

## Cloud Climatologies from Global Climate Models—A Comparison of CMIP5 and CMIP6 Models with Satellite Data

AXEL LAUER<sup>a</sup>, LISA BOCK<sup>a</sup>, BIRGIT HASSLER<sup>a</sup>, MARC SCHRÖDER<sup>b</sup>, AND MARTIN STENDEL<sup>b</sup>

<sup>a</sup> *Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany*

<sup>b</sup> *Deutscher Wetterdienst (DWD), Offenbach, Germany*

(Manuscript received 17 March 2022, in final form 14 September 2022)

**ABSTRACT:** Simulating clouds with global climate models is challenging as the relevant physics involves many nonlinear processes covering a wide range of spatial and temporal scales. As key components of the hydrological cycle and the climate system, an evaluation of clouds from models used for climate projections is an important prerequisite for assessing the confidence in the results from these models. Here, we compare output from models contributing to phase 6 of the Coupled Model Intercomparison Project (CMIP6) with satellite data and with results from their predecessors (CMIP5). We use multiproduct reference datasets to estimate the observational uncertainties associated with different sensors and with internal variability on a per-pixel basis. Selected cloud properties are also analyzed by region and by dynamical regime and thermodynamic conditions. Our results show that for parameters such as total cloud cover, cloud water path, and cloud radiative effect, the CMIP6 multimodel mean performs slightly better than the CMIP5 ensemble mean in terms of mean bias, pattern correlation, and relative root-mean square deviation. The intermodel spread in CMIP6, however, is not reduced compared to CMIP5. Compared with *CALIPSO-ICECLOUD* data, the CMIP5/6 models overestimate cloud ice, particularly in the lower and middle troposphere, partly due to too high ice fractions for given temperatures. This bias is reduced in the CMIP6 multimodel mean. While many known biases such as an underestimation in cloud cover in stratocumulus regions remain in CMIP6, we find that the CMIP5 problem of too few but too reflective clouds over the Southern Ocean is significantly improved.

**KEYWORDS:** Climate models; Clouds; Coupled models; Model evaluation/performance

### 1. Introduction

In meteorology, clouds are defined as a visible aggregate or aerosol of microscopic water droplets and/or ice particles in the atmosphere above Earth's surface. As a key component of the hydrological cycle, clouds play an important role in weather and climate. By reflecting incoming solar radiation and by absorbing outgoing longwave radiation, clouds have a large impact on the Earth's radiation budget called cloud radiative effect (cre) (e.g., Ramanathan et al. 1989). In addition, clouds play an important role in tropospheric chemistry through their involvement in multiphase reactions such as oxidation of SO<sub>2</sub> to SO<sub>4</sub> and scavenging of atmospheric trace gases and aerosol particles (e.g., Herkes and Collett 2015). Furthermore, cloud–climate feedbacks have important implications for climate sensitivity (Bjorndal et al. 2020; Bony et al. 2015; Zelinka et al. 2020) and thus for the amplitude and pace of future climate change. Despite their pivotal role in weather and climate, clouds are still quite challenging to simulate with global climate models as the relevant physics includes many nonlinear processes on temporal and spatial scales covering several orders of magnitude (Bony et al. 2015). The numerical representation of clouds in global climate models therefore requires a high degree of parameterization, including formation, growth, and sedimentation of water particles. This makes a careful evaluation of clouds in global climate models an important part when assessing the skills of the models in

reproducing observed climate as well as building confidence in projections of future climate change.

In this paper, we document and evaluate the mean cloud properties from models contributing to the latest phase of the Coupled Model Intercomparison Project (CMIP6) with satellite data and compare the results with output from the predecessor phase CMIP5. We focus on the cloud properties cloud fraction, cloud liquid water and cloud ice water, and total cloud water path (tcwp) as well as on the shortwave (swcre) and longwave (lwcre) cloud radiative effects at the top of the atmosphere. Section 2 describes the CMIP model ensembles, the model simulations evaluated, and the satellite data used for comparison with the models. Section 3 describes the methods applied including a brief overview on the Earth System Model Evaluation Tool (ESMValTool) used for the analyses. Results are presented in section 4 including climatologies, interannual variability, seasonal cycle, cloud properties by dynamical regime, and analyses of the cloud regimes in the Southern Ocean, southeastern Pacific, and Pacific intertropical convergence zone (ITCZ). A summary of the findings and conclusions is presented in section 5.

### 2. Models, model simulations, and satellite data

#### a. CMIP models

In this study we investigate the performance of the models participating in the current phase of CMIP (CMIP6; Eyring et al. 2016) by comparing results from the “historical” simulations [in which forcings due to both natural causes such as volcanic eruptions and solar variability and human factors such

Corresponding author: Axel Lauer, axel.lauer@dlr.de

DOI: 10.1175/JCLI-D-22-0181.1

© 2022 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy ([www.ametsoc.org/PUBSReuseLicenses](http://www.ametsoc.org/PUBSReuseLicenses)).

TABLE 1. List of models analyzed. If more than one ensemble member is available, only the first ensemble member “r1i1p1” (CMIP5, historical) or “r1i1p1f1” (CMIP6, historical) is analyzed. Models providing cloud liquid water path instead of total cloud water path are marked with an asterisk (\*); see [section 3c](#).

Model name	Generation	Atmosphere horizontal resolution, vertical levels	References
ACCESS1.0	CMIP5	1.3° × 1.9°, L38	Bi et al. (2013)
ACCESS1.3	CMIP5	1.3° × 1.9°, L38	Bi et al. (2013)
BCC-CSM1.1	CMIP5	2.8° × 2.8°, L26	Wu (2012), Wu et al. (2010)
BCC-CSM1.1(m)	CMIP5	1.1° × 1.1°, L26	Wu (2012), Wu et al. (2010)
BNU-ESM	CMIP5	2.8° × 2.8°, L26	Ji et al. (2014)
CanESM2	CMIP5	2.8° × 2.8°, L35	Arora et al. (2011)
CCSM4*	CMIP5	1.3° × 0.9°, L26	Gent et al. (2011)
CESM1(BGC)	CMIP5	1.3° × 0.9°, L26	Long et al. (2013)
CESM1(CAM5)*	CMIP5	1.3° × 0.9°, L30	Hurrell et al. (2013)
CESM1(FASTCHEM)	CMIP5	1.3° × 0.9°, L27	Cameron-Smith et al. (2006), Hurrell et al. (2013)
CESM1(WACCM)	CMIP5	2.5° × 1.9°, L66	Hurrell et al. (2013)
CMCC-CESM	CMIP5	3.8° × 3.4°, L39	Fogli et al. (2009), Vichi et al. (2011)
CMCC-CM	CMIP5	0.8° × 0.7°, L31	Fogli et al. (2009), Scoccimarro et al. (2011)
CSIRO-Mk3.6.0*	CMIP5	1.9° × 1.9°, L18	Rotstayn et al. (2010)
FGOALS-g2	CMIP5	2.8° × 3°, L26	Li et al. (2013)
FIO-ESM	CMIP5	2.9° × 2.9°, L26	Qiao et al. (2013)
GFDL-CM3	CMIP5	2.5° × 2°, L48	Donner et al. (2011)
GFDL-ESM2G	CMIP5	2.5° × 2°, L24	Dunne et al. (2012)
GFDL-ESM2M	CMIP5	2.5° × 2°, L24	Dunne et al. (2012)
GISS-E2-H-CC	CMIP5	2.5° × 2°, L40	Schmidt et al. (2006)
GISS-E2-H*	CMIP5	2.5° × 2°, L40	Schmidt et al. (2006)
GISS-E2-R-CC	CMIP5	2.5° × 2°, L40	Schmidt et al. (2006)
GISS-E2-R	CMIP5	2.5° × 2°, L40	Schmidt et al. (2006)
HadGEM2-CC	CMIP5	1.9° × 1.3°, L60	Collins et al. (2011)
HadGEM2-ES	CMIP5	1.9° × 1.3°, L38	Collins et al. (2011)
INM-CM4	CMIP5	2° × 1.5°, L21	Volodin et al. (2010)
IPSL-CM5A-LR*	CMIP5	3.8° × 1.9°, L39	Dufresne et al. (2013)
IPSL-CM5A-MR*	CMIP5	2.5° × 1.3°, L39	Dufresne et al. (2013)
IPSL-CM5B-LR*	CMIP5	3.8° × 1.9°, L39	Hourdin et al. (2013)
MIROC4h	CMIP5	0.6° × 0.6°, L56	Sakamoto et al. (2012)
MIROC5	CMIP5	1.4° × 1.4°, L40	Watanabe et al. (2010)
MIROC-ESM-CHEM*	CMIP5	2.8° × 2.8°, L80	Watanabe et al. (2011)
MIROC-ESM*	CMIP5	2.8° × 2.8°, L80	Watanabe et al. (2011)
MPI-ESM-LR*	CMIP5	1.9° × 1.9°, L47	Giorgetta et al. (2013), Stevens et al. (2013)
MPI-ESM-MR*	CMIP5	1.9° × 1.9°, L95	Giorgetta et al. (2013), Stevens et al. (2013)
MPI-ESM-P*	CMIP5	1.9° × 1.9°, L47	Giorgetta et al. (2013), Stevens et al. (2013)
MRI-CGCM3	CMIP5	1.1° × 1.1°, L48	Yukimoto et al. (2012)
MRI-ESM1	CMIP5	1.1° × 1.1°, L48	Adachi et al. (2013)
NorESM1-ME	CMIP5	2.5° × 1.9°, L26	Bentsen et al. (2013)
NorESM1-M	CMIP5	2.5° × 1.9°, L26	Bentsen et al. (2013)
AWI-ESM-1-1-LR*	CMIP6	0.9° × 0.9°, L95	Rackow et al. (2018), Sidorenko et al. (2015)
BCC-CSM2-MR	CMIP6	1.1° × 1.1°, L46	Wu et al. (2019)
CAMS-CSM1-0*	CMIP6	1.1° × 1.1°, L31	Rong et al. (2018)
CESM2-FV2	CMIP6	2.5° × 1.9°, L32	Danabasoglu et al. (2020)
CESM2	CMIP6	1.3° × 0.9°, L32	Danabasoglu et al. (2020)
CESM2-WACCM	CMIP6	1.3° × 0.9°, L32	Danabasoglu et al. (2020), Gettelman et al. (2019)
CNRM-CM6-1-HR	CMIP6	0.3° × 0.3°, L91	Voltaire et al. (2019)
CNRM-CM6-1	CMIP6	1.4° × 1.4°, L91	Voltaire et al. (2019)
CNRM-ESM2-1	CMIP6	1.4° × 1.4°, L91	Séférian et al. (2019)
FGOALS-g3	CMIP6	2° × 2.3°, L26	Li et al. (2020)
GFDL-CM4	CMIP6	1° × 1°, L33	Held et al. (2019)
GFDL-ESM4	CMIP6	1° × 1°, L49	Dunne et al. (2020)
GISS-E2-1-G	CMIP6	2.5° × 2.5°, L40	Rind et al. (2020)
GISS-E2-1-H	CMIP6	2.5° × 2.5°, L40	Rind et al. (2020)
HadGEM3-GC31-LL	CMIP6	1.9° × 1.3°, L85	Kuhlbrodt et al. (2018)
ISPL-CM6A-LR*	CMIP6	2.5° × 1.3°, L79	Boucher et al. (2020)

TABLE 1. (Continued)

Model name	Generation	Atmosphere horizontal resolution, vertical levels	References
KACE-1-0-G	CMIP6	$1.9^\circ \times 1.3^\circ$ , L85	Lee et al. (2020a)
MIROC6	CMIP6	$1.4^\circ \times 1.4^\circ$ , L81	Tatebe et al. (2019)
MIROC-ES2L	CMIP6	$2.8^\circ \times 2.8^\circ$ , L40	Hajima et al. (2020)
MPI-ESM1-2-HAM*	CMIP6	$1.9^\circ \times 1.9^\circ$ , L47	Mauritsen et al. (2019)
MPI-ESM1-2-HR*	CMIP6	$0.9^\circ \times 0.9^\circ$ , L95	Müller et al. (2018)
MPI-ESM1-2-LR*	CMIP6	$1.9^\circ \times 1.9^\circ$ , L47	Mauritsen et al. (2019)
MRI-ESM2-0	CMIP6	$1.1^\circ \times 1.1^\circ$ , L80	Yukimoto et al. (2019)
NESM3	CMIP6	$1.9^\circ \times 1.9^\circ$ , L47	Cao et al. (2018)
NorESM2-LM	CMIP6	$2.5^\circ \times 1.9^\circ$ , L32	Seland et al. (2020)
SAM0-UNICON*	CMIP6	$1.3^\circ \times 0.9^\circ$ , L30	Park et al. (2019)
TaiESM1	CMIP6	$1.3^\circ \times 0.9^\circ$ , L30	Lee et al. (2020b)
UKESM1-0-LL	CMIP6	$1.9^\circ \times 1.3^\circ$ , L85	Sellar et al. (2019)

as CO<sub>2</sub> concentration (e.g., fossil fuel burning), aerosols, and land use were included over the time period 1850–2014] with the corresponding CMIP5 (Taylor et al. 2012) model runs (historical simulations for 1850–2005 conducted with the best record of natural and anthropogenic climate forcing) and satellite data. We used one ensemble member (typically the first member) per model. The 20-yr means are calculated by using the last 20 years of the historical simulation (i.e., 1986–2005 for CMIP5 and 1995–2014 for CMIP6). While this choice of model years is somewhat arbitrary, we found that it has very little impact on the multiyear ensemble averages. This is not surprising as ESMs are not expected to reproduce the exact observed phase of climate modes largely controlling present-day variability of clouds but rather their statistical properties.

A particular focus of this work is to assess how much progress has been made over the recent years by the model ensembles as a whole. This is done by evaluating multimodel means from the CMIP5 and CMIP6 model ensembles rather than trying to trace individual models as a number of new models contribute to CMIP6 that have not been part of CMIP5 and because some individual models changed so substantially that it would be difficult to compare them with older model versions. Table 1 gives an overview of the 35 CMIP5 and 28 CMIP6 models analyzed here.

### b. Satellite data

In the following, each satellite and reanalysis dataset that is used in the comparison with the model simulations is briefly described. An overview of each of the datasets is provided in Table 2 (for variable definitions see Table 3).

#### 1) CALIPSO

The *Cloud–Aerosol Lidar and Infrared Pathfinder Satellite Observation* (CALIPSO) combines an active lidar instrument (CALIOP) with passive infrared and visible imagers to obtain global vertically resolved properties of thin clouds and aerosols. CALIPSO and *CloudSat* (see below) are highly complementary and part of the A-Train satellite constellation consisting of several satellites flying in formation (Stephens et al. 2018). Here, we use the GCM-Oriented CALIPSO Cloud Product (CALIPSO-GOCCP; Chepfer et al. 2010) providing 3D observations of

the cloud fraction and the CALIPSO version 1.00 lidar Level 3 Ice Cloud Data Product (CALIPSO-ICECLOUD; NASA/LARC/SD/ASDC 2018) providing histograms of the 3D cloud ice content. To calculate the grid box average ice water content for comparison with the models, we follow the procedure described in the CALIPSO-ICECLOUD Data Quality Summary available at [https://www-calipso.larc.nasa.gov/resources/calipso\\_users\\_guide/qs/cal\\_lid\\_l3\\_ice\\_cloud\\_v1-00.php](https://www-calipso.larc.nasa.gov/resources/calipso_users_guide/qs/cal_lid_l3_ice_cloud_v1-00.php) (last access 8 July 2021). For more details, we refer to the CALIPSO data user's guide and the references therein available at [https://www-calipso.larc.nasa.gov/resources/calipso\\_users\\_guide/data\\_quality/level\\_3\\_icecloud\\_summary\\_v100.php](https://www-calipso.larc.nasa.gov/resources/calipso_users_guide/data_quality/level_3_icecloud_summary_v100.php) (last access 8 July 2021).

#### 2) CERES-EBAF

The Clouds and Earth's Radiant Energy Systems (CERES) Energy Balanced and Filled (EBAF) dataset provides global monthly-mean top-of-atmosphere (TOA) and surface longwave (LW), shortwave (SW), and net radiative fluxes under clear and all-sky conditions. CERES instruments are flown on NASA's *Terra* and *Aqua* satellites. For the generation of the CERES-EBAF dataset, an objective constraint algorithm is used in order to adjust SW and LW TOA fluxes within their ranges of uncertainty to remove the inconsistency between average global net TOA flux and the observed ocean heating rate. In addition, gaps in the CERES clear-sky TOA flux maps are filled using clear-sky fluxes from the Moderate Resolution Imaging Spectrometer (MODIS) (Loeb et al. 2009, 2012). Here, we use monthly mean CERES-EBAF Ed 4.1 data that are provided on a  $1^\circ \times 1^\circ$  grid (NASA/LARC/SD/ASDC 2019).

#### 3) CLARA-A2.1

The satellite-derived climate data record CLARA (CM SAF Cloud, Albedo and Radiation dataset from AVHRR data) is based on data from the Advanced Very High Resolution Radiometer (AVHRR) and covers the time period 1982–2018. In this study, we use monthly data from edition 2.1 (Karlsson et al. 2020), which is an extended version of CLARA-A2 (Karlsson et al. 2017), for the cloud liquid water

TABLE 2. Observationally based datasets used in the comparison with model simulations. For variable definitions, see Table 3; for details on the estimated uncertainty range (calculated if two or more datasets for are available for the variable), see section 3d.

Dataset	Type	Variable(s)	Resolution	Years	Estimated average uncertainty	Reference
CALIPSO-GOCCP	Satellite	clcalipso	$2^\circ \times 2^\circ \times L40$	2007–15	—	Chepfer et al. (2010)
CALIPSO-ICECLOUD	Satellite	cli	$2.5^\circ \times 2.0^\circ \times L27$	2007–15	—	NASA/LARC/SD/ASDC (2018)
CERES-EBAF	Satellite	lwcre swcre	$1^\circ \times 1^\circ$	2001–21	$2 \text{ W m}^{-2}$ $5 \text{ W m}^{-2}$	Loeb et al. (2009), Loeb et al. (2012), NASA/LARC/SD/ASDC (2019)
CLARA-AVHRR	Satellite	clivi clt	$0.25^\circ \times 0.25^\circ$	1982–2018	$0.03 \text{ kg m}^{-2}$ 4%	Karlsson et al. (2020), Karlsson et al. (2017)
CloudSat-L2	Satellite	lwp clivi clw	$2^\circ \times 2^\circ \times L40$	2006–17	$0.04 \text{ kg m}^{-2}$ $0.02 \text{ kg m}^{-2}$	Stephens et al. (2018), Stephens et al. (2002)
ERA5	Reanalysis	lwp ta wap cl clivi clt lwp	$0.25^\circ \times 0.25^\circ$ , L37	1979–2019	— — — $0.03 \text{ kg m}^{-2}$ 8% $0.03 \text{ kg m}^{-2}$	Hersbach et al. (2019)
ERA-Interim	Reanalysis	clivi clt lwp prw	$0.75^\circ \times 0.75^\circ$ , L19	1979–2018	$0.03 \text{ kg m}^{-2}$ 9% $0.04 \text{ kg m}^{-2}$ $1.7 \text{ kg m}^{-2}$	Dee et al. (2011)
ESACCI-CLOUD	Satellite	clivi clt lwp lwcre swcre	$0.5^\circ \times 0.5^\circ$	1992–2016 (1982–2016)	$0.01 \text{ kg m}^{-2}$ 3% $0.03 \text{ kg m}^{-2}$ $3 \text{ W m}^{-2}$ $7 \text{ W m}^{-2}$	Stengel et al. (2020)
CM SAF / CCI TCWV-global (COMBI)	Satellite	prw	$0.5^\circ \times 0.5^\circ$	2003–17	$2.0 \text{ kg m}^{-2}$	Beta version of the combined microwave and near-infrared imager based TCWV data record (COMBI), official release is planned in 2022 by CM SAF, for questions contact <a href="mailto:contact.cmsaf@dwd.de">contact.cmsaf@dwd.de</a>
ISCCP-FH	Satellite	prw lwcre swcre	$1^\circ \times 1^\circ$	1984–2016	$2.0 \text{ kg m}^{-2}$ $2 \text{ W m}^{-2}$ $4 \text{ W m}^{-2}$	Rossow and NOAA CDR Program (2017), Young et al. (2018)
MAC-LWP	Satellite	lwp	$1^\circ \times 1^\circ$	1988–2016	$0.03 \text{ kg m}^{-2}$	Elsaesser et al. (2017)
MODIS	Satellite	clivi clt lwp	$1^\circ \times 1^\circ$	2003–18	$0.02 \text{ kg m}^{-2}$ 5% $0.02 \text{ kg m}^{-2}$	Platnick et al. (2003)
PATMOS-x	Satellite	clt	$0.1^\circ \times 0.1^\circ$	1982–2016	5%	Heidinger et al. (2014)

and ice water path as well as for total cloud fraction provided on a global  $0.25^\circ \times 0.25^\circ$  grid. The data and corresponding documentation are available at [https://doi.org/10.5676/EUM\\_SAF\\_CM/CLARA\\_AVHRR/V002\\_01](https://doi.org/10.5676/EUM_SAF_CM/CLARA_AVHRR/V002_01).

#### 4) CLOUDSAT

CloudSat uses a cloud profiling radar to provide retrievals of global vertically resolved cloud liquid water content from space (Stephens et al. 2002). In this study, we use 3D cloud

liquid water content data from individual CloudSat orbits (level 2 data: 2B\_CWC\_RO “radar only”) that are binned onto a  $2^\circ \times 2^\circ$  grid with 40 vertical levels. Individual orbits and days are averaged to monthly means. CloudSat is sensitive to both cloud and precipitation liquid and ice particles. Estimates of cloud liquid water from CloudSat are therefore expected to overestimate the actual cloud liquid water content (e.g., Waliser et al. 2009). Similar to Li et al. (2018), cases when precipitation is present are therefore not included in



TABLE 3. Variables (CMOR names) and derived variables used. Derived variables are marked with an asterisk (\*).

Variable	Name	Units	Comment
cl	3D cloud area fraction	%	Interpolated to pressure levels
clcalipso	3D cloud area fraction as seen from <i>CALIPSO</i>	%	Output from the satellite simulation software for model assessment (COSM; <a href="#">Bodas-Salcedo et al. 2011</a> )
cli	3D cloud ice mass fraction	kg kg <sup>-1</sup>	Interpolated to pressure levels
clivi	Cloud ice water path (iwp)	kg m <sup>-2</sup>	The more common acronym iwp is used in the text synonymously with CMOR variable clivi
clt	Total cloud area fraction	%	Also called total cloud cover
clw	3D cloud liquid water mass fraction	kg kg <sup>-1</sup>	Interpolated to pressure levels
clwvi	Cloud total water path	kg m <sup>-2</sup>	Mass of condensed water (liquid + ice) per area, same as tcwp
iwp	Ice water path	kg m <sup>-2</sup>	Same as clivi
lwcre*	TOA longwave cloud radiative effect	W m <sup>-2</sup>	Calculated as lwcre = rlutcs - rlut
lwp*	Cloud liquid water path	kg m <sup>-2</sup>	Calculated as lwp = clwvi - clivi (see <a href="#">section 3b</a> )
prw	Water vapor path (wvp)	kg m <sup>-2</sup>	The more common acronym wvp is used in the text synonymously with CMOR variable prw
rlut	TOA outgoing longwave radiation	W m <sup>-2</sup>	At the top of the atmosphere
rlutcs	TOA outgoing clear-sky longwave radiation	W m <sup>-2</sup>	At the top of the atmosphere
rsut	TOA outgoing shortwave radiation	W m <sup>-2</sup>	At the top of the atmosphere
rsutcs	TOA outgoing clear-sky shortwave radiation	W m <sup>-2</sup>	At the top of the atmosphere
swcre*	TOA shortwave cloud radiative effect	W m <sup>-2</sup>	Calculated as swcre = rsutcs - rsut
ta	Air temperature	K	
tcwp*	Total cloud water path	kg m <sup>-2</sup>	Mass of condensed water (liquid + ice) per area, calculated as tcwp = lwp + iwp (see <a href="#">section 3b</a> )
ts	Surface temperature	K	“Skin” temperature (i.e., SST for open ocean)
wap	Lagrangian tendency of air pressure	Pa s <sup>-1</sup>	
wvp	Water vapor path	kg m <sup>-2</sup>	Same as prw

this analysis. Precipitating pixels are identified using the *CloudSat* level 2 data for column precipitation (2C\_PRECIP\_COLUMN). For comparison with the CMIP models, *CloudSat* data are converted from kg m<sup>-3</sup> to kg kg<sup>-1</sup> using the air density from the U.S. Standard Atmosphere 1976. In addition, the in-cloud values from *CloudSat* are converted to the grid box averages provided by the models assuming that missing values in the *CloudSat* orbits are actually cloud-free (i.e., zero cloud liquid water). For more details on *CloudSat*, refer to [Stephens et al. \(2002\)](#) and [Stephens et al. \(2018\)](#).

### 5) ESA CCI

For this study we use two datasets from the European Space Agency’s Climate Change Initiative (ESA CCI; e.g., [Hollmann et al. 2013](#)), ESACCI-CLOUD and CM SAF/CCI TCWV-global (COMBI).

The version 3 cloud datasets of ESA CCI CLOUD are based on data from the passive imager satellite sensors AVHRR and (A)ATSR, and are named AVHRR-AMv3, AVHRR-PMv3 ([Stengel et al. 2020](#)), and ATSR2-AATSRv3 ([Poulsen et al. 2020](#)). In this study, Level 3C data from AVHRR-AMv3 and AVHRR-PMv3 are averaged to include both AVHRR morning (AM) and afternoon (PM) satellite measurements, resulting in a combined dataset with monthly mean data at a horizontal resolution of 0.5° × 0.5° for the overlapping time period (1992–2016) of both source datasets.

Variables used in this study are total cloud cover, liquid water path, and ice water path. Known limitations of the AVHRR data include an underrepresentation of optically very thin clouds and sparse temporal sampling as well as satellite drift affecting in particular data before the year 2001 ([Ji and Brown 2017](#)). For details, refer to [Stengel et al. \(2020\)](#).

The dataset CM SAF/CCI TCWV-global (COMBI) combines microwave and near-infrared imager based TCWV over the ice-free ocean as well as over land, coastal ocean, and sea ice. COMBI is available on a global grid with a spatial resolution of 0.5° × 0.5° and 0.05° × 0.05° and a daily and monthly temporal resolution (here we use the 0.5° and monthly resolutions). The data record relies on microwave observations from SSM/I, SSMIS ([Fennig et al. 2020](#)), AMSR-E, and TMI and on near-infrared observations from the Medium Resolution Imaging Spectrometer (MERIS; third reprocessing), MODIS-Terra (collection 6.1), and the Ocean and Land Color Instrument (OLCI; first reprocessing). Details of the retrieval for microwave imagers can be found in [Andersson et al. \(2010\)](#) and [Graw et al. \(2017\)](#) and for near-infrared imagers in [Diedrich et al. \(2015\)](#), [Fischer et al. \(2021\)](#), and [Lindstrot et al. \(2012\)](#).

### 6) ISCCP-FH

From the third-generation version of the International Satellite Cloud Climatology Project (ISCCP) we use TOA

radiative fluxes (cloud radiative effects) and water vapor path data for this study (ISCCP-FH; Rossow and NOAA CDR Program 2017). The ISCCP H-series products use data derived from geostationary and polar-orbiting satellite imaging radiometers with common visible and infrared channels. They supersede their predecessors from the D-series. Key improvements include, for instance, an improved quality control, higher spatial resolution input and output products, and updated documentation and metadata (Young et al. 2018). The ISCCP-FH data used here are provided as monthly means on a  $1^\circ \times 1^\circ$  grid. We would like to note that variable wvp included in the ISCCP-FH dataset is one of the “sub-product parameters” and originally comes from ISCCP-HGG.

### 7) MAC-LWP

The Multisensor Advanced Climatology of Liquid Water Path (MAC-LWP; Elsaesser et al. 2017) is an updated and enhanced version of the University of Wisconsin (UWisc; O'Dell et al. 2008) cloud liquid water path climatology. The data consist of monthly mean liquid water path over the ocean provided on a  $1^\circ \times 1^\circ$  grid covering the time period 1988–2016. MAC-LWP uses input data from SSM/I, TMI, AMSR-E, *WindSat*, SSMIS, AMSR-2, and GMI. A known limitation of this dataset is the increased uncertainty in regions with substantial precipitation as the microwave emission signal from cloud water is similar to that of precipitation. In contrast to many other climatological LWP datasets, MAC-LWP explicitly accounts for the diurnal cycle of cloud liquid water, despite being limited to ice-free ocean regions only.

### 8) MODIS

Here, we use data from the Moderate Resolution Imaging Spectroradiometer (MODIS) L3 Atmosphere Product, Collection 6.1 covering the time period 2003–18. For details on the instrument, retrievals, and derived cloud products, we refer to Platnick et al. (2003), Platnick et al. (2017), and Marchant et al. (2016), for the atmosphere L3 monthly product to Platnick et al. (2015). In this study, we use the variables clivi, clt, and lwp.

### 9) PATMOS-x

The Advanced Very High Resolution Radiometer (AVHRR) Pathfinder Atmospheres – Extended (PATMOS-x) dataset is a satellite-based long-term record providing properties of tropospheric clouds and aerosols, Earth's surface, Earth's radiation budget, and relevant ancillary variables (Heidinger et al. 2014). PATMOS-x is composed of data from 17 different sensors flown on a number of polar-orbiting NOAA and MetOp satellites. Here, we use data for clt covering the time period 1982–2016 on a  $0.1^\circ \times 0.1^\circ$  grid. For details on the dataset, processing, and limitations, we refer to Heidinger et al. (2014).

### c. Reanalysis data

The two reanalysis datasets used in this study are briefly described in the following. A summary is provided in Table 2.

### 1) ERA5

The European Centre for Medium-Range Weather Forecasts (ECMWF) fifth-generation reanalysis ERA5 is the successor of the widely used ERA-Interim reanalysis (Dee et al. 2011). ERA5 is based on four-dimensional variational (4D-Var) data assimilation and uses Cycle 41r2 of the Integrated Forecasting System (IFS) (Copernicus Climate Change Service 2017). In this study, ERA5 data served on the Copernicus Climate Change Service Climate Data Store (CDS) (Copernicus Climate Change Service 2017) are used. Atmospheric variables are provided on the CDS interpolated to 37 pressure levels ranging from 1000 hPa near the surface to 1 hPa (ECMWF 2020). The ERA5 datasets used here have a horizontal resolution of  $0.25^\circ \times 0.25^\circ$  grid. For more details on the ERA5 dataset we refer to Hersbach et al. (2020). Here, we use monthly means of the variables ta, wap, cl, clivi, clt, and lwp.

### 2) ERA-INTERIM

The reanalysis ERA-Interim (Dee et al. 2011) is a global dataset providing a multitude of atmospheric variables over the time period January 1979 through August 2019. ERA-Interim was produced by ECMWF and has been widely used in climate studies. It has been superseded by the ERA5 reanalysis. Similar to ERA5, ERA-Interim uses 4D-Var data assimilation but is based on the older Cycle 31r2 of the ECMWF IFS. The dataset has a spatial resolution of  $0.75^\circ \times 0.75^\circ$  and is provided on 60 vertical levels from the surface up to 0.1 hPa. For more details, please refer to Dee et al. (2011) or, for example, <https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim> (last access: 1 September 2021).

## 3. Methods

### a. Earth system model evaluation tool

All analyses in this study are made with the open-source community diagnostics and performance metrics tool for evaluation of Earth system models “Earth System Model Evaluation Tool” (ESMValTool; Eyring et al. 2020; Lauer et al. 2020; Righi et al. 2020; Weigel et al. 2021). The ESMValTool allows for consistent processing of all datasets (e.g., re-gridding to common grids, masking of land/sea and missing values, vertical interpolation, etc.) and traceability and reproducibility of the results by providing provenance records for all results. With the ESMValTool “recipes” (i.e., configuration files defining input data, preprocessing steps, and diagnostics to be applied) recipe\_lauer22jclim\_fig\*.yaml, figures shown in this paper can be reproduced. For use with the ESMValTool, the variables and metadata of the input datasets have to be formatted following the CMOR (Climate Model Output Rewriter; Taylor et al. 2006) tables and definitions [e.g., <https://github.com/PCMDI/cmip6-cmor-tables/tree/master/Tables> (PCMDI 2019) for CMIP6]. While the CMIP6 and CMIP6 model output follow this standard, some observational data have to be reformatted according to this standard. For this, the ESMValTool contains a set of scripts that

are provided with the ESMValTool source code and are publicly available on GitHub (<https://github.com/ESMValGroup/ESMValTool>). These scripts give exact downloading and processing instructions to create the observational datasets used in this publication.

### b. Regridding, masking, and vertical interpolation

To average different datasets, calculate differences or spatial correlations with a reference dataset, data have been linearly regridded to a regular  $2^\circ \times 2^\circ$  latitude–longitude grid using the regridding function of the ESMValTool preprocessor (Righi et al. 2020). Application of land/sea masks and area selection were also performed with the ESMValTool preprocessor. Vertically resolved variables such as cloud liquid water or cloud ice water content from the CMIP models are typically provided on model specific hybrid levels. To calculate differences, data have been vertically interpolated to the pressure levels defined by the CMIP6 coordinate “plev27” (27 pressure levels from 1000 to 100 hPa) or the height levels defined by “alt40” (40 height levels up to  $\sim 19$  km) (e.g., <https://github.com/PCMDI/cmip6-cmor-tables/tree/master/Tables>, last access 7 November 2020) depending on the variable.

### c. Calculation of liquid water path and total water path

Cloud liquid water path (lwp) from the CMIP models is calculated by subtracting the ice water path (clwi) from the vertically integrated total cloud water path (clwvi = ice + liquid). Some models erroneously provide lwp only instead of total cloud water path. Subtracting clwi from clwvi would then result in significant negative and thus unphysical values. Most of the affected CMIP5 models are listed on the CMIP5 errata page (<https://pcmdi.llnl.gov/mips/cmip5/errata.html>; last access on 8 October 2021). Similarly, some CMIP6 models are affected by the same issue. All CMIP5 and CMIP6 models affected and used in this study are marked in Table 1 with an asterisk. For these models we presume that variable clwvi provides liquid water path data only.

For the same reason, we do not use variable clwvi directly as total cloud water path, but calculate total cloud water path (tcwp) for all models as  $tcwp = lwp + clwi$ .

### d. Multiobservational datasets and estimation of observational uncertainties

For evaluation of the climatologies from the models, multiobservational means  $\bar{X}$  are calculated as average over the time ( $t$ ) means  $\bar{x}_n = (1/T_n) \sum_{t=1}^{T_n} x_{t,n}$  of all  $N_{\text{obs}}$  observational datasets ( $\bar{x}_n$ ) whenever two or more observational datasets are available for a given variable  $x$ :

$$\bar{X} = \frac{1}{N_{\text{obs}}} \sum_{n=1}^{N_{\text{obs}}} \bar{x}_n. \quad (1)$$

Here, all observational datasets are given the same weight, independent of their record length. In most cases, the record length used is the maximum record available (see Table 2) (i.e., of similar length or longer than for the models) but is

TABLE 4. Observational datasets used to calculate the multiobservational datasets per variable. Reanalysis datasets are not included when calculating multiobs products.

Variable	Observational datasets
clt	CLARA-AVHRR, ESACCI-CLOUD, MODIS, PATMOS-x
iwp	CLARA-AVHRR, <i>CloudSat</i> -L2, ESACCI-CLOUD, MODIS
lwcre	CERES-EBAF, ESACCI-CLOUD, ISCCP-FH
lwp	CLARA-AVHRR, <i>CloudSat</i> -L2, ESACCI-CLOUD, MAC-LWP, MODIS
swcre	CERES-EBAF, ESACCI-CLOUD, ISCCP-FH
wvp	ESACCI-WATERVAPOUR, ISCCP-FH

allowed to be shorter if fewer years of data are available. This allows for estimating the observational uncertainty on a per-pixel basis by calculating the average standard deviation of individual annual means against the multiobservational climatology. The standard deviation  $\sigma_n$  of dataset  $n$  is calculated over the individual years ( $t$ ) and using the multiobservational mean  $\bar{X}$ :

$$\sigma_n = \sqrt{\frac{1}{T_n - 1} \sum_{t=1}^{T_n} (\bar{X} - x_{t,n})^2}. \quad (2)$$

This standard deviation is then averaged over all observational datasets using the same weight for each dataset:

$$\bar{\sigma} = \sqrt{\frac{1}{N_{\text{obs}}} \sum_{n=1}^{N_{\text{obs}}} \sigma_n^2}. \quad (3)$$

Constructed this way,  $\bar{\sigma}$  includes estimates for the uncertainty introduced by natural interannual variability as well as deviations of the individual observational datasets from the multiobservational mean. Considering interannual variability is important in this context as individual years from ESMs do not necessarily correspond to the same observed years as relevant climate modes are not necessarily in the same phase in the models and in the real world.

In the figures presented here, areas are stippled if the absolute values of the differences between the model climatology and the multiobservational mean do not exceed  $\bar{\sigma}$ .

The estimates of the global average uncertainty of the observational datasets given in Table 2 are calculated as the square root of the average of the area-weighted squared differences between the values of the respective dataset and the multiobservational mean climatology. The observational datasets used to calculate a multiobservational dataset for a given variable are listed in Table 4.

### e. Temporal variability and seasonal cycle amplitude

The temporal (interannual) variability at each grid cell is estimated as the temporal standard deviation of the deseasonalized monthly means ( $\sigma_{\text{temp}}$ ). The monthly mean values of a

variable ( $x$ ) are normalized by the average over the entire time series ( $\bar{x}$ ). The (relative) temporal variability of the multimodel ensembles and of the multiobservational reference ( $\bar{\sigma}_{\text{temp}}$ ) is calculated by averaging over all  $N$  individual model or observational datasets:

$$\bar{\sigma}_{\text{temp}} = \frac{1}{N} \sum_{i=1}^N \frac{\sigma_{\text{temp},i}}{\bar{x}_i} \times 100\%. \quad (4)$$

The (relative) seasonal cycle amplitude ( $A_{\text{seas}}$ ) is calculated as difference between the minimum ( $x_{\text{min}}$ ) and maximum value ( $x_{\text{max}}$ ) of the average seasonal cycle normalized by the climatological annual mean value ( $\bar{x}$ ). As for the interannual variability, the seasonal cycle amplitude of the multimodel ensembles and of the multiobservational reference ( $\bar{A}_{\text{seas}}$ ) is calculated by averaging over all  $N$  individual model or observational datasets:

$$\bar{A}_{\text{seas}} = \frac{1}{N} \sum_{i=1}^N \left| \frac{x_{\text{max},i} - x_{\text{min},i}}{\bar{x}_i} \right| \times 100\%. \quad (5)$$

#### 4. Comparison of CMIP6 with CMIP5 and satellite data

Modeled clouds and satellite observations are usually difficult to compare because observations are affected by the satellite instrument's sensitivity, the temporal and spatial sampling, and the vertical overlap of the cloud layers, while the clouds in climate models are assumed to be plane-parallel and are of coarse horizontal and vertical resolution. Ideally, a satellite simulator such as COSP [Cloud Feedback Model Intercomparison Project (CFMIP) Observation Simulator Package; [Bodas-Salcedo et al. 2011](#)] is used during the model simulation to mimic the satellite viewing geometry, temporal sampling, and specific instrument characteristics such as cut-off values. Many CMIP5 and CMIP6 historical simulations, however, have been run without such a satellite simulator, or variables of interest such as cloud water path or 3D cloud liquid and ice water content are not available. Alongside with the evaluation of satellite simulator output of 3D cloud fraction [section 4a(3)] we also use multiproduct observational datasets (see section 3d) that provide a per-pixel estimation of observational uncertainties by different sensors and internal variability when output from satellite simulators was not available. Using multiple independent datasets for comparison with the models is considered one possibility to at least partly reduce limitations of observational datasets caused by observational uncertainties and internal variability when no other information is available (e.g., [Flato et al. 2013](#)). This method is, however, not able to reduce sampling biases introduced when only datasets from polar-orbiting satellites are available (twice daily sampling). If no satellite simulators are available for the models, the only way to reduce this kind of sampling bias would be a correction applied to the satellite observations by the data providers as done, for example, for the sampling-bias-corrected AIRS Obs4MIPs V2.1 dataset ([Tian and Hearty 2020](#)).

#### a. Multiyear annual mean climatologies

##### 1) GEOGRAPHICAL DISTRIBUTION

[Figure 1](#) shows maps of the multiyear annual mean climatologies of total cloud cover (clt), liquid (lwp) and ice water path (iwp), longwave (lwcre) and shortwave cloud radiative effect (swcre), and the water vapor path (wvp) for the reference dataset (multiobservational mean, herein called the "multiobs mean") and the differences between the two CMIP ensembles and the reference. This allows for a comparison of typical regional features in the geographical distribution of these parameters. [Table 5](#) summarizes global values of mean, bias, pattern correlation, and root-mean square deviation for all mentioned parameters.

The observed distribution of clt shows high values in the midlatitude storm track regions associated with frontal systems, the intertropical convergence zone (ITCZ) associated with frequent and strong convection, and the stratocumulus decks off the west coasts of the continents in the subtropical subsidence regions. The CMIP6 models underestimate total cloud cover over the Southern Ocean and in stratocumulus regions by about 5% less than the CMIP5 models. Overestimation of clt north and south of the equator has increased by about 5% in CMIP6 compared with CMIP5 to up to 15%. The overestimation in clt south of the equator is also seen in other quantities such as lwp or lwcre. This double-ITCZ bias has found to be only slightly reduced from CMIP5 to CMIP6 ([Tian and Dong 2020](#)). There is also a stronger overestimation in high latitudes of up to more than 30% in CMIP6, which has also been reported for single models (e.g., Nor-ESM; [Seland et al. 2020](#)), and which is not seen in CMIP5 to the same extent. However, many satellite products have shortcomings in detecting optically thin clouds (e.g., [Karlsson et al. 2017](#)). [Stengel et al. \(2018\)](#) found that removing modeled clouds with optical thicknesses below 1 can reduce the cloud fraction by several tens of percent, which indicates that clt in at least some observational datasets used here can be assumed to be significantly biased low in the polar regions. This makes the comparison between observations and models challenging in these regions even though the differences between the multimodel mean (MMM) and the multiobs mean are larger than the estimated  $\pm 1\sigma$  uncertainty in these regions.

Cloud liquid water path measurements from satellites show large uncertainties resulting in a large spread among different datasets. On global annual average we find about 20% less lwp in CMIP6 than in CMIP5, which is in better agreement with the multiobs mean. Most differences of the MMMs from the multiobs mean are, however, below the uncertainty estimate of  $\pm 1$  sigma of the individual observed time series to the multiobs mean [Eq. (3)] except in the tropical regions near or within the ITCZ and in some polar regions. The pattern correlation (Pearson product-moment coefficient of linear correlation between two fields) of the MMM with the multiobs reference dataset increased from 0.71 in CMIP5 to 0.76 in CMIP6. The bias of the modeled liquid water path toward low values at low latitudes and high values in the extratropics has also been reported by single model results (e.g., Nor-ESM; [Seland et al. 2020](#)).



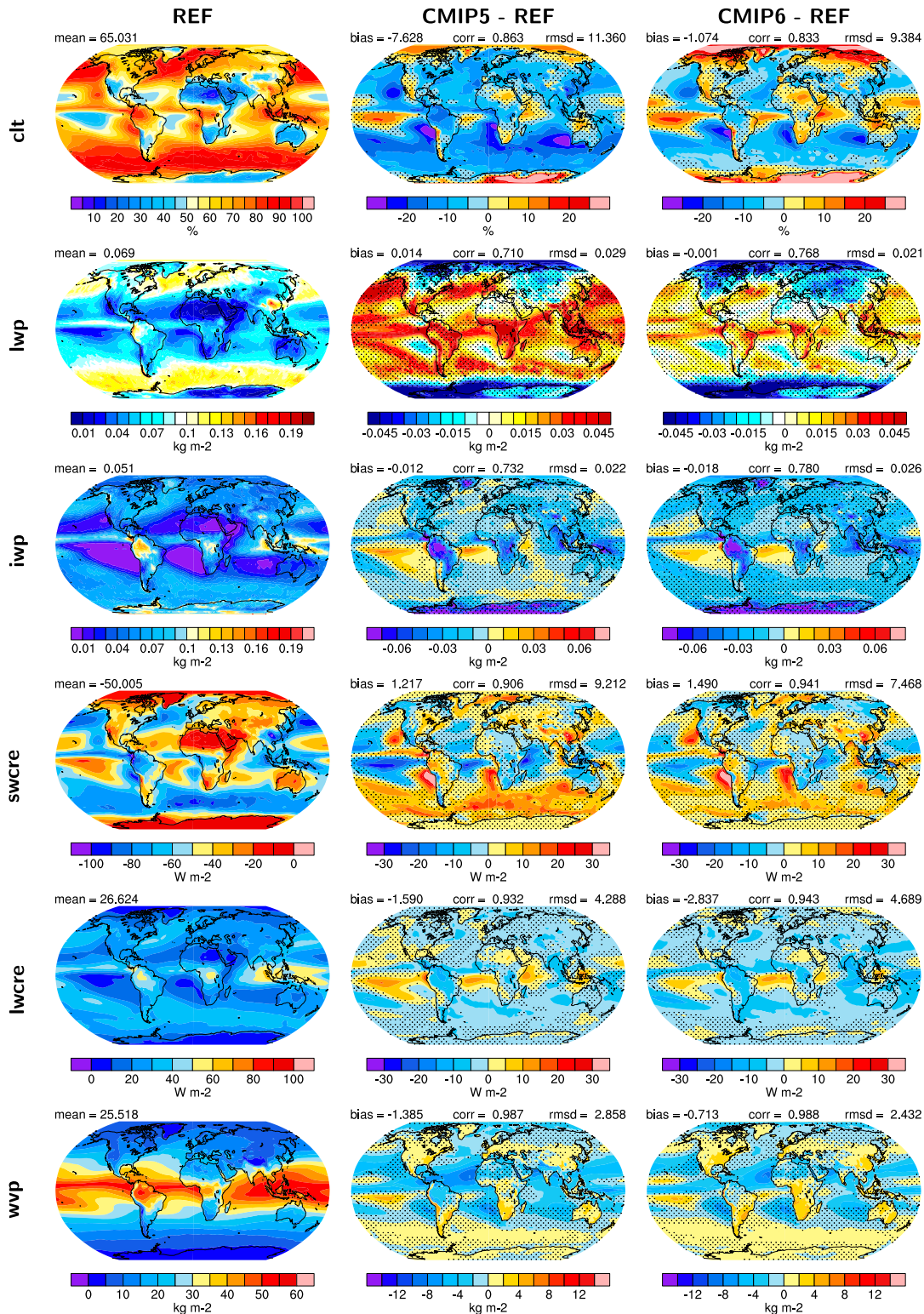


FIG. 1. The 20-yr annual averages (models) of total cloud cover (clt), liquid water path (lwp), ice water path (iwp), TOA shortwave (swcre) and longwave (lwcre) cloud radiative effects, and water vapor path (wvp), shown from top to bottom. (left) Observational reference datasets (multiobs means) and the differences of the (center) CMIP5 and (right) CMIP6 multimodel means to the observational reference dataset. Differences that are smaller than the observational uncertainty estimate calculated from interannual variability and variability across individual observational datasets compared with the multiobs mean [Eq. (3)] are stippled.



TABLE 5. Range (minimum and maximum) of the 20-yr global annual mean, average bias, pattern correlation, and root-mean-square deviation (RMSD) from individual observations and individual CMIP5 and CMIP6 models compared with the multiobs reference. For the models, also the 10% and 90% quantiles are given in parentheses.

	Mean			Bias			Correlation			RMSD		
	obs	CMIP5	CMIP6	obs	CMIP5	CMIP6	obs	CMIP5	CMIP6	obs	CMIP5	CMIP6
clt (%)	62–67	46–72 (50–63)	54–70 (58–68)	–3 to 2	–19 to 7 (–15 to –2)	–12 to 5 (–4 to 3)	0.96 to 0.99	0.35–0.89 (0.40–0.83)	0.60–0.88 (0.62–0.83)	3–5	10–21 (11–18)	9–16 (10–14)
lwp ( $\text{g m}^{-2}$ )	36–105	32–137 (40–130)	28–140 (35–96)	–33 to 36	–36 to 68 (–29 to 61)	–41 to 72 (–34 to 27)	0.49–0.94	0.05–0.84 (0.27–0.81)	0.37–0.87 (0.39–0.80)	22–47	28–110 (32–77)	22–89 (23–59)
iwp ( $\text{g m}^{-2}$ )	36–61	8–102 (15–70)	5–108 (9–52)	–15 to 10	–43 to 51 (–35 to 19)	–46 to 58 (–42 to 2)	0.74–0.93	0.37–0.71 (0.40–0.68)	0.36–0.77 (0.40–0.75)	17–27	23–100 (29–57)	23–108 (24–48)
swcre ( $\text{W m}^{-2}$ )	–56 to –46	–55 to –41 (–54 to –45)	–60 to –43 (–60 to –45)	–6 to 4	–5 to 9 (–4 to 5)	–10 to 7 (–10 to 5)	0.98–0.99	0.53–0.91 (0.56–0.86)	0.74–0.93 (0.75–0.92)	4–7	11–20 (11–18)	8–18 (9–14)
lwcre ( $\text{W m}^{-2}$ )	25–28	19–32 (19–30)	21–30 (21–26)	–2 to 1	–8 to 6 (–8 to 3)	–6 to 3 (–5 to –1)	0.98–0.99	0.64–0.91 (0.73–0.89)	0.76–0.93 (0.76–0.92)	2–3	5–11 (5–10)	5–9 (5–8)
wvp ( $\text{kg m}^{-2}$ )	25–26	22–26 (23–26)	22–28 (23–27)	–0.7 to 0.7	–3.4 to 0.7 (–2.9 to 0.0)	–3.5 to 2.1 (–3.0 to 1.1)	0.99	0.97–0.99 (0.97–0.98)	0.97–0.99 (0.97–0.99)	2.0	2.6–5.1 (2.8–4.6)	2.5–4.6 (2.7–4.3)

The observational uncertainty in ice water path is similarly high as the one for lwp as indicated by the large spread between different satellite products (see Table 5). Compared to CMIP5, we find a reduction by about 15% in the global annual MMM from CMIP6 to  $32 \text{ g m}^{-2}$ . This value is slightly below the range of observational datasets analyzed ( $36\text{--}61 \text{ g m}^{-2}$ ). Pattern correlation in CMIP6, however, increased from 0.73 in CMIP5 to 0.78.

Compared with CMIP5, the CMIP6 MMM of the short-wave cloud (sw) radiative effect shows an improvement in the overestimation of sw cooling north and south of ITCZ by about  $5\text{--}10 \text{ W m}^{-2}$ . Over the Southern Ocean, the underestimation of the observed sw cooling in the CMIP6 models is improved by about  $5 \text{ W m}^{-2}$ . This improvement is at least partly based on advances in model parameterizations and was already shown for several individual models (e.g., Danabasoglu et al. 2020; Kawai et al. 2019; Madeleine et al. 2020; Voldoire et al. 2019). In contrast, there is rather little change in the underestimation of swcre in the subtropical stratocumulus regions between CMIP5 and CMIP6 as described, for example, for the IPSL model by Madeleine et al. (2020). Biases of up to  $30\text{--}40 \text{ W m}^{-2}$  remain for example in the southeastern Pacific. Pattern correlation of swcre improved slightly from 0.91 in CMIP5 to 0.94 in CMIP6. This agrees very well with results of Jian et al. (2020).

Similar to swcre, there is an overestimation in the longwave cloud radiative effect north and south of the ITCZ in the models. This bias is improved in CMIP6 by about  $5 \text{ W m}^{-2}$  compared with the CMIP5 MMM, already seen for CNRM-CM6-1 (Voldoire et al. 2019) and CESM2 (Danabasoglu et al. 2020). Otherwise, there are very few changes from CMIP5 to CMIP6. Pattern correlations are high and amount 0.93 in CMIP5 and 0.94 in CMIP6.

The geographical distribution of water vapor path in CMIP6 shows a small improvement with smaller biases over the tropical ocean of about  $1\text{--}2 \text{ kg m}^{-2}$ , a smaller overestimation of wvp north and south of the ITCZ, and about  $1\text{--}2 \text{ kg m}^{-2}$  less underestimation over the tropical Atlantic. The global annual average of the CMIP6 MMM ( $24.8 \text{ kg m}^{-2}$ ) is slightly larger than in CMIP5 ( $24.1 \text{ kg m}^{-2}$ ) and now within range of analyzed observational datasets ( $24.8\text{--}26.2 \text{ kg m}^{-2}$ ). The pattern correlations of the MMMs are high and amount to about 0.99 for both CMIP5 and CMIP6.

For some of the parameters investigated here (lwp, iwp, swcre, wvp), we find that the CMIP6 intermodel spread is larger than in CMIP5 (see Table 5), while the average root-mean-square deviation is decreased (clt, lwp, swcre, wvp). We would expect this increased intermodel spread to be at least partly caused by an increased complexity of the CMIP6 models compared to CMIP5 introducing additional degrees of freedom.

Biases in simulated sea surface temperatures (SSTs) can affect simulated cloud properties. We therefore also analyzed some results from simulations using the atmosphere components of the CMIP models and prescribed observed SSTs (the so-called AMIP runs) for the same time periods (not shown). Similar to Lauer and Hamilton (2013), who investigated CMIP3 and CMIP5 models, we found that in general the skill

of the AMIP multimodel means (MMMs) in reproducing the observed cloud climatologies measured by global means, biases, pattern correlations, and RMSDs does not systematically differ from the MMM obtained from the historical simulations. This is the case for both CMIP5 and CMIP6. While there are some improvements of specific biases such as, for example, the double-ITCZ visible in the geographical distributions of lwp and iwp, the AMIP models do not systematically outperform the coupled models in reproducing observed mean cloud properties.

## 2) TAYLOR DIAGRAMS

The Taylor diagrams shown in Fig. 2 give the spatial standard deviation and linear pattern correlation of the 20-yr annual means of a variable from individual models and observational climatologies with the corresponding multiobs mean reference dataset (see also Table 4). The spatial standard deviation is given as the ratio between the value for the individual dataset and the value from the reference dataset. The reference dataset is represented in each panel by the filled black circle on the  $x$  axis at  $x = 1$ . In the polar coordinate system used by the Taylor diagrams, the linear distance between the individual dataset and the reference dataset is proportional to the (centered) root-mean-square error (rmse) (Taylor 2001) and can be estimated using the green circles centered on the reference (black) dots. Using the multiobs mean as reference data allows for also including the individual observational datasets as an estimate of the observational uncertainty.

For all six variables investigated, the performance of the MMMs from CMIP5 and CMIP6 in reproducing the observed annual mean distribution is quite similar. For clt, the pattern correlation of the MMM decreased slightly from 0.86 in CMIP5 to 0.83 in CMIP6 while the spread of the individual models' correlations narrowed down from 0.35–0.89 in CMIP5 to 0.60–0.88 in CMIP6. For comparison, also data from the reanalyses ERA-Interim and ERA5 are included. Their pattern correlations are 0.90 and 0.89, respectively. Correlations of the individual observational datasets range between 0.96 and 0.99.

The models' skill in reproducing the geographical distribution of shortwave and longwave cloud radiative effects as well as water vapor path is high compared to the other variables investigated. Correlations range from 0.91 for swcre for CMIP5 (CMIP6 = 0.94) to 0.99 for wvp (CMIP5 and CMIP6). In contrast, the intermodel spread as well as the interobservation spread for lwp and iwp are very high regarding the pattern correlations as well as the ratio of the spatial standard deviations. For lwp, the pattern correlations of the MMMs (CMIP5 = 0.71, CMIP6 = 0.76) are within the range given by individual observational datasets, between 0.49 (MAC-LWP) and 0.94 (MODIS), which makes an assessment of the models' performance difficult and confirms the challenges of obtaining accurate global maps of lwp from satellite retrievals. For iwp, the spread of pattern correlations for the individual observational datasets ranges from 0.74 (CLARA-AVHRR) to 0.94 (ESACCI-CLOUD). For comparison, the pattern

correlation of the CMIP5 MMM with the multiobs reference is 0.73 and 0.78 for CMIP6. Again, because of the large spread among the different observational datasets, this improvement in CMIP6 is difficult to assess. More accurate measurements of the global iwp distribution would be needed.

## 3) ZONAL MEANS

The vertical distribution of clouds plays an important role in their effect on the radiation budget. A negative net cloud radiative effect calculated as the sum of swcre and lwcre is a cooling effect, a positive sign is a net warming effect of clouds. While for optically thick low-level clouds the net cloud radiative effect is typically dominated by swcre, resulting in a cooling net effect, lwcre becomes more important with increasing cloud-top height as the temperature difference between the cloud top and the surface increases.

Figure 3 shows a comparison of the zonal means of the 3D cloud fraction from CALIPSO-GOCCP (Chepfer et al. 2010) and the CMIP5 and CMIP6 MMMs of the subset of models providing the corresponding output (clcalipso) from the satellite simulation software COSP. The observations show large cloud fractions in the lowermost troposphere (0–2 km) over the Southern Ocean and in Northern Hemispheric middle and high latitudes. The 3D cloud fraction decreases with height with the values in midlatitudes, being higher and appearing separated from the adjacent subtropical latitudes with low cloud fractions. Another local maximum in the observed zonal mean distribution of the 3D cloud fraction is observed in the upper troposphere (12–15 km) in the central tropics with the highest values in the ITCZ. This geographical distribution is qualitatively reproduced by the models. While both model generations underestimate cloud fraction in the lower and middle troposphere by roughly 5%, and by up to 30% (absolute differences) in the boundary layer over the Southern Ocean, cloud fractions are overestimated by 3%–5% in a band extending about 2–3 km below the tropopause region across most latitudes. This overestimation of the cloud fraction is more pronounced in the CMIP6 MMM than in CMIP5, already indicated for single models (e.g., the IPSL model; Madeleine et al. 2020). An exception to this is the tropical tropopause region, in which the CMIP5 models underestimate the observed cloud fraction by up to 5%. This underestimation is not present in the CMIP6 MMM.

As COSP output for clcalipso was only available from 9 CMIP5 and 7 CMIP6 models, we also calculated zonal means of the native (unprocessed) 3D cloud fraction (cl) that was available from 38 CMIP5 and 35 CMIP6 models. These are compared to ERA5 reanalysis results as no directly comparable observations are available. The zonal mean cloud fraction from ERA5 is qualitatively similar to the one from CALIPSO, but shows generally higher values of up to 10% throughout the domain with exceptions being the lowermost troposphere over the Southern Ocean (–20%) and the northern midlatitudes (–15%), as well as the middle troposphere in the ITCZ region (–5%). Both model generations generally overestimate the ERA5 cloud fraction by about 0%–5% (CMIP5) and 0%–10% (CMIP6). Particularly in CMIP5, there is no clearly

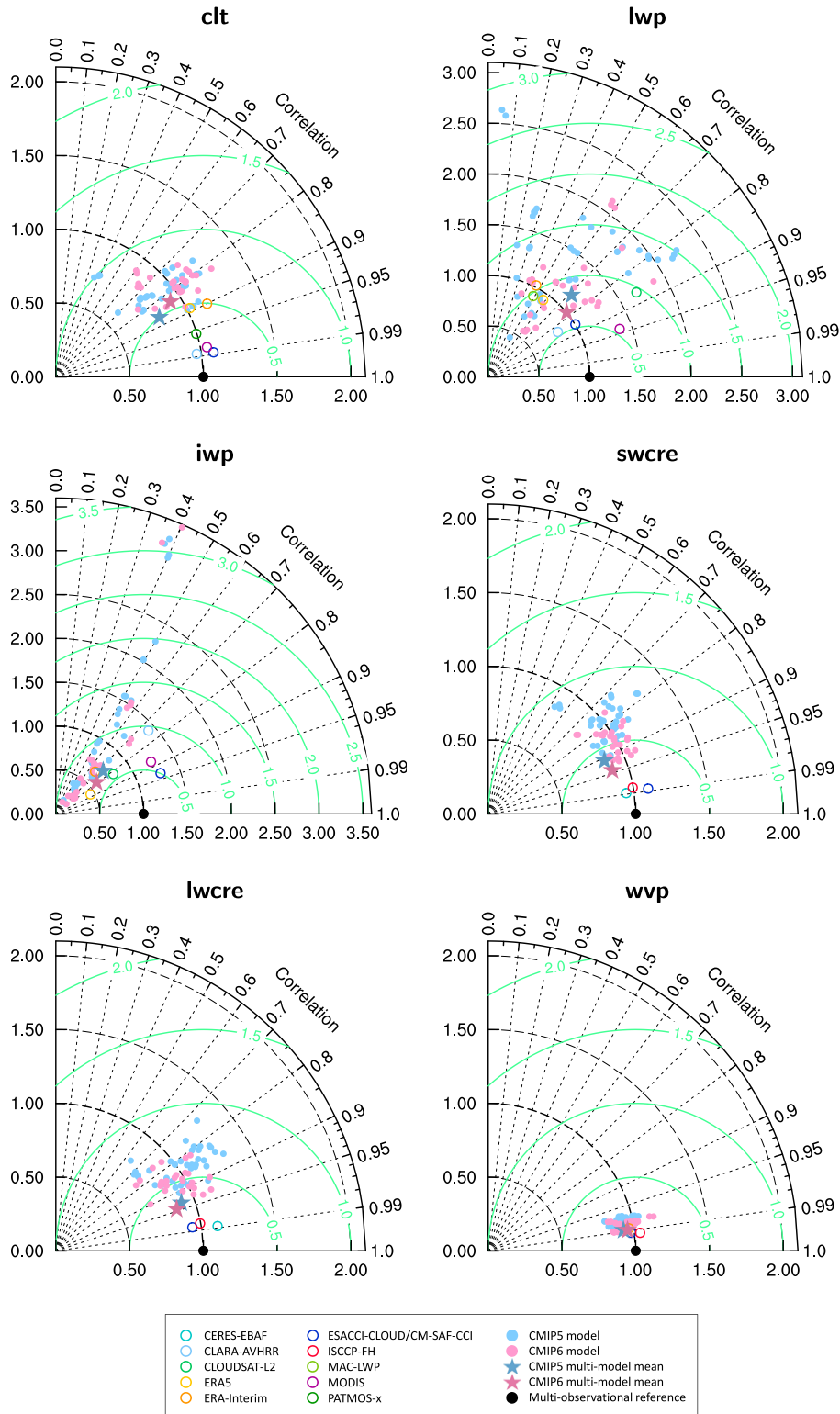


FIG. 2. Taylor diagrams showing the 20-yr annual average performance of the CMIP5 (light blue) and CMIP6 (light red) models for total cloud cover (clt), liquid water path (lwp), ice water path (iwp), TOA shortwave (swcre) and longwave (lwcre) cloud radiative effects, and water vapor path (wvp). Individual models are shown by the filled circles, the multimodel means by the stars, and the individual observational datasets by the colored circles. The multiobs reference is shown by the filled black circle on the  $x$  axis at  $x = 1$ .

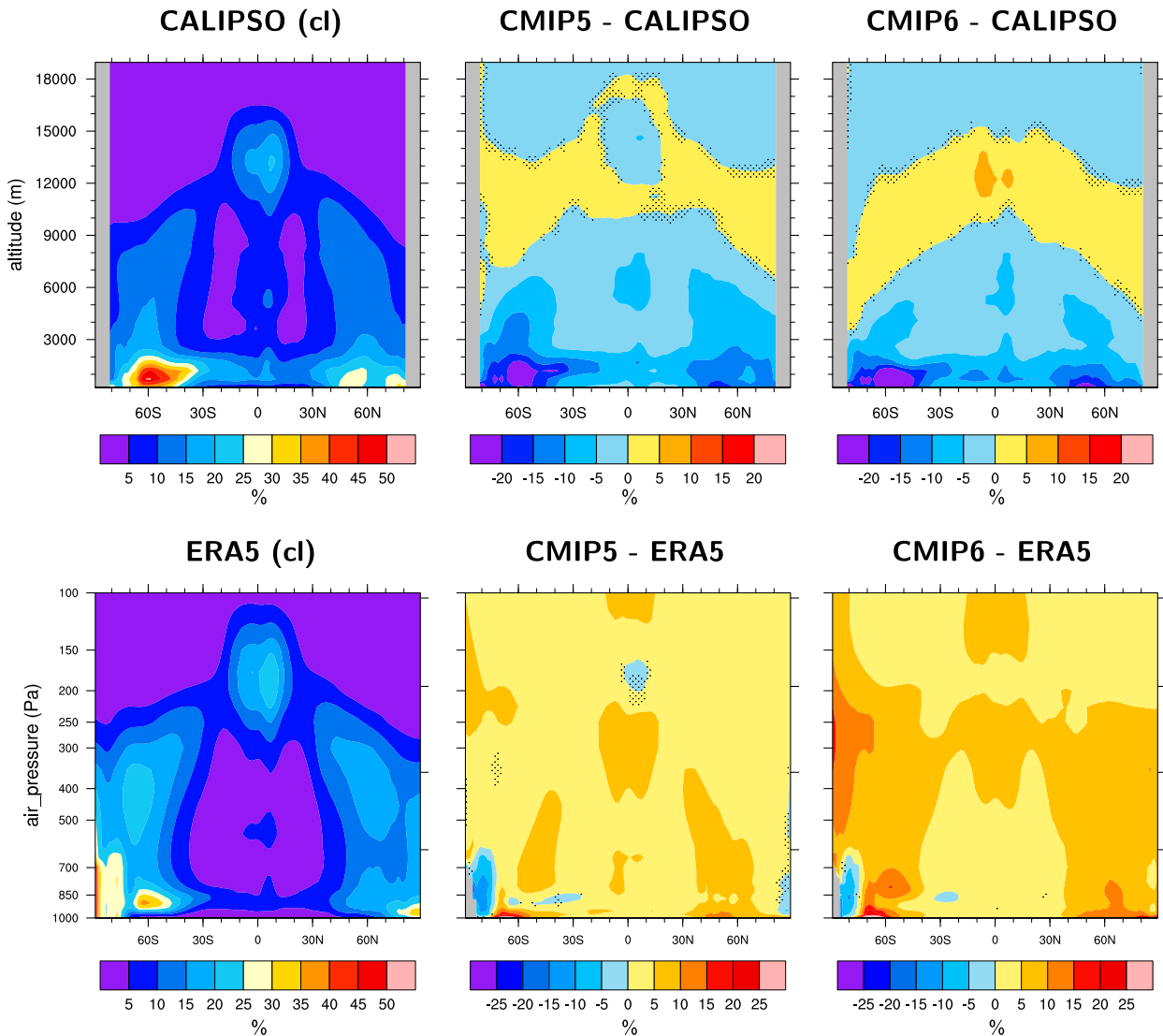


FIG. 3. Multiyear annual averages of zonal mean cloud fraction (%). The top row shows (left) data from *CALIPSO*-GOCCP, as well as differences of the (center) CMIP5 (9 models) and (right) CMIP6 (7 models) multimodel mean (models providing COSP output) compared with *CALIPSO*-GOCCP. The bottom row shows (left) data from ERA5, as well as differences of the (center) CMIP5 (38 models) and (right) CMIP6 (35 models) multimodel mean compared with ERA5. Stippled differences are not statistically significant at a 95% confidence level.

outstanding pattern in the bias of cloud fraction visible related to specific regions or cloud types.

The vertical distribution of cloud liquid water and cloud ice water from observations from *CloudSat* and from *CALIPSO*-ICECLOUD, respectively, and the deviation of the CMIP5 and CMIP6 MMMs from the observations are shown in Fig. 4. Zonal mean cloud liquid water (clw) observations show maxima centered in the lowermost troposphere at about 800–900 hPa and 60°S and 60°N. Bands of high values are extending from these locations to about 10°N (the ITCZ region) and 400 hPa. Another band of high values is found in the lowermost troposphere at about 850 hPa across almost all

latitudes. Both model generations overestimate clw in the boundary below the 900-hPa layer by up to  $0.03 \text{ g kg}^{-1}$  (CMIP5) and  $0.02 \text{ g kg}^{-1}$  (CMIP6). Here, *CloudSat* shows almost no cloud liquid water. We would like to note that *CloudSat*'s cloud profiling radar is known to have difficulties in detecting low-level clouds because of the inherent difficulty the sensor has in detecting such clouds due to surface clutter (Marchand et al. 2008).

In contrast, the models underestimate clw in both bands with observed high liquid water values extending from about 900 hPa at 60°S and 60°N to about 10°N at 400 hPa. The CMIP5 MMM underestimates clw in these bands by about

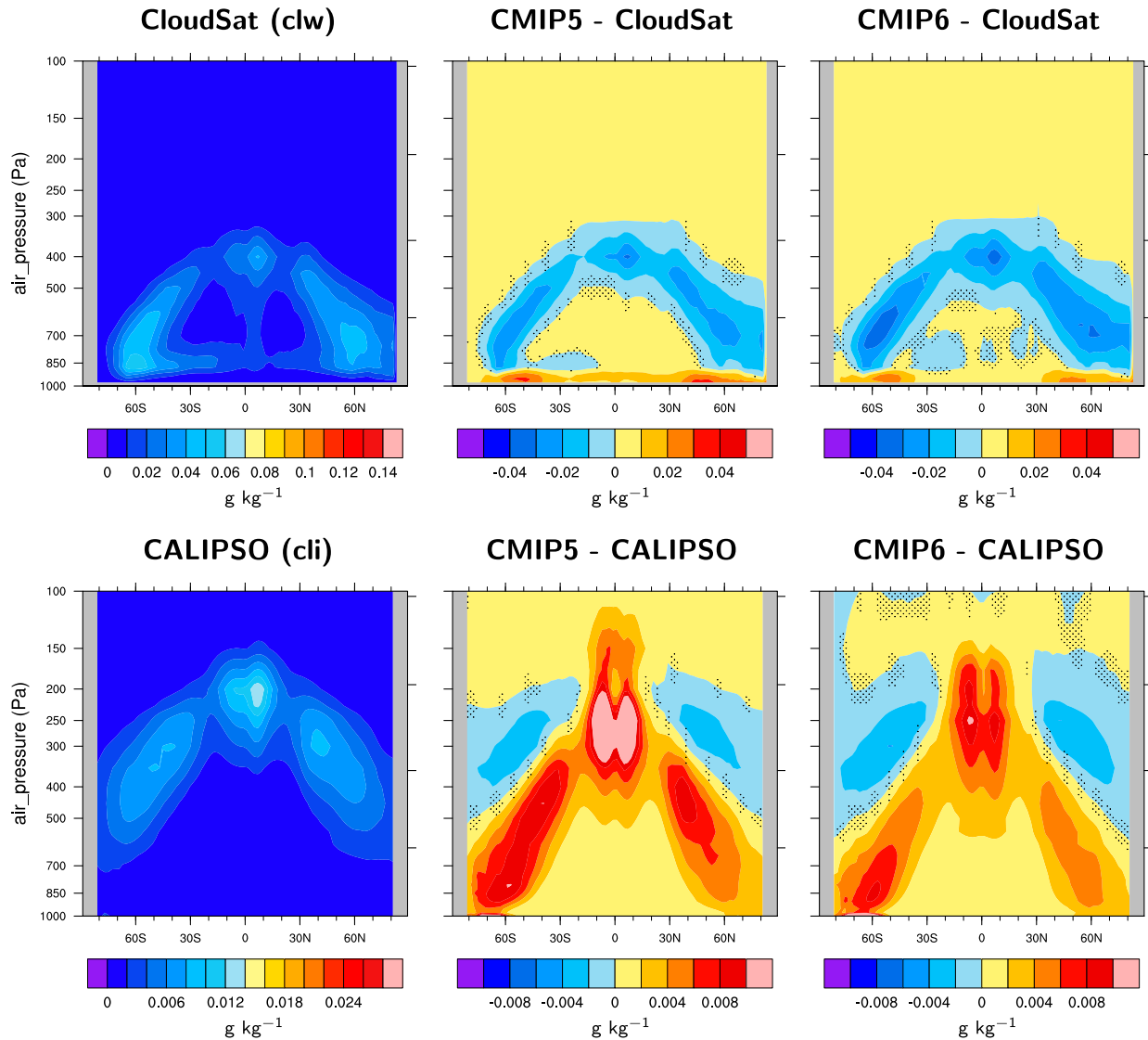


FIG. 4. The top row shows the multiyear annual average of (left) the zonal mean cloud liquid water content ( $\text{g kg}^{-1}$ ) from *CloudSat*-L2, and differences of the (center) CMIP5 (32 models) and (right) CMIP6 (28 models) multimodel mean compared with *CloudSat*. The bottom row shows (left) the zonal mean cloud ice water content ( $\text{g kg}^{-1}$ ) from *CALIPSO*-ICECLOUD, and differences of the (center) CMIP5 (32 models) and (right) CMIP6 (27 models) multimodel mean compared with the *CALIPSO* data. Stippled differences are not statistically significant at a 95% confidence level.

50% in the midlatitudes and up to about 75% in the tropics. This underestimation is even more pronounced in the CMIP6 MMM, amounting to about 75% in the midlatitudes and up to 90% in the tropics. It is noteworthy that the *CloudSat* values used here do not include pixels with precipitating clouds as *CloudSat* is known to be sensitive to both cloud water and precipitation [see section 2b(5)]. The *CloudSat* clw values are therefore rather regarded as a lower estimate, making the underestimation (above the boundary layer) of clw in the models even more pronounced.

For cloud ice water content (cli), the *CALIPSO*-ICECLOUD observations show a maximum in the zonal means at about

$10^{\circ}\text{N}$  and 200 hPa (ITCZ) as well as two bands of high values in northern and southern midlatitudes at around 300 hPa. The CMIP5 models overestimate cli in the lower and middle troposphere in midlatitudes by up to  $0.01 \text{ g kg}^{-1}$ , in the upper tropical troposphere at about 250 hPa by up to  $0.02 \text{ g kg}^{-1}$ . This bias is reduced in the CMIP6 MMM showing deviations in the lower and middle troposphere in midlatitudes up to  $0.008 \text{ g kg}^{-1}$  in the Southern Hemisphere and up to  $0.005 \text{ g kg}^{-1}$  in the Northern Hemisphere. This improvement was already seen for the IPSL models (Madeleine et al. 2020). In the upper tropical troposphere, the CMIP6 MMM bias is reduced to  $0.01 \text{ g kg}^{-1}$ . In contrast to this overestimation, both model generations show an



underestimation of  $cli$  in the upper troposphere in the middle and high latitudes by about  $0.003 \text{ g kg}^{-1}$ . We would like to note that the vertical integral of the 3D *CALIPSO*-ICECLOUD data (not shown) result in about a factor of 2 higher  $iwp$  values than the multiobs mean [see section 4a(1)]. This suggests that the overestimation of  $cli$  by the models is not simply caused by rather low  $cli$  values in the *CALIPSO*-ICECLOUD dataset due to saturation effects of the lidar instrument.

#### 4) ICE FRACTION

To investigate the pattern of too much cloud ice in the lower and middle troposphere and too little ice in the upper troposphere in the models [section 4a(3)] and to better understand the differences between CMIP5 and CMIP6, we calculate average cloud ice fraction as a function of temperature. Similar to Zelinka et al. (2020), we use monthly means for the calculation of the ice fraction but instead of vertical integrals of cloud liquid water and cloud ice, we use the full 3D fields. For each 3D grid cell containing cloud water, the ice fraction is calculated as  $cli/(cli + clw)$  and then binned by temperature from the same grid cell. We use  $10^{-6} \text{ kg kg}^{-1}$  as a threshold for the minimum cloud water to calculate the ice fraction within the corresponding grid cell. Grid cells with less cloud water (sum of ice + liquid) are ignored. As an observationally based estimate, we do the same calculation using  $cli$  from *CALIPSO*-ICECLOUD,  $clw$  from *CloudSat*, and air temperature ( $ta$ ) from ERA5. Figure 5 shows the results for nine individual CMIP5/CMIP6 model pairs that we identified by almost identical model names (BCC-CSM1/2, CESM1/2-WACCM, GFDL-CM3/4, GISS-E2-H/GISS-E2-1-H, HadGEM2-CC/HadGEM3-GC31-LL, MICRO-ESM/MIROC-ES2L, MPI-ESM-LR/MPI-ESM1-2-LR, MRI-ESM1/MRI-ESM2-0, NorESM1-M/NorESM2-LM) and the MMM of all 32 CMIP5 and 28 CMIP6 models with the required data available. The observations show a very steep gradient in ice fraction from  $>90\%$  at 245 K to about 10% at 260 K. The models show a wide spread in the relation between temperature and ice fraction. A number of models and in particular the MMMs show a strongly overestimated ice fraction in the temperature range 250–270 K and an underestimation at temperatures below about 245 K. The intermodel spread is, however, as shown by the shaded area indicating the  $\pm 1$  standard deviation range in Fig. 5, very large and includes the observations at most temperatures except around 260 K. While some models (BCC-CSM, CESM-WACCM, GISS-E2, MRI-ESM) show large differences between the CMIP5 and the CMIP6 version, the models GFDL-CM, HadGEM, MPI-ESM, and NorESM show rather little difference between the two model generations. Particularly the CMIP6 versions of BCC-CSM, CESM-WACCM, and MRI-ESM are now closer to the observations [also shown in, e.g., Frey and Kay (2018) and Kawai et al. (2019)]. While the average ice fraction per temperature bin is higher in some CMIP6 models compared with their CMIP5 counterparts (BCC-CSM, CESM-WACCM, GFDL-CM) it is lower in other models (GISS-E2, HadGEM, MRI-ESM) as shown in Bodas-Salcedo et al. (2019). Across the whole model ensemble (MMM), both model ensembles show few differences in the mean and intermodel spread

across most temperatures. An exception is the temperature range 245–255 K, in which the CMIP6 models predict a cloud ice fraction that is about 4%–5% smaller than in CMIP5 (absolute differences). This suggests that the ice fraction for a given air temperature, which we expect to be to a large degree influenced by cloud microphysics, is probably not the only reason explaining the difference in the zonal means of  $clw$  and  $cli$  between CMIP5 and CMIP6. Other reasons contributing to the differences in simulated cloud ice could be differences in the simulated temperature fields. An analysis of the annual mean zonal average temperature (not shown) shows that the models simulate slightly lower temperatures compared with ERA5 in regions where cloud ice is overestimated by the models. Similar to Tian et al. (2013), who reported a tropospheric cold bias in many CMIP5 models, we find that temperatures are underestimated in CMIP5 in northern and southern midlatitudes at around 450 hPa by about 2 K, and in the upper tropical troposphere at around 250 hPa by about 1 K. This cold bias is improved by about 0.5 K in the CMIP6 MMM, which might also contribute to a small degree to the reduced cloud ice in the CMIP6 MMM in these parts of the troposphere.

In addition to the global averages, we also calculated the ice fraction for the Southern Ocean (latitude belt  $30^{\circ}$ – $65^{\circ}$ S) and Pacific ITCZ ( $135^{\circ}$ E– $85^{\circ}$ W,  $0^{\circ}$ – $12^{\circ}$ N) (not shown). While we found the results for the Southern Ocean to differ very little from the global averages, the models show higher ice fractions in the Pacific ITCZ. Here, the CMIP5 MMM shows an increase in ice fraction of up to about 10% at 260 K and the CMIP6 MMM of up to about 15% at 250 K. For the temperature range of  $\sim 250$ – $270$  K, this increases the overestimation of the ice fraction in this region compared with the global averages. This suggests that an important contribution to the overestimation of modeled cloud ice in this temperature range is related to convective clouds.

Our results suggest that the overestimation of the cloud ice fraction over a large temperature range ( $\sim 250$ – $270$  K) in the models probably explains part of the overestimation of cloud ice in the lower and middle troposphere as cloud ice can either be formed or persist in the models at temperatures higher than shown in the observation (i.e., at lower altitudes in the troposphere). We tested this hypothesis by applying the observed ice fraction to the model data whenever the modeled ice fraction for a given temperature was larger than the observed one. We did this by multiplying the modeled ice concentration by the ratio of ice fraction( $T$ )<sub>obs</sub> and ice fraction( $T$ )<sub>mod</sub> whenever the observed ice fraction is smaller than the modeled one for a given temperature. This results in a decrease of the positive bias in simulated zonally averaged MMM cloud ice in the lower and middle troposphere in midlatitudes by roughly 70%–80%. Another reason for the overestimation in cloud ice could be that cloud ice is not removed quickly enough from the atmosphere (e.g., through sedimentation).

#### b. Interannual variability

The interannual variability is estimated by subtracting the average seasonal cycle from the time series of monthly means and then calculating the relative temporal standard deviation

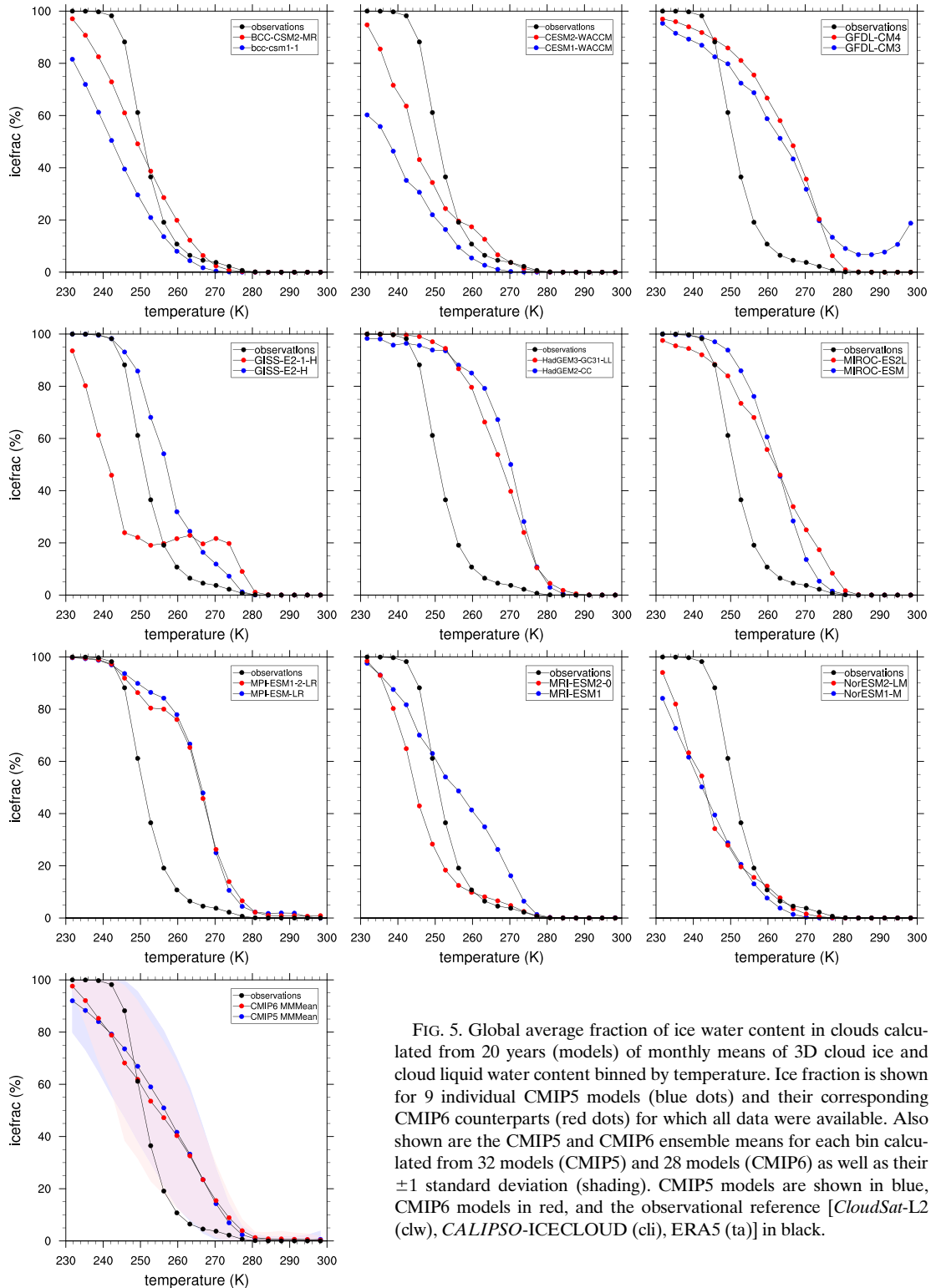


FIG. 5. Global average fraction of ice water content in clouds calculated from 20 years (models) of monthly means of 3D cloud ice and cloud liquid water content binned by temperature. Ice fraction is shown for 9 individual CMIP5 models (blue dots) and their corresponding CMIP6 counterparts (red dots) for which all data were available. Also shown are the CMIP5 and CMIP6 ensemble means for each bin calculated from 32 models (CMIP5) and 28 models (CMIP6) as well as their  $\pm 1$  standard deviation (shading). CMIP5 models are shown in blue, CMIP6 models in red, and the observational reference [*CloudSat-L2* (clw), *CALIPSO-ICECLOUD* (cli), ERA5 (ta)] in black.

[Eq. (4) and section 3e]. The observations show in Fig. 6 a high interannual variability of all cloud parameters in the Pacific ITCZ as well the tropical ocean regions and the South Pacific convergence zone (SPCZ). Particularly in the midlatitudes,

the continents often show a higher interannual variability than the ocean regions at the same latitude. Other regions of high variability are for example the Sahara in North Africa. This is a result of very small cloud amounts in this region so that even small

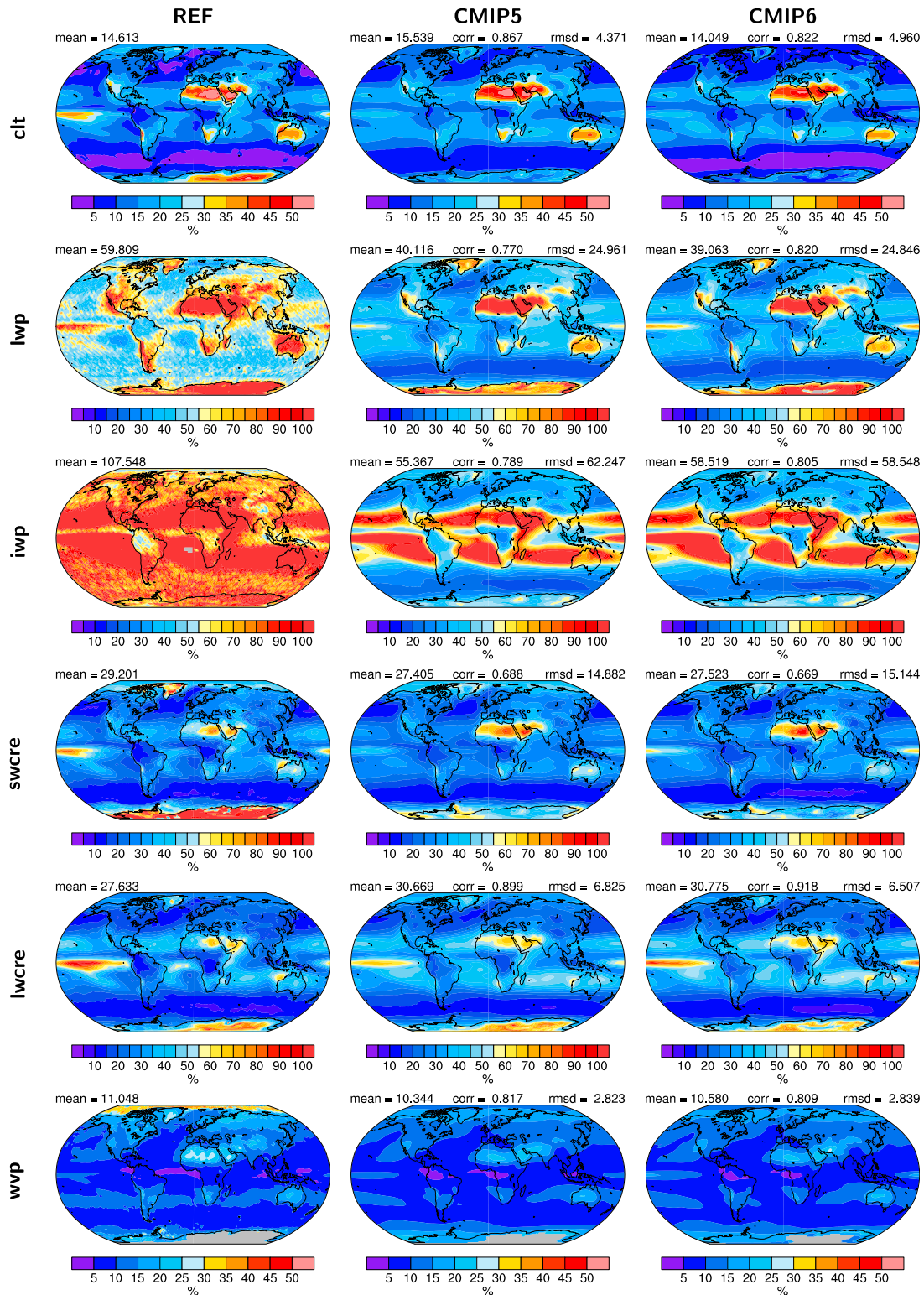


FIG. 6. Relative temporal standard deviation of 20 years (models) of total cloud cover (clt), liquid water path (lwp), ice water path (iwp), TOA shortwave (swcre) and longwave (lwcre) cloud radiative effects, and water vapor path (wvp), shown from top to bottom, from the (center) CMIP5 and (right) CMIP6 models compared with (left) multiobs averages. The temporal standard deviations are calculated from monthly anomalies after subtracting the climatological mean seasonal cycle. These have then been averaged over all models [Eq. (4)].

changes lead to large relative changes. This general pattern is qualitatively reproduced by the models. However, a common feature across different variables is an underestimation of the observed interannual variability in the Pacific ITCZ by the models. For CMIP5, this underestimation ranges between about 20% to 30% for *lwcre* and *swcre* and around 30% to 40% for *lwp*, and it has been reduced by about 10% for these three variables in CMIP6. For all eight variables investigated here, the differences between the CMIP5 and CMIP6 MMMs throughout most of the domain are, however, rather small (typically <5%). It is noteworthy that the global average interannual variabilities of *wvp* and *clt* are small (<15%) in both the observations and the models. The observed variabilities of *lwp* and *iwp*, however, are high, about 60% and 108%, respectively. In both cases, the average variability is strongly underestimated in both model generations, amounting to about 40% for *lwp* and >60% for *iwp*. While part of these differences can probably be attributed to the large uncertainties in *lwp* and *iwp* data obtained from space, this could also point to the total cloud water path in the models not being sensitive enough to changes in dynamics and thermodynamic structure of the atmosphere as this underestimation is found in many different cloud regimes and geographical regions. The observed interannual variability of *lwcre* and *swcre* lies in between the ones for *clt* and for cloud water path with about 28% (*lwcre*) and 27% (*swcre*). Both model ensembles show similar global averages of about 31% for *lwcre* and about 27% for *swcre*.

### c. Seasonal cycle

As a measure for the strength of the seasonal cycle, we calculate its amplitude as the relative difference between minimum and maximum monthly mean values within the average seasonal cycle [Eq. (4) and section 3e]. Figure 7 shows the seasonal cycle amplitude for the six parameters *clt*, *lwp*, *ipw*, *swcre*, *lwcre*, and *wvp*. For total cloud cover, the observed seasonal cycle amplitude is about 40%–60% over large parts of the tropical and subtropical oceans, between 10% and 20% over midlatitude oceans, and <10% in large parts of the Southern Ocean. The amplitudes are higher particularly over large parts of low- and midlatitude continental areas. While this pattern is qualitatively reproduced by the models, the CMIP5 models overestimate the seasonal cycle amplitude by about 10%, particularly over the tropical ocean and the central parts of the Southern Ocean. This overestimation is slightly reduced in the CMIP6 MMM by about 5%. In northern India, northern Australia, and southern Africa, the models underestimate the observed seasonal cycle amplitude of *clt*. While there is little difference in these negative biases between CMIP5 and CMIP6 in India and Australia, the underestimation in southern Africa is about 10% stronger in CMIP6 than in CMIP5. The observed seasonal cycle amplitude of *lwp* and *iwp* in midlatitude ocean regions and in the Southern Ocean in particular is underestimated by the models by about 10%–20%. This negative bias is slightly reduced in the CMIP6 MMM by about 5%–10%. The observed seasonal cycle amplitudes are overestimated by the models over the tropical ocean by about 10% for *swcre* and by up to 20%–30% for *lwcre*. This bias is slightly larger (~5%–10%) in

CMIP6 than in the CMIP5 MMM. Similar to *clt*, the seasonal cycle amplitude of *wvp* is overestimated by the models particularly over tropical and subtropical oceans by about 10%. This did not change significantly from CMIP5 to CMIP6.

### d. Dynamical regime

The relative occurrence of specific cloud types is sensitive to the large-scale circulation (dynamical regime) and local thermodynamic conditions (Bony et al. 2004). Bony and Dufresne (2005) used monthly mean 500-hPa vertical velocity ( $\omega_{500}$ ) as a proxy for the large-scale circulation over tropical oceans (latitude belt from 30°S to 30°N) to calculate composites of climate variables. Similar to Williams et al. (2003), we calculate two-dimensional composites of cloud properties binned by 500-hPa vertical velocity as a proxy for the large-scale circulation and binned by SST as a proxy for local thermodynamic conditions. We used all grid cells over the ocean. As observational reference, we use ESACCI-CLOUD data for total cloud cover and total cloud water path as well as ERA5 for SST and  $\omega_{500}$ .

Figure 8 shows such composites for total cloud cover (top) and total cloud water path (middle). The CMIP5 and CMIP6 ensemble means have been calculated by averaging the two-dimensional distributions from each individual model participating in CMIP5 and CMIP6, respectively. In this analysis, the ITCZ can be roughly characterized by SSTs > 300 K and  $\omega_{500} < -3 \text{ Pa min}^{-1}$ . The observations show an increase in *clt* from about 75% to >90% with updraft velocity. This increase is qualitatively reproduced by the CMIP5 models, but absolute values are underestimated by several percent with maximum cloud cover values of about 85%. This is improved in CMIP6 particularly for strong updraft velocities. The *tcwp* increases from about 0.2 to 0.3  $\text{kg m}^{-2}$  with increasing updraft velocities in the observations, which is overestimated in CMIP5 models and is better matched in CMIP6.

In the subtropical stratocumulus regions, roughly characterized in this analysis by SSTs between 290 and 295 K and subsidence (vertical velocities of  $>0 \text{ Pa min}^{-1}$ ), observations show that total cloud cover is particularly sensitive to SST and decreases roughly from 65% at 290 K to 55% at 295 K. This sensitivity is underestimated in CMIP5 (decrease from roughly 55% at 290 K to 50% at 295 K). The CMIP6 MMM also underestimates this gradient in *clt* with SST but shows a larger sensitivity of total cloud cover to SST (about 60% at 290 K to 50% at 295 K) that is in better agreement with observations. In both model ensembles, absolute *clt* values are underestimated compared with the observations. Absolute changes in *tcwp* are small, as *tcwp* values in this regime are typically rather small, and also decrease with increasing SST with little difference between CMIP5 and CMIP6.

Extratropical conditions are characterized by SSTs being typically below 290–295 K depending on the season. Here, we find an increase in total cloud cover with updraft velocities and a decrease with increasing SST. This behavior is qualitatively reproduced by the models but the CMIP5 MMM underestimates *clt* values by about 10% throughout most of the extratropical regime. The CMIP6 MMM is in better agreement



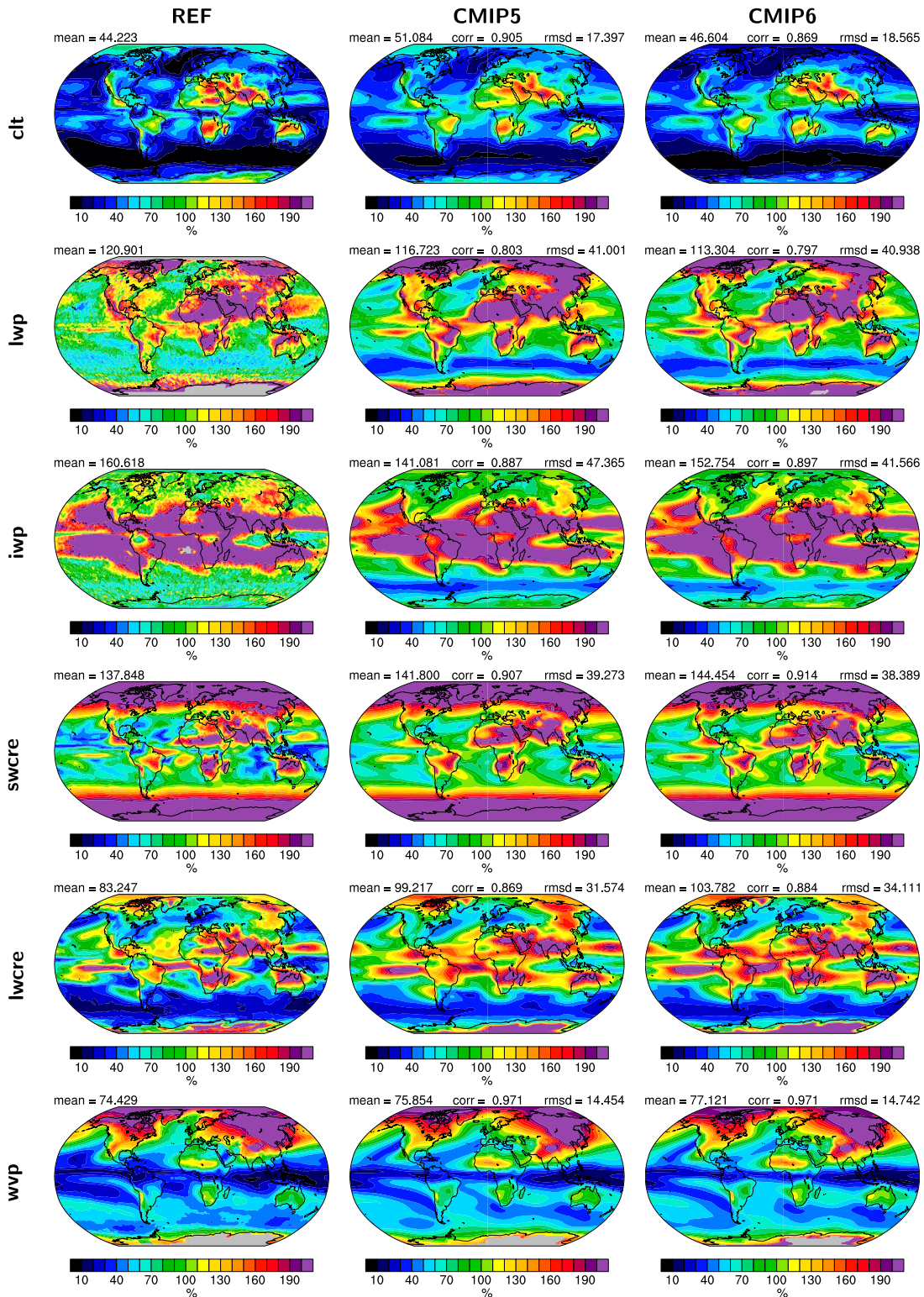


FIG. 7. Relative amplitude of the mean seasonal cycle averaged over 20 years (models) of total cloud cover (clt), liquid water path (lwp), ice water path (iwp), shortwave (swcre) and longwave (lwcre) cloud radiative effects, and water vapor path (wvp), shown from top to bottom, from the (center) CMIP5 and (right) CMIP6 models compared with (left) multi-obs averages [Eq. (5)].



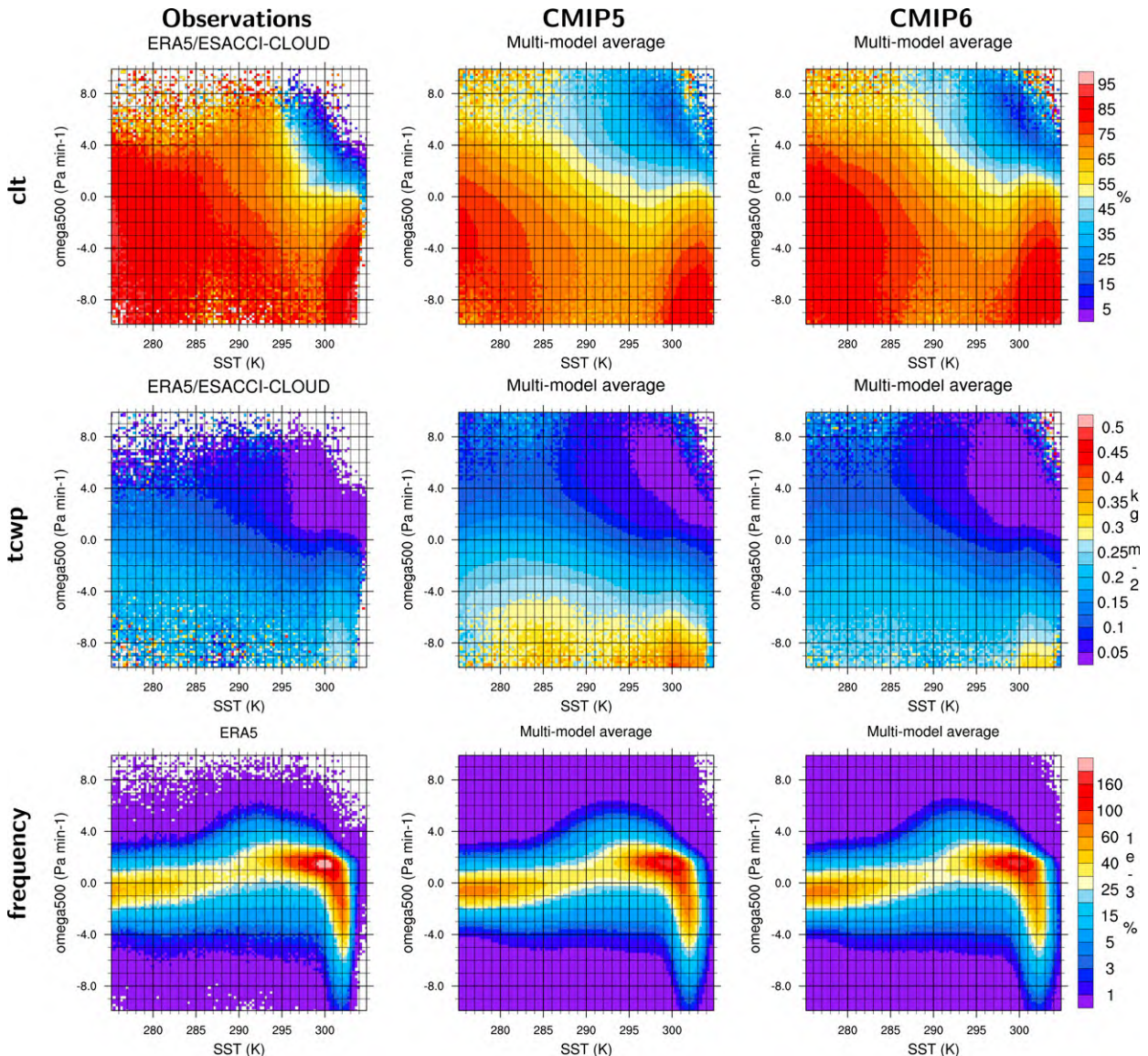


FIG. 8. Two-dimensional distribution of average (top) total cloud cover (clt) and (middle) total cloud water path (tcwp) binned by SST ( $x$  axis) and vertical velocity at 500 hPa ( $\omega_{500}$ ;  $y$  axis) averaged over 20 years and all grid cells over the ocean. (bottom) The relative frequency of occurrence of the SST- $\omega_{500}$  combinations.

with observations showing higher absolute values. The maximum in clt at moderate vertical velocities ( $<4 \text{ Pa min}^{-1}$ ) and low SSTs ( $<280 \text{ K}$ ) is underestimated in both model generations, but again with improved absolute values in CMIP6. Total cloud water path in the extratropical regime is more sensitive to vertical velocity than to SST. While this is qualitatively reproduced by the CMIP models, particularly the CMIP5 MMM strongly overestimates the sensitivity of tcwp to vertical velocity. Here, CMIP6 models are in better agreement with the observations regarding absolute values as well as the gradient in tcwp with vertical velocity.

The relative frequency of occurrence of the different SST- $\omega_{500}$  combinations is shown in the bottom row of Fig. 8. The reanalysis

data show maxima over cold ocean surfaces ( $T < 285 \text{ K}$ ) combined with weakly ascending conditions ( $-2 \text{ Pa min}^{-1} < \omega_{500} < 0$ ) associated particularly with mid- and high latitudes, over warm ocean surfaces ( $290 \text{ K} < T < 300 \text{ K}$ ) combined with weak subsidence ( $0 < \omega_{500} < 3 \text{ Pa min}^{-1}$ ) associated with the subtropical subsidence regions, and over the ocean with  $T > 295 \text{ K}$  combined with weak to strong ascending motion ( $-6 \text{ Pa min}^{-1} < \omega_{500} < 0$ ) associated with the rising branch of the Hadley cell in the tropics. The two multimodel ensembles qualitatively reproduce the shape of this distribution. The biggest relative differences are found in the tropics with the CMIP5 models underestimating the frequency of occurrence of SST- $\omega_{500}$  combinations of  $T \sim 302 \text{ K}$  and  $-5 \text{ Pa min}^{-1} < \omega_{500} < -2 \text{ Pa min}^{-1}$

by about 50%, the CMIP6 models by about 40%. In contrast, for strong ascending motions ( $\omega_{500} < -6 \text{ Pa min}^{-1}$ ) in this temperature range the CMIP5 models overestimate the frequency of occurrence by about roughly 50%–100%. In CMIP6, this overestimation is even slightly more pronounced. This overestimation is probably partly related to the so-called double-ITCZ bias (e.g., [Tian and Dong 2020](#)) in some of the models that can lead to two ascending branches of the Hadley cell in the tropics, which are visible in zonal annual means of the vertical velocity (not shown). As this problem in the zonal annual mean distribution of the vertical velocity is found to be reduced in CMIP6 (not shown), this suggests that, for example, convection might be too strong or too sensitive to high SSTs.

#### e. Regional cloud analyses

##### 1) SOUTHERN OCEAN

As a key region for connecting the surface ocean with the deep ocean, the Southern Ocean plays an important role in climate through its major contribution to the global ocean heat uptake and storage (e.g., [Sallée 2018](#)). The Southern Ocean is known to be a challenging region for ESMs with many models underestimating reflected shortwave radiation at the top of the atmosphere as a result of too few clouds (e.g., [Trenberth and Fasullo 2010](#)). Other models are simulating clouds over the Southern Ocean that are too reflective, which is thought to be caused by an overestimation of cloud ice and a corresponding underestimation of supercooled cloud liquid water (e.g., [Bodas-Salcedo et al. 2012](#); [Lohmann and Neubauer 2018](#)).

To investigate the relationship between cloud radiative effect, total cloud water path, and cloud cover, we apply a diagnostic similar to [Lauer et al. \(2018\)](#), who binned monthly mean cloud radiative effect by cloud cover. Here we apply the same kind of binning also to tcwp. As the incoming solar radiation is at a maximum in austral summer and the effect of errors in model clouds on radiation are the largest, we focus on the months December, January, and February (DJF). [Figure 9](#) shows 20-yr seasonal (DJF) averages of swcre and of tcwp versus total cloud fraction averaged over the Southern Ocean. The Southern Ocean is defined as all ocean grid cells in the latitude belt 30°–65°S. For cloud cover values ranging between about 20% and 75%, the CMIP5 MMM overestimates swcre (i.e., more negative values) compared to the ESACCI-CLOUD observations. In contrast, the CMIP6 MMM shows improved agreement with the observations up to clt values of about 70%. Also, swcre is underestimated (i.e., values not negative enough) in both model ensembles for total cloud cover values over 80%. Comparison of the total cloud water path per total cloud cover value suggests that this improved agreement is at least partially caused by a reduction in tcwp in the CMIP6 models, which is in better agreement with the ESACCI-CLOUD data than in CMIP5. The frequency distributions of simulated clt values, also shown in [Fig. 9](#), indicate that the models tend to overestimate the frequency of values below about 80% and strongly underestimate very high values (above 80%), that are the most frequent in the observations. While this underestimation of very high clt

values is still present in CMIP6, this bias is significantly reduced. The cumulative frequency of occurrence of clt values above 80% is 0.65 in the ESACCI-CLOUD data compared to 0.27 in the CMIP5 MMM and 0.47 in CMIP6.

These findings suggest that the CMIP5 problem of “too few, too bright” clouds over the Southern Ocean (e.g., [Trenberth and Fasullo 2010](#)) is reduced in CMIP6: the frequency of occurrence of high total cloud cover values in CMIP6 is closer to the observations than in CMIP5, and at the same time swcre for a given total cloud cover is in better agreement with ESACCI-CLOUD data than CMIP5 over a wide range of cloud cover values. This is at least partly caused by an improvement in the simulated total cloud water path, which is in closer agreement with ESACCI-CLOUD data in CMIP6 than in CMIP5. The average ratio of simulated to observed tcwp across all total cloud cover bins in CMIP5 is 2.1 while this is improved to an average ratio of 1.4 in CMIP6.

##### 2) SOUTHEASTERN PACIFIC

The southeastern Pacific (SEP) off the west coasts of northern Chile, Peru, and Ecuador is home of the largest and most persistent subtropical stratocumulus regime in the world (e.g., [Bretherton et al. 2010](#)). The stratocumulus clouds are formed and maintained by complex interactions between a relatively cool ocean surface and a moist marine boundary layer capped by a strong inversion with warm and dry air aloft (e.g., [Lin et al. 2014](#)). Because of their low cloud tops, these clouds have a cooling effect on climate by reflecting a large fraction of the incoming solar radiation while contributing only very little to the TOA longwave cloud radiative effect. Despite their importance in climate (e.g., [Wood 2012](#)), stratocumulus clouds have been notoriously difficult to simulate with global climate models (e.g., [Caldwell et al. 2013](#); [Lauer and Hamilton 2013](#)). Common model problems in this region are an underestimation of cloud cover, too high cloud tops, and an unrealistic cloud albedo ([Jian et al. 2020](#)). Particularly the underestimation in cloud cover results in an overly weak shortwave cloud radiative effect, leading to too warm sea surface temperatures ([Lin et al. 2014](#)).

We investigate the relationship between cloud radiative effect, total cloud water path, and cloud cover over the SEP defined as rectangular ocean region covering 85°–95°W, 25°–5°S. As stratocumulus clouds play an important role in climate throughout the year, we focus on annual averages. Associated results are shown in [Fig. 10](#).

The observations show an almost linear increase in tcwp with cloud fraction, which is reproduced by the multimodel ensemble means up to monthly mean clt values of about 80%. Beyond about 80% (CMIP5) and 90% (CMIP6), the intermodel spread indicated by the red shaded areas shows a very large variability of simulated tcwp among the individual models. On average over all total cloud cover bins, the CMIP5 MMM overestimates observed tcwp values by about 75%, the CMIP6 MMM by about 40%. As tcwp is an important factor determining the shortwave cloud radiative effect, this decrease in average tcwp per total cloud cover class in CMIP6 is also reflected in the corresponding swcre. On average over all

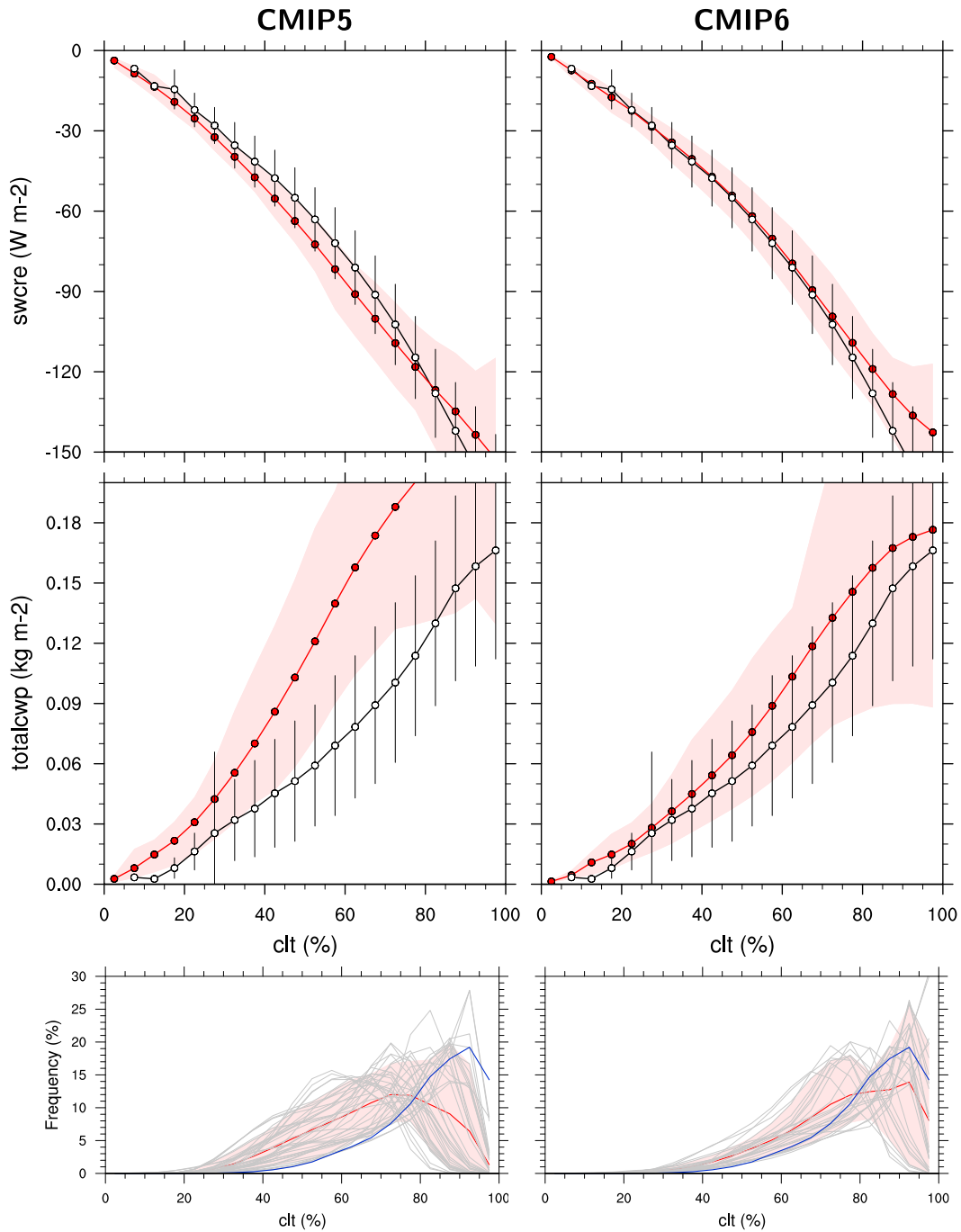


FIG. 9. The 20-yr seasonal average (DJF) of (top) TOA cloud shortwave radiative effect ( $\text{W m}^{-2}$ ) and (middle) total cloud water path (cwp) vs total cloud fraction (clt; %) averaged over the Southern Ocean defined as latitude belt  $30^{\circ}$ – $65^{\circ}$ S (ocean grid cells only). Shown are (left) CMIP5 and (right) CMIP6 multimodel means as red filled circles and lines. The black circles and lines show observational estimates obtained from ESACCI-CLOUD data. The red shaded areas represent the range between the 10th and 90th percentiles of the results from all individual models within each model ensemble. (bottom) The frequency distribution of monthly mean total cloud cover with the red curve showing the multimodel average, the blue curve the ESACCI-CLOUD data, and the thin gray lines the individual models. The red shading shows  $\pm 1$  standard deviation of the intermodel spread.

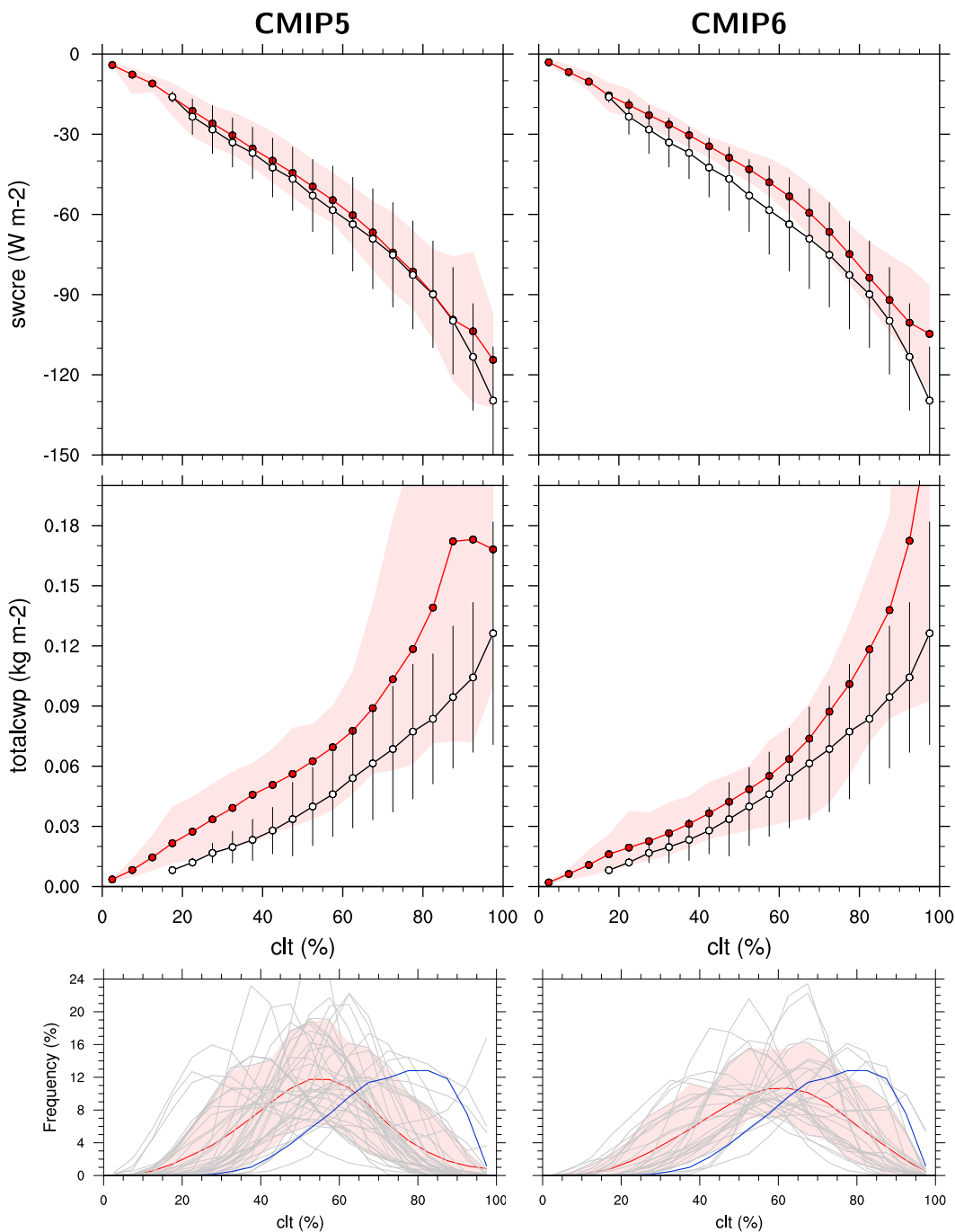


FIG. 10. As in Fig. 9, but for annual averages and over the southeast Pacific (SEP) defined as the region 85°–95°W, 25°–5°S.

total cloud cover bins, amplitudes of swcre are underestimated in the CMIP5 MMM by about 5% whereas this underestimation amounts to about 15% in the CMIP6 MMM. Similar to the Southern Ocean, the frequencies of high monthly mean clt values are increased in CMIP6 compared to CMIP5 and thus in better agreement with observations. For CMIP5, the cumulative frequencies of cloud cover larger than

60% is 37%, in CMIP6 48%, compared to 79% in the observations. This means that the occurrence of stratocumulus clouds in the southeast Pacific is still underestimated in CMIP6 even though there are clear improvements compared to CMIP5. In contrast, swcre for a given total cloud fraction is stronger underestimated in CMIP6 than in CMIP5 which can be partly attributed to a decrease in simulated twcp per cloud



cover class. This agrees with the results from [Jian et al. \(2020\)](#), who also saw that the AMIP6 models underestimate cloud albedos and have more difficulties than AMIP5 models in capturing the linear relationship of cloud fraction and planetary albedo over the marine stratocumulus regions, and especially in the SEP.

### 3) PACIFIC ITCZ

The ITCZ is a low pressure region in the tropics generated by convergence of the trade winds. As an important part of the global circulation, the ITCZ forms the upward branch of the Hadley cell that redistributes heat from the tropics to mid-latitudes. Frequent occurrence of convection, including deep convection, is common in the ITCZ (e.g., [Waliser and Gautier 1993](#)). The ITCZ is a challenging region for global climate models with biases in the simulated clouds and convection. A well-known bias through different generations of global climate models including CMIP6 models is the so-called double ITCZ, an unrealistic zone over the southeastern Pacific and South Atlantic at about 10°S parallel to the equator that somewhat resembles a second ITCZ ([Tian and Dong 2020](#)). To capture the ITCZ over the eastern and western Pacific, we here define the region “Pacific ITCZ” used for the analysis as rectangular region covering 135°E–85°W, 0°–12°N. Associated analysis results are shown in [Fig. 11](#).

When averaged over all total cloud cover bins, the MMMs of CMIP5 and CMIP6 show larger tcwp values by a factor of 1.5 and 1.1, respectively, than the observational reference. Similar to the SEP, amplitudes of swcre are underestimated in CMIP5 by about 3% when averaged over all cloud cover bins, in CMIP6 by about 17%. Again, the reduction in swcre amplitude per cloud cover in CMIP6 can be at least partially attributed to the corresponding decrease in tcwp. In contrast to the Southern Ocean and the SEP, the frequencies of high monthly mean cloud cover values are overestimated in the CMIP6 MMM. The frequency of occurrence of a monthly mean total cloud cover of 60% and larger is 62% in the observations and 70% in CMIP6 compared with 58% in CMIP5. Possible reasons for this could be a more active convection in CMIP6 or longer cloud lifetimes.

## 5. Summary and conclusions

In this study, we compared results from historical simulations from models contributing to CMIP6 with satellite data and with output from the previous phase CMIP5. Whenever possible, reference datasets based on multiple satellite products were used for evaluation of the models in order to allow for an estimation of the observational uncertainties associated with different sensors and with internal variability on a per-pixel basis. A particular focus of this work was to assess how much progress has been made over the recent years by the model ensembles as a whole by comparing the MMMs from CMIP5 and CMIP6 with satellite data and with each other. In total, 35 models contributing to CMIP5 and 28 models contributing to CMIP6 models were used for the analysis.

Our results show that for many cloud parameters investigated here such as total cloud cover, cloud water path, and

cloud radiative effect, the CMIP6 MMM performs slightly better than CMIP5 ensemble mean in terms of mean bias, pattern correlation, and relative root-mean-square deviation. The intermodel spread in CMIP6, however, is not reduced or is even slightly larger than in CMIP5. This is probably a result of the increased model complexity and thus an increased degree of freedom in the models. Compared with *CALIPSO-ICECLOUD* data, the CMIP5/6 models overestimate cloud ice particularly in the lower and middle troposphere. This bias in the cloud ice climatology is reduced in CMIP6. Our results suggest that the overestimation of the cloud ice fraction over a relatively large temperature range (~250–270 K) in the models can probably explain part of the cloud ice bias in the lower and middle troposphere as cloud ice is probably formed in the models at too high temperatures. Another reason for the overestimation in cloud ice could be that cloud ice is not removed quickly enough from the atmosphere, e.g., through sedimentation, resulting in cloud ice being too persistent (i.e., too long residence times in the atmosphere) in this temperature range.

While many known biases such as an underestimation in cloud cover in stratocumulus regions or unrealistic cloud distributions in the tropics due to a double-ITCZ in some models remain a problem in CMIP6, we find that the CMIP5 problem of too few but too reflective clouds over the Southern Ocean is significantly reduced in CMIP6. Here, CMIP6 models simulate more frequently high total cloud cover values, which are in better agreement with satellite data than the frequency distribution of cloud cover from the CMIP5 models. At the same time, the shortwave cloud radiative effect for a given total cloud cover value in CMIP6 is reduced and in better agreement with satellite data compared with CMIP5.

While we found little change in interannual variability and amplitude of seasonal cycle of the investigated variables between CMIP5 and CMIP6, cloud cover and total cloud water path from CMIP6 are in better agreement with observations over a wide range of dynamical regimes characterized by 500-hPa vertical velocity and sea surface temperature. In CMIP6, total cloud cover is found to be increased in regions with high SSTs and strong upward motion (ITCZ) while showing a reduced sensitivity of the total cloud water path with updraft velocity at 500 hPa, which is in better agreement with observations than in CMIP5. In contrast, the sensitivity of total cloud cover to SST in the subtropical stratocumulus regions is underestimated in the models, with a slight improvement in CMIP6 compared to CMIP5. Total cloud water path in the extratropical regime is more sensitive to vertical velocity than to SST and is overestimated by the models. CMIP6 models are found to be in better agreement with the observations regarding absolute values as well as the gradient in tcwp with vertical velocity than the CMIP5 models. The results further show that the models are overestimating the frequency of occurrence of high updraft values in regions with high SSTs by about 50%–100%. In CMIP6, this overestimation is even slightly more pronounced than in CMIP5, and is probably partly related to the double-ITCZ bias in some of the models that can lead to two ascending branches of the Hadley cell in the tropics. As this problem in the zonal annual



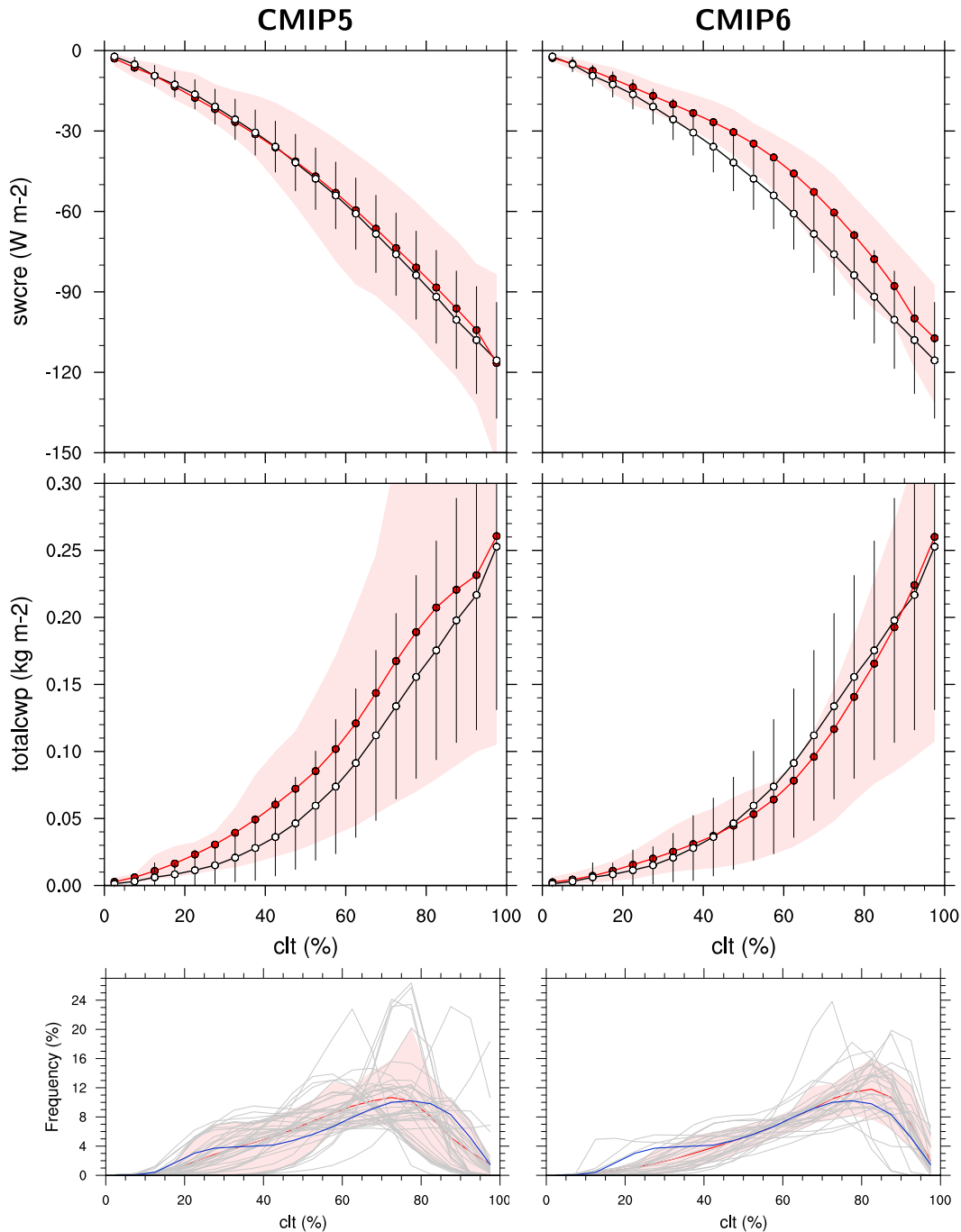


FIG. 11. As in Fig. 9, but for annual averages and averaged over the Pacific ITCZ region defined as  $135^{\circ}\text{E}$ – $85^{\circ}\text{W}$ ,  $0^{\circ}$ – $12^{\circ}\text{N}$ .

mean distribution of the vertical velocity is found to be reduced in CMIP6, this suggests that, for example, convection might be too strong or too sensitive to high SSTs.

We note that despite the increased complexity of the CMIP6 models and the resulting increase in degrees of freedom, the MMM climatology of many investigated cloud parameters has improved in some regions compared with satellite data and

results from CMIP5. Among the most notable improvements is an improved agreement of cloud amount and reflectivity of clouds over the Southern Ocean. Many challenges remain, however, such as simulating clouds in the ITCZ or in the stratocumulus regions. Clouds remain a very challenging task for global climate models and further evaluation studies are therefore important for future improvements of cloud parameterizations in

these models. For this, more widespread application of satellite simulators including covering more instruments (e.g., Eliasson et al. 2020, 2019) in upcoming model intercomparison projects such as CMIP7 would clearly help to evaluate the models and improve our understanding of differences and model deficiencies.

*Acknowledgments.* The development of ESMValTool is supported by several projects. The diagnostic development of ESMValTool v2 for this paper was supported by the ESA Climate Change Initiative Climate Model User Group (ESA CCI CMUG). We acknowledge the World Climate Research Program's (WCRP's) Working Group on Coupled Modelling (WGCM), which is responsible for CMIP, and we thank the climate modelling groups for producing and making available their model output in the framework of ESGF. The CMIP data of this study were replicated and made available for this study by the DKRZ.

This work used resources of the Deutsches Klimarechenzentrum (DKRZ) granted by its Scientific Steering Committee (WLA) under project ID bd0854.

*CALIPSO-GOCCP* data were obtained from the service ClimServ of the Institut Pierre Simon Laplace (IPSL). *CALIPSO-ICECLOUD* data were obtained from the NASA Langley Research Center Atmospheric Science Data Center. The work performed was done using CLARA-AHRR data from the EUMETSAT Satellite Application Facility on Climate Monitoring (CM SAF), source: <https://www.cmsaf.eu/>. *CloudSat-L2* data have been obtained from the *CloudSat* Data Processing Center run by the Cooperative Institute for Research in the Atmosphere (CIRA) at Colorado State University. This manuscript contains modified Copernicus Climate Change Service (2021) information with ERA5 data retrieved from the Climate Data Store (neither the European Commission nor ECMWF is responsible for any use that may be made of the Copernicus information or data it contains). This manuscript contains modified (time averaged) ERA-Interim data, copyright European Centre for Medium-Range Weather Forecasts (ECMWF), source: [www.ecmwf.int](http://www.ecmwf.int); ERA-Interim is published under a Creative Commons Attribution 4.0 International (CC BY 4.0; <https://creativecommons.org/licenses/by/4.0/>). ECMWF does not accept any liability whatsoever for any error or omission in the data, their availability, or for any loss or damage arising from their use. The ESA Climate Change Initiative (CCI) and *Cloud\_cci* project are kindly acknowledged. The combined microwave and near-infrared imager based product COMBI (CM SAF/CCI TCWV-global) was initiated, funded, and provided by the Water Vapour project of the ESA CCI, with contributions from Brockmann Consult, Spectral Earth, Deutscher Wetterdienst, and the EUMETSAT Satellite Climate Facility on Climate Monitoring (CM SAF). The combined MW and NIR product will be owned by EUMETSAT CM SAF and will be released by CM SAF via <https://wui.cmsaf.eu>. The ISCCP H-Series Cloud Properties CDR used in this study was acquired from the NOAA National Centers for Environmental Information (NCEI; formerly NCDC) (<https://www.ncei.noaa.gov>). This CDR was developed by William B. Rossow and

colleagues supported by NOAA's CDR Program. The MAC-LWP dataset is provided by the Goddard Earth Sciences Data and Information Services Center (GES DISC) funded by NASA's Science Mission Directorate (SMD). The AVHRR Clouds Properties-PATMOS-x CDR used in this study was acquired from NOAA's National Climatic Data Center (<http://www.ncdc.noaa.gov>). This CDR was originally developed by A. Heidinger and colleagues for NOAA's CDR Program. The CERES-EBAF data were obtained from the NASA Langley Research Center Atmospheric Science Data Center.

The dataset MODIS used in this work was obtained from the obs4MIPs [<https://esgf-node.llnl.gov/projects/obs4mips/> (accessed on 2 November 2021)] project hosted on the Earth System Grid Federation [<https://esgf.llnl.gov> (accessed on 2 November 2021)].

We thank Sabine Brinkop (DLR) for helpful comments on the manuscript.

*Data availability statement.* CMIP5 and CMIP6 data are available freely and publicly from the Earth System Grid Federation (ESGF). Observations used in the evaluation are detailed in Table 2. The observational datasets are not distributed with the ESMValTool that is restricted to the code as open source software. Observational datasets that are available through the Observations for Model Intercomparisons Project (obs4MIPs; <https://esgf-node.llnl.gov/projects/obs4mips/>) can be downloaded freely from the ESGF and directly used in the ESMValTool. For all other observational datasets, the ESMValTool provides a collection of scripts (NCL and Python) with exact downloading and processing instructions to recreate the datasets used in this publication.

*Code availability statement.* ESMValTool v2 is released under the Apache License, version 2.0. The latest release of ESMValTool v2 is publicly available on Zenodo at <https://doi.org/10.5281/zenodo.3387139>. The source code of the ESMValCore package, which is installed as a dependency of the ESMValTool v2, is also publicly available on Zenodo at <https://doi.org/10.5281/zenodo.3401363>. ESMValTool and ESMValCore are developed on the GitHub repositories available at <https://github.com/ESMValGroup>.

## REFERENCES

- Adachi, Y., and Coauthors, 2013: Basic performance of a new Earth system model of the Meteorological Research Institute (MRI-ESM1). *Pap. Meteor. Geophys.*, **64**, 1–19, <https://doi.org/10.2467/mripapers.64.1>.
- Andersson, A., K. Fennig, C. Klepp, S. Bakan, H. Graßl, and J. Schulz, 2010: The Hamburg Ocean atmosphere parameters and fluxes from satellite data—HOAPS-3. *Earth Syst. Sci. Data*, **2**, 215–234, <https://doi.org/10.5194/essd-2-215-2010>.
- Arora, V. K., and Coauthors, 2011: Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases. *Geophys. Res. Lett.*, **38**, L05805, <https://doi.org/10.1029/2010GL046270>.
- Bentsen, M., and Coauthors, 2013: The Norwegian Earth system model, NorESM1-M—Part 1: Description and basic

- evaluation of the physical climate. *Geosci. Model Dev.*, **6**, 687–720, <https://doi.org/10.5194/gmd-6-687-2013>.
- Bi, D., and Coauthors, 2013: The ACCESS coupled model: Description, control climate and evaluation. *Aust. Meteor. Oceanogr. J.*, **63**, 41–64, <https://doi.org/10.22499/2.6301.004>.
- Bjordal, J., T. Storelvmo, K. Alterskjaer, and T. Carlsen, 2020: Equilibrium climate sensitivity above 5°C plausible due to state-dependent cloud feedback. *Nat. Geosci.*, **13**, 718–721, <https://doi.org/10.1038/s41561-020-00649-1>.
- Bodas-Salcedo, A., and Coauthors, 2011: COSP: Satellite simulation software for model assessment. *Bull. Amer. Meteor. Soc.*, **92**, 1023–1043, <https://doi.org/10.1175/2011BAMS2856.1>.
- , K. D. Williams, P. R. Field, and A. P. Lock, 2012: The surface downwelling solar radiation surplus over the Southern Ocean in the Met Office model: The role of midlatitude cyclone clouds. *J. Climate*, **25**, 7467–7486, <https://doi.org/10.1175/JCLI-D-11-00702.1>.
- , J. P. Mulcahy, T. Andrews, K. D. Williams, M. A. Ringer, P. R. Field, and G. S. Elsaesser, 2019: Strong dependence of atmospheric feedbacks on mixed-phase microphysics and aerosol–cloud interactions in HadGEM3. *J. Adv. Model. Earth Syst.*, **11**, 1735–1758, <https://doi.org/10.1029/2019MS001688>.
- Bony, S., and J. L. Dufresne, 2005: Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. *Geophys. Res. Lett.*, **32**, L20806, <https://doi.org/10.1029/2005GL023851>.
- , —, H. Le Treut, J.-J. Morcrette, and C. Senior, 2004: On dynamic and thermodynamic components of cloud changes. *Climate Dyn.*, **22**, 71–86, <https://doi.org/10.1007/s00382-003-0369-6>.
- , and Coauthors, 2015: Clouds, circulation and climate sensitivity. *Nat. Geosci.*, **8**, 261–268, <https://doi.org/10.1038/ngeo2398>.
- Boucher, O., and Coauthors, 2020: Presentation and evaluation of the IPSL-CM6A-LR climate model. *J. Adv. Model. Earth Syst.*, **12**, e2019MS002010, <https://doi.org/10.1029/2019MS002010>.
- Bretherton, C. S., R. Wood, R. C. George, D. Leon, G. Allen, and X. Zheng, 2010: Southeast Pacific stratocumulus clouds, precipitation and boundary layer structure sampled along 20°S during VOCALS-REx. *Atmos. Chem. Phys.*, **10**, 10639–10654, <https://doi.org/10.5194/acp-10-10639-2010>.
- Caldwell, P. M., Y. Zhang, and S. A. Klein, 2013: CMIP3 subtropical stratocumulus cloud feedback interpreted through a mixed-layer model. *J. Climate*, **26**, 1607–1625, <https://doi.org/10.1175/JCLI-D-12-00188.1>.
- Cameron-Smith, P., J.-F. Lamarque, P. Connell, C. Chuang, and F. Vitt, 2006: Toward an Earth system model: Atmospheric chemistry, coupling, and petascale computing. *J. Phys. Conf. Ser.*, **46**, 343–350, <https://doi.org/10.1088/1742-6596/46/1/048>.
- Cao, J., and Coauthors, 2018: The NUIST Earth System Model (NESM) version 3: Description and preliminary evaluation. *Geosci. Model Dev.*, **11**, 2975–2993, <https://doi.org/10.5194/gmd-11-2975-2018>.
- Chepfer, H., S. Bony, D. Winker, G. Cesana, J. L. Dufresne, P. Minnis, C. J. Stubenrauch, and S. Zeng, 2010: The GCM-Oriented CALIPSO cloud product (CALIPSO-GOCCP). *J. Geophys. Res.*, **115**, D00H16, <https://doi.org/10.1029/2009JD012251>.
- Collins, W. J., and Coauthors, 2011: Development and evaluation of an Earth-system model—HadGEM2. *Geosci. Model Dev.*, **4**, 1051–1075, <https://doi.org/10.5194/gmd-4-1051-2011>.
- Copernicus Climate Change Service, 2017: ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. C3S and CDS, accessed 2 November 2021, <https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation>.
- Danabasoglu, G., and Coauthors, 2020: The Community Earth System Model version 2 (CESM2). *J. Adv. Model. Earth Syst.*, **12**, e2019MS001916, <https://doi.org/10.1029/2019MS001916>.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, <https://doi.org/10.1002/qj.828>.
- Diedrich, H., R. Preusker, R. Lindstrot, and J. Fischer, 2015: Retrieval of daytime total columnar water vapour from MODIS measurements over land surfaces. *Atmos. Meas. Tech.*, **8**, 823–836, <https://doi.org/10.5194/amt-8-823-2015>.
- Donner, L. J., and Coauthors, 2011: The dynamical core, physical parameterizations, and basic simulation characteristics of the atmospheric component AM3 of the GFDL global coupled model CM3. *J. Climate*, **24**, 3484–3519, <https://doi.org/10.1175/2011JCLI3955.1>.
- Dufresne, J.-L., and Coauthors, 2013: Climate change projections using the IPSL-CM5 Earth System model: From CMIP3 to CMIP5. *Climate Dyn.*, **40**, 2123–2165, <https://doi.org/10.1007/s00382-012-1636-1>.
- Dunne, J. P., and Coauthors, 2012: GFDL’s ESM2 global coupled climate–carbon Earth system models. Part I: Physical formulation and baseline simulation characteristics. *J. Climate*, **25**, 6646–6665, <https://doi.org/10.1175/JCLI-D-11-00560.1>.
- , and Coauthors, 2020: The GFDL Earth system model version 4.1 (GFDL-ESM 4.1): Overall coupled model description and simulation characteristics. *J. Adv. Model. Earth Syst.*, **12**, e2019MS002015, <https://doi.org/10.1029/2019MS002015>.
- ECMWF, 2020: ERA5 data documentation. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), accessed 20 July 2020, <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels-monthly-means?tab=overview>.
- Eliasson, S., K. G. Karlsson, E. van Meijgaard, J. F. Meirink, M. Stengel, and U. Willén, 2019: The Cloud\_cci simulator v1.0 for the Cloud\_cci climate data record and its application to a global and a regional climate model. *Geosci. Model Dev.*, **12**, 829–847, <https://doi.org/10.5194/gmd-12-829-2019>.
- , —, and U. Widen, 2020: A simulator for the CLARA-A2 cloud climate data record and its application to assess EC-Earth polar cloudiness. *Geosci. Model Dev.*, **13**, 297–314, <https://doi.org/10.5194/gmd-13-297-2020>.
- Elsaesser, G. S., C. W. O’Dell, M. D. Lebsock, R. Bennartz, T. J. Greenwald, and F. J. Wentz, 2017: The Multisensor Advanced Climatology of Liquid Water Path (MAC-LWP). *J. Climate*, **30**, 10193–10210, <https://doi.org/10.1175/JCLI-D-16-0902.1>.
- Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the Coupled Model Intercomparison Project phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, **9**, 1937–1958, <https://doi.org/10.5194/gmd-9-1937-2016>.
- , and Coauthors, 2020: Earth System Model Evaluation Tool (ESMValTool) v2.0—An extended set of large-scale diagnostics for quasi-operational and comprehensive evaluation of Earth system models in CMIP. *Geosci. Model Dev.*, **13**, 3383–3438, <https://doi.org/10.5194/gmd-13-3383-2020>.
- Fennig, K., M. Schröder, A. Andersson, and R. Hollmann, 2020: A fundamental climate data record of SMMR, SSM/I, and SSMIS brightness temperatures. *Earth Syst. Sci. Data*, **12**, 647–681, <https://doi.org/10.5194/essd-12-647-2020>.
- Fischer, J., R. Preusker, M. Schröder, and O. Danne, 2021: Water Vapour Climate Change Initiative (WV\_cci)—CCI+ Phase 1.

- ATBD Part I: MERIS-MODIS-OLCI L2 Products, ESA, 43 pp., [https://climate.esa.int/media/documents/Water\\_Vapour\\_cci\\_D2.2\\_ATBD\\_Part\\_1\\_-\\_MERIS-MODIS-OLCI\\_L2\\_products\\_v2.1.pdf](https://climate.esa.int/media/documents/Water_Vapour_cci_D2.2_ATBD_Part_1_-_MERIS-MODIS-OLCI_L2_products_v2.1.pdf).
- Flato, G., and Coauthors, 2013: Evaluation of climate models. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 741–866.
- Fogli, P. G., and Coauthors, 2009: INGV-CMCC Carbon (ICC): A carbon cycle Earth system model. CMCC Research Paper No. 61, 31 pp., [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1517282M](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1517282M).
- Frey, W. R., and J. E. Kay, 2018: The influence of extratropical cloud phase and amount feedbacks on climate sensitivity. *Climate Dyn.*, **50**, 3097–3116, <https://doi.org/10.1007/s00382-017-3796-5>.
- Gent, P. R., and Coauthors, 2011: The Community Climate System Model version 4. *J. Climate*, **24**, 4973–4991, <https://doi.org/10.1175/2011JCLI4083.1>.
- Gottelman, A., and Coauthors, 2019: The Whole Atmosphere Community Climate Model version 6 (WACCM6). *J. Geophys. Res. Atmos.*, **124**, 12 380–12 403, <https://doi.org/10.1029/2019JD030943>.
- Giorgetta, M. A., and Coauthors, 2013: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *J. Adv. Model. Earth Syst.*, **5**, 572–597, <https://doi.org/10.1002/jame.20038>.
- Graw, K., A. Andersson, M. Schröder, and K. Fennig, 2017: Algorithm Theoretical baseline document—HOAPS version 4.0. CM SAF, 83 pp., <https://www.cen.uni-hamburg.de/en/icdc/data/atmosphere/docs-atmo/saf-cm-dwd-val-hoaps4-1-2.pdf>.
- Hajima, T., and Coauthors, 2020: Development of the MIROC-ES2L Earth system model and the evaluation of biogeochemical processes and feedbacks. *Geosci. Model Dev.*, **13**, 2197–2244, <https://doi.org/10.5194/gmd-13-2197-2020>.
- Heidinger, A. K., M. J. Foster, A. Walther, and X. Zhao, 2014: The Pathfinder Atmospheres—Extended AVHRR climate dataset. *Bull. Amer. Meteor. Soc.*, **95**, 909–922, <https://doi.org/10.1175/BAMS-D-12-00246.1>.
- Held, I. M., and Coauthors, 2019: Structure and performance of GFDL’s CM4.0 climate model. *J. Adv. Model. Earth Syst.*, **11**, 3691–3727, <https://doi.org/10.1029/2019MS001829>.
- Herckes, P., and J. L. Collett, 2015: Tropospheric chemistry and composition cloud chemistry. *Encyclopedia of Atmospheric Sciences*, 2nd ed. G. R. North, J. Pyle, and F. Zhang, Eds., Academic Press, 218–225.
- Hersbach, H., and Coauthors, 2019: ERA5 monthly averaged data on pressure levels from 1979 to present. Copernicus Climate Change Service (C3S) and Climate Data Store (CDS), accessed 10 March 2022, <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels-monthly-means?tab=overview>.
- , and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, <https://doi.org/10.1002/qj.3803>.
- Hollmann, R., and Coauthors, 2013: The ESA climate change initiative satellite data records for essential climate variables. *Bull. Amer. Meteor. Soc.*, **94**, 1541–1552, <https://doi.org/10.1175/BAMS-D-11-00254.1>.
- Hourdin, F., and Coauthors, 2013: LMDZ5B: The atmospheric component of the IPSL climate model with revised parameterizations for clouds and convection. *Climate Dyn.*, **40**, 2193–2222, <https://doi.org/10.1007/s00382-012-1343-y>.
- Hurrell, J. W., and Coauthors, 2013: The Community Earth System Model: A framework for collaborative research. *Bull. Amer. Meteor. Soc.*, **94**, 1339–1360, <https://doi.org/10.1175/BAMS-D-12-00121.1>.
- Ji, D., and Coauthors, 2014: Description and basic evaluation of Beijing Normal University Earth system model (BNU-ESM) version 1. *Geosci. Model Dev.*, **7**, 2039–2064, <https://doi.org/10.5194/gmd-7-2039-2014>.
- Ji, L., and J. F. Brown, 2017: Effect of NOAA satellite orbital drift on AVHRR-derived phenological metrics. *Int. J. Appl. Earth Obs. Geoinf.*, **62**, 215–223, <https://doi.org/10.1016/j.jag.2017.06.013>.
- Jian, B., J. Li, Y. Zhao, Y. He, J. Wang, and J. Huang, 2020: Evaluation of the CMIP6 planetary albedo climatology using satellite observations. *Climate Dyn.*, **54**, 5145–5161, <https://doi.org/10.1007/s00382-020-05277-4>.
- Karlsson, K.-G., and Coauthors, 2017: CLARA-A2: The second edition of the CM SAF cloud and radiation data record from 34 years of global AVHRR data. *Atmos. Chem. Phys.*, **17**, 5809–5828, <https://doi.org/10.5194/acp-17-5809-2017>.
- , and Coauthors, 2020: CLARA-A2.1: CM SAF cLOUD, Albedo and surface radiation dataset from AVHRR data—Edition 2.1. EUMETSAT SAF on climate monitoring, accessed 28 October 2020, [https://doi.org/10.5676/EUM\\_SAF\\_CM/CLARA\\_AVHRR/V002\\_01](https://doi.org/10.5676/EUM_SAF_CM/CLARA_AVHRR/V002_01).
- Kawai, H., S. Yukimoto, T. Koshiro, N. Oshima, T. Tanaka, H. Yoshimura, and R. Nagasawa, 2019: Significant improvement of cloud representation in the global climate model MRI-ESM2. *Geosci. Model Dev.*, **12**, 2875–2897, <https://doi.org/10.5194/gmd-12-2875-2019>.
- Kuhlbrodt, T., and Coauthors, 2018: The low-resolution version of HadGEM3 GC3.1: Development and evaluation for global climate. *J. Adv. Model. Earth Syst.*, **10**, 2865–2888, <https://doi.org/10.1029/2018MS001370>.
- Lauer, A., and K. Hamilton, 2013: Simulating clouds with global climate models: A comparison of CMIP5 results with CMIP3 and satellite data. *J. Climate*, **26**, 3823–3845, <https://doi.org/10.1175/JCLI-D-12-00451.1>.
- , and Coauthors, 2018: Process-level improvements in CMIP5 models and their impact on tropical variability, the Southern Ocean, and monsoons. *Earth Syst. Dyn.*, **9**, 33–67, <https://doi.org/10.5194/esd-9-33-2018>.
- , and Coauthors, 2020: Earth system model evaluation tool (ESMValTool) v2.0—Diagnostics for emergent constraints and future projections from Earth system models in CMIP. *Geosci. Model Dev.*, **13**, 4205–4228, <https://doi.org/10.5194/gmd-13-4205-2020>.
- Lee, J., and Coauthors, 2020a: Evaluation of the Korea Meteorological Administration Advanced Community Earth-system model (K-ACE). *Asia-Pac. J. Atmos. Sci.*, **56**, 381–395, <https://doi.org/10.1007/s13143-019-00144-7>.
- Lee, W.-L., and Coauthors, 2020b: Taiwan Earth System Model version 1: Description and evaluation of mean state. *Geosci. Model Dev.*, **13**, 3887–3904, <https://doi.org/10.5194/gmd-13-3887-2020>.
- Li, J. F., S. Lee, H.-Y. Ma, G. Stephens, and B. Guan, 2018: Assessment of the cloud liquid water from climate models and reanalysis using satellite observations. *Terr. Atmos. Oceanic Sci.*, **29**, 653–678, <https://doi.org/10.3319/TAO.2018.07.04.01>.
- Li, L., and Coauthors, 2013: The Flexible Global Ocean-Atmosphere-Land System Model, grid-point version 2:



- FGOALS-g2. *Adv. Atmos. Sci.*, **30**, 543–560, <https://doi.org/10.1007/s00376-012-2140-6>.
- , and Coauthors, 2020: The Flexible Global Ocean-Atmosphere-Land System Model grid-point version 3 (FGOALS-g3): Description and evaluation. *J. Adv. Model. Earth Syst.*, **12**, e2019MS002012, <https://doi.org/10.1029/2019MS002012>.
- Lin, J.-L., T. Qian, and T. Shinoda, 2014: Stratocumulus clouds in southeastern Pacific simulated by eight CMIP5-CFMP global climate models. *J. Climate*, **27**, 3000–3022, <https://doi.org/10.1175/JCLI-D-13-00376.1>.
- Lindstrot, R., R. Preusker, H. Diedrich, L. Doppler, R. Bennartz, and J. Fischer, 2012: 1D-Var retrieval of daytime total columnar water vapour from MERIS measurements. *Atmos. Meas. Tech.*, **5**, 631–646, <https://doi.org/10.5194/amt-5-631-2012>.
- Loeb, N. G., B. A. Wielicki, D. R. Doelling, G. L. Smith, D. F. Keyes, S. Kato, N. Manalo-Smith, and T. Wong, 2009: Toward optimal closure of the Earth's top-of-atmosphere radiation budget. *J. Climate*, **22**, 748–766, <https://doi.org/10.1175/2008JCLI2637.1>.
- , J. M. Lyman, G. C. Johnson, R. P. Allan, D. R. Doelling, T. Wong, B. J. Soden, and G. L. Stephens, 2012: Observed changes in top-of-the-atmosphere radiation and upper-ocean heating consistent within uncertainty. *Nat. Geosci.*, **5**, 110–113, <https://doi.org/10.1038/ngeo1375>.
- Lohmann, U., and D. Neubauer, 2018: The importance of mixed-phase and ice clouds for climate sensitivity in the global aerosol-climate model ECHAM6-HAM2. *Atmos. Chem. Phys.*, **18**, 8807–8828, <https://doi.org/10.5194/acp-18-8807-2018>.
- Long, M. C., K. Lindsay, S. Peacock, J. K. Moore, and S. C. Doney, 2013: Twentieth-century oceanic carbon uptake and storage in CESM1(BGC). *J. Climate*, **26**, 6775–6800, <https://doi.org/10.1175/JCLI-D-12-00184.1>.
- Madeleine, J. B., and Coauthors, 2020: Improved representation of clouds in the atmospheric component LMDZ6A of the IPSL-CM6A Earth system model. *J. Adv. Model. Earth Syst.*, **12**, e2020MS002046, <https://doi.org/10.1029/2020MS002046>.
- Marchand, R., G. G. Mace, T. Ackerman, and G. Stephens, 2008: Hydrometeor detection using *Cloudsat*—An Earth-orbiting 94-GHz cloud radar. *J. Atmos. Oceanic Technol.*, **25**, 519–533, <https://doi.org/10.1175/2007JTECHA1006.1>.
- Marchant, B., S. Platnick, K. Meyer, G. T. Arnold, and J. Riedi, 2016: MODIS Collection 6 shortwave-derived cloud phase classification algorithm and comparisons with CALIOP. *Atmos. Meas. Tech.*, **9**, 1587–1599, <https://doi.org/10.5194/amt-9-1587-2016>.
- Mauritsen, T., and Coauthors, 2019: Developments in the MPI-M Earth system model version 1.2 (MPI-ESM1.2) and its response to increasing CO<sub>2</sub>. *J. Adv. Model. Earth Syst.*, **11**, 998–1038, <https://doi.org/10.1029/2018MS001400>.
- Müller, W. A., and Coauthors, 2018: A higher-resolution version of the Max Planck Institute Earth system model (MPI-ESM1.2-HR). *J. Adv. Model. Earth Syst.*, **10**, 1383–1413, <https://doi.org/10.1029/2017MS001217>.
- NASA/LARC/SD/ASDC, 2018: CALIPSO Lidar Level 3 ice cloud data, standard V1-00. NASA, accessed 8 April 2020, <https://catalog.data.gov/dataset/calipso-lidar-level-3-ice-cloud-data-standard-v1-00>.
- , 2019: CERES Energy Balanced and Filled (EBAF) TOA and surface monthly means data in netCDF edition 4.1. NASA, accessed 1 July 2022, [https://doi.org/10.5067/TERRA-AQUA/CERES/EBAF\\_L3B.004.1](https://doi.org/10.5067/TERRA-AQUA/CERES/EBAF_L3B.004.1).
- O'Dell, C. W., F. J. Wentz, and R. Bennartz, 2008: Cloud liquid water path from satellite-based passive microwave observations: A new climatology over the global oceans. *J. Climate*, **21**, 1721–1739, <https://doi.org/10.1175/2007JCLI1958.1>.
- Park, S., J. Shin, S. Kim, E. Oh, and Y. Kim, 2019: Global climate simulated by the Seoul National University Atmosphere Model version 0 with a unified convection scheme (SAM0-UNICON). *J. Climate*, **32**, 2917–2949, <https://doi.org/10.1175/JCLI-D-18-0796.1>.
- PCMDI, 2019: CMIP6-cmor-tables. GitHub, <https://github.com/PCMDI/cmip6-cmor-tables/tree/master/Tables>.
- Platnick, S., M. D. King, S. A. Ackerman, W. P. Menzel, B. A. Baum, J. C. Riedi, and R. A. Frey, 2003: The MODIS cloud products: Algorithms and examples from Terra. *IEEE Trans. Geosci. Remote Sens.*, **41**, 459–473, <https://doi.org/10.1109/TGRS.2002.808301>.
- , and Coauthors, 2015: MODIS atmosphere L3 monthly product. NASA MODIS adaptive processing system, Goddard Space Flight Center, [http://doi.org/10.5067/MODIS/MOD08\\_M3.006](http://doi.org/10.5067/MODIS/MOD08_M3.006).
- , and Coauthors, 2017: The MODIS cloud optical and microphysical products: Collection 6 updates and examples from Terra and Aqua. *IEEE Trans. Geosci. Remote Sens.*, **55**, 502–525, <https://doi.org/10.1109/TGRS.2016.2610522>.
- Poulsen, C. A., and Coauthors, 2020: Cloud\_cci ATSR-2 and AATSR data set version 3: A 17-year climatology of global cloud and radiation properties. *Earth Syst. Sci. Data*, **12**, 2121–2135, <https://doi.org/10.5194/essd-12-2121-2020>.
- Qiao, F., Z. Song, Y. Bao, Y. Song, Q. Shu, C. Huang, and W. Zhao, 2013: Development and evaluation of an Earth system model with surface gravity waves. *J. Geophys. Res. Oceans*, **118**, 4514–4524, <https://doi.org/10.1002/jgrc.20327>.
- Rackow, T., H. F. Goessling, T. Jung, D. Sidorenko, T. Semmler, D. Barbi, and D. Handorf, 2018: Towards multi-resolution global climate modeling with ECHAM6-FESOM. Part II: Climate variability. *Climate Dyn.*, **50**, 2369–2394, <https://doi.org/10.1007/s00382-016-3192-6>.
- Ramanathan, V., R. D. Cess, E. F. Harrison, P. Minnis, B. R. Barkstrom, E. Ahmad, and D. Hartmann, 1989: Cloud-radiative forcing and climate: Results from the Earth Radiation Budget Experiment. *Science*, **243**, 57–63, <https://doi.org/10.1126/science.243.4887.57>.
- Righi, M., and Coauthors, 2020: Earth System Model Evaluation Tool (ESMValTool) v2.0—Technical overview. *Geosci. Model Dev.*, **13**, 1179–1199, <https://doi.org/10.5194/gmd-13-1179-2020>.
- Rind, D., and Coauthors, 2020: GISS model E2.2: A climate model optimized for the middle atmosphere—Model structure, climatology, variability, and climate sensitivity. *J. Geophys. Res. Atmos.*, **125**, e2019JD032204, <https://doi.org/10.1029/2019JD032204>.
- Rong, X., and Coauthors, 2018: The CAMS climate system model and a basic evaluation of its climatology and climate variability simulation. *J. Meteor. Res.*, **32**, 839–861, <https://doi.org/10.1007/s13351-018-8058-x>.
- Rossov, W. B., and NOAA CDR Program, 2017: NOAA Climate Data Record (CDR) of Cloud Properties-ISCCP, version 1, NOAA, National Centers for Environmental Information (NCEI), accessed 8 November 2019, <https://doi.org/10.7289/V5QZ281S>.
- Rotstayn, L. D., M. A. Collier, M. R. Dix, Y. Feng, H. B. Gordon, S. P. O'Farrell, I. N. Smith, and J. Syktus, 2010: Improved simulation of Australian climate and ENSO-related rainfall variability in a global climate model with an interactive aerosol treatment. *Int. J. Climatol.*, **30**, 1067–1088, <https://doi.org/10.1002/joc.1952>.



- Sakamoto, T. T., and Coauthors, 2012: MIROC4h—A new high-resolution atmosphere–ocean coupled general circulation model. *J. Meteor. Soc. Japan*, **90**, 325–359, <https://doi.org/10.2151/jmsj.2012-301>.
- Sallée, J. B., 2018: Southern Ocean warming. *Oceanography*, **31**, 52–62, <https://doi.org/10.5670/oceanog.2018.215>.
- Schmidt, G. A., and Coauthors, 2006: Present-day atmospheric simulations using GISS ModelE: Comparison to in situ, satellite, and reanalysis data. *J. Climate*, **19**, 153–192, <https://doi.org/10.1175/JCLI3612.1>.
- Scoccimarro, E., and Coauthors, 2011: Effects of tropical cyclones on ocean heat transport in a high-resolution coupled general circulation model. *J. Climate*, **24**, 4368–4384, <https://doi.org/10.1175/2011JCLI4104.1>.
- Séférian, R., and Coauthors, 2019: Evaluation of CNRM Earth system model, CNRM-ESM2-1: Role of Earth system processes in present-day and future climate. *J. Adv. Model. Earth Syst.*, **11**, 4182–4227, <https://doi.org/10.1029/2019MS001791>.
- Seland, Ø., and Coauthors, 2020: Overview of the Norwegian Earth System Model (NorESM2) and key climate response of CMIP6 DECK, historical, and scenario simulations. *Geosci. Model Dev.*, **13**, 6165–6200, <https://doi.org/10.5194/gmd-13-6165-2020>.
- Sellar, A. A., and Coauthors, 2019: UKESM1: Description and evaluation of the UK Earth System Model. *J. Adv. Model. Earth Syst.*, **11**, 4513–4558, <https://doi.org/10.1029/2019MS001739>.
- Sidorenko, D., and Coauthors, 2015: Towards multi-resolution global climate modeling with ECHAM6-FESOM. Part I: Model formulation and mean climate. *Climate Dyn.*, **44**, 757–780, <https://doi.org/10.1007/s00382-014-2290-6>.
- Stengel, M., C. Schlundt, S. Stapelberg, O. Sus, S. Eliasson, U. Willen, and J. F. Meirink, 2018: Comparing ERA-Interim clouds with satellite observations using a simplified satellite simulator. *Atmos. Chem. Phys.*, **18**, 17601–17614, <https://doi.org/10.5194/acp-18-17601-2018>.
- , and Coauthors, 2020: Cloud\_cci Advanced Very High Resolution Radiometer post meridiem (AVHRR-PM) dataset version 3: 35-year climatology of global cloud and radiation properties. *Earth Syst. Sci. Data*, **12**, 41–60, <https://doi.org/10.5194/essd-12-41-2020>.
- Stephens, G. L., and Coauthors, 2002: The CloudSat mission and the A-Train: A new dimension of space-based observations of clouds and precipitation. *Bull. Amer. Meteor. Soc.*, **83**, 1771–1790, <https://doi.org/10.1175/BAMS-83-12-1771>.
- , and Coauthors, 2018: *CloudSat* and *CALIPSO* within the A-Train: Ten years of actively observing the Earth system. *Bull. Amer. Meteor. Soc.*, **99**, 569–581, <https://doi.org/10.1175/BAMS-D-16-0324.1>.
- Stevens, B., and Coauthors, 2013: Atmospheric component of the MPI-M Earth system model: ECHAM6. *J. Adv. Model. Earth Syst.*, **5**, 146–172, <https://doi.org/10.1002/jame.20015>.
- Tatebe, H., and Coauthors, 2019: Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geosci. Model Dev.*, **12**, 2727–2765, <https://doi.org/10.5194/gmd-12-2727-2019>.
- Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, **106**, 7183–7192, <https://doi.org/10.1029/2000JD900719>.
- , C. Doutriaux, and J.-Y. Peterschmitt, 2006: Climate Model Output Rewriter (CMOR). [https://pcmdi.github.io/cmor-site/media/pdf/cmor\\_users\\_guide.pdf](https://pcmdi.github.io/cmor-site/media/pdf/cmor_users_guide.pdf).
- , R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, <https://doi.org/10.1175/BAMS-D-11-00094.1>.
- Tian, B. J., and X. Y. Dong, 2020: The double-ITCZ bias in CMIP3, CMIP5, and CMIP6 models based on annual mean precipitation. *Geophys. Res. Lett.*, **47**, e2020GL087232, <https://doi.org/10.1029/2020GL087232>.
- , and T. Hearty, 2020: Estimating and removing the sampling biases of the AIRS Obs4MIPs V2 data. *Earth Space Sci.*, **7**, e2020EA001438, <https://doi.org/10.1029/2020EA001438>.
- , E. J. Fetzer, B. H. Kahn, J. Teixeira, E. Manning, and T. Hearty, 2013: Evaluating CMIP5 models using AIRS tropospheric air temperature and specific humidity climatology. *J. Geophys. Res. Atmos.*, **118**, 114–134, <https://doi.org/10.1029/2012JD018607>.
- Trenberth, K. E., and J. T. Fasullo, 2010: Simulation of present-day and twenty-first-century energy budgets of the southern oceans. *J. Climate*, **23**, 440–454, <https://doi.org/10.1175/2009JCLI3152.1>.
- Vichi, M., E. Manzini, P. G. Fogli, A. Alessandri, L. Patara, E. Scoccimarro, S. Masina, and A. Navarra, 2011: Global and regional ocean carbon uptake and climate change: Sensitivity to a substantial mitigation scenario. *Climate Dyn.*, **37**, 1929–1947, <https://doi.org/10.1007/s00382-011-1079-0>.
- Voldoire, A., and Coauthors, 2019: Evaluation of CMIP6 DECK experiments with CNRM-CM6-1. *J. Adv. Model. Earth Syst.*, **11**, 2177–2213, <https://doi.org/10.1029/2019MS001683>.
- Volodin, E. M., N. A. Dianskii, and A. V. Gusev, 2010: Simulating present-day climate with the INMCM4.0 coupled model of the atmospheric and oceanic general circulations. *Izv., Atmos. Ocean. Phys.*, **46**, 414–431, <https://doi.org/10.1134/S000143381004002X>.
- Waliser, D. E., and C. Gautier, 1993: A satellite-derived climatology of the ITCZ. *J. Climate*, **6**, 2162–2174, [https://doi.org/10.1175/1520-0442\(1993\)006<2162:ASDCOT>2.0.CO;2](https://doi.org/10.1175/1520-0442(1993)006<2162:ASDCOT>2.0.CO;2).
- , and Coauthors, 2009: Cloud ice: A climate model challenge with signs and expectations of progress. *J. Geophys. Res.*, **114**, D00A21, <https://doi.org/10.1029/2008JD010015>.
- Watanabe, M., and Coauthors, 2010: Improved climate simulation by MIROC5: Mean states, variability, and climate sensitivity. *J. Climate*, **23**, 6312–6335, <https://doi.org/10.1175/2010JCLI3679.1>.
- Watanabe, S., and Coauthors, 2011: MIROC-ESM 2010: Model description and basic results of CMIP5-20c3m experiments. *Geosci. Model Dev.*, **4**, 845–872, <https://doi.org/10.5194/gmd-4-845-2011>.
- Weigel, K., and Coauthors, 2021: Earth System Model Evaluation Tool (ESMValTool) v2.0—Diagnostics for extreme events, regional and impact evaluation, and analysis of Earth system models in CMIP. *Geosci. Model Dev.*, **14**, 3159–3184, <https://doi.org/10.5194/gmd-14-3159-2021>.
- Williams, K. D., M. A. Ringer, and C. A. Senior, 2003: Evaluating the cloud response to climate change and current climate variability. *Climate Dyn.*, **20**, 705–721, <https://doi.org/10.1007/s00382-002-0303-3>.
- Wood, R., 2012: Stratocumulus clouds. *Mon. Wea. Rev.*, **140**, 2373–2423, <https://doi.org/10.1175/MWR-D-11-00121.1>.
- Wu, T. W., 2012: A mass-flux cumulus parameterization scheme for large-scale models: Description and test with observations. *Climate Dyn.*, **38**, 725–744, <https://doi.org/10.1007/s00382-011-0995-3>.
- , and Coauthors, 2010: The Beijing Climate Center atmospheric general circulation model: Description and its

- performance for the present-day climate. *Climate Dyn.*, **34**, 123, <https://doi.org/10.1007/s00382-008-0487-2>.
- , and Coauthors, 2019: The Beijing Climate Center Climate System Model (BCC-CSM): The main progress from CMIP5 to CMIP6. *Geosci. Model Dev.*, **12**, 1573–1600, <https://doi.org/10.5194/gmd-12-1573-2019>.
- Young, A. H., K. R. Knapp, A. Inamdar, W. Hankins, and W. B. Rossow, 2018: The International Satellite Cloud Climatology Project H-series climate data record product. *Earth Syst. Sci. Data*, **10**, 583–593, <https://doi.org/10.5194/essd-10-583-2018>.
- Yukimoto, S., and Coauthors, 2012: A new global climate model of the Meteorological Research Institute: MRI-CGCM3-model description and basic performance. *J. Meteor. Soc. Japan*, **90A**, 23–64, <https://doi.org/10.2151/jmsj.2012-A02>.
- , and Coauthors, 2019: The Meteorological Research Institute Earth system model version 2.0, MRI-ESM2.0: Description and basic evaluation of the physical component. *J. Meteor. Soc. Japan*, **97**, 931–965, <https://doi.org/10.2151/jmsj.2019-051>.
- Zelinka, M. D., T. A. Myers, D. T. McCoy, S. Po-Chedley, P. M. Caldwell, P. Ceppi, S. A. Klein, and K. E. Taylor, 2020: Causes of higher climate sensitivity in CMIP6 models. *Geophys. Res. Lett.*, **47**, e2019GL085782, <https://doi.org/10.1029/2019GL085782>.