ADVANCING EARTH AND SPACE SCIENCE

Earth and Space Science

RESEARCH ARTICLE
10.1029/2022EA002758

Special Section:
Monitoring the Earth radiation budget and its implication to climate simulations: Recent Advances and Discussions

Key Points:

- This study investigates the regional energy budgets in CMIP6 models, and compares them with reference data sets and CMIP5 models
- In most models, some energy budget components are out of the references' uncertainty range in some regions
- The regional scale energy budgets simulated by the CMIP6 models are overall improved compared to CMIP5

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:
D. Li,
dli@marum.de

Citation:
Li, D., Folini, D., & Wild, M. (2023). Assessment of top of atmosphere, atmospheric and surface energy budgets in CMIP6 models on regional scales. *Earth and Space Science*, *10*, e2022EA002758. https://doi.org/10.1029/2022EA002758

Received 30 NOV 2022
Accepted 7 MAR 2023

Author Contributions:
Conceptualization: Doris Folini, Martin Wild
Data curation: Donghao Li
Formal analysis: Donghao Li
Project Administration: Martin Wild
Supervision: Doris Folini, Martin Wild
Visualization: Donghao Li

Assessment of Top of Atmosphere, Atmospheric and Surface Energy Budgets in CMIP6 Models on Regional Scales

Donghao Li[1,2], Doris Folini[1], and Martin Wild[1]

[1]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland, [2]Now at MARUM—Center for Marine Environmental Sciences and Faculty of Geosciences, University of Bremen, Bremen, Germany

Abstract We examine top of atmosphere (TOA), atmospheric, and surface energy budget components of 53 CMIP6 models for the period 2000–2009 on regional scales with respect to two reference data sets: the NASA Energy and Water cycle Study (NEWS), from which we adopt the regional decomposition, and the Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF). Focusing on regional scale, CMIP6 models tend to have more energy entering or less energy leaving the climate systems at TOA over the Northern Hemisphere land, Southern Hemisphere ocean and the polar regions compared to CERES EBAF, while the contrary applies in other regions. Atmospheric net shortwave and longwave fluxes both tend to be underestimated in CMIP6 models as compared to EBAF, with substantial regional differences. Regional surface radiative fluxes as reported by NEWS and EBAF can differ substantially. Nevertheless, robust regional biases exist in CMIP6. Surface upward shortwave radiation is overestimated by 43 (81%) models over Eurasia. For almost all surface radiative flux components over the North Atlantic and Indian Ocean, there are at least 21 (40%) models which fall outside the NEWS uncertainty range. Latent heat flux is overestimated over most of the land and ocean regions while firm conclusions for sensible heat flux remain elusive due to discrepancies between different reference data sets. Compared to CMIP5, there is an overall improvement in CMIP6 on regional scale. Still, substantial deficiencies and spreads on regional scale remain, which are potentially translated into an inadequate simulation of atmospheric dynamics or hydrological cycle.

Plain Language Summary The Earth's energy budget is the principal determinant of climate on our planet and is shaped by the balance between the solar energy absorbed by the Earth and the thermal energy that leaves Earth to space. The ability of Earth System Models (ESMs) to simulate the Earth's energy budget on global scale has been widely assessed in previous studies. In the present study, their ability to simulate the energy budgets on regional scales are examined in detail and compared with reference data sets and their previous model generation. Our study finds that models tend to receive more energy or lose less energy at the top of atmosphere over Northern Hemispheric land, Southern Hemispheric ocean, and polar regions. The net energy fluxes within the atmosphere are also underestimated with substantial differences between regions. In most models, some energy fluxes at the surface of the Earth are even out of the references' uncertainty range in regions such as Eurasia, the North Atlantic, and Indian Oceans. Nevertheless, compared to their previous generation, the state-of-the-art ESMs investigated in our study show an overall improvement in their regional energy budgets, but deficiencies in atmospheric dynamics and hydrological cycle on regional scales should not be neglected.

1. Introduction

The Earth's climate, as a mainly solar powered system, is determined by the vertical and horizontal components of the Earth's energy budget at the top of atmosphere (TOA), atmosphere, and surface. Initial attempts have been conducted since the early twentieth century to observationally detect and numerically characterize Earth's energy budget (e.g., Abbot & Fowle, 1908; Brooks, 1932; Budyko et al., 1962; Dines, 1917; Hunt et al., 1986; Lettau, 1954).

Since the 1960s, the study of Earth's energy budget at the TOA has been revolutionized with the advent of satellite observations (e.g., House et al., 1986; Raschke & Bandeen, 1970; Raschke et al., 1973; Suomi, 1958) among which the most prominent ones are the Earth Radiation Budget Experiment (ERBE, Barkstrom, 1984; Barkstrom et al., 1990), the Clouds and the Earth's Radiant Energy System (CERES, Loeb et al., 2009, 2018; Wielicki et al., 1996), and the Solar Radiation and Climate Experiment (SORCE, Anderson & Cahalan, 2005). However,

Writing – original draft: Donghao Li
Writing – review & editing: Doris Folini, Martin Wild

the satellite missions are not able to directly measure the energy budget within the atmosphere and at the surface, where the surface-based measurement networks such as the Global Energy Balance Archive (GEBA, Gilgen et al., 1998; Ohmura et al., 1989; Wild et al., 2017), the Baseline Surface Radiation Network (BSRN, Driemel et al., 2018; Ohmura et al., 1998), and the Argo Program (Jayne et al., 2017; Roemmich et al., 2009) play an important role. In parallel, the rapid development of computing power in the 1970s contributed to the maturation of global climate models (GCMs) (Edwards, 2000), which further provided the basis for the rise of reanalyses such as ERA-15 (Gibson et al., 1997), ERA-Interim (Dee et al., 2011), ERA5 (Hersbach et al., 2020), NCEP (Kalnay et al., 1996; Saha et al., 2010), and JRA-25 (Onogi et al., 2007). Following these advances, in the past 30 years, new estimates of Earth's energy budget have been continuously published (e.g., Jung et al., 2019; Kiehl & Trenberth, 1997; L'Ecuyer et al., 2015; Stephens et al., 2012; Thomas et al., 2020; Trenberth et al., 2009; Wild et al., 1998, 2013, 2015). An essential element in this overall context is the partitioning of net surface radiation into sensible and latent heat turbulent fluxes, with the latter directly linking to the Earth's water cycle.

Classified by time scale, the Earth's annual energy budget sets the tone for a more elaborate study of seasonal or monthly variability (e.g., Kato, Rose, et al., 2021; L'Ecuyer et al., 2015; Trenberth & Fasullo, 2013c; Thomas et al., 2020; van den Broeke et al., 2011). In terms of spatial scale, the investigation of Earth's global energy budget (e.g., Kiehl & Trenberth, 1997; Stephens et al., 2012; Trenberth et al., 2009; Wild et al., 2013) provides a first-order understanding of the energy distribution within the climate system and the vertical energy transfer mechanism. The global biases can be further split into regional biases using regional divisions such as inter-hemispheric differences (Lembo et al., 2019), land-sea differences (Wild et al., 2015), continental and ocean basin differences (Jung et al., 2019; L'Ecuyer et al., 2015; Thomas et al., 2020). Proceeding from global to regional scales is of interest as the latter set the stage for horizontal energy transport, which has a vital role in regulating Earth's regional energy budget. Various related studies on cross-equatorial energy transport (Loeb et al., 2016; Stephens et al., 2016), meridional energy transport (Fasullo & Trenberth, 2008b), ocean to land energy transport (Fasullo & Trenberth, 2008a; Liu et al., 2020), as well as continental and ocean basin energy transport (Kato, Loeb, et al., 2021; Trenberth & Fasullo, 2013b, 2013c, 2017) exist. These studies highlight the key role of regional scale energy (im-)balances for dynamical aspects of the Earth system, via vertical and horizontal moisture and energy transport, and the need to assess associated regional scale biases in climate models.

In the context of global warming, human health is at risk due to the impact of regional climate change (e.g., Patz et al., 2005). Although regional climate models (RCMs) are more suitable to provide regional climate projections in detail than GCMs, RCMs tend to inherit uncertainties in the boundary conditions from GCMs through downscaling (Rummukainen, 2010). Thus, credible GCMs are essential for reliable regional climate change projections (Xie et al., 2015). An accurate representation of Earth's regional energy budget is fundamental for a skillful GCM to provide projections of regional climate change. This reinforces the necessity to assess the model performance in simulating regional energy budgets.

The estimates of the magnitudes of the energy budget components are always accompanied with uncertainties. Although the satellite-based data sets and reanalyses provide a full-scale view of the Earth's energy budget, they generally lack uncertainty quantification on regional scales. While surface-based measurements could take this concern into account, they are hampered by their limited spatial coverage. As a result, combined approaches such as multi-ensemble and multi-product are used. However, closure problems related to large atmospheric or surface energy imbalances arise when independent data sets are combined. With the aim to address these issues, L'Ecuyer et al. (2015) and Rodell, Beaudoing, et al. (2015) explicitly couple the energy and water cycles through reconciling the satellite-based data sets under NASA Energy and Water Cycle Study (NEWS), while accounting for the uncertainty in each component. Their study covers 16 land and ocean regions across the globe over the period 2000–2009. Thomas et al. (2020) further developed the NEWS solution by considering spatial covariances and introducing additional constrains from ocean reanalyses.

As the Coupled Model Intercomparison Project (CMIP) has reached its sixth phase (CMIP6, Eyring et al., 2016), the NEWS data product provides a solid basis to assess the CMIP6 model performance in simulating the Earth's regional annual energy budget. We complement this surface data product with TOA and additional surface data products from CERES as our references. Our analysis retains the geographical decomposition into regions as given by the NEWS data product. The focus on this study is thus on the assessment of the regional scale representation of the energy budgets in the CMIP6 models, thereby complementing the analysis of the global energy budgets in the CMIP6 models by Wild (2020).

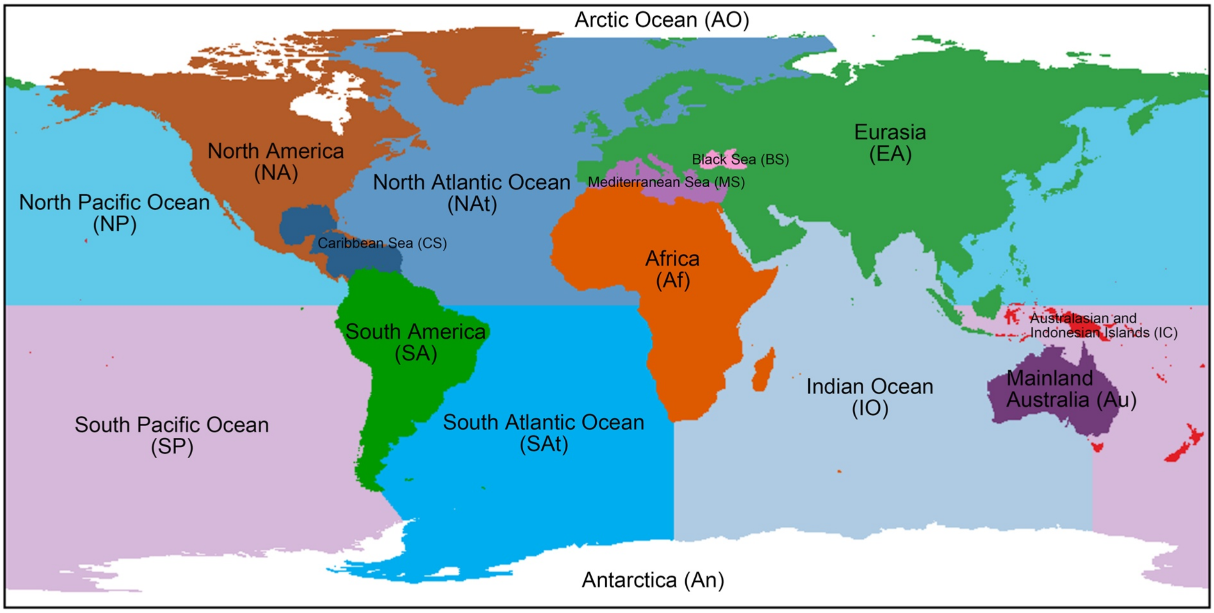


Figure 1. Geographical distribution of 16 land and ocean regions as originally determined for the NEWS data product (L'Ecuyer et al., 2015; Rodell, Beaudoing, et al., 2015; Rodell, L'Ecuyer, et al., 2015) with their corresponding name abbreviations as used in the following figures and tables.

A description of data and methods is featured in Section 2. In Section 3, we focus on comparing the regional multi-annual mean values of all the (all-sky) components of the Earth's energy budget between our references and CMIP6 models. We also compare the annual regional energy budget components as represented in CMIP6 models with the ones of its previous generation CMIP5. Summary and conclusions are presented in Section 4.

2. Data and Methods

We examine the energy budget components of CMIP models on regional scales with respect to a range of reference data products. The models and data products all come with their own spatial and temporal characteristics. To compare the different data, we single out one of the reference data products, the NEWS product, and adopt its geographical regions, spatial resolution, and temporal coverage.

2.1. NEWS Annual Climatology Version 1.0 Data Product

We primarily use the mean values and uncertainty ranges from NEWS Annual Climatology of the 1st decade of the 21st Century Data Product Version 1.0 (L'Ecuyer et al., 2015; Rodell, Beaudoing, et al., 2015; Rodell, L'Ecuyer, et al., 2015) as reference for the surface energy budget components over 16 continents and ocean basins as well as over global land, global oceans and the globe (Figure 1). Besides the data as such, we adopt from this data set the 16 geographical regions illustrated in Figure 1 and Table 1, the $0.25° \times 0.25°$ spatial discretization, and the temporal coverage of 2000–2009.

The satellite-based data sets used to derive the NEWS energy budget estimates are mainly evaluated for the period 2000–2009, while individual data sets cover slightly different periods starting no earlier than 1998 and ending no later than 2010. Detailed input data set information regarding the energy and water cycle is listed in Table 1 of L'Ecuyer et al. (2015) and Rodell, Beaudoing, et al. (2015), respectively. Based on the characteristics of the selected data sets for different components, prior to the constraint, the authors carefully used various methods to estimate the uncertainty, such as comparisons against direct observations, standard deviation of independent data sets, sensitivity approaches, error propagation, etc. Thereafter, they utilized the inverse modeling method (Kalnay, 2003; Rodgers, 2000) to simultaneously impose energy and water balance constraints on the components of the annual energy and water cycles, while explicitly accounting for the uncertainty in each component with the assumption that the uncertainty is random and Gaussian.

Table 1
The Areas of 16 Land and Ocean Regions and Their Percentages of the Global Land or Ocean Area As Well As Their Percentages of the Global Area

Region	Abbr.	Area ($10^6 km^2$)	% Of global land/ocean	% Of globe
Eurasia	EA	53.2	36%	10%
Africa	Af	29.9	20%	5.8%
North America	NA	24.0	16%	4.7%
South America	SA	17.7	12%	3.5%
Antarctica	An	12.7	8.7%	2.5%
Mainland Australia	Au	7.56	5.2%	1.5%
Australasian and Indonesian Islands/Island Continent	IC	1.48	1.0%	0.29%
South Pacific	SP	99.9	27%	20%
North Pacific	NP	81.8	22%	16%
Indian Ocean	IO	75.4	21%	15%
South Atlantic	SAt	46.5	13%	9.1%
North Atlantic	NAt	43.5	12%	8.5%
Arctic Ocean	AO	10.2	2.8%	2.0%
Caribbean Sea	CS	4.35	1.2%	0.85%
Mediterranean Sea	MS	2.60	0.71%	0.51%
Black Sea	BS	0.472	0.13%	0.09%
Global Land	–	146.7	100%	29%
Global Ocean	–	364.6	100%	71%
Globe	–	511.2	–	100%

2.2. CERES EBAF TOA Edition 4.1 Data Product

Our reference for TOA radiative fluxes is calculated from the Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) TOA Edition 4.1 Data Product (Loeb et al., 2018; NASA/LARC/SD/ASDC, 2019b) covering the period from March 2000 to present. The instruments of CERES are carried on satellites in sun-synchronous orbits and measure filtered radiances in three different wavelengths channels: shortwave (0.3 μm–5 μm), total (0.3 μm–200 μm) and window (8 μm–12 μm). The longwave radiances are determined by subtracting shortwave radiances from the total radiances. TOA radiative fluxes are defined at an optimal reference level of 20 km (Loeb et al., 2002). To produce the radiation in the CERES EBAF TOA Edition 4.1 Data Product, first, CERES Single Scanner Footprint (SSF) TOA/Surface Fluxes Edition 4A Data Products which contains total solar irradiance (TSI) data mainly from the SORCE (Kopp & Lean, 2011) are used as input to produce the CERES SSF 1° (SSF1deg) and Synoptic 1° (SYN1deg) Ed4A Data Products. Then, monthly mean unadjusted outgoing longwave radiation (OLR) is calculated from the CERES SYN1deg daily OLR, while monthly mean unadjusted outgoing shortwave radiation (OSR) is calculated by applying predetermined empirical diurnal correction ratios to the CERES SSF1deg daily OSR. Finally, an objective constrainment algorithm (Loeb et al., 2009) is used to adjust OLR and OSR within their uncertainty range to match their TOA imbalance with the Earth's energy imbalance result provided by Johnson et al. (2016) inferred from the in-situ Argo ocean measurements, which results in the CERES EBAF TOA Edition 4.1 Data Product.

Loeb et al. (2018) estimated the overall uncertainties (1σ) of monthly OSR and OLR in $1° \times 1°$ grid box by combining all known uncertainty sources (the EBAF diurnal correction, radiance-to-flux conversion error (Su et al., 2015), and CERES instrument calibration uncertainty) assuming their independency from each other. Monthly $1° \times 1°$ gridded OSR and OLR uncertainties (1σ) are both approximately 3 Wm^{-2}. These uncertainties are not, however, representative for some regions. For example, the uncertainties of OSR and OLR are higher in the terrestrial convective regions and marine stratocumulus regions because of the resulting strong diurnal cycles (Loeb et al., 2018; Taylor, 2012).

2.3. CERES EBAF Surface Edition 4.1 Data Product

We use the CERES EBAF Surface Edition 4.1 Data Product (Kato et al., 2018; NASA/LARC/SD/ASDC, 2019a) covering the period from March 2000 to present as an additional reference for surface radiation. This is also a $1° \times 1°$ satellite-derived data product, making use of the information contained in the CERES SYN1deg-Month Edition 4A Data Product (Rutan et al., 2015; Wielicki et al., 1996) and the CERES EBAF TOA Edition 4.1 Data Product (Loeb et al., 2018) as input for radiation. The surface radiation is constrained based on bias correction (in upper-tropospheric temperature and specific humidity, and cloud fraction) and a Lagrange multiplier process, to reduce the difference between the CERES SYN1deg-Month TOA radiation and CERES EBAF TOA radiation (Kato et al., 2013, 2018).

To estimate the monthly $1° \times 1°$ gridded uncertainty (1σ) of upward shortwave radiation (USR), upward longwave radiation (ULR), downward shortwave radiation (DSR), and downward longwave radiation (DLR) over land, oceans, Antarctica (60°–90°S) and the Arctic Ocean (60°–90°N) in the EBAF Surface data set, Kato et al. (2018) used the root mean square differences (RMSDs) between EBAF Surface monthly Data and surface observation to represent the monthly uncertainties. Then they compared the results with the monthly uncertainties inferred by Kato et al. (2012, 2013) and the monthly uncertainties inferred using perturbation method as in Zhang et al. (1995). Limited by the number and spatial distribution of observation sites, it is not possible to directly evaluate the uncertainty on a larger spatial scale. Kato et al. (2018) made the best use of the available sites through grouping them randomly and equally, then they used fitted lines to represent the RMSD as a function of the number of sites in different sets of groups, and made assumptions that the order of 100 (10^4) sites approximately corresponds to the uncertainty of monthly zonal (global) mean radiation. After combining with the uncertainties given by Kato et al. (2013), the final results are listed in Table 8 of Kato et al. (2018).

We combine the CERES EBAF TOA Edition 4.1 Data Product and the CERES EBAF Surface Edition 4.1 Data Product to calculate the CERES product reference estimates of the atmospheric radiative components, namely the atmospheric net shortwave radiation (atmospheric net SW radiation, the SW radiation absorbed by the atmosphere) and atmospheric net longwave radiation (atmospheric net LW radiation, the LW radiation emitted by the atmosphere to the outer space).

2.4. CMIP6 and CMIP5 Models

The CMIP6 historical simulations (Eyring et al., 2016) are externally driven by both natural and anthropogenic forcings based on observational data sets covering the period 1850–2014. We select 53 models which contain all energy budget components required for our study from CMIP6 historical simulations over the period 2000–2009. Different models contain different numbers of ensemble simulations. Therefore, to equally weight each model in the multi-model ensemble mean (MEM), we first calculate ensemble means within each model as the model mean value, and then average all model mean values to get a MEM.

We also draw a comparison between the CMIP6 and CMIP5 historical simulations over the period 2000–2005 because the CMIP5 historical simulation only reach up to 2005. The comparison of MEM is based on a two-sample *t*-test for equal means at 0.05 significance level. To take individual model into account, we use the mean absolute bias between individual CMIP6/CMIP5 models' ensemble means and our reference data sets. In terms of the degree of consistency among different CMIP6/CMIP5 models, we look at the multi-model standard deviation of ensemble means and the inter-model spread of ensemble means.

The 53 CMIP6 and 46 CMIP5 models used in our study are listed in Table 2 and Table 3, respectively.

2.5. Spatial Decomposition and Time Span

We apply the same spatial decomposition with the resolution of $0.25° \times 0.25°$ as in the NEWS data product and investigate the Earth's energy budget in 16 land and ocean regions as depicted in Figure 1. The area of each region, its percentage of global land or ocean area, as well as its percentage of the area of the entire globe are listed in Table 1. As the spatial resolution of the CERES EBAF data product is $1° \times 1°$ and the spatial resolution differs among the various CMIP6 and CMIP5 models, we use a second-order conservative remap method (Jones, 1999) to remap those data onto $0.25° \times 0.25°$ grids so that they match with the NEWS map mask.

To be consistent with the NEWS data product, we focus on the 10-year annual means of the model-calculated energy flux fields covering the period from 2000 to 2009. Since the CERES EBAF TOA and Surface data

Table 2
The 53 CMIP6 Models Used in Our Study With Their Corresponding Institution, Number of Ensembles and Horizontal Grid Resolution in Terms of Number of Longitudinal and Latitudinal Gridpoints

Institution	Model name	Number of ensembles	Horizontal grid (lon × lat)
Commonwealth Scientific and Industrial Research Organisation, Australia	ACCESS-CM2	3	192 × 144
	ACCESS-ESM1-5	10	192 × 145
Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Germany	AWI-CM-1-1-MR	5	384 × 192
	AWI-ESM-1-1-LR	1	192 × 96
Beijing Climate Center, China	BCC-CSM2-MR	3	320 × 160
	BCC-ESM1	3	128 × 64
Chinese Academy of Meteorological Sciences, China	CAMS-CSM1-0	3	320 × 160
Chinese Academy of Sciences, China	CAS-ESM2-0	4	256 × 128
National Center for Atmospheric Research, Climate and Global Dynamics Laboratory, USA	CESM2-FV2	3	144 × 96
	CESM2-WACCM-FV2	3	144 × 96
	CESM2-WACCM	3	288 × 192
	CESM2	11	288 × 192
Department of Earth System Science, Tsinghua University, China	CIESM	3	288 × 192
Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici, Italy	CMCC-CM2-HR4	1	288 × 192
	CMCC-CM2-SR5	1	288 × 192
Centre National de Recherches Meteorologiques, France; Centre Europeen de Recherche et de Formation Avancee en Calcul Scientifique, France	CNRM-CM6-1-HR	1	Reduced Gaussian grids with 181,724 grid points over 360 latitude circles
	CNRM-CM6-1	29	Reduced Gaussian grids with 24,572 grid points over 128 latitude circles
	CNRM-ESM2-1	9	Reduced Gaussian grids with 24,572 grid points over 128 latitude circles
Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada, Canada	CanESM5-CanOE	3	128 × 64
	CanESM5	65	128 × 64
E3SM Project	E3SM-1-0	5	Cubed sphere spectral-element grid 90 × 90 × 6 longitude/latitude/cubeface
	E3SM-1-1-ECA	1	Cubed sphere spectral-element grid 90 × 90 × 6 longitude/latitude/cubeface
	E3SM-1-1	1	Cubed sphere spectral-element grid 90 × 90 × 6 longitude/latitude/cubeface
EC-Earth Consortium	EC-Earth3-Veg-LR	3	320 × 160
	EC-Earth3-Veg	6	512 × 256
	EC-Earth3	73	512 × 256
Chinese Academy of Sciences, China	FGOALS-f3-L	3	360 × 180
	FGOALS-g3	6	180 × 80
National Oceanic and Atmospheric Administration, Geophysical Fluid Dynamics Laboratory, USA	GFDL-CM4	1	360 × 180
	GFDL-ESM4	1	360 × 180
Goddard Institute for Space Studies, USA	GISS-E2-1-G-CC	1	144 × 90
	GISS-E2-1-G	44	144 × 90
	GISS-E2-1-H	23	144 × 90
Met Office Hadley Centre, UK	HadGEM3-GC31-LL	4	192 × 144

Table 2
Continued

Institution	Model name	Number of ensembles	Horizontal grid (lon × lat)
	HadGEM3-GC31-MM	4	432 × 324
	UKESM1-0-LL	18	192 × 144
Centre for Climate Change Research, Indian Institute of Tropical Meteorology Pune, India	IITM-ESM	1	192 × 94
Institute for Numerical Mathematics, Russian Academy of Science, Russia	INM-CM4-8	1	180 × 120
	INM-CM5-0	10	180 × 120
Institut Pierre Simon Laplace, France	IPSL-CM6A-LR	32	144 × 143
National Institute of Meteorological Sciences/Korea Meteorological Administration, Climate Research Division, Republic of Korea	KACE-1-0-G	3	192 × 144
Japan Agency for Marine-Earth Science and Technology, Japan; Atmosphere and Ocean Research Institute, The University of Tokyo, Japan; National Institute for Environmental Studies, Japan; RIKEN Center for Computational Science, Japan	MIROC6	50	256 × 128
	MIROC-ES2L	10	128 × 64
Max Planck Institute for Meteorology, Germany	MPI-ESM1-2-HR	10	384 × 192
	MPI-ESM1-2-LR	10	192 × 96
	MPI-ESM-1-2-HAM	2	192 × 96
Meteorological Research Institute, Japan	MRI-ESM2-0	6	320 × 160
Nanjing University of Information Science and Technology, China	NESM3	5	192 × 96
NorESM Climate modeling Consortium	NorCPM1	30	144 × 96
	NorESM2-LM	3	144 × 96
	NorESM2-MM	3	288 × 192
Seoul National University, Republic of Korea	SAM0-UNICON	1	288 × 192
Research Center for Environmental Changes, Academia Sinica, Taiwan	TaiESM1	1	288 × 192

products start from March 2000, we set the end of our investigation period of the CERES EBAF data products to February 2010. The first decade of the 21st century excludes the strong El Niño events of 1997–1998 and 2015–2016, and falls within the hiatus phase of the increase in global mean surface temperature (Trenberth & Fasullo, 2013a). The hiatus can be attributed to the natural variability associated with the switch in the sign of Pacific Decadal Oscillation which compensates the rise of GMST caused by the increase of greenhouse gases (Trenberth, 2015). As the CMIP models are freely evolving, it is highly unlikely that they show a hiatus due to internal variability at precisely the same years as the hiatus we observed in the real world. The quantitative effect of the hiatus on the different energy budget components at global and, especially, regional scales is a topic of ongoing research which we do not further pursue in this study.

3. Results and Discussion

We start with a discussion of the TOA energy fluxes in Section 3.1, before turning to the surface radiative and turbulent heat fluxes in Sections 3.2 and 3.3, respectively. The net energy imbalance at the surface is further examined in Section 3.4. Then, we investigate the atmospheric radiative fluxes and finally dive into the comparison between CMIP6 and CMIP5 models. In each subsection, we typically proceed from large to small spatial scales, from global means to global land or ocean means, from large regions to small regions. The sign convention of the energy components used in our study is that the upward flux has a negative sign while the downward flux is positive. The biases between models and reference data sets, as well as differences among reference data sets are calculated with this sign convention. Assessments of overestimation and underestimation as well as statements referring to higher or lower flux values are all based on the absolute magnitude of the energy fluxes. An overview of quantitative biases between CMIP6 models and our reference data sets is given in Figures 2 and 3. Maps providing a geographical impression of the differences between NEWS and CERES EBAF are given in

Table 3
The 46 CMIP5 Models Used in Our Study With Their Corresponding Institution, Number of Ensembles and Horizontal Grid Resolution in Terms of Number of Longitudinal and Latitudinal Gridpoints

Institution	Model name	Number of ensembles	Horizontal grid (lon × lat)
Commonwealth Scientific and Industrial Research Organization and Bureau of Meteorology, Australia	ACCESS1-0	3	192 × 145
	ACCESS1-3	3	192 × 145
Beijing Climate Center, China Meteorological Administration, China	BCC-CSM1-1-M	3	320 × 160
	BCC-CSM1-1	3	128 × 64
Beijing Normal University, China	BNU-ESM	1	128 × 64
US National Centre for Atmospheric Research, USA	CCSM4	8	288 × 192
National Science Foundation; Department of Energy; National Center for Atmospheric Research, USA	CESM1-BGC	1	288 × 192
	CESM1-CAM5-1-FV2	4	144 × 96
	CESM1-CAM5	3	288 × 192
	CESM1-FASTCHEM	3	288 × 192
	CESM1-WACCM	7	144 × 96
Centro Euro-Mediterraneo per I Cambiamenti Climatici, Italy	CMCC-CESM	1	96 × 48
	CMCC-CMS	1	192 × 96
	CMCC-CM	1	480 × 240
Centre National de Recherches Meteorologiques, France; Centre Europeen de Recherche et Formation Avancees en Calcul Scientifique, France	CNRM-CM5-2	1	256 × 128
	CNRM-CM5	10	256 × 128
Commonwealth Scientific and Industrial Research Organisation in collaboration with the Queensland Climate Change Centre of Excellence, Australia	CSIRO-Mk3-6-0	10	192 × 96
	CSIRO-Mk3L-1-2	3	64 × 56
Canadian Centre for Climate Modelling and Analysis, Canada	CanESM2	5	128 × 64
LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences, China	FGOALS-g2	5	128 × 60
	FGOALS-s2	2	128 × 108
National Oceanic and Atmospheric Administration, Geophysical Fluid Dynamics Laboratory, USA	GFDL-CM3	5	144 × 90
	GFDL-ESM2G	1	144 × 90
	GFDL-ESM2M	1	144 × 90
NASA Goddard Institute for Space Studies, USA	GISS-E2-H-CC	1	144 × 90
	GISS-E2-H	18	144 × 90
	GISS-E2-R-CC	1	144 × 90
	GISS-E2-R	26	144 × 90
Met Office Hadley Centre, UK	HadCM3	10	96 × 73
	HadGEM2-CC	3	192 × 145
Met Office Hadley Centre, UK; Instituto Nacional de Pesquisas Espaciais, Brasil	HadGEM2-ES	5	192 × 145
Institute for Numerical Mathematics, Russian Academy of Science, Russia	INM-CM4	1	180 × 120
Institut Pierre Simon Laplace, France	IPSL-CM5A-LR	6	96 × 96
	IPSL-CM5A-MR	3	144 × 143
	IPSL-CM5B-LR	1	96 × 96
University of Tokyo, National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology, Japan	MIROC4h	3	640 × 320
	MIROC5	5	256 × 128
	MIROC-ESM-CHEM	1	128 × 64
	MIROC-ESM	3	128 × 64

Table 3
Continued

Institution	Model name	Number of ensembles	Horizontal grid (lon × lat)
Max Planck Institute for Meteorology, Germany	MPI-ESM-LR	3	192 × 96
	MPI-ESM-MR	3	192 × 96
	MPI-ESM-P	2	192 × 96
Meteorological Research Institute, Japan	MRI-CGCM3	5	320 × 160
	MRI-ESM1	1	320 × 160
Norwegian Climate Centre, Norway	NorESM1-ME	1	144 × 96
	NorESM1-M	3	144 × 96

Figure S1 in Supporting Information S1. Likewise, maps giving a geographical view of CMIP6 regional biases can be found in Figures S2–S5 in Supporting Information S1, again in the supplementary material. An overview of the reference values is given in Table 4.

3.1. TOA Radiative Flux

As shown in Figure 2, the CMIP6 MEM incoming solar radiation (ISR) largely agrees with EBAF over all regions with absolute biases of at most 0.1 Wm^{-2}. The CMIP6 multi-model standard deviations are less than 1 Wm^{-2} over all regions (Table S1 in Supporting Information S1). These suggest that the TOA energy input in the climate system in CMIP6 models is essentially accurate on regional scale. It then follows that the biases of TOA net SW radiation are mostly determined by the biases of OSR. CMIP6 multi-model land mean and ocean mean OSR are both higher than those of EBAF (Figure 2, Table S1). Besides, 36 out of 53 CMIP6 models overestimate the global mean OSR with 12 of them out of the uncertainty range of EBAF (Table S2 in Supporting Information S1). On the regional scale, there are consistent overestimations of OSR among individual models over most regions (Figure 3). Exceptions are South America, the South Atlantic Ocean, and the Black Sea.

The biases we find are in line with the literature, notably regarding cloud radiative effects and cloud fraction. The overestimation of MEM OSR can be ascribed to the overestimation of MEM SW cloud radiation

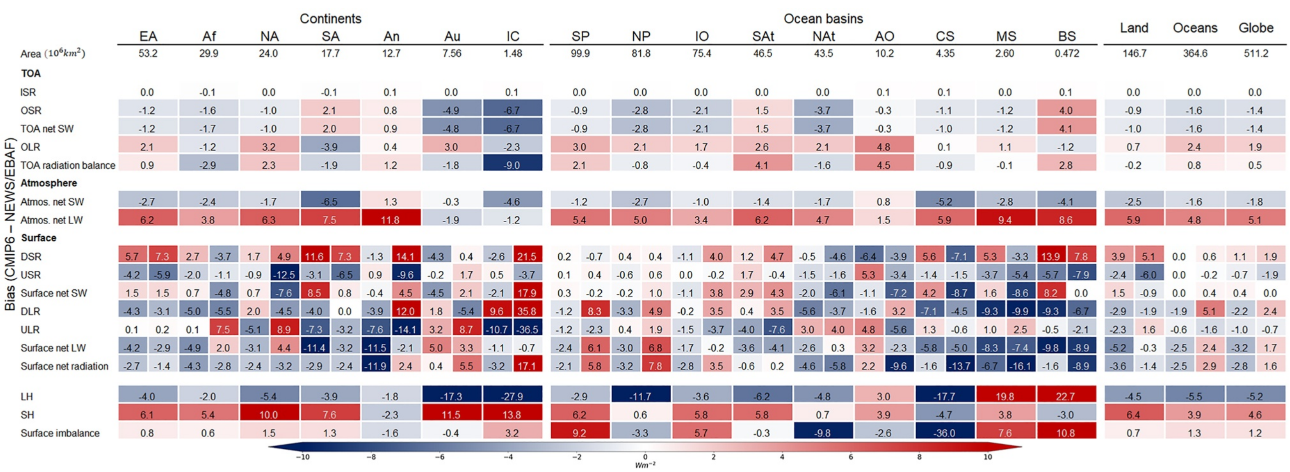


Figure 2. Biases (in Wm^{-2}) between CMIP6 MEM (2000–2009), the NEWS Annual Climatology data product, and the CERES EBAF TOA and Surface data products. The overall structure is the same as in Table 4. From top to bottom, the first panel of the figure represents the TOA biases (CMIP6—EBAF TOA). The second panel represents the atmospheric biases (CMIP6—EBAF). In the third panel, in each region, the left column represents the surface biases (CMIP6—EBAF Surface) and the right column represents the surface biases (CMIP6—NEWS). The values in the lowermost panel represent the biases (CMIP6—NEWS) in the turbulent heat fluxes and the surface imbalance.

10.1029/2022EA002758

Figure 3. Percentage of CMIP6 models with positive biases (individual model's ensemble mean—NEWS/EBAF). The overall structure is the same as Figure 2. In each cell, the percentage of models with positive biases is given and indicated in pink color. The skyblue portion in each cell indicates the percentage of models with negative or zero biases.

effect (CRE) as compared to EBAF (Jian et al., 2020). The overestimations of MEM SWCRE are substantial over the Intertropical Convergence Zone (ITCZ) (Jian et al., 2020), which influences the overestimation of MEM OSR over the Pacific Ocean, the Atlantic Ocean, the Indian Ocean, and Australasian and Indonesian Islands. The overestimation of MEM OSR over Mainland Australia is also in line with Jian et al.'s (2020) finding that overestimations of the cloud albedo forcing (CAF) exist in the CMIP6 MEM over the arid regions in the Southern Hemisphere. Nevertheless, there are notable exceptions over the Black Sea and two large neighboring regions: South America and the South Atlantic, where a majority of models underestimate OSR (Figure 3, Figure S2 in Supporting Information S1). It is difficult to attribute the deficient MEM OSR over South America merely via the distribution of CAF biases and clear sky planetary albedo biases in Jian et al.'s (2020) results, but the temporal correlation of clear sky planetary albedo between the CMIP6 MEM and EBAF shows significant weak or negative values over the land convection region of South America, which points to deficiencies in reproducing the annual cycle of the clear sky planetary albedo by the CMIP6 models (Jian et al., 2020). Over the South Atlantic Ocean, the contrasting underestimation of MEM OSR to the aforementioned overestimation over ITCZ is in line with the underestimated MEM SWCRE over the marine stratocumulus region off the coast of Namibia (Jian et al., 2020, 2021), together with the underestimated SWCRE over the Southern Ocean (40°S-70°S) within the regime representing the stratocumulus cloud (Schuddeboom & McDonald, 2021). The underestimations of cloud fraction (Tselioudis et al., 2021) and cloud albedo (Jian et al., 2021) both contribute to the weak SWCRE of the stratocumulus clouds in CMIP6 models.

Turning to OLR on the global scale, less energy leaves the Earth system in the CMIP6 MEM as compared to EBAF (Figure 2, Table S1 in Supporting Information S1) (see also Wild, 2020). 39 out of 53 CMIP6 models underestimate the global mean OLR with 11 of them out of the uncertainty range of EBAF (Table S2 in Supporting Information S1). OLR is consistently underestimated by most models over most ocean regions (Figures 2 and 3). Land regions show a more mixed picture. For example, OLR is grossly underestimated in North America by most models and in the MEM, while the opposite is true for South America (Figures 2 and 3).

Different causes are examined in the literature. Following Miao et al. (2021), CMIP6 models underestimate LWCRE over the globe primarily due to the underestimation in the relative frequency of occurrence (RFO) of the clouds with high in-cloud ice water path (ICIWP). The lower OLR could then be ascribed to the underestimation of clear-sky OLR (Wild, 2020).

The negative biases of MEM TOA net SW radiation and the positive biases of MEM OLR compensate each other over most regions and lead to smaller biases of the TOA radiation balance in absolute magnitude (Figure 2). On the other hand, over Africa, Antarctica, Australasian and Indonesian Islands, and the South

Atlantic Ocean, these two fluxes result in enhanced biases in the TOA radiation balance. Most prominently over the South Atlantic Ocean, as the result of the superimposition of the deficiencies in both OSR and OLR, 52 out of 53 CMIP6 models have higher TOA radiation balance than EBAF (Figure 3). The heterogeneous pattern of the TOA radiation balance biases on the regional scale is compensated when averaging over land and oceans, which leads to smaller biases with absolute values less than 1 Wm^{-2} as compared to EBAF (Figure 2).

The biases between models and observation can be put into perspective by considering regional differences in the TOA radiation balance: a net energy gain at the TOA in some regions is transported by the atmosphere and/or ocean to other regions where a net energy loss at TOA occurs. The CMIP6 multi-model land mean TOA imbalance is -17.6 Wm^{-2} (Table S1 in Supporting Information S1). The negative sign indicates a net radiative energy loss over land which is compensated by the net atmospheric energy transport from ocean to land. We subtract the CMIP6 multi-model land mean surface imbalance of 0.7 Wm^{-2} from the CMIP6 multi-model land mean TOA radiation balance of -17.6 Wm^{-2} to get the net atmospheric energy transport with the value of -18.3 Wm^{-2} (Table S1 in Supporting Information S1) (neglecting the atmospheric heat storage). Trenberth and Fasullo (2013c) determined the annual total atmospheric energy transport from ocean to land with the value of 2.5 PW, using ERA-Interim reanalysis data covering the period 1979–2010. Dividing this value by the global land area of $146.6 \times 10^{12} m^2$ applied in our study, we get a transport energy flux of -17.1 Wm^{-2}. A new estimate combining satellite and ocean data by Liu et al. (2020) provides a slightly higher value of 2.78 PW covering the period 1985–2010 due to different calculation method and different land surface heat uptake assumptions, and a value of 2.74 PW (-18.7 Wm^{-2}) covering the period 2000–2010 which is within 1 Wm^{-2} from the CMIP6 MEM. Similarly, Wild et al. (2015) determined a net land-ocean energy transport of 2.8 PW (-19.0 Wm^{-2}) based on CERES-EBAF data and found a good agreement with the corresponding transport in the CMIP5 multi-model mean. The good agreement between CMIP6 MEM value and independent estimates suggests that the CMIP6 MEM precisely reproduce the annual atmospheric energy transport from ocean to land.

Turning to individual regions, the most substantial bias in the TOA radiation balance is found on the Australasian and Indonesian Islands with a value of -9.0 Wm^{-2} in the MEM (Figure 2), which amounts to about 19% of its absolute value (Table 4), and with 45 out of 53 CMIP6 models showing negative biases as compared to EBAF (Figure 3). This contrasts with a general positive bias over the oceans in the Southern Hemisphere (the South Pacific Ocean and the South Atlantic Ocean), in terms of MEM as well as in terms of the fraction of individual models showing positive biases (Figures 2 and 3). Moreover, the South Pacific Ocean and the South Atlantic Ocean together cover about 58% of the area of the Southern Hemisphere (Table 1), so the positive biases of TOA radiation balance in the CMIP6 models over these two ocean basins could largely contribute to the difference in the TOA radiation balance between the Northern and the Southern Hemisphere (Loeb et al., 2016; Stephens et al., 2016; Trenberth & Zhang, 2019), which could have an influence on the cross-equatorial energy transport and the mean position of the ITCZ (Frierson et al., 2013; Marshall et al., 2014) in the CMIP6 models.

3.2. Surface Radiative Flux

3.2.1. Comparison Between the NEWS Annual Climatology and CERES EBAF Surface Data Products

We first compare the surface radiative fluxes between our main reference data sets, the NEWS Annual Climatology Data Product and the CERES EBAF Surface Data Product. The goal here is not to evaluate either data set but to explore the target range before comparing these reference data sets with the CMIP6 models below. The comparison of the NEWS and EBAF regional scale surface radiation estimates is summarized in Figure 4. Note that while for global mean, global land, and global ocean, both NEWS and EBAF provide uncertainty estimates, only NEWS provides uncertainty estimates on regional scales.

There is overall good agreement among both data sets on large scales over the globe, global land and global oceans, for all individual radiative components (Figure 4). The only two prominent disagreements are the land mean USR and the ocean mean DLR where EBAF is out of the uncertainty range of NEWS. For the ocean mean DLR, the uncertainty ranges of the two data sets do not contain each other's annual mean value (Table 8 of Kato et al. (2018) and Table S1 in Supporting Information S1).

10.1029/2022EA002758

Table 4
Magnitudes of the Energy Budget Components (in Wm^{-2}*) as Given by the NEWS Annual Climatology Data Product As Well As the CERES EBAF TOA and Surface Data Products*

		Continents													
		EA		Af		NA		SA		An		Au		IC	
NEWS/EBAF (Wm^{-2})	Area(10^6 km^2)	53.2		29.9		24.0		17.7		12.7		7.56		1.48	
	TOA														
	ISR	307.4		397.0		282.4		394.0		186.0		378.1		394.1	
	OSR	−106		−116		−104		−120		−128		−94		−123	
	TOA net SW	202		281		178		274		58		284		271	
	OLR	−231		−262		−221		−242		−159		−271		−223	
	TOA radiation balance	−29		19		−43		32		−101		13		48	
	Atmosphere														
	Atmos. net SW	71		95		62		96		29		80		96	
	Atmos. net LW	−169		−180		−169		−193		−124		−176		−180	
	Surface														
	DSR	166	165	241	248	155	152	207	211	140	124	245	241	195	171
	USR	−35	−34	−55	−56	−39	−27	−29	−26	−111	−100	−42	−44	−20	−16
	Surface net SW	131	131	186	192	116	124	178	186	29	24	203	197	175	155
	DLR	301	300	370	370	281	287	381	377	145	129	353	360	384	358
	ULR	−364	−364	−452	−460	−333	−347	−430	−434	−180	−173	−448	−454	−428	−402
	Surface net LW	−62	−63	−83	−90	−52	−60	−49	−57	−35	−44	−95	−93	−43	−44
	Surface net radiation	69	68	104	102	64	65	129	129	−6	−20	109	104	132	111
	LH	−34		−45		−33		−77		−1		−27		−75	
	SH	−34		−58		−32		−52		21		−77		−36	
	Surface imbalance	0		0		0		0		0		0		0	

Note. The columns from left to right represent seven continents, nine ocean basins and the spatial mean values over land, oceans and the globe. The regions are sorted from left to right in descending order of area size (in $10^6 km^2$) as provided in the first row. Rows contain different energy budget components. From top to bottom, the table is divided into four panels. The first panel of the table represents the TOA components from the CERES EBAF TOA product. The second panel represents the atmospheric components calculated by combining the CERES EBAF TOA and Surface products. In the third panel, the left and right columns in each region represent the surface components from CERES EBAF Surface product and the NEWS Annual Climatology data product, respectively. The values in the lowermost panel represent the turbulent heat fluxes and surface imbalance from the NEWS Annual Climatology data product.

On the level of individual regions, good agreement between NEWS and EBAF exists across all surface radiative components for Eurasia, Mainland Australia, the Indian Ocean, and the North Atlantic Ocean. Meanwhile, prominent differences are found in the Pacific Ocean (DLR and surface net LW radiation), in the South Atlantic Ocean (USR and ULR), as well as over the Americas (all radiative fluxes except DSR and surface net radiation). In terms of components, there is generally good agreement between NEWS and EBAF for DSR.

We can only speculate on the reason for the aforementioned differences. The difference of land mean USR is mainly caused by the differences over North America, South America, and Antarctica. Over the two Americas, where the two data sets are compatible in terms of DSR but not for USR, it seems plausible that surface albedo plays a role in the disagreement, in North America possibly via snow cover. Dwelling further on the two Americas, significant differences occur not only in the SW components but also in the LW components. However, the higher surface net SW radiation and the lower surface net LW radiation compensate each other, which results in small differences in the surface net radiation with an absolute magnitude of less than 1 Wm^{-2}.

Over Antarctica, the NEWS's lower USR is mostly caused by its lower DSR. The difference in their LW counterpart is similar to that in the SW. It is noteworthy that the disagreement in DLR over Antarctica is even more

10.1029/2022EA002758

Ocean basins																							
SP		NP		IO		SAt		NAt		AO		CS		MS		BS		Land		Oceans		Globe	
99.9		81.8		75.4		46.5		43.5		10.2		4.35		2.60		0.472		146.7		364.6		511.2	
347.2		371.7		351.3		333.5		332.5		187.6		393.6		339.7		313.0		326.1		346.0		340.3	
−95		−98		−94		−98		−88		−103		−81		−71		−86		−111		−95		−99	
253		274		258		235		244		84		313		269		227		215		251		241	
−244		−246		−245		−241		−243		−199		−263		−258		−239		−233		−243		−240	
9		28		13		−6		1		−115		49		10		−11		−17		8		0.8	
77		87		79		74		77		44		94		76		72		74		79		77	
−194		−198		−192		−193		−191		−164		−210		−185		−177		−171		−193		−187	
189	190	198	198	192	187	178	175	179	183	99	97	231	244	206	215	166	172	187	186	187	186	187	186
−13	−14	−11	−12	−13	−13	−17	−15	−12	−12	−59	−50	−12	−12	−13	−11	−11	−9	−46	−42	−14	−14	−23	−22
175	176	187	186	179	174	161	160	167	171	41	47	218	231	193	204	155	164	141	143	173	172	164	164
359	349	383	375	359	355	339	336	358	357	232	228	407	404	343	344	330	327	312	311	359	353	345	341
−408	−407	−432	−433	−411	−409	−388	−384	−410	−411	−268	−258	−460	−458	−417	−419	−392	−390	−374	−378	−408	−407	−398	−399
−50	−58	−48	−58	−52	−54	−48	−48	−52	−55	−36	−30	−53	−54	−74	−75	−62	−63	−62	−67	−50	−55	−53	−58
126	118	138	127	127	120	113	112	115	116	5	17	165	177	119	129	93	101	79	76	123	117	110	106
−99		−105		−106		−83		−98		−10		−125		−113		−87		−38		−98		−81	
−20		−18		−21		−19		−20		−7		−12		−24		−22		−38		−19		−25	
−2		4		−6		10		−3		0		40		−8		−8		0		0.6		0.4	

substantial because the mean value of either data set falls outside the uncertainty range of the other data set (Table 8 of Kato et al. (2018) and Table S1). It should be kept in mind, however, that precise estimates of surface energy budget components for Antarctica remain challenging. Kato et al. (2018) compared EBAF with four sites in Antarctica. They inferred mean biases of monthly mean DLR and DSR of 3.1 Wm^{-2} and −4.1 Wm^{-2} respectively, which are both within the uncertainty of surface observations. However, the root mean square errors (RMSEs) of monthly mean DLR and DSR are 11.7 Wm^{-2} and 20.1 Wm^{-2}, respectively. This is caused by the large temporal and spatial variability of surface radiation over polar regions. The poor agreement between the two data sets over the Australasian and Indonesian Islands might be linked to the heterogeneity of the region which could not be resolved by the original $1° \times 1°$ grid box of EBAF.

The substantial disagreement in ocean mean DLR in Figure 4 can be further tracked to the Pacific Ocean, where the EBAF values exceed the NEWS values by nearly 10 Wm^{-2}. According to Kato et al. (2018), there is a significant improvement in EBAF edition 4 as compared to its previous edition in the nighttime DLR over the Pacific Ocean, where the hourly mean bias of nighttime DLR amounts to only 1 Wm^{-2}. There also exist other significant disagreements over the ocean basins, mostly in USR and ULR. For instance, the magnitude of the difference of ULR over the Arctic Ocean is about one order higher than over other ocean basins, considering its relatively low

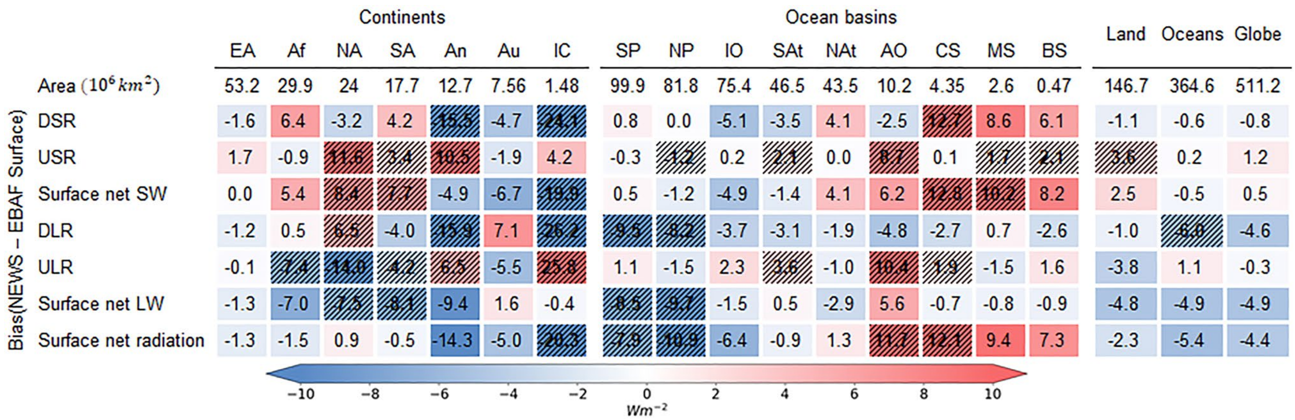


Figure 4. Differences (NEWS—EBAF in Wm^{-2}) between the NEWS Annual Climatology data product and the CERES EBAF Surface data product. Hatching indicates EBAF being out of the uncertainty range of NEWS.

absolute value (Table 4). Another significant disagreement is the substantially higher DSR over the Caribbean Sea in the NEWS data set, which further contributes to the excessive surface imbalance (Table 4) in NEWS over the Caribbean Sea as indicated by Thomas et al. (2020). These disagreements limit the model evaluation in these regions.

3.2.2. Comparison Between CMIP6 Models and Reference Data Sets

In this section, we structure the presentation of our findings as follows: (a) we proceed from global mean, global land and ocean mean where both NEWS and EBAF provide uncertainty range, to regional mean where we have the uncertainty range of NEWS and the annual mean value of EBAF. (b) In view of the disagreement for at least some regions and radiative components (indicated by the same hatching of cells in Figures 4–6), we consider for this comparison two pairs of uncertainty ranges depending on the spatial scale: the uncertainty range of NEWS and EBAF on global, global land and ocean scale (Figures 5 and 7); the uncertainty range of NEWS (Figure 5) as well as the largest range bounded by the end values of NEWS uncertainty range and the mean value of EBAF (Figure 6) on regional scale. (c) In the assessment, we start from the components in the regions where the CMIP6 MEMs are out of the uncertainty range decided by (a) and (b) (indicated by "±" in the corresponding cells of Figures 5 and 6), then we dwell upon the ensemble means of individual CMIP6 models.

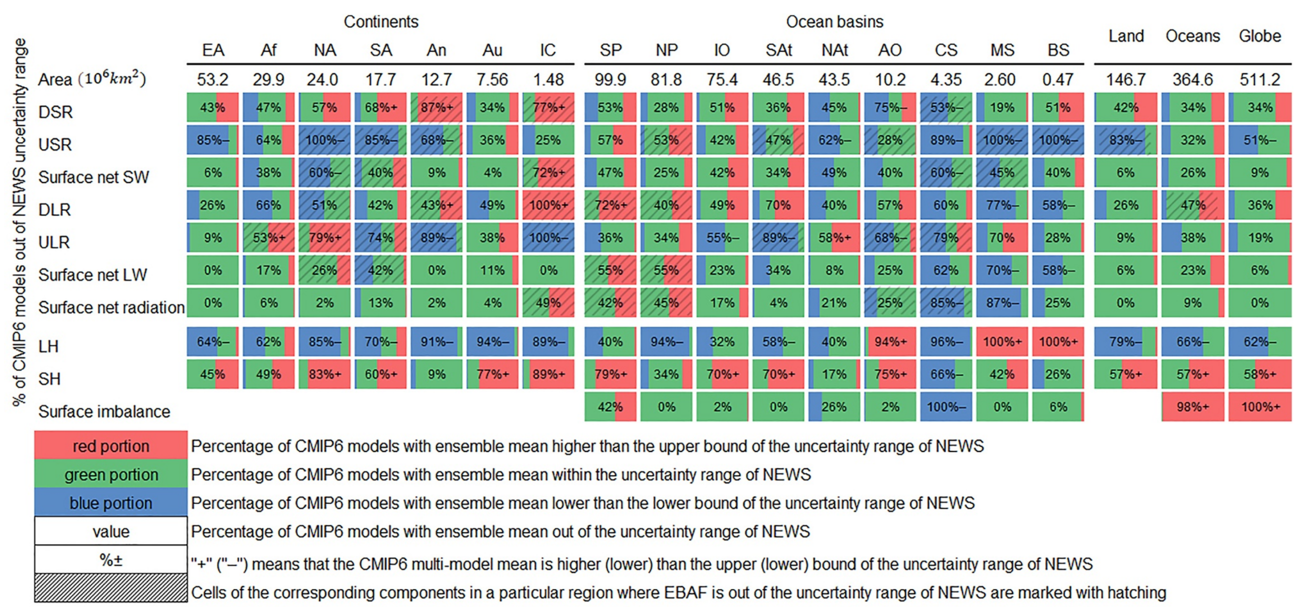


Figure 5. Comparison of the surface energy budget components between the CMIP6 models and the uncertainty range of the NEWS Annual Climatology data product.

10.1029/2022EA002758

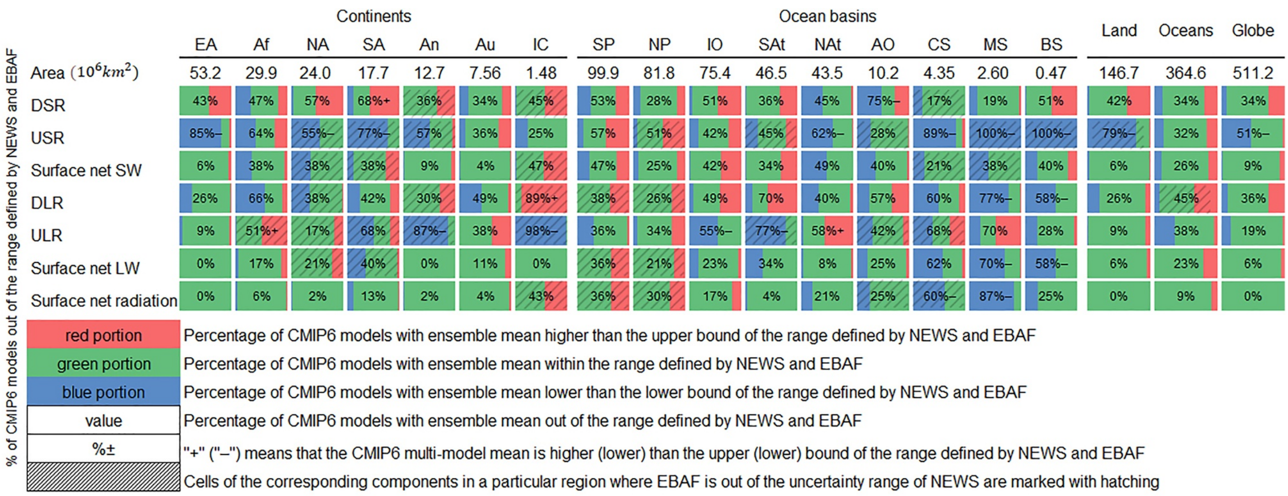


Figure 6. Comparison of the surface radiative components between CMIP6 models and the largest range bounded by the end values of the uncertainty ranges of the NEWS Annual Climatology data product and the annual mean value of the CERES EBAF Surface data product.

Over the global land, the most prominent difference occurs in USR where 43 out of 53 CMIP6 models provide high values of USR which are out of the uncertainty range of NEWS (Figure 5, Table S3 in Supporting Information S1). Even though the uncertainty range of EBAF in global land mean USR shifts toward higher value and is twice in size as compared to that of NEWS, there are still nine out of 53 CMIP6 models which have high values of USR that are out of the uncertainty range of EBAF (Figure 7, Table S2 in Supporting Information S1). This can be ascribed to the high CMIP6 MEM USR over Eurasia (overestimated by 43 out of 53 CMIP6 models as shown in Figure 5 and Table S3 in Supporting Information S1) and the two Americas (Figures 5 and 6), which is possibly caused by the excessive DSR over these regions (Figures 2 and 5). While over Eurasia, there are almost twice as many CMIP6 models that overestimate the USR as those that overestimate the DSR, which suggests that the higher albedo is also a contributing factor. The land mean DLRs of CMIP6 models are slightly underestimated as depicted in Figures 2 and 5, which can be attributed to the underestimation of DLR over the two largest continents in our study (Eurasia and Africa). Comparing to the uncertainty range of EBAF, even more CMIP6 models overestimate global land mean DSR (28 out of 53 models) and underestimate global land mean DLR (20 out of 53 models) than comparing to NEWS (Figure 7, Table S2 in Supporting Information S1). The ULRs over South America and Antarctica are significantly overestimated in CMIP6 models with MEMs as well as 35 and

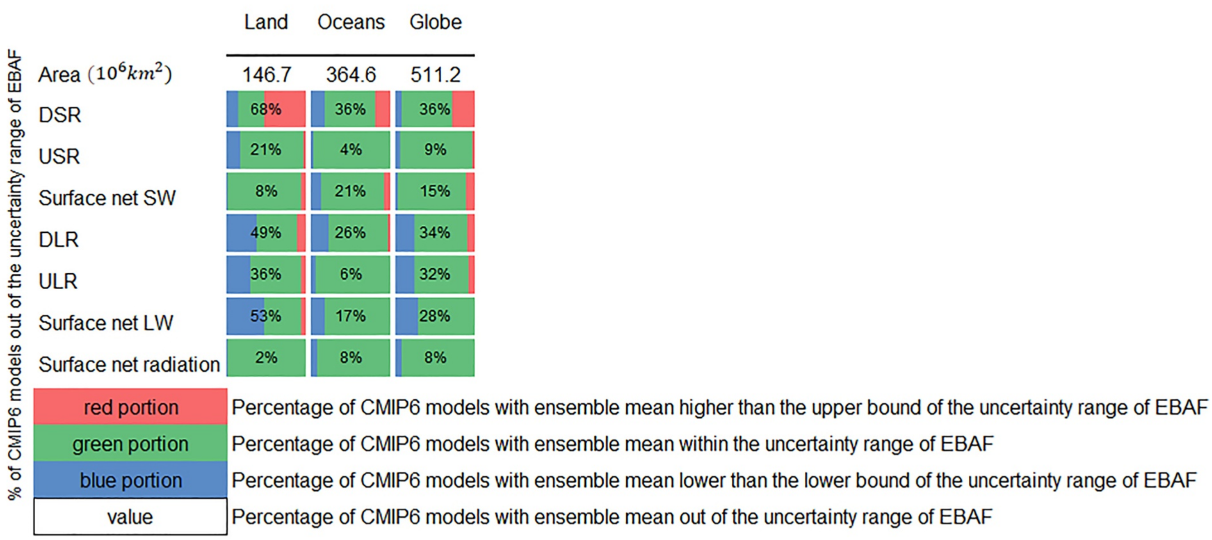


Figure 7. Comparison of the surface radiative components between CMIP6 models and the uncertainty range of EBAF.

46 models out of the uncertainty range, respectively (Figures 2 and 6, Table S4 in Supporting Information S1). Although the CMIP6 MEM ULR over Africa is also out of the uncertainty range and 26 out of 53 CMIP6 models have higher absolute values (Figure 6 and Table S4 in Supporting Information S1), the bias between CMIP6 MEM and EBAF is only 0.1 Wm^{-2}, which is even one order lower than the annual uncertainty of EBAF over global land, so we consider it as an agreement between CMIP6 models and reference data sets.

Over the global oceans, because of the compensating effect of positively and negatively biased models (Figure 5), the CMIP6 multi-model ocean mean SW components (DSR, USR, and surface net SW radiation) are all within 1 Wm^{-2} from the NEWS (Figure 2). On the regional scale, the compensating effect remains over the Pacific Ocean. 24 and 18 out of 53 CMIP6 models overestimate DSR over the Indian Ocean and the South Atlantic Ocean, respectively (Figure 5, Table S3 in Supporting Information S1), while the opposite is true over the North Atlantic Ocean where 22 CMIP6 models underestimate DSR. Moreover, 32 CMIP6 models overestimate USR over the North Atlantic Ocean so that the CMIP6 MEM USR is also out of the uncertainty range of NEWS in this region. As for the LW components, the global ocean mean DLR of the CMIP6 models shows an opposite pattern when compared to NEWS and EBAF (Figures 6 and 7), which is primarily due to the significant difference between NEWS and EBAF over the Pacific ocean (Section 3.2.1). Figure 5 shows a similar pattern regarding DLR over the Indian Ocean and the Atlantic Ocean compared to the one of DSR. The global ocean mean ULR of the CMIP6 models are biased toward overestimation (Figures 2 and 5). On the regional scale, the overestimation of ULR can be tracked to the Ocean basins mainly located in the Southern Hemisphere (the South Pacific Ocean, the Indian Ocean and the South Atlantic Ocean). Especially over the Indian Ocean and the South Atlantic Ocean, the CMIP6 MEM ULRs are overestimated and out of the uncertainty range (Figures 5 and 6). According to the Stefan-Boltzmann law, the overestimation of ULR suggests an overestimation of sea surface temperature (SST) in the Southern Hemisphere in the CMIP6 models. Considering our finding of the overestimated ULR over South America and Antarctica, the surface skin temperature in the CMIP6 models is overestimated over most of the Southern Hemisphere. On the contrary, the ULR over the North Atlantic Ocean is rather underestimated and out of the uncertainty range (Figure 5) both in terms of MEM and individual models (29 out of 53) (Table S3 in Supporting Information S1).

3.3. Turbulent Heat Flux

The partitioning of surface net radiative energy between latent heat (LH) and sensible heat (SH) flux is of key relevance as it ties the radiative fluxes to the water cycle. Models are known to struggle in this respect (Li et al., 2021; Mueller & Seneviratne, 2014; Wang et al., 2021; Wild, 2020), and turbulent heat flux products are known to come with substantial uncertainties (Brunke et al., 2002, 2011; Rannik et al., 2016). Our analysis is overall in line with these common places. Our primary reference data set here is again NEWS. However, as indicated originally in L'Ecuyer et al. (2015) and also in Thomas et al. (2020), the SH fluxes in the satellite-based input data sets for the NEWS results are higher than other estimates based on reanalyses and models. Therefore, considering the possibly substantial variation between different turbulent heat flux data sets, we also compare our results with respect to other reference data sets.

The LH fluxes are overall overestimated on global land, global ocean and regional scale in the CMIP6 models (Figures 2 and 5). The only exceptions are the Arctic Ocean and the two smallest ocean basins (the Mediterranean Sea and the Black Sea). On the contrary, the SH fluxes are overall underestimated when comparing CMIP6 models with the NEWS data set (Figures 2 and 5). Several exceptions are Antarctica, the North Pacific Ocean, the North Atlantic Ocean, the Caribbean Sea and the Black Sea.

Over the land regions, the overall overestimation of LH flux in CMIP6 models is in line with Li et al. (2021) and Wang et al. (2021). Note that Antarctica is excluded in their studies. As for the terrestrial SH flux, the bias between the CMIP6 multi-model median and the mean value of multiple land turbulent flux products shows a mixed pattern (Figure 5 of Li et al. (2021)). One result on the continental scale from Li et al. (2021) is that the CMIP6 multi-model median overestimates SH flux over South America. Our results in Figure 5 and also our CMIP6 multi-model median in Table S1 rather suggest that the SH fluxes over South America are underestimated by the CMIP6 models. This demonstrates that even opposite conclusions can be reached with different turbulent heat flux products as reference data sets, and it shows the need for further improvement of corresponding data products. Further comparison between CMIP6 models and land turbulent flux products need to be pursued in the future.

For the ocean regions, we took as an additional reference data set the average of multiple ocean turbulent flux products from Tables 2 and 3 of Thomas et al. (2020). For LH flux, the comparison gives similar result as with

NEWS: CMIP6 models tend to overestimate the LH flux over the globe and global ocean, which can be attributed to the overestimation over large ocean basins. However, as for SH flux, the underestimation over the globe and the global ocean obtained with respect to NEWS turns into a better agreement between most of the models and our additional reference. Our additional reference generally has lower SH flux over all the compared ocean basins than NEWS, so the CMIP6 models which underestimate SH flux as compared to NEWS tend to agree with our new reference. Meanwhile, those CMIP6 models that are in agreement with NEWS overestimate SH flux when compared to the Thomas et al. (2020) reference.

In summary, the comparisons of LH flux over both land and ocean regions between CMIP6 models and different reference data sets are consistent in the sense that CMIP6 models tend to overestimate the LH, whereas the SH flux counterpart involves ambiguousness as different turbulent heat flux data sets vary substantially. Following Wild (2020) who found substantial CMIP6 inter-model spread of SH flux on global scale, in our study on regional scales, the CMIP6 inter-model spread of SH flux amounts to at least 50% of the CMIP6 MEM (Table S5 in Supporting Information S1).

3.4. Surface Imbalance

Being the combination of all surface energy flux components, the surface imbalance inherits associated issues as described in more detail in the previous sections. Consequently, we restrict ourselves in the following mostly to summarizing the biases as such, without giving much room to potential underlying causes.

We compared the surface imbalance of CMIP6 models with the uncertainty range of NEWS. The NEWS data set constrains the surface imbalance over land regions to zero and the surface imbalance over ocean regions to match with the ocean heat content with the value of $0.6 \pm 0.4\ \mathrm{Wm^{-2}}$ measured by Argo array (Lyman et al., 2010; Willis et al., 2009). As shown in Figure 5, all CMIP6 models provide higher surface imbalance over the globe than the upper bound of the uncertainty range of NEWS. Over the global ocean, all except only one model is within the uncertainty range of NEWS. The CMIP6 multi-model ocean mean surface imbalance with the value of $1.9\ \mathrm{Wm^{-2}}$ (Table S1 in Supporting Information S1) is out of the uncertainty range of the ocean heat content measured by Argo array.

On regional scale, we calculate the uncertainty range of the surface imbalance based on error propagation, assuming that the energy budget components are independent. Then, the uncertainty range of the surface imbalance over the individual ocean regions is typically higher than that over the global ocean. Thus, over most of the ocean regions, the CMIP6 models are within the uncertainty range of NEWS (Figure 5). Exceptions are the South Pacific Ocean, the North Atlantic Ocean, and the Caribbean Sea.

Over the South Pacific Ocean, the high surface imbalance of many of the CMIP6 models may be attributed to the high DLR and low SH as compared to NEWS, but two issues remain. First, the NEWS and EBAF do not agree with each other in DLR over the South Pacific Ocean (Section 3.2.1) where EBAF has higher DLR than NEWS. Note that Figure 10 of Kato et al. (2018) shows that EBAF overestimates DLR over the South Pacific Ocean as compared to observations taken from buoys, but the comparison is limited to tropical regions and may not be representative for the whole region. Therefore, we could not simply reach the conclusion that CMIP6 models overestimate the DLR over the South Pacific Ocean. Second, as indicated in Section 3.3, NEWS tends to overestimate SH and if we change the reference to Thomas et al. (2020), 39 out of 53 CMIP6 models are within the uncertainty range over the South Pacific Ocean, so we cannot conclude either that CMIP6 models underestimate the SH over the South Pacific Ocean.

As for the disagreement over the North Atlantic Ocean and the Caribbean Sea, Thomas et al. (2020) compared NEWS with the surface imbalance calculated through combining CERES TOA data set with ERA-Interim reanalysis, and they found that the surface imbalances over the North Atlantic Ocean, the Caribbean Sea as well as the Arctic Ocean from the NEWS solution is inconsistent with the fact that these ocean regions receive a substantial amount of heat from the ocean overturning circulation, so these regions should be losing more heat on the surface than the NEWS solution suggests. In this case, the comparisons of the surface imbalance between CMIP6 models and NEWS over those three ocean regions are not tenable anymore.

3.5. Atmospheric Radiative Flux

Our reference data set for the atmospheric radiative fluxes is based on the combination of the CERES EBAF TOA and CERES EBAF Surface data sets.

As shown in Figure 2, CMIP6 MEM atmospheric net SW radiations have lower values than EBAF over global land, global ocean and the globe. We further compared atmospheric net SW radiations of individual CMIP6 models with the uncertainty range inferred from Kato et al. (2018) on global scale (Table S2 in Supporting Information S1). The result indicates that 48 out of 53 CMIP6 models simulate atmospheric net SW radiations that are within the uncertainty range of EBAF, while only five CMIP6 models underestimate global atmospheric net SW radiations. On regional scale, except over the polar regions (Antarctica and the Arctic Ocean), CMIP6 MEM atmospheric net SW radiations are lower than EBAF. There is, however, a considerable spread of biases among regions, ranging from -6.5 Wm^{-2} over South America to 1.3 Wm^{-2} over Antarctica (Figure 2). Over South America, the low CMIP6 MEM atmospheric net SW radiation leads to an overestimated CMIP6 MEM DSR (Figure 5). The polar regions have the largest CMIP6 relative inter-model spread in atmospheric net SW radiation compared to all other land and ocean regions (Table S5 in Supporting Information S1). This suggests that the slightly higher atmospheric net SW radiations over the polar regions in CMIP6 MEM than in EBAF, by around 1 Wm^{-2}, are a result of compensating model biases.

CMIP6 MEM atmospheric net LW radiations also have lower values than EBAF over global land, global ocean and the globe. We again compared atmospheric net LW radiations of individual CMIP6 models with the uncertainty range inferred from Kato et al. (2018) on global scale (Table S2 in Supporting Information S1). The result is that 20 out of 53 CMIP6 models underestimate the atmospheric net LW radiations and are outside the uncertainty range, while the other 33 CMIP6 models are within the uncertainty range. Consistent low values of CMIP6 MEM atmospheric net LW radiations prevail over all ocean and land regions except for Mainland Australia as well as the Australasian and Indonesian Islands (Figure 2). Regional differences are, again, pronounced. For example, South America has a bias of 7.5 Wm^{-2}, while the bias for Africa is only 3.8 Wm^{-2}, and Australia has even a negative bias of -1.9 Wm^{-2}. To what degree these regionally different model biases affect regional energy budgets and modeling of associated atmospheric dynamics is a topic for future research.

3.6. Comparison Between CMIP6 and CMIP5 Models

So far we discussed the energy budget components from global to regional scales in CMIP6 historical simulations. In this section, we examine how the CMIP6 models compare in this respect with the preceding model generation CMIP5. In doing so, we compare the CMIP6 and CMIP5 energy budget components over the same time period from 2000 to 2005 (Section 2.4), as the CMIP5 historical simulation extend only till 2005. We used the CMIP6 historical simulations to ascertain that the MEM is robust against the choice of either averaging period (2000–2005 or 2000–2009). We thereby found that differences due to the two different periods are typically smaller than 0.1 Wm^{-2}, with a maximum difference of 0.7 Wm^{-2} for ULR over the Arctic Ocean and can therefore be neglected.

The results of the comparison between the CMIP6 and CMIP5 energy budget components are shown in Figure 8. From this figure, it is apparent that statistically significant changes of the MEMs of the two model generations are largely absent in some regions (global land, the Americas) and for some flux components (atmosphere net SW radiation), whereas substantial differences are present in other regions (global ocean, the Arctic Ocean, the South Pacific Ocean) and for other flux components (ISR, DLR, SH). We address these statistically significant differences (bold face numbers in Figure 8) in some more detail in the following, proceeding from the TOA to the surface.

The most obvious feature of the TOA components in Figure 8 is the difference between the ISR, in which the CMIP6 MEMs give slightly lower values in all the regions and a uniform lower value of 0.8 Wm^{-2} over global land, global ocean, and the globe as compared to the CMIP5 MEMs. This is caused by the updated estimate of the TSI as given by Kopp and Lean (2011). According to the measurements from SORCE, they determine the TSI of 1360.8 ± 0.5 Wm^{-2}, which is lower than the previous widely used value of 1365.4 ± 0.7 Wm^{-2} (Lee III et al., 1995).

There is no significant difference between CMIP6 and CMIP5 MEM in other TOA components over global land, global ocean and the globe. On regional scale, there are differences in MEM OSR over Australia and the Oceans in the Northern Hemisphere (the North Pacific Ocean, the North Atlantic Ocean and the Arctic Ocean), which all help to narrow the biases between CMIP6 MEM and EBAF (Figure 2). By contrast, the difference in MEM OLR over the Arctic Ocean enlarges the bias between CMIP6 MEM and EBAF (Figure 2).

Difference (CMIP6 – CMIP5)	Continents							Ocean basins											
	EA	Af	NA	SA	An	Au	IC	SP	NP	IO	SAt	NAt	AO	CS	MS	BS	Land	Ocean	Globe
Area ($10^6 km^2$)	53.2	29.9	24.0	17.7	12.7	7.56	1.48	99.9	81.8	75.4	46.5	43.4	10.2	4.35	2.60	0.472	146.6	364.5	511.2
TOA																			
ISR	**-0.7**	**-1.0**	**-0.7**	**-0.9**	**-0.3**	**-0.8**	**-0.9**	**-0.8**	**-0.9**	**-0.8**	**-0.7**	**-0.8**	**-0.5**	**-0.9**	**-0.8**	**-0.7**	**-0.8**	**-0.8**	**-0.8**
OSR	-1.0	0.5	-0.4	1.0	-0.1	**3.2**	-1.8	1.5	**2.1**	0.2	0.8	**2.1**	**-3.2**	2.2	2.6	2.3	-0.1	1.2	0.9
TOA net SW	-1.8	-0.5	-1.1	0.1	-0.4	**2.4**	-2.6	0.7	1.2	-0.5	0.1	1.3	**-3.6**	1.3	1.8	1.6	-0.9	0.4	0.1
OLR	0.9	0.0	0.8	1.8	-0.1	0.0	**3.2**	-0.6	-0.8	-0.3	0.6	0.4	**1.8**	0.3	-0.7	-0.3	0.7	-0.2	0.0
TOA radiation balance	-0.9	-0.5	-0.3	1.9	-0.5	**2.3**	0.6	0.2	0.4	**-0.8**	0.7	**1.8**	-1.8	1.6	1.1	1.3	-0.2	0.2	0.1
Atmosphere																			
Atmos. net SW	0.6	1.6	-0.2	1.4	0.2	0.8	1.1	0.6	0.8	1.0	0.8	0.2	0.3	1.1	-1.1	-0.9	0.7	0.7	0.7
Atmos. net LW	-0.9	0.9	-1.0	0.1	**-2.5**	0.1	0.8	**-2.5**	**-2.0**	**-2.9**	-1.6	-0.9	**-2.7**	-1.3	-0.8	0.3	-0.5	**-2.2**	**-1.7**
Surface																			
DSR	-2.8	-1.1	-1.0	-2.6	0.1	-2.2	-3.7	-0.7	-0.2	**-2.9**	-2.1	0.4	**-4.1**	-1.0	2.2	2.3	-1.9	-1.2	-1.4
USR	0.4	-0.9	0.1	1.3	-0.8	**3.7**	0.0	**0.9**	0.6	**1.3**	**1.3**	0.7	0.2	**1.3**	0.7	0.2	0.3	**0.9**	0.7
Surface net SW	**-2.4**	-2.0	-0.9	-1.3	-0.6	1.6	-3.7	0.1	0.4	-1.6	-0.7	1.1	**-3.9**	0.3	2.9	2.5	-1.6	-0.3	-0.6
DLR	**2.9**	2.0	2.1	1.9	**4.1**	1.7	1.7	**2.0**	1.6	**3.0**	**3.6**	**2.3**	**6.9**	**3.7**	1.9	1.0	2.5	**2.5**	**2.5**
ULR	-1.1	**-3.0**	-0.4	-0.2	-1.7	-1.9	0.7	0.0	-0.3	-0.4	-1.4	-0.9	-2.4	**-2.1**	-1.8	-1.6	-1.3	-0.5	-0.8
Surface net LW	1.8	-0.9	1.7	1.7	**2.5**	-0.1	**2.4**	**2.0**	1.3	**2.6**	**2.2**	1.4	**4.5**	1.6	0.1	-0.6	1.2	**2.0**	**1.7**
Surface net radiation	-0.6	-3.0	0.8	0.3	1.9	1.4	-1.3	**2.1**	1.6	1.0	1.5	**2.5**	0.6	1.8	**3.0**	2.0	-0.4	**1.7**	1.1
LH	1.0	**3.0**	1.1	-1.7	-0.3	0.1	3.7	0.1	0.5	0.5	-1.7	-1.1	0.4	**-4.1**	**-3.3**	-0.2	1.0	-0.2	0.2
SH	-0.4	-0.2	-1.7	1.3	**-2.0**	-1.5	**-3.5**	**-1.4**	**-1.6**	-1.2	**-1.4**	**-1.6**	-0.7	-0.4	-1.3	**-3.2**	-0.6	**-1.4**	**-1.2**
Surface imbalance	0.1	-0.2	0.3	-0.1	-0.4	0.0	-1.1	**0.8**	0.5	0.3	**-1.6**	-0.2	0.4	**-2.7**	**-1.6**	-1.4	0.0	0.1	0.1

Figure 8. Difference (CMIP6 – CMIP5) between CMIP6 MEM (2000–2005) and CMIP5 MEM (2000–2005) in Wm^{-2}. The overall structure is similar to Figure 2. The values of the components in the corresponding regions where CMIP6 MEM (2000–2005) and CMIP5 MEM (2000–2005) are significantly different at 95% confidence level are marked in bold and larger font.

For the atmospheric components, significant differences between the two model generations exist only for atmospheric net LW radiation (not for net SW radiation) and primarily over the oceans—global, Pacific, Indian and Arctic Ocean—as well as over Antarctica. These upward adjustments in absolute magnitude of the CMIP6 MEM in atmospheric net LW radiation help to narrow the bias between the CMIP6 MEM and EBAF (Figure 2), yet still not enough to completely get rid of the deficient CMIP6 MEM atmospheric net LW radiation. A substantial enhancement of the atmospheric SW absorption under cloud-free conditions in the CMIP6 MEM compared to CMIP5 has been noted in Wild (2020) (significant at the 95% confidence level).

As for the surface components over land regions, there is no significant difference in the global land means between the two model generations in all components. Nevertheless, on regional scales, the significant downward adjustment in the absolute magnitude of the CMIP6 MEM surface net SW radiation over Eurasia and the significant upward adjustment in the absolute magnitude of the CMIP6 MEM DLR over Eurasia as well as the CMIP6 MEM ULR over Africa reduce the bias between CMIP6 MEM and EBAF (Figure 2).

For the surface energy components over ocean regions, we find differences between CMIP6 and CMIP5 MEM mainly, although not exclusively, over the oceans in the Southern Hemisphere (Figure 8). Over the global ocean, differences in MEM USR and MEM DLR propagate to surface net LW radiation and surface net radiation. The downward adjustment in the absolute magnitude of global ocean means USR in the CMIP6 MEM lowers the biases between CMIP6 MEM and EBAF as well as NEWS (Figure 2). We recall, however, that these small biases in the CMIP6 MEM result at least in part from compensational effects among CMIP6 models (Section 3.2.2), where eight models overestimate and nine models underestimate global ocean mean USR (Table S3 in Supporting Information S1). The downward adjustments in the absolute magnitude of CMIP6 MEM USR on regional scale are found mostly over the oceans in the Southern Hemisphere (the South Pacific Ocean, the Indian Ocean and the South Atlantic Ocean). Since the two reference data sets, NEWS and EBAF, disagree with each other in global ocean mean DLR, it is hard to tell whether the upward adjustment in the absolute magnitude of CMIP6 MEM DLR helps to improve the performance of CMIP6 models over most ocean regions. However, Wild et al. (2015) noted a small underestimation of the CMIP5 DLR over oceans compared to the available direct observations from buoys and maritime BSRN sites, which supports a slight upward adjustment as seen in the CMIP6 MEM DLR. On regional scales, at least over the North Atlantic Ocean and the Caribbean Sea, the upward adjustments in the CMIP6 MEM DLR help to narrow the biases between the CMIP6 MEM and EBAF as well as NEWS. Meanwhile, the change in CMIP6 MEM DSR over the Arctic Ocean enlarges the biases between CMIP6 MEM and EBAF as well as NEWS (Figure 2).

The reduction in the absolute magnitude of the CMIP6 MEM LH compared to its CMIP5 counterpart over Africa reduces the bias between the CMIP6 MEM and NEWS (Figure 2), and makes Africa the only land region where the CMIP6 MEM LH is within the uncertainty range of NEWS (Figure 5). There are also significant differences in SH over large ocean regions but no conclusion can be easily made because of the discrepancies in the reference data sets (Section 3.3).

We conclude this section on the comparison of CMIP5 and CMIP6 by slightly broadening the view beyond the discussion of MEMs by taking individual models into account. On the one hand, we utilize the mean absolute bias between individual CMIP6/CMIP5 models' ensemble means and our reference data sets to represent the accuracy of CMIP models on regional scales. On the other hand, we determine for both CMIP5 and CMIP6 model generations the multi-model standard deviation of ensemble means as well as the inter-model spread of ensemble means for the different energy flux components to assess the degree of consistency between the different CMIP6/CMIP5 models. We further condense our results over 16 regions into two tables by taking the area-weighted means of the same metrics over seven land regions (Table 5) and nine ocean regions (Table 6). It is noteworthy to mention that even if we use arithmetic means or medians instead of area-weighted means over land or ocean regions, we still get similar results as in Tables 5 and 6.

From this analysis, we take that the CMIP6 models generally have lower mean absolute biases against our reference data sets, lower mean standard deviations and lower inter-model spreads in all Earth's energy budget components on regional scale than the CMIP5 models. One may interpret this as an improvement from CMIP5 to CMIP6 in the sense that the models tend to agree better with the reference estimates and are more consistent amongst themselves with respect to their representation of the energy budget components on regional scales.

However, a number of challenges remain. CMIP6 DSR on regional scale averaged over seven land regions, as one crucial component on Earth's surface, still suffers from a higher mean absolute bias, mean standard deviation and mean inter-model spread than any other component over land, and, more importantly, roughly half of all models fall out of the uncertainty range on regional scale over land (Figure 6). The mean standard deviation and mean inter-model spreads of LH on regional scale over land (mainly over the four largest land regions based on Table S1 in Supporting Information S1) are even increased in CMIP6 as compared to CMIP5, which points to deficiencies in simulating LH in the CMIP6 models as discussed in Section 3.4. It is also noteworthy to mention that, unlike in most other regions, over the polar regions (Antarctica and the Arctic Ocean) the CMIP6 inter-model spreads of most energy budget components are higher than CMIP5. Over Antarctica, the CMIP6 components with higher inter-model spreads than CMIP5 are ISR, OSR, TOA net SW radiation, TOA radiation balance, atmospheric net SW radiation, USR, surface net SW radiation, DLR, ULR, surface net radiation and surface imbalance. Over the Arctic Ocean, the CMIP6 components with higher inter-model spreads than CMIP5 are ISR, OSR, TOA net SW radiation, OLR, TOA radiation balance, atmospheric net SW radiation, atmospheric net LW radiation, DLR, ULR, surface net radiation, SH and surface imbalance. Over the polar regions, many energy budget components in CMIP6 models also have higher mean absolute biases against our reference data sets than in CMIP5 models. Over Antarctica, these are OSR, TOA net SW radiation, atmospheric net SW radiation, USR, ULR, LH, surface imbalance, so the higher biases exist mainly in the upward fluxes. Over the Arctic Ocean, these are OLR, atmospheric net LW radiation, DSR.

4. Summary and Conclusions

We examined the energy budget in CMIP6 historical experiments from 53 models, averaged over the period 2000–2009, from global to regional scales with regard to satellite-derived estimates, primarily but not exclusively the NEWS and CERES EBAF data products. Results are put into perspective by providing, on the one hand, some comparisons between CMIP6 and the two aforementioned reference data sets. On the other hand, we quantify changes from CMIP5 to CMIP6.

Table 5
Overall Performance of CMIP6 and CMIP5 Models With Respect to Their Representation of Regional Scale Energy Budget Components Over Seven Land Regions

	Mean absolute bias against NEWS		Mean absolute bias against EBAF		Mean standard deviation		Mean spread	
Averaged over 7 land regions	CMIP6	CMIP5	CMIP6	CMIP5	CMIP6	CMIP5	CMIP6	CMIP5
TOA								
ISR			0.3	1.0	0.7	0.9	4.5	4.4
OSR			4.3	5.8	5.1	6.9	24.5	29.5
TOA net SW			4.3	5.9	5.1	6.8	24.4	27.9
OLR			3.9	4.3	4.2	4.8	19.5	21.6
TOA radiation balance			3.3	4.4	3.8	4.7	19.0	19.0
Atmosphere								
Atmos. net SW			3.6	4.8	3.5	4.2	17.5	17.7
Atmos. net LW			7.3	8.2	5.2	5.5	23.6	28.2
Surface								
DSR	9.3	12.2	8.0	10.9	7.8	11.3	32.1	50.8
USR	7.4	7.9	4.8	5.2	5.3	5.7	25.3	26.3
Surface net SW	5.6	7.8	5.2	7.8	5.7	8.6	27.2	39.0
DLR	7.5	8.3	7.1	8.3	7.4	7.8	32.4	34.5
ULR	7.7	8.0	5.8	5.6	6.2	6.2	27.3	30.8
Surface net LW	5.0	6.4	7.3	8.5	5.3	7.0	24.8	31.0
Surface net radiation	4.7	5.5	5.6	6.4	5.1	6.2	24.5	29.4
LH	6.2	6.6			5.7	5.0	31.0	20.7
SH	7.5	8.5			5.1	6.3	22.4	31.2
Surface imbalance	1.1	1.0			1.0	0.9	6.1	4.0
mean of all components					4.8	5.8	22.7	26.2
mean of TOA components			3.2	4.3	3.8	4.8	18.4	20.5
mean of atmos. components			5.5	6.5	4.3	4.8	20.5	22.9
mean of surface radiative components	6.8	8.0	6.2	7.5	6.1	7.5	27.7	34.5
mean of surface components	6.2	7.2			5.5	6.5	25.3	29.8

Note. The metrics used for the comparison are mean absolute bias between individual CMIP6/CMIP5 models (2000–2005) and NEWS (2000–2009), mean absolute bias between individual CMIP6/CMIP5 models (2000–2005) and EBAF (2000 March–2006 February), standard deviation of CMIP6/CMIP5 models (2000–2005) and inter-model spread of CMIP6/CMIP5 models (2000–2005). All these metrics are first calculated over seven individual land regions and the final results shown above are the area-weighted means of the same metrics over seven land regions. Units in Wm^{-2}.

The global mean SW net energy input at TOA is underestimated and falls outside of the CERES EBAF uncertainty range by 11 CMIP6 models. The underestimation in CMIP6 models tends to be pronounced in the Northern Hemisphere, while there is an overestimation of TOA net SW radiation over South America, the South Atlantic Ocean, and Antarctica. OLR in CMIP6 models tends to be too low over most regions, with the notable exceptions of South America and Africa. The TOA radiation balance in many CMIP6 models is characterized by more energy entering or less energy leaving the climate system over Northern Hemisphere land, Southern Hemisphere oceans, and polar regions compared to CERES EBAF. Less energy enters or more energy leaves over Southern Hemisphere land and Northern Hemisphere oceans. This spatial pattern in TOA imbalance biases sets the stage not only for deficiencies in cross-equatorial energy transports in models, but also for other shortcomings in the modeled circulation.

For the atmospheric components, the majority of the CMIP6 models is within the uncertainty range of EBAF. There is, however, a tendency toward underestimation of net atmospheric absorption in the SW, except for polar regions, and even more so in the atmospheric net LW radiation, with the exception of Mainland Australia. The substantial regional differences (such as too low SW absorption over South America or too low net LW radiation

Table 6
Same as Table 5 but for Nine Ocean Regions

	Mean absolute bias against NEWS		Mean absolute bias against EBAF		Mean standard deviation		Mean spread	
Averaged over 9 ocean regions	CMIP6	CMIP5	CMIP6	CMIP5	CMIP6	CMIP5	CMIP6	CMIP5
TOA								
ISR			0.3	1.0	0.6	0.9	4.4	4.2
OSR			3.4	4.6	4.2	4.7	20.6	21.4
TOA net SW			3.5	4.2	4.4	4.6	21.2	18.9
OLR			3.1	3.5	3.5	3.6	17.2	15.3
TOA radiation balance			2.4	2.8	2.2	2.7	10.1	12.8
Atmosphere								
Atmos. net SW			2.3	3.0	2.6	2.9	12.4	13.2
Atmos. net LW			5.5	7.5	4.8	4.4	20.8	24.5
Surface								
DSR	5.2	6.2	4.5	5.3	5.8	6.2	27.6	24.2
USR	1.6	2.1	1.6	2.1	2.1	2.4	10.7	10.4
Surface net SW	4.9	5.7	4.0	4.8	4.9	5.7	22.6	23.4
DLR	6.4	5.4	5.1	5.9	5.4	5.0	22.3	24.1
ULR	4.2	4.2	3.3	3.4	3.5	3.8	14.4	18.3
Surface net LW	5.2	4.8	3.9	5.2	4.1	3.7	16.1	18.4
Surface net radiation	6.1	5.8	3.7	5.6	3.9	4.7	17.4	19.8
LH	6.6	7.2			4.3	5.5	18.4	22.8
SH	4.8	5.7			3.4	2.7	15.5	17.0
Surface imbalance	6.4	6.4			2.1	2.4	10.2	13.4
Mean of all components					3.6	3.9	16.6	17.8
Mean of TOA components			2.5	3.2	3.0	3.3	14.7	14.5
Mean of atmos. components			3.9	5.3	3.7	3.7	16.6	18.9
Mean of surface radiative components	4.8	4.9	3.7	4.6	4.2	4.5	18.7	19.8
Mean of surface components	5.1	5.4			3.9	4.2	17.5	19.2

Note. Units in Wm^{-2}.

over the South Atlantic Ocean as compared to their neighboring regions) potentially again impact atmospheric dynamics.

Turning to surface radiative fluxes, the differences between the NEWS and CERES EBAF products, our main references, show a strong heterogeneity with regard to region and flux component. Particularly large differences exist in the two reference data sets over the Americas, where CERES EBAF falls outside the uncertainty range of NEWS for all fluxes except DSR. Similar challenges exist for the polar regions. The radiative flux components USR, ULR, and DLR estimates disagree in many regions between the two reference data sets. In contrast, the two references are consistent for all flux components in Eurasia, Mainland Australia, the North Atlantic Ocean, and the Indian Ocean. With regard to the CMIP6-simulated regional surface energy budget components, we find five major deficiencies. First, around half of the CMIP6 models overestimate DSR over Eurasia and the two Americas, while the ocean regions show a mixed picture in the sense that CMIP6 models almost equally overestimate and underestimate DSR over the Pacific Ocean, resulting in little bias in MEM DSR, but 18 models overestimate DSR over the South Atlantic Ocean and 22 models underestimate DSR over the North Atlantic Ocean. These compensating effects over the ocean regions, which exist not only in individual regions but also across regions lead to accurate CMIP6 MEM DSR averaged over the entire global ocean. Second and third, CMIP6 models tend to overestimate USR and underestimate DLR over the largest two land regions (Eurasia and Africa).

Fourth, CMIP6 models tend to overestimate ULR over the largest two ocean regions in the Southern Hemisphere (the South Pacific Ocean and the Indian Ocean). Last but not least, in the North Atlantic Ocean and the Indian Ocean, at least 40% of all CMIP6 models lie outside our reference uncertainty range for any surface radiative flux components except surface net LW radiation and surface net radiation. Thorough quantifications of model deficiencies over the Americas, the Pacific Ocean, and the polar regions are hampered by the disagreement of the reference data in these regions. Better consistency of the references here is clearly desirable, and an independent validation with direct surface observations is a worthwhile future step.

Likewise, better accuracy of reference data set is desirable for latent and sensible heat fluxes. What can be concluded here is that LH is overestimated in almost all regions by the majority of CMIP6 models. Regional differences in the biases are, however, substantial (e.g., there is 11.7 Wm^{-2} bias in LH over the North Pacific Ocean, yet only 2.9 Wm^{-2} bias in LH over the South Pacific Ocean between CMIP6 MEM and our reference data set).

Comparing the model generations CMIP6 and CMIP5 with respect to their regional scale representation of the energy budget components, we find a general reduction of the biases between CMIP6 models and our reference data sets compared to CMIP5. The inter-model spread also decreased from CMIP5 to CMIP6 except over the polar regions, where the inter-model spread is larger in CMIP6 than CMIP5 for most energy budget components.

Future work should, ideally, make use of more consistent reference estimates notably over the Americas, the Pacific Ocean and the polar regions. More consistent reference data for LH and SH would likewise be desirable. Also, a thorough evaluation of the various data products with direct observations would be required. On the modeling side on regional scale, challenges lie with DSR over South America and the Arctic Ocean, USR over Eurasia and the North Atlantic Ocean, ULR over the Indian Ocean and the North Atlantic Ocean, and with the partitioning of LH and SH in general. The heterogeneity of these challenges in terms of region and flux component, as documented in this paper, suggests potentially equally heterogeneous underlying physical causes. To improve the situation, dedicated physical studies at regional scales seem indispensable and a logical next step.

Data Availability Statement

The data sets used in our study can be downloaded in the following links: NEWS Annual Climatology Version 1.0 Data Product: https://dx.doi.org/10.5067/7VZB10AK8S3D (L'Ecuyer et al., 2015; Rodell, Beaudoing, et al., 2015; Rodell, L'Ecuyer, et al., 2015). CERES EBAF TOA Edition 4.1 Data Product: https://doi.org/10.5067/TERRA-AQUA/CERES/EBAF-TOA_L3B004.1 (Loeb et al., 2018). CERES EBAF Surface Edition 4.1 Data Product: https://doi.org/10.5067/Terra-Aqua/CERES/EBAF_L3B.004.1 (Kato et al., 2018). CMIP6 historical simulations: https://esgf-node.llnl.gov/search/cmip6/ In this link, select Experiment ID as "historical", select Table ID as "Amon", then click "Search". CMIP5 historical simulations: https://esgf-node.llnl.gov/search/cmip5/ In this link, select Experiment ID as "historical", select CMIP Table as "Amon," then click "Search."

Acknowledgments
We would like to thank Tristan L'Ecuyer for providing the digitized maps defining the regions used in our study, which have been developed for the NASA/NEWS project. The NASA/Langley Research Center is highly acknowledged for their continuous efforts to provide outstanding satellite-derived energy flux products. Our research on the Earth's energy budget has been supported by a sequence of Swiss National Science Foundation Grants (Grant 200021_135395, 200020_159938, and 200020_188601) and by funding from the Federal Office of Meteorology and Climatology MeteoSwiss within the framework of GCOS Switzerland. We acknowledge the World Climate Research Program which, through its Working Group on Coupled Modeling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF. Open Access funding enabled and organized by Projekt DEAL.

References

Abbot, C. G., & Fowle, F. E. (1908). Determination of the intensity of the solar radiation outside the Earth's atmosphere, otherwise termed "the Solar Constant Radiation". *Annals of the Astrophysical Observatory of the Smithonian Institute*, *2*(Part 1), 11–124.

Anderson, D. E., & Cahalan, R. F. (2005). The solar radiation and climate experiment (SORCE) mission for the NASA Earth observing system (EOS). In G. Rottman, T. Woods, & V. George (Eds.), *The solar radiation and climate experiment (SORCE): Mission description and early results* (pp. 3–6). Springer. https://doi.org/10.1007/0-387-37625-9_1

Barkstrom, B. R. (1984). The Earth radiation budget experiment (ERBE). *Bulletin of the American Meteorological Society*, *65*(11), 1170–1185. https://doi.org/10.1175/1520-0477(1984)065<1170:TERBE>2.0.CO;2

Barkstrom, B. R., Harrison, E. F., Lee, R. B., III, & Team, E. S. (1990). Earth Radiation budget experiment. *EOS, Transactions American Geophysical Union*, *71*(9), 297–304. https://doi.org/10.1029/EO071i009p00297

Brooks, C. F. (1932). What the atmosphere does to solar radiation. *Bulletin of the American Meteorological Society*, *13*(12), 217–220. https://doi.org/10.1175/1520-0477-13.12.217

Brunke, M. A., Wang, Z., Zeng, X., Bosilovich, M., & Shie, C.-L. (2011). An assessment of the uncertainties in ocean surface turbulent fluxes in 11 reanalysis, satellite-derived, and combined global datasets. *Journal of Climate*, *24*(21), 5469–5493. https://doi.org/10.1175/2011JCLI4223.1

Brunke, M. A., Zeng, X., & Anderson, S. (2002). Uncertainties in sea surface turbulent flux algorithms and data sets. *Journal of Geophysical Research*, *107*(C10), 51–521. https://doi.org/10.1029/2001JC000992

Budyko, M. I., Yefimova, N. A., Aubenok, L. I., & Strokina, L. A. (1962). The heat balance of the surface of the Earth. *Soviet Geography*, *3*(5), 3–16. https://doi.org/10.1080/00385417.1962.10769936

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597. https://doi.org/10.1002/qj.828

Dines, W. H. (1917). The heat balance of the atmosphere. *Quarterly Journal of the Royal Meteorological Society*, *43*(182), 151–158. https://doi.org/10.1002/qj.49704318203

Driemel, A., Augustine, J., Behrens, K., Colle, S., Cox, C., Cuevas-Agulló, E., et al. (2018). Baseline surface radiation network (BSRN): Structure and data description (1992–2017). *Earth System Science Data*, *10*(3), 1491–1501. https://doi.org/10.5194/essd-10-1491-2018

Edwards, P. N. (2000). A brief history of atmospheric general circulation modeling. *International Geophysics Series*, *70*, 67–90.

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

Fasullo, J. T., & Trenberth, K. E. (2008a). The annual cycle of the energy budget. Part I: Global mean and land–ocean exchanges. *Journal of Climate*, *21*(10), 2297–2312. https://doi.org/10.1175/2007JCLI1935.1

Fasullo, J. T., & Trenberth, K. E. (2008b). The annual cycle of the energy budget. Part II: Meridional structures and poleward transports. *Journal of Climate*, *21*(10), 2313–2325. https://doi.org/10.1175/2007JCLI1936.1

Frierson, D. M. W., Hwang, Y.-T., Fučkar, N. S., Seager, R., Kang, S. M., Donohoe, A., et al. (2013). Contribution of ocean overturning circulation to tropical rainfall peak in the Northern Hemisphere. *Nature Geoscience*, *6*(11), 940–944. https://doi.org/10.1038/ngeo1987

Gibson, J. K., Kallberg, P., Uppala, S., Hernandez, A., Nomura, A., & Serrano, E. (1997). *ERA description*. ECMWF Re-Analysis Project Report Series 1 (p. 74). ECMWF.

Gilgen, H., Wild, M., & Ohmura, A. (1998). Means and trends of shortwave irradiance at the surface estimated from global energy balance archive data. *Journal of Climate*, *11*(8), 2042–2061. https://doi.org/10.1175/1520-0442(1998)011<2042:MATOSI>2.0.CO;2

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

House, F. B., Gruber, A., Hunt, G. E., & Mecherikunnel, A. T. (1986). History of satellite missions and measurements of the Earth Radiation Budget (1957–1984). *Reviews of Geophysics*, *24*(2), 357–377. https://doi.org/10.1029/RG024i002p00357

Hunt, G. E., Kandel, R., & Mecherikunnel, A. T. (1986). A history of presatellite investigations of the Earth's Radiation Budget. *Reviews of Geophysics*, *24*(2), 351–356. https://doi.org/10.1029/RG024i002p00351

Jayne, S. R., Roemmich, D., Zilberman, N., Riser, S. C., Johnson, K. S., Johnson, G. C., & Piotrowicz, S. R. (2017). The Argo Program: Present and future. *Oceanography*, *30*(2), 18–28. https://doi.org/10.5670/oceanog.2017.213

Jian, B., Li, J., Wang, G., Zhao, Y., Li, Y., Wang, J., et al. (2021). Evaluation of the CMIP6 marine subtropical stratocumulus cloud albedo and its controlling factors. *Atmospheric Chemistry and Physics*, *21*(12), 9809–9828. https://doi.org/10.5194/acp-21-9809-2021

Jian, B., Li, J., Zhao, Y., He, Y., Wang, J., & Huang, J. (2020). Evaluation of the CMIP6 planetary albedo climatology using satellite observations. *Climate Dynamics*, *54*(11), 5145–5161. https://doi.org/10.1007/s00382-020-05277-4

Johnson, G. C., Lyman, J. M., & Loeb, N. G. (2016). Improving estimates of Earth's energy imbalance. *Nature Climate Change*, *6*(7), 639–640. https://doi.org/10.1038/nclimate3043

Jones, P. W. (1999). First- and second-order conservative remapping schemes for grids in spherical coordinates. *Monthly Weather Review*, *127*(9), 2204–2210. https://doi.org/10.1175/1520-0493(1999)127<2204:FASOCR>2.0.CO;2

Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., et al. (2019). The FLUXCOM ensemble of global land-atmosphere energy fluxes. *Scientific Data*, *6*(1), 74. https://doi.org/10.1038/s41597-019-0076-8

Kalnay, E. (2003). *Atmospheric modeling, data Assimilation and Predictability*. Cambridge University Press.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., et al. (1996). The NCEP/NCAR 40-year reanalysis Project. *Bulletin of the American Meteorological Society*, *77*(3), 437–472. https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2

Kato, S., Loeb, N. G., Fasullo, J. T., Trenberth, K. E., Lauritzen, P. H., Rose, F. G., et al. (2021). Regional energy and water budget of a precipitating atmosphere over ocean. *Journal of Climate*, *34*(11), 4189–4205. https://doi.org/10.1175/JCLI-D-20-0175.1

Kato, S., Loeb, N. G., Rose, F. G., Doelling, D. R., Rutan, D. A., Caldwell, T. E., et al. (2013). Surface irradiances consistent with CERES-derived top-of-atmosphere shortwave and longwave irradiances. *Journal of Climate*, *26*(9), 2719–2740. https://doi.org/10.1175/JCLI-D-12-00436.1

Kato, S., Loeb, N. G., Rutan, D. A., Rose, F. G., Sun-Mack, S., Miller, W. F., & Chen, Y. (2012). Uncertainty estimate of surface irradiances computed with MODIS-CALIPSO-and CloudSat-derived cloud and aerosol properties. *Surveys in Geophysics*, *33*(3–4), 395–412. https://doi.org/10.1007/s10712-012-9179-x

Kato, S., Rose, F. G., Chang, F.-L., Painemal, D., & Smith, W. L. (2021b). Evaluation of regional surface energy budget over ocean derived from satellites. *Frontiers in Marine Science*, *8*, 1264. https://doi.org/10.3389/fmars.2021.688299

Kato, S., Rose, F. G., Rutan, D. A., Thorsen, T. J., Loeb, N. G., Doelling, D. R., et al. (2018). Surface irradiances of edition 4.0 clouds and the Earth's radiant energy system (CERES) energy balanced and filled (EBAF) data product. *Journal of Climate*, *31*(11), 4501–4527. https://doi.org/10.1175/JCLI-D-17-0523.1

Kiehl, J. T., & Trenberth, K. E. (1997). Earth's annual global mean energy budget. *Bulletin of the American Meteorological Society*, *78*(2), 197–208. https://doi.org/10.1175/1520-0477(1997)078<0197:EAGMEB>2.0.CO;2

Kopp, G., & Lean, J. L. (2011). A new, lower value of total solar irradiance: Evidence and climate significance. *Geophysical Research Letters*, *38*(1). https://doi.org/10.1029/2010GL045777

L'Ecuyer, T. S., Beaudoing, H. K., Rodell, M., Olson, W., Lin, B., Kato, S., et al. (2015). The observed state of the energy budget in the early twenty-first century. *Journal of Climate*, *28*(21), 8319–8346. https://doi.org/10.1175/JCLI-D-14-00556.1

Lee, R. B., III, Gibson, M. A., Wilson, R. S., & Thomas, S. (1995). Long-term total solar irradiance variability during sunspot cycle 22. *Journal of Geophysical Research*, *100*(A2), 1667–1675. https://doi.org/10.1029/94JA02897

Lembo, V., Folini, D., Wild, M., & Lionello, P. (2019). Inter-hemispheric differences in energy budgets and cross-equatorial transport anomalies during the 20th century. *Climate Dynamics*, *53*(1), 115–135. https://doi.org/10.1007/s00382-018-4572-x

Lettau, H. (1954). A study of the mass, momentum and energy budget of the atmosphere. *Archiv Für Meteorologie, Geophysik Und Bioklimatologie, Serie A*, *7*(1), 133–157. https://doi.org/10.1007/BF02277912

Li, J., Miao, C., Wei, W., Zhang, G., Hua, L., Chen, Y., & Wang, X. (2021). Evaluation of CMIP6 global climate models for simulating land surface energy and water fluxes during 1979–2014. *Journal of Advances in Modeling Earth Systems*, *13*(6). https://doi.org/10.1029/2021MS002515

Liu, C., Allan, R. P., Mayer, M., Hyder, P., Desbruyères, D., Cheng, L., et al. (2020). Variability in the global energy budget and transports 1985–2017. *Climate Dynamics*, *55*(11), 3381–3396. https://doi.org/10.1007/s00382-020-05451-8

Loeb, N. G., Doelling, D. R., Wang, H., Su, W., Nguyen, C., Corbett, J. G., et al. (2018). Clouds and the Earth's radiant energy system (CERES) energy balanced and filled (EBAF) top-of-atmosphere (TOA) edition-4.0 data product. *Journal of Climate*, *31*(2), 895–918. https://doi.org/10.1175/JCLI-D-17-0208.1

Loeb, N. G., Kato, S., & Wielicki, B. A. (2002). Defining top-of-the-atmosphere flux reference level for Earth radiation budget studies. *Journal of Climate*, *15*(22), 3301–3309. https://doi.org/10.1175/1520-0442(2002)015<3301:DTOTAF>2.0.CO;2

Loeb, N. G., Wang, H., Cheng, A., Kato, S., Fasullo, J. T., Xu, K.-M., & Allan, R. P. (2016). Observational constraints on atmospheric and oceanic cross-equatorial heat transports: Revisiting the precipitation asymmetry problem in climate models. *Climate Dynamics*, *46*(9), 3239–3257. https://doi.org/10.1007/s00382-015-2766-z

Loeb, N. G., Wielicki, B. A., Doelling, D. R., Smith, G. L., Keyes, D. F., Kato, S., et al. (2009). Toward optimal closure of the Earth's top-of-atmosphere radiation budget. *Journal of Climate*, *22*(3), 748–766. https://doi.org/10.1175/2008JCLI2637.1

Lyman, J. M., Good, S. A., Gouretski, V. V., Ishii, M., Johnson, G. C., Palmer, M. D., et al. (2010). Robust warming of the global upper ocean. *Nature*, *465*(7296), 334–337. https://doi.org/10.1038/nature09043

Marshall, J., Donohoe, A., Ferreira, D., & McGee, D. (2014). The ocean's role in setting the mean position of the Inter-Tropical Convergence Zone. *Climate Dynamics*, *42*(7), 1967–1979. https://doi.org/10.1007/s00382-013-1767-z

Miao, H., Wang, X., Liu, Y., & Wu, G. (2021). A regime-based investigation into the errors of CMIP6 simulated cloud radiative effects using satellite observations. *Geophysical Research Letters*, *48*(18), e2021GL095399. https://doi.org/10.1029/2021GL095399

Mueller, B., & Seneviratne, S. I. (2014). Systematic land climate and evapotranspiration biases in CMIP5 simulations. *Geophysical Research Letters*, *41*(1), 128–134. https://doi.org/10.1002/2013GL058055

NASA/LARC/SD/ASDC. (2019a). *CERES energy balanced and filled (EBAF) TOA and surface monthly means data in netCDF edition 4.1*. NASA Langley Atmospheric Science Data Center DAAC. https://doi.org/10.5067/TERRA-AQUA/CERES/EBAF_L3B.004.1

NASA/LARC/SD/ASDC. (2019b). *CERES energy balanced and filled (EBAF) TOA monthly means data in netCDF edition 4.1*. NASA Langley Atmospheric Science Data Center DAAC. https://doi.org/10.5067/TERRA-AQUA/CERES/EBAF-TOA_L3B004.1

Ohmura, A., Dutton, E. G., Forgan, B., Fröhlich, C., Gilgen, H., Hegner, H., et al. (1998). Baseline surface radiation network (BSRN/WCRP): New precision radiometry for climate research. *Bulletin of the American Meteorological Society*, *79*(10), 2115–2136. https://doi.org/10.1175/1520-0477(1998)079<2115:BSRNBW>2.0.CO;2

Ohmura, A., Gilgen, H., & Wild, M. (1989). Global energy balance archive GEBA. Retrieved from https://www.osti.gov/etdeweb/biblio/6903161

Onogi, K., Tsutsui, J., Koide, H., Sakamoto, M., Kobayashi, S., Hatsushika, H., et al. (2007). The JRA-25 reanalysis. *Journal of the Meteorological Society of Japan. Series II*, *85*(3), 369–432. https://doi.org/10.2151/jmsj.85.369

Patz, J. A., Campbell-Lendrum, D., Holloway, T., & Foley, J. A. (2005). Impact of regional climate change on human health. *Nature*, *438*(7066), 7066–7317. https://doi.org/10.1038/nature04188

Rannik, Ü., Peltola, O., & Mammarella, I. (2016). Random uncertainties of flux measurements by the eddy covariance technique. *Atmospheric Measurement Techniques*, *9*(10), 5163–5181. https://doi.org/10.5194/amt-9-5163-2016

Raschke, E., & Bandeen, W. R. (1970). The radiation balance of the planet Earth from radiation measurements of the satellite Nimbus II. *Journal of Applied Meteorology and Climatology*, *9*(2), 215–238. https://doi.org/10.1175/1520-0450(1970)009<0215:TRBOTP>2.0.CO;2

Raschke, E., Haar, T. H. V., Bandeen, W. R., & Pasternak, M. (1973). The annual radiation balance of the Earth-atmosphere system during 1969–70 from Nimbus 3 measurements. *Journal of the Atmospheric Sciences*, *30*(3), 341–364. https://doi.org/10.1175/1520-0469(1973)030<0341:TARBOT>2.0.CO;2

Rodell, M., Beaudoing, H. K., L'Ecuyer, T. S., Olson, W. S., Famiglietti, J. S., Houser, P. R., et al. (2015). The observed state of the water cycle in the early twenty-first century. *Journal of Climate*, *28*(21), 8289–8318. https://doi.org/10.1175/JCLI-D-14-00555.1

Rodell, M., L'Ecuyer, T. S., Beaudoing, H. K., & NASA/GSFC/HSL (2015). NASA Energy and Water cycle Study (NEWS) Annual Climatology of the 1st decade of the 21st Century, Version 1.0 [Dataset]. NASA Goddard Earth Sciences Data and Information Services Center. https://doi.org/10.5067/7VZB10AK8S3D

Rodgers, C. D. (2000). *Inverse methods for atmospheric sounding: Theory and practice* (Vol. 2). World Scientific Press. https://doi.org/10.1142/3171

Roemmich, D., Johnson, G. C., Riser, S., Davis, R., Gilson, J., Owens, W. B., et al. (2009). The Argo Program: Observing the global ocean with profiling floats. *Oceanography*, *22*(2), 34–43. https://doi.org/10.5670/oceanog.2009.36

Rummukainen, M. (2010). State-of-the-art with regional climate models. *WIREs Climate Change*, *1*(1), 82–96. https://doi.org/10.1002/wcc.8

Rutan, D. A., Kato, S., Doelling, D. R., Rose, F. G., Nguyen, L. T., Caldwell, T. E., & Loeb, N. G. (2015). CERES synoptic product: Methodology and validation of surface radiant flux. *Journal of Atmospheric and Oceanic Technology*, *32*(6), 1121–1143. https://doi.org/10.1175/JTECH-D-14-00165.1

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., et al. (2010). The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, *91*(8), 1015–1058. https://doi.org/10.1175/2010BAMS3001.1

Schuddeboom, A. J., & McDonald, A. J. (2021). The Southern Ocean radiative bias, cloud compensating errors, and equilibrium climate sensitivity in CMIP6 models. *Journal of Geophysical Research: Atmospheres*, *126*(22), e2021JD035310. https://doi.org/10.1029/2021JD035310

Stephens, G. L., Hakuba, M. Z., Hawcroft, M., Haywood, J. M., Behrangi, A., Kay, J. E., & Webster, P. J. (2016). The curious nature of the hemispheric symmetry of the Earth's water and energy balances. *Current Climate Change Reports*, *2*(4), 135–147. https://doi.org/10.1007/s40641-016-0043-9

Stephens, G. L., Li, J., Wild, M., Clayson, C. A., Loeb, N., Kato, S., et al. (2012). An update on Earth's energy balance in light of the latest global observations. *Nature Geoscience*, *5*(10), 691–696. https://doi.org/10.1038/ngeo1580

Su, W., Corbett, J., Eitzen, Z., & Liang, L. (2015). Next-generation angular distribution models for top-of-atmosphere radiative flux calculation from CERES instruments: Validation. *Atmospheric Measurement Techniques*, *8*(8), 3297–3313. https://doi.org/10.5194/amt-8-3297-2015

Suomi, V. E. (1958). The radiation balance of the Earth from a satellite. *Annals of the IGY*, *1*, 331–340.

Taylor, P. C. (2012). Tropical outgoing longwave radiation and longwave cloud forcing diurnal cycles from CERES. *Journal of the Atmospheric Sciences*, *69*(12), 3652–3669. https://doi.org/10.1175/JAS-D-12-088.1

Thomas, C. M., Dong, B., & Haines, K. (2020). Inverse modeling of global and regional energy and water cycle fluxes using Earth observation data. *Journal of Climate*, *33*(5), 1707–1723. https://doi.org/10.1175/JCLI-D-19-0343.1

Trenberth, K. E. (2015). Has there been a hiatus? *Science*, *349*(6249), 691–692. https://doi.org/10.1126/science.aac9225

Trenberth, K. E., & Fasullo, J. T. (2013a). An apparent hiatus in global warming? *Earth's Future*, *1*(1), 19–32. https://doi.org/10.1002/2013EF000165

Trenberth, K. E., & Fasullo, J. T. (2013b). North American water and energy cycles. *Geophysical Research Letters*, *40*(2), 365–369. https://doi.org/10.1002/grl.50107

Trenberth, K. E., & Fasullo, J. T. (2013c). Regional energy and water cycles: Transports from ocean to land. *Journal of Climate*, *26*(20), 7837–7851. https://doi.org/10.1175/JCLI-D-13-00008.1

Trenberth, K. E., & Fasullo, J. T. (2017). Atlantic meridional heat transports computed from balancing Earth's energy locally. *Geophysical Research Letters*, *44*(4), 1919–1927. https://doi.org/10.1002/2016GL072475

Trenberth, K. E., Fasullo, J. T., & Kiehl, J. (2009). Earth's global energy budget. *Bulletin of the American Meteorological Society*, *90*(3), 311–324. https://doi.org/10.1175/2008BAMS2634.1

Trenberth, K. E., & Zhang, Y. (2019). Observed interhemispheric meridional heat transports and the role of the Indonesian throughflow in the Pacific Ocean. *Journal of Climate*, *32*(24), 8523–8536. https://doi.org/10.1175/JCLI-D-19-0465.1

Tselioudis, G., Rossow, W. B., Jakob, C., Remillard, J., Tropf, D., & Zhang, Y. (2021). Evaluation of clouds, radiation, and precipitation in CMIP6 models using global weather states derived from ISCCP-H cloud property data. *Journal of Climate*, *34*(17), 7311–7324. https://doi.org/10.1175/JCLI-D-21-0076.1

van den Broeke, M. R., Smeets, C. J. P. P., & van de Wal, R. S. W. (2011). The seasonal cycle and interannual variability of surface energy balance and melt in the ablation zone of the west Greenland ice sheet. *The Cryosphere*, *5*(2), 377–390. https://doi.org/10.5194/tc-5-377-2011

Wang, Z., Zhan, C., Ning, L., & Guo, H. (2021). Evaluation of global terrestrial evapotranspiration in CMIP6 models. *Theoretical and Applied Climatology*, *143*(1), 521–531. https://doi.org/10.1007/s00704-020-03437-4

Wielicki, B. A., Barkstrom, B. R., Harrison, E. F., Lee, R. B., Smith, G. L., & Cooper, J. E. (1996). Clouds and the Earth's radiant energy system (CERES): An Earth observing system experiment. *Bulletin of the American Meteorological Society*, *77*(5), 853–868. https://doi.org/10.1175/1520-0477(1996)077<0853:CATERE>2.0.CO;2

Wild, M. (2020). The global energy balance as represented in CMIP6 climate models. *Climate Dynamics*, *55*(3–4), 553–577. https://doi.org/10.1007/s00382-020-05282-7

Wild, M., Folini, D., Hakuba, M. Z., Schär, C., Seneviratne, S. I., Kato, S., et al. (2015). The energy balance over land and oceans: An assessment based on direct observations and CMIP5 climate models. *Climate Dynamics*, *44*(11–12), 3393–3429. https://doi.org/10.1007/s00382-014-2430-z

Wild, M., Folini, D., Schär, C., Loeb, N., Dutton, E. G., & König-Langlo, G. (2013). The global energy balance from a surface perspective. *Climate Dynamics*, *40*(11–12), 3107–3134. https://doi.org/10.1007/s00382-012-1569-8

Wild, M., Ohmura, A., Gilgen, H., Roeckner, E., Giorgetta, M., & Morcrette, J.-J. (1998). The disposition of radiative energy in the global climate system: GCM-calculated versus observational estimates. *Climate Dynamics*, *14*(12), 853–869. https://doi.org/10.1007/s003820050260

Wild, M., Ohmura, A., Schär, C., Müller, G., Folini, D., Schwarz, M., et al. (2017). The global energy balance archive (GEBA) version 2017: A database for worldwide measured surface energy fluxes. *Earth System Science Data*, *9*(2), 601–613. https://doi.org/10.5194/essd-9-601-2017

Willis, J. K., Lyman, J. M., Johnson, G. C., & Gilson, J. (2009). In situ data biases and recent ocean heat content variability. *Journal of Atmospheric and Oceanic Technology*, *26*(4), 846–852. https://doi.org/10.1175/2008JTECHO608.1

Xie, S.-P., Deser, C., Vecchi, G. A., Collins, M., Delworth, T. L., Hall, A., et al. (2015). Towards predictive understanding of regional climate change. *Nature Climate Change*, *5*(10), 921–930. https://doi.org/10.1038/nclimate2689

Zhang, Y.-C., Rossow, W. B., & Lacis, A. A. (1995). Calculation of surface and top of atmosphere radiative fluxes from physical quantities based on ISCCP data sets: 1. Method and sensitivity to input data uncertainties. *Journal of Geophysical Research: Atmospheres*, *100*(D1), 1149–1165. https://doi.org/10.1029/94JD02747

Climate Dynamics (2020) 55:553–577
https://doi.org/10.1007/s00382-020-05282-7

The global energy balance as represented in CMIP6 climate models

Martin Wild[1]

Received: 25 December 2019 / Accepted: 2 May 2020 / Published online: 25 May 2020

Abstract

A plausible simulation of the global energy balance is a first-order requirement for a credible climate model. Here I investigate the representation of the global energy balance in 40 state-of-the-art global climate models participating in the Coupled Model Intercomparison Project phase 6 (CMIP6). In the CMIP6 multi-model mean, the magnitudes of the energy balance components are often in better agreement with recent reference estimates compared to earlier model generations on a global mean basis. However, the inter-model spread in the representation of many of the components remains substantial, often on the order of 10–20 Wm^{-2} globally, except for aspects of the shortwave clear-sky budgets, which are now more consistently simulated by the CMIP6 models. The substantial inter-model spread in the simulated global mean latent heat fluxes in the CMIP6 models, exceeding 20% (18 Wm^{-2}), further implies also large discrepancies in their representation of the global water balance. From a historic perspective of model development over the past decades, the largest adjustments in the magnitudes of the simulated present-day global mean energy balance components occurred in the shortwave atmospheric clear-sky absorption and the surface downward longwave radiation. Both components were gradually adjusted upwards over several model generations, on the order of 10 Wm^{-2}, to reach 73 and 344 Wm^{-2}, respectively in the CMIP6 multi-model means. Thereby, CMIP6 has become the first model generation that largely remediates long-standing model deficiencies related to an overestimation in surface downward shortwave and compensational underestimation in downward longwave radiation in its multi-model mean.

1 Introduction

The global energy balance fundamentally constrains the energy content of Earth's climate system as well as its internal distribution. For more than a century, scientists have attempted to quantify the magnitudes of the components of the global energy balance (i.e., the energy balance averaged over the Earth's sphere and over the year). Early attempts had to rely on a sparse number of observations taken at the surface and from balloon measurements combined with numerous assumptions, and the uncertainties in the global estimates were accordingly large (e.g., Abbot and Fowle 1908; Dines 1917). It was only with the advent of space-based measurements that the shortwave (solar) and longwave (thermal) energy exchanges between Earth and space could finally be quantified adequately, particularly through the Earth Radiation Budget Experiment (ERBE, Barkstrom et al. 1990) in the late 1980s and the more recent Clouds and Earth's Radiant Energy System (CERES, Wielicki et al. 1996) mission since the beginning of the 2000s. These data have extensively been used for the assessment of the Top of Atmosphere (TOA) radiation budgets and cloud radiative effects in global climate models (GCMs) (e.g., Potter et al. 1992; Cess and Potter 1987; Potter and Cess 2004; Wild and Roeckner 2006; Trenberth and Fasullo 2010; Wang and Su 2013; Li et al. 2013; Dolinar et al. 2014). However, the distribution of the radiative energy within the climate system and at the Earth's surface remained less well known also in the age of space-born measurements, since satellite measurements could provide only limited constraints on these aspects of the global energy balance. Thus, published estimates on the magnitudes of the global mean surface energy budget components still largely varied also in the satellite age, typically on the order of 10–20 Wm^{-2} or more (e.g., Ohmura and Gilgen 1993; Kiehl and Trenberth 1997; Wild et al. 1998, 2013; Hatzianastassiou et al. 2005; Trenberth et al. 2009; Stephens et al. 2012). Accordingly, throughout the history of model development, GCMs showed considerable discrepancies in their perception of the global energy

✉ Martin Wild
martin.wild@env.ethz.ch

[1] ETH Zurich, Institute for Atmospheric and Climate Science, 8001 Zurich, Switzerland

balance, particularly at the Earth's surface. The inter-model spread in the magnitudes of the individual components of the surface energy balance was known to be considerable since the earliest attempts of systematic model intercomparisons (Gutowski et al. 1991; Randall et al. 1992; Wild et al. 1995; Garratt and Prata 1996; Gleckler and Weare 1997; Li et al. 1997), whereas the agreement in their corresponding TOA components has been better. The latter was a consequence of the general practice to tune the GCMs to match their TOA flux magnitudes to the well-accepted space-born reference values, which became available since the late 1980s from ERBE and since the 2000s with even higher accuracy from CERES. No similar consensus reference values that could have served as tuning targets were available for the surface components, since these estimates historically showed large discrepancies as outlined above. However, with progress in the satellite-derived estimates of surface fluxes, as well as the availability of high accuracy radiation measurements from worldwide surface networks such as the Baseline Surface Radiation network (BSRN, Ohmura et al. 1998; Driemel et al. 2018), recent independently derived estimates of the global mean surface radiative components converged to within 4 Wm^{-2} (Wild 2017).

Comparisons with direct observations at the surface revealed a tendency of the GCMs to overestimate the downward shortwave radiation at the surface, and underestimate the downward longwave radiation, a long-standing problem that has persisted over several decades and generations of GCM development (Wild et al. 1995, 2013; Li et al. 1997; Cusack et al. 1998; Bodas-Salcedo et al. 2008; Wild 2008; Tang et al. 2019).

In the present study I will discuss the representation of the global energy balance in the latest generation of climate models participating in the sixth phase of the Coupled Model Intercomparison Project (CMIP6, Eyring et al. 2016), which will provide the basis for the upcoming Intergovernmental Panel on Climate Change (IPCC) 6th Assessment Report (AR6). The spatiotemporal focus will be on the global climatological annual mean, which will give a first order impression on the current model generations' abilities to capture the overall energy distribution in the climate system. Their simulated global energy budgets will be intercompared and opposed to recently emerging reference estimates in the following. An adequate representation of the global mean energy budget provides a necessary, though not sufficient condition for a credible climate model.

2 Data

At the time of the revision of this manuscript (March 2000), data from simulations performed by 40 GCMs appropriate for the present analysis have become available from CMIP6. Details on the modeling groups participating in CMIP6 can be found on the CMIP6 webpages of the Program for Climate Model Diagnosis and Intercomparison (PCMDI) (https://pcmdi.llnl.gov/CMIP6/).

The model-output variables under consideration for this study are the shortwave and longwave radiative fluxes at the surface and the TOA under both all-sky and clear-sky conditions, as well as the non-radiative fluxes of surface sensible and latent heat. They stem from the "historical all forcings" experiments of CMIP6, which aim at simulating the climate evolution since preindustrial times as realistic as possibly, considering all major natural and anthropogenic forcings, namely changes in solar output, atmospheric greenhouse gases, aerosol loadings (tropospheric and stratospheric volcanic), and land use (Eyring et al. 2016). These simulations cover the period 1850–2014. The global energy budgets of the CMIP6 models discussed in this study have been determined as averages over the final 15 years of these simulations (2000-2014) and shall represent present-day conditions at the beginning of the new millennium. To allow for a comparison with the previous model generation CMIP5 evaluated in Wild et al. (2013, 2015, 2019), I also determined the CMIP6 budgets for the averaging period 2000–2004 used in these former studies. The end year of 2004 was chosen in these studies since the corresponding historical simulation of the CMIP5 models only reached up to the year 2005 at the most. For the global mean budgets, the differences induced by the different averaging periods (2000–2014 versus 2000–2004) were, however, insignificant (<0.3 Wm^{-2}) for most components, with the exception of the longwave upward and downward radiation at the surface, which were enhanced by 0.6 and 0.8 Wm^{-2} in the 2000–2014 averaging period, due to the slightly stronger greenhouse forcing and associated warming. I further also investigated the interannual variability in the global annual mean energy budget components of the CMIP6 models, which turned out to be very small, with standard-deviations typically on the order of 0.2–0.3 Wm^{-2} for the global annual mean all-sky budget components, and even somewhat smaller for the respective clear-sky budgets. This further indicates that the exact length of the averaging period is not critical for the present analysis.

From many of the CMIP6 models, multiple realizations of the historic all forcings experiments with slightly differing initial conditions are available (ensemble simulations). The choice of the specific ensemble member is not critical, since their global multi-annual mean energy budgets do not differ significantly. Therefore, only one ensemble member from each model is included in the present analysis. Not all energy budget components were available from all models, therefore the number of models included in the analyses slightly varies depending on the energy balance component under investigation, as indicated in Table 1. The conclusions drawn in this study, however, were found to be very robust

Table 1 Global annual mean estimates of the magnitudes of various energy balance components under clear-sky and all-sky conditions at the TOA, within the atmosphere and at the surface, representative for present-day climate

Energy balance component	Reference Estimates Wm^{-2}	# CMIP6 models	CMIP6 mean Wm^{-2}	CMIP6 spread Wm^{-2}	CMIP6 stdev. Wm^{-2}	CMIP5 mean Wm^{-2}	CMIP5 spread Wm^{-2}	CMIP5 stdev. Wm^{-2}
TOA								
SW down TOA	340[a], 340[b], 340[c]	37	**340.2**	5.3	0.9	**341.3**	3.4	0.8
SW up all-sky TOA	– 99[a], – 100[b], – 102[c]	38	**– 100.6**	13.1	2.7	**– 102.0**	12.6	3.1
SW absorbed all-sky TOA	241[a], 240[b], 238[c]	37	239.5	14.5	2.9	239.2	11.2	3.0
SW up clear-sky TOA	– 53[a], – 53[b]	37	– 53.0	7.7	1.9	– 52.6	11.2	2.3
SW absorbed clear-sky TOA	287[a], 287[b]	37	**287.3**	7.1	1.8	**288.6**	10.6	2.1
SW CRE TOA	– 46[a], – 47[b]	37	– 47.8	19.2	3.6	– 49.3	14.0	3.5
LW up (OLR) all-sky TOA	– 240[a], – 239[b], – 238[c]	40	– 238.3	15.6	2.8	– 238.0	11.7	2.9
LW up (OLR) clear-sky TOA	– 268[a], – 267[b]	38	– 262.4	12.5	2.6	– 263.3	12.9	3.3
LW CRE TOA	28[a], 28[b]	38	24.1	10.4	2.3	24.9	12.6	3.5
Net CRE TOA	– 18[a], – 19[b]	37	– 23.6	13.5	3.3	– 24.1	15.5	3.9
Imbalance TOA	0.7[a]	37	1.1	4.5	0.8	1.2	n.a.	n.a.
Atmosphere								
SW absorbed all-sky atmos.	80[b]. 74[c], 77[d]	37	**76.0**	8.9	2.0	**74.4**	9.9	2.8
SW absorbed clear-sky atmos.	73[b], 73[d]	36	**72.8**	8.6	1.8	**70.1**	11.3	2.9
SW CRE atmos.	7[b], 4[d]	36	**3.2**	4.0	1.1	**4.3**	8.8	1.6
LW net all-sky atmos.	– 183[b], – 180[c], – 187[d]	37	**– 182.1**	17.2	4.2	**– 179.8**	22.5	3.8
LW net clear-sky atmos.	– 183[b], – 184[d]	33	**– 180.9**	15.1	3.0	**– 179.1**	15.0	2.9
LW CRE atmos.	0[b], – 3[d]	33	– 1.3	9.8	2.9	– 0.7	19.5	3.5
Net CRE atmos.	7[b], 1[d]	33	**1.9**	10.0	2.6	**3.6**	18.9	4.1
Surface								
SW down all-sky surface	185[b], 186[c], 187[d]	38	**187.4**	20.8	4.5	**189.6**	15.8	4.7
SW up all-sky surface	– 25[b], – 22[c], – 23[d]	37	– 23.9	9.4	2.0	– 24.6	10.5	2.3
SW absorbed all-sky surface	160[b], 164[c], 164[d]	37	**163.4**	12.1	3.0	**165.0**	12.2	3.8
SW down clear-sky surface	247[b], 244[d]	37	**244.8**	15.4	2.8	**249.7**	13.3	3.6
SW up clear-sky surface	33[b], 30[d]	36	30.2	12.7	2.3	31.1	12.8	2.9
SW absorbed clear-sky surface	214[b], 214[d]	36	**214.6**	11.0	2.2	**218.5**	15.5	3.6
SW CRE surface	– 54[b], – 50[d]	36	**– 51.2**	20.4	4.0	**– 53.5**	16.7	4.1
LW down all-sky surface	342[b], 341[c], 344[d]	38	**343.8**	20.3	5.2	**340.1**	18.5	4.3
LW up all-/clear-sky surface	398[b], 399[c], 398[d]	37	– 399.9	11.7	3.0	– 398.7	10.7	2.6
LW net all-sky surface	– 56[b], – 58[c], – 54[d]	37	**– 56.2**	14.0	3.6	**– 58.6**	15.7	3.2
LW down clear-sky surface	314[b], 314[d]	33	**318.0**	22.5	5.1	**314.5**	25.8	5.5
LW net clear-sky surface	– 84[b], – 84[d]	33	**– 81.7**	16.1	3.5	**– 83.9**	15.9	3.7
LW CRE surface	28[b], 30[d]	33	25.5	7.5	2.2	25.3	13.3	3.3
Net CRE surface	– 26[b],– 20[d]	33	**– 25.4**	15.3	3.6	**– 28.2**	24.4	4.4
Net radiation surface	104[b], 106[c], 110[d]	37	107.2	13.1	3.1	106.2	17.2	3.9
Latent heat flux	– 82[b], – 81[c]	38	– 85.3	18.0	3.5	– 85.8	13.9	3.9
Sensible heat flux	– 21[b], – 25[c]	39	**– 20.1**	13.2	2.7	**– 18.9**	13.1	2.6
Surface Imbalance	0.6[b], 0.5[c]	36	1.5	1.2	0.3	1.5	n.a.	n.a.

Given are recent reference estimates, together with the CMIP6 and CMIP5 model-calculated estimates in terms of their multi-model means, their inter-model spreads as well as their standard deviations

CMIP6 results from present study, CMIP5 results from Wild et al. (2019)

Units Wm^{-2}

Reference estimates from Loeb et al. (2018) ([a]), Wild et al. (2015, 2019) ([b]), L'Ecuyer et al. (2015) ([c]) and Kato et al. (2018) ([d])

Bold values indicate CMIP6 and CMIP5 multi-model means which are significantly different at the 95% confidence level and do not critically depend on the exact number of models. The submitted version of this manuscript was based on a lower number of models available at the time (25 models), but the conclusions remained virtually identical in the present revised manuscript, despite the consideration of 50% additional models that became available in the meantime.

The reference values for the magnitudes of the TOA components stem from the Energy Balanced and Filled (EBAF) data set Edition 4.0 for the period 2001–2010 that resulted from the CERES mission (Loeb et al. 2018). In this mission, filtered radiances in the shortwave (between 0.3 and 5 µm), total (0.3 and 200 µm), and window (8 and 12 µm) regions are measured on board of the NASA satellites Terra and Aqua, with longwave radiances determined as differences between total and shortwave channel radiances. The uncertainty of the outgoing longwave flux at the TOA as measured by CERES due to the uncertainty in calibration is ~ 3.7 W m^{-2} (2 σ), whereas the uncertainty in the shortwave reflected flux is ~ 2% (2 σ), or equivalently 2 Wm^{-2} (Loeb et al. 2009). The CERES EBAF data set is gap-filled and adjusts the shortwave and longwave TOA fluxes within their range of uncertainty to be consistent with independent estimates of the global heating rate based upon in situ ocean observations (Loeb et al. 2018).

As references for the surface components, I use a number of recent estimates which are derived by independent approaches. Kato et al. (2018) developed an algorithm that forces computed TOA fluxes to match with the abovementioned CERES-EBAF TOA fluxes by adjusting surface, cloud, and atmospheric properties. Surface irradiances as provided in the CERES-EBAF surface product are subsequently adjusted using radiative kernels. L'Ecuyer et al. (2015) made use of a variety of satellite-derived products, and reintroduced energy and water cycle closure information lost in the development of these independently derived products through a variational method that explicitly accounts for the relative accuracies in all component fluxes. Wild et al. (2013, 2015, 2019) made use of the information contained in the direct flux measurements taken at worldwide surface observation sites and took into account the associated bias structure of a large number of GCMs to infer best estimates for the magnitudes of the global mean surface energy balance components. After decades of large discrepancies in published reference estimates for the global surface energy budget components, the abovementioned recent independent approaches provide estimates that converge to within a few Wm^{-2} on a global mean basis (Wild 2017). This increases the confidence in these references and enhances their usefulness as guidance in the assessment of the CMIP6 global mean energy budget components as discussed in the following.

3 Results—all-sky budgets

3.1 Shortwave components

The global annual mean incoming shortwave radiation at the TOA in 37 CMIP6 models is shown in Fig. 1, with the quantification of the associated multi-model mean, range and standard deviation of model estimates given in Table 1. It is evident, that most models use a solar constant near 1361 Wm^{-2} (four times the values presented in Fig. 1, which represent the incoming shortwave radiation at the TOA per square meter on the Earth's sphere, whereas the solar constant relates to the same quantity but per square meter on the cross-section of the Earth's sphere). This is consistent

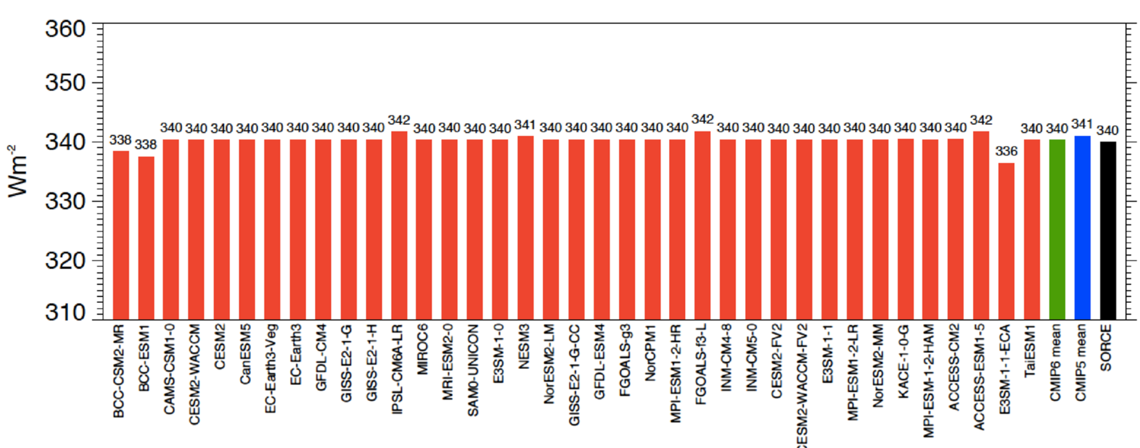


Fig. 1 Global annual mean incoming shortwave radiation at the TOA as simulated by 37 individual CMIP6 models (red bars), by the CMIP6 multi-model mean (green bar), and the CMIP5 multi-model mean (blue bar). Reference estimate from the NASA Solar Radiation and Climate Experiment (SORCE, Kopp and Lean 2011) (black bar). Values can be multiplied by a factor of four to infer the solar constants used in the CMIP6 models. Units Wm^{-2}

with current best estimates from space-based observations of 1361 Wm^{-2} (Kopp and Lean 2011) provided by the NASA Solar Radiation and Climate Experiment (SORCE). There remain, however, a few models which still use a solar constant that deviates substantially from the 1361 Wm^{-2}. The highest global mean incoming shortwave radiation at the TOA used in a CMIP6 model corresponds to a solar constant of 1367 Wm^{-2}, the lowest to 1346 Wm^{-2}. It is further interesting to note from Table 1 that the multi-model mean incoming shortwave radiation at the TOA is lower by 0.9 Wm^{-2} in CMIP6 than in the preceding model generation CMIP5 also presented in Table 1. This signifies that on average the solar constant used in the CMIP6 models is lower by 3.6 Wm^{-2} than in CMIP5 (again considering a factor of four), enforced by the developments in the measurement technologies that accounted for a lower value of the solar constant (Kopp and Lean 2011). Note that the difference in the multi-model mean estimates of the incoming shortwave radiation at the TOA in CMIP6 and CMIP5 is statistically significant at the 95% confidence level, as denoted by bold values in Table 1. The statistical significance at the 95% level of the differences between the CMIP5 and CMIP6 multi-model means in Table 1 has been determined by gaussian error propagation rules from the standard deviations of the individual models in CMIP5 and CMIP6.

The global annual mean shortwave absorption in the total climate system (TOA), within the atmosphere and at the Earth's surface of 37 CMIP6 climate models is shown in Fig. 2, with the statistical summary given in Table 1. The individual models vary in their simulated global mean shortwave budgets with standard deviations near 3 Wm^{-2} both at the TOA and the surface (Table 1). Table 1 further shows that the inter-model spread in these budgets in the CMIP6 models is as large as in the preceding model generation CMIP5, despite the slightly lower number of CMIP6 models providing the shortwave budgets (37 models) compared to CMIP5 (43 models, Wild et al. 2015).

Compared to the reference values, the multi-model mean TOA shortwave absorption, at 239.5 Wm^{-2} globally, closely matches the satellite-based reference estimates near 240 ± 2 Wm^{-2} (Table 1). This is favored by the fact that the various modelling groups aim at tuning their TOA energy fluxes to match the CERES-EBAF reference estimates on a global mean basis. Individual models, however, still differ by up to 9 Wm^{-2} from these reference estimates (Fig. 2). Given the tuning efforts undertaken by all modelling groups, this is surprising, as well as the fact that 9 out of 37 CMIP6 models simulate a TOA shortwave absorption outside the 2-sigma observational uncertainty ranges (± 2 Wm^{-2}) of the CERES reference values (tuning targets) given in Loeb et al. (2009).

Also at the surface, the multi-model mean shortwave absorption is, at 163.4 Wm^{-2} globally, close to recent reference estimates of 160–164 Wm^{-2} (Wild et al. 2015; L'Ecuyer et al. 2015; Kato et al. 2018), again with substantial deviations by some individual models. Still, two-thirds of the model-calculated estimates fall within the range given by the above references. The global multi-model mean surface shortwave absorption in CMIP6 is lower by 1.6 Wm^{-2} than in CMIP5 (165 Wm^{-2}) (statistically significant, Table 1). The lower multi-model mean absorption at the surface in CMIP6 is mostly due to a somewhat higher atmospheric shortwave absorption. The global multi-model mean atmospheric shortwave absorption in CMIP6 amounts to 76.0 Wm^{-2}, compared to the corresponding value of 74.4 Wm^{-2} in CMIP5 (difference statistically significant, Table 1). The higher atmospheric absorption in CMIP6 leads also to a global mean downward shortwave radiation at the Earth's surface, which is, at 187.4 Wm^{-2}, lower by more than 2 Wm^{-2} compared to CMIP5 (statistically significant, Table 1), and thereby in closer agreement with recent reference estimates (Table 1). But note also the large spread in the global mean downward shortwave radiation at the Earth's surface amongst the various CMIP6 models in Fig. 3 (upper panel), which amounts to as much as 21 Wm^{-2}. This spread is more than 8 Wm^{-2} larger than the spread in the corresponding surface absorbed shortwave radiation (Table 1). This implies that the surface albedos in some of the CMIP6 models partly compensate for the discrepancies in the simulated incoming shortwave radiation at the Earth's surface, with a tendency for higher and lower surface albedos in models with high and low incoming shortwave radiation, respectively (correlation coefficient 0.73).

3.2 Longwave components

Global annual mean estimates of the net longwave radiation at the TOA (outgoing longwave radiation, OLR), within the atmosphere and at the surface as simulated by the various CMIP6 models are shown in Fig. 4. The spread amongst the models amounts to 15.6, 17.2, and 14.0 Wm^{-2}, with standard deviations of 2.8, 4.2 and 3.6 Wm^{-2} for the OLR, the net atmosphere and net surface longwave radiation, respectively (Table 1). As for the shortwave budgets discussed above, also for the longwave budgets of the CMIP6 models this implies no convergence in their individual estimates compared to CMIP5 (Table 1). The inter-model spread in the simulated global mean OLR is even considerably larger in CMIP6 than in CMIP5, and also in terms of standard deviations, the CMIP6 models differ as much or more in their longwave budgets as their CMIP5 counterparts. In terms of absolute magnitudes, the CMIP6 multi-model mean, at 238.3 Wm^{-2} nearly matches the CMIP5 multi-model mean estimate, and is close to the satellite-based reference values of 240 ± 3 Wm^{-2} (Table 1). This is again largely a reflection of the tuning of the models to match the CERES values. Still, individual CMIP6 models do deviate by up to 11 Wm^{-2}

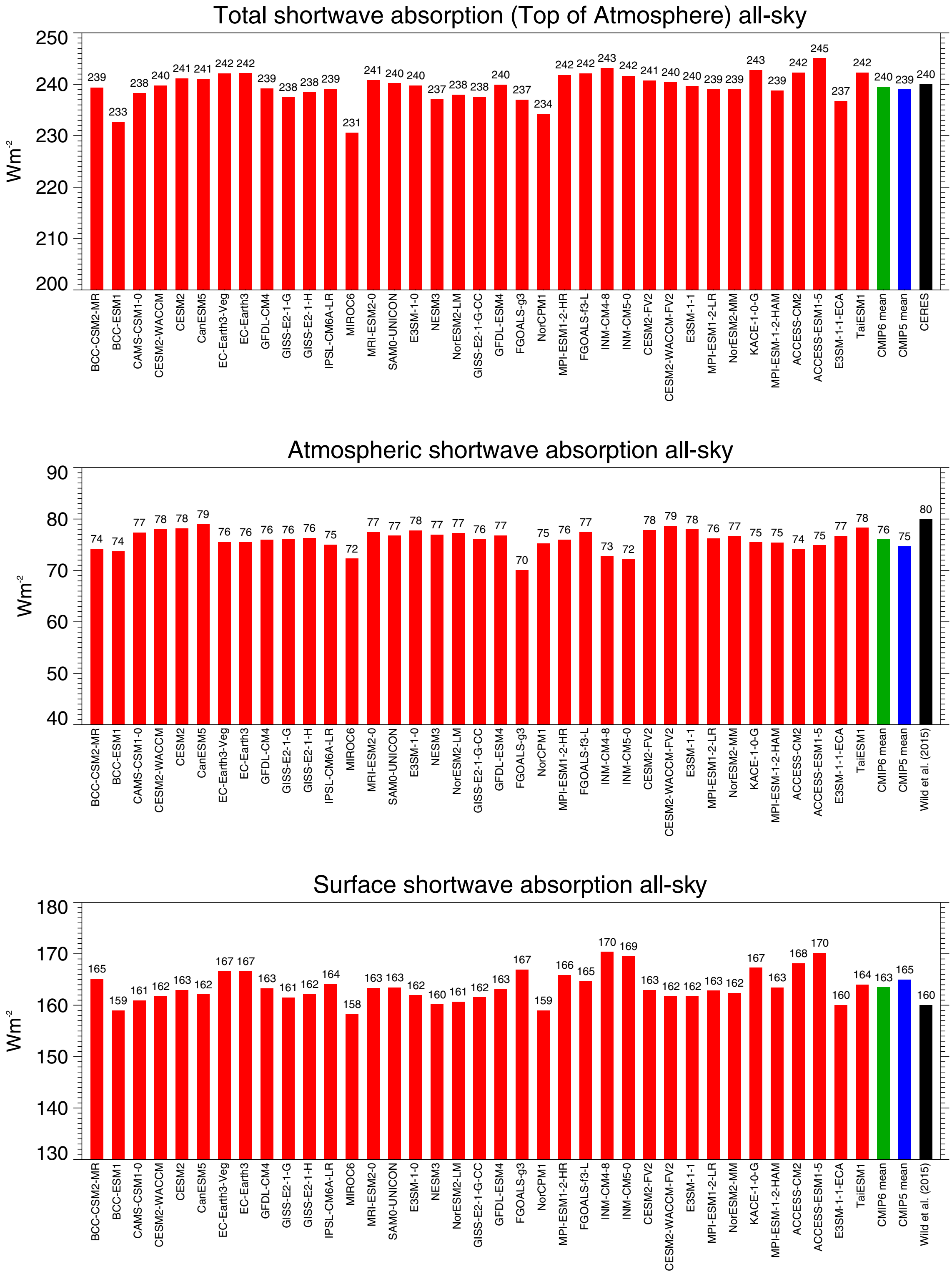


Fig. 2 Global annual mean shortwave all-sky radiation budgets representative for present-day climate. Shortwave radiation absorbed at the surface (lower panel), within the atmosphere (middle panel), and in the total climate system (TOA, upper panel), as simulated by 37 individual CMIP6 models (red bars). CMIP6 and CMIP5 multi-model means given by green and blue bars, respectively. Reference estimates from CERES (Loeb et al. 2018) and Wild et al. (2015) (black bars). Units Wm^{-2}

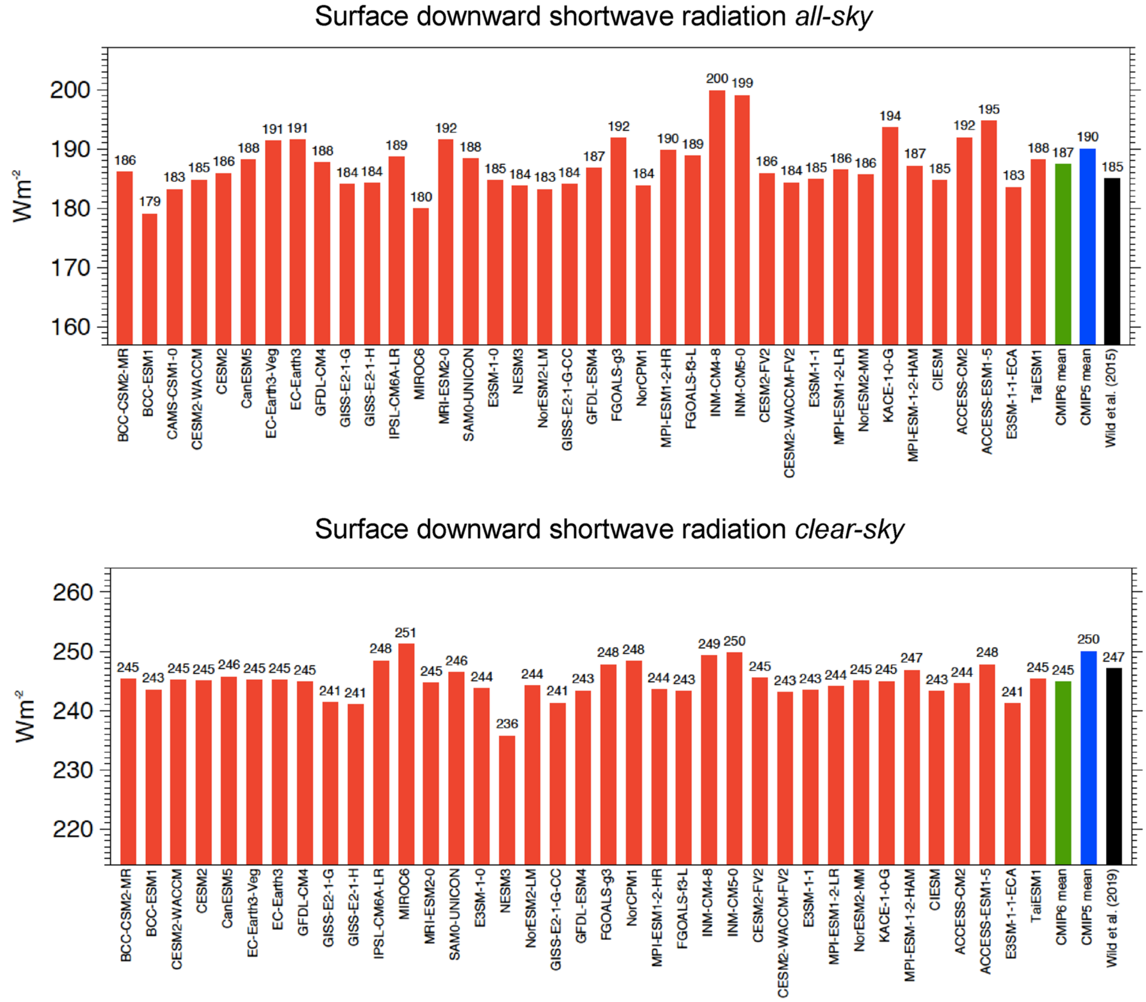


Fig. 3 Global annual mean downward shortwave radiation at Earth's surface representative for present-day climate under all-sky (upper panel) and clear-sky conditions (lower panel), as simulated by various CMIP6 models (red bars). CMIP6 and CMIP5 multi-model means given by green and blue bars, respectively. All-sky and clear-sky reference estimates from Wild et al. (2015, 2019), respectively (black bars). Clear-sky fluxes determined using Method II according to Cess and Potter (1987). Units Wm^{-2}

from this reference value (Fig. 4, upper panel). Specifically, 8 out of 40 CMIP6 models simulate a global mean OLR outside the 2-sigma observational uncertainty given in Loeb et al. (2009) for the CERES reference value.

The global mean net surface longwave budget in the multi-model mean in CMIP6 is, at $-56.2\ Wm^{-2}$, more than $2\ Wm^{-2}$ less negative than in CMIP5 ($-58.6\ Wm^{-2}$) (statistically significant, Table 1), i.e. the surface longwave cooling in CMIP6 is less effective than in the CMIP5 multi-model mean (Table 1). This is largely caused by a $3.7\ Wm^{-2}$ higher surface downward longwave radiation in the CMIP6 multi-model mean compared to CMIP5 (statistically significant, Table 1), which is not compensated by the $1.2\ Wm^{-2}$ higher multi-model mean surface upward longwave radiation in CMIP6 (Table 1). The higher global mean downward longwave radiation in the CMIP6 models, at $343.8\ Wm^{-2}$ in the multi-model mean comes now very close to the reference estimates given in Tables 1 and 3 (see discussion in Sect. 6). Yet note that, similarly to the downward shortwave radiation (Sect. 3.1), the spread in the global mean downward longwave radiation amongst the individual CMIP6 models remains considerable, covering as much as $20\ Wm^{-2}$ (Fig. 5, upper panel, Table 1).

3.3 Net radiation balance and non-radiative fluxes

If the Earth's climate system is in equilibrium, the shortwave radiation absorbed by the climate system should match the outgoing longwave radiation at the TOA on a global annual mean basis. Currently, with anthropogenic climate change, the climate system is slightly out of balance, with less longwave radiation emitted out to space than absorbed by our planet, so that energy is accumulating in the climate system, leading to global warming (Hansen et al. 2005). This

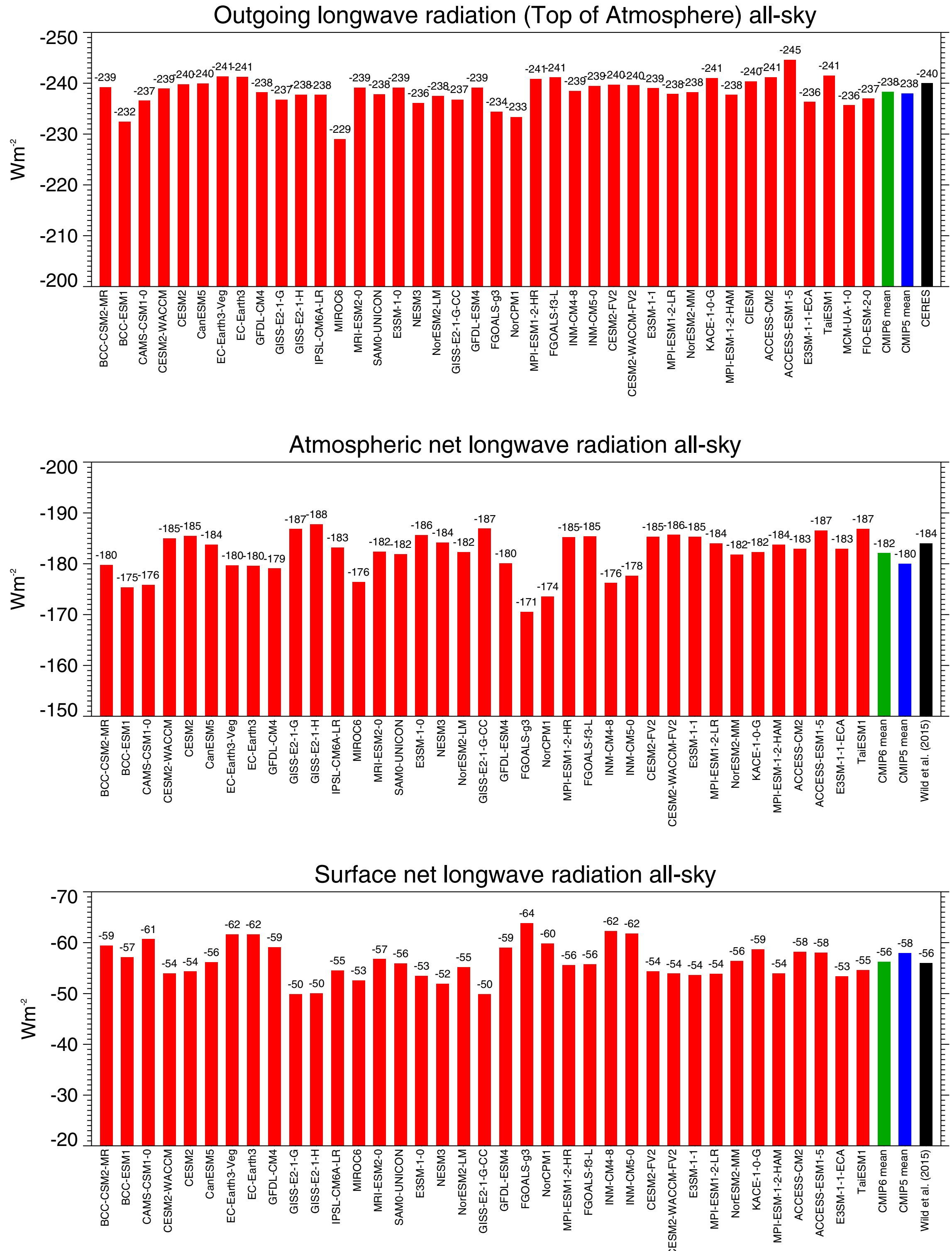


Fig. 4 Global annual mean longwave all-sky radiation budgets representative for present-day climate. Net longwave radiation at the surface (lower panel), within the atmosphere (middle panel), and emitted to space (upper panel) as simulated by various CMIP6 models (red bars). CMIP6 and CMIP5 multi-model means given by green and blue bars, respectively. Reference estimates from CERES (Loeb et al. 2018) and Wild et al. (2015) (black bars). Units Wm^{-2}

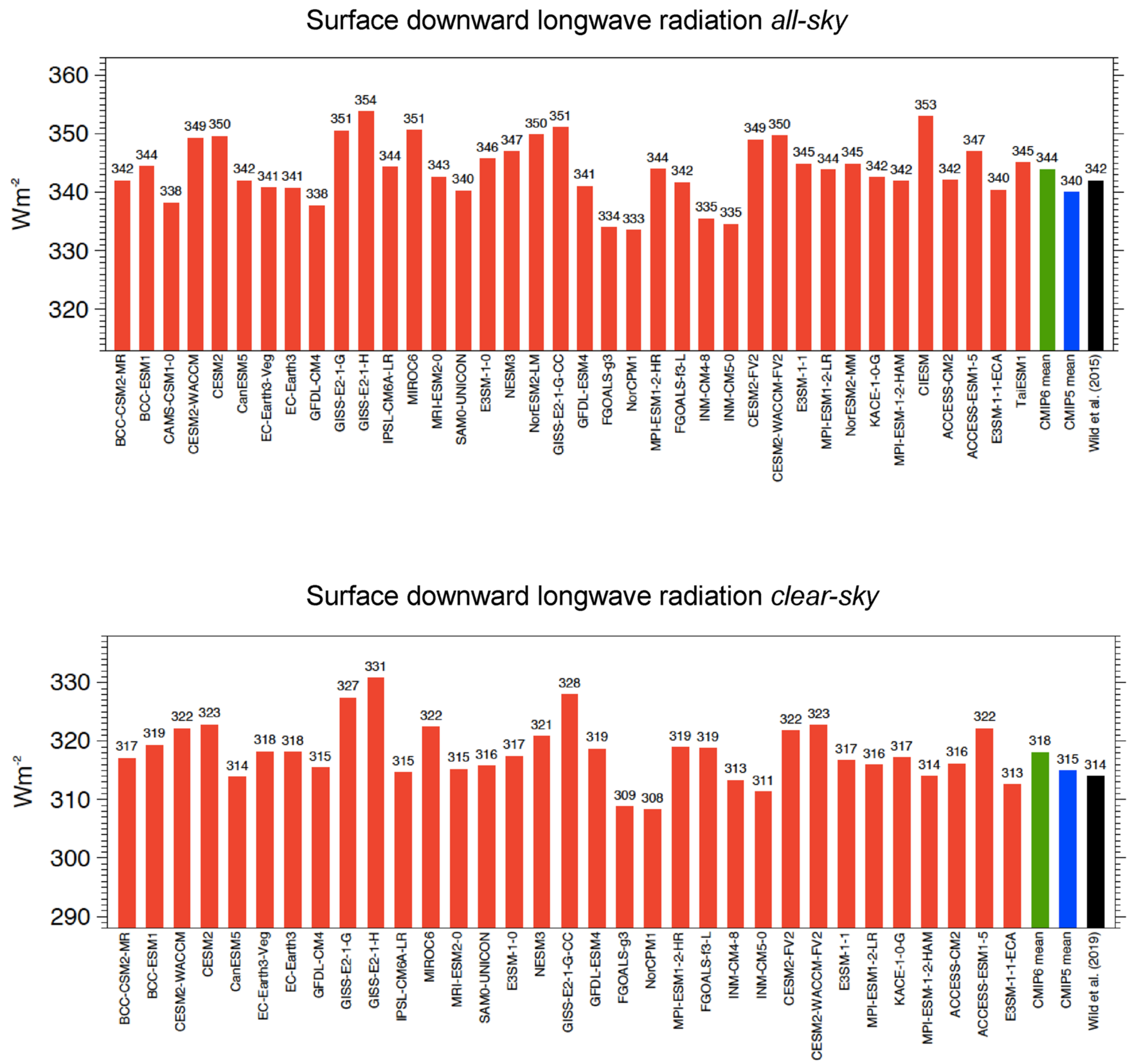


Fig. 5 Global annual mean downward longwave radiation at Earth's surface for present-day climate under all-sky (upper panel) and clear-sky conditions (lower panel), as simulated by various CMIP6 models (red bars). CMIP6 and CMIP5 multi-model means given by green and blue bars, respectively. All-sky and clear-sky reference estimates from Wild et al. (2015, 2019), respectively (black bars). Clear-sky fluxes are determined using Method II according to Cess and Potter (1987). Units Wm^{-2}

imbalance is estimated to be slighly less than 1 Wm^{-2} on a global mean basis, based on measurements of changes in the heat content of the oceans (Hansen et al. 2005; von Schuckmann et al. 2016; Johnson et al. 2016). These measurements stem from a global array of more than 4000 free-drifting profiling floats, known as ARGO, that record the temperature and salinity of the upper 2000 m of the oceans since the early 2000s, which allows for the first time a continuous monitoring of the change in the energy content in the oceans. Since more than 90% of the energy accumulation induced by the TOA radiation imbalance is stored in the world's oceans due to their large heat capacities, their change in the energy content is considered a good measure of the radiative imbalance at the TOA (e.g., Hansen et al. 2005; von Schuckmann et al. 2016; Johnson et al. 2016). Most of the CMIP6 models show a positive TOA imbalance of different magnitudes over the averaging period 2000-2014 considered here, with a multi-model mean of 1.1 Wm^{-2} not too far away from the reference estimates, such as the 0.7 Wm^{-2} given by Johnson et al. (2016) (Fig. 6, upper panel). Since energy might not be 100% preserved in some of the numerical schemes used in the climate models (Hourdin et al. 2017), not too much weight should be placed on the exact magnitudes of these simulated values. While most models show imbalances reasonably close to the reference estimates, the imbalances cover still a range of more than 4 Wm^{-2}, and some of the models show unrealistically high imbalances, pointing to problems in energy conservation in these models.

The surface net radiation (also known as surface radiation balance) consists of the absorbed shortwave radiation and

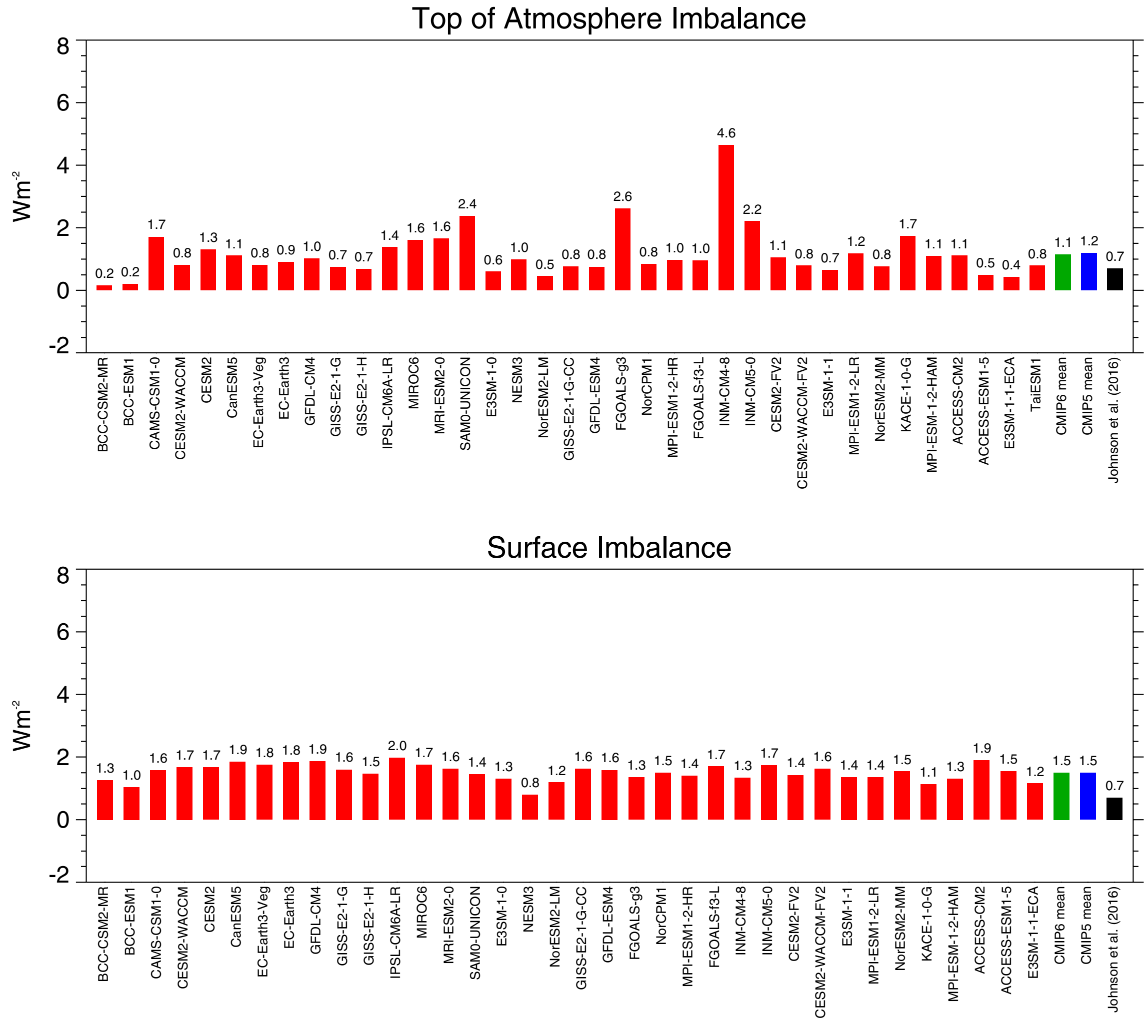


Fig. 6 Global annual mean energy imbalance at the TOA (upper panel) and at the Earth's surface (lower panel) for present-day conditions as simulated by various CMIP6 models (red bars). CMIP6 and CMIP5 multi-model means given by green and blue bars, respectively. Reference estimates from Johnson et al. (2016) (black bars). TOA energy imbalance determined as difference between absorbed shortwave radiation in the climate system (Fig. 2, upper panel) and the longwave emission to space (Fig. 4, upper panel). Surface imbalance determined as difference between surface net radiation (Fig. 7, upper panel) and the sum of surface sensible and latent heat fluxes (Fig. 7, middle/lower panels). Units Wm^{-2}

the net longwave cooling at the Earth's surface. It provides the energy available for the non-radiative fluxes of the surface energy balance, particularly the surface sensible and latent heat fluxes.

The global mean surface net radiation in the various CMIP6 models is shown in Fig. 7 (upper panel), together with their global mean latent (middle panel) and sensible heat fluxes (lower panel). The globally averaged surface net radiation in the CMIP6 models is, at 107.2 Wm^{-2}, slightly higher than the corresponding value of CMIP5 (106.2 Wm^{-2}). However, compared to CMIP5, the CMIP6 multi-model mean estimate is composed of a lower surface shortwave absorption, which is overcompensated by a lower surface net longwave cooling due to the higher downward longwave radiation. The surface net radiation in the CMIP6 global multi-model mean is still somewhat higher than the estimates provided by Wild et al. (2015) and L'Ecuyer et al. (2015) (Table 1). The spread and standard deviation in the global mean surface net radiation amongst the 37 individual CMIP6 models is, with 13 Wm^{-2} and 3.1 Wm^{-2} respectively, also still substantial, but somewhat smaller than in CMIP5.

The latent heat flux is an interesting quantity, since it makes the link between the global energy and water balance. The latent heat flux is the energy equivalent of evaporation, which in the global annual mean equals precipitation. Thus, differences in the magnitudes of the global mean latent heat flux in the various models reflect also differences in global

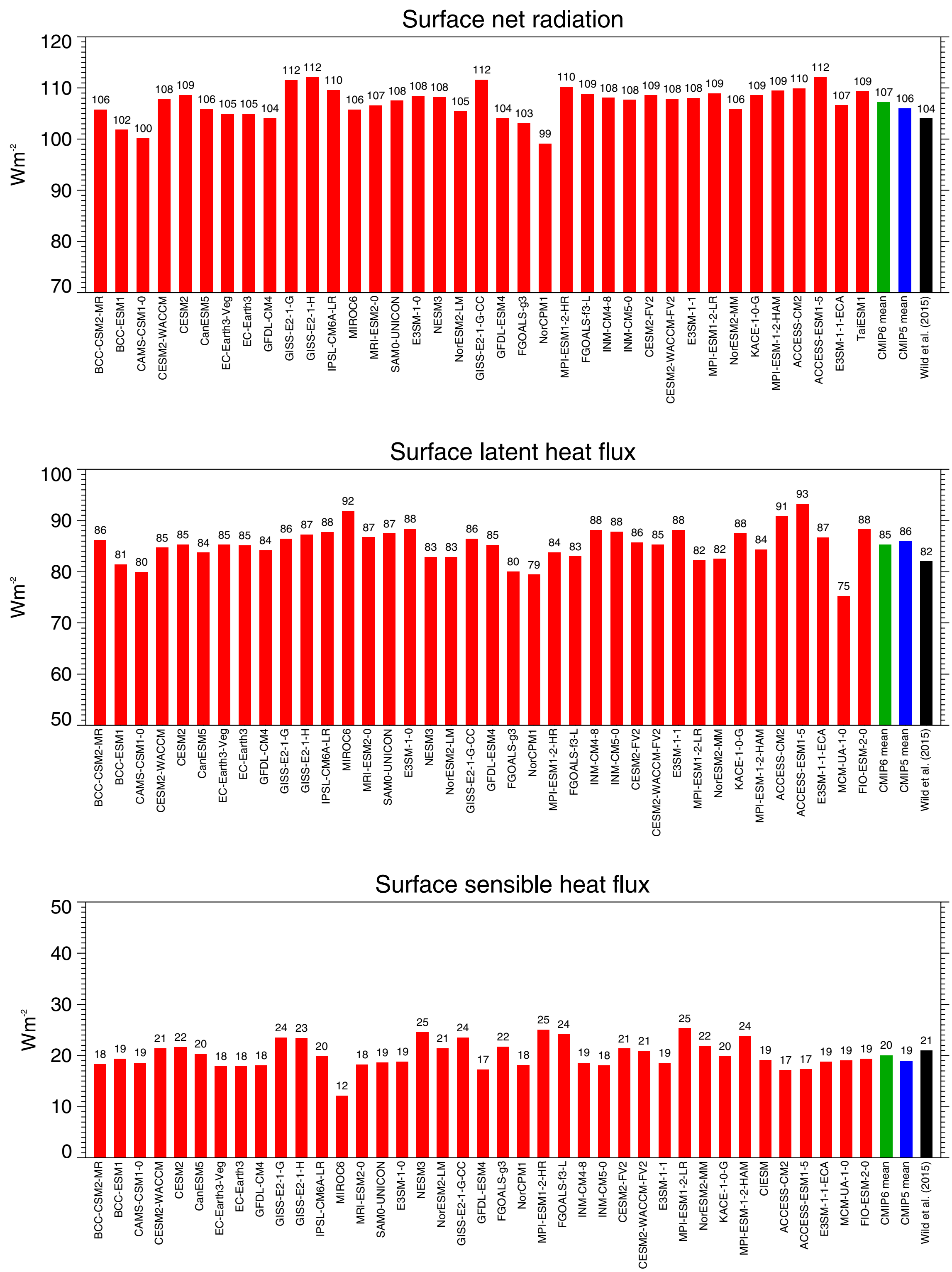


Fig. 7 Global annual mean surface net radiation (upper panel), latent heat fluxes (middle panel) and sensible heat fluxes (lower panel) representative for present-day climate as calculated by various CMIP6 models (red bars). CMIP6 and CMIP5 multi-model means given by green and blue bars, respectively. Reference estimates from Wild et al. (2015) (black bars). Units Wm^{-2}

evaporation and precipitation, and therefore in the intensity of the global water cycle. The multi-model mean latent heat flux is, at 85.3 Wm^{-2}, slightly above the recently published reference estimates (Table 1). Reference estimates for the global mean latent heat flux can be inferred from observational-based global precipitation estimates. However, these estimates are still afflicted with considerable uncertainties.

The individual CMIP6 models on the other hand differ in their simulated global mean latent heat fluxes by up to 18 Wm^{-2}, which corresponds to a spread of as much as 21%, considering the multi-model mean latent heat flux of 85 Wm^{-2} (Fig. 7, middle panel). This implies that the simulated global mean precipitation between the individual CMIP6 models also must have the same spread of 21%, or, in other words, the intensity of the global water cycle simulated by the different CMIP6 models varies in range of more than 20%). This is even larger than amongst the 43 CMIP5 models, where the intensity of the water cycle in terms of their global latent heat fluxes varied in a range of 16% (14 Wm^{-2}) (Table 1). Thus, there is no indication that the considerable discrepancies in the quantitative representation of the global water cycle in the various models reduce in CMIP6.

The global mean sensible heat flux is poorly constrained from an observational perspective. The CMIP6 models, with a multi-model mean sensible heat flux of 20.1 Wm^{-2} globally, are close to the estimate in Wild et al. (2015) of 21 Wm^{-2} as well as related estimates from reanalyses (Trenberth et al. 2009; Wild et al. 2013 and references therein), yet somewhat lower than the estimates given in Stephens et al. (2012) and L'Ecuyer et al. (2015) (Table 1). However, the global mean sensible heat fluxes in individual CMIP6 models vary in a range of 13 Wm^{-2}, which corresponds to a spread of as much as 65% (Fig. 7, lower panel, Table 1). This wide spread reflects the considerable uncertainties still inherent in the quantification of the sensible heat fluxes in climate models.

In addition, the global annual mean energy imbalance at the Earth's surface of the CMIP6 models is shown in Fig. 6 (lower panel), which refers to the difference between the surface net radiation and the sum of the surface sensible and latent heat fluxes, and which is closely related to the TOA energy imbalance discussed above. Most of this energy goes into the oceans, while a small fraction is stored in the terrestrial sub-surfaces and used for the melting of snow and ice. All models show a positive surface imbalance as expected with increasing greenhouse-gas forcing, with values mostly between 1 and 2 Wm^{-2}, and a multi-model mean of 1.5 Wm^{-2} (Table 1, Fig. 6, lower panel). This is slightly higher than the reference values which are somewhat below 1 Wm^{-2} (Hansen et al. 2005; von Schuckmann et al. 2016; Johnson et al. 2016), again potentially due to imperfect energy conservation in the models (Hourdin et al. 2017). The potential lack of precise energy conservation in the individual models may also be the reason that the TOA and surface imbalances are not obviously correlated across models.

4 Results—clear-sky budgets

4.1 Shortwave components

Shown in Fig. 8 are the global annual mean shortwave budgets in the absence of clouds ("clear-sky") of various CMIP6 models at the TOA (upper panel), within the atmosphere (middle panel) and at the surface (lower panel). The cloud-free fluxes in the climate models are determined according to the so-called "Method II" (Cess and Potter 1987; Potter et al. 1992), i.e. the clear-sky fluxes are determined at every model-timestep, irrespective of the presence or absence of clouds. Thus, clear-sky fluxes are also calculated during cloudy conditions in the models, just by removing the clouds in the radiative transfer calculations, but otherwise retaining the atmospheric conditions prevailing during these cloudy conditions. Observational reference estimates which consider only "true" cloud-free conditions (Method I according to Cess and Potter (1987), have therefore to be slightly adjusted to match the clear-sky definition as used in the model world (see Wild et al. 2019).

The shortwave clear-sky TOA budget determines the amount of shortwave radiation absorbed in the cloud-free climate system. In the CMIP6 global multi-model mean, this amounts to 287.3 Wm^{-2}, which perfectly matches the observational reference value from CERES (Loeb et al. 2018), slightly adjusted to satisfy Method II as described in Wild et al. (2019) to account for the different clear-sky definitions in models and observations as outlined in the paragraph above. Again the agreement between simulated and observed fluxes is partly an outcome of the tuning process of the models. The CMIP6 multi-model mean clear-sky shortwave TOA absorption is somewhat smaller than in CMIP5 by 1.3 Wm^{-2}, indicative of a slightly higher clear-sky planetary albedo in the CMIP6 multi-model mean (statistically significant, Table 1). The inter-model spread and standard deviation of the clear-sky shortwave TOA absorption amongst the CMIP6 models are almost half of the corresponding ones under all-sky conditions, as might be expected when the complicating cloud-effects are excluded in the flux calculations.

The absorption of shortwave radiation in the cloud-free atmosphere in the multi-model mean is, at 72.8 Wm^{-2} globally, higher by 2.7 Wm^{-2} than in the CMIP5 models (statistically significant, Table 1). This brings the CMIP6 multi-model mean in almost perfect match with the reference estimate of 73 Wm^{-2} determined in independent

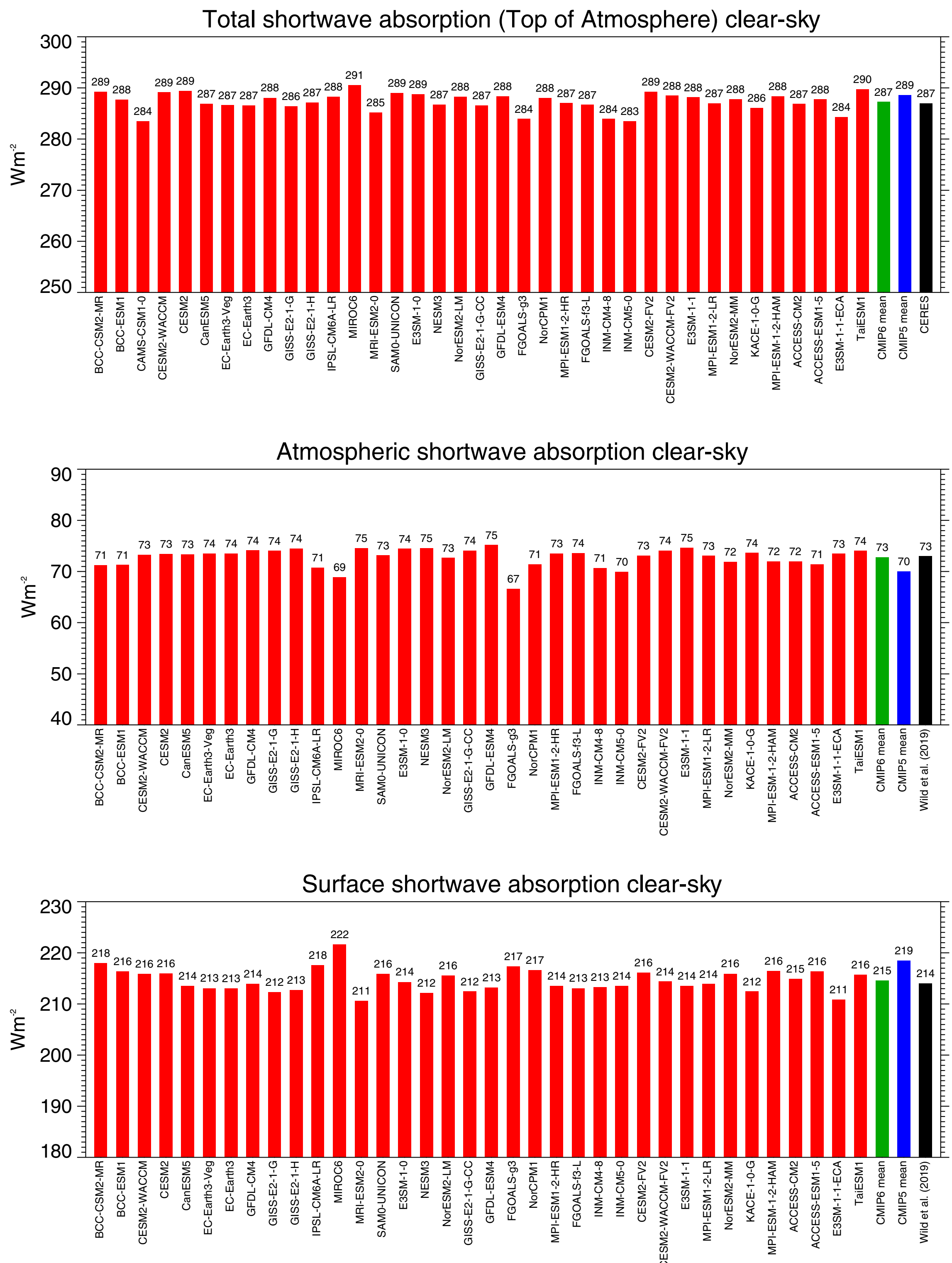


Fig. 8 Global annual mean shortwave clear-sky radiation budgets representative for present-day climate. Shortwave clear-sky radiation absorbed at the surface (lower panel), within the atmosphere (middle panel), and in the total climate system (TOA, upper panel) as simulated by various CMIP6 models (red bars). CMIP6 and CMIP5 multi-model means given by green and blue bars, respectively. Reference estimates from CERES (Loeb et al. 2018) and Wild et al. (2019) (black bars). Units Wm^{-2}

approaches by Wild et al. (2015) and Kato et al. (2018) (Table 1). It is noteworthy that not only the multi-model mean but also many individual models closely match the reference values of 73 Wm^{-2}. 33 out of 36 models determine the atmospheric clear-sky shortwave absorption to within 2 Wm^{-2} from these reference values (Fig. 8, middle panel). This is even more remarkable, as this quantity has been notoriously underestimated over generations of GCMs, as further discussed in Sect. 6. The shortwave clear-sky budgets simulated in the various CMIP6 models are generally more consistent than in CMIP5, as evident in smaller spreads and standard deviations (Table 1). This is in contrast to most other components of the global energy balance which typically show no reduction in terms of inter-model spreads and standard deviations from CMIP5 to CMIP6.

The absorption of shortwave radiation at the Earth's surface under cloud-free conditions is in the CMIP6 multi-model mean at 214.6 Wm^{-2} globally almost 4 Wm^{-2} lower than in CMIP5 (statistically significant, Table 1). This is primarily caused by the higher clear-sky shortwave atmospheric absorption (by 2.7 Wm^{-2}), as well as by the slightly lower overall (net TOA) clear-sky shortwave absorption (by 1.3 Wm^{-2}) as mentioned above and seen in Table 1. The CMIP6 multi-model mean clear-sky shortwave absorption is also in near perfect match with the two independently derived reference estimates of Kato et al. (2018) and Wild et al. (2019), both consistently at 214 Wm^{-2}, and thus no longer indicates an overestimation as noted in the CMIP5 models (Table 1, Wild et al. 2019) and in previous model generations. Again it is remarkable, that 29 out of 36 CMIP6 models simulate a global mean clear-sky surface shortwave absorption that is within 2 Wm^{-2} of the above reference estimates (Fig. 8, lower panel).

The lower clear-sky surface shortwave absorption in the CMIP6 models is also in line with a substantially lower surface downward shortwave clear-sky radiation in these models, which is, at 244. 8 Wm^{-2} lower by almost 5 Wm^{-2} than in CMIP5 (statistically significant, Table 1). This lower surface downward shortwave clear-sky radiation in the CMIP6 multi-model mean leads then again to a better agreement with the reference estimates of Wild et al. (2019) and Kato et al. (2018) (Table 1).

Overall, the global mean shortwave radiation budget under cloud-free conditions in CMIP6 is in remarkable agreement with recent reference estimates, not only in its multi-model mean which is within 1 Wm^{-2} of the reference values for the total (TOA), atmosphere and surface absorption, but also in the majority of the individual models which are in close agreement with these references. This indicates a clear improvement compared to previous model generations in these quantities, and increases confidence both in the model-calculated and reference estimates of the shortwave clear-sky budgets.

4.2 Longwave components

The global mean longwave budget under cloud-free conditions of the various CMIP6 models is presented in Fig. 9, with the clear-sky OLR in the upper panel, and the longwave clear-sky budget in the atmosphere and at the surface in the middle and lower panels, respectively.

The CMIP6 multi-model-mean clear-sky OLR is, at – 262.4 Wm^{-2} globally, lower by 1 Wm^{-2} compared to CMIP5. Quantitatively, both these amounts are a fair bit smaller than the latest CERES Ed 4.0 reference estimate (– 268 Wm^{-2}, Loeb et al. 2018), slightly adjusted to – 267 Wm^{-2} to conform with Method II (Wild et al. 2019). As in CMIP5, the lower model values might have been favored by earlier CERES product releases (Ed 2.8 and Ed2 SYN1deg-Month) with somewhat smaller clear-sky OLR estimates, which may have been used as target estimates in the model tuning process.

The net longwave cooling of the cloud-free atmosphere is, at – 180.9 Wm^{-2}, somewhat stronger in the CMIP6 multi-model mean than in CMIP5, particularly due to a stronger clear-sky emission towards the surface (clear-sky surface downward longwave radiation), which is higher by 3.5 Wm^{-2} in the global multi-model mean (statistically significant, Table 1). Accordingly, the global multi-model mean net longwave cooling at the Earth's surface is weaker in CMIP6 compared to CMIP5 by 2.2 Wm^{-2}, since the slightly higher surface longwave upward radiation in CMIP6 of 1.2 Wm^{-2} cannot compensate for the 3.5 Wm^{-2} additional energy that the surface obtains from the enhanced downward longwave clear-sky emission in CMIP6 (Table 1, Fig. 5, lower panel). The discrepancies amongst the simulated surface net longwave clear-sky budgets in the various CMIP6 models remain substantial (Fig. 9, lower panel), and are substantially larger both in terms of spread and standard deviation compared to their shortwave counterparts, i.e. the surface shortwave clear-sky absorption, despite their smaller absolute amounts (cf. Fig. 8 lower panel, Table 1).

In terms of absolute values, the downward longwave clear-sky radiation is, at 318.0 Wm^{-2} now larger than the independent reference estimates of Wild et al. (2019) and Kato et al. (2018), both at 314 Wm^{-2}. Note also the particularly large spread in the downward longwave clear-sky radiation amongst the 37 CMIP6 models (22.5 Wm^{-2}, Fig. 5 lower panel), which is thus the quantity with the largest spread of all CMIP6 energy balance components discussed in this study. This already applied for the CMIP5 models (Wild et al. 2019). Also, as in CMIP5 and in earlier model intercomparison projects, the spread amongst the simulated

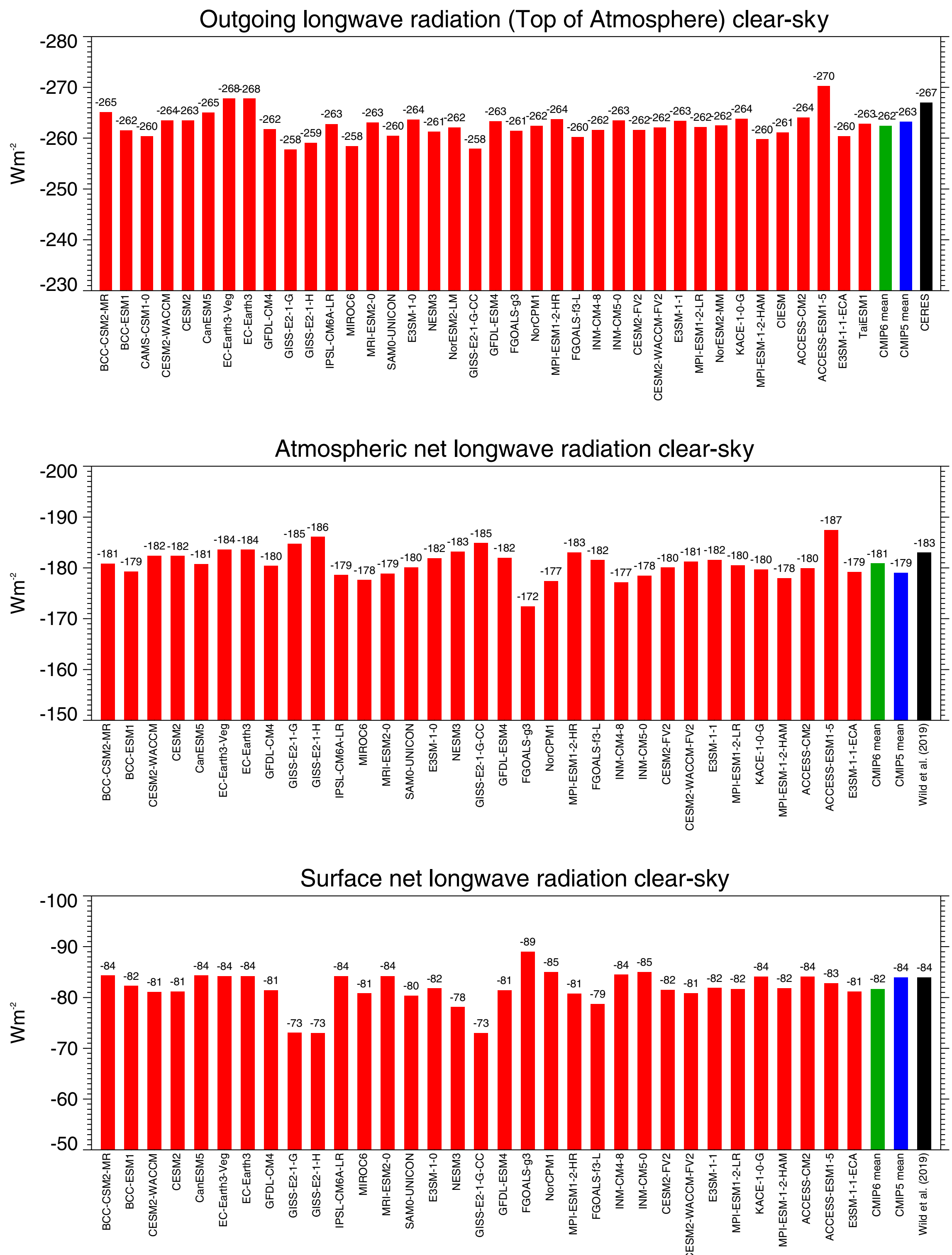


Fig. 9 Global annual mean longwave clear-sky radiation budgets representative for present-day climate. Net clear-sky longwave radiation at the surface (lower panel), within the atmosphere (middle panel), and emitted to space (upper panel) as simulated by various CMIP6 models (red bars). CMIP6 and CMIP5 multi-model means given by green and blue bars, respectively. Reference estimates from CERES (Loeb et al. 2018) and Wild et al. (2019) (black bars). Units Wm^{-2}

global mean downward longwave clear-sky radiation in the various CMIP6 models is larger (22.5 Wm^{-2}) than in their all-sky counterparts (20.3 Wm^{-2}) (Fig. 5 and Table 1). This confirms findings based on earlier model generations, that the simulated clouds tend to mask rather than to enhance the notable discrepancies which exist between these clear-sky flux estimates in the various models (Wild 2008, 2019). This indicates that the downward longwave radiation from the cloud-free atmosphere is largely contributing to the spread noted in the (all-sky) downward longwave radiation across the various CMIP6 models.

Overall, under cloud-free conditions, the longwave budgets in the CMIP6 models still show substantial discrepancies and are not as consistently simulated as their shortwave counterparts, as reflected in considerably larger standard deviations and inter-model spreads (Table 1).

5 Results—global cloud radiative effects

The quantification of both all-sky and clear-sky budgets allows an estimation of the effects that clouds exert globally on the energy flows in the various GCMs. In the following, the global cloud radiative effects (CRE) on the shortwave, longwave and net budgets are discussed as they apply at the TOA, within the atmosphere and at the Earth's surface.

5.1 TOA cloud radiative effects

The TOA shortwave absorption in the CMIP6 multi-model mean under clear-sky and all-sky conditions, at 287.3 and 239.5 Wm^{-2}, respectively, differs by 47.8 Wm^{-2} globally. This implies that the overall effect of clouds in the CMIP6 models is to reduce the absorption of shortwave radiation in the climate system by – 47.8 Wm^{-2} (TOA shortwave CRE). This is in close agreement with the CERES EBAF reference estimate (Loeb et al. 2018), adjusted according to Method II for an exact comparison with climate models, of – 47 Wm^{-2} (Wild et al. 2019). However, the spread in the TOA shortwave CRE amongst the individual CMIP6 models is again substantial, ranging from – 41 to – 60 Wm^{-2} globally (Fig. 10 upper panel). This range is larger than in the CMIP5 models, despite the somewhat smaller number of models considered in CMIP6 (Table 1). Still two-third of the CMIP6 models simulate a global mean TOA shortwave CRE within 2 Wm^{-2} of the reference estimate.

Similarly, the difference between the global mean OLR under clear-sky and all-sky conditions in the CMIP6 multi-model mean, at – 262.4 Wm^{-2} and – 238.3 Wm^{-2}, respectively, differs by 24.1 Wm^{-2}. This implies that clouds globally reduce the longwave emission to space by 24.1 Wm^{-2} (TOA longwave CRE) in the CMIP6 multi-model mean, causing a gain of energy for the climate system of slightly lower amount than in the CMIP5 multi-model mean (Table 1, Fig. 11 upper panel). The TOA longwave CRE in both CMIP6 and CMIP5 multi-model means is weaker than in the CERES reference estimate adjusted for Method II (28 Wm^{-2}, Table 1), due to the lower clear-sky OLR in the models as discussed in the previous section. The global mean TOA longwave CRE in the individual CMIP6 models ranges from 19 to 29 Wm^{-2} (Fig. 11 upper panel).

In terms of the net effect of clouds on the energy content of the climate system (TOA net CRE), the enhanced shortwave reflection of – 47.8 Wm^{-2} thus globally dominates over the longwave energy gain of 24.1 Wm^{-2} in the CMIP6 multi-model mean, which implies an overall energy reduction of – 23.7 Wm^{-2} for the climate system (TOA net CRE), close to the corresponding value of the CMIP5 multi-model mean (Table 1, Fig. 12 upper panel). This overall energy loss due to clouds is stronger than indicated in the corresponding CERES satellite reference estimates on the order of 5 Wm^{-2}, primarily due to the weaker trapping of longwave outgoing radiation, plus a slightly stronger shortwave reflection back to space in the CMIP6 models (Table 1). The global mean TOA net CRE in the individual CMIP6 models ranges from – 17 to – 31 Wm^{-2} (Fig. 12 upper panel). Thus also most of the individual models simulate a more negative TOA net CRE than the reference estimates suggest.

5.2 Atmospheric cloud radiative effects

The presence of clouds slightly enhances the shortwave absorption in the atmospheric column in all CMIP6 models (Fig. 10, middle panel). The CMIP6 multi-model mean atmospheric shortwave CRE is, at 3.2 Wm^{-2} globally, somewhat weaker than the CMIP5 multi-model mean estimate (statistically significant, Table 1).

The atmospheric cloud effect in the longwave is marginal in the CMIP6 multi-model mean, at -1.3 Wm^{-2} globally (Table 1), as in CMIP5. Individual CMIP6 model estimates vary in a range from – 6 to + 4 Wm^{-2} (Fig. 11, middle panel). This leaves a global mean net effect of clouds on the atmospheric column absorption of 1.9 Wm^{-2} in the CMIP6 multi-model global mean (3.6 Wm^{-2} in CMIP5, difference statistically significant, Table 1). The net effect of clouds is thus a slight enhancement of the atmospheric energy content globally. This slight enhancement is found in half of the individual CMIP6 models and reaches up to 8 Wm^{-2}, while the other half shows a near zero effect or a slight reduction (Fig. 12 middle panel).

5.3 surface cloud radiative effects

The effect of clouds on the absorption of shortwave radiation at the Earth's surface (surface shortwave CRE) in the CMIP6 multi-model mean is a global mean reduction of

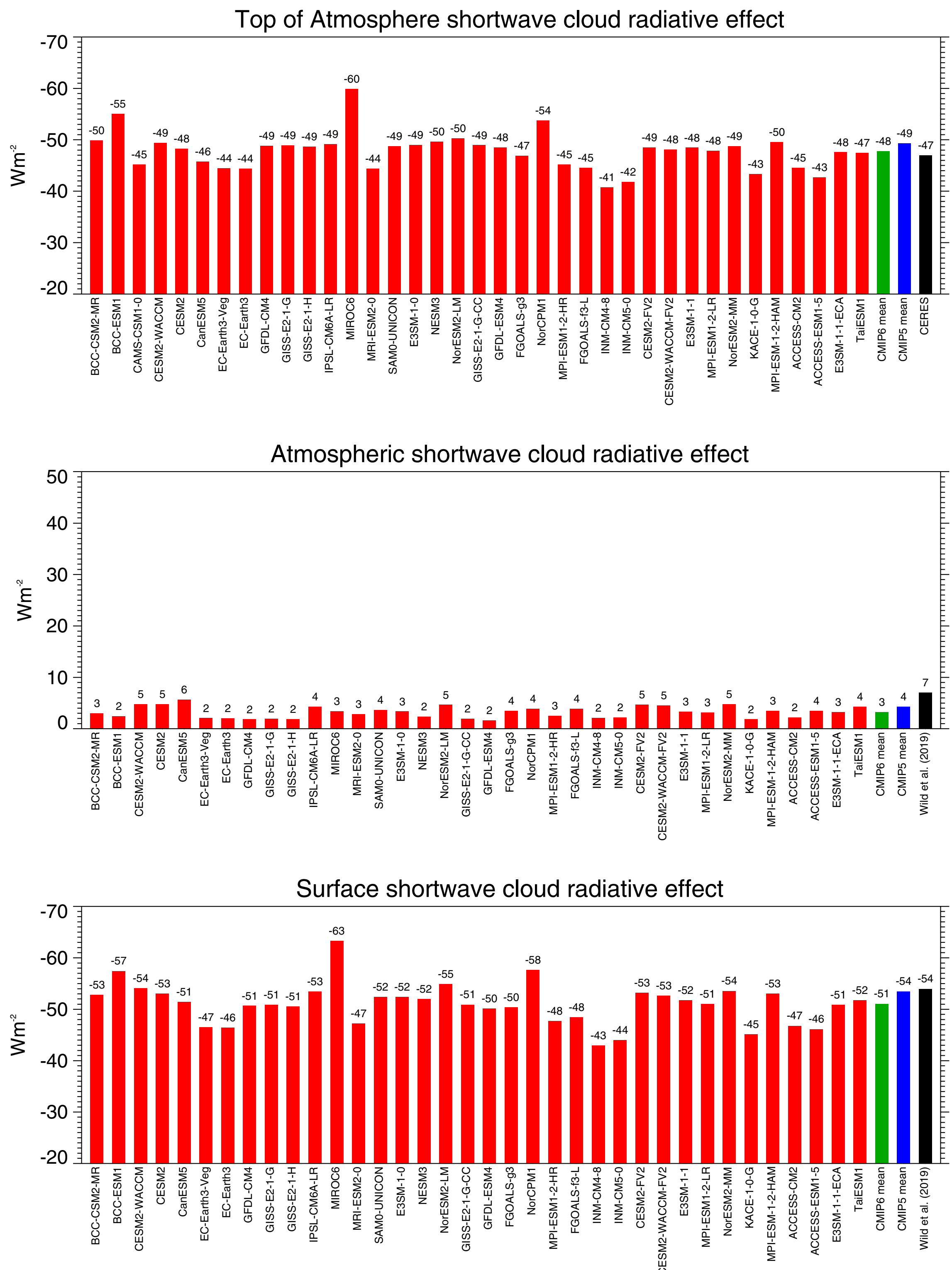


Fig. 10 Global annual mean shortwave cloud radiative effects at the TOA (upper panel), within the atmosphere (middle panel) and at the surface (lower panel) representative for present-day climate, as simulated by various CMIP6 models (red bars). Cloud radiative effects determined as differences between the respective all-sky (Fig. 2) and clear-sky (Fig. 8) shortwave radiation budgets of the individual CMIP6 models. CMIP6 and CMIP5 multi-model means given by green and blue bars, respectively. Reference estimates from CERES (Loeb et al. 2018) and Wild et al. (2019) (black bars). Units Wm^{-2}

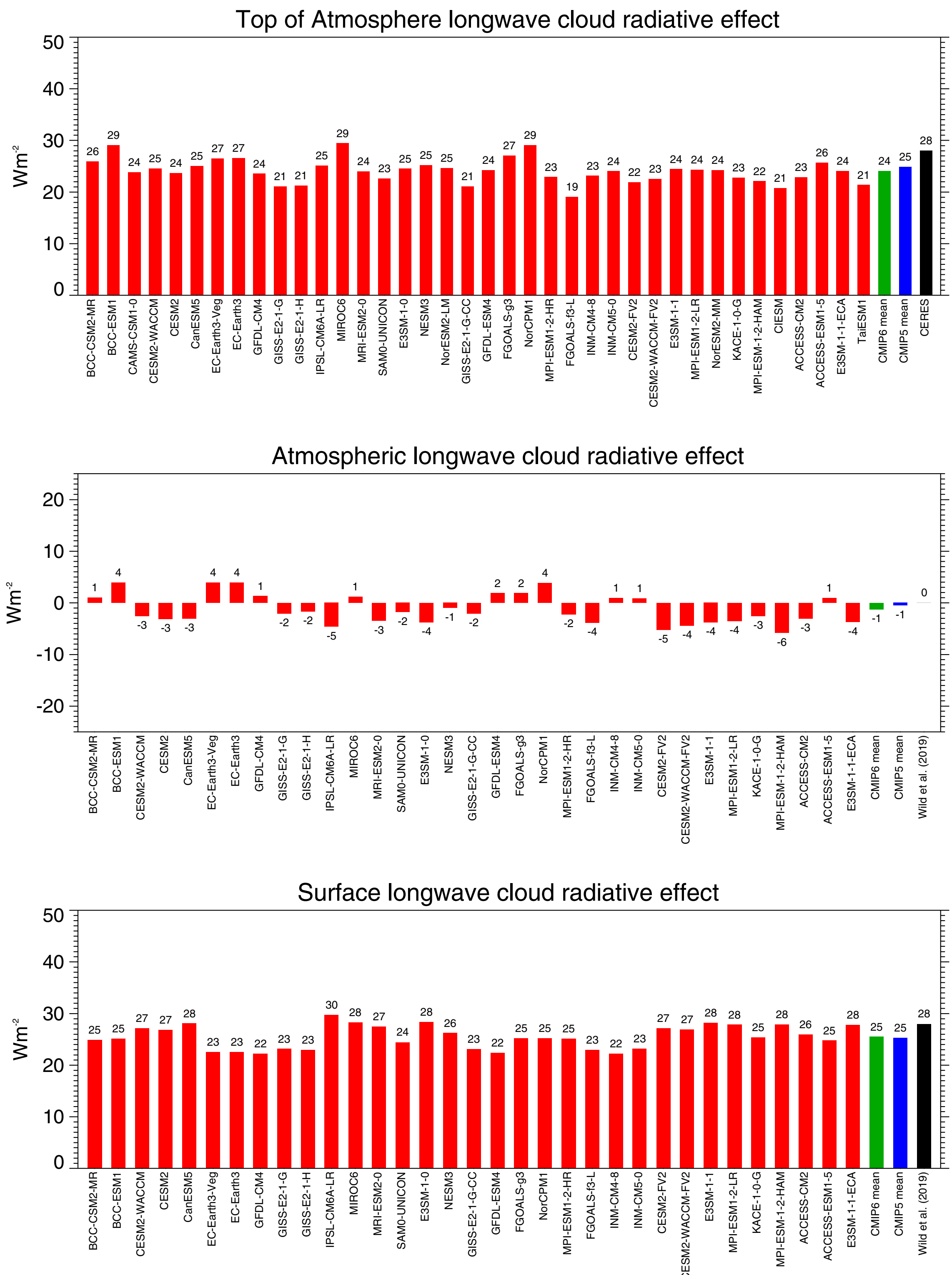


Fig. 11 Global annual mean longwave cloud radiative effects at the TOA (upper panel), within the atmosphere (middle panel) and at the surface (lower panel) representative for present-day climate, as simulated by various CMIP6 models (red bars). Cloud radiative effects determined as differences between the respective all-sky (Fig. 4) and clear-sky (Fig. 9) longwave radiation budgets of the individual CMIP6 models. CMIP6 and CMIP5 multi-model means given by green and blue bars, respectively. Reference estimates from CERES (Loeb et al. 2018) and Wild et al. (2019) (black bars). Units Wm^{-2}

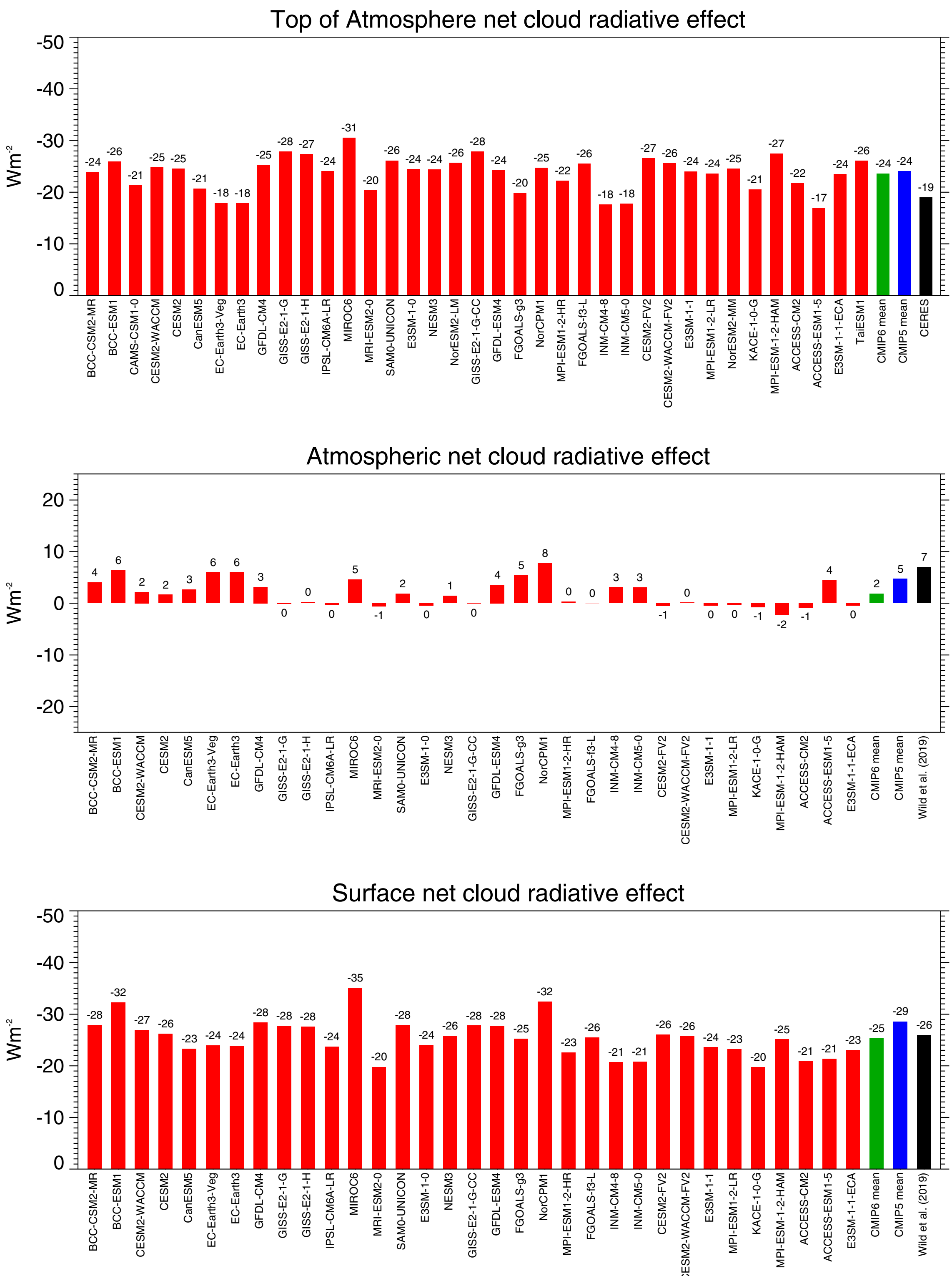
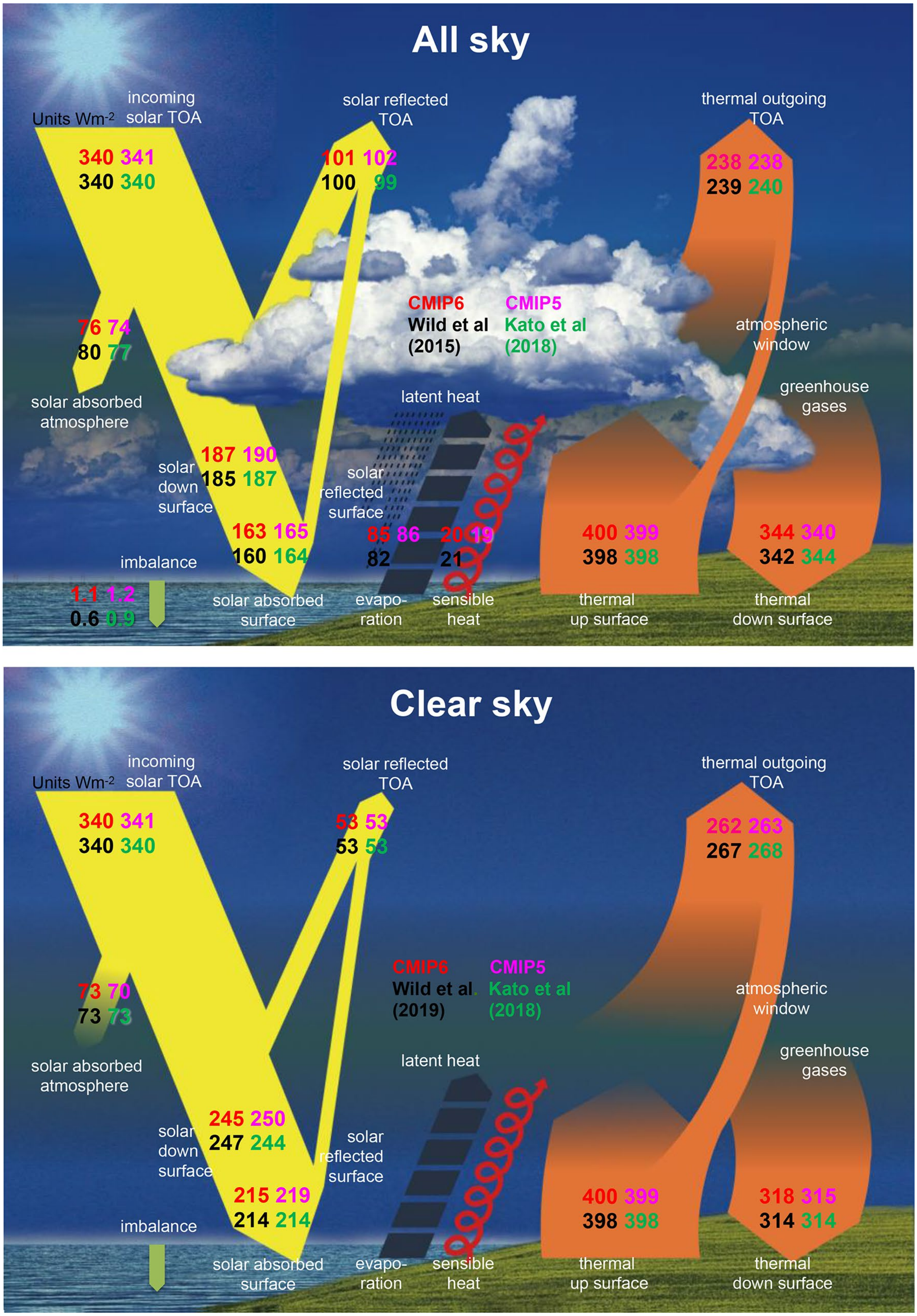


Fig. 12 Global annual mean net (shortwave + longwave) cloud radiative effects at the TOA (upper panel), within the atmosphere (middle panel) and at the surface (lower panel) representative for present-day climate, as simulated by various CMIP6 models (red bars). Net cloud radiative effects defined as differences between the respective all-sky and clear-sky net radiation budgets of the individual CMIP6 models. CMIP6 and CMIP5 multi-model means given by green and blue bars, respectively. Reference estimates from CERES (Loeb et al. 2018) and Wild et al. (2019) (black bars). Units Wm^{-2}


All sky
Units Wm-2
incoming solar TOA
340 341
340 340
solar reflected TOA
101 102
100 99
thermal outgoing TOA
238 238
239 240
CMIP6
Wild et al (2015)
CMIP5
Kato et al (2018)
76 74
80 77
solar absorbed atmosphere
atmospheric window
greenhouse gases
latent heat
187 190
185 187
solar down surface
solar reflected surface
163 165
160 164
85 86
82
20 19
21
400 399
398 398
344 340
342 344
imbalance
1.1 1.2
0.6 0.9
solar absorbed surface
evapo-ration
sensible heat
thermal up surface
thermal down surface
Clear sky
Units Wm-2
incoming solar TOA
340 341
340 340
solar reflected TOA
53 53
53 53
thermal outgoing TOA
262 263
267 268
CMIP6
Wild et al (2019)
CMIP5
Kato et al (2018)
73 70
73 73
solar absorbed atmosphere
atmospheric window
greenhouse gases
latent heat
245 250
247 244
solar down surface
solar reflected surface
215 219
214 214
400 399
398 398
318 315
314 314
imbalance
solar absorbed surface
evapo-ration
sensible heat
thermal up surface
thermal down surface

◂**Fig. 13** Comparison of different global annual mean energy balance estimates for present-day climate under "all-sky" (upper panel) and "clear-sky" (lower panel) conditions, as simulated in the CMIP6 multi-model mean (upper left (red) values) and in the CMIP5 multi-model mean (upper right (pink) values), and as estimated by Wild et al. (2015, 2019) (lower left (black) values) and Kato et al. (2018) (lower right (green) values). Values attached to arrows correspond to energy fluxes in Wm^{-2} in the direction given by the arrows. Averaging periods for CMIP5 and Wild et al. (2015, 2019): 2000–2004; CMIP6: 2000–2014; Kato et al. (2018): 2005–2015

− 51.2 Wm^{-2} (from 214.6 Wm^{-2} clear-sky absorption to 163.4 Wm^{-2} all-sky absorption). This magnitude falls within the reference estimates given in Table 1. The global mean surface shortwave CRE in the CMIP6 multi-model mean is weaker than in its CMIP5 counterpart (statistically significant, Table 1), due to the fact that the surface clear-sky shortwave absorption is more reduced than the all-sky absorption in the CMIP6 compared to the CMIP5 multi-model mean. Again the spread of the global estimates in the individual CMIP6 models is remarkable, covering a range of 20 Wm^{-2} (Fig. 10, bottom panel).

The effect of clouds on the longwave surface balance is to reduce the surface cooling by 25.5 Wm^{-2} globally in the CMIP6 multi-model mean, nearly matching its CMIP5 counterpart. This effect is somewhat smaller than the reference estimates indicate (Table 1), which are near to the upper bound of the individual model estimates given in Fig. 11 (bottom panel). Both spread and standard deviation in the surface longwave CRE of the CMIP6 models are substantially reduced compared to CMIP5.

As a net effect at the Earth's surface (surface net CRE), the presence of clouds reduces the available energy by − 25.4 Wm^{-2} in the CMIP6 multi-model mean globally, since the energy gain for the surface in the longwave does not compensate the energy loss in the shortwave. The global mean surface net CRE is weaker in the multi-model mean in CMIP6 than in CMIP5 (statistically significant, Table 1), due to the weaker shortwave CRE as discussed above, and comes close to the reference estimate in Wild et al. (2019). The spread of the global mean surface net CRE in the individual CMIP6 models is illustrated in Fig. 12 (bottom panel).

6 Discussion and conclusions

The global energy budget components of up to 40 newly available GCMs participating in CMIP6 have been assessed both under all-sky and clear-sky conditions, covering TOA, surface and atmospheric budgets. On a global multi-model mean basis, the simulated energy balance components in CMIP6 are in the majority close to recent reference estimates, often closer than any preceding model generation, and particularly close in case of the shortwave clear-sky budgets. This is also evident from Fig. 13, which summarizes the CMIP6 and CMIP5 multi-model mean magnitudes of the various global energy balance components in graphical form and compares them with two recent reference estimates. The good agreement of the CMIP6 multi-model means with the reference estimates is not only evident in the TOA components where the reference estimates are commonly used as tuning targets, but increasingly also in other quantities not directly considered in the tuning process (Fig. 13). Note that this does not necessarily apply for the individual CMIP6 models. Despite the tuning efforts applied in model development to match particularly the simulated TOA global mean fluxes with the observational space-based references, 9 (8) CMIP6 models still simulate a global mean shortwave TOA absorption (OLR) outside the 2-sigma observational uncertainty given in Loeb et al. (2009).

In terms of the surface energy budget, a prominent and persistent model bias consisted for many years in a too large shortwave irradiance at the Earth's surface, which was partly compensated by a overly small downward longwave radiation, leading to a superficially correct surface net radiation in the global mean due to this error cancellation, an issue noted already back in the 1990s (Wild et al. 1995). This excessive insolation and compensational lack of downward longwave radiation has not only been found under all-sky conditions, but similarly also under clear-skies (Wild et al. 1995, 2006; Wild 2008). The excessive surface insolation has therefore been related to a lack of absorption in the cloud-free atmosphere in the models. It is interesting to note that the amount of shortwave radiation absorbed within the cloud-free atmosphere under present-day conditions as simulated by climate models has been gradually adjusted upwards from one model generation to the next during the history of GCM development. This is documented in Table 2, which shows the evolution of multi-model global means of shortwave absorption in the cloud-free atmosphere over several generations of GCMs, from early models representing the status in the late 1980s/early 1990s, up to the most recent model generation CMIP6. The model-representation of shortwave absorption in the cloud-free atmosphere increased during this development process on the order of 10 Wm^{-2} (15% of its absolute value), thereby contributing to counteract the excessive surface insolation bias. This upward adjustment brings the shortwave absorption in the cloud-free atmosphere of the CMIP6 multi-model mean now also in close agreement with the recent independently derived reference estimates of Kato et al. (2018) and Wild et al. (2019) of 73 Wm^{-2}, also given in Table 2 and Fig. 13 for comparison. Another independent reference estimate amounts to 72 Wm^{-2} based on a combination of global satellite-derived data sets for aerosols, water vapor and total ozone and a Monte Carlo Aerosol-Cloud-Radiation (MACR) model (Kim and Ramanathan 2008), and thus gives further quantitative support

Table 2 Historic evolution of the quantitative representation of present-day global annual mean shortwave atmospheric absorption under clear-sky conditions in multi-model means of different generations of climate models covering 30 years of model development

Model Generation	# of models	Multi-model mean (Wm^{-2})	References
Pre-AMIP (late 1980s)	7	63	Wild et al. (1998)
AMIPII (1990s)	20	67	Wild et al. (2006)
CMIP3 (early 2000s)	14	69	Wild et al. (2006)
CMIP5 (late 2000s)	43	70	Wild et al. (2019)
CMIP6 (late 2010s)	36	73	This study
Recent reference estimates		73	Wild et al. (2019)
		73	Kato et al. (2018)
		72	Kim and Ramanathan (2008)

For comparison also recent reference estimates are added

Units Wm^{-2}

Table 3 Historic evolution of the quantitative representation of present-day global annual mean downward longwave radiation in multi-model means of different generations of climate models covering 30 years of model development

Model Generation	# Of models	Multi-model mean ($W\ m^{-2}$)	References
Pre-AMIP (late 1980s)	6	327	Wild et al. (1995)
	11	329	Wild et al. (2001)
AMIPII (1990s)	20	336	Wild (2008)
CMIP3 (early 2000s)	20	337	Wild (2008)
CMIP5 (late 2000s)	22	338	Wild et al. (2013)
	43	340	Wild et al. (2015)
CMIP6 (late 2010s)	38	344	This study
Recent reference estimates		342	Wild et al. (2013, 2015)
		342	Wang and Dickinson (2013)
		341	L'Ecuyer et al. (2015)
		344	Kato et al. (2018)

for the magnitudes of the above reference estimates. It is also remarkable that the global mean shortwave absorption in the cloud-free atmosphere simulated by the CMIP6 models is not only close to these recent reference estimates in their multi-model mean, but also in the individual models, most of them deviating less than 2 Wm^{-2} from the reference estimates (see Sect. 4.1). The gradual upward adjustment in the simulated present-day shortwave absorption in the cloud-free atmosphere over the history of model development has been favored by the inclusion of absorbing aerosol in the radiation codes of the models [the early models did only consider sulfur-based scattering aerosols, or did not consider aerosols at all, e.g., Cusack et al. (1998)]. Also, atmospheric water vapor absorption has been underestimated by the early radiation codes, and has increased during the evolution of model development, based on newer assessments of the spectroscopic absorption coefficients and improved formulations of the near-infrared water vapor continuum (Wild et al. 1998; Morcrette 2002; Pincus et al. 2015; Paynter and Ramaswamy 2012; Radel et al. 2015; Paynter and Ramaswamy 2014). This has also been noted in the Continual Intercomparison of Radiation Codes (CIRC, Oreopoulos and Mlawer 2010; Oreopoulos et al. 2012) as well as in preceding radiation code intercomparison projects (Fouquart et al. 1991; Barker et al. 2003). Therein also some missing, yet well-established radiation physics, such as the neglection of N_2O and CH_4 absorption in some of the earlier radiation codes has been identified (Collins et al. 2006), which has been taken into account in the meantime in modern radiation codes.

Another persistent issue in the model-calculated surface energy budgets over the history of GCM model development has been the abovementioned underestimation of downward longwave radiation when compared to surface observations, as we first noted in Wild et al. (1995). Uncertainties in the formulation of the water vapor continuum have been contributing to this underestimation (Iacono et al. 2000; Wild et al. 2001). During the course of model development over the past 30 years, the simulated present-day downward longwave radiation has overall been gradually adjusted upwards from one model generation to the next, as indicated in Table 3. Thereby, considerable progress has been made in reducing these biases during the course of model developments (Ma et al. 2014; Wild et al. 2015, 2019). Note that the early model generations are representative of a slightly earlier period (1980s/1990s) than the one used for CMIP5 and CMIP6 (early 2000s), and thus are expected to have a slightly smaller downward longwave radiation due to the somewhat weaker greenhouse forcing in the earlier period. However, this effect can only account for a minor fraction of the differences in the downward longwave radiation between the different model generations. The multi-model global mean downward longwave radiation in the CMIP6 models, at 343.8 Wm^{-2}, is now in near perfect agreement with recent independent reference estimates, also given in Table 3. Note that the slightly lower reference value given in Wild et al. (2013, 2015), at 342 Wm^{-2}, is derived for the period 2000–2004, which converted to the model analysis period 2000–2014 would be higher by about 0.8 Wm^{-2} due to somewhat stronger greenhouse forcing and warming on average over this period (see Sect. 2), and thus even closer to the CMIP6 multi-model mean.

Therefore, the long-standing tendency in the present-day GCM surface energy budgets to compensate an excessive surface shortwave radiation with a too small downward longwave radiation globally, is now to a large degree remediated in the CMIP6 multi-model mean.

While the global surface radiation budget in the CMIP6 multi-model mean seems now to be quite realistic, and probably more realistic in terms of its multi-model mean than in any preceding model generation, further development work needs to be done by some of the individual modelling groups to converge to this level as well. Indeed the inter-model spread amongst the magnitudes of the global energy balance components in the individual CMIP6 models is still unsatisfactorily large, typically on the order of 10–20 Wm^{-2}. The substantial inter-model spread of 18 Wm^{-2} in the simulated global mean surface latent heat flux further points to considerable discrepancies not only in the representation of the global energy cycle, but also of the global water cycle in the CMIP6 models. All these discrepancies have generally not decreased from the previous model generation CMIP5 to the latest model generation CMIP6, and the inter-model spreads and standard deviations remain similar. Thus, there is no clear sign of convergence in the energy budget estimates of current state-of the art climate models. An exception state the clear-sky shortwave budgets, which are now not only similarly represented in the majority of the CMIP6 models in terms of their global means, but also closely match recent reference estimates.

The substantial discrepancies in the representation of some of the energy balance components between the various CMIP6 models noted here on a global annual mean basis are worrisome as the inter-model spread will undoubtedly further increase on regional, seasonal and diurnal scales. This has major implications for the simulation of regional climates, which cannot be excepted to reach a high degree of consistency amongst the different models under these conditions. Convergence in the representation of the energy budgets by the various models on a global mean basis is therefore a necessary, but not sufficient prerequisite for consistent simulations of regional energy budgets and climates.

Acknowledgements I am grateful to Prof. Atsumu Ohmura for many stimulating discussions on the global energy balance, as well as to Prof. Christoph Schär and Dr. Doris Folini for their continuous support. My research on the global energy balance has been supported by a sequence of Swiss National Science Foundation Grants (Grant No 200021_135395, 200020_159938, 200020_188601) and by funding from the Federal Office of Meteorology and Climatology MeteoSwiss within the framework of GCOS Switzerland. The global energy balance diagrams shown in this paper were designed by Barbara Schär. I acknowledge the World Climate Research Program which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. I thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF. I dedicate this paper to my dear colleague and friend Dr. Chuck Long, who passed away during the writing of this manuscript, and whose lifelong engagement for accurate measurements of the surface radiation budget has been crucial for a better quantification of the global energy balance.

References

Abbot CG, Fowle FE (1908) Radiation and terrestrial temperature. In: Annals of the Astrophysical Observatory of the Smithsonian Institution, vol 2, pp 125–189

Barker HW, Stephens GL, Partain PT, Bergman JW, Bonnel B, Campana K, Clothiaux EE, Clough S, Cusack S, Delamere J, Edwards J, Evans KF, Fouquart Y, Freidenreich S, Galin V, Hou Y, Kato S, Li J, Mlawer E, Morcrette JJ, O'Hirok W, Raisanen P, Ramaswamy V, Ritter B, Rozanov E, Schlesinger M, Shibata K, Sporyshev P, Sun Z, Wendisch M, Wood N, Yang F (2003) Assessing 1d atmospheric solar radiative transfer models: interpretation and handling of unresolved clouds. J Clim 16(16):2676–2699

Barkstrom BR, Harrison EF, Lee RB III (1990) Earth radiation budget experiment. EOS 71:297–305

Bodas-Salcedo A, Ringer MA, Jones A (2008) Evaluation of the surface radiation budget in the atmospheric component of the Hadley Centre global environmental model (HADGEM1). J Clim 21(18):4723–4748. https://doi.org/10.1175/2008jcli2097.1

Cess RD, Potter GL (1987) Exploratory studies of cloud radiative forcing with a general circulation model. Tellus 39A:460–473. https://doi.org/10.1111/j.1600-0870.1987.tb00321.x

Collins WD, Ramaswamy V, Schwarzkopf MD, Sun Y, Portmann RW, Fu Q, Casanova SEB, Dufresne JL, Fillmore DW, Forster PMD, Galin VY, Gohar LK, Ingram WJ, Kratz DP, Lefebvre MP, Li J, Marquet P, Oinas V, Tsushima Y, Uchiyama T, Zhong WY (2006) Radiative forcing by well-mixed greenhouse gases: estimates from climate models in the intergovernmental panel on climate change (IPCC) fourth assessment report (A4). J Geophys Res Atmos. https://doi.org/10.1029/2005jd006713 **(Artn D14317)**

Cusack S, Slingo A, Edwards JM, Wild M (1998) The radiative impact of a simple aerosol climatology on the Hadley Centre atmospheric GCM. Q J R Meteorol Soc 124(551):2517–2526

Dines H (1917) The heat balance of the atmosphere. Quart J Roy Meteorol Soc 43:151–158

Dolinar E, Dong X, Xi B, Jiang J, Su H (2014) Evaluation of CMIP5 simulated clouds and TOA radiation budgets using NASA satellite observations. Clim Dyn. https://doi.org/10.1007/s00382-014-2158-9

Driemel A, Augustine JA, Behrens K, Colle S, Cox C, Cuevas-Agulló E, Denn FM, Duprat T, Fukuda M, Grobe H, Haeffelin M, Hyett N, Ijima O, Kallis A, Knap W, Kustov V, Long CN, Longenecker D, Lupi A, Maturilli M, Mimouni M, Ntsangwane L, Ogihara H, Olano X, Olefs M, Omori M, Passamani L, Pereira EB, Schmithüsen H, Schumacher S, Sieger R, Tamlyn J, Vogt R, Vuilleumier L, Xia X, Ohmura A, König-Langlo G (2018) Baseline Surface Radiation Network (BSRN): Structure and data description (1992–2017). Earth Syst Sci Data. https://doi.org/10.5194/essd-2018-8

Eyring V, Bony S, Meehl GA, Senior CA, Stevens B, Stouffer RJ, Taylor KE (2016) Overview of the Coupled Model Intercomparison Project phase 6 (CMIP6) experimental design and organization. Geosci Model Dev 9(5):1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Fouquart Y, Bonnel B, Ramaswamy V (1991) Intercomparing shortwave radiation codes for climate studies. J Geophys Res Atmos 96(D5):8955–8968

Garratt JR, Prata AJ (1996) Downwelling longwave fluxes at continental surfaces—a comparison of observations with GCM simulations and implications for the global land surface radiation budget. J Clim 9(3):646–655

Gleckler PJ, Weare BC (1997) Uncertainties in global ocean surface heat flux climatologies derived from ship observations. J Clim 10(11):2764–2781

Gutowski WJ, Gutzler DS, Wang WC (1991) Surface-energy balances of 3 general-circulation models—implications for simulating regional climate change. J Clim 4(2):121–134

Hansen J, Nazarenko L, Ruedy R, Sato M, Willis J, Del Genio A, Koch D, Lacis A, Lo K, Menon S, Novakov T, Perlwitz J, Russell G, Schmidt GA, Tausnev N (2005) Earth's energy imbalance: confirmation and implications. Science 308(5727):1431–1435. https://doi.org/10.1126/science.1110252

Hatzianastassiou N, Matsoukas C, Fotiadi A, Pavlakis KG, Drakakis E, Hatzidimitriou D, Vardavas I (2005) Global distribution of Earth's surface shortwave radiation budget. Atmos Chem Phys 5:2847–2867

Hourdin F, Mauritsen T, Gettelman A et al (2017) The art and science of climate model tuning. Bull Am Meteorol Soc 98:589–602. https://doi.org/10.1175/BAMS-D-15-00135.1

Iacono MJ, Mlawer EJ, Clough SA, Morcrette JJ (2000) Impact of an improved longwave radiation model, RRTM, on the energy budget and thermodynamic properties of the NCAR community climate model, CCM3. J Geophys Res Atmos 105(D11):14873–14890

Johnson GC, Lyman JM, Loeb NG (2016) Improving estimates of Earth's energy imbalance. Nat Clim Change 6:639–640. https://doi.org/10.1038/nclimate3043

Kato S, Rose FG, Rutan DA, Thorsen TJ, Loeb NG, Doelling DR, Huang X, Smith WL, Su WY (2018) Surface irradiances of Edition 4.0 Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced And Filled (EBAF) data product. J Clim 31:4501–4527. https://doi.org/10.1175/JCLI-D-17-0523.1

Kiehl JT, Trenberth KE (1997) Earth's annual global mean energy budget. Bull Am Meteorol Soc 78(2):197–208

Kim DY, Ramanathan V (2008) Solar radiation budget and radiative forcing due to aerosols and clouds. J Geophys Res Atmos 113(D2):D02203. https://doi.org/10.1029/2007jd008434

Kopp G, Lean JL (2011) A new, lower value of total solar irradiance: evidence and climate significance. Geophys Res Lett 38:L01706. https://doi.org/10.1029/2010gl04577

L'Ecuyer TS, Beaudoing HK, Rodell M, Olson W, Lin B, Kato S, Clayson CA, Wood E, Sheffield J, Adler R, Huffman G, Bosilovich M, Gu G, Robertson F, Houser PR, Chambers D, Famiglietti JS, Fetzer E, Liu WT, Gao X, Schlosser CA, Clark E, Lettenmaier DP, Hilburn K (2015) The observed state of the energy budget in the early twenty-first century. J Clim 28(21):8319–8346. https://doi.org/10.1175/Jcli-D-14-00556.1

Li ZQ, Moreau L, Arking A (1997) On solar energy disposition: a perspective from observation and modeling. Bull Am Meteorol Soc 78(1):53–70

Li JLF, Waliser DE, Stephens G, Lee S, L'Ecuyer T, Kato S, Loeb N, Ma HY (2013) Characterizing and understanding radiation budget biases in CMIP3/CMIP5 GCMs, contemporary GCM, and reanalysis. J Geophys Res Atmos 118(15):8166–8184. https://doi.org/10.1002/Jgrd.50378

Loeb NG, Wielicki BA, Doelling DR, Smith GL, Keyes DF, Kato S, Manalo-Smith N, Wong T (2009) Toward optimal closure of the Earth's top-of-atmosphere radiation budget. J Clim 22(3):748–766. https://doi.org/10.1175/2008jcli2637.1

Loeb NG, Doelling DR, Wang HL, Su WY, Nguyen C, Corbett JG, Liang LS, Mitrescu C, Rose FG, Kato S (2018) Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced And Filled (EBAF) top-of-atmosphere (TOA) Edition-4.0 data product. J Clim 31(2):895–918. https://doi.org/10.1175/jcli-d-17-0208.1

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

Morcrette JJ (2002) Assessment of the ECMWF model cloudiness and surface radiation fields at the ARM SGP site. Mon Weather Rev 130(2):257–277

Ohmura A, Gilgen H (1993) Reevaluation of the global energy balance. In: McBean GA, Hantel M (eds) Interactions between global climate subsystems—the legacy of hann, vol 75. Geophysical monograph series. American Geophysical Union, Washington, pp 93–110

Ohmura A, Dutton EG, Forgan B, Frohlich C, Gilgen H, Hegner H, Heimo A, Konig-Langlo G, McArthur B, Muller G, Philipona R, Pinker R, Whitlock CH, Dehne K, Wild M (1998) Baseline surface radiation network (BSRN/WCRP): new precision radiometry for climate research. Bull Am Meteorol Soc 79(10):2115–2136. https://doi.org/10.1175/1520-0477(1998)079%3c2115:Bsrnbw%3e2.0.Co;2

Oreopoulos L, Mlawer E (2010) The continual intercomparison of radiation codes (CIRC) assessing anew the quality of GCM radiation algorithms. Bull Am Meteorol Soc 91(3):305–310. https://doi.org/10.1175/2009bams2732.1

Oreopoulos L, Mlawer E, Delamere J, Shippert T, Cole J, Fomin B, Iacono M, Jin ZH, Li JN, Manners J, Raisanen P, Rose F, Zhang YC, Wilson MJ, Rossow WB (2012) The continual intercomparison of radiation codes: results from phase I. J Geophys Res Atmos 117:D06118. https://doi.org/10.1029/2011jd016821

Paynter D, Ramaswamy V (2012) Variations in water vapor continuum radiative transfer with atmospheric conditions. J Geophys

Res Atmos. https://doi.org/10.1029/2012jd017504 **(Artn D16310)**
Paynter D, Ramaswamy V (2014) Investigating the impact of the shortwave water vapor continuum upon climate simulations using GFDL global models. J Geophys Res Atmos 119(18):10720–10737. https://doi.org/10.1002/2014jd021881
Pincus R, Mlawer EJ, Oreopoulos L, Ackerman AS, Baek S, Brath M, Buehler SA, Cady-Pereira KE, Cole JNS, Dufresne JL, Kelley M, Li JN, Manners J, Paynter DJ, Roehrig R, Sekiguchi M, Schwarzkopf DM (2015) Radiative flux and forcing parameterization error in aerosol-free clear skies. Geophys Res Lett 42(13):5485–5492. https://doi.org/10.1002/2015gl064291
Potter GL, Cess RD (2004) Testing the impact of clouds on the radiation budgets of 19 atmospheric general circulation models. J Geophys Res Atmos 109(D2):D02106. https://doi.org/10.1029/2003jd004018
Potter GL, Slingo JM, Morcrette JJ, Corsetti L (1992) A modeling perspective on cloud radiative forcing. J Geophys Res Atmos 97(D18):20507–20518. https://doi.org/10.1029/92jd01909
Radel G, Shine KP, Ptashnik IV (2015) Global radiative and climate effect of the water vapour continuum at visible and near-infrared wavelengths. Q J R Meteorol Soc 141(688):727–738. https://doi.org/10.1002/qj.2385
Randall DA, Cess RD, Blanchet JP, Boer GJ, Dazlich DA, Delgenio AD, Deque M, Dymnikov V, Galin V, Ghan SJ, Lacis AA, Letreut H, Li ZX, Liang XZ, Mcavaney BJ, Meleshko VP, Mitchell JFB, Morcrette JJ, Potter GL, Rikus L, Roeckner E, Royer JF, Schlese U, Sheinin DA, Slingo J, Sokolov AP, Taylor KE, Washington WM, Wetherald RT, Yagai I, Zhang MH (1992) Intercomparison and interpretation of surface-energy fluxes in atmospheric general-circulation models. J Geophys Res Atmos 97(D4):3711–3724
Stephens GL, Li JL, Wild M, Clayson CA, Loeb N, Kato S, L'Ecuyer T, Stackhouse PW, Lebsock M, Andrews T (2012) An update on Earth's energy balance in light of the latest global observations. Nat Geosci 5(10):691–696. https://doi.org/10.1038/Ngeo1580
Tang C, Morel B, Wild M, Pohl B, Abiodun B, Bessafi M (2019) Numerical simulation of surface solar radiation over southern Africa. Part 1: evaluation of regional and global climate models. Clim Dyn 52(1–2):457–477. https://doi.org/10.1007/s00382-018-4143-1
Trenberth KE, Fasullo JT (2010) Simulation of present-day and twenty-first-century energy budgets of the southern oceans. J Clim 23(2):440–454. https://doi.org/10.1175/2009jcli3152.1
Trenberth KE, Fasullo JT, Kiehl J (2009) Earth's global energy budget. Bull Am Meteor Soc 90(3):311. https://doi.org/10.1175/2008bams2634.1
von Schuckmann K, Palmer MD, Trenberth KE, Cazenave A, Chambers D, Champollion N, Hansen J, Josey SA, Loeb N, Mathieu PP, Meyssignac B, Wild M (2016) An imperative to monitor Earth's energy imbalance. Nat Clim Change 6(2):138–144. https://doi.org/10.1038/Nclimate2876
Wang KC, Dickinson RE (2013) Global atmospheric downward longwave radiation at the surface from ground-based observations, satellite retrievals, and reanalyses. Rev Geophys 51(2):150–185. https://doi.org/10.1002/Rog.20009
Wang HL, Su WY (2013) Evaluating and understanding top of the atmosphere cloud radiative effects in Intergovernmental Panel on Climate Change (IPCC) fifth Assessment Report (AR5) Coupled Model Intercomparison Project phase 5 (CMIP5) models using satellite observations. J Geophys Res Atmos 118(2):683–699. https://doi.org/10.1029/2012jd018619
Wielicki BA, Barkstrom BR, Harrison EF, Lee RB, Smith GL, Cooper JE (1996) Clouds and the Earth's Radiant Energy System (CERES): an Earth observing system experiment. Bull Am Meteor Soc 77(5):853–868
Wild M (2008) Short-wave and long-wave surface radiation budgets in GCMs: a review based on the IPCC-AR4/CMIP3 models. Tellus A 60(5):932–945. https://doi.org/10.1111/J.1600-0870.2008.00342.X
Wild M (2017) Towards global estimates of the surface energy budget. Curr Clim Change Rep 3(1):87–97. https://doi.org/10.1007/s40641-017-0058-x
Wild M, Roeckner E (2006) Radiative fluxes in ECHAM5. J. Clim 19:3792–3809
Wild M, Ohmura A, Gilgen H, Roeckner E (1995) Validation of general-circulation model radiative fluxes using surface observations. J Clim 8(5):1309–1324
Wild M, Ohmura A, Gilgen H, Roeckner E, Giorgetta M, Morcrette JJ (1998) The disposition of radiative energy in the global climate system: GCM-calculated versus observational estimates. Clim Dyn 14(12):853–869
Wild M, Ohmura A, Gilgen H, Morcrette JJ, Slingo A (2001) Evaluation of downward longwave radiation in general circulation models. J Clim 14(15):3227–3239
Wild M, Long CN, Ohmura A (2006) Evaluation of clear-sky solar fluxes in GCMs participating in AMIP and IPCC-AR4 from a surface perspective. J Geophys Res Atmos. https://doi.org/10.1029/2005jd006118 **(Artn D01104)**
Wild M, Folini D, Schar C, Loeb N, Dutton EG, Konig-Langlo G (2013) The global energy balance from a surface perspective. Clim Dyn 40(11–12):3107–3134. https://doi.org/10.1007/s00382-012-1569-8
Wild M, Folini D, Hakuba MZ, Schar C, Seneviratne SI, Kato S, Rutan D, Ammann C, Wood EF, Konig-Langlo G (2015) The energy balance over land and oceans: an assessment based on direct observations and CMIP5 climate models. Clim Dyn 44(11–12):3393–3429. https://doi.org/10.1007/s00382-014-2430-z
Wild M, Hakuba MZ, Folini D, Dörig-Ott P, Schär C, Kato S, Long CN (2019) The cloud-free global energy balance and inferred cloud radiative effects: an assessment based on direct observations and climate models. Clim Dyn 52:4787–4812. https://doi.org/10.1007/s00382-018-4413-y