JGR Atmospheres

RESEARCH ARTICLE
10.1029/2019JD032321

Key Points:

- Temperature, water vapor, and zonal wind speed show improvements in CMIP6 with amplitudes of many long-standing biases smaller than CMIP3/5
- High-resolution models show significant improvements in their historical CMIP6 simulations for temperature and precipitation mean biases
- Spread in effective climate sensitivity in CMIP6 is larger than in previous phases

Correspondence to:
L. Bock,
lisa.bock@dlr.de

Citation:
Bock, L., Lauer, A., Schlund, M., Barreiro, M., Bellouin, N., Jones, C., et al. (2020). Quantifying progress across different CMIP phases with the ESMValTool. *Journal of Geophysical Research: Atmospheres*, *125*, e2019JD032321. https://doi.org/10.1029/2019JD032321

Received 29 DEC 2019
Accepted 7 OCT 2020
Accepted article online 18 OCT 2020

Quantifying Progress Across Different CMIP Phases With the ESMValTool

L. Bock[1], A. Lauer[1], M. Schlund[1], M. Barreiro[2], N. Bellouin[3], C. Jones[4], G. A. Meehl[5], V. Predoi[6], M. J. Roberts[7], and V. Eyring[1,8]

[1]Deutsches Zentrum für Luft-und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany, [2]Departamento de Ciencias de la Atmósfera, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay, [3]Department of Meteorology, University of Reading, Reading, UK, [4]National Centre for Atmospheric Science, University of Leeds, Leeds, UK, [5]National Center for Atmospheric Research, Boulder, CO, USA, [6]NCAS Computational Modelling Services (CMS), University of Reading, Reading, UK, [7]Met Office Hadley Centre, Exeter, UK, [8]Institute of Environmental Physics (IUP), University of Bremen, Bremen, Germany

Abstract More than 40 model groups worldwide are participating in the Coupled Model Intercomparison Project Phase 6 (CMIP6), providing a new and rich source of information to better understand past, present, and future climate change. Here, we use the Earth System Model Evaluation Tool (ESMValTool) to assess the performance of the CMIP6 ensemble compared to the previous generations CMIP3 and CMIP5. While CMIP5 models did not capture the observed pause in the increase in global mean surface temperature between 1998 and 2013, the historical CMIP6 simulations agree well with the observed recent temperature increase, but some models have difficulties in reproducing the observed global mean surface temperature record of the second half of the twentieth century. While systematic biases in annual mean surface temperature and precipitation remain in the CMIP6 multimodel mean, individual models and high-resolution versions of the models show significant reductions in many long-standing biases. Some improvements are also found in the vertical temperature, water vapor, and zonal wind speed distributions, and root-mean-square errors for selected fields are generally smaller with reduced intermodel spread and higher average skill in the correlation patterns relative to observations. An emerging property of the CMIP6 ensemble is a higher effective climate sensitivity with an increased range between 2.3 and 5.6 K. A possible reason for this increase in some models is improvements in cloud representation resulting in stronger shortwave cloud feedbacks than in their predecessor versions.

1. Introduction

Climate model simulations are coordinated as part of the World Climate Research Programme's (WCRP) Coupled Model Intercomparison Project (CMIP) since the early 1990s (Eyring, Bony, et al., 2016; Meehl et al., 1997, 2000, 2005, 2007; Taylor et al., 2012). The main objective of CMIP is to better understand past, present, and future climate variability and change arising from natural, unforced variability and in response to changes in radiative forcing in a multimodel context. Model simulations are defined and carried out by the participating modeling groups under common forcings and forcing scenarios. CMIP defines not only common model simulations but also aims at making a wide range of model output available to the research community in order to better learn from a large model ensemble. In this sense, CMIP3 marked a paradigm shift in the climate science community by making model output from state-of-the-art climate change simulations broadly accessible (Meehl et al., 2007). CMIP model simulations and associated publications analyzing the multimodel data set constitute the state of the climate science and thus have been assessed by the Intergovernmental Panel on Climate Change (IPCC) Assessment Reports. CMIP3 supported the Fourth Assessment Report (AR4) (Solomon et al., 2007).

The next phase of CMIP was CMIP5 with an integrated set of experiments (Taylor et al., 2012) and was used in numerous peer-reviewed studies as well as providing the basis for the IPCC Fifth Assessment Report (AR5) (Stocker et al., 2013). Flato et al. (2013) pointed out that there were significant variations in skill across the CMIP5 ensemble when measured against reanalyses and observations, and that some systematic biases remained over several generations of CMIP model ensembles. The current phase of CMIP, CMIP6 (Eyring, Bony, et al., 2016), therefore includes the quantification and understanding of systematic biases as one of its three central scientific questions. There are 23 CMIP6-Endorsed Model Intercomparison Projects (MIPs) that define an additional set of simulations targeting other specific scientific questions. For example, the new High-Resolution Model Intercomparison Project (HighResMIP, Haarsma et al., 2016) that we also evaluate here assesses the robustness of improvements in the representation of important climate processes using weather-resolving global model resolutions (~25 km or finer).

An important question to answer is how these new simulations compare to previous generations of CMIP ensembles, and whether systematic biases detected earlier are reduced or remain. A thorough assessment of the models' skill in reproducing observed past and present climate is also an essential prerequisite to assess and interpret the results from model projections (Eyring et al., 2019). Known systematic biases include (i) a too strong intertropical convergence zone (ITCZ) in the Southern Hemisphere (SH) which often persists through the seasonal cycle; problems simulating the Walker circulation and the associated dry Amazon bias also seen in many models; (ii) poor simulation of tropical and subtropical low-level clouds, particularly the persistent stratocumulus decks over the eastern parts of ocean basins, which are related to too warm sea surface temperatures (SSTs); (iii) an overly deep tropical thermocline; (iv) the tendency to simulate land surfaces too warm and too dry during summertime, and (v) the northward shift in the position of the SH atmospheric jet which leads to poor simulation of the surface wind stress over the Southern Ocean and to errors in the vertical structure of the water masses in the Southern Ocean (Stouffer et al., 2017).

Even though not sufficient, a systematic evaluation of models results by comparisons with observations and reanalysis data is commonly seen as an important prerequisite to building up confidence in the models' future climate projections (Flato et al., 2013). This more general assessment of model performance is complimented with approaches that use observations to constrain the uncertainty in multimodel projections or feedbacks with observations (Eyring et al., 2019). This aspect is not covered here and requires further analysis on how the different generations of CMIP ensembles compare in this respect. Some initial studies exist, for example, Schlund et al. (2020) compare emergent constraints on effective climate sensitivity (ECS) between CMIP5 and CMIP6 and Tokarska et al. (2020) constrain future warming based on the ability of the models to reproduce past temperature trends.

Many CMIP6 modeling groups already reported improvements in their model's ability to simulate past and present-day climate compared to their CMIP5 predecessor versions (Andrews et al., 2019; Danabasoglu et al., 2020; Gettelman et al., 2019; Mulcahy et al., 2018; Swart et al., 2019). Typically, model developments consist of including more detailed Earth system processes as well as improvements in existing parameterizations or higher horizontal and vertical resolution. However, a systematic assessment of the CMIP6 ensemble in comparison to previous generations is still missing. In order to evaluate how well the models perform, we compare the performance of the CMIP3, CMIP5, and CMIP6 ensemble by evaluating the historical simulations forced by common boundary conditions in each phase with observations or reanalysis data. We apply the Earth System Model Evaluation Tool (ESMValTool) Version 2 (Eyring et al., 2020; Righi et al., 2020) for a consistent assessment of the CMIP ensemble across phases. The ESMValTool is a community-developed open-source diagnostic and performance metrics tool to evaluate Earth system models (ESMs) with observations and reanalysis data.

In section 2, the model ensembles and observations used in this study as well as the ESMValTool are described. The surface temperature record of the three model ensembles CMIP3, CMIP5, and CMIP6 is discussed in section 3 and the multimodel mean biases of some important climate variables such as temperature, precipitation and selected meteorological variables such as zonal wind are compared across the model ensembles in section 4. An overview on the general model performance in comparison with observations by applying performance metrics and pattern correlations are shown in section 5. In section 6 we discuss the high ECS in some CMIP6 models compared with results from previous phases of CMIP and close with a summary in section 7.

2. Models and Observations

In this study we use model simulations from CMIP Phases 3, 5 and 6, organized by the WCRP CMIP Panel under the auspices of the Working Group on Coupled Modelling (WGCM). The model data (see Tables 1–3) from CMIP3, CMIP5, and CMIP6 are freely available on servers of the Earth System Grid Federation (ESGF), which is an international collaboration that manages the decentralized database of CMIP output. In order to

Table 1
CMIP6 Models Used in This Study

Model(s)	Institute	Reference(s)
ACCESS-CM2 ACCESS-ESM 1-5	Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australian Research Council Centre of Excellence for Climate System Science (ARCCSS)	
AWI-CM-1-1-MR AWI-ESM-1-1-LR	Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Germany	Rackow et al. (2018); Sidorenko et al. (2015)
BCC-CSM2-MR	Beijing Climate Center, China	Wu et al. (2019)
BCC-ESM 1	Meteorological Administration, China	
CAMS-CSM1-0	Chinese Academy of Meteorological Sciences, China	Rong et al. (2018)
CanESM5 CanESM5-CanOE	Canadian Center for Atmospheric Research, Canada	Swart et al. (2019)
CESM2 CESM2-FV2 CESM2-WACCM CESM2-WACCM-FV2	National Science Foundation (NSF), Department of Energy (DOE), National Center for Atmospheric Research (NCAR), USA	
CNRM-CM6-1 CNRM-CM6-HR CNRM-ESM 2-1	Météo-France/Centre National de Recherches Météorologiques (CNRM) and Centre Européen de Recherches et de Formation Avancée en Calcul Scientifique (CERFACS), France	Séférian et al. (2019); Voldoire et al. (2019)
E3SM-1-0 E3SM-1-1 E3SM-1-1-ECA	E3SM-Project, USA	Golaz et al. (2019)
EC-Earth3-Veg	EC-Earth consortium, Europe	Wyser et al. (2019)
FGOALS-f3-L FGOALS-g3	Chinese Academy of Meteorological Sciences, China	
FIO-ESM-2-0	First Institute of Oceanography, Ministry of Natural Resources, China (FIO), Qingdao National Laboratory for Marine Science and Technology, China (QNLM)	Song et al. (2019)
GFDL-CM4 GFDL-ESM 4	National Oceanic and Atmospheric Administration (NOAA) /Geophysical Fluid Dynamics Laboratory (GFDL), USA	Dunne et al. (2019); Held et al. (2019)
GISS-E2-1-G GISS-E2-1-G-CC GISS-E2-1-H	National Aeronautics and Space Administration (NASA), Goddard Institute for Space Studies (GISS), USA	
HadGEM3-GC31-LL HadGEM3-GC31-MM	Met Office Hadley Centre, UK	Kuhlbrodt et al. (2018); Williams et al. (2018)
INM-CM4-8 INM-CM5-0	Institute for Numerical Mathematics, Russian Academy of Science, Russia	Volodin et al. (2017a); Volodin et al. (2017b); Volodin et al. (2018)
IPSL-CM6A-LR	L'Institut Pierre-Simon Laplace (IPSL), France	Boucher et al. (2020)
KACE-1-0-G	National Institute of Meteorological Sciences/Korea Meteorological Administration, Climate Research Division, Republic of Korea	Lee et al. (2019)
MCM-UA-1-0	Department of Geosciences, University of Arizona, USA	Delworth et al. (2002)
MIROC6 MIROC-ES2L	Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Atmosphere and Ocean Research Institute (AORI), University of Tokyo, and National Institute for Environmental Studies (NIES), Japan	Hajima et al. (2020); Tatebe et al. (2019)
MPI-ESM-1-2-HAM	HAMMOZ-Consortium: ETH Zurich, Switzerland; Max Planck Institut fur Meteorologie, Germany; Forschungszentrum Julich, Germany; University of Oxford, UK; Finnish Meteorological Institute, Finland; Leibniz Institute for Tropospheric Research, Germany; Center for Climate Systems Modeling (C2SM) at ETH Zurich, Switzerland	Neubauer et al. (2019)
MPI-ESM 1-2-HR MPI-ESM 1-2-LR	Max Planck Institute, Germany	Mauritsen et al. (2019); Muller et al. (2018)
MRI-ESM 2-0	Meteorological Research Institute (MRI), Japan	Yukimoto et al. (2019)
NESM3	Nanjing University of Information Science and Technology, China	Cao et al. (2018)
NorCPM1 NorESM2-LM NorESM2-MM	NorESM Climate modeling Consortium, Norway	Bentsen et al. (2013)
SAM0-UNICON	Seoul National University, Republic of Korea	Park et al. (2019)
TaiESM1	Research Center for Environmental Changes, Academia Sinica, Taiwan	
UKESM1-0-LL	Met Office Hadley Centre, UK	Sellar et al. (2019)

Table 2
CMIP5 Models Used in This Study

Model(s)	Institute	Reference(s)
ACCESS1.0 ACCESS1.3	Australian Commonwealth Scientific and Industrial Research Organization (CSIRO) Marine and Atmospheric Research, Bureau of Meteorology (BoM), Australia	Bi et al. (2013)
BCC-CSM1.1 BCC-CSM1.1-M	Beijing Climate Center, China Meteorological Administration, China	Wu et al. (2013); Zhou et al. (2014)
BNU-ESM	Beijing Normal University (BNU), China	Ji et al. (2014)
CanCM4 CanESM2	Canadian Center for Atmospheric Research, Canada	Arora et al. (2011)
CCSM4	National Center for Atmospheric Research (NCAR), USA	Gent et al. (2011)
CESM1-BGC CESM1-CAM5 CESM1-CAM5-1-FV2 CESM1-FASTCHEM CESM1-WACCM	National Science Foundation (NSF), Department of Energy (DOE), National Center for Atmospheric Research (NCAR), USA	Marsh et al. (2013); Meehl et al. (2013)
CMCC-CESM CMCC-CM CMCC-CMS	Centro Euro-Mediterraneo per I Cambiamenti Climatici (CMCC), Italy	(Fogli et al., 2009)
CNRM-CM5 CNRM-CM5-2	Météo-France/Centre National de Recherches Météorologiques (CNRM) and Centre Européen de Recherches et de Formation Avancée en Calcul Scientifique (CERFACS), France	Voldoire et al. (2013)
CSIRO-Mk3.6.0	Australian Commonwealth Scientific and Industrial Research Organization (CSIRO) Marine and Atmospheric Research, Queensland Climate Change Centre of Excellence (QCCCE), Australia	Rotstayn et al. (2010)
EC-EARTH	EC-Earth consortium, Europe	Hazeleger et al. (2012)
FGOALS-g2 FGOALS-s2	National Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics (LASG) /Institute of Atmospheric Physics, China	Li et al. (2013); http://www.lasg.ac.cn/FGOALS/CMIP5/
FIO-ESM	First Institute of Oceanography (FIO), State Oceanic Administration (SOA), China	Zhou et al. (2014)
GFDL-CM3 GFDL-ESM 2G GFDL-ESM 2M GFDL-CM2p1	National Oceanic and Atmospheric Administration (NOAA) /Geophysical Fluid Dynamics Laboratory (GFDL), USA	Donner et al. (2011); http://nomads.gfdl.noaa.gov/
GISS-E2-H GISS-E2-H-CC GISS-E2-R GISS-E2-R-CC	National Aeronautics and Space Administration (NASA), Goddard Institute for Space Studies (GISS), USA	Schmidt et al. (2006)
HadCM3 HadGEM2-AO HadGEM2-CC HadGEM2-ES	Met Office Hadley Centre, UK	W. J. Collins et al. (2011)
INM-CM4	Institute for Numerical Mathematics (INM), Russia	Volodin et al. (2010)
IPSL-CM5A-LR IPSL-CM5A-MR IPSL-CM5B-LR	L'Institut Pierre-Simon Laplace (IPSL), France	Dufresne et al. (2013); Hourdin et al. (2013)
MIROC-ESM MIROC-ESM-CHEM MIROC4h MIROC5	Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Atmosphere and Ocean Research Institute (AORI), University of Tokyo, and National Institute for Environmental Studies (NIES), Japan	M. Watanabe et al. (2010); Watanabe et al. (2011)
MPI-ESM-LR MPI-ESM-MR MPI-ESM-P	Max Planck Institute, Germany	Roeckner et al. (2006); Wetzel et al. (2006)
MRI-CGCM3 MRI-ESM 1	Meteorological Research Institute (MRI), Japan	Yukimoto et al. (2012)
NorESM1-M NorESM1-ME	Norwegian Climate Centre, Norway	Bentsen et al. (2013)

Table 3
CMIP3 Models Used in This Study

Model(s)	Institute	Reference(s)
bccr_bcm2_0	Bjerknes Centre for Climate Research (BCCR), University of Bergen (UiB)	http://bjerknes.uib.no
cccma_cgcm3_1	Canadian Centre for Climate Modelling and Analysis, Canada	Mcfarlane et al. (1992)
cccma_cgcm3_1_t63		
csiro_mk3_0	Australian Commonwealth Scientific and Industrial Research Organization (CSIRO), Australia	H. B. Gordon et al. (2002)
gfdl_cm2_0	National Oceanic and Atmospheric Administration (NOAA) /Geophysical Fluid Dynamics Laboratory (GFDL), USA	(2004)
gfdl_cm2_1		
giss_aom	NASA Goddard Institute for Space Studies (GISS), USA	Bleck (2002); Russell et al. (1995); Schmidt et al. (2006)
giss_model_e_h		
giss_model_e_r		
iap_fgoals1_0_g	Institute of Atmospheric Physics (IAP), China	Yu et al. (2004)
ingv_echam4	Instituto Nazionale di Geofisica e Vulcanologia, Italy	
inmcm3_0	Institute for Numerical Mathematics (INM), Russia	Alekseev et al. (1998); Galin et al. (2003)
ipsl_cm4	L'Institut Pierre-Simon Laplace (IPSL), France	Hourdin et al. (2006)
miroc3_2_hires	Center for Climate Research (University of Tokyo), Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Japan	Hasumi and Emori (2004)
miroc3_2_medres		
mpi_echam5	Max Planck Institute, Germany	Roeckner et al. (2003)
mri_cgcm2_3_2a	Meteorological Research Institute (MRI), Japan	Yukimoto et al. (2006)
ncar_ccsm3_0	National Center for Atmospheric Research (NCAR), USA	Collins et al. (2006); Washington et al. (2000)
ncar_pcm1		
ukmo_hadcm3	Hadley Centre for Climate Prediction and Research/Met Office, UK	Gordon et al. (2000); Martin et al. (2006); Pope et al. (2000)
ukmo_hadgem1		

assess the models' improvements in reproducing observed climate parameters, we use several observational data sets and reanalyses summarized in Table 5.

In CMIP5 most of the models had a higher spatial resolution with 0.5° to 4° for the atmosphere component and 0.2° to 2° for the ocean component than the CMIP3 models (Taylor et al., 2012). In CMIP6 the spread of the models' spatial resolutions shifts again to finer grids. For the first time, a new “input4MIPs” activity (https://pcmdi.llnl.gov/mips/input4MIPs/) has been initiated in CMIP6 encourage adoption of common data standards and to create an archive of the forcing data sets and boundary conditions needed for the CMIP6 simulations available via ESGF. Many of the new forcing data sets are improved versions of the ones used in CMIP5 (see summary available at http://goo.gl/r8up31).

2.1. CMIP6

CMIP6 consists of three main elements (Eyring, Bony, et al., 2016): (1) a set of common experiments, the DECK (Diagnostic, Evaluation and Characterization of Klima) and CMIP historical simulations (1850 to near present) that are used here to document the basic characteristics of the models across different phases of CMIP; (2) common standards, coordination, infrastructure, and documentation that facilitates the distribution of model output and the characterization of the model ensemble; and (3) an ensemble of CMIP-Endorsed MIPs that are specific to a particular phase of CMIP (now CMIP6) and that build on the DECK and CMIP historical simulations to address a large range of specific scientific questions and help fill the scientific gaps of previous CMIP phases. CMIP6 models have an increased degree of freedom by including more processes and couplings, primarily aimed at being able to better simulate future feedbacks (e.g., nitrogen effects of terrestrial carbon uptake or permafrost processes). In this study we use the CMIP6 Carbon Dioxide (CO_2) concentration driven historical simulations (*historical*) over the time period 1850–2014 (Table 1). Common forcing data sets are defined for the CMIP6 historical simulations are largely based on observations and include: land-use changes (Ma et al., 2019), emissions and concentrations of long-lived greenhouse gases (Meinshausen et al., 2017) and of short-lived species (Hoesly et al., 2018), stratospheric aerosol from volcanoes (Zanchettin et al., 2016), biomass burning emissions (van Marle et al., 2017), and solar forcing (Matthes et al., 2017). It should, however, be noted that forcings in the model simulations can differ according to the complexity of the model. For example, some models are forced with a parameterization of anthropogenic aerosol optical properties and an associated Twomey effect (Stevens et al., 2017),

Table 4
HighResMIP Models Used in This Study

Model(s)	Institute	Reference(s)
CMCC-CM2-VHR4 CMCC-CM2-HR4	Centro Euro-Mediterraneo per I Cambiamenti Climatici (CMCC), Italy	Cherchi et al. (2019)
CNRM-CM6-1-HR CNRM-CM6-1	Météo-France/Centre National de Recherches Météorologiques (CNRM) and Centre Européen de Recherches et de Formation Avancée en Calcul Scientifique (CERFACS), France	Voldoire et al. (2019)
ECMWF-IFS-HR ECMWF-IFS-LR	European Centre for Medium-Range Weather Forecasting (ECMWF)	C. D. Roberts et al. (2018)
HadGEM3-GC31-HM HadGEM3-GC31-LL	Met Office Hadley Centre, UK; University of Reading, UK; Natural Environment Research Council (NERC), UK	M. J. Roberts et al. (2019)
MPI-ESM 1-2-XR MPI-ESM 1-2-HR	Max Planck Institute, Germany	Gutjahr et al. (2019)

Note. In each column the name of the high-resolution version (first line) and the corresponding low-resolution version (second line) is given.

while others treat aerosols interactively and therefore prescribe emissions of aerosols and their precursors instead (Hoesly et al., 2018). We only consider one ensemble member per model (“r1i1p1f1,” if available). In order to calculate ECS, we also use the simulations forced by an abrupt quadrupling of CO_2 (*abrupt-4* × *CO2*) and the preindustrial control simulations (*piControl*).

2.2. CMIP5

For CMIP5 (Taylor et al., 2012), we use the results from up to 48 models (Table 2) for the historical simulations depending on data availability for a specific variable. The historical simulations are twentieth-century simulations covering the time period 1850–2005 and are performed using the then best available record of natural and anthropogenic climate forcing (Cionni et al., 2011; Lamarque et al., 2010). In case there are multiple ensemble members available for a given model, we only consider the first ensemble member “r1i1p1” in our analysis. Again, we use the idealized abrupt four times CO_2 and the preindustrial control simulations to calculate ECS from the models.

2.3. CMIP3

The CMIP3 model simulations analyzed are the twentieth century runs (1860–1999) with natural and anthropogenic forcings (20C3M experiments). Again, in case there are multiple ensemble members available for a given model, we only analyze the first ensemble member “run1.” In total, there are up to 22 CMIP3 models considered in our analyses depending on data availability for a specific variable (Table 3).

2.4. HighResMIP

The HighResMIP (Haarsma et al., 2016) applies, for the first time, a multimodel approach to systematically investigate the impact of horizontal resolution on the results of global ESMs. A coordinated set of experiments over the time period 1950–2014 has been designed to assess both, a standard and an enhanced horizontal resolution simulation, in the atmosphere and ocean of each participating model (Table 4). To make the highest-resolution models computationally affordable, some compromises were necessary. The experiment design incorporates only a short (30–50 year) spin-up from 1950 initial conditions before control and historic-future simulations. Therefore, a direct comparison to the CMIP6 historical simulations that start in 1850 is not always possible. In this study, we therefore compare the lower-resolution and high-resolution model versions within HighResMIP, both starting in 1950, in order to assess possible improvements due to higher horizontal model resolution. In HighResMIP, physical models with few ESM components are used, and the aerosol optical properties are specified over time using the MACv2-SP scheme (Stevens et al., 2017).

2.5. Observations and Reanalysis Data

The observations and reanalysis data used for the model evaluation and assessment of the progress made during the different phases of CMIP are summarized in Table 5 including the type of observation, variables used, time period covered, and main reference(s). Where available, we use observational data sets from the

Table 5
Observations and Reanalyses Used in This Study

Data set	Type	Variable	Time period	Reference
AIRS	satellite	specific humidity (hus)[a]	2003–2010	Susskind et al. (2006); Tian et al. (2013)
CERES-EBAF	satellite	TOA outgoing shortwave radiation (rsut)[a], TOA outgoing longwave radiation (rlut)[a], TOA shortwave cloud radiative effect (swcre)[a], TOA longwave cloud radiative effect (lwcre)[a]	2001–2015	Loeb et al. (2012)
ERA5	reanalysis	near-surface temperature (tas)	1980–2014	Copernicus Climate Change Service (C3S) (2017)
ERA-Interim	reanalysis	specific humidity (hus)[b], sea level pressure (psl)[b], temperature, eastward wind (ua)[a], northward wind (va)[a], temperature (ta)[a], near-surface air temperature (tas)[a], geopotential height (zg)[a]	1980–2014	Dee et al. (2011)
ESACCI-CLOUD	satellite	total cloud cover (clt)[a]	1982–2014	Stengel et al. (2017)
ESACCI-SST	satellite	surface temperature (ts)[a]	1992–2010	Merchant et al. (2014)
GHCN	ground	precipitation (pr)[b]	1980–2014	Vose et al. (1992)
GPCP-SG	satellite and rain gauge	precipitation (pr)[a]	1980–2014	Adler et al. (2003); Huffman and Bolvin (2012)
HadCRUT4	station	near-surface temperature (tas)	1980–2014	Morice et al. (2012)
HadISST	station	surface temperature (ts)[b]	1980–2014	Rayner et al. (2003)
JRA-55	reanalysis	sea level pressure (psl)[a]	1980–2014	Harada et al. (2016); Kobayashi et al. (2015)
NCEP	reanalysis	temperature (ta)[b], eastward wind (ua)[b], northward wind (va)[b], near-surface air temperature (tas)[b], geopotential height (zg)[b]	1980–2014	Kalnay et al. (1996)
PATMOS-x	satellite	total cloud cover (clt)[a]	1982–2014	Heidinger et al. (2014)

[a] Reference observational data sets for this variable in Figures 6 and 7. [b] Alternate observational data sets for this variable in Figures 6 and 7.

observations for Model Intercomparison Projects (obs4MIPs; Ferraro et al., 2015; Waliser et al., 2020), which can be downloaded freely from the ESGF and, because they are provided in the same file format including all relevant meta data as the output from the CMIP6 models, and can be used directly with the ESMValTool.

2.6. ESMValTool

The ESMValTool is a community diagnostics and performance metrics tool specifically developed for evaluation of ESMs contributing to CMIP (Eyring et al., 2020; Righi et al., 2020). ESM results from single or multiple models can be compared with their predecessor versions and against observations. The diagnostics available in the ESMValTool cover a wide range of scientific themes focusing on selected essential climate variables, a range of known systematic biases common to ESMs, meteorology, clouds, tropospheric aerosols, ocean variables, land processes, etc. All diagnostics are grouped in sets of standard “recipes” for each scientific topic reproducing diagnostics or performance metrics that have demonstrated their importance in ESM evaluation in the peer-reviewed literature. The main aim of the ESMValTool is to facilitate and improve ESM evaluation beyond the state-of-the-art and to support activities within CMIP and at individual modeling centers. This includes provision of well-documented diagnostics and source code as well as ensuring reproducibility and traceability of the results (provenance). The ESMValTool is an open source project and can be found on GitHub at https://github.com/ESMValGroup/ESMValTool with contributions from the community very welcome. Contributions could include, but are not limited to, documentation improvements, bug reports, new or improved diagnostic code, scientific and technical code reviews, infrastructure improvements, mailing list and chat participation, community help/building, education, and outreach. For more information on contributing to the ESMValTool, general guidelines, code style, etc., we refer to the ESMValTool user's guide available at https://docs.esmvaltool.org website. A general overview on the ESMValTool is given by Eyring, Righi, et al. (2016), technical details of the latest version (v2.0) can be found in Righi et al. (2020), diagnostics and metrics newly added to v2.0 are described in three companion papers (Eyring et al., 2020; Lauer et al., 2020; Weigel et al., 2020). The ESMValTool is fully integrated into the ESGF infrastructure at the Deutsches Klimarechenzentrum (DKRZ) where all the model output and the observations are stored in a local replica and the tool is run. All diagnostics used for this paper will be made available in the ESMValTool after acceptance of this publication and the figures can be reproduced with the newly added recipe “recipe_bock20jgr.yml.”

3. Surface Temperature Record

Figure 1 shows the time series of anomalies in annual global mean near-surface temperature simulated by CMIP3, CMIP5, and CMIP6 models. The time period 1850–1900 has been used as reference period to calculate the temperature anomalies (1870–1900 for CMIP3 models starting in 1870). The reference data set for comparison with the models is HadCRUT4 (Morice et al., 2012). In general, the models of all CMIP phases are able to reproduce the observed temperature record reasonably well in a range of ±0.9°C showing an increase in global-averaged annual mean near-surface temperature since the year 1850 including an accelerated warming beginning in the 1970s and a temporary cooling that follows large volcanic eruptions such as Krakatoa in 1883 or Agung in 1963. The temperature changes since the late nineteenth century are driven by a number of factors, including increasing atmospheric greenhouse gas concentrations, changes in aerosol amounts, changes in solar activity, volcanic eruptions, and changes in land use. Natural variability also plays an important role particularly on shorter timescales such as for the observed slowdown (“hiatus”) in the observed increase in global surface temperature warming rates during the time period 1998–2013 (Meehl et al., 2014), although ocean heat content continued to increase over the same period (Yin et al., 2018).

The CMIP3 multimodel mean already captured the observed surface temperature change quite well with a warming for the years 1990 to 1999 in the range of 0.45°C to 0.73°C compared to 0.38°C to 0.74°C for the observations. Even though there are some outliers leading to a rather large intermodel spread. A similarly large spread exists in mean absolute temperatures simulated by CMIP3 models, and that spread persists in CMIP5 and CMIP6 (see insets in Figure 1).

Figure 2 shows the intermodel spread of the three CMIP ensembles as ±1 standard deviation around the multimodel means in comparison to the uncertainty estimates of the global temperature anomalies from HadCRUT4. The observed uncertainty estimates are the 5% and 95% percentiles of the confidence interval of the combined effects of uncertainties from measurement and sampling as well as bias and coverage (Morice et al., 2012). All models have been sampled according to the temporal and spatial data availability from HadCRUT4 and therefore include similar sampling and coverage uncertainties as the observations. The intermodel spread for temperature anomalies, which are less uncertain in observations than absolute values (P. D. Jones et al., 1999), are slightly reduced in CMIP5 and CMIP6 with standard deviations of 0.16°C and 0.17°C, respectively, after the reference period compared to 0.19°C for CMIP3. Particularly from the second half of the twentieth century onward, the intermodel spreads in all CMIP enselbles are larger than the HadCRUT4 uncertainty estimates and do not narrow down with time. This suggests that besides natural variability, model uncertainty is an important contribution to the intermodel spread in all three CMIP phases. Since the intermodel spread does not change substantially among the different CMIP phases, this further suggests that model uncertainties remain to be important factors determining the intermodel spread throughout the observed time period.

There were discussions focussing on the observed reduction in the rate of surface temperature warming during the hiatus period which was apparently not reproduced by the CMIP5 models (Flato et al., 2013; Meehl et al., 2014). It has subsequently been shown that the slowdown in the rate of global warming in the early 2000s likely predominantly due to internal variability from the negative phase of the Interdecadal Pacific Oscillation (IPO) in the Pacific (England et al., 2014; Fyfe et al., 2016; Xie & Kosaka, 2017) with some contributions from aerosol forcing from a collection of moderate sized volcanic eruptions (Santer et al., 2015) and perhaps partly from anthropogenic aerosol forcing (D. M. Smith et al., 2016) though such a role for anthropogenic aerosols is still being debated (Oudar et al., 2018). Thus, uninitialized climate models averaged across multiple ensemble members to remove the effects of internal variability cannot be expected, by definition, to reproduce, in such a multimodel mean, a phase of internal variability in the single realization of the observations. However, a small number of CMIP5 model realizations were, by chance, able to

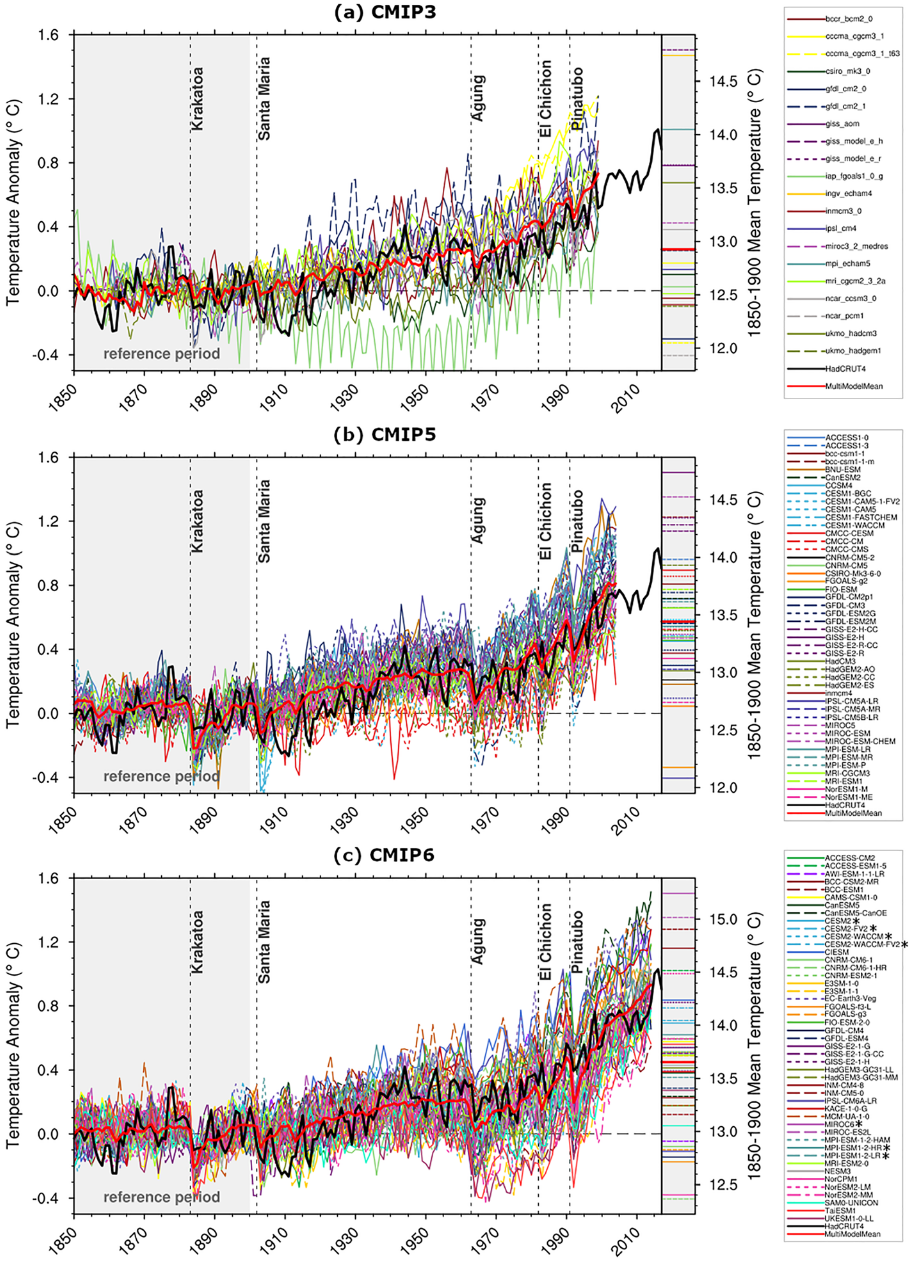

Figure 1. Observed and simulated time series of the anomalies in annual and global mean surface temperature. For CMIP3, CMIP5, and CMIP6, all anomalies are differences from the 1850–1900 time mean of each individual time series. For the CMIP3 model simulations starting later than 1850 the reference period is defined as the available time period between 1850 and 1900. The reference period is indicated by gray shading. The thin lines show individual climate model simulations from (a) CMIP3, (b) CMIP5, and (c) CMIP6, the thick red lines show the multimodel means. The observational data (thick black lines) are the Hadley Centre/Climatic Research Unit gridded surface temperature data set Version 4 (HadCRUT4; Morice et al., 2012). All models have been subsampled using the HadCRUT4 observational data mask (see Jones et al., 2013). Inset: The global mean surface temperature for the reference period 1850–1900 of the subsampled fields. CMIP6 models, which are masked with an asterisk are either tuned to reproduce observed warming directly, or indirectly by tuning equilibrium climate sensitivity.

simulate the internally generated slowdown that happened to occur at the same time as shown by the observations, and those simulations also were characterized by a negative phase of the IPO (Meehl et al., 2014). This strongly suggests that the models do indeed include the processes that can produce decadal slowdowns or accelerations, but this presents a challenge for interpreting multimodel ensemble averages when comparing to observed decadal-timescale variability from the single realization of the observations. As the historical CMIP6 simulations extend beyond the hiatus period, we found that there is again a convergence between the time series of the multimodel mean and the observed temperature record until the year 2014. But the CMIP6 multimodel mean tends to simulate reduced warming over the period 1950–1990 (with a mean bias of −0.07°C) which is probably at least partly related to an overestimation of the cooling in response to large increases in anthropogenic emissions of primary aerosol and precursors in the 1950s in some models (Andrews et al., 2019; Dittus et al., 2020; Flynn & Mauritsen, 2020; Hoesly et al., 2018). The lack of simulated warming in that period (Figure 1) could be caused by a high aerosol effective radiative forcing (ERF) in these models. Dittus et al. (2020) supports that explanation by varying the strength of aerosol ERF in the CMIP6 version of the HadGEM3 climate model. They find that temperature trends over the period 1951–1980 are significantly more sensitive to the strength of aerosol ERF than the 30 previous (1921–1950) and following (1981–2010) years, when temperature trends where driven by greenhouse gas increases. Aerosol ERF measures imbalances in the Earth's energy budget due to anthropogenic aerosols, including aerosol-radiation interactions and aerosol-cloud interactions and their rapid adjustments (Sherwood et al., 2015). Several models reduced the strength of their simulated aerosol radiative forcing during their development phase to ensure that total anthropogenic radiative forcing remained positive (Danabasoglu et al., 2020; Mulcahy et al., 2018). Potentially as a result of overly sensitive aerosol-cloud-radiation coupling, individual CMIP6 models may underestimate the observed global temperature anomalies in the 1960s to 1980s by up to 0.5°C, while being much closer to the observations during the rest of the historical period.

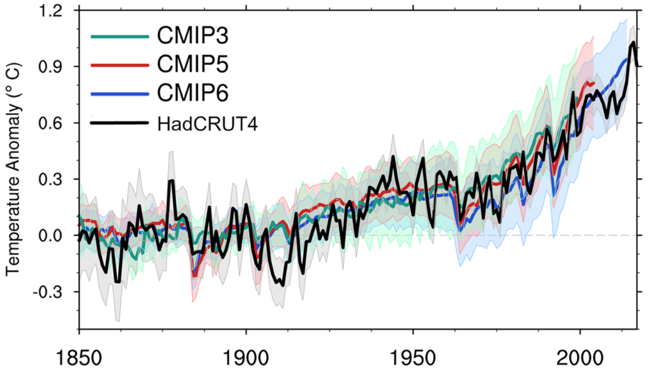

Figure 2. Observed and simulated time series of the anomalies in annual and global mean surface temperature as in Figure 1; all anomalies are calculated by subtracting the 1850–1900 time mean from the time series. Displayed are the multimodel means of all three CMIP ensembles with shaded range of the respective standard deviation. In black the HadCRUT4 data set (HadCRUT4; Morice et al., 2012). Gray shading shows the 5% to 95% confidence interval of the combined effects of all the uncertainties described in the HadCRUT4 error model (measurement and sampling, bias, and coverage uncertainties) (Morice et al., 2012).

By correlating each model's aerosol ERF for 2014 (C. J. Smith et al., 2020) with its simulated warming trend between 1945 to 1970, we find some evidence to support the hypothesis that CMIP6 models with particularly strong negative aerosol forcing show a larger surface cooling trend in the midtwentieth to late twentieth century, with this relationship most clear when temperature trends for the NH extratropics are considered. We note that the C. J. Smith et al. (2020) aerosol ERF for 2014 is not always representative of the aerosol ERF experienced by models over the time period 1945–1970 because models could have different aerosol ERF histories. We do not, however, expect this to have a large impact on the strength or sign of the relation found between aerosol ERF and temperature trend as preliminary results from the RFMIP piClim-histaer simulation suggest that the aerosol ERF values for midcentury and present-day typically scale rather similarly among the models. In addition to the forcing itself, details of how individual models respond to this negative forcing also plays a role in determining their overall historical temperature record. The very high warming rates in the last part of the twentieth century of some models such as CanESM5 and UK-ESM, as well as their strong cooling after volcanic eruptions, are reflected in very large climate sensitivity values (see further discussion in section 6).

When evaluating model simulations of historical temperature change, it is important to keep in mind that good agreement with the long-term twentieth century trend of observed surface temperature changes is expected for models that are directly or indirectly tuned to reproduce observed twentieth-century warming (Hourdin et al., 2017; Mauritsen et al., 2012). Tuning itself means an objective process of parameter estimation to fit a predefined set of observations (Hourdin et al., 2017). However, the tuning is not time-dependent so the decadal variability of the time evolution of global temperature relies on how the models respond to external forcings such as volcanic eruptions, solar variability, and time-evolving anthropogenic aerosols. Thus, there is no significant difference in the multimodel mean anomaly time series of near-surface temperature obtained for models that have been tuned toward the observed warming rates or for models that have not (not shown). The anomaly time series for surface temperature for the tuned models (marked with asterisks in the legend of Figure 1) is too cold in the second half of the twentieth century, just like models that are not tuned to twentieth century warming.

4. Systematic Biases

Climate models are known to exhibit a number of different and partly long-standing biases in reproducing observed climate (Stouffer et al., 2017). In order to be able to address one of the scientific key question in CMIP6, "What are the origins and consequences of systematic model biases?" (Eyring, Bony, et al., 2016), a first step is to identify which of the systematic model biases are still present in the CMIP6 historical simulations. A second step is then to assess potential progress and improvements in the models' performance compared with older model generations that contributed to CMIP3 and CMIP5 throughout the last two decades. Here, we are not specifically aiming at tracking the performance of individual models but rather the performance of generations of climate models. We therefore compare multimodel means of CMIP Phases 3, 5, and 6 against observations and against each other in order to identify still existing biases and assess potential progress in reproducing the observed climate state of the last decades.

4.1. Surface Temperature

One of the prognostic variables of climate models that is most commonly used and downloaded from the CMIP archives is surface temperature. Figure 3a shows that the CMIP6 multimodel mean is able to simulate the key characteristics of the observed global surface temperature pattern. The dominant feature of the climatology (1995–2014) is the zonal gradient from high temperatures at the equator to low values at the poles. High-elevation regions like the Himalayas, the Andes or Antarctica are significantly cooler than the latitudinal average temperature. Seasonal changes in temperatures are also generally well reproduced (Flato et al., 2013).

All CMIP ensembles reproduce the large-scale annual mean patterns from the reference reanalysis data set ERA5 quite well (pattern correlations of 0.99 or larger for all CMIP models, see Figure 7). The global mean bias improves from CMIP3 (−0.451°C) to values near zero for CMIP5 and CMIP6 (Figures 3b–3d). And also the global mean root-mean-square difference (RMSD) decreases continously for the different CMIP ensembles but in some regions there are some long-standing biases (Figures 3b–3d). These biases include too high surface temperatures in the upwelling regions of subtropical oceans of up to several °C. One possible reason for this warm bias is an underestimation of the stratocumulus cloud fraction in these regions. Biases in the high-elevation regions are also still apparent in CMIP6 but typically somewhat smaller than in CMIP3. This also applies to biases along the edge of the North Atlantic sea ice field. The positive temperature bias over the Southern Ocean, however, seems to have gotten worse in time (Hyder et al., 2018) with the CMIP6 multimodel mean showing larger biases than in the two previous CMIP phases. Regional absolute biases in surface temperature of up to 6°C as seen in CMIP5 and some pre-CMIP6 models (Lauer et al., 2018) are still present in CMIP6.

There are many different, model-dependent causes for biases in modeled surface temperature. Common causes include biases in downward shortwave radiation at the surface because of errors in simulated cloud properties (Hyder et al., 2018; Lauer et al., 2018), errors in oceanic circulation (Kuhlbrodt et al., 2018), errors in the simulation of trade winds (Lauer et al., 2018), and errors in surface albedo and moisture propagated from the vegetation schemes (Séférian et al., 2016). Even though the multimodel mean of the surface temperature bias shows only small improvements in CMIP6, some individual models made significant progress (Danabasoglu et al., 2020).

It is noteworthy that some of the long-standing biases seem to be related to horizontal model resolution. After increasing the horizontal resolution, as done in HighResMIP (Figures 3e and 3f), some of the biases were reduced or even disappeared compared to the mean bias of the corresponding lower-resolution versions of the same models simulating the same time period. Both, the global mean bias and RMSD decrease with higher horizontal resolution. There is a clear improvement in many of these regional biases, particularly in the stratocumulus regions where typical biases found in the high-resolution models of HighResMIP are below 1°C compared to up to 3–4°C found in the multimodel mean of the low-resolution models of HighResMIP (M. J. Roberts et al., 2019). Improvements can be seen in the upwelling regions off the west coasts of South America and Africa and also over the northern Atlantic (Caldwell et al., 2019; Docquier et al., 2019). The cold bias along the equator in the Pacific Ocean with too cold SSTs extending too far west (Lauer et al., 2018) disappeared in the high resolution versions (Roberts et al., 2018, 2019). A notable exception to biases improvements from higher resolution is again the Southern Ocean, where

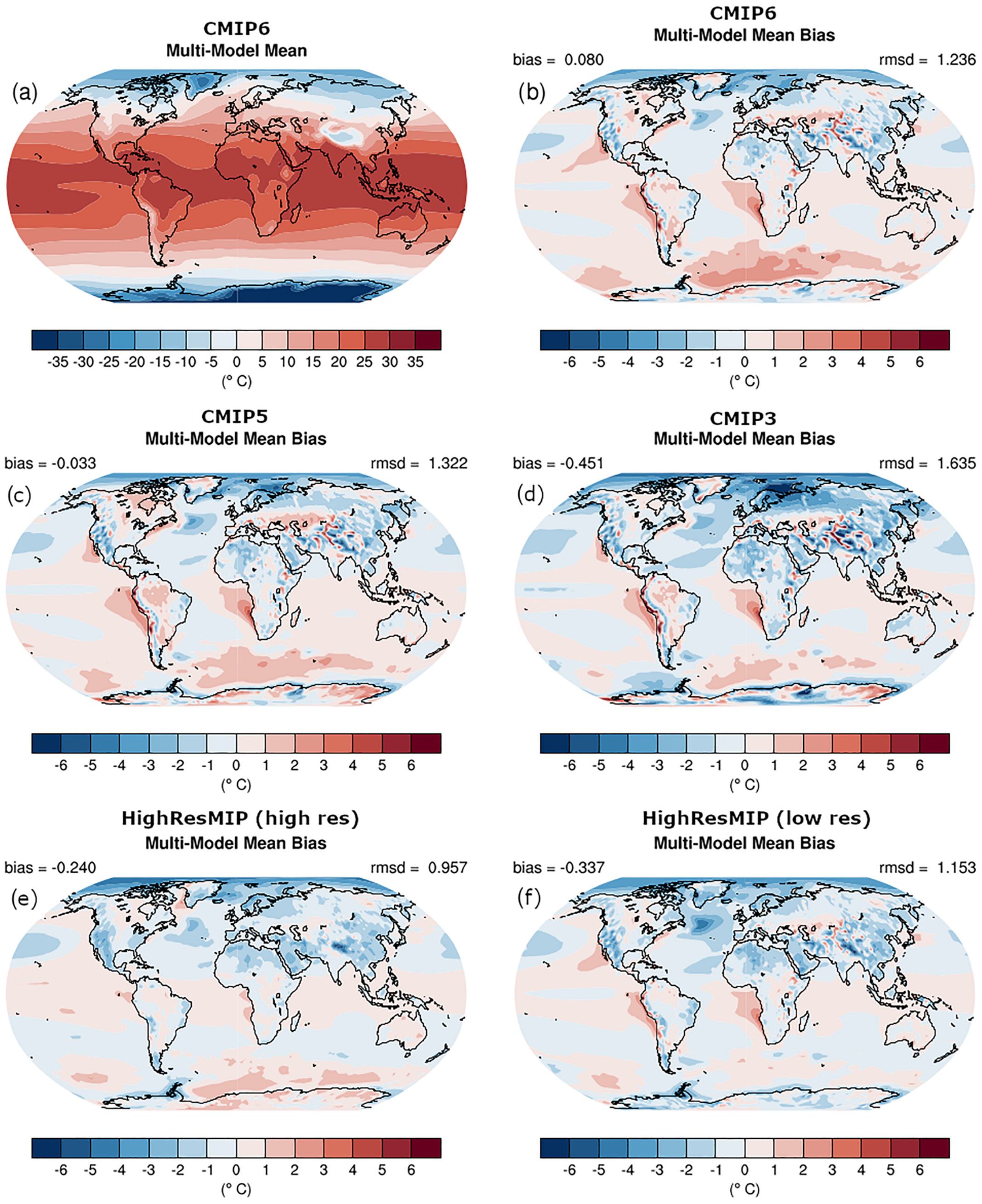

Figure 3. Annual mean near-surface (2 m) air temperature (°C). (a) Multimodel (ensemble) mean constructed with one realization of CMIP6 historical experiments for the period 1995–2014. Multimodel-mean bias of (b) CMIP6 (1995–2014), (c) CMIP5 (1985–2004), (d) CMIP3 (1980–1999), (e) high-resolution, and (f) low-resolution simulations of the HighResMIP ensemble (1995–2014) compared to the corresponding time period of the climatology from ERA5 (Copernicus Climate Change Service (C3S), 2017).

biases increase at these eddy-permitting ocean resolutions. It should, however, be noted that the most models in the current group of HighResMIP models include a NEMO-based ocean, so there is little ocean model diversity. Also, because of the shorter simulation period starting from observed initial conditions in 1950 in HighResMIP, compared to starting in 1850 from a preindustrial spun-up state for the CMIP6 historical simulations, better agreement of the HighResMIP simulations with observations can be expected. Because of this, the performance of the HighResMIP simulations is not directly comparable to the one of the CMIP6 historical simulations.

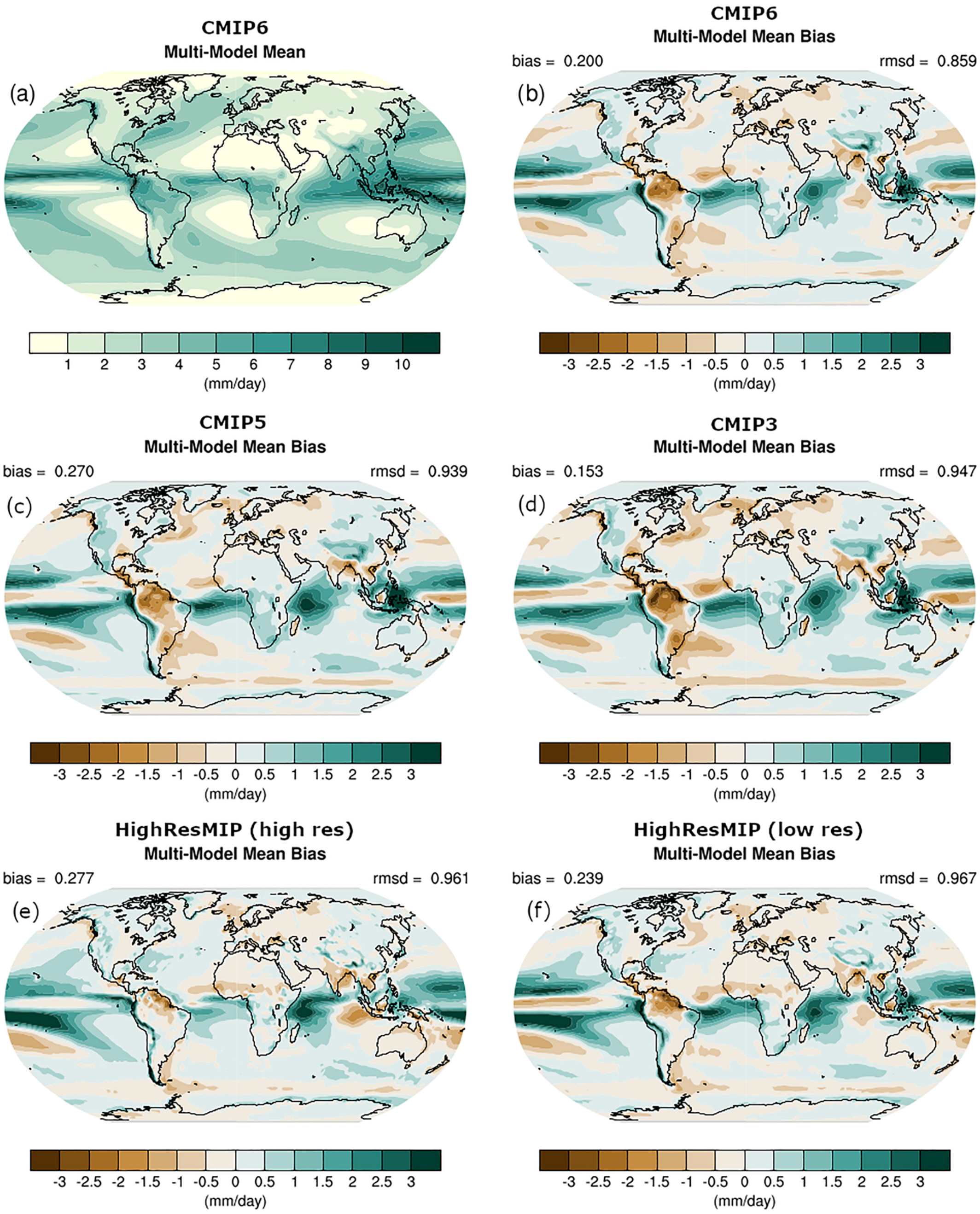

Figure 4. Same as Figure 3 but for annual mean precipitation rate (mm day^{-1}). Data from the Global Precipitation Climatology Project (GPCP) Version 2.3 (Adler et al., 2003) are used as a reference.

4.2. Precipitation

The multimodel mean of the CMIP6 ensemble shows the well-known large-scale features of the global precipitation pattern (Figure 3a). Precipitation near the equator is high due to frequently occurring deep convection connected with the Intertropical Convergence Zone (ITCZ). In the subtropical subsidence regions precipitation rates are low and increase again in midlatitudes due to precipitation by frontal systems (midlatitude storm tracks). The cold temperatures and the associated low water vapor saturation ratio at the poles leads to a relatively low amount of precipitation in high latitudes. Pattern correlations between the modeled and observed geographical distribution of annual mean precipitation range between 0.69 and 0.87 for CMIP3, 0.79 to 0.88 for CMIP5, and 0.80 and 0.92 for CMIP6 models (Figure 7).

The comparison of the multimodel mean with the global precipitation data set from the Global Precipitation Climatology Project (GPCP; Adler et al., 2003) shows that the global mean climatology of the CMIP ensembles is a bit too wet but the global RMSD decreases from CMIP3 to CMIP6 (Figures 4b–4d). But there are some long-standing systematic model biases throughout the different CMIP phases. The largest precipitation biases of up to 3.5 mm day^{-1} are found in the tropics. They include the occurrence of a double ITCZ in the tropical Pacific and a southward shifted ITCZ in the equatorial Atlantic with rather little progress from CMIP3 to CMIP6. A double ITCZ is often driven by incorrect simulation of the meridional gradients in SST across the equator (Oueslati & Bellon, 2015) and thus a complex problem of the coupled atmosphere-ocean system. In general, the amplitude and geographical pattern of the precipitation biases in CMIP6 are quite similar to those from CMIP5. There is some improvement, however, in CMIP6 compared with CMIP3 and CMIP5 in the overly intense Indian Ocean ITCZ and the too dry South American continent (excluding the Andes) by about 1 mm day^{-1}. There have been also progressive improvements in the extratropical representation of precipitation from CMIP3 to CMIP6. The CMIP6 models have an improved zonal tilt of the North Atlantic winter storm track (Priestley et al., 2020), which may have contributed to the decrease in the dry bias over that region. Also, the equatorward and zonal mean bias in the SH midlatitudes has been largely reduced. These improvements have been attributed to model horizontal resolution in the NH and to model physics in the SH (Priestley et al., 2020).

The multimodel mean bias of the high resolution versions in comparison to the multimodel mean bias of their corresponding low resolution counterparts in HighResMIP (Figures. 4e and 4f) shows some improvements. There is a strong decrease in the precipitation bias in the tropical Atlantic by about 1–2 mm day^{-1} as well as a near disappearance of the dry bias in the equatorial Pacific. A possible explanation for this improvement could be that, together with the improved SST biases (Figures 3e and 3f), the seasonal mean circulation and ITCZ migration are better represented with higher horizontal resolution (Vannière et al., 2019) leading to smaller biases.

4.3. Meteorology

Figure 5 shows climatological annual means of zonally averaged temperature, specific humidity and zonal wind (u-component) from the CMIP multimodel means compared with data from the ERA-Interim reanalysis (Dee et al., 2011). Prominent, well-known biases in the simulated vertical temperature distribution throughout all CMIP phases include a cold bias of several K in the extratropical upper troposphere at around 200 hPa and a warm bias of about 1–2 K in the tropics at about 100 hPa (John & Soden, 2007). The cold bias is somewhat reduced from maximum values of around 8 K in CMIP3 to about 6 K in CMIP6. Additionally, the warm bias in the upper tropical troposphere is reduced in its extent and magnitude from CMIP3 (up to 3 K) to CMIP6 (up to 2 K). The same is true for the cold bias in the lower stratosphere in the tropics with a reduction from about 3 K in CMIP3 to about 1 K in CMIP6. An improvement from CMIP5 to CMIP6 can also be seen in the cold bias throughout most of the troposphere in the SH that is reduced from 1–2 K in CMIP5 to about 1 K in CMIP6.

The concentration of water vapor in the atmosphere spans several orders of magnitude. Therefore, Figure 5 shows relative biases of simulated specific humidity instead of absolute differences in order to facilitate assessment of the performance of the CMIP multimodel means to reproduce the observationally based reference data set ERA-Interim. Consistent with the cold bias in the extratropical upper troposphere at around 200 hPa in both hemispheres, water vapor is underestimated in the CMIP models. While there is little change in this dry bias from CMIP3 to CMIP5, this bias is clearly improved in CMIP6 with bias values now ranging between −10% and −30% down from −20% to more than −45%. Similarly, the wet bias in middle- and high-latitude upper troposphere/lower stratosphere is improved from CMIP3 through CMIP5 to CMIP6. Throughout most of the stratosphere, the CMIP3 multimodel mean shows a strong dry bias (−40% to −130%). This bias has been reduced in CMIP5 with the dry bias now being mostly confined to high latitudes and even further reduced in the CMIP6 multimodel mean with values now mostly below −20% to −30% compared to ERA-Interim (Figure 5, middle column).

The simulated multimodel mean zonal wind speed (u-component) from CMIP6 models shows a reduction in the positive bias in stratosphere in midlatitudes in both hemispheres (bias up to 4–5 m s^{-1}) compared to CMIP3 (bias up to 9 m s^{-1}) and also compared to CMIP5 (bias up to 6 m s^{-1}). The negative bias in zonal wind speed found above the tropical tropopause of up to several m s^{-1} in CMIP5 is also clearly reduced in the

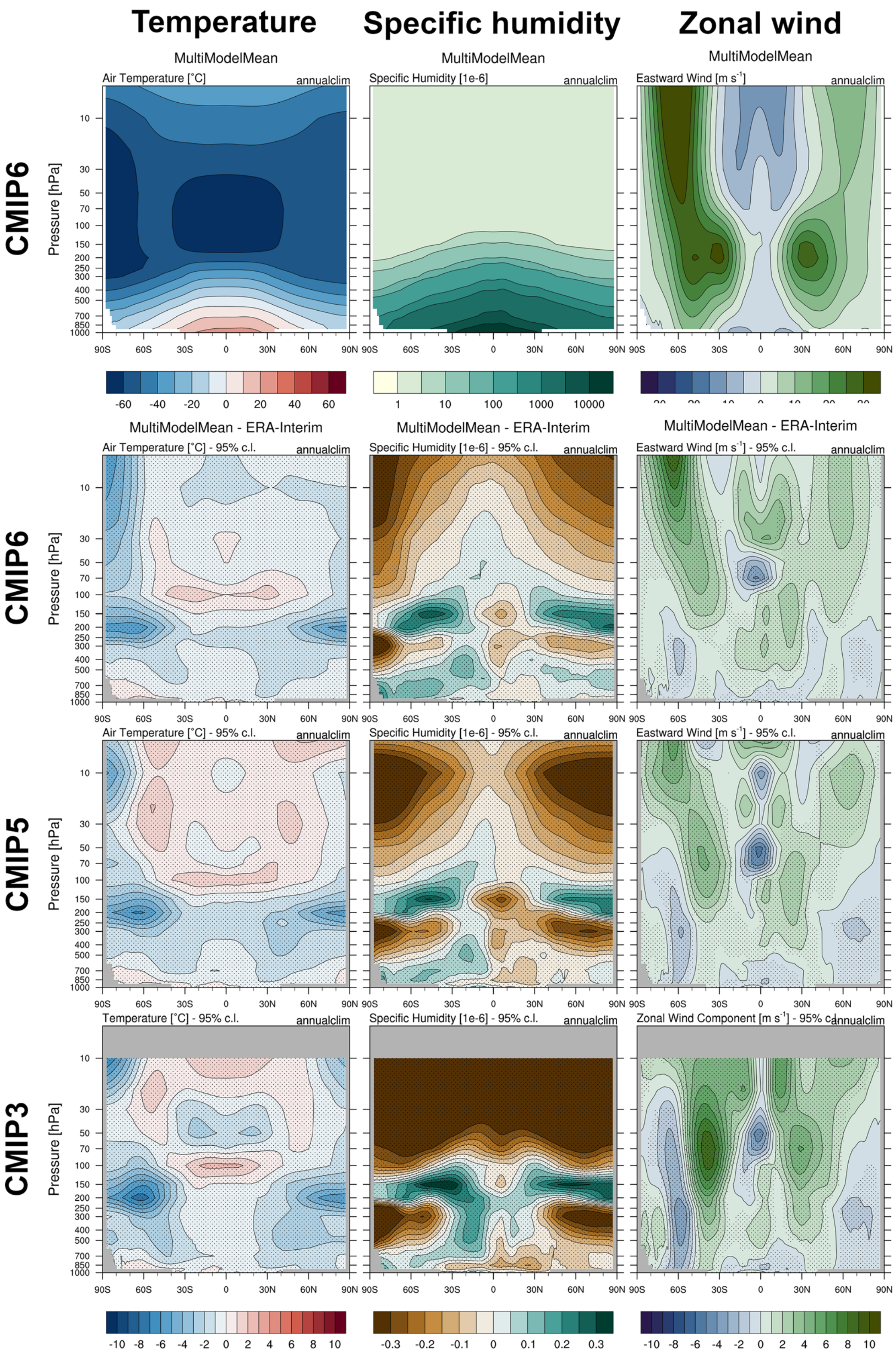

Figure 5. Multimodel mean of zonal averages for temperature (C°, left), specific humidity (10^{-6}, middle) and zonal wind (m s^{-1}, right) from CMIP6 (1995–2004) (upper row). Also shown are the absolute (temperature and zonal wind) and the relative (specific humidity) deviations from ERA-Interim for (from top to bottom) CMIP6 (1995–2004), CMIP5 (1995–2004) and CMIP3 (1980–1999) multimodel means. Stippled areas show differences that are statistically significant at a 95% confidence level.

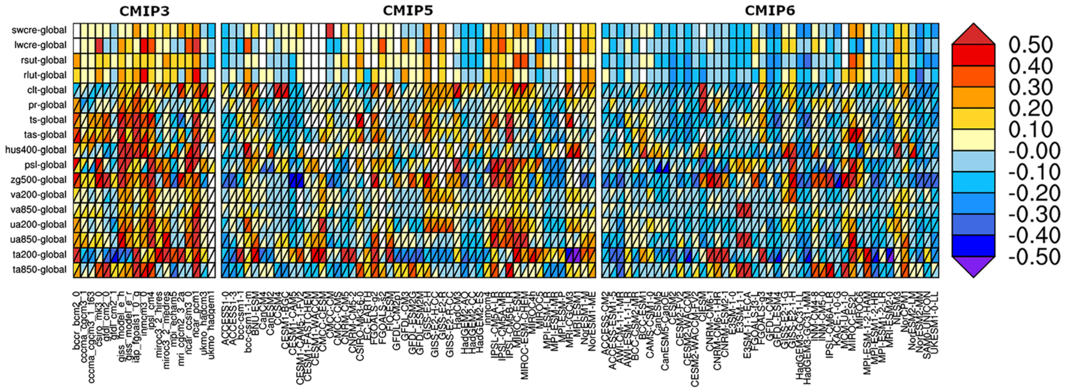

Figure 6. Relative space-time root-mean-square deviation (RMSD) calculated from the climatological seasonal cycle of the CMIP3, CMIP5, and CMIP6 simulations (1980–1999) compared to observational data sets (Table 5). A relative performance is displayed, with blue shading being better and red shading worse than the median RMSD of all model results of all ensembles. A diagonal split of a grid square shows the relative error with respect to the reference data set (lower right triangle) and the alternative data set (upper left triangle) which are marked in Table 5. White boxes are used when data are not available for a given model and variable. Updated and expanded from Figure 9.7 of Flato et al. (2013).

CMIP6 multimodel mean by about 0.1 m s^{-1}. Reasons for that could be the increased vertical resolution in the upper (tropical) troposphere and lower stratosphere and better resolved processes in the stratosphere, for example, gravity wave parameterizations (Manzini et al., 2014). In addition, the better representation of zonal mean specific humidity (Figure 5, middle column) might contribute to the improvement of the zonal wind speed climatology in CMIP6.

5. Quantification of Model Performance Across the CMIP6 Ensemble and CMIP Phases

In this section, the performance of the three different generations of climate models from CMIP3, CMIP5, and CMIP6 is assessed across different variables using multiple diagnostic fields. For every diagnostic field considered, model performance is compared to one or multiple observational reference data sets, and the quality of the simulation is summarized in a single number such as correlation coefficient or a RMSD. The general model improvements can then be quantified by simultaneously assessing a number of different performance indices. The use of such performance metrics in the model development phase potentially introduces the risk of tuning models to reproduce a set of metrics ignoring deficiencies elsewhere. This is why such performance metrics should mostly be seen as a possible starting point for more in-depth process-oriented evaluation that allows the identification of compensating errors (Eyring et al., 2005).

Performance metrics such as a portrait diagram shown in Figure 6 or a summary plot of the pattern correlations for different variables as shown in Figure 7 offer the possibility to quickly get an overview on model performance and can be either used as a starting point for more in-depth evaluation of individual variables or climate parameters with observations (Flato et al., 2013) or as one possible summary of overall model performance. Figure 6 is an extended and updated version of Figure 9.7 of Flato et al. (2013) that is based on Gleckler et al. (2008). It shows the normalized relative space-time RMSD of the climatological seasonal cycle from model simulations compared with observations for selected variables. Here, RMSD values are normalized with the centered median RMSD, that is, by substracting the median RMSD from the RMSD of an individual model and then dividing by the median RMSD. The median RMSD for each variable used for normalization is calculated across all models from all CMIP phases to make the grading of the models directly comparable across CMIP3, CMIP5, and CMIP6. Thus, positive and negative values are possible with positive values indicating a model performance worse than the median RMSD and negative values a performance better than the median RMSD. Here, all RMSD values are averaged over the whole globe. Where available, the model results are not only compared to one observational (ly based) reference data set but also to a second alternative data set to get an estimate of the observational uncertainty. This is indicated by diagonally divided boxes in Figure 6. All model data are masked according to data availability from the reference data sets and averaged over the same years with observational data available.

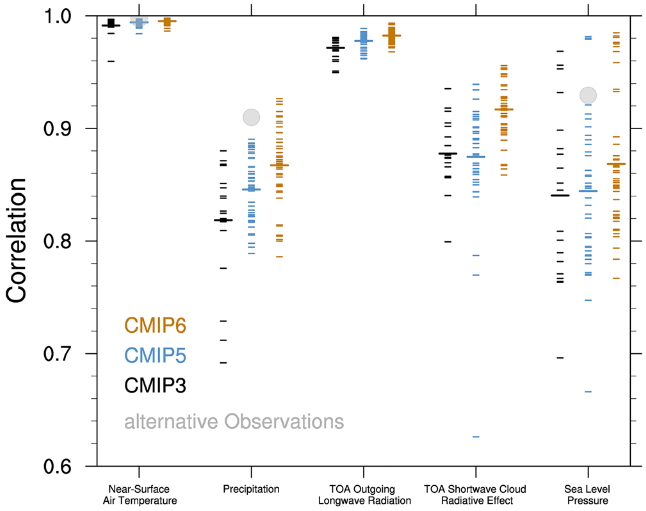

Figure 7. Centered pattern correlations between models and observations for the annual mean climatology over the period 1980–1999. Results are shown for individual CMIP3 (black), CMIP5 (blue), and CMIP6 (brown) models as short lines, along with the corresponding ensemble averages (long lines). The correlations are shown between the models and the reference observational data set listed in Table 5. In addition, the correlation between the reference and alternate observational data sets are shown (solid gray circles, marked in Table 5). To ensure a fair comparison across a range of model resolutions, the pattern correlations are computed after regridding all data sets to a resolution of 2.5° in longitude and 2.5° in latitude. Only one realization is used from each model from the CMIP3, CMIP5, and CMIP6 historical simulations.

Figure 6 shows that model performance varies across the models and across the variables, with no single model outperforming the other models for all variables. Nevertheless, we see model families of which members are performing quite similar, for example, the CMIP6 GFDL or CMIP6 GISS models. This is, however, not true for all model families with, for example, CMIP6 models MIROC-ES2L and MIROC6 showing quite different performances.

In general, there are clear improvements from CMIP3 to CMIP6 with the majority of CMIP3 models showing on average more red (positive values) boxes (CMIP3 ensemble median RMSD over all diagnostics = 0.127; 25%/75% percentiles = 0.003/0.283) than CMIP5 (CMIP5 median RMSD = 0.022; 25%/75% percentiles = −0.069/0.146) and the CMIP6 models showing the most blue (negative values) boxes (CMIP6 median RMSD = −0.064; 25%/75% percentiles = −0.146/0.048). Radiation fields have already shown improvements from CMIP3 to CMIP5 and this development continues in CMIP6 as the models fit quite well to the CERES-EBAF observations. The same applies to total cloud cover (clt) and precipitation (pr). The seasonal cycle of near-surface air temperature is not represented extremely well in CMIP3 (median RMSD = 0.191) but there were a lot improvements through CMIP5 (median RMSD = 0.014) to CMIP6 (median RMSD = −0.069). Moreover, the dynamical fields, sea level pressure (psl) and the geopotential height at 500 hPa (zg_{500}) show improvements from CMIP3 (median RMSD for zg_{500} = 0.357) to CMIP6 (median RMSD for zg_{500} = −0.121) even though some individual models still have problems in specific regions. Also, wind fields simulated by the CMIP6 models are in better agreement with observations than those from previous CMIP phases (see also section 4.3). The results for the temperature fields in 200 and 850 hPa show quite a large range in the RMSD for the different models in CMIP3 (median RMSD = 0.166), CMIP5 (median RMSD = 0.017) and also in CMIP6 (median RMSD = −0.050).

Using centered pattern correlations for selected fields (here: near-surface air temperature; precipitation; outgoing top of the atmosphere, TOA; longwave radiation; TOA shortwave cloud radiative forcing; and sea level pressure), Figure 7 shows significant improvements from the CMIP3 ensemble to the CMIP6 ensemble. Little progress was found for fields that were already quite well simulated such as near-surface air temperature and TOA outgoing longwave radiation. For precipitation, the intermodel spread is reduced from CMIP3 to CMIP5 and CMIP6, particularly because the worst performing models improved significantly. Additionally, there is a continuous improvement of the pattern correlation from CMIP3 to CMIP6 in all variables. The short-wave cloud radiative effect shows large improvements in CMIP6 regarding the correlation and also the multimodel spread. In CMIP3 and CMIP5, shortwave cloud radiative effect was relatively poorly simulated with a large intermodel spread. Concerning sea level pressure, there is an improvement from CMIP5 to CMIP6 but the wide intermodel spread has not been reduced significantly.

6. Effective Climate Sensitivity

Since the release of the first CMIP6 simulations one of the most discussed topics is the higher ECS reported in some of the models (Forster et al., 2019; Meehl et al., 2020). ECS is an important metric for assessing the future warming sensitivity of the climate system to increasing concentrations of CO_2, which is an important constraint on the total amount of greenhouse gases, in particular CO_2, that can be emitted before a given global mean warming target is exceeded. ECS provides a single number, defined as the change in global mean surface air temperature resulting from a doubling of atmospheric CO_2 concentration compared to preindustrial conditions, once the climate has reached a new equilibrium (Gregory et al., 2004). For this study we used the common assumption by the Gregory method of extrapolating the relationship between the changes in near-surface temperature and the changes in the net downward radiation flux at TOA (Gregory et al., 2004). This method is unable to represent nonlinearities in the climate response and tends to

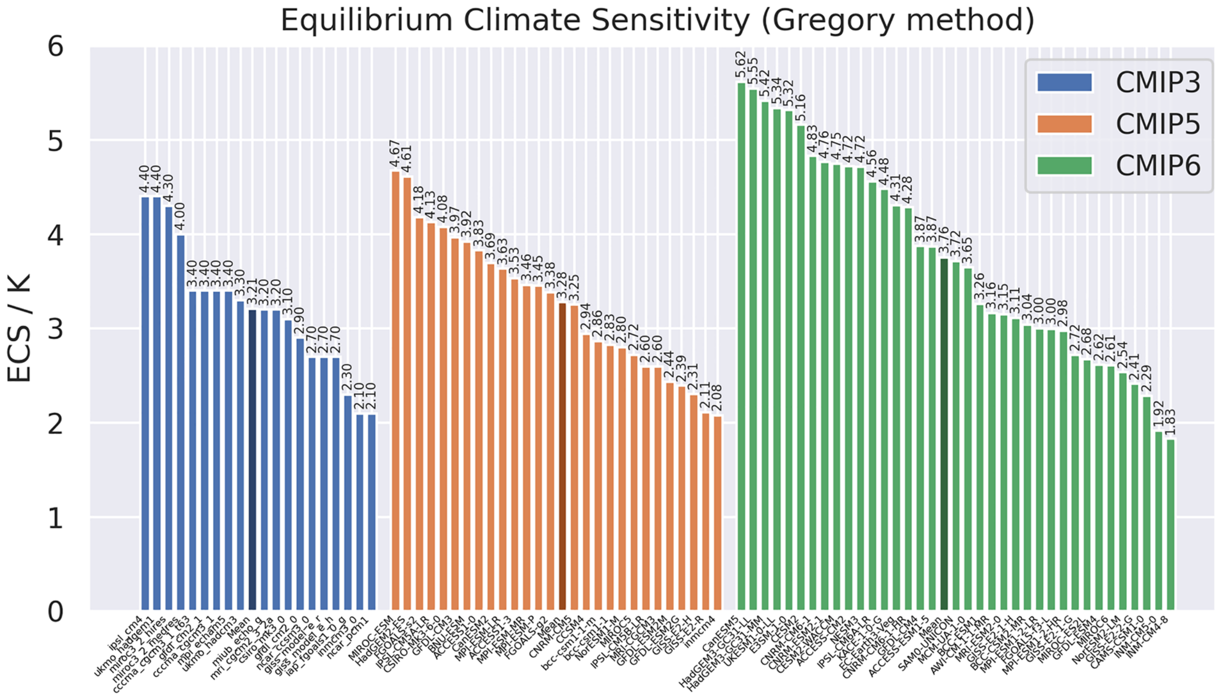

Figure 8. Effective climate sensitivity (ECS) calculated for CMIP3 (blue), CMIP5 (orange), and CMIP6 (green) models using the method from Gregory et al. (2004). The ensemble means are indicated by a darker shading of the corresponding bars.

underestimate the true ECS obtained from equilibrating the climate models (Rugenstein et al., 2020). However, since only a small subset of the CMIP models provides the long-running simulations necessary for the calculation of the true ECS, we use the Gregory method for an approximate, yet consistent ECS assessment for all climate models. The ESMValTool offers the flexibility to adjust the ECS calculation for example by changing the first year and the length of the time period used for calculating the slope of the Gregory relationship. This allows to repeat this study with different settings if needed.

The modeled range of ECS of 2.1 to 4.4 K in CMIP3, which was quite similar in CMIP5 (2.1 to 4.7 K), has increased to 1.8 to 5.6 K in CMIP6 (Figure 8). Consistent with this, the ECS multimodel mean has also increased from 3.2 K in CMIP3 and CMIP5 to almost 3.8 K in CMIP6. The increased range in ECS in CMIP6 suggests an increased uncertainty in this metric compared to previous CMIP phases, which might lead to reduced trust in the models' projections of future climate by some stakeholders and decision makers. It is therefore critically important to understand the reasons for the increased span in ECS given by the latest generation of CMIP models. In addition to Meehl et al. (2020), several modeling groups have already published studies confirming higher ECS values in their CMIP6 models (Andrews et al., 2019; Gettelman et al., 2019; Wyser et al., 2019).

Numerous improvements to the underpinning physical, chemical, and biological processes have been developed and implemented in the new CMIP6 models. These result in models that are capable to represent the coupled climate system in more detail. Some of these improvements influence the ECS in the models (Forster et al., 2019). Meehl et al. (2020) give possible explanations for the occurrence of high ECS values in some of the models, with coupled cloud microphysical and aerosol developments potentially being a common factor. Besides cloud feedbacks, other main contributors to ECS are, for example, the water vapor—lapse rate feedback and the snow/ice albedo feedback. Cloud feedbacks play a particularly important role because (i) they remain the largest contributor to the spread of ECS across models (Flato et al., 2013; Zelinka et al., 2020) and (ii) a number of models have specifically increased the degree of complexity/detail with respect to mixed phase clouds (Bodas-Salcedo et al., 2019; Gettelman et al., 2019; Mulcahy et al., 2020; Williams et al., 2020). Further studies are required to better understand the higher ECSs in CMIP6 relative to CMIP5.

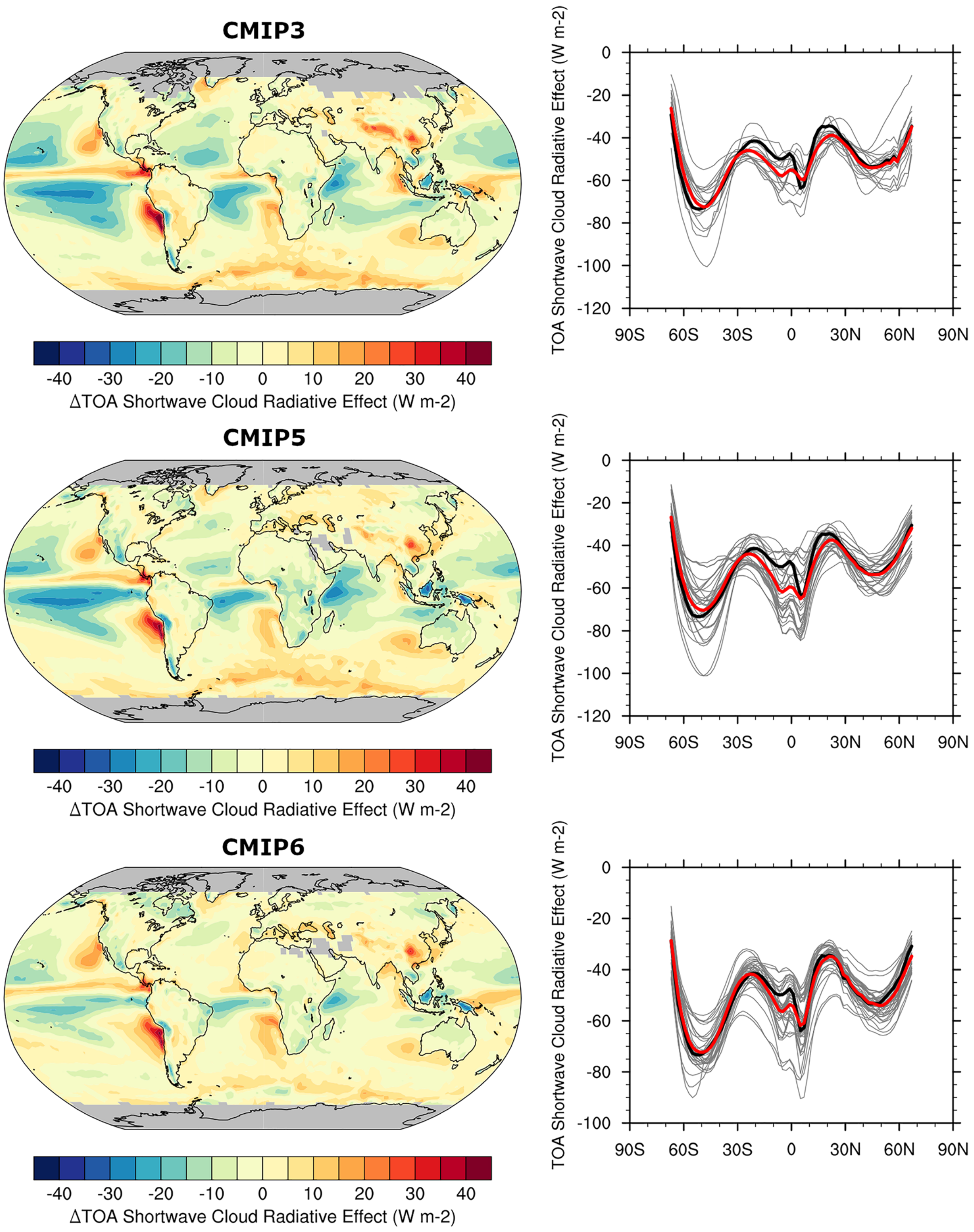

Figure 9. Annual mean shortwave cloud radiative effect from CMIP3 (1980–1999), CMIP5 (1986–2005), and CMIP6 (1995–2014) compared against the Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) Version 2.7 data set (Loeb et al., 2012). Left column: Geographical distributions of the differences between the multimodel means and CERES-EBAF (bias). Right column: Zonal averages from the individual models (gray lines), the multimodel-mean (red lines) and the observational data set (black lines).

Improvements to the representation mixed phase clouds in some CMIP6 models reduces a long-standing model bias of too little supercooled liquid water (and conversely too large amount of ice crystals) in low-level, midlatitude clouds, particularly over the relatively pristine Southern Ocean (Bodas-Salcedo et al., 2016; McCoy et al., 2016). These developments also improve the representation of both cloud microphysical structure and cloud radiative impacts in these regions (Hyder et al., 2018; Kay et al., 2016). Earlier models, such as

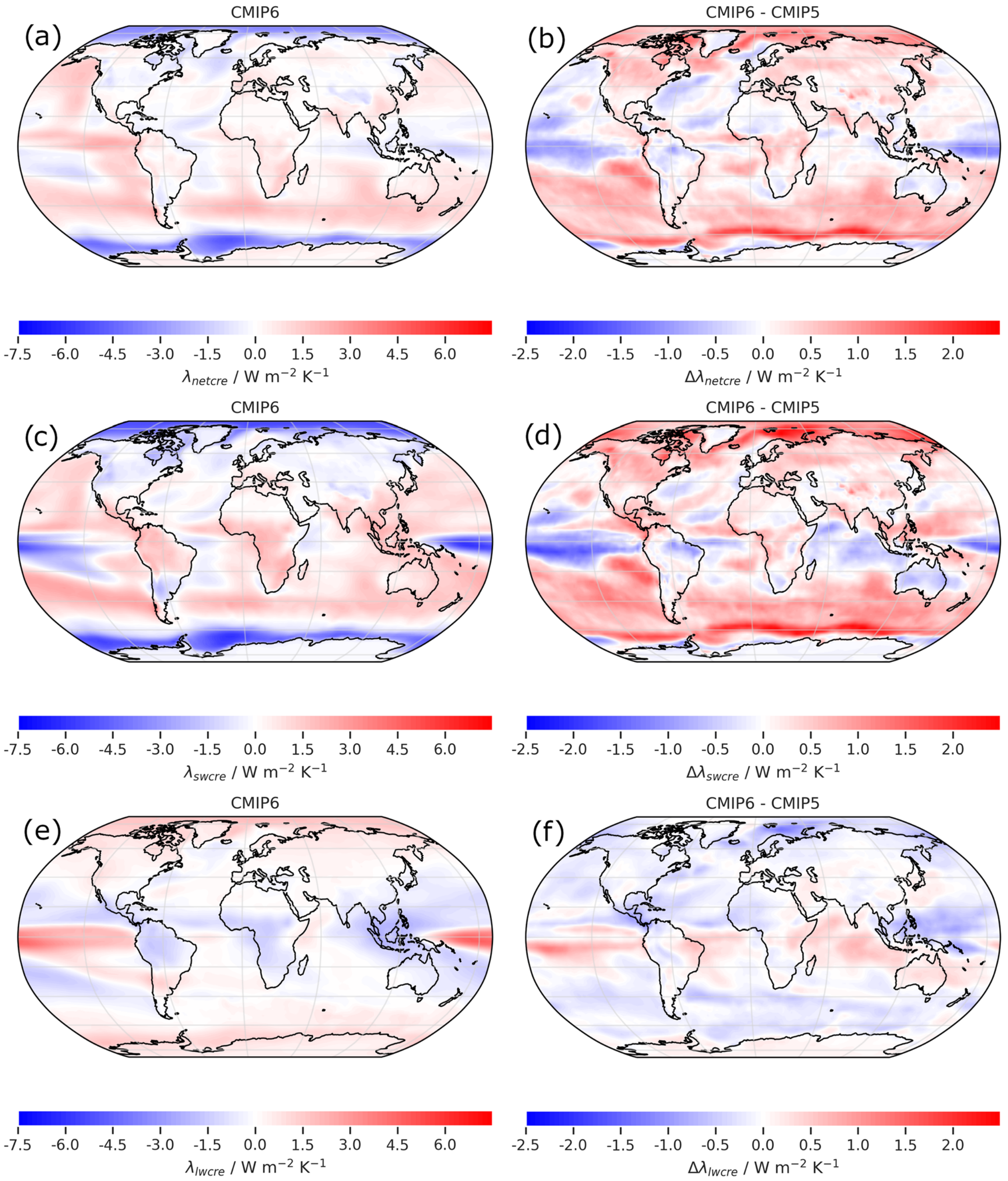

Figure 10. Cloud (a) net, (c) shortwave, and (e) longwave feedback parameter for CMIP6 (multimodel mean) and the differences to the CMIP5 multimodel mean for cloud (b) net, (d) shortwave, and (f) longwave cloud feedback.

in CMIP5 and CMIP3, exhibited a large negative SW cloud feedback over the Southern Ocean as predominantly ice clouds melted to become liquid clouds as the simulated climate warmed (McCoy et al., 2015). For a given water content, a cloud consisting of (physically smaller) liquid droplets will be more reflective to solar radiation than the "same" cloud composed of (larger) ice crystals. Furthermore, a predominantly liquid cloud will also tend to precipitate less than a cloud composed of both ice and liquid, resulting in more water staying in the liquid cloud. Earlier (CMIP3/CMIP5) models exhibited a widespread (erroneous) tendency over the Southern Ocean to go from predominantly ice clouds in the present-day period to liquid clouds in the future. This cloud phase change provided a relatively strong negative shortwave feedback on warming (through a reduction in cloud reflectivity in the simulated future). This negative feedback is removed (or significantly reduced) in those CMIP6 models that simulate predominantly liquid clouds for the present day over the Southern Ocean. This negative SW cloud feedback acts to balance other (mainly tropical and subtropical) positive cloud feedbacks, reducing the overall global net cloud feedback (Zelinka et al., 2016). The size of this negative (cloud phase change) feedback has long been questioned due to the known systematic cloud phase bias seen in many models (McCoy et al., 2015; Tan et al., 2016). Improving the microphysical structure of mixed-phase clouds acts to reduce this negative SW feedback as the climate warms, increasing the net global cloud feedback (Bodas-Salcedo et al., 2019; Tan et al., 2016) and the resulting ECS in those models (Andrews et al., 2019; Gettelman et al., 2019).

Figure 9 shows that in CMIP6, the simulated TOA shortwave cloud radiative effect agrees better with observations than in previous CMIP phases (Figure 9). The main improvement in CMIP6 compared with previous phases of CMIP is a reduced (less negative) bias in the tropics and over the Southern Ocean. The latter collocated with the aforementioned cloud phase negative shortwave feedback.

The geographical distribution of the net cloud feedback parameter, defined as changes in the sum of shortwave and longwave cloud radiative effect per degree of surface warming is dominated in many regions by the shortwave component (Figures 10a and 10c). The sign change at around 60°S seen in the shortwave cloud feedback is indicative of where models are switching, in their preindustrial and present-day experiments, from simulating clouds almost totally composed of liquid droplets to clouds with an increasing ice component. With increasing latitude there is an increasing ice component in model clouds that will support a negative shortwave feedback on warming.

Figure 10d supports the results of Zelinka et al. (2020) that there is an increase in the shortwave cloud feedback parameter over the Southern Ocean in CMIP6 compared with CMIP5 (in many regions understood as a decrease, or even sign change, in the size of a negative shortwave cloud feedback). Zelinka et al. (2020) found that the distribution of net cloud feedback is shifted toward larger positive values in CMIP6 due to a stronger positive (reduced negative) low-level cloud feedback, mainly in the extratropics. The CMIP6 models show weaker increases in extratropical low-level cloud cover and associated liquid water content with increasing surface temperature than previous model generations. This primarily arises from an increase in the liquid condensate fraction (LFC) simulated in these clouds for the preindustrial and present-day periods (Zelinka et al., 2020), leading to the aforementioned reduction in cloud phase change on warming. A higher cloud feedback contributes to an increase in climate sensitivity and could be one possible explanation for the high climate sensitivity values of some CMIP6 models.

7. Summary

In this study, we evaluated multimodel ensembles from three different phases of CMIP, namely CMIP3, CMIP5 and CMIP6. Improvements or changes in model performance from one CMIP phase to the next are typically a combination of different factors such as an increasing spatial and vertical resolution, a more complete and also a more detailed representation of individual ESM components and the inclusion of additional Earth system processes that could be added in recent years as increasing computing power became available. In addition, also input data including prescribed emissions and forcings were continuously refined and further developed (Eyring, Bony, et al., 2016; Taylor et al., 2012). These changes in combination with modifications of the experiment design over time make a direct one by one comparison of the model results among different CMIP phases difficult if not impossible. We therefore focused on ensemble average results as one possible representation of the state-of-the-art climate modeling at the time of a particular CMIP phase in order to assess the general progress in the field over the last two decades. For this we compared the model results from CMIP3, 5 and 6 for present-day climate with observations that serve as one possible benchmark for the overall model performance. The main aim was to assess the different generations of climate models as a whole instead of tracking the progress made by individual models. For this, we analyzed data from the historical CMIP6 simulations published to the ESGF in comparison with observations and reanalyses as well as with results from CMIP3 and CMIP5. Additionally, we evaluated some results from HighResMIP to assess the potential improvements achieved by increasing the horizontal model resolution.

To analyze how the performance of different generations of CMIP models compared to observations has changed relative to each other, we have used the ESMValTool for the production of all figures in this study. It enables a comprehensive evaluation of the models and ensures as an open source software provenance and traceability. One of the topics widely discussed even outside of the climate science community was the apparent "failure" of the CMIP5 models to reproduce the warming hiatus seen in observations of the global mean warming rates from 1998 to 2013. Because of the high attention this topic received, there were even potential implications on the public perception of the trustworthyness of climate models and climate projections in general. It has been shown that the hiatus was likely predominantly a result of internal climate variability with the phase of the IPO playing an important role. The uninitialized historical CMIP5 model runs cannot be expected to reproduce the exact timing of effects caused by internal variability as seen in observations. In fact, a small number of CMIP5 model simulations were, by chance, in a negative IPO phase at the right time and able to simulate the observed pause of the increase in global warming rates. Now, CMIP6 models show the observed accelerated temperature increase in recent years and agree quite well with the observed mean global warming in the 2010s. Some CMIP6 models, however, also show a cooling in the second half of the twentieth century and a too large increase in near-surface temperatures in the last years which might be related to a too strong aerosol-related ERF. This needs to be further investigated in order to fully understand the driving mechanisms of this potentially overestimated sensitivity to the prescribed aerosol emissions.

The CMIP6 results currently available show that the latest generation of CMIP models have a similar or even slightly higher skill in reproducing observed large-scale mean surface temperature and precipitation patterns as their CMIP3 and CMIP5 predecessors. CMIP6 models have an increased degree of freedom by including more processes and couplings, primarily aimed at being able to better simulate future feedbacks (e.g., nitrogen effects of terrestrial carbon uptake or permafrost processes). All these additions make the models better "fit for purpose," if the purpose is simulating future global change. But the increased degree of freedom has the potential to increase model biases. A reduction of some of the long-standing systematic model biases for instance over high-elevated regions, the North Atlantic and Southern Ocean, and upwelling regions is found particularly in the high horizontal resolution models contributing to HighResMIP. Other biases however, notably in Southern Ocean, seem to be more stubborn. Vertical distributions of key variables such as temperature, water vapor and zonal wind speed also show improvements throughout the three different CMIP phases. While most of the long-standing model biases are still present in CMIP6, their amplitude is often smaller than in CMIP3 and CMIP5.

The performance metrics (portrait diagram) and the correlation patterns of some important fields such as TOA radiative fluxes, temperature, precipitation, and sea level pressure show some overall improvements across the different CMIP ensembles with a reduced intermodel spread and higher average skill of the CMIP6 ensemble (RMSD, pattern correlation).

A maybe surprising result from CMIP6 is the high ECS in some of the models resulting in an even larger spread in ECS than the large range of values obtained from previous generations of climate models. This has been already discussed in first studies and the exact probably model-specific reasons need to be understood in detail as the increased spread in ECS potentially shows an increased uncertainty in this important climate metric. First studies suggest that the causes might be improvements in the representation of mixed-phased clouds, which leads to changes in cloud feedbacks and in the shortwave component of the cloud feedback in particular. It is noteworthy that cloud-radiation interactions and in particular the shortwave cloud radiative effect in CMIP6 models are closer to observations than in previous generations of climate models. As ECS depends on numerous and interacting feedbacks, improvements in one specific variable or physical process can potentially lead to less error compensation and thus more spread in such complex quantities as ECS. A realistic representation of clouds, however, remains a challenge in current climate modeling. Here, further model improvements, stemming from higher resolution (Palmer & Stevens, 2019) or completely novel approaches to parameterize clouds and convection in climate models such as, for instance, machine learning based cloud parametrizations (Gentine et al., 2018; Rasp et al., 2018) are required to make further progress toward more realistic simulations of clouds with climate models.

Data Availability Statement

CMIP model data are available freely and publicly from the Earth System Grid Federation (ESGF, https://esgf.llnl.gov) and listed in Tables 1–3. Observations used in the evaluation are detailed in Table 5 of the manuscript. Observational data sets available through the observations for Model Intercomparisons Project (obs4MIPs, https://esgf-node.llnl.gov/projects/obs4mips/) can be downloaded freely from the ESGF and used directly with the ESMValTool. For all other observational data sets, the ESMValTool provides a collection of scripts (NCL and Python) with downloading and processing instructions to recreate the data sets used in this publication.

Acknowledgments

Funding for this study was provided by the European Union's Framework Programme Horizon 2020 for Research and Innovation under Grant Agreement No. 641816, the Coordinated Research in Earth Systems and Climate: Experiments, kNowledge, Dissemination and Outreach (CRESCENDO) project, and under Grant Agreement No. 641727, the PRIMAVERA project. Funding was also received from the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung) CMIP6-DICAD-TP2 project and from the European Space Agency (ESA) under its Climate Change Initiative (CCI) support for the Climate Modelling User Group (CMUG). Portions of this study were supported by the Regional and Global Model Analysis (RGMA) component of the Earth and Environmental System Modeling Program of the U.S. Department of Energy's Office of Biological & Environmental Research (BER) via National Science Foundation IA 1844590. This work also was supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977. The corresponding recipe that can be used to reproduce the figures of this paper will be included in the ESMValTool v2.0. The ESMValTool v2.0 is released under the Apache License, Version 2.0. The latest release of ESMValTool v2.0 is publicly available on Zenodo at https://doi.org/10.5281/zenodo.3970975 (Andela et al., 2020). The ESMValTool and its workflow manager/preprocessor (ESMValCore) are developed on the GitHub repositories available at https://github.com/ESMValGroup (last access: 5 October 2020). We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF. We thank Kevin Debeire (DLR) for his help with the portrait diagram and the entire ESMValTool development team for the great work. The computational resources of the Deutsches Klimarechenzentrum (DKRZ, Hamburg) were essential for developing and testing the ESMValTool used to produce the figures in this paper as well as to do the analysis presented here. We would also like to acknowledge the

References

(2004). The New GFDL Global Atmosphere and Land Model AM2–LM2: Evaluation with Prescribed SST Simulations. *Journal of Climate*, *17*(24), 4641–4673. https://doi.org/10.1175/jcli-3223.1

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P. P., Janowiak, J., et al. (2003). The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present). *Journal of Hydrometeorology*, *4*(6), 1147–1167. https://doi.org/10.1175/1525-7541(2003)004<1147:Tvgpcp>2.0.Co;2

Alekseev, V. A., Volodin, E. M., Galin, V. Y., Dymnikov, V. P., & Lykossov, V. N. (1998). Modelling of the present-day climate by the RAS atmospheric model "DNM GCM.", Institute of Numerical Mathematics. (p. 200).

Andela, B., Brötz, B., de Mora, L., Drost, N., Eyring, V., Koldunov, N., et al. (2020). ESMValTool (version v2.0.0). https://doi.org/10.5281/zenodo.3970975

Andrews, T., Andrews, M. B., Bodas-Salcedo, A., Jones, G. S., Kulhbrodt, T., Manners, J., et al. (2019). Forcings, feedbacks and climate sensitivity in HadGEM3-GC3. 1 and UKESM1. *Journal of Advances in Modeling Earth Systems*, *11*, 4377–4394. https://doi.org/10.1029/2019MS001866

Arora, V. K., Scinocca, J. F., Boer, G. J., Christian, J. R., Denman, K. L., Flato, G. M., et al. (2011). Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases. *Geophysical Research Letters*, *38*, L05805. https://doi.org/10.1029/2010GL046270

Bentsen, M., Bethke, I., Debernard, J. B., Iversen, T., Kirkevåg, A., Seland, Ø., et al. (2013). The Norwegian Earth System Model, NorESM1-M—Part 1: Description and basic evaluation of the physical climate. *Geoscientific Model Development*, *6*(3), 687–720. https://doi.org/10.5194/gmd-6-687-2013

Bi, D. H., Dix, M., Marsland, S. J., O'Farrell, S., Rashid, H., Uotila, P., et al. (2013). The ACCESS coupled model: Description, control climate and evaluation. *Australian Meteorological and Oceanographic*, *63*(1), 41–64. https://doi.org/10.22499/2.6301.004

Bleck, R. (2002). An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates. *Ocean Model*, *4*(1), 55–88. https://doi.org/10.1016/S1463-5003(01)00012-9

Bodas-Salcedo, A., Hill, P. G., Furtado, K., Williams, K. D., Field, P. R., Manners, J. C., et al. (2016). Large contribution of supercooled liquid clouds to the solar radiation budget of the Southern Ocean. *Journal of Climate*, *29*(11), 4213–4228.

Bodas-Salcedo, A., Mulcahy, J. P., Andrews, T., Williams, K. D., Ringer, M. A., Field, P. R., & Elsaesser, G. S. (2019). Strong dependence of atmospheric feedbacks on mixed-phase microphysics and aerosol-cloud interactions in HadGEM3. *Journal of Advances in Modeling Earth Systems*, *11*, 1735–1758. https://doi.org/10.1029/2019MS001688

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., et al. (2020). Presentation and Evaluation of the IPSL-CM6A-LR Climate Model. *Journal of Advances in Modeling Earth Systems*, *12*. https://doi.org/10.1029/2019ms002010

Caldwell, P. M., Mametjanov, A., Tang, Q., van Roekel, L. P., Golaz, J. C., Lin, W., et al. (2019). The DOE E3SM coupled model version 1: Description and results at high resolution. *Journal of Advances in Modeling Earth Systems*, *11*, 4095–4146. https://doi.org/10.1029/2019MS001870

Cao, J., Wang, B., Young-Min, Y., Ma, L., Li, J., Sun, B., et al. (2018). The NUIST Earth System Model (NESM) version 3: Description and preliminary evaluation. *Geoscientific Model Development*, *11*(7), 2975–2993. https://doi.org/10.5194/gmd-11-2975-2018

Cherchi, A., Fogli, P. G., Lovato, T., Peano, D., Iovino, D., Gualdi, S., et al. (2019). Global mean climate and main patterns of variability in the CMCC-CM2 coupled model. *Journal of Advances in Modeling Earth Systems*, *11*, 185–209. https://doi.org/10.1029/2018MS001369

Cionni, I., Eyring, V., Lamarque, J. F., Randel, W. J., Stevenson, D. S., Wu, F., et al. (2011). Ozone database in support of CMIP5 simulations: Results and corresponding radiative forcing. *Atmospheric Chemistry and Physics*, *11*(21), 11,267–11,292. https://doi.org/10.5194/acp-11-11267-2011

Collins, W. D., Bitz, C. M., Blackmon, M. L., Bonan, G. B., Bretherton, C. S., Carton, J. A., et al. (2006). The Community Climate System Model version 3 (CCSM3). *Journal of Climate*, *19*(11), 2122–2143. https://doi.org/10.1175/JCLI3761.1

Collins, W. J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., et al. (2011). Development and evaluation of an Earth-System model-HadGEM2. *Geoscientific Model Development*, *4*(4), 1051–1075. https://doi.org/10.5194/gmd-4-1051-2011

Copernicus Climate Change Service (C3S) (2017). ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate, edited, Copernicus Climate Change Service Climate Data Store (CDS). https://cds.climate.copernicus.eu/cdsapp#!/home

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., et al. (2020). The Community Earth System Model Version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, *12*. https://doi.org/10.1029/2019ms001916

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

Delworth, T., Stouffer, R., Dixon, K., Spelman, M., Knutson, T., Broccoli, A., et al. (2002). Review of simulations of climate variability and change with the GFDL R30 coupled climate model. *Climate Dynamics*, *19*(7), 555–574. https://doi.org/10.1007/s00382-002-0249-5

Dittus, A. J., Hawkins, E., Wilcox, L. J., Sutton, R. T., Smith, C. J., Andrews, M. B., & Forster, P. M. (2020). Sensitivity of Historical Climate Simulations to Uncertain Aerosol Forcing. *Geophysical Research Letters*, *47*. https://doi.org/10.1029/2019GL085806

Docquier, D., Grist, J. P., Roberts, M. J., Roberts, C. D., Semmler, T., Ponsoni, L., et al. (2019). Impact of model resolution on Arctic sea ice and North Atlantic Ocean heat transport. *Climate Dynamics*, *53*(7–8), 4989–5017. https://doi.org/10.1007/s00382-019-04840-y

Donner, L. J., Wyman, B. L., Hemler, R. S., Horowitz, L. W., Ming, Y., Zhao, M., et al. (2011). The dynamical core, physical parameterizations, and basic simulation characteristics of the atmospheric component AM3 of the GFDL global coupled model CM3. *Journal of Climate*, *24*(13), 3484–3519. https://doi.org/10.1175/2011jcli3955.1

Dufresne, J. L., Foujols, M. A., Denvil, S., Caubel, A., Marti, O., Aumont, O., et al. (2013). Climate change projections using the IPSL-CM5 Earth System Model: From CMIP3 to CMIP5. *Climate Dynamics*, *40*(9–10), 2123–2165. https://doi.org/10.1007/s00382-012-1636-1

obs4MIPs activity, a project initiated by the National Aeronautical and Space Administration (NASA) and U.S. Department of Energy (DOE), with governance provided by the World Climate Research Program's (WCRP) Data Advisory Council (WDAC) as well as the Earth System Federation Grid.

Dunne, J., Horowitz, L., Held, I., Krasting, J., John, J., Malyshev, S., et al. (2019). The GFDL Earth System Model version 4.1 (GFDL-ESM 4. 1): Model description and simulation characteristics. *Journal of Advances in Modeling Earth Systems*, *11*, 3167–3211. https://doi.org/10.1029/2019MS002015

England, M. H., McGregor, S., Spence, P., Meehl, G. A., Timmermann, A., Cai, W., et al. (2014). Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus. *Nature Climate Change*, *4*(3), 222–227. https://doi.org/10.1038/nclimate2106

Eyring, V., Bock, L., Lauer, A., Righi, M., Schlund, M., Andela, B., et al. (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. *Geoscientific Model Development*, *13*(7), 3383–3438. https://doi.org/10.5194/gmd-13-3383-2020

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

Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., et al. (2019). Taking climate model evaluation to the next level. *Nature Climate Change*, *9*(2), 102–110. https://doi.org/10.1038/s41558-018-0355-y

Eyring, V., Harris, N. R. P., Rex, M., Shepherd, T. G., Fahey, D. W., Amanatidis, G. T., et al. (2005). A strategy for process-oriented validation of coupled chemistry–climate models. *Bulletin of the American Meteorological Society*, *86*(8), 1117–1134. https://doi.org/10.1175/bams-86-8-1117

Eyring, V., Righi, M., Lauer, A., Evaldsson, M., Wenzel, S., Jones, C., et al. (2016). ESMValTool (v1.0)—A community diagnostic and performance metrics tool for routine evaluation of Earth system models in CMIP. *Geoscientific Model Development*, *9*(5), 1747–1802. https://doi.org/10.5194/gmd-9-1747-2016

Ferraro, R., Waliser, D. E., Gleckler, P., Taylor, K. E., & Eyring, V. (2015). Evolving Obs4MIPs to support Phase 6 of the Coupled Model Intercomparison Project (CMIP6). *Bulletin of the American Meteorological Society*, *96*(8), Es131–Es133. https://doi.org/10.1175/Bams-D-14-00216.1

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., et al. (2013). Evaluation of climate models. In *Climate change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 741–866). Cambridge: Cambridge University Press.

Flynn, C. M., & Mauritsen, T. (2020). On the climate sensitivity and historical warming evolution in recent coupled model ensembles. *Atmospheric Chemistry and Physics Discussions*, *2020*, 1–26. https://doi.org/10.5194/acp-2019-1175

Fogli, P. G., Manzini, E., Vichi, M., Alessandri, A., Patara, L., Gualdi, S., et al. (2009). INGV-CMCC carbon: A carbon cycle Earth System Model, CMCC online RP0061: http://www.cmcc.it/publications/rp0061-ingv-cmcc-carbon-icc-a-carbon-cycle-earth-system-model

Forster, P. M., Maycock, A. C., McKenna, C. M., & Smith, C. J. (2019). Latest climate models confirm need for urgent mitigation. *Nature Climate Change*, *10*(1), 7–10.

Fyfe, J. C., Meehl, G. A., England, M. H., Mann, M. E., Santer, B. D., Flato, G. M., et al. (2016). Making sense of the early-2000s warming slowdown. *Nature Climate Change*, *6*(3), 224–228. https://doi.org/10.1038/nclimate2938

Galin, V. Y., Volodin, E. M., & Smyshliaev, S. P. (2003). Atmosphere general circulation model of INM RAS with ozone dynamics. *Russian Meteorology and Hydrology*, *5*, 13–22.

Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., et al. (2011). The Community Climate System Model version 4. *Journal of Climate*, *24*(19), 4973–4991. https://doi.org/10.1175/2011jcli4083.1

Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., & Yacalis, G. (2018). Could machine learning break the convection parameterization deadlock? *Geophysical Research Letters*, *45*, 5742–5751. https://doi.org/10.1029/2018GL078202

Gettelman, A., Hannay, C., Bacmeister, J. T., Neale, R. B., Pendergrass, A. G., Danabasoglu, G., et al. (2019). High climate sensitivity in the Community Earth System Model version 2 (CESM2). *Geophysical Research Letters*, *46*, 8329–8337. https://doi.org/10.1029/2019GL083978

Gleckler, P. J., Taylor, K. E., & Doutriaux, C. (2008). Performance metrics for climate models. *Journal of Geophysical Research*, *113*, D06104. https://doi.org/10.1029/2007JD008972

Golaz, J. C., Caldwell, P. M., van Roekel, L. P., Petersen, M. R., Tang, Q., Wolfe, J. D., et al. (2019). The DOE E3SM coupled model version 1: Overview and evaluation at standard resolution. *Journal of Advances in Modeling Earth Systems*, *11*, 2089–2129. https://doi.org/10.1029/2018MS001603

Gordon, C., Cooper, C., Senior, C. A., Banks, H., Gregory, J. M., Johns, T. C., et al. (2000). The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dynamics*, *16*(2–3), 147–168. https://doi.org/10.1007/s003820050010

Gordon, H. B., Rotstayn, L. D., McGregor, J. L., Dix, M. R., Kowalczyk, E. A., O'Farrell, S. P., et al. (2002). The CSIRO Mk3 Climate System Model, CSIRO Atmospheric Research Technical Paper(60).

Gregory, J. M., Ingram, W., Palmer, M., Jones, G., Stott, P., Thorpe, R., et al. (2004). A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical Research Letters*, *31*, L03205. https://doi.org/10.1029/2003GL018747

Gutjahr, O., Putrasahan, D., Lohmann, K., Jungclaus, J. H., von Storch, J. S., Brüggemann, N., et al. (2019). Max Planck Institute Earth System Model (MPI-ESM 1. 2) for the High-Resolution Model Intercomparison Project (HighResMIP). *Geoscientific Model Development*, *12*(7), 3241–3281. https://doi.org/10.5194/gmd-12-3241-2019

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., et al. (2016). High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6. *Geoscientific Model Development*, *9*(11), 4185–4208. https://doi.org/10.5194/gmd-9-4185-2016

Hajima, T., Watanabe, M., Yamamoto, A., Tatebe, H., Noguchi, M. A., Abe, M., et al. (2020). Development of the MIROC-ES2L Earth system model and the evaluation of biogeochemical processes and feedbacks. *Geoscientific Model Development*, *13*(5), 2197–2244. https://doi.org/10.5194/gmd-13-2197-2020

Harada, Y., Kamahori, H., Kobayashi, C., Endo, H., Kobayashi, S., Ota, Y., et al. (2016). The JRA-55 reanalysis: Representation of atmospheric circulation and climate variability. *Journal of the Meteorological Society of Japan*, *94*(3), 269–302.

Hasumi, H., & Emori, S. (2004). K-1 coupled model (MIROC) description, Center for Climate System Research K-1 Tech. Rep.(1). (p. 34).

Hazeleger, W., Wang, X., Severijns, C., Ştefănescu, S., Bintanja, R., Sterl, A., et al. (2012). EC-Earth V2.2: Description and validation of a new seamless earth system prediction model. *Climate Dynamics*, *39*(11), 2611–2629. https://doi.org/10.1007/s00382-011-1228-5

Heidinger, A. K., Foster, M. J., Walther, A., & Zhao, X. (2014). The pathfinder atmospheres–extended AVHRR climate dataset. *Bulletin of the American Meteorological Society*, *95*(6), 909–922.

Held, I. M., Guo, H., Adcroft, A., Dunne, J. P., Horowitz, L. W., Krasting, J., et al. (2019). Structure and performance of GFDL's CM4.0 climate model. *Journal of Advances in Modeling Earth Systems*, *11*, 3691–3727. https://doi.org/10.1029/2019MS001829

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., et al. (2018). Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS). *Geoscientific Model Development*, *11*(PNNL-SA-123932(1), 369–408. https://doi.org/10.5194/gmd-11-369-2018

Hourdin, F., Grandpeix, J. Y., Rio, C., Bony, S., Jam, A., Cheruy, F., et al. (2013). LMDZ5B: The atmospheric component of the IPSL climate model with revisited parameterizations for clouds and convection. *Climate Dynamics*, *40*(9–10), 2193–2222. https://doi.org/10.1007/s00382-012-1343-y

Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J. C., Balaji, V., Duan, Q., et al. (2017). The art and science of climate model tuning. *Bulletin of the American Meteorological Society*, *98*(3), 589–602. https://doi.org/10.1175/Bams-D-15-00135.1

Hourdin, F., Musat, I., Bony, S., Braconnot, P., Codron, F., Dufresne, J. L., et al. (2006). The LMDZ4 general circulation model: Climate performance and sensitivity to parametrized physics with emphasis on tropical convection. *Climate Dynamics*, *27*(7–8), 787–813. https://doi.org/10.1007/s00382-006-0158-0

Huffman, G. J., & Bolvin, D. T. (2012). GPCP version 2.2 SG combined precipitation data set documentation, edited, NASA GSFC, available at ftp://precip.gsfc.nasa.gov/pub/gpcp-v2.2/doc/V2.2_doc.pdf, last access: January 2016.

Hyder, P., Edwards, J. M., Allan, R. P., Hewitt, H. T., Bracegirdle, T. J., Gregory, J. M., et al. (2018). Critical Southern Ocean climate model biases traced to atmospheric model cloud errors. *Nature Communications*, *9*.

Ji, D., Wang, L., Feng, J., Wu, Q., Cheng, H., Zhang, Q., et al. (2014). Description and basic evaluation of Beijing Normal University Earth System Model (BNU-ESM) version 1. *Geoscientific Model Development*, *7*(5), 2039–2064. https://doi.org/10.5194/gmd-7-2039-2014

John, V. O., & Soden, B. J. (2007). Temperature and humidity biases in global climate models and their impact on climate feedbacks. *Geophysical Research Letters*, *34*, L18704. https://doi.org/10.1029/2007GL030429

Jones, G. S., Stott, P. A., & Christidis, N. (2013). Attribution of observed historical near–surface temperature variations to anthropogenic and natural causes using CMIP5 simulations. *Journal of Geophysical Research: Atmospheres*, *118*, 4001–4024. https://doi.org/10.1002/jgrd.50239

Jones, P. D., New, M., Parker, D. E., Martin, S., & Rigor, I. G. (1999). Surface air temperature and its changes over the past 150 years. *Reviews of Geophysics*, *37*(2), 173–199. https://doi.org/10.1029/1999RG900002

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., et al. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, *77*(3), 437–471. https://doi.org/10.1175/1520-0477(1996)077<0437:Tnyrp>2.0.Co;2

Kay, J. E., Bourdages, L., Miller, N. B., Morrison, A., Yettella, V., Chepfer, H., & Eaton, B. (2016). Evaluating and improving cloud phase in the Community Atmosphere Model version 5 using spaceborne lidar observations. *Journal of Geophysical Research: Atmospheres*, *121*, 4162–4176. https://doi.org/10.1002/2015JD024699

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., et al. (2015). The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan*, *93*(1), 5–48. https://doi.org/10.2151/jmsj.2015-001

Kuhlbrodt, T., Jones, C. G., Sellar, A., Storkey, D., Blockley, E., Stringer, M., et al. (2018). The low-resolution version of HadGEM3 GC3. 1: Development and evaluation for global climate. *Journal of Advances in Modeling Earth Systems*, *10*, 2865–2888. https://doi.org/10.1029/2018MS001370

Lamarque, J.-F., Bond, T. C., Eyring, V., Granier, C., Heil, A., Klimont, Z., et al. (2010). Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: Methodology and application. *Atmospheric Chemistry and Physics*, *10*(15), 7017–7039. https://doi.org/10.5194/acp-10-7017-2010

Lauer, A., Eyring, V., Bellprat, O., Bock, L., Gier, B. K., Hunter, A., et al. (2020). Earth System Model Evaluation Tool (ESMValTool) v2.0—Diagnostics for emergent constraints and future projections from Earth system models in CMIP. *Geoscientific Model Development*, *13*(9), 4205–4228. https://doi.org/10.5194/gmd-13-4205-2020

Lauer, A., Jones, C., Eyring, V., Evaldsson, M., Stefan, H. A., Makela, J., et al. (2018). Process-level improvements in CMIP5 models and their impact on tropical variability, the Southern Ocean, and monsoons. *Earth System Dynamics*, *9*(1), 33–67. https://doi.org/10.5194/esd-9-33-2018

Lee, J., Kim, J., Sun, M.-A., Kim, B.-H., Moon, H., Sung, H. M., et al. (2019). Evaluation of the Korea Meteorological Administration Advanced Community Earth-System model (K-ACE). *Asia-Pacific Journal of Atmospheric Sciences*, *56*(3), 381–395. https://doi.org/10.1007/s13143-019-00144-7

Li, L. J., Lin, P., Yu, Y., Wang, B., Zhou, T., Liu, L., et al. (2013). The flexible global ocean-atmosphere-land system model, grid-point version 2: FGOALS-g2. *Advances in Atmospheric Sciences*, *30*(3), 543–560. https://doi.org/10.1007/s00376-012-2140-6

Loeb, N. G., Lyman, J. M., Johnson, G. C., Allan, R. P., Doelling, D. R., Wong, T., et al. (2012). Observed changes in top-of-the-atmosphere radiation and upper-ocean heating consistent within uncertainty. *Nature Geoscience*, *5*(2), 110–113. https://doi.org/10.1038/Ngeo1375

Ma, L., Hurtt, G. C., Chini, L. P., Sahajpal, R., Pongratz, J., Frolking, S., et al. (2019). Global transition rules for translating land-use change (LUH2) to land-cover change for CMIP6 using GLM2. *Geoscientific Model Development Discussion*, *2019*, 1–30. https://doi.org/10.5194/gmd-2019-146

Manzini, E., Karpechko, A. Y., Anstey, J., Baldwin, M. P., Black, R. X., Cagnazzo, C., et al. (2014). Northern winter climate change: Assessment of uncertainty in CMIP5 projections related to stratosphere-troposphere coupling. *Journal of Geophysical Research: Atmospheres*, *119*, 7979–7998. https://doi.org/10.1002/2013JD021403

Marsh, D. R., Mills, M. J., Kinnison, D. E., Lamarque, J. F., Calvo, N., & Polvani, L. M. (2013). Climate change from 1850 to 2005 simulated in CESM1(WACCM). *Journal of Climate*, *26*(19), 7372–7391.

Martin, G. M., Ringer, M. A., Pope, V. D., Jones, A., Dearden, C., & Hinton, T. J. (2006). The physical properties of the atmosphere in the new Hadley Centre Global Environmental Model (HadGEM1), Part I: Model description and global climatology. *Journal of Climate*, *19*(7), 1274–1301. https://doi.org/10.1175/Jcli3636.1

Matthes, K., Funke, B., Andersson, M. E., Barnard, L., Beer, J., Charbonneau, P., et al. (2017). Solar forcing for CMIP6 (v3.2). *Geoscientific Model Development*, *10*(6), 2247–2302. https://doi.org/10.5194/gmd-10-2247-2017

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., et al. (2019). Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM 1.2) and Its Response to Increasing CO2. *Journal of Advances in Modeling Earth Systems*, *11*, 998–1038. https://doi.org/10.1029/2018MS001400

Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., et al. (2012). Tuning the climate of a global model. *Journal of Advances in Modeling Earth Systems*, *4*, M00A01. https://doi.org/10.1029/2012MS000154

McCoy, D. T., Hartmann, D. L., Zelinka, M. D., Ceppi, P., & Grosvenor, D. P. (2015). Mixed-phase cloud physics and Southern Ocean cloud feedback in climate models. *Journal of Geophysical Research: Atmospheres*, *120*, 9539–9554. https://doi.org/10.1002/2015JD023603

McCoy, D. T., Tan, I., Hartmann, D. L., Zelinka, M. D., & Storelvmo, T. (2016). On the relationships among cloud cover, mixed-phase partitioning, and planetary albedo in GCMs. *Journal of Advances in Modeling Earth Systems*, *8*, 650–668. https://doi.org/10.1002/2015MS000589

Mcfarlane, N. A., Boer, G. J., Blanchet, J. P., & Lazare, M. (1992). The Canadian Climate Center 2nd-generation general-circulation model and its equilibrium climate. *Journal of Climate*, *5*(10), 1013–1044. https://doi.org/10.1175/1520-0442(1992)005<1013:Tcccsg>2.0.Co;2

Meehl, G. A., Boer, G. J., Covey, C., Latif, M., & Stouffer, R. J. (1997). Intercomparison makes for a better climate model. *Eos*, *78*(41), 445–451. https://doi.org/10.1029/97EO00276

Meehl, G. A., Boer, G. J., Covey, C., Latif, M., & Stouffer, R. J. (2000). The Coupled Model Intercomparison Project (CMIP). *Bulletin of the American Meteorological Society*, *81*(2), 313–318. https://doi.org/10.1175/1520-0477(2000)081<0313:TCMIPC>2.3.CO;2

Meehl, G. A., Covey, C., McAvaney, B., Latif, M., & Stouffer, R. J. (2005). Overview of the Coupled Model Intercomparison Project. *Bulletin of the American Meteorological Society*, *86*(1), 89–96. https://doi.org/10.1175/Bams-86-1-89

Meehl, G. A., Covey, C., Taylor, K. E., Delworth, T., Stouffer, R. J., Latif, M., et al. (2007). The WCRP CMIP3 multimodel dataset: A new era in climate change research. *Bulletin of the American Meteorological Society*, *88*(9), 1383–1394. https://doi.org/10.1175/BAMS-88-9-1383

Meehl, G. A., Senior, C. A., Eyring, V., Flato, G., Lamarque, J. F., Stouffer, R. J., et al. (2020). Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth system models. *Science Advances*, *6*(26), eaba1981. https://doi.org/10.1126/sciadv.aba1981

Meehl, G. A., Teng, H., & Arblaster, J. M. (2014). Climate model simulations of the observed early-2000s hiatus of global warming. *Nature Climate Change*, *4*(10), 898–902. https://doi.org/10.1038/nclimate2357

Meehl, G. A., Washington, W. M., Arblaster, J. M., Hu, A. X., Teng, H. Y., Kay, J. E., et al. (2013). Climate change projections in CESM1 (CAM5) compared to CCSM4. *Journal of Climate*, *26*(17), 6287–6308.

Meinshausen, M., Vogel, E., Nauels, A., Lorbacher, K., Meinshausen, N., Etheridge, D. M., et al. (2017). Historical greenhouse gas concentrations for climate modelling (CMIP6). *Geoscientific Model Development*, *10*, 2057–2116.

Merchant, C. J., Embury, O., Roberts-Jones, J., Fiedler, E., Bulgin, C. E., Corlett, G. K., et al. (2014). Sea surface temperature datasets for climate applications from Phase 1 of the European Space Agency Climate Change Initiative (SST CCI). *Geoscience Data Journal*, *1*(2), 179–191.

Morice, C. P., Kennedy, J. J., Rayner, N. A., & Jones, P. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *Journal of Geophysical Research*, *117*, D08101. https://doi.org/10.1029/2011JD017187

Mulcahy, J. P., Johnson, C., Jones, C. G., Povey, A. C., Scott, C. E., Sellar, A., et al. (2020). Description and evaluation of aerosol in UKESM1 and HadGEM3-GC3.1 CMIP6 historical simulations. Geoscientific Model Development Discussions, 2020, 1–59, doi:https://doi.org/10.5194/gmd-2019-357

Mulcahy, J. P., Jones, C., Sellar, A., Johnson, B., Boutle, I., Jones, A., et al. (2018). Improved aerosol processes and effective radiative forcing in HadGEM3 and UKESM1. *Journal of Advances in Modeling Earth Systems*, *10*, 2786–2805. https://doi.org/10.1029/2018MS001464

Muller, W. A., Jungclaus, J. H., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., et al. (2018). A higher-resolution version of the Max Planck Institute Earth System Model (MPI-ESM 1.2-HR). *Journal of Advances in Modeling Earth Systems*, *10*, 1383–1413. https://doi.org/10.1029/2017MS001217

Neubauer, D., Ferrachat, S., Siegenthaler-Le Drian, C., Stier, P., Partridge, D. G., Tegen, I., et al. (2019). The global aerosol–climate model ECHAM6.3–HAM2.3—Part 2: Cloud evaluation, aerosol radiative forcing, and climate sensitivity. *Geoscientific Model Development*, *12*(8), 3609–3639. https://doi.org/10.5194/gmd-12-3609-2019

Oudar, T., Kushner, P. J., Fyfe, J. C., & Sigmond, M. (2018). No impact of anthropogenic aerosols on early 21st century global temperature trends in a large initial-condition ensemble. *Geophysical Research Letters*, *45*, 9245–9252. https://doi.org/10.1029/2018GL078841

Oueslati, B., & Bellon, G. (2015). The double ITCZ bias in CMIP5 models: Interaction between SST, large-scale circulation and precipitation. *Climate Dynamics*, *44*(3), 585–607. https://doi.org/10.1007/s00382-015-2468-6

Palmer, T., & Stevens, B. (2019). The scientific challenge of understanding and estimating climate change. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(49), 24,390–24,395. https://doi.org/10.1073/pnas.1906691116

Park, S., Shin, J., Kim, S., Oh, E., & Kim, Y. (2019). Global climate simulated by the Seoul National University atmosphere model version 0 with a unified convection scheme (SAM0-UNICON). *Journal of Climate*, *32*(10), 2917–2949.

Pope, V. D., Gallani, M. L., Rowntree, P. R., & Stratton, R. A. (2000). The impact of new physical parametrizations in the Hadley Centre climate model: HadAM3. *Climate Dynamics*, *16*(2–3), 123–146. https://doi.org/10.1007/s003820050009

Priestley, M. D. K., Ackerley, D., Catto, J. L., Hodges, K. I., McDonald, R. E., & Lee, R. W. (2020). An overview of the Extratropical storm tracks in CMIP6 historical simulations. *Journal of Climate*, *33*(15), 6315–6343. https://doi.org/10.1175/jcli-d-19-0928.1

Rackow, T., Goessling, H. F., Jung, T., Sidorenko, D., Semmler, T., Barbi, D., & Handorf, D. (2018). Towards multi-resolution global climate modeling with ECHAM6-FESOM. Part II: Climate variability. *Climate Dynamics*, *50*(7–8), 2369–2394.

Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences*, *115*(39), 9684–9689. https://doi.org/10.1073/pnas.1810286115

Rayner, N., Parker, D. E., Horton, E., Folland, C. K., Alexander, L. V., Rowell, D., et al. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, *108*(D14), 4407. https://doi.org/10.1029/2002JD002670

Righi, M., Andela, B., Eyring, V., Lauer, A., Predoi, V., Schlund, M., et al. (2020). Earth System Model Evaluation Tool (ESMValTool) v2.0 – technical overview. *Geoscientific Model Development*, *13*(3), 1179–1199. https://doi.org/10.5194/gmd-13-1179-2020

Roberts, C. D., Senan, R., Molteni, F., Boussetta, S., Mayer, M., & Keeley, S. P. (2018). Climate model configurations of the ECMWF Integrated Forecasting System (ECMWF-IFS cycle 43r1) for HighResMIP. *Geoscientific Model Development*, *11*(9), 3681–3712.

Roberts, M. J., Baker, A., Blockley, E. W., Calvert, D., Coward, A., Hewitt, H. T., et al. (2019). Description of the resolution hierarchy of the global coupled HadGEM3-GC3. 1 model as used in CMIP6 HighResMIP experiments. *Geoscientific Model Development*, *12*(12), 4999–5028.

Roeckner, E., Bäuml, G., Bonaventura, L., Brokopf, R., Esch, M., Giorgetta, M., et al. (2003). The atmospheric general circulation model ECHAM5. Part I: Model description, Max Planck Institute Rep.(349).

Roeckner, E., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kornblueh, L., et al. (2006). Sensitivity of simulated climate to horizontal and vertical resolution in the ECHAM5 atmosphere model. *Journal of Climate*, *19*(16), 3771–3791.

Rong, X. Y., Li, J., Chen, H., Xin, Y., Su, J., Hua, L., et al. (2018). The CAMS climate system model and a basic evaluation of its climatology and climate variability simulation. *Journal of Meteorological Research*, *32*(6), 839–861. https://doi.org/10.1007/s13351-018-8058-x

Rotstayn, L. D., Collier, M. A., Dix, M. R., Feng, Y., Gordon, H. B., O'Farrell, S. P., et al. (2010). Improved simulation of Australian climate and ENSO-related rainfall variability in a global climate model with an interactive aerosol treatment. *International Journal of Climatology*, *30*(7), 1067–1088. https://doi.org/10.1002/joc.1952

Rugenstein, M., Bloch-Johnson, J., Gregory, J., Andrews, T., Mauritsen, T., Li, C., et al. (2020). Equilibrium Climate Sensitivity Estimated by Equilibrating Climate Models. *Geophysical Research Letters*, *47*. https://doi.org/10.1029/2019GL083898

Russell, G. L., Miller, J. R., & Rind, D. (1995). A coupled atmosphere-ocean model for transient climate change studies. *Atmosphere-Ocean*, *33*(4), 683–730. https://doi.org/10.1080/07055900.1995.9649550

Santer, B. D., Solomon, S., Bonfils, C., Zelinka, M. D., Painter, J. F., Beltran, F., et al. (2015). Observed multivariable signals of late 20th and early 21st century volcanic activity. *Geophysical Research Letters*, *42*, 500–509. https://doi.org/10.1002/2014GL062366

Schlund, M., Lauer, A., Gentine, P., Sherwood, S. C., & Eyring, V. (2020). Emergent constraints on equilibrium climate sensitivity in CMIP5: Do they hold for CMIP6? *Earth System Dynamics Discussions*, *2020*, 1–40. https://doi.org/10.5194/esd-2020-49

Schmidt, G. A., Ruedy, R., Hansen, J. E., Aleinov, I., Bell, N., Bauer, M., et al. (2006). Present-day atmospheric simulations using GISS ModelE: Comparison to in situ, satellite, and reanalysis data. *Journal of Climate*, *19*(2), 153–192. https://doi.org/10.1175/Jcli3612.1

Séférian, R., Delire, C., Decharme, B., Voldoire, A., Salas y Melia, D., Chevallier, M., et al. (2016). Development and evaluation of CNRM Earth system model—CNRM-ESM 1. *Geoscientific Model Development*, *9*(4), 1423–1453. https://doi.org/10.5194/gmd-9-1423-2016

Séférian, R., Nabat, P., Michou, M., Saint-Martin, D., Voldoire, A., Colin, J., et al. (2019). Evaluation of CNRM Earth-System model, CNRM-ESM 2–1: Role of Earth system processes in present-day and future climate. *Journal of Advances in Modeling Earth Systems*, *11*, 4182–4227. https://doi.org/10.1029/2019MS001791

Sellar, A. A., Jones, C. G., Mulcahy, J., Tang, Y., Yool, A., Wiltshire, A., et al. (2019). UKESM1: Description and evaluation of the UK Earth System Model. *Journal of Advances in Modeling Earth Systems*, *11*, 4513–4558. https://doi.org/10.1029/2019MS001739

Sherwood, S. C., Bony, S., Boucher, O., Bretherton, C., Forster, P. M., Gregory, J. M., & Stevens, B. (2015). Adjustments in the forcing-feedback framework for understanding climate change. *Bulletin of the American Meteorological Society*, *96*(2), 217–228. https://doi.org/10.1175/Bams-D-13-00167.1

Sidorenko, D., Rackow, T., Jung, T., Semmler, T., Barbi, D., Danilov, S., et al. (2015). Towards multi-resolution global climate modeling with ECHAM6–FESOM. Part I: model formulation and mean climate. *Climate Dynamics*, *44*(3–4), 757–780.

Smith, C. J., Kramer, R. J., Myhre, G., Alterskjær, K., Collins, W., Sima, A., et al. (2020). Effective radiative forcing and adjustments in CMIP6 models. *Atmospheric Chemistry and Physics Discussions*, *2020*, 1–37. https://doi.org/10.5194/acp-2019-1212

Smith, D. M., Booth, B. B., Dunstone, N. J., Eade, R., Hermanson, L., Jones, G. S., et al. (2016). Role of volcanic and anthropogenic aerosols in the recent global surface warming slowdown. *Nature Climate Change*, *6*(10), 936.

Solomon, S., Manning, M., Marquis, M., & Qin, D. (2007). *Climate change 2007—The physical science basis: Working group I contribution to the fourth assessment report of the IPCC*. Cambridge: Cambridge university press.

Song, Z., Bao, Y., & Qiao, F. J. C. C. R. (2019). Introduction of FIO-ESM v2. 0 and its participation plan in CMIP6 experiments. *Climate Change Research*, *15*, 558–565.

Stengel, M., Stapelberg, S., Sus, O., Schlundt, C., Poulsen, C., Thomas, G., et al. (2017). Cloud property datasets retrieved from AVHRR, MODIS, AATSR and MERIS in the framework of the Cloud_cci project. *Earth System Science Data*, *9*(2), 881–904.

Stevens, B., Fiedler, S., Kinne, S., Peters, K., Rast, S., Müsse, J., et al. (2017). MACv2-SP: A parameterization of anthropogenic aerosol optical properties and an associated Twomey effect for use in CMIP6. *Geoscientific Model Development*, *10*, 433–452.

Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., et al. (2013). *Climate change 2013: The physical science basis, edited*. Cambridge: Cambridge University Press.

Stouffer, R. J., Eyring, V., Meehl, G. A., Bony, S., Senior, C., Stevens, B., & Taylor, K. E. (2017). Cmip5 Scientific Gaps and Recommendations for Cmip6. *Bulletin of the American Meteorological Society*, *98*(1), 95–105. https://doi.org/10.1175/Bams-D-15-00013.1

Susskind, J., Barnet, C., Blaisdell, J., Iredell, L., Keita, F., Kouvaris, L., et al. (2006). Accuracy of geophysical parameters derived from Atmospheric Infrared Sounder/Advanced Microwave Sounding Unit as a function of fractional cloud cover. *Journal of Geophysical Research*, *111*, D09S17. https://doi.org/10.1029/2005JD006272

Swart, N. C., Cole, J. N., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., et al. (2019). The Canadian Earth System Model version 5 (CanESM5. 0.3). *Geoscientific Model Development*, *12*(11), 4823–4873.

Tan, I., Storelvmo, T., & Zelinka, M. D. (2016). Observational constraints on mixed-phase clouds imply higher climate sensitivity. *Science*, *352*(6282), 224–227. https://doi.org/10.1126/science.aad5300

Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., et al. (2019). Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geoscientific Model Development*, *12*(7), 2727–2765. https://doi.org/10.5194/gmd-12-2727-2019

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498.

Tian, B. J., Fetzer, E. J., Kahn, B. H., Teixeira, J., Manning, E., & Hearty, T. (2013). Evaluating CMIP5 models using AIRS tropospheric air temperature and specific humidity climatology. *Journal of Geophysical Research: Atmospheres*, *118*, 114–134. https://doi.org/10.1029/2012JD018607

Tokarska, K. B., Stolpe, M. B., Sippel, S., Fischer, E. M., Smith, C. J., Lehner, F., & Knutti, R. (2020). Past warming trend constrains future warming in CMIP6 models. *Science Advances*, *6*(12), eaaz9549. https://doi.org/10.1126/sciadv.aaz9549

van Marle, M. J. E., Kloster, S., Magi, B. I., Marlon, J. R., Daniau, A. L., Field, R. D., et al. (2017). Historic global biomass burning emissions for CMIP6 (BB4CMIP) based on merging satellite observations with proxies and fire models (1750–2015). *Geoscientific Model Development*, *10*(9), 3329–3357. https://doi.org/10.5194/gmd-10-3329-2017

Vannière, B., Demory, M.-E., Vidale, P. L., Schiemann, R., Roberts, M. J., Roberts, C. D., et al. (2019). Multi-model evaluation of the sensitivity of the global energy budget and hydrological cycle to resolution. *Climate Dynamics*, *52*(11), 6817–6846.

Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., et al. (2019). Evaluation of CMIP6 DECK experiments with CNRM-CM6-1. *Journal of Advances in Modeling Earth Systems*, *11*, 2177–2213. https://doi.org/10.1029/2019MS001683

Voldoire, A., Sanchez-Gomez, E., Salas y Mélia, D., Decharme, B., Cassou, C., Sénési, S., et al. (2013). The CNRM-CM5.1 global climate model: Description and basic evaluation. *Climate Dynamics*, *40*(9–10), 2091–2121. https://doi.org/10.1007/s00382-011-1259-y

Volodin, E. M., Dianskii, N. A., & Gusev, A. V. (2010). Simulating present-day climate with the INMCM4.0 coupled model of the atmospheric and oceanic general circulations. *Izvestiya Atmospheric and Oceanic Physics*, *46*(4), 414–431. https://doi.org/10.1134/S000143381004002x

Volodin, E. M., Mortikov, E. V., Kostrykin, S. V., Galin, V. Y., Lykosov, V. N., Gritsun, A. S., et al. (2017a). Simulation of modern climate with the new version of the INM RAS climate model. *Izvestiya Atmospheric and Oceanic Physics*, *53*(2), 142–155. https://doi.org/10.1134/S0001433817020128

Volodin, E. M., Mortikov, E. V., Kostrykin, S. V., Galin, V. Y., Lykossov, V. N., Gritsun, A. S., et al. (2017b). Simulation of the present-day climate with the climate model INMCM5. *Climate Dynamics*, *49*(11–12), 3715–3734. https://doi.org/10.1007/s00382-017-3539-7

Volodin, E. M., Mortikov, E. V., Kostrykin, S. V., Galin, V. Y., Lykossov, V. N., Gritsun, A. S., et al. (2018). Simulation of the modern climate using the INM-CM48 climate model. *Russian Journal of Numerical Analysis and Mathematical Modelling*, *33*(6), 367–374. https://doi.org/10.1515/rnam-2018-0032

Vose, R. S., Schmoyer, R. L., Steurer, P. M., Peterson, T. C., Heim, R., Karl, T. R., & Eischeid, J. K. (1992). The Global Historical Climatology Network: Long-term monthly temperature, precipitation, sea level pressure, and station pressure data*Rep*., Oak Ridge National Lab., TN (United States). Carbon Dioxide Information.

Waliser, D., Gleckler, P. J., Ferraro, R., Taylor, K. E., Ames, S., Biard, J., et al. (2020). Observations for Model Intercomparison Project (Obs4MIPs): Status for CMIP6. *Geoscientific Model Development*, *13*(7), 2945–2958. https://doi.org/10.5194/gmd-13-2945-2020

Washington, W. M., Weatherly, J. W., Meehl, G. A., Semtner Jr, A. J., Bettge, T. W., Craig, A. P., et al. (2000). Parallel climate model (PCM) control and transient simulations. *Climate Dynamics*, *16*(10–11), 755–774. https://doi.org/10.1007/s003820000079

Watanabe, M., Suzuki, T., O'ishi, R., Komuro, Y., Watanabe, S., Emori, S., et al. (2010). Improved climate simulation by MIROC5. Mean States, Variability, and Climate Sensitivity. *Journal of Climate*, *23*(23), 6312–6335. https://doi.org/10.1175/2010JCLI3679.1

Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., Okajima, H., et al. (2011). MIROC-ESM 2010: Model description and basic results of CMIP5-20c3m experiments. *Geoscientific Model Development*, *4*(4), 845–872. https://doi.org/10.5194/gmd-4-845-2011

Weigel, K., Bock, L., Gier, B. K., Lauer, A., Righi, M., Schlund, M., et al. (2020). 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. Discuss. https://doi.org/10.5194/gmd-2020-244

Wetzel, P., Maier-Reimer, E., Botzet, M., Jungclaus, J., Keenlyside, N., & Latif, M. (2006). Effects of ocean biology on the penetrative radiation in a coupled climate model. *Journal of Climate*, *19*(16), 3973–3987.

Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., et al. (2018). The Met Office global coupled model 3.0 and 3.1 (GC3.0 and GC3.1) Configurations. *Journal of Advances in Modeling Earth Systems*, *10*, 357–380. https://doi.org/10.1002/2017MS001115

Williams, K. D., Hewitt, A. J., & Bodas-Salcedo, A. (2020). Use of Short-Range Forecasts to Evaluate Fast Physics Processes Relevant for Climate Sensitivity. *Journal of Advances in Modeling Earth Systems*, *12*. https://doi.org/10.1029/2019ms001986

Wu, T. W., Li, W., Ji, J., Xin, X., Li, L., Wang, Z., et al. (2013). Global carbon budgets simulated by the Beijing Climate Center Climate System Model for the last century. *Journal of Geophysical Research: Atmospheres*, *118*, 4326–4347. https://doi.org/10.1002/jgrd.50320

Wu, T. W., Lu, Y., Fang, Y., Xin, X., Li, L., Li, W., et al. (2019). The Beijing Climate Center Climate System Model (BCC-CSM): The main progress from CMIP5 to CMIP6. *Geoscientific Model Development*, *12*(4), 1573–1600. https://doi.org/10.5194/gmd-12-1573-2019

Wyser, K., van Noije, T., Yang, S., von Hardenberg, J., O'Donnell, D., & Döscher, R. (2019). On the increased climate sensitivity in the EC-Earth model from CMIP5 to CMIP6, Geoscientific Model Development Discussion. https://doi.org/10.5194/gmd-2019-282

Xie, S.-P., & Kosaka, Y. (2017). What caused the global surface warming hiatus of 1998–2013? *Current Climate Change Reports*, *3*(2), 128–140. https://doi.org/10.1007/s40641-017-0063-0

Yin, J., Overpeck, J., Peyser, C., & Stouffer, R. (2018). Big jump of record warm global mean surface temperature in 2014–2016 related to unusually large oceanic heat releases. *Geophysical Research Letters*, *45*, 1069–1078. https://doi.org/10.1002/2017GL076500

Yu, Y. Q., Zhang, X. H., & Guo, Y. F. (2004). Global coupled ocean-atmosphere general circulation models in LASG/IAP. *Advances in Atmospheric Sciences*, *21*(3), 444–455.

Yukimoto, S., Adachi, Y., Hosaka, M., Sakami, T., Yoshimura, H., Hirabara, M., et al. (2012). A new global climate model of the Meteorological Research Institute: MRI-CGCM3-model description and basic performance. *Journal of the Meteorological Society of Japan*, *90a*(0), 23–64. https://doi.org/10.2151/jmsj.2012-A02

Yukimoto, S., Kawai, H., Koshiro, T., Oshima, N., Yoshida, K., Urakawa, S., et al. (2019). The Meteorological Research Institute Earth System Model version 2.0, MRI-ESM 2.0: Description and Basic Evaluation of the Physical Component. *Journal of the Meteorological Society of Japan*, *97*(5), 931–965. https://doi.org/10.2151/jmsj.2019-051

Yukimoto, S., Noda, A., Kitoh, A., Hosaka, M., Yoshimura, H., Uchiyama, T., et al. (2006). Present-day climate and climate sensitivity in the Meteorological Research Institute coupled GCM version 2.3 (MRI-CGCM2.3). *Journal of the Meteorological Society of Japan*, *84*(2), 333–363. https://doi.org/10.2151/jmsj.84.333

Zanchettin, D., Khodri, M., Timmreck, C., Toohey, M., Schmidt, A., Gerber, E. P., et al. (2016). The Model Intercomparison Project on the climatic response to Volcanic forcing (VolMIP): Experimental design and forcing input data for CMIP6. *Geoscientific Model Development*, *9*(8), 2701–2719. https://doi.org/10.5194/gmd-9-2701-2016

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., et al. (2020). Causes of Higher Climate Sensitivity in CMIP6 Models. *Geophysical Research Letters*, *47*. https://doi.org/10.1029/2019GL085782

Zelinka, M. D., Zhou, C., & Klein, S. A. (2016). Insights from a refined decomposition of cloud feedbacks. *Geophysical Research Letters*, *43*, 9259–9269. https://doi.org/10.1002/2016GL069917

Zhou, T. J., Chen, X. L., Dong, L., Wu, B., Man, W. M., Zhang, L. X., et al. (2014). Chinese contribution to CMIP5: An overview of five Chinese models' performances. *Journal of Meteorological Research*, *28*(4), 481–509.