Assessing Surface Solar Radiation Fluxes in the CMIP Ensembles

ALEXANDER LOEW

Ludwig-Maximilians-Universität München, Munich, and Max Planck Institute for Meteorology, Hamburg, Germany

AXEL ANDERSSON

Deutscher Wetterdienst, Offenbach, Germany

JÖRG TRENTMANN AND MARC SCHRÖDER

Satellite-Based Climate Monitoring, Deutscher Wetterdienst, Offenbach, Germany

(Manuscript received 18 July 2014, in final form 30 November 2015)

ABSTRACT

Earth system models are indispensable tools in climate studies. The Coupled Model Intercomparison Project (CMIP) is a coordinated effort of the Earth system modeling community to intercompare existing models. An accurate simulation of surface solar radiation fluxes is of major importance for the accuracy of simulations of the near-surface climate in Earth system models. The present study provides a quantitative assessment of the accuracy and multidecadal changes of surface solar radiation fluxes for model results from two phases of CMIP. The entire archives of phase 5 of CMIP (CMIP5) and its predecessor phase 3 (CMIP3) are analyzed for present-day climate conditions. A relative model ranking is provided, and its uncertainty is quantified using different global observational records. It is shown that the choice of an observational dataset can have a major influence on relative model ranking between CMIP models. However the multidecadal variability of surface solar radiation fluxes, also known as global "dimming" or "brightening," is largely underestimated by the CMIP models.

1. Introduction

Earth system models (ESMs) are indispensable tools for studying the climate of the past and present and for providing projections of Earth's potential future climate. An assessment of the accuracy of ESMs to predict present-day climate conditions and their variability is therefore essential. The Coupled Model Intercomparison Project (CMIP) coordinates multimodel climate experiments using state-of-the-art ESMs. Experiments are conducted by contributing international climate modeling centers. The CMIP experiments are essential for comparing the quality of models from different research institutions and contribute to international climate assessments like the Intergovernmental Panel on Climate Change (IPCC) reports (Flato et al. 2013). The present study analyzes results of

two phases of CMIP [phases 3 and 5 of CMIP (CMIP3 and CMIP5, respectively)] (Meehl et al. 2007; Taylor et al. 2012).

Model evaluation using quantitative model performance metrics has become a widely used practice in different research communities (Abramowitz et al. 2008; Gleckler et al. 2008; Reichler and Kim 2008; Blyth et al. 2011; Abramowitz 2012; Luo et al. 2012; Gettelman et al. 2012; Luo et al. 2012; Anav et al. 2013; Goddard et al. 2013; Brovkin et al. 2013; Eyring et al. 2016). A standardized and regular model evaluation procedure is supposed to be implemented in future CMIPs as well (Meehl et al. 2014).

To evaluate either model processes or the general capabilities of models to represent climate mean states and their variability, the assessment of different geophysical variables is important. In particular, geophysical information derived from satellite observations is used for that purpose, as it provides global data.

Considerable effort has been devoted to improve the atmospheric radiative transfer codes and their parameterizations in CMIP models (e.g., Stevens et al. 2013).

Corresponding author address: Alexander Loew, Ludwig-Maximilians-Universität München, Luisenstr. 37, 80333 Munich, Germany.
E-mail: alexander.loew@lmu.de

The surface solar radiation fluxes are the major energy input for Earth's surface. An assessment of their accuracy is therefore an indicator of the quality of the atmospheric radiative transfer codes used in ESMs. An analysis of the accuracy of the surface solar radiation fluxes is therefore of major importance. Li et al. (2013) provided a first analysis of the accuracy of shortwave and longwave radiation fluxes in CMIP5 simulations, focusing on top-of-the-atmosphere fluxes. Individual assessments of surface water and energy flux components of single ESMs have been reported as well (Hagemann et al. 2013; Brovkin et al. 2013).

Allen et al. (2013) used in situ observations of the Global Energy Balance Archive to evaluate CMIP5 surface solar downwelling radiation fluxes for selected regions. They investigated how CMIP5 models reproduce regionally observed decrease (dimming) or increase (brightening) of surface solar downwelling radiation fluxes. It was found that ESM typically underestimated the multidecadal trends observed at the ground stations.

However, no thorough assessment of surface solar radiation fluxes of CMIP3 and CMIP5 models using state-of-the-art global observational datasets has been conducted so far. The present study therefore provides a complementary assessment of the surface solar radiation fluxes, as it (i) provides a global analysis using multiple observational data records and (ii) analyzes the downwelling as well as upwelling radiation fluxes. Global and long-term satellite products are used as reference. In particular, the following research questions are addressed in the study:

- How well do CMIP models simulate the surface radiation fluxes at climatological time scales? The study is focused on providing skill scores for different surface solar radiation fluxes, which provide information on relative model performance compared to chosen reference datasets. The utilized skill scores provide a quantitative measure of how well a model simulates the climatological mean seasonal cycle of chosen geophysical variables.
- Do CMIP5 models provide improved surface solar radiation fluxes compared to CMIP3? We compare results from the large ensemble of CMIP5 models with their predecessors of CMIP3 to assess if the accuracy in simulated surface solar radiation fluxes has changed between the two phases of CMIP.
- How robust are model rankings, given uncertainties in the observational products? Models can be ranked according to their skill for representing the temporal and spatial patterns of geophysical variables. Such a ranking is, however, also dependent on the uncertainties

of the reference data. We therefore analyze how uncertainties in observations affect CMIP model assessments of surface solar radiation fluxes.
- How well do CMIP models reproduce the multidecadal trends in surface radiation fluxes? Systematic changes in the surface solar radiation have been reported in the literature, known as global dimming and brightening (Wild 2009). We investigate how these multidecadal changes are represented in CMIP simulations as well as long-term satellite radiation products.

We first briefly introduce the CMIP models and datasets used in section 2. The data (pre) processing steps and evaluation methods are introduced thereafter (section 3). Results of the analysis are provided in section 4.

2. Data

a. CMIP5 model data

Data from CMIP5 simulations were obtained through the Earth System Grid Federation node of the German Climate Computing Centre (http://esgf-data.dkrz.de). The analysis in the present study focuses on two core CMIP5 experiments (AMIP and historical), which are supposed to represent present climate conditions and are therefore expected to be best comparable with satellite observations (Taylor et al. 2012):

- Historical (twentieth century) experiments comprise fully coupled (ocean–atmosphere–land) simulations with prescribed greenhouse gas concentrations and are supposed to represent present climate conditions. The period 1979–2005 is analyzed in the present study.
- AMIP simulations correspond to an atmosphere–land-only setup. The ocean state (sea surface temperature and sea ice distribution) is prescribed from observational data. The period 1979–2008 is used in the present study.

For each model and experiment, several realizations (ensemble members) are available. The spread among the ensemble members accounts for the internal variability of each model. All CMIP5 models that fulfilled the following criteria were selected: (i) surface downwelling $R^{\downarrow}$ and upwelling $R^{\uparrow}$ shortwave radiation fluxes were available; (ii) more than one ensemble member was available; and (iii) simulations for either the AMIP or the historical experiments were provided. Based on these requirements, a total of 54 different models were selected for the present study, which corresponds to 90% of all models providing input to CMIP5. A summary of the used models is provided in Table A1.

TABLE 1. Datasets used for surface radiation assessment.

Variable	Dataset (abbreviation)	Period	Horizontal resolution	Reference
Surface downwelling solar radiation flux $R^{\downarrow}$ (W m^{-2})	CLARA-A1 (C-SIS)	1982–2009	0.25° × 0.25°	Karlsson et al. (2013)
	SRB3.0	1983–2007	1° × 1°	Stackhouse et al. (2011); Zhang et al. (2013)
	ISCCP	1989–2009	1° × 1°	Zhang et al. (2004)
	CERES EBAF Ed2.7 (CERES2.7)	2000–13	1° × 1°	Loeb et al. (2009, 2012)
Surface upward solar radiation flux $R^{\uparrow}$ (W m^{-2})	CERES EBAF Ed2.7 (CERES2.7)	2000–13	1° × 1°	Loeb et al. (2009, 2012)
	SRB3.0	1983–2007	1° × 1°	Stackhouse et al. (2011)
	ISCCP	1989–2009	1° × 1°	Zhang et al. (2004)

b. CMIP3 model data

The predecessor of CMIP5 was CMIP3 (Meehl et al. 2007). CMIP3 data are used in this study in particular to assess if the representation of the surface solar radiation fluxes in CMIP5 has been improved.

For CMIP3, we used all models that provided data for the AMIP experiment and also provided the required surface solar upward and downward radiation fluxes. We focused on AMIP simulations, as these prescribe sea surface temperature and sea ice distributions from observations, which should result in the most realistic surface–atmosphere fluxes. A total of nine models is used. A summary of the selected CMIP3 models is provided in Table A1 as well.

c. Observations

Different observational datasets of surface downwelling $R^{\downarrow}$ and upwelling $R^{\uparrow}$ all-sky and broadband (0.25, . . . , 4 μm) surface solar radiation fluxes are used for the assessment. The surface solar net radiation flux ($R_N = R^{\downarrow} - R^{\uparrow}$) is used as an additional metric, as it determines the amount of available energy for surface processes. The following observational datasets are used:

- The International Satellite Cloud Climatology Project (ISCCP) provides information on surface solar radiation fluxes at spatial scales of approximately 280 × 280 km^2 and at 3-hourly temporal resolution. Further details on the estimation of surface fluxes for ISCCP are provided by Zhang et al. (2004). Data from 1989 to 2009 were used for the present analysis.
- The NASA Global Energy and Water Cycle Experiment (GEWEX) Surface Radiation Budget (SRB) project aims at the production of long-term datasets of shortwave and longwave surface and top-of-atmosphere radiation fluxes. It provides 3-hourly to monthly global products with a spatial resolution of 1° × 1°. The fluxes are calculated using cloud parameters from ISCCP and meteorological fields from the NASA GMAO reanalysis dataset. Monthly means of the shortwave surface radiation flux products from

SRB, version 3.0 (SRB3.0 ;Stackhouse et al. 2011; Zhang et al. 2013) are used in the present study.
- Clouds and the Earth's Radiant Energy System (CERES), version 2.7 (Ed2.7), surface solar radiation fluxes are derived from measurements onboard the EOS *Terra* and *Aqua* satellites (Loeb et al. 2012). The CERES surface fluxes are obtained from the CERES Energy Balanced and Filled (EBAF) surface radiation product (Kato et al. 2013).
- The Satellite Application Facility on Climate Monitoring (CM SAF) Cloud, Albedo and Radiation dataset from AVHRR data (CLARA-A1) provides a 28-yr (1982–2009) record of cloud parameters and radiation fluxes. Details on the product and algorithms used are provided by Karlsson et al. (2013).

Major differences between the observations are due to different retrieval algorithms and the origin of the satellite input data (i.e., the type of sensor from which the data products are derived). The GEWEX radiation flux assessment provides a comprehensive review and intercomparison of these flux products (Raschke et al. 2012). The dataset details and references are summarized in Table 1.

d. Analysis regions

The comparison of models and observations is either done for global or regional scales. For the latter, we used 26 regions as predefined in the IPCC Special Report on Extreme Events (SREX). A map of the defined regions is shown in Fig. B1 and the labels for the regions used subsequently in the manuscript are given in appendix B.

3. Methods

a. Data preprocessing

The evaluation of the CMIP models with a set of observations requires careful data preprocessing and a coherent framework for intercomparison. The following processing steps are consistently applied to all model simulations and observations:

1) Ensemble mean calculation: CMIP model simulations are provided as an ensemble of realizations for each experiment. The ensemble mean is used for further data analysis.

2) Monthly means: All model simulations and observations are aggregated to monthly means.

3) Spatial remapping: The observations and model data are remapped to a common grid (horizontal T63 spectral truncation, ~1.8° × 1.8°) using an energy conservative approach with weights proportional to the grid cell sizes.

4) Consistent data mask: The observational datasets typically contain gaps in space and time. These gaps were therefore also applied to the model fields to avoid any sampling-related biases in the comparisons between models and observations.

5) Climatological means: Climatological means are calculated for each grid point from the entire time period of each dataset and are used for further analysis. The effect of different lengths of the individual data records was tested with a sensitivity analysis, which showed only a minor impact on the biases between models and observations.

All data analysis and model–data intercomparisons are performed using an open-source and flexible data analysis framework for the analysis of geospatial data and the benchmarking of models (Loew 2015a).

b. Evaluation skill scores

In the following, we define the skill score metrics that are used to assess the CMIP models. We focus on global skill scores, which provide a first insight into general model performance. Details on spatial and temporal model biases are not provided within this paper as a result of the multitude of models analyzed but can be found in separate reports (see appendix E for details).

1) GLOBAL SKILL SCORES

As a global metric, we use a temporally and area-weighted root-mean-square difference E between observations y and model simulations x. It is derived from climatological monthly means, as defined in Gleckler et al. (2008):

$$E^2 = \sum_i^{nlat} \sum_j^{nlon} \sum_k^{12} w_{i,j,k}(y_{i,j,k} - x_{i,j,k})^2, \quad (1)$$

where w is weight, accounting for different area sizes of model grid cells (nlat, nlon) and time in months. Subscripts i, j, and k are for latitude, longitude, and time, respectively. For each model, the weighted root-mean-square difference (RMSD) is calculated using Eq. (1),

and a relative ranking of all models is then obtained by calculating a normalized error metric for each model as

$$E' = \frac{E - \overline{E}}{\overline{E}}, \quad (2)$$

where $\overline{E}$ is the median value of E from all models (CMIP3 + CMIP5). This results in a relative skill measure to compare the different models against each other. If E' has a value of 0.3, the error of the respective model is 30% higher than on average, whereas the absolute error E provides a measure of the absolute deviation between models and observations.

As E' is calculated for each observational dataset independently, we define two additional skill scores to characterize 1) the overall consistency between models and observations and 2) the stability of E' estimates for different observations. Based on Reichler and Kim (2008), the average skill score for each variable is defined as

$$I^2 = \frac{1}{m} \sum_i^m E'_i, \quad (3)$$

which provides the average error across different observations, where m corresponds to the number of observational datasets (Reichler and Kim 2008). Note that, while Reichler and Kim (2008) used I^2 as a metric to summarize model skill across different geophysical variables, we use the same approach here to quantify the skill for a single variable, but across different observational datasets; I^2 is therefore a measure for the overall agreement of a model with all used observational datasets. In addition, we define also a coefficient of variation as

$$c_v = \frac{\sqrt{\mathrm{Var}(E')}}{|I^2|}, \quad (4)$$

which provides additional information of the spread of I^2 and is therefore an indicator for the consistency of error estimates from different observational datasets. The variance $\mathrm{Var}(E')$ is estimated across the different observational datasets for each variable.

2) CONSISTENCY OF MODEL RANKING

Models can be ranked according to the skill score defined by Eq. (2) for each observation. If all observational datasets provide the same coherent picture of a geophysical field, the resulting model ranking is expected to be independent of the observational reference. However, as the reference datasets differ themselves, the relative model ranking would also depend on the choice of the reference itself. The Spearman rank
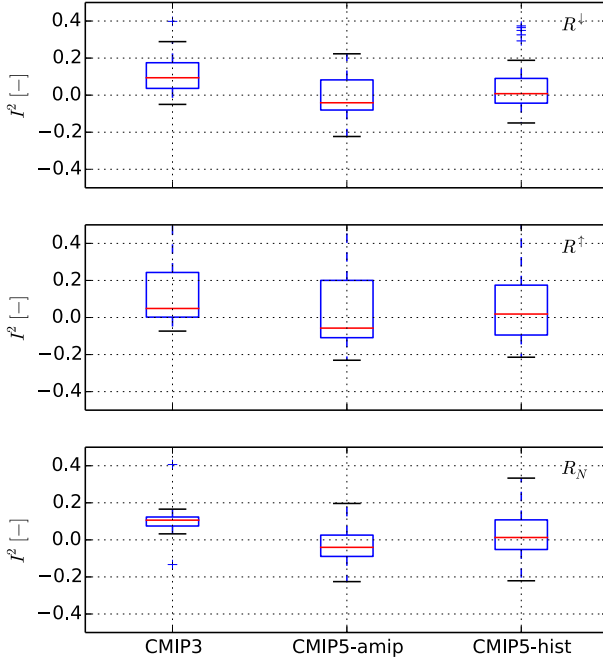
FIG. 1. Multivariate model skill score I^2 for (top) $R^\downarrow$, (middle) $R^\uparrow$, and (bottom) R_N radiation fluxes for CMIP5 and CMIP3 models and AMIP and historical experiments. Boxes correspond to the interquartile range (25%–75%), while whiskers correspond to extreme data values within 1.5 times the IQR.

FIG. 2. As in Fig. 1, but for distribution of c_v [Eq. (6)].

correlation coefficient r_s is used as a measure for the similarity of model rankings obtained using different observational datasets.

Given datasets A and B and the corresponding skill scores for a model x_i from Eq. (2) $[E'_A(x_i)$ and $E'_B(x_i)]$, the ranks $\text{rg}[E'_A(x_i)]$ and $\text{rg}[E'_B(x_i)]$ are obtained. The Spearman rank correlation coefficient r_s is then defined as

$$r_s = \frac{\sum_i^M \{\text{rg}[E'_A(x_i)] - \text{RG}_A\}\{\text{rg}[E'_B(x_i)] - \text{RG}_B\}}{\sqrt{\sum_i^M \{\text{rg}[E'_A(x_i)] - \text{RG}_A\}^2} \sqrt{\sum_i^M \{\text{rg}[E'_B(x_i)] - \text{RG}_B\}^2}}, \tag{5}$$

where $\text{RG}_A = M^{-1} \sum_i^M \text{rg}[E_A(x_i)]$ is the average model rank using observational record A. Similarly RG_B corresponds to the mean model rank, using B as observations and M as the number of models. In the case that the used observational datasets provide a robust representation of the investigated geophysical field, r_s would correspond to unity.

3) MULTIDECADAL VARIABILITY

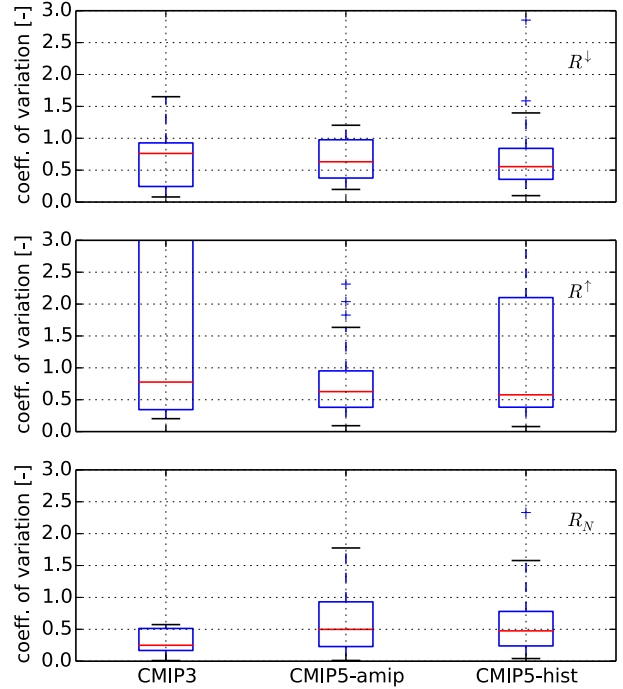A linear regression analysis is further used to assess the similarity in multidecadal trends of surface solar

radiation fluxes. First, an anomaly time series x' is constructed from the entire time series of a dataset by removing the mean seasonal cycle for each grid cell. Regional means are then calculated for each of the IPCC SREX regions, and the long-term trend is estimated by fitting a linear model in time t

$$x'(t) = mt + c + \varepsilon, \tag{6}$$

with random residuals ε, using a least squares approach. The Pearson product moment correlation coefficient ρ and its significance (p value) are calculated. The period 1989–2007 is chosen for the analysis, which corresponds to the same period as analyzed by Allen et al. (2013). Trends are estimated for the SRB3.0, ISCCP, and

TABLE 2. Spearman rank correlation coefficients r_s of model ranks using different observational datasets for AMIP (historical in parentheses) simulations.

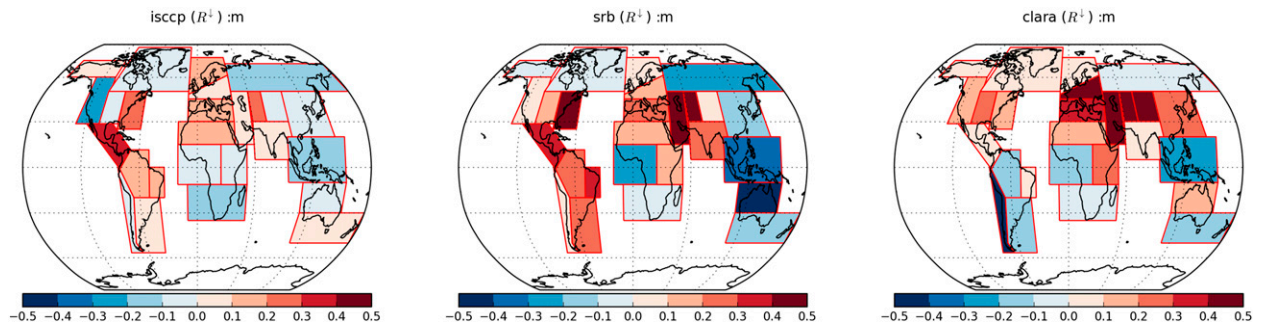|		ISCCP	SRB3.0	CLARA-A1
$R^\downarrow$	CERES2.7	0.82 (0.93)	0.92 (0.94)	0.64 (0.78)
	ISCCP	—	0.84 (0.95)	0.55 (0.83)
	CLARA-A1	—	0.77 (0.91)	—
$R^\uparrow$	CERES2.7	0.85 (0.87)	0.59 (0.6)	—
	ISCCP	—	0.62 (0.68)	—
R_N	CERES2.7	0.86 (0.87)	0.94 (0.91)	—
	ISCCP	—	0.80 (0.80)	—

FIG. 3. Multidecadal regional trends of $R^\downarrow$ (W m^{-2} a^{-1}) for different observational datasets: (left) ISSCP, (center) SRB3.0, and (right) CLARA-A1. Trends are significant ($p < 0.05$) for $|r| > 0.13$.

CLARA-A1 dataset. The CERES2.7 data are excluded from the analysis as they do not cover the entire time period. It is, however, emphasized that trend estimates from multidecadal satellite data records need to be interpreted very carefully. The generation of climate data records from satellite data is a challenging task that requires a careful intercalibration of observations from different platforms and sensors. Otherwise, the datasets might contain spurious trends (Loew 2013). Raschke et al. (2012) provide a comprehensive intercomparison of the different existing global satellite radiation products and their known issues.

4. Results and discussion

a. Overview: CMIP3 versus CMIP5

Results of the multimodel comparison are analyzed in the following. Figure 1 shows the distribution of the overall skill score metric I^2 [see Eq. (3)] for the different surface solar radiation fluxes and experiments. The boxes correspond to the interquartile range (IQR; 25%–75%) of the I^2 values from all models, whereas the whiskers correspond to 1.5 times the IQR.

In general, CMIP5 models have lower I^2 values than CMIP3 simulations, which indicates an overall better
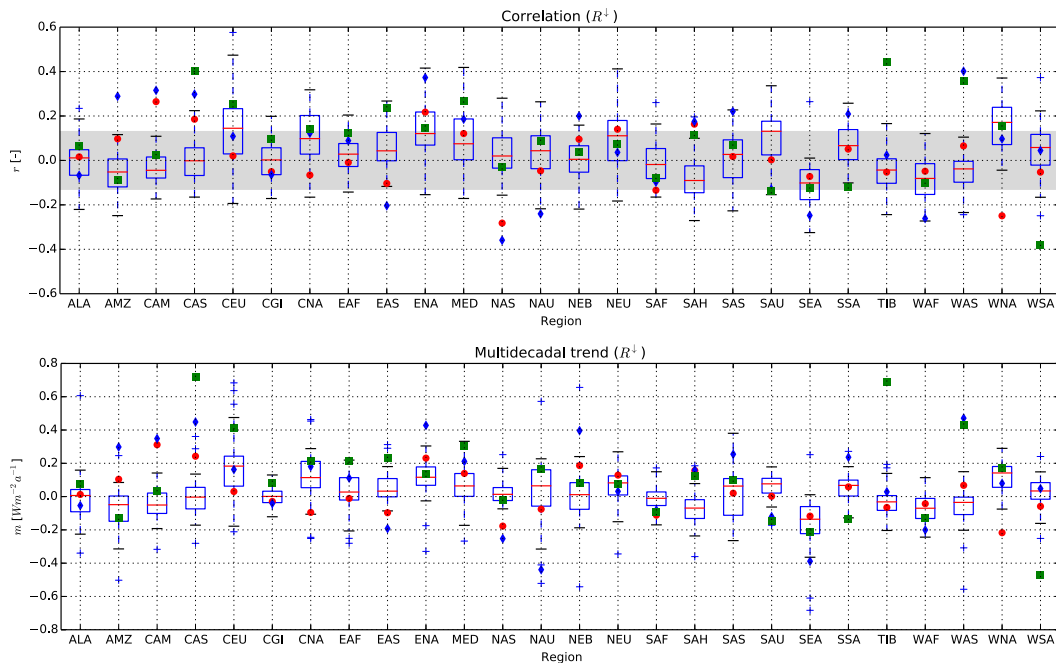
FIG. 4. Multidecadal regional trends of $R^\downarrow$: (top) The Pearson product moment correlation coefficient and (bottom) multidecadal trend (W m^{-2} a^{-1}) for IPCC regions are shown. Shaded areas correspond to correlations that are not significant ($p > 0.05$). Box plots represent model results from AMIP experiment, where the box corresponds to the IQR. The following datasets are represented: ISCCP (red dots), SRB3.0 (blue diamonds), and CLARA-A1 (green rectangle).

FIG. 5. As in Fig. 4, but for multidecadal regional trends of $R^{\uparrow}$.
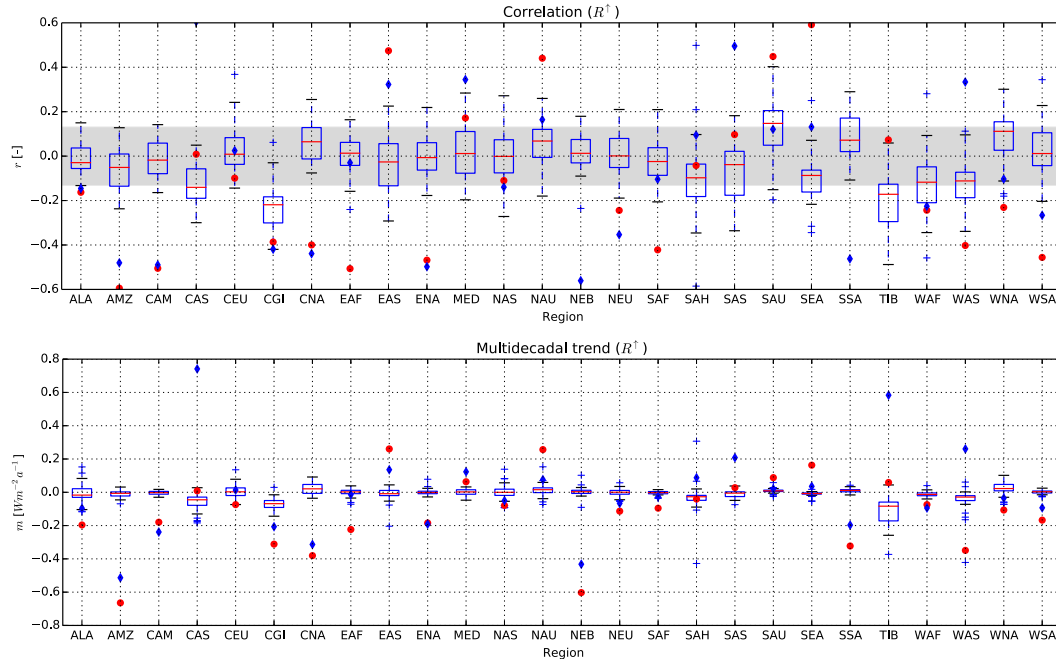
performance of the new model generation. Best results (lowest I^2) are obtained for CMIP5 AMIP experiments with median values for I^2 below zero.

The distributions of I^2 for the downwelling and net solar radiation fluxes are symmetric, while the results for the upwelling solar radiation flux are positively skewed, which indicates a larger spread and uncertainty of the upwelling flux estimates compared to the observations.

The different CMIP models are ranked in accordance to their normalized RMSD E' for each observation. This ranking is summarized in Tables C1 and C2 for all variables and experiments. In general, the multimodel mean outperforms the individual models across the different observations and variables for both the AMIP and historical simulations. This indicates that the ensemble of different models is generally in better agreement with the observations than individual models, which is consistent with previous findings (Flato et al. 2013).

Details of the spatial and temporal deviations of each model compared to the observational datasets are provided in additional reports (see appendix E for details). The global mean absolute error is, on average, 7.5, 5.9, and 6.5 W m^{-2} for $R^{\downarrow}$, $R^{\uparrow}$, and R_N, respectively, and ranges between 4 and 10 W m^{-2}.

b. Consistency of model ranking

It is expected that the spread in E', caused by the different observations, will also affect the relative model ranking. While I^2 provides a general metric for the overall performance of a model across different observational datasets, the c_v [Eq. (6)] is an initial measure of the variability of E'. The distributions of c_v are provided in Fig. 2 for the different experiments and CMIP phases. The variability of E' is typically on the order of 50% for all fluxes and experiments. Largest uncertainties in E' are observed for $R^{\uparrow}$, where the distribution of c_v is largely skewed and the uncertainty of the upward solar

TABLE 3. Observed brightening trends (W m^{-2} decade^{-1}) reported by Allen et al. (2013) and estimated from satellite observations for different regions. Results are provided for IPCC regions defined in Fig. B1.

Region	Allen et al. (2013)	IPCC region	ISCCP	SRB3.0	CLARA-A1
China	2.6 ± 2.2	EAS	−1.0	−1.9	2.3
Japan	4.4 ± 2.0				
Europe	3.5 ± 1.9	NEU	1.3	0.3	0.8
		CEU	0.3	1.6	4.1
		MED	1.4	2.1	3.0

TABLE A1. CMIP3 and CMIP5 models for AMIP and historical experiments.

Modeling center	Model	CMIP5		CMIP3
		AMIP	Historical	AMIP
BCC	BCC_CSM1.1	x	x	
	BCC_CSM1.1(m)	x	x	
CCCma	CanAM4	x		
	CanCM4		x	
	CanESM2		x	
CMCC	CMCC-CESM		x	
	CMCC-CM	x	x	
	CMCC-CMS		x	
Centre National de Recherches	CNRM-CM5	x	x	
Météorologiques–Centre	CNRM-CM5.2		x	
Européen de Recherche et de				
Formation Avancée en Calcul				
Scientific (CNRM–CERFACS)				
CSIRO–BoM	ACCESS1.0	x	x	
	ACCESS1.3	x	x	
CSIRO–Queensland Climate	CSIRO Mk3.6.0	x	x	
Change Centre of Excellence				
(QCCCE)				
EC-EARTH	EC-EARTH		x	
First Institute of Oceanography (FIO)	FIO-ESM		x	
College of Global Change and Earth	BNU-ESM	x	x	
System Science (GCESS)				
Institute of Numerical Mathematics	INM-CM4.0	x	x	
(INM)	INM-CM3.0			x
IPSL	IPSL-CM5A-LR	x	x	
	IPSL-CM5A-MR	x	x	
	IPSL-CM5B-LR	x	x	
	IPSL-CM4			x
LASG–Center of Earth System	FGOALS-g2	x	x	
Science (CESS)				
LASG–IAP	FGOALS-s2	x		
	FGOALS-g1.0			x
Model for Interdisciplinary Research on	MIROC4h		x	
Climate (MIROC) consortium	MIROC5	x	x	
	MIROC-ESM		x	
	MIROC-ESM-CHEM		x	
	MIROC3.2 (hires)			x
	MIROC3.2 (medres)			x
Met Office Hadley Centre (MOHC)	HadCM3		x	
	HadGEM2-A	x		
	HadGEM2-CC		x	
	HadGEM2-ES		x	
	HadGEM1			x
MPI for Meteorology (MPI-M)	MPI-ESM-LR	x	x	
	MPI-ESM-MR	x	x	
	MPI-ESM-P		x	
	ECHAM5			x
Meteorological Research Institute	MRI-AGCM3.2H	x		
(MRI)	MRI-AGCM3.2S	x		
	MRI-CGCM3	x	x	
	MRI-ESM1		x	
NASA GISS	GISS-E2-H		x	
	GISS-E2-H-CC		x	
	GISS-E2-R	x	x	
	GISS-E2-R-CC		x	

TABLE A1. (*Continued*)

Modeling center	Model	CMIP5 AMIP	CMIP5 Historical	CMIP3 AMIP
NCAR	CCSM4	x	x	
	CCSM3			x
	PCM, version 1 (PCM1)			x
Norwegian Climate Centre (NCC)	NorESM1-M	x	x	
	NorESM1-ME		x	
National Institute of Meteorological Research (NIMR)/Korea Meteorological Administration (KMA)	HadGEM2-AO		x	
NOAA/GFDL	GFDL CM3	x	x	
	GFDL-ESM2G		x	
	GFDL-ESM2M		x	
	HiRAM-C180	x		
	HiRAM-C360	x		
National Science Foundation (NSF), DOE, and NCAR	CESM1with biogeochemistry [CESM1(BGC)]		x	
	CESM1(CAM5)	x	x	
	CESM1(FASTCHEM)		x	
	CESM1(WACCM)		x	

radiation flux can be more than twice l^2. This variability also affects the relative model ranking.

The Spearman rank correlation coefficient r_s [Eq. (5)] is a measure for the consistency of the model ranks obtained using different observational datasets. Table 2 summarizes the rank correlations between all pairs of observations. For none of the combinations, the same relative model ranking is obtained ($r_s = 1$). A good agreement is nevertheless found for many of them, for both the AMIP and historical experiments. Best agreement for the surface downweling solar radiation fluxes is observed between the CERES2.7 dataset and the ISCCP (SRB) dataset with $r_s = 0.82$ (0.92). The CLARA-A1 radiation dataset shows, in general, a smaller agreement with the other radiation datasets with respect to the model ranking. Uncertainties in model ranking are, in general, larger for $R^\uparrow$ with $0.59 \leq r_s \leq 0.85$.

c. Multidecadal variability

The multidecadal trend maps for $R^\downarrow$ are provided in Fig. 3. Similar maps for $R^\uparrow$ are given in Fig. D1. The correlation coefficients r and temporal regression slopes m are provided in Fig. 4 and Fig. 5 for each region. The boxes indicate IQR, and the trend estimates from the observations are provided as markers. Results are discussed for each variable in the following.

1) SURFACE SOLAR DOWNWARD RADIATION FLUX

Significant temporal trends ($p \leq 0.05$) are obtained only for regions with $|r| \geq 0.13$, which is only the case for a limited number of regions. The trends of the CMIP models are typically smaller than those obtained from the satellite datasets. This is consistent with results from Allen et al. (2013), who also identified an underestimation of dimming and brightening trends in CMIP models. While the trends from CMIP are typically not statistically significant, the satellite observations show significant positive or negative trends for various regions (see appendix B for region acronyms).

An increase (brightening) is observed for all datasets over Europe and the Mediterranean (NEU, CEU, and MED), northern Africa (SAH), the Arabian Peninsula (WAS), and central and southern Asia (CAS, TIB, and SAS), including parts of Central and North America (CAM and ENA). A reduction of solar radiation (dimming) is observed in all datasets for Siberia (NAS), Southeast Asia (SEA), and western and southern Africa (WAF and SAF). However, the temporal trends of the satellite datasets have, in many cases, different magnitudes and even opposite signs and are statistically not significant ($|r| < 0.13$). These trends need to be interpreted very carefully because of the risk of spurious trends caused by, for example, changes in the observing system (cf. section 3).

Allen et al. (2013) provide an independent study of long-term trends (1987–2009) in surface solar irradiance for Europe, China, Japan, and India. We therefore compare the satellite trend estimates with their results for China, Japan, and Europe for the same time period (Table 3). A comparison for India is not possible, as Allen et al. (2013) only provide information on dimming trends prior to 1989. For China and Japan, the increase in surface solar radiation is 2.6 ± 2.2 and 4.4 ± 2.0 W m^{-2} decade^{-1} for the ground measurements.
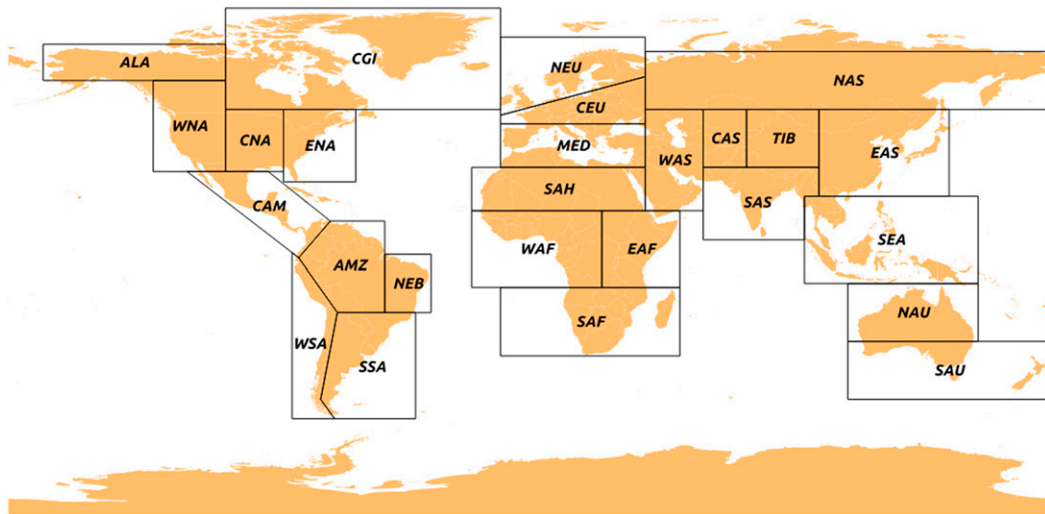
FIG. B1. Definition of regions used for the analysis. The regions are based on those of IPCC (2012), in which the latitude–longitude coordinates of individual regions are also defined.

While the ISCCP and SRB3.0 datasets show even a decrease in $R^{\downarrow}$ for the same time period, the CLARA-A1 dataset shows an increase of $2.3\,\mathrm{W\,m^{-2}\,decade^{-1}}$, which is much closer to the reported in situ observations. The European region is divided into three subregions (NEU, CEU, and MED). Allen et al. (2013) reported a positive trend of $3.5\,\pm\,1.95\,\mathrm{W\,m^{-2}\,decade^{-1}}$. The derived trends from the satellite observations are, in general, much lower, except for the CLARA-A1 dataset, which shows positive trends for central and southern Europe (CEU and MED) that are comparable to the findings of Allen et al. (2013).

While the CMIP models in general underestimate the long-term trend, they at least show positive trends for the aforementioned European and Mediterranean regions (NEU, CEU, and MED). This initial comparison is still very limited, as the reliability of the trend estimates from satellite datasets needs further investigation and cross-comparison with in situ data. The trends for CMIP historical experiments are comparable to those obtained for the AMIP experiments and are provided in Fig. 5.

2) SURFACE UPWELLING SOLAR RADIATION FLUX

As the upward solar radiation flux is the product of the downwelling solar radiation flux and the surface albedo $(R^{\uparrow} = \alpha R^{\downarrow})$, it depends on the temporal variation of $R^{\downarrow}$ and the albedo. The surface albedo is typically simulated in different ways in ESMs. While it was common practice for CMIP3 to prescribe the surface conditions using climatological mean values, most of the land surface schemes in the current ESMs use an interactive albedo

scheme, which simulates the surface albedo as a function of the model state. A comparison of current albedo schemes used in ESMs and their accuracy was investigated by Loew et al. (2014).

ISCCP and SRB3.0 show, in general, a decrease in $R^{\uparrow}$ in North and South America with a maximum decrease of more than $5\,\mathrm{W\,m^{-2}\,decade^{-1}}$ over the Amazon (Fig. 5, bottom), while an increase is mainly observed in eastern Asia. Large differences between the simulated decadal changes of $R^{\uparrow}$ and the satellite records are observed. While the models show no significant trend for most of the IPCC regions, the ISCCP dataset and SRB3.0 show significant changes for many areas. The observed change is often several times larger than those of the models. The spread of trend values among the models is small, and the magnitude of trends is minor.

5. Conclusions

This study provides an initial assessment of the surface solar radiation fluxes for the CMIP3 and CMIP5 ensembles. The accuracy of the CMIP model simulations to describe the mean state and variability of the investigated variables ($R^{\uparrow}$, $R^{\downarrow}$, and R_N) is evaluated. The consistency of using different observational records for the assessment has been investigated.

- Accuracy of CMIP models, when the multimodel mean, in general, outperforms the individual models. This is consistent with previous findings (e.g., Flato et al. 2013). The accuracy of different models shows a large spread on the order of $\pm20\%$.

TABLE C1. Model ranks for CMIP historical experiments. The first three ranks are indicated in boldface.

Model	$R^{\downarrow}$				$R^{\uparrow}$			R_N		
	C-SIS	CERES2.7	ISCCP	SRB3.0	CERES2.7	SRB3.0	ISCCP	CERES2.7	SRB3.0	ISCCP
BCC_CSM1.1	4	27	19	23	28	4	30	8	21	13
BCC_CSM1.1(m)	17	34	32	35	27	**2**	32	25	35	30
BNU-ESM	13	25	26	21	44	44	44	29	34	21
CanCM4	34	17	22	25	32	39	22	18	15	15
CanESM2	36	21	24	26	36	40	33	22	14	14
CMCC-CESM	42	48	47	45	41	24	35	47	47	47
CMCC-CM	6	19	12	7	37	25	28	16	20	24
CMCC-CMS	22	36	27	30	33	19	24	31	37	31
CNRM-CM5.2	35	33	34	31	39	37	34	32	27	34
CNRM-CM5	31	28	28	24	19	15	11	33	22	35
ACCESS1.0	20	5	5	6	**2**	5	4	4	6	**2**
ACCESS1.3	37	20	13	27	5	21	**2**	23	29	6
CSIRO Mk3.6.0	45	40	41	44	43	41	43	41	42	33
FIO-ESM	38	32	35	32	34	38	38	24	30	29
EC-EARTH	**2**	22	18	12	—	—	—	—	—	—
INM-CM4.0	30	29	15	19	47	45	47	28	26	11
HadGEM2-ES	8	**3**	**3**	**3**	6	7	8	**3**	4	4
IPSL-CM5A-LR	47	49	46	47	40	20	37	44	44	36
IPSL-CM5A-MR	46	47	45	46	38	18	41	43	43	32
IPSL-CM5B-LR	43	44	38	43	42	32	42	37	40	23
FGOALS-g2	12	23	17	9	29	35	31	26	16	20
MIROC-ESM-CHEM	49	46	49	49	31	43	40	46	46	46
MIROC-ESM	48	45	48	48	30	42	39	45	45	45
MIROC4h	44	30	40	42	9	30	29	30	39	22
MIROC5	39	31	39	38	16	28	23	36	41	39
HadCM3	28	8	11	15	25	29	26	14	18	10
HadGEM2-CC	14	4	4	4	10	13	10	6	5	5
HadGEM2-ES	7	**2**	**2**	**2**	7	6	9	**2**	**2**	**3**
MPI-ESM-LR	18	11	14	20	11	16	15	19	19	25
MPI-ESM-MR	21	10	10	18	8	12	14	13	12	16
MPI-ESM-P	24	12	16	22	13	9	19	21	17	26
MRI-CGCM3	41	43	44	41	46	47	46	35	25	40
MRI-ESM1	40	42	43	40	45	46	45	34	23	37
GISS-E2-H-CC	27	39	36	36	18	17	**3**	39	36	41
GISS-E2-H	15	41	37	39	24	8	7	42	38	44
GISS-E2-R-CC	25	38	33	34	17	10	5	40	32	43
GISS-E2-R	23	37	31	33	20	11	6	38	31	42
CCSM4	9	14	20	10	22	36	21	9	8	17
NorESM1-M	19	26	30	28	**3**	14	13	15	24	27
NorESM1-ME	26	24	29	29	12	22	16	20	28	28
HadGEM2-AO	16	6	6	5	—	—	—	—	—	—
GFDL CM3	29	9	9	17	23	23	27	7	11	7
GFDL-ESM2G	5	15	8	14	35	27	36	17	13	9
GFDL-ESM2M	**3**	13	7	8	26	**3**	25	10	9	8
CESM1(BGC)	11	18	25	13	15	31	17	12	10	19
CESM1(CAM5)	32	7	23	16	4	33	12	5	**3**	12
CESM1(FASTCHEM)	10	16	21	11	21	34	20	11	7	18
CESM1(WACCM)	33	35	42	37	14	26	18	27	33	38
Model mean	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**

- CMIP5 models have improved compared to CMIP3 models. For all investigated surface radiation fluxes, the CMIP5 models show better skills in simulating the spatiotemporal fields of surface solar radiation fluxes.

- AMIP experiments are better than historical experiments. The CMIP5 AMIP experiments show better agreement with the observations than the historical experiments. The latter have nevertheless smaller errors than the CMIP3 AMIP simulations.

TABLE C2. As in Table C1, but for CMIP3 and CMIP5 AMIP experiments.

Model	$R^{\downarrow}$				$R^{\uparrow}$			R_N		
	C-SIS	CERES2.7	ISCCP	SRB3.0	CERES2.7	SRB3.0	ISCCP	CERES2.7	SRB3.0	ISCCP
CMIP3										
IAP FGOALS-g1.0	28	29	25	26	19	16	21	29	29	31
INM-CM3.0	7	21	11	8	31	25	26	24	22	23
IPSL-CM4	34	36	35	35	35	34	32	35	35	26
MIROC3.2 (hires)	35	24	31	33	12	28	22	28	30	27
MIROC3.2 (medres)	36	35	36	36	26	33	30	36	36	36
MPI ECHAM5	10	28	21	22	24	6	14	25	26	32
CCSM3.0	12	26	27	27	18	**3**	16	23	27	30
PCM1	18	33	28	28	23	5	20	31	31	28
HadGEM1	27	17	32	18	25	27	27	11	12	4
CMIP5										
BCC_CSM1.1	9	27	26	25	33	30	34	16	20	21
BCC_CSM1.1(m)	8	19	19	24	32	32	35	12	15	17
BNU-ESM	5	20	17	16	29	26	24	21	24	20
CanAM4	21	11	14	15	28	31	25	18	14	16
CMCC-CM	**3**	10	6	4	21	7	8	8	8	13
CNRM-CM5	30	32	34	30	16	9	5	33	32	35
ACCESS1.0	17	6	5	7	**2**	**2**	**3**	6	6	**3**
ACCESS1.3	24	12	9	17	6	8	**2**	19	21	8
CSIRO Mk3.6.0	26	13	13	19	30	29	29	17	17	10
INM-CM4.0	22	14	10	9	36	36	33	13	18	9
IPSL-CM5A-LR	37	37	37	37	37	37	37	37	37	37
IPSL-CM5A-MR	33	34	30	34	22	11	28	34	34	19
IPSL-CM5B-LR	32	30	23	31	27	21	23	26	28	14
FGOALS-g2	15	18	16	13	15	17	15	22	16	18
FGOALS-s2	29	31	29	29	17	15	11	32	25	33
MIROC5	31	25	33	32	7	24	10	30	33	34
HadGEM2-A	11	5	4	5	5	4	4	5	5	2
MPI-ESM-LR	19	9	12	12	10	19	12	10	13	12
MPI-ESM-MR	20	8	8	10	8	20	13	9	11	11
MRI-CGCM3	25	22	24	20	34	35	31	20	10	24
GISS-E2-R	23	23	20	23	9	14	1	27	19	29
CCSM4	6	15	18	14	**3**	18	6	14	9	22
NorESM1-M	16	16	22	21	4	13	9	15	23	25
GFDL CM3	13	4	7	6	14	22	17	4	7	7
HiRAM-C180	4	**3**	**3**	**3**	11	12	18	**3**	4	6
HiRAM-C360	**2**	**2**	**2**	**2**	13	10	19	**2**	**2**	5
CESM1(CAM5)	14	7	15	11	**1**	23	7	7	**3**	15
Model mean	**1**	**1**	**1**	**1**	20	**1**	36	**1**	**1**	**1**

- Choice of observational reference matters since the relative model ranking depends on the choice of the observational reference dataset. It is therefore recommended that a multitude of observations is always used when evaluating model simulations. This allows one to account for the variability between the observational records.
- CMIP models underestimate multidecadal trends in surface radiation fluxes. The CMIP model ensemble underestimates observed multidecadal trends in surface solar radiation fluxes. Significant changes observed in in situ and satellite observations are not reproduced by the CMIP models.

- Satellite trend estimates are uncertain because the satellite data records differ in the magnitudes, and in some regions even the sign of regional trends of surface solar radiation fluxes. Hence, trend estimates from the individual datasets should be interpreted very carefully (cf. section 3). An initial comparison for Europe and China with published trend estimates from Allen et al. (2013) indicated that the used CLARA-A1 surface radiation dataset showed multidecadal trends similar to the ground observations.

The main purpose of this study is to provide a general overview about the accuracy of CMIP models to
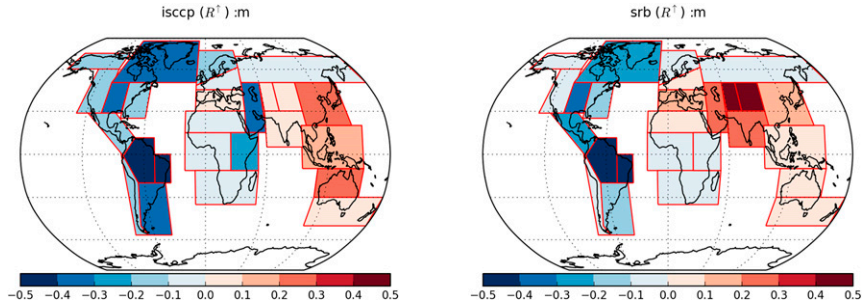
FIG. D1. Multidecadal regional trends of $R^{\uparrow}$ (W m^{-2} a^{-1}) for (left) ISCCP and (right) SRB3.0 observations.

represent climate mean states of surface solar radiation fluxes and their long-term variability. Providing more detailed information on the spatial and temporal biases of individual models is beyond the scope of this paper. Detailed information of individual model deviations is therefore not provided here but is available in separate reports (see appendix E for details).

A more detailed analysis of the long-term trends using in situ, satellite, and model data of different regions of the globe should be subject to further investigations. The current comparison was limited to Europe and China, while previous studies only focused on the comparison of selected satellite and in situ data (Hinkelman et al. 2009; Riihelä et al. 2015; Müller et al. 2015) or in situ data and models (Allen et al. 2013). As only a single time period (1989–2007) was used in this study, to be consistent with Allen et al. (2013), it is recommended that future studies should analyze trends of subsequent time periods, as dimming and brightening trends might compensate each other (Hinkelman et al. 2009).

This study has also shown that the choice of the observational reference has an impact on the relative model ranking. It is recommended that, whenever available, different observational datasets are used for model evaluation purposes to quantify this additional variability. In addition, robust quantitative information on observational uncertainties and guidance for users to choose observations for particular applications is needed. Both issues are subjects of ongoing research (e.g., Hollmann et al. 2013; Gregow et al. 2015) and are of high importance to minimize the impact of observation uncertainties in future ESM evaluation exercises.
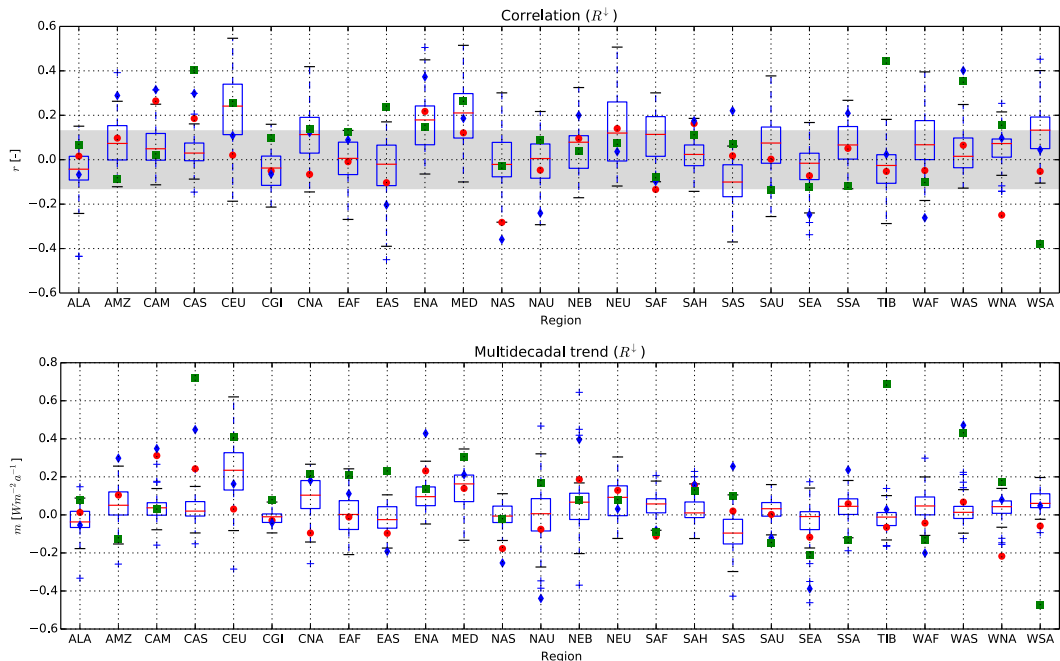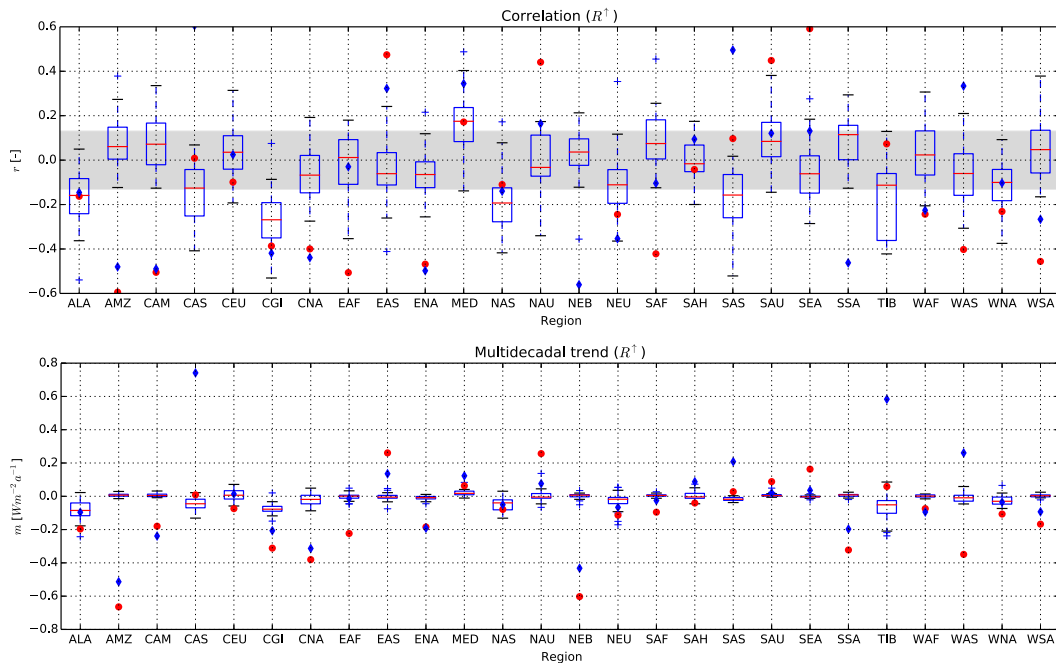

FIG. D2. As in Fig. 4, but for historical experiments.

FIG. D3. As in Fig. 5, but for historical experiments.

APPENDIX A

Models

CMIP models and experiments used in this study are in Table A1. Expansions of acronyms are available at http://www.ametsoc.org/PubsAcronymList.

APPENDIX B

Regions

Figure B1 shows the regions from the IPCC SPEX (IPCC 2012). The defined regions are Alaska/ northwestern Canada (ALA); eastern Canada, Greenland, and Iceland (CGI); western North America (WNA); central North America (CNA); eastern North America (ENA); Central America and Mexico (CAM); Amazon (AMZ); Northeast Brazil (NEB); west coast South America (WSA); southeastern South America (SSA); northern Europe (NEU); central Europe (CEU); southern Europe and Mediterranean (MED); Sahara (SAH); western Africa (WAF); eastern Africa (EAF); southern Africa (SAF); northern Asia (NAS); western Asia (WAS); central Asia (CAS); Tibetan Plateau (TIB); eastern Asia (EAS); southern Asia (SAS); southeastern Asia (SEA); northern Australia (NAU); and southern Australia/New Zealand (SAU).

APPENDIX C

Model Ranking

CMIP model rankings for historical and AMIP experiments are summarized in Tables C1 and C2, respectively. Acronyms for the datasets are as in Table 1.

APPENDIX D

Multidecadal Variability

The spatial distribution of the observed multidecadal trends for $R^\uparrow$ is shown in Fig. D1. Regional distributions of trends for $R^\uparrow$ ($R^\downarrow$) similar to Fig. 4 (Fig. 5) are given in Fig. D2 (Fig. D3).

APPENDIX E

Further Information

This paper summarizes the assessment of the CMIP3 and CMIP5 surface solar radiation fluxes but does not provide detailed information on individual model deviations as well as temporal and spatial deviations between individual models and observations.

As the software used for the evaluation of the different models is publicly available [Python Climate Model Benchmarking Suite (pyCMBS); Loew 2015a], as are all datasets, the interested reader might reproduce all results. The output of the software is comprehensive reports, providing detailed information on individual model deviations. The reports corresponding to the processing of the present paper have been published separately and are accessible in Loew (2015b).

REFERENCES

Abramowitz, G., 2012: Towards a public, standardized, diagnostic benchmarking system for land surface models. *Geosci. Model Dev.*, **5**, 819–827, doi:10.5194/gmd-5-819-2012.

——, R. Leuning, M. Clark, and A. Pitman, 2008: Evaluating the performance of land surface models. *J. Climate*, **21**, 5468–5481, doi:10.1175/2008JCLI2378.1.

Allen, R. J., J. R. Norris, and M. Wild, 2013: Evaluation of multi-decadal variability in CMIP5 surface solar radiation and inferred underestimation of aerosol direct effects over Europe, China, Japan, and India. *J. Geophys. Res. Atmos.*, **118**, 6311–6336, doi:10.1002/jgrd.50426.

Anav, A., G. Murray-Tortarolo, P. Friedlingstein, S. Sitch, S. Piao, and Z. Zhu, 2013: Evaluation of land surface models in reproducing satellite derived leaf area index over the high-latitude Northern Hemisphere. Part II: Earth system models. *Remote Sens.*, **5**, 3637–3661, doi:10.3390/rs5083637.

Blyth, E., D. B. Clark, R. Ellis, C. Huntingford, S. Los, M. Pryor, M. Best, and S. Sitch, 2011: A comprehensive set of benchmark tests for a land surface model of simultaneous fluxes of water and carbon at both the global and seasonal scale. *Geosci. Model Dev.*, **4**, 255–269, doi:10.5194/gmd-4-255-2011.

Brovkin, V., L. Boysen, T. Raddatz, V. Gayler, A. Loew, and M. Claussen, 2013: Evaluation of vegetation cover and land-surface albedo in MPI-ESM CMIP5 simulations. *J. Adv. Model. Earth Syst.*, **5**, 48–57, doi:10.1029/2012MS000169.

Eyring, V., and Coauthors, 2016: ESMValTool (v1.0)—A community diagnostic and performance metrics tool for routine evaluation of Earth system models in CMIP. *Geosci. Model Dev.*, **9**, 1747–1802, doi:10.5194/gmd-9-1747-2016.

Flato, G., and Coauthors, 2013: Evaluation of climate models. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 741–866. [Available online at https://www.ipcc.ch/pdf/assessment-report/ar5/wg1/WG1AR5_Chapter09_FINAL.pdf.]

Gettelman, A., and Coauthors, 2012: A community diagnostic tool for chemistry climate model validation. *Geosci. Model Dev.*, **5**, 1061–1073, doi:10.5194/gmd-5-1061-2012.

Gleckler, P. J., K. E. Taylor, and C. Doutriaux, 2008: Performance metrics for climate models. *J. Geophys. Res.*, **113**, D06104, doi:10.1029/2007JD008972.

Goddard, L., and Coauthors, 2013: A verification framework for interannual-to-decadal predictions experiments. *Climate Dyn.*, **40**, 245–272, doi:10.1007/s00382-012-1481-2.

Gregow, H., and Coauthors, 2015: User awareness concerning feedback data and input observations used in reanalysis systems. *Adv. Sci. Res.*, **12**, 63–67, doi:10.5194/asr-12-63-2015.

Hagemann, S., A. Loew, and A. Andersson, 2013: Combined evaluation of MPI-ESM land surface water and energy fluxes. *J. Adv. Model. Earth Syst.*, **5**, 259–286, doi:10.1029/2012MS20008.

Hinkelman, L. M., P. W. Stackhouse, B. A. Wielicki, T. Zhang, and S. R. Wilson, 2009: Surface insolation trends from satellite and ground measurements: Comparisons and challenges. *J. Geophys. Res.*, **114**, D00D20, doi:10.1029/2008JD011004.

Hollmann, R., and Coauthors, 2013: The ESA Climate Change Initiative: Satellite data records for essential climate variables. *Bull. Amer. Meteor. Soc.*, **94**, 1541–1552, doi:10.1175/BAMS-D-11-00254.1.

IPCC, 2012: *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*, C. B. Field et al., Eds., Cambridge University Press, 582 pp.

Karlsson, K.-G., and Coauthors, 2013: CLARA-A1: A cloud, albedo, and radiation dataset from 28 yr of global AVHRR data. *Atmos. Chem. Phys.*, **13**, 5351–5367, doi:10.5194/acp-13-5351-2013.

Kato, S., N. G. Loeb, F. G. Rose, D. R. Doelling, D. A. Rutan, T. E. Caldwell, L. Yu, and R. A. Weller, 2013: Surface irradiances consistent with CERES-derived top-of-atmosphere shortwave and longwave irradiances. *J. Climate*, **26**, 2719–2740, doi:10.1175/JCLI-D-12-00436.1.

Li, J.-L. F., D. E. Waliser, G. Stephens, S. Lee, T. L'Ecuyer, S. Kato, N. Loeb, and H.-Y. Ma, 2013: Characterizing and understanding radiation budget biases in CMIP3/CMIP5 GCMs, contemporary GCM, and reanalysis. *J. Geophys. Res.*, **118**, 8166–8184, doi:10.1002/jgrd.50378.

Loeb, N. G., B. A. Wielicki, D. R. Doelling, G. L. Smith, D. F. Keyes, S. Kato, N. Manalo-Smith, and T. Wong, 2009: Toward optimal closure of the earth's top-of-atmosphere radiation budget. *J. Climate*, **22**, 748–766, doi:10.1175/2008JCLI2637.1.

——, S. Kato, W. Su, T. Wong, F. G. Rose, D. R. Doelling, J. R. Norris, and X. Huang, 2012: Advances in understanding top-of-atmosphere radiation variability from satellite observations. *Surv. Geophys.*, **33**, 359–385, doi:10.1007/s10712-012-9175-1.

Loew, A., 2013: Terrestrial satellite climate data records: How long is long enough? A test case for the Sahel. *Theor. Appl. Climatol.*, **115**, 427–440, doi:10.1007/s00704-013-0880-6.

——, 2015a: pycmbs, version 1.1.0. Zenodo, accessed 11 May 2015, doi:10.5281/zenodo.17486.

——, 2015b: Comprehensive CMIP5 model evaluation reports: Auxiliary material. Open Data LMU, doi:10.5282/ubm/data.70.

——, P. M. van Bodegom, J.-L. Widlowski, J. Otto, T. Quaife, B. Pinty, and T. Raddatz, 2014: Do we (need to) care about canopy radiation schemes in DGVMs? Caveats and potential impacts. *Biogeosciences*, **11**, 1873–1897, doi:10.5194/bg-11-1873-2014.

Luo, Y. Q., and Coauthors, 2012: A framework for benchmarking land models. *Biogeosciences*, **9**, 3857–3874, doi:10.5194/bg-9-3857-2012.

Meehl, G. A., C. Covey, K. E. Taylor, T. Delworth, R. J. Stouffer, M. Latif, B. McAvaney, and J. F. B. Mitchell, 2007: The WCRP CMIP3 multimodel dataset: A new era in climate change research. *Bull. Amer. Meteor. Soc.*, **88**, 1383–1394, doi:10.1175/BAMS-88-9-1383.

——, R. Moss, and K. Taylor, 2014: Climate model intercomparisons: Preparing for the next phase. *Eos, Trans. Amer. Geophys. Union*, **95**, 77–78, doi:10.1002/2014EO090001.

Müller, R., U. Pfeifroth, C. Träger-Chatterjee, J. Trentmann, and R. Cremer, 2015: Digging the METEOSAT treasure—3 decades of solar surface radiation. *Remote Sens.*, **7**, 8067–8101, doi:10.3390/rs70608067.

Raschke, E., S. Kinne, and P. W. Stackhouse, 2012: GEWEX Radiative Flux Assessment (RFA) volume 1: Assessment. WCRP Tech. Rep. 19/2012, 273 pp. [Available online at http://www.wcrp-climate.org/documents/GEWEX%20RFA-Volume%201-report.pdf.]

Reichler, T., and J. Kim, 2008: How well do coupled models simulate today's climate? *Bull. Amer. Meteor. Soc.*, **89**, 303–311, doi:10.1175/BAMS-89-3-303.

Riihelä, A., T. Carlund, J. Trentmann, R. Müller, and A. Lindfors, 2015: Validation of CM SAF surface solar radiation datasets over Finland and Sweden. *Remote Sens.*, **7**, 6663–6682, doi:10.3390/rs70606663.

Stackhouse, P. W., S. K. Gupta, S. J. Cox, T. Zhang, J. C. Mikovitz, and L. M. Hinkelman, 2011: 24.5-year data set released. *GEWEX News*, Vol. 21, No. 1, International GEWEX Project Office, Silver Spring, MD, 10–12.

Stevens, B., and Coauthors, 2013: Atmospheric component of the MPI-M Earth System Model: ECHAM6. *J. Adv. Model. Earth Syst.*, **5**, 146–172, doi:10.1002/jame.20015.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, doi:10.1175/BAMS-D-11-00094.1.

Wild, M., 2009: Global dimming and brightening: A review. *J. Geophys. Res.*, **114**, D00D16, doi:10.1029/2008JD011470.

Zhang, T., P. W. Stackhouse, S. K. Gupta, S. J. Cox, J. Colleen Mikovitz, and L. M. Hinkelman, 2013: The validation of the GEWEX SRB surface shortwave flux data products using BSRN measurements: A systematic quality control, production and application approach. *J. Quant. Spectrosc. Radiat. Transfer*, **122**, 127–140, doi:10.1016/j.jqsrt.2012.10.004.

Zhang, Y., W. B. Rossow, A. A. Lacis, V. Oinas, and M. I. Mishchenko, 2004: Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global data sets: Refinements of the radiative transfer model and the input data. *J. Geophys. Res.*, **109**, D19105, doi:10.1029/2003JD004457.