Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO2

Thorsten Mauritsen, Jürgen Bader, Tobias Becker, Jörg Behrens, Matthias Bittner, Renate Brokopf, Victor Brovkin, Martin Claussen, Traute Crueger, Monika Esch, et al.

JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2018MS001400

Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO_2

Thorsten Mauritsen[1,2], Jürgen Bader[1], Tobias Becker[1], Jörg Behrens[3], Matthias Bittner[1], Renate Brokopf[1], Victor Brovkin[1], Martin Claussen[1,4], Traute Crueger[1], Monika Esch[1], Irina Fast[3], Stephanie Fiedler[2], Dagmar Fläschner[1], Veronika Gayler[1], Marco Giorgetta[1], Daniel S. Goll[5], Helmuth Haak[1], Stefan Hagemann[1,6], Christopher Hedemann[1], Cathy Hohenegger[1], Tatiana Ilyina[1], Thomas Jahns[3], Diego Jimenéz-de-la-Cuesta[1], Johann Jungclaus[1], Thomas Kleinen[1], Silvia Kloster[1], Daniela Kracher[1], Stefan Kinne[1], Deike Kleberg[1], Gitta Lasslop[1,7], Luis Kornblueh[1], Jochem Marotzke[1], Daniela Matei[1], Katharina Meraner[1], Uwe Mikolajewicz[1], Kameswarrao Modali[1], Benjamin Möbis[1,8,9], Wolfgang A. Müller[1], Julia E. M. S. Nabel[1], Christine C. W. Nam[1,10], Dirk Notz[1], Sarah-Sylvia Nyawira[1,11], Hanna Paulsen[1], Karsten Peters[3], Robert Pincus[1,2,12,13], Holger Pohlmann[1], Julia Pongratz[1,3,14], Max Popp[1,4,15], Thomas Jürgen Raddatz[1], Sebastian Rast[1], Rene Redler[1], Christian H. Reick[1], Tim Rohrschneider[1], Vera Schemann[1,5,16], Hauke Schmidt[1], Reiner Schnur[1], Uwe Schulzweida[1], Katharina D. Six[1], Lukas Stein[1], Irene Stemmler[1], Bjorn Stevens[1], Jin-Song von Storch[1], Fangxing Tian[1,6,17], Aiko Voigt[1,7,8,18,19], Philipp Vrese[1], Karl-Hermann Wieners[1], Stiig Wilkenskjeld[1], Alexander Winkler[1] and Erich Roeckner[1]

[1]Max Planck Institute for Meteorology, Hamburg, Germany, [2]Department of Meteorology, Stockholm University, Stockholm, Sweden, [3]Deutsche Klimarechenzentrum GmbH, Hamburg, Germany, [4]Centrum für Erdsystemforschung und Nachhaltigkeit, Hamburg, Germany, [5]LSCE CEA-CNRS-UVSQ, Saclay, Gif sur Yvette, France, [6]Institute of Coastal Research, Helmholtz-Zentrum Geesthacht, Geesthacht, Germany, [7]Senckenberg Biodiversity and Climate Research Centre, Frankfurt am Main, Germany, [8]School of Earth, Atmosphere and Environment of Monash University, Melbourne, Victoria, Australia, [9]Deceased 14 January 2018, [10]Institute for Meteorology, University of Leipzig, Leipzig, Germany, [11]International Centre for Tropical Agriculture, ICIPE Duduville Campus, Nairobi, Kenya, [12]Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO, USA, [13]Physical Sciences Division, NOAA Earth System Research Lab, Boulder, CO, USA, [14]Department of Geography, Ludwig-Maximilians-Universität München, München, Germany, [15]Laboratoire de Météorologie Dynamique/Institute Pierre-Simon Laplace, CNRS, Sorbonne Université, Paris, France, [16]Institute for Geophysics and Meteorology, University of Cologne, Cologne, Germany, [17]National Centre for Atmospheric Science, University of Reading, Reading, UK, [18]Institute of Meteorology and Climate Research-Department Troposphere Research, Karlsruhe Institute of Technology, Karlsruhe, Germany, [19]Lamont-Doherty Earth Observatory, Columbia University, New York, NY, USA

Key Points:
- An updated version of the Max Planck Institute for Meteorology Earth System Model (MPI-ESM1.2) is presented
- The model includes both code corrections and parameterization improvements
- Despite this, the model maintains an equilibrium climate sensitivity, which rises with warming

Correspondence to:
T. Mauritsen,
thorsten.mauritsen@misu.su.se

Citation:
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-ESM1.2) and its response to increasing CO_2. *Journal of Advances in Modeling Earth Systems*, *11*, 998–1038. https://doi.org/10.1029/2018MS001400

Received 8 JUN 2018
Accepted 6 JAN 2019
Accepted article online 13 JAN 2019
Published online 16 APR 2019

Abstract A new release of the Max Planck Institute for Meteorology Earth System Model version 1.2 (MPI-ESM1.2) is presented. The development focused on correcting errors in and improving the physical processes representation, as well as improving the computational performance, versatility, and overall user friendliness. In addition to new radiation and aerosol parameterizations of the atmosphere, several relatively large, but partly compensating, coding errors in the model's cloud, convection, and turbulence parameterizations were corrected. The representation of land processes was refined by introducing a multilayer soil hydrology scheme, extending the land biogeochemistry to include the nitrogen cycle, replacing the soil and litter decomposition model and improving the representation of wildfires. The ocean biogeochemistry now represents cyanobacteria prognostically in order to capture the response of nitrogen fixation to changing climate conditions and further includes improved detritus settling and numerous other refinements. As something new, in addition to limiting drift and minimizing certain biases, the instrumental record warming was explicitly taken into account during the tuning process. To this end, a very high climate sensitivity of around 7 K caused by low-level clouds in the tropics as found in an intermediate model version was addressed, as it was not deemed possible to match observed warming otherwise. As a result, the model has a climate sensitivity to a doubling of CO_2 over preindustrial conditions of 2.77 K, maintaining the previously identified highly nonlinear global mean response to increasing CO_2 forcing, which nonetheless can be represented by a simple two-layer model.

1. Introduction

The Max Planck Institute for Meteorology has a history of developing versatile state-of-the-art climate models (Roeckner et al., 1989), and in the present study, we describe the development of the latest version of the institute's Earth System Model (MPI-ESM1.2) over its predecessor (MPI-ESM; Giorgetta et al., 2013). MPI climate models see broad applications supporting research both within the institute and around the world: The model code is freely available for research purposes and participates in several collaborative model comparisons, such as the upcoming sixth phase of the Coupled Model Intercomparison Project (CMIP6; Eyring et al., 2016).

In order to be useful, a climate model should among other things yield a reasonable analogy to the Earth's climate. What is a reasonable analogy in this regard depends on the problem at hand and so must be determined on a case-by-case basis. Thus, for a general-purpose model, such as MPI-ESM1.2, we naturally seek a compromise that foremost satisfies a majority of the research needs at our institute but certainly not all. For instance, individuals or partners maintain slow or computationally expensive components such as interactive ice sheets or prognostic atmospheric aerosol and chemistry in separate versions of the model. Traditionally, model mean state biases have been in the focus of model advances, and clearly, large-scale biases in models participating in CMIP have decreased steadily over time (Reichler & Kim, 2008). For this, modelers have focused on improving the representation of subgrid scale process parameterizations, increasing the model resolution, as well as refining the model tuning.

But other aspects of a model can contribute to its usefulness, for instance, the ability to conserve energy and moisture, or something as simple as having a code that functions the way it was intended. Often this is taken for granted but is nevertheless not always the case, as the sheer complexity of models, which often comprise hundreds of thousands of lines of code inevitably, leads to programming errors. For instance, a large fraction of climate models exhibit signs of leaking energy, as they are either stationary at a nonzero radiation imbalance, or as they cool while at the same time having a positive imbalance (Mauritsen et al., 2012). Energy leakages in climate models though common are in any case undesirable but are mostly problematic if the magnitude depends on state such that an artificial feedback to climate change occurs. This was indeed the case for a series of errors in earlier versions of MPI-ESM, and it was feared that previously identified nonlinearities (Heinemann et al., 2009; Meraner et al., 2013) were merely artifacts of coding errors, which we shall investigate at the end of this study. Likewise, parameterizations are usually built upon a certain idea, or an empirical relationship, but problems in the numerical code implementation may lead to behavior not originally intended. Such coding error need not per se lead to larger mean state biases but could hinder the user of the model from understanding why the model does what it does. The here-described updates to the atmospheric component of MPI-ESM1.2 particularly address issues of this kind.

Finally, transparency of decisions made during model development is a prerequisite for most scientific use of climate models. In particular, it is important to know for which properties the model results were tuned, for example, global mean temperature, winds, or sea ice; and there is little point in evaluating a model against observations for such properties (Mauritsen et al., 2012). For MPI-ESM1.2 we let the instrumental-record warming be an explicit target of the development (to be described in a companion paper). We decided to do so, in part, because there was an agreement across the institute that the new climate model would be more useful for several purposes, including decadal prediction, if it matched observed warming, but also in part to challenge our understanding of the controls on past warming.

MPI-ESM1.2 is planned to be the last release in the series of coupled climate models based on the HOPE, later renamed to MPIOM, ocean models (Maier-Reimer et al., 1982) and the ECHAM spectral dynamical core atmosphere models (Roeckner et al., 1989); see Stevens et al. (2013) for a historical overview of ECHAM model versions. However, the physical process parameterizations of the atmosphere, the land, and ocean biogeochemistry components have been transferred and further developed (Giorgetta et al., 2018) within the new ICON model framework developed in a collaboration with the German Weather Service (DWD). The new coupled Earth system model, ICON-ESM, consists of a nonhydrostatic atmosphere dynamical core (Zängl et al., 2015) and a newly developed ocean model component (Korn, 2017) both discretized on the icosahedral grid. In this way, also the next-generation ICON-ESM model based on the ICON model framework will build on decades of experience, development, and improvements, the latest of which are described in this article.

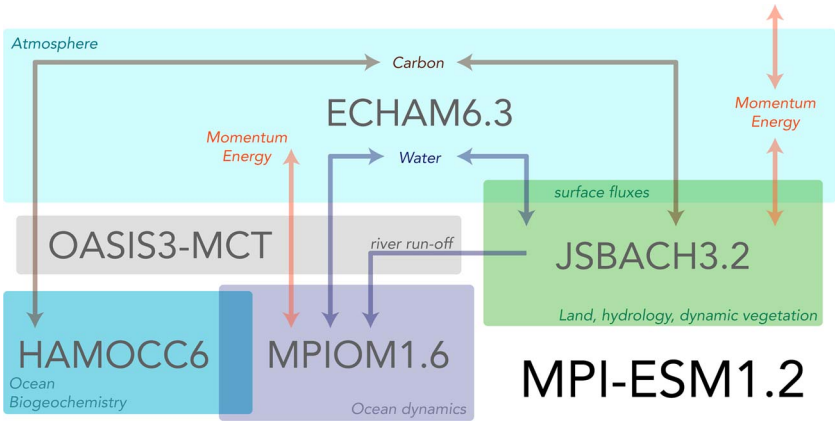

Figure 1. Schematic overview of the components of MPI-ESM1.2 and how these are coupled. The atmosphere ECHAM6.3 is directly coupled with the land surface model JSBACH3.2, whereas the ocean biogeochemistry model HAMOCC6 is directly coupled to the ocean dynamic model MPIOM1.6. These two major model component blocks are in turn coupled through the OASIS3-MCT coupler software.

In the following, we will describe the major configurations of MPI-ESM1.2 in terms of resolutions in the atmosphere and ocean (section 2). Then we describe the changes made to the atmosphere (section 3), most of which where introduced already in the intermediate MPI-ESM1.1 grand ensemble model. Changes made to the ocean component are in section 4, the ocean biogeochemistry in section 5, the land component in section 6, and technical improvements in section 7. We then inspect some properties of the coupled climate model that we found particularly interesting in section 8.

2. Model Configurations

The MPI-ESM1.2 model consists of four model components and a coupler, which are connected as it was done in the predecessor MPI-ESM (Figure 1, Giorgetta et al., 2013). The ocean dynamical model, MPIOM1.6, directly advects tracers of the ocean biogeochemistry model, HAMOCC6. The atmosphere model, ECHAM6.3, is directly coupled to the land model, JSBACH3.2, through surface exchange of mass, momentum, and heat. These two major model blocks are then coupled via the OASIS3-MCT coupler (Craig et al., 2017). The individual model components can also be operated in stand-alone modes.

The model is applied to a number of scientific and practical problems, each of which offer their own challenges in terms of representing processes or phenomena and in terms of their computational demands, which is by far mostly controlled by horizontal resolution in the atmosphere and ocean. To this end, five different coupled model configurations were created (the coarse resolution CR, low resolution LR, higher resolution HR, ocean-eddy resolving ER, and very high resolution XR; see Table 1), which span more than a factor thousand in computational cost. As such, the different model configurations have been developed with varying purposes, goals, and demands, and they have been finalized at disparate instances during the past years. Also, therefore, some updates and bug fixes are only included in the latest release of MPI-ESM1.2-LR, and in general, any comparison across the configurations should carefully consider the differences that are not limited to resolution.

For several generations of climate models developed at the Max Planck Institute for Meteorology, the workhorse atmospheric horizontal resolution has featured a spectral truncation at T63 or approximately 200-km grid spacing, corresponding to that of MPI-ESM1.2-LR (Table 1); a fact that is sometimes viewed as a lack of progress. However, with modern computers it is possible to run this configuration with 45–85 model years per physical day with fairly small computational cost (section 7), a fact that opens up new possibilities to experiment which were previously out of reach, for example, conduct large ensembles or run long simulations (section 8.3). We find that scientific users of the model experiment more freely using the MPI-ESM1.2-LR model, when not having to worry much about the computing time budgets or data storage. Further, model configurations that are well known and characterized are usually easier to learn from.

The higher-resolution MPI-ESM1.2-HR is configured with grid spacings of 40 km in the ocean and 100 km in the atmosphere with twice as many atmospheric vertical levels, which together results in it being com-

Table 1
Overview of the Named Configurations of MPI-ESM1.2, Including Grids of the Atmosphere and Oceans, Differently Included Bug Fixes, Features, and Tuning

Configuration	CR	LR	HR	ER	XR
Atmospheric triangular truncation (section 3)	T31	T63	T127	T127	T255
Atmosphere approximate grid spacing (km)	400	200	100	100	50
Atmospheric vertical levels	31	47	95	95	95
Ocean grid (section 4)	GR3.0	GR1.5	TP0.4	TP6 m	TP0.4
Ocean approximate grid spacing (km)	300	150	40	10	40
Ocean vertical levels	40	40	40	40	40
Coupling frequency (Tian et al., 2017)	Daily	Daily	Hourly	Hourly	Hourly
Maximum throughput (years/day; #cores)	345 (*264*)	85 (*960*)	22 (*2,592*)		
Stable coupled physical climate (section 8.1)	[a]Yes	Yes	[a]Yes		
Stable carbon cycle (sections 5.3, 6.7, and 8.1)		Yes			
Ocean mixing bug fix (section 4)		Yes			
Tuned historical warming		Yes	Yes		
Verified quasi-biennial oscillation (Krismer et al., 2013)			Yes		
Ocean-eddy resolving (Li & von Storch, 2013)				Yes	
CMIP6 participation planned		Yes	Yes		Yes

Note. The maximum throughput are the number of simulated years per physical day, run with monthly mean value output, at the point where adding more processors to the computation does no longer lead to appreciable increase in execution speed on the currently available DKRZ Mistral supercomputer. The processor used were 12-core Intel Xeon E5-2680 v3 Haswell architecture with a base frequency of 2.5 GHz. The ER and HR configurations have not been scaled out.

[a]The HR and CR configurations were tuned and spun up in earlier versions of the model. Users of these resolutions in later model versions with slightly different boundary conditions and the ocean mixing bug fix (section 4) must conduct a new spin-up and possibly fine tune the model to suit their needs (section 8.1).

putationally about 20 times more expensive than MPI-ESM1.2-LR. Although the improvements in terms of mean state bias reductions are relatively modest (Appendix B, Hertwig et al., 2014), the model does have advantages over MPI-ESM1.2-LR such as improved midlatitude storm track dynamics, atmospheric blocking (Müller et al., 2018), and the ability to represent the quasi-biennial oscillation (Krismer et al., 2013). The model is therefore thought to be better suited for studies involving, for example, initialized prediction, teleconnections, or midlatitude dynamics.

Concerning the land carbon and vegetation the MPI-ESM1.2-LR and MPI-ESM1.2-HR configurations differ: While for MPI-ESM1.2-HR the vegetation distribution is prescribed by a map, it is dynamically computed in MPI-ESM1.2-LR. Likewise, in historical and scenario simulations, in MPI-ESM1.2-HR land use change prescribed by a sequence of maps, while in MPI-ESM1.2-LR land use is computed from a sequence of land use transitions (section 6.3). Moreover, in contrast to MPI-ESM1.2-LR in MPI-ESM1.2-HR no attempt has been made to run the land carbon cycle to equilibrium; hence, the global carbon cycle is not considered stable (Table 1).

An even higher-resolution version, MPI-ESM1.2-XR with further enhanced atmospheric horizontal resolution has been devised for participation in HighResMIP (Haarsma et al., 2016). The MPI-ESM1.2-ER configuration is devised to serve the study of explicitly resolved ocean eddies. This model features 100-km grid spacing in the atmosphere but more importantly 10 km in the ocean allowing ocean eddies to emerge (Li & von Storch, 2013). Due to the slow and expensive integration of MPI-ESM1.2-ER, this model is not tuned and spun up to the same standard as the other configurations (section 8.1). Thus, users of the model must find ways to deal with drifts (Hasselmann et al., 1993) and other effects of the ocean being out of equilibrium. The ER and XR model configurations have among other things been applied to the study of the ocean heat uptake (Rimac et al., 2015; von Storch et al., 2016), Lorentz energy cycle (von Storch et al., 2012) and internal tides (Li et al., 2015, 2017) and to investigate the influence of resolution on biases in precipitation (Siongco et al., 2014, 2017) and sea surface temperature (SST) biases (Milinski et al., 2016) in the tropical Atlantic.

A particularity of the HR, ER, and XR model configurations is that instead of daily exchange between the atmosphere and ocean blocks, hourly coupling is applied. Beginning to resolve the diurnal cycle of upper ocean temperatures has first-order impacts on fluxes in some regions (Tian et al., 2017), but this also has impacts on the longer timescales, for instance, it enhances the asymmetry between El Niño and La Niña (Tian et al., 2018). The computational overhead to hourly coupling, however, was deemed too high for implementation in the CR and LR configurations.

For cases where a high throughput is essential, for example, for conducting multimillenial simulations (Li et al., 2013; Mikolajewicz et al., 2018; Ziemen et al., 2014), teaching, or development and testing purposes, we have created the MPI-ESM1.2-CR configuration. In addition to using only half the horizontal resolution in both the ocean and the atmosphere, the CR version also has a lowered model top from 0.01 to 10 hPa and correspondingly fewer vertical levels when compared to MPI-ESM1.2-LR. Having fewer vertical levels allows making slightly longer time steps and yields a higher compute parallelization, together resulting in throughput in excess of 300 simulated years per day. Tuning the model at this coarse resolution was, however, not easy. Particular challenges associated with coarse atmospheric resolution were to obtain sufficient precipitation on tropical land, which is essential to be able to represent, for example, the Amazon forests with dynamic vegetation, and also the atmospheric circulation around the Greenland ice sheet was difficult to represent. As a consequence of this retuning, relative to MPI-ESM1.2-LR, the equilibrium climate sensitivity rose to around 4 K (Mikolajewicz et al., 2018), and thus, without applying compensating anthropogenic aerosol cooling, the model is unlikely to be consistent with the instrumental-record warming, even if this has not been investigated.

In addition to the coupled model configurations, the components of MPI-ESM1.2 can be run in stand-alone modes and have been applied in both Earth-like and idealized settings. Examples of Earth-like settings are the Atmospheric Model Intercomparison Project configuration (Gates et al., 1998), in which observed SSTs are prescribed (Appendix B), and the Ocean Model Intercomparison Project (OMIP) configuration (Griffies et al., 2016), in which fluxes of heat, momentum, and fresh water at the air-sea boundary are derived from a prescribed atmospheric state. Likewise, the land component JSBACH3.2 can also be forced with prescribed meteorological fields allowing direct comparison with other land-component models (Lawrence et al., 2016; van den Hurk et al., 2016). The atmospheric model can further be run in the idealized aquaplanet configuration (Neale & Hoskins, 2001; Medeiros et al., 2008), which is regularly used to better understand fundamental differences between different models and to investigate how the model output depends on different parameterizations (e.g., Möbis & Stevens, 2012, Voigt et al., 2014). ECHAM6.3 can also be used in an even more idealized configuration, the global radiative-convective equilibrium, which is run with spatially homogeneous insolation, zero ocean energy transport, and an inertial nonrotating frame of reference. The radiative-convective equilibrium configuration has been used to improve our understanding of tropical convection (Becker et al., 2017; Popke et al., 2013).

3. Revisions of the Atmospheric Model Component (ECHAM6.3)

The modifications discussed in this section are confined to the physical parameterizations of ECHAM6.3, which apply to all configurations. The main goal was to remove a number of bugs and deficiencies apparent in predecessor versions such as the lack of energy conservation, the poor representation of boundary layer clouds, and inconsistent treatments of fractional cloudiness, condensational growth, and cloud-radiation interactions. Ultimately, all physical parameterizations are affected, that is, radiative transfer, cumulus convection, stratiform clouds, vertical diffusion, land surface processes, and gravity wave drag. On the other hand, the adiabatic core, the horizontal diffusion, the transport of atmospheric constituents, and the model configurations remained unchanged. For a detailed description of ECHAM6, see Stevens et al. (2013). This model description comprises both versions, ECHAM6.0 and ECHAM6.1, being identical except for technical changes like code optimization and some fine tuning. An intermediate version, ECHAM6.2, containing already most of the changes done in ECHAM6.3, was never released as it revealed an extremely high climate sensitivity of about 7 K with respect to a doubling of atmospheric CO_2. At the time we deemed it unlikely that we would be capable of reproducing the instrumental-record warming and so undertook a retuning of ECHAM6.2 resulting in a climate sensitivity of about 3 K in ECHAM6.3. The effort is described in detail in a companion paper. The description below is of changes in ECHAM version 6.3.05, though some of the experiments displayed were conducted using earlier versions of the model.

3.1. Fractional Cloud Cover

In ECHAM6.1, subgrid scale cloudiness is represented using the scheme of Sundqvist et al. (1989), such that the cloud fraction, f_{cld}, is calculated diagnostically as a function of relative humidity, η, once a threshold value, η_{crit}, is exceeded:

$$f_{\mathrm{cld}}(\eta) = \max\left[0, 1 - \sqrt{1 - \min\left(1, \frac{\eta - \eta_{\mathrm{crit}}}{1 - \eta_{\mathrm{crit}}}\right)}\right]. \tag{1}$$

The threshold relative humidity is a function of pressure (or height) introduced in ECHAM4 (Roeckner et al., 1996):

$$\eta_{\mathrm{crit}} = a_1 + (a_2 - a_1)\exp\left[1 - \left(\frac{p_{\mathrm{s}}}{p}\right)^{a_3}\right]. \tag{2}$$

The parameter set, originally $\{a_1, a_2, a_3\} = \{0.7, 0.9, 4.0\}$, determines the vertical profile of η_{crit}, such that $\eta_{\mathrm{crit}} = a_2$ at the surface and $\eta_{\mathrm{crit}} = a_1$ aloft. The shape of the profile is determined by the parameter a_3. With the choice of $a_3 = 4$ the decrease of η_{crit} with height is faster than in earlier versions of ECHAM, which applied $a_3 = 2$, so that now $\eta_{\mathrm{crit}} \approx a_1$ is reached already at a height of about 3 km above the surface.

In addition to the criterion $\eta > \eta_{\mathrm{crit}}$ for cloud generation, the sum of cloud liquid water and cloud ice, $r_{\mathrm{liq}} + r_{\mathrm{ice}}$, as produced by the condensation scheme and detrained from the moist convection scheme has to be positive in the respective grid box. Otherwise, $f_{\mathrm{cld}} = 0$, even for $\eta > \eta_{\mathrm{crit}}$. This assumption is problematic because f_{cld} is a weighting factor in the condensation scheme (Roeckner et al., 1996) so that the formation of a new cloud is inhibited unless supersaturation ($q > q_{\mathrm{sat}}$) occurs in the grid box. Then, $f_{\mathrm{cld}} = 1$, and the excess humidity ($q - q_{\mathrm{sat}}$) is used for generating cloud liquid water or cloud ice. As a result, the model tends to produce a binary distribution of cloud fraction with preference of $f_{\mathrm{cld}} = 0/1$. In ECHAM6.3, the criterion $r_{\mathrm{liq}} + r_{\mathrm{ice}} > 0$ is only applied in the radiative transfer scheme to avoid clouds without condensate, whereas $\eta > \eta_{\mathrm{crit}}$ allows condensational growth as soon as the condensate rate is positive. This new approach leads to a substantial increase in fractional cloud cover of the order of 10%, resulting in a decrease of the top-of-atmosphere net radiation by almost 15 W/m^2. Although the model changes discussed in section 3.4 partially compensate for this deficit, several tuning parameters had to be adjusted in order to restore the top-of-atmosphere radiation balance in the new model (see section 3.6).

A persistent problem in atmospheric general circulation models is the underrepresentation of marine stratocumulus clouds, abundant below subsidence inversions in the cold upwelling regions of the tropical and subtropical oceans. This deficiency is often caused by inadequate physics representations combined with too coarse vertical resolution to capture the relatively thin stratocumulus decks and the cloud top mixing processes, which crucially influence them (Mellado, 2017). In ECHAM6.3, an attempt is made to improve the low-bias simulation of marine stratocumulus (Stevens et al., 2013, Figure 5) by introducing a parameterization of vertical subgrid scale cloudiness such that $f_{\mathrm{cld}} = 1$ is achieved already at a relative humidity $\eta < 1$. For this purpose, a reference humidity, $\eta_{\mathrm{ref}} \leq 1$, is introduced to rescale the relative humidity when calculating the cloud fraction as in equation (1), namely, by replacing η by η/η_{ref}. This formulation attempts to correct for the fact that the gridbox mean relative humidity in vicinity of inversions is not a good predictor of the cloud fraction in the vertical.

In general $\eta_{\mathrm{ref}} = 1$ except when, in the height range [300 m; 2 km] over the ocean, the ratio of the temperature lapse rate, $\Gamma = \partial_z T$, and the dry adiabatic one, $\Gamma_{\mathrm{d}} = -g \cdot c_{pd}$, falls below a given threshold, that is, $\Gamma/\Gamma_{\mathrm{d}} \leq 0.25$. In this case, the reference humidity is gradually reduced with decreasing lapse rate according to

$$\eta_{\mathrm{ref}} = \min\left[1, \eta_{\mathrm{sc}} + \max\left(0, \Gamma/\Gamma_{\mathrm{d}}\right)\right] \tag{3}$$

with $\eta_{\mathrm{sc}} = 0.7$, so that $0.7 \leq \eta_{\mathrm{ref}} \leq 0.95$ denotes the range of reference values at which $f_{\mathrm{cld}} = 1$ is achieved already, depending on the ratio $\Gamma/\Gamma_{\mathrm{d}}$. In the extreme case of a real temperature inversion ($\Gamma \geq 0$), cloud fraction is generated starting already at a relative humidity of 70%. However, equation (3) is applied only in the model layer beneath the inversion. In all other model layers, $\eta_{\mathrm{ref}} = 1$. A similar scheme was developed already for ECHAM6.1 but never applied because it was inefficient due to the inadequate use of f_{cld} in the condensation scheme discussed above.

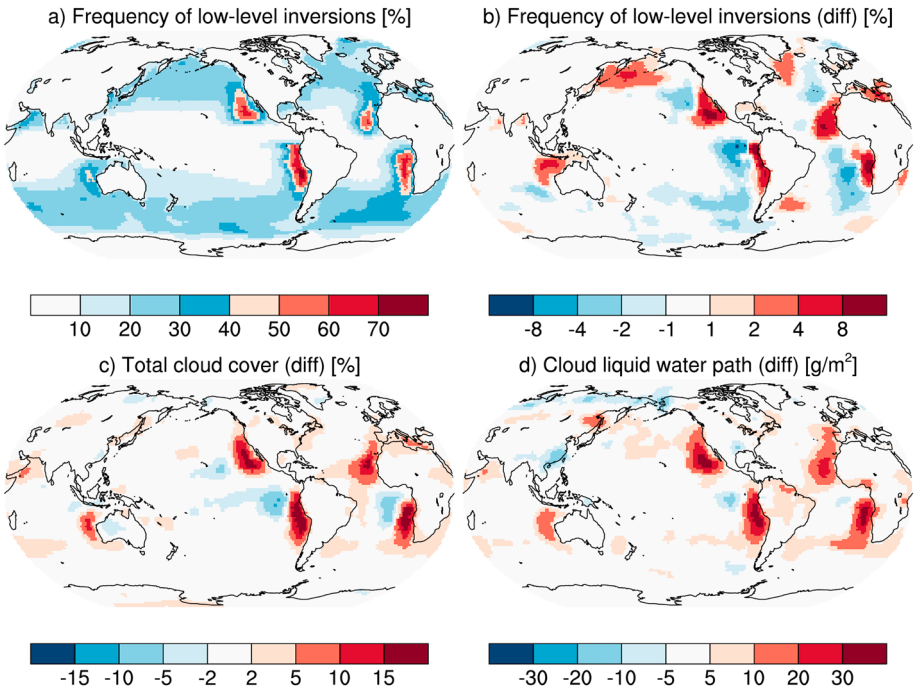

Figure 2. Representing stratocumulus clouds with equation 3. Shown are results from two simulations: one with η_{sc} equal 0.7 minus one with 1.0. The runs use observed sea surface temperatures and sea ice concentrations for the period 1979–1988. (a) The frequency of low-level inversions counted when the temperature lapse rate in the boundary layer falls below a specified fraction (0.25) of the dry adiabatic one. (b) Difference in the frequency of low-level inversions between the two parameter settings. (c) The change in the total cloud cover. (d) The change in cloud liquid water path.

As expected, low-level inversions are most frequently simulated over cold ocean currents to the west of the continents in the subtropics, Figure 2a. The responses of total cloud cover, Figure 2c, and cloud liquid water path, Figure 2d, to decreasing η_{sc} from 1.0 to 0.7 in the model are characterized by marked enhancements along the coasts and somewhat weaker decreases to the west of the respective maxima. These dipole patterns are also reflected in the changing frequency of inversions, Figure 2b, and are presumably driven by dynamical processes such as changes in the low-level moisture convergence. On the global scale, the water cycle is enhanced over both ocean and land, total cloud cover, and cloud liquid water path are higher by 1% and 4%, respectively, the top-of-atmosphere radiation budget is decreased by 1.3 W/m^2 and the net surface radiation by 0.9 W/m^2. The net atmospheric cooling of 0.4 W/m^2 is caused largely by the longwave cloud radiative cooling of 0.5 W/m^2 due to the increase in marine stratocumulus. This additional atmospheric cooling contributes to the increase in the globally averaged precipitation. A further mechanism contributing to the enhanced hydrological cycle appears to be the enhanced ventilation of the boundary layer resulting in an increase in the latent heat flux at the surface by 0.8 W/m^2, despite the radiative cooling of the surface. Overall, including the representation of stratocumulus reduces biases in reflected shortwave radiation to space in the respective regions (see Appendix B).

3.2. Radiative Transfer Scheme (PSrad)

In ECHAM6.3 a new representation of radiative transfer is introduced, following the PSrad implementation of the RRTMG (Pincus & Stevens, 2013). PSrad uses the gas optics from the Rapid Radiative Transfer Model for General circulation models (RRTMG) version 4.84 for 140 quasi-monochromatic spectral integration points (pseudo-wave numbers) in 16 bands in the long wave, and RRTMG version 3.8 for 112 spectral integration points in 14 shortwave bands. Optical properties for each pseudo-wave number are constructed using the correlated-*k* approach (Mlawer et al., 1997). The computation of cloud optical properties for a given cloud is unchanged from ECHAM6.1. Although written mostly from scratch PSrad closely follows RRTMG's algorithms (Iacono et al., 2008; Mlawer et al., 1997): Longwave calculations consider only emission and absorption, while shortwave calculations use two-stream calculations to compute layer transmittance and reflectance, and adding in order to compute the transport of radiation between layers.

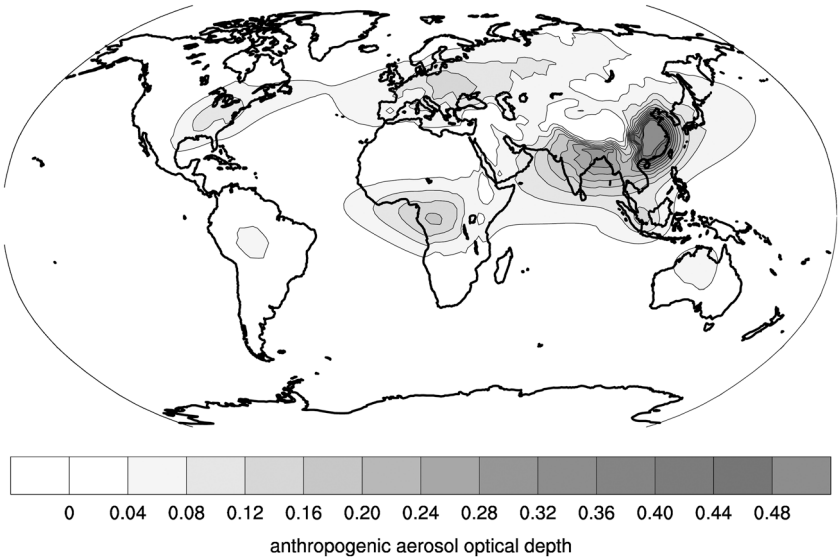

Figure 3. Anthropogenic aerosol in ECHAM6.3. Shown is the annual mean of anthropogenic aerosol optical depth at 550 nm for present-day conditions from MACv2-SP.

PSrad treats fractional cloudiness using the Monte Carlo Independent Column Approximation (see Pincus et al., 2003). Subcolumns are created by randomly sampling the vertical distribution of clouds within a grid column consistent with the cloud fraction at each level and the maximum-random assumption for the vertical overlap (e.g., Räisänen et al., 2004). Clouds are assumed to be horizontally homogeneous within each subcolumn and also within each column, that is, the mean in-cloud condensate amount and particle size are used to compute optical properties for all cloudy subcolumns. In ECHAM6.3 gases (including water vapor), aerosols, and the surface are also assumed uniform within each column. The implementation of Monte Carlo Independent Column Approximation in ECHAM6.3 resolves an earlier problem with the implementation of fractional cloudiness impact on radiation in ECHAM6.1 (Stevens et al., 2013).

PSrad also implements Monte Carlo Spectral Integration (MCSI; Pincus & Stevens, 2009, 2013), which allows the broadband radiation calculation to be approximated with a randomly chosen subset of spectral integration points, reducing the computational cost, but introducing unbiased random noise. This opens the possibility of adjusting the spectral and temporal sampling independently. "Teams" of spectral points (Pincus & Stevens, 2013) keep the maximum error in surface fluxes to levels that do not dramatically degrade the simulation.

PSrad was built to explore the possibilities of MCSI. The awkwardness of retrofitting the k distribution from RRTMG to use MCSI makes the implementation of PSrad more computationally expensive than a scheme in which all spectral points are always used in order. As it turns out, experience in both free-running climate simulations (Pincus & Stevens, 2013) and short-term weather forecasts (Bozzo et al., 2014) suggests that the total error in sparse radiation calculations, relative to computations at every grid point and time step, is determined by the degree of sparsity regardless whether the radiation are sparse in time (infrequent) or in spectral space. For this reason MCSI is not enabled in the base simulations described here, and a later developed radiative transfer scheme applied in ICON-ESM does not use MCSI.

3.3. Tropospheric Aerosol

A major change in ECHAM6.3, as of version 6.3.03, is the treatment of radiative effects of anthropogenic aerosol in the radiation calculation. The old climatology of anthropogenic aerosol has been replaced with the newly developed MACv2-SP parameterization (Fiedler et al., 2017; Stevens et al., 2017), which has been designed for the usage in the framework of CMIP6, CFMIP, and other model intercomparisons (Eyring et al., 2016; Pincus et al., 2016). MACv2-SP prescribes the anthropogenic aerosol optical properties, namely, the aerosol optical depth, asymmetry and single scattering albedo, for inducing aerosol-radiation interactions as a function of geographical position, height above ground level, time, and wavelength. To this end, MACv2-SP approximates the observationally constrained present-day spatial distributions of the monthly mean anthropogenic aerosol optical depth, τ_a, from the climatology of MPI-M (MACv2, updated over Kinne et al., 2013)

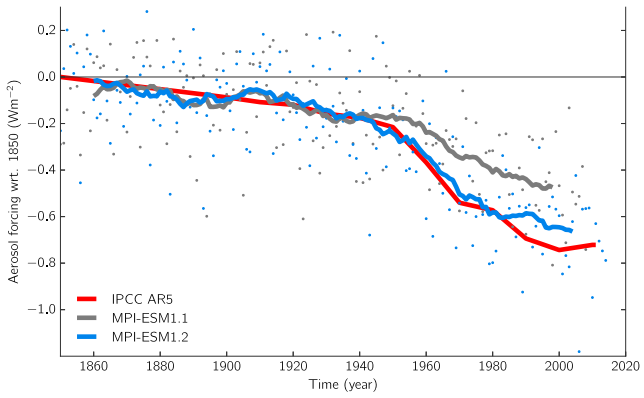

Figure 4. Historical aerosol forcing for MPI-ESM1.1 and MPI-ESM1.2. The aerosol effective radiative forcing is estimated as the anomalies of the radiation imbalance in sstClim experiments, following Hansen et al. (2005) and Pincus et al. (2016), wherein the historical evolution of aerosols are prescribed. In calculating the effective radiative forcing, we compensate for land cooling. A 21-year running mean is applied to smooth out the apparent year-to-year fluctuations in forcing (Fiedler et al., 2017). For comparison, the red line shows the best estimate from the IPCC AR5 report (Myhre et al., 2013), wherein for comparability we have subtracted the value in 1850. IPCC = Intergovernmental Panel on Climate Change.

with mathematical functions. The spatial distribution of τ_a is constructed by a superposition of nine pairs of rotated Gaussian plumes in the horizontal dimension and beta functions in the vertical direction. Figure 3 illustrates the annual mean distribution of τ_a in the midvisible from MACv2-SP. The spectral dependency of the optical properties is represented by using the Ångström exponent for adjusting the midvisible values for other wavelengths in the shortwave radiation parameterization (section 3.2). Here the anthropogenic aerosol optical properties are mixed with those of the natural aerosol. The latter is prescribed as a monthly mean climatology of natural aerosol optical properties (Kinne et al., 2013), representing the bulk optical properties of naturally occurring aerosol species, such as desert dust and sea spray, and is identical to that used in the predecessors MPI-ESM and MPI-ESM1.1 models.

The temporal scaling in MACv2-SP represents both an annual cycle and year-to-year changes in τ_a. Stepwise linear functions approximate the monthly mean annual cycle of MACv2 that is particularly pronounced in the regions with seasonally active biomass burning. The CMIP6 emission inventory per country for the industrial era is used to scale τ_a for the years 1850 to 2014 (Stevens et al., 2017). It is herein assumed that changes in anthropogenic aerosol scale with the temporal development of SO_2 and NH_3 emissions and use their relative contribution to the radiative forcing at present day for weighting the impact of these emissions on τ_a. MACv2-SP further interprets the gridded CMIP6 scenarios for ScenarioMIP (O'Neill et al., 2016) for capturing τ_a from 2015 to 2100 (manuscript in preparation), using the same assumptions, but emissions are averaged around the center of each plume instead of integrating emissions over sets of countries.

MACv2-SP also prescribes aerosol-cloud interactions in the form of a Twomey effect (Twomey, 1977). These are induced by increasing the cloud-droplet number concentration in the shortwave radiation calculation using the factor, η_N:

$$\eta_N = 1 + \frac{\mathrm{d}N}{N} = \frac{ln\left[1,000\left(\tau_a + \tau_{bg}\right) + 1\right]}{ln\left[1,000\tau_{bg} + 1\right]}. \tag{4}$$

The mathematical formulation is such that the Twomey effect is strongest for small background aerosol burden and large anthropogenic pollution as suggested by satellite data (Stevens et al., 2017). Given the spatially different τ_a, also, η_N varies with space and time. It is herein noteworthy that the background aerosol, τ_{bg}, for parameterizing the Twomey effect is simplified and independent from the natural aerosol climatology used for aerosol-radiation interaction:

$$\tau_{bg}(\varphi, \lambda, t) = \tau_{pl}(\varphi, \lambda, t) + \tau_{gl}. \tag{5}$$

The plume background, τ_{pl}, is constructed at the positions of the plume centers of the anthropogenic aerosol. Additionally, a global background, $\tau_{gl} = 0.02$, is prescribed that can be altered for inducing different strengths of Twomey effects (Fiedler et al., 2017). We implement the Twomey effect by multiplying the factor η_N by the background cloud-droplet number concentration in the radiation transfer calculation but do not perturb that used in the cloud microphysics.

Figure 4 shows the aerosol effective radiative forcing from atmosphere-only simulations for the historical period in MPI-ESM1.1 and MPI-ESM1.2. Here the SST is prescribed to preindustrial conditions (sstClim) and the evolution of anthropogenic aerosols alone exerts a negative impact on the top-of-atmosphere radiation balance, which can be interpreted as an effective radiative forcing (Hansen et al., 2005; Pincus et al., 2016). The here-estimated near-present-day anthropogenic aerosol cooling of little more than -0.6 W/m^2 is slightly stronger than the -0.50 W/m^2 found by Fiedler et al. (2017) for the year 2005, which may for instance be due to the different experiment protocols, analysis methods, and/or the different underlying SSTs. The new simple plume climatology leads to slightly enhanced cooling of around -0.2 W/m^2 in the latter parts of the

Table 2
Global Annual Mean Atmospheric Energy Budget Errors Calculated as the Differences Between Vertically Integrated Heating Rates and Surface Fluxes in ECHAM, With and Without the Advection of Cloud Condensate, Respectively

Dynamical core	With advection	Without advection
Error 1	−1.6	−1.6
Error 2	1.1	1.3
Error 3	0.3	0.3
Total ECHAM6.1	−0.03	0.15
Total ECHAM6.3	−0.03	0.00

Note. Note that errors 1, 2, and 3 were isolated in single simulations by including them step by step in the revised model. Total error represents the budget errors with all three errors included in one model simulation. Units are in watts per square meter.

twentieth century over the climatology used previously (Kinne et al., 2013), which did not include a Twomey effect, and the evolution is in line with the best estimate from the IPCC AR5 report (Myhre et al., 2013).

3.4. Convective Mass Flux, Convective Detrainment, and Turbulent Transfer

Changes in various parts of ECHAM6.1 were required because of inconsistent formulations in the convection scheme, in the grid-scale stratiform cloud condensation scheme, and in the turbulence transfer scheme, causing violations of the atmospheric energy budget (Stevens et al., 2013). For a detailed discussion of the budget errors and their eliminations, see Appendix A.

Error 1 of about −1.6 W/m^2 globally is caused by an inappropriate discretization of the condensate fluxes in the convection scheme (Nordeng, 1994; Tiedtke, 1989). The problem arises at the melting level of the updraft where the preexisting liquid gets frozen without releasing the associated latent energy, thereby generating a spurious cooling in the column. This error is eliminated in ECHAM6.3 by using an appropriate discretization of the respective convective heat flux.

Error 2 of about +1.1 W/m^2 globally is caused by an inconsistent treatment of the convective detrainment. In the convection scheme, the phase of the cloud condensate is a function of temperature, with the ice/liquid phase generated simply at temperatures below/above the melting point. The condensate detrained from the updraft is used as a source term in the stratiform cloud condensation scheme. By inconsistently passing the phase of the detrained condensate a spurious energy source is generated. In ECHAM6.3 the fraction of liquid and ice is explicitly passed to the stratiform cloud condensation scheme, thus avoiding a spurious redefinition of the condensates phase.

Error 3 of about +0.3 W/m^2 globally is caused by an inconsistent use of the specific heat capacity of water vapor within the turbulence transfer scheme. In solving the vertical diffusion equation for the dry static energy, the specific heat capacity $c_p(q_s)$ is applied as lower boundary condition, where q_s is the specific humidity at the surface. Over water and ice, $q_s = q_{sat}$ is assumed whereas over land q_s is a function of soil wetness, vegetation index, and specific humidity at the lowest model level. The sensible heat flux at the surface, however, is inconsistent with these boundary conditions, as it is formulated in terms of $c_p(q_{lml})$, where q_{lml} ($< q_{sat}$ in general) is the specific humidity at the lowest model level. The mismatch of globally about 8 W/m^2 is almost compensated by an inconsistent derivation of the temperature tendency from the tendency of the dry static energy used in the turbulence transfer scheme. Furthermore, temperature changes due to melting of snow are not considered in the calculation of the sensible heat flux. In ECHAM6.3, all of these errors are eliminated so that identity between atmospheric heating by turbulence transport and the surface heat flux according to equation (A8) is achieved in every column and at every time step.

Errors 1 and 2 can also be analysed columnwise in model simulations neglecting the advection of cloud condensate in the stratiform cloud condensation scheme. Then, the storage terms $\partial r_{ice}/\partial t$ and $\partial r_{liq}/\partial t$ are determined solely by local sources and sinks due to phase changes, convective detrainment, and precipitation formation. Interestingly, the global water fluxes and also the regional distributions of cloud liquid water and cloud ice are hardly affected by this simplification (not shown). Table 2 shows the global annual mean atmospheric energy budget errors in model simulations with and without the advection of cloud con-

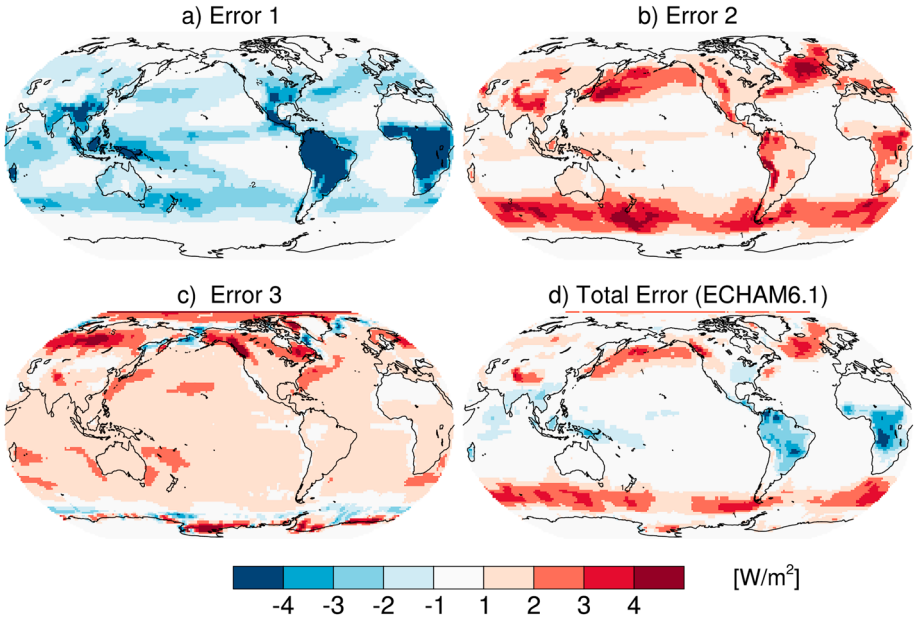

Figure 5. Maps of the impact of the three coding errors described in section 3.4 on the atmospheric energy budget, as well as the total error in the predecessor ECHAM6.1 model. The estimates shown here ignore the advection of cloud condensate.

densate, respectively. The total error in ECHAM6.1 of −0.03 W/m^2 is relatively small, which could explain why it remained undetected for many years. The problem, however, became apparent when the model was run into states far away from present-day conditions (Meraner et al., 2013; Popp et al., 2015, 2016; Voigt et al., 2011) resulted errors of several watts per square meter. Incidentally, the total error is almost identical to that of the revised model, ECHAM6.3, because errors of positive and negative signs almost cancel each other. Given the rather short integration time of only 1 year, the total error found in ECHAM6.3, with the advection of cloud condensate included, is due to incomplete cancellation on the regional scale, between the vertically integrated heating associated with phase changes of the water components on the one hand and the latent heat release by precipitation on the other hand (see equation (A10)). In the simulation without advection, the total error in ECHAM6.3 is virtually 0. The conservation properties were confirmed to hold also in warmer and colder states.

The regional error distributions, based on the simulations without advection of cloud condensate, are shown in Figure 5. The errors due to the inadequate discretization of the convective mass flux (Figure 5a) stand out particularly over the tropical continents. Smaller errors are found in the convective regimes over the low-latitude oceans and also in the midlatitude storm tracks. The negative sign in these regions implies that the vertically integrated heating due to phase changes of the water components (vapor, liquid, ice, rain, and snow) is smaller than the latent energy realized by rain and snow reaching the surface (see equation (A10)). On the other hand, the error caused by the inadequate treatment of the convective detrainment (Figure 5b) has a positive sign because the redefinition of the phase of the convective detrainment in the stratiform cloud condensation scheme ignores the energy required for the spurious melting of cloud ice generated in the convection scheme. This error is prominent particularly in the midlatitude storm tracks where it actually overcompensates the errors evident in Figure 5a. The positive sign of the error 3 implies that the surface heat flux is smaller than the diagnosed heating in the vertical column (Figure 5c). Due to error compensation (see the discussion of error 3 in Appendix A), the resulting error is small, except over snow-covered land where errors of up to 3 W/m^2 are found. The total error pattern in ECHAM6.1 (Figure 5d) is characterized by spurious atmospheric cooling at low latitudes, governed by error 1, and spurious atmospheric heating in the storm track regions governed by error 2. In ECHAM6.3 the individual errors as well as the total are either exactly 0 (error 3) or close to 0 (errors 1 and 2) at the grid point scale (not shown). Small residual errors can be explained by a nonzero storage of the cloud condensate, given the short integration time of 1 year in the here-presented experiments.

The fact that the atmosphere physics parameterizations now conserve energy does by no means guarantee conservation properties of the overall model system. Indeed, there are additional energy leakages in the atmosphere dynamical core, interpolation errors in the coupler near coastlines, neglect of the temperature of precipitation falling from the lowest model level into the ocean, and other minor errors. The overall leakage of MPI-ESM1.2-LR is largely unchanged at around 0.45 W/m^2 as inferred from the top-of-atmosphere radiation balance in long control simulations. Some of these errors are reduced with higher resolution, but none of them vary appreciably with warming and so do not induce artificial feedbacks, which was a particular concern in relation to past studies finding strongly increasing climate sensitivity in warmer climates (Heinemann et al., 2009; Meraner et al., 2013). Therefore, in section 8.3 we verify the nonlinearity properties of the model are retained after removing the coding errors.

3.5. Additional Modifications

In addition to the model revisions presented in sections 3.1 through 3.4, a few more features are changed: To achieve consistency with the rest of the model, a minor change was done in the convection scheme by replacing the specific heat of dry air by the specific heat of moist air.

The sea ice surface albedo and thermodynamics are calculated as part of the atmosphere whereas ice growth, melt, and advection is handled by the ocean in MPI-ESM1.2 (for details on this, see section 4). A bug in the sea ice melt pond scheme, relevant only in coupled model simulations, was identified and corrected (Roeckner et al., 2012): In the MPI-ESM (Giorgetta et al., 2013), melt ponds were accidentally ignored in the calculation of sea ice melt. Thus, the interaction between pond evolution, ice albedo, and ice melt was missing in the predecessor model. In MPI-ESM1.2 the melting of sea ice depends on the respective albedos of snow covered ice, bare ice, and melt ponds and on the respective fractional areas as well.

The subgrid scale mountain drag parameterization in ECHAM6.1 comprises the momentum transport from Earth's surface to the atmosphere accomplished by orographic gravity waves and, second, by the drag exerted by the subgrid scale mountains when the flow is blocked at low levels (Lott, 1999; Miller et al., 1989; Palmer et al., 1986). The strength of the gravity wave drag from unresolved orographic sources can be determined by parameters tuning the blocked flow drag amplitude (C_d) and the gravity wave stress amplitude (G), using the terminology of Lott (1999). In ECHAM6.1, the scheme was not activated at every grid point but only when two threshold values were exceeded characterizing rough terrain, that is, the standard deviation derived from the subgrid scale orography and the difference between the peak elevation and the mean elevation in the respective grid area. Thus, in addition to C_d and G, two more resolution-dependent tuning parameters were introduced in ECHAM6.1. This is avoided in ECHAM6.3 where the routine is activated at every grid point so that the results are now independent of the choice of the thresholds.

A related modification was made to the roughness length for momentum. In ECHAM6.1 this is a function of the subgrid scale orography and can be as large as 20 m. In ECHAM6.3 the roughness length for heat and momentum depends only on characteristics of the vegetation. Consequently, the momentum transport through vertical diffusion is systematically reduced over rough terrain, though compensated to some extent by an enhanced low-level flow blocking accomplished by a larger C_d discussed above.

Furthermore, a bug in the definition of the land surface humidity was eliminated, the land surface albedo was slightly modified using MODIS data, and the land ice albedo (Greenland, Antarctica) has been decreased by about 5%, in better accordance with satellite data.

3.6. Atmospheric Model Tuning

A major retuning of the model was required because all modifications taken together caused a decrease of the global top-of-atmosphere radiation budget by about 10 W/m^2 due mainly to the corrected cloud fraction scheme (section 3.1) and also because the model climate sensitivity had roughly doubled to around 7 K, which would have prevented a reasonable match to the instrumental-record warming. If the latter had not been addressed, the model's historical warming would have roughly exceeded that observed by a factor of 2. When reducing the historical warming in a model there are essentially three options: reduce forcing, increase deep ocean heat uptake efficiency, or reduce the climate sensitivity. The forcing can be reduced by increasing aerosol cooling by enhancing the indirect effect, but at the time (2014–2015) we did not have such a parameterization in the model, which was developed after that (section 3.3). Further, ocean heat uptake already exceed that observed (Giorgetta et al., 2013), and so we were left with reducing the climate sensitivity. Since the predecessor MPI-ESM model warmed slightly more than observed, and it had a sensitivity of 3.5

K, we decided to aim at an equilibrium climate sensitivity of around 3 K. The reduction of the model's sensitivity was primarily achieved by increasing the entrainment rate for shallow convection by a factor of 10, from 3×10^{-4} m^{-1} in ECHAM6.1 to 3×10^{-3} m^{-1} in ECHAM6.3, with the purpose to reduce tropical low-level cloud feedback. But also other convective cloud parameters, mixed-phase cloud processes, and the representation of stratocumulus were found to be important.

After the corrections to the cloud fraction scheme (section 3.1), it turned out that the most efficient way to compensate for this radiation deficit is to reduce cloudiness by modifying the free parameters in the cloud fraction scheme. This was done by changing the profile parameters of the critical relative humidity (cf. equation (2)) from

$$\{a_1, a_2, a_3\} = \{0.7, 0.9, 4.0\} \quad \text{to} \quad \{0.75, 0.968, 1.0\}, \tag{6}$$

so that cloud formation is systematically shifted to higher relative humidities. In addition, the optical thickness of low-level clouds is now parameterized as a function of cloud type. In the radiative transfer scheme, clouds are treated as plane parallel homogeneous layers defined by the cloud water path, CWP (liquid and/or ice). However, since real clouds are never homogeneous, it can be easily shown that the calculated reflectivity, for a given CWP, is always too high. This well-known bias can be attenuated to some extent by introducing a reduction factor (Cahalan et al., 1994), such that the CWP used in the radiation scheme is a fraction, $f_{\text{hom}} < 1$, of that predicted by the cloud scheme, $\text{CWP} = f_{\text{hom}}\text{CWP}_{\text{cld}}$. In ECHAM6.1 f_{hom} was set to 0.77 for liquid water clouds and 0.8 for ice clouds. In ECHAM6.3 $f_{\text{hom}} = 0.8$ for all cloud types, except for isolated shallow cumulus clouds, where $f_{\text{hom}} = 0.4$ is applied as long as the CWP above cloud top is smaller than 20% of the total CWP. Such a distinction was suggested by Bäuml et al. (2004) who derived reduction factors for two cloud fields using data from cloud resolving simulations: a nocturnal marine stratocumulus case and a trade wind cumulus field.

A particular concern during the tuning of ECHAM6.3 was obtaining sufficient precipitation on tropical land, specifically in the Amazon and Sahel. This was long an issue in the model but with the increasing focus on the carbon cycle and the usage of dynamic vegetation the situation could no longer be ignored. The problem applies foremost to the CR and LR configurations. Investigating the controls of precipitation on land in atmosphere-only simulations is complicated because a simple reduction in cloudiness will lead to warming land surface temperatures, but the prescribed SSTs do not respond. Thus, in the tropics a direct thermal cell will drive more atmospheric moisture transport inland and thereby cause more precipitation on land, and therefore care must be taken to compensate the radiation balance. During our investigations, we found the problem to be robust to most changes, and the only parameter with a discernible effect is to increase critical relative humidity aloft (a_1, equation (6)). This action, however, moves the parameter further from observational inference that favor even lower values (Quaas, 2012) and increases the equilibrium climate sensitivity away from the target. Thus, a fairly modest increase of a_1 was chosen. Additionally, the resistance to river-runoff in the hydrological discharge model was increased to further enhance the availability of groundwater to vegetation.

The tuning of the subgrid-scale orographic drag parameterization was further refined. For tuning, the strength of low-level blocking (C_d) and gravity wave stress (G) is used (section 3.5). In previous versions of the ECHAM model, these were set equal to simplify the tuning. However, we found that C_d is particularly important for controlling the flow over the Southern Ocean, whereas G was most effective in the Northern Hemisphere. A similar experience for another model is described in Pithan et al. (2016). By controlling the two parameters individually, it was possible to reduce sea level pressure and tropospheric zonal wind biases in the Southern Hemisphere substantially (see Appendix B).

4. Revisions of the Ocean Dynamics Model Component (MPIOM1.6)

Both the updated MPI-ESM1.1 and MPI-ESM1.2 model versions include the ocean component MPIOM1.6.3 and a sea ice model. These model components remained largely unchanged with respect to the MPI-ESM model version used during CMIP5 (for details, see Jungclaus et al., 2013; Notz et al., 2013). MPIOM is formulated on an Arakawa-C grid in the horizontal and on z levels in the vertical direction using the hydrostatic and Boussinesq approximations. Subgrid-scale parameterizations include lateral mixing on isopycnals (Redi, 1982) and tracer transports by unresolved eddies (Gent et al., 1995). Vertical mixing is realized by a

combination of the Richardson number-dependent scheme of Pacanowski and Philander (1981) and directly wind-driven turbulent mixing in the mixed layer (for details see Jungclaus et al., 2013; Marsland et al., 2003).

Variants of MPI-ESM1.2 differ in their ocean horizontal grids, where the MPI-ESM1.2-LR and the coarser MPI-ESM1.2-CR setups apply a bipolar grid, the HR, and higher-resolution configuration use a tripolar formulation, Table 1. The bipolar grids (GR1.5 in MPI-ESM1.2-LR and GR3.0 in MPI-ESM1.2-CR) feature one grid pole under Greenland and one under Antarctica. This choice causes regionally enhanced resolutions in the deepwater formation regions and the overflows across the Greenland-Scotland ridge. The tripolar grids allow for more uniform resolutions and are available in the TP04 version with eddy-permitting 0.4° (as applied in MPI-ESM1.2-HR) and in the eddy-resolving TP6 m with a resolution of 6 min (MPI-ESM1.2-XR), as applied in the simulations of von Storch et al. (2012). In the vertical, 40 levels are unevenly placed in the water column, with the first 20 levels distributed over the top 700 m. The bottom topography is represented by a partial-step formulation (Wolff et al., 1997).

The MPIOM1.6.3 has been improved over earlier versions in terms of online diagnostics, meeting the requirements for OMIP (Griffies et al., 2016), and the flexibility of the model output. During the later stages of the preparation of MPI-ESM1.2, a coding error was discovered in the vertical viscosity scheme of MPIOM. Fixing this bug results in a considerable reduction of the SST biases in upwelling regions as well as in the Southern Ocean (Figures 6a and 6c) but also caused a slight surface cooling drift of about 0.2–0.3 K globally (Figure 6b). The latter required a small retuning of the model. Owing to various deadlines, MPI-ESM1.2-HR had already begun running its DECK runs and it was therefore deemed too late to implement the bug fix into MPI-ESM1.2-HR, but the LR version received this correction.

The sea ice model consists of code both in MPIOM and in ECHAM. In ECHAM, a simplified thermodynamic sea ice model is incorporated to provide at each atmospheric time step a physically consistent surface temperature in ice-covered regions. This part of the sea ice model contains a melt-pond scheme, which divides the surface of the sea ice into snow, bare ice, and melt pond, with each their individual albedo (Pedersen et al., 2009). In MPI-ESM1.2, these melt ponds are now fully activated, in contrast to an erroneous only partial activation in MPI-ESM (section 3.5, Roeckner et al., 2012). The atmospheric part of the sea ice model integrates all surface fluxes and provides at each coupling time step the gridded bulk surface flux into sea ice to the sea ice model in MPIOM. This surface flux is then used in MPIOM to calculate the sea ice surface energy balance and related changes in ice thickness. The thermodynamic description of sea ice is based on a simple zero-layer, mono-category formulation (Semtner, 1976). The differentiation of thermodynamic sea ice growth or melt between lateral processes that change ice concentration and vertical processes that change ice thickness is parameterised and can be adjusted by two tuning parameters (for details, see Mauritsen et al., 2012; Notz et al., 2013). These two parameters are used to tune the preindustrial Arctic sea ice volume of MPI-ESM1.2 to an annual average of roughly 20–25,000 km^3. The values of the tuning parameters are slightly adjusted in MPI-ESM1.2 relative to MPI-ESM to compensate for an otherwise too low sea ice volume.

For sea ice dynamics, MPI-ESM1.2 uses a viscoplastic rheology following Hibler (1979). This part of the sea ice model is also fully contained within MPIOM. At each coupling interval, MPIOM relates the updated sea ice thickness and sea ice concentration to ECHAM, which is then used there until the next coupling instance.

5. Revisions of the Ocean Carbon Cycle Model Component (HAMOCC6)

Processes represented in the Hamburg Ocean Carbon Cycle model are extended relative to that described in Ilyina et al. (2013) and parameters used in empirical relationships of existing processes are updated. As a new feature HAMOCC6 now resolves nitrogen-fixing cyanobacteria as an additional prognostic phytoplankton class (Paulsen et al., 2017). This replaces the diagnostic formulation of N_2 fixation applied in MPI-ESM-LR and allows the model to capture the response of N_2 fixation and ocean biogeochemistry to changing climate conditions. Updates of existing parameterized processes follow recommendations of the C4MIP and OMIP protocols (Jones et al., 2016; Orr et al., 2017).

5.1. Marine Nitrogen Fixation by Cyanobacteria

The parameterization of prognostic cyanobacteria as additional phytoplankton class is based on the physiological characteristics of the cyanobacterium *Trichodesmium* (Paulsen et al., 2017). Cyanobacteria in the model differ from bulk phytoplankton by their ability to grow on both nitrate (NO_3) and dinitrogen (N_2). Fur-

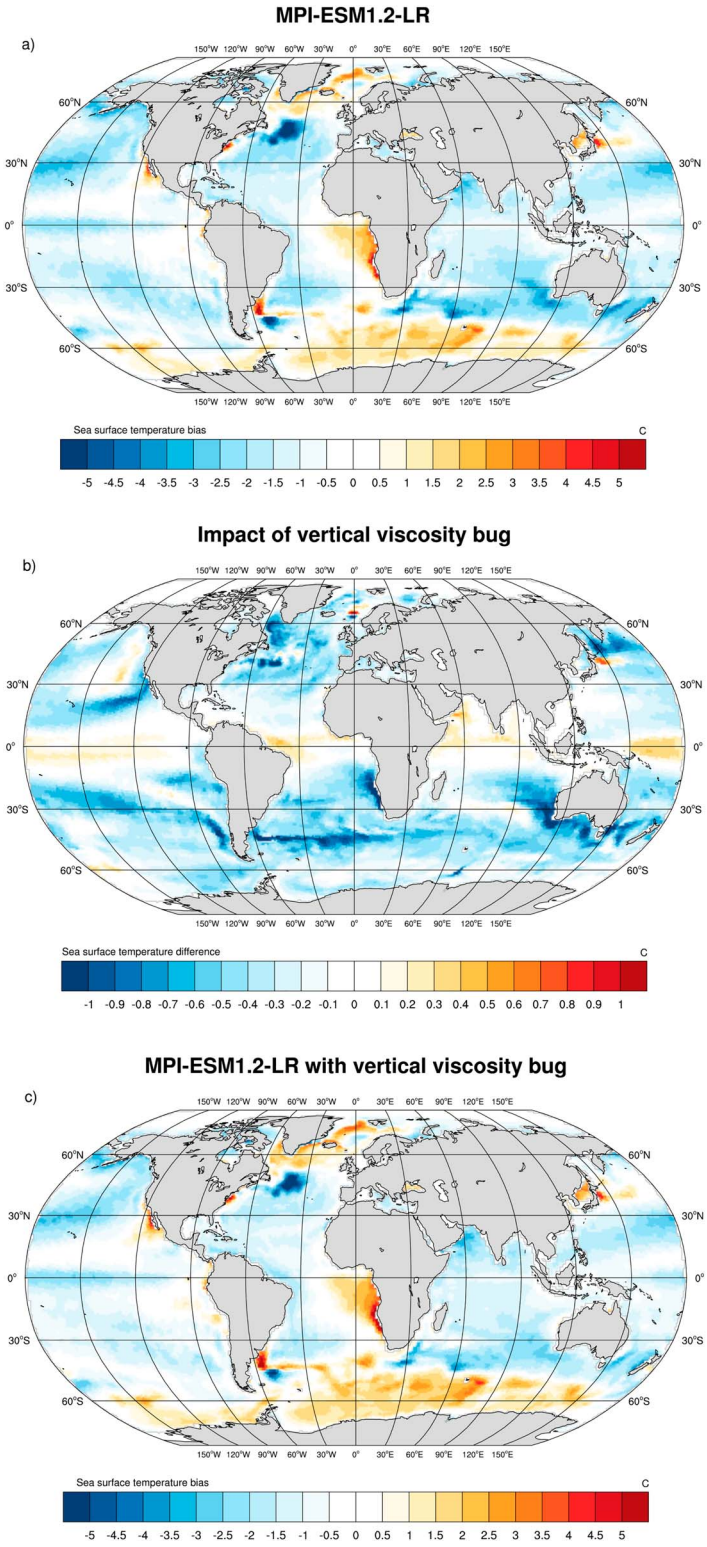

Figure 6. Sea surface temperature biases (50-year average) in preindustrial control simulations with respect to observed climatology (Steele et al., 2001): (a) MPI-ESM1.2-LR after applying the vertical viscosity bug fix and (c) the otherwise identical MPI-ESM1.2-LR but including the coding bug. (b) The difference between (a) and (c); (b) has a different color scale.

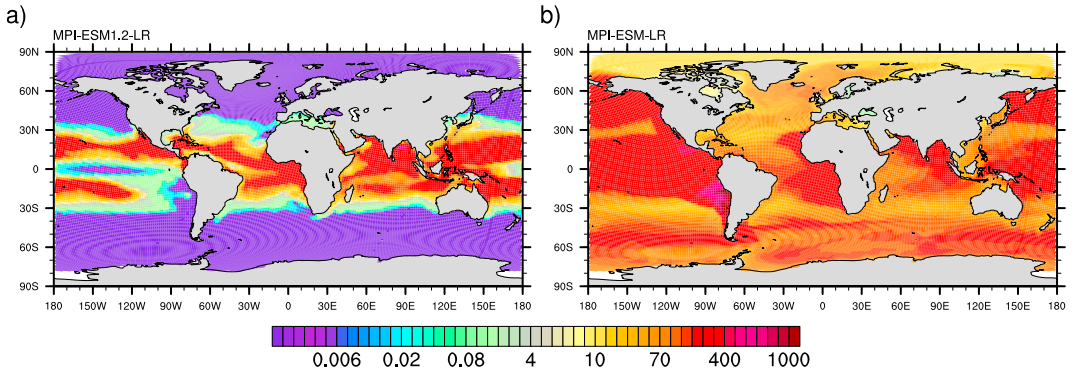

Figure 7. Dinitrogen (N_2) fixation ($\mu mol\, N \cdot m^{-3} \cdot day^{-1}$) in preindustrial control simulations at the sea surface in (a) the MPI-ESM1.2 version and (b) in the predecessor MPI-ESM model version used during CMIP5.

thermore, cyanobacteria have a slower maximum growth rate, are limited to a specific optimum temperature range, face a stronger iron limitation, and are positively buoyant. This way, modeled N_2 fixation is evolving in response to the combined effect of temperature, light, and nutrient distributions, that is, environmental conditions shaping cyanbacteria's ecological niche.

The implementation of prognostic cyanobacteria modifies the spatial distribution of N_2 fixation. In the diagnostic formulation, N_2 fixation was prescribed by a constant rate being active only in case of nitrate concentrations below the Redfield ratio (Ilyina et al., 2013). This resulted in unrealistically high N_2 fixation rates in high latitudes and a strong coupling of N_2 fixation to the nitrate-depleted regions overlying denitrification sites, that is, in the eastern tropical Pacific, eastern tropical Atlantic, and in the northern Indian Ocean (Figure 7b). In MPI-ESM1.2 the abundance of cyanobacteria is the premise for N_2 fixation. Thus, fixation is restricted to the warm tropical and subtropical ocean, roughly between 40° S and 40° N, a result of temperature-limited cyanobacteria growth (Figure 7a). Furthermore, the availability of nutrients (iron and phosphate) and the competition with phytoplankton shape the areas of fixation. A detailed description of the parameterization and an evaluation of the representation of cyanobacteria and N_2 fixation against observations is given in Paulsen et al. (2017).

In MPI-ESM1.2, one modification was made to change the fate of dead cyanobacteria. The major fraction is considered as detritus and thus being affected by gravitational sinking. A small nonsinking fraction is considered as dissolved organic matter (DOM), which in Paulsen et al. (2017) had a decay time of several years. In MPI-ESM1.2, this decay rate was reduced to several months, equivalent to that of the already existing DOM pool (being fed by phytoplankton). As sensitivity studies did not show considerable differences due to this rate parameter change, we combined both DOM pools to one at the benefit of lower computational cost.

5.2. Additional Changes and Bug Fixes

Previously, in HAMOCC, plankton dynamics were calculated only within the upper 100 m of the ocean, assuming low light availability would impede plankton growth below. This neglects grazing activity below 100 m and introduces temporally too high phytoplankton concentrations at depths after mixing below 100 m. Therefore, we now extend biological process calculations to the whole water column.

Detritus settling was modified by replacing the constant uniform settling rate of 5 m/day by a vertically varying settling rate based on observed particle fluxes (Martin et al., 1987). Martin and coauthors measured POC fluxes in the open ocean, found low spatial variability, and fitted the data to a normalized power law function. We implemented this fit by a depth-dependent settling velocity, which is 3.5 m/day within the upper 100 m and increases linearly below, up to 80 m/day at 6,000 m.

In HAMOCC6 we consider remineralization of organic material on dissolved oxygen, nitrate, nitrous oxide, and sulfate. Thereby, nutrients and dissolved inorganic carbon are released to the water column. Remineralization, as well, introduces changes to alkalinity, which depend on the composition of the organic material (i.e., the prescribed Redfield ratio in organic material) and the oxidation pathway. Previously, we assumed that alkalinity decreases during remineralization on sulfate. This would imply an instantaneous oxidation of hydrogen sulfide (H_2S) being produced by sulfate reduction. However, as sulfate reduction occurs only under nearly anaerobic conditions H_2S oxidation is unlikely to occur. Therefore, we introduced an addi-

Table 3
Parameter Setup for HAMOCC6 in Two Configurations of the MPI-ESM1.2 Model

Configuration	HR	LR
Parameter		
Grazing rate (day^{-1})	0.7	1.0
Initial slope P-I curve ($W^{-1} \cdot m^2 \cdot day^{-1}$)		
Bulk phytoplankton	0.04	0.025
Cyanobacteria	0.15	0.03
Cyanobacteria mortality rate (day^{-1})	0.04	0.10
Cyanobacteria half saturation constant for iron limitation ($mmol/m^3$)	0.6	0.8
Weathering rates		
Organic material ($kmol \cdot P \cdot s^{-1}$)[a]	5	4
DIC, alkalinity ($kmol \cdot C \cdot s^{-1}$)[b]	537	428
Silicate ($kmol \cdot Si \cdot s^{-1}$)	0	100

Note. Added as DOM, thus, all of N, P, Fe, O_2, and C are added following the composition of organic material in HAMOCC6.
DIC and alkalinity are added in the ratio 1:2.

tional prognostic tracer (H_2S), which is produced during remineralization on sulfate and decays only in oxygenated water. The purpose of H_2S is to track alkalinity changes and to account for the local alkalinity increase due to sulfate reduction in oxygen minimum zones. However, this revision has only a minor impact (less than 1%0) on local alkalinity concentrations. This updated sulfate cycle is only implemented into MPI-ESM1.2-LR, while the MPI-ESM1.2-HR version did not receive this new feature (Table 3).

With respect to remineralization we also reformulated the oxygen dependence for aerobic and anaerobic processes. In MPI-ESM there was a definite oxygen threshold value that limited aerobic remineralization (Ilyina et al., 2013). The corresponding turnover rate was independent of the ambient oxygen concentration in the water. To allow for a gradual transition between aerobic and anaerobic zones, we introduced a Monod kinetic type oxygen limitation Γ_{oxy} with a half saturation value of 10 μmol O_2/L. Thus, the aerobic remineralization rate decreases with decreasing dissolved oxygen concentration. As a complement, anaerobic remineralization rates are modulated by $(1 - \Gamma_{oxy})$. All maximum remineralization rates and the critical oxygen concentration at which anaerobic processes start to occur are kept as in MPI-ESM.

Among other processes, primary production is a function of iron availability, which is mainly controlled by atmospheric dust deposition. In MPI-ESM1.2, the dust deposition climatology from Mahowald et al. (2006) is replaced with that of Mahowald et al. (2005). The latter has an overall lower input of dust and a slightly different spatial pattern, which we find preferable. Especially, the high dust deposition in the eastern Pacific of the Mahowald et al. (2006) climatology, which lead to excessive growth of cyanobacteria in the eastern equatorial Pacific, is alleviated in Mahowald et al. (2005), and hence, we reverted to the older climatology.

In addition to scientifically motivated model refinement and bug fixes, we modified the parameter setup of several processes to comply to the agreements of protocols for model setup within CMIP6. In line with the OMIP protocol (Orr et al., 2017), we updated the formulation of the carbonate chemistry. In particular, we use the total pH scale and equilibrium constants recommended by (Dickson, 2010; Dickson et al., 2007). Furthermore, total alkalinity now additionally considers the alkalinity from phosphoric and silicic acid systems. Following the suggestions by the OMIP protocol (Orr et al., 2017), we updated the gas-exchange parameterization. Now we use the gas transfer velocity formulation and parameter setup of Wanninkhof (2014). This includes updated Schmidt number parameterizations for CO_2, O_2, DMS, and N_2O. To allow for simulations with transient anthropogenic atmospheric nitrogen deposition within C4MIP (Jones et al., 2016), we implemented nitrogen deposition as an additional nitrate source. Wet and dry deposition of all nitrogen species (NO_y+NH_x) is directly used to update NO_3, assuming instantaneous unlimited oxidation of NH_x in sea water. The corresponding H+ change is accounted for as an alkalinity decrease. Gridded coupled chemistry model intercomparison N-deposition fields of version 1.0 are used as provided via the CMIP6 input database (https://esgf-node.llnl.gov/projects/input4mips/).

5.3. Tuning of HAMOCC

The MPI-ESM1.2-HR and MPI-ESM1.2-LR configurations differ in the atmosphere-ocean coupling frequency (section 2). Primarily, therefore, a different tuning of biological parameters in HAMOCC6 is required, as due to light limitation of plankton growth source-sink dynamics differ when using daily mean light versus considering a day-night cycle. The default tuning of the model assumes daily mean light. Using this parameter set in HR would lead to a too strong depletion of the phytoplankton standing stock during night. The general tuning strategy for HR is to increase growth during day and reduce mortality or grazing pressure to ensure a reasonable biological carbon pump and nitrogen fixation.

Global values of N_2 fixation (MPI-ESM1.2-HR approximately 67 Tg N/year and MPI-ESM1.2-LR approximately 82 Tg N/year) are in line with the range of observationally based estimates of about 70 to 200 Tg N/year (e.g., Karl et al., 2002; Grokopf et al., 2012). N_2 fixation compensates N loss by denitrification on longer times scale. Thus, the lower fixation rate in HR fits well to the lower export of carbon in MPI-ESM1.2-HR (approximately 6 Gt C/year) compared to MPI-ESM1.2-LR (approximately 6.7 Gt C/year), both lying within the reported range of export fluxes (3–20 Gt C/year, Najjar et al., 2007).

HAMOCC6 includes a prognostic sediment with a burial layer, which constitutes an ultimate loss of nutrients and carbon to the system. This loss is compensated for by constant weathering rates, which are derived from a linear regression of the long-term (approximately 100 years) temporal evolution of the sediment inventory during the spin-up in a stable climate state.

6. Revisions of the Land Physics and Biogeochemistry Components (JSBACH3.2)

JSBACH is the land component of MPI-ESM1.2. It provides the lower boundary conditions for the atmosphere and includes components to describe the dynamics of the land biogeochemistry in interaction with global climate. The updated JSBACH3.20 used in MPI-ESM1.2 has undergone a host of changes compared to the predecessor applied in MPI-ESM. Some process descriptions have been replaced by more refined formulations (soil hydrology, soil carbon decomposition, and fire), a few other processes known to be of relevance for the Earth system have been added (land nitrogen cycle and carbon storage of wood products), and finally the description of land use change was adapted to conform with the revised land use change data prepared for CMIP6.

6.1. Soil Hydrology

JSBACH inherits its land physics from ECHAM5 (Roeckner et al., 2006). With the recent revision, the single soil water bucket scheme was replaced by a multilayer description based on the Richards equation to account for the vertical movement of water in the soils (Hagemann & Stacke, 2015). The layer thicknesses are chosen identically to those employed for the heat conduction in the soil (0.065, 0.254, 0.913, 2.902, and 5.7 m). The multilayer hydrology model uses a revised data set for water holding capacities having lower capacities in arid and semiarid areas and increased capacities in areas with tropical rain forest. While the water bucket scheme represented only the root zone moisture, the multilayer scheme now accounts also for soil water storage below the root zone. In regions where the rooting depth is smaller than soil depth, this additional buffer for soil water leads to shorter dry periods (Hagemann & Stacke, 2015). Another consequence of the shift to a layered hydrology concerns the behavior of bare soil evaporation. While in the bucket model, bare soil evaporation depended on the total soil moisture content, that is, the filling of the upper 10 cm of the bucket, in the multilayer hydrology model, it depends more realistically only on the moisture in the uppermost soil layer. The consequence is that bare soil evaporation responds more promptly to precipitation events.

6.2. Soil and Litter Decomposition

The version of JSBACH applied in the predecessor MPI-ESM contained a rather simple model for decomposition of dead organic matter (soil and litter) with most parameters unconstrained by observations. Accordingly, they were used to tune the land carbon cycle to reproduce the historical development of atmospheric CO_2 in coupled simulations irrespective of the resulting soil and litter carbon stocks. To improve the latter, decomposition is now described by the YASSO model, which is based on a compilation of litter decomposition data and soil carbon data without any unconstrained parameters (Tuomi et al., 2011). In YASSO, decomposed organic matter is divided into four solubility classes (acid soluble, water soluble,

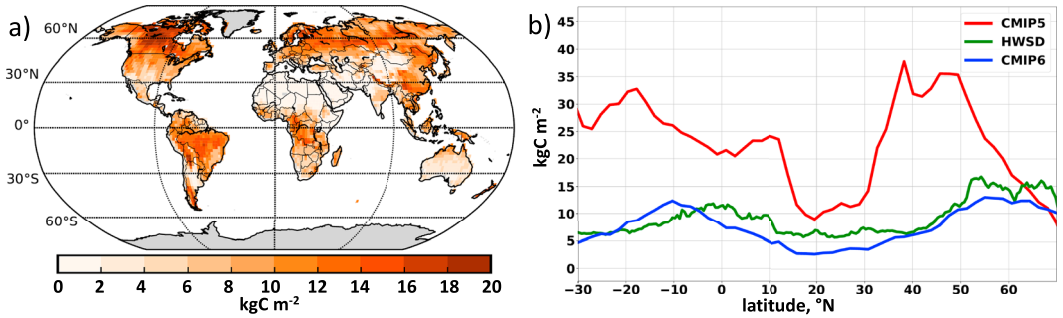

Figure 8. Evaluation of soil carbon distribution. (a) Density of soil and litter carbon (kg C/m^2), in the CMIP6 historical simulation with MPI-ESM1.2-LR, at year 2005. (b) Zonally averaged density of soil and litter carbon in the MPI-ESM-LR (red) and MPI-ESM1.2-LR (blue) model versions comparing to soil carbon in the Harmonized World Soil Database (HWSD, green).

ethanol soluble, and nonsoluble) and one additional class for humic and slowly decomposing substances. Decomposition is simulated as a first-order decay process depending on air temperature, water availability, and litter size (Goll et al., 2015). In JSBACH3.2 this pool structure is applied separately to the decomposition of organic matter originating from woody and nonwoody plant remains, and also separately for aboveground and belowground decomposition, resulting in a total of 18 carbon pools for dead organic matter.

The soil carbon distribution in simulations with the predecessor MPI-ESM exhibited maxima in continental dry lands due to its parameterization of decomposition that depended on soil moisture. Total carbon storage was considerably overestimated in comparison with observations from the Harmonized World Soil Database (Figure 8b). On the contrary, the soil carbon model YASSO applied in MPI-ESM1.2 simulates plausible soil density patterns with relative maxima in tropical and boreal forests (Figure 8a), and the distribution is comparable with observations (Figure 8b). A detailed evaluation of the YASSO decomposition model in terms of simulating the spatial variability of soil carbon stocks using estimates from the Harmonized World Soil Database is provided by Goll et al. (2015). JSBACH as applied in MPI-ESM1.2 does not include permafrost carbon, and maximum of soil carbon storage in high northern latitudes, and hence, is not well reproduced by the model. These processes are applied to future implementations in ICON-ESM.

6.3. Land Use Change

Throughout history, land cover has changed with the agricultural expansion. This has consequences for the surface radiation balance and in particular for the carbon cycle, because especially by the large-scale replacement of forests by agricultural land CO_2 is released to the atmosphere. This is not only a result of a cutback in carbon stored in vegetation but also a consequence of a long-term decline of soil carbon associated with agriculture. To account for such processes in simulations, forcing data are required describing how agricultural usage of land changes over time. For CMIP6, an updated version of such data was released by the Land Use Harmonization (LUH) project (Hurtt et al., 2011). This version, LUH2v2h (Hurtt et al., 2016), comprises 12 land use states and transitions among most of them. The land use implementation in JSBACH, as described in (Reick et al., 2013), is based on three types: natural vegetation, crops, and pasture. To use the LUH2v2h data in JSBACH, its states and transitions between need to be mapped in a preprocessing step: All four LUH2v2h natural vegetation types, as well as rangelands, are aggregated to the JSBACH natural vegetation type; all five LUH2v2h crop types are aggregated to crops; and the LUH2v2h types pasture and urban are aggregated to the JSBACH pasture type. Subsequently crops, pasture, and natural vegetation are scaled proportionally to sum up to unity. Concerning wood harvest, all five LUH2v2h wood harvest categories are aggregated in the preprocessing to a single value because JSBACH does not distinguish between different types of wood harvest.

In the LUH2v2h data set, there is no separate category for deserts and other bare land; these surface types are included in the natural vegetation, while they are treated as different from vegetated surfaces in JSBACH. Accordingly, for usage in JSBACH the LUH2v2h extent of natural vegetation has to be reduced. For this reduction, the bare land fraction of a grid cell needs to be known, but JSBACH calculates this bare land fraction interactively during runtime by the dynamical vegetation component. Hence, these bare land fractions are not available for a preprocessing of the LUH2v2h data. Previously, we largely ignored this problem by

Table 4
Fractions of Carbon Fluxes From Wood Harvest, ALCC, and Crop Biomass Allocated to Litter Carbon and the Seven Product Pools (3 for Wood Harvest, 3 for ALCC, and 1 for Crop Harvest)

From	To	Tropicalforest	Extra- tropical forest	Shrubland	Grassland	Crops
Wood Harvest	Litter	0.300	0.300	0.3	—	—
	Fuel	0.630	0.280	0.7	—	—
	Paper	0.028	0.168	0	—	—
	Construction	0.042	0.252	0	—	—
ALCC	Fuel wood	0.597	0.5970	0.8	1	—
	Paper	0.403	0.2985	0.2	0	—
	Construction	0	0.1045	0	0	—
Crop biomass	Litter	—	—	—	—	0.5
	Harvest	—	—	—	—	0.5

Note. ALCC = anthropogenic land cover change.

equally scaling down the extent of all vegetation types, but we now employ another compromise: Taking the preindustrial bare land fraction from a separate JSBACH simulation, we corrected during preprocessing the LUH2v2h grid cell fraction of natural vegetation but kept crop and pasture extent unchanged. With these input data, during simulation time fractions of natural vegetation but also of agricultural land are further modified by the interactively calculated bare land fraction. Since there are no major shifts in the bare land fraction during historical times, by this approach the extent of agricultural lands actually used in our simulations still matches the cropland and pasture area from LUH2v2h and the underlying data from the Food and Agriculture Organization. A drawback is that in scenarios with strong climate change in certain areas where the bare land fraction also undergoes strong changes, agricultural area assumed in the simulations may differ considerably from the LUH2v2h data.

Concerning the land use effects on the land carbon cycle, JSBACH3.2 has been extended by the implementation of product pools, where carbon is respired to the atmosphere independent of climate conditions. JSBACH3.2 has now seven product pools all filled from above ground carbon. More precisely, one pool is filled by crop harvest, calculated as half of the crop litter flux (as determined by phenological leaf losses). Carbon from wood harvest is stored in a set of three pools denoted as "fuel wood," "paper," and "construction" and vegetation carbon lost by anthropogenic land cover change (ALCC) is stored in a similar set of three pools. The LUH2v2h wood harvest is reduced by a slash fraction of 30%, which is consistent with the processing of the LUH2v2h data. This slash is passed to the YASSO litter pools. The other 70% of the wood harvest is distributed to the three wood harvest pools according to the values assumed in the BLUE model (Hansis et al., 2015), which are based on Houghton et al. (1983) and the assumption that harvest from shrublands can only be used for the purpose of fuel burning. Vegetation carbon affected by ALCC is apportioned to the three ALCC pools based on Houghton et al. (1983). Table 4 summarizes the fractional allocation to litter carbon and the product pools. The carbon decays in all product pools with constant rates and is respired to the atmosphere. The e-folding time is 1 year for the crop harvest pool, and 1 year (fuel wood), 10 years (paper), and 100 years (construction) for both the wood harvest and the ALCC pools.

6.4. Fire

The model land biosphere is affected by fire, which directly impacts land carbon storage, for example, aboveground litter, living biomass, and the vegetation distribution. Earlier versions of JSBACH included a simple fire parameterization based on litter availability and air humidity (Reick et al., 2013) and assumed globally constant impacts per burned area on carbon pools and vegetation. In JSBACH3.20, fires are now represented by the process-based SPITFIRE model (Lasslop et al., 2014; Thonicke et al., 2010). SPITFIRE determines fire occurrence in terms of the burned area, computes the carbon and nitrogen emissions, and carbon and nitrogen redistribution on the affected areas. Vegetation distribution is affected through fire-caused tree mortality and nitrogen losses. The computation of burned area is based on a combination of ignitions by humans and lightning, fire spread rate and fire duration. Main drivers for the spread rate are fuel availability, type, and moisture. The fire duration is additionally limited by population density (Hantson et al., 2015) and for technical reasons to last for 12 hr. The completeness of biomass combustion depends on fuel moisture, the tree

mortality on fire intensity and residence time of the fire. A detailed description of the model equations and a comparison to other global fire models can be found in Rabin et al. (2017).

The main tuning factor for the total global burned area is the number of ignitions. Both the numbers of lightning ignitions (derived from remote sensing) and human ignitions (derived from population density) are highly uncertain. The balance between types of ignitions is constrained by the global spatial patterns of burned area: A higher relative amount of lightning ignitions leads to a higher contribution of boreal regions to the total burned area. With the implementation of YASSO, the amount of carbon in the litter pools increased, especially the nonwoody litter of woody plant functional types. This led to too high fuel loads in forested areas so that in the model, it is now assumed that only a prescribed fraction of the litter actively contributes to fire spread and that the other part of the fuel is too compact to burn.

6.5. Nitrogen Cycle

The inclusion of nitrogen constraints on plant growth into land surface models was shown to reduce the carbon-climate feedback and the carbon-concentration feedback (Zaehle, 2013). A representation of the terrestrial nitrogen cycle and its interactions with the carbon cycle was introduced into JSBACH3.2 (Goll et al., 2017) based on a previous model version (Parida, 2010). The nitrogen cycling in JSBACH3.2 is driven by the nitrogen demand of the carbon cycle assuming constant carbon-to-nitrogen ratios of organic pools, with the exception of nonwoody litter (Parida, 2010). Total land nitrogen storage changes due to the processes linking the terrestrial nitrogen cycle with the atmosphere (biological nitrogen fixation and denitrification) and the land hydrosphere (leaching), which are modeled either as substrate limited or, in the case of biological nitrogen fixation, as demand driven due to the ample supply of N_2 from the atmosphere (Parida, 2010). The interactions between nitrogen availability, plant productivity, and decomposition are based on the concept of CO_2-induced nutrient limitation (Goll et al., 2012). The model reproduces the spatial variability in the reactive nitrogen loss pathways as derived from a compilation of $\delta^{15}N$ isotopic data (Goll et al., 2017). Idealized simulations furthermore demonstrated that the inclusion of such a nitrogen cycle leads only to a minor dampening of the carbon-concentration feedback and has almost no effect on the carbon-climate feedback (Goll et al., 2017).

Because organic matter is composed of both carbon and nitrogen, the land nitrogen cycle is affected in a similar way as the carbon cycle by vegetation dynamics, fire, crop harvest, wood harvest, and land cover change. Generally, the corresponding nitrogen fluxes are obtained from the associated carbon fluxes using the fixed carbon-to-nitrogen ratios. Exceptions are the biomass fluxes to the product pools. The associated nitrogen is assumed to be lost for the land biosphere as a diagnostic flux to the atmosphere. By contrast, the nitrogen released by the decay of biomass in the crop harvest pool is ascribed to the soil mineral nitrogen pool to compensate for the lack of fertilizer application in JSBACH3.2.

6.6. Interactions Between Fire, Vegetation Dynamics, and Nitrogen Cycle

The updated processes implemented in JSBACH3.2 lead to a number of interactions previously absent. In particular, an interesting interplay between the fire dynamics, land nitrogen cycling, and the dynamics of vegetation distribution was found. As a result of biomass burning, organic nitrogen is lost not only to the soils but also to the atmosphere. The resulting reduced nitrogen availability reduces plant productivity so that the dynamic vegetation module renders regions prone to wildfires less hospitable to vegetation and thus increases the bare land fraction. This loss of nitrogen, productivity, and vegetation cover is only incompletely compensated by the prescribed nitrogen deposition. Presumably, a regional redeposition of nitrogen emissions from combusted organic material (Bauters et al., 2018) is missing in the model. Overall, these complex interactions result in less vegetation and less fire so that in MPI-ESM1.2-LR, the burned area is underestimated compared to observations.

6.7. Adjustment of Vegetation and Land Carbon/Nitrogen Parameters

Carbon cycle parameters were adjusted in JSBACH3.2 to improve the internal consistency of the model and the realism of its results. The timescale of carbon in the wood pools is now plant functional type specific and the same as for establishment and mortality of the respective forest types in the dynamic vegetation computations. Additionally, the maximum content of the wood pool of extratropical forest is increased by 20%. Such a limitation was introduced to avoid unrealistically high stand volume under extraordinary good growth conditions (e.g., at very high CO_2 concentration); however, the original value for the maximum content appeared to be too restrictive already at standard circumstances.

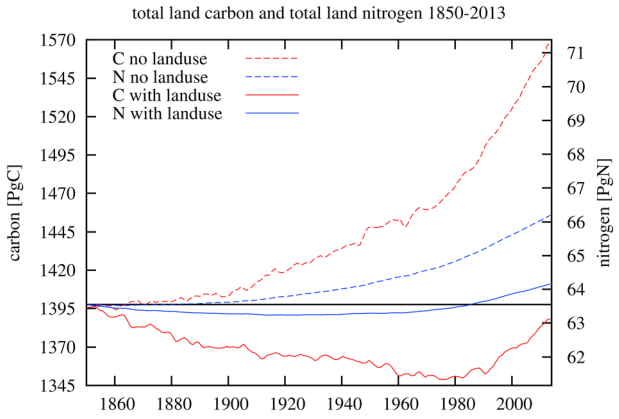

Figure 9. Global total land carbon (red) and global total land nitrogen (blue) in historical simulations with land use change (solid lines) and without land use change (dashed lines). The black horizontal line indicates the average preindustrial value of the global total land carbon and nitrogen, respectively.

Another change was done in the calculation of the fraction of absorbed photosynthetically active radiation. Here JSBACH3.2 accounts for gaps in the canopy by a clumping factor, which, according to Loew et al. (2014), results in an underestimation of the fraction of absorbed photosynthetically active radiation in a variety of canopy scenarios. Therefore, the clumping is reduced for thin canopies and not applied for $LAI > 3$.

Overall, with the implementation of product pools and YASSO, the carbon cycle model has fewer uncertain parameters. For the product pools the carbon community generally uses a set of parameter values (see Table 4) originating from (Houghton et al., 1983) to have comparable results across different land carbon cycle models. In YASSO the decomposition rates and pathways are determined from measurements and are thus integral parts of the model.

Accordingly, the carbon cycle model is run, including dynamic vegetation, without further adjustments of the parameters close to equilibrium in a multimillennial preindustrial simulation. Then, the nitrogen model is initialized with this carbon cycle state by means of the fixed carbon-to-nitrogen ratios and a standard value for the mineral nitrogen pool. After an equilibration period of again several thousand years, the leaching rate was adjusted to obtain leaching and denitrification rates both of the order of 50 Mt N/year and a nitrogen limitation of the order of 1% of global net primary productivity (NPP). A subsequent simulation showed that by the mechanism described in section 6.6, the vegetation cover in regions with frequent fires is reduced excessively. Therefore, the nitrogen loss is reduced by returning half of the nitrogen of the burned biomass to the soil mineral nitrogen pool. Another equilibration run of several millennia followed. All these extensive simulations were performed in the off-line mode of JSBACH3.2 to save computing time, using climate data generated by a coupled atmosphere/ocean simulation of MPI-ESM1.2-LR with preindustrial boundary conditions. The resulting vegetation-carbon-nitrogen state was then transferred to the Earth System model including all components for a final equilibration run of more than 3,000 years. In the end, a satisfactory equilibrium of the land carbon cycle is obtained with a drift in total land carbon of less than 1 Pg C per century.

6.8. Land Carbon and Nitrogen Cycle Results

At the preindustrial equilibrium the land vegetation has a global gross primary productivity of 121 Pg C/year and a global NPP of 62 PgC/year. In the last 20 years of a historical simulation, it increases to a gross primary productivity of 152 Pg C/year and a global NPP of 74.5 Pg C/year mainly as a result of CO_2 fertilization. Figure 9 shows a loss of carbon and nitrogen in the period 1850–1960 and a reversal to a carbon and nitrogen gain in the recent decades. This time evolution of global land carbon is in line with estimates of Ciais et al. (2013) that are based on the ocean carbon inventory and anthropogenic fossil fuel emissions. Until the end of a historical simulation without land use change total land carbon rises to 180 Pg C more than is the case in a simulation that includes the impacts of land use change. This difference in land carbon can be considered as accumulated land use change emissions and is also in line with numbers presented in Ciais et al. (2013). The agreement with common estimates of land carbon uptake and land use change emissions is obtained without any particular parameter adjustments. By contrast, in the previous model setup, the fraction of biomass affected by land use and directly emitted as CO_2 to the atmosphere was adapted to achieve these objectives.

7. Technical Improvements

A number of improvements were made to the infrastructure to enhance computational performance and usability of the model system. This includes both small and large changes to the code, some of which are outlined here.

Running a model with a spectral dynamical core on modern massively parallel computers is challenging because of all-to-all processor communication when transferring to and from spectral and grid-point space. Therefore, communication becomes a bottleneck at often relatively modest processor numbers. A new communication library, YAXT, was implemented which, depending on the experiment, can improve

the scalability of the model up to approximately twice as many processors compared to earlier versions, and as a result can considerably improve the maximum throughput, which can be achieved on a given machine. Likewise, considerable performance gains were achieved by choosing more optimal settings for the compilers, extending the use of vectorization, and switching to using optimized mathematical libraries. With the transition from coupling using OASIS3 (Valcke, 2013) to OASIS3-MCT (Craig et al., 2017), we removed another performance bottleneck. The data exchange between compute processes is now performed in parallel and directly from processor to processor rather than by passing the coupling fields via separate coupling processes.

Another expensive part of running a climate model is the output. Previously, the strategy was to run the model once with extensive output, for example, every 6 hr, whereafter monthly or yearly means were calculated. With the relative cost of computing processors going down faster than the cost of storage, this strategy was revised such that it is now often advantageous to calculate the temporal means in memory while the model is running and only outputting, for example, monthly means or whatever a specific scientific purpose requires. This may mean that the same experiment will be run multiple times with tailored output.

For cases wherein substantial output is desired, in particular at high resolutions, the actual writing of data can substantially slow down the execution of the model. This is because the output was previously done using a serial output scheme whereby one of the compute processors would also be used to write output, and meanwhile, the other processors would be idle. To alleviate this issue, an asynchronous output scheme wherein some processors are dedicated to writing while the others can continue to compute has been implemented. The overhead in practice is the resources added for dedicated output tasks, plus the time to copy decomposed data for output to a memory buffer on the compute tasks, plus the effect of additional background data transfers from compute to output tasks. This holds as long as the time-dedicated tasks need for output is less than the time the compute tasks need to enter the next output step.

Together with advances in high performance computers, these technical improvements of the ECHAM6.3 model have substantially improved the scalability of the model to a larger number of processors (Amdahl, 1967), such that longer and/or higher-resolution experiments can now be conducted. For instance, the standard LR resolution with 47 levels (T63L47) can now be scaled to about 85 years per physical day in the current German Climate Computing Center (DKRZ) machine *Mistral* using its fastest Intel Haswell cores with a base clock frequency of 2.5 GHz, compared to about 17 years per day for the predecessor model used during CMIP5 run on the previous generation IBM Power6 575-based machine *Blizzard*. Some of the performance gain is due to using a newer computing architecture and a faster network, but most is attained from model infrastructural improvements. Since ECHAM is the slowest part of the coupled model, these performance increases are mirrored in the MPI-ESM1.2 model. By default, though, a typical setup of the coupled model yields around 45 years per day with 504 Intel Broadwell processing cores with 2.1-GHz base clock frequency.

A system to generate bash scripts to run the model was developed in Python. The idea is to separate system-specific and experiment-specific settings for the user. That is, the same template for an experiment can be used on different machines without the user having to necessarily know about system-specific settings, close to optimal parallelization, etc. For a number of users, there was a desire to have access to these more technical settings in a transparent way. Therefore, the new script-generating system creates single scripts with all settings, both experiment and system specific, such that both novice and experienced users of the model can work effectively. Additionally, the script generation system creates scripts to postprocess, monitor, and quickly plot standard graphics from the experiment. By version-controlling the quick-plotting code with the model code, it is ensured that the scripts are tested and maintained.

8. Coupled MPI-ESM1.2 Experiments

The last step in a model development is the tuning and spin-up of the coupled model. Here the main goal is to minimize drift by fine-tuning the radiation balance, and only to a limited extent are other issues addressed at this stage. The following will demonstrate how the spin-up and final tuning is conducted at the Max Planck Institute for Meteorology. Furthermore, we inspect a set of historical simulations and verify that the match with the instrumental record is improved relative to MPI-ESM (see also section 3.6). We then investigate the models response to weak and strong CO_2 increases and find that a previously identified nonlinear behavior (Heinemann et al., 2009; Meraner et al., 2013) is retained despite the removal of a series of state-dependent

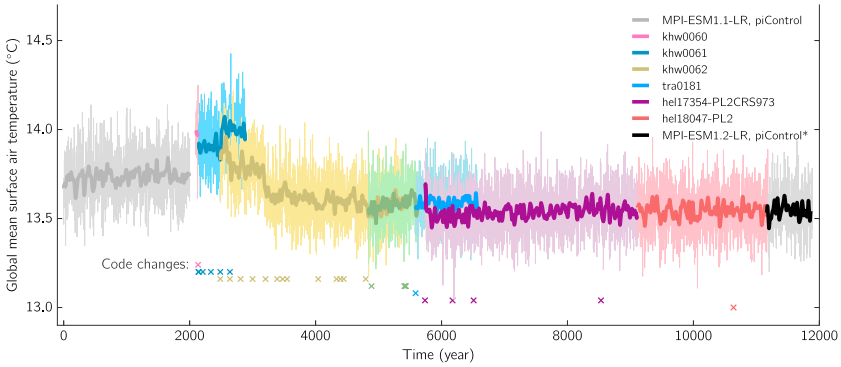

Figure 10. Timeseries illustrating the spin-up process with multiple consecutive runs that often overlap. The experiments khw0060 to hel18047-PL2 are the spin-up runs named according to the institutes naming scheme. Thin lines are annual means, thick lines are smoothed with a timescale of 25 years. Code changes mostly pertaining to the tuning of the physical climate and the carbon cycle, here marked with crosses, are conducted on the fly while the model is spun up. Data from a short simulation preceding khw0060 was lost and so could not be plotted. *The realization of the MPI-ESM1.2-LR piControl experiment displayed here will differ from the run that will eventually be published in CMIP6 because of a change in the calendar.

energy leakages associated with coding errors (section 3.4). The study is rounded off by providing an explanation for the models somewhat surprising nearly constant temperature feedback and finally by showing that the overall nonlinear behavior can be represented by a simple two-layer model.

8.1. Model Spin-Up and Control Simulation

It is usually convenient—and sometimes essential—that a coupled model control simulation is in a quasi-stationary state, that is, without a drift. Therefore, the spin-up of a new Max Planck Institute coupled climate model usually continues from the ocean state of the predecessor, in this case MPI-ESM1.1, and is continued while fine-tuning until a quasi-stationary state is reached. This procedure has been applied also for MPI-ESM1.1, which was spun up from a state obtained from a control simulation using MPI-ESM, which in turn was tuned and spun up from a state obtained from ECHAM5/MPIOM, meaning that the full spin-up since the ocean state was first initialized from an observed state is well in excess of 20,000 years. Although different atmospheric model components are used in each case, and so tuning and spin-up is needed, their preferred ocean states are far more similar to each other than they are to the observed ocean, thereby leading to faster equilibration. In our experience the physical state spin-up can thus be reduced, which is important because the final carbon cycle tuning and spin-up is mainly conducted after the physical state spin-up is completed.

We illustrate the spin-up procedure for MPI-ESM1.2-LR (Figure 10). Shown is a series of spin-up experiments conducted, sometimes with a bit of overlap, but otherwise in sequential order. Although the atmospheric physics and tuning parameters were largely unchanged between MPI-ESM1.1 and MPI-ESM1.2 (the reader is reminded that the former is a prerelease of the latter), there were changes in atmospheric composition and land properties which together resulted in a slight global warming. Around simulation year 2,500 of the combined spin-up displayed in Figure 10, the before-mentioned bug fix in the ocean mixing was introduced (section 4). Subsequently, it was decided to cool the model to about 13.6 ° C starting from run khw0062, because it was not possible to obtain acceptable amounts of Arctic sea ice at the warmer state around 14 ° C (run khw0061) using regular sea ice tuning parameters (Mauritsen et al., 2012). The global cooling was achieved by lowering the near-surface critical relative humidity (section 3.1). A further minor change was made to the ocean component around year 5,400 and paired with a small change to the critical relative humidity parameter in run hel17354-PL2CRS973, after which no more changes to the physical climate were conducted. The carbon cycle was subsequently fine tuned as described in sections 5.3 and 6.7.

8.2. Historical Experiments

Beginning with the development of ECHAM6.3 used in MPI-ESM1.1 and MPI-ESM1.2, it was decided to use the historical warming as an explicit development and tuning target. The instrumental-record warming can be matched in a model using forcing, climate sensitivity, and ocean heat uptake (Kiehl, 2007), and in the case of MPI-ESM1.1, we chose to tune the climate sensitivity down to improve the models' fidelity.

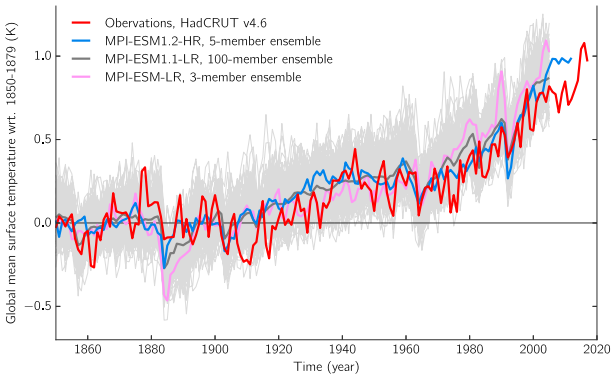

Figure 11. Historical simulations with MPI-ESM1.2-HR, compared to observations and the predecessor MPI-ESM1.1-LR grand ensemble of simulations, and the MPI-ESM-LR model used during CMIP5. Individual simulations from the grand ensemble are shown as thin gray lines and the ensemble means as thick lines. All lines are normalized to the 1850–1879 mean.

Therefore, the close match of MPI-ESM1.2 to observations should not be taken as a sign of quality but rather a signature that the development and tuning was successful. Indeed, as was the case for MPI-ESM1.1, also, MPI-ESM1.2 matches well the observed warming; at the time of writing, we have a five-member ensemble of MPI-ESM1.2-HR available (Figure 11). The better match is an improvement over the MPI-ESM-LR model used during CMIP5 which clearly warmed too fast. Here it should be remembered that observations only provide us with a single realization, and so there is no point in tuning the ensemble mean to match the observed as we can only require that a large ensemble of model realizations encompass the observed warming.

Upon closer inspection MPI-ESM1.2-HR is actually a closer match to the observed warming than the predecessor model from the 1970s up until around the year 2000 and the widely discussed hiatus period. The MPI-ESM1.2-HR five-member ensemble is little less than 0.1 K cooler than the MPI-ESM1.1 hundred-member ensemble on average during this period. This is no coincidence but rather an effect of the new simple-plume aerosol climatology (section 3.3, Figure 4). We see that the aerosol cooling differ by about 0.2 $W \cdot m^{-2} \cdot K^{-1}$ in the second half of the twentieth century. This is enough to cause around 0.05–0.10 K of cooling during the transient.

It deserves mention that the MPI-ESM, MPI-ESM1.1, and MPI-ESM1.2 climate models do not have a background volcanic stratospheric aerosol prescribed in their spin-up and control simulations. As a result, when starting a historical simulation from a state that is in equilibrium with zero volcanic stratospheric aerosols the model will start to cool as volcanoes are activated (Gregory et al., 2013). On average the volcanic forcing from 1850–2005 is around −0.32 W/m^2 (Myhre et al., 2013). This extra volcanic cooling in the models adds on top of the anthropogenic tropospheric aerosol cooling to slightly weaken historic warming, such that effectively the combined anthropogenic plus volcanic aerosol cooling in MPI-ESM1.2 is about −0.9 W/m^2 relative to the preindustrial control simulation; somewhat larger than the IPCC AR5 best estimate of −0.7 W/m^2 at present-day conditions relative to 1850 (Figure 4, Myhre et al., 2013) but well within the uncertainty range. Much of this cooling will however be realized at the surface within a few decades, and so the plotted surface temperature evolution relative to 1850–1879 (Figure 11) is only weakly affected by this choice.

8.3. Climate Sensitivity Experiments

The equilibrium climate sensitivity (ECS) of a model to external forcing is the long-term response to a doubling of atmospheric CO_2. Beginning with CMIP5, it became common to measure ECS by instantaneously quadrupling CO_2 in a coupled model run for 150 years (abrupt4xCO2) and then using linear regression between radiation imbalance and global mean surface temperature to estimate the long-term warming response (Andrews et al., 2012; Gregory et al., 2004). The result is then divided by 2 to be equivalent to a single doubling of atmospheric CO_2. The assumption underlaying this procedure is that a quadrupling of CO_2 is a small enough forcing such that the response is still linear, giving the advantage that the signal-to-noise ratio is better for a large forcing. Using this de facto standard method, the ECS of MPI-ESM1.2-LR is 2.83 K. Occasionally, the model's diagnosed 2-m temperature is used instead of the surface temperature, which here results in an ECS of 3.01 K.

Previous experience is, though, that ECS of models often depends on the strength of forcing so as to rise in warmer states (Heinemann et al., 2009; Meraner et al., 2013), and with time as the deep oceans equilibrate (Senior & Mitchell, 2000). To explore the impacts of such state- and time-dependent feedbacks, we apply 2, 4, 8, and 16 times preindustrial CO_2 as instantaneous forcing and run them for 1,000 years each (Figure 12). MPI-ESM1.2 exhibits considerable deviations from a simple linear relationship between the global means of the radiation imbalance and surface temperature. To avoid the influence of nonlinearity within the first few decades, we estimate the equilibrium climate sensitivities in each run using linear extrapolation from the years 100 to 1,000. The resulting ECS of MPI-ESM1.2-LR for doubled CO_2 is then 2.77 K, only slightly smaller than that estimated using the standard method.

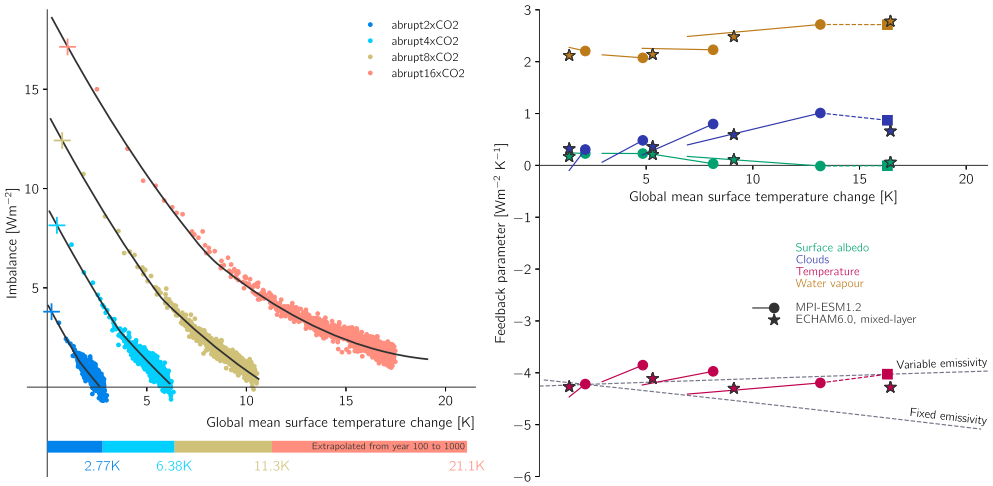

Figure 12. Idealized runs with abruptly increased atmospheric CO_2 concentrations relative to preindustrial levels. The left panel shows yearly means of radiation imbalance and global mean surface temperature. Each simulation is 1,000 years, and the equilibrium response (lower bar) is estimated by linear extrapolation of the years 100 to 1,000. There is a strong state dependency of equilibrium climate sensitivity in MPI-ESM1.2, as was also the case in earlier versions of the model (Heinemann et al., 2009; Meraner et al., 2013). Crosses show runs with fixed sea surface temperatures taken from the control simulation, but increased atmospheric CO_2 used to estimate effective radiative forcing (Hansen et al., 2005). Black lines show results from a two-layer model with coefficients chosen to roughly match the MPI-ESM1.2 behavior; see text for details. Right panel shows a feedback parameter analysis of the runs. Each colored line represents an experiment with the leftmost point being calculated over the years 1–20, the round dot for years 100–300, and the square for years 300–1,000 (only 16xCO_2). Stars represent values tabulated in Meraner et al. (2013) for the ECHAM6.0 model coupled to a mixed-layer ocean; here the temperature is midway between consecutive doublings. Dashed gray lines show two theoretical expressions for the temperature feedback.

With stronger forcing the ECS for each consecutive doubling rises dramatically to 3.6, 4.9, and nearly 10 K for the last doubling from 8 to 16 times preindustrial CO_2 (Figure 12). We may think of the slope of the curves in the left panel of Figure 12 as a feedback parameter, which consists of contributions from individual feedback mechanisms due to changes in surface albedo, clouds, water vapor, and temperature. To this end, we perform a partial radiative perturbation feedback analysis to explore the cause of this rise in ECS (Figure 12, right panel). The more elaborate partial radiative perturbation method is necessary due to the nonlinearities that occur under strong forcing, for instance, due to the rise of the tropopause (Meraner et al., 2013). In the updated model, the rise in ECS is due to strengthening water vapor and cloud feedbacks and is partially counteracted by a diminishing surface albedo feedback as snow and ice disappears at around 8–10 K above preindustrial.

8.4. Temperature Feedback State Dependence

Perhaps contrary to expectation, the temperature feedback does not strengthen to become more negative in a warmer climate in MPI-ESM1.2 (Figure 12). We can understand the nearly constant temperature feedback by considering the full energy imbalance (N) of the Earth:

$$N = \frac{S_o}{4}(1-\alpha) - e\sigma T_s^4, \tag{7}$$

where S_o is the solar constant, α planetary albedo, e planetary emissivity, $\sigma = 5.67 \cdot 10^{-8}\ \mathrm{W \cdot m^{-2} \cdot K^{-4}}$ and T_s a representative surface temperature. Assuming e and α are independent of T_s, we can calculate the feedback parameter:

$$\lambda_T = \frac{\partial N}{\partial T_s} = -4e\sigma T_s^3, \tag{8}$$

yielding the well-known third-power dependence of the Planck feedback parameter, which is plotted in Figure 12 as "Fixed emissivity," meaning e was determined for the 2xCO_2 case and then held fixed. Clearly, this curve does not reflect the behavior of the temperature feedback in MPI-ESM1.2. Returning to the energy balance equation (7), we see that at any stationary state ($N = 0$), there is a unique relationship between S_o, α, and e and the stationary state T_s. Assuming S_o and α are constant, then $e \propto T_s^{-4}$ at stationary states.

Table 5
Parameters and Forcing Used in a Simple Two-Layer Model (Equation (9)) to Match the MPI-ESM1.2 Behavior (Figure 12)

Parameter	Value	Unit
Feedback parameter (λ):	-1.65	$W{\cdot}m^{-2}{\cdot}K^{-1}$
Quadratic parameter (a):	0.04	$Wm^{-2}K^{-2}$
Deep ocean heat uptake coefficient (γ):	0.8	$W{\cdot}m^{-2}{\cdot}K^{-1}$
Heat uptake efficacy (ϵ):	1.2	
Upper ocean thickness:	50	m
Deep ocean thickness:	1,200	m
Effective radiative forcing:		
F_{2x}	4.10	Wm^{-2}
F_{4x}	8.85	Wm^{-2}
F_{8x}	13.50	Wm^{-2}
F_{16x}	18.60	Wm^{-2}

This makes sense if we remember that increasing CO_2 lowers planetary emissivity, which in turn leads to a warmer surface. Inserting this in equation (8) gives

$$\lambda_T \propto T_s^{-1},$$

which yields a slightly weakening temperature feedback with warming plotted in Figure 12 as the case of "Variable emissivity," exhibiting a striking similarity with the MPI-ESM1.2 model behavior. The argument does not depend on an infrared water vapor feedback, but it does depend on a possible shortwave feedback, that is, $d\alpha/dT_s \neq 0$: In the presence of a positive shortwave feedback e needs not decrease as fast with T_s as is the case with zero shortwave feedback, and so the temperature feedback parameter λ_T may be closer to constant. If the temperature feedback of MPI-ESM1.2 behaved more like the case of fixed emissivity, then the climate sensitivity of the model would not have risen nearly as much as it does. The decreasing planetary emissivity at warmer stationary states is therefore an important part of the explanation for why MPI-ESM1.2 exhibits a rise in ECS with warming. The simplicity of the argument provides hope that it could be applicable to other models and to the real Earth.

8.5. A Simple Representation of Global Mean Temperature in MPI-ESM1.2

It is possible to mimic the behavior of the comprehensive climate model using a fairly simple two-layer model representing time- and state-dependent feedback. Simple model representations of the climate system are instrumental in theoretical work, and also see widespread applications, for example, in economy-climate models. We modify the model of Held et al. (2010) and Winton et al. (2010), which represents time-dependent feedback, by adding a quadratic term to represent state dependence as suggested by Bloch-Johnson et al. (2015). In this model the surface temperature T deviation from a basic state is representative of a small heat reservoir C, including the atmosphere, land, and ocean mixed layer, and it is coupled to a large heat reservoir C_d with temperature T_d representative of the deep ocean:

$$\begin{aligned} C\frac{dT}{dt} &= F + \lambda T + aT^2 - \epsilon\gamma(T - T_d) \\ C_d\frac{dT_d}{dt} &= \gamma(T - T_d), \end{aligned} \tag{9}$$

where F is an external forcing, λ is the feedback parameter, a is a quadratic term parameter, ϵ is the ocean heat uptake efficacy, and γ is the deep ocean heat uptake coefficient. The imbalance of the system can be calculated as the rate of change of heat in the system: $N = C\frac{dT}{dt} + C_d\frac{dT_d}{dt} = F + \lambda T + aT^2 - (\epsilon - 1)(T - T_d)$. The model exhibits time-dependent feedback whenever $\epsilon \neq 1$ and state-dependent feedback when $a \neq 0$.

We provide a set of parameters in Table 5 that were determined manually. There is no unique way to determine these parameters, and here we emphasized first matching the equilibria up to $8xCO_2$ (giving λ and a) and subsequently tried to match the temporal behavior of imbalance and temperature on decadal to millennial timescales (providing γ, ϵ, C, and C_d). On shorter timescales, it is difficult to replicate the behavior

of MPI-ESM1.2 perhaps because land surface temperatures can respond faster than the upper ocean. A feature of the quadratic model is that if forced hard eventually it enters a thermal runaway where $dN/dT_s > 0$. This happens if forced beyond $F_b = \frac{\lambda^2}{4a} \approx 17.0$ W/m^2, which is less than the estimated forcing from 16xCO_2 (Table 5). Thus, the modified two-layer model should be useful for emulating the MPI-ESM1.2 model's transient behavior from preindustrial to and slightly beyond 8xCO_2, that is, under typical centennial forcing scenarios.

9. Summary

We have presented the latest version of the Max Planck Institute climate models, MPI-ESM1.2, which is also planned to be the last coupled model based on the ECHAM and MPIOM model components as we move to the ICON model framework. The MPI-ESM1.2 model release thus aims at serving the institutes scientific modeling needs in a few years to come, as researchers gradually transition to using the new modeling system, and many of the improvements described here are transferred into and further refined in ICON-ESM. Additionally, MPI-ESM1.2 is going to be used to participate in the currently ongoing CMIP6 activities and other model intercomparison projects.

The model has seen updates and improvements in all its components including a new tropospheric aerosol, a new radiative transfer, a multilayer soil hydrology, updated fire and soil representations, prognostic ocean cyanobacteria that fixates nitrogen, and improved settling of detritus, to mention just a few. Emphasis during the development was also given to the computational performance, versatility, and general user friendliness of the model; the collective sum of which has in many ways revolutionized how scientists work with the model in that the technical improvements allow them to do things that were previously not possible. Furthermore, there has been a focus on removing coding errors, in particular regarding energy conservation in the atmosphere physics but also a number of other coding errors were removed most notably pertaining to ocean mixing, sea ice melt ponds, and the cloud fraction scheme. Removing such coding errors from the climate model rank higher than other improvements, because scientifically it is frequently essential that the code does what it was intended to do.

The climate model was tuned, foremost to avoid drifts in the mean state variables but also to alleviate a series of model biases thought to be important by the development team. Now compared to an earlier description of this process, the instrumental record warming is considered an explicit target of the model development. To achieve the goal of matching observed twentieth century warming the climate sensitivity of an intermediate model version was reduced from around 7 to 3 K. We verified using a 100-member ensemble with MPI-ESM1.1 that after this change, the model is statistically consistent with observed warming, though it on average warms around 0.1 K more than observed. Further improvement was subsequently achieved in MPI-ESM1.2 after the introduction of an aerosol indirect effect. Ongoing research is revisiting high climate sensitivity versions of the model paired with stronger aerosol cooling to explore the bounds of our understanding.

A particular concern on the development team was whether the model's previously identified rapid rise of ECS to CO_2 with warming was possibly an artifact of the coding errors that had been discovered during the development. This concern was not without merit as the strength of the energy leakage was dependent on temperature and so could constitute a false climate change feedback. To this end, a set of 1,000-year simulations were conducted and found to be consistent with previous results that the rise in sensitivity is due to water vapor and clouds. Furthermore, the surprising stability of the temperature feedback is explained to be an effect of planetary emissivity decreasing in a warming climate counteracting the pure T^3 Planck feedback dependency. The global response and nonlinearity of the model can be fairly accurately represented by a simple quadratic two-layer model.

10. Closing Remarks on Three Decades of Model Development

The MPI-ESM1.2 model in many ways is the culmination of more than three decades of development by hundreds of scientists. But foremost, it is the fruition of the careers of two especially talented and creative individuals, Erich Roeckner and the late Ernst Maier-Reimer. Erich Roeckner led the development of the ECHAM atmospheric model from its conception as an outgrowth of the European Centre for Medium-Range Weather Forecasts model in the late 1980s. His passion for model development and his comprehensive understanding of the climate system were evident in the quality of the simulations by every release of what

has long been one of the world's most successful and influential models of the atmospheric general circulation and climate. Even after his retirement in 2009, he made very important contributions to the finalization of ECHAM6.3 upon which MPI-ESM1.2 is based. Ernst Maier-Reimer pioneered the development of modern dynamical ocean models already in the 1980s, one of which is now MPIOM, he was an early pioneer, and about as singular as one can become in our field, in the development of three-dimensional ocean carbon models, simply doing what others would only talk about (Hasselmann, 2013), models that are now known as HAMOCC. Common to both Erich and Ernst is that their developments were guided by a deep understanding of Earth's changing climate, an understanding that enabled them to prioritize among the infinity of processes that comprise the Earth system and to create models they could use to challenge and advance their understanding; something which has influenced us all.

Appendix A: Energy Conservation in the Atmosphere Physics Parameterizations

The long-term global heat budget of the atmosphere is given by the difference between the top-of-atmosphere (*toa*) net radiation

$$R^{toa} = S_d^{toa} - S_u^{toa} - L_u^{toa}, \tag{A1}$$

where S_d, S_u, and L_u are the respective downwelling and upwelling shortwave and longwave radiative fluxes, and the net surface heat flux

$$R^{sfc} = S_d^{sfc} - S_u^{sfc} + L_d^{sfc} - L_u^{sfc} - SH - \left(L_v E_v + L_s E_s + L_f Sn\right), \tag{A2}$$

where the radiative fluxes are defined now at the surface (*sfc*), and the sensible heat flux (*SH*), vaporization (E_v), sublimation (E_s), and snowfall at the surface (*Sn*) act to cool the surface relative to a state of liquid. In ECHAM $L_v = 2.5008 \cdot 10^6$ J/kg is the latent heat of vaporization, $L_s = 2.8345 \cdot 10^6$ J/kg is the latent heat of sublimation and $L_f = L_s - L_v$ is the latent heat of fusion. The atmospheric heat budget (*atm*) is then given by the difference between equations (A1) and (A2),

$$\begin{aligned} H^{atm} = {} & S_d^{toa} - S_u^{toa} - \left(S_d^{sfc} - S_u^{sfc}\right) - L_u^{toa} - \left(L_d^{sfc} - L_u^{sfc}\right) \\ & + \quad SH + \left(L_v E_v + L_s E_s + L_f Sn\right), \end{aligned} \tag{A3}$$

or by combining the radiative fluxes into net shortwave and net longwave radiative heating of the atmosphere,

$$H^{atm} = S_n^{atm} + L_n^{atm} + SH + \left(L_v E_v + L_s E_s + L_f Sn\right). \tag{A4}$$

Atmospheric heating must equal the integrated column heating by physical processes, in ECHAM consisting of radiation (*rad*), turbulent transfer (*turb*), and the combined cumulus convection and cloud processes (*cc*):

$$H^{atm} = \int_0^{p=p_s} c_p(q) \left[\left(\frac{\partial T}{\partial t}\right)_{rad} + \left(\frac{\partial T}{\partial t}\right)_{turb} + \left(\frac{\partial T}{\partial t}\right)_{cc} \right] \frac{dp}{g}, \tag{A5}$$

where T is temperature, g is gravity, p_s is the surface pressure, and the specific heat capacity at constant pressure (c_p) is a function of specific humidity (q) such that $c_p(q) = c_{pd} + (c_{pv} - c_{pd})q$, whereby the specific heat of water vapor (c_{pv}) is larger than the specific heat of dry air (c_{pd}). It is further convenient to rewrite the heating terms by moist processes:

$$L_v E_v + L_s E_s + L_f Sn = L_v Ra + L_s Sn,$$

by invoking conservation of water. Here *Ra* denotes rain reaching the surface by convective and stratiform cloud processes, such that the fluxes can be assigned to heating rates by physical processes:

$$\int_0^{p=p_s} c_p(q) \left(\frac{\partial T}{\partial t}\right)_{rad} \frac{\mathrm{d}p}{g} = S_n^{atm} + L_n^{atm}, \tag{A6}$$

$$\int_0^{p=p_s} c_p(q) \left(\frac{\partial T}{\partial t}\right)_{turb} \frac{\mathrm{d}p}{g} = SH, \tag{A7}$$

$$\int_0^{p=p_s} c_p(q)\left(\frac{\partial T}{\partial t}\right)_{cc}\frac{\mathrm{d}p}{g} = L_v Ra + L_s Sn; \tag{A8}$$

relationships that must hold if the model physical process parameterizations are energy conserving.

Further, in the ECHAM model's leapfrog time integration scheme, the specific heat capacity is defined as a function of the specific humidity at the previous time step, q^{n-1}, and the temperature tendencies are discretized by centered differences in time such that $\partial T/\partial t \rightarrow (\tilde{T}-T^{n-1})/2\Delta t$, where Δt is the time step length and $\tilde{T}$ is the updated temperature calculated by the respective individual physical process parameterizations: radiation, turbulence, cumulus convection, and stratiform cloud.

Error 1 is caused by an inappropriate discretization of the mass flux in the convection scheme (Tiedtke, 1989) in the term contributing the convective heating, which is calculated on a vertically staggered grid. In earlier versions of the model, the discretization was done as follows:

$$\frac{\partial}{\partial p}\left(L_{v,s}M_u l_u\right)_k \rightarrow \left(L_{v,s}\right)_k \frac{\left(M_u l_u\right)_{k+1/2} - \left(M_u l_u\right)_{k-1/2}}{p_{k+1/2} - p_{k-1/2}}, \tag{A9}$$

where M_u is the mass flux in the convective updraft, l_u is the condensate of liquid and ice in the updraft, $L_{v,s}$ is the relevant latent heat. Here k denote full model levels where temperature and latent heat is defined, whereas convective mass flux, condensate, and pressure is defined at half levels ($k + 1/2$ and $k - 1/2$). Level zero is at the top of the atmosphere, and k increases toward the surface.
The problem with the former discretization (A9) arises at the melting point where $T_k \geq 0°\mathrm{C}$ with $L = L_v$ and $T_{k-1} < 0°\mathrm{C}$ with $L = L_s$ and so the latent heat cannot be assumed to be constant as done previously. In this case the vertical integral of the left hand side of (A8) yields a spurious cooling in the column. The problem can be easily alleviated by defining the latent heat at the same half levels where the mass flux is defined:

$$\frac{\partial}{\partial p}\left(L_{v,s}M_u l_u\right)_k \rightarrow \frac{\left(L_{v,s}M_u l_u\right)_{k+1/2} - \left(L_{v,s}M_u l_u\right)_{k-1/2}}{p_{k+1/2} - p_{k-1/2}}. \tag{A10}$$

Error 2 is caused by an inconsistent treatment of the convective detrainment of cloud condensate within the stratiform cloud scheme. In the convection scheme the phase of the cloud condensate is a simple function of temperature being liquid phase at temperatures $T \geq 0°\mathrm{C}$ above the melting point and ice phase at temperatures $T < 0°\mathrm{C}$ below the melting point. The convective detrainment of cloud condensate was passed as a single variable to the stratiform cloud scheme, where it is treated as a source term. However, here the phase of the cloud condensate depends not only on temperature but also on the availability of ice crystals (Lohmann & Roeckner, 1996). For example, supercooled cloud droplets are allowed to exist in the temperature range $-38°\mathrm{C} \leq T < 0°\mathrm{C}$ if the coexisting concentration of ice crystals falls below a tunable threshold parameter value. As done in ECHAM6.1, applying this criterion not only in the cloud generation by condensation/deposition but also to the detrained convective condensate results in a spurious heat source. Thereby, cloud ice generated by the convection scheme at temperatures $T < 0°\mathrm{C}$ could inadvertently be converted into cloud liquid without taking into account the heat required to melt the ice; a heat that was previously released in the convection scheme. In the updated model version ECHAM6.3, this problem is avoided by defining the phases of the convective detrainment within the convection scheme and passing them both to the stratiform cloud scheme.

Error 3 is caused by an inconsistent use of the specific heat capacity in the turbulent transfer scheme. Here the vertical heat transfer is expressed in terms of the dry static energy $s = c_p T + gz$, where $c_p T$ is the specific enthalpy and gz the geopotential energy. Thus, equation (A7) can be expressed as

$$\int_0^{p=p_s} \left(\frac{\partial s}{\partial t}\right)_{turb}\frac{\mathrm{d}p}{g} = SH \tag{A11}$$

and, in discretized form

$$\sum_0^{k=nk}\left(\frac{\tilde{s}-s^{n-1}}{2\Delta t}\right)_k \frac{\Delta p_k}{g} = SH, \tag{A12}$$

where $\tilde{s}$ is the dry static energy as updated by the turbulence parameterization, s^{n-1} is the previous time step dry static energy, nk is the number of vertical levels, and Δp_k is the pressure thickness of layer k.

When (A12) is checked for consistency, the global mean error is much larger at about 8 W/m^2 than the 0.3 W/m^2 reported in Table 2. This large error is caused by different definitions of $c_p(q)$ at the left- and right-hand sides of the equation. While $c_p(q) = c_p(q_{sfc})$ is used in solving the diffusion equation, $c_p(q) = c_p(q_{nk})$ is used for calculating the sensible heat flux, SH. Here q_{sfc} is the surface specific humidity, which is saturated over water and ice, but otherwise depends on soil wetness, vegetation, and lowest model level humidity over land, and q_{nk} is the specific humidity at the lowest model level, typically around 30 m above the surface. Since $q_{sfc} > q_{nk}$ in by far the most cases, the sensible heat flux is systematically smaller than the column heating.

The large error was nearly cancelled by another error. This arose when converting the dry static energy tendency $\partial s/\partial t$ to temperature tendency $\partial T/\partial t$ required for the model integration, which was erroneously formulated as follows:

$$\frac{\tilde{s} - s^{n-1}}{2\Delta t} \rightarrow \frac{\tilde{T} - T^{n-1}}{2\Delta t} = \frac{\frac{\tilde{s}}{c_p(\tilde{q})} - \frac{s^{n-1}}{c_p(q^{n-1})}}{2\Delta t}. \tag{A13}$$

By multiplying with $c_p(q^{n-1})$, vertically integrating and noting that for turbulent processes in most cases $\tilde{q} > q^{n-1}$ because of surface evaporation, it is easy to realize that this leads to an energy inconsistency:

$$\sum_{0}^{k=nk} \left(\frac{\frac{c_p(q^{n-1})}{c_p(\tilde{q})}\tilde{s} - s^{n-1}}{2\Delta t} \right)_k \frac{\Delta p_k}{g} < \sum_{0}^{k=nk} \left(\frac{\tilde{s} - s^{n-1}}{2\Delta t} \right)_k \frac{\Delta p_k}{g}. \tag{A14}$$

The problem can be avoided by changing the definition of $\tilde{T}$ in (A13) such that

$$\frac{\tilde{T} - T^{n-1}}{2\Delta t} = \frac{\tilde{s} - s^{n-1}}{c_p(q^{n-1})2\Delta t} \tag{A15}$$

as is done also in other process parameterizations in ECHAM. Together with the corrected surface boundary condition, the turbulence parameterization errors could be completely eliminated at every grid point and every time step in ECHAM6.3.

Appendix B: Evaluation of Land-Atmosphere Simulations

Here we present results from simulations with prescribed SSTs. Two experiments were performed using observed SSTs 1976–2008 from the Atmospheric Model Intercomparison Project. For the first simulation a low resolution (MPI-ESM1.2-LR) version of ECHAM6.3, T63L47, was run and in the second one a high-resolution (MPI-ESM1.2-HR) version, T127L95. A detailed evaluation and comparison of the coupled models MPI-ESM1.2-LR and MPI-ESM1.2-HR is conducted by Müller et al. (2018) and so is omitted here.

B1. The Climatological Mean State

An overview of a set of global, annual mean climate variables simulated by ECHAM6.1 is compared to the same fields as simulated by ECHAM6.3 is shown in Table B1. The effect of increased resolution configuration is similar in both model versions, though generally less pronounced in the new model. We find that overall, there is much less sensitivity of the ECHAM6.3 models climate to resolution and time step length compared to earlier versions. We speculate that this is a consequence of having removed the coding errors discussed in section 3.

By design, all model versions reproduce the global mean top-of-atmosphere radiation fluxes within the uncertainty limits of the satellite data since the are tuned (Mauritsen et al., 2012). Cloudiness in ECHAM6.3 is shifted slightly from high-level to low-level clouds (not shown), evident in the reduced ice water path and higher outgoing longwave radiation on the one hand and increased liquid water path and larger downwelling longwave radiation at the surface on the other hand. Nevertheless, the planetary albedo is not enhanced because clouds in the shallow cumulus regime are made less bright by the change of the reduction factor f_{hom} from 0.77 in ECHAM6.1 to 0.4 in ECHAM6.3 (section 3.6). A difference between the two model versions

Table B1
Global Annual Mean Climate Variables Derived From Experiments Using Observed Sea Surface Temperatures and Sea Ice Concentrations in Comparison With Observational Estimates (Stephens et al., 2012)

Variable	6.1-LR	6.1-HR	6.3-LR	6.3-HR	Observed
Total cloud cover (%)	63.3	61.4	62.6	62.1	
Ice water path (g/m^2)	31.1	32.3	29.5	30.1	
Liquid water path (g/m^2)	59.6	56.0	62.9	62.6	
Water vapor path (kg/m^2)	25.4	25.0	25.6	25.2	
Top-of-atmosphere budget:					
Incident shortwave (W/m^2)	340.4	340.4	340.4	340.4	340.2 ± 0.1
Reflected shortwave (W/m^2)	103.0	98.4	101.6	98.7	100 ± 2
Planetary albedo (%)	30.3	28.9	29.8	29.0	29.4 ± 2
Absorbed shortwave (W/m^2)	237.4	242.0	238.9	241.7	240.2 ± 2
Outgoing longwave (W/m^2)	236.9	240.5	238.9	241.5	239.7 ± 3.3
Net (W/m^2)	0.50	1.46	−0.00	0.26	0.6 ± 0.4
Surface budget:					
Incident shortwave (W/m^2)	186.0	192.1	185.6	189.6	188 ± 6
Reflected shortwave (W/m^2)	24.8	25.6	24.3	24.8	23 ± 3
Absorbed shortwave (W/m^2)	161.2	166.5	161.3	164.8	165 ± 6
Downwelling longwave (W/m^2)	344.2	341.3	346.8	345.5	345.6 ± 9
Upwelling longwave (W/m^2)	399.0	398.9	399.5	399.4	398 ± 5
Net longwave (W/m^2)	−54.8	−57.5	−52.7	−53.9	−52.4
Sensible upward heat flux (W/m^2)	19.6	19.6	24.4	24.0	24 ± 7
Latent upward heat flux (W/m^2)	86.0	87.4	84.0	86.1	88 ± 10
Bowen ratio (sensible/latent)	0.23	0.22	0.29	0.28	0.27
Snow melt (W/m^2)	0.78	0.78	0.76	0.76	
Net (W/m^2)	0.01	1.20	−0.51	0.00	0.6 ± 17
Atmosphere budget:					
Absorbed shortwave (W/m^2)	76.2	75.5	77.5	76.9	75 ± 10
Emitted longwave (W/m^2)	182.2	183.0	186.2	187.6	188 ± 12.5
Net radiation (W/m^2)	−105.9	−107.5	−108.7	−110.7	−113
Net atmosphere (W/m^2)	0.50	0.25	0.50	0.26	0
Precipitation:					
Global (mm/day)	2.98	3.03	2.91	2.99	2.61
Oceans (mm/day)	3.49	3.55	3.36	3.43	2.80
Land (mm/day)	2.00	2.04	2.05	2.17	2.16
Large scale (mm/day)	1.06	1.16	1.03	1.12	
Convective (mm/day)	1.92	1.87	1.87	1.87	

Note. Shown are the respective model results for the 30-year period 1979–2008. The observational climatologies are based on years 2000–2010. The precipitation climatologies are from GPCP 1979–2008 (Huffman et al., 2009) for the land surface, from HOAPS 1988–2005 (Andersen et al., 2010) for the ice-free ocean, and from GPCP + HOAPS combined for the global mean. For calculating the means, the land/sea mask at T127 (HR) resolution was utilized, corresponding to a grid spacing of about 100 km. Net fluxes are positive in the downward direction at the top of atmosphere and surface, and into the atmosphere for the atmospheric budget.

is the increase in the sensible heat flux, resulting in an increased Bowen ratio by more than 25%. By considering the atmospheric energy budget, which is equilibrated on yearly and longer timescales, this explains why the global precipitation fluxes are systematically smaller in ECHAM6.3 despite the enhanced radiative cooling of the atmosphere of about 3 W/m^2.

Total cloud cover is reduced, ice water path larger, and liquid water path smaller; the water vapor path is reduced, and both the top-of-atmosphere shortwave and longwave fluxes are enhanced when comparing HR to LR. In the HR model resolutions the surface downward shortwave radiation is larger in both versions:

Table B2
Climate Variables Contributing to the Performance Index

Variables	Domain	Data source	Period
Mean sea level pressure	Global	ERA-Interim	1979–2008
Surface eastward wind stress	Ocean	ERA-Interim	1979–2008
Surface northward wind stress	Ocean	ERA-Interim	1979–2008
Temperature (850 hPa)	Global	ERA-Interim	1979–2008
Stationary waves (geopotential at 500 hPa)	Global	ERA-Interim	1979–2008
Zonal mean temperature (up to 10 hPa)	Global	ERA-Interim	1979–2008
Zonal mean zonal wind (up to 10 hPa)	Global	ERA-Interim	1979–2008
Surface air temperature	Land	HadCRU	1979–2008
Precipitation	Land	GPCP	1979–2008
Precipitation	Ocean	HOAPS	1988–2005
Column water vapor content	Global	NVAP	1988–1999
Outgoing longwave radiation	Global	CERES	2001–2013
Reflected shortwave radiation	Global	CERES	2001-2013

Note. Data sources refer to Dee et al. (2011) for ERA-Interim, Jones et al. (2012) for HadCRU, Huffman et al. (2009) for GPCP, Andersson et al. (2010) for HOAPS, Randel et al. (1996) for NVAP, and Loeb et al. (2009) for CERES.

by 6 W/m^2 in ECHAM6.1 and 4.0 W/m^2 in ECHAM6.3. This energy surplus is compensated by reduced longwave downwelling radiation (less low-level clouds) and enhanced latent heat fluxes, resulting in more precipitation, whereas the sensible heat fluxes are hardly affected. Although the differences between model versions and model resolutions appear systematic, it is impossible to decide on a better or worse simulation because all model results fall into the range of the rather large observational uncertainties at the surface.

Compared to the Hamburg Ocean Atmosphere Parameters and fluxes from Satellite data (HOAPS) climatology (Andersson et al., 2010), all model versions overestimate precipitation by roughly 20%, whereas the global land precipitation is generally well simulated, in comparison to GPCP (Huffman et al., 2009). There is a long-running discrepancy between measurements of precipitation and inference from atmospheric and surface energy budgets, with the latter suggesting more precipitation than the former. A global mean precipitation of 2.61 mm/day corresponds to a latent heat flux of about 75 W/m^2, which is outside the range of 88 $\pm$ 10 W/m^2 proposed by Stephens et al. (2012). The simulated latent heat fluxes, on the other hand, are close to the central value, so that the global precipitation might not necessarily be overestimated by the models. In our experience, it is difficult to change global mean precipitation in ECHAM, and the rates around 3 mm/day are a robust feature that is hardly dependent on tuning parameters or parameterizations.

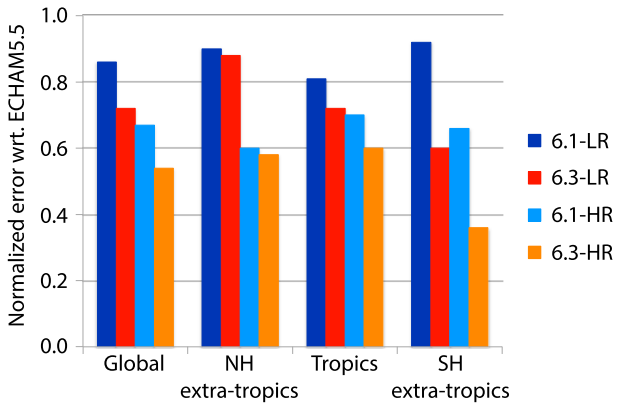

Figure B1. Modified Reichler and Kim (2008) standardized error based on analysis fields summarized in Table B2. Lower is better and a value of one corresponds to the performance index of ECHAM5.5 run at T63L31 resolution. Here the extratropical Northern Hemisphere is 30° N to 90° N, tropics 30° S to 30° N, and the Southern Hemisphere extratropics 30° S to 90° S.

Following Reichler and Kim (2008) model biases in the annual mean climatology were quantified using a performance index derived from a set of selected climate variables (see Table B2). Verifying data used for the comparison are surface observations (temperature and precipitation over land), satellite data (precipitation over ocean, water vapor, and radiative fluxes) and reanalysis products. To construct the index, the squared difference between simulation and observation was normalized by the interannual variance of the observed field and then averaged over the whole globe and over three subdomains (Northern Hemisphere, Southern Hemisphere, and tropics). These normalized errors are scaled by those of a reference model, ECHAM5 (Roeckner et al., 2006), version 5.5 running at a resolution of T63L31. This normalization ensures that different climate variables receive similar weights when combining their errors into a performance index defined by the average over all climate variables shown in Table B2. A performance index of unity denotes the average per-

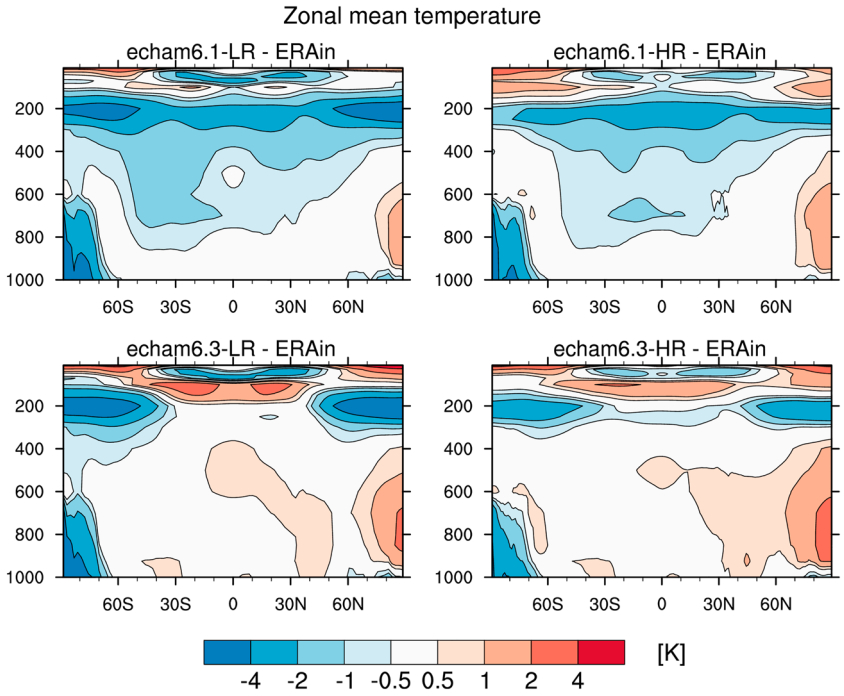

Figure B2. Zonal mean annual average temperature biases in ECHAM6.1 and ECHAM6.3 with LR and HR resolutions. Shown is the bias relative to the ERA-Interim reanalysis.

formance of ECHAM5.5, and values less than unity denote improved performance relative to ECHAM5.5, as found for all model versions shown in Figure B1. Here it is notable that the new model performs better for a given resolution in the tropics and Southern Hemisphere extratropics. Increasing resolution leads to smaller biases, but interestingly ECHAM6.3-LR is in many respects nearly on par with ECHAM6.1-HR showing that the improvements to the physics and tuning of the model has had an appreciable impact on reducing these model biases.

The extratropical improvements are basically caused by the changed patterns of the atmospheric temperature errors (Figure B2) and the associated changes in the zonal wind errors (Figure B3). In ECHAM6.1 the upper tropospheric cold bias of more than 2 K, centered around 200 hPa, is rather uniformly distributed across all latitudes. In ECHAM6.3, on the other hand, the cold bias is almost eliminated at lower latitudes, whereas it persists at higher latitudes. The general tropospheric warming in ECHAM6.3 relative to ECHAM6.1 is caused by the elimination of the budget errors evident in ECHAM6.1 (see Figure 5d) displaying a spurious heat deficit predominantly over the tropical continents due to the inappropriate discretization of the convective heat fluxes (see section 3.4, Appendix A and Figure 5a).

The upper-tropospheric warming in ECHAM6.3 together with a slightly larger polar cooling generates stronger upper-level westerlies in both hemispheres, extending down to the surface (Figure B3). Thus, the characteristic easterly wind bias throughout the troposphere around 60° S, evident in both versions of ECHAM6.1 and primarily driven by an equatorward shift of the westerlies, is replaced by a somewhat smaller westerly wind bias in ECHAM6.3-LR. In ECHAM6.3-HR the westerly wind bias is largely confined to the upper troposphere and lower stratosphere. In the Northern Hemisphere, on the other hand, all model versions show similar error patterns but, in contrast to the Southern Hemisphere, the errors in ECHAM6.3 are larger. In the tropics, the upper-tropospheric westerly wind error is reduced in the new model, particularly in the HR version.

The biases in mean sea level pressure (Figure B4) are largely consistent with those of the near-surface zonal wind biases (Figure B3). For example, the easterly bias near 60° S in ECHAM6.1 can be linked to a dipole pattern over the South Pacific, characterized by a high-pressure bias close to Antarctica and a low-pressure bias further north. This pattern evolves independent of resolution but is absent in the new model, which, on the other hand, simulates a larger high-pressure bias over much of the North Pacific. This error is much

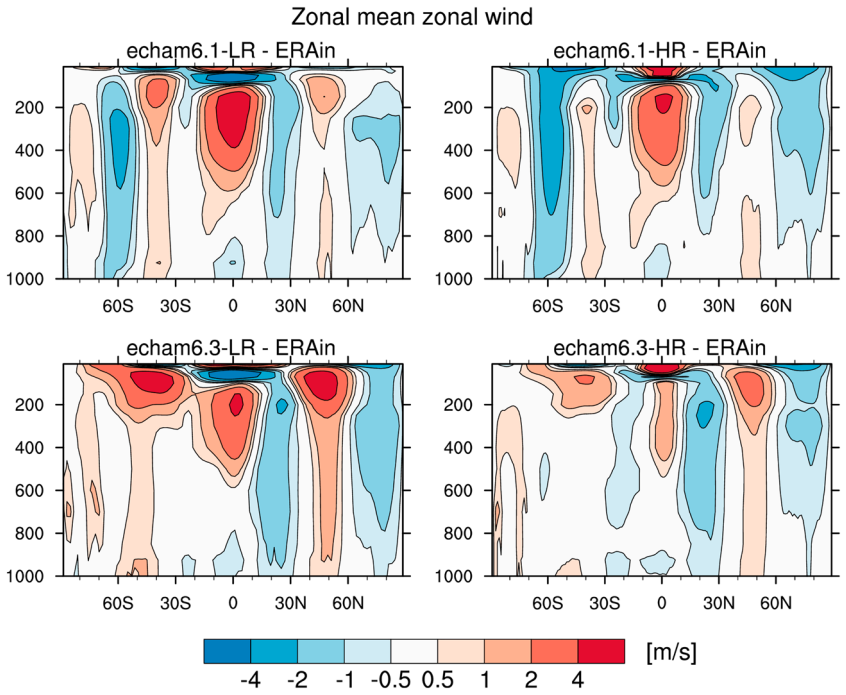

Figure B3. Zonal mean zonal wind biases, otherwise as Figure B2.

smaller in ECHAM6.1-LR and practically absent in ECHAM6.1-HR. A recurrent feature is the dipole pattern of too low pressure over parts of Europe and too high pressure over northern Africa, the latter being more evident in the LR versions.

Keeping in mind the large uncertainties in the satellite derived estimates of precipitation, all model versions tend to overestimate the precipitation over the oceans (Figure B5). One of the exceptions is the tropical East Pacific where the position of the Intertropical Convergence Zone is too far north in the models and

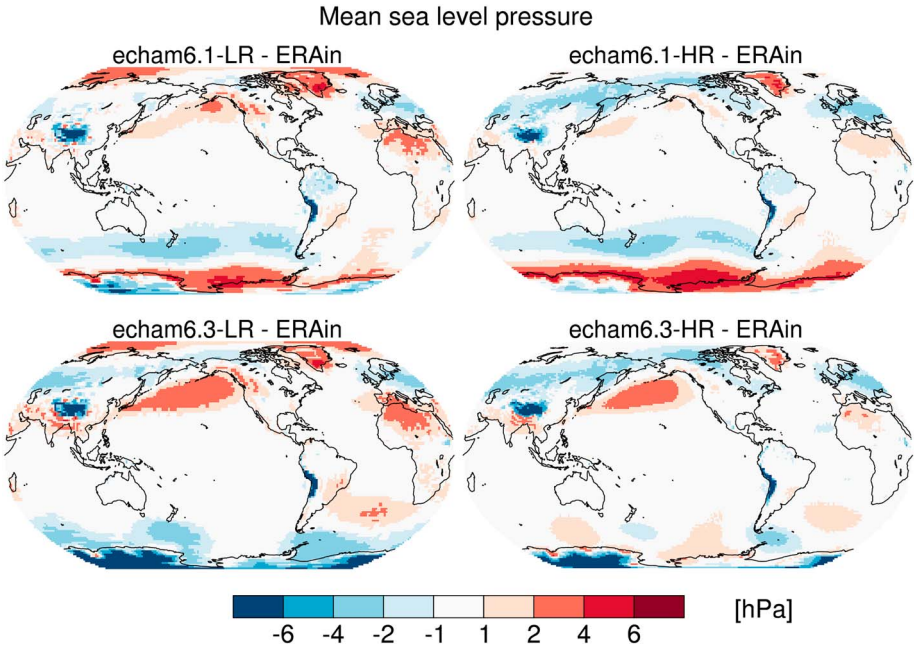

Figure B4. Maps of annually averaged mean sea level pressure biases in ECHAM6.1 and ECHAM6.3 in the LR and HR resolutions, relative to the ERA-Interim reanalysis.

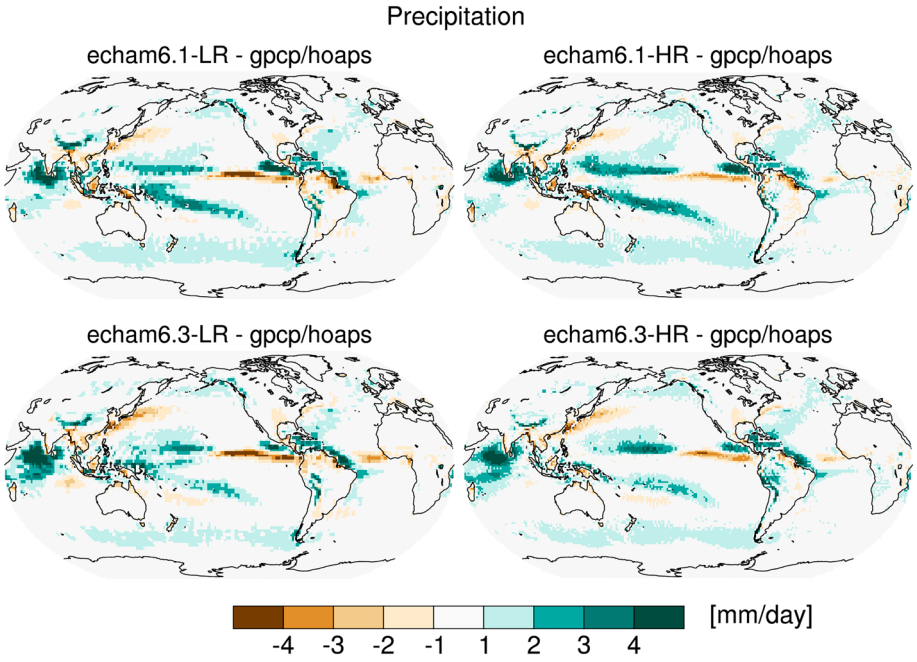

Figure B5. Biases of annually averaged precipitation with respect to HOAPS (ocean) and GPCP (land).

another one the low bias along the Kuroshio Current near China and Japan. In all simulations, both the patterns and the biases are hardly distinguishable among each other, and the differences between the model versions are generally much smaller than the respective biases and observational uncertainty. Over land, where the observations are more reliable, the biases are much smaller, but again similar in all of the four model versions. In the tropics, a low bias over land goes often along with a high bias in the surrounding ocean areas (e.g., northeastern Brazil, Indonesia, and Borneo). As analyzed by Siongco et al. (2014), all versions show the typical west Atlantic precipitation bias typical of low-resolution models.

Related difference patterns with respect to CERES are also found for the top-of-atmosphere reflected shortwave radiation, (Figure B6). The most pronounced feature is the low reflectance in the tropical West Pacific. This is attributed to an abundance of high-level clouds and lack of middle- and low-level clouds, relative to observed amounts seen in the CALIPSO-GOCCP data set (not shown), which causes outgoing shortwave radiation to be underestimated. In contrast, reflectance over parts of the subtropical and midlatitude oceans is generally too high, which is due to the commonly referred to "too few, too bright" low-level cloud bias (Nam et al., 2012). The low-level cloud cover in ECHAM6.3 has increased over that in ECHAM6.1; however, it remains underestimated when compared to observations. Despite this underestimation in low-level cloud fraction, the shortwave reflectance is overestimated. This is probably caused by the overestimation of high-level cloud fraction.

The impact of reduced optical thickness in the shallow cumulus regime can be seen by comparing the biases in ECHAM6.1-LR, where a reduction factor of 0.77 is applied for the optical thickness and ECHAM6.3-LR where a reduction factor of 0.4 is applied (see section 3.6), resulting in less reflectance in most areas, despite low clouds being more abundant in ECHAM6.3. At higher resolution the reduced reflectance due to the lower optical thickness in the shallow convection regime is almost compensated by larger reflectance due to enhanced low-level cloudiness. In fact, both ECHAM6.1-HR and ECHAM6.3-HR reflect slightly too much sunlight over parts of the subtropical and midlatitude oceans, albeit the bias is generally smaller than in the LR models. In these regions where high-level clouds are rare, the outgoing longwave radiation (not shown) is largely SST driven so that the errors are small and generally within the uncertainty limits of the CERES data. In ECHAM6.1 the reflectance by marine stratocumulus over eastern boundary currents is underestimated, especially to the west of South America where the reflectance is up to 30 W/m^2 too low in both model versions. In ECHAM6.3 the low biased cloud fractions are smaller due to the enhancement of the marine stratocumulus (see also section 3.1 and Figure 2).

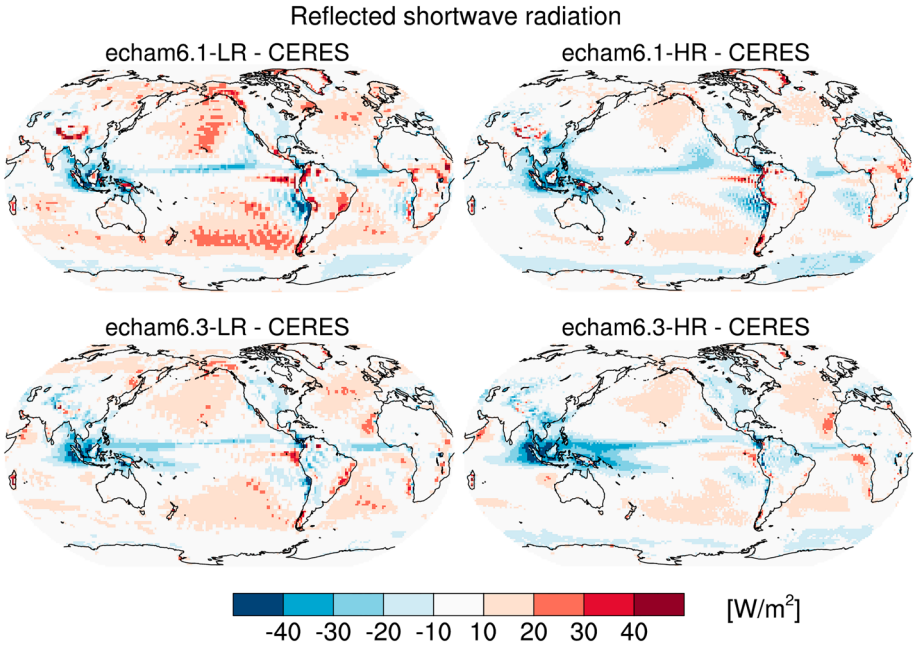

Figure B6. Model biases of annually averaged reflected shortwave radiation at the top of the atmosphere with respect to CERES.

Acknowledgments
This work was supported by the Max Planck Society (MPG) and the German Science Foundation (DFG). Computational resources were made available by Deutsches Klimarechenzentrum (DKRZ) through support from BMBF and by the Swiss National Supercomputing Centre (CSCS). T. M. received funding from the European Research Council (ERC) Consolidator Grant 770765. T. C. has received funding by CliSAP, an excellence cluster at the University of Hamburg supported by the German Science Foundation. D. S. G. is funded by the IMBALANCE-P project of the European Research Council (ERC-2013-SyG-610028). J. P. was supported by the German Research Foundations Emmy Noether program. The primary data for this study, that is, the source code for MPI-ESM1.2 and its components, is available on https://www.mpimet.mpg.de-/en/science/models/mpi-esm/, and output from a large number of experiments will be available through the Earth System Grid Federation (ESGF).

References

Amdahl, G. M. (1967). Validity of the single processor approach to achieving large-scale computing capabilities. *AFIPS Conference Proceedings*, *30*, 483–485. https://doi.org/10.1145/1465482.1465560

Andersson, A., Fennig, K., Klepp, C., Bakan, S., Graßl, H., & Schulz, J. (2010). The Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data—HOAPS-3. *Earth System Science Data*, *2*, 215–234. https://doi.org/10.5194/essd-2-215-2010

Andrews, T., Gregory, J. M., Webb, M. J., & Taylor, K. E. (2012). Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. *Geophysical Research Letters*, *39*, L09712. https://doi.org/10.1029/2012GL051607

Bäuml, G., Chlond, A., & Roeckner, E. (2004). Estimating the PPH-bias for simulations of convective and stratiform clouds. *Atmospheric Research*, *72*, 317–328.

Bauters, M., Drake, T., Verbeeck, H., Bode, S., Herve-Fernandez, P., Zito, P., et al. (2018). High fire-derived nitrogen deposition on Central African forests. *Proceedings of the National Academy of Sciences*, *115*(3), 549–554. https://doi.org/10.1073/pnas.1714597115

Becker, T., Stevens, B., & Hohenegger, C. (2017). Imprint of the convective parameterization and sea-surface temperature on large-scale convective self-aggregation. *Journal of Advances in Modeling Earth Systems*, *9*, 1488–1505. https://doi.org/10.1002/2016MS000865

Bloch-Johnson, J., Pierrehumbert, R. T., & Abbot, D. S. (2015). Feedback temperature dependence determines the risk of high warming. *Geophysical Research Letters*, *42*, 4973–4980. https://doi.org/10.1002/2015gl064240

Bozzo, A., Pincus, R., Sandu, I., & Morcrette, J.-J. (2014). Impact of a spectral sampling technique for radiation on ECMWF weather forecasts. *Journal of Advances in Modeling Earth Systems*, *6*, 1288–1300. https://doi.org/10.1002/2014MS000386

Cahalan, R. F., Ridgway, W., Wiscombe, W. J., Bell, T. L., & Snider, J. B. (1994). The albedo of fractal stratocumulus clouds. *Journal of the Atmospheric Sciences*, *51*(16), 2434–2455.

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., et al. (2013). Carbon and Other Biogeochemical Cycles. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, et al. (Eds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK and New York: Cambridge University Press.

Craig, A., Valcke, S., & Coquart, L. (2017). Development and performance of a new version of the OASIS coupler, OASIS3-MCT3.0. *Geoscientific Model Development*, *10*, 3297–3308. https://doi.org/10.5194/gmd-10-3297-2017

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*, 553–597. https://doi.org/10.1002/qj.828

Dickson, A. G. (2010). The carbon dioxide system in seawater: Equilibrium chemistry and measurements. In U. Riebesell, V. J. Fabry, & L. Hansson (Eds.), *Guide to best practices for ocean acidification research and data reporting* (pp. 17–40). Luxembourg: Publications Office of the European Union.

Dickson, A. G., Sabine, C. L., & Christian, J. R. (2007). Guide to best practices for ocean CO2 measurements, PICES special publication 3 (191 pp.). Retrieved from http://aquaticcommons.org/1443/

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

Fiedler, S., Stevens, B., & Mauritsen, T. (2017). On the sensitivity of anthropogenic aerosol forcing to model-internal variability and parameterizing a Twomey effect. *Journal of Advances in Modeling Earth Systems*, *9*, 1942–2466. https://doi.org/10.1002/2017MS000932

Gates, W. L., Boyle, J., Covey, C., Dease, C., Doutriaux, C., Drach, R., et al. (1998). An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). *Bulletin of the American Meteorological Society*, *73*, 1962–1970.

Gent, P. R., Willebrand, J., McDougall, T., & McWilliams, J. C. (1995). Parameterizing eddy induced tracer transports in ocean circulation models. *Journal of Physical Oceanography*, *25*, 463–474.

Giorgetta, M. A., Brokopf, R., Crueger, T., Esch, M., Fiedler, S., Hohenegger, C., et al. (2018). The ICON atmosphere model with MPI physics. Part 1: Model description. *Journal of Advances in Modeling Earth Systems*, *10*, 1613–1637. https://doi.org/10.1029/2017MS001242

Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., et al. (2013). Climate and carbon cycle changes demo 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *Journal of Advances in Modeling Earth Systems*, *5*, 572–597. https://doi.org/10.1002/jame.20038

Goll, D. S., Brovkin, V., Liski, J., Raddatz, T., Thum, T., & Todd-Brown, K. E. (2015). Strong dependence of CO_2 emissions from anthropogenic land cover change on initial land cover and soil carbon parametrization. *Global Biogeochemical Cycles*, *29*, 1511–1523. https://doi.org/10.1002/2014GB004988

Goll, D. S., Brovkin, V., Parida, B. R., Reick, C. H., Kattge, J., Reich, P. B., & Niinemets, Ü. (2012). Nutrient limitation reduces land carbon uptake in simulations with a model of combined carbon, nitrogen and phosphorus cycling. *Biogeosciences*, *9*, 3547–3569.

Goll, D. S., Winkler, A. J., Raddatz, T., Dong, N., Prentice, I. C., Ciais, P., & Brovkin, V. (2017). Carbon–nitrogen interactions in idealized simulations with JSBACH (version 3.10). *Geoscientific Model Development*, *10*(5), 2009–2030.

Gregory, J. M., Bi, D., Collier, M. A., Dix, M. R., Hirst, A. C., Hu, A., et al. (2013). Climate models without preindustrial volcanic forcing underestimate historical ocean thermal expansion. *Geophysical Research Letters*, *40*, 1600–1604. https://doi.org/10.1002/grl.50339

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

Griffies, S. M., Danabasoglu, G., Durack, P. J., Adcroft, A. J., Balaji, V., Böning, C. W., et al. (2016). OMIP contribution to CMIP6: Experimental and diagnostic protocol for the physical component of the Ocean Model Intercomparison Project. *Geoscientific Model Development*, *9*, 3231–3296. https://doi.org/10.5194/gmd-9-3231-2016

Grokopf, T, Mohr, W., Baustian, T., Schunck, H., Gill, D., Kuypers, M. M. M., et al. (2012). Doubling of marine dinitrogen-fixation rates based on direct measurements. *Nature*, *488*, 361–364.

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*, 4185–4208. https://doi.org/10.5194/gmd-9-4185-2016

Hagemann, S., & Stacke, T. (2015). Impact of the soil hydrology scheme on simulated soil moisture memory. *Climate Dynamics*, *44*, 1731–1750. https://doi.org/10.1007/s00382-014-2221-6

Hansen, J., Sato, M., Ruedy, R., Nazarenko, L., Lacis, A., Schmidt, G. A., et al. (2005). Efficacy of climate forcings. *Journal of Geophysical Research*, *110*, D18104. https://doi.org/10.1029/2005JD005776

Hansis, E., Davis, S., & Pongratz, J. (2015). Relevance of methodological choices for accounting of land use change carbon fluxes. *Global Biogeochemical Cycles*, *29*, 1230–1246. https://doi.org/10.1002/2014GB004997

Hantson, S., Lasslop, G., Kloster, S., & Chuvieco, E. (2015). Anthropogenic effects on global mean fire size. *International Journal of Wildland Fire*, *24*, 589–596.

Hasselmann, K. (2013). Ernst Maier-Reimer: The discovery of silence. *Nature Geoscience*, *6*, 809. https://doi.org/10.1038/ngeo1953

Hasselmann, K., Saussen, R., Maier-Reimer, E., & Voss, R. (1993). On the cold start problem in transient simulations with coupled atmosphere-ocean models. *Climate Dynamics*, *9*, 53–61.

Heinemann, M., Jungclaus, J. H., & Marotzke, J. (2009). Warm Paleocene/Eocene climate as simulated in ECHAM5/MPI-OM. *Climate of the Past*, *5*, 785–802. https://doi.org/10.5194/cp-5-785-2009

Held, I. M., Winton, M., Takahashi, K., Delworth, T., Zeng, F., & Vallis, G. K. (2010). Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing. *Journal of Climate*, *23*, 2418–2427.

Hertwig, E., von Storch, J.-S., Handorf, D., Dethloff, K., Fast, I., & Krismer, T. (2014). Effect of horizontal resplution on ECHAM6-AMIP performance. *Climate Dynamics*, *45*, 185–211.

Hibler, W. D. (1979). A thermodynamic sea ice model. *Journal of Physical Oceanography*, *9*, 815–846.

Houghton, R. A., Hobbie, J. E., Melillo, J. M., Moore, B., Peterson, B. J., Shaver, G. R., & Woodwell, G. M. (1983). Changes in the carbon content of terrestrial biota and soils between 1860 and 1980: A net release of CO_2 to the atmosphere. *Ecological Monographs*, *53*, 253–262. https://doi.org/10.2307/1942531

Huffman, G. J., Adler, R. F., Bolvin, D. T., & Gu, G. (2009). Improving the global precipitation record: GPCP version 2.1. *Geophysical Research Letters*, *36*, L17808. https://doi.org/10.1029/2009GL040000

Hurtt, G., Chini, L., Frolking, S., & Sahajpal, R. (2016). For a short description of the LUH2v2h dataset see. Retrieved from http://gsweb1vh2.umd.edu/LUH2/LUH2_v2h_README.pdf

Hurtt, G. C., Chini, L. P., Frolking, S., Betts, R. A., Feddema, J., Fischer, G., et al. (2011). Harmonization of land-use scenarios for the period 1500-2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary land. *Climatic Change*, *109*, 117. https://doi.org/10.1007/s10584-011-0153-2

Iacono, M. J., Delamare, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research*, *113*, D13103. https://doi.org/10.1029/2008JD00994

Ilyina, T., Six, K. D., Segschneider, J., Maier-Reimer, E., Li, H., & Nunez-Riboni, I. (2013). Global ocean biogeochemistry model HAMOCC: Model architecture and performance as component of the MPI Earth system model in different CMIP5 experimental realizations. *Journal of Advances in Modeling Earth Systems*, *5*, 287–315. https://doi.org/10.1029/2012MS000178

Jones, C. D., Arora, V., Friedlingstein, P., Bopp, L., Brovkin, V., Dunne, J., et al. (2016). C4MIP—the Coupled Climate-Carbon Cycle Model Intercomparison Project: Experimental protocol for CMIP6. *Geoscientific Model Development*, *9*(8), 2853–2880.

Jones, P. D., Lister, D. H., Osborn, T. J., Harpham, C., Salmon, M., & Morice, C. P. (2012). Hemispheric and large-scale land surface air temperature variations: An extensive revision and an update to 2010. *Journal of Geophysical Research*, *117*, D05127. https://doi.org/10.1029/2011JD017139

Jungclaus, J., Fischer, N., Haak, H., Lohmann, K., Marotzke, J., Matei, D., et al. (2013). Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model. *Journal of Advances in Modeling Earth Systems*, *5*, 422–446. https://doi.org/10.1002/jame.20023

Karl, D., Michaels, A., Bergman, B., Capone, D., Carpenter, E., Letelier, R., et al. (2002). Dinitrogen fixation in the world's oceans. *Biogeochemistry*, *57*(1), 47–98.

Kiehl, J. T. (2007). Twentieth century climate model response and climate sensitivity. *Geophysical Research Letters*, *34*, L22710. https://doi.org/10.1029/2007GL031383

Kinne, S., O'Donnel, D., Stier, P., Kloster, S., Zhang, K., Schmidt, H., et al. (2013). MAC-v1: A new global aerosol climatology for climate studies. *Journal of Advances in Modeling Earth Systems*, *5*, 704–740. https://doi.org/10.1002/jame.20035

Korn, P. (2017). Formulation of an unstructured grid model for global ocean dynamics. *Journal of Computational Physics*, *339*, 525–552. https://doi.org/10.1016/j.jcp.2017.03.009

Krismer, T. R., Giorgetta, M. A., & Esch, M. (2013). Seasonal aspects of the quasi-biennial oscillation in MPI-ESM and ERA-40. *Journal of Advances in Modeling Earth Systems*, *5*, 406–421. https://doi.org/10.1002/jame.20024

Lasslop, G., Thonicke, K., & Kloster, S. (2014). SPITFIRE within the MPI Earth system model: Model development and evaluation. *Journal of Advances in Modeling Earth Systems*, *6*, 740–755. https://doi.org/10.1002/2013MS000284

Lawrence, D. M., Hurtt, G. C., Arneth, A., Brovkin, V., Calvin, K. V., Jones, A. D., et al. (2016). The Land Use Model Intercomparison Project (LUMIP) contribution to CMIP6: Rationale and experimental design. *Geoscientific Model Development*, *9*, 2973–2998. https://doi.org/10.5194/gmd-9-2973-2016

Li, H., & von Storch, J.-S. (2013). On the fluctuations of buoyancy fluxes simulated in a 1/10° OGCM. *Journal of Physical Oceanography*, *43*, 1270–1287.

Li, C., von Storch, J.-S., & Marotzke, J. (2013). Deep-ocean heat uptake and equilibrium climate response. *Climate Dynamics*, *40*, 1071–1089. https://doi.org/10.1007/s00382-012-1350-z

Li, Z., von Storch, J.-S., & Müller, M. (2015). The M2 internal tide simulated by a 1/10° OGCM. *Journal of Physical Oceanography*, *45*, 3119–3135.

Li, Z., von Storch, J.-S., & Müller, M. (2017). The K_1 internal tide simulated by a 1/10° OGCM. *Ocean Modelling*, *113*, 145–156.

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

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

Lohmann, U., & Roeckner, E. (1996). Design and performance of a new cloud microphysical scheme developed for the ECHAM general circulation model. *Climate Dynamics*, *12*, 557–572.

Lott, F. (1999). Alleviation of stationary biases in a GCM through a mountain drag parameterization scheme and a simple representation of mountain lift forces. *Monthly Weather Review*, *127*, 788–801.

Mahowald, N., Baker, A., Bergametti, G., Brooks, N., Duce, R., Jickells, T., et al. (2005). The atmospheric global dust cycle and iron inputs to the ocean. *Global Biogeochemical Cycles*, *19*, GB4025. https://doi.org/10.1029/2004GB002402

Mahowald, N. M., Muhs, D. R., Levis, S., Rasch, P. J., Yoshioka, M., Zender, C. S., & Luo, C. (2006). Change in atmospheric mineral aerosols in response to climate: Last glacial period, preindustrial, modern, and doubled carbon dioxide climates. *Journal of Geophysical Research*, *111*, D10202. https://doi.org/10.1029/2005JD006653

Maier-Reimer, E., Hasselmann, K., Olbers, D., & Willebrand, J. (1982). An ocean circulation model for climate studies. Technical Report, Max-Planck-Institute for Meteorology.

Marsland, S. J., Haak, H., Jungclaus, J. H., Latif, M., & Rö ske, F. (2003). The Max-Planck-Institute global ocean/sea ice model with orthogonal curvilinear coordinates. *Ocean Modelling*, *5*, 91–127.

Martin, J. H., Knauer, G. A., Karl, D. M., & Broenkow, W. W. (1987). VERTEX: Carbon cycling in the northeast Pacific. *Deeep Sea Research*, *34*(2), 267–285.

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

Medeiros, B., Stevens, B., Held, I. M., Zhao, M., Williamson, D. L., Olson, J. G., & Bretherton, C. S. (2008). Aquaplanets, climate sensitivity, and low clouds. *Journal of Climate*, *21*, 4974–4991. https://doi.org/10.1175/2008JCLI1995.1

Mellado, J. P. (2017). Cloud-top entrainment in stratocumulus clouds. *Annual Review of Fluid Mechanics*, *49*, 145–69.

Meraner, K, Mauritsen, T., & Voigt, A. (2013). Robust increase in equilibrium climate sensitivity under global warming. *Geophysical Research Letters*, *40*, 5944–5948. https://doi.org/10.1002/2013GL058118

Mikolajewicz, U., Ziemen, F., Cioni, G., Claussen, M., Fraedrich, K., Heidkamp, M., et al. (2018). The climate of a retrograde rotating Earth. *Earth System Dynamics*, *9*, 1191–1215. https://doi.org/10.5194/esd-2018-31

Milinski, S., Bader, J., Haak, H., Siongco, A. C., & Jungclaus, J. H. (2016). High atmospheric horizontal resolution eliminates the wind-driven coastal warm bias in the southeastern tropical Atlantic. *Geophysical Research Letters*, *43*, 10,455–10,462. https://doi.org/10.1002/2016GL070530

Miller, M. J., Palmer, T. N., & Swinbank, R. (1989). Parametrization and influence of subgridscale orography in general circulation and numerical weather prediction models. *Meteorology and Atmospheric Physics*, *40*, 84–109.

Mlawer, E. J., Taubmann, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research*, *102*(D14), 16,663–16,682.

Möbis, B., & Stevens, B. (2012). Factors controlling the position of the intertropical convergence zone on an aquaplanet. *Journal of Advances in Modeling Earth Systems*, *4*, M00A04. https://doi.org/10.1029/2012MS000199

Müller, W. A., Jungclaus, J. H., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., et al. (2018). A high-resolution version of the Max Planck Institute Earth System Model MPI-ESM1.2-HR. *Journal of Advances in Modeling Earth Systems*, *10*, 1383–1413.

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., et al. (2013). Anthropogenic and natural radiative forcing. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Doschung, et al. (Eds.), *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. 659–740). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.018

Najjar, R. G., Jin, X., Louanchi, F., Aumont, O., Caldeira, K., Doney, S. C., et al. (2007). Impact of circulation on export production, dissolved organic matter, and dissolved oxygen in the ocean: Results from Phase II of the Ocean Carbon-cycle Model Intercomparison Project (OCMIP-2). *Global Biogeochemical Cycles*, *21*, GB3007. \ignorespaceshttps://doi.org/10.1029/2006GB002857

Nam, C., Bony, S., Dufresne, J.-L., & Chepfer, H. (2012). The 'too few, too bright' tropical low-cloud problem in CMIP5 models. *Geophysical Research Letters*, *39*, L21801. https://doi.org/10.1029/2012GL053421

Neale, R. B., & Hoskins, B. J. (2001). A standard test for AGCMs including their physical parametrizations: I: The proposal. *Atmospheric Science Letters*, *1*, 108–114. https://doi.org/10.1006/asle.2000.0019

Nordeng, T. E. (1994). Extended versions of the convection parametrization scheme at ECMWF and their impact on the mean and transient activity of the model in the tropics. ECMWF Tech. Memo. No. 206 (41 pp.).

Notz, D., Haumann, F. A., Haak, H., Jungclaus, J. H., & Marotzke, J. (2013). Arctic sea-ice evolution as modeled by MPI-ESM. *Journal of Advances in Modeling Earth Systems*, *5*, 173–194. https://doi.org/10.1002/jame20016

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., et al. (2016). The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, *9*, 3461–3482. https://doi.org/10.5194/gmd-9-3461-2016

Orr, J. C., Najjar, R. G., Aumont, O., Bopp, L., Bullister, J. L., Danabasoglu, G., et al. (2017). Biogeochemical protocols and diagnostics for the CMIP6 Ocean Model Intercomparison Project (OMIP). *Geoscientific Model Development*, *10*(6), 2169–2199.

Pacanowski, R. C., & Philander, S. G. H. (1981). Parameterization of vertical mixing in numerical models of tropical oceans. *Journal of Physical Oceanography*, *11*, 1443–1451.

Palmer, T. N., Shutts, G. J., & Swinbank, R. (1986). Alleviation of systematic westerly bias in general circulation and numerical weather prediction models through an orographic gravity wave drag parameterization. *Quarterly Journal of the Royal Meteorological Society*, *112*, 2056–2066.

Parida, B. (2010). The influence of plant nitrogen availability on the global carbon cycle and N_2O emissions, Reports on Earth System Science. http://www.mpimet.mpg.de/fileadmin/publikationen/Reports/WEBBzE92.pdf

Paulsen, H., Ilyina, T., Six, K. D., & Stemmler, I. (2017). Incorporating a prognostic representation of marine nitrogen fixers into the global ocean biogeochemical model HAMOCC. *Journal of Advances in Modeling Earth Systems*, *9*, 438–464. https://doi.org/10.1002/2016MS000737

Pedersen, C. A., Roeckner, E., Lüthje, M., & Winther, J.-G. (2009). A new sea ice albedo scheme including melt ponds for ECHAM5 general circulation model. *Journal of Geophysical Research*, *114*, D08101. https://doi.org/10.1029/2008JD010440

Pincus, R., Barker, H. W., & Morcrette, J. J. (2003). A fast, flexible, approximate technique for computing radiative transfer in inhomogeneous cloud fields. *Journal of Geophysical Research*, *108*, 438–464. https://doi.org/10.1029/2002JD003322

Pincus, R., Forster, P. M., & Stevens, B. (2016). The Radiative Forcing Model Intercomparison Project (RFMIP): Experimental protocol for CMIP6. *Geoscientific Model Development Discuss*, *2016*, 1–18. https://doi.org/10.5194/gmd-2016-88

Pincus, R., & Stevens, B. (2009). Monte Carlo spectral integration: A consistent approximation for radiative transfer in large eddy simulations. *Journal of Advances in Modeling Earth*, *1*, 1. https://doi.org/10.3894/JAMES.2009.1.1

Pincus, R., & Stevens, B. (2013). Paths to accuracy for radiation parameterizations in atmospheric models. *Journal of Advances in Modeling Earth Systems*, *5*, 225–233. https://doi.org/10.1002/jame.20027

Pithan, F., Shepherd, T. G., Zappa, G., & Sandu, I. (2016). Climate model biases in jet streams, blocking and storm tracks resulting from missing orographic drag. *Geophysical Research Letters*, *43*, 7231–7240. https://doi.org/10.1002/2016GL069551

Popke, D., Stevens, B., & Voigt, A. (2013). Climate and climate change in a radiative convective equilibrium version of ECHAM6. *Journal of Advances in Modeling Earth Systems*, *5*, 1–14. https://doi.org/10.1029/2012MS000191

Popp, M., Schmidt, H., & Marotzke, J. (2015). Initiation of a runaway greenhouse in a process parameterizations cloudy column. *Journal of Climate*, *72*, 452–471.

Popp, M., Schmidt, H., & Marotzke, J. (2016). Transition to a moist greenhouse with CO_2 and solar forcing. *Nature Communications*, *7*, 10627. https://doi.org/10.1038/ncomms10627

Quaas, J. (2012). Evaluating the critical relative humidity as a measure of subgrid-scale variability of humidity in general circulation model cloud cover parameterizations using satellite data. *Journal of Geophysical Research*, *117*, D09208. https://doi.org/10.1029/2012JD017495

Rabin, S. S., Melton, J. R., Lasslop, G., Bachelet, D., Forrest, M., Hantson, S., et al. (2017). The Fire Modeling Intercomparison Project (FireMIP), phase 1: Experimental and analytical protocols with detailed model descriptions. *Geoscientific Model Development*, *10*, 1175–1197.

Räisänen, P., Barker, H. W., Khairoutdinov, M., Li, J., & Randall, D. (2004). Stochastic generation of subgrid-scale cloudy columns for large-scale models. *Quarterly Journal of the Royal Meteorological Society*, *130*, 2047–2068.

Randel, D. L., Vonder Haar, T. H., Ringerud, M. A., Stephens, G. L., Greenwald, T. J., & Combs, C. L. (1996). A new global water vapor dataset. *Bulletin of the American Meteorological Society*, *77*, 1233–1246.

Redi, M. H. (1982). Oceanic isopycanal mixing by coordinate rotation. *Journal of Physical Oceanography*, *12*, 1154–1158.

Reichler, T., & Kim, J. (2008). How well do climate models simulate today's climate? *Bulletin of the American Meteorological Society*, *89*(3), 303–312. https://doi.org/10.1175/BAMS-89-3-303

Reick, C. H., Raddatz, T., Brovkin, V., & Gayler, V. (2013). Representation of natural and anthropogenic land cover change in MPI-ESM. *Journal of Advances in Modeling Earth Systems*, *5*, 459–482. https://doi.org/10.1002/jame.20022

Rimac, A., Von Storch, J.-S., & Eden, C. (2015). The total energy flux leaving the ocean's mixed layer. *Journal of Physical Oceanography*, *46*, 1885–1900. https://doi.org/10.1175/JPO-D-15-0115

Roeckner, E., Arpe, L., Bengtsson, L., Christoph, M., Clauseen, L., Dümenil, L., et al. (1996). *The atmospheric general circulation model ECHAM4: Model description and simulation of present-day climate report 218*. Hamburg: Max Planck Institut für Meteorologie.

Roeckner, E., Dümenil, L., Kirk, E., Lunkeit, F., Ponater, M., Rockel, B., et al. (1989). *The Hamburg version of the ECMWF model (ECHAM), Tech. Rep. 13*. Geneva, Switzerland: World Meteorological Organization.

Roeckner, E. R., Brokopf, M., 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*, 3771–3791. https://doi.org/10.1175/JCLI3824.1

Roeckner, E., Mauritsen, T., Esch, M., & Brokopf, R. (2012). Impact of melt ponds on Arctic sea ice in past and future climates as simulated by MPI-ESM. *Journal of Advances in Modeling Earth Systems*, *4*, M00A02. https://doi.org/10.1029/2012MS000157

Semtner, A. J. (1976). A model for the thermodynamic growth of sea ice in numerical investigations of climate. *Journal of Physical Oceanography*, *6*, 379–389.

Senior, C., & Mitchell, J. F. B. (2000). The time-dependence of climate sensitivity. *Geophysical Research Letters*, *27*(17), 2685–2688.

Siongco, A. C., Hohenegger, C., & Stevens, B. (2014). The Atlantic ITCZ bias in CMIP5 models. *Climate Dynamics*, *45*, 1169–1180. https://doi.org/10.1007/s00382-014-2366-3

Siongco, A. C., Hohenegger, C., & Stevens, B. (2017). Sensitivity of the summertime tropical Atlantic precipitation distribution to convective parameterization and model resolution in ECHAM6. *Journal of Geophysical Research: Atmospheres*, *122*, 2579–2594. https://doi.org/10.1002/2016JD026093

Steele, M., Morley, R., & Ermold, W. (2001). PHC: A global ocean hydrography with a high-quality Arctic Ocean. *Journal Climate*, *14*, 2079–2087.

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

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*(1), 433–452. https://doi.org/10.5194/gmd-10-433-2017

Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., et al. (2013). Atmospheric component of the MPI-M Earth System Model: ECHAM6. *Journal of Advances in Modeling Earth Systems*, *5*, 146–172. https://doi.org/10.1002/jame.20015

Sundqvist, H., Berge, E., & Kristjansson, J. E. (1989). Condensation and cloud parameterization studies with a mesoscale numerical weather prediction model. *Monthly Weather Review*, *117*, 1641–1657.

Thonicke, K., Spessa, A., Prentice, I. C., Harrison, S. P., Dong, L., & Carmona-Moreno, C. (2010). The influence of vegetation, fire spread and fire behaviour on biomass burning and trace gas emissions: Results from a process-based model. *Biogeosciences*, *7*, 1991–2011.

Tian, F., von Storch, J.-S., & Hertwig, E. (2017). *Air-sea fluxes in a climate model using hourly coupling between the atmospheric and the oceanic components* (Vol. 48, pp. 2819–2836). https://doi.org/10.1007/s00382-016-3228-y

Tian, F., von Storch, J.-S., & Hertwig, E. (2018). Impact of SST diurnal cycle on ENSO asymmetry. https://doi.org/10.1007/s00382-018-4271-7

Tiedtke, M. (1989). A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Monthly Weather Review*, *117*, 1779–1800.

Tuomi, M., Laiho, R., Repo, A., & Liski, J. (2011). Wood decomposition model for boreal forests. *Ecological Modelling*, *222*(3), 709–718.

Twomey, S. (1977). The influence of pollution on the shortwave albedo of clouds. *Journal of the Atmospheric Sciences*, *34*, 1149–1152.

Valcke, S. (2013). The OASIS3 coupler: A European climate modelling community software. *Geoscientific Model Development*, *6*, 373–388. https://doi.org/10.5194/gmd-6-373-2013

van den Hurk, B., Kim, H., Krinner, G., Seneviratne, S. I., Derksen, C., Oki, T., et al. (2016). LS3MIP (v1.0) contribution to CMIP6: The Land Surface, Snow and Soil moisture Model Intercomparison Project—Aims, setup and expected outcome. *Geoscientific Model Development*, *9*, 2809–2832. https://doi.org/10.5194/gmd-9-2809-2016

Voigt, A., Abbot, D. S., Pierrehumbert, R. T., & Marotzke, J. (2011). Initiation of a Marinoan Snowball Earth in a state-of-the-art atmosphere-ocean general circulation model. *Climate of the Past*, *7*, 249–263. https://doi.org/10.5194/cp-7-249-2011

Voigt, A., Bony, S., Dufresne, J.-L., & Stevens, B. (2014). The radiative impact of clouds on the shift of the Intertropical Convergence Zone. *Geophysical Research Letters*, *41*, 4308–4315. https://doi.org/10.1002/2014GL060354

von Storch, J.-S., Eden, C., Fast, I., Haak, H., Hernández-Deckers, D., Maier-Reimer, E., et al. (2012). An estimate of the Lorenz energy cycle for the world ocean based on the 1/10° STORM/NCEP simulation. *Journal of Physical Oceanography*, *42*, 2185–2205.

von Storch, J.-S., Haak, H., Hertwig, E., & Fast, I. (2016). Vertical heat and salt fluxes due to resolved and parameterized meso-scale eddies. *Ocean Modelling*, *108*, 1–19.

Wanninkhof, R. (2014). Relationship between wind speed and gas exchange over the ocean revisited. *Limnology and Oceanography: Methods*, *12*(6), 351–362. https://doi.org/10.4319/lom.2014.12.351

Winton, M., Takahashi, K., & Held, I. M. (2010). Importance of ocean heat uptake efficacy to transient cliamte change. *Journal Climate*, *23*, 2333–2344.

Wolff, J. O., Maier-Reimer, E. E., & Legutke, S. (1997). The Hamburg Ocean Primitive Equation Model HOPE, technical report no. 13 German Climate Computer Center (DKRZ) Hamburg Germany.

Zaehle, S. (2013). Terrestrial nitrogen–carbon cycle interactions at the global scale. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *368*(1621), 20130125.

Zängl, G., Reinert, D., Ripodas, P., & Baldauf, M. (2015). The ICON (ICOsahedral non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core. *Quarterly Journal of the Royal Meteorological Society*, *141*, 563–579. https://doi.org/10.1002/qj.2378

Ziemen, F. A., Rodehacke, C. B., & Mikolajewicz, U. (2014). Coupled ice sheet–climate modeling under glacial and pre-industrial boundary conditions. *Climate of the Past*, *10*, 1817–1836. https://doi.org/10.5194/cp-10-1817-2014