Geoscientific
Model Development

Model evaluation paper

AWI-CM3 coupled climate model: description and evaluation experiments for a prototype post-CMIP6 model

Jan Streffing[1,2], **Dmitry Sidorenko**[1], **Tido Semmler**[1], **Lorenzo Zampieri**[3], **Patrick Scholz**[1], **Miguel Andrés-Martínez**[1], **Nikolay Koldunov**[1], **Thomas Rackow**[4,1], **Joakim Kjellsson**[5], **Helge Goessling**[1], **Marylou Athanase**[1], **Qiang Wang**[1], **Jan Hegewald**[1], **Dmitry V. Sein**[1,6], **Longjiang Mu**[7,1], **Uwe Fladrich**[8], **Dirk Barbi**[9,1], **Paul Gierz**[1], **Sergey Danilov**[1,2], **Stephan Juricke**[1,2], **Gerrit Lohmann**[1,10], and **Thomas Jung**[1,10]

[1]Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, Germany
[2]Department of Mathematics & Logistics, Jacobs University Bremen, Campus Ring 1, 28759 Bremen, Germany
[3]National Center for Atmospheric Research, 1850 Table Mesa Dr, Boulder, CO 80305, United States of America
[4]European Centre for Medium-Range Weather Forecasts, Robert-Schuman-Platz 3, 53175 Bonn, Germany
[5]GEOMAR Helmholtz Centre for Ocean Research Kiel, Wischhofstraße 1–3, 24148 Kiel, Germany
[6]Shirshov Institute of Oceanology, RAS, Moscow, Russia
[7]Pilot National Laboratory for Marine Science and Technology, Qingdao, China
[8]Swedish Meteorological and Hydrological Institute, Folkborgsvägen 17, 60176 Norrköping, Sweden
[9]University IT and Data Center, Rhenish Friedrich Wilhelm University of Bonn, Regina-Pacis-Weg 3, 53113 Bonn, Germany
[10]Department of Physics, University of Bremen, Bibliothekstraße 1, 28359 Bremen, Germany

Correspondence: Jan Streffing (jan.streffing@awi.de) and Dmitry Sidorenko (dmitry.sidorenko@awi.de)

Abstract. We developed a new version of the Alfred Wegener Institute Climate Model (AWI-CM3), which has higher skills in representing the observed climatology and better computational efficiency than its predecessors. Its ocean component FESOM2 (Finite-volumE Sea ice–Ocean Model) has the multi-resolution functionality typical of unstructured-mesh models while still featuring a scalability and efficiency similar to regular-grid models. The atmospheric component OpenIFS (CY43R3) enables the use of the latest developments in the numerical-weather-prediction community in climate sciences. In this paper we describe the coupling of the model components and evaluate the model performance on a variable-resolution (25–125 km) ocean mesh and a 61 km atmosphere grid, which serves as a reference and starting point for other ongoing research activities with AWI-CM3. This includes the exploration of high and variable resolution and the development of a full Earth system model as well as the creation of a new sea ice prediction system. At this early development stage and with the given coarse to medium resolutions, the model already features above-CMIP6-average skills (where CMIP6 denotes Coupled Model Intercomparison Project phase 6) in representing the climatology and competitive model throughput. Finally we identify remaining biases and suggest further improvements to be made to the model.

1 Introduction

The evolution of coupled climate models between phases of the Coupled Model Intercomparison Project (CMIP) is advancing our ability to simulate the Earth's climate and to quantify humankind's past and future impact.

The Alfred Wegener Institute (AWI) contributed to CMIP6 with the atmosphere–ocean general circulation model (AOGCM) AWI-CM1.1-MR (Semmler et al., 2020) as well as the Earth system model (ESM) AWI-ESM1.1-LR (Danek et al., 2020), built upon AWI-CM1 through using an additional dynamic vegetation module. The AOGCM ver-

sion also contributed to the HighResMIP of CMIP6 (Rackow et al., 2022) with AWI-CM1.1-LR and AWI-CM1.1-HR.

Experience from HighResMIP shows that, with respect to model accuracy, high-resolution climate models are about one CMIP generation ahead of their standard-resolution counterparts (Bock et al., 2020). Moreover, atmospheric modeling studies indicate that a number of key processes ranging from orographic drag (Pithan et al., 2016) to atmospheric blocking (Schiemann et al., 2017; Davini et al., 2017), storm tracks (Willison et al., 2015; Baker et al., 2019), and precipitation (van Haren et al., 2015), as well as mean sea level pressure (Hertwig et al., 2015), can be improved by increased horizontal resolution. In many of these cases the operational resolutions of the current numerical-weather-prediction (NWP) systems (~ 10 km) would be sufficient but are often not reached by climate models running more commonly at ~ 100 km resolution due to computational cost and time-to-solution limitations.

While the scientific evaluation of CMIP6 is still ongoing, we turn our attention to lessons learned and begin the development of our next-generation climate model. Our CMIP6 model, AWI-CM1, has reached the limits of scalability, both in the design of its numerical cores and in the peripheries, such as data structures and input–output (IO) schemes. As a first step forward, AWI embarked on a mission to create a Finite-volumE Sea ice–Ocean Model (FESOM2) with a finite-volume numerical core instead of finite elements (Danilov et al., 2017; Scholz et al., 2019; Koldunov et al., 2019a). In the vertical dimension the arbitrary Lagrangian Eulerian (ALE) framework in FESOM2 allows the vertical grids to follow the isopycnals with reduced numerical mixing and to follow bottom topography with improved representation of bottom boundary layers. FESOM2 thus bears resemblance to the MPAS model described by Petersen et al. (2015). Following the development of FESOM2, a new climate model was assembled, coupling this upgraded ocean model with the same atmospheric model as before, ECHAM6 (Sidorenko et al., 2019). This model, dubbed AWI-CM2, is however practically limited to an atmospheric resolution of about 100 km grid spacing, with an absolute upper limit of 50 km.

Highly relevant atmosphere–ocean coupled processes such as local energy transfer, ocean warm-layer formation, and diurnal cycles require not only high resolution in the ocean component but an atmosphere that can adequately react to the ocean in an eddying regime (Ma et al., 2016; Renault et al., 2016). We therefore couple FESOM2 to the OpenIFS atmospheric model to develop our new AOGCM AWI-CM3. OpenIFS is based on the Integrated Forecasting System (IFS) numerical-weather-prediction suite. IFS and OpenIFS are highly scalable and have been used in the Centre of Excellence in Simulation of Weather and Climate in Europe (ESiWACE) (Zeman et al., 2021). They constitute the highest atmospheric resolution contribution to HighResMIP (Haarsma et al., 2016) and have a long history of experimental (Jung

et al., 2012) and operational (Malardel et al., 2016) high-resolution atmospheric modeling.

While it might be tempting to apply the coupled model at the highest possible spatial resolution right from the start, cost and time considerations dictate that initial development and evaluation best be done at lower resolution. Furthermore many future applications of AWI-CM3, especially as the basis for a paleoclimate-capable full Earth system model (ESM), will likely not employ particularly high resolutions due to long simulation periods and/or a large number of tracers. Finally, conducting this first model development phase at relatively low resolutions enables a fair comparison between this model and the old AWI-CM1.1, as well as other well-established climate models. We therefore present in detail the capabilities and scientific applicability of the lower-resolution AWI-CM3 here, with a glance at its higher-resolution performance. What we consider low resolution for the OpenIFS atmosphere TCo159L91 (61 km) is already beyond the practical limits of our previous AWI-CM1 and AWI-CM2 models with ECHAM6 (100 km). The higher-resolution simulation that we briefly touch on features a TCo319L137 (31 km) atmosphere as well as a 5–27 km ocean, making the simulation more finely resolved than the HighResMIP contribution with AWI-CM1.1 HR.

2   Model components

The AWI-CM3 coupled system encompasses two major components with the atmosphere (OpenIFS) and ocean (FESOM2), as well as an auxiliary component, the runoff mapper. The fourth component (XML Input/Output Server – XIOS), running in parallel, handles the output from the atmospheric model.

2.1   OpenIFS 43R3 atmosphere

For its atmospheric component AWI-CM3 uses OpenIFS, which is based on ECMWF's Integrated Forecast System (IFS) (ECMWF, 2017a, b, c). With the readily available OpenIFS, ECMWF aims to provide cutting-edge performance from the world of operational numerical weather forecasting to the research community while in turn presenting a test bed for developments that can feed back into weather forecasting.

As such, OpenIFS contains the hydrostatic dynamical core and the physical parameterizations, as well as the H-TESSEL hydrology model (Balsamo et al., 2009) and the WAM wave model (Komen et al., 1996) from the ECMWF IFS suite. The two-way OpenIFS–WAM coupling makes the surface roughness calculation dependent on the wave state, which in turn influences the calculation of momentum and sensible heat fluxes. WAM wave fields are currently not directly coupled to FESOM2 for, e.g., wave-induced vertical mixing calculations. H-TESSEL provides column model type soil
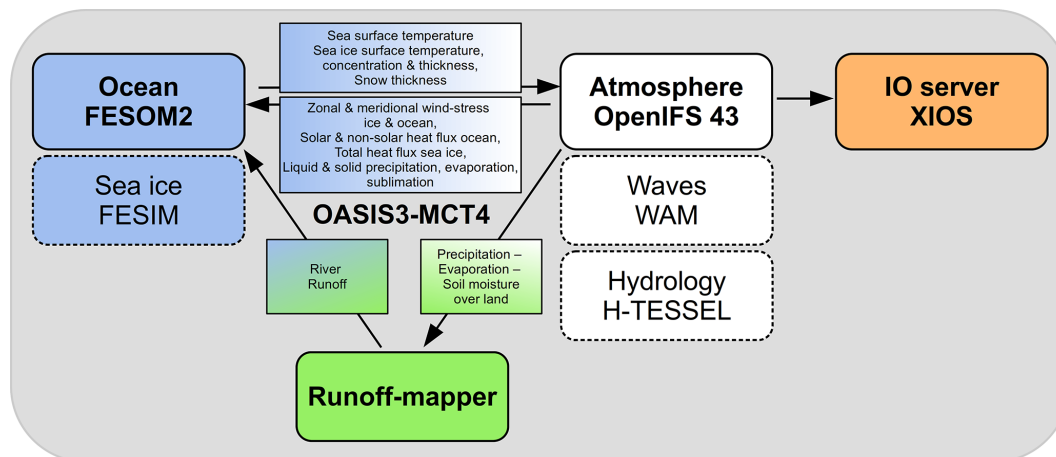
Figure 1. Schematic of the coupling between AWI-CM3 model components. Two-hourly parallel communication in AWI-CM3 is implemented via the OASIS3-MCT_4.0 library. Heat, mass, and momentum fluxes are sent from OpenIFS to FESOM2, while ocean and ice surface state variables are sent back. The precipitation minus evaporation over land is handled with a runoff mapper that generates river runoff at basin discharge points. OpenIFS output is written in parallel with optional online postprocessing via XIOS.

moisture computations, while horizontal water transport on land is simulated using a separate runoff mapper, as shown in Fig. 1. In contrast to the operational IFS NWP system, the 4D-Var data assimilation, the non-hydrostatic core, the adjoint/tangent-linear versions, the Météo-France IO server, and the subroutine level implementation of the NEMO ocean model have been removed from OpenIFS.

The defining feature of the OpenIFS numerical core is its semi-Lagrangian (Ritchie et al., 1995) semi-implicit (Robert et al., 1972) advection scheme (Ritchie, 1987; Ritchie et al., 1995), which allows for advection distances in excess of the Courant–Friedrichs–Lewy (CFL) condition. At the same grid resolution and with the same characteristic fluid velocities, this permits OpenIFS simulations to be numerically stable at much longer time steps than equivalent Eulerian integrations. What the model saves in terms of computing resources can be invested instead into higher model resolution, larger ensembles, or longer simulations. We use the cycle 43R3V1 version of OpenIFS released in September 2020 that is based on IFS CY43R3, which constituted the operational NWP system at ECMWF between July 2017 and June 2018. In contrast to ECHAM6, OpenIFS allows for the use of both full and reduced Gaussian grids (Hortal and Simmons, 1991), thus reducing grid cell shape distortion near the poles.

OpenIFS is available at a wide variety of horizontal resolutions, ranging from a TQ21 (626 km) toy model to the TCo1279 (9 km) operational NWP. Even higher experimental horizontal resolutions require only minor source code changes. Three options exist for the representation of the highest-resolution (truncation) spherical harmonics in grid point space: the linear-truncation TL grids with two grid points for the smallest spherical harmonics, the quadratic TQ with three grid points, and the cubic octahedral TCo with four grid points (Malardel et al., 2016). Of these, the TCo

grids are the most accurate and applicable to coupled climate simulations as the coupling takes place in grid point space. The vertical-resolution choice is much more limited than the horizontal one, as each horizontal resolution is typically paired with one optimal vertical resolution, for which the model parameterizations have been tuned.

The main experiments we present were performed at a resolution of TCo159 (61 km) with 91 vertical layers (TCo159L91), with some experiments at TCo319L137 (31 km). Further resolutions we successfully tested computationally are TCo95L91 (100 km) and TCo639L137 (16 km). All grids have a ceiling pressure level of 0.01 hPa, with the L137 grids resolving the vertical space in between more finely than the L91 grids.

2.2 FESOM2 ocean

The ocean dynamics of AWI-CM3 are simulated by the Finite-volumE Sea ice–Ocean Model (FESOM2), the second version of the global unstructured-mesh ocean model developed at AWI (Danilov et al., 2017).

FESOM2 is formulated with a finite-volume dynamical core using ALE vertical coordinates. FESOM2 is a global unstructured-mesh ocean model that has computational performance comparable to structured-mesh models. Unstructured meshes allow for local mesh refinements without sharp resolution boundaries, as encountered by classical nesting models. In practice, the mesh can be designed to follow the patterns of local sea surface height variability or to scale corresponding to the local Rossby radius of deformation (Sein et al., 2017). FESOM2 contains the embedded FESIM sea ice model (Danilov et al., 2015). For the coupled model presented here, FESIM was modified such that it calculates prognostically the sea ice surface temperature, while in the

previous AOGCMs AWI-CM1 and AWI-CM2 this computation was performed in the atmospheric component.

For the experiments presented here we employ a mesh called CORE2, which has about 127 000 surface nodes, as shown in Koldunov et al. (2019a). The mesh has 47 vertical layers and horizontal resolution varying from 25 to 125 km depending on latitude and distance to coastlines. We tested a second mesh called DART with shorter simulations. The DART mesh has 3.1 million surface nodes with 80 vertical layers and horizontal resolution of 5 to 27 km. We show the spacial distribution of horizontal resolution for both meshes in Fig. 2. In the vertical, the first two layers of both meshes have 5 m resolution, and subsequent layers are in 10 m intervals till 100 m depth. Below 100 m depth the CORE2 vertical resolution decreases downwards more rapidly than the DART resolution.

The low-resolution CORE2 mesh for FESOM2 is not eddy-resolving, and we employ the Gent–McWilliams (GM) parameterization to include the effect of mesoscale eddies on temperature, salinity, and tracers (Gent and Mcwilliams, 1990). For all runs presented here, FESOM2 was configured with the "zstar" vertical coordinate, where the total change in sea surface height is distributed over all layers in the vertical, except the partial cell layer at the bottom (Scholz et al., 2019). Such a vertical setup reduces erroneous numerical mixing in the vertical (Adcroft and Campin, 2004). Parameterization of physical vertical mixing is achieved via the K-profile parameterization (KPP) scheme after Large et al. (1994). In the Southern Ocean we additionally apply vertical mixing within the Monin–Obukhov length scale calculated based on heat flux, freshwater flux, wind stress, sea ice concentration, and sea ice velocity, as developed by Timmermann and Beckmann (2004) based on Lemke (1987). For the horizontal viscosity, FESOM2 applies the kinematic backscatter scheme of Juricke et al. (2020) with a backscatter coefficient of 1.5. This scheme dissipates kinetic energy on small scales but reinjects kinetic energy on large scales, resulting in overall reduced dissipation. Detailed information on the available mixing-scheme options in FESOM2 can be found in Scholz et al. (2022a). For the experiments shown below, the net evaporation − precipitation − runoff ($E - P - R$) integrated over the global ocean is forced to 0 at each time step by subtracting the residual.

The FESIM sea ice model is integrated directly in FESOM2 source code on a module level. The sea ice computations are done on the ocean surface grid, and the dynamics use an adaptive elastic–viscous–plastic solver (Kimmritz et al., 2016; Koldunov et al., 2019b). For the thermodynamics FESIM implements a 0-layer scheme after Parkinson and Washington (1979). The standalone ocean model FESOM2 allows for the optional use of the Icepack sea ice thermodynamics module (Hunke et al., 2020; Zampieri et al., 2021), representing a potential future upgrade for AWI-CM3.

Bodies of water that are cut off by land from the world oceans, such as the Caspian Sea and the Great Lakes, are not simulated via FESOM2 but are included as lakes via the OpenIFS lake module.

For the low-resolution CORE2 mesh in particular, several ocean basins with narrow outflow channels are not included. These are the White Sea, Persian Gulf, Black Sea, and Gulf of Ob. Such narrow inlets on a coarse mesh would reduce the smallest horizontal grid spacing and thus incur a smaller time step in order to still fulfill the CFL condition globally. The higher-resolution DART mesh does not omit these basins.

2.3 Runoff mapper

In addition to the two major components, AWI-CM3 includes a river routing scheme. It receives from OpenIFS the difference between precipitation, evaporation, and soil moisture over land ($P - E - S$) and uses a map of river basins to deliver the water to discharge points along the coastline. The current river routing component has no water storage and thus acts instantaneously. Separation of the routing component from the atmosphere and ocean is a design decision shared with EC-Earth, where the flexibility and ease of modification are core ideas. This design keeps open the option of swiftly replacing the basic runoff mapper with a more sophisticated hydrological model, such as mHM (Samaniego et al., 2010) or CaMa-Flood (Yamazaki et al., 2011), in the future.

2.4 XIOS parallel IO server

The fourth model component of AWI-CM3 is a technical helper. ECMWF removes the parallel IO server developed by Météo-France for IFS when generating a new release of OpenIFS. This leaves OpenIFS with only the possibility of providing sequentially written GRIB file output.

While this sequential IO scheme has been used, for example, by Döscher et al. (2022), it can often reach data-throughput limits in practical applications. Furthermore, while GRIB files are common in the NWP community, the climate modeling community often uses NetCDF files. Thus, the sequential output has to be converted for most analysis tools after each simulation.

With increasing model resolution and improved use of MPI/OMP hybrid parallelization, the sequential IO overhead constitutes an ever-growing fraction of the computational cost of running OpenIFS. Recently Yepes-Arbós et al. (2022) implemented the parallel XML Input/Output Server (XIOS) 2.5 into OpenIFS for this reason, and we make use of it for AWI-CM3 to reduce the computational cost and increase the integration speed.

While XIOS takes the file writing out of the critical path of the simulation, the overhead cost of XIOS is non-zero. The main reason for this is that XIOS works only in grid point space and therefore requires spectral fields to be transformed inside OpenIFS before they can be sent to XIOS for writing. Nevertheless it provides a significant reduction in computing cost, as shown in Table 2. Furthermore, XIOS enables
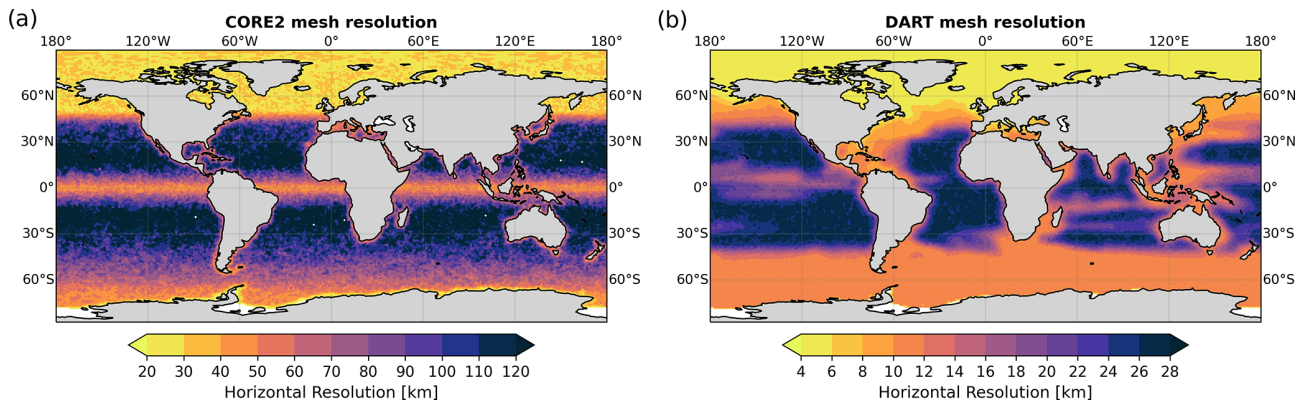
Figure 2. (a) Horizontal resolution of the CORE2 mesh used for the main simulations presented in this work. **(b)** As **(a)** but for the DART mesh with higher resolution, which is based on a mix of the local Rossby radius of deformation and local sea surface height variability. The DART mesh is used in Sect. 5.1 for an outlook on high-resolution applications.

online data postprocessing, such as vertical and horizontal interpolation, as well as temporal operators for maximum, minimum, mean, etc. In doing so, XIOS reduces the number of times that files have to be written and read from disk, saving storage space and reducing the number of job steps in the workflow. Finally XIOS allows for output directly in NetCDF, facilitating the use of AWI-CM3 model output.

FESOM2 contains its own bespoke parallel IO routines, and the integration of an external dedicated IO library is currently not envisioned.

3 Coupled model description

The coupled climate model is constructed by combining and building on the approaches of Hazeleger et al. (2010) and Sidorenko et al. (2015). The OpenIFS version 43R3 is common between a number of AOGCMs and ESMs with regular ocean grids currently under development, including EC-Earth4 of the EC-Earth consortium and GEOMAR's FOCI-OpenIFS (Kjellsson et al., 2020). Indeed the basic functionality of the coupling interface, as well as the future ESM component integration, will be shared between EC-Earth4 and AWI-CM3. Nevertheless, some differences in the coupling strategy exist, and the setup developed for AWI-CM3 shall be detailed further here.

The ESM-Tools (version 3) infrastructure software (Barbi et al., 2021) was used to manage the configuration, compiling, and runtime scripts of the coupled model, as well as to ensure simulation reproducibility.

Coupling strategy

The surface heat, mass, and momentum fluxes are calculated within OpenIFS and supplied to FESOM2. Here the state variables for ocean and sea ice surface are updated accordingly. The runoff mapper calculates its river routing after

the atmosphere component computes and provides the $P - E$ over land for a given coupling time step.

We employ so-called concurrent coupling, with surface condition updates that are considered to be numerically independent between ocean and atmosphere. The temporal exchange of the ocean and atmosphere surface conditions takes place at the least common multiple of the ocean and atmosphere time steps. For the TCo159L91-CORE2 simulations the time steps for coupling, the atmospheric model, and the oceanic model are 120, 60, and 40 min, respectively, while for the TCo319L137-DART simulation they are 60, 15, and 4 min. In the production mode both the atmosphere and the ocean components compute their own surface update at time t_n based on time-lagged information at t_{n-1} of the other component. The physical inconsistencies resulting from this double-sided-lag method are small compared to those stemming from, e.g., spatial and temporal truncation and are generally accepted in the climate modeling community as they allow for parallel execution of model components (Lemarié et al., 2015; Marti et al., 2021).

AWI-CM3 can also be run in a sequential atmosphere-first mode, updating the ocean at time step t_n with atmospheric fluxes from t_n. In this mode climate models can get very close to what would be a converged solution of an iterative coupling at the atmosphere–ocean interface (Marti et al., 2021). Integration of a Schwarz iterative method for fully converged surface coupling is not planned due to the high computational cost compared to small reduction in model error.

On the technical side all three components of the coupled model are compiled into their own respective executables, and a parallel communication library, OASIS3-MCT_4.0 (Craig et al., 2017), is integrated into each one. The AOGCM setup realized is sketched in Fig. 1. Since FESOM2 has previously been coupled to the atmospheric model ECHAM6 (Sidorenko et al., 2019), an interface for the data exchange already existed, and the grouping of fluxes in OpenIFS has been modeled after the grouping in ECHAM6. OpenIFS

CY43R3 had not been coupled via OASIS before, but an older related model OpenIFS CY40R1 was coupled via OASIS as a test case in EC-Earth3. The coupling of the interface for OpenIFS CY43R3 is inspired by this predecessor. Future releases of OpenIFS will be published with this coupling interface already included.

4 Climatological performance

In this section we will outline the ability of the new coupled system to reach a stable equilibrium with constant greenhouse gas and solar forcing from the year 1850. Thereafter we test to what degree the model can simulate the climate as observed over the period 1850 to 2014, with a particular focus on the last 25 years, when the observational coverage is most dense and reliable. Finally this is followed by a characterization of the response to two idealized future CO_2 emission scenarios, one with a sudden $4\times$ increase in CO_2, and the other with a constant increase of 1 % per year, starting from 1850 values.

4.1 Spinup drift (SPIN)

A 700-year-long spinup of AWICM3 was carried out under constant greenhouse gas and solar forcing from the year 1850, starting from winter Polar Science Center Hydrographic Climatology (PHC3) (Steele et al., 2001). The forcing fields were collected from the input4MIPs data server (https://esgf-node.llnl.gov/search/input4mips/, last access: 6 November 2019). Aerosol fields were kept at present-day levels as the integration of the emissions-based aerosols into OpenIFS CY43R3 through the EC-Earth consortium with tracing via the M7 model (Vignati et al., 2004) is still ongoing. This implies a somewhat colder pre-industrial state, in particular in regions of the Northern Hemisphere that are cooled in present-day observations due to industrial aerosol emissions.

During the first 500 years of the spinup simulation we noted positive global ocean temperature trends throughout nearly the entire water column, as can be seen in the Hovmöller diagram of Fig. 3b. Evaluation of the top-of-atmosphere (TOA) and surface (SFC) net heat fluxes in the atmospheric model revealed a near-constant radiative imbalance of $2\,\mathrm{W\,m^{-2}}$, depicted in Fig. 3a. A partial solution is to switch the OpenIFS mass fixer from dry mass to total mass conservation (Malardel et al., 2019) with the McGregor scheme (McGregor, 2005). We implemented this solution starting from the year 1651. Subsequently the radiative imbalance reduced to $+0.7\,\mathrm{W\,m^{-2}}$, which is within the range of imbalance of CMIP6 models (Wild, 2020).

Further experiments showed that decreasing the OpenIFS time step from 60 to 30 min reduced the imbalance to $+0.4\,\mathrm{W\,m^{-2}}$. An additional reduction to 15 min time steps showed no more improvements. We surmise that the time-

step-dependent component of the error implicates the semi-Lagrangian trajectory algorithm operating close to the stability limit for the given atmospheric resolution of TCo159. For our analysis we judged this additional error an acceptable price for a doubling in model integration speed, and we thus keep the 60 min time step. The time-step-independent flux imbalance of $+0.4\,\mathrm{W\,m^{-2}}$ will be targeted in future model development and tuning efforts.

In the global-mean ocean temperature Hovmöller diagram (Fig. 3) we can see that after switching to total mass conservation, the reduction in spurious heat production in the atmosphere led to a stabilization of the ocean temperatures in the upper 1000 m. The trend in the deep ocean, strongest at 4500 m depth, on the other hand has hardly slowed down and, thus, likely has a different origin. One candidate currently under investigation is the topography-influenced equilibrium depth of the Strait of Gibraltar overflow. Alternatively we speculate that overestimated mixing from the KPP mixing scheme (Large et al., 1994) might be the reason. The bias pattern is very similar to that of the previous FESOM2–ECHAM6 (AWI-CM2) coupled model (Sidorenko et al., 2019).

Another potential contributor to the accumulation of heat at depth is a consistent positive shortwave radiation bias of OpenIFS in the Southern Hemisphere. Some of the spuriously heated Southern Ocean surface water becomes entrained into the Antarctic Intermediate Water (AAIW). In recent years, the Southern Ocean shortwave radiation bias has been the subject of research, which led to its reduction by $5–10\,\mathrm{W\,m^{-2}}$ (Forbes et al., 2016). We have already backported these improvements originally developed for the operational IFS CY45 model into OpenIFS CY43R3 prior to our spinup simulation. Even with the improvements, a positive shortwave downward radiation bias of up to $5–10\,\mathrm{W\,m^{-2}}$ between 45–60° S remains and can be seen in Fig. 6e. Idealized experiments have shown that removal of the remaining shortwave radiation bias would cool the Southern Ocean by roughly 1 K within a decade, with potentially larger improvements on longer timescales (not shown).

4.2 Pre-industrial control (PICT)

As evident from the Hovmöller diagram (Fig. 3b), the spinup run is not yet in equilibrium at depths greater than 1000 m. A small residual drift can also be found at the ocean surface and in the atmosphere, which can be seen as a consequence of the still-drifting deep ocean. Based on experience with other ocean models we can estimate that a 3000–5000-year-long simulation would be needed for the model to reach full equilibrium (Rackow et al., 2018). Instead, we run a pre-industrial control experiment which serves as a reference for correcting the historical-period simulation with respect to the remaining trends. The pre-industrial control run thus extends the spinup run by 165 years with the same year 1850 greenhouse gas and solar forcing. We construct a simple linear re-
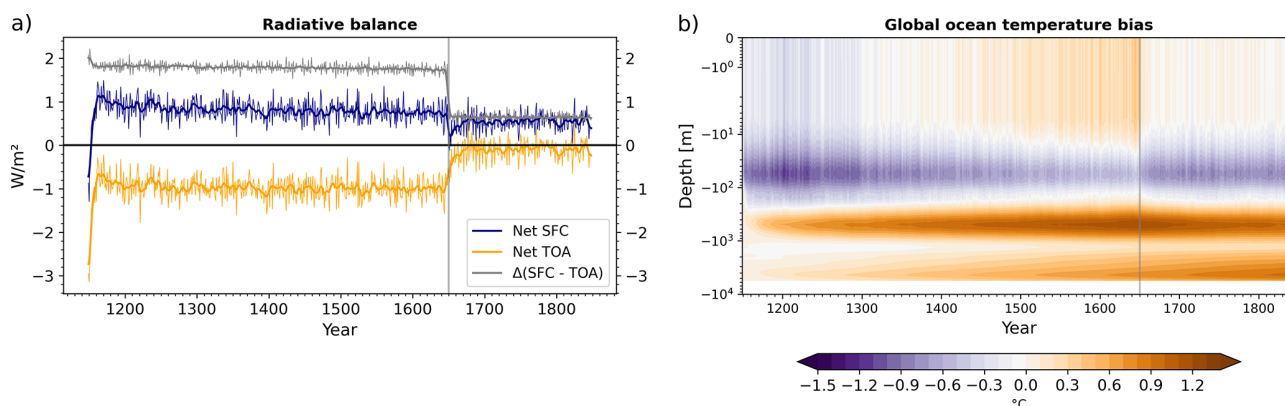
Figure 3. (a) Net radiative imbalance at the top-of-atmosphere (TOA) and at the surface (SFC) in the spinup simulation. Positive (negative) values indicate downward (upward) net heat flux. The difference results from spurious energy production in the atmosphere. **(b)** Semi-logarithmic depth Hovmöller diagram of the evolution of the global-mean ocean temperature bias over the spinup period with respect to PHC3 climatological values (Steele et al., 2001). Switching from the dry to total mass fixer after 500 years reduced spurious heat production in the atmosphere and halted the ocean warming trend in the upper 1000 m. Warming at depth continued.

gression model for the remaining drifts in PICT and subtract these when we analyze the response to historical forcing in the historical simulation.

4.3 Climatological performance during the historical period (HIST)

In the following we first characterize the most prominent bias patterns of the last 25 years of the historical simulation with regards to reanalysis and satellite data products. We then take a look at the climate response to the historical forcing in comparison to the pre-industrial control.

For a succinct overview of the model performance we calculated climate model performance indices, based on Reichler and Kim (2008) and shown in Table 1. For the 4 seasons, 7 regions and 12 key variables, the absolute error in climatology of the last 25 years in AWI-CM3 historic simulation as a fraction of the absolute error averaged over 30 CMIP6 contributing models can be seen. A complete list of the CMIP6 models serving as the evaluation set is given in Appendix A. The list of observational datasets used to calculate all mean absolute errors is also given in Appendix A. For all model and observational datasets, where available, the time period from December 1989 until November 2014 is considered. As the performance of AWI-CM3 is expressed as a fraction of the errors in the CMIP6 average performance and values below 1 indicate better performance, values above 1 point to larger model errors.

For the majority of variables, seasons, and regions our post-CMIP6 prototype model is already performing better than the average CMIP6 model. The lead is especially large for cloud cover (clt) and 500 hPa geopotential height (zg). For surface meridional wind (vas), surface zonal winds (uas), 300 hPa zonal wind (ua), TOA outgoing longwave radiation (rlut), and precipitation (pr), the bias compared to obser-

vations is mostly below average. The very important near-surface (2 m) air temperature (tas) is simulated well in the Arctic (60–90° N) and reasonably well in the northern mid-latitudes (30–60° N) and tropics (30° S–30° N). In the southern mid-latitudes (30–60° S) and Antarctic (60–90° S) the air temperature bias is relatively high. A look at the sea ice concentration (siconc) reveals that the respective errors are far above average in the Antarctic. The problem is particularly severe in the austral winter season. This bias is in co-occurrence with a large mixed-layer-depth bias in the Antarctic. We note that the mixed-layer depth from reanalysis we used here as the reference could be associated with uncertainties too, as it is based on the implementation of mixing parameterizations in the reanalysis model. The standard deviation of sea surface height (zos) shows reasonable variability in the ocean currents in the middle and high latitudes, with problems found in the Niño 3.4 region (5° S–5° N, 170–120° W). On the other hand, the standard deviation of temperature is best represented here, with deficiencies in the Antarctic cold season and northern mid-latitude summer.

A simple average over all individual performance indices gives AWI-CM3 a score of 0.931, with 15 out of the 30 considered CMIP6 models performing better. The overall index of the AWI-CM3 prototype simulation is improved compared to its CMIP6 predecessor with similar resolution and computational cost (AWI-ESM1.1-LR, 1.044). The performance of our medium-resolution AWI-CM1.1-MR CMIP6 contribution (Semmler et al., 2020) is better (0.894), but this model configuration is 20 times more expensive to run than the simulations presented here. Preliminary tests with higher resolution at equal computational cost indicate that AWI-CM3 can achieve better climatological performance than AWI-CM1.1-MR, as we will show in Sect. 5.1. A major contributing factor is the faster dynamic cores and better computational scalabil-

Table 1. Performance indices after Reichler and Kim (2008) that give the absolute error in climatology of the last 25 years in AWI-CM3 historic simulation as a fraction of the absolute error averaged over CMIP6 models. Values below (above) 1 correspond to below (above) CMIP6 average biases. The underlying observations against which all models were evaluated are OSI SAF OSI-450 (Lavergne et al., 2019) – sea ice concentration (siconc); MODIS Atmosphere L2 Cloud Product (Platnick et al., 2015) – cloud cover (clt); Global Precipitation Climatology Project (GPCP) Monthly analysis (Adler et al., 2018) – precipitation (pr); Clouds and the Earth's Radiant Energy System (CERES) (Wielicki et al., 1996) – TOA outgoing longwave radiation (rlut); ECMWF reanalysis ERA5 (Hersbach et al., 2020) – near-surface air temperature (tas), eastward near-surface wind (uas), northward near-surface wind (vas), 300 hPa eastward wind (ua), and 500 hPa geopotential height; NOAA Jason-1, Jason-2, and CryoSat-2 combined – sea surface height (zos); HadISST2 (Titchner and Rayner, 2014) – sea surface temperature (tos); and C-GLORSv7 (Storto et al., 2016) – mixed-layer depth (mlotst). A full list of the considered CMIP6 models is given in Appendix A.

AWI-CM3 CMPI: 0.931

	arctic MAM	arctic JJA	arctic SON	arctic DJF	northmid MAM	northmid JJA	northmid SON	northmid DJF	tropics MAM	tropics JJA	tropics SON	tropics DJF	nino34 MAM	nino34 JJA	nino34 SON	nino34 DJF	southmid MAM	southmid JJA	southmid SON	southmid DJF	antarctic MAM	antarctic JJA	antarctic SON	antarctic DJF
siconc	1.07	0.92	0.74	0.98	0.67	1.07	1.23	0.84									0.97	1.15	1.23	1.16	1.79	3.49	3.50	1.68
tas	1.08	0.42	0.42	0.80	0.63	0.97	0.92	0.69	0.79	0.62	0.70	0.75	0.35	1.06	0.70	0.30	1.64	1.35	1.44	1.60	1.66	1.59	1.23	0.80
clt	0.90	1.16	1.19	1.07	0.70	0.76	0.66	0.78	0.85	0.78	0.61	0.68	0.91	0.32	0.56	0.74	0.80	0.99	0.87	0.79	0.97	0.97	0.83	0.82
pr	0.77	0.87	1.02	1.07	0.87	1.22	1.10	0.91	1.11	1.00	0.90	0.84	1.38	0.91	0.96	1.00	1.17	0.73	1.10	1.08	1.11	1.19	1.04	0.79
rlut	1.02	0.88	0.61	0.49	1.01	0.67	0.63	0.90	1.21	1.04	0.95	0.92	1.52	0.86	0.80	1.04	0.54	0.53	0.59	0.63	1.30	1.44	0.69	0.80
uas	0.65	0.85	0.64	0.91	0.70	0.84	0.56	0.67	0.98	0.80	0.81	0.70	1.34	1.12	0.40	0.70	0.79	0.67	0.78	0.41	0.50	0.45	0.40	0.41
vas	0.62	0.80	0.69	0.82	0.73	0.74	0.74	0.65	0.95	0.80	0.81	0.74	1.21	1.25	0.81	0.65	0.88	0.93	0.68	0.47	0.52	0.50	0.39	0.44
300hPa ua	0.75	0.99	0.95	1.23	0.94	1.23	0.86	1.12	0.95	0.73	0.77	0.82	0.60	0.85	0.38	0.30	0.67	0.91	1.02	0.46	0.91	0.85	0.75	0.76
500hPa zg	0.29	0.62	1.31	1.05	0.41	0.85	0.78	0.63	0.38	0.27	0.61	0.31	0.26	0.22	0.67	0.31	0.94	0.67	0.46	0.74	0.62	0.34	0.18	0.77
SD zos	0.66	0.42	0.61	0.61	0.88	0.93	0.90	0.87	1.02	1.02	1.06	1.07	1.35	1.44	1.71	1.70	1.02	1.05	1.08	1.04	0.95	0.82	0.97	0.97
SD tos	1.06	1.04	0.97	1.06	0.81	1.77	1.49	0.97	1.07	1.10	0.89	1.20	0.28	0.19	0.42	0.52	1.06	0.96	1.01	1.12	0.84	1.73	1.67	0.77
mlotst	1.25	0.55	0.70	1.09	2.32	0.64	0.86	1.85	1.41	1.02	1.54	1.01	0.61	0.66	0.89	0.62	0.51	1.89	2.50	1.08	1.96	2.85	2.81	3.64

ity of the model components, allowing for higher resolutions at equal computational cost.

With this overview in mind we will limit our model bias analysis to problematic areas, knowing that for the variables we do not focus on we achieved good model performance.

4.3.1 Sea ice and mixed-layer depth

The sea ice thickness in the Arctic contains realistic values for both the end of summer (EOS) and the end of winter (EOW), PICT and HIST runs (Fig. 4). In the central Arctic, PICT sea ice thickness ranges from 3.5 m at the EOW to 2.5 m at the EOS. The mean over the last 25 years of the HIST simulation reveals a reduction in sea ice thicknesses to 2.5 and 1.5 m in these two seasons. In both simulations, the maxima of EOW ice thickness can be found in the East Siberian Sea and along the coastline of northern Greenland. Although, in some years, a local maximum of sea ice thickness was observed along the East Siberian Sea coast, in the multi-year-mean field it should not be as large as pre-

sented by the model. Similar issues are common in many CMIP6 models and exist even in the PIOMAS reanalysis (Watts et al., 2021). One clear bias in comparison to observations is a too wide tongue of sea ice extending eastwards in the Greenland Sea during the winter months. A similar feature can be seen in our previous model versions. The annual cycle of sea ice extent is represented well with 16×10^6 to 7×10^6 km² in the last 25 years, compared to observational values of 15×10^6 to 6×10^6 km² by Walsh et al. (2019).

The Antarctic sea ice biases require the most improvement as follows from the metrics presented in Table 1. In both the PICT and HIST simulations the sea-ice-covered area is strongly underestimated during austral winter (Fig. 4c and g). Notable are two spots of low sea ice thickness in the Weddell Sea and in the eastern Ross Sea. Both areas feature low EOW mean sea ice thickness of less than 20 cm during the PICT run and are partially ice-free during the last 25 years of HIST. These locations feature persistent large-scale polynyas. In reality, polynyas were observed in the Weddell Sea (e.g., during the winters 1974 to 1976); however, the frequency of
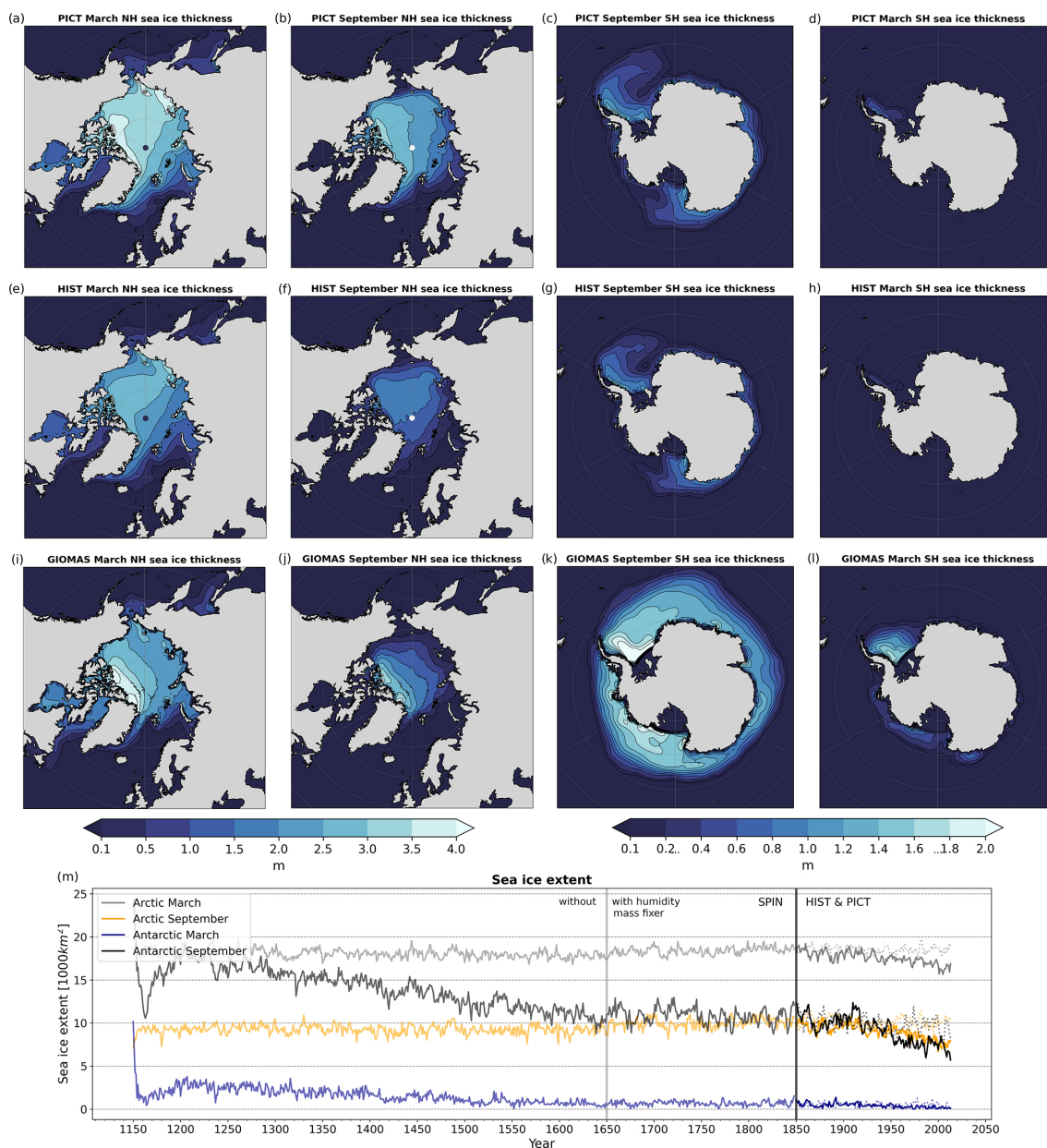
Figure 4. (a–d) Mean sea ice volume per unit area over last 25 years of the PICT simulation. **(e–h)** Same as top row but for the HIST simulation. **(i–l)** Same as top row but for the GIOMAS sea ice thickness reanalysis product. (Zhang and Rothrock, 2003). **(m)** Evolution of sea ice extent over SPIN, as well as HIST and PICT simulations. Mass fixer switched from dry to total in years 1650. HIST and PICT forcing applied from 1850 onwards.

their occurrence (not presented in this paper) is clearly over-estimated in our model.

In order to gain an understanding of the reasons for the per-sisting polynya, we investigate the mean mixed-layer depth (MLD), defined as the depth at which the potential density differs by $0.125 \, \mathrm{kg \, m^{-3}}$ from the surface density (Monterey and Levitus, 1997). The MLD shown in Fig. 5a features large values in the Weddell and increased values in the Ross seas that are co-located with low sea ice concentration values seen

in Fig. 5b. We therefore speculate that the large MLD could be one of the reasons for the underestimated wintertime sea ice, as the ocean heat from the warmer Circumpolar Deep Water (CDW) can be mixed up to reach sea ice from be-low. The salinity profile over the highlighted region confirms that the surface layer is more saline in our model than in the PHC3 climatology. The exact reason for the overestimated MLD is not clear, and understanding whether the salinity bias is the main cause is among our planned future research.
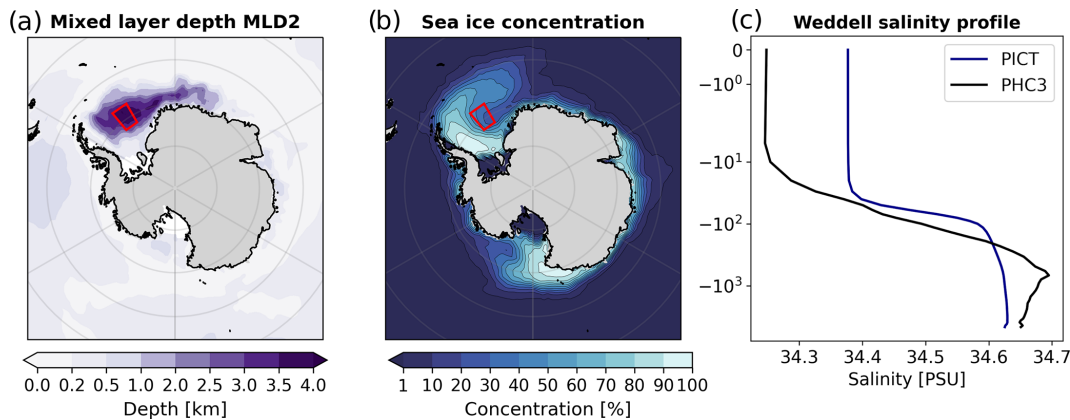
Figure 5. (a) Austral winter mean mixed-layer depth (MLD) where the potential density over depth differs by $0.125\,\mathrm{kg\,m^{-3}}$ from the surface density (Monterey and Levitus, 1997) averaged over the last 25 years of the PICT simulation. **(b)** High MLD values in the Weddell and Ross seas contribute to the persistent polynyas visible in the mean austral winter sea ice concentration over the same time period. **(c)** Salinity profile over the region marked in blue in comparison to PHC3 climatology (Steele et al., 2001).

We hypothesize that in the subsequent austral summer the reduced sea ice cover results in a strong positive sea ice–albedo feedback, further heating up the surface ocean. We examine this feedback in the following section.

4.3.2 Surface temperature and fluxes

The near-surface (2 m) air temperature bias in comparison to ERA5 largely shows a pattern similar in sign (Fig. 6a) and at somewhat higher amplitude compared to the CMIP6 multi-model mean bias shown in Bock et al. (2020). In the tropical Pacific we find a slight cold bias located at the Equator, flanked by equally sized warm biases, which is typical of a too strong double ITCZ found in many coupled models and associated with characteristic precipitation and short-wave radiation biases (see below). The upwelling regions off the west coasts of southern Africa and South America feature well-known warm biases of up to 4 °C resulting from insufficient upwelling, too little stratocumulus cloud cover, and thus too much downwelling short-wave radiation with our low-resolution ocean model. Further north and south the subtropical gyres feature cold biases on the order of −1 to −2 °C.

In the mid-latitudes, the most prominent bias is the misplacement of the Gulf Stream, which fails to represent the correct northwest corner detachment (Fig. 6a). This is a well-known bias in OGCMs of coarser-than-10 km resolution in the North Atlantic. For the FESOM2 model, a study by Sein et al. (2017) shows that this problem can be mitigated by increasing the ocean model resolution in the region.

At high latitudes we find a strong warm bias of more than +8 °C over areas of the Southern Ocean (SO) adjacent to Antarctica and a moderate one over Antarctica of +3 to +4 °C, as well as a cold bias of around −2 °C in the Arctic. The cold bias in the Arctic likely stems from the ERA5 reanalysis, which misses snow cover on sea ice (Batrak and

Müller, 2019), rather than from AWI-CM3. We thus focus on the southern warm biases in our analysis, some of which are well known for IFS-based climate models and can also be seen in Döscher et al. (2022) and Roberts et al. (2018). We hypothesize that these biases are caused by the direct effect of heat released from a spuriously deep mixed layer (as noted in Sect. 4.3.1), a positive ice–albedo feedback to the resulting reduced sea ice cover, and a remaining positive net shortwave downward heat flux bias between 45–60° S.

Nearly all of the near-surface temperature biases (Fig. 6a) can be associated with co-located net surface shortwave radiation biases (Fig. 6c). These can be largely explained by surface downward radiation biases (Fig. 6e), which in turn result from total cloud cover fraction biases (Fig. 6f). This linkage pattern is well known in the modeling community (Satoh et al., 2019) and will likely remain a dominant source for surface temperature biases until deep-convection-resolving atmospheric models become readily available for multi-decadal to centennial climate simulations.

There are two notable exceptions to this causal chain in our simulations. Firstly, we found a cold bias over the Greenland Sea where the surface downward shortwave radiation bias is positive. The cold bias in the region is the result of a lobe of sea ice drifting into the area from the Greenland coast during winter. This results in an overestimation of surface albedo and a reduction in the net surface shortwave radiation.

The second and similar exception is the previously mentioned warm bias south of 60° S in the SO. While the model overestimates the cloud and thus underestimates surface downward shortwave radiation in the region, the sea ice fraction in this area is too low. The surface net short-wave radiation and the near-surface temperature biases are therefore positive. As the SO surface net shortwave radiation bias is negative, the low sea ice concentrations can not originally be caused by shortwave biases. Indeed, the near-surface
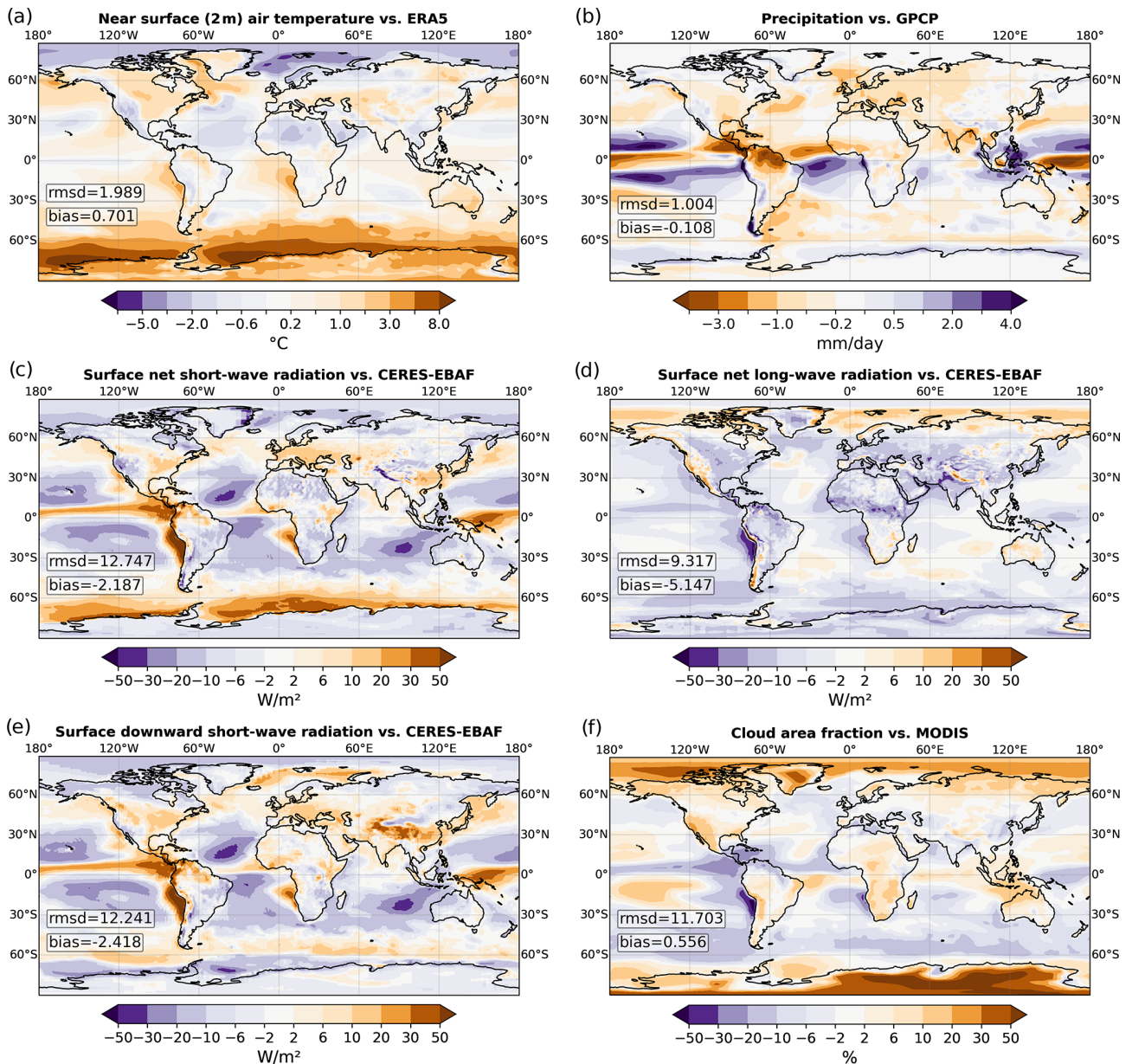
Figure 6. Annual mean **(a)** near-surface (2 m) air temperature bias with respect to ERA5, **(b)** precipitation bias with respect to GPCP, **(c)** surface net shortwave radiation bias with respect to CERES (2001–2014), **(d)** surface net longwave radiation bias with respect to CERES (2001–2014), **(e)** surface downward shortwave radiation bias with respect to CERES (2001–2014), and **(f)** total cloud cover fraction bias with respect to MODIS. For each variable rmsd gives the area-weighted root mean square distance between observations and model data, and bias gives the area-weighted mean distance.

(2 m) air temperature and sea ice concentration biases south of 60° S peak during the austral winter season, while the biases in the southern mid-latitudes peak during austral summer (Table 1). Correcting the largest biases south of 60° S in our model will probably necessitate work on non-solar heat fluxes and mixed-layer depths at high latitudes.

The precipitation biases shown in Fig. 6b feature the canonical double ITCZ bias in the tropics. Notably the mid-latitudes receive too little precipitation, especially over west-

ern boundary currents, Europe, and North America east of the Rocky Mountains.

Figure 7a depicts the zonal-mean temperature bias averaged over the last 25 years of the HIST simulation with respect to ERA5 for the same period. The previously seen Southern Ocean surface warm bias is well visible up to 500 hPa. Around the tropopause height the model exhibits a prominent cold bias of up to −4 °C. In the high stratosphere meanwhile we find a large warm bias of up to 8 °C.
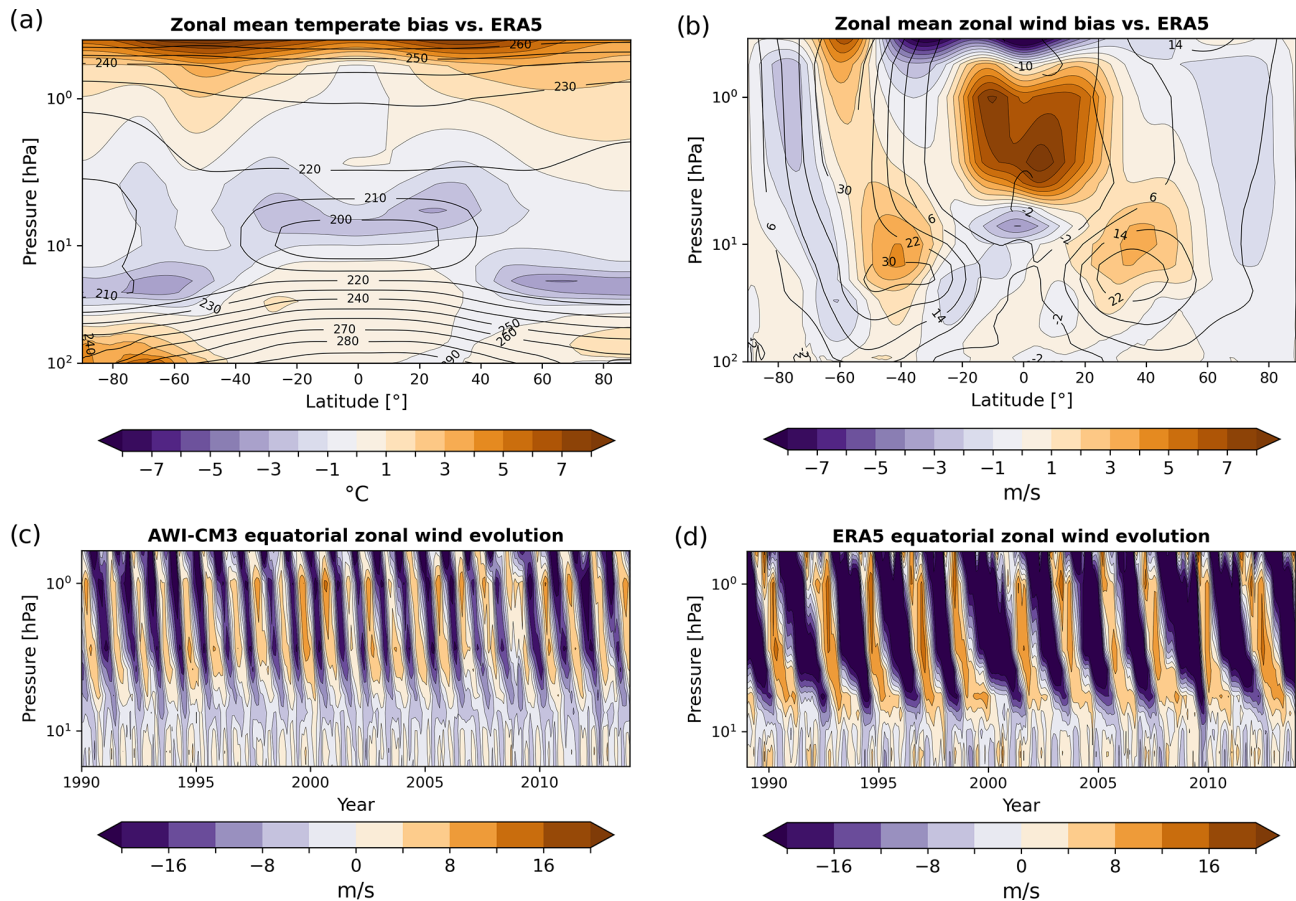
Figure 7. AWI-CM3 HIST **(a)** zonal-mean air temperature bias with respect to ERA5 averaged over the last 25 years; **(b)** zonal-mean zonal-wind bias with respect to ERA5 averaged over the last 25 years; **(c)** near-Equator zonal wind over time and height for AWI-CM3 HIST simulation; and **(d)** same as **(c)** but for ERA5 reanalysis, showing the Quasi-Biennial Oscillation.

The stratospheric bias will be reduced with ECMWF's next minor release of OpenIFS, cy43r3v2, which enables the use of spectral solar insolation instead of total solar irradiance (not shown).

The zonal-mean zonal wind in Fig. 7b broadly shows overestimation of the subtropical jet strength and height, as well as underestimation of the polar jet streams. The most notable feature though is the misrepresentation of equatorial stratospheric winds.

A closer look at the near-Equator (10° N–10° S) zonal- and meridional-mean wind speeds over height and time in Fig. 7c reveals that the Quasi-Biennial Oscillation (QBO) is in fact an annual oscillation in the last 25 years of the HIST simulation. Contrasted with the same winds in the ERA5 reanalysis of the equivalent time period, the QBO frequency error is apparent. The main reason for this behavior is a lack of tuning for the gravity wave flux parameterization, as was confirmed by tests conducted with OpenIFS at GEOMAR (not shown). Tuning of the gravitational wave flux parameterization will be addressed in future model releases.

4.3.3 Ocean temperature

On the ocean side of the coupling interface, we find mean sea surface temperature biases with respect to PHC3 that are nearly identical to the aforementioned near-surface air temperature biases (Fig. 8). At a depth of 100 m the bias looks similar to that at the surface in high latitudes where mixed-layer depths are large. In the tropics and subtropics this depth range is dominated by cold biases, which could be partially due to missing vertical mixing associated with Langmuir circulation in the current version of FESOM2. It is known that including a parameterization for this mixing can effectively alleviate the cold bias in the near-surface ocean in the mid-latitudes and tropics (Wang et al., 2019; Ali et al., 2019), and the integration of a Langmuir circulation parameterization in vertical mixing schemes is planned.

At a depth of 1000 m the Atlantic shows a strong warm bias, which we speculate results partially from spuriously warm SO surface water entrained into the AAIW. Ultimately the whole bias probably stems from multiple yet-to-be-identified sources in both hemispheres. The cold bias
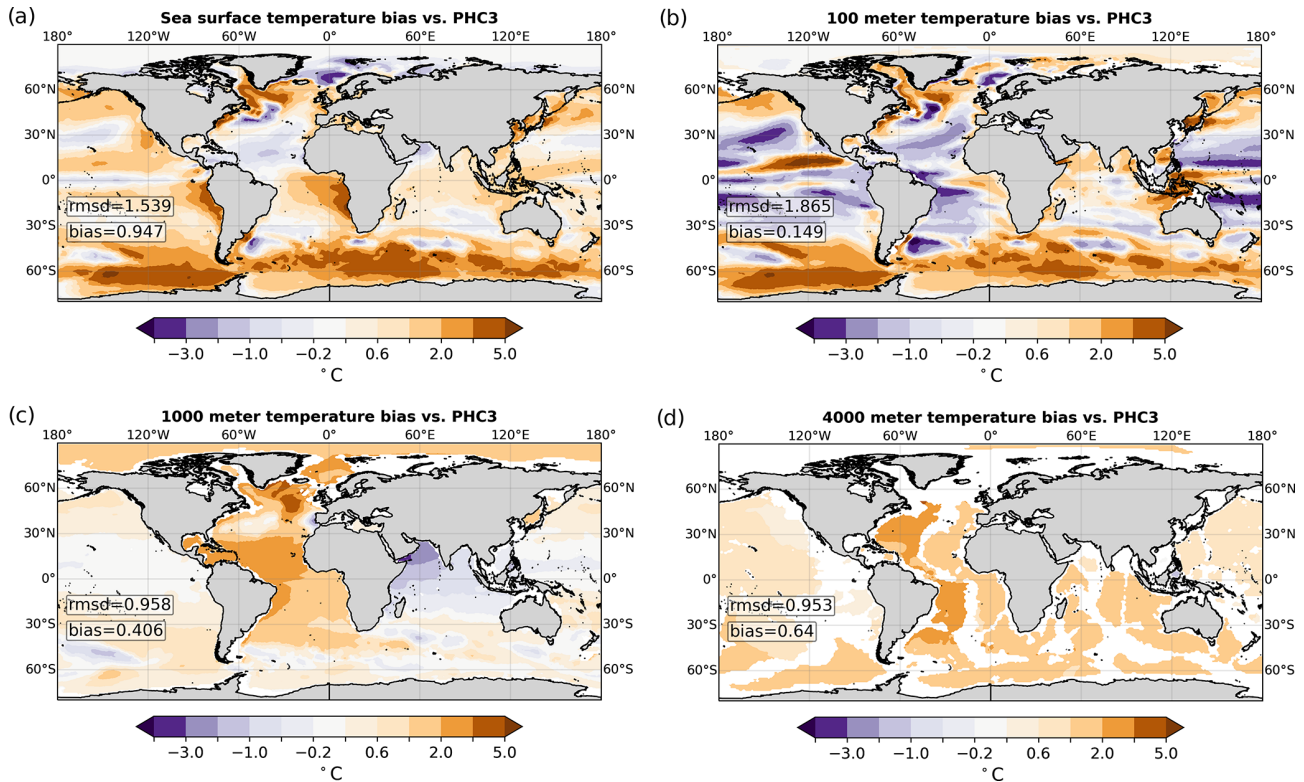
Figure 8. Mean ocean temperature bias of the last 25 years of the historical simulation with respect to PHC3 at (**a**) 0, (**b**) 100, (**c**) 1000, and (**d**) 4000 m depth. For each depth rmsd gives the area-weighted root mean square distance between observations and model data, and bias gives the area-weighted mean distance. At the surface the ocean-resolution-specific warm biases in upwelling regions and western boundary currents can be seen, in addition to a pronounced Southern Ocean warm bias. At depth, a warm bias dominates in the Atlantic.

found in the Indian Ocean at this depth is likely the result of insufficient warm outflow from the Persian Gulf because the narrow strait is not well resolved on the FESOM2 CORE2 mesh.

AWI-CM3 shows a pronounced warm bias of 3–5 K in the Atlantic subpolar gyre. Sidorenko et al. (2015) noted a similar bias when analyzing AWI-CM1 simulations. This bias is shared between many climate models which contributed to CMIP. Hence, a similar drift in ocean hydrography is also described in Sterl et al. (2012), Delworth et al. (2006, 2012), and Jungclaus et al. (2013). These authors discuss different factors that may be responsible for the bias. Sterl et al. (2012) show that overestimation of the Mediterranean outflow can significantly increase the deep-ocean salinity bias. Delworth et al. (2012) attribute this anomaly to the insufficient eddy transport required to compensate for the wind-driven subduction in the subtropical gyres. They show that moving towards an eddy-resolving setting or a parameterization of the eddy stirring reduces the temperature biases significantly. Jungclaus et al. (2013) suggest that part of the problem arises from the improper interbasin exchange between the Indian and South Atlantic oceans.

At 4000 m depth we observe the warm biased Atlantic water mass spreading into the whole global ocean. As the merid-

ional circulation at this depth is northwards in the Antarctic Bottom Water (AABW) cell, the origin of the warm bias is presumably insufficient cold AABW formation in the Southern Ocean on this coarse mesh. The slowness of the AABW circulation in the Atlantic coincides with the slow but continued global-mean warming trends at great depths seen in Fig. 3. Biases presented in Fig. 8 explain the Hovmöller diagram in Fig. 3, which we addressed in Sect. 4.1. The three sets of anomalies in the Hovmöller diagram at depths 100, 1000, and 4000 m stem from the biases in the mid-latitudes, the North Atlantic, and the entire ocean, respectively.

4.3.4 El Niño–Southern Oscillation

To assess the model capabilities in representing climate variability, we investigate the representation of the El Niño–Southern Oscillation (ENSO). Using the entire HIST simulation and the EOF toolbox (Dawson, 2016), Fig. 9a depicts the spatial extent of the correlation between the first principal component and the input dataset at each point in space. Also depicted is the Niño 3.4 box (5° N to 5° S, 120 to 170° W) upon which we based our further analysis.

We calculated the area mean SST within the box, applied a linear detrending, subtracted the mean seasonal cycle, and

finally computed a 3-month running mean. The resulting Niño 3.4 index time series is shown in Fig. 9c. Comparison with the Niño 3.4 index based on the HadISST observational estimates (Fig. 9d) reveals that our simulated amplitude of ENSO, with a range of $+1.9$ to $-1.6\,°C$, is lower than that observed. Indeed, the histograms in Figs. 9e and 9f reveal missing tails of the distribution in the simulation.

While the overall strength of ENSO is underestimated by AWI-CM3, the normalized power spectrum density (PSD) in Fig. 9b shows that the frequencies of the observed HadISST ENSO are well reproduced. We find several peaks concentrated between periods of 2.8 and 12 years. Further analysis (not shown) indicated that ENSO phase locking is rather weak in the model version presented and that both ENSO amplitude and phase locking can be improved through reduction in equatorial Pacific precipitation and temperature biases. Understanding how to do so without negatively impacting model performance in other regions is work in progress.

Future improvements for the ENSO amplitude could also potentially be achieved by activating the OpenIFS internal stochastic parameterization schemes, as shown by Yang et al. (2019). Indeed, HighResMIP simulations performed by Roberts et al. (2018) with IFS CY43 that employed the stochastically perturbed parameterization tendencies (SPPT) scheme (Buizza et al., 1999) had an ENSO amplitude more in line with observational estimates. Since these studies used the SPPT scheme, they required additional humidity mass fixers for climate-length integrations.

Alternatively, in order to use the new – and from a mass conservation perspective more favorable – stochastically perturbed parameterization (SPP) scheme (Ollinaho et al., 2017), either the full version of SPP has to be backported from IFS CY47 to OpenIFS CY43R3, or the new version OpenIFS CY47 has to be released. Noteworthy is also that CMIP6 simulations done with CNRM-CM6-1, featuring the IFS atmosphere with the ARPEGE physics package (not in use AWI-CM3, as we employ ECMWF physics) and no stochastic parameterizations, also had a high ENSO amplitude. However the frequency of ENSO was too focused with a sharp single peak in the 3–4-year band (Voldoire et al., 2019).

Finally experiments with stochastic coupling at the atmosphere–ocean interface conducted with AWI-CM1 yielded improved ENSO phase locking (Rackow and Juricke, 2020), providing another potential development avenue.

4.4 Impact of historical forcing

4.4.1 Air temperature

After we have established that AWI-CM3 behaves reasonably for much of the globe and many important climate parameters, we will now characterize the impact of historical greenhouse gas and solar forcing on the evolution of several of these variables.

The global-mean near-surface (2 m) air temperature increases under HIST forcing, as seen in Fig. 10a. Since our SPIN experiment did not establish a full equilibrium in the deep ocean, we analyze the air temperature change in our pre-industrial control experiment PICT and obtain the residual drift of $0.00091\,°C\,yr^{-1}$ via linear regression. Over the 165-year-long simulation this amounts to a drift of $0.15\,°C$ in the PICT run.

In Fig. 10b we have corrected both runs by deducting the linear trend. The resulting temperature change represents a global warming over the period 1850 to 2014 of $1.4\,°C$, most of which comes in a steep rise between 1960 and 2000. Comparison with observational values from HadCRUT5 (Morice et al., 2021) shows that AWI-CM3 gets the timing of historical warming spikes right; however the strength is overestimated by 40 % (1.0 vs. $1.4\,°C$).

The main reason for the stronger-than-observed historical warming is that the increase in global aerosol emissions, which partially masks the warming induced by well-mixed greenhouse gases, is not incorporated in the version of OpenIFS (CY43R3) used here. The integration of the tropospheric aerosol component M7 into OpenIFS CY43R3 is still ongoing within the EC-Earth consortium (personal communication with the EC-Earth aerosol working group, 2022). The sixth IPCC report (Masson-Delmotte et al., 2021) concludes that the anthropogenic aerosols direct and indirect contributions to the effective radiative forcing amount to about $-0.5\,W\,m^{-2}$. Compared to the total anthropogenic forcing of $+1.5\,W\,m^{-2}$, we foresee that, once transient aerosols are included in OpenIFS CY43R3, AWI-CM3 historical global warming will be much closer to the observed value.

The map of temperature changes in Fig. 11a shows that under historic well-mixed greenhouse gas and solar forcing, AWI-CM3 simulates a temperature increase of approximately $1–2\,°C$ in the tropics, Antarctic, and large parts of Eurasia. Mid-latitude oceans in both hemispheres see a weaker warming of $0.6–1\,°C$. Larger warming of $2–3\,°C$ can be found in the Middle East, North America, the Caucasus, Australia, and South Africa. Sea-ice-covered regions in the Arctic experience the strongest air temperature increases, with the Arctic (65–90° N) warming by $3–8\,°C$. Defining an Arctic amplification index (AAI) as the ratio of warming north of 65° N to the whole Northern Hemisphere warming (Davy et al., 2018; Johannessen et al., 2016), the resulting AAI is 2.87. Note that, in contrast to observations as well as full-forcing CMIP simulations, there is no trace of a warming hole in the North Atlantic south of Greenland. Whether this might be due to the missing transient aerosol forcing in our simulation is unclear. Earlier studies have linked the warming hole rather to a weakening of the Atlantic meridional overturning circulation (AMOC; Keil et al., 2020), but although our historic simulation does exhibit such an AMOC weakening (see Fig. 12), no warming hole forms.
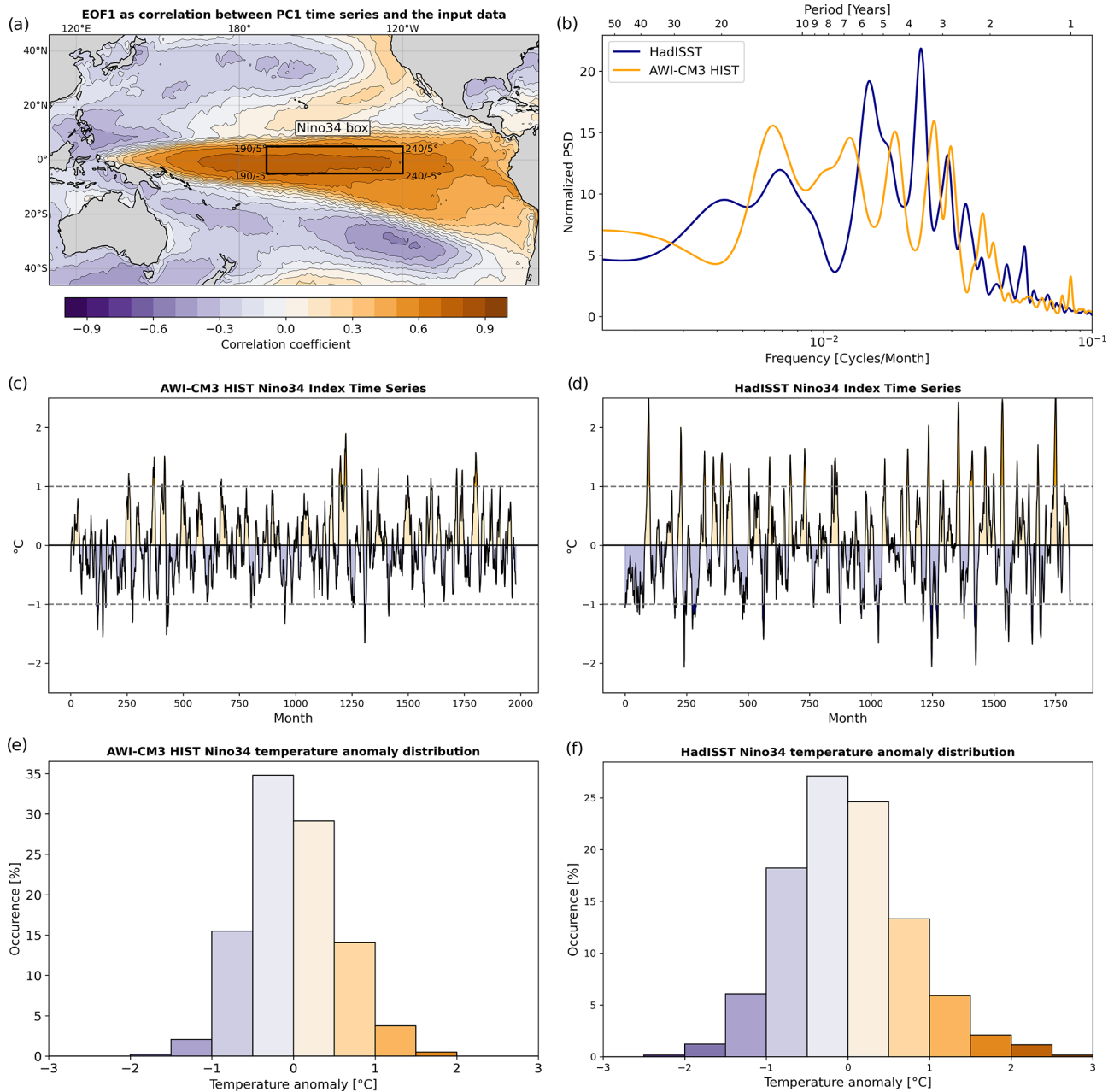
Figure 9. (a) Extent of the Niño 3.4 box, as well as the first empirical orthogonal function expressed as the correlation between the first principal component and the input dataset at each point in space. Calculated from the entire HIST simulation using the EOF toolbox (Dawson, 2016). **(b)** Normalized power spectrum density of simulated ENSO within the Niño 3.4 box in comparison to HadISST observational estimates for the years 1850 to 2014. **(c, d)** Niño 3.4 SST index defined as the detrended seasonal cycle removed and 3-month running mean time series of SST averaged over the Niño 3.4 box. The results from AWI-CM3 HIST and HadISST observational estimates are shown, respectively. **(e, f)** As **(c)** and **(d)** but depicting occurrence of temperature anomalies in bins with a width of 0.5 °C.

4.4.2 Precipitation

Figure 11b shows the simulated changes in the precipitation pattern resulting from historic well-mixed greenhouse gas and solar forcing. The most important features are the following: the high latitudes nearly uniformly receive more precipitation; the monsoonal precipitation in North Africa and China intensifies; the ITCZ is enhanced in the western Pacific and more focused on the Equator in the eastern Pacific and in the Atlantic; considerable parts of the subtropics tend to receive less precipitation. These patterns are largely consistent with precipitation changes simulated in CMIP6 models
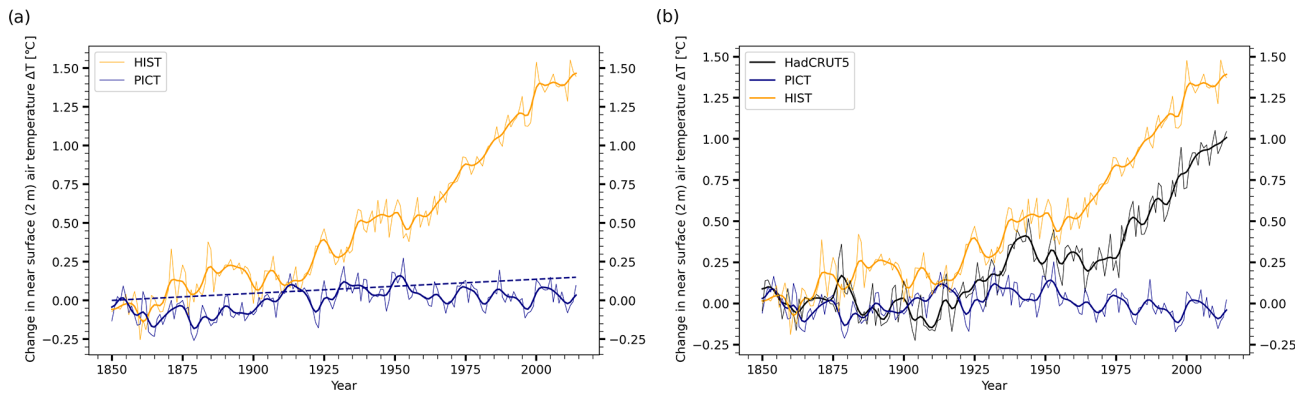
Figure 10. (a) Near-surface (2 m) air temperature changes over the historic period and in the pre-industrial control simulation. Thick lines show the 10-year running means. The pre-industrial control simulation shows a residual drift of 0.00091 °C yr^{-1} obtained from linear regression. **(b)** Both simulations have been corrected by subtracting the residual drift. The resulting simulated global warming over the period 1850 to 2014 is 1.4 °C. In comparison to observational estimates from HadCRUT5 (Morice et al., 2021) warming is overestimated by 0.4 K, likely due to fixed aerosols in the AWI-CM3 prototype.
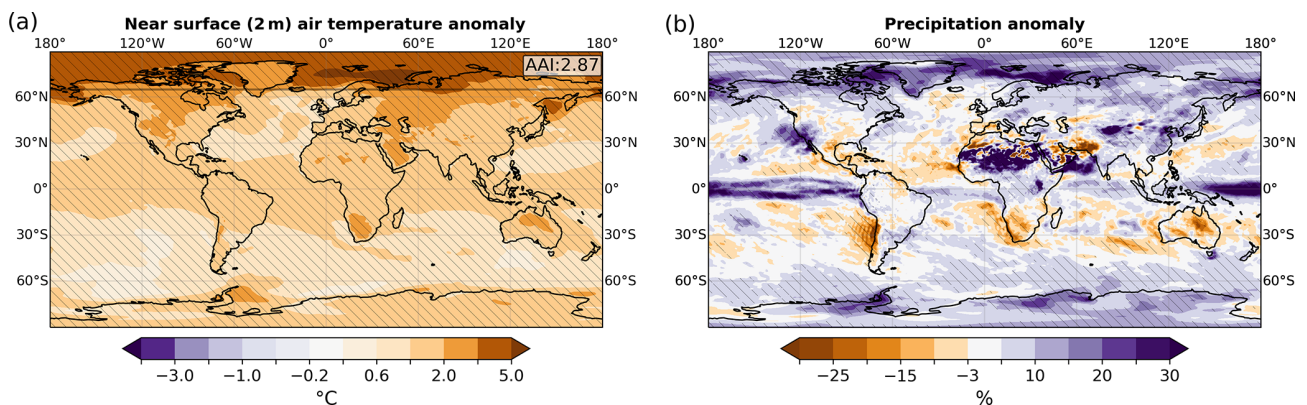
Figure 11. (a) Temperature anomaly between last 25 years of the historic simulations and the equivalent period of the pre-industrial control run. The Arctic amplification index, defined as the ratio of warming north of 65° N expressed as a fraction of global warming, is 2.87. Significant temperature anomalies are hatched, and the 95% significance is obtained via bootstrap testing. **(b)** Same as **(a)** but for relative precipitation changes.

where transient aerosols are included, although precipitation increases over the Indian Ocean and northern Central Africa are not as pronounced as in the CMIP6 model mean and more pronounced over the Indonesian warm pool (compare with Fig. SPM.5c in Masson-Delmotte et al., 2021).

4.4.3 Ocean circulation

The AWI-CM3 simulations reasonably reproduce the canonical pattern of the AMOC streamfunction, with an upper cell consisting of northward surface flow as well as southward return flow of North Atlantic Deep Water and a lower cell representing the northward flow of AABW (Fig. 12a and b). During the spinup simulation, the maximum of the northward transport between 30–45° N in AWI-CM3 fluctuates at around 20 Sv (Sv denotes sverdrups; see Fig. 12c). Initially the AMOC gradually slowed down to 18 Sv while the up-

per ocean was spuriously warming, as described in Sect. 4.1. After applying the total mass fixer starting 500 years into the SPIN experiment the AMOC is recovered to 20 Sv along with the cooling of the upper ocean (Fig. 3b).

Figure 12d shows a rather strong decline in the AMOC strength over the last 70 years of the HIST simulation, while the multi-decadal natural variability in the AMOC is still high. As we do not have additional ensemble members, we do not make strong statements about the impact of historic forcing on the AMOC. However, we can conclude that our simulated AMOC response to historical forcing is consistent with recent synthesis based on observations (Caesar et al., 2022), and the ability for rapid AMOC state changes, as described in Ollinaho et al. (2017), does seem to exist in AWI-CM3 since our low resolution is indeed higher than their high-resolution case. While the upper cell diminishes con-
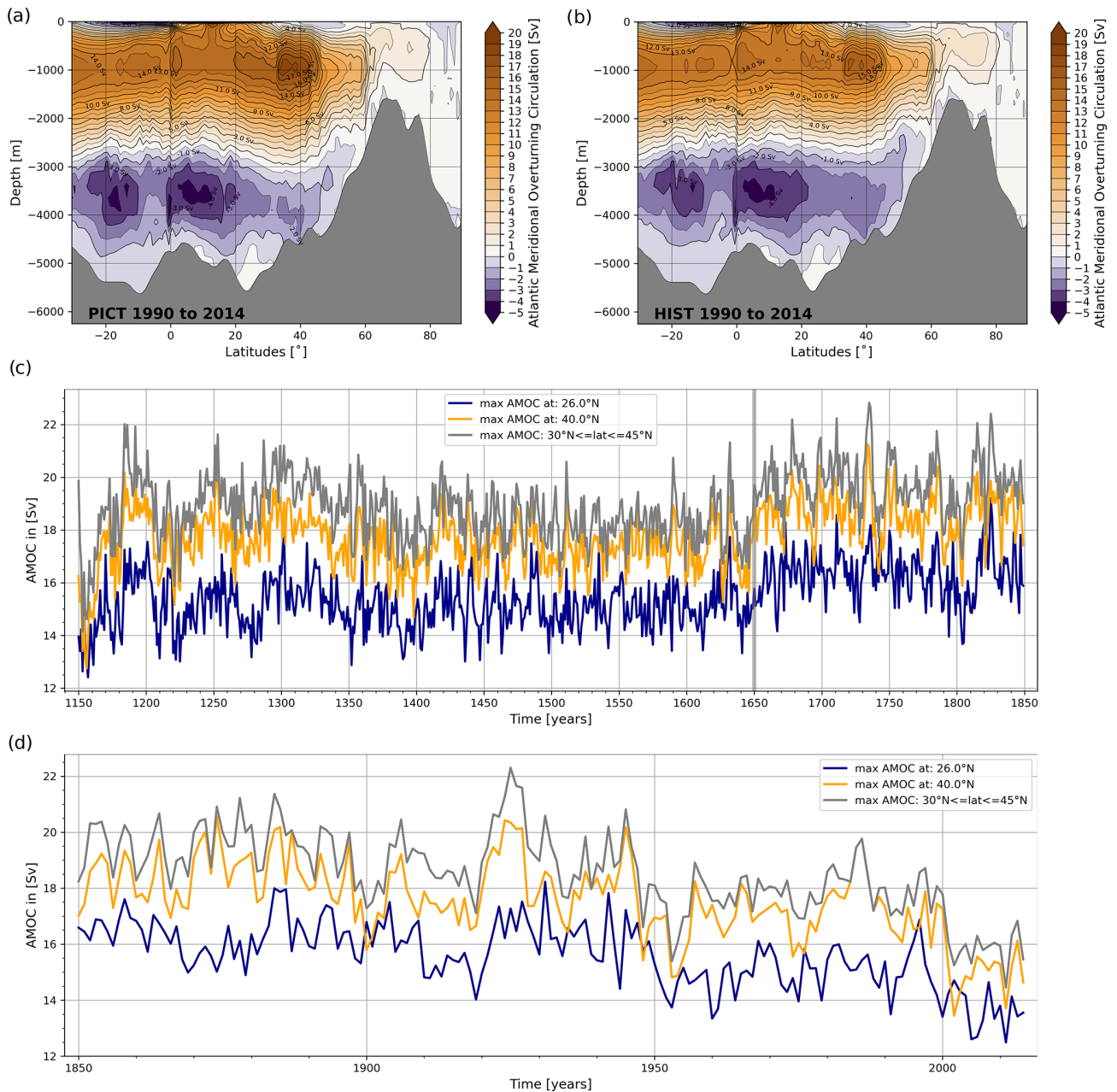
Figure 12. (a) Atlantic meridional overturning circulation (AMOC) streamfunction averaged over the last 25 years in the pre-industrial control (PICT) simulation. (b) As (a) but for the historic (HIST) simulation. (c) Evolution of the AMOC maximum at different latitudes throughout the spinup simulation (SPIN) that precedes both PICT and HIST. (d) Evolution of the AMOC during HIST.

siderably under historic forcing, the lower one is mostly unaffected.

AMOC variability across all latitudes is well correlated, pointing to the same (northern) origin of the signal. In the HIST run the decrease in AMOC is found across all latitudes. The correlation is high but not perfect (see years between 1920 and 1930, for instance). This indicates that the recirculating cell, associated with physical and numerical mixing in the model caused by the advection operator, changes in

strength (Sidorenko et al., 2021). Further conclusions would require extra analysis in the density space and the isopycnal framework (as in Sidorenko et al., 2021), which we did not activate in this run. However, this extra analysis was done by Sidorenko et al. (2021), and they concluded that the North Atlantic Oscillation (NAO) was the main driver of AMOC variability.

Simulated volume transport fluxes were computed across several major ocean straits (Table C2). Historical values aver-

aged over 1985–2014 match well the observation-based estimates for most straits. Two exceptions are the Nares and Davis straits where the simulated transport is lower than in the observational estimates. Indeed, these very narrow straits are likely not well resolved on the CORE2 mesh. In the Arctic, the poleward inflow through the Barents Sea Opening and Bering Strait is somewhat overestimated and counterbalanced by a southward transport at the Fram Strait larger than in the observational estimates (yet all remain within the uncertainty range). The simulated Antarctic Circumpolar Current (ACC) transport through the Drake Passage is within the range of observations. While being lower than the most recent measurements by Donohue et al. (2016), our AWI-CM3 estimate is comparable to the CMIP5 multi-model mean (MMM) and larger than the CMIP6 MMM (Beadling et al., 2020).

Historical transports through major straits are thus satisfactorily reproduced in the present AWI-CM3 configuration. Future configurations with an increased-resolution, eddy-resolving ocean are expected to provide even more accurate results within narrow straits and in eddy-rich areas such as the Southern Ocean, where mesoscale activity is key in accurately depicting the ACC behavior (e.g., Rackow et al., 2022).

4.5 Climate sensitivity experiments

4.5.1 Equilibrium climate sensitivity (ECS)

To investigate the ability of AWI-CM3 to simulate warmer climate states, we conducted the 4xCO2 experiment which prescribes a sudden permanent increase in the CO_2 concentrations to 4 times (1137.27 ppm) the base value (284.32 ppm) from 1850. It can be used to analyze the model's inherent climate sensitivity. As shown by Gregory et al. (2004) the relationship between the change in net downward radiative flux and the change in near-surface (2 m) air temperature can be described by a linear regression model. The Gregory plot in Fig. 13a was computed from the 4xCO2 experiment in comparison to the pre-industrial control simulation. The first axis intersection point determines the instantaneous radiative forcing $F = 7.06\,\mathrm{W\,m^{-2}}$ resulting from the quadrupling of CO_2. The linear regression line intersects the second axis at an equilibrium temperature difference $\Delta T = 6.49\,°\mathrm{C}$, resulting in a climate response parameter of AWI-CM3 of $\alpha = \frac{\Delta T}{F} = 0.92\,\mathrm{K\,W^{-1}\,m^{-2}}$. Equilibrium climate sensitivity (ECS) is given by doubling rather than quadrupling CO_2 concentrations and is thus $\mathrm{ECS} = \frac{\Delta T}{2} = 3.2\,°\mathrm{C}$. With this ECS value the AWI-CM3 prototype finds itself near the center of the range predicted by CMIP6 models (1.8–5.6 °C) (Meehl et al., 2020; Scafetta, 2021) and is the same as AWI-CM1, which had a value of 3.2 °C as well.

4.5.2 Transient climate response (TCR)

We obtain the transient climate response of AWI-CM3 TCo159L91-CORE2 from an experiment that features increased CO_2 forcing by 1 % per year, starting from the 1850 value of 285 ppm. Radiative forcing thus applied results in a near-surface (2 m) air temperature increase of 2.1 °C after 70 years, when CO_2 concentrations had doubled, as shown in Fig. 13b. As with the ECS, the TCR of AWI-CM3 is also near the center of the CMIP6 model distribution of 1.8–5.6 °C (Scafetta, 2021). Compared to the predecessor model AWI-CM1, the TCR did not change. Interestingly, a further doubling to a total of 4 times the 1850 CO_2 concentrations until the year 1990 results in another 2.6 °C increase and thus a larger global-mean near-surface (2 m) air temperature rise than during the first period.

5 Computational performance

The computational performance of a climate model can be measured according to a variety of criteria. Balaji et al. (2017) provide a good overview of what can be considered the computational performance, but in our analysis we will focus on just two aspects, the simulated years per day (SYPD) and the computational cost measured in core hours per simulated year (CHSY). Systematic and rigorous experiment design would require that we identify all the degrees of freedom and vary them in all combinations. Unfortunately the number of degrees of freedom is large, including atmosphere and ocean vertical and horizontal resolution; atmospheric spectral and grid point resolution; and the number of cores allocated for MPI and/or OpenMP parallelization for FESOM2, OpenIFS43, and XIOS, as well as the amount of model output for each of the main components. Testing all combinations is impractical. Instead we present results for setups that have been optimized empirically, involving not only the use of analytical tools such as Dr.Hook (Saarinen et al., 2005), LUCIA (Maisonnave et al., 2020), and the XIOS internal statistics but also educated guesswork. It may well be that better configurations exist, but AWI-CM3 can at a minimum perform to the level presented here.

Tables 2 and 3 list the SYPD and CHSY values we achieved when optimization is performed for speed and cost, respectively. Note that the scaling limit of TCo319L137-DART has not been reached, and simulations with upwards of 10 SYPD are likely feasible.

5.1 High-resolution outlook

While we document mainly the first CMIP-prototype simulations of AWI-CM3 and its strengths and weaknesses at low resolution, we also tested higher-resolution configurations. In Table 4 we show the performance of a 31 km atmosphere with 137 vertical layers, coupled to a high-resolution ocean with a mesh that features 3.1 million surface nodes and
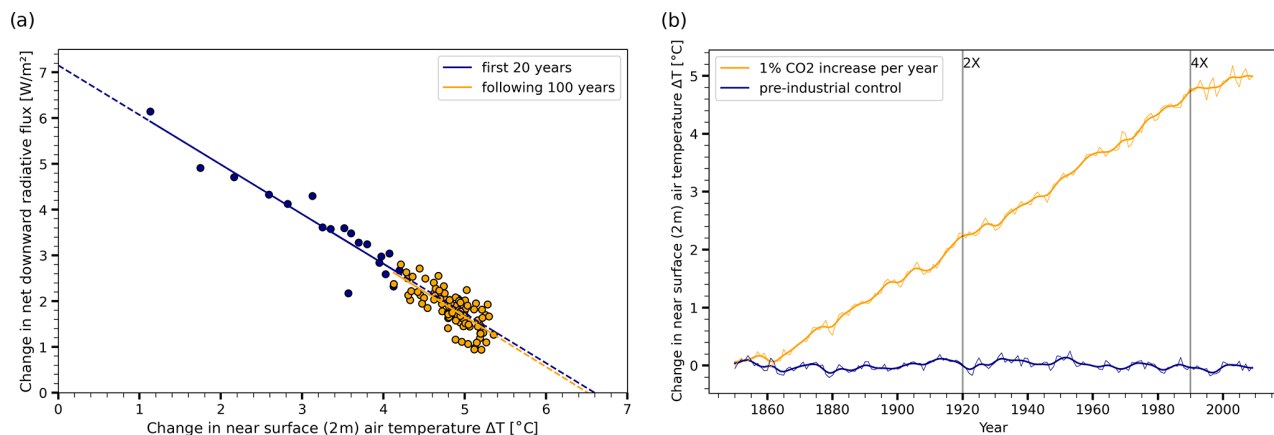
(a)

(b)

Figure 13. (a) Gregory regression plot from the abrupt 4xCO2 experiment in comparison to the pre-industrial control. We construct the linear regression of near-surface (2 m) air temperature change (ΔT) against the downward net radiative flux change (ΔF). From its axis intersection points we obtain the radiative forcing $\Delta F = 7.06\,\mathrm{W\,m^{-2}}$, and the equilibrium temperature difference $\Delta T = 6.49\,°\mathrm{C}$. The climate response parameter of AWI-CM3 is then $\alpha = \frac{\Delta T}{\Delta F} = 0.92\,\mathrm{K\,W^{-1}}$, and the equilibrium climate sensitivity $\mathrm{ECS} = \frac{\Delta T}{2} = 3.24\,°\mathrm{C}$. The first 20 years already results in a linear regression almost identical to the following 100 years. **(b)** As Fig. 10b but for the experiment with a 1 % increase in CO_2 per year. Vertical lines indicate doubling and quadrupling of CO_2 concentrations. The estimated transient climate response is 2.1 °C.

Table 2. Computational performance optimized for integration speed. Simulated years per day (SYPD) and core hours per simulated year (CHSY) of AWI-CM3 for three atmospheric grids, two ocean meshes, and two IO schemes are shown. Values were measured on HPC system JUWELS (at Jülich Supercomputing Centre) with Intel Xeon Platinum 8168 CPU, 2×24 cores, and 2.7 GHz processors, accepting sub-linear strong scaling. The atmospheric grids TCo95L91, TCo159L91, and TCo319L137 have a grid point resolution of 100, 61, and 31 km and 91, 91, and 137 vertical layers, respectively. The CORE2 mesh (Fig. 2a) has 47 vertical layers, $\sim 127\,000$ surface nodes, and a peak resolution of 20 km in northern middle to high latitudes and up to 125 km in subtropical gyres. The horizontal resolution of the DART mesh follows the local Rossby radius, peaks at a highest resolution of 5 km, and has a maximum spacing of 27 km, with ~ 3.1 million surface nodes and 80 vertical layers.

Atmosphere grid	Ocean mesh	IO scheme	Cores	SYPD	CHSY
TCo95L91	CORE2L47	Sequential	2593	124.05	501
TCo95L91	CORE2L47	XIOS parallel	2689	134.24	480
TCo159L91	CORE2L47	Sequential	2593	60.74	1024
TCo159L91	CORE2L47	XIOS parallel	2833	68.7	988
TCo319L137	DARTL80	XIOS parallel	15 680	7.90	47 528

80 layers. This configuration with the name TCo319L137-DART has been run for 50 years under constant 1990 forcing, with some of the insights gained while performing the set of experiments at low resolution already taken into account.

Key improvements in the TCo319L137-DART setup compared to TCo159L91-CORE2 are the eddy-resolving ocean in the western boundary currents, with eddy permittance in the ACC region, as well as better representation of orography and related effects in the atmospheric model. In lower-latitude regions the atmosphere has sufficient resolution to resolve and react to ocean eddies. Both the atmosphere and the ocean feature more vertical layers, allowing for better representation of vertical processes.

The ability of the TCo319L137-DART simulation to reproduce a climate as observed during the years 1990 to 2014 is better than that of the TCo159-CORE2 runs we have analyzed so far. The improvement is almost universal with only
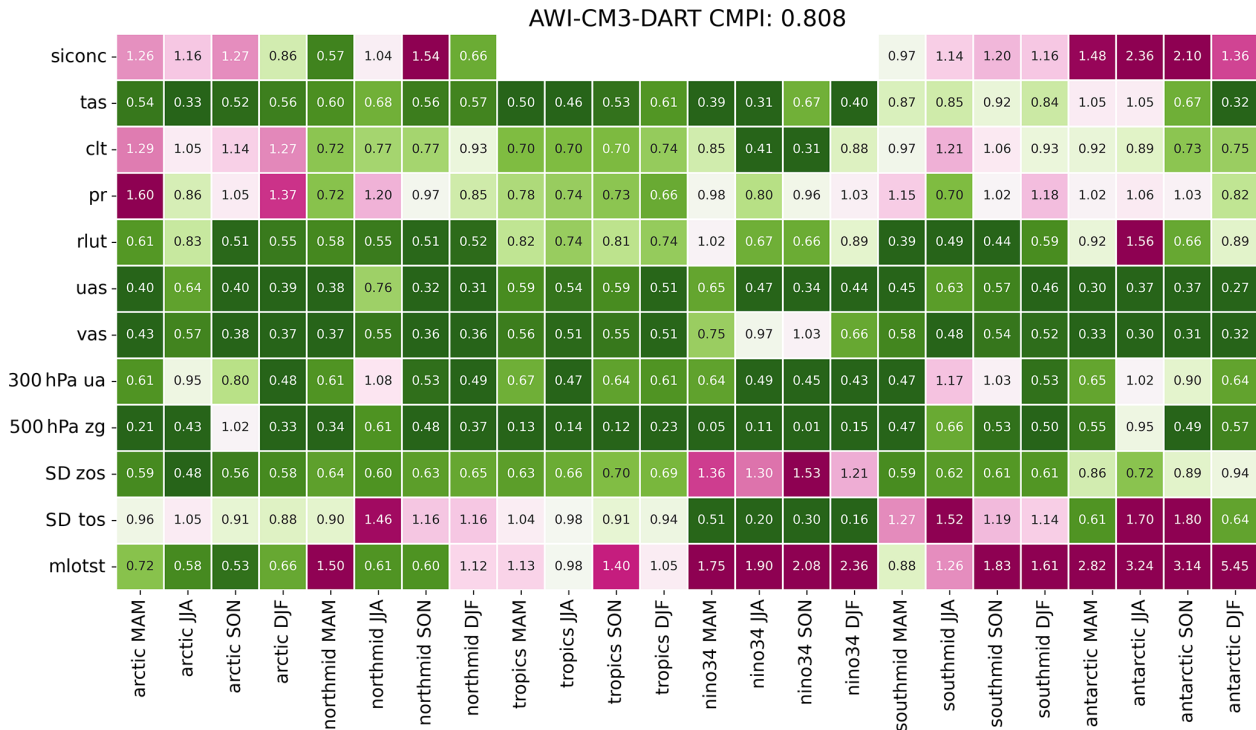
the sea ice concentrations and mixed-layer depths still showing below-average performance compared to CMIP6 models. A TCo159-CORE2 simulation of the same length and with the same forcing (not shown) showed performance almost identical to the last 25 years of the HIST experiment discussed above. We conclude that the improvement is related to the model resolution and continued model development, rather than to the shorter run length or different forcing.

Obviously the improved climatological performance comes at a cost, as detailed in Sect. 5. Every simulated year with TCo319L137-DART costs 35 times the CHSY and is performed at 15-times-lower SYPD. More detailed exploration of the higher-resolution capabilities of AWI-CM3 will be the subject of future work.

Table 3. As Table 2 but optimized for computational cost instead of integration speed by limiting total core numbers within the limit of linear strong scaling.

Atmosphere grid	Ocean mesh	IO scheme	Cores	SYPD	CHSY
TCo95L91	CORE2L47	XIOS parallel	721	52.34	330
TCo159L91	CORE2L47	XIOS parallel	1297	42.71	736
TCo319L137	DARTL80	XIOS parallel	6769	4.61	35 220

Table 4. Same as Table 1 but for the last 10 years of a 50-year-long TCo319L137-DART simulation under constant 1990 forcing. Biases compared to observations are not only smaller than in our TCo159L91-CORE2 simulations but also smaller than those of all but two (HadGEM3MM and NOAA-GFDL) of the CMIP6 models listed in Appendix A (performance indices for CMIP6 models not shown).

AWI-CM3-DART CMPI: 0.808

	arctic MAM	arctic JJA	arctic SON	arctic DJF	northmid MAM	northmid JJA	northmid SON	northmid DJF	tropics MAM	tropics JJA	tropics SON	tropics DJF	nino34 MAM	nino34 JJA	nino34 SON	nino34 DJF	southmid MAM	southmid JJA	southmid SON	southmid DJF	antarctic MAM	antarctic JJA	antarctic SON	antarctic DJF
siconc	1.26	1.16	1.27	0.86	0.57	1.04	1.54	0.66									0.97	1.14	1.20	1.16	1.48	2.36	2.10	1.36
tas	0.54	0.33	0.52	0.56	0.60	0.68	0.56	0.57	0.50	0.46	0.53	0.61	0.39	0.31	0.67	0.40	0.87	0.85	0.92	0.84	1.05	1.05	0.67	0.32
clt	1.29	1.05	1.14	1.27	0.72	0.77	0.77	0.93	0.70	0.70	0.70	0.74	0.85	0.41	0.31	0.88	0.97	1.21	1.06	0.93	0.92	0.89	0.73	0.75
pr	1.60	0.86	1.05	1.37	0.72	1.20	0.97	0.85	0.78	0.74	0.73	0.66	0.98	0.80	0.96	1.03	1.15	0.70	1.02	1.18	1.02	1.06	1.03	0.82
rlut	0.61	0.83	0.51	0.55	0.58	0.55	0.51	0.52	0.82	0.74	0.81	0.74	1.02	0.67	0.66	0.89	0.39	0.49	0.44	0.59	0.92	1.56	0.66	0.89
uas	0.40	0.64	0.40	0.39	0.38	0.76	0.32	0.31	0.59	0.54	0.59	0.51	0.65	0.47	0.34	0.44	0.45	0.63	0.57	0.46	0.30	0.37	0.37	0.27
vas	0.43	0.57	0.38	0.37	0.37	0.55	0.36	0.36	0.56	0.51	0.55	0.51	0.75	0.97	1.03	0.66	0.58	0.48	0.54	0.52	0.33	0.30	0.31	0.32
300 hPa ua	0.61	0.95	0.80	0.48	0.61	1.08	0.53	0.49	0.67	0.47	0.64	0.61	0.64	0.49	0.45	0.43	0.47	1.17	1.03	0.53	0.65	1.02	0.90	0.64
500 hPa zg	0.21	0.43	1.02	0.33	0.34	0.61	0.48	0.37	0.13	0.14	0.12	0.23	0.05	0.11	0.01	0.15	0.47	0.66	0.53	0.50	0.55	0.95	0.49	0.57
SD zos	0.59	0.48	0.56	0.58	0.64	0.60	0.63	0.65	0.63	0.66	0.70	0.69	1.36	1.30	1.53	1.21	0.59	0.62	0.61	0.61	0.86	0.72	0.89	0.94
SD tos	0.96	1.05	0.91	0.88	0.90	1.46	1.16	1.16	1.04	0.98	0.91	0.94	0.51	0.20	0.30	0.16	1.27	1.52	1.19	1.14	0.61	1.70	1.80	0.64
mlotst	0.72	0.58	0.53	0.66	1.50	0.61	0.60	1.12	1.13	0.98	1.40	1.05	1.75	1.90	2.08	2.36	0.88	1.26	1.83	1.61	2.82	3.24	3.14	5.45

6 Conclusions

We developed a new coupled climate model AWI-CM3, by coupling the AWI ocean model FESOM2, the ECMWF NWP atmosphere model OpenIFS CY43R3, a small runoff-mapper model, and the XIOS parallel IO library. The coupling exchange is achieved via the OASIS3-MCT_4.0 library.

We ran a set of experiments closely resembling the Coupled Model Intercomparison Project phase 6 (CMIP6) Diagnostic, Evaluation and Characterization of Klima (DECK) simulations to evaluate the representation of the climatological state and the computational performance of the new model. From the experiments we found that, when activating the humidity mass fixer for OpenIFS, the model was able to reach a near equilibrium under constant 1850 well-mixed greenhouse gas and solar forcing. After 700 years of spinup we branched off four experiments, a historic simulation (165 years); a pre-industrial control run (165 years);

and two idealized experiments, one with a sudden increase to 4xCO2 (120 years) and the other with a 1 % CO_2 (150 years) increase per year.

Climate sensitivity experiments with AWI-CM3 obtained an equilibrium climate sensitivity and transient climate response of 3.2 and 2.1 °C, respectively, both of which are near the center of the CMIP6 spreads.

Using the last 25 years of the historical simulation we established that a low-resolution version of AWI-CM3 provides above-CMIP6-average performance for representing the climatological state of precipitation, wind speeds, the cloud fraction, 500 hPa geopotential height, and air temperature north of 30° S. We found that the Southern Ocean sea ice concentration and thickness were severely underestimated, leading to large positive near-surface air temperature biases in this region, and traced the sea ice biases to spuriously large mixed-layer depth, a positive ice–albedo feedback, and bi-

ases in shortwave downward radiation associated with the cloud fraction between 45–60° S.

AWI-CM3 is capable of realistically simulating the Atlantic meridional overturning circulation (AMOC) in terms of both the shape and the strength of the streamfunction, as well as of reproducing a decreasing trend in the historical period consistent with observations. While the model produces an El Niño–Southern Oscillation (ENSO) at realistic frequencies, the amplitude of ENSO is currently underestimated.

Our recommendations for resolving the Southern Ocean sea ice concentration and thickness biases in future versions of AWI-CM3 include re-tuning of the vertical mixing scheme and the inclusion of coupling minor fluxes that are at least 1 order of magnitude smaller than the ones that have been exchanged so far. We have identified the coupling of ocean surface currents, rain temperature, enthalpy of snow falling into the ocean, enthalpy of melting icebergs, and basal melt flux as promising candidates to reduce model biases. Based on the literature we suggest porting and activating the stochastically perturbed parameterization scheme as a potential way to improve ENSO amplitude.

The advanced computational efficiency and scalability of AWI-CM3, combined with very solid model and coupling physics implementation, will eventually enable us to perform full DECK and scenario simulations at resolutions of 5–25 km, which were previously reserved for the high end of the HighResMIP protocol. We provide a preview with a shorter high-resolution simulation, indicating that most AWI-CM3 climatological biases at future operational resolution will be about half those of the average CMIP6 model.

Appendix A: List of CMIP6 models for climate model performance index calculation

ACCESS-CM2, AWI-CM-1-1-MR, BCC-CSM2-MR, CAMS-CSM1-0, CAS-ESM2-0, CanESM5, CIESM, CESM2, CMCC-CM2-SR5, CNRM-CM6-1-HR, FGOALS-f3-L, FIO-ESM-2-0, E3SM-1-1, EC-Earth3, GFDL-CM4, GISS-E2-1-G, HadGEM3-GC31-MM, ICON-ESM-LR, IITM-ESM, INM-CM5-0, IPSL-CM6A-LR, KIOST-ESM, NESM3, NorESM2-MM, MCM-UA-1-0, MIROC6, MPI-ESM1-2-LR, MRI-ESM2-0, SAM0-UNICON, and TaiESM1.

Appendix B: Additional tables

Table B1. HIST biases as in Table 1 but using NCEP/DOE Reanalysis II by Kistler et al. (2001) instead of ERA5 for near-surface air temperature (tas), eastward near-surface wind (uas), northward near-surface wind (vas), 300 hPa eastward wind (ua), and 500 hPa geopotential height. Note that ERA5 was created using IFS CY41R2, a model closely related to the OpenIFS CY43R3 atmosphere employed in AWI-CM3.

AWI-CM3 CMPI: 0.969

	arctic MAM	arctic JJA	arctic SON	arctic DJF	northmid MAM	northmid JJA	northmid SON	northmid DJF	tropics MAM	tropics JJA	tropics SON	tropics DJF	nino34 MAM	nino34 JJA	nino34 SON	nino34 DJF	southmid MAM	southmid JJA	southmid SON	southmid DJF	antarctic MAM	antarctic JJA	antarctic SON	antarctic DJF
siconc	1.07	0.92	0.74	0.98	0.67	1.07	1.23	0.84									0.97	1.15	1.23	1.16	1.79	3.49	3.50	1.68
tas	1.04	0.64	0.79	0.78	0.75	0.97	0.97	0.94	0.93	0.98	0.94	0.92	0.95	1.55	1.10	0.77	1.26	1.07	1.11	1.33	1.18	1.24	1.17	0.87
clt	0.90	1.16	1.19	1.07	0.70	0.76	0.66	0.78	0.85	0.78	0.61	0.68	0.91	0.32	0.56	0.74	0.80	0.99	0.87	0.79	0.97	0.97	0.83	0.82
pr	0.77	0.87	1.02	1.07	0.87	1.22	1.10	0.91	1.11	1.00	0.90	0.84	1.38	0.91	0.96	1.00	1.17	0.73	1.10	1.08	1.11	1.19	1.04	0.79
rlut	1.02	0.88	0.61	0.49	1.01	0.67	0.63	0.90	1.21	1.04	0.95	0.92	1.52	0.86	0.80	1.04	0.54	0.53	0.59	0.63	1.30	1.44	0.69	0.80
uas	0.86	0.98	0.85	1.15	0.83	0.96	0.80	0.91	1.03	0.92	0.93	0.86	1.76	1.43	0.76	1.01	0.63	0.51	0.58	0.45	0.82	0.76	0.74	0.69
vas	0.82	0.91	0.95	1.07	0.92	0.88	0.98	1.00	1.07	0.96	0.90	0.89	1.18	1.30	0.96	0.72	0.94	0.85	0.74	0.84	0.93	0.94	0.88	0.94
300 hPa ua	0.77	1.02	0.96	1.27	0.92	1.22	0.86	1.15	0.89	0.76	0.83	0.74	0.62	0.84	0.44	0.18	0.76	0.96	1.01	0.50	0.91	0.98	0.72	0.60
500 hPa zg	0.29	0.58	1.22	1.15	0.40	0.69	0.61	0.68	0.29	0.21	0.51	0.24	0.23	0.22	0.66	0.27	0.87	0.70	0.46	0.69	0.85	0.65	0.33	0.82
SD zos	0.66	0.42	0.61	0.61	0.88	0.93	0.90	0.87	1.02	1.02	1.06	1.07	1.35	1.44	1.71	1.70	1.02	1.05	1.08	1.04	0.95	0.82	0.97	0.97
SD tos	1.06	1.04	0.97	1.06	0.81	1.77	1.49	0.97	1.07	1.10	0.89	1.20	0.28	0.19	0.42	0.52	1.06	0.96	1.01	1.12	0.84	1.73	1.67	0.77
mlotst	1.25	0.55	0.70	1.09	2.32	0.64	0.86	1.85	1.41	1.02	1.54	1.01	0.61	0.66	0.89	0.62	0.51	1.89	2.50	1.08	1.96	2.85	2.81	3.64

Table B2. TCo319L137-DART biases as in Table 4 but using NCEP/DOE Reanalysis II by Kistler et al. (2001) instead of ERA5 for near-surface air temperature (tas), eastward near-surface wind (uas), northward near-surface wind (vas), 300 hPa eastward wind (ua), and 500 hPa geopotential height. Note that ERA5 was created using IFS CY41R2, a model closely related to the OpenIFS CY43R3 atmosphere employed in AWI-CM3.

AWI-CM3-DART CMPI: 0.893

	arctic MAM	arctic JJA	arctic SON	arctic DJF	northmid MAM	northmid JJA	northmid SON	northmid DJF	tropics MAM	tropics JJA	tropics SON	tropics DJF	nino34 MAM	nino34 JJA	nino34 SON	nino34 DJF	southmid MAM	southmid JJA	southmid SON	southmid DJF	antarctic MAM	antarctic JJA	antarctic SON	antarctic DJF
siconc	1.26	1.16	1.27	0.86	0.57	1.04	1.54	0.66									0.97	1.14	1.20	1.16	1.48	2.36	2.10	1.36
tas	0.63	0.83	0.95	0.57	0.83	0.82	0.83	0.84	0.93	0.94	0.91	0.94	1.10	0.89	1.09	1.30	0.84	0.95	0.98	0.83	0.91	0.97	0.92	0.73
clt	1.29	1.05	1.14	1.27	0.72	0.77	0.77	0.93	0.70	0.70	0.70	0.74	0.85	0.41	0.31	0.88	0.97	1.21	1.06	0.93	0.92	0.89	0.73	0.75
pr	1.60	0.86	1.05	1.37	0.72	1.20	0.97	0.85	0.78	0.74	0.73	0.66	0.98	0.80	0.96	1.03	1.15	0.70	1.02	1.18	1.02	1.06	1.03	0.82
rlut	0.61	0.83	0.51	0.55	0.58	0.55	0.51	0.52	0.82	0.74	0.81	0.74	1.02	0.67	0.66	0.89	0.39	0.49	0.44	0.59	0.92	1.56	0.66	0.89
uas	0.77	0.90	0.82	0.82	0.67	0.97	0.71	0.69	0.73	0.79	0.77	0.75	0.96	0.81	0.29	0.71	0.65	0.82	0.76	0.67	0.82	0.79	0.78	0.75
vas	0.84	0.91	0.83	0.85	0.75	0.83	0.84	0.89	0.81	0.88	0.86	0.80	0.75	0.92	1.15	0.84	0.87	0.79	0.81	0.87	0.94	0.92	0.96	0.99
300 hPa ua	0.58	0.94	0.76	0.50	0.59	1.08	0.50	0.51	0.61	0.51	0.64	0.63	0.52	0.46	0.51	0.40	0.44	1.21	1.04	0.55	0.65	1.21	0.97	0.50
500 hPa zg	0.25	0.43	0.93	0.44	0.43	0.47	0.30	0.46	0.23	0.26	0.17	0.37	0.08	0.11	0.03	0.19	0.40	0.72	0.56	0.44	0.77	1.25	0.78	0.62
SD zos	0.59	0.48	0.56	0.58	0.64	0.60	0.63	0.65	0.63	0.66	0.70	0.69	1.36	1.30	1.53	1.21	0.59	0.62	0.61	0.61	0.86	0.72	0.89	0.94
SD tos	0.96	1.05	0.91	0.88	0.90	1.46	1.16	1.16	1.04	0.98	0.91	0.94	0.51	0.20	0.30	0.16	1.27	1.52	1.19	1.14	0.61	1.70	1.80	0.64
mlotst	0.72	0.58	0.53	0.66	1.50	0.61	0.60	1.12	1.13	0.98	1.40	1.05	1.75	1.90	2.08	2.36	0.88	1.26	1.83	1.61	2.82	3.24	3.14	5.45

Appendix C: Table of observational datasets for climate model performance index calculation and ocean transports

Table C1. Table of observational datasets for climate model performance index calculation.

Variable	Long name	Dataset	Time range
tas	Near-surface (2 m) air temperature	ERA5 reanalysis	1 Nov 1989 to 30 Nov 2014
uas	Near-surface (10 m) zonal wind speed	ERA5 reanalysis	1 Nov 1989 to 30 Nov 2014
vas	Near-surface (10 m) meridional wind speed	ERA5 reanalysis	1 Nov 1989 to 30 Nov 2014
300 hPa ua	300 hPa zonal wind speed	ERA5 reanalysis	1 Nov 1989 to 30 Nov 2014
300 hPa zg	500 hPa geopotential height	ERA5 reanalysis	1 Nov 1989 to 30 Nov 2014
pr	Precipitation flux	GPCP v2.3	1 Nov 1989 to 30 Nov 2014
siconc	Sea ice area fraction	OSI SAF OSI-450	1 Nov 1989 to 30 Nov 2014
rlut	TOA outgoing longwave flux	CERES-EBAF	15 Mar 2000 to 30 Jun 2014
clt	Cloud area fraction	MODIS Atmosphere L2	15 Mar 2000 to 30 Nov 2014
SD tos	SD of ocean surface temperature	HadISST	1 Nov 1989 to 30 Nov 2014
SD zos	SD of sea surface height	Jason-1, Jason-2, CryoSat	1 Jan 2002 to 30 Nov 2014
mlotst	Mixed-layer depth	C-GLORSv7 reanalysis	1 Nov 1989 to 30 Nov 2014

Table C2. Transport fluxes (Sv) through a number of straits and channels as simulated by AWI-CM3 in comparison to observational estimates. Analysis period was for 1990–2014 of the historic simulation.

Transport (Sv)	AWI-CM3 HIST	Observations	References of observations
Fram Strait	−2.96	−2.0 ± 2.7	Schauer et al. (2008)
Davis Strait	−0.42	−1.6 ± 0.5	Curry et al. (2014)
Bering Strait	1.19	0.83 ± 0.66, 1.0 ± 0.05	Roach et al. (1995), Woodgate (2018)
Nares Strait	−0.31	−0.57 ± 0.09, −0.8 ± 0.3	Münchow and Melling (2008), Münchow et al. (2006)
Barents Sea Opening	2.46	2.0	Smedsrud et al. (2010)
Drake Passage	148.63	136.7 ± 6.9, 173.3 ± 10.7	Meredith et al. (2011), Donohue et al. (2016)
Mozambique Channel	−19.59	−16 ± 8.9	Ridderinkhof et al. (2010)

References

Adcroft, A. and Campin, J.-M.: Rescaled height coordinates for accurate representation of free-surface flows in ocean circulation models, Ocean Model., 7, 269–284, https://doi.org/10.1016/j.ocemod.2003.09.003, 2004.

Adler, R. F., Sapiano, M. R. P., Huffman, G. J., Wang, J.-J., Gu, G., Bolvin, D., Chiu, L., Schneider, U., Becker, A., Nelkin, E., Xie, P., Ferraro, R., and Shin, D.-B.: The Global Precipitation Climatology Project (GPCP) Monthly Analysis (New Version 2.3) and a Review of 2017 Global Precipitation, Atmosphere, 9, 138, https://doi.org/10.3390/atmos9040138, 2018.

Ali, A., Christensen, K. H., Øyvind Breivik, Malila, M., Raj, R. P., Bertino, L., Chassignet, E. P., and Bakhoday-Paskyabi, M.: A comparison of Langmuir turbulence parameterizations and key wave effects in a numerical model of the North Atlantic and Arctic Oceans, Ocean Model., 137, 76–97, https://doi.org/10.1016/j.ocemod.2019.02.005, 2019.

Baker, A. J., Schiemann, R., Hodges, K. I., Demory, M.-E., Mizielinski, M. S., Roberts, M. J., Shaffrey, L. C., Strachan, J., and Vidale, P. L.: Enhanced Climate Change Response of Win-

tertime North Atlantic Circulation, Cyclonic Activity, and Precipitation in a 25-km-Resolution Global Atmospheric Model, J. Climate, 32, 7763–7781, https://doi.org/10.1175/JCLI-D-19-0054.1, 2019.

Balaji, V., Maisonnave, E., Zadeh, N., Lawrence, B. N., Biercamp, J., Fladrich, U., Aloisio, G., Benson, R., Caubel, A., Durachta, J., Foujols, M.-A., Lister, G., Mocavero, S., Underwood, S., and Wright, G.: CPMIP: measurements of real computational performance of Earth system models in CMIP6, Geosci. Model Dev., 10, 19–34, https://doi.org/10.5194/gmd-10-19-2017, 2017.

Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M., and Betts, A. K.: A Revised Hydrology for the ECMWF Model: Verification from Field Site to Terrestrial Water Storage and Impact in the Integrated Forecast System, J. Hydrometeorol., 10, 623–643, https://doi.org/10.1175/2008JHM1068.1, 2009.

Barbi, D., Wieters, N., Gierz, P., Andrés-Martínez, M., Ural, D., Chegini, F., Khosravi, S., and Cristini, L.: ESM-Tools version 5.0: a modular infrastructure for stand-alone and coupled Earth system modelling (ESM), Geosci. Model Dev., 14, 4051–4067, https://doi.org/10.5194/gmd-14-4051-2021, 2021.

Barbi, D., Gierz, P., Andrés-Martínez, M., Ural, D., and Cristini, L.: esm_tools_release3_as_used_by_AWI-CM3_paper (3.1), Zenodo [code], https://doi.org/10.5281/zenodo.6335309, 2022.

Batrak, Y. and Müller, M.: On the warm bias in atmospheric reanalyses induced by the missing snow over Arctic sea-ice, Nat. Commun., 10, 4170, https://doi.org/10.1038/s41467-019-11975-3, 2019.

Beadling, R. L., Russell, J. L., Stouffer, R. J., Mazloff, M., Talley, L. D., Goodman, P. J., Sallée, J. B., Hewitt, H. T., Hyder, P., and Pandde, A.: Representation of Southern Ocean Properties across Coupled Model Intercomparison Project Generations: CMIP3 to CMIP6, J. Climate, 33, 6555–6581, https://doi.org/10.1175/JCLI-D-19-0970.1, 2020.

Bock, L., Lauer, A., Schlund, M., Barreiro, M., Bellouin, N., Jones, C., Meehl, G. A., Predoi, V., Roberts, M. J., and Eyring, V.: Quantifying Progress Across Different CMIP Phases With the ESMValTool, J. Geophys. Res.-Atmos., 125, e2019JD032321, https://doi.org/10.1029/2019JD032321, 2020.

Buizza, R., Milleer, M., and Palmer, T. N.: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system, Q. J. Roy. Meteor. Soc., 125, 2887–2908, https://doi.org/10.1002/qj.49712556006, 1999.

Caesar, L., McCarthy, G. D., Thornalley, D. J. R., Cahill, N., and Rahmstorf, S.: Reply to: Atlantic circulation change still uncertain, Nat. Geosci., 15, 168–170, https://doi.org/10.1038/s41561-022-00897-3, 2022.

Craig, A., Valcke, S., and Coquart, L.: Development and performance of a new version of the OASIS coupler, OASIS3-MCT_3.0, Geosci. Model Dev., 10, 3297–3308, https://doi.org/10.5194/gmd-10-3297-2017, 2017.

Curry, B., Lee, C. M., Petrie, B., Moritz, R. E., and Kwok, R.: Multiyear Volume, Liquid Freshwater, and Sea Ice Transports through Davis Strait, 2004–10, J. Phys. Oceanogr., 44, 1244–1266, https://doi.org/10.1175/JPO-D-13-0177.1, 2014.

Danek, C., Shi, X., Stepanek, C., Yang, H., Barbi, D., Hegewald, J., and Lohmann, G.: AWI AWI-ESM1.1LR model output prepared for CMIP6 CMIP, Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.9301, 2020.

Danilov, S., Wang, Q., Timmermann, R., Iakovlev, N., Sidorenko, D., Kimmritz, M., Jung, T., and Schröter, J.: Finite-Element Sea Ice Model (FESIM), version 2, Geosci. Model Dev., 8, 1747–1761, https://doi.org/10.5194/gmd-8-1747-2015, 2015.

Danilov, S., Sidorenko, D., Wang, Q., and Jung, T.: The Finite-volumE Sea ice–Ocean Model (FESOM2), Geosci. Model Dev., 10, 765–789, https://doi.org/10.5194/gmd-10-765-2017, 2017.

Davini, P., Corti, S., D'Andrea, F., Rivière, G., and von Hardenberg, J.: Improved Winter European Atmospheric Blocking Frequencies in High-Resolution Global Climate Simulations, J. Adv. Model. Earth Sy., 9, 2615–2634, https://doi.org/10.1002/2017MS001082, 2017.

Davy, R., Chen, L., and Hanna, E.: Arctic amplification metrics, International J. Climatol., 38, 4384–4394, https://doi.org/10.1002/joc.5675, 2018.

Dawson, A.: EOFs: A library for EOF analysis of meteorological, oceanographic, and climate data, J. Open Res. Softw., 4, e14, https://doi.org/10.5334/jors.122, 2016.

Delworth, T. L., Broccoli, A. J., Rosati, A., Stouffer, R. J., Balaji, V., Beesley, J. A., Cooke, W. F., Dixon, K. W., Dunne, J., Dunne, K. A., Durachta, J. W., Findell, K. L., Ginoux, P., Gnanadesikan, A., Gordon, C. T., Griffies, S. M., Gudgel, R., Harrison, M. J., Held, I. M., Hemler, R. S., Horowitz, L. W., Klein, S. A., Knutson, T. R., Kushner, P. J., Langenhorst, A. R., Lee, H.-C., Lin, S.-J., Lu, J., Malyshev, S. L., Milly, P. C. D., Ramaswamy, V., Russell, J., Schwarzkopf, M. D., Shevliakova, E., Sirutis, J. J., Spelman, M. J., Stern, W. F., Winton, M., Wittenberg, A. T., Wyman, B., Zeng, F., and Zhang, R.: GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics, J. Climate, 19, 643–674, https://doi.org/10.1175/JCLI3629.1, 2006.

Delworth, T. L., Rosati, A., Anderson, W., Adcroft, A. J., Balaji, V., Benson, R., Dixon, K., Griffies, S. M., Lee, H.-C., Pacanowski, R. C., Vecchi, G. A., Wittenberg, A. T., Zeng, F., and Zhang, R.: Simulated Climate and Climate Change in the GFDL CM2.5 High-Resolution Coupled Climate Model, J. Climate, 25, 2755–2781, https://doi.org/10.1175/JCLI-D-11-00316.1, 2012.

Donohue, K. A., Tracey, K. L., Watts, D. R., Chidichimo, M. P., and Chereskin, T. K.: Mean Antarctic Circumpolar Current transport measured in Drake Passage, Geophys. Res. Lett., 43, 11760–11767, https://doi.org/10.1002/2016GL070319, 2016.

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arsouze, T., Bergman, T., Bernardello, R., Boussetta, S., Caron, L.-P., Carver, G., Castrillo, M., Catalano, F., Cvijanovic, I., Davini, P., Dekker, E., Doblas-Reyes, F. J., Docquier, D., Echevarria, P., Fladrich, U., Fuentes-Franco, R., Gröger, M., v. Hardenberg, J., Hieronymus, J., Karami, M. P., Keskinen, J.-P., Koenigk, T., Makkonen, R., Massonnet, F., Ménégoz, M., Miller, P. A., Moreno-Chamarro, E., Nieradzik, L., van Noije, T., Nolan, P., O'Donnell, D., Ollinaho, P., van den Oord, G., Ortega, P., Prims, O. T., Ramos, A., Reerink, T., Rousset, C., Ruprich-Robert, Y., Le Sager, P., Schmith, T., Schrödner, R., Serva, F., Sicardi, V., Sloth Madsen, M., Smith, B., Tian, T., Tourigny, E., Uotila, P., Vancoppenolle, M., Wang, S., Wårlind, D., Willén, U., Wyser, K., Yang, S., Yepes-Arbós, X., and Zhang, Q.: The EC-Earth3 Earth system model for the Coupled Model Intercomparison Project 6, Geosci. Model Dev., 15, 2973–3020, https://doi.org/10.5194/gmd-15-2973-2022, 2022.

ECMWF: IFS Documentation CY43R3 – Part III: Dynamics and numerical procedures, no. 3 in IFS Documentation, ECMWF, https://doi.org/10.21957/8l7miod5m, 2017a.

ECMWF: IFS Documentation CY43R3 – Part IV: Physical processes, no. 4 in IFS Documentation, ECMWF, https://doi.org/10.21957/efyk72kl, 2017b.

ECMWF: IFS Documentation CY43R3 – Part VII: ECMWF wave model, no. 7 in IFS Documentation, ECMWF, https://doi.org/10.21957/mxz9z1gb, 2017c.

Forbes, R., Geer, A., Lonitz, K., and Ahlgrimm, M.: Reducing systematic errors in cold-air outbreaks, ECMWF newsletter, 146, 17–22, 2016.

Gent, P. R. and Mcwilliams, J. C.: Isopycnal mixing in ocean circulation models, J. Phys. Oceanogr., 20, 150–155, 1990.

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., Lowe, J. A., Johns, T. C., and Williams, K. D.: A new method for diagnosing radiative forcing and climate sensitivity, Geophys. Res. Lett., 31, https://doi.org/10.1029/2003GL018747, 2004.

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N. S., Guemas, V., von Hardenberg, J., Hazeleger, W., Kodama, C., Koenigk, T., Leung, L. R., Lu, J., Luo, J.-J., Mao, J., Mizielinski, M. S., Mizuta, R., Nobre, P., Satoh, M., Scoccimarro, E., Semmler, T., Small, J., and von Storch, J.-S.: High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6, Geosci. Model Dev., 9, 4185–4208, https://doi.org/10.5194/gmd-9-4185-2016, 2016.

Hazeleger, W., Severijns, C., Semmler, T., Ştefănescu, S., Yang, S., Wang, X., Wyser, K., Dutra, E., Baldasano, J. M., Bintanja, R., Bougeault, P., Caballero, R., Ekman, A. M. L., Christensen, J. H., van den Hurk, B., Jimenez, P., Jones, C., Kållberg, P., Koenigk, T., McGrath, R., Miranda, P., van Noije, T., Palmer, T., Parodi, J. A., Schmith, T., Selten, F., Storelvmo, T., Sterl, A., Tapamo, H., Vancoppenolle, M., Viterbo, P., and Willén, U.: EC-Earth: a seamless earth-system prediction approach in action, B. Am. Meteorol. Soc., 91, 1357–1364, https://doi.org/10.1175/2010BAMS2877.1, 2010.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Q. J. Roy. Meteor. Soc., 146, 1999–2049, 2020.

Hertwig, E., von Storch, J.-S., Handorf, D., Dethloff, K., Fast, I., and Krismer, T.: Effect of horizontal resolution on ECHAM6-AMIP performance, Clim. Dynam., 45, 185–211, https://doi.org/10.1007/s00382-014-2396-x, 2015.

Hortal, M. and Simmons, A. J.: Use of Reduced Gaussian Grids in Spectral Models, Mon. Weather Rev., 119, 1057–1074, https://doi.org/10.1175/1520-0493(1991)119<1057:UORGGI>2.0.CO;2, 1991.

Hunke, E., Allard, R., Bailey, D. A., Blain, P., Craig, A., Dupont, F., DuVivier, A., Grumbine, R., Hebert, D., Holland, M., Jeffery, N., Lemieux, J.-F., Osinski, R., Rasmussen, T., Ribergaard, M., Roach, L., Roberts, A., Turner, M., and Winton,

M.: CICE-Consortium/Icepack: Icepack 1.2.1, Zenodo [code], https://doi.org/10.5281/zenodo.3712299, 2020.

Johannessen, O. M., Kuzmina, S. I., Bobylev, L. P., and Miles, M. W.: Surface air temperature variability and trends in the Arctic: new amplification assessment and regionalisation, Tellus A, 68, 28234, https://doi.org/10.3402/tellusa.v68.28234, 2016.

Jung, T., Miller, M. J., Palmer, T. N., Towers, P., Wedi, N., Achuthavarier, D., Adams, J. M., Altshuler, E. L., Cash, B. A., Kinter, J. L., Marx, L., Stan, C., and Hodges, K. I.: High-Resolution Global Climate Simulations with the ECMWF Model in Project Athena: Experimental Design, Model Climate, and Seasonal Forecast Skill, J. Climate, 25, 3155–3172, https://doi.org/10.1175/JCLI-D-11-00265.1, 2012.

Jungclaus, J. H., Fischer, N., Haak, H., Lohmann, K., Marotzke, J., Matei, D., Mikolajewicz, U., Notz, D., and von Storch, J. S.: Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model, J. Adv. Model. Earth Sy., 5, 422–446, https://doi.org/10.1002/jame.20023, 2013.

Juricke, S., Danilov, S., Koldunov, N., Oliver, M., Sein, D. V., Sidorenko, D., and Wang, Q.: A kinematic kinetic energy backscatter parametrization: From implementation to global ocean simulations, J. Adv. Model. Earth Sy., 12, e2020MS002175, https://doi.org/10.1029/2020MS002175, 2020.

Keil, P., Mauritsen, T., Jungclaus, J., Hedemann, C., Olonscheck, D., and Ghosh, R.: Multiple drivers of the North Atlantic warming hole, Nat. Clim. Change, 10, 667–671, https://doi.org/10.1038/s41558-020-0819-8, 2020.

Kimmritz, M., Danilov, S., and Losch, M.: The adaptive EVP method for solving the sea ice momentum equation, Ocean Model., 101, 59–67, https://doi.org/10.1016/j.ocemod.2016.03.004, 2016.

Kistler, R., Kalnay, E., Collins, W., Saha, S., White, G., Woollen, J., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., van den Dool, H., Jenne, R., and Fiorino, M.: The NCEP–NCAR 50-Year Reanalysis: Monthly Means CD-ROM and Documentation, B. Am. Meteorol. Soc., 82, 247–268, https://doi.org/10.1175/1520-0477(2001)082<0247:TNNYRM>2.3.CO;2, 2001.

Kjellsson, J., Streffing, J., Carver, G., and Köhler, M.: From weather forecasting to climate modelling using OpenIFS, ECMWF Newsletter, 164, 38–41, 2020.

Koldunov, N. V., Aizinger, V., Rakowsky, N., Scholz, P., Sidorenko, D., Danilov, S., and Jung, T.: Scalability and some optimization of the Finite-volumE Sea ice–Ocean Model, Version 2.0 (FESOM2), Geosci. Model Dev., 12, 3991–4012, https://doi.org/10.5194/gmd-12-3991-2019, 2019a.

Koldunov, N. V., Danilov, S., Sidorenko, D., Hutter, N., Losch, M., Goessling, H., Rakowsky, N., Scholz, P., Sein, D., Wang, Q., and Jung, T.: Fast EVP Solutions in a High-Resolution Sea Ice Model, J. Adv. Model. Earth Sy., 11, 1269–1284, https://doi.org/10.1029/2018MS001485, 2019b.

Komen, G., Cavaleri, L., Donelan, M., Hasselmann, K., Hasselmann, S., and Janssen, P.: Dynamics and modelling of ocean waves, Cambridge University Press, UK, 554, 1271–1285, 1996.

Large, W. G., McWilliams, J. C., and Doney, S. C.: Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization, Rev. Geophys., 32, 363–403, https://doi.org/10.1029/94RG01872, 1994.

Lavergne, T., Sørensen, A. M., Kern, S., Tonboe, R., Notz, D., Aaboe, S., Bell, L., Dybkjær, G., Eastwood, S., Gabarro, C., Heygster, G., Killie, M. A., Brandt Kreiner, M., Lavelle, J., Saldo, R., Sandven, S., and Pedersen, L. T.: Version 2 of the EUMETSAT OSI SAF and ESA CCI sea-ice concentration climate data records, The Cryosphere, 13, 49–78, https://doi.org/10.5194/tc-13-49-2019, 2019.

Lemarié, F., Blayo, E., and Debreu, L.: Analysis of Ocean-atmosphere Coupling Algorithms: Consistency and Stability, Procedia Computer Science, 51, 2066–2075, https://doi.org/10.1016/j.procs.2015.05.473, 2015.

Lemke, P.: A coupled one-dimensional sea ice-ocean model, J. Geophys. Res.-Oceans, 92, 13164–13172, https://doi.org/10.1029/JC092iC12p13164, 1987.

Ma, X., Jing, Z., Chang, P., Liu, X., Montuoro, R., Small, R. J., Bryan, F. O., Greatbatch, R. J., Brandt, P., Wu, D., Lin, X., and Wu, L.: Western boundary currents regulated by interaction between ocean eddies and the atmosphere, Nature, 535, 533–537, https://doi.org/10.1038/nature18640, 2016.

Maisonnave, E., Coquart, L., and Piacentini, A.: A better diagnostic of the load imbalance in OASIS based coupled systems, Tech. rep., Technical Report, TR/CMGC/20/176, CECI, UMR CER-FACS/CNRS No5318, France, 2020.

Malardel, S., Wedi, N., Deconinck, W., Diamantakis, M., Kühnlein, C., Mozdzynski, G., Hamrud, M., and Smolarkiewicz, P.: A new grid for the IFS, ECMWF newsletter, 146, 23–28, 2016.

Malardel, S., Diamantakis, M., Agusti-Panareda, A., and Flemming, J.: Dry mass versus total mass conservation in the IFS, ECMWF, https://doi.org/10.21957/s3ho2aq29, 2019.

Marti, O., Nguyen, S., Braconnot, P., Valcke, S., Lemarié, F., and Blayo, E.: A Schwarz iterative method to evaluate ocean–atmosphere coupling schemes: implementation and diagnostics in IPSL-CM6-SW-VLR, Geosci. Model Dev., 14, 2959–2975, https://doi.org/10.5194/gmd-14-2959-2021, 2021.

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J., Maycock, T., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B. (Eds.): Summary for Policymakers, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Tech. rep., Intergovernmental Panel on Climate Change, https://doi.org/10.1017/9781009157896, 2021.

McGregor, J. L.: C-CAM: Geometric aspects and dynamical formulation, CSIRO Atmospheric Research Dickson ACT, 2005.

Meehl, G. A., Senior, C. A., Eyring, V., Flato, G., Lamarque, J.-F., Stouffer, R. J., Taylor, K. E., and Schlund, M.: Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth system models, Sci. Adv., 6, eaba1981, https://doi.org/10.1126/sciadv.aba1981, 2020.

Meredith, M. P., Woodworth, P. L., Chereskin, T. K., Marshall, D. P., Allison, L. C., Bigg, G. R., Donohue, K., Heywood, K. J., Hughes, C. W., Hibbert, A., Hogg, A. M., Johnson, H. L., Jullion, L., King, B. A., Leach, H., Lenn, Y.-D., Morales Maqueda, M. A., Munday, D. R., Naveira Garabato, A. C., Provost, C., Sallée, J.-B., and Sprintall, J.: Sustained Monitoring of the Southern Ocean at Drake Passage: Past Achievements and Future Priorities, Rev. Geophys., 49, RG4005, https://doi.org/10.1029/2010RG000348, 2011.

Meurdesoif, Y.: XIOS 2.0 (Revision 1297), Zenodo [code], https://doi.org/10.5281/zenodo.4905653, 2017.

Monterey, G., and S. Levitus, Seasonal Variability of Mixed Layer Depth for the World Ocean, NOAA Atlas NESDIS, vol. 14, 100 pp., Natl. Oceanic and Atmos. Admin., Silver Spring, Md., 1997.

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., Dunn, R. J. H., Osborn, T. J., Jones, P. D., and Simpson, I. R.: An Updated Assessment of Near-Surface Temperature Change From 1850: The HadCRUT5 Data Set, J. Geophys. Res.-Atmos., 126, e2019JD032361, https://doi.org/10.1029/2019JD032361, 2021.

Münchow, A. and Melling, H.: Ocean current observations from Nares Strait to the west of Greenland: Interannual to tidal variability and forcing, J. Marine Res., 66, 801–833, https://doi.org/10.1357/002224008788064612, 2008.

Münchow, A., Melling, H., and Falkner, K. K.: An Observational Estimate of Volume and Freshwater Flux Leaving the Arctic Ocean through Nares Strait, J. Phys. Oceanogr., 36, 2025–2041, https://doi.org/10.1175/JPO2962.1, 2006.

Ollinaho, P., Lock, S.-J., Leutbecher, M., Bechtold, P., Beljaars, A., Bozzo, A., Forbes, R. M., Haiden, T., Hogan, R. J., and Sandu, I.: Towards process-level representation of model uncertainties: stochastically perturbed parametrizations in the ECMWF ensemble, Q. J. Roy. Meteor. Soc., 143, 408–422, https://doi.org/10.1002/qj.2931, 2017.

Parkinson, C. L. and Washington, W. M.: A large-scale numerical model of sea ice, J. Geophys. Res.-Oceans, 84, 311–337, https://doi.org/10.1029/JC084iC01p00311, 1979.

Petersen, M. R., Jacobsen, D. W., Ringler, T. D., Hecht, M. W., and Maltrud, M. E.: Evaluation of the arbitrary Lagrangian–Eulerian vertical coordinate method in the MPAS-Ocean model, Ocean Model., 86, 93–113, https://doi.org/10.1016/j.ocemod.2014.12.004, 2015.

Pithan, F., Shepherd, T. G., Zappa, G., and Sandu, I.: Climate model biases in jet streams, blocking and storm tracks resulting from missing orographic drag, Geophys. Res. Lett., 43, 7231–7240, https://doi.org/10.1002/2016GL069551, 2016.

Platnick, S., Ackerman, S., King, M., Meyer, K., Menzel, W., Holz, R., Baum, B., and Yang, P.: MODIS atmosphere L2 cloud product (06_L2), NASA MODIS Adaptive Processing System, Goddard Space Flight Center, https://doi.org/10.5067/MODIS/MOD06_L2.006, 2015.

Rackow, T. and Juricke, S.: Flow-dependent stochastic coupling for climate models with high ocean-to-atmosphere resolution ratio, Q. J. Roy. Meteor. Soc., 146, 284–300, https://doi.org/10.1002/qj.3674, 2020.

Rackow, T., Goessling, H. F., Jung, T., Sidorenko, D., Semmler, T., Barbi, D., and Handorf, D.: Towards multi-resolution global climate modeling with ECHAM6-FESOM. Part II: climate variability, Clim. Dynam., 50, 2369–2394, https://doi.org/10.1007/s00382-016-3192-6, 2018.

Rackow, T., Danilov, S., Goessling, H. F., Hellmer, H. H., Sein, D. V., Semmler, T., Sidorenko, D., and Jung, T.: Delayed Antarctic sea-ice decline in high-resolution climate change simulations, Nat. Commun., 13, 637, https://doi.org/10.1038/s41467-022-28259-y, 2022.

Reichler, T. and Kim, J.: How Well Do Coupled Models Simulate Today's Climate?, B. Am. Meteorol. Soc., 89, 303–312, https://doi.org/10.1175/BAMS-89-3-303, 2008.

Renault, L., Molemaker, M. J., Gula, J., Masson, S., and McWilliams, J. C.: Control and Stabilization of the Gulf Stream by Oceanic Current Interaction with the Atmosphere, J. Phys. Oceanogr., 46, 3439–3453, https://doi.org/10.1175/JPO-D-16-0115.1, 2016.

Ridderinkhof, H., van der Werf, P. M., Ullgren, J. E., van Aken, H. M., van Leeuwen, P. J., and de Ruijter, W. P. M.: Seasonal and interannual variability in the Mozambique Channel from moored current observations, J. Geophys. Res.-Oceans, 115, C6, https://doi.org/10.1029/2009JC005619, 2010.

Ritchie, H.: Semi-Lagrangian Advection on a Gaussian Grid, Mon. Weather Rev., 115, 608–619, https://doi.org/10.1175/1520-0493(1987)115<0608:SLAOAG>2.0.CO;2, 1987.

Ritchie, H., Temperton, C., Simmons, A., Hortal, M., Davies, T., Dent, D., and Hamrud, M.: Implementation of the Semi-Lagrangian Method in a High-Resolution Version of the ECMWF Forecast Model, Mon. Weather Rev., 123, 489–514, https://doi.org/10.1175/1520-0493(1995)123<0489:IOTSLM>2.0.CO;2, 1995.

Roach, A. T., Aagaard, K., Pease, C. H., Salo, S. A., Weingartner, T., Pavlov, V., and Kulakov, M.: Direct measurements of transport and water properties through the Bering Strait, J. Geophys. Res.-Oceans, 100, 18443–18457, https://doi.org/10.1029/95JC01673, 1995.

Robert, A., Henderson, J., and Turnbull, C.: An Implicit Time Integration Scheme for Baroclinic Models of the Atmosphere, Mon. Weather Rev., 100, 329–335, https://doi.org/10.1175/1520-0493(1972)100<0329:AITISF>2.3.CO;2, 1972.

Roberts, C. D., Senan, R., Molteni, F., Boussetta, S., Mayer, M., and Keeley, S. P. E.: Climate model configurations of the ECMWF Integrated Forecasting System (ECMWF-IFS cycle 43r1) for HighResMIP, Geosci. Model Dev., 11, 3681–3712, https://doi.org/10.5194/gmd-11-3681-2018, 2018.

Saarinen, S., Hamrud, M., Salmond, D., and Hague, J.: Dr. Hook instrumentation tool, Tech. rep., European Centre for Medium-Range Weather Forecasts, 2005.

Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization ofhttps://doi.org/10.1029/2008WR007327, 2010.

Satoh, M., Stevens, B., Judt, F., Khairoutdinov, M., Lin, S.-J., Putman, W. M., and Düben, P.: Global Cloud-Resolving Models, Current Climate Change Reports, 5, 172–184, https://doi.org/10.1007/s40641-019-00131-0, 2019.

Scafetta, N.: Testing the CMIP6 GCM Simulations versus Surface Temperature Records from 1980–1990 to 2011–2021: High ECS Is Not Supported, Climate, 9, 161, https://doi.org/10.3390/cli9110161, 2021.

Schauer, U., Beszczynska-Möller, A., Walczowski, W., Fahrbach, E., Piechura, J., and Hansen, E.: Variation of Measured Heat Flow Through the Fram Strait Between 1997 and 2006, pp. 65–85, Springer Netherlands, Dordrecht, https://doi.org/10.1007/978-1-4020-6774-7_4, 2008.

Schiemann, R., Demory, M.-E., Shaffrey, L. C., Strachan, J., Vidale, P. L., Mizielinski, M. S., Roberts, M. J., Matsueda, M., Wehner, M. F., and Jung, T.: The Resolution Sensitivity of Northern Hemisphere Blocking in Four 25-km Atmospheric Global Circulation Models, J. Climate, 30, 337–358, https://doi.org/10.1175/JCLI-D-16-0100.1, 2017.

Scholz, P., Sidorenko, D., Gurses, O., Danilov, S., Koldunov, N., Wang, Q., Sein, D., Smolentseva, M., Rakowsky, N., and Jung, T.: Assessment of the Finite-volumE Sea ice-Ocean Model (FESOM2.0) – Part 1: Description of selected key model elements and comparison to its predecessor version, Geosci. Model Dev., 12, 4875–4899, https://doi.org/10.5194/gmd-12-4875-2019, 2019.

Scholz, P., Sidorenko, D., Danilov, S., Wang, Q., Koldunov, N., Sein, D., and Jung, T.: Assessment of the Finite-VolumE Sea ice–Ocean Model (FESOM2.0) – Part 2: Partial bottom cells, embedded sea ice and vertical mixing library CVMix, Geosci. Model Dev., 15, 335–363, https://doi.org/10.5194/gmd-15-335-2022, 2022a.

Scholz, P., Sidorenko, D., Gurses, O., Danilov, S., Koldunov, N., Wang, Q., Sein, D., Smolentseva, M., Rakowsky, N., and Jung, T.: FESOM 2.0 AWI-CM3 version 3.0 (AWI-CM3 version 3.0), Zenodo [code], https://doi.org/10.5281/zenodo.6335383, 2022b.

Sein, D. V., Koldunov, N. V., Danilov, S., Wang, Q., Sidorenko, D., Fast, I., Rackow, T., Cabos, W., and Jung, T.: Ocean Modeling on a Mesh With Resolution Following the Local Rossby Radius, J. Adv. Model. Earth Sy., 9, 2601–2614, https://doi.org/10.1002/2017MS001099, 2017.

Semmler, T., Danilov, S., Gierz, P., Goessling, H. F., Hegewald, J., Hinrichs, C., Koldunov, N., Khosravi, N., Mu, L., Rackow, T., Sein, D. V., Sidorenko, D., Wang, Q., and Jung, T.: Simulations for CMIP6 With the AWI Climate Model AWI-CM-1-1, J. Adv. Model. Earth Sy., 12, e2019MS002009, https://doi.org/10.1029/2019MS002009, 2020.

Sidorenko, D., Rackow, T., Jung, T., Semmler, T., Barbi, D., Danilov, S., Dethloff, K., Dorn, W., Fieg, K., Goessling, H. F., Handorf, D., Harig, S., Hiller, W., Juricke, S., Losch, M., Schröter, J., Sein, D. V., and Wang, Q.: Towards multi-resolution global climate modeling with ECHAM6–FESOM. Part I: model formulation and mean climate, Clim. Dynam., 44, 757–780, https://doi.org/10.1007/s00382-014-2290-6, 2015.

Sidorenko, D., Goessling, H., Koldunov, N., Scholz, P., Danilov, S., Barbi, D., Cabos, W., Gurses, O., Harig, S., Hinrichs, C., Juricke, S., Lohmann, G., Losch, M., Mu, L., Rackow, T., Rakowsky, N., Sein, D., Semmler, T., Shi, X., Stepanek, C., Streffing, J., Wang, Q., Wekerle, C., Yang, H., and Jung, T.: Evaluation of FESOM2.0 Coupled to ECHAM6.3: Preindustrial and HighResMIP Simulations, J. Adv. Model. Earth Sy., 11, 3794–3815, https://doi.org/10.1029/2019MS001696, 2019.

Sidorenko, D., Danilov, S., Streffing, J., Fofonova, V., Goessling, H. F., Scholz, P., Wang, Q., Androsov, A., Cabos, W., Juricke, S., Koldunov, N., Rackow, T., Sein, D. V., and Jung, T.: AMOC Variability and Watermass Transformations in the AWI Climate Model, J. Adv. Model. Earth Sy., 13, e2021MS002582, https://doi.org/10.1029/2021MS002582, 2021.

Smedsrud, L. H., Ingvaldsen, R., Nilsen, J. E. Ø., and Skagseth, Ø.: Heat in the Barents Sea: transport, storage, and surface fluxes, Ocean Sci., 6, 219–234, https://doi.org/10.5194/os-6-219-2010, 2010.

Steele, M., Morley, R., and Ermold, W.: PHC: A global ocean hydrography with a high-quality Arctic Ocean, J. Climate, 14, 2079–2087, https://doi.org/10.1175/1520-0442(2001)014<2079:PAGOHW>2.0.CO;2, 2001.

Sterl, A., Bintanja, R., Brodeau, L., Gleeson, E., Koenigk, T., Schmith, T., Semmler, T., Severijns, C., Wyser, K., and Yang, S.: A look at the ocean in the EC-Earth climate model, Clim. Dy-

nam., 39, 2631–2657, https://doi.org/10.1007/s00382-011-1239-2, 2012.

Storto, A., Masina, S., and Navarra, A.: Evaluation of the CMCC eddy-permitting global ocean physical reanalysis system (C-GLORS, 1982–2012) and its assimilation components, Q. J. Roy. Meteorol. Soc., 142, 738–758, https://doi.org/10.1002/qj.2673, 2016.

Streffing, J.: AWI-CM3 version 3.0 historic, 4XCO$_2$, 1 % CO$_2$ data, Zenodo [data set], https://doi.org/10.5281/zenodo.6337593, 2022a.

Streffing, J.: AWI-CM3 version 3.0 pre-industrial control data, Zenodo [data set], https://doi.org/10.5281/zenodo.6337571, 2022b.

Streffing, J.: AWI-CM3 version 3.0 spinup simulation data, Zenodo [data set], https://doi.org/10.5281/zenodo.6337627, 2022c.

Streffing, J.: JanStreffing/2020_AWICM3_GMD_PAPER: Revision 1 (1.1), Zenodo [code], https://doi.org/10.5281/zenodo.6653826, 2022d.

Streffing, J. and Fladich, U.: Modifications to use OpenIFS CY43R3V1 for AWI-CM3 version 3.0, Zenodo [code], https://doi.org/10.5281/zenodo.6335498, 2022.

Timmermann, R. and Beckmann, A.: Parameterization of vertical mixing in the Weddell Sea, Ocean Model., 6, 83–100, https://doi.org/10.1016/S1463-5003(02)00061-6, 2004.

Titchner, H. A. and Rayner, N. A.: The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations, J. Geophys. Res.-Atmos., 119, 2864–2889, https://doi.org/10.1002/2013JD020316, 2014.

van Haren, R., Haarsma, R. J., Van Oldenborgh, G. J., and Hazeleger, W.: Resolution Dependence of European Precipitation in a State-of-the-Art Atmospheric General Circulation Model, J. Climate, 28, 5134–5149, https://doi.org/10.1175/JCLI-D-14-00279.1, 2015.

Vignati, E., Wilson, J., and Stier, P.: M7: An efficient size-resolved aerosol microphysics module for large-scale aerosol transport models, J. Geophys. Res.-Atmos., 109, https://doi.org/10.1029/2003JD004485, 2004.

Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., Colin, J., Guérémy, J.-F., Michou, M., Moine, M.-P., Nabat, P., Roehrig, R., Salas y Mélia, D., Séférian, R., Valcke, S., Beau, I., Belamari, S., Berthet, S., Cassou, C., Cattiaux, J., Deshayes, J., Douville, H., Ethé, C., Franchistéguy, L., Geoffroy, O., Lévy, C., Madec, G., Meurdesoif, Y., Msadek, R., Ribes, A., Sanchez-Gomez, E., Terray, L., and Waldman, R.: Evaluation of CMIP6 DECK Experiments With CNRM-CM6-1, J. Adv. Model. Earth Sy., 11, 2177–2213, https://doi.org/10.1029/2019MS001683, 2019.

Walsh, J., Chapman, W., Fetterer, F., and Stewart, J.: Gridded monthly sea ice extent and concentration, 1850 onward, Version 2, Boulder, Colorado USA, NSIDC: National Snow and Ice Data Center, https://doi.org/10.7265/jj4s-tq79, 2019.

Wang, S., Wang, Q., Shu, Q., Scholz, P., Lohmann, G., and Qiao, F.: Improving the Upper-Ocean Temperature in an Ocean Climate Model (FESOM 1.4): Shortwave Penetration Versus Mixing Induced by Nonbreaking Surface Waves, J. Adv. Model. Earth Sy., 11, 545–557, https://doi.org/10.1029/2018MS001494, 2019.

Watts, M., Maslowski, W., Lee, Y. J., Kinney, J. C., and Osinski, R.: A Spatial Evaluation of Arctic Sea Ice and Regional Limitations in CMIP6 Historical Simulations, J. Climate, 34, 6399–6420, https://doi.org/10.1175/JCLI-D-20-0491.1, 2021.

Wielicki, B. A., Barkstrom, B. R., Harrison, E. F., Lee, R. B., Smith, G. L., and Cooper, J. E.: Clouds and the Earth's Radiant Energy System (CERES): An Earth Observing System Experiment, B. Am. Meteorol. Soc., 77, 853–868, https://doi.org/10.1175/1520-0477(1996)077<0853:CATERE>2.0.CO;2, 1996.

Wild, M.: The global energy balance as represented in CMIP6 climate models, Clim. Dynam., 55, 553–577, 2020.

Willison, J., Robinson, W. A., and Lackmann, G. M.: North Atlantic Storm-Track Sensitivity to Warming Increases with Model Resolution, J. Climate, 28, 4513–4524, https://doi.org/10.1175/JCLI-D-14-00715.1, 2015.

Woodgate, R. A.: Increases in the Pacific inflow to the Arctic from 1990 to 2015, and insights into seasonal trends and driving mechanisms from year-round Bering Strait mooring data, Prog. Oceanogr., 160, 124–154, https://doi.org/10.1016/j.pocean.2017.12.007, 2018.

Wyser, K.: EC-Earth community runoff-mapper scheme, Zenodo [code], https://doi.org/10.5281/zenodo.6335474, 2022.

Yamazaki, D., Kanae, S., Kim, H., and Oki, T.: A physically based description of floodplain inundation dynamics in a global river routing model, Water Resour. Res., 47, https://doi.org/10.1029/2010WR009726, 2011.

Yang, C., Christensen, H. M., Corti, S., von Hardenberg, J., and Davini, P.: The impact of stochastic physics on the El Niño Southern Oscillation in the EC-Earth coupled model, Clim. Dynam., 53, 2843–2859, https://doi.org/10.1007/s00382-019-04660-0, 2019.

Yepes-Arbós, X., van den Oord, G., Acosta, M. C., and Carver, G. D.: Evaluation and optimisation of the I/O scalability for the next generation of Earth system models: IFS CY43R3 and XIOS 2.0 integration as a case study, Geosci. Model Dev., 15, 379–394, https://doi.org/10.5194/gmd-15-379-2022, 2022.

Zampieri, L., Kauker, F., Fröhle, J., Sumata, H., Hunke, E. C., and Goessling, H. F.: Impact of Sea-Ice Model Complexity on the Performance of an Unstructured-Mesh Sea-Ice/Ocean Model under Different Atmospheric Forcings, J. Adv. Model. Earth Sy., 13, e2020MS002438, https://doi.org/10.1029/2020MS002438, 2021.

Zeman, C., Wedi, N. P., Dueben, P. D., Ban, N., and Schär, C.: Model intercomparison of COSMO 5.0 and IFS 45r1 at kilometer-scale grid spacing, Geosci. Model Dev., 14, 4617–4639, https://doi.org/10.5194/gmd-14-4617-2021, 2021.

Zhang, J. and Rothrock, D. A.: Modeling Global Sea Ice with a Thickness and Enthalpy Distribution Model in Generalized Curvilinear Coordinates, Mon. Weather Rev., 131, 845–861, https://doi.org/10.1175/1520-0493(2003)131<0845:MGSIWA>2.0.CO;2, 2003.