 and made it impossible to explore the long-term variabilities in the SULR. Overall, the extent of the increase was obviously larger than that of the decrease, resulting in an increment in SULR from 1850 to 2014 throughout the world.



**Figure 10.** The annual variations in anomalous global mean SULR of the ERA5, CERES EBAF, and the BMA results of the CMIP6 and CMIP5 GCMs.

The spatial variation in the annual mean SULR trend worldwide for CMIP6 GCMs based on the BMA method is illustrated in Figure 11. The annual mean SULR showed an increasing trend in the majority of regions around the world during 1850–2014, especially in the Arctic, with the maximum increase rate being approximately  $0.91 \text{ W m}^{-2}$  per decade. Additionally, in the Brazilian Plateau, Sahara Desert, and Arabian Peninsula, the annual SULR values exhibited large increments higher than  $0.40 \text{ W m}^{-2}$  per decade. Decreasing SULR trends appeared only on the North China Plain, with a maximum decrease rate of  $-0.08 \text{ W m}^{-2}$  per decade in 1850–2014.



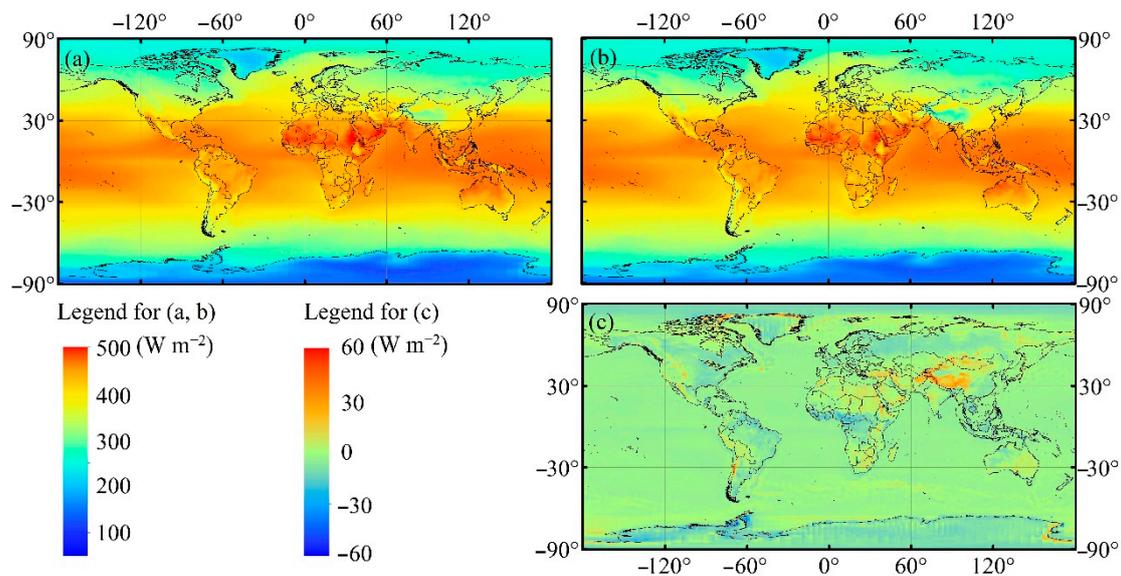
**Figure 11.** Map of the decadal trend in the SULR estimation driven by the BMA method based on 51 CMIP6 GCMs during 1850–2014 (in units of  $\text{W m}^{-2}$  per decade). The solid dots indicate grids at 95% confidence.

## 5. Discussion

The abilities of 51 CMIP6 GCMs to simulate SULR were validated with the SULR observations derived from the SURFRAD, BSRN, and FLUXNET in this study. Various validation metrics have been applied for the assessment of GCMs. In the present research, we used six metrics to evaluate the performance of GCMs in simulating SULR. Out of the six metrics, five were utilized earlier in the assessment of GCMs. It was found that the rankings of the GCMs varied significantly according to these five metrics, so it is difficult to make determinations based on the contradictory results. Moreover, it is questionable to validate the ability of GCMs using an individual metric since single metrics cannot denote the overall accuracy. The GPI combines the results acquired using the five metrics to identify a solution with less human influence, so it can represent the overall model accuracy. In a previous study, Qin et al. [68] used the global performance indicator (GPI), which is calculated through the RMSE, RRMSE, mean absolute bias error (MAE), the relatively mean absolute bias error (MAER), and R, to validate the overall model accuracy. The compromise programming indicator (CPI), similar to GPI, was applied by Iqbal et al. [27] to rank the CMIP6 GCMs by estimating the cumulative effects of multiple metrics. Therefore, we believe that it was a good choice to use GPI to validate the CMIP6 GCMs.

Figure 8 shows that the monthly mean SULR estimated by the SMA method was higher than ERA5, CERES EBAF, and half of the 51 CMIP6 GCMs, while the BMA results exhibited lower SULR values in comparison with the ERA5, CERES EBAF, and the majority of the 51 CMIP6 GCMs. This is due to the fact that the CMIP6 GCMs showed an obvious tendency to overestimate the SULR. The SMA method provided SULR estimations through a simple weighted averaging of individual GCMs, and there still exists the problem of overestimation. However, the BMA method significantly improved the overestimation of SULR, owing to the bias correction procedure using the ground-measured SULR. Therefore, the BMA method showed lower SULR values than the SMA method. Compared to the individual GCMs and SMA method, the BMA method performed better in simulating the SULR, which was consistent with the validation results in Section 4.1.1.

The spatial variations in the SULR estimations from the CERES EBAF and ERA5 during 2000–2005 are exhibited in Figure 12. Both CERES EBAF and ERA5 showed SULR values in the range of 120–520  $W m^{-2}$ , with SULR values higher than 460  $W m^{-2}$  in some land surface areas over tropical regions, such as Central Africa, the Arabian Peninsula, the Indian Peninsula, and North Australia. The Antarctica was an area with lower SULR values, equal to approximately 120–250  $W m^{-2}$  and 120–260  $W m^{-2}$  in the ERA5 and CERES EBAF, respectively. The SULR was also lower in the Arctic, at values of 190–290  $W m^{-2}$  and 200–290  $W m^{-2}$  in the ERA5 and CERES EBAF, respectively. Figure 12 also shows the spatial pattern of the biases between the CERES EBAF and ERA5 for the annual mean SULR; these biases ranged from  $-44$  to 51  $W m^{-2}$ . In the Qinghai–Tibet Plateau and Andes, the positive biases were generally higher than 20  $W m^{-2}$ , and significant negative biases greater than 10  $W m^{-2}$  appeared in Antarctica and West Africa. Generally, the spatial patterns were similar in ERA5 and CERES. However, the differences in SULR were higher over the land surface than over the oceans, especially in high-altitude and high-latitude regions.



**Figure 12.** Spatial variation in the CERES EBAF SULR estimations (a), ERA5 SULR estimations (b), and the biases between CERES EBAF and ERA5 (c) from 2000 to 2005 (in units of  $W m^{-2}$ ).

When CMIP6 and CMIP5 GCMs were compared, both the spatial and annual variation showed some discrepancies; these differences may be due to the better ability of the CMIP6 GCMs to understand the effect of clouds in climate sensitivity, improve the climate prediction in the near term, and analyze the cryosphere changes caused by climate extremes and climate changes [27]. The quantifications of the radiative forcing from different external forcing factors, such as aerosol forcing and greenhouse gases (GHGs), may also result in differences between the CMIP6 and CMIP5 GCMs [29]. The CMIP6 GCMs filled some scientific gaps of the CMIP5 GCMs, including an improvement in the estimation of the radiative forcing in the historical simulations, better identification of the climate response to aerosol forcing over the historical period, and excellent computation of the effect of short-term land use and forcing factors on climate [70]. These improvements in the CMIP6 GCMs compared to the CMIP5 GCMs lead to the discrepancy.

## 6. Conclusions

This research validated the ability of 51 CMIP6 GCMs to simulate SULR values with ground measurements and examined how well the MME methods performed in estimating the SULR with CMIP6 GCMs. The validation datasets included 133 sites (10,598 samples) from SURFRAD, BSRN, and FLUXNET during 1992–2014. Large differences were found in the monthly SULR among all 51 CMIP6 GCMs through a comparison between the CMIP6 GCM SULR simulations and the ground-measured SULR. The bias (Rbias), RMSE (RRMSE), R, and GPI values for the individual CMIP6 GCMs at 133 sites ranged from  $-3$  to  $24 W m^{-2}$  ( $-0.8\%$  to  $6.6\%$ ),  $22$  to  $38 W m^{-2}$  ( $6.3\%$  to  $10.8\%$ ),  $0.91$  to  $0.96$ , and  $-40$  to  $9$ , respectively. The CMIP6 GCMs showed an obvious tendency to overestimate the SULR. Forty-six out of the 51 CMIP6 GCMs overestimated the SULR and showed positive biases for all 133 sites. Among the CMIP6 GCMs, E3SM-1-1-ECA agreed best with the SULR measurements from all 133 sites, exhibiting a bias of  $1.28 W m^{-2}$ , an RMSE of  $22.56 W m^{-2}$ , an R of  $0.951$ , and a maximum GPI value of  $8.383$ . FGOALS-g3 performed the worst, with the lowest GPI value at 133 sites, and its SULR simulations had poor performance for both the FLUXNET and BSRN sites. The BMA method performed better in simulating the SULR than the individual GCMs and SMA method, with a bias of  $0 W m^{-2}$ , an RMSE of  $19.29 W m^{-2}$ , and an R of  $0.97$  for the 133 sites. This method increased the R by approximately  $0.02$  and reduced the RMSE by approximately  $6 W m^{-2}$  on average in comparison with the individual GCMs. Therefore, we recommend using the BMA method to estimate SULR by merging CMIP6 GCMs.

The SULR simulations derived from the CMIP6 GCMs were compared to their counterparts from the CMIP5 GCMs according to the SULR measurements collected at 61 sites (3052 samples) in 1992–2005. The results illustrated that the CMIP6 GCMs did not show better performance when simulating SULR than the CMIP5 GCMs, with an average bias (Rbias) of  $4 \text{ W m}^{-2}$  (1.2%), an average RMSE (RRMSE) of  $26 \text{ W m}^{-2}$  (7.6%), and an average R of 0.95 for the 61 sites. Overall, the CMIP6 GCMs did not improve the overestimation of SULR compared to the CMIP5 GCMs. The SULR values simulated by the CMIP6 GCMs were also validated using the CERES EBAF. The RMSE and biases of the GCMs with respect to the CERES EBAF were lower than those between the GCMs and ground measurements. Similar to the validation results obtained using the SULR observations, FGOALS-g3 showed the poorest ability to simulate SULR in comparison with CERES EBAF.

Based on the CMIP6 GCMs, the BMA method was utilized to estimate a gridded global SULR dataset ( $1^\circ \times 1^\circ$ ) from 2000 to 2005. We discussed the spatial pattern and monthly variations in SULR values throughout the world using the SULR dataset produced in the present study. Generally, the SULR gradually decreased from tropical to polar regions, ranging from 130 to  $490 \text{ W m}^{-2}$ . At the same latitude, the SULR was lower in high-elevation areas than in low-elevation areas. The maximum monthly mean SULR was observed in July at values of  $401 \text{ W m}^{-2}$ , and the minimum SULR value occurred in January at  $383 \text{ W m}^{-2}$ .

In terms of the global annual mean SULR, our best estimation for the CMIP6 GCMs using the BMA method was  $392 \text{ W m}^{-2}$  during 2000–2014. The SULR varied from  $386$  to  $393 \text{ W m}^{-2}$  from 1850 to 2014, exhibiting an increasing tendency of  $0.2 \text{ W m}^{-2}$  per decade ( $p < 0.05$ ). Regarding the spatial variations in the tendency of the annual mean SULR, the SULR showed increasing trends in the majority of regions worldwide during 1850–2014, especially in the Arctic, with the maximum increase rate being approximately  $0.91 \text{ W m}^{-2}$  per decade. Decreasing SULR trends appeared only on the North China Plain, with a maximum decrease rate of  $-0.08 \text{ W m}^{-2}$  per decade in 1850–2014.

**Author Contributions:** Conceptualization, X.Z. (Xiaotong Zhang); methodology, J.X.; validation, J.X.; data curation, J.X.; writing—original draft preparation, J.X.; writing—review and editing, X.Z. (Xiaotong Zhang), J.X., C.F., S.Y., S.G., K.J., Y.Y., X.X., B.J., J.C., X.Z. (Xiang Zhao); supervision, X.Z. (Xiaotong Zhang). All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Natural Science Foundation of China under grants 4217010304 and 42090012.

**Data Availability Statement:** The simulated SULR from CMIP6 and CMIP5 GCMs was downloaded from <https://esgf-node.llnl.gov/search/cmip6/> (accessed on 11 October 2021) and <https://esgf-node.llnl.gov/search/cmip5/> (accessed on 11 October 2021). The ground-measured SULR was obtained from SURFRAD (<https://www.esrl.noaa.gov/gmd/grad/surfrad/>, accessed on 11 October 2021), BSRN (<https://bsrn.awi.de/>, accessed on 11 October 2021), and FLUXNET (<https://fluxnet.fluxdata.org/data/fluxnet2015-dataset/>) (accessed on 11 October 2021). The ERA5 and CERES EBAF data applied in this work are available at <https://cds.climate.copernicus.eu/> (accessed on 11 October 2021) and <https://ceres.larc.nasa.gov/> (accessed on 11 October 2021).

**Acknowledgments:** We thank the researchers and institutions who provided the data utilized in this research and made them available to the public. We also thank the editors and anonymous reviewers for their invaluable suggestions and comments on the original manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

Table A1. 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	DE-RuS	50.87°N	6.45°E
BSRN	BAR	71.32°N	156.61°W	FLUXNET	DE-SfN	47.81°N	11.33°E
BSRN	CAB	51.97°N	4.93°E	FLUXNET	DE-Spw	51.89°N	14.03°E
BSRN	DOM	75.10°S	123.38°E	FLUXNET	DE-Tha	50.96°N	13.57°E
BSRN	GOB	23.56°S	15.04°E	FLUXNET	DK-Sor	55.49°N	11.64°E
BSRN	GVN	70.65°S	8.25°W	FLUXNET	FI-Hyy	61.85°N	24.29°E
BSRN	NYA	78.93°N	11.93°E	FLUXNET	FI-Lom	68.00°N	24.21°E
BSRN	PAY	46.82°N	6.94°E	FLUXNET	FR-Gri	48.84°N	1.95°E
BSRN	SPO	89.98°S	24.80°W	FLUXNET	FR-LBr	44.72°N	0.77°W
BSRN	SYO	69.01°S	39.59°E	FLUXNET	FR-Pue	43.74°N	3.60°E
BSRN	TAT	36.06°N	140.13°E	FLUXNET	GF-Guy	5.28°N	52.92°W
BSRN	TIK	71.59°N	128.92°E	FLUXNET	IT-BCi	40.52°N	14.96°E
BSRN	TOR	58.25°N	26.46°E	FLUXNET	IT-CA1	42.38°N	12.03°E
SURFRAD	BND	40.05°N	88.37°W	FLUXNET	IT-CA2	42.38°N	12.03°E
SURFRAD	TBL	40.12°N	105.24°W	FLUXNET	IT-CA3	42.38°N	12.02°E
SURFRAD	DRA	36.62°N	116.02°W	FLUXNET	IT-Col	41.85°N	13.59°E
SURFRAD	FPK	48.31°N	105.10°W	FLUXNET	IT-Isp	45.81°N	8.63°E
SURFRAD	GWN	34.25°N	89.87°W	FLUXNET	IT-La2	45.95°N	11.29°E
SURFRAD	PSU	40.72°N	77.93°W	FLUXNET	IT-Lav	45.96°N	11.28°E
SURFRAD	SXF	43.73°N	96.62°W	FLUXNET	IT-MBo	46.01°N	11.05°E
FLUXNET	AT-Neu	47.12°N	11.32°E	FLUXNET	IT-Noe	40.61°N	8.15°E
FLUXNET	AU-Ade	13.08°S	131.12°E	FLUXNET	IT-Ren	46.59°N	11.43°E
FLUXNET	AU-ASM	22.28°S	133.25°E	FLUXNET	IT-Ro1	42.41°N	11.93°E
FLUXNET	AU-Cpr	34.00°S	140.59°E	FLUXNET	IT-Ro2	42.39°N	11.92°E
FLUXNET	AU-Cum	33.62°S	150.72°E	FLUXNET	IT-SR2	43.73°N	10.29°E
FLUXNET	AU-DaP	14.06°S	131.32°E	FLUXNET	IT-SRo	43.73°N	10.28°E
FLUXNET	AU-DaS	14.16°S	131.39°E	FLUXNET	IT-Tor	45.84°N	7.58°E
FLUXNET	AU-Dry	15.26°S	132.37°E	FLUXNET	JP-MBF	44.39°N	142.32°E
FLUXNET	AU-Emr	23.86°S	148.47°E	FLUXNET	JP-SMF	35.26°N	137.08°E
FLUXNET	AU-Fog	12.55°S	131.31°E	FLUXNET	NL-Hor	52.24°N	5.07°E
FLUXNET	AU-Gin	31.38°S	115.71°E	FLUXNET	NL-Loo	52.17°N	5.74°E
FLUXNET	AU-GWW	30.19°S	120.65°E	FLUXNET	RU-Che	68.61°N	161.34°E
FLUXNET	AU-How	12.49°S	131.15°E	FLUXNET	RU-Cok	70.83°N	147.49°E
FLUXNET	AU-Lox	34.47°S	140.66°E	FLUXNET	RU-Fyo	56.46°N	32.92°E
FLUXNET	AU-RDF	14.56°S	132.48°E	FLUXNET	SJ-Adv	78.19°N	15.92°E
FLUXNET	AU-Rig	36.65°S	145.58°E	FLUXNET	SJ-Blv	78.92°N	11.83°E
FLUXNET	AU-Rob	17.12°S	145.63°E	FLUXNET	US-AR1	36.43°N	99.42°W
FLUXNET	AU-Stp	17.15°S	133.35°E	FLUXNET	US-AR2	36.64°N	99.60°W
FLUXNET	AU-TTE	22.29°S	133.64°E	FLUXNET	US-ARM	36.61°N	97.49°W
FLUXNET	AU-Tum	35.66°S	148.15°E	FLUXNET	US-GBT	41.37°N	106.24°W
FLUXNET	AU-Wac	37.43°S	145.19°E	FLUXNET	US-GLE	41.37°N	106.24°W
FLUXNET	AU-Whr	36.67°S	145.03°E	FLUXNET	US-Los	46.08°N	89.98°W
FLUXNET	AU-Wom	37.42°S	144.09°E	FLUXNET	US-Me2	44.45°N	121.56°W
FLUXNET	AU-Ync	34.99°S	146.29°E	FLUXNET	US-Me6	44.32°N	121.61°W
FLUXNET	BE-Bra	51.31°N	4.52°E	FLUXNET	US-MMS	39.32°N	86.41°W
FLUXNET	BE-Lon	50.55°N	4.75°E	FLUXNET	US-Ne1	41.17°N	96.48°W
FLUXNET	BR-Sa3	3.02°S	54.97°W	FLUXNET	US-Ne2	41.16°N	96.47°W
FLUXNET	CA-Qfo	49.69°N	74.34°W	FLUXNET	US-Ne3	41.18°N	96.44°W
FLUXNET	CA-SF1	54.49°N	105.82°W	FLUXNET	US-NR1	40.03°N	105.55°W
FLUXNET	CA-SF2	54.25°N	105.88°W	FLUXNET	US-ORv	40.02°N	83.02°W
FLUXNET	CA-SF3	54.09°N	106.01°W	FLUXNET	US-Prr	65.12°N	147.49°W
FLUXNET	CH-Cha	47.21°N	8.41°E	FLUXNET	US-SRG	31.79°N	110.83°W
FLUXNET	CH-Dav	46.82°N	9.86°E	FLUXNET	US-SRM	31.82°N	110.87°W
FLUXNET	CH-Fru	47.12°N	8.54°E	FLUXNET	US-Syv	46.24°N	89.35°W



12. Qin, B.; Cao, B.; Li, H.; Bian, Z.; Hu, T.; Du, Y.; Yang, Y.; Xiao, Q.; Liu, Q. Evaluation of six high-spatial resolution clear-sky surface upward longwave radiation estimation methods with MODIS. *Remote Sens.* **2020**, *12*, 1834. [[CrossRef](#)]
13. Wang, W.; Liang, S. Estimation of high-spatial resolution clear-sky longwave downward and net radiation over land surfaces from MODIS data. *Remote Sens. Environ.* **2009**, *113*, 745–754.
14. Yang, J.; Gong, P.; Fu, R.; Zhang, M.; Chen, J.; Liang, S.; Xu, B.; Shi, J.; Dickinson, R. Erratum: The role of satellite remote sensing in climate change studies. *Nat. Clim. Chang.* **2013**, *3*, 1001. [[CrossRef](#)]
15. Delia Garcia, R.; Cuevas, E.; Ramos, R.; Eugenia Cachorro, V.; Redondas, A.; Moreno-Ruiz, J.A. Description of the Baseline Surface Radiation Network (BSRN) station at the Izaña Observatory (2009–2017): Measurements and quality control/assurance procedures. *Geosci. Instrum. Meth.* **2019**, *8*, 77–96. [[CrossRef](#)]
16. Ohmura, A.; Dutton, E.; Forgan, B.; Frohlich, C.; Gilgen, H.; Hegner, H.; Heimo, A.; König-Langlo, G.; McArthur, B.; Müller, G.; et al. Baseline Surface Radiation Network (BSRN/WCRP): New precision radiometry for climate research. *Bull. Am. Meteorol. Soc.* **1998**, *79*, 2115–2136. [[CrossRef](#)]
17. Augustine, J.; DeLuisi, J.; Long, C. SURFRAD—A national surface radiation budget network for atmospheric research. *Bull. Am. Meteorol. Soc.* **2000**, *81*, 2341–2357. [[CrossRef](#)]
18. Garratt, J. Observed screen (air) and GCM surface/screen temperatures: Implications for outgoing longwave fluxes at the surface. *J. Clim.* **1995**, *8*, 1360–1368. [[CrossRef](#)]
19. Wild, M. The global energy balance as represented in CMIP6 climate models. *Clim. Dynam.* **2020**, *55*, 553–577. [[CrossRef](#)]
20. Wild, M.; Folini, D.; Hakuba, M.Z.; Schär, C.; Seneviratne, S.I.; Kato, S.; Rutan, D.; Ammann, C.; Wood, E.F.; König-Langlo, G. The energy balance over land and oceans: An assessment based on direct observations and CMIP5 climate models. *Clim. Dynam.* **2015**, *44*, 3393–3429.
21. Di Biagio, C.; Pelon, J.; Blanchard, Y.; Loyer, L.; Hudson, S.R.; Walden, V.P.; Raut, J.C.; Kato, S.; Mariage, V.; Granskog, M.A. Toward a better surface radiation budget analysis over sea ice in the high Arctic Ocean: A comparative study between satellite, reanalysis, and local-scale observations. *J. Geophys. Res.-Atmos.* **2021**, *126*. [[CrossRef](#)]
22. Hinkelman, L.M. The global radiative energy budget in MERRA and MERRA-2: Evaluation with respect to CERES EBAF data. *J. Clim.* **2019**, *32*, 1973–1994.
23. Loeb, N.G.; Kato, S.; Rose, F.G.; Doelling, D.R.; Rutan, D.A.; Caldwell, T.E.; Yu, L.; Weller, R.A. Surface irradiances consistent with CERES-Derived Top-of-Atmosphere shortwave and longwave irradiances. *J. Clim.* **2013**, *26*, 2719–2740.
24. Kato, S.; Rose, F.G.; Rutan, D.A.; Thorsen, T.J.; Loeb, N.G.; Doelling, D.R.; Huang, X.; Smith, W.L.; Su, W.; Ham, S.-H. Surface irradiances of Edition 4.0 Clouds and the Earth’s Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) data product. *J. Clim.* **2018**, *31*, 4501–4527.
25. Perez, J.; Menendez, M.; Mendez, F.J.; Losada, I.J. Evaluating the performance of CMIP3 and CMIP5 global climate models over the North-East Atlantic region. *Clim. Dynam.* **2014**, *43*, 2663–2680. [[CrossRef](#)]
26. Kim, Y.-H.; Min, S.-K.; Zhang, X.; Sillmann, J.; Sandstad, M. Evaluation of the CMIP6 multi-model ensemble for climate extreme indices. *Weather Clim. Extrem.* **2020**, *29*, 100269.
27. Iqbal, Z.; Shahid, S.; Ahmed, K.; Ismail, T.; Ziarh, G.F.; Chung, E.-S.; Wang, X. Evaluation of CMIP6 GCM rainfall in mainland Southeast Asia. *Atmos. Res.* **2021**, *254*, 105525.
28. Taylor, K.E.; Stouffer, R.J.; Meehl, G.A. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **2012**, *93*, 485–498. [[CrossRef](#)]
29. Eyring, V.; Bony, S.; Meehl, G.A.; Senior, C.A.; Stevens, B.; Stouffer, R.J.; Taylor, K.E. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* **2016**, *9*, 1937–1958.
30. Bağçacı, S.Ç.; Yucel, I.; Duzenli, E.; Yilmaz, M.T. Intercomparison of the expected change in the temperature and the precipitation retrieved from CMIP6 and CMIP5 climate projections: A Mediterranean hot spot case, Turkey. *Atmos. Res.* **2021**, *256*, 105576.
31. Lu, Z.; Zhao, T.; Zhou, W. Evaluation of the Antarctic circumpolar wave simulated by CMIP5 and CMIP6 models. *Atmosphere* **2020**, *11*, 931.
32. Shen, Z.; Duan, A.; Li, D.; Li, J. Assessment and ranking of climate models in Arctic Sea ice cover simulation: From CMIP5 to CMIP6. *J. Clim.* **2021**, *34*, 3609–3627. [[CrossRef](#)]
33. Marotzke, J.; Jakob, C.; Bony, S.; Dirmeyer, P.A.; O’Gorman, P.A.; Hawkins, E.; Perkins-Kirkpatrick, S.; Nowicki, S.; Paulavets, K.; Seneviratne, S.; et al. Climate research must sharpen its view. *Nat. Clim. Chang.* **2017**, *7*, 89–91. [[CrossRef](#)] [[PubMed](#)]
34. Akinsanola, A.A.; Ongoma, V.; Kooperman, G.J. Evaluation of CMIP6 models in simulating the statistics of extreme precipitation over Eastern Africa. *Atmos. Res.* **2021**, *254*, 105509.
35. Ma, Q.; Wang, K.; Wild, M. Evaluations of atmospheric downward longwave radiation from 44 coupled general circulation models of CMIP5. *J. Geophys. Res.-Atmos.* **2014**, *119*, 4486–4497. [[CrossRef](#)]
36. Kusunoki, S.; Arakawa, O. Are CMIP5 models better than CMIP3 models in simulating precipitation over East Asia? *J. Clim.* **2015**, *28*, 5601–5621. [[CrossRef](#)]
37. Li, J.L.F.; Waliser, D.E.; Stephens, G.; Lee, S.; L’Ecuyer, T.; Kato, S.; Loeb, N.; Ma, H.-Y. Characterizing and understanding radiation budget biases in CMIP3/CMIP5 GCMs, contemporary GCM, and reanalysis. *J. Geophys. Res.-Atmos.* **2013**, *118*, 8166–8184.
38. Wild, M.; Folini, D.; Schär, C.; Loeb, N.; Dutton, E.G.; König-Langlo, G. The global energy balance from a surface perspective. *Clim. Dynam.* **2013**, *40*, 3107–3134. [[CrossRef](#)]

39. Wild, M.; Hakuba, M.Z.; Folini, D.; Dorig-Ott, P.; Schar, C.; Kato, S.; Long, C.N. The cloud-free global energy balance and inferred cloud radiative effects: An assessment based on direct observations and climate models. *Clim. Dynam.* **2019**, *52*, 4787–4812. [[CrossRef](#)]
40. Wild, M. Short-wave and long-wave surface radiation budgets in GCMs: A review based on the IPCC-AR4/CMIP3 models. *Tellus A* **2008**, *60*, 932–945. [[CrossRef](#)]
41. Freidenreich, S.M.; Ramaswamy, V. Analysis of the biases in the downward shortwave surface flux in the GFDL CM2.1 general circulation model. *J. Geophys. Res.* **2011**, *116*. [[CrossRef](#)]
42. Wild, M. Discrepancies between model-calculated and observed shortwave atmospheric absorption in areas with high aerosol loadings. *J. Geophys. Res.-Atmos.* **1999**, *104*, 27361–27371.
43. Mackie, A.; Wild, M.; Brindley, H.; Folini, D.; Palmer, P. Observed and CMIP5-simulated radiative flux variability over West Africa. *Earth Space Sci.* **2020**, *7*. [[CrossRef](#)]
44. Li, J.L.F.; Lee, W.-L.; Yu, J.-Y.; Hulley, G.; Fetzer, E.; Chen, Y.-C.; Wang, Y.-H. The impacts of precipitating hydrometeors radiative effects on land surface temperature in contemporary GCMs using satellite observations. *J. Geophys. Res.-Atmos.* **2016**, *121*, 67–79. [[CrossRef](#)]
45. Li, J.L.F.; Richardson, M.; Hong, Y.; Lee, W.-L.; Wang, Y.-H.; Yu, J.-Y.; Fetzer, E.; Stephens, G.; Liu, Y. Improved simulation of Antarctic Sea ice due to the radiative effects of falling snow. *Environ. Res. Lett.* **2017**, *12*, 084010. [[CrossRef](#)]
46. Bhat, K.S.; Haran, M.; Terando, A.; Keller, K. Climate projections using Bayesian model averaging and space–time dependence. *J. Agr. Biol. Envir. St.* **2011**, *16*, 606–628. [[CrossRef](#)]
47. Raftery, A.E.; Gneiting, T.; Balabdaoui, F.; Polakowski, M. Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Weather Rev.* **2005**, *133*, 1155–1174. [[CrossRef](#)]
48. Duan, Q.; Ajami, N.K.; Gao, X.; Sorooshian, S. Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Adv. Water Reso.* **2007**, *30*, 1371–1386.
49. Zhang, H.; Chu, P.-S.; He, L.; Unger, D. Improving the CPC’s ENSO forecasts using Bayesian model averaging. *Clim. Dynam.* **2019**, *53*, 3373–3385.
50. Yao, Y.; Liang, S.; Li, X.; Liu, S.; Chen, J.; Zhang, X.; Jia, K.; Jiang, B.; Xie, X.; Munier, S.; et al. Assessment and simulation of global terrestrial latent heat flux by synthesis of CMIP5 climate models and surface eddy covariance observations. *Agric. For. Meteorol.* **2016**, *223*, 151–167.
51. Zhang, W.; Zhang, X.; Li, W.; Hou, N.; Wei, Y.; Jia, K.; Yao, Y.; Cheng, J. Evaluation of Bayesian multimodel estimation in surface incident shortwave radiation simulation over high latitude areas. *Remote Sens.* **2019**, *11*, 1776.
52. Wu, H.; Zhang, X.; Liang, S.; Yang, H.; Zhou, G. Estimation of clear-sky land surface longwave radiation from MODIS data products by merging multiple models. *J. Geophys. Res.* **2012**, *117*, D22107.
53. Martinez, A.; Iglesias, G. Wind resource evolution in Europe under different scenarios of climate change characterised by the novel Shared Socioeconomic Pathways. *Energ. Convers. Manag.* **2021**, *234*, 113961. [[CrossRef](#)]
54. Landerer, F.W.; Gleckler, P.J.; Lee, T. Evaluation of CMIP5 dynamic sea surface height multi-model simulations against satellite observations. *Clim. Dynam.* **2014**, *43*, 1271–1283.
55. Pastorello, G.; Trotta, C.; Canfora, E.; Chu, H.; Christianson, D.; Cheah, Y.-W.; Poindexter, C.; Chen, J.; Elbashandy, A.; Humphrey, M.; et al. The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. *Sci. Data* **2020**, *7*, 225. [[CrossRef](#)]
56. Verma, M.; Fisher, J.; Mallick, K.; Ryu, Y.; Kobayashi, H.; Guillaume, A.; Moore, G.; Ramakrishnan, L.; Hendrix, V.; Wolf, S.; et al. Global surface net-radiation at 5 km from MODIS Terra. *Remote Sens.* **2016**, *8*, 739.
57. Graham, R.M.; Cohen, L.; Ritzhaupt, N.; Segger, B.; Graversen, R.G.; Rinke, A.; Walden, V.P.; Granskog, M.A.; Hudson, S.R. Evaluation of six atmospheric reanalyses over Arctic Sea ice from winter to early summer. *J. Clim.* **2019**, *32*, 4121–4143. [[CrossRef](#)]
58. Hamal, K.; Sharma, S.; Baniya, B.; Khadka, N.; Zhou, X. Inter-annual variability of winter precipitation over Nepal coupled with ocean-atmospheric patterns during 1987–2015. *Front. Earth Sci.* **2020**, *8*, 161. [[CrossRef](#)]
59. Flynn, C.J.; Eloranta, E.W.; Cadeddu, M.; Fridlind, A.M.; Bromwich, D.H.; Wang, S.-H.; Verlinde, J.; Silber, I. Cloud influence on ERA5 and AMPS surface downwelling longwave radiation biases in West Antarctica. *J. Clim.* **2019**, *32*, 7935–7949.
60. Zhang, X.; Liang, S.; Wang, G.; Yao, Y.; Jiang, B.; Cheng, J. Evaluation of the reanalysis surface incident shortwave radiation products from NCEP, ECMWF, GSFC, and JMA using satellite and surface observations. *Remote Sens.* **2016**, *8*, 225.
61. Loeb, N.G.; Doelling, D.R.; Wang, H.; Su, W.; Nguyen, C.; Corbett, J.G.; Liang, L.; Mitrescu, C.; Rose, F.G.; Kato, S. Clouds and the Earth’s Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) Top-of-Atmosphere (TOA) edition-4.0 data product. *J. Clim.* **2018**, *31*, 895–918. [[CrossRef](#)]
62. 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. Assessment of the clouds and the Earth’s Radiant Energy System (CERES) Flight Model 5 (FM5) instrument performance and stability. In *Earth Observing Systems XIX, Proceedings of the SPIE Optical Engineering + Applications Conference, San Diego, CA, USA, 18–20 August 2014*; Butler, J., Xiong, X., Gu, X., Eds.; SPIE: Washington, DC, USA, 2014.
63. Smith, N.P.; Wilson, R.; Szewczyk, Z.P.; Thomas, S.; Priestley, K.J. Early trends on the Clouds and the Earth’s Radiant Energy System (CERES) Flight Model 6 (FM6) instrument’s performance. In *Earth Observing Systems XXIII, Proceedings of the SPIE Optical Engineering + Applications Conference, San Diego, CA, USA, 21–23 August 2018*; Butler, J., Xiong, X., Gu, X., Eds.; SPIE: Washington, DC, USA, 2018.

64. 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. Potential faster Arctic Sea ice retreat triggered by snowflakes' greenhouse effect. *Cryosphere* **2019**, *13*, 969–980. [[CrossRef](#)]
65. Jia, B.; Cai, X.; Zhao, F.; Liu, J.; Chen, S.; Luo, X.; Xie, Z.; Xu, J. Potential future changes of terrestrial water storage based on climate projections by ensemble model simulations. *Adv. Water Resour.* **2020**, *142*, 103635. [[CrossRef](#)]
66. Guo, Y.; Cheng, J.; Liang, S. Comprehensive assessment of parameterization methods for estimating clear-sky surface downward longwave radiation. *Theor. Appl. Climatol.* **2019**, *135*, 1045–1058.
67. Medina, H.; Tian, D. Comparison of probabilistic post-processing approaches for improving numerical weather prediction-based daily and weekly reference evapotranspiration forecasts. *Hydrol. Earth Syst. Sci.* **2020**, *24*, 1011–1030. [[CrossRef](#)]
68. Qin, W.; Wang, L.; Gueymard, C.A.; Bilal, M.; Lin, A.; Wei, J.; Zhang, M.; Yang, X. Constructing a gridded direct normal irradiance dataset in China during 1981–2014. *Renew. Sust. Energ. Rev.* **2020**, *131*, 110004.
69. Taylor, K.E. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.-Atmos.* **2001**, *106*, 7183–7192. [[CrossRef](#)]
70. Stouffer, R.J.; Eyring, V.; Meehl, G.A.; Bony, S.; Senior, C.; Stevens, B.; Taylor, K.E. CMIP5 scientific gaps and recommendations for CMIP6. *Bull. Am. Meteorol. Soc.* **2017**, *98*, 95–105. [[CrossRef](#)]