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 Sein[1], Longjiang Mu[6,1],
Uwe Fladrich[7], Dirk Barbi[8,1], Paul Gierz[1], Sergey Danilov[1,2], Stephan Juricke[1,2], Gerrit Lohmann[1,9], and
Thomas Jung[1,9]

[1]Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, Germany
[2]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]Pilot National Laboratory for Marine Science and Technology, Qingdao, China
[7]Swedish Meteorological and Hydrological Institute, Folkborgsvägen 17, SE-60176 Norrköping, Sweden
[8]Rhenish Friedrich Wilhelm University of Bonn, Regina-Pacis-Weg 3, 53113 Bonn, Germany
[9]University of Bremen, Bibliothekstraße 1, 28359 Bremen, Germany

Correspondence: Jan Streffing (Jan.Streffing@awi.de), 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 has the multi-resolution functionality typical for unstructured-mesh models while still featuring a scalability and efficiency similar to regular-grid models. The atmospheric component OpenIFS (CY43R3) enables the use of latest developments in the

5   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 on-going research activities with AWI-CM3. This includes the exploration of high and variable resolution, 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

10 CMIP6-average skills 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
15   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 AWI-CM1.1-MR (Semmler et al., 2020) as well as AWI-ESM1.1-LR (Danek et al., 2020), and 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 model-
20   ing 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), precipitation (van Haren et al., 2015), as well as mean sea level pressure (Hertwig et al., 2015), can be improved by increased horizontal resolution.

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
25   its numerical cores and in the peripheries, such as data structures, and IO schemes. As a first step forward, AWI embarked on a mission to create a FESOM version 2.0 with a finite volume numerical core instead of finite elements (Danilov et al., 2017; Scholz et al., 2019; Koldunov et al., 2019). Following that, a new coupled model was developed, featuring FESOM2 with the same atmospheric model as before, ECHAM6 (Sidorenko et al., 2019a). 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.

30   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 model version 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
35   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 modelling.

While it might be tempting to apply the coupled model at the highest possible spatial resolution right from the start, cost and time concerns dictate that initial development and evaluation best be done at lower resolution. Furthermore many future
40   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 high-resolution capabilities. What we consider low
45   resolution for the OpenIFS atmosphere TCo159L91 (61km) is already beyond the practical limits of our previous AWI-CM1 and AWI-CM2 models with ECHAM6 (100 km).
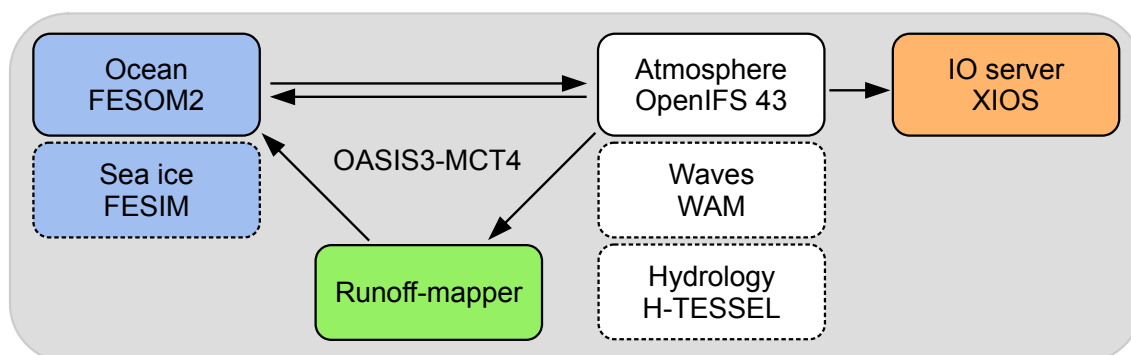
Figure 1. Schematic of the coupling between AWI-CM3 model components. Two hourly parallel communication in AWI-CM3 is implemented via the OASIS3-MCT 4.0 library. Heat, mass and momentum fluxes are sent from OpenIFS to FESOM2, while ocean and ice surface state variables are sent back. The precipitation 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.

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 (XIOS), running in parallel, handles the output
50 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 testbed for developments
55 that can feed back into weather forecasting.

As such, OpenIFS contains the hydrostatic dynamical core, 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. In contrast to the operational IFS Numerical Weather Prediction (NWP) system, the 4D-var data assimilation, the non-hydrostatic core, the adjoint/tangent-linear versions, the Meteo France IO server, and the subroutine level implementation of the NEMO ocean
60 model have been removed from OpenIFS.

The defining feature of the numerical core is its semi-Lagrangian semi-implicit 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 timesteps than equivalent Eulerian integrations. What the model saves in terms of computing resources can be invested
65   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.

OpenIFS is available at a wide variety of horizontal resolutions, ranging from a TQ21 (626km) toy model to the TCo1279
(9km) operational NWP. Even higher experimental horizontal resolutions require only minor source code changes. Three op-
70   tions exist for the representation of highest (truncation) resolution spherical harmonics in grid point space. The linear truncation
TL grids with two gridpoints for the smallest spherical harmonics, the quadratic TQ with three 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 cli-
mate 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
75   parameterizations have been tuned.

The main experiments we present were performed at a resolution of TCo159 (61km) with 91 vertical layers (TCo159L91),
with some experiments at TCo319L137 (31km). Further resolutions we successfully tested computationally are TCo95L91
(100km) and TCo639L137 (16km).

2.2   FESOM2 Ocean

80   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 the Alfred Wegener Institute (Danilov et al., 2017).

FESOM2 is formulated with a finite volume dynamical core using arbitrary Lagrangian Eulerian (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
85   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 (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 previous AWI-CM1 and AWI-CM2 versions the sea ice thermodynamics was solved in the
atmospheric component.

90   For the experiments presented here we employ the mesh called CORE2, which has about 127k surface nodes as shown
in Koldunov et al. (2019). 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 27km. We show the spacial distribution of
horizontal resolution for both meshes in Figure 2.

95   The low-resolution CORE2 mesh for FESOM2 is not eddy-resolving, and we employ the Gent-McWilliams (GM) parame-
terization 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 SSH is distributed
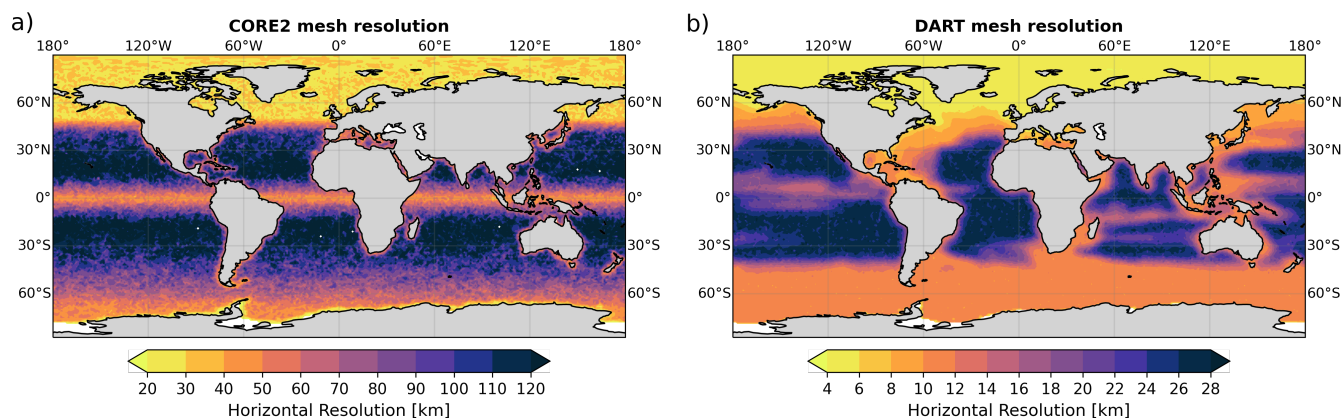
Figure 2. a) Horizontal resolution of the CORE2 mesh used for the main simulations presented in this work. **b)** As a) but for the DART mesh with higher resolution, which is based on a mix of local Rossby radius of deformation and local sea surface height variability. The DART mesh is used in Section 5.1 for an outlook of high resolution applications.

over all layers in the vertical, except the layer that touches the bottom (Scholz et al., 2019). Furthermore, the net evaporation - precipitation - runoff (E-P-R) integrated over the global ocean is forced to 0 at each timestep for the experiments shown.

100   As a design choice FESOM2 requires the world's oceans to be described as one continuous basin. A split into multiple meshes would result in more complex data structures, and ocean dynamics tend to play a lesser role in large lakes compared to the world oceans. Therefore, 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, and have to be 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 105   are: the White Sea, Persian Gulf, Black Sea, and the Gulf of Ob. Such narrow inlets on a coarse mesh would reduce the smallest horizontal grid spacing, and thus incur a smaller timestep in order to still fulfill the CFL condition globally.

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 110   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 Meteo France for IFS when generating a new release of OpenIFS. This leaves OpenIFS with only the possibility to provide sequentially written GRIB file output.

While this sequential IO scheme has been used, for example, by Döscher et al. (2021), it can often reach data-throughput limits in practical applications. Furthermore, while GRIB files are common in the NWP community, the climate modelling community uses NetCDF files as industry standard. 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. (2021) 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 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 of computing cost, as shown in table 2. Furthermore, XIOS enables 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 work flow. Finally XIOS allows for output directly in NetCDF, facilitating the use of AWI-CM3 model output.

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.

Computed in FESOM2	Computed in OpenIFS	Computed in Runoff-mapper
Sea surface temperature	Zonal wind stress, ocean	River runoff
Sea ice surface temperature	Meridional wind stress, ocean	
Sea ice concentration	Zonal wind stress, sea ice	
Sea ice thickness	Meridional wind stress, sea ice	
Snow thickness	Non-solar heat flux ocean	
	Solar heat flux	
	Total heat flux sea ice	
	Liquid precipitation	
	Solid precipitation	
	Evaporation	
	Sublimation	
	Precipitation - evaporation - soil moisture	

Table 1. Coupling fields sent by the respective components of the coupled model. Precipitation - evaporation - soil moisture over land is send to the Runoff-mapper and from there, as River runoff, to the FESOM2. All other fields are exchanged between OpenIFS and FESOM2 directly.

3.1 Coupling strategy

The surface heat, mass and momentum fluxes are calculated within OpenIFS and supplied to FESOM2. Here the state variables
145 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 is taking place at the least common multiple of the ocean and atmosphere timesteps. For the TCo159L91-CORE2 simulations the timesteps for cou-
150 pling, the atmospheric model and the oceanic model are 120, 60 and 40 minutes respectively, while for the TCo319L137-DART simulation they are 60, 15 and 4 minutes. In the production mode both the atmosphere and 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 modelling community, as they allow for parallel execution of model components
155 (Lemarié et al., 2015; Marti et al., 2021).

AWI-CM3 can also be run in a sequential atmosphere-first mode, updating the ocean at timestep t_n with atmospheric fluxes from t_n. In this mode climate models can get very close to a converged solution 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.

160   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 realized AOGCM
setup is sketched in Figure 1. Since FESOM2 has previously been coupled to the atmospheric model ECHAM6 (Sidorenko
et al., 2019b), 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
165   CY40R1 was coupled via OASIS as a test case in EC-Earth 3. The coupling of interface for OpenIFS CY43R3 is inspired by
this predecessor. Future releases of OpenIFS will be published with this coupling interface already included. A full breakdown
of the fields computed in and sent by each component is given in table 1.

4   Climatological Performance

In this section we will outline the ability of the new coupled system to reach a stable equilibrium with constant greenhouse
170   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 4x increase of CO_2, and the other with a constant increase of 1% per year, starting from 1850 values.

4.1   Spinup drift (SPIN)

175   A 700-year long spinup of AWICM3 was carried out under constant greenhouse gas and solar forcing from the year 1850.
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
180   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 Figure 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 $2\,W/m^2$, depicted in
Figure 3a. A partial solution is to switch the OpenIFS mass fixer from dry mass to total mass conservation (Malardel et al.,
185   2019) with the McGregor scheme (McGregor, 2005). We implemented this solution starting from year 1651. Subsequently the
radiative imbalance reduced to $+0.7\,W/m^2$, which is within the range of imbalance of CMIP6 models (Wild, 2020).

Further experiments showed that decreasing the OpenIFS timestep from 60 minutes to 30 minutes reduced the imbalance to
$+0.4\,W/m^2$. An additional reduction to 15 minutes timesteps showed no more improvements. We surmise that the timestep-
dependent component of the error implicates the semi-lagrangian trajectory algorithm operating close to the stability limit for
190   the given atmospheric resolution of TCo159. For our analysis we judged this additional error an acceptable price for a doubling
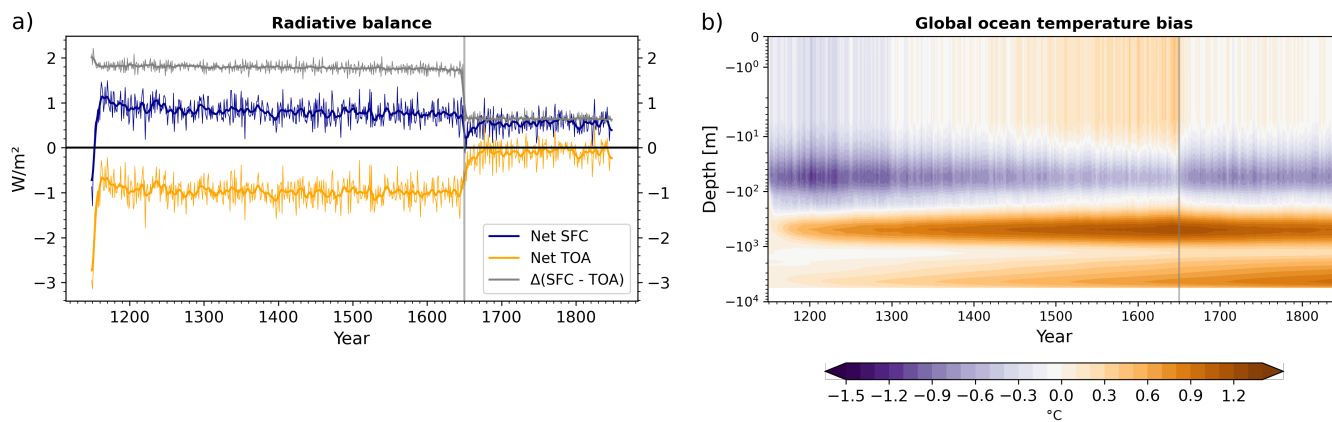
Figure 3. a) Net radiative imbalance at the top of atmosphere (TOA) and at the surface (SFC) in the spinup simulation. Positive (negative) values indicate downwards (upwards) net heat flux. The difference results from spurious energy production in the atmosphere. **b)** Semi-logarithmic depth Hovmöller 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 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 meters. Warming at depth continued.

in model integration speed, that is, we keep the 60 min timestep. The timestep independent flux imbalance of $+0.4\,W/m^2$ will be targeted in future model development and tuning efforts.

In the global mean ocean temperature Hovmöller diagram (Figure 3) we can see that after switching to total mass conservation the reduction of spurious heat production in the atmosphere led to a stabilization of the upper 1000 meters for ocean

195 temperature. The trend in the deep ocean, strongest at 4500 meters 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 Gibraltar Straight 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., 2019b).

200 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 gets 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 $5-10\,W/m^2$ (Forbes et al., 2016). We have already back-ported these improvements originally developed for the operational IFS CY45 model into OpenIFS CY43R3 prior to our spinup simulation. Even with

205 the improvements, a positive shortwave downward radiation bias of up to $5-10\,W/m^2$ between 45-60°S remains, and can be seen in Figure 7 e). 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 (Figure 3b), the spinup run is not yet in equilibrium at depths greater than 1000 meters.
210 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 OGCM 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
215 solar forcing. We construct a simple linear regression 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
220 to the pre-industrial control.

For a succinct overview of the model performance we calculated climate model performance indices, based on the ideas of Reichler and Kim (2008) and shown in Figure 4. For the four seasons, seven regions and nine key variables, the mean absolute errors of AWI-CM3 with respect to observations are expressed as a fraction of the average of the mean absolute errors of 32 CMIP6 contributing models. A complete list of the CMIP6 models serving as the evaluation set is given in Appendix B.
225 The list of observational datasets used to calculate all mean absolute errors is also given in Appendix B. For all model and observational datasets, where available, the time period from December 1989 until November 2014 of the historic period is considered. As the performance of AWI-CM3 is expressed as a fraction of the errors of the CMIP6 average performance, values below 1 indicate better performance, values above 1 point to larger model errors.

For the vast majority of variables, seasons and regions our post-CMIP6 prototype model is already performing better than
230 the average CMIP6 model. The lead is especially large for cloud cover (clt), 500 hPa geopotential height (zg) and surface zonal winds (uas). For surface meridional wind (vas), 300 hPa zonal wind (ua), TOA outgoing longwave radiation (rlut), and precipitation (pr) the bias to observations is still below average. The very important near surface (2m) air temperature (tas) is simulated well in the Arctic (60°-90°N) and reasonably well in the northern mid-latitudes (30°-60°N), tropics (30°S-30°N) and inner tropics (15°S-15°N). In the southern mid-latitudes (30°-60°S) and Antarctic (60°-90°S) the air temperature bias is
235 relatively high. Finally, 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.

A simple average over all individual performance indices gives AWI-CM3 a score of 0.854, placing it at rank 13 out of the 32 considered CMIP6 models. The overall index of the AWI-CM3 prototype simulation is comparable to the similarly expensive low-resolution CMIP6 run of AWI-ESM1.1-LR (0.887). The performance of our medium-resolution AWI-CM1.1-MR CMIP6
240 contribution is better (0.774), but this model configuration is 20 times more expensive to run than the simulations presented
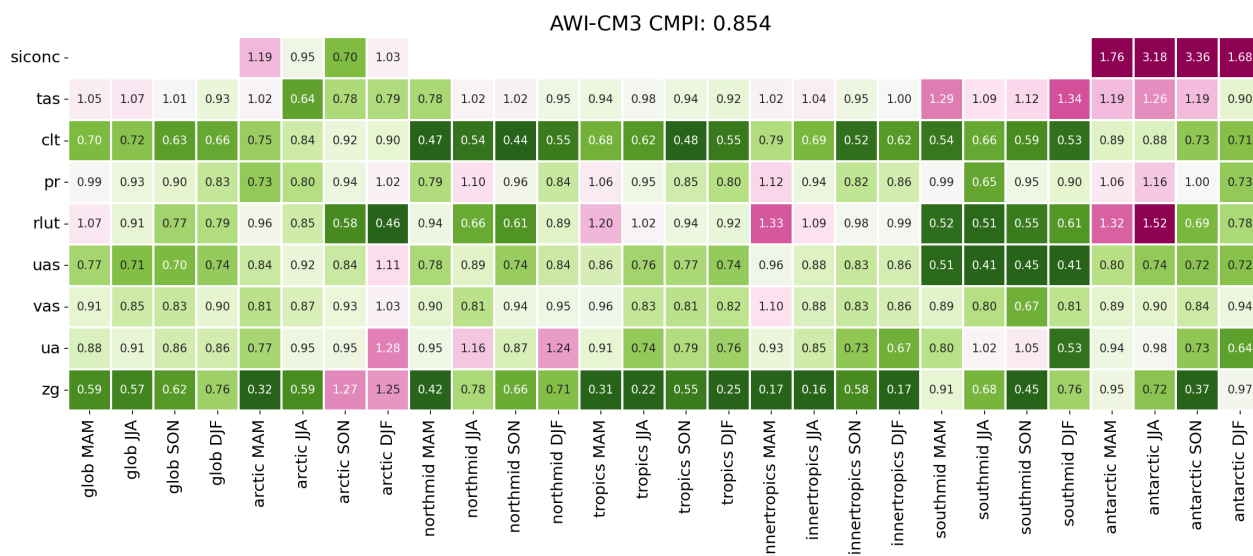
Figure 4. Mean absolute error of the climatology for the last 25 years of the AWI-CM3 historic simulations with respect to satellite observation and reanalysis data, expressed as a fraction of the errors averaged over CMIP6 models. Values below (above) one correspond to below (above) CMIP6 average biases. The underlying observations against which all models were evaluated are OSISAF 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); NCEP-DOE Reanalysis 2 (Kistler et al., 2001): near-surface air temperature (tas), eastward near-surface wind (uas), northward near-surface wind (vas), 300 hPa eastward wind (ua), 500 hPa geopotential height. A full list of the considered CMIP6 models is given in Appendix B.

here. Preliminary tests with higher resolution at equal computational cost indicate that the AWI-CM3 simulation can achieve better climatological performance than AWI-CM1.1-MR, as we will show in Section 5.1. A major contributing factor are the faster dynamic cores and better computational scalability of the model components, allowing for higher resolutions at equal computational cost.

245   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 end of summer (EOS) and winter (EOW), PICT and HIST runs (Figure 5). In the central Arctic, PICT sea ice thickness ranges from 3.5m at the EOW to 2.5m at the EOS. The mean over 250   the last 25 years of the HIST simulation reveals a reduction of sea ice thicknesses to 2.5m and 1.5m 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 presented 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

255 wide tongue of sea ice extending eastward 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 to 7 million square kilometers in the last 25 years, compared to observational values of 15 to 6 million square kilometers by Walsh et al. (2019).

The Antarctic sea ice biases require the most improvement as follows from the metrics presented in Figure 4. In both the PICT and HIST simulations the sea ice covered area is strongly underestimated during austral winter (Figure 5i). Notable are

260 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 20cm 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 their occurrence (not presented in this paper) is clearly overestimated in our model.

In order to gain an understanding towards the reasons for the persisting polynya, we investigate the mean mixed layer depth

265 (MLD), defined as the depth where the local buoyancy gradient matches the maximum buoyancy gradient (Large et al., 1997). The MLD shown in Figure 6 features large MLD values in the Weddell and increased values in the Ross seas. 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 below. The reason for the overestimated MLD is not clear and understanding it is among our planned future research.

270 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. Alternatively, the issue could also be triggered by the positive air temperature bias, leading to increased MLD. This more complex causal chain will be the subject of future work.

4.3.2 Surface temperature and fluxes

275 The near surface (2m) air temperature bias in comparison to ERA5 largely shows a pattern similar in sign (Figure 7a), 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 percipitation 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

280 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 North West Corner detachment (Figure 7a). This is a well known bias in OGCMs of coarser than 10 km resolution in the North

285 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.
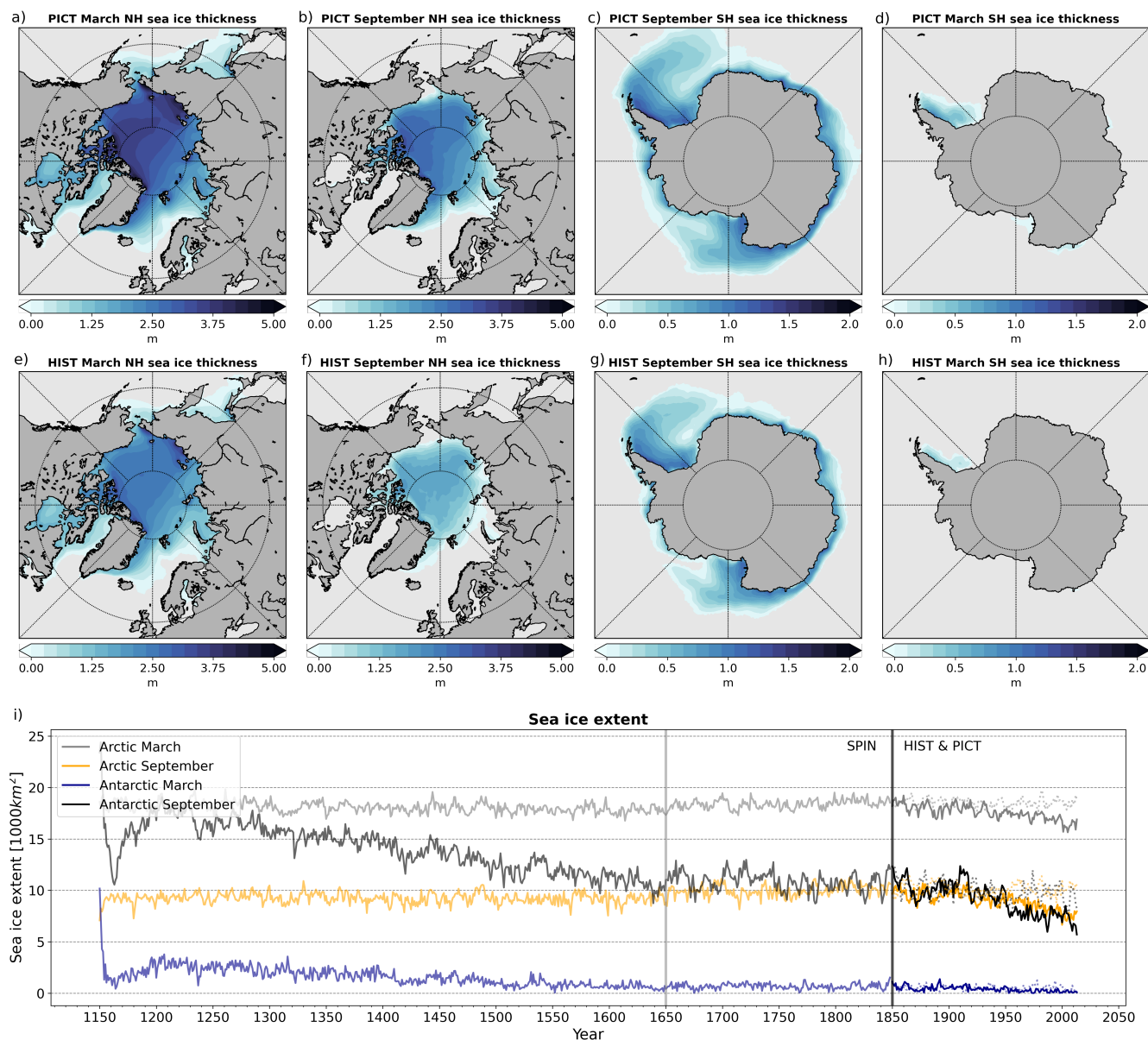
Figure 5. a-d) Mean sea ice volume per unit area over last 25 years of PICT simulation. **e-h):** Same as top row but for the HIST simulation. **i):** Evolution of sea ice extent over SPIN, as well as HIST and PICT simulations. Mass fixer switched from dry to total in years 1650. HIST and PICT forcing applied from 1850 onwards.

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. Some of these biases are well known for IFS based climate models and can also be seen in Döscher et al. (2021) and Roberts et al. (2018). We hypothesize that these biases are caused by the direct effect
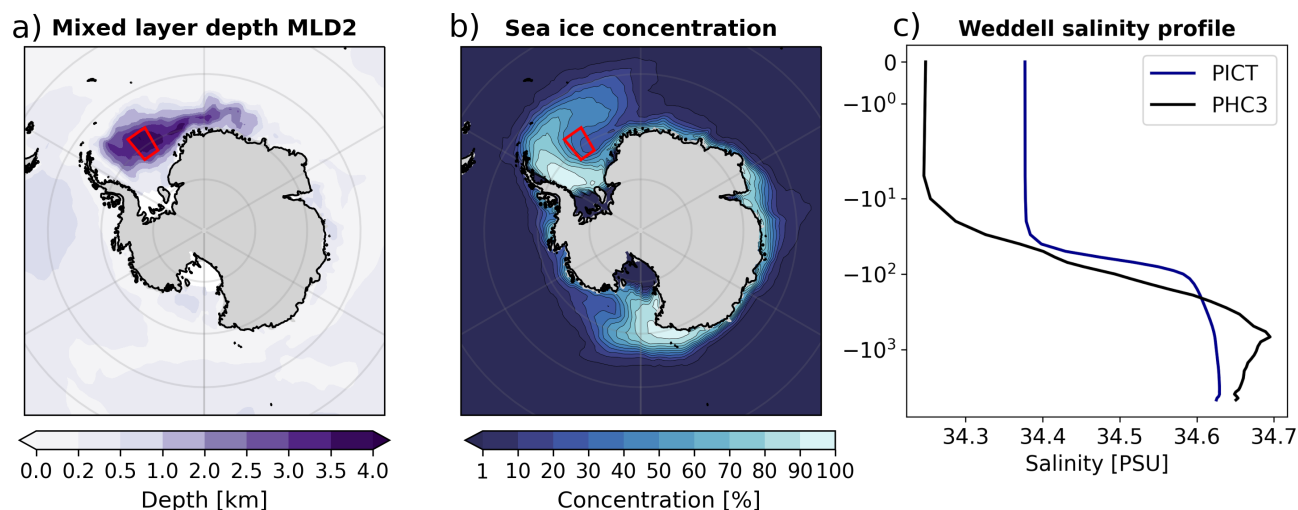
Figure 6. a) Austral winter mean mixed layer depth (MLD) according to where the local buoyancy gradient matches the maximum buoyancy gradient (Large et al., 1997) during the last 25 years of the PICT simulation. **b)** High MLD values in the Weddell and Ross seas contribute to the persistent polynyas visible in the mean austral winter sea ice concentration over the same time period. **c)** Salinity profile over the region marked in red in comparison to PHC3 observational data (Steele et al., 2001)

290   of heat released from a spuriously deep mixed layer (as noted in Section 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 (Figure 7a) can be associated with co-located net surface shortwave radiation biases (Figure 7c). These can be largely explained by surface downward radiation biases (Figure 7e), which in turn result from total cloud cover fraction biases (Figure 7f). This linkage pattern is well known in the modelling community (Satoh et al.,
295   2019) and will likely remain a dominant source for surface temperature biases until deep convection resolving atmospheric models become readily available for multidecadal 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
300   reduction of 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 shortwave 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.
305   Indeed, the near-surface (2m) 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 (Figure 4). Correcting the largest biases south of 60°S in our model will likely necessitate work on non-solar heat fluxes and mixed layer depths at high-latitudes.
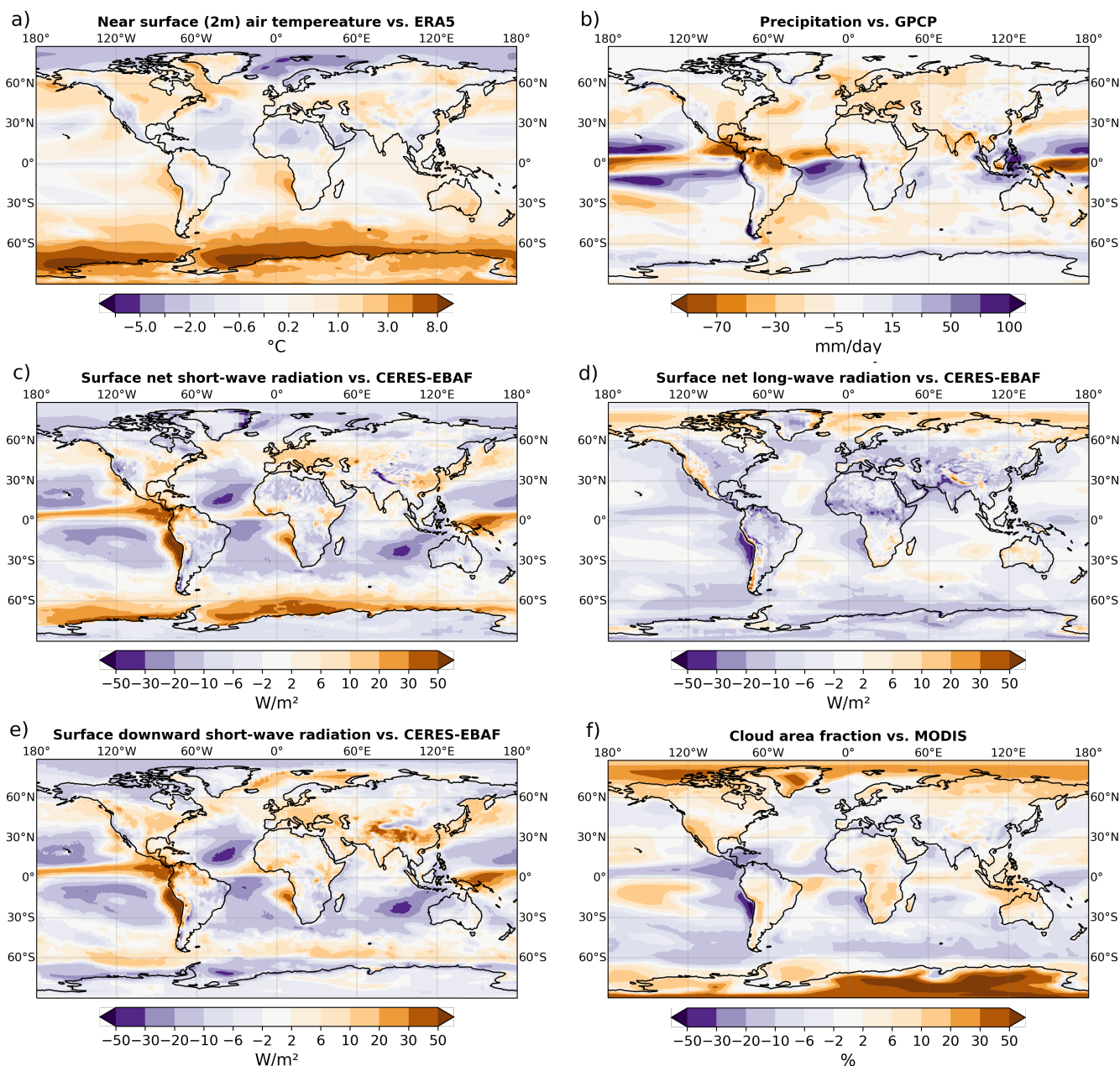
Figure 7. a) Near surface (2m) air temperature bias with respect to ERA5. **b)** Precipitation bias with respect to GPCP. **c)** Surface net shortwave radiation with respect to CERES (2001-2014). **d)** Surface net longwave radiation with respect to CERES (2001-2014). **e)** Surface downward shortwave radiation bias with respect to CERES (2001-2014). **f)** Total cloud cover fraction bias with respect to MODIS.

The precipitation biases shown in Figure 7b feature the canonical double ITCZ bias in the Tropics. Notably the mid-latitudes receive too little precipitation, especially over western boundary currents, Europe and North America east of the Rocky Mountains.
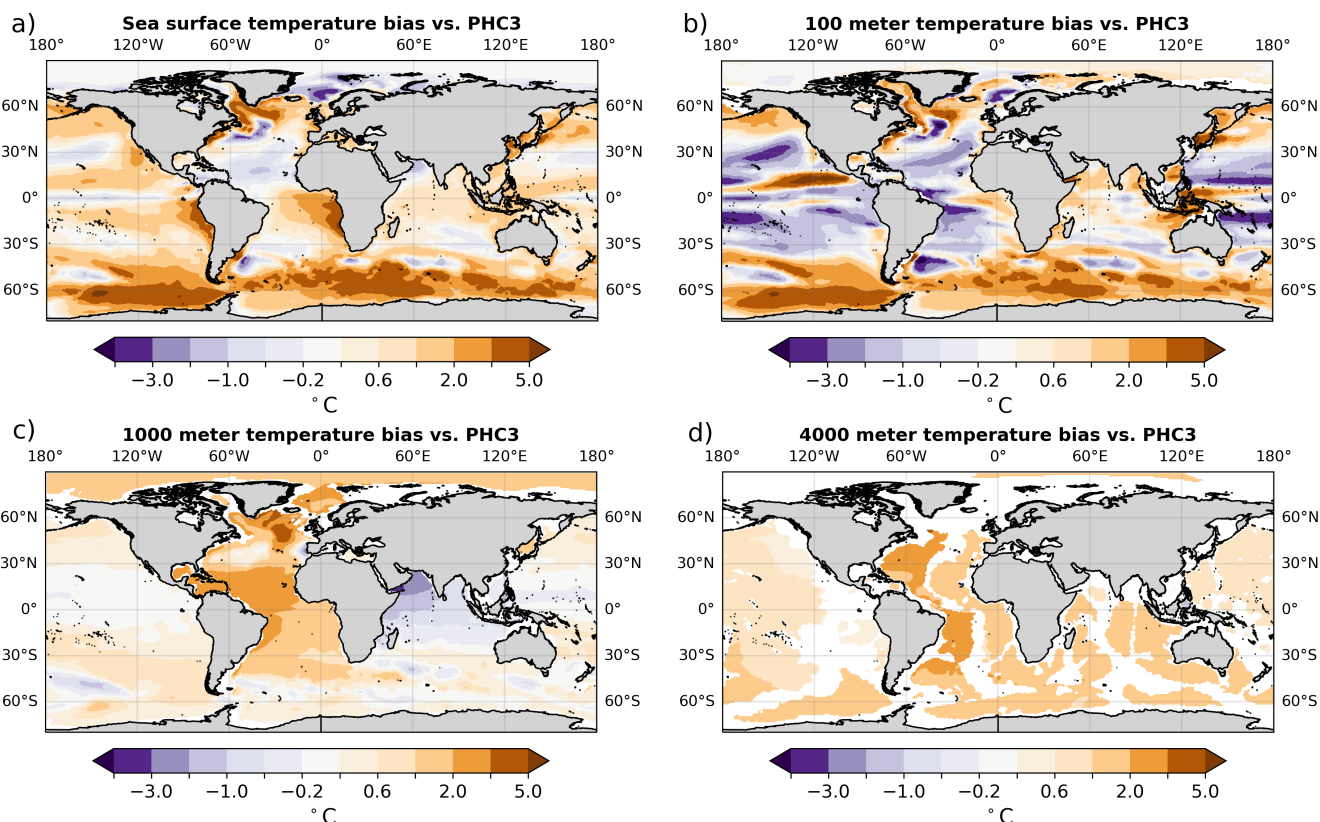
Figure 8. Mean ocean temperature bias of the last 25 years of the historical simulation with respect to PHC3 at **a)** 0, **b)** 100, **c)** 1000 and **d)** 4000 meters depth. At the surface the ocean-resolution-specific warm biases in upwelling regions and western boundary currents can be seen, in addition to a pronounced Southern Ocean warm bias. At depth, a warm bias is dominating in the Atlantic.

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 (Figure 8). At a depth of 100 meters 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).

At a depth of 1000 meters 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 found in the Indian Ocean at this depth is likely the result of insufficient warm outflow

315

320

from the Persian Gulf because the narrow strait is not well resolved on the FESOM2 CORE2 mesh. At 4000 meters depth we observe the warm biased Atlantic water mass spreading into the whole global ocean. As the meridional circulation at this depth is northward 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

325 slow but continued global-mean warming trends at great depths seen in Figure 3. Biases presented in 8 explain the Hovmöller diagram in Fig 3 which we addressed in Section 4.1. The three sets of anomalies in the Hovmöller diagram at depths 100m, 1000m and 4000m stem from the biases in the mid latitudes, the North Atlantic and in 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 Nino Southern
330 Oscillation (ENSO). To this end, we selected the Nino3.4 box (5°N to 5°S, 120°W to 170°W) and calculated the area mean SST within the box. We then applied a linear detrending, subtracted the mean seasonal cycle, and finally computed a three month running mean. The resulting Nino3.4 index timeseries is shown in Figure 9c. Comparison with the Nino3.4 index based on the HadISST observational estimates (Figure 9d) reveals that our simulated amplitude of ENSO, with a range of +1.9 to -1.6°C, is lower than the observed. Indeed, the histograms in Figures 9e and 9f reveal missing tails of the distribution in
335 the simulation. While the overall strength of ENSO is underestimated by AWI-CM3, the normalized power spectrum density (PSD) in Figure 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.

Future improvements for the ENSO amplitude could be achieved by activating the OpenIFS internal stochastic parametrization schemes, as shown by Yang et al. (2019). Indeed, HighresMIP simulations performed by Roberts et al. (2018) with IFS
340 CY43 that employed the Stochastically Perturbated Parametrisation 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 Parametrisations (SPP) scheme (Ollinaho et al., 2017), the full version of SPP has to either be backported from IFS CY47 to OpenIFS CY43R3, or the new version OpenIFS CY47 has to be released.
345 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).

4.4 Impact of historical forcing

350 ### 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 (2m) air temperature increases under HIST forcing, as seen in Figure 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

355 control experiment PICT, and obtain the residual drift of $0.00091°C$/Year via linear regression. Over the 165-year-long simulation this amounts to a drift of 0.15°C in the PICT run.

In Figure 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

360 right, however the strength is overestimated by 40% ($1.0°C$ 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 (private communication with the EC-Earth aerosol working group). The sixth IPCC report (Masson-Delmotte

365 et al., 2021) concludes that the anthropogenic aerosols direct and indirect contributions to the effective radiative forcing amount to about $-0.5W/m^2$. Comparing to the total anthropogenic forcing of $+1.5W/m^2$, we foresee that, once transient aerosols will have been included in OpenIFS CY43R3, AWI-CM3 historical global warming will be much closer to the observed value.

The map of temperature changes in Figure 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

370 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 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

375 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 AMOC (Keil et al., 2020), but although our historic simulation does exhibit such an AMOC weakening (see Figure 12), no warming hole forms.

4.4.2 Precipitation

Figure 11b shows the simulated changes in the precipitation pattern resulting from historic well-mixed greenhouse gas and

380 solar forcing. The most important features are that the high-latitudes nearly uniformly receive more precipitation, the monsoon in North Africa and China intensifies and the ITCZ in the Pacific is enhanced in the west and weakened in the east.

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

385 the northward flow of AABW (Figures 12 a,b). During the spinup simulation the maximum of the northward transport between

30-45°N in AWI-CM3 fluctuates at around 20 Sv (see Figure 12c.). Initially the AMOC gradually slowed down to 18 Sv while the upper ocean was spuriously warming, as described in Section 4.1. After applying the total mass fixer starting after 500 years into the SPIN experiment the AMOC is recovered to 20 Sv along with the cooling of the upper ocean (Figure 3c).

Figure 12d shows a rather strong decline of the AMOC strength over the last 70 years of the HIST simulation, while the multi decadal natural variability of 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 considerably 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 HIST run the decrease of 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 spurious transformations, 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 didn't activate in this run. However, this extra analysis was done by Sidorenko et al. (2021) and they concluded that the NAO was the main driver of AMOC variability.

Simulated volume transport fluxes were additionally computed across several major ocean straits (Table C2). Historical values averaged 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 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 increased, eddy-resolving ocean resolution are expected to provide even more accurate results (i) within narrow straits and (ii) in eddy-rich areas such as the Southern Ocean, where mesoscale activity is key in accurately depicting the ACC behaviour (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 $4xCO_2$ experiment which prescribes a sudden permanent increase of the CO2 concentrations to 4 times (1137.27 ppm) the base value (284.32 ppm) from

1850. It can be used to analyse the model's inherent climate sensitivity. As described by Gregory et al. (2004) the relationship between the change in net downward radiative flux and the change in near surface (2m) air temperature can be described by a linear regression model. The Gregory plot in Figure 13 was computed from the 4xCO$_2$ experiment in comparison to the pre-industrial control simulation. The first axis intersection point determines the instantaneous radiative forcing $F = 7.06\frac{W}{m^2}$ resulting from the quadrupling of CO2. The linear regression line intersects the second axis at an Equilibrium Temperature difference $\Delta T = 6.49\,°C$, resulting in a climate response parameter of AWI-CM3 of $\alpha = \frac{\Delta T}{F} = 0.92\frac{K}{W}$. Equilibrium Climate Sensitivity (ECS) is given by doubling rather than quadrupling CO2 concentrations, and is thus ECS=$\frac{\Delta T}{2} = 3.2\,°C$. With this ECS value the AWI-CM3 prototype finds itself near the center of the range predicted by CMIP6 models ($1.8 - 5.6\,°C$) (Scafetta, 2021), and just slightly above AWI-CM1, which had a value of $3.2\,°C$.

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 (2m) air temperature increase of 2.1°C after 70 years, when CO$_2$ concentrations have doubled, as shown in Figure 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) and slightly higher than that of AWI-CM1, at 2.1°C. Interestingly, a further doubling to a total of four 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 (2m) 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, 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 etc. 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 in table 2.

Tables 2 and 3 list the Simulated Years Per Day (SYPD) and core hours per simulated year (CHSY) values we achieved when optimization is done 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.

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	6769	4.61	35220

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@fz-juelich.de with Intel Xeon Platinum 8168 CPU, 2× 24 cores, 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 (Figure 2a) has 47 vertical layers, 127.000 surface nodes, a peak resolution of 20km in Northern mid to high latitudes and up to 125km in Subtropical Gyres. The horizontal resolution of the DART mesh follows the local Rossby radius, peaks at a highest resolution of 5km and has a maximum spacing of 27km, with ∼3.1 million surface nodes and 80 vertical layers.

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

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.

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 Figure 14 we show the performance of a 31km atmosphere with 137 vertical layers, coupled to a high-resolution ocean mesh with 3.1 million surface nodes and 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.

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 analysed so far. The improvement is almost universal with only the sea ice concentrations 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 the shorter run length or different forcing.

Obviously the improved climatological performance comes at a cost, as detailed in Section 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
465 higher resolution capabilities of AWI-CM3 will be subject of future work.

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.

470   We ran a set of experiments closely resembling the Coupled Model Intercomparison Project phase 6 (CMIP6) 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 (165y), a pre-industrial control run (165y), and two idealized experiments, one with a sudden
475   increase to $4xCO_2$ (120y) and the other with 1% CO_2 (150y) increase per year.

Climate sensitivity experiments with AWI-CM3 obtained an Equilibrium Climate Sensitivity and Transient Climate Response of 3.2°C 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, cloud fraction, 500 hPa
480   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 biases in shortwave downward radiation associated with cloud fraction between 45-60°S.

AWI-CM3 is capable of realistically simulating the Atlantic meridional overturning circulation (AMOC) in terms of both
485   the shape and strength of the streamfunction, as well as 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 power 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 fluxes that are at least one magnitude
490   smaller than the ones 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 further reduce biases. Based on literature we suggest porting and activating the Stochastically Perturbed Parametrisations scheme as a potential way to improve ENSO amplitude.

The advanced computational efficiency and scalability of AWI-CM3, combined with a very solid model and coupling physics
495   implementation, will eventually enable us to perform full DECK and scenario simulations at resolutions of 5-25km, that

were previously reserved for the high end of the HighResMIP protocol. We provide a preview with a shorter high-resolution simulation, indicating that AWI-CM3 climatological biases at future operational resolution will be about half those of the average CMIP6 model.

Code and data availability. The ocean model FESOM2 source code is available on Zenodo at $10.5281/zenodo.6335383$ and at $https : $
500 $//github.com/FESOM/fesom2/releases/tag/AWI-CM3\_v3.0$. OpenIFS is not publicly available but rather subject to licecing by ECMWF. However licences are readily given free of charge to any academic or research institute. All modifications required to enable AWI-CM3 simulations with OpenIFS CY43R3V1 as provided by ECMWF can be obtained on Zenodo at: $10.5281/zenodo.6335498$. The OASIS coupler is available upon registration at: https://oasis.cerfacs.fr/en/downloads/. The XIOS source code is available on Zenodo ($10.5281/zenodo.4905653$ Meurdesoif, 2017) and on the official repository ($http : //forge.ipsl.jussieu.fr/ioserver$, last ac-
505 cess: 4 March 2022). The runoff mapper scheme is available on Zenodo at $10.5281/zenodo.6335474$. The compile and runtime engine esm-tools is available on Zenodo at: $10.5281/zenodo.6335309$. All data required to reproduce the plots shown here can be found at: $10.5281/zenodo.6337593$, $10.5281/zenodo.6337571$, and $10.5281/zenodo.6337627$. The processing and plotting scripts for the reproduction of the shown analysis can be found at: $10.5281/zenodo.6335530$. Documentation of AWI-CM3 and a user guide can be found at: $awi-cm3-documentation.readthedocs.io$

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

AWI-ESM-1-1-LR, IITM-ESM, MIROC6, CMCC-CM2-SR5, CanESM5, INM-CM5-0, MCM-UA-1-0, MPI-ESM-1-2-HAM, KACE-1-0-G, GFDL-CM4, CNRM-CM6-1, FGOALS-f3-L, UKESM1-0-LL, BCC-CSM2-MR, FGOALS-g3, EC-Earth3, HadGEM3-GC31-MM, IPSL-CM6A-LR, GISS-E2-1-G, E3SM-1-1, CESM2, AWI-CM-1-1-MR, ACCESS-ESM1-5, CAMS-CSM1-0, KIOST-ESM, NorESM2-LM, SAM0-UNICON, NESM3, CIESM, TaiESM1, CAS-ESM2-0, MRI-ESM2-0, ICON-ESM-LR

515 Appendix C: Table of observational datasets for climate model performance index calculation

Appendix B: Additional figures

Author contributions. J. Streffing: Conceptualization, coupling methodology, coding, low resolution numerical experiments, analysis, and writing original draft preparation; D. Sidorenko, J. Kjellsson, U. Fladrich and L. Zampieri: Coupling methodology, coding; N. Koldunov, D. Sidorenko and T. Rackow: development of analysis methods for unstructured grids; H. Goessling, Q. Wang, L. Mu and S. Danilov, T.
520 Jung: Conceptualization; T.Semmler and D. Sein: high resolution numerical experiments; M. Athanase: Analysis and writing original draft of transports Section; M. Andrés-Martínez, D. Barbi and P. Gierz: compile, runtime optimization; S. Danilov, S. Juricke, G. Lohmann, and T. Jung: supervision, funding acquisition. All authors discussed, read, edited, and approved the article. All authors have read and agreed to the published version of the manuscript.

Variable	Longname	Dataset	Time range
tas	Near surface (2m) air temperature	NCEP2 Reanalysis	1989/11/01 to 2014/11/30
uas	Near surface (10m) zonal wind speed	NCEP2 Reanalysis	1989/11/01 to 2014/11/30
vas	Near surface (10m) meridional wind speed	NCEP2 Reanalysis	1989/11/01 to 2014/11/30
Ua (300 hPa)	300 hPa zonal wind speed	NCEP2 Reanalysis	1989/11/01 to 2014/11/30
Zg (500 hPa)	500 hPa geopotential height	NCEP2 Reanalysis	1989/11/01 to 2014/11/30
pr	Precipitation flux	GPCP v2.3	1989/11/01 to 2014/11/30
siconc	Sea ice area fraction	OSISAF OSI-450	1989/11/01 to 2014/11/30
rlut	TOA outgoing longwave flux	CERES-EBAF	2000/03/15 to 2014/06/30
clt	Cloud area fraction	MODIS Atmosphere L2	2000/03/15 to 2014/11/30

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

Competing interests. At least one of the (co-)authors is a member of the editorial board of Geoscientific Model Development.

525 *Acknowledgements.* This paper is a contribution to the projects L4, S1 and S2 of the Collaborative Research Centre TRR 181 "Energy Transfers in Atmosphere and Ocean" funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Project 274762653. We thank the Jülich Supercomputing Centre for providing a share of the JUWELS ESM Partition under the compute projects chhb19, chhb20 and cesmtst as well as the data storage project hhb19. The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time through the John von Neumann Institute for
530 Computing (NIC) on the GCS Supercomputer JUWELS at Jülich Supercomputing Centre (JSC). We also thank the Helmholtz ESM-Project for technical support. H. Goessling and M. Athanase acknowledge funding by the Federal Ministry of Education and Research of Germany in the framework of SSIP (grant01LN1701A).

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)

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

References

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.,
535   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, 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 Modelling, 137, 76–97, https://doi.org/https://doi.org/10.1016/j.ocemod.2019.02.005, 2019.

540   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 Wintertime North Atlantic Circulation, Cyclonic Activity, and Precipitation in a 25-km-Resolution Global Atmospheric Model, Journal of 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
545   system models in CMIP6, Geoscientific Model Development, 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, Journal of Hydrometeorology, 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
550   modular infrastructure for stand-alone and coupled Earth system modelling (ESM), Geoscientific Model Development, 14, 4051–4067, https://doi.org/10.5194/gmd-14-4051-2021, 2021.

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, Journal of Climate, 33, 6555 – 6581, https://doi.org/10.1175/JCLI-D-19-0970.1, 2020.

555 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, Journal of Geophysical Research: Atmospheres, 125, e2019JD032 321, https://doi.org/https://doi.org/10.1029/2019JD032321, e2019JD032321 2019JD032321, 2020.

Buizza, R., Milleer, M., and Palmer, T. N.: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system, Quarterly Journal of the Royal Meteorological Society, 125, 2887–2908, https://doi.org/https://doi.org/10.1002/qj.49712556006, 1999.

560 Caesar, L., McCarthy, G. D., Thornalley, D. J. R., Cahill, N., and Rahmstorf, S.: Reply to: Atlantic circulation change still uncertain, Nature Geoscience, 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, Geoscientific Model Development, 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
565   Strait, 2004–10, Journal of Physical Oceanography, 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, 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.

570 Danilov, S., Sidorenko, D., Wang, Q., and Jung, T.: The Finite-volumE Sea ice–Ocean Model (FESOM2), Geoscientific Model Development, 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, Journal of Advances in Modeling Earth Systems, 9, 2615–2634, https://doi.org/10.1002/2017MS001082, 2017.

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

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, Geophysical Research Letters, 43, 11,760–11,767, https://doi.org/https://doi.org/10.1002/2016GL070319, 2016.

580 Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arneth, A., Arsouze, T., Bergmann, T., Bernadello, R., Bousetta, 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.,
585 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 Climate Model Intercomparison Project 6, Geoscientific Model Development Discussions, 2021, 1–90, https://doi.org/10.5194/gmd-2020-446, 2021.

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arneth, A., Arsouze, T., Bergmann, T., Bernadello, R., Bousetta, S., Caron, L.-P., et al.: The EC-earth3 Earth system model for the climate model intercomparison project 6, Geoscientific Model Development Discussions, pp.
590 1–90, 2021.

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.

595 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, Journal of Physical Oceanography, 20, 150–155, 1990.

600 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, Geophysical Research Letters, 31, https://doi.org/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,
605 M., Scoccimarro, E., Semmler, T., Small, J., and von Storch, J.-S.: High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6, Geoscientific Model Development, 9, 4185–4208, https://doi.org/10.5194/gmd-9-4185-2016, 2016.

Hazeleger, W., Severijns, C., Semmler, T., Ştefănescu, S., Yang, S., Wang, X., Wyser, K., Dutra, E., Baldasano, J. M., Bintanja, R., et al.: EC-Earth: a seamless earth-system prediction approach in action, Bulletin of the American Meteorological Society, 91, 1357–1364, 2010.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al.:
610 The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 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, Climate Dynamics, 45, 185–211, https://doi.org/10.1007/s00382-014-2396-x, 2015.

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: Dynamic Meteorology and Oceanography, 68, 28 234,
615 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, Journal of Climate, 25, 3155 – 3172, https://doi.org/10.1175/JCLI-D-11-00265.1, 2012.

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

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, Bulletin of the American Meteorological Society, 82, 247 – 268, https://doi.org/10.1175/1520-0477(2001)082<0247:TNNYRM>2.3.CO;2, 2001.

625   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), Geoscientific Model Development, 12, 3991–4012, https://doi.org/10.5194/gmd-12-3991-2019, 2019.

630   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, Reviews of Geophysics, 32, 363–403, https://doi.org/https://doi.org/10.1029/94RG01872, 1994.

Large, W. G., Danabasoglu, G., Doney, S. C., and McWilliams, J. C.: Sensitivity to Surface Forcing and Boundary Layer Mixing in a 635   Global Ocean Model: Annual-Mean Climatology, Journal of Physical Oceanography, 27, 2418 – 2447, https://doi.org/10.1175/1520-0485(1997)027<2418:STSFAB>2.0.CO;2, 1997.

Lavergne, T., Sørensen, A. M., Kern, S., Tonboe, R., Notz, D., Aaboe, S., Bell, L., Dybkjær, G., Eastwood, S., Gabarro, C., et al.: Version 2 of the EUMETSAT OSI SAF and ESA CCI sea-ice concentration climate data records, The Cryosphere, 13, 49–78, 2019.

Lemarié, F., Blayo, E., and Debreu, L.: Analysis of Ocean-atmosphere Coupling Algorithms: Consistency and Stability, Procedia Computer 640   Science, 51, 2066–2075, https://doi.org/https://doi.org/10.1016/j.procs.2015.05.473, international Conference On Computational Science, ICCS 2015, 2015.

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.

645   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 CERFACS/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, 321, 2016.

Malardel, S., Diamantakis, M., Agusti-Panareda, A., and Flemming, J.: Dry mass versus total mass conservation in the IFS, European Centre 650   for Medium-Range Weather Forecasts, 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, Geoscientific Model Development, 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, 655   K., Lonnoy, E., Matthews, J., Maycock, T., Waterfield, T., Yelekçi, O., Yu, R., , and (eds.), B. Z.: 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, 2021.

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

Meredith, M. P., Woodworth, P. L., Chereskin, T. K., Marshall, D. P., Allison, L. C., Bigg, G. R., Donohue, K., Heywood, K. J.,
660   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, Reviews of Geophysics, 49, https://doi.org/https://doi.org/10.1029/2010RG000348, 2011.

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,
665   I. R.: An Updated Assessment of Near-Surface Temperature Change From 1850: The HadCRUT5 Data Set, Journal of Geophysical Research: Atmospheres, 126, e2019JD032 361, https://doi.org/https://doi.org/10.1029/2019JD032361, e2019JD032361 2019JD032361, 2021.

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

670   Münchow, A., Melling, H., and Falkner, K. K.: An Observational Estimate of Volume and Freshwater Flux Leaving the Arctic Ocean through Nares Strait, Journal of Physical Oceanography, 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, Quarterly Journal of the Royal Meteorological Society, 143, 408–422, https://doi.org/https://doi.org/10.1002/qj.2931, 2017.

675   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, Geophysical Research Letters, 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, 2015.

Rackow, T., Goessling, H. F., Jung, T., Sidorenko, D., Semmler, T., Barbi, D., and Handorf, D.: Towards multi-resolution global climate
680   modeling with ECHAM6-FESOM. Part II: climate variability, Climate Dynamics, 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, Nature Communications, 13, 637, https://doi.org/10.1038/s41467-022-28259-y, 2022.

685   Reichler, T. and Kim, J.: How Well Do Coupled Models Simulate Today's Climate?, Bulletin of the American Meteorological Society, 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, Journal of Physical Oceanography, 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 in-
690   terannual variability in the Mozambique Channel from moored current observations, Journal of Geophysical Research: Oceans, 115, https://doi.org/https://doi.org/10.1029/2009JC005619, 2010.

Ritchie, H.: Semi-Lagrangian Advection on a Gaussian Grid, Monthly Weather Review, 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
695   in a High-Resolution Version of the ECMWF Forecast Model, Monthly Weather Review, 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, Journal of Geophysical Research: Oceans, 100, 18 443–18 457,
       https://doi.org/https://doi.org/10.1029/95JC01673, 1995.

700 Roberts, C. D., Senan, R., Molteni, F., Boussetta, S., Mayer, M., and Keeley, S. P.: Climate model configurations of the ECMWF Integrated
       Forecasting System (ECMWF-IFS cycle 43r1) for HighResMIP, Geoscientific model development, 11, 3681–3712, 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 of a grid-based hydrologic model at the mesoscale, Water
705   Resources Research, 46, https://doi.org/https://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, https://doi.org/10.3390/cli9110161, 2021.

710 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,
       Journal of Climate, 30, 337–358, https://doi.org/10.1175/JCLI-D-16-0100.1, 2017.

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

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, Journal of Advances in Modeling Earth Systems, 9, 2601–2614,
720   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, Journal of Advances in Mod-
       eling Earth Systems, 12, e2019MS002 009, https://doi.org/https://doi.org/10.1029/2019MS002009, e2019MS002009 2019MS002009,
       2020.

725 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, Climate Dynamics, 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,
730   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, Journal of Advances in Modeling Earth Systems, 11, 3794–3815, https://doi.org/https://doi.org/10.1029/2019MS001696, 2019a.

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, Journal of Advances in Modeling Earth Systems, 11, 3794–3815, https://doi.org/10.1029/2019MS001696, 2019b.

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, Journal of Advances in Modeling Earth Systems, 13, e2021MS002 582, https://doi.org/https://doi.org/10.1029/2021MS002582, e2021MS002582 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 Science, 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, Journal of Climate, 14, 2079–2087, 2001.

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, Journal of 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, Journal of Geophysical Research: Atmospheres, 109, https://doi.org/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, Journal of Advances in Modeling Earth Systems, 11, 2177–2213, https://doi.org/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.[Sea Ice Area], Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center [Accessed: Jan 9, 2020], 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, Journal of Advances in Modeling Earth Systems, 11, 545–557, https://doi.org/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, Journal of 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, Bulletin of the American Meteorological Society, 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, Climate Dynamics, 55, 553–577, 2020.

Willison, J., Robinson, W. A., and Lackmann, G. M.: North Atlantic Storm-Track Sensitivity to Warming Increases with Model Resolution, Journal of 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, Progress in Oceanography, 160, 124–154, https://doi.org/https://doi.org/10.1016/j.pocean.2017.12.007, 2018.

770 Yamazaki, D., Kanae, S., Kim, H., and Oki, T.: A physically based description of floodplain inundation dynamics in a global river routing model, Water Resources Research, 47, https://doi.org/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, Climate Dynamics, 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
775 of Earth system models: IFS CY43R3 and XIOS 2.0 integration as a case study, Geoscientific Model Development Discussions, pp. 1–23, 2021.

Zeman, C., Wedi, N., Dueben, P., Ban, N., and Schaer, C.: Model intercomparison of COSMO 5.0 and IFS 45r1 at kilometer-scale grid spacing, Geoscientific Model Development, 14, 4617–4639, https://doi.org/10.5194/gmd-14-4617-2021, 2021.
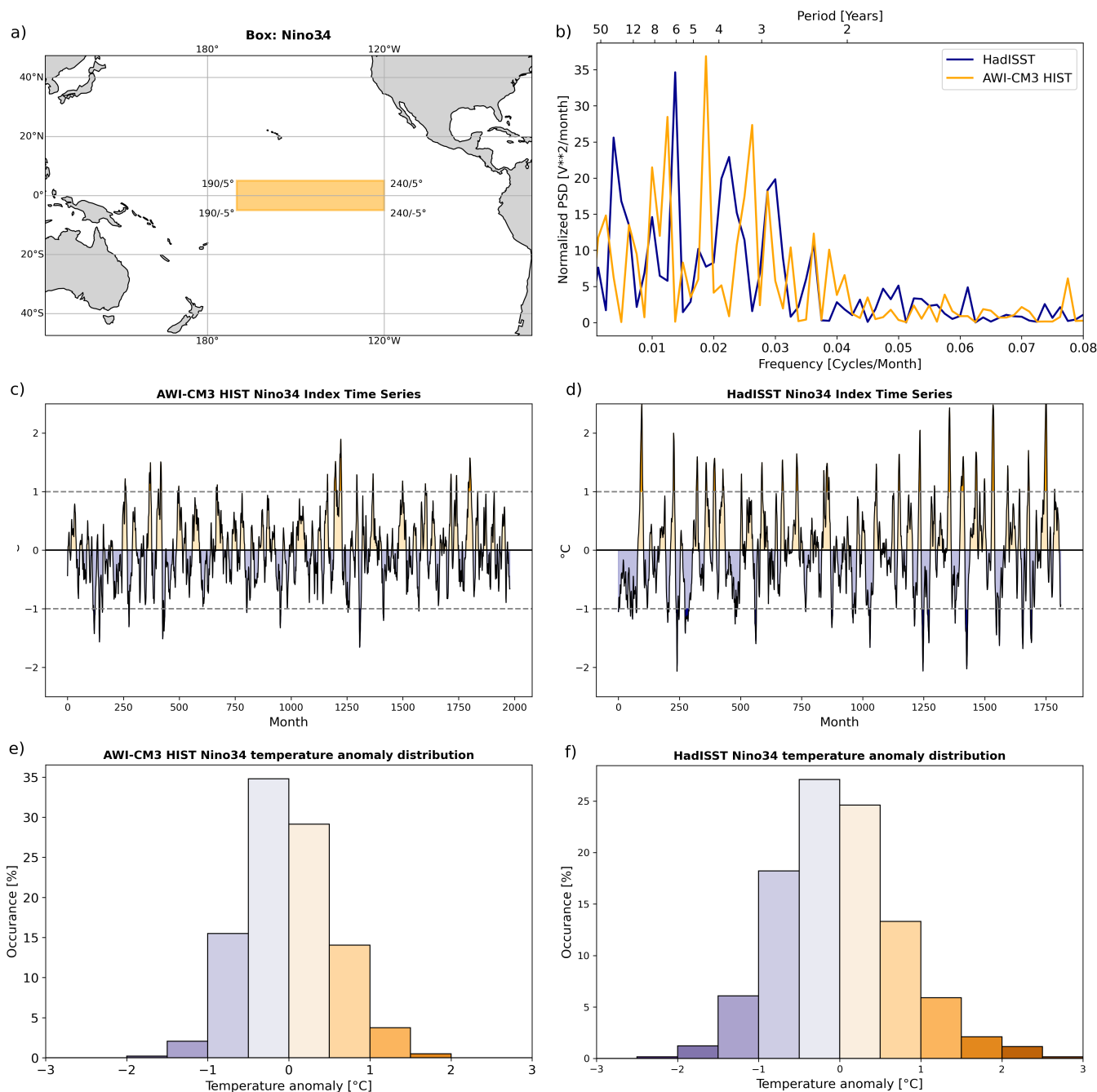
Figure 9. a) Extent of the Nino3.4 box. **b)** Normalized power spectrum density of simulated ENSO in comparison to HadISST observational estimates for the years 1850 to 2014. **c)** & **d)** Nino3.4 SST index defined as the detrended, seasonal cycle removed and 3-months running mean timeseries of SST averaged in the Nino3.4 box. The results from AWI-CM3 HIST and HadISST observational estimates are shown, respectively. **e)** & **f)** As c) & d) but depicting occurrence of temperature anomalies in bins with a $0.5°C$ range.
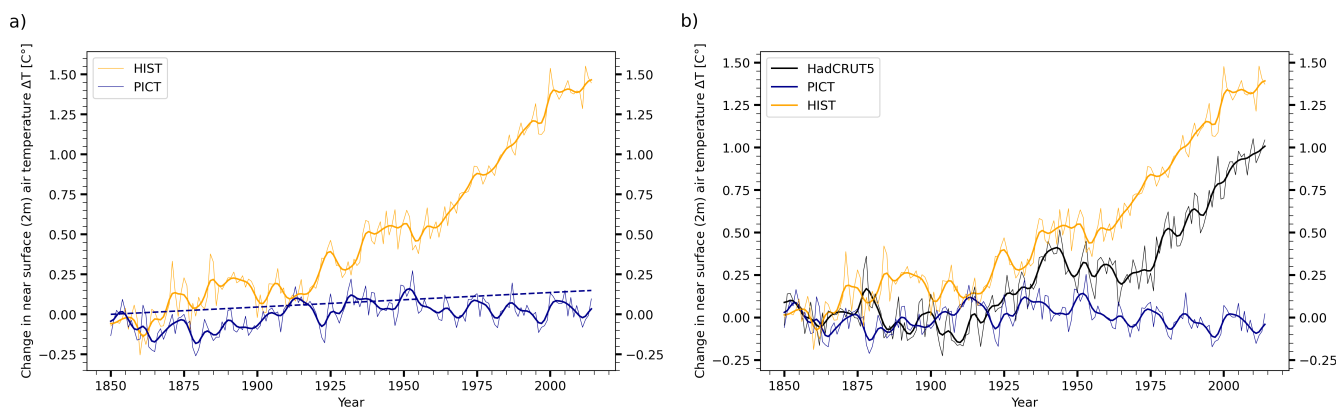
Figure 10. a) Near surface (2m) air temperature changes over the historic period and in the pre-industrial control simulation. Thick lines show the 10 year running means. The pre-industrial control simulation shows a residual drift of $0.00091 \frac{^{\circ}C}{\text{Year}}$ obtained from linear regression. **b)** Both simulations have been corrected by subtracting the residual drift. The resulting simulated global warming over the period 1850 to 2014 is $1.4^{\circ}C$. In comparison to observational estimates from HadCRUT5 (Morice et al., 2021) warming is overestimated by $0.4\,\text{K}$, likely due to fixed aerosols in the AWI-CM3 prototype.
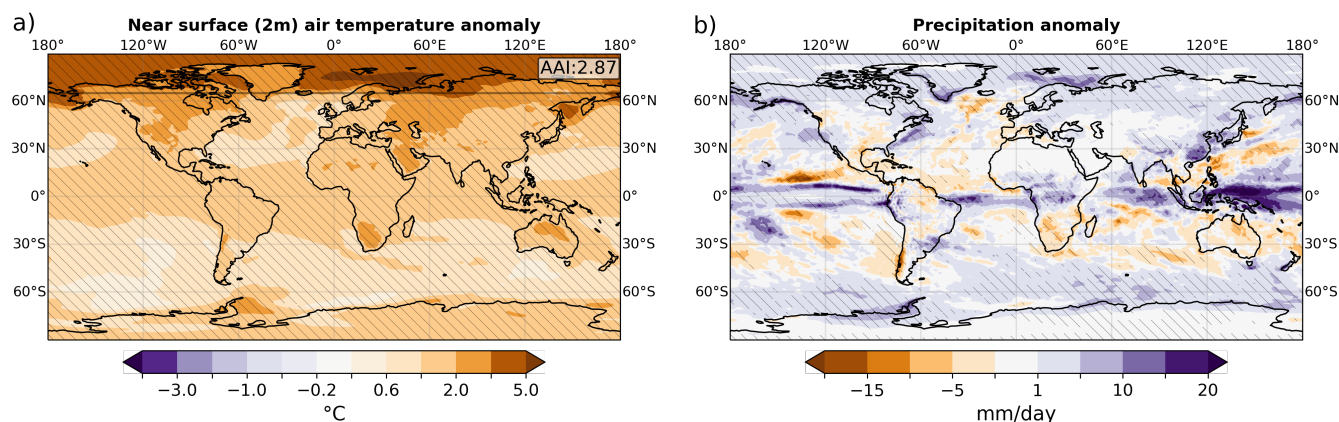
Figure 11. a) Temperature anomaly between last 25 years of the historic simulations and the equivalent period of the pre-industrial control run. The Arctic Amplification index, defined as the ratio of warming north of 65°N expressed as a fraction of global warming, is 2.87. Significant temperature anomalies are hatched and the 95% significance is obtained via bootstrap test. **b)** Same as a) but for precipitation changes.
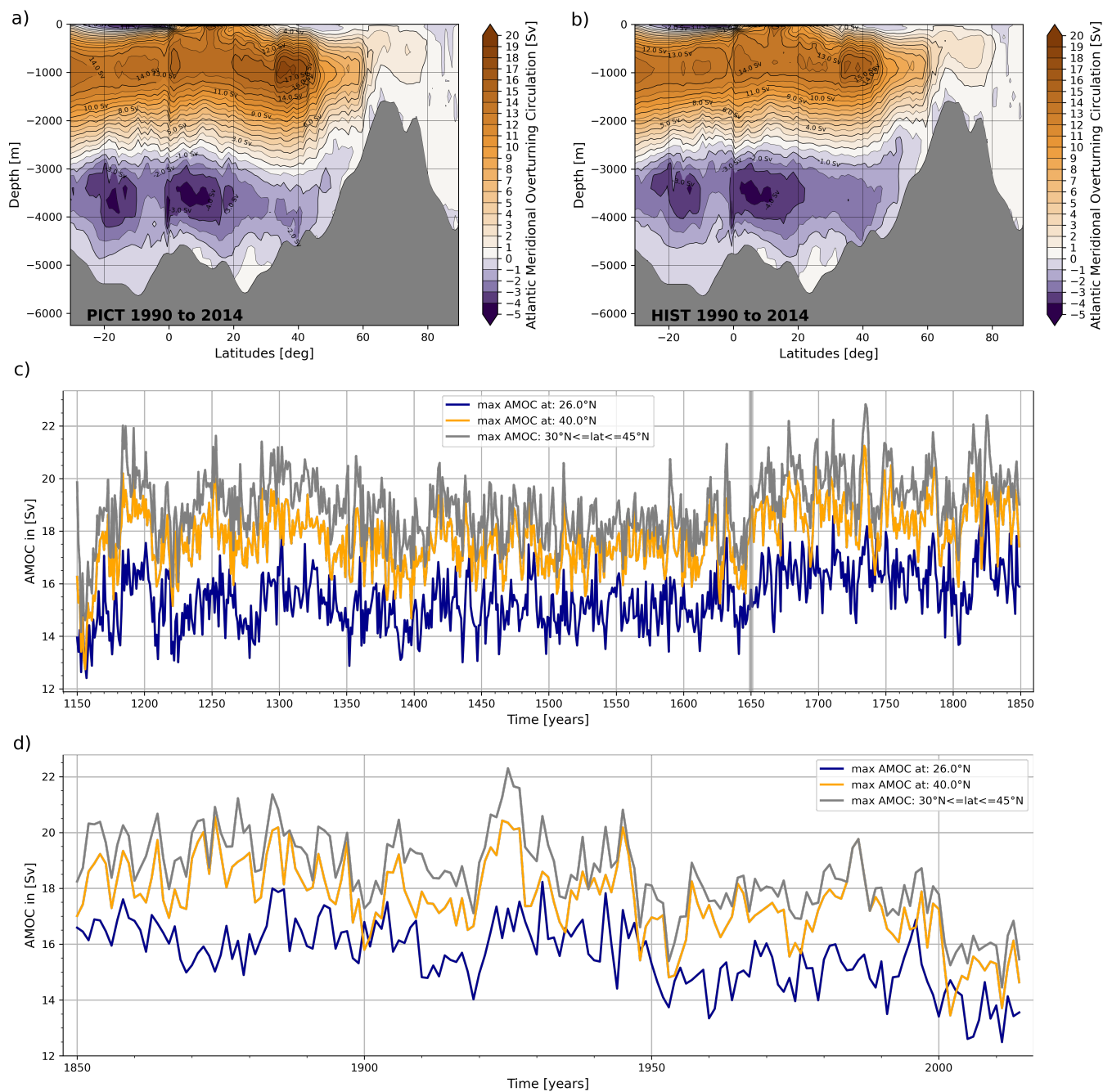
Figure 12. a) Atlantic Meridional Overtuning Circulation (AMOC) streamfunction averaged over the last 25 years in the pre-industrial control (PICT) simulation. **b)** as a) but for the historic (HIST) simulation. **c)** Evolution of the AMOC maximum at different latitudes throughout the spinup simulation (SPIN) that precedes both PICT and HIST. **d)** Evolution of the AMOC during HIST.
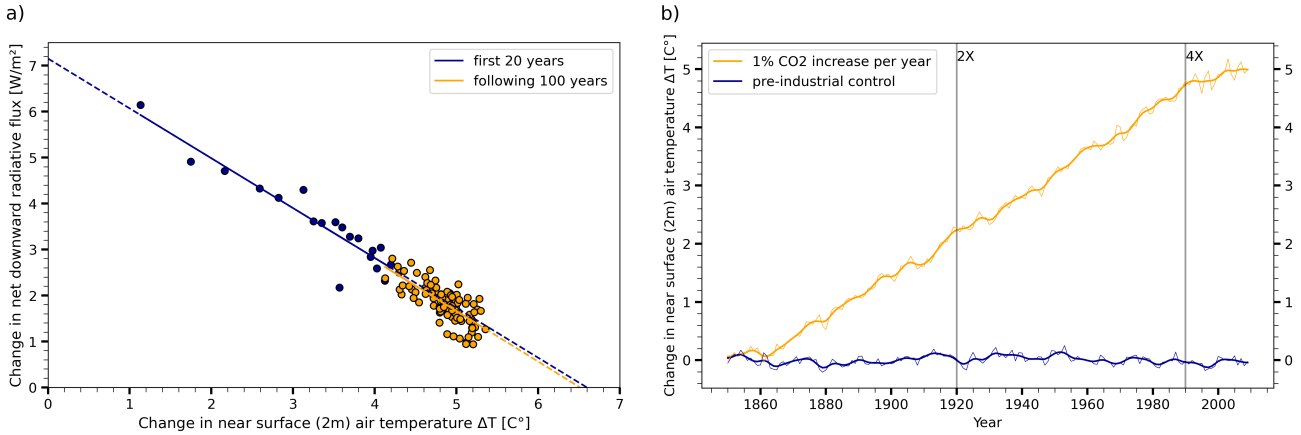
Figure 13. a) Gregory regression plot from the abrupt $4xCO_2$ experiment in comparison to the pre-industrial control. We construct the linear regression of near surface (2m) air temperature (ΔT) change against the downward net radiative flux change (ΔF). From its axis intersection points we obtain the radiative forcing $\Delta F = 7.06\,\frac{W}{m^2}$, and the Equilibrium temperature difference $\Delta T = 6.49\,°C$. The climate response parameter of AWI-CM3 is then $\alpha = \frac{\Delta T}{\Delta F} = 0.92\,\frac{K}{W}$ and the Equilibrium Climate Sensitivity ECS$= \frac{\Delta T}{2} = 3.24\,°C$. The first 20 years already result in a linear regression almost identical to the following 100 years. **b)** As Figure 10 b), but for the experiment with 1% increase of CO_2 per year. Vertical lines indicate doubling and quadrupling of CO_2 concentrations. The estimated Transient Climate Response is 2.1°C.
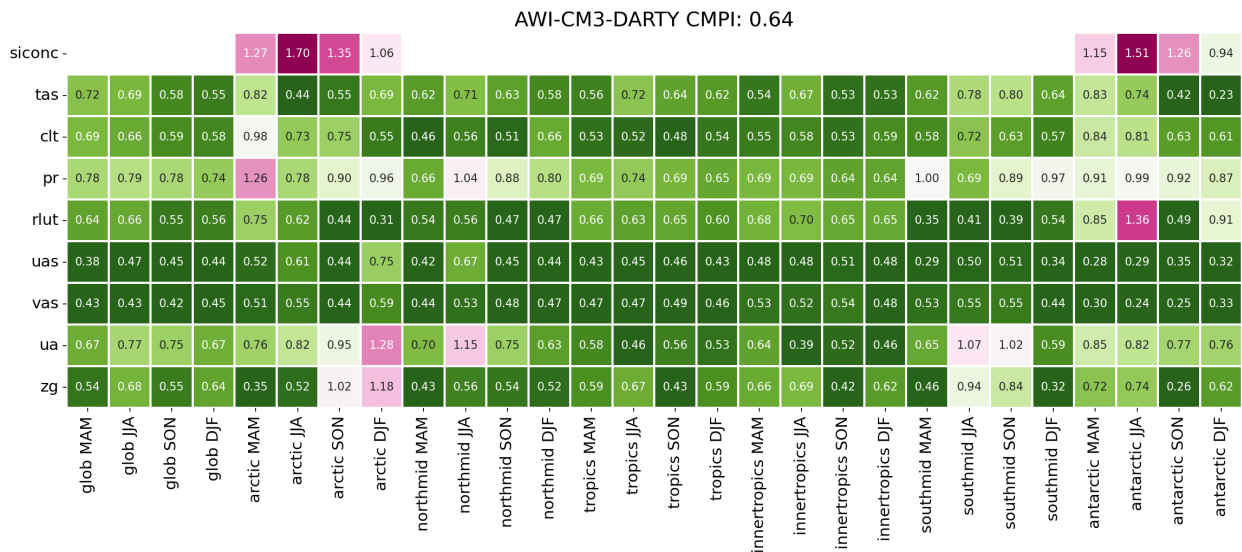
Figure 14. Same as figure 4 but for the last 10 years of a 50-year-long TCo319L137-DART simulation under constant 1990 forcing. Biases to observations are not only smaller than in our TCo159L91-CORE2 simulations, but also smaller than those of every single of the CMIP6 models listed in B (not shown).
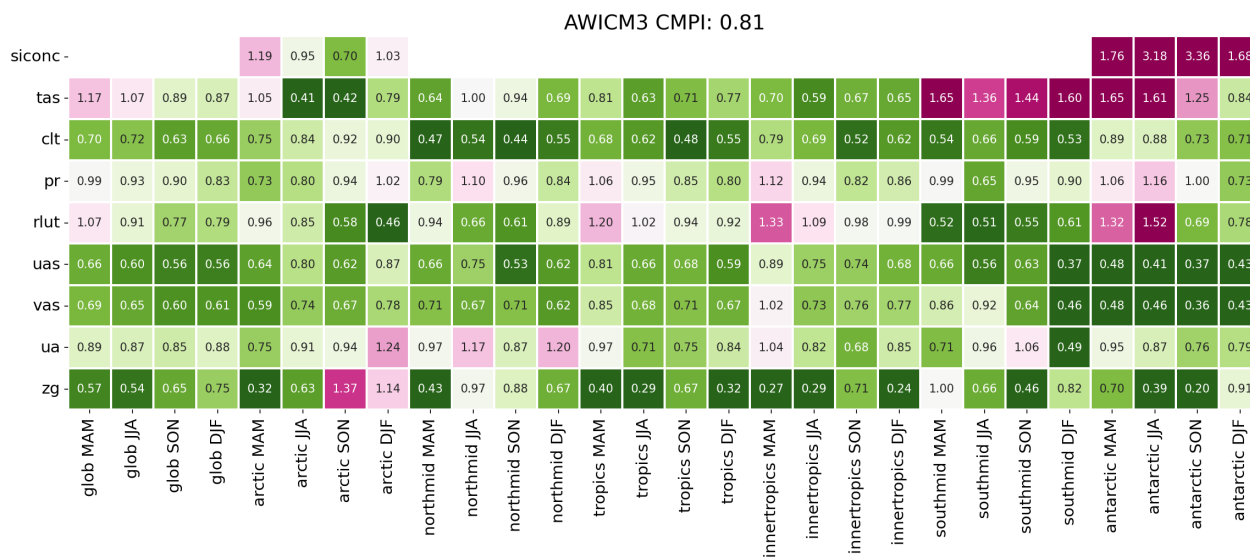
Figure B1. As figure 4, but using ECMWF re-analysis ERA5 Hersbach et al. (2020) instead of NCEP2 for near-surface air temperature (tas), eastward near-surface wind (uas), northward near-surface wind (vas), 300 hPa eastward wind (ua), 500 hPa geopotential height. Note that ERA5 was created using IFS CY41R2, a model closely related to the OpenIFS CY43R3 atmosphere employed in AWI-CM3.